

A Study on the Efficacy of Sentiment Analysis in Author Attribution

A thesis

presented to

the faculty of the Department of Computer and Information Science

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Master of Science in Computer and Information Sciences

by

Michael Schneider

August 2015

Jay Jarman, Ph.D., Chair

Phil Pfeffer, Ph.D.

Chris Wallace, Ph.D.

Keywords: NLP, Sentiment Analysis, Author Attribution, Data Mining

ProQuest Number: 10291092

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10291092

Published by ProQuest LLC (2016). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

ABSTRACT

A Study on the Efficacy of Sentiment Analysis in Author Attribution

by

Michael Schneider

The field of authorship attribution seeks to characterize an author's writing style well enough to determine whether he or she has written a text of interest. One subfield of authorship attribution, stylometry, seeks to find the necessary literary attributes to quantify an author's writing style. The research presented here sought to determine the efficacy of sentiment analysis as a new stylometric feature, by comparing its performance in attributing authorship against the performance of traditional stylometric features. Experimentation, with a corpus of sci-fi texts, found sentiment analysis to have a much lower performance in assigning authorship than the traditional stylometric features.

Copyright by Michael Schneider 2015

TABLE OF CONTENTS

ABSTRACT	2
LIST OF TABLES	15
LIST OF FIGURES	17
1 INTRODUCTION	18
2 OVERVIEW OF AUTHORSHIP ATTRIBUTION	20
2.1 Definition of AA	20
2.2 Problems in Authorship Attribution	21
2.3 History of Text Features	24
3 PROPOSED AUTHORSHIP ATTRIBUTION PROCESS	26
3.1 Process Overview	26
3.2 Step 1: Corpus Preparation	27
3.2.1 Corpus Creation	27
3.2.2 Two-Authors and Multiple-Authors Corpora	30
3.2.3 Corpus Sub-Sections	31
3.3 Step 2: Feature Extraction	33
3.4 Step 3: Token Frequencies	35
3.5 Step 4: Data Mining	37
3.6 Step 5: Predictive Model	38
4 RESULTS	39
4.1 Overview of Results	39
4.1.1 Z-Score Results	40
4.1.2 TF/IDF Results	43

4.1.3	TF Results	46
5	CONCLUSION AND FUTURE RESEARCH	49
5.1	Methodological Quality	49
5.2	Analysis of Results	50
	BIBLIOGRAPHY	54
	APPENDIX	57
APPENDIX A	Number of Author Entries in Each Corpus	57
APPENDIX B	List of Failed Sentences for Alchemy API	60
APPENDIX C	List of Top “X” Words for Entire Corpus	62
APPENDIX D	Results for Experiment Data Models	71
D.1	KNN POS Results - Two Authors	71
D.2	KNN POS Results - Multiple Authors	103
D.3	Naive Bayes POS Results - Two Authors	123
D.4	Naive Bayes POS Results - Multiple Authors	155
D.5	KNN Sentiment Frequency Results - Two Authors	175
D.6	KNN POS Results - Multiple Authors	191
D.7	Naive Bayes POS Results - Two Authors	201
D.8	Naive Bayes POS Results - Multiple Authors	217
	VITA	227

LIST OF TABLES

1	Entire Corpus - Number of Entries Per Document	57
2	Foundation Corpus (1980's Era) - Number of Entries Per Document .	58
3	Foundation Corpus (1950's Era) - Number of Entries Per Document .	58
4	Galactic Empire Corpus - Number of Entries Per Document	59
5	Robot Corpus - Number of Entries Per Document	59
6	Entire Corpus - All Words - Top 50 Tokens	62
7	Entire Corpus - All Words - Top 100 Tokens	63
8	Entire Corpus - All Words - Top 150 Tokens	64
9	Entire Corpus - Adjectives - Top 50 Tokens	65
10	Entire Corpus - Adjectives - Top 100 Tokens	65
11	Entire Corpus - Adjectives - Top 150 Tokens	66
12	Entire Corpus - Nouns - Top 50 Tokens	67
13	Entire Corpus - Nouns - Top 100 Tokens	67
14	Entire Corpus - Nouns - Top 150 Tokens	68
15	Entire Corpus - Verbs - Top 50 Tokens	69
16	Entire Corpus - Verbs - Top 100 Tokens	69
17	Entire Corpus - Verbs - Top 150 Tokens	70
18	KNN POS Results for Entire Corpus(500 Sentences) - Two Authors .	71
19	KNN POS Results for Entire Corpus(1000 Sentences) - Two Authors	73
20	KNN POS Results for Entire Corpus(1500 Sentences) - Two Authors	75
21	KNN POS Results for Foundation Corpus [1980's Era] (500 Sentence) - Two Authors	77

22	KNN POS Results Foundation Corpus [1980's Era] (1000 Sentence) - Two Authors	79
23	KNN POS Results Foundation Corpus [1980's Era] (1500 Sentence) - Two Authors	81
24	KNN POS Results Foundation Corpus [1950's Era] (500 Sentence) - Two Authors	83
25	KNN POS Results Foundation Corpus [1950's Era] (1000 Sentence) - Two Authors	85
26	KNN POS Results Foundation Corpus [1950's Era] (1500 Sentence) - Two Authors	87
27	KNN POS Results Galactic Empire Corpus (500 Sentence) - Two Authors	89
28	KNN POS Results Galactic Empire Corpus (1000 Sentence) - Two Authors	91
29	KNN POS Results Galactic Empire Corpus (1500 Sentence) - Two Authors	93
30	KNN POS Results Robot Corpus (500 Sentence) - Two Authors . . .	95
31	KNN POS Results Robot Corpus (1000 Sentence) - Two Authors . .	97
32	KNN POS Results Robot Corpus (1500 Sentence) - Two Authors . .	99
33	Part-Of-Speech Statistics for KNN - Two Authors	101
34	KNN POS Results for Entire Corpus(500 Sentences) - Multiple Authors	103
35	KNN POS Results for Entire Corpus(1000 Sentences) - Multiple Authors	105
36	KNN POS Results for Entire Corpus(1500 Sentences) - Multiple Authors	107

37	KNN POS Results for Foundation Corpus [1980's Era](500 Sentences)	
	- Multiple Authors	109
38	KNN POS Results for Foundation Corpus [1980's Era](1000 Sentences)	
	- Multiple Authors	111
39	KNN POS Results for Foundation Corpus [1980's Era](1500 Sentences)	
	- Multiple Authors	113
40	KNN POS Results for Foundation Corpus [1950's Era](500 Sentences)	
	- Multiple Authors	115
41	KNN POS Results for Foundation Corpus [1950's Era](1000 Sentences)	
	- Multiple Authors	117
42	KNN POS Results for Foundation Corpus [1950's Era](1500 Sentences)	
	- Multiple Authors	119
43	Part-Of-Speech Statistics for KNN - Multiple Authors	121
44	Naive Bayes POS Results for Entire Corpus(500 Sentences) - Two Authors	123
45	Naive Bayes POS Results for Entire Corpus(1000 Sentences) - Two Authors	125
46	Naive Bayes POS Results for Entire Corpus(1500 Sentences) - Two Authors	127
47	Naive Bayes POS Results for Foundation Corpus [1980's Era] (500 Sentences) - Two Authors	129
48	Naive Bayes POS Results for Foundation Corpus [1980's Era] (1000 Sentences) - Two Authors	131

49	Naive Bayes POS Results for Foundation Corpus [1980's Era] (1500 Sentences) - Two Authors	133
50	Naive Bayes POS Results for Foundation Corpus [1950's Era] (500 Sentences) - Two Authors	135
51	Naive Bayes POS Results for Foundation Corpus [1950's Era] (1000 Sentences) - Two Authors	137
52	Naive Bayes POS Results for Foundation Corpus [1950's Era] (1500 Sentences) - Two Authors	139
53	Naive Bayes POS Results for Galactic Empire Corpus (500 Sentences) - Two Authors	141
54	Naive Bayes POS Results for Galactic Empire Corpus (1000 Sentences) - Two Authors	143
55	Naive Bayes POS Results for Galactic Empire Corpus (1500 Sentences) - Two Authors	145
56	Naive Bayes POS Results for Robot Corpus (500 Sentences) - Two Authors	147
57	Naive Bayes POS Results for Robot Corpus (1000 Sentences) - Two Authors	149
58	Naive Bayes POS Results for Robot Corpus (1500 Sentences) - Two Authors	151
59	Part-Of-Speech Statistics for Naive Bayes - Two Authors	153
60	Naive Bayes POS Results for Entire Corpus(500 Sentences) - Multiple Authors	155

61	Naive Bayes POS Results for Entire Corpus(1000 Sentences) - Multiple Authors	157
62	Naive Bayes POS Results for Entire Corpus(1500 Sentences) - Multiple Authors	159
63	Naive Bayes POS Results for Foundation Corpus [1980's Era] (500 Sentences) - Multiple Authors	161
64	Naive Bayes POS Results for Foundation Corpus [1980's Era] (1000 Sentences) - Multiple Authors	163
65	Naive Bayes POS Results for Foundation Corpus [1980's Era] (1500 Sentences) - Multiple Authors	165
66	Naive Bayes POS Results for Foundation Corpus [1950's Era] (500 Sentences) - Multiple Authors	167
67	Naive Bayes POS Results for Foundation Corpus [1950's Era] (1000 Sentences) - Multiple Authors	169
68	Naive Bayes POS Results for Foundation Corpus [1950's Era] (1500 Sentences) - Multiple Authors	171
69	Part-Of-Speech Statistics for Naive Bayes - Multiple Authors	173
70	KNN Sentiment Frequency Results for Entire Corpus (500 Sentence) - Two Authors	175
71	KNN Sentiment Frequency Results for Entire Corpus (1000 Sentence) - Two Authors	176
72	KNN Sentiment Frequency Results for Entire Corpus (1500 Sentence) - Two Authors	177

73	KNN Sentiment Frequency Results for Foundation Corpus [1980's Era] (500 Sentence) - Two Authors	178
74	KNN Sentiment Frequency Results for Foundation Corpus [1980's Era] (1000 Sentence) - Two Authors	179
75	KNN Sentiment Frequency Results for Foundation Corpus [1980's Era] (1500 Sentence) - Two Authors	180
76	KNN Sentiment Frequency Results for Foundation Corpus [1950's Era] (500 Sentence) - Two Authors	181
77	KNN Sentiment Frequency Results for Foundation Corpus [1950's Era] (1000 Sentence) - Two Authors	182
78	KNN Sentiment Frequency Results for Foundation Corpus [1950's Era] (1500 Sentence) - Two Authors	183
79	KNN Sentiment Frequency Results for Galactic Empire Corpus (500 Sentence) - Two Authors	184
80	KNN Sentiment Frequency Results for Galactic Empire Corpus (1000 Sentence) - Two Authors	185
81	KNN Sentiment Frequency Results for Galactic Empire Corpus (1500 Sentence) - Two Authors	186
82	KNN Sentiment Frequency Results for Robot Corpus (500 Sentence) - Two Authors	187
83	KNN Sentiment Frequency Results for Robot Corpus (1000 Sentence) - Two Authors	188

84	KNN Sentiment Frequency Results for Robot Corpus (1500 Sentence)	
	- Two Authors	189
85	Sentiment Frequency Statistics for KNN - Two Authors	190
86	KNN Sentiment Frequency Results for Entire Corpus(500 Sentences)	
	- Multiple Authors	191
87	KNN Sentiment Frequency Results for Entire Corpus(1000 Sentences)	
	- Multiple Authors	192
88	KNN Sentiment Frequency Results for Entire Corpus(1500 Sentences)	
	- Multiple Authors	193
89	KNN Sentiment Frequency Results for Foundation Corpus [1980's Era](500 Sentences) - Multiple Authors	194
90	KNN Sentiment Frequency Results for Foundation Corpus [1980's Era](1000 Sentences) - Multiple Authors	195
91	KNN Sentiment Frequency Results for Foundation Corpus [1980's Era](1500 Sentences) - Multiple Authors	196
92	KNN Sentiment Frequency Results for Foundation Corpus [1950's Era](500 Sentences) - Multiple Authors	197
93	KNN Sentiment Frequency Results for Foundation Corpus [1950's Era](1000 Sentences) - Multiple Authors	198
94	KNN Sentiment Frequency Results for Foundation Corpus [1950's Era](1500 Sentences) - Multiple Authors	199
95	Sentiment Frequency Statistics for KNN - Multiple Authors	200

96	Naive Bayes Sentiment Frequency Results for Entire Corpus(500 Sentences) - Two Authors	201
97	Naive Bayes Sentiment Frequency Results for Entire Corpus(1000 Sentences) - Two Authors	202
98	Naive Bayes Sentiment Frequency Results for Entire Corpus(1500 Sentences) - Two Authors	203
99	Naive Bayes Sentiment Frequency Results for Foundation Corpus [1980's Era](500 Sentences) - Two Authors	204
100	Naive Bayes Sentiment Frequency Results for Foundation Corpus [1980's Era](1000 Sentences) - Two Authors	205
101	Naive Bayes Sentiment Frequency Results for Foundation Corpus [1980's Era](1500 Sentences) - Two Authors	206
102	Naive Bayes Sentiment Frequency Results for Foundation Corpus [1950's Era](500 Sentences) - Two Authors	207
103	Naive Bayes Sentiment Frequency Results for Foundation Corpus [1950's Era](1000 Sentences) - Two Authors	208
104	Naive Bayes Sentiment Frequency Results for Foundation Corpus [1950's Era](1500 Sentences) - Two Authors	209
105	Naive Bayes Sentiment Frequency Results for Galactic Empire Corpus (500 Sentences) - Two Authors	210
106	Naive Bayes Sentiment Frequency Results for Galactic Empire Corpus (1000 Sentences) - Two Authors	211

107	Naive Bayes Sentiment Frequency Results for Galactic Empire Corpus (1500 Sentences) - Two Authors	212
108	Naive Bayes Sentiment Frequency Results for Robot Corpus (500 Sentences) - Two Authors	213
109	Naive Bayes Sentiment Frequency Results for Robot Corpus (1000 Sentences) - Two Authors	214
110	Naive Bayes Sentiment Frequency Results for Robot Corpus (1500 Sentences) - Two Authors	215
111	Sentiment Frequency Statistics for Naive Bayes - Two Authors	216
112	Naive Bayes Sentiment Frequency Results for Entire Corpus(500 Sentences) - Multiple Authors	217
113	Naive Bayes Sentiment Frequency Results for Entire Corpus(1000 Sentences) - Multiple Authors	218
114	Naive Bayes Sentiment Frequency Results for Entire Corpus(1500 Sentences) - Multiple Authors	219
115	Naive Bayes Sentiment Frequency Results for Foundation Corpus [1980's Era](500 Sentences) - Multiple Authors	220
116	Naive Bayes Sentiment Frequency Results for Foundation Corpus [1980's Era](1000 Sentences) - Multiple Authors	221
117	Naive Bayes Sentiment Frequency Results for Foundation Corpus [1980's Era](1500 Sentences) - Multiple Authors	222
118	Naive Bayes Sentiment Frequency Results for Foundation Corpus [1950's Era](500 Sentences) - Multiple Authors	223

119	Naive Bayes Sentiment Frequency Results for Foundation Corpus [1950's Era](1000 Sentences) - Multiple Authors	224
120	Naive Bayes Sentiment Frequency Results for Foundation Corpus [1950's Era](1500 Sentences) - Multiple Authors	225
121	Sentiment Frequency Statistics for Naive Bayes - Multiple Authors	226

LIST OF FIGURES

1	Process Corpus	26
2	Form the corpus	27
3	Form Two-Authors and Multiple-Authors Corpora	30
4	Split Corpus into Sub-Sections	31
5	Feature Extraction	33
6	Split Corpus into Sub-Sections	35
7	Split Corpus into Sub-Sections	37
8	Average F-Score for Z-Score Frequencies using the KNN algorithm on the Two Author Corpora	41
9	Average F-Score for Z-Score Frequencies using the Naive Bayes algo- rithm on the Two Author Corpora	41
10	Average F-Score for Z-Score Frequencies using the KNN algorithm on the Multiple Author Corpora	42
11	Average F-Score for Z-Score Frequencies using the Naive Bayes algo- rithm on the Multiple Author Corpora	42
12	Average F-Score for TF/IDF using the KNN algorithm on the Two Author Corpora	44
13	Average F-Score for TF/IDF using the Naive Bayes algorithm on the Two Author Corpora	44
14	Average F-Score for TF/IDF using the KNN algorithm on the Multiple Author Corpora	45

15	Average F-Score for TF/IDF using the Naive Bayes algorithm on the Multiple Author Corpora	45
16	Average F-Score for Term Frequencies using the KNN algorithm on the Two Author Corpora	47
17	Average F-Score for Term Frequencies using the Naive Bayes algorithm on the Two Author Corpora	47
18	Average F-Score for Term Frequencies using the KNN algorithm on the Multiple Author Corpora	48
19	Average F-Score for Term Frequencies using the Naive Bayes algorithm on the Multiple Author Corpora	48

1 INTRODUCTION

The field of authorship attribution (AA) seeks to characterize an author’s writing style well enough to determine whether he or she has written a text of interest. The field originated in academic debates over the authorship of literary works. Over time, the range of applications for AA has broadened to include more common applications such as computer forensics (e.g. correctly attributing emails to criminals) and detecting plagiarism [15].

There is still no universally accepted method for defining an author’s writing style. At one time, academia tried to use “expert opinion” to analyze and define an author’s writing style, but this subjective approach led to heated and unsettled debates amongst literary scholars over authorship and the definition of writing style [6]. Modern AA seeks to break away from the subjective debate and find an objective and verifiable approach to defining an author’s writing style.

One modern AA approach to characterizing style, stylometry, uses the frequency of literary features to correctly distinguish between 2 or more authors. The research presented here sought to determine the effectiveness of sentiment analysis for author attribution, relative to more common stylometric metrics such as word and punctuation frequency. Sentiment analysis seeks to define if author is expressing a positive, negative, or neutral emotion in their writing [12]. In the experiments conducted for this study, sentiment analysis proved inferior to the other metrics, which had consistently high performance in authorship attribution. The results found here, however, are only indicative of sentiment analysis’s effectiveness for authorship attribution on the studied corpus. More research is necessary to determine the effectiveness of

sentiment analysis, along with the other studied features, for other corpora.

The balance of this study is divided into five sections. Section 2 further defines authorship attribution and details previous work in the field. A few of the most common text features are described, along with some of the hurdles that have impeded AA research. Section 3 describes the methodology for these experiments, from the selection of a corpus to the extraction of text features from said corpus to the creation of a predictive data model from the extracted features. Section 4 describes the experiments' results. Section 5 draws conclusions from these results and suggests directions for future research. Section 6 concludes with a series of appendices, most of which provide a detailed presentation of the study's experimental data.

2 OVERVIEW OF AUTHORSHIP ATTRIBUTION

2.1 Definition of AA

Authorship Attribution (AA), according to Joula, is “any attempt to infer the characteristics of the creator of a piece of linguistic data” [8]. Joula divides AA into three main areas:

- **Closed class attribution.** A text’s author is known to belong to a pool of candidate authors. The problem is to determine which candidate is the most likely author.
- **Open class attribution.** A text’s author might belong to the pool of candidates. The problem is to determine if the author is in the pool of candidate authors.
- **Profiling.** A text’s author is known. The problem is to infer the author’s characteristics from the text: e.g., the author’s age, ethnicity, education level, and/or gender.

Stamatos, [15], by contrast, identifies four common tasks of modern AA:

- **Author verification**, or determining if a specific author wrote a given text. This task resembles but is not quite the same as Joula’s attribution tasks.
- **Plagiarism detection**, or determining if one text has been plagiarized from another. This is outside of Joula’s categories.
- **Author Characterization**, which corresponds to Joula’s Profiling.

- **Stylistic Inconsistencies (SI)**, or identifying irregularities in a text that are inconsistent with the author’s other writings. These irregularities could indicate multiple authorship (the text had multiple authors) or authorship under duress (the author was being threatened while writing the text).

2.2 Problems in Authorship Attribution

The Bibliographies of stylistics contain thousands of titles, there is no lack of observed facts; however, the polysemy of concepts, the impression of methods, the uncertainty about the very goal of this research hardly make for a prosperous discipline. -Todorov [16]

Todorov’s observation is representative of various critiques of the current status of authorship attribution. Progress in AA, according to Rudman [13], has been retarded by the community’s inability to agree on the effectiveness of stylometric features as well as the sheer number of stylometric features being proposed¹. Rudman lists 7 problems that he believes must be addressed to make progress in AA research.

- **Rushed Work.** Researchers have produced questionable studies due to limited time and/or resources.
- **Lack of Novel Research.** Too many studies are being repeated and many previous works have been forgotten or ignored.

1. Rudman’s work, which dates to 1990, identified over 1,000 different stylometric features that were proposed by researchers [13]

- **Inappropriate.** Too many studies use tests from outside linguistics that make unfounded assumptions about how languages behave. Rudman cites the use of the Effron-Thisted test, which comes from butterfly collecting. This test assumes that the odds of sampling a particular individual are independent of when the sample is taken. That assumption, in the context of textual analysis, equates to assuming that all of a text’s words are evenly distributed throughout that text. This second assumption, according to Rudman, has no known basis in linguistics.
- **Data Cleansing.** Little attention has been given to the problems posed by textual modification. Source texts can change over time due to errors in transmission. Changes in language over time can also affect a text’s perceived meaning. A third source of error involves the effect of translation and editing—e.g., modernization—on the original text’s meaning and style.
- **Multiple Disciplines.** Rudman sees a need to marshal cross-disciplinary expertise in linguistics, statistics, and computer science to understand and correctly analyze results.
- **Ignorance.** Some research is far too shallow to sustain the conclusions that have been drawn from the results.
- **Error Reporting.** “Cherry picking” of data and skewing results is far too common in AA research.

In [6], Grieve details another problem with AA research: academic squabbling, as exemplified by a spat between Don Foster on the one hand and Ward Elliot & Robert Valenza on the other. In 1989, Foster used the frequencies of unstopped lines, feminine lines, and open lines to attribute the anonymous poem “The Funeral Elegy” to William Shakespeare. In and of itself, Foster’s work was not necessarily controversial; many works have been attributed and later unattributed to Shakespeare. However, a subsequent refutation of Foster’s work by Elliot and Valenza led to a decade’s worth of sparring over Foster’s work in *Computers and the Humanities* featuring publications with titles like “The Professor Doth Protest Too Much, Methinks” and “So Many Hardballs So Few Over the Plate”. It ended in 2002 when Don Foster accepted Monsarrat and Vickers’s separate attributions of “The Funeral Elegy” to the Elizabethan author John Ford

A second such controversy involved, the now infamous, research by the Reverend Andrew W. Morton on *positional stylometry*: a strategy for textual attribution that assumes that an author consistently uses related textual features throughout his or her writing [6]. Positional stylometry yielded visually impressive results, thanks to the use of cusum graphs, also called QSUM charts, to depict consistent authorship as parallel lines and changes in authorship as divergences from parallelism [4]. These charts were criticized by researchers like David Holmes, who held that “the underlying assumption regarding the consistency of habits within the utterances of one person is false” [7]. Holmes further argued that Morton and Michaelson had not adequately defined how parallel a set of graphs must be and at what point a divergence becomes significant to indicate a change in authorship. Holmes and other critics warned against

the increasing use of cusum charts in the British legal system.

Cusum charts fell into disfavor in 1993, when during a live television interview Morton failed to correctly distinguish a convicted felon’s writings from writings by England’s Chief Justice. After his interview, Morton’s cusum charts lost appeal in the research community and were dropped from use in the British legal system [6].

2.3 History of Text Features

Joula [8] divides features used for AA into structural, character, and lexical features. Structural features characterize a text’s physical properties at the level of its words, without regard to word meaning: e.g., average word length and average sentence length ². Character features characterize a text at the level of its characters: e.g., frequency of a given character or an n-gram— a sequence of n characters like the trigram “ion”. Lexical features characterize word choice: e.g., word frequencies, diversity of vocabulary.

Joula identifies the following features as the most commonly researched for each category:

- **Structural:** word length and sentence length
- **Character:** character frequency, upper-case character frequency, lower-case character frequency, n-gram frequency and punctuation frequency

2. Joula considered word length and sentence length as lexical features, but I find they fit more closely with Zheng et al’s classification as structural features

- **Lexical:** vocabulary diversity; word n-gram frequency; local word frequency, or the number of times a word is used in a text; language word frequency, or the frequency with which a word appears in a language; and function word frequency, or the frequency of articles and words that connect dependent clauses.

Zheng et al, [18] identify a fourth class of features: content-specific features, or words are specific to a topic or domain of interest.³ Zheng et al’s example is the acronym “obo”, short for or best offer, which was prevalent in their corpus’ online comments.

While structural features yielded mild improvements to Zheng et al.’s models, structural features have gone in and out of vogue. Mendenhall, one of the original proponents of sentence and word length, found structural features to be far too sensitive to changes in topic and genre [11]. This conclusion has since been supported by other researchers [17, 14, 6].

After structural features, the most highly studied features are lexical features, most notably word frequencies. Mosteller and Wallace made word frequencies famous in their 1964 study of the Federalist Papers, which used function word frequency to distinguish between the writing styles of presumed authors John Jay, Alexander Hamilton, and James Madison. Their final attributions of authorship are still considered the “most likely” author attribution for the contested Federalist Papers [6].

3. Zheng et al refer to “Lexical Features” as being both character-level features and “Word-based” features, but considered function words to be in a different category, syntactic, which included punctuation mark frequency.

Other studies have found high correlations between word frequency and authorship [6, 18, 8].

3 PROPOSED AUTHORSHIP ATTRIBUTION PROCESS

3.1 Process Overview

The five-step process used in this study for authorship attribution is depicted in in Figure 1.

- In step 1, documents from various authors are gathered to create a corpus.
- In Step 2, the stylometric features are extracted from the corpus.
- In Step 3, three different frequency metrics are calculated for the stylometric features extracted from the corpus documents. These frequency metrics are then aggregated into data sets which are passed onto step 4.

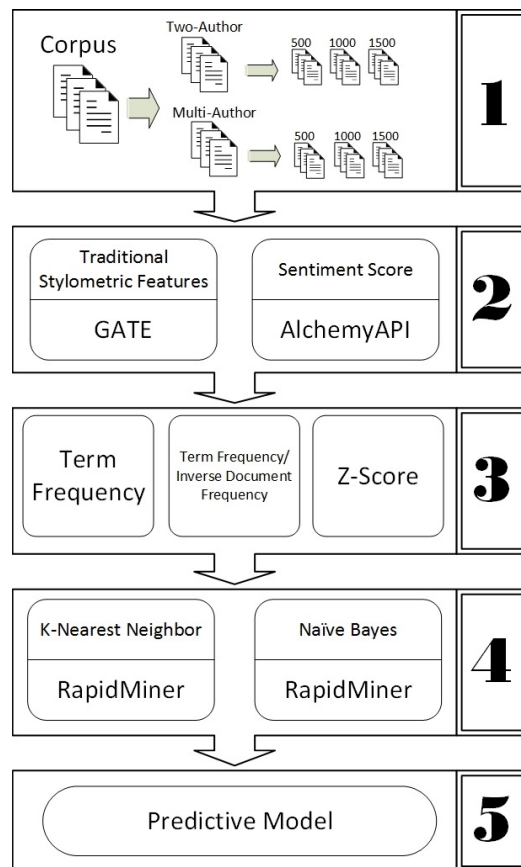


Figure 1: Process Corpus

- In step 4, the data tables are processed by two different data mining algorithms, K-Nearest Neighbor and Naive Bayes.
- These two algorithms each create a predictive model that can be used to infer the author of each of a test corpus's documents. The accuracy of a model's inferences is then used to judge how well each stylometric feature performs in authorship attribution.

3.2 Step 1: Corpus Preparation

3.2.1 Corpus Creation

The corpus for this research was formed from the works of four different science fiction authors. This focus on a single genre is in keeping with best practices in AA research: i.e., choosing texts that share as many attributes as possible in order to reduce the influence of extra-

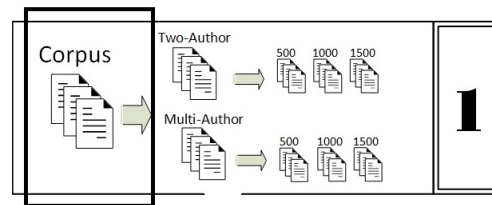


Figure 2: Form the corpus

neous, domain-related factors on the final data model. Due to the diversity of the science-fiction genre, the corpus was further narrowed to stories from the Isaac Asimov universe. Asimov was a prolific 20th century author who created a common setting for nearly all of his stories. Later authors used Asimov's universe as a setting for new stories as well as stories based on Asimov's earlier works. This narrowing of scope to Asimov's universe addresses two additional criteria for AA corpora: similarity of

topic and, with one exception (detailed later), target audience. The choice of corpus weakly addresses a fourth consideration, similarity of time period, in that publication dates range from the 1950's when Asimov was creating his universe through the 1990's, after Asimov's death.⁴ While this difference in time periods is less than ideal, I saw this as less important than maintaining consistency of genre, topic, and audience. Asimov's literary universe has been categorized into three series: Foundation, Galactic Empire, and Robots. These three series of Asimov's universe are all tied together over a massive literary time line. The works of these three series were used to form three disjoint corpora, which, when combined, make up the corpus for this study:

Foundation Corpus

The Foundation corpus consists of 6 books that Isaac Asimov wrote for his Foundation series and the three books that posthumously continued his series. These three continuation books were written by three separate authors, Greg Benford, Greg Bear, and David Brin. The titles and publication date of the books are as follows:

Asimov **1950's Era** *Foundation* 1951, *Foundation and Empire* 1952,
Second Foundation 1953,

1980's Era *Foundation's Edge* 1982,

Foundation and Earth 1986, *Prelude to Foundation* 1988

Benford *Foundation's Fear* 1997

4. Each author was given permission to write in Asimov's Universe either by Asimov, before his passing, or posthumously by the Asimov Estate.

Bear *Foundation and Chaos* 1998

Brin *Foundation's Triumph* 1999

Galactic Corpus

The Galactic corpus is more loosely knit than the Foundation Series. Its stories span multiple millennia in Asimov's world and focus on his "Galactic Empire". Roger Allen MacBride's Caliban Trilogy is part of this corpus. This trilogy takes place between the events of the Galactic Empire series and the Robots series. While the Caliban Trilogy is more closely tied to the "Robot" series in plot, it was placed in the Galactic corpus because it is the closest series in the Asimov's universe time line to the Galactic Empire. The works chosen from Asimov and MacBride are as follows:

Asimov *Pebble in the Sky* 1950, *The Stars, Like Dust* 1951,
 The Currents of Space 1952

MacBride *Caliban* 1993, *Inferno* 1994, *Utopia* 1996

Robot Corpus

The Robot Corpus compares Asimov's works which emphasized the three laws of robotics and the works of William F Wu. These works refer to similar characters and rely heavily on a shared fictional vocabulary (ex. positronic brain). A major difference between Asimov's and Wu's writings are their targeted audiences. Unlike the study's other authors who targeted their work for adult readers, Wu wrote his books for young adults. The works from Asimov and from Wu that

comprise this corpus are as follows:

Asimov *Caves of Steel* 1953, *Naked Sun* 1955,
 Robots of Dawn 1983, *Robots and Empire* 1985,
Wu *Predator* 1993, *Maurader* 1993, *Warrior* 1993,
 Dictator 1994

3.2.2 Two-Authors and Multiple-Authors Corpora

After the documents for each corpus have been selected, each corpus needs to be modified to fit into two categories, two-authors and multiple authors. As other studies have shown, the accuracy of stylometric features for AA declines as more authors are placed in the candidate pool [6, 18].

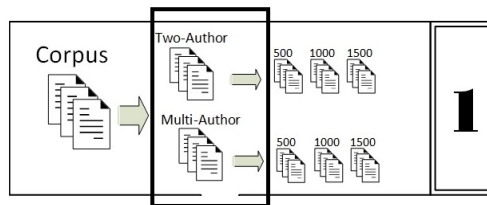


Figure 3: Form Two-Authors and Multiple-Authors Corpora

One approach to improving AA performance is to compare a single author’s works to a combined pool of works by other authors. In this study, the single author is Isaac Asimov, who is then compared to all other authors. In practice, all documents not written by Isaac Asimov are marked as written by “Not Asimov”. For example, the 1950’s Era corpus has 4 candidate authors, Asimov, Benford, Bear, and Brin. For the two-authors version of this corpus, the works of Brin, Bear, and Benford are all marked as “Not Asimov”. The multiple-authors version of this same corpus leaves all works unaltered and marked with their respective author. By creating two

different versions of each corpus, two-authors and multiple authors, this research aims to see how increasing the number of candidate authors affects the performance of the stylometric features.⁵

3.2.3 Corpus Sub-Sections

The last part of preparing the corpus for stylometric feature extraction involves breaking each corpus document into smaller documents. Each of the authors in this study wrote only a few works, or in some cases only one work,

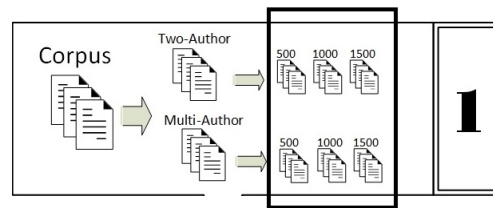


Figure 4: Split Corpus into Sub-Sections

that can be included in the corpus. With only a few examples to work from, the final predictive data model created from this corpus may “over-fit” its predictions to the corpus documents⁶. Splitting each document into multiple documents artificially increases the number of documents attributed to each author, reducing the risk of overfitting.

To create these smaller collections of documents, each document was partitioned

5. The Robot Corpus and the Galactic Corpus do not have a multiple author versions, because they only contain two possible authors. The only corpora that have multiple candidate authors are the 1950’s and 1980’s Era Foundation Corpora and the Full Corpus, which contains all of the works studied here by the 6 candidate authors.

6. An “over-fit” model is a model appears to have high performance in testing, but when given new documents, outside the training corpus, the over-fit model will have lower accuracy in correctly attributing the authorship of the new documents.

into sub-documents containing a set number of sentences (500, 1000, 1500). For each k -sentence document, this partitioning yielded a collection of $\lceil k/500 \rceil$, $\lceil k/1000 \rceil$, or $\lceil k/1500 \rceil$ sub-documents. Each resulting sub-document contained a contiguous block of sentences from the original document, with the sentences in their original order. For example, a split of a 3,000 sentence document into 1,000 sections would yield 3 text files, containing sentences 1-1000, 1001-2000, and 2001-3000 from the original document, respectively.⁷ The 500, 1000, 1500 sentence sub-document sizes were chosen at the researcher’s discretion, in the absence of a clear understanding as to how large a document should be to obtain adequate predictive data models that avoid being over-fit.⁸

Corpus documents were partitioned with the help of a utility from the GATE application suite. GATE, the General Architecture for Text Engineering, is an open source natural language processing toolset [3]. The GATE utility, ANNIE (A Nearly-New Information Extraction system), is a pipeline of natural language processing programs that tokenize texts and annotate tokens’ parts of speech.⁹ In this step of the processes, a Java application uses ANNIE to parse a document into its sentences. The Java application then writes blocks of k sentences into successive sub-documents.

7. To see the number of subsections created for each text in each of corpora, see Appendix A.

8. Kim Luyckx and Walter Daelemans did similar work by testing to see how limiting the amount of data in a training set affected a model’s performance [10] This is different from the approach in this study, in that they limited the amount of data, where as here the amount of data is the same, it is just being splintered into smaller pieces.

9. “The tokenizer splits the text into very simple tokens such as numbers, punctuation and words of different types” [5]

Unfortunately, while parsing the sentences from the corpus documents, ANNIE replaced single quotes with <apos>, double quotes with <quot>, and characters like é, that have multi-byte representations, with ❖. This insertion of HTML tags and ❖ symbols was inconsistent, due in all likelihood to the use of three different input formats for corpus documents. To cleanse this bad data, a Java function was created to convert all <apos> and <quot> tags to their original unicode representation and to remove ❖'s from the processed text. The effect of ❖ removal on the results was not determined in this study.

3.3 Step 2: Feature Extraction

After the documents have been cleansed, the stylometric features can be extracted. For this study, two types of stylometric features were extracted from each corpus document, traditional stylometric features and sentiment scores. Sentiment is analyzed using

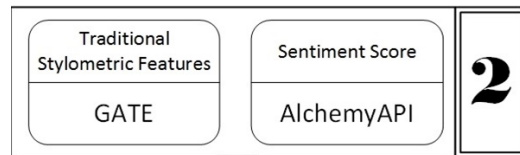


Figure 5: Feature Extraction

AlchemyAPI. AlchemyAPI¹⁰ is a licensed, web-based tool that processes text documents via HTTP Post requests and returns a sentiment score for the posted document. This score is a number between -1 and 1 with 6 digits of precision. This study used

10. AlchemyAPI is a commercial tool, but does offer an academic license, which was used for this study. The academic license allowed for an increased number in HTTP Post requests, which allowed for a quicker pace in calculating the sentiment scores for all of the corpus documents.

the Java application to Post each sentence in a corpus document to AlchemyAPI and record the returned sentiment score. The Java application collected these sentiment scores into one large list for each document.

AlchemyAPI calculates sentiment scores using a proprietary formula. It is unknown how the scores are calculated or if the values calculated will be the same for repeated posts of a given text to AlchemyAPI. Because of this uncertainty about the scores repeatability, each sentiment score was rounded to the 3rd, 2nd, and 1st decimal places on the assumption that any change to a score, over time, would likely be limited to the least significant digit(s). Unfortunately, this assumption cannot be tested because of AlchemyAPI's "No Compete" clause in their license agreement, which forbids comparing AlchemyAPI sentiment scores to competitor scores and previously calculated AlchemyAPI scores.

Tests were conducted to assess how the rounding of sentiment scores affected their performance in authorship attribution. Each sentiment score was rounded to three, two, and one decimal, and tallies kept for how often each rounded value appears in each document. After calculating and counting the sentiment scores for the corpus documents, the Java application used ANNIE to extract the traditional stylometric features that were being compared against sentiment analysis.

This study focuses on three traditional stylometric features: words, marks of punctuation, and parts of speech (adjectives, articles, nouns, and verbs). These are all variations of the traditional "word frequency" metric, one of the most successful and heavily studied stylometric metrics [6]. To extract these features from a document, the

Java application uses ANNIE¹¹ to obtain a list of each of a document’s punctuation marks and words, along with each word’s parts of speech. The JAVA application steps through this list to count how often each punctuation mark appears, each word appears and how often each word is associated with a given part of speech.

3.4 Step 3: Token Frequencies

The traditional feature tallies described in the previous section are used to calculate three different types of frequency measurements:

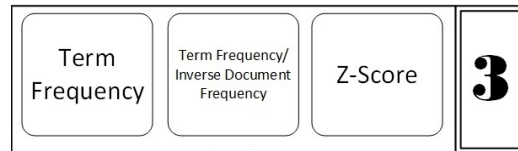


Figure 6: Split Corpus into Sub-Sections

- **Term Frequency.** - TF measures the relative frequency of a word in a host document, as determined by the word’s occurrences divided by document’s total number of sentences.

11. The parts-of-speech tagger, which annotates each word or symbol given to it, uses a built in lexicon and rule set [5] While ANNIE supports the use of other lexicons and rule sets, this study uses the default ANNIE settings.

- **Term Frequency - Inverse Document Frequency.** - TF/IDF divides a term's local frequency by the term's document frequency. This measure assigns greater weight to infrequently used words, highlighting those words that are characteristic of a given author. For example, if only one author is fond of using the word "inconceivable", TF/IDF assigns a greater weight to the rarefied word "inconceivable" and a lesser weight to more commonly shared words, such as "the" or "he". TF/IDF creates a greater range in the frequency scores, which is believed to better differentiate authors' styles [2]. The inverse document frequency for a given word w is calculated as $\log(N/df_w)$, where N is the number of documents in a corpus and df_w are the number of documents in the corpus that contain the word w . The TF/IDF for a given word w can then be calculated as $TF_w \times \log(N/df_w)$. [2]
- **Z-Score.** A Z-score relates a word's term frequency to its mean term frequency and the standard deviation of its term frequency in all corpus documents [1, 9]. Z-score, like TF/IDF, tries to relate the frequency of a document's words to their frequency in all documents, but does so with a different calculation:

$$z_w = \frac{TF_w - MeanTF_w}{StDevTF_w}$$

- TF_w is the term frequency of the given word w in document X
- $MeanTF_w$ is the mean term frequency for w in all of the corpus's documents
- $StDevTF_w$ is the standard deviation for the term frequency of the given word w in all documents of the corpus

The resulting ordered lists of words were categorized by type, to obtain lists of adjectives, nouns, verbs, articles, and punctuation in addition to the complete word list. The lists of adjectives, nouns, verbs, and all words were subsetted once again, to obtain lists of the 50, 100, and 150 most frequent words, by word type.¹² Top 50, 100, and 150 lists for articles and marks of punctuation were unnecessary, since these had fewer than 50 tokens to choose from.

After calculating the frequency metrics for the tokens in each corpus document, the data sets are then converted into .csv files to be processed by the data mining application.¹³

3.5 Step 4: Data Mining

RapidMiner is an open source data mining application that creates predictive data models from the gathered data sets using data mining algorithms. This study uses RapidMiner’s built-in implementations of the Naïve Bayes and K-

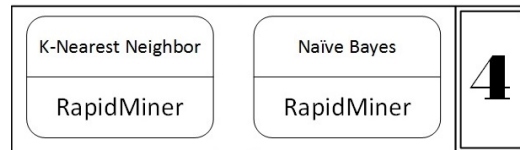


Figure 7: Split Corpus into Sub-Sections

Nearest Neighbor (KNN) algorithms to create a Naïve Bayes predictive data model and KNN predictive data model for each data set. RapidMiner-created training sets

12. The list of the top “x” number of tokens used differs between the corpora. The list of the top “x” number of words used for the entire corpus can be found in appendix section C.

13. The combination of different corpora, corpora sub-sections, number of stylometric features, and different frequency metrics being calculated led to a total of 1,512 different data sets to be processed by RapidMiner.

for model generation which were validated against RapidMiner-created test sets in order to judge model performance. Cross-validation was used to create the training sets and tests sets from the initial data set.

While RapidMiner is a mature software system, its recent change to a “business source” pricing model caused problems for this study. When RapidMiner released version 6 in 2013 it set the previous version, 5, to be open source and labeled it as the “community edition”. RapidMiner version 5, unlike version 6, limits processing to one core. This restriction, which is documented on the company’s website customer support forum, was only found late in this study. The delays created by this restriction were severe enough to prevent a third set of support-vector-machine models from being computed in time for this study.

3.6 Step 5: Predictive Model

The effectiveness of predictive data models was determined from the models’ f-scores. F-scores are calculated using the precision and recall of a data model’s predictions.¹⁴ The formula for computing an f-score is as follows:

$$F - Score = 2 * \frac{(Precision * Recall)}{(Precision + Recall)}$$

14. There does not appear to be a consensus on using accuracy vs f-score in the research studied. For this reason, all experimental results have been included in the appendix section D, in case future research would prefer to compare their results against accuracy, precision, or recall.

4 RESULTS

4.1 Overview of Results

The effectiveness of sentiment analysis as a means of author attribution was evaluated by comparing the accuracy of its attributions to the accuracy of attributions based on word and punctuation frequency. The textual bases for these comparisons, the documents created from the Asimov universe corpora, were described in Section 3. The attributions' results were compared by averaging the f-scores of their predictive models across the two-authors and multiple-authors corpora.

The presentation that follows consists of three sections, which compare the results from sentiment analysis to the results from frequency analysis, as scored by term frequency, term frequency/inverse document frequency, and z-score, respectively. Each of these three first-level sections consists of two second-level sections: one that presents data for the two author comparisons, and one that presents data for the multiple author comparisons. Finally, each of these second-level sections consists of two third-level sections: one that focuses on traditional features and one that focuses on sentiment analysis.

Performance comparisons are presented by superimposing the results from a related set of traditional stylometric analyses and sentiment analyses on a common graph. This presentation is somewhat unorthodox, in that these datasets assume incomparable scales for x-axis values. For the traditional features, the x-axis shows a decreasing progression in the number of tokens being used. For sentiment analyses, the x-axis shows a decrease in the number of decimal places. This departure from

convention was justified on the grounds that the resulting graphs' y-axes highlight the relative effectiveness of the different feature types.

4.1.1 Z-Score Results

Figures 8 and 9 show average z-scores for KNN- and Naïve-Bayes-based models of the two author corpora. Figures 10 and 11 show these scores for the multi-author corpora.

- **Traditional Features.** In z-score-based comparisons of KNN and Bayesian models of traditional stylometric attributes, the KNN models fared better, with one exception. Using all available words to generate the models yielded a striking difference in performance between KNN and Naive Bayes. Reducing the number of words dramatically improved KNN performance, whereas reducing the number of words reduced the Bayes models' average f-score for traditional metrics.
- **Sentiment Scores.** Traditional stylometric features outperformed sentiment scores. KNN models of sentiment scores outperformed Naive Bayes models of these scores. Reducing the sentiment score's precision improved the average F-Score for both KNN and Naive Bayes. The change to a multiple author corpora reduced the average f-score for sentiment scores to almost single digits.



Figure 8: Average F-Score for Z-Score Frequencies using the KNN algorithm on the Two Author Corpora



Figure 9: Average F-Score for Z-Score Frequencies using the Naive Bayes algorithm on the Two Author Corpora

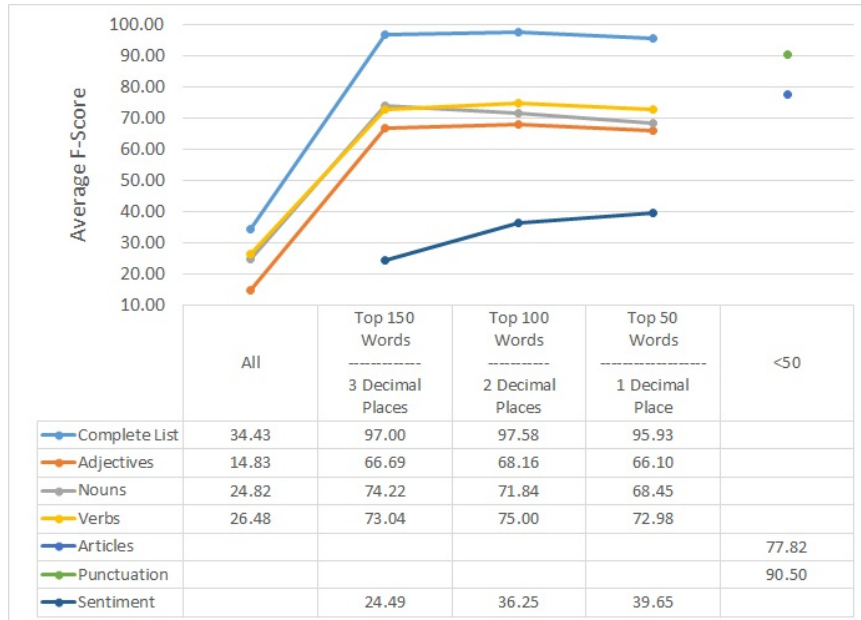


Figure 10: Average F-Score for Z-Score Frequencies using the KNN algorithm on the

Multiple Author Corpora

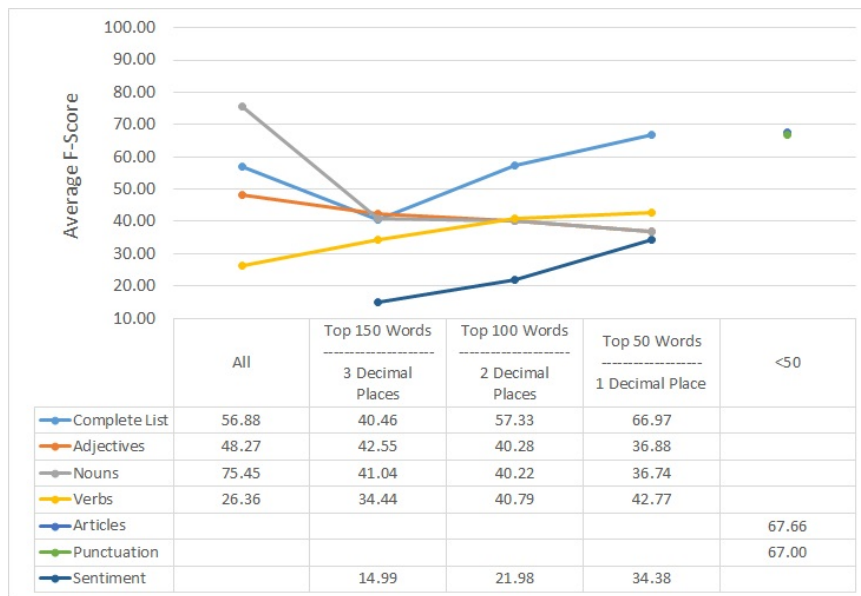


Figure 11: Average F-Score for Z-Score Frequencies using the Naive Bayes algorithm

on the Multiple Author Corpora

4.1.2 TF/IDF Results

Figures 12 and 13 show Term-Frequency/Inverse Document Frequency (TF/IDF) scores for KNN- and Naïve-Bayes-based models of the two author corpora. Figures 14 and 15 show these scores for the multi-author corpora.

- **Traditional Features.** TF/IDF performance was fairly consistent across all levels of granularity, especially with KNN. Naive Bayes had high average f-scores, with Adjectives and Verbs performing slightly lower than other parts-of-speech features. As with z-score, punctuation performed well, outperforming articles in average f-score. KNN had a higher average performance with z-score, while Naive Bayes had a better average performance using TF/IDF.
- **Sentiment Scores.** Traditional models outperformed sentiment-based models. Sentiment-based models performed roughly as well as z-scores and TD/IDF scores, with only a slightly higher performance for the multiple corpora.

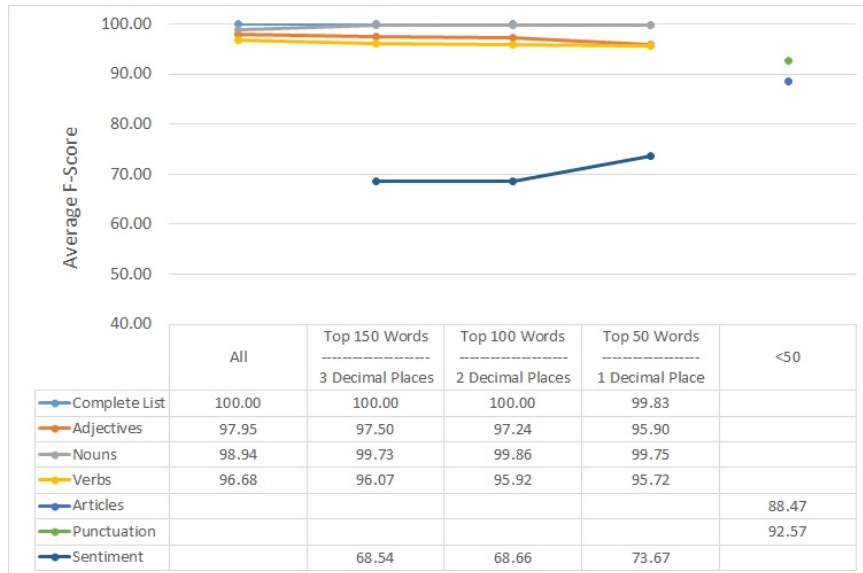


Figure 12: Average F-Score for TF/IDF using the KNN algorithm on the Two Author Corpora

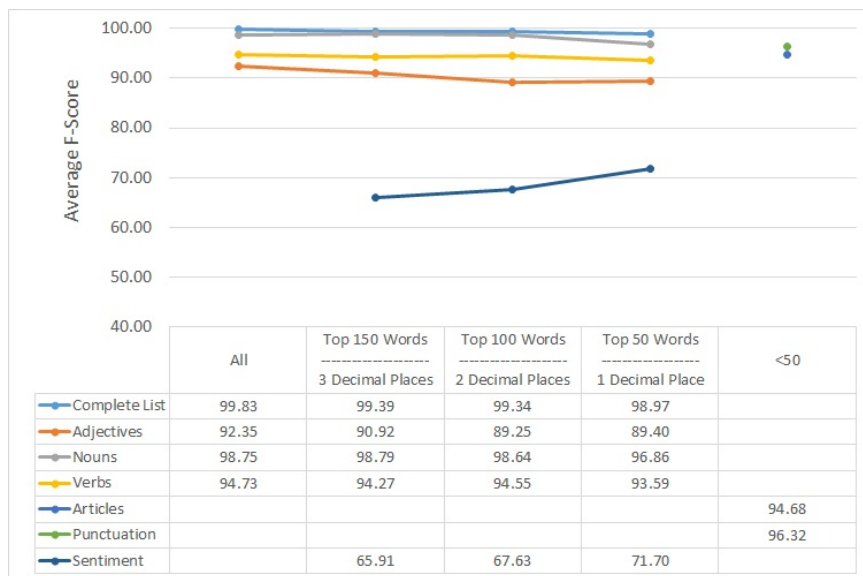


Figure 13: Average F-Score for TF/IDF using the Naive Bayes algorithm on the Two Author Corpora



Figure 14: Average F-Score for TF/IDF using the KNN algorithm on the Multiple Author Corpora

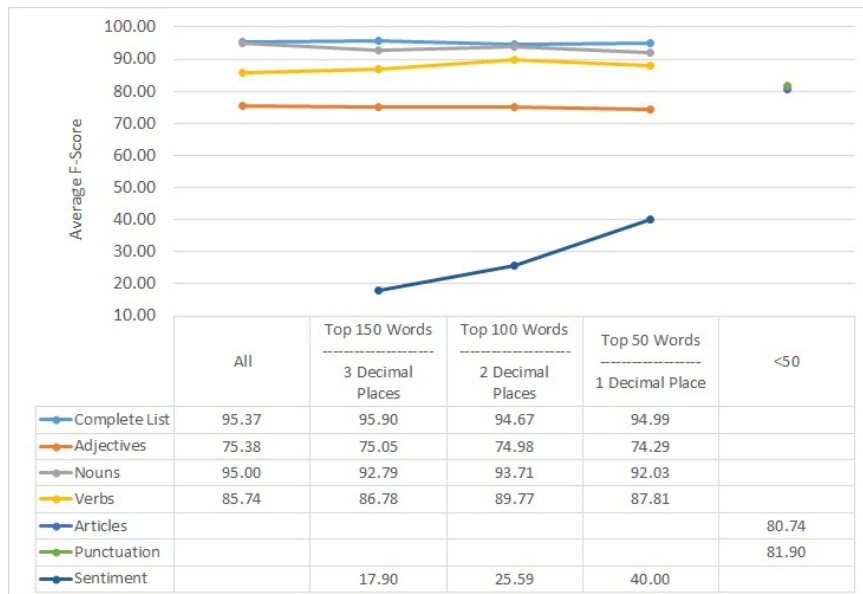


Figure 15: Average F-Score for TF/IDF using the Naive Bayes algorithm on the Multiple Author Corpora

4.1.3 TF Results

Figures 16 and 17 show Term-Frequency (TF) scores for KNN- and Naïve-Bayes-based models of the two-author corpora. Figures 18 and 19 show these scores for the multi-author corpora.

Term Frequency results for the two-authors corpora can be found in figure 16 for KNN and in figure 17 for Naive Bayes. Results for the multi-authors corpora are shown in figure 18 for KNN and in figure 19 for Naive Bayes.

- **Traditional Features.** Traditional stylometric features performed relatively the same as TF/IDF for the two author and multiple author corpora.
- **Sentiment Scores.** The minor improvement in TF-based sentiment scores relative to TF/IDF scores was deemed insignificant. This included minor improvements in the multiple author corpora for 3 and 2 decimal place options.



Figure 16: Average F-Score for Term Frequencies using the KNN algorithm on the Two Author Corpora



Figure 17: Average F-Score for Term Frequencies using the Naive Bayes algorithm on the Two Author Corpora

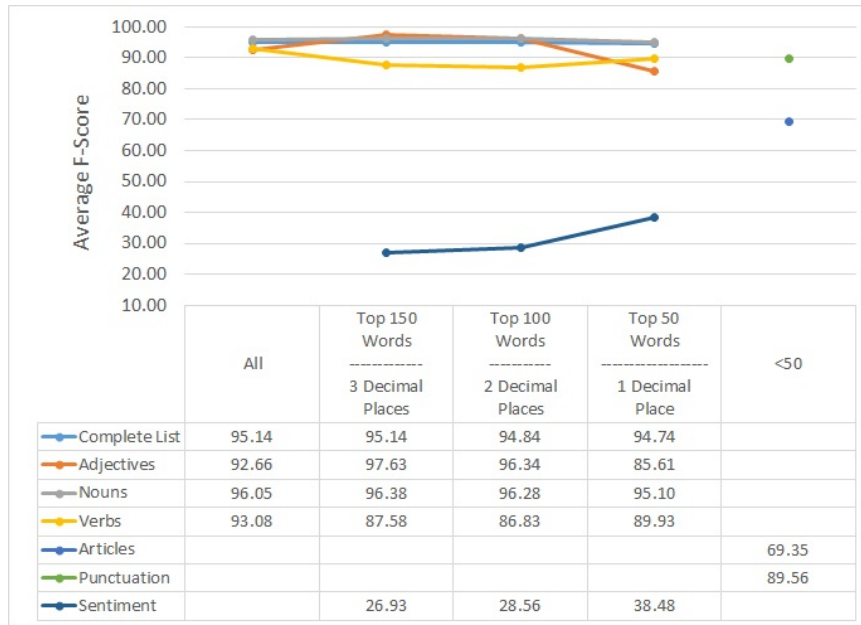


Figure 18: Average F-Score for Term Frequencies using the KNN algorithm on the

Multiple Author Corpora

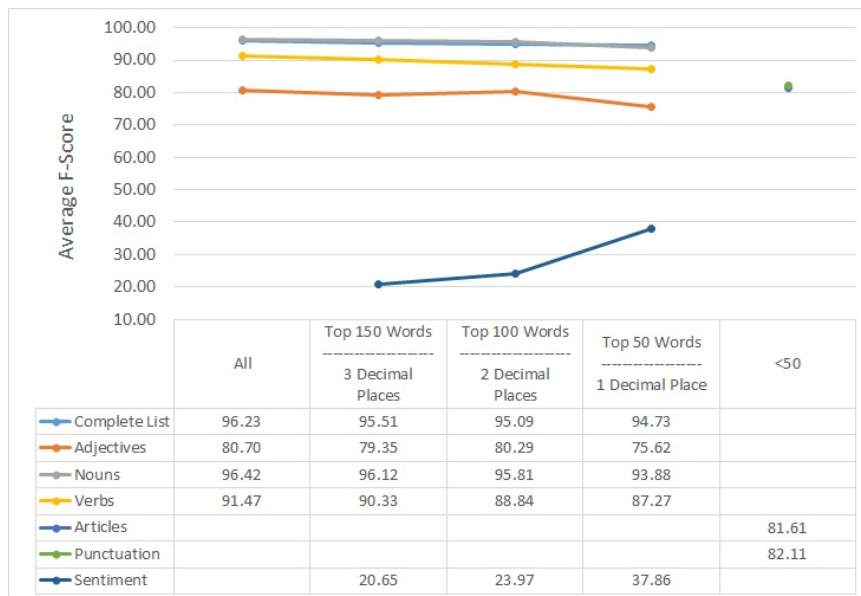


Figure 19: Average F-Score for Term Frequencies using the Naive Bayes algorithm

on the Multiple Author Corpora

5 CONCLUSION AND FUTURE RESEARCH

5.1 Methodological Quality

The strategies for carrying out this research were informed to varying degrees by Rudman's concerns about AA research quality (see 2.2). The following is an assessment of the extent to which this work meets Rudman's standards:

- **Pacing.** The time devoted to the study, I believe, was adequate for what was intended as a preliminary study of Sentiment Analysis for Authorship Attribution.
- **Literature review.** The review that was undertaken for this work, while adequate for an M.S. thesis, is quite incomplete relative to the overall body of research in AA. This includes, for example, works in languages other than English that I neither accessed nor read.
- **Statistical methods.** F-scores, TF/IDF, and TF metrics are in general use among linguists. Z-scores are a standard statistical measurement, like standard deviation or mean, and would not be considered by Rudman as a borrowed statistic.
- **Data cleansing.** While I have made a good-faith effort to remove the errors from my corpus, the amount of work needed to retain nonstandard characters and hand-verify the cleaned documents exceeded the available time for this work. More work should be devoted to finding errors in corpus texts.

- **Prior training in appropriate disciplines.** While I have a basic understanding of linguistics and English, this research could possibly have benefitted from my having a higher degree of expertise in both fields.
- **Breadth of knowledge.** As noted above, I have read some of the seminal works for authorship attribution but have possibly missed research in other countries and in other languages.
- **Characterization of the research’s limitations.** The discussions of this study’s methodologies, together with this section of the thesis, are intended in part to expose decisions and actions that could have yielded errors in my results. I hope attempt at full transparency will help future researchers to avoid similar mistakes.

5.2 Analysis of Results

Algorithms KNN generally performed better than Naive Bayes, except in the case of punctuation and articles which often fared better with Naive Bayes. While KNN operates well for one the simplest data mining algorithms, other, more complex, classification algorithms are now in common use, such as Support Vector Machines. Future research needs to determine if the more complex algorithms will perform better, as shown in [6], as well as the extent to which their higher calculation cost warrant their use.

Frequency Type TF-based assessments exhibited consistently higher average f-scores than z-score- and TF/IDF-based assessments. TF scores, moreover, were relatively easy to calculate. The importance of this lower calculation cost for term frequency cannot be overstated, in that it applies to initial calculations as well as recalculations of corpus scores. Each text’s scores are independent and this may help with the study of future corpora that contain thousands or even millions of entries. Frequency metrics should be rated by their cost of computation as well as their accuracy. This will help to better match frequency metrics to suitable “real world” applications that balance accuracy with computational performance.

Granularity With one exception, lowering granularity for the traditional features decreased the models’ accuracy. This one exception involved the z-score-based measure of KNN model accuracy, relative to the transition from all tokens to the top 150 tokens. For sentiment score frequencies, lowering the granularity always increased performance. Future research needs to further examine these changes in granularity. For the traditional features, this study was limited to comparisons based on lists of top 50, 100, and 150 tokens. Future research should assess a broader spectrum of granularities to obtain a clearer picture of how token counts affect performance. As for the sentiment scores, future research should study the full range of rounding choices, from rounding to integers (-1, 0, 1) to no round off.

Stylometric Features Traditional features far outperformed sentiment score frequencies. While this result does not rule out the potential use of sentiment scores in AA, it suggests their relative ineffectiveness for this corpus. Sentiment score frequency might fare better on other types of text. Besides changing the corpus texts, this study did not experiment with combinations of stylometric features. Perhaps sentiment scores would have improved the accuracy of traditional stylometric features' models. This would be especially useful if it could improve the accuracy of the models formed from the multiple authors corpora.

Corpus Selection While the Asimov corpus provided a worthwhile learning experience, these results do not readily generalize to other corpora. The lack of standard corpora for investigating new stylometric features is, in my opinion, the major problem in stylometric research. A set of standard corpora should cover the breadth of English literature. A possible starting point for defining the corpora would be existing taxonomies of written works, such as the Dewey Decimal or Library of Congress Classification system. I plan to focus on one category of a classification system and slowly build my corpus to include new texts that fit the different categories of the chosen classification system.

Changes in Sentence Count While preparing the corpus for sentiment extract, this study divided the corpus documents into smaller sub-documents, based on 500, 1000, and 1500 sentences. This study failed to investigate how sub-document size of sentences affected the performance of the predictive data

models, because the final results were flawed. In creating the sub-documents, this study did not keep a uniform number of sub-documents in each k-sentence group. This means that when a large work is split up into smaller pieces, 500 vs 1,500, it will have more pieces to contribute to the new sub-document corpus. This study only kept the number of documents per author per corpus uniform, and allowed the 500, 1000, and 1,500 sub-document corpora to differ in the number of documents they contained. For this reason, it is unknown if the increases and decreases in model performance can be attributed to the change in the number of sentences in each document or to the change in the number of documents in each corpus. Future research should study the effect of increasing and decreasing the size of each corpus document, while keeping the number of documents contained in each corpus uniform.

BIBLIOGRAPHY

- [1] John Burrows. ‘delta’: A measure of stylistic difference and a guide to likely authorship. *Literary and Linguistic Computing*, 17(3):267–287, 2002.
- [2] Gobinda Chowdhury. *Introduction to modern information retrieval*. Facet publishing, 2010.
- [3] Hamish Cunningham, Diana Maynard, Kalina Bontcheva, and Valentin Tablan. GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL’02)*, 2002.
- [4] Jill M Farrington. How to be a literary detective: Authorship attribution. *A brief introduction to cusum analysis. available online at <http://members.aol.com/qsums/QsumIntroduction.html>*, 2004.
- [5] GATE. Chapter 6: Annie: a nearly-new information extraction system.
- [6] Jack William Grieve. *Quantitative authorship attribution: A history and an evaluation of techniques*. PhD thesis, Department of Linguistics-Simon Fraser University, 2005.
- [7] David I Holmes. The evolution of stylometry in humanities scholarship. *Literary and linguistic computing*, 13(3):111–117, 1998.
- [8] Patrick Juola. Authorship attribution. *Foundations and Trends in information Retrieval*, 1(3):233–334, 2006.

- [9] Anthony Kenny. The computation of style: An introduction to statistics for students of literature and humanities. 1982.
- [10] Kim Luyckx and Walter Daelemans. Authorship attribution and verification with many authors and limited data. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, pages 513–520. Association for Computational Linguistics, 2008.
- [11] Thomas Corwin Mendenhall. The characteristic curves of composition. *Science*, pages 237–249, 1887.
- [12] Tetsuya Nasukawa and Jeonghee Yi. Sentiment analysis: Capturing favorability using natural language processing. In *Proceedings of the 2nd international conference on Knowledge capture*, pages 70–77. ACM, 2003.
- [13] Joseph Rudman. The state of authorship attribution studies: Some problems and solutions. *Computers and the Humanities*, 31(4):351–365, 1997.
- [14] Michael William Smith. Recent experience and new developments of methods for the determination of authorship. *ALLC BULL.*, 11(3):73–82, 1983.
- [15] Efstathios Stamatatos. A survey of modern authorship attribution methods. *Journal of the American Society for information Science and Technology*, 60(3):538–556, 2009.
- [16] Tzvetan Todorov. The place of style in the structure of the text. In *Literary Style: A Symposium*, volume 32, 1971.

- [17] Carrington Bonsor Williams. *Style and vocabulary: numerical studies*. Griffin, 1970.
- [18] Rong Zheng, Jiexun Li, Hsinchun Chen, and Zan Huang. A framework for authorship identification of online messages: Writing-style features and classification techniques. *Journal of the American Society for Information Science and Technology*, 57(3):378–393, 2006.

APPENDIX

APPENDIX A Number of Author Entries in Each Corpus

Table 1: Entire Corpus - Number of Entries Per Document

Author	Title	# of Entries (500 Sentences)	# of Entries (1000 Sentences)	# of Entries (1500 Sentences)
I. Asimov	Forward the Foundation	12	6	4
I. Asimov	Foundation	6	3	2
I. Asimov	Pebble in the Sky	6	3	2
I. Asimov	Second Foundation	6	3	2
I. Asimov	The Caves of Steel	6	3	2
I. Asimov	The Currents of Space	6	3	2
I. Asimov	The Robots of Dawn	12	6	4
I. Asimov	The Stars Like Dust	6	3	2
G. Bear	Foundation and Chaos	12	6	4
G. Benford	Foundation's Fear	12	6	4
D. Brin	Foundation's Triumph	12	6	4
R. M. Allen	Caliban	12	6	4
W. F. Wu	Predator	6	3	2
W. F. Wu	Warrior	6	3	2

Table 2: Foundation Corpus (1980's Era) - Number of Entries Per Document

Author	Title	# of Entries (500 Sentences)	# of Entries (1000 Sentences)	# of Entries (1500 Sentences)
I. Asimov	Foundation's Edge	13	6	4
I. Asimov	Foundation and Earth	13	6	4
I. Asimov	Prelude to Foundation	13	6	4
G. Benford	Foundation's Fear	13	6	4
G. Bear	Foundation and Chaos	13	6	4
D. Brin	Foundation's Triumph	13	6	4

Table 3: Foundation Corpus (1950's Era) - Number of Entries Per Document

Author	Title	# of Entries (500 Sentences)	# of Entries (1000 Sentences)	# of Entries (1500 Sentences)
I. Asimov	Foundation	8	4	2
I. Asimov	Foundation and Empire	8	4	2
I. Asimov	Second Foundation	8	4	2
G. Benford	Foundation's Fear	8	4	2
G. Bear	Foundation and Chaos	8	4	2
D. Brin	Foundation's Triumph	8	4	2

Table 4: Galactic Empire Corpus - Number of Entries Per Document

Author	Title	# of Entries (500 Sentences)	# of Entries (1000 Sentences)	# of Entries (1500 Sentences)
I. Asimov	The Stars, Like Dust	7	3	2
I. Asimov	The Currents of Space	7	3	2
I. Asimov	Pebble in the Sky	7	3	2
R. M. Allen	Caliban	7	3	2
R. M. Allen	Inferno	7	3	2
R. M. Allen	Utopia	7	3	

Table 5: Robot Corpus - Number of Entries Per Document

Author	Title	# of Entries (500 Sentences)	# of Entries (1000 Sentences)	# of Entries (1500 Sentences)
I. Asimov	The Caves of Steel	7	3	2
I. Asimov	The Naked Sun	7	3	2
I. Asimov	The Robots of Dawn	7	3	2
I. Asimov	Robots and Empire	7	3	2
W. F. Wu	Predator	7	3	2
W. F. Wu	Marauder	7	3	2
W. F. Wu	Warrior	7	3	2
W. F. Wu	Dictator	7	3	2

APPENDIX B List of Failed Sentences for Alchemy API

Prelude To Foundation - Asimov

"State functions are tiring, Sire," murmured Demerzel.

Robots and Empire - Asimov

I do not wish to go, Partner Elijah.

It read: "Ceterum censeo, delenda est Carthago."

The Caves of Steel - Asimov

Dr. Gerrigel extended his hand and shook Daneel's.

The Star's Like Dust - Asimov

I am a man, a mature man.

Inferno - Macbride

I am afraid I don't understand, sir.

Fredda blinked, shook her head, stared at Kresh.

Utopia - Macbride

I ah-ah-ah, sir, I ah-" "Yes?

STOP COMIT + PUT 500,000 TDC N PBI ACCT 18083-19109 ORE BEDDL

WIL DI.

Yes, sir, I am afraid I did.

Invader - Wu

Inside, several men sat at tables drinking.

Predators - Wu

Doc Nystrom don't program no idiot robots.

APPENDIX C List of Top “X” Words for Entire Corpus

Table 6: Entire Corpus - All Words - Top 50 Tokens

A	BY	HIS	ONE	THIS
ALL	COULD	I	OUT	TO
AN	DO	IF	SAID	WAS
AND	FOR	IN	SHE	WE
ARE	FROM	IS	SO	WERE
AS	HAD	IT	THAT	WHAT
AT	HAVE	NO	THE	WILL
BE	HE	NOT	THEM	WITH
BEEN	HER	OF	THERE	WOULD
BUT	HIM	ON	THEY	YOU

Table 7: Entire Corpus - All Words - Top 100 Tokens

A	EVEN	ITS	OUR	TIME
ABOUT	FIRST	JUST	OUT	TO
ALL	FOR	KNOW	OVER	TWO
AN	FROM	LIKE	OWN	UP
AND	HAD	MAN	SAID	US
ANY	HAS	ME	SEE	VERY
ARE	HAVE	MIGHT	SHE	WAS
AS	HE	MORE	SO	WAY
AT	HER	MUCH	SOME	WE
BACK	HERE	MUST	THAN	WELL
BE	HIM	MY	THAT	WERE
BEEN	HIS	NO	THE	WHAT
BEFORE	HOW	NOT	THEIR	WHEN
BUT	HUMAN	NOW	THEM	WHICH
BY	I	OF	THEN	WHO
CAN	IF	ON	THERE	WILL
COULD	IN	ONE	THEY	WITH
DID	INTO	ONLY	THINK	WOULD
DO	IS	OR	THIS	YOU
DOWN	IT	OTHER	THROUGH	YOUR

Table 8: Entire Corpus - All Words - Top 150 Tokens

A	EYES	LET	OVER	THOUGH
ABOUT	FELT	LIKE	OWN	THOUGHT
AFTER	FIRST	LITTLE	PEOPLE	THROUGH
AGAIN	FOR	LONG	RIGHT	TIME
ALL	FROM	LOOKED	ROBOT	TO
AM	GET	MADE	ROBOTS	TOO
AN	GO	MAKE	ROOM	TWO
AND	HAD	MAN	SAID	UP
ANY	HARI	MAY	SEE	US
ARE	HAS	ME	SEEMED	VERY
AS	HAVE	MIGHT	SELDON	WANT
AT	HE	MORE	SHE	WAS
AWAY	HER	MOST	SO	WAY
BACK	HERE	MUCH	SOME	WE
BALEY	HIM	MUST	SOMETHING	WELL
BE	HIMSELF	MY	STILL	WERE
BEEN	HIS	NEVER	SUCH	WHAT
BEFORE	HOW	NO	TAKE	WHEN
BEING	HUMAN	NOT	THAN	WHERE
BUT	HUNTER	NOTHING	THAT	WHICH
BY	I	NOW	THE	WHO
CAN	IF	OF	THEIR	WHY
COME	IN	ON	THEM	WILL
COULD	INTO	ONCE	THEN	WITH
COURSE	IS	ONE	THERE	WITHOUT
DANEEL	IT	ONLY	THESE	WOULD
DID	ITS	OR	THEY	YEARS
DO	JUST	OTHER	THINK	YET
DOWN	KNEW	OUR	THIS	YOU
EVEN	KNOW	OUT	THOSE	YOUR

Table 9: Entire Corpus - Adjectives - Top 50 Tokens

ABLE	FULL	LIGHT	OLD	SECOND
ANCIENT	GOOD	LITTLE	OPEN	SEVERAL
BEST	GREAT	LONG	OTHER	SHORT
BETTER	HARD	MANY	OWN	SINGLE
CERTAIN	HUMAN	MORE	PAST	SMALL
CLEAR	IMPORTANT	MOST	PERSONAL	SUCH
DEEP	LARGE	MUCH	POSSIBLE	SURE
DIFFERENT	LAST	NECESSARY	REAL	WHOLE
FEW	LEAST	NEW	RIGHT	YOUNG
FIRST	LESS	NEXT	SAME	TRUE

Table 10: Entire Corpus - Adjectives - Top 100 Tokens

ABLE	DIFFERENT	LEAST	OPEN	SEVERAL
AFRAID	DIFFICULT	LESS	ORDINARY	SHORT
ANCIENT	EARLY	LIGHT	OTHER	SIMPLE
AWARE	EASY	LIKELY	OWN	SINGLE
BAD	EMPTY	LITTLE	PAST	SMALL
BEST	ENTIRE	LONG	PERSONAL	SOCIAL
BETTER	FEW	LOW	PHYSICAL	STRAIGHT
BIG	FIRST	MANY	POLITICAL	STRANGE
BLACK	FREE	MENTAL	POSITRONIC	STRONG
BLUE	FULL	MORE	POSSIBLE	SUCH
BRIGHT	GOOD	MOST	PRESENT	SUDDEN
CAPABLE	GREAT	MUCH	PRIVATE	SURE
CERTAIN	GREATER	NATURAL	READY	THIN
CLEAR	HARD	NECESSARY	REAL	TINY
COLD	HIGH	NEW	RED	WHITE
COMPLETE	HUMAN	NEXT	RIGHT	WHOLE
DANGEROUS	IMPORTANT	NODDED	SAFE	WORSE
DARK	IMPOSSIBLE	OBVIOUS	SAME	WRONG
DEAD	LARGE	OLD	SECOND	YOUNG
DEEP	LAST	ONLY	SECRET	TRUE

Table 11: Entire Corpus - Adjectives - Top 150 Tokens

AFRAID	LAST	SPECIAL	WORSE	RIGHT
SILENT	DIFFICULT	MORE	ONLY	YOUNG
POSSIBLE	MENTAL	GENTLE	EARLY	NEARBY
BRIGHT	QUIET	HIGH	WARM	TRUE
SURE	SECOND	TALL	FREE	SAME
NECESSARY	SEVERAL	FURTHER	BETTER	QUICK
NEW	SHARP	MERE	WILLING	RED
DIFFERENT	POSITRONIC	ODD	DANGEROUS	SECRET
LOW	CAPABLE	OLDER	ANCIENT	STRANGE
GERMAN	PRIVATE	HUMAN	DEEP	IMPORTANT
WRONG	UNUSUAL	TINY	PRESENT	ORDINARY
LITTLE	OLD	LIKELY	VAST	OPEN
PAST	BLACK	LOWER	SUDDEN	SINGLE
CLEAR	WIDE	PHYSICAL	NODDED	ENTIRE
CERTAIN	LOCAL	POWERFUL	TIRED	PERSONAL
READY	EASY	POOR	NEXT	PARTICULAR
COLD	IMPOSSIBLE	MUCH	MILITARY	BAD
FINAL	MANY	SIMPLE	NORMAL	STRONG
AWARE	WHOLE	SOFT	SERIOUS	LEAST
FULL	EMPTY	THIRD	MOST	ROBOTIC
OWN	BLUE	HUGE	USEFUL	LARGE
GREATER	WHITE	FEW	OTHER	NATURAL
SMALL	REAL	GREAT	USUAL	GOOD
DEAD	ABLE	VERY	COMPLETE	LONG
THIN	GENERAL	MAIN	LIGHT	PUBLIC
HARD	CLOSER	SOCIAL	BEST	DISTANT
DARK	ALIVE	CAREFUL	POLITICAL	BASIC
LESS	ORIGINAL	PROPER	SAFE	INTERESTED
COMMON	FAINT	ENOUGH	SORRY	BIG
SUCH	STRAIGHT	OBVIOUS	SHORT	FIRST

Table 12: Entire Corpus - Nouns - Top 50 Tokens

ANYTHING	EYES	MAN	PLANET	THING
CASE	FACE	MATTER	POINT	THINGS
CITY	FACT	MEN	RIGHT	TIME
COURSE	GALAXY	MIND	ROBOT	VOICE
DANEEL	HAND	MOMENT	ROBOTS	WAY
DAY	HANDS	NIGHT	ROOM	WOMAN
DOOR	HARI	NOTHING	SELDON	WORK
DR	HEAD	PART	SIDE	WORLD
EARTH	LAW	PEOPLE	SOMETHING	WORLDS
EMPIRE	LIFE	PLACE	SPACE	YEARS

Table 13: Entire Corpus - Nouns - Top 100 Tokens

AIR	DR	HUNTER	ORDER	SOMEONE
ANYONE	EARTH	IMPERIAL	OTHERS	SOMETHING
ANYTHING	EMPEROR	INFORMATION	PART	SPACE
BALEY	EMPIRE	ISHIHARA	PEOPLE	STEVE
BIRON	END	JANE	PLACE	THING
BIT	EYES	KIND	PLANET	THINGS
BODY	FACE	KRESH	POINT	TIME
BRAIN	FACT	LAW	POWER	TIMES
CALIBAN	FASTOLFE	LAWS	PSYCHOHISTORY	TRANTOR
CAR	FATHER	LIFE	QUESTION	VOICE
CASE	FOUNDATION	LODOVIK	R	WALL
CHANCE	GALAXY	MAN	RAYCH	WAY
CHEN	GISKARD	MATTER	REASON	WAYNE
CITY	HAND	MEN	RIGHT	WOMAN
COURSE	HANDS	MIND	ROBOT	WORD
DANEEL	HARI	MINISTER	ROBOTS	WORDS
DAY	HEAD	MOMENT	ROOM	WORK
DAYS	HISTORY	NAME	SELDON	WORLD
DOOR	HUMANITY	NIGHT	SHIP	WORLDS
DORS	HUMANS	NOTHING	SIDE	YEARS

Table 14: Entire Corpus - Nouns - Top 150 Tokens

AIR	DONALD	HUMANITY	NOTHING	SITUATION
ALVAR	DOOR	HUMANS	OFFICE	SOCIETY
ANYONE	DORS	HUNTER	ORDER	SOMEONE
ANYTHING	DR	IDEA	OTHERS	SOMETHING
ARM	EARTH	IMPERIAL	PALACE	SORT
ARMS	ELIJAH	INFORMATION	PART	SPACE
AURORA	EMPEROR	ISHIHARA	PEOPLE	STEVE
BALEY	EMPIRE	JANE	PLACE	TABLE
BEING	END	JUDY	PLAN	THING
BEINGS	EVERYTHING	KIND	PLANCH	THINGS
BIRON	EYES	KNOWLEDGE	PLANET	TIME
BIT	FACE	KRESH	POINT	TIMES
BODY	FACT	LAW	POWER	TRANTOR
BRAIN	FASTOLFE	LAWS	PSYCHOHISTORY	TROUBLE
CALIBAN	FATHER	LIFE	QUESTION	USE
CAR	FEET	LODOVIK	R	VOICE
CASE	FLOOR	MAN	RAYCH	VOLTAIRE
CHAIR	FOUNDATION	MARCIA	REASON	WALL
CHANCE	GALACTIC	MATTER	REST	WANDA
CHEN	GALAXY	MC	RIGHT	WAY
CITY	GISKARD	MEN	ROBOT	WAYNE
CLEON	HAIR	MIND	ROBOTS	WOMAN
COMMISSIONER	HAND	MINISTER	ROOM	WORD
CONTROL	HANDS	MINUTES	SECOND	WORDS
COURSE	HARDIN	MOMENT	SELDON	WORK
DANEEL	HARI	MR	SENSE	WORLD
DANGER	HEAD	MULE	SHIP	WORLDS
DAY	HISTORY	NAME	SIDE	YEAR
DAYS	HOME	NIGHT	SINTER	YEARS
DEATH	HOURS	NONE	SIR	YUGO

Table 15: Entire Corpus - Verbs - Top 50 Tokens

AM	DOES	HAS	MADE	THINK
ARE	DONE	HAVE	MAKE	THOUGHT
BE	FELT	IS	PUT	TOLD
BEEN	FIND	KEEP	SAID	TOOK
BEING	FOUND	KNEW	SAW	TURNED
BELIEVE	GET	KNOW	SAY	USED
CAME	GO	LEFT	SEE	WANT
COME	GOING	LET	SEEMED	WAS
DID	GOT	LOOK	TAKE	WENT
DO	HAD	LOOKED	TELL	WERE

Table 16: Entire Corpus - Verbs - Top 100 Tokens

AM	FELT	IS	PUT	THINK
ARE	FIND	KEEP	REACHED	THOUGHT
ASK	FOUND	KEPT	REMEMBER	TOLD
ASKED	GAVE	KNEW	SAID	TOOK
BE	GET	KNOW	SAT	TRIED
BEEN	GIVE	KNOWN	SAW	TRY
BEGAN	GIVEN	LEAVE	SAY	TRYING
BEING	GO	LEFT	SEE	TURNED
BELIEVE	GOING	LET	SEEM	UNDERSTAND
BROUGHT	GONE	LIKE	SEEMED	USE
CALLED	GOT	LOOK	SEEN	USED
CAME	HAD	LOOKED	SET	WAITING
COME	HAPPENED	LOOKING	SPEAK	WALKED
COMING	HAS	LOST	SPOKE	WANT
DID	HAVE	MADE	STOOD	WANTED
DO	HAVING	MAKE	STOPPED	WAS
DOES	HEAR	MEAN	SUPPOSE	WENT
DOING	HEARD	MOVED	TAKE	WERE
DONE	HELD	NEED	TAKEN	WORK
FEEL	HELP	PASSED	TELL	WORKED

Table 17: Entire Corpus - Verbs - Top 150 Tokens

ADDED	EXPECT	KEPT	READ	TAKING
ALLOW	FEEL	KNEW	REALIZED	TALK
ALLOWED	FELT	KNOW	REMAINED	TALKING
AM	FIND	KNOWN	REMEMBER	TELL
ARE	FOLLOWED	LEARNED	RETURNED	THINK
ASK	FORCED	LEAVE	RUNNING	THOUGHT
ASKED	FOUND	LED	SAID	TOLD
BE	GAVE	LEFT	SAT	TOOK
BECAME	GET	LET	SAW	TRIED
BECOME	GETTING	LIFTED	SAY	TRY
BEEN	GIVE	LIKE	SEE	TRYING
BEGAN	GIVEN	LIVE	SEEM	TURNED
BEING	GO	LOOK	SEEMED	UNDERSTAND
BELIEVE	GOING	LOOKED	SEEMS	USE
BRING	GONE	LOOKING	SEEN	USED
BROUGHT	GOT	LOST	SENT	USING
CALL	HAD	MADE	SET	WAIT
CALLED	HAPPEN	MAKE	SHOOK	WAITED
CAME	HAPPENED	MAKING	SMILED	WAITING
CAUGHT	HAS	MEAN	SPEAK	WALKED
CLOSED	HAVE	MEANT	SPOKE	WANT
COME	HAVING	MET	STARED	WANTED
COMING	HEAR	MOVED	STARTED	WAS
DECIDED	HEARD	MOVING	STEPPED	WATCHED
DID	HELD	NEED	STOOD	WENT
DO	HELP	NEEDED	STOP	WERE
DOES	HOPE	OPENED	STOPPED	WISH
DOING	IMAGINE	PASSED	SUPPOSE	WORK
DONE	IS	PUT	TAKE	WORKED
ENTERED	KEEP	REACHED	TAKEN	WORKING

APPENDIX D Results for Experiment Data Models

D.1 KNN POS Results - Two Authors

Table 18: KNN POS Results for Entire Corpus(500 Sentences) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	51.67	51.67	51.70	51.68	75.00	75.00	75.00	75.00
Z Score-150 Words	75.83	75.83	76.73	76.28	100.00	100.00	100.00	100.00
Z Score-100 Words	77.50	77.50	82.24	79.80	100.00	100.00	100.00	100.00
Z Score-50 Words	69.17	69.17	73.19	71.12	99.17	99.17	99.18	99.17
TF/IDF	86.67	86.67	88.78	87.71	100.00	100.00	100.00	100.00
TF/IDF-150 Words	85.00	85.00	87.68	86.32	100.00	100.00	100.00	100.00
TF/IDF-100 Words	84.17	84.17	85.85	85.00	100.00	100.00	100.00	100.00
TF/IDF-50 Words	83.33	83.33	85.25	84.28	99.17	99.17	99.18	99.17
Term Frequency	90.00	90.00	91.67	90.83	100.00	100.00	100.00	100.00
Term Frequency-150 Words	92.50	92.50	93.48	92.99	100.00	100.00	100.00	100.00
Term Frequency-100 Words	90.00	90.00	91.67	90.83	100.00	100.00	100.00	100.00
Term Frequency-50 Words	85.83	85.83	87.08	86.45	99.17	99.17	99.18	99.17
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	85.83	85.83	86.66	86.24	60.00	60.00	77.78	67.74
Z Score -150 Words	85.83	85.83	86.08	85.95	81.67	81.67	81.99	81.83
Z Score-100 Words	87.50	87.50	87.59	87.54	84.17	84.17	84.41	84.29
Z Score-50 Words	85.00	85.00	85.16	85.08	80.83	80.83	81.05	80.94
TF/IDF	85.00	85.00	85.63	85.31	96.67	96.67	96.72	96.69
TF/IDF-150 Words	99.17	99.17	99.18	99.17	94.17	94.17	94.78	94.47
TF/IDF-100 Words	99.17	99.17	99.18	99.17	95.00	95.00	95.05	95.02
TF/IDF-50 Words	97.50	97.50	97.62	97.56	94.17	94.17	94.28	94.22

Term Frequency	100.00	100.00	100.00	100.00	97.50	97.50	97.51	97.50
Term Frequency-150 Words	99.17	99.17	99.18	99.17	94.17	94.17	94.48	94.32
Term Frequency-100 Words	100.00	100.00	100.00	100.00	95.00	95.00	95.00	95.00
Term Frequency-50 Words	97.50	97.50	97.51	97.50	94.17	94.17	94.48	94.32

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	91.67	91.67	92.42	92.04
TF/IDF	85.00	85.00	85.63	85.31
Term Frequency	84.17	84.17	84.64	84.40

	Punctuation			
	Accuracy	Recall	Precision	F-Score
Z Score	91.67	91.67	91.71	91.69
TF/IDF	85.00	85.00	85.04	85.02
Term Frequency	85.53	85.83	86.08	85.95

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	77.40	14.47	92.04	51.68
Z Score-150 Words	86.02	8.77	100.00	76.28
Z Score-100 Words	87.91	7.50	100.00	79.80
Z Score-50 Words	84.08	10.08	99.17	71.12
TF/IDF	90.01	6.04	100.00	85.02
TF/IDF-150 Words	94.99	5.43	100.00	86.32
TF/IDF-100 Words	94.80	5.96	100.00	85.00
TF/IDF-50 Words	93.81	5.78	99.17	84.28
Term Frequency	93.12	6.41	100.00	84.40
Term Frequency-150 Words	96.62	3.02	100.00	92.99
Term Frequency-100 Words	96.46	3.84	100.00	90.83
Term Frequency-50 Words	94.36	4.89	99.17	86.45

Table 19: KNN POS Results for Entire Corpus(1000 Sentences) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	55.00	55.00	76.32	63.93	93.33	93.33	93.53	93.43
Z Score-150 Words	88.33	88.33	90.54	89.42	98.33	98.33	98.39	98.36
Z Score-100 Words	86.67	86.67	89.47	88.05	98.33	98.33	98.39	98.36
Z Score-50 Words	81.67	81.67	86.59	84.06	98.33	98.33	98.39	98.36
TF/IDF	96.67	96.67	96.88	96.77	100.00	100.00	100.00	100.00
TF/IDF-150 Words	95.00	95.00	95.45	95.22	100.00	100.00	100.00	100.00
TF/IDF-100 Words	95.00	95.00	95.45	95.22	100.00	100.00	100.00	100.00
TF/IDF-50 Words	93.33	93.33	94.12	93.72	100.00	100.00	100.00	100.00
Term Frequency	95.00	95.00	95.45	95.22	100.00	100.00	100.00	100.00
Term Frequency-150 Words	95.00	95.00	95.45	95.22	100.00	100.00	100.00	100.00
Term Frequency-100 Words	96.67	96.67	96.88	96.77	100.00	100.00	100.00	100.00
Term Frequency-50 Words	93.33	93.33	94.12	93.72	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	75.00	75.00	83.33	78.95	65.00	65.00	79.41	71.49
Z Score -150 Words	80.00	80.00	85.71	82.76	95.00	95.00	95.45	95.22
Z Score-100 Words	90.00	90.00	90.00	90.00	91.67	91.67	92.86	92.26
Z Score-50 Words	80.00	80.00	80.00	80.00	93.33	93.33	93.53	93.43
TF/IDF	100.00	100.00	100.00	100.00	96.67	96.67	96.88	96.77
TF/IDF-150 Words	100.00	100.00	100.00	100.00	95.00	95.00	95.45	95.22
TF/IDF-100 Words	100.00	100.00	100.00	100.00	95.00	95.00	95.45	95.22
TF/IDF-50 Words	100.00	100.00	100.00	100.00	95.00	95.00	95.45	95.22
Term Frequency	100.00	100.00	100.00	100.00	98.33	98.33	98.39	98.36
Term Frequency-150 Words	100.00	100.00	100.00	100.00	95.00	95.00	95.45	95.22
Term Frequency-100 Words	100.00	100.00	100.00	100.00	95.00	95.00	95.45	95.22
Term Frequency-50 Words	98.33	98.33	98.39	98.36	95.00	95.00	95.45	95.22

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	95.00	95.00	95.05	95.02
TF/IDF	88.33	88.33	88.38	88.35
Term Frequency	86.67	86.67	86.67	86.67

Punctuation			
Accuracy	Recall	Precision	F-Score
90.00	90.00	90.18	90.09
88.33	88.33	88.38	88.35
88.33	88.33	88.38	88.35

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	82.15	11.63	95.02	63.93
Z Score-150 Words	91.44	5.95	98.36	82.76
Z Score-100 Words	92.17	3.87	98.36	88.05
Z Score-50 Words	88.96	7.29	98.36	80.00
TF/IDF	95.04	4.91	100.00	88.35
TF/IDF-150 Words	97.61	2.39	100.00	95.22
TF/IDF-100 Words	97.61	2.39	100.00	95.22
TF/IDF-50 Words	97.24	2.81	100.00	93.72
Term Frequency	94.77	5.39	100.00	86.67
Term Frequency-150 Words	97.61	2.39	100.00	95.22
Term Frequency-100 Words	98.00	2.07	100.00	95.22
Term Frequency-50 Words	96.83	2.48	100.00	93.72

Table 20: KNN POS Results for Entire Corpus(1500 Sentences) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	52.50	52.50	75.64	61.98	100.00	100.00	100.00	100.00
Z Score-150 Words	72.50	72.50	82.26	77.07	100.00	100.00	100.00	100.00
Z Score-100 Words	72.50	72.50	82.26	77.07	100.00	100.00	100.00	100.00
Z Score-50 Words	77.50	77.50	78.13	77.81	100.00	100.00	100.00	100.00
TF/IDF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-100 Words	97.50	97.50	97.62	97.56	100.00	100.00	100.00	100.00
TF/IDF-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	97.50	97.50	97.62	97.56	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	70.00	70.00	76.67	73.18	67.50	67.50	80.30	73.35
Z Score -150 Words	95.00	95.00	95.00	95.00	92.50	92.50	92.61	92.55
Z Score-100 Words	97.50	97.50	97.62	97.56	97.50	97.50	97.62	97.56
Z Score-50 Words	90.00	90.00	90.00	90.00	97.50	97.50	97.62	97.56
TF/IDF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	95.00	95.00	95.00	95.00	100.00	100.00	100.00	100.00

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	92.50	92.50	93.48	92.99
TF/IDF	85.00	85.00	85.00	85.00
Term Frequency	82.50	82.50	82.58	82.54

Punctuation			
Accuracy	Recall	Precision	F-Score
95.00	95.00	95.00	95.00
92.50	92.50	92.61	92.55
92.50	92.50	92.61	92.55

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	82.75	13.93	100.00	61.98
Z Score-150 Words	91.16	8.56	100.00	77.07
Z Score-100 Words	93.05	9.28	100.00	77.07
Z Score-50 Words	91.34	8.64	100.00	77.81
TF/IDF	96.26	5.72	100.00	85.00
TF/IDF-150 Words	100.00	0.00	100.00	100.00
TF/IDF-100 Words	99.39	1.06	100.00	97.56
TF/IDF-50 Words	100.00	0.00	100.00	100.00
Term Frequency	95.85	6.54	100.00	82.54
Term Frequency-150 Words	100.00	0.00	100.00	100.00
Term Frequency-100 Words	100.00	0.00	100.00	100.00
Term Frequency-50 Words	98.14	2.07	100.00	95.00

Table 21: KNN POS Results for Foundation Corpus [1980's Era] (500 Sentence) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	50.00	50.00	25.00	33.33	50.63	50.00	25.32	33.62
Z Score-150 Words	87.34	87.31	87.45	87.38	100.00	100.00	100.00	100.00
Z Score-100 Words	84.81	84.65	87.28	85.94	100.00	100.00	100.00	100.00
Z Score-50 Words	88.61	88.53	89.28	88.90	98.73	98.75	98.75	98.75
TF/IDF	93.83	93.67	93.96	93.81	100.00	100.00	100.00	100.00
TF/IDF-150 Words	93.67	93.59	94.44	94.01	100.00	100.00	100.00	100.00
TF/IDF-100 Words	93.67	93.62	93.96	93.79	100.00	100.00	100.00	100.00
TF/IDF-50 Words	92.41	92.34	92.89	92.61	100.00	100.00	100.00	100.00
Term Frequency	88.61	88.46	90.82	89.62	100.00	100.00	100.00	100.00
Term Frequency-150 Words	97.47	97.44	97.62	97.53	100.00	100.00	100.00	100.00
Term Frequency-100 Words	94.94	94.90	95.08	94.99	100.00	100.00	100.00	100.00
Term Frequency-50 Words	96.20	96.19	96.25	96.22	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	50.00	50.00	25.00	33.33	50.00	50.00	25.00	33.33
Z Score -150 Words	77.22	76.96	82.65	79.70	89.87	89.75	91.67	90.70
Z Score-100 Words	82.28	82.12	84.55	83.32	86.08	86.09	86.09	86.09
Z Score-50 Words	70.89	71.03	72.04	71.53	91.14	91.09	91.41	91.25
TF/IDF	98.73	98.75	98.75	98.75	98.73	98.75	98.75	98.75
TF/IDF-150 Words	98.73	98.75	98.75	98.75	98.73	98.75	98.75	98.75
TF/IDF-100 Words	98.73	98.75	98.75	98.75	98.73	98.75	98.75	98.75
TF/IDF-50 Words	98.73	98.75	98.75	98.75	98.73	98.75	98.75	98.75
Term Frequency	100.00	100.00	100.00	100.00	98.73	98.75	98.75	98.75
Term Frequency-150 Words	100.00	100.00	100.00	100.00	97.47	97.50	97.56	97.53
Term Frequency-100 Words	100.00	100.00	100.00	100.00	97.47	97.47	97.47	97.47
Term Frequency-50 Words	100.00	100.00	100.00	100.00	97.47	97.50	97.56	97.53

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	86.08	78.85	81.85	80.32
TF/IDF	75.95	64.76	71.73	68.07
Term Frequency	75.95	68.65	68.83	68.74

Punctuation			
Accuracy	Recall	Precision	F-Score
100.00	100.00	100.00	100.00
94.94	94.90	95.08	94.99
96.20	96.19	96.25	96.22

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	52.32	27.35	100.00	33.33
Z Score-150 Words	89.45	7.28	100.00	79.70
Z Score-100 Words	88.84	6.54	100.00	83.32
Z Score-50 Words	87.61	9.97	98.75	71.53
TF/IDF	92.40	11.10	100.00	68.07
TF/IDF-150 Words	97.88	2.29	100.00	94.01
TF/IDF-100 Words	97.82	2.38	100.00	93.79
TF/IDF-50 Words	97.53	2.88	100.00	92.61
Term Frequency	92.22	11.09	100.00	68.74
Term Frequency-150 Words	98.76	1.24	100.00	97.53
Term Frequency-100 Words	98.11	2.08	100.00	94.99
Term Frequency-50 Words	98.44	1.63	100.00	96.22

Table 22: KNN POS Results Foundation Corpus [1980's Era] (1000 Sentence) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	50.00	50.00	25.00	33.33	50.00	50.00	25.00	33.33
Z Score-150 Words	94.44	94.44	95.00	94.72	100.00	100.00	100.00	100.00
Z Score-100 Words	94.44	94.44	94.44	94.44	100.00	100.00	100.00	100.00
Z Score-50 Words	91.67	91.67	92.86	92.26	100.00	100.00	100.00	100.00
TF/IDF	97.22	97.22	97.37	97.29	100.00	100.00	100.00	100.00
TF/IDF-150 Words	97.22	97.22	97.37	97.29	100.00	100.00	100.00	100.00
TF/IDF-100 Words	97.22	97.22	97.37	97.29	100.00	100.00	100.00	100.00
TF/IDF-50 Words	91.67	91.67	91.80	91.73	100.00	100.00	100.00	100.00
Term Frequency	94.44	94.44	95.00	94.72	100.00	100.00	100.00	100.00
Term Frequency-150 Words	94.44	94.44	95.00	94.72	100.00	100.00	100.00	100.00
Term Frequency-100 Words	97.22	97.22	97.37	97.29	100.00	100.00	100.00	100.00
Term Frequency-50 Words	94.44	94.44	95.00	94.72	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	50.00	50.00	25.00	33.33	50.00	50.00	25.00	33.33
Z Score -150 Words	94.44	94.44	95.00	94.72	100.00	100.00	100.00	100.00
Z Score-100 Words	97.22	97.22	97.37	97.29	100.00	100.00	100.00	100.00
Z Score-50 Words	88.89	88.89	89.38	89.13	100.00	100.00	100.00	100.00
TF/IDF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00
TF/IDF	91.67	91.67	91.67	91.67
Term Frequency	91.67	91.67	91.67	91.67

Punctuation			
Accuracy	Recall	Precision	F-Score
100.00	100.00	100.00	100.00
100.00	100.00	100	100.00
97.22	97.22	97.37	97.29

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	55.56	31.43	100.00	33.33
Z Score-150 Words	97.36	2.64	100.00	94.72
Z Score-100 Words	97.93	2.30	100.00	94.44
Z Score-50 Words	95.35	4.78	100.00	89.13
TF/IDF	98.17	3.04	100.00	91.73
TF/IDF-150 Words	99.32	1.17	100.00	97.29
TF/IDF-100 Words	99.32	1.17	100.00	97.29
TF/IDF-50 Words	97.93	3.58	100.00	91.73
Term Frequency	97.29	3.15	100.00	91.73
Term Frequency-150 Words	98.68	2.29	100.00	94.72
Term Frequency-100 Words	99.32	1.17	100.00	97.29
Term Frequency-50 Words	98.68	2.29	100.00	94.72

Table 23: KNN POS Results Foundation Corpus [1980's Era] (1500 Sentence) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	50.00	50.00	25.00	33.33	50.00	50.00	25.00	33.33
Z Score-150 Words	91.67	91.67	92.86	92.26	100.00	100.00	100.00	100.00
Z Score-100 Words	91.67	91.67	92.86	92.26	100.00	100.00	100.00	100.00
Z Score-50 Words	87.50	87.50	90.00	88.73	100.00	100.00	100.00	100.00
TF/IDF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency	95.83	95.83	96.15	95.99	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	50.00	50.00	25.00	33.33	50.00	50.00	25.00	33.33
Z Score -150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Z Score-100 Words	95.83	95.83	96.15	95.99	100.00	100.00	100.00	100.00
Z Score-50 Words	83.33	83.33	83.33	83.33	100.00	100.00	100.00	100.00
TF/IDF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00
TF/IDF	91.67	91.67	91.67	91.67
Term Frequency	91.67	91.67	91.67	91.67

Punctuation			
Accuracy	Recall	Precision	F-Score
100.00	100.00	100.00	100.00
95.83	95.83	96.15	95.99
95.83	95.83	96.15	95.99

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	55.56	31.43	100.00	33.33
Z Score-150 Words	98.07	3.35	100.00	92.26
Z Score-100 Words	97.06	3.22	100.00	92.26
Z Score-50 Words	93.02	7.24	100.00	83.33
TF/IDF	97.94	3.16	100.00	91.67
TF/IDF-150 Words	100.00	0.00	100.00	100.00
TF/IDF-100 Words	100.00	0.00	100.00	100.00
TF/IDF-50 Words	100.00	0.00	100.00	100.00
Term Frequency	97.27	3.08	100.00	91.67
Term Frequency-150 Words	100.00	0.00	100.00	100.00
Term Frequency-100 Words	100.00	0.00	100.00	100.00
Term Frequency-50 Words	100.00	0.00	100.00	100.00

Table 24: KNN POS Results Foundation Corpus [1950's Era] (500 Sentence) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	50.00	50.00	25.00	33.33	50.00	50.00	25.00	33.33
Z Score-150 Words	77.08	77.08	81.52	79.24	100.00	100.00	100.00	100.00
Z Score-100 Words	87.50	87.50	88.57	88.03	100.00	100.00	100.00	100.00
Z Score-50 Words	85.42	85.42	87.02	86.21	100.00	100.00	100.00	100.00
TF/IDF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	95.83	95.83	96.15	95.99	100.00	100.00	100.00	100.00
TF/IDF-100 Words	95.83	95.83	96.15	95.99	100.00	100.00	100.00	100.00
TF/IDF-50 Words	93.75	93.75	93.83	93.79	100.00	100.00	100.00	100.00
Term Frequency	93.75	93.75	93.83	93.79	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	91.92	97.92	98.00	97.96	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	54.17	54.17	76.09	63.29	50.00	50.00	25.00	33.33
Z Score -150 Words	95.83	95.83	96.15	95.99	85.42	85.42	88.71	87.03
Z Score-100 Words	81.25	81.25	81.30	81.27	89.58	89.58	90.21	89.89
Z Score-50 Words	72.92	72.92	82.43	77.38	89.58	89.58	89.65	89.61
TF/IDF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	97.92	97.92	98.00	97.96	100.00	100.00	100.00	100.00
TF/IDF-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency	97.92	97.92	98.00	97.96	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	95.83	95.83	95.83	95.83
TF/IDF	95.83	95.83	95.83	95.83
Term Frequency	95.83	95.83	95.83	95.83

Punctuation			
Accuracy	Recall	Precision	F-Score
97.92	97.92	98.00	97.96
95.83	95.83	95.83	95.83
97.92	97.92	98.00	97.96

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	59.51	28.48	97.96	33.33
Z Score-150 Words	90.57	8.05	100.00	79.24
Z Score-100 Words	89.80	6.71	100.00	81.27
Z Score-50 Words	88.30	8.10	100.00	77.38
TF/IDF	98.61	1.97	100.00	95.83
TF/IDF-150 Words	98.49	1.67	100.00	95.99
TF/IDF-100 Words	99.00	1.74	100.00	95.99
TF/IDF-50 Words	98.45	2.69	100.00	93.79
Term Frequency	97.59	2.21	100.00	93.79
Term Frequency-150 Words	100.00	0.00	100.00	100.00
Term Frequency-100 Words	100.00	0.00	100.00	100.00
Term Frequency-50 Words	99.49	0.88	100.00	97.96

Table 25: KNN POS Results Foundation Corpus [1950's Era] (1000 Sentence) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	50.00	50.00	25.00	33.33	54.17	54.17	76.09	63.29
Z Score-150 Words	91.67	91.67	92.86	92.26	100.00	100.00	100.00	100.00
Z Score-100 Words	95.83	95.83	96.15	95.99	100.00	100.00	100.00	100.00
Z Score-50 Words	95.83	95.83	96.15	95.99	95.83	95.83	96.15	95.99
TF/IDF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	62.50	62.50	78.57	69.62	50.00	50.00	25.00	33.33
Z Score -150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Z Score-100 Words	91.67	91.67	92.86	92.26	95.83	95.83	96.15	95.99
Z Score-50 Words	95.83	95.83	96.15	95.99	87.50	87.50	87.76	87.63
TF/IDF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00
TF/IDF	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00

Punctuation			
Accuracy	Recall	Precision	F-Score
100.00	100.00	100.00	100.00
91.67	91.67	91.67	91.67
95.83	95.83	96.15	95.99

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	66.60	27.28	100.00	33.33
Z Score-150 Words	98.07	3.35	100.00	92.26
Z Score-100 Words	96.06	2.74	100.00	92.26
Z Score-50 Words	93.90	3.62	95.99	87.63
TF/IDF	98.61	3.10	100.00	91.67
TF/IDF-150 Words	100.00	0.00	100.00	100.00
TF/IDF-100 Words	100.00	0.00	100.00	100.00
TF/IDF-50 Words	100.00	0.00	100.00	100.00
Term Frequency	99.33	1.49	100.00	95.99
Term Frequency-150 Words	100.00	0.00	100.00	100.00
Term Frequency-100 Words	100.00	0.00	100.00	100.00
Term Frequency-50 Words	100.00	0.00	100.00	100.00

Table 26: KNN POS Results Foundation Corpus [1950's Era] (1500 Sentence) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	50.00	50.00	25.00	33.33	75.00	75.00	83.33	78.95
Z Score-150 Words	66.67	66.67	80.00	72.73	100.00	100.00	100.00	100.00
Z Score-100 Words	91.67	91.67	92.86	92.26	100.00	100.00	100.00	100.00
Z Score-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-50 Words	91.67	91.67	92.86	92.26	100.00	100.00	100.00	100.00
Term Frequency	91.67	91.67	92.86	92.26	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	91.67	91.67	92.86	92.26	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	75.00	75.00	83.33	78.95	58.33	58.33	77.27	66.48
Z Score -150 Words	91.67	91.67	92.86	92.26	91.67	91.67	92.86	92.26
Z Score-100 Words	83.33	83.33	83.33	83.33	83.33	83.33	87.50	85.36
Z Score-50 Words	91.67	91.67	92.86	92.26	91.67	91.67	92.86	92.26
TF/IDF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	91.67	91.67	92.86	92.26	100.00	100.00	100.00	100.00

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00
TF/IDF	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00

Punctuation			
Accuracy	Recall	Precision	F-Score
91.67	91.67	92.86	92.26
83.33	83.33	83.33	83.33
75.00	75.00	75.71	75.35

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	74.99	21.46	100.00	33.33
Z Score-150 Words	89.31	10.08	100.00	72.73
Z Score-100 Words	90.24	6.54	100.00	83.33
Z Score-50 Words	96.13	3.87	100.00	92.26
TF/IDF	97.22	6.21	100.00	83.33
TF/IDF-150 Words	100.00	0.00	100.00	100.00
TF/IDF-100 Words	100.00	0.00	100.00	100.00
TF/IDF-50 Words	98.07	3.35	100.00	92.26
Term Frequency	94.60	9.06	100.00	75.35
Term Frequency-150 Words	100.00	0.00	100.00	100.00
Term Frequency-100 Words	100.00	0.00	100.00	100.00
Term Frequency-50 Words	96.13	3.87	100.00	92.26

Table 27: KNN POS Results Galactic Empire Corpus (500 Sentence) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	64.29	64.29	79.17	70.96	100.00	100.00	100.00	100.00
Z Score-150 Words	92.86	92.86	92.95	92.90	100.00	100.00	100.00	100.00
Z Score-100 Words	90.48	90.48	90.85	90.66	100.00	100.00	100.00	100.00
Z Score-50 Words	88.10	88.10	88.89	88.49	100.00	100.00	100.00	100.00
TF/IDF	95.24	95.24	95.65	95.44	100.00	100.00	100.00	100.00
TF/IDF-150 Words	95.24	95.24	95.65	95.44	100.00	100.00	100.00	100.00
TF/IDF-100 Words	95.24	95.24	95.65	95.44	100.00	100.00	100.00	100.00
TF/IDF-50 Words	95.24	95.24	95.65	95.44	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-150 Words	97.62	97.62	97.73	97.67	100.00	100.00	100.00	100.00
Term Frequency-100 Words	95.24	95.24	95.65	95.44	100.00	100.00	100.00	100.00
Term Frequency-50 Words	97.62	97.62	97.73	97.67	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	76.19	76.19	83.87	79.85	100.00	100.00	100.00	100.00
Z Score -150 Words	95.24	95.24	95.65	95.44	90.48	90.48	92.00	91.23
Z Score-100 Words	97.62	97.62	97.73	97.67	92.86	92.86	93.75	93.30
Z Score-50 Words	95.24	95.24	95.65	95.44	95.24	95.24	95.65	95.44
TF/IDF	100.00	100.00	100.00	100.00	83.33	83.33	85.34	84.32
TF/IDF-150 Words	100.00	100.00	100.00	100.00	83.33	83.33	84.03	83.68
TF/IDF-100 Words	100.00	100.00	100.00	100.00	80.95	80.95	80.95	80.95
TF/IDF-50 Words	100.00	100.00	100.00	100.00	83.33	83.33	84.03	83.68
Term Frequency	100.00	100.00	100.00	100.00	88.10	88.10	88.18	88.14
Term Frequency-150 Words	100.00	100.00	100.00	100.00	85.74	85.71	86.04	85.87
Term Frequency-100 Words	100.00	100.00	100.00	100.00	80.95	80.95	81.24	81.09
Term Frequency-50 Words	100.00	100.00	100.00	100.00	80.95	80.95	80.95	80.95

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	90.48	90.48	90.85	90.66
TF/IDF	83.33	83.33	83.41	83.37
Term Frequency	80.95	80.95	80.95	80.95

Punctuation			
Accuracy	Recall	Precision	F-Score
97.62	97.62	97.73	97.67
95.24	95.24	95.65	95.44
95.24	95.24	95.65	95.44

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	89.86	10.99	100.00	70.96
Z Score-150 Words	94.90	3.31	100.00	91.23
Z Score-100 Words	95.41	3.65	100.00	90.66
Z Score-50 Words	94.85	4.11	100.00	88.49
TF/IDF	93.10	6.81	100.00	83.37
TF/IDF-150 Words	94.78	6.67	100.00	83.68
TF/IDF-100 Words	94.10	7.82	100.00	80.95
TF/IDF-50 Words	94.78	6.67	100.00	83.68
Term Frequency	94.09	7.24	100.00	80.95
Term Frequency-150 Words	95.89	5.86	100.00	85.87
Term Frequency-100 Words	94.13	7.75	100.00	81.09
Term Frequency-50 Words	94.66	7.97	100.00	80.95

Table 28: KNN POS Results Galactic Empire Corpus (1000 Sentence) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	88.89	88.89	90.91	89.89	100.00	100.00	100.00	100.00
Z Score-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Z Score-100 Words	94.44	94.44	95.00	94.72	100.00	100.00	100.00	100.00
Z Score-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	66.67	66.67	80.00	72.73	100.00	100.00	100.00	100.00
Z Score -150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Z Score-100 Words	100.00	100.00	100.00	100.00	94.44	94.44	95.00	94.72
Z Score-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF	100.00	100.00	100.00	100.00	94.44	94.44	95.00	94.72
TF/IDF-150 Words	100.00	100.00	100.00	100.00	83.33	83.33	83.75	83.54
TF/IDF-100 Words	100.00	100.00	100.00	100.00	83.33	83.33	83.75	83.54
TF/IDF-50 Words	100.00	100.00	100.00	100.00	77.78	77.78	79.22	78.49
Term Frequency	100.00	100.00	100.00	100.00	94.44	94.44	95.00	94.72
Term Frequency-150 Words	100.00	100.00	100.00	100.00	88.89	88.89	88.89	88.89
Term Frequency-100 Words	100.00	100.00	100.00	100.00	88.89	88.89	90.91	89.89
Term Frequency-50 Words	100.00	100.00	100.00	100.00	83.33	83.33	83.75	83.54

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00
TF/IDF	83.33	83.33	87.50	85.36
Term Frequency	77.78	77.78	79.22	78.49

Punctuation			
Accuracy	Recall	Precision	F-Score
88.89	88.89	90.91	89.89
94.44	94.44	95.00	94.72
94.44	94.44	95.00	94.72

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	92.08	9.77	100.00	72.73
Z Score-150 Words	100.00	0.00	100.00	100.00
Z Score-100 Words	97.36	2.64	100.00	94.72
Z Score-50 Words	100.00	0.00	100.00	100.00
TF/IDF	95.80	5.23	100.00	85.36
TF/IDF-150 Words	95.88	7.13	100.00	83.54
TF/IDF-100 Words	95.88	7.13	100.00	83.54
TF/IDF-50 Words	94.62	9.31	100.00	78.49
Term Frequency	94.66	7.60	100.00	78.49
Term Frequency-150 Words	97.22	4.81	100.00	88.89
Term Frequency-100 Words	97.47	4.38	100.00	89.89
Term Frequency-50 Words	95.88	7.13	100.00	83.54

Table 29: KNN POS Results Galactic Empire Corpus (1500 Sentence) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	83.33	83.33	87.50	85.36	100.00	100.00	100.00	100.00
Z Score-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Z Score-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Z Score-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	58.33	58.33	77.27	66.48	100.00	100.00	100.00	100.00
Z Score -150 Words	100.00	100.00	100.00	100.00	83.33	83.33	87.50	85.36
Z Score-100 Words	91.67	91.67	92.86	92.26	83.33	83.33	87.50	85.36
Z Score-50 Words	100.00	100.00	100.00	100.00	58.33	58.33	77.27	66.48
TF/IDF	100.00	100.00	100.00	100.00	75.00	75.00	83.33	78.95
TF/IDF-150 Words	100.00	100.00	100.00	100.00	83.33	83.33	87.50	85.36
TF/IDF-100 Words	100.00	100.00	100.00	100.00	83.33	83.33	87.50	85.36
TF/IDF-50 Words	100.00	100.00	100.00	100.00	83.33	83.33	87.50	85.36
Term Frequency	100.00	100.00	100.00	100.00	83.33	83.33	83.33	83.33
Term Frequency-150 Words	100.00	100.00	100.00	100.00	83.33	83.33	87.50	85.36
Term Frequency-100 Words	100.00	100.00	100.00	100.00	83.33	83.33	83.33	83.33
Term Frequency-50 Words	100.00	100.00	100.00	100.00	75.00	75.00	75.71	75.35

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00
TF/IDF	91.67	91.67	92.86	92.26
Term Frequency	75.00	75.00	75.71	75.35

Punctuation			
Accuracy	Recall	Precision	F-Score
83.33	83.33	87.50	85.36
91.67	91.67	92.86	92.26
91.67	91.67	92.86	92.26

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	89.53	12.21	100.00	66.48
Z Score-150 Words	96.34	6.34	100.00	85.36
Z Score-100 Words	94.41	6.10	100.00	85.36
Z Score-50 Words	91.62	14.52	100.00	66.48
TF/IDF	93.91	7.53	100.00	78.95
TF/IDF-150 Words	96.34	6.34	100.00	85.36
TF/IDF-100 Words	96.34	6.34	100.00	85.36
TF/IDF-50 Words	96.34	6.34	100.00	85.36
Term Frequency	91.82	9.52	100.00	75.35
Term Frequency-150 Words	96.34	6.34	100.00	85.36
Term Frequency-100 Words	95.83	7.22	100.00	83.33
Term Frequency-50 Words	93.84	10.67	100.00	75.35

Table 30: KNN POS Results Robot Corpus (500 Sentence) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	50.00	50.00	25.00	33.33	50.00	50.00	25.00	33.33
Z Score-150 Words	96.43	96.43	96.43	96.43	100.00	100.00	100.00	100.00
Z Score-100 Words	94.64	94.64	94.70	94.67	100.00	100.00	100.00	100.00
Z Score-50 Words	92.86	92.86	93.08	92.97	100.00	100.00	100.00	100.00
TF/IDF	98.21	98.21	98.28	98.24	100.00	100.00	100.00	100.00
TF/IDF-150 Words	98.21	98.21	98.28	98.24	100.00	100.00	100.00	100.00
TF/IDF-100 Words	98.21	98.21	98.28	98.24	100.00	100.00	100.00	100.00
TF/IDF-50 Words	94.64	94.64	94.70	94.67	98.21	98.21	98.28	98.24
Term Frequency	98.21	98.21	98.28	98.24	100.00	100.00	100.00	100.00
Term Frequency-150 Words	96.43	96.43	96.67	96.55	100.00	100.00	100.00	100.00
Term Frequency-100 Words	96.43	96.43	96.67	96.55	100.00	100.00	100.00	100.00
Term Frequency-50 Words	94.64	94.64	95.16	94.90	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	50.00	50.00	25.00	33.33	94.64	94.64	95.16	94.90
Z Score -150 Words	85.71	85.71	88.89	87.27	87.50	87.50	90.00	88.73
Z Score-100 Words	96.43	96.43	96.67	96.55	80.36	80.36	85.90	83.04
Z Score-50 Words	96.43	96.43	96.43	96.43	94.64	94.64	95.16	94.90
TF/IDF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00
TF/IDF	92.86	92.86	92.86	92.86
Term Frequency	92.86	92.86	92.86	92.86

Punctuation			
Accuracy	Recall	Precision	F-Score
98.21	98.21	98.28	98.24
98.21	98.21	98.28	98.24
98.21	98.21	98.28	98.24

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	65.52	32.23	100.00	33.33
Z Score-150 Words	93.11	5.29	100.00	87.27
Z Score-100 Words	93.56	6.37	100.00	83.04
Z Score-50 Words	96.07	2.58	100.00	92.97
TF/IDF	98.22	2.52	100.00	92.86
TF/IDF-150 Words	99.56	0.76	100.00	98.24
TF/IDF-100 Words	99.56	0.76	100.00	98.24
TF/IDF-50 Words	98.23	2.18	100.00	94.67
Term Frequency	98.22	2.52	100.00	92.86
Term Frequency-150 Words	99.14	1.49	100.00	96.55
Term Frequency-100 Words	99.14	1.49	100.00	96.55
Term Frequency-50 Words	98.72	2.21	100.00	94.90

Table 31: KNN POS Results Robot Corpus (1000 Sentence) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	57.14	50.00	28.57	36.36	57.14	50.00	28.57	36.36
Z Score-150 Words	95.24	94.44	96.15	95.29	100.00	100.00	100.00	100.00
Z Score-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Z Score-50 Words	95.24	95.83	95.00	95.41	100.00	100.00	100.00	100.00
TF/IDF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	57.14	50.00	28.57	36.36	57.14	50.00	28.57	36.36
Z Score -150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Z Score-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Z Score-50 Words	100.00	100.00	100.00	100.00	95.24	95.83	95.00	95.41
TF/IDF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00
TF/IDF	80.95	80.56	80.56	80.56
Term Frequency	85.71	84.72	86.06	85.38

Punctuation			
Accuracy	Recall	Precision	F-Score
95.24	95.83	95.00	95.41
85.71	83.33	90.00	86.54
85.71	83.33	90.00	86.54

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	56.81	28.95	100.00	36.36
Z Score-150 Words	98.82	2.04	100.00	95.29
Z Score-100 Words	100.00	0.00	100.00	100.00
Z Score-50 Words	97.71	2.29	100.00	95.41
TF/IDF	94.52	7.94	100.00	80.56
TF/IDF-150 Words	100.00	0.00	100.00	100.00
TF/IDF-100 Words	100.00	0.00	100.00	100.00
TF/IDF-50 Words	100.00	0.00	100.00	100.00
Term Frequency	95.32	6.63	100.00	85.38
Term Frequency-150 Words	100.00	0.00	100.00	100.00
Term Frequency-100 Words	100.00	0.00	100.00	100.00
Term Frequency-50 Words	100.00	0.00	100.00	100.00

Table 32: KNN POS Results Robot Corpus (1500 Sentence) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	53.33	50.00	26.67	34.79	53.33	50.00	26.67	34.79
Z Score-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Z Score-100 Words	93.33	92.86	94.44	93.64	100.00	100.00	100.00	100.00
Z Score-50 Words	86.67	85.70	90.00	87.80	100.00	100.00	100.00	100.00
TF/IDF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	93.33	93.75	93.75	93.75	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	53.33	50.00	26.67	34.79	93.33	92.86	94.44	93.64
Z Score -150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Z Score-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Z Score-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00
TF/IDF	80.95	80.56	80.56	80.56
Term Frequency	85.71	84.72	86.06	85.38

Punctuation			
Accuracy	Recall	Precision	F-Score
100.00	100.00	100.00	100.00
93.33	92.86	94.44	93.64
100.00	100.00	100.00	100.00

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	66.33	31.62	100.00	34.79
Z Score-150 Words	100.00	0.00	100.00	100.00
Z Score-100 Words	98.41	2.75	100.00	93.64
Z Score-50 Words	96.95	5.28	100.00	87.80
TF/IDF	96.71	5.08	100.00	86.61
TF/IDF-150 Words	100.00	0.00	100.00	100.00
TF/IDF-100 Words	100.00	0.00	100.00	100.00
TF/IDF-50 Words	100.00	0.00	100.00	100.00
Term Frequency	97.77	4.99	100.00	86.61
Term Frequency-150 Words	100.00	0.00	100.00	100.00
Term Frequency-100 Words	100.00	0.00	100.00	100.00
Term Frequency-50 Words	98.44	2.71	100.00	93.75

Table 33: Part-Of-Speech Statistics for KNN - Two Authors

	Adjectives				All Words			
	Average F-Score	Standard Deviation F-Score	Max F-Score	Min F-Score	Average F-Score	Stand. Dev. F-Score	Max F-Score	Min F-Score
Z Score	48.55	50.00	89.89	33.33	63.25	29.08	100.00	33.33
Z Score-150 Words	89.73	100.00	100.00	72.73	99.89	0.41	100.00	98.36
Z Score-100 Words	91.17	92.86	100.00	77.07	99.89	0.41	100.00	98.36
Z Score-50 Words	89.98	85.70	100.00	71.12	99.48	1.07	100.00	95.99
TF/IDF	97.95	100.00	100.00	87.71	100.00	0.00	100.00	100.00
TF/IDF-150 Words	97.50	100.00	100.00	86.32	100.00	0.00	100.00	100.00
TF/IDF-100 Words	97.24	100.00	100.00	85.00	100.00	0.00	100.00	100.00
TF/IDF-50 Words	95.90	100.00	100.00	84.28	99.83	0.47	100.00	98.24
Term Frequency	96.71	100.00	100.00	89.62	100.00	0.00	100.00	100.00
Term Frequency-150 Words	98.31	100.00	100.00	92.99	100.00	0.00	100.00	100.00
Term Frequency-100 Words	98.13	100.00	100.00	90.83	100.00	0.00	100.00	100.00
Term Frequency-50 Words	96.35	93.75	100.00	86.45	99.94	0.21	100.00	99.17
	Nouns				Verbs			
	Average F-Score	Stan. Dev. F-Score	Max F-Score	Min F-Score	Average F-Score	Stan. Dev. F-Score	Max F-Score	Min F-Score
Z Score	58.25	20.45	86.24	33.33	64.71	27.50	100.00	33.33
Z Score -150 Words	93.94	6.67	100.00	79.70	93.66	5.99	100.00	81.83
Z Score-100 Words	93.00	6.29	100.00	81.27	92.52	6.19	100.00	83.04
Z Score-50 Words	90.44	8.86	100.00	71.53	92.33	8.63	100.00	66.48
TF/IDF	98.94	3.65	100.00	85.31	96.68	6.19	100.00	78.95
TF/IDF-150 Words	99.73	0.59	100.00	97.96	96.07	6.19	100.00	83.54
TF/IDF-100 Words	99.86	0.36	100.00	98.75	95.92	6.57	100.00	80.95
TF/IDF-50 Words	99.75	0.66	100.00	97.56	95.72	6.95	100.00	78.49
Term Frequency	99.86	0.51	100.00	97.96	97.39	4.86	100.00	83.33
Term Frequency-150 Words	99.94	0.21	100.00	99.17	96.48	5.24	100.00	85.36
Term Frequency-100 Words	100.00	0.00	100.00	100.00	96.13	6.15	100.00	81.09

Term Frequency-50 Words	98.88	2.23	100.00	92.26	95.13	7.94	100.00	75.35
-------------------------	-------	------	--------	-------	-------	------	--------	-------

	Articles			
	Average F-Score	Stand. Dev. F-Score	Max F-Score	Min F-Score
Z Score	96.46	5.42	100.00	80.32
TF /IDF	88.47	7.81	100	68.07
Term Frequency	86.75	8.57	100	68.74

	Punctuation			
	Average F-Score	Stand. Dev. F-Score	Max F-Score	Min F-Score
Z Score	95.57	4.53	100.00	85.36
TF /IDF	92.57	4.67	100.00	83.33
Term Frequency	92.86	6.23	100.00	75.35

D.2 KNN POS Results - Multiple Authors

Table 34: KNN POS Results for Entire Corpus(500 Sentences) - Multiple Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	50.00	16.67	8.33	11.11	50.83	18.06	10.63	13.38
Z Score-150 Words	55.00	27.22	31.70	29.29	90.83	91.39	86.65	88.96
Z Score-100 Words	59.17	39.72	54.56	45.97	94.17	90.28	95.30	92.72
Z Score-50 Words	52.50	20.83	36.40	26.50	95.83	93.06	94.37	93.71
TF/IDF	82.50	74.17	87.02	80.08	99.17	98.61	98.72	98.66
TF/IDF-150 Words	81.67	73.89	85.60	79.32	99.17	98.61	99.73	99.17
TF/IDF-100 Words	80.83	70.28	86.73	77.64	98.33	97.22	97.44	97.33
TF/IDF-50 Words	79.17	65.28	92.60	76.58	97.50	95.83	95.92	95.87
Term Frequency	85.83	76.39	90.08	82.67	100.00	100.00	100.00	100.00
Term Frequency-150 Words	90.83	84.72	94.92	89.53	100.00	100.00	100.00	100.00
Term Frequency-100 Words	88.33	80.56	94.17	86.83	98.33	97.22	97.44	97.33
Term Frequency-50 Words	82.50	70.83	92.52	80.23	97.50	95.83	97.06	96.44
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	52.50	21.94	13.15	16.44	59.17	31.94	25.84	28.57
Z Score -150 Words	59.17	34.17	67.34	45.34	60.00	33.33	65.17	44.10
Z Score-100 Words	62.50	41.94	66.13	51.33	60.00	37.78	59.89	46.33
Z Score-50 Words	64.17	42.50	71.04	53.18	56.67	28.89	44.68	35.09
TF/IDF	100.00	100.00	100.00	100.00	90.83	84.72	91.01	87.75
TF/IDF-150 Words	98.33	97.22	98.44	97.83	90.00	87.78	87.93	87.85
TF/IDF-100 Words	97.50	95.83	97.35	96.58	90.00	85.56	86.98	86.26
TF/IDF-50 Words	95.00	91.67	97.44	94.47	89.17	86.39	86.82	86.60
Term Frequency	100.00	100.00	100.00	100.00	93.33	92.22	92.60	92.41
Term Frequency-150 Words	99.17	99.72	98.72	99.22	90.83	90.28	87.92	89.08
Term Frequency-100 Words	100.00	100.00	100.00	100.00	90.83	88.06	88.07	88.06

Term Frequency-50 Words	95.83	94.17	95.11	94.64	88.33	80.56	89.52	84.80
-------------------------	-------	-------	-------	-------	-------	-------	-------	-------

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	80.83	68.06	83.81	75.12
TF/IDF	75.00	65.00	72.57	68.58
Term Frequency	76.67	66.67	77.98	71.88

	Punctuation			
	Accuracy	Recall	Precision	F-Score
Z Score	89.17	84.17	87.90	85.99
TF/IDF	80.00	76.67	87.72	81.82
Term Frequency	81.67	71.67	88.43	79.17

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	38.44	30.45	85.99	11.11
Z Score-150 Words	51.92	22.30	88.96	29.29
Z Score-100 Words	59.09	19.53	92.72	45.97
Z Score-50 Words	52.12	25.87	93.71	26.50
TF/IDF	86.15	10.92	100.00	68.58
TF/IDF-150 Words	91.04	8.06	99.17	79.32
TF/IDF-100 Words	89.46	8.10	97.33	77.64
TF/IDF-50 Words	88.38	7.68	95.87	76.58
Term Frequency	87.69	10.59	100.00	71.88
Term Frequency-150 Words	94.46	5.16	100.00	89.08
Term Frequency-100 Words	93.06	5.70	100.00	86.83
Term Frequency-50 Words	89.03	6.74	96.44	80.23

Table 35: KNN POS Results for Entire Corpus(1000 Sentences) - Multiple Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	50.00	16.67	8.33	11.11	53.33	22.22	17.95	19.86
Z Score-150 Words	90.00	83.33	94.84	88.71	96.67	96.67	95.24	95.95
Z Score-100 Words	88.33	80.56	93.89	86.72	98.33	97.22	97.62	97.42
Z Score-50 Words	78.33	66.11	87.84	75.44	95.00	91.67	93.91	92.78
TF/IDF	95.00	91.67	98.48	94.95	100.00	100.00	100.00	100.00
TF/IDF-150 Words	95.00	91.67	98.48	94.95	100.00	100.00	100.00	100.00
TF/IDF-100 Words	93.33	88.89	96.10	92.35	100.00	100.00	100.00	100.00
TF/IDF-50 Words	93.33	88.89	98.04	93.24	100.00	100.00	100.00	100.00
Term Frequency	96.67	94.44	98.96	96.65	100.00	100.00	100.00	100.00
Term Frequency-150 Words	98.33	97.22	99.46	98.33	100.00	100.00	100.00	100.00
Term Frequency-100 Words	95.00	91.67	96.58	94.06	100.00	100.00	100.00	100.00
Term Frequency-50 Words	93.33	88.89	98.04	93.24	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	53.33	26.67	15.51	19.61	65.00	41.67	43.14	42.39
Z Score -150 Words	83.33	74.44	93.27	82.80	90.00	87.78	92.20	89.94
Z Score-100 Words	83.33	74.44	87.22	80.32	91.67	86.11	93.72	89.75
Z Score-50 Words	75.00	60.56	91.24	72.80	85.00	75.00	96.15	84.27
TF/IDF	100.00	100.00	100.00	100.00	95.00	91.67	96.58	94.06
TF/IDF-150 Words	100.00	100.00	100.00	100.00	91.67	88.33	92.72	90.47
TF/IDF-100 Words	100.00	100.00	100.00	100.00	90.00	85.56	91.43	88.40
TF/IDF-50 Words	98.33	97.22	99.46	98.33	90.00	85.56	89.95	87.70
Term Frequency	100.00	100.00	100.00	100.00	95.00	91.67	96.58	94.06
Term Frequency-150 Words	100.00	100.00	100.00	100.00	93.33	88.89	96.10	92.35
Term Frequency-100 Words	98.33	97.22	99.46	98.33	91.67	88.33	92.72	90.47
Term Frequency-50 Words	98.33	97.22	99.46	98.33	88.33	82.78	88.65	85.61

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	93.33	88.89	96.10	92.35
TF/IDF	80.00	73.33	77.16	75.20
Term Frequency	78.33	70.56	75.70	73.04

Punctuation			
Accuracy	Recall	Precision	F-Score
90.00	87.78	90.85	89.29
86.67	86.67	85.36	86.01
85.00	83.89	83.19	83.54

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	45.77	33.25	92.35	11.11
Z Score-150 Words	89.35	4.67	95.95	82.80
Z Score-100 Words	88.55	6.15	97.42	80.32
Z Score-50 Words	81.32	7.86	92.78	72.80
TF/IDF	91.70	8.74	100.00	75.20
TF/IDF-150 Words	96.36	3.97	100.00	90.47
TF/IDF-100 Words	95.19	5.01	100.00	88.40
TF/IDF-50 Words	94.82	4.80	100.00	87.70
Term Frequency	91.21	9.84	100.00	73.04
Term Frequency-150 Words	97.67	3.14	100.00	92.35
Term Frequency-100 Words	95.71	3.72	100.00	90.47
Term Frequency-50 Words	94.30	5.60	100.00	85.61

Table 36: KNN POS Results for Entire Corpus(1500 Sentences) - Multiple Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	50.00	16.67	8.33	11.11	50.00	16.67	9.52	12.12
Z Score-150 Words	75.00	65.00	57.14	60.82	97.50	95.83	96.67	96.25
Z Score-100 Words	67.50	59.17	59.58	59.37	97.50	95.83	96.67	96.25
Z Score-50 Words	70.00	56.67	62.00	59.22	100.00	100.00	100.00	100.00
TF/IDF	100.00	100.00	100.00	100.00	97.50	95.83	96.67	96.25
TF/IDF-150 Words	100.00	100.00	100.00	100.00	97.50	95.83	96.67	96.25
TF/IDF-100 Words	100.00	100.00	100.00	100.00	97.50	95.83	96.67	96.25
TF/IDF-50 Words	100.00	100.00	100.00	100.00	97.50	95.83	96.67	96.25
Term Frequency	97.50	95.83	99.21	97.49	97.50	95.83	96.67	96.25
Term Frequency-150 Words	100.00	100.00	100.00	100.00	97.50	95.83	96.67	96.25
Term Frequency-100 Words	100.00	100.00	100.00	100.00	97.50	95.83	96.67	96.25
Term Frequency-50 Words	100.00	10.00	100.00	18.18	97.50	95.83	96.67	96.25
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	52.50	20.83	13.01	16.02	67.50	45.83	45.83	45.83
Z Score -150 Words	72.50	60.83	57.22	58.97	92.50	90.83	93.61	92.20
Z Score-100 Words	77.50	62.50	50.40	55.80	95.00	91.67	95.87	93.72
Z Score-50 Words	75.00	65.00	66.65	65.81	90.00	90.00	90.23	90.11
TF/IDF	100.00	100.00	100.00	100.00	95.00	91.67	91.67	91.67
TF/IDF-150 Words	100.00	100.00	100.00	100.00	95.00	91.67	91.67	91.67
TF/IDF-100 Words	100.00	100.00	100.00	100.00	95.00	91.67	91.67	91.67
TF/IDF-50 Words	100.00	100.00	100.00	100.00	95.00	91.67	91.67	91.67
Term Frequency	97.50	95.83	96.67	96.25	95.00	91.67	91.67	91.67
Term Frequency-150 Words	100.00	100.00	100.00	100.00	95.00	91.67	91.67	91.67
Term Frequency-100 Words	100.00	100.00	100.00	100.00	95.00	91.67	91.67	91.67
Term Frequency-50 Words	100.00	100.00	100.00	100.00	95.00	91.67	91.67	91.67

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	87.50	79.17	86.26	82.56
TF/IDF	67.50	59.17	57.09	58.11
Term Frequency	65.00	55.00	55.70	55.35

Punctuation			
Accuracy	Recall	Precision	F-Score
92.50	90.83	91.11	90.97
87.50	85.83	87.90	86.85
87.50	85.83	87.90	86.85

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	43.10	33.10	90.97	11.11
Z Score-150 Words	77.06	17.24	96.25	58.97
Z Score-100 Words	76.29	18.76	96.25	55.80
Z Score-50 Words	78.79	16.81	100.00	59.22
TF/IDF	88.81	14.49	100.00	58.11
TF/IDF-150 Words	96.98	3.43	100.00	91.67
TF/IDF-100 Words	96.98	3.43	100.00	91.67
TF/IDF-50 Words	96.98	3.43	100.00	91.67
Term Frequency	87.31	14.74	97.49	55.35
Term Frequency-150 Words	96.98	3.43	100.00	91.67
Term Frequency-100 Words	96.98	3.43	100.00	91.67
Term Frequency-50 Words	76.52	33.81	100.00	18.18

Table 37: KNN POS Results for Foundation Corpus [1980's Era](500 Sentences) - Multiple Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	50.63	25.00	12.66	16.81	50.63	25.00	12.66	16.81
Z Score-150 Words	68.35	51.92	79.24	62.73	96.20	94.23	95.31	94.77
Z Score-100 Words	74.68	64.51	83.77	72.89	97.47	96.15	98.81	97.46
Z Score-50 Words	70.89	36.56	77.09	49.60	88.61	82.69	92.71	87.41
TF/IDF	88.61	85.29	92.54	88.77	100.00	100.00	100.00	100.00
TF/IDF-150 Words	89.87	84.62	95.83	89.88	98.73	98.08	98.21	98.14
TF/IDF-100 Words	88.61	82.69	95.41	88.60	98.73	98.08	98.21	98.14
TF/IDF-50 Words	87.34	80.77	95.00	87.31	100.00	100.00	100.00	100.00
Term Frequency	86.08	78.85	91.84	84.85	100.00	100.00	100.00	100.00
Term Frequency-150 Words	93.67	91.68	95.59	93.59	100.00	100.00	100.00	100.00
Term Frequency-100 Words	91.14	89.13	92.90	90.98	100.00	100.00	100.00	100.00
Term Frequency-50 Words	89.87	87.21	92.03	89.56	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	50.63	25.00	12.66	16.81	50.63	25.00	12.66	16.81
Z Score -150 Words	77.22	70.58	81.35	75.58	69.62	53.85	79.17	64.10
Z Score-100 Words	74.68	66.73	73.89	70.13	67.09	56.49	62.81	59.48
Z Score-50 Words	64.56	60.43	70.18	64.94	74.68	61.54	85.16	71.45
TF/IDF	98.73	99.38	98.21	98.79	96.20	96.83	94.64	95.72
TF/IDF-150 Words	98.73	99.38	98.21	98.79	92.41	91.06	91.84	91.45
TF/IDF-100 Words	97.47	97.45	96.67	97.06	93.67	92.98	92.80	92.89
TF/IDF-50 Words	96.20	95.53	94.64	95.08	91.14	89.13	89.06	89.09
Term Frequency	100.00	100.00	100.00	100.00	96.20	95.53	94.48	95.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	96.20	95.53	95.51	95.52
Term Frequency-100 Words	100.00	100.00	100.00	100.00	94.94	93.61	94.72	94.16
Term Frequency-50 Words	100.00	100.00	100.00	100.00	93.67	91.68	92.95	92.31

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	86.08	78.85	81.85	80.32
TF/IDF	75.95	64.76	71.73	68.07
Term Frequency	75.95	68.65	68.83	68.74

Punctuation			
Accuracy	Recall	Precision	F-Score
97.47	96.15	97.60	96.87
94.94	92.31	93.39	92.85
94.94	94.90	93.10	93.99

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	40.74	34.18	96.87	16.81
Z Score-150 Words	74.30	12.83	94.77	62.73
Z Score-100 Words	74.99	13.91	97.46	59.48
Z Score-50 Words	68.35	13.57	87.41	49.60
TF/IDF	90.70	10.78	100.00	68.07
TF/IDF-150 Words	94.57	3.95	98.79	89.88
TF/IDF-100 Words	94.17	3.77	98.14	88.60
TF/IDF-50 Words	92.87	5.02	100.00	87.31
Term Frequency	90.43	10.94	100.00	68.74
Term Frequency-150 Words	97.28	2.81	100.00	93.59
Term Frequency-100 Words	96.28	3.88	100.00	90.98
Term Frequency-50 Words	95.47	4.64	100.00	89.56

Table 38: KNN POS Results for Foundation Corpus [1980's Era](1000 Sentences) - Multiple Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	50.00	25.00	12.50	16.67	50.00	25.00	12.50	16.67
Z Score-150 Words	77.78	66.67	92.31	77.42	100.00	100.00	100.00	100.00
Z Score-100 Words	86.11	79.17	94.47	86.15	100.00	100.00	100.00	100.00
Z Score-50 Words	94.44	94.44	94.44	94.44	94.44	91.67	97.50	94.50
TF/IDF	97.22	95.83	98.68	97.23	100.00	100.00	100.00	100.00
TF/IDF-150 Words	97.22	95.83	98.68	97.23	100.00	100.00	100.00	100.00
TF/IDF-100 Words	97.22	95.83	98.68	97.23	100.00	100.00	100.00	100.00
TF/IDF-50 Words	94.44	94.44	95.04	94.74	100.00	100.00	100.00	100.00
Term Frequency	91.67	87.50	96.43	91.75	100.00	100.00	100.00	100.00
Term Frequency-150 Words	97.22	95.83	98.68	97.23	100.00	100.00	100.00	100.00
Term Frequency-100 Words	97.22	95.83	98.68	97.23	100.00	100.00	100.00	100.00
Term Frequency-50 Words	94.44	91.67	97.50	94.50	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	50.00	25.00	12.50	16.67	50.00	25.00	12.50	16.67
Z Score -150 Words	94.44	91.67	97.50	94.50	83.33	75.00	83.04	78.82
Z Score-100 Words	91.67	87.50	96.43	91.75	86.11	79.17	88.68	83.66
Z Score-50 Words	88.89	88.89	86.67	87.77	83.33	75.00	87.05	80.58
TF/IDF	100.00	100.00	100.00	100.00	97.22	95.83	96.43	96.13
TF/IDF-150 Words	100.00	100.00	100.00	100.00	94.44	91.67	95.11	93.36
TF/IDF-100 Words	100.00	100.00	100.00	100.00	94.44	91.67	95.11	93.36
TF/IDF-50 Words	100.00	100.00	100.00	100.00	91.67	87.50	90.95	89.19
Term Frequency	100.00	100.00	100.00	100.00	97.22	95.83	96.43	96.13
Term Frequency-150 Words	100.00	100.00	100.00	100.00	94.44	91.67	95.11	93.36
Term Frequency-100 Words	100.00	100.00	100.00	100.00	94.44	91.67	95.11	93.36
Term Frequency-50 Words	100.00	100.00	100.00	100.00	91.67	87.50	90.95	89.19

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	94.44	91.67	95.11	93.36
TF/IDF	77.78	72.22	72.34	72.28
Term Frequency	80.56	76.39	77.72	77.05

Punctuation			
Accuracy	Recall	Precision	F-Score
91.67	87.50	93.33	90.32
91.67	87.50	93.93	90.60
94.44	91.67	95.11	93.36

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	41.72	35.45	93.36	16.67
Z Score-150 Words	87.68	9.77	100.00	77.42
Z Score-100 Words	90.39	6.28	100.00	83.66
Z Score-50 Words	89.32	5.74	94.50	80.58
TF/IDF	92.71	9.66	100.00	72.28
TF/IDF-150 Words	97.65	2.72	100.00	93.36
TF/IDF-100 Words	97.65	2.72	100.00	93.36
TF/IDF-50 Words	95.98	4.47	100.00	89.19
Term Frequency	93.05	7.79	100.00	77.05
Term Frequency-150 Words	97.65	2.72	100.00	93.36
Term Frequency-100 Words	97.65	2.72	100.00	93.36
Term Frequency-50 Words	95.92	4.49	100.00	89.19

Table 39: KNN POS Results for Foundation Corpus [1980's Era](1500 Sentences) - Multiple Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	50.00	25.00	12.50	16.67	100.00	100.00	100.00	100.00
Z Score-150 Words	91.67	87.50	96.43	91.75	100.00	100.00	100.00	100.00
Z Score-100 Words	91.67	87.50	96.43	91.75	100.00	100.00	100.00	100.00
Z Score-50 Words	83.33	75.00	68.75	71.74	100.00	100.00	100.00	100.00
TF/IDF	95.83	93.75	98.08	95.87	95.83	93.75	95.00	94.37
TF/IDF-150 Words	100.00	100.00	100.00	100.00	95.83	93.75	95.00	94.37
TF/IDF-100 Words	100.00	100.00	100.00	100.00	95.83	93.75	95.00	94.37
TF/IDF-50 Words	100.00	100.00	100.00	100.00	95.83	93.75	95.00	94.37
Term Frequency	95.83	93.75	98.08	95.87	95.83	93.75	95.00	94.37
Term Frequency-150 Words	100.00	100.00	100.00	100.00	95.83	93.75	95.00	94.37
Term Frequency-100 Words	100.00	100.00	100.00	100.00	95.83	93.75	95.00	94.37
Term Frequency-50 Words	100.00	100.00	100.00	100.00	95.83	93.75	95.00	94.37
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	50.00	25.00	12.50	16.67	50.00	25.00	12.50	16.67
Z Score -150 Words	100.00	100.00	100.00	100.00	91.67	91.67	91.67	91.67
Z Score-100 Words	87.50	81.25	91.43	86.04	95.83	93.75	95.00	94.37
Z Score-50 Words	70.83	56.25	90.79	69.46	91.67	87.50	87.50	87.50
TF/IDF	100.00	100.00	100.00	100.00	91.67	87.50	87.50	87.50
TF/IDF-150 Words	100.00	100.00	100.00	100.00	91.67	87.50	87.50	87.50
TF/IDF-100 Words	100.00	100.00	100.00	100.00	91.67	87.50	87.50	87.50
TF/IDF-50 Words	100.00	100.00	100.00	100.00	91.67	87.50	87.50	87.50
Term Frequency	100.00	100.00	100.00	100.00	91.67	87.50	87.50	87.50
Term Frequency-150 Words	100.00	100.00	100.00	100.00	91.67	87.50	87.50	87.50
Term Frequency-100 Words	100.00	100.00	100.00	100.00	91.67	87.50	87.50	87.50
Term Frequency-50 Words	100.00	100.00	100.00	100.00	91.67	87.50	87.50	87.50

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	91.67	87.50	91.83	89.61
TF/IDF	75.00	70.83	68.98	69.89
Term Frequency	79.17	77.08	78.68	77.87

Punctuation			
Accuracy	Recall	Precision	F-Score
95.83	93.75	98.08	95.87
95.83	93.75	98.08	95.87
95.83	93.75	98.08	95.87

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	55.91	39.36	100.00	16.67
Z Score-150 Words	95.85	4.15	100.00	91.67
Z Score-100 Words	93.04	5.02	100.00	86.04
Z Score-50 Words	82.18	12.42	100.00	69.46
TF/IDF	90.58	9.97	100.00	69.89
TF/IDF-150 Words	95.47	5.14	100.00	87.50
TF/IDF-100 Words	95.47	5.14	100.00	87.50
TF/IDF-50 Words	95.47	5.14	100.00	87.50
Term Frequency	91.91	7.30	100.00	77.87
Term Frequency-150 Words	95.47	5.14	100.00	87.50
Term Frequency-100 Words	95.47	5.14	100.00	87.50
Term Frequency-50 Words	95.47	5.14	100.00	87.50

Table 40: KNN POS Results for Foundation Corpus [1950's Era](500 Sentences) - Multiple Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	50.00	25.00	12.50	16.67	52.08	28.12	37.77	32.24
Z Score-150 Words	79.17	70.83	87.35	78.23	97.92	96.88	97.22	97.05
Z Score-100 Words	64.58	46.88	48.96	47.90	100.00	100.00	100.00	100.00
Z Score-50 Words	70.83	56.25	78.72	65.61	95.83	93.75	96.22	94.97
TF/IDF	95.83	93.75	98.08	95.87	100.00	100.00	100.00	100.00
TF/IDF-150 Words	95.83	93.75	98.08	95.87	100.00	100.00	100.00	100.00
TF/IDF-100 Words	93.75	90.62	94.51	92.52	100.00	100.00	100.00	100.00
TF/IDF-50 Words	93.75	92.71	95.22	93.95	100.00	100.00	100.00	100.00
Term Frequency	93.75	90.62	95.30	92.90	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	97.92	96.88	99.00	97.93	100.00	100.00	100.00	100.00
Term Frequency-50 Words	95.83	93.75	95.88	94.80	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	52.08	28.12	37.77	32.24	50.00	25.00	12.50	16.67
Z Score -150 Words	77.08	65.62	83.18	73.36	72.92	63.54	74.41	68.55
Z Score-100 Words	72.92	64.46	78.58	70.82	79.17	70.83	83.72	76.74
Z Score-50 Words	64.58	46.88	70.38	56.28	79.17	72.92	85.00	78.50
TF/IDF	100.00	100.00	100.00	100.00	93.75	90.62	93.18	91.88
TF/IDF-150 Words	97.92	98.96	97.22	98.08	93.75	90.62	93.18	91.88
TF/IDF-100 Words	95.83	97.92	94.44	96.15	95.83	93.75	95.00	94.37
TF/IDF-50 Words	97.92	96.88	97.22	97.05	97.92	96.88	97.22	97.05
Term Frequency	100.00	100.00	100.00	100.00	97.92	96.88	97.22	97.05
Term Frequency-150 Words	100.00	100.00	100.00	100.00	95.83	93.75	95.00	94.37
Term Frequency-100 Words	100.00	100.00	100.00	100.00	93.75	90.62	93.18	91.88
Term Frequency-50 Words	100.00	100.00	100.00	100.00	95.83	93.75	95.00	94.37

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	81.25	71.88	75.79	73.78
TF/IDF	77.08	67.71	68.54	68.12
Term Frequency	77.08	71.88	73.12	72.49

Punctuation			
Accuracy	Recall	Precision	F-Score
95.83	93.75	93.75	93.75
93.75	94.79	93.36	94.07
93.75	94.79	93.36	94.07

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	44.22	29.25	93.75	16.67
Z Score-150 Words	79.30	10.81	97.05	68.55
Z Score-100 Words	73.86	18.54	100.00	47.90
Z Score-50 Words	73.84	14.53	94.97	56.28
TF/IDF	91.66	10.93	100.00	68.12
TF/IDF-150 Words	96.46	3.02	100.00	91.88
TF/IDF-100 Words	95.76	2.76	100.00	92.52
TF/IDF-50 Words	97.01	2.14	100.00	93.95
Term Frequency	92.75	9.45	100.00	72.49
Term Frequency-150 Words	98.59	2.44	100.00	94.37
Term Frequency-100 Words	97.45	3.33	100.00	91.88
Term Frequency-50 Words	97.29	2.71	100.00	94.37

Table 41: KNN POS Results for Foundation Corpus [1950's Era](1000 Sentences) - Multiple Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	50.00	25.00	12.50	16.67	54.17	31.25	38.04	34.31
Z Score-150 Words	83.33	75.00	88.75	81.30	100.00	100.00	100.00	100.00
Z Score-100 Words	75.00	62.50	81.25	70.65	95.83	93.75	95.00	94.37
Z Score-50 Words	91.67	91.67	91.67	91.67	100.00	100.00	100.00	100.00
TF/IDF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	66.67	50.00	40.00	44.44	50.00	25.00	12.50	16.67
Z Score -150 Words	83.33	83.33	87.50	85.36	83.33	79.17	83.65	81.35
Z Score-100 Words	79.17	77.08	80.95	78.97	79.17	68.75	92.65	78.93
Z Score-50 Words	66.67	75.00	73.21	74.09	70.83	64.58	71.73	67.97
TF/IDF	95.83	93.75	95.00	94.37	95.83	93.75	95.00	94.37
TF/IDF-150 Words	91.67	87.50	88.75	88.12	95.83	93.75	95.00	94.37
TF/IDF-100 Words	87.50	81.25	84.74	82.96	95.83	93.75	95.00	94.37
TF/IDF-50 Words	83.33	79.17	81.25	80.20	95.83	93.75	95.00	94.37
Term Frequency	100.00	100.00	100.00	100.00	95.83	93.75	95.00	94.37
Term Frequency-150 Words	100.00	100.00	100.00	100.00	95.83	93.75	95.00	94.37
Term Frequency-100 Words	100.00	100.00	100.00	100.00	95.83	93.75	95.00	94.37
Term Frequency-50 Words	95.83	93.75	95.00	94.37	95.83	93.75	95.00	94.37

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	83.33	75.00	81.25	78.00
TF/IDF	83.33	79.17	80.00	79.58
Term Frequency	83.33	79.17	83.04	81.06

Punctuation			
Accuracy	Recall	Precision	F-Score
95.83	93.75	95.00	94.37
91.67	91.67	91.67	91.67
95.83	93.75	98.08	95.87

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	47.41	29.47	94.37	16.67
Z Score-150 Words	87.00	7.68	100.00	81.30
Z Score-100 Words	80.73	8.57	94.37	70.65
Z Score-50 Words	83.43	12.93	100.00	67.97
TF/IDF	93.33	6.87	100.00	79.58
TF/IDF-150 Words	95.62	4.90	100.00	88.12
TF/IDF-100 Words	94.33	6.96	100.00	82.96
TF/IDF-50 Words	93.64	8.10	100.00	80.20
Term Frequency	95.22	6.71	100.00	81.06
Term Frequency-150 Words	98.59	2.44	100.00	94.37
Term Frequency-100 Words	98.59	2.44	100.00	94.37
Term Frequency-50 Words	97.19	2.81	100.00	94.37

Table 42: KNN POS Results for Foundation Corpus [1950's Era](1500 Sentences) - Multiple Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	75.00	62.50	66.67	64.52	50.00	25.00	12.50	16.67
Z Score-150 Words	100.00	100.00	100.00	100.00	58.33	37.50	25.00	30.00
Z Score-100 Words	100.00	100.00	100.00	100.00	66.67	50.00	54.17	52.00
Z Score-50 Words	100.00	100.00	100.00	100.00	75.00	62.50	58.93	60.66
TF/IDF	83.33	75.00	58.33	65.62	91.67	87.50	96.43	91.75
TF/IDF-150 Words	83.33	75.00	58.33	65.62	100.00	100.00	100.00	100.00
TF/IDF-100 Words	83.33	75.00	58.33	65.62	100.00	100.00	100.00	100.00
TF/IDF-50 Words	83.33	75.00	58.33	65.62	91.67	87.50	96.43	91.75
Term Frequency	83.33	75.00	58.33	65.62	91.67	87.50	96.43	91.75
Term Frequency-150 Words	83.33	75.00	58.33	65.62	100.00	100.00	100.00	100.00
Term Frequency-100 Words	83.33	75.00	58.33	65.62	100.00	100.00	100.00	100.00
Term Frequency-50 Words	83.33	75.00	58.33	65.62	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	66.67	50.00	40.00	44.44	58.33	37.50	38.64	38.06
Z Score -150 Words	66.67	58.33	47.02	52.07	66.67	50.00	43.75	46.67
Z Score-100 Words	75.00	70.83	54.17	61.39	58.33	54.17	50.00	52.00
Z Score-50 Words	83.33	75.00	68.75	71.74	75.00	70.83	54.17	61.39
TF/IDF	75.00	62.50	50.00	55.56	91.67	87.50	91.67	89.54
TF/IDF-150 Words	75.00	62.50	54.17	58.04	91.67	87.50	91.67	89.54
TF/IDF-100 Words	75.00	62.50	50.60	55.92	66.67	50.00	50.00	50.00
TF/IDF-50 Words	75.00	62.50	58.93	60.66	66.67	50.00	50.00	50.00
Term Frequency	83.33	75.00	62.50	68.18	91.67	87.50	91.67	89.54
Term Frequency-150 Words	83.33	75.00	62.50	68.18	66.67	50.00	50.00	50.00
Term Frequency-100 Words	83.33	75.00	62.50	68.18	66.67	50.00	50.00	50.00
Term Frequency-50 Words	83.33	75.00	63.10	68.54	91.67	87.50	91.67	89.54

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	58.33	37.50	33.33	35.29
TF/IDF	66.67	50.00	43.75	46.67
Term Frequency	66.67	50.00	43.75	46.67

Punctuation			
Accuracy	Recall	Precision	F-Score
83.33	75.00	79.17	77.03
83.33	83.33	83.33	83.33
83.33	83.33	83.33	83.33

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	46.00	19.77	77.03	16.67
Z Score-150 Words	57.18	26.02	100.00	30.00
Z Score-100 Words	66.35	19.80	100.00	52.00
Z Score-50 Words	73.45	15.94	100.00	60.66
TF/IDF	72.08	17.22	91.75	46.67
TF/IDF-150 Words	78.30	17.09	100.00	58.04
TF/IDF-100 Words	67.89	19.36	100.00	50.00
TF/IDF-50 Words	67.01	15.36	91.75	50.00
Term Frequency	74.18	15.78	91.75	46.67
Term Frequency-150 Words	70.95	18.16	100.00	50.00
Term Frequency-100 Words	70.95	18.16	100.00	50.00
Term Frequency-50 Words	80.92	14.37	100.00	65.62

Table 43: Part-Of-Speech Statistics for KNN - Multiple Authors

	Adjectives				All Words			
	Average F-Score	Stand. Dev. F-Score	Max F-Score	Min F-Score	Average F-Score	Stand. Dev. F-Score	Max F-Score	Min F-Score
Z Score	14.83	2.63	16.81	11.11	34.43	27.87	100.00	12.12
Z Score-150 Words	66.69	22.03	91.75	29.29	97.00	3.45	100.00	88.96
Z Score-100 Words	68.16	16.65	91.75	45.97	97.58	2.56	100.00	92.72
Z Score-50 Words	66.10	19.70	94.44	26.50	95.93	4.17	100.00	87.41
TF/IDF	93.84	5.93	100.00	80.08	94.99	10.56	100.00	65.62
TF/IDF-150 Words	95.25	6.48	100.00	79.32	94.84	10.50	100.00	65.62
TF/IDF-100 Words	94.26	7.11	100.00	77.64	94.64	10.43	100.00	65.62
TF/IDF-50 Words	93.06	7.11	100.00	76.58	94.68	10.49	100.00	65.62
Term Frequency	92.66	5.43	100.00	82.67	95.14	10.62	100.00	65.62
Term Frequency-150 Words	97.63	3.51	100.00	89.53	95.14	10.62	100.00	65.62
Term Frequency-100 Words	96.34	4.49	100.00	86.83	94.84	10.52	100.00	65.62
Term Frequency-50 Words	85.61	24.57	100.00	18.18	94.74	10.50	100.00	65.62
	Nouns				Verbs			
	Average F-Score	Stan. Dev. F-Score	Max F-Score	Min F-Score	Average F-Score	Stan. Dev. F-Score	Max F-Score	Min F-Score
Z Score	24.82	11.53	44.44	16.02	26.48	11.76	45.83	16.67
Z Score -150 Words	74.22	17.74	100.00	45.34	73.04	17.43	92.20	44.10
Z Score-100 Words	71.84	12.96	91.75	51.33	75.00	17.09	94.37	46.33
Z Score-50 Words	68.45	9.64	87.77	53.18	72.98	16.05	90.11	35.09
TF/IDF	94.30	13.81	100.00	55.56	92.07	3.08	96.13	87.50
TF/IDF-150 Words	93.43	13.01	100.00	58.04	90.90	2.18	94.37	87.50
TF/IDF-100 Words	92.07	13.76	100.00	55.92	86.54	13.23	94.37	50.00
TF/IDF-50 Words	91.75	12.45	100.00	60.66	85.91	13.10	97.05	50.00
Term Frequency	96.05	9.92	100.00	68.18	93.08	2.93	97.05	87.50
Term Frequency-150 Words	96.38	9.97	100.00	68.18	87.58	13.51	95.52	50.00
Term Frequency-100 Words	96.28	9.95	100.00	68.18	86.83	13.22	94.37	50.00

Term Frequency-50 Words	95.10	9.64	100.00	68.54	89.93	3.33	94.37	84.80
-------------------------	-------	------	--------	-------	-------	------	-------	-------

	Articles			
	Average F-Score	Stand. Dev. F-Score	Max F-Score	Min F-Score
Z Score	77.82	16.51	93.36	35.29
TF /IDF	67.39	9.18	79.58	46.67
Term Frequency	69.35	10.59	81.06	46.67

Punctuation			
Average F-Score	Stand. Dev. F-Score	Max F-Score	Min F-Score
90.5	5.76	96.87	77.03
89.23	4.64	95.87	81.82
89.56	6.00	77.03	79.17

D.3 Naive Bayes POS Results - Two Authors

Table 44: Naive Bayes POS Results for Entire Corpus(500 Sentences) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	99.17	99.17	99.18	99.17	100.00	100.00	100.00	100.00
Z Score-150 Words	72.50	72.50	72.66	72.58	84.17	84.17	84.95	84.56
Z Score-100 Words	77.50	77.50	77.57	77.53	88.33	88.33	89.03	88.68
Z Score-50 Words	71.67	71.67	76.68	74.09	98.33	98.33	98.39	98.36
TF/IDF	85.00	85.00	85.16	85.08	97.50	97.50	97.51	97.50
TF/IDF-150 Words	81.67	81.67	81.70	81.68	95.00	95.00	95.00	95.00
TF/IDF-100 Words	83.33	83.33	83.67	83.50	94.17	94.17	94.18	94.17
TF/IDF-50 Words	80.83	80.83	81.05	80.94	95.00	95.00	95.00	95.00
Term Frequency	91.67	91.67	92.09	91.88	100.00	100.00	100.00	100.00
Term Frequency-150 Words	88.33	88.33	88.72	88.52	95.83	95.83	95.85	95.84
Term Frequency-100 Words	85.00	85.00	85.35	85.17	96.67	96.67	96.72	96.69
Term Frequency-50 Words	82.50	82.50	82.73	82.61	95.00	95.00	95.00	95.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00	78.33	78.33	84.88	81.47
Z Score -150 Words	73.33	73.33	75.12	74.21	63.33	63.33	64.65	63.98
Z Score-100 Words	73.33	73.33	74.00	73.66	60.00	60.00	65.62	62.68
Z Score-50 Words	70.83	70.83	74.42	72.58	74.17	74.17	74.34	74.25
TF/IDF	91.67	91.67	91.85	91.76	89.17	89.17	89.44	89.30
TF/IDF-150 Words	90.83	90.83	91.40	91.11	86.67	86.67	86.71	86.69
TF/IDF-100 Words	91.67	91.67	92.09	91.88	85.83	85.83	86.33	86.08
TF/IDF-50 Words	87.50	87.50	87.76	87.63	85.00	85.00	85.35	85.17
Term Frequency	100.00	100.00	100.00	100.00	90.83	90.83	90.94	90.88
Term Frequency-150 Words	98.33	98.33	98.39	98.36	86.67	86.67	87.04	86.85
Term Frequency-100 Words	95.83	95.83	96.15	95.99	86.67	86.67	86.71	86.69

Term Frequency-50 Words	92.50	92.50	93.09	92.79	85.00	85.00	85.16	85.08
-------------------------	-------	-------	-------	-------	-------	-------	-------	-------

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	92.50	92.50	92.61	92.55
TF/IDF	90.83	92.83	91.40	92.11
Term Frequency	90.83	92.83	91.40	92.11

	Punctuation			
	Accuracy	Recall	Precision	F-Score
Z Score	73.33	73.33	81.11	77.02
TF/IDF	89.17	89.17	89.44	89.30
Term Frequency	88.33	88.33	89.03	88.68

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	91.70	9.26	100.00	77.02
Z Score-150 Words	73.83	7.31	84.56	63.98
Z Score-100 Words	75.64	9.29	88.68	62.68
Z Score-50 Words	79.82	10.72	98.36	72.58
TF/IDF	90.84	3.76	97.50	85.08
TF/IDF-150 Words	88.62	4.97	95.00	81.68
TF/IDF-100 Words	88.91	4.30	94.17	83.50
TF/IDF-50 Words	87.19	5.11	95.00	80.94
Term Frequency	93.93	4.44	100.00	88.68
Term Frequency-150 Words	92.39	4.83	98.36	86.85
Term Frequency-100 Words	91.14	5.24	96.69	85.17
Term Frequency-50 Words	88.87	5.16	95.00	82.61

Table 45: Naive Bayes POS Results for Entire Corpus(1000 Sentences) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Z Score-150 Words	60.00	60.00	61.25	60.62	100.00	100.00	100.00	100.00
Z Score-100 Words	83.33	83.33	83.94	83.63	96.67	96.67	96.67	96.67
Z Score-50 Words	93.33	93.33	93.53	93.43	96.67	96.67	96.88	96.77
TF/IDF	83.33	83.33	84.72	84.02	100.00	100.00	100.00	100.00
TF/IDF-150 Words	80.00	80.00	81.25	80.62	98.33	98.33	98.39	98.36
TF/IDF-100 Words	83.33	83.33	83.94	83.63	98.33	98.33	98.39	98.36
TF/IDF-50 Words	80.00	80.00	81.25	80.62	95.00	95.00	95.05	95.02
Term Frequency	90.00	90.00	90.18	90.09	100.00	100.00	100.00	100.00
Term Frequency-150 Words	81.67	81.67	82.57	82.12	98.33	98.33	98.39	98.36
Term Frequency-100 Words	81.67	81.67	81.99	81.83	96.67	96.67	96.88	96.77
Term Frequency-50 Words	78.33	78.33	78.36	78.34	96.67	96.67	96.67	96.67
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00	81.67	81.67	86.59	84.06
Z Score -150 Words	78.33	78.33	79.14	78.73	88.33	88.33	89.43	88.88
Z Score-100 Words	85.00	85.00	86.00	85.50	91.67	91.67	91.71	91.69
Z Score-50 Words	83.33	83.33	83.33	83.33	85.00	85.00	85.04	85.02
TF/IDF	96.67	96.67	96.88	96.77	91.67	91.67	91.71	91.69
TF/IDF-150 Words	96.67	96.67	96.88	96.77	90.00	90.00	90.18	90.09
TF/IDF-100 Words	95.00	95.00	95.05	95.02	90.00	90.00	90.18	90.09
TF/IDF-50 Words	90.00	90.00	90.00	90.00	90.00	90.00	90.18	90.09
Term Frequency	100.00	100.00	100.00	100.00	91.67	91.67	91.71	91.69
Term Frequency-150 Words	98.33	98.33	98.39	98.36	90.00	90.00	90.18	90.09
Term Frequency-100 Words	98.33	98.33	98.39	98.36	90.00	90.00	90.18	90.09
Term Frequency-50 Words	96.67	96.67	96.88	96.77	88.33	88.33	88.72	88.52

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	93.33	93.33	93.53	93.43
TF/IDF	91.67	91.67	92.09	91.88
Term Frequency	91.67	91.67	92.09	91.88

Punctuation			
Accuracy	Recall	Precision	F-Score
81.67	81.67	86.59	84.06
90.00	90.00	90.72	90.36
90.00	90.00	90.18	90.09

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	93.59	7.13	100.00	84.06
Z Score-150 Words	82.06	14.48	100.00	60.62
Z Score-100 Words	89.37	5.16	96.67	83.63
Z Score-50 Words	89.64	5.62	96.77	83.33
TF/IDF	92.45	5.03	100.00	84.02
TF/IDF-150 Words	91.46	6.99	98.36	80.62
TF/IDF-100 Words	91.78	5.55	98.36	83.63
TF/IDF-50 Words	88.93	5.21	95.02	80.62
Term Frequency	93.96	4.33	100.00	90.09
Term Frequency-150 Words	92.23	6.75	98.36	82.12
Term Frequency-100 Words	91.76	6.52	98.36	81.83
Term Frequency-50 Words	90.08	7.56	96.77	78.34

Table 46: Naive Bayes POS Results for Entire Corpus(1500 Sentences) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Z Score-150 Words	65.00	65.00	65.62	65.31	100.00	100.00	100.00	100.00
Z Score-100 Words	62.50	62.50	62.79	62.64	100.00	100.00	100.00	100.00
Z Score-50 Words	62.50	62.50	64.25	63.36	97.50	97.50	97.62	97.56
TF/IDF	85.00	85.00	85.00	85.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	82.50	82.50	84.67	83.57	97.50	97.50	97.62	97.56
TF/IDF-100 Words	62.50	62.50	62.79	62.64	97.50	97.50	97.62	97.56
TF/IDF-50 Words	77.50	77.50	78.13	77.81	97.50	97.50	97.62	97.56
Term Frequency	90.00	90.00	90.00	90.00	97.50	97.50	97.62	97.56
Term Frequency-150 Words	80.00	80.00	80.30	80.15	97.50	97.50	97.62	97.56
Term Frequency-100 Words	87.50	87.50	88.36	87.93	97.50	97.50	97.62	97.56
Term Frequency-50 Words	82.50	82.50	82.58	82.54	97.50	97.50	97.62	97.56
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00	77.50	77.50	84.48	80.84
Z Score -150 Words	62.50	62.50	63.33	62.91	65.00	65.00	65.15	65.07
Z Score-100 Words	80.00	80.00	80.30	80.15	77.50	77.50	77.57	77.53
Z Score-50 Words	80.00	80.00	80.30	80.15	95.00	95.00	95.45	95.22
TF/IDF	95.00	95.00	95.45	95.22	97.50	97.50	97.62	97.56
TF/IDF-150 Words	95.00	95.00	95.45	95.22	92.50	92.50	93.48	92.99
TF/IDF-100 Words	95.00	95.00	95.45	95.22	95.00	95.00	95.45	95.22
TF/IDF-50 Words	90.00	90.00	90.40	90.20	95.00	95.00	95.00	95.00
Term Frequency	100.00	100.00	100.00	100.00	97.50	97.50	97.62	97.56
Term Frequency-150 Words	100.00	100.00	100.00	100.00	95.00	95.00	95.45	95.22
Term Frequency-100 Words	97.50	97.50	97.62	97.56	95.00	95.00	95.45	95.22
Term Frequency-50 Words	92.50	92.50	93.48	92.99	92.50	92.50	93.48	92.99

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	97.50	97.50	97.62	97.56
TF/IDF	90.00	90.00	91.67	90.83
Term Frequency	90.00	90.00	91.67	90.83

Punctuation			
Accuracy	Recall	Precision	F-Score
85.00	85.00	88.46	86.70
92.50	92.50	92.61	92.55
92.50	92.50	92.61	92.55

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	94.18	7.61	100.00	80.84
Z Score-150 Words	73.32	15.43	100.00	62.91
Z Score-100 Words	80.08	13.30	100.00	62.64
Z Score-50 Words	84.07	13.70	97.56	63.36
TF/IDF	93.53	4.86	100.00	85.00
TF/IDF-150 Words	92.34	5.31	97.56	83.57
TF/IDF-100 Words	87.66	14.48	97.56	62.64
TF/IDF-50 Words	90.14	7.59	97.56	77.81
Term Frequency	94.75	3.79	100.00	90.00
Term Frequency-150 Words	93.23	7.74	100.00	80.15
Term Frequency-100 Words	94.57	3.95	97.56	87.93
Term Frequency-50 Words	91.52	5.51	97.56	82.54

Table 47: Naive Bayes POS Results for Foundation Corpus [1980's Era] (500 Sentences) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	94.94	95.00	95.35	95.17	97.47	97.50	97.56	97.53
Z Score-150 Words	63.29	63.46	64.48	63.97	94.94	94.94	94.94	94.94
Z Score-100 Words	70.89	71.06	72.66	71.85	98.73	98.75	98.75	98.75
Z Score-50 Words	77.22	77.08	78.39	77.73	97.47	97.47	97.47	97.47
TF/IDF	86.08	86.06	86.10	86.08	100.00	100.00	100.00	100.00
TF/IDF-150 Words	83.54	83.49	83.75	83.62	100.00	100.00	100.00	100.00
TF/IDF-100 Words	83.54	83.49	83.75	83.62	100.00	100.00	100.00	100.00
TF/IDF-50 Words	83.54	83.46	84.12	83.79	98.73	98.75	98.75	98.75
Term Frequency	91.14	91.12	91.17	91.14	100.00	100.00	100.00	100.00
Term Frequency-150 Words	87.34	87.31	87.45	87.38	100.00	100.00	100.00	100.00
Term Frequency-100 Words	87.34	87.34	87.34	87.34	100.00	100.00	100.00	100.00
Term Frequency-50 Words	86.08	86.06	86.10	86.08	98.73	98.75	98.75	98.75
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	96.20	96.25	96.43	96.34	81.01	81.25	86.11	83.61
Z Score -150 Words	78.48	78.46	78.50	78.48	67.09	66.96	67.78	67.37
Z Score-100 Words	72.15	72.08	72.37	72.22	63.29	63.04	65.42	64.21
Z Score-50 Words	62.03	61.79	63.63	62.70	75.95	75.77	78.15	76.94
TF/IDF	97.47	97.50	97.56	97.53	93.67	93.65	93.71	93.68
TF/IDF-150 Words	98.73	98.75	98.75	98.75	93.67	93.65	93.71	93.68
TF/IDF-100 Words	97.47	97.50	97.56	97.53	92.41	92.37	92.54	92.45
TF/IDF-50 Words	97.47	97.50	97.56	97.53	93.67	93.62	93.96	93.79
Term Frequency	98.73	98.75	98.75	98.75	93.67	93.65	93.71	93.68
Term Frequency-150 Words	98.73	98.75	98.75	98.75	93.67	93.65	93.71	93.68
Term Frequency-100 Words	98.73	98.75	98.75	98.75	92.41	92.37	92.54	92.45
Term Frequency-50 Words	98.73	98.75	98.75	98.75	93.67	93.65	93.71	93.68

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	97.47	97.50	97.56	97.53
TF/IDF	96.20	96.22	96.22	96.22
Term Frequency	96.20	96.22	96.22	96.22

Punctuation			
Accuracy	Recall	Precision	F-Score
97.47	97.50	97.56	97.53
100.00	100.00	100.00	100.00
98.73	98.72	98.78	98.75

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	94.62	5.00	97.53	83.61
Z Score-150 Words	76.19	12.08	94.94	63.97
Z Score-100 Words	76.76	13.09	98.75	64.21
Z Score-50 Words	78.71	12.37	97.47	62.70
TF/IDF	95.58	4.78	100.00	86.08
TF/IDF-150 Words	94.01	6.45	100.00	83.62
TF/IDF-100 Words	93.40	6.27	100.00	83.62
TF/IDF-50 Words	93.46	5.88	98.75	83.79
Term Frequency	96.42	3.14	100.00	91.14
Term Frequency-150 Words	94.95	4.97	100.00	87.38
Term Frequency-100 Words	94.64	5.09	100.00	87.34
Term Frequency-50 Words	94.31	5.19	98.75	86.08

Table 48: Naive Bayes POS Results for Foundation Corpus [1980's Era] (1000 Sentences) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	94.44	94.44	95.00	94.72	94.44	94.44	95.00	94.72
Z Score-150 Words	72.22	72.22	72.50	72.36	100.00	100.00	100.00	100.00
Z Score-100 Words	88.89	88.89	89.38	89.13	100.00	100.00	100.00	100.00
Z Score-50 Words	94.44	94.44	95.00	94.72	97.22	97.22	97.37	97.29
TF/IDF	88.89	88.89	89.38	89.13	100.00	100.00	100.00	100.00
TF/IDF-150 Words	88.89	88.89	89.38	89.13	100.00	100.00	100.00	100.00
TF/IDF-100 Words	88.89	88.89	89.38	89.13	100.00	100.00	100.00	100.00
TF/IDF-50 Words	86.11	86.11	87.14	86.62	100.00	100.00	100.00	100.00
Term Frequency	88.89	88.89	89.38	89.13	100.00	100.00	100.00	100.00
Term Frequency-150 Words	88.89	88.89	89.38	89.13	100.00	100.00	100.00	100.00
Term Frequency-100 Words	88.89	88.89	89.38	89.13	100.00	100.00	100.00	100.00
Term Frequency-50 Words	86.11	86.11	87.14	86.62	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	94.44	94.44	95.00	94.72	69.44	69.44	81.03	74.79
Z Score -150 Words	97.22	97.22	97.37	97.29	80.56	80.56	80.65	80.60
Z Score-100 Words	88.89	88.89	89.38	89.13	97.22	97.22	97.37	97.29
Z Score-50 Words	91.67	91.67	91.80	91.73	94.44	94.44	94.44	94.44
TF/IDF	100.00	100.00	100.00	100.00	97.22	97.22	97.37	97.29
TF/IDF-150 Words	100.00	100.00	100.00	100.00	97.22	97.22	97.37	97.29
TF/IDF-100 Words	100.00	100.00	100.00	100.00	97.22	97.22	97.37	97.29
TF/IDF-50 Words	97.22	97.22	97.37	97.29	97.22	97.22	97.37	97.29
Term Frequency	100.00	100.00	100.00	100.00	97.22	97.22	97.37	97.29
Term Frequency-150 Words	97.22	97.22	97.37	97.29	97.22	97.22	97.37	97.29
Term Frequency-100 Words	97.22	97.22	97.37	97.29	97.22	97.22	97.37	97.29
Term Frequency-50 Words	94.44	94.44	95.00	94.72	97.22	97.22	97.37	97.29

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00
TF/IDF	97.22	97.22	97.37	97.29
Term Frequency	97.22	97.22	97.37	97.29

Punctuation			
Accuracy	Recall	Precision	F-Score
91.67	91.67	91.80	91.73
100.00	100.00	100.00	100.00
100.00	100.00	100.00	100.00

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	91.78	7.98	100.00	74.79
Z Score-150 Words	87.56	11.50	100.00	72.36
Z Score-100 Words	93.89	4.85	100.00	89.13
Z Score-50 Words	94.55	1.97	97.29	91.73
TF/IDF	97.29	3.84	100.00	89.13
TF/IDF-150 Words	96.61	4.45	100.00	89.13
TF/IDF-100 Words	96.61	4.45	100.00	89.13
TF/IDF-50 Words	95.30	5.13	100.00	86.62
Term Frequency	97.29	3.84	100.00	89.13
Term Frequency-150 Words	95.93	4.08	100.00	89.13
Term Frequency-100 Words	95.93	4.08	100.00	89.13
Term Frequency-50 Words	94.66	5.00	100.00	86.62

Table 49: Naive Bayes POS Results for Foundation Corpus [1980's Era] (1500 Sentences) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	95.83	95.83	96.15	95.99	95.83	95.83	96.15	95.99
Z Score-150 Words	79.17	79.17	81.11	80.13	100.00	100.00	100.00	100.00
Z Score-100 Words	91.67	91.67	92.86	92.26	100.00	100.00	100.00	100.00
Z Score-50 Words	91.67	91.67	92.86	92.26	100.00	100.00	100.00	100.00
TF/IDF	91.67	91.67	92.86	92.26	100.00	100.00	100.00	100.00
TF/IDF-150 Words	91.67	91.67	92.86	92.26	100.00	100.00	100.00	100.00
TF/IDF-100 Words	91.67	91.67	92.86	92.26	100.00	100.00	100.00	100.00
TF/IDF-50 Words	87.50	87.50	90.00	88.73	100.00	100.00	100.00	100.00
Term Frequency	95.83	95.83	96.15	95.99	100.00	100.00	100.00	100.00
Term Frequency-150 Words	95.83	95.83	96.15	95.99	100.00	100.00	100.00	100.00
Term Frequency-100 Words	91.67	91.67	92.86	92.26	100.00	100.00	100.00	100.00
Term Frequency-50 Words	91.67	91.67	92.86	92.26	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	95.83	95.83	96.15	95.99	79.17	79.17	85.29	82.12
Z Score -150 Words	91.67	91.67	92.86	92.26	79.17	79.17	79.37	79.27
Z Score-100 Words	87.50	87.50	90.00	88.73	91.67	91.67	92.86	92.26
Z Score-50 Words	79.17	79.17	79.37	79.27	100.00	100.00	100.00	100.00
TF/IDF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	95.83	95.83	96.15	95.99	100.00	100.00	100.00	100.00

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00
TF/IDF	95.83	95.83	96.15	95.99
Term Frequency	95.83	95.83	96.15	95.99

Punctuation			
Accuracy	Recall	Precision	F-Score
91.67	91.67	92.86	92.26
100.00	100.00	100.00	100.00
100.00	100.00	100.00	100.00

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	93.72	5.65	100.00	82.12
Z Score-150 Words	87.91	8.66	100.00	79.27
Z Score-100 Words	93.31	4.12	100.00	88.73
Z Score-50 Words	92.88	8.47	100.00	79.27
TF/IDF	98.04	2.97	100.00	92.26
TF/IDF-150 Words	98.07	3.35	100.00	92.26
TF/IDF-100 Words	98.07	3.35	100.00	92.26
TF/IDF-50 Words	97.18	4.88	100.00	88.73
Term Frequency	98.66	1.89	100.00	95.99
Term Frequency-150 Words	99.00	1.74	100.00	95.99
Term Frequency-100 Words	98.07	3.35	100.00	92.26
Term Frequency-50 Words	97.06	3.22	100.00	92.26

Table 50: Naive Bayes POS Results for Foundation Corpus [1950's Era] (500 Sentences) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Z Score-150 Words	68.75	68.75	71.82	70.25	93.75	93.75	93.83	93.79
Z Score-100 Words	68.75	68.75	69.05	68.90	97.92	97.92	98.00	97.96
Z Score-50 Words	64.58	64.58	68.46	66.46	95.83	95.83	95.83	95.83
TF/IDF	91.67	91.67	91.96	91.81	100.00	100.00	100.00	100.00
TF/IDF-150 Words	93.75	93.75	93.83	93.79	100.00	100.00	100.00	100.00
TF/IDF-100 Words	91.67	91.67	91.96	91.81	100.00	100.00	100.00	100.00
TF/IDF-50 Words	91.67	91.67	91.96	91.81	100.00	100.00	100.00	100.00
Term Frequency	91.67	91.67	91.96	91.81	100.00	100.00	100.00	100.00
Term Frequency-150 Words	91.67	91.67	91.96	91.81	100.00	100.00	100.00	100.00
Term Frequency-100 Words	95.83	95.83	95.83	95.83	100.00	100.00	100.00	100.00
Term Frequency-50 Words	93.75	93.75	93.83	93.79	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00	83.33	83.33	87.50	85.36
Z Score -150 Words	85.42	85.42	85.48	85.45	77.08	77.08	77.51	77.29
Z Score-100 Words	68.75	68.75	71.82	70.25	89.58	89.58	89.65	89.61
Z Score-50 Words	72.92	72.92	76.67	74.75	77.08	77.08	81.52	79.24
TF/IDF	100.00	100.00	100.00	100.00	95.83	95.83	95.83	95.83
TF/IDF-150 Words	100.00	100.00	100.00	100.00	97.92	97.92	98.00	97.96
TF/IDF-100 Words	100.00	100.00	100.00	100.00	97.92	97.92	98.00	97.96
TF/IDF-50 Words	97.92	97.92	98.00	97.96	97.92	97.92	98.00	97.96
Term Frequency	97.92	97.92	98.00	97.96	97.92	97.92	98.00	97.96
Term Frequency-150 Words	97.92	97.92	98.00	97.96	97.92	97.92	98.00	97.96
Term Frequency-100 Words	100.00	100.00	100.00	100.00	97.92	97.92	98.00	97.96
Term Frequency-50 Words	97.92	97.92	98.00	97.96	97.92	97.92	98.00	97.96

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	93.75	93.75	93.83	93.79
TF/IDF	97.92	97.92	98.00	97.96
Term Frequency	97.92	97.92	98.00	97.96

Punctuation			
Accuracy	Recall	Precision	F-Score
77.08	77.08	81.52	79.24
95.83	95.83	96.15	95.99
93.75	93.75	94.44	94.09

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	93.07	8.12	100.00	79.24
Z Score-150 Words	81.70	8.81	93.79	70.25
Z Score-100 Words	81.68	12.47	97.96	68.90
Z Score-50 Words	79.07	10.71	95.83	66.46
TF/IDF	96.93	2.83	100.00	91.81
TF/IDF-150 Words	97.94	2.54	100.00	93.79
TF/IDF-100 Words	97.44	3.35	100.00	91.81
TF/IDF-50 Words	96.93	3.07	100.00	91.81
Term Frequency	96.63	2.78	100.00	91.81
Term Frequency-150 Words	96.93	3.07	100.00	91.81
Term Frequency-100 Words	98.45	1.73	100.00	95.83
Term Frequency-50 Words	97.43	2.26	100.00	93.79

Table 51: Naive Bayes POS Results for Foundation Corpus [1950's Era] (1000 Sentences) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	95.83	95.83	96.15	95.99	100.00	100.00	100.00	100.00
Z Score-150 Words	83.33	83.33	83.33	83.33	91.67	91.67	91.67	91.67
Z Score-100 Words	79.17	79.17	79.37	79.27	95.83	95.83	96.15	95.99
Z Score-50 Words	79.17	79.17	81.11	80.13	91.67	91.67	91.67	91.67
TF/IDF	95.83	95.83	96.15	95.99	100.00	100.00	100.00	100.00
TF/IDF-150 Words	95.83	95.83	96.15	95.99	100.00	100.00	100.00	100.00
TF/IDF-100 Words	95.83	95.83	96.15	95.99	100.00	100.00	100.00	100.00
TF/IDF-50 Words	95.83	95.83	96.15	95.99	100.00	100.00	100.00	100.00
Term Frequency	95.83	95.83	96.15	95.99	100.00	100.00	100.00	100.00
Term Frequency-150 Words	95.83	95.83	96.15	95.99	100.00	100.00	100.00	100.00
Term Frequency-100 Words	95.83	95.83	96.15	95.99	100.00	100.00	100.00	100.00
Term Frequency-50 Words	95.83	95.83	96.15	95.99	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00	83.33	83.33	87.50	85.36
Z Score -150 Words	79.17	79.17	79.37	79.27	87.50	87.50	90.00	88.73
Z Score-100 Words	100.00	100.00	100.00	100.00	91.67	91.67	92.86	92.26
Z Score-50 Words	83.33	83.33	84.29	83.81	79.17	79.17	79.37	79.27
TF/IDF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00
TF/IDF	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00

Punctuation			
Accuracy	Recall	Precision	F-Score
87.50	87.50	87.76	87.63
100.00	100.00	100.00	100.00
95.83	95.83	96.15	95.99

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	94.83	6.10	100.00	85.36
Z Score-150 Words	85.75	4.79	91.67	79.27
Z Score-100 Words	91.88	7.78	100.00	79.27
Z Score-50 Words	83.72	4.90	91.67	79.27
TF/IDF	99.33	1.49	100.00	95.99
TF/IDF-150 Words	99.00	1.74	100.00	95.99
TF/IDF-100 Words	99.00	1.74	100.00	95.99
TF/IDF-50 Words	99.00	1.74	100.00	95.99
Term Frequency	98.66	1.89	100.00	95.99
Term Frequency-150 Words	99.00	1.74	100.00	95.99
Term Frequency-100 Words	99.00	1.74	100.00	95.99
Term Frequency-50 Words	99.00	1.74	100.00	95.99

Table 52: Naive Bayes POS Results for Foundation Corpus [1950's Era] (1500 Sentences) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Z Score-150 Words	58.33	58.33	58.57	58.45	83.33	83.33	83.33	83.33
Z Score-100 Words	50.00	50.00	50.00	50.00	91.67	91.67	92.86	92.26
Z Score-50 Words	58.33	58.33	61.11	59.69	91.67	91.67	92.86	92.26
TF/IDF	91.67	91.67	92.86	92.26	100.00	100.00	100.00	100.00
TF/IDF-150 Words	91.67	91.67	92.86	92.26	100.00	100.00	100.00	100.00
TF/IDF-100 Words	83.33	83.33	87.50	85.36	100.00	100.00	100.00	100.00
TF/IDF-50 Words	83.33	83.33	87.50	85.36	100.00	100.00	100.00	100.00
Term Frequency	91.67	91.67	92.86	92.26	100.00	100.00	100.00	100.00
Term Frequency-150 Words	91.67	91.67	92.86	92.26	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	83.33	83.33	87.50	85.36	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Z Score -150 Words	75.00	75.00	83.33	78.95	75.00	75.00	75.71	75.35
Z Score-100 Words	50.00	50.00	50.00	50.00	66.67	66.67	68.75	67.69
Z Score-50 Words	66.67	66.67	80.00	72.73	83.33	83.33	83.33	83.33
TF/IDF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-50 Words	91.67	91.67	92.86	92.26	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00
TF/IDF	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00

Punctuation			
Accuracy	Recall	Precision	F-Score
91.67	91.67	92.86	92.26
83.33	83.33	87.50	85.36
83.33	83.33	87.50	85.36

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	98.71	2.88	100.00	92.26
Z Score-150 Words	74.02	9.42	83.33	58.45
Z Score-100 Words	64.99	17.32	92.26	50.00
Z Score-50 Words	77.00	12.15	92.26	59.69
TF/IDF	96.27	5.64	100.00	85.36
TF/IDF-150 Words	98.07	3.35	100.00	92.26
TF/IDF-100 Words	96.34	6.34	100.00	85.36
TF/IDF-50 Words	94.41	6.10	100.00	85.36
Term Frequency	96.27	5.64	100.00	85.36
Term Frequency-150 Words	98.07	3.35	100.00	92.26
Term Frequency-100 Words	100.00	0.00	100.00	100.00
Term Frequency-50 Words	96.34	6.34	100.00	85.36

Table 53: Naive Bayes POS Results for Galactic Empire Corpus (500 Sentences) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Z Score-150 Words	80.95	80.92	82.12	81.52	92.86	92.86	92.95	92.90
Z Score-100 Words	73.81	73.81	73.86	73.83	100.00	100.00	100.00	100.00
Z Score-50 Words	57.14	57.14	57.78	57.46	97.62	97.62	97.73	97.67
TF/IDF	95.24	95.24	95.65	95.44	100.00	100.00	100.00	100.00
TF/IDF-150 Words	95.24	95.24	95.65	95.44	100.00	100.00	100.00	100.00
TF/IDF-100 Words	95.24	95.24	95.65	95.44	100.00	100.00	100.00	100.00
TF/IDF-50 Words	95.24	95.24	95.65	95.44	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-150 Words	97.62	97.62	97.73	97.67	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	95.24	95.24	95.65	95.44	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Z Score -150 Words	73.81	73.81	73.86	73.83	57.14	57.14	57.78	57.46
Z Score-100 Words	78.57	78.57	79.17	78.87	66.67	66.67	66.67	66.67
Z Score-50 Words	97.62	97.62	97.73	97.67	80.95	80.95	86.21	83.50
TF/IDF	100.00	100.00	100.00	100.00	83.33	83.33	83.41	83.37
TF/IDF-150 Words	100.00	100.00	100.00	100.00	88.10	88.10	88.89	88.49
TF/IDF-100 Words	100.00	100.00	100.00	100.00	83.33	83.33	83.41	83.37
TF/IDF-50 Words	100.00	100.00	100.00	100.00	83.33	83.33	83.41	83.37
Term Frequency	100.00	100.00	100.00	100.00	90.48	90.48	92.00	91.23
Term Frequency-150 Words	100.00	100.00	100.00	100.00	92.86	92.86	93.75	93.30
Term Frequency-100 Words	100.00	100.00	100.00	100.00	90.48	90.48	90.85	90.66
Term Frequency-50 Words	100.00	100.00	100.00	100.00	88.10	88.10	88.18	88.14

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	92.86	92.86	92.95	92.90
TF/IDF	95.24	95.24	95.65	95.44
Term Frequency	97.62	97.62	97.73	97.67

Punctuation			
Accuracy	Recall	Precision	F-Score
97.62	97.62	97.73	97.67
97.62	97.62	97.73	97.67
100.00	100.00	100.00	100.00

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	98.43	2.61	100.00	92.90
Z Score-150 Words	76.43	12.88	92.90	57.46
Z Score-100 Words	79.84	12.42	100.00	66.67
Z Score-50 Words	84.08	16.42	97.67	57.46
TF/IDF	95.32	5.66	100.00	83.37
TF/IDF-150 Words	95.98	4.71	100.00	88.49
TF/IDF-100 Words	94.70	6.80	100.00	83.37
TF/IDF-50 Words	94.70	6.80	100.00	83.37
Term Frequency	98.15	3.21	100.00	91.23
Term Frequency-150 Words	97.74	2.73	100.00	93.30
Term Frequency-100 Words	97.67	4.04	100.00	90.66
Term Frequency-50 Words	95.90	4.85	100.00	88.14

Table 54: Naive Bayes POS Results for Galactic Empire Corpus (1000 Sentences) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Z Score-150 Words	77.78	77.78	77.78	77.78	88.89	88.89	88.89	88.89
Z Score-100 Words	77.78	77.78	79.22	78.49	100.00	100.00	100.00	100.00
Z Score-50 Words	88.89	88.89	88.89	88.89	100.00	100.00	100.00	100.00
TF/IDF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Z Score -150 Words	83.33	83.33	83.75	83.54	77.78	77.78	79.22	78.49
Z Score-100 Words	88.89	88.89	90.91	89.89	88.89	88.89	90.91	89.89
Z Score-50 Words	100.00	100.00	100.00	100.00	94.44	94.44	95.00	94.72
TF/IDF	100.00	100.00	100.00	100.00	88.89	88.89	88.89	88.89
TF/IDF-150 Words	100.00	100.00	100.00	100.00	83.33	83.33	83.75	83.54
TF/IDF-100 Words	100.00	100.00	100.00	100.00	83.33	83.33	83.75	83.54
TF/IDF-50 Words	100.00	100.00	100.00	100.00	77.78	77.78	77.78	77.78
Term Frequency	100.00	100.00	100.00	100.00	88.89	88.89	88.89	88.89
Term Frequency-150 Words	100.00	100.00	100.00	100.00	83.33	83.33	83.75	83.54
Term Frequency-100 Words	100.00	100.00	100.00	100.00	77.78	77.78	77.78	77.78
Term Frequency-50 Words	100.00	100.00	100.00	100.00	72.22	72.22	72.50	72.36

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00
TF/IDF	88.89	88.89	88.89	88.89
Term Frequency	88.89	88.89	88.89	88.89

Punctuation			
Accuracy	Recall	Precision	F-Score
100.00	100.00	100.00	100.00
100.00	100.00	100.00	100.00
100.00	100.00	100.00	100.00

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	100.00	0.00	100.00	100.00
Z Score-150 Words	82.18	4.47	88.89	77.78
Z Score-100 Words	89.57	7.61	100.00	78.49
Z Score-50 Words	95.90	4.59	100.00	88.89
TF/IDF	96.30	5.24	100.00	88.89
TF/IDF-150 Words	95.88	7.13	100.00	83.54
TF/IDF-100 Words	95.88	7.13	100.00	83.54
TF/IDF-50 Words	95.56	8.89	100.00	77.78
Term Frequency	95.56	5.44	100.00	88.89
Term Frequency-150 Words	95.88	7.13	100.00	83.54
Term Frequency-100 Words	94.45	9.62	100.00	77.78
Term Frequency-50 Words	93.09	11.97	100.00	72.36

Table 55: Naive Bayes POS Results for Galactic Empire Corpus (1500 Sentences) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Z Score-150 Words	75.00	75.00	75.71	75.35	100.00	100.00	100.00	100.00
Z Score-100 Words	75.00	75.00	75.71	75.35	91.67	91.67	92.86	92.26
Z Score-50 Words	91.67	91.67	92.86	92.26	91.67	91.67	92.86	92.26
TF/IDF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Z Score -150 Words	100.00	100.00	100.00	100.00	91.67	91.67	92.86	92.26
Z Score-100 Words	66.67	66.67	68.75	67.69	66.67	66.67	68.75	67.69
Z Score-50 Words	83.33	83.33	87.50	85.36	58.33	58.33	58.57	58.45
TF/IDF	100.00	100.00	100.00	100.00	83.33	83.33	83.33	83.33
TF/IDF-150 Words	100.00	100.00	100.00	100.00	83.33	83.33	83.33	83.33
TF/IDF-100 Words	100.00	100.00	100.00	100.00	91.67	91.67	92.86	92.26
TF/IDF-50 Words	100.00	100.00	100.00	100.00	83.33	83.33	83.33	83.33
Term Frequency	100.00	100.00	100.00	100.00	91.67	91.67	92.86	92.26
Term Frequency-150 Words	100.00	100.00	100.00	100.00	83.33	83.33	83.33	83.33
Term Frequency-100 Words	100.00	100.00	100.00	100.00	91.67	91.67	92.86	92.26
Term Frequency-50 Words	100.00	100.00	100.00	100.00	83.33	83.33	83.33	83.33

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	91.67	91.67	92.86	92.26
TF/IDF	83.33	83.33	83.33	83.33
Term Frequency	75.00	75.00	75.71	75.35

Punctuation			
Accuracy	Recall	Precision	F-Score
100.00	100.00	100.00	100.00
100.00	100.00	100.00	100.00
100.00	100.00	100.00	100.00

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	97.42	3.65	100.00	92.26
Z Score-150 Words	91.90	10.06	100.00	75.35
Z Score-100 Words	75.75	10.03	92.26	67.69
Z Score-50 Words	82.08	13.93	92.26	58.45
TF/IDF	94.44	7.86	100.00	83.33
TF/IDF-150 Words	95.83	7.22	100.00	83.33
TF/IDF-100 Words	98.07	3.35	100.00	92.26
TF/IDF-50 Words	95.83	7.22	100.00	83.33
Term Frequency	94.60	9.06	100.00	75.35
Term Frequency-150 Words	95.83	7.22	100.00	83.33
Term Frequency-100 Words	98.07	3.35	100.00	92.26
Term Frequency-50 Words	95.83	7.22	100.00	83.33

Table 56: Naive Bayes POS Results for Robot Corpus (500 Sentences) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Z Score-150 Words	66.07	66.07	66.60	66.33	94.64	94.64	95.16	94.90
Z Score-100 Words	75.00	75.00	75.00	75.00	94.64	94.64	94.70	94.67
Z Score-50 Words	82.14	82.14	82.81	82.47	96.43	96.43	96.67	96.55
TF/IDF	87.50	87.50	88.74	88.12	100.00	100.00	100.00	100.00
TF/IDF-150 Words	87.50	87.50	88.74	88.12	100.00	100.00	100.00	100.00
TF/IDF-100 Words	87.50	87.50	88.74	88.12	100.00	100.00	100.00	100.00
TF/IDF-50 Words	85.74	85.74	87.43	86.58	98.21	98.21	98.28	98.24
Term Frequency	98.21	98.21	98.28	98.24	100.00	100.00	100.00	100.00
Term Frequency-150 Words	94.64	94.64	94.70	94.67	100.00	100.00	100.00	100.00
Term Frequency-100 Words	94.64	94.64	94.77	94.70	100.00	100.00	100.00	100.00
Term Frequency-50 Words	91.07	91.07	91.55	91.31	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00	98.21	98.21	98.28	98.24
Z Score -150 Words	80.36	80.36	81.36	80.86	82.14	82.14	82.31	82.22
Z Score-100 Words	89.29	89.29	90.10	89.69	80.36	80.36	80.71	80.53
Z Score-50 Words	94.64	94.64	95.16	94.90	92.86	92.86	92.86	92.86
TF/IDF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00
TF/IDF	96.43	96.43	96.43	96.43
Term Frequency	96.43	96.43	96.43	96.43

Punctuation			
Accuracy	Recall	Precision	F-Score
87.50	87.50	87.94	87.72
98.21	98.21	98.28	98.24
98.21	98.21	98.28	98.24

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	97.66	4.49	100.00	87.72
Z Score-150 Words	81.08	10.12	94.90	66.33
Z Score-100 Words	84.97	7.67	94.67	75.00
Z Score-50 Words	91.70	5.48	96.55	82.47
TF/IDF	97.13	4.24	100.00	88.12
TF/IDF-150 Words	97.03	5.15	100.00	88.12
TF/IDF-100 Words	97.03	5.15	100.00	88.12
TF/IDF-50 Words	96.21	5.61	100.00	86.58
Term Frequency	98.82	1.33	100.00	96.43
Term Frequency-150 Words	98.67	2.31	100.00	94.67
Term Frequency-100 Words	98.68	2.29	100.00	94.70
Term Frequency-50 Words	97.83	3.76	100.00	91.31

Table 57: Naive Bayes POS Results for Robot Corpus (1000 Sentences) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Z Score-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Z Score-100 Words	86.67	86.61	86.61	86.61	100.00	100.00	100.00	100.00
Z Score-50 Words	86.67	86.61	86.61	86.61	95.24	95.83	95.00	95.41
TF/IDF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	93.33	92.86	94.44	93.64	100.00	100.00	100.00	100.00
TF/IDF-100 Words	93.33	92.86	94.44	93.64	100.00	100.00	100.00	100.00
TF/IDF-50 Words	93.33	92.86	94.44	93.64	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	86.67	86.61	86.61	86.61	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Z Score -150 Words	66.67	66.67	66.36	66.51	71.43	69.44	71.43	70.42
Z Score-100 Words	76.19	75.00	75.96	75.48	76.19	72.22	85.29	78.21
Z Score-50 Words	90.48	88.89	92.86	90.83	100.00	100.00	100.00	100.00
TF/IDF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00
TF/IDF	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00

Punctuation			
Accuracy	Recall	Precision	F-Score
90.48	90.28	90.28	90.28
95.24	94.44	96.15	95.29
95.24	94.44	96.15	95.29

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	98.38	3.62	100.00	90.28
Z Score-150 Words	84.23	15.83	100.00	66.51
Z Score-100 Words	85.07	9.54	100.00	75.48
Z Score-50 Words	93.21	5.00	100.00	86.61
TF/IDF	99.21	1.76	100.00	95.29
TF/IDF-150 Words	98.41	2.75	100.00	93.64
TF/IDF-100 Words	98.41	2.75	100.00	93.64
TF/IDF-50 Words	98.41	2.75	100.00	93.64
Term Frequency	99.21	1.76	100.00	95.29
Term Frequency-150 Words	100.00	0.00	100.00	100.00
Term Frequency-100 Words	100.00	0.00	100.00	100.00
Term Frequency-50 Words	96.65	5.80	100.00	86.61

Table 58: Naive Bayes POS Results for Robot Corpus (1500 Sentences) - Two Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Z Score-150 Words	100.00	100.00	100.00	100.00	93.33	92.86	94.44	93.64
Z Score-100 Words	86.67	86.61	86.61	86.61	100.00	100.00	100.00	100.00
Z Score-50 Words	86.67	86.61	86.61	86.61	100.00	100.00	100.00	100.00
TF/IDF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	93.33	92.86	94.44	93.64	100.00	100.00	100.00	100.00
TF/IDF-100 Words	93.33	92.86	94.44	93.64	100.00	100.00	100.00	100.00
TF/IDF-50 Words	93.33	92.86	94.44	93.64	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	86.67	86.61	86.61	86.61	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Z Score -150 Words	100.00	100.00	100.00	100.00	73.33	73.32	75.00	74.15
Z Score-100 Words	86.67	87.50	88.89	88.19	93.33	92.86	94.44	93.64
Z Score-50 Words	86.67	86.61	86.61	86.61	100.00	100.00	100.00	100.00
TF/IDF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-50 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	100.00	100.00	100.00	100.00	93.33	92.86	94.44	93.64

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00
TF/IDF	93.33	93.75	93.75	93.75
Term Frequency	93.33	93.75	93.75	93.75

Punctuation			
Accuracy	Recall	Precision	F-Score
80.00	79.46	80.56	80.01
100.00	100.00	100.00	100.00
100.00	100.00	100.00	100.00

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	96.67	7.45	100.00	80.01
Z Score-150 Words	91.95	10.60	100.00	74.15
Z Score-100 Words	92.11	5.25	100.00	86.61
Z Score-50 Words	93.31	6.70	100.00	86.61
TF/IDF	98.96	2.33	100.00	93.75
TF/IDF-150 Words	98.41	2.75	100.00	93.64
TF/IDF-100 Words	98.41	2.75	100.00	93.64
TF/IDF-50 Words	98.41	2.75	100.00	93.64
Term Frequency	98.96	2.33	100.00	93.75
Term Frequency-150 Words	100.00	0.00	100.00	100.00
Term Frequency-100 Words	100.00	0.00	100.00	100.00
Term Frequency-50 Words	95.06	5.53	100.00	86.61

Table 59: Part-Of-Speech Statistics for Naive Bayes - Two Authors

	Adjectives				All Words			
	Average F-Score	Stand. Dev. F-Score	Max F-Score	Min F-Score	Average F-Score	Stand. Dev. F-Score	Max F-Score	Min F-Score
Z Score	98.74	2.00	100.00	94.72	99.22	1.65	100.00	94.72
Z Score-150 Words	75.20	12.12	100.00	58.45	94.58	5.47	100.00	83.33
Z Score-100 Words	76.74	10.52	92.26	50.00	97.15	3.53	100.00	88.68
Z Score-50 Words	79.75	12.36	94.72	57.46	96.61	2.65	100.00	91.67
TF/IDF	92.35	5.78	100.00	84.02	99.83	0.62	100.00	97.50
TF/IDF-150 Words	90.92	6.04	100.00	80.62	99.39	1.37	100.00	95.00
TF/IDF-100 Words	89.25	8.91	100.00	62.64	99.34	1.55	100.00	94.17
TF/IDF-50 Words	89.40	6.80	100.00	77.81	98.97	1.72	100.00	95.00
Term Frequency	95.10	4.19	100.00	89.13	99.84	0.61	100.00	97.56
Term Frequency-150 Words	93.05	6.27	100.00	80.15	99.45	1.20	100.00	95.84
Term Frequency-100 Words	94.01	6.10	100.00	81.83	99.40	1.21	100.00	96.69
Term Frequency-50 Words	89.57	6.33	100.00	78.34	99.20	1.50	100.00	95.00
	Nouns				Verbs			
	Average F-Score	Stan. Dev. F-Score	Max F-Score	Min F-Score	Average F-Score	Stan. Dev. F-Score	Max F-Score	Min F-Score
Z Score	99.14	1.75	100.00	94.72	90.39	9.07	100.00	74.79
Z Score -150 Words	82.15	10.87	100.00	62.91	76.10	9.62	92.26	57.46
Z Score-100 Words	79.96	11.88	100.00	50.00	80.79	11.95	97.29	62.68
Z Score-50 Words	83.76	10.04	100.00	62.70	86.48	11.40	100.00	58.45
TF/IDF	98.75	2.36	100.00	91.76	94.73	5.86	100.00	83.33
TF/IDF-150 Words	98.79	2.47	100.00	91.11	94.27	6.14	100.00	83.33
TF/IDF-100 Words	98.64	2.48	100.00	91.88	94.55	6.03	100.00	83.37
TF/IDF-50 Words	96.86	4.31	100.00	87.63	93.59	7.41	100.00	77.78
Term Frequency	99.78	0.58	100.00	97.96	96.10	4.01	100.00	88.89
Term Frequency-150 Words	99.38	0.92	100.00	97.29	94.75	5.92	100.00	83.33
Term Frequency-100 Words	99.20	1.26	100.00	95.99	94.69	6.25	100.00	77.78

Term Frequency-50 Words	98.00	2.60	100.00	92.79	92.87	7.81	100.00	72.36
-------------------------	-------	------	--------	-------	-------	------	--------	-------

	Articles			
	Average F-Score	Stand. Dev. F-Score	Max F-Score	Min F-Score
Z Score	97.34	3.19	100	92.26
TF/IDF	94.68	4.49	100	83.33
Term Frequency	94.29	6.07	100	46.67

	Punctuation			
	Average F-Score	Stand. Dev. F-Score	Max F-Score	Min F-Score
Z Score	89.09	6.65	100	77.02
TF/IDF	96.32	4.62	100	85.36
Term Frequency	95.65	4.68	100	85.36

D.4 Naive Bayes POS Results - Multiple Authors

Table 60: Naive Bayes POS Results for Entire Corpus(500 Sentences) - Multiple Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	61.67	36.11	75.10	48.77	59.17	31.94	59.17	41.49
Z Score-150 Words	58.33	42.78	56.11	48.55	50.83	38.06	41.85	39.87
Z Score-100 Words	49.17	43.06	39.67	41.30	71.67	70.56	72.59	71.56
Z Score-50 Words	35.83	37.50	37.33	37.41	95.00	92.78	96.04	94.38
TF/IDF	67.50	46.94	57.21	51.57	99.17	98.61	99.73	99.17
TF/IDF-150 Words	66.67	45.56	57.11	50.69	98.33	97.22	99.46	98.33
TF/IDF-100 Words	67.50	46.94	73.88	57.41	97.50	96.94	98.06	97.50
TF/IDF-50 Words	66.67	45.56	57.11	50.69	97.50	96.94	98.06	97.50
Term Frequency	75.83	60.83	92.20	73.30	100.00	100.00	100.00	100.00
Term Frequency-150 Words	74.17	58.06	75.28	65.56	99.17	98.61	99.73	99.17
Term Frequency-100 Words	73.33	56.67	91.82	70.08	98.33	97.22	99.46	98.33
Term Frequency-50 Words	68.33	49.44	73.21	59.02	96.67	96.67	96.88	96.77
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	75.00	58.33	94.44	72.12	50.83	18.06	25.07	21.00
Z Score -150 Words	55.00	36.11	45.81	40.39	45.83	21.94	20.71	21.31
Z Score-100 Words	48.33	36.11	49.78	41.86	37.50	22.50	22.49	22.49
Z Score-50 Words	37.50	40.28	35.23	37.59	39.17	36.39	34.37	35.35
TF/IDF	95.00	91.67	98.48	94.95	87.50	93.61	87.11	90.24
TF/IDF-150 Words	93.33	90.00	95.10	92.48	88.33	83.89	86.93	85.38
TF/IDF-100 Words	92.50	89.72	93.79	91.71	89.17	84.17	87.34	85.73
TF/IDF-50 Words	90.83	85.83	94.35	89.89	85.83	80.83	83.21	82.00
Term Frequency	99.17	99.72	98.72	99.22	92.50	87.50	92.33	89.85
Term Frequency-150 Words	99.17	99.72	98.72	99.22	90.00	84.44	92.60	88.33
Term Frequency-100 Words	95.83	98.61	95.10	96.82	88.33	82.78	86.40	84.55

Term Frequency-50 Words	94.17	94.72	92.72	93.71	85.83	80.83	83.50	82.14
-------------------------	-------	-------	-------	-------	-------	-------	-------	-------

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	87.50	81.39	84.36	82.85
TF/IDF	87.50	79.17	86.97	82.89
Term Frequency	88.83	80.56	87.40	83.84

	Punctuation			
	Accuracy	Recall	Precision	F-Score
Z Score	64.17	75.83	73.18	74.48
TF/IDF	86.67	87.78	86.21	86.99
Term Frequency	87.50	86.94	86.28	86.61

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	56.78	21.63	82.85	21.00
Z Score-150 Words	37.53	9.98	48.55	21.31
Z Score-100 Words	44.30	17.56	71.56	22.49
Z Score-50 Words	51.18	24.96	94.38	35.35
TF/IDF	84.30	15.55	99.17	51.57
TF/IDF-150 Words	81.72	18.49	98.33	50.69
TF/IDF-100 Words	83.08	15.40	97.50	57.41
TF/IDF-50 Words	80.02	17.80	97.50	50.69
Term Frequency	88.80	9.17	100.00	73.30
Term Frequency-150 Words	88.07	13.73	99.22	65.56
Term Frequency-100 Words	87.45	11.36	98.33	70.08
Term Frequency-50 Words	82.91	14.83	96.77	59.02

Table 61: Naive Bayes POS Results for Entire Corpus(1000 Sentences) - Multiple Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	58.33	30.56	59.09	40.29	53.33	22.22	41.95	29.05
Z Score-150 Words	55.00	29.44	37.99	33.17	70.00	50.00	77.08	60.65
Z Score-100 Words	56.67	38.89	59.31	46.98	81.67	69.44	95.53	80.42
Z Score-50 Words	65.00	48.33	68.20	56.57	83.33	72.22	90.60	80.37
TF/IDF	68.33	49.44	57.51	53.17	100.00	100.00	100.00	100.00
TF/IDF-150 Words	70.00	52.22	57.73	54.84	100.00	100.00	100.00	100.00
TF/IDF-100 Words	68.33	49.44	55.75	52.41	98.33	97.22	99.46	98.33
TF/IDF-50 Words	68.33	49.44	55.75	52.41	100.00	100.00	100.00	100.00
Term Frequency	75.00	58.33	61.11	59.69	100.00	100.00	100.00	100.00
Term Frequency-150 Words	71.67	55.00	56.82	55.90	100.00	100.00	100.00	100.00
Term Frequency-100 Words	68.33	49.44	55.75	55.90	100.00	100.00	100.00	100.00
Term Frequency-50 Words	68.33	49.44	55.75	52.41	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	70.00	50.00	77.08	60.65	50.00	16.67	8.33	11.11
Z Score -150 Words	55.00	36.11	38.50	37.27	86.67	34.44	42.46	38.03
Z Score-100 Words	61.67	36.11	76.47	49.06	66.67	44.44	82.64	57.80
Z Score-50 Words	60.00	33.33	59.62	42.76	75.00	60.56	79.03	68.57
TF/IDF	98.33	97.22	99.46	98.33	93.33	88.89	92.85	90.83
TF/IDF-150 Words	98.33	97.22	99.46	98.33	91.67	88.33	90.83	89.56
TF/IDF-100 Words	98.33	97.22	99.46	98.33	90.00	83.33	89.60	86.35
TF/IDF-50 Words	96.67	94.44	98.96	96.65	88.33	85.00	87.08	86.03
Term Frequency	100.00	100.00	100.00	100.00	91.67	88.33	90.83	89.56
Term Frequency-150 Words	98.33	99.44	97.62	98.52	91.67	88.33	90.83	89.56
Term Frequency-100 Words	98.33	99.44	97.62	98.52	90.00	85.56	88.99	87.24
Term Frequency-50 Words	98.33	97.22	97.62	97.42	88.33	82.78	87.33	84.99

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	83.33	72.22	83.73	77.55
TF/IDF	88.33	80.56	89.55	84.82
Term Frequency	90.00	83.33	94.84	88.71

Punctuation			
Accuracy	Recall	Precision	F-Score
70.00	78.89	76.95	77.91
91.67	92.78	92.60	92.69
91.67	92.78	93.30	93.04

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	49.43	24.82	77.91	11.11
Z Score-150 Words	42.28	10.77	60.65	33.17
Z Score-100 Words	58.56	13.26	80.42	46.98
Z Score-50 Words	62.07	13.97	80.37	42.76
TF/IDF	86.64	15.77	100.00	53.17
TF/IDF-150 Words	85.68	18.24	100.00	54.84
TF/IDF-100 Words	83.85	18.80	98.33	52.41
TF/IDF-50 Words	83.77	18.83	100.00	52.41
Term Frequency	88.50	13.64	100.00	59.69
Term Frequency-150 Words	85.99	17.83	100.00	55.90
Term Frequency-100 Words	85.41	17.74	100.00	55.90
Term Frequency-50 Words	83.70	18.94	100.00	52.41

Table 62: Naive Bayes POS Results for Entire Corpus(1500 Sentences) - Multiple Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	52.50	20.83	25.21	22.81	52.50	20.83	25.21	22.81
Z Score-150 Words	57.50	29.17	59.52	39.15	50.00	16.67	8.33	11.11
Z Score-100 Words	55.00	28.33	42.93	34.13	55.00	25.00	42.11	31.37
Z Score-50 Words	52.50	20.83	15.36	17.68	70.00	50.00	77.08	60.65
TF/IDF	72.50	54.17	60.75	57.27	97.50	95.83	99.21	97.49
TF/IDF-150 Words	72.50	54.17	60.75	57.27	97.50	95.83	99.21	97.49
TF/IDF-100 Words	70.00	50.00	60.42	54.72	97.50	95.83	99.21	97.49
TF/IDF-50 Words	70.00	50.00	60.42	54.72	97.50	95.83	99.21	97.49
Term Frequency	72.50	54.17	60.75	57.27	97.50	95.83	99.21	97.49
Term Frequency-150 Words	75.00	58.33	61.11	59.69	97.50	95.83	99.21	97.49
Term Frequency-100 Words	75.00	58.33	61.11	59.69	97.50	95.83	99.21	97.49
Term Frequency-50 Words	72.50	54.17	60.75	57.27	97.50	95.83	99.21	97.49
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	62.50	37.50	59.52	46.01	50.00	16.67	8.33	11.11
Z Score -150 Words	50.00	16.67	8.33	11.11	45.00	18.33	12.50	14.86
Z Score-100 Words	52.50	20.83	25.21	22.81	55.00	28.33	50.46	36.29
Z Score-50 Words	55.00	25.00	42.11	31.37	57.50	29.17	50.93	37.09
TF/IDF	97.50	95.83	99.21	97.49	92.50	90.83	88.33	89.56
TF/IDF-150 Words	95.00	91.67	98.48	94.95	95.00	91.67	91.67	91.67
TF/IDF-100 Words	95.00	91.67	98.48	94.95	92.50	90.83	88.33	89.56
TF/IDF-50 Words	92.50	87.50	97.83	92.38	95.00	91.67	91.67	91.67
Term Frequency	100.00	100.00	100.00	100.00	95.00	91.67	91.67	91.67
Term Frequency-150 Words	100.00	100.00	100.00	100.00	92.50	90.83	88.33	89.56
Term Frequency-100 Words	100.00	100.00	100.00	100.00	92.50	90.83	88.33	89.56
Term Frequency-50 Words	100.00	100.00	100.00	100.00	92.50	90.83	88.33	89.56

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	70.00	50.00	61.43	55.13
TF/IDF	87.50	79.17	90.33	84.38
Term Frequency	87.50	79.17	90.33	84.38

Punctuation			
Accuracy	Recall	Precision	F-Score
70.00	56.67	76.39	65.07
82.50	74.17	87.67	80.36
80.00	70.00	87.18	77.65

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	37.16	19.45	65.07	11.11
Z Score-150 Words	19.06	11.70	39.15	11.11
Z Score-100 Words	31.15	5.12	36.29	22.81
Z Score-50 Words	36.70	15.52	60.65	17.68
TF/IDF	84.43	13.67	97.49	57.27
TF/IDF-150 Words	85.35	16.34	97.49	57.27
TF/IDF-100 Words	84.18	17.25	97.49	54.72
TF/IDF-50 Words	84.06	17.09	97.49	54.72
Term Frequency	84.74	14.42	100.00	57.27
Term Frequency-150 Words	86.69	16.06	100.00	59.69
Term Frequency-100 Words	86.69	16.06	100.00	59.69
Term Frequency-50 Words	86.08	17.07	100.00	57.27

Table 63: Naive Bayes POS Results for Foundation Corpus [1980's Era] (500 Sentences) - Multiple Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	81.01	71.15	93.18	80.69	89.87	84.62	95.83	89.88
Z Score-150 Words	45.57	51.06	49.23	50.13	74.68	68.03	70.80	69.39
Z Score-100 Words	50.63	50.96	48.86	49.89	91.14	87.84	92.32	90.02
Z Score-50 Words	48.10	60.10	56.52	58.26	96.20	95.53	96.86	96.19
TF/IDF	78.48	69.90	89.35	78.44	98.73	98.08	99.39	98.73
TF/IDF-150 Words	79.75	70.53	91.13	79.52	98.73	98.08	99.39	98.73
TF/IDF-100 Words	77.22	67.98	89.02	77.09	98.73	98.08	99.39	98.73
TF/IDF-50 Words	79.75	71.83	89.42	79.67	97.47	97.45	96.67	97.06
Term Frequency	91.14	86.54	96.28	91.15	100.00	100.00	100.00	100.00
Term Frequency-150 Words	86.08	78.85	94.61	86.01	100.00	100.00	100.00	100.00
Term Frequency-100 Words	88.61	83.99	93.82	88.63	100.00	100.00	100.00	100.00
Term Frequency-50 Words	83.54	77.60	90.82	83.69	97.47	97.45	97.45	97.45
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	98.73	98.08	99.39	98.73	64.56	46.15	39.71	42.69
Z Score -150 Words	54.43	50.24	49.28	49.76	46.84	41.30	41.64	41.47
Z Score-100 Words	48.10	47.12	46.39	46.75	45.57	51.06	46.58	48.72
Z Score-50 Words	40.51	45.96	44.28	45.10	46.84	51.68	55.84	53.68
TF/IDF	97.47	98.75	96.43	97.58	88.61	85.29	86.35	85.82
TF/IDF-150 Words	97.47	98.75	96.43	97.58	91.14	87.84	88.68	88.26
TF/IDF-100 Words	97.47	98.75	96.67	97.70	91.14	89.13	88.93	89.03
TF/IDF-50 Words	96.20	96.83	95.65	96.24	87.34	83.37	84.48	83.92
Term Frequency	100.00	100.00	100.00	100.00	91.14	89.13	90.32	89.72
Term Frequency-150 Words	98.73	99.38	98.21	98.79	91.14	89.13	89.13	89.13
Term Frequency-100 Words	98.73	99.38	98.21	98.79	88.61	85.29	86.63	85.95
Term Frequency-50 Words	97.47	98.75	96.43	97.58	88.61	83.99	86.04	85.00

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	86.08	81.44	80.54	80.99
TF/IDF	87.34	84.66	84.79	84.72
Term Frequency	87.34	84.66	84.79	84.72

Punctuation			
Accuracy	Recall	Precision	F-Score
68.35	72.69	75.14	73.89
93.67	90.38	91.67	91.02
93.67	90.38	91.67	91.02

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	77.81	17.56	98.73	42.69
Z Score-150 Words	52.69	10.25	69.39	41.47
Z Score-100 Words	58.85	18.04	90.02	46.75
Z Score-50 Words	63.31	19.56	96.19	45.10
TF/IDF	89.38	7.20	98.73	78.44
TF/IDF-150 Words	91.02	7.78	98.73	79.52
TF/IDF-100 Words	90.64	8.68	98.73	77.09
TF/IDF-50 Words	89.22	7.58	97.06	79.67
Term Frequency	92.77	5.54	100.00	84.72
Term Frequency-150 Words	93.48	6.03	100.00	86.01
Term Frequency-100 Words	93.34	6.14	100.00	85.95
Term Frequency-50 Words	90.93	6.60	97.58	83.69

Table 64: Naive Bayes POS Results for Foundation Corpus [1980's Era] (1000 Sentences) - Multiple Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	72.22	58.33	66.07	61.96	88.89	83.33	95.45	88.98
Z Score-150 Words	66.67	55.56	56.55	56.05	72.22	58.33	91.07	71.11
Z Score-100 Words	58.33	43.06	43.15	43.10	88.89	83.33	95.45	88.98
Z Score-50 Words	63.89	45.83	51.67	48.58	88.89	83.33	95.45	88.98
TF/IDF	88.89	83.33	95.45	88.98	100.00	100.00	100.00	100.00
TF/IDF-150 Words	86.11	79.17	94.57	86.19	100.00	100.00	100.00	100.00
TF/IDF-100 Words	86.11	79.17	94.57	86.19	100.00	100.00	100.00	100.00
TF/IDF-50 Words	86.11	79.17	94.57	86.19	100.00	100.00	100.00	100.00
Term Frequency	91.67	87.50	96.43	91.75	100.00	100.00	100.00	100.00
Term Frequency-150 Words	91.67	87.50	96.43	91.75	100.00	100.00	100.00	100.00
Term Frequency-100 Words	88.89	83.33	95.45	88.98	100.00	100.00	100.00	100.00
Term Frequency-50 Words	83.33	75.00	93.75	83.33	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	97.22	95.83	98.68	97.23	61.11	41.67	39.06	40.32
Z Score -150 Words	69.44	54.17	79.17	64.33	50.00	38.89	45.33	41.86
Z Score-100 Words	72.22	58.33	76.75	66.28	66.67	50.00	70.83	58.62
Z Score-50 Words	66.67	52.78	54.49	53.62	66.67	50.00	72.92	59.32
TF/IDF	100.00	100.00	100.00	100.00	91.67	87.50	89.52	88.50
TF/IDF-150 Words	100.00	100.00	100.00	100.00	97.22	95.83	98.68	97.23
TF/IDF-100 Words	100.00	100.00	100.00	100.00	91.67	87.50	89.52	88.50
TF/IDF-50 Words	100.00	100.00	100.00	100.00	94.44	91.67	94.52	93.07
Term Frequency	100.00	100.00	100.00	100.00	94.44	91.67	94.52	93.07
Term Frequency-150 Words	100.00	100.00	100.00	100.00	91.67	87.50	89.52	88.50
Term Frequency-100 Words	100.00	100.00	100.00	100.00	91.67	87.50	89.52	88.50
Term Frequency-50 Words	97.22	98.61	96.43	97.51	94.44	91.67	94.52	93.07

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	88.89	83.33	83.10	83.21
TF/IDF	94.44	91.67	95.11	93.36
Term Frequency	88.89	83.33	86.18	84.73

Punctuation			
Accuracy	Recall	Precision	F-Score
83.33	83.33	82.86	83.09
100.00	100.00	100.00	100.00
100.00	100.00	100.00	100.00

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	75.80	19.11	97.23	40.32
Z Score-150 Words	58.34	10.91	71.11	41.86
Z Score-100 Words	64.25	16.54	88.98	43.10
Z Score-50 Words	62.62	15.68	88.98	48.58
TF/IDF	95.14	5.10	100.00	88.50
TF/IDF-150 Words	95.86	5.69	100.00	86.19
TF/IDF-100 Words	93.67	6.38	100.00	86.19
TF/IDF-50 Words	94.82	5.73	100.00	86.19
Term Frequency	94.93	5.70	100.00	84.73
Term Frequency-150 Words	95.06	5.07	100.00	88.50
Term Frequency-100 Words	94.37	5.63	100.00	88.50
Term Frequency-50 Words	93.48	6.36	100.00	83.33

Table 65: Naive Bayes POS Results for Foundation Corpus [1980's Era] (1500 Sentences) - Multiple Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	66.67	50.00	40.00	44.44	87.50	81.25	95.00	87.59
Z Score-150 Words	58.33	37.50	65.00	47.56	50.00	25.00	12.50	16.67
Z Score-100 Words	62.50	43.75	64.29	52.07	50.00	25.00	12.50	16.67
Z Score-50 Words	50.00	25.00	12.50	16.67	58.33	37.50	63.64	47.19
TF/IDF	95.83	93.75	98.08	95.87	95.83	93.75	95.00	94.37
TF/IDF-150 Words	91.67	87.50	96.43	91.75	100.00	100.00	100.00	100.00
TF/IDF-100 Words	91.67	87.50	96.43	91.75	95.83	93.75	95.00	94.37
TF/IDF-50 Words	91.67	87.50	96.43	91.75	95.83	93.75	95.00	94.37
Term Frequency	95.83	93.75	98.08	95.87	100.00	100.00	100.00	100.00
Term Frequency-150 Words	95.83	93.75	98.08	95.87	95.83	93.75	95.00	94.37
Term Frequency-100 Words	95.83	93.75	98.08	95.87	95.83	93.75	95.00	94.37
Term Frequency-50 Words	87.50	81.25	95.00	87.59	95.83	93.75	95.00	94.37
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	100.00	100.00	100.00	100.00	62.50	43.75	39.29	41.40
Z Score -150 Words	62.50	47.92	59.03	52.90	66.67	50.00	59.54	54.35
Z Score-100 Words	54.17	35.42	39.47	37.34	54.17	35.42	45.72	39.92
Z Score-50 Words	50.00	25.00	12.50	16.67	54.17	31.25	26.79	28.85
TF/IDF	100.00	100.00	100.00	100.00	91.67	87.50	87.50	87.50
TF/IDF-150 Words	100.00	100.00	100.00	100.00	91.67	87.50	87.50	87.50
TF/IDF-100 Words	100.00	100.00	100.00	100.00	91.67	87.50	87.50	87.50
TF/IDF-50 Words	100.00	100.00	100.00	100.00	91.67	87.50	87.50	87.50
Term Frequency	100.00	100.00	100.00	100.00	91.67	87.50	87.50	87.50
Term Frequency-150 Words	100.00	100.00	100.00	100.00	91.67	87.50	87.50	87.50
Term Frequency-100 Words	100.00	100.00	100.00	100.00	91.67	87.50	87.50	87.50
Term Frequency-50 Words	100.00	100.00	100.00	100.00	91.67	87.50	87.50	87.50

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	75.00	62.50	67.92	65.10
TF/IDF	87.50	81.25	88.08	84.53
Term Frequency	87.50	85.42	87.92	86.65

Punctuation			
Accuracy	Recall	Precision	F-Score
75.00	70.83	81.55	75.81
91.67	87.50	93.08	90.20
95.83	93.75	98.08	95.87

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	69.06	21.34	100.00	41.40
Z Score-150 Words	42.87	15.34	54.35	16.67
Z Score-100 Words	36.50	12.73	52.07	16.67
Z Score-50 Words	27.34	12.49	47.19	16.67
TF/IDF	92.08	5.23	100.00	84.53
TF/IDF-150 Words	94.81	5.40	100.00	87.50
TF/IDF-100 Words	93.40	4.53	100.00	87.50
TF/IDF-50 Words	93.40	4.53	100.00	87.50
Term Frequency	94.31	5.39	100.00	86.65
Term Frequency-150 Words	94.43	4.50	100.00	87.50
Term Frequency-100 Words	94.43	4.50	100.00	87.50
Term Frequency-50 Words	92.36	5.22	100.00	87.50

Table 66: Naive Bayes POS Results for Foundation Corpus [1950's Era] (500 Sentences) - Multiple Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	75.00	62.50	91.67	74.33	79.17	68.75	92.65	78.93
Z Score-150 Words	56.25	40.62	49.24	44.52	70.83	60.42	63.67	62.00
Z Score-100 Words	50.00	33.33	31.10	32.18	87.50	83.33	88.78	85.97
Z Score-50 Words	47.92	44.79	46.77	45.76	85.42	78.12	90.83	84.00
TF/IDF	89.58	84.38	95.69	89.68	100.00	100.00	100.00	100.00
TF/IDF-150 Words	91.67	87.50	96.43	91.75	100.00	100.00	100.00	100.00
TF/IDF-100 Words	91.67	87.50	96.43	91.75	100.00	100.00	100.00	100.00
TF/IDF-50 Words	89.58	84.38	95.69	89.68	100.00	100.00	100.00	100.00
Term Frequency	93.75	90.62	97.22	93.80	100.00	100.00	100.00	100.00
Term Frequency-150 Words	95.83	93.75	98.08	95.87	100.00	100.00	100.00	100.00
Term Frequency-100 Words	95.83	93.75	98.08	95.87	100.00	100.00	100.00	100.00
Term Frequency-50 Words	93.75	90.62	97.22	93.80	97.92	96.88	99.00	97.93
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	95.83	93.75	98.08	95.87	56.25	34.38	38.33	36.25
Z Score -150 Words	64.58	48.96	55.97	52.23	54.17	41.67	54.13	47.09
Z Score-100 Words	60.42	44.79	44.80	44.79	54.17	45.83	40.74	43.14
Z Score-50 Words	58.33	52.08	53.09	52.58	50.00	47.92	54.76	51.11
TF/IDF	97.92	98.96	97.22	98.08	95.83	93.75	95.88	94.80
TF/IDF-150 Words	100.00	100.00	100.00	100.00	93.75	90.62	92.30	91.45
TF/IDF-100 Words	100.00	100.00	100.00	100.00	97.92	96.88	99.00	97.93
TF/IDF-50 Words	97.92	98.96	97.22	98.08	91.67	87.50	89.28	88.38
Term Frequency	100.00	100.00	100.00	100.00	97.92	96.88	99.00	97.93
Term Frequency-150 Words	100.00	100.00	100.00	100.00	97.92	96.88	97.22	97.05
Term Frequency-100 Words	100.00	100.00	100.00	100.00	97.92	96.88	97.22	97.05
Term Frequency-50 Words	97.92	98.96	97.22	98.08	89.58	84.38	86.50	85.43

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	77.08	67.71	70.67	69.16
TF/IDF	91.67	87.50	89.87	88.67
Term Frequency	89.58	84.38	87.81	86.06

Punctuation			
Accuracy	Recall	Precision	F-Score
79.17	70.83	78.58	74.50
95.83	93.75	98.08	95.87
95.83	93.75	95.88	94.80

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	71.51	17.86	95.87	36.25
Z Score-150 Words	51.46	6.69	62.00	44.52
Z Score-100 Words	51.52	20.47	85.97	32.18
Z Score-50 Words	58.36	15.02	84.00	45.76
TF/IDF	94.52	4.13	100.00	88.67
TF/IDF-150 Words	95.80	4.20	100.00	91.45
TF/IDF-100 Words	97.42	3.38	100.00	91.75
TF/IDF-50 Words	94.04	5.07	100.00	88.38
Term Frequency	95.43	4.81	100.00	86.06
Term Frequency-150 Words	98.23	1.82	100.00	95.87
Term Frequency-100 Words	98.23	1.82	100.00	95.87
Term Frequency-50 Words	93.81	5.14	98.08	85.43

Table 67: Naive Bayes POS Results for Foundation Corpus [1950's Era] (1000 Sentences) - Multiple Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	66.67	50.00	40.00	44.44	66.67	50.00	65.00	56.52
Z Score-150 Words	58.33	37.50	63.64	47.19	50.00	25.00	12.50	16.67
Z Score-100 Words	58.33	41.67	51.97	46.25	54.17	31.25	38.04	34.31
Z Score-50 Words	54.17	31.25	38.04	34.31	54.17	31.25	38.04	34.31
TF/IDF	91.67	87.50	96.43	91.75	100.00	100.00	100.00	100.00
TF/IDF-150 Words	91.67	87.50	96.43	91.75	100.00	100.00	100.00	100.00
TF/IDF-100 Words	91.67	87.50	96.43	91.75	100.00	100.00	100.00	100.00
TF/IDF-50 Words	91.67	87.50	96.43	91.75	100.00	100.00	100.00	100.00
Term Frequency	91.67	87.50	96.43	91.75	100.00	100.00	100.00	100.00
Term Frequency-150 Words	91.67	87.50	96.43	91.75	100.00	100.00	100.00	100.00
Term Frequency-100 Words	95.83	93.75	98.08	95.87	100.00	100.00	100.00	100.00
Term Frequency-50 Words	91.67	87.50	96.43	91.75	100.00	100.00	100.00	100.00
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	91.67	87.50	96.43	91.75	50.00	25.00	12.50	16.67
Z Score -150 Words	62.50	43.75	45.83	44.77	54.17	31.25	38.04	34.31
Z Score-100 Words	54.17	31.25	38.04	34.31	58.33	37.50	51.79	43.50
Z Score-50 Words	54.17	31.25	38.04	34.31	54.17	31.25	38.04	34.31
TF/IDF	100.00	100.00	100.00	100.00	95.83	93.75	95.00	94.37
TF/IDF-150 Words	95.83	93.75	98.08	95.87	100.00	100.00	100.00	100.00
TF/IDF-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TF/IDF-50 Words	95.83	93.75	95.00	94.37	95.83	93.75	95.00	94.37
Term Frequency	100.00	100.00	100.00	100.00	95.83	93.75	95.00	94.37
Term Frequency-150 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-100 Words	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Term Frequency-50 Words	100.00	100.00	100.00	100.00	95.83	93.75	95.00	94.37

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	79.17	68.75	85.42	76.18
TF/IDF	79.17	68.75	73.51	71.05
Term Frequency	91.67	87.50	91.83	89.61

Punctuation			
Accuracy	Recall	Precision	F-Score
70.83	56.25	60.42	58.26
83.33	75.00	93.75	83.33
83.33	75.00	93.75	83.33

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	57.30	23.67	91.75	16.67
Z Score-150 Words	35.73	12.03	47.19	16.67
Z Score-100 Words	39.59	5.37	46.25	34.31
Z Score-50 Words	34.31	0.00	34.31	34.31
TF/IDF	90.08	10.21	100.00	71.05
TF/IDF-150 Words	96.90	3.42	100.00	91.75
TF/IDF-100 Words	97.94	3.57	100.00	91.75
TF/IDF-50 Words	95.12	3.01	100.00	91.75
Term Frequency	93.18	5.86	100.00	83.33
Term Frequency-150 Words	97.94	3.57	100.00	91.75
Term Frequency-100 Words	98.97	1.79	100.00	95.87
Term Frequency-50 Words	96.53	3.59	100.00	91.75

Table 68: Naive Bayes POS Results for Foundation Corpus [1950's Era] (1500 Sentences) - Multiple Authors

	Adjectives				All Words			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	50.00	25.00	12.50	16.67	50.00	25.00	12.50	16.67
Z Score-150 Words	50.00	25.00	12.50	16.67	50.00	25.00	12.50	16.67
Z Score-100 Words	50.00	25.00	12.50	16.67	50.00	25.00	12.50	16.67
Z Score-50 Words	50.00	25.00	12.50	16.67	50.00	25.00	12.50	16.67
TF/IDF	83.33	75.00	68.75	71.74	83.33	75.00	63.10	68.54
TF/IDF-150 Words	83.33	75.00	68.75	71.74	83.33	75.00	63.10	68.54
TF/IDF-100 Words	83.33	75.00	68.75	71.74	83.33	75.00	58.33	65.62
TF/IDF-50 Words	83.33	75.00	68.75	71.74	83.33	75.00	63.10	68.54
Term Frequency	83.33	75.00	68.75	71.74	83.33	75.00	63.10	68.54
Term Frequency-150 Words	83.33	75.00	68.75	71.74	83.33	75.00	63.10	68.54
Term Frequency-100 Words	83.33	75.00	68.75	71.74	83.33	75.00	58.33	65.62
Term Frequency-50 Words	83.33	75.00	68.75	71.74	83.33	75.00	63.10	68.54
	Nouns				Verbs			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	50.00	25.00	12.50	16.67	50.00	25.00	12.50	16.67
Z Score -150 Words	50.00	25.00	12.50	16.67	50.00	25.00	12.50	16.67
Z Score-100 Words	50.00	25.00	15.00	18.75	50.00	25.00	12.50	16.67
Z Score-50 Words	50.00	25.00	12.50	16.67	50.00	25.00	12.50	16.67
TF/IDF	83.33	75.00	63.10	68.54	66.67	50.00	50.00	50.00
TF/IDF-150 Words	75.00	62.50	50.60	55.92	66.67	50.00	50.00	50.00
TF/IDF-100 Words	75.00	62.50	58.93	60.66	83.33	75.00	93.75	83.33
TF/IDF-50 Words	75.00	62.50	58.93	60.66	83.33	75.00	93.75	83.33
Term Frequency	83.33	75.00	63.10	68.54	91.67	87.50	91.67	89.54
Term Frequency-150 Words	83.33	75.00	63.10	68.54	83.33	75.00	93.75	83.33
Term Frequency-100 Words	83.33	75.00	62.50	68.18	83.33	75.00	83.93	79.21
Term Frequency-50 Words	75.00	62.50	58.93	60.66	83.33	75.00	93.75	83.33

	Articles			
	Accuracy	Recall	Precision	F-Score
Z Score	50.00	25.00	15.00	18.75
TF/IDF	66.67	50.00	54.76	52.27
Term Frequency	66.67	50.00	42.26	45.81

Punctuation			
Accuracy	Recall	Precision	F-Score
50.00	25.00	16.67	20.00
50.00	25.00	12.50	16.67
50.00	25.00	12.50	16.67

	Corpus Statistics			
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score
Z Score	17.57	1.33	20.00	16.67
Z Score-150 Words	16.67	0.00	16.67	16.67
Z Score-100 Words	17.19	0.90	18.75	16.67
Z Score-50 Words	16.67	0.00	16.67	16.67
TF/IDF	54.63	18.92	71.74	16.67
TF/IDF-150 Words	61.55	8.91	71.74	50.00
TF/IDF-100 Words	70.34	8.47	83.33	60.66
TF/IDF-50 Words	71.07	8.15	83.33	60.66
Term Frequency	60.14	23.22	89.54	16.67
Term Frequency-150 Words	73.04	6.09	83.33	68.54
Term Frequency-100 Words	71.19	5.12	79.21	65.62
Term Frequency-50 Words	71.07	8.15	83.33	60.66

Table 69: Part-Of-Speech Statistics for Naive Bayes - Multiple Authors

	Adjectives				All Words			
	Average F-Score	Stand. Dev. F-Score	Max F-Score	Min F-Score	Average F-Score	Stand. Dev. F-Score	Max F-Score	Min F-Score
Z Score	48.27	20.16	80.69	16.67	56.88	28.55	89.88	16.67
Z Score-150 Words	42.55	11.03	56.05	16.67	40.46	24.06	71.11	11.11
Z Score-100 Words	40.28	10.44	52.07	16.67	57.33	30.07	90.02	16.67
Z Score-50 Words	36.88	15.82	58.26	16.67	66.97	27.05	96.19	16.67
TF/IDF	75.38	16.63	95.87	51.57	95.37	9.64	100.00	68.54
TF/IDF-150 Words	75.05	16.02	91.75	50.69	95.90	9.71	100.00	68.54
TF/IDF-100 Words	74.98	15.68	91.75	52.41	94.67	10.41	100.00	65.62
TF/IDF-50 Words	74.29	16.47	91.75	50.69	94.99	9.53	100.00	68.54
Term Frequency	80.70	14.46	95.87	57.27	96.23	9.82	100.00	68.54
Term Frequency-150 Words	79.35	15.21	95.87	55.90	95.51	9.70	100.00	68.54
Term Frequency-100 Words	80.29	15.16	95.87	55.90	95.09	10.57	100.00	65.62
Term Frequency-50 Words	75.62	15.00	93.80	52.41	94.73	9.42	100.00	68.54
	Nouns				Verbs			
	Average F-Score	Stan. Dev. F-Score	Max F-Score	Min F-Score	Average F-Score	Stan. Dev. F-Score	Max F-Score	Min F-Score
Z Score	75.45	27.60	100.00	16.67	26.36	12.77	42.69	11.11
Z Score -150 Words	41.04	16.34	64.33	11.11	34.44	13.10	54.35	14.86
Z Score-100 Words	40.22	13.46	66.28	18.75	40.79	13.42	58.62	16.67
Z Score-50 Words	36.74	12.84	53.62	16.67	42.77	15.50	68.57	16.67
TF/IDF	95.00	9.48	100.00	68.54	85.74	12.93	94.80	50.00
TF/IDF-150 Words	92.79	13.26	100.00	55.92	86.78	13.72	100.00	50.00
TF/IDF-100 Words	93.71	11.99	100.00	60.66	89.77	5.25	100.00	83.33
TF/IDF-50 Words	92.03	11.53	100.00	60.66	87.81	4.19	94.37	82.00
Term Frequency	96.42	9.86	100.00	68.54	91.47	3.01	97.93	87.50
Term Frequency-150 Words	96.12	9.77	100.00	68.54	90.33	4.78	100.00	83.33
Term Frequency-100 Words	95.81	9.82	100.00	68.18	88.84	5.93	100.00	79.21

Term Frequency-50 Words	93.88	11.89	100.00	60.66	87.27	4.01	94.37	82.14
-------------------------	-------	-------	--------	-------	-------	------	-------	-------

	Articles			
	Average F-Score	Stand. Dev. F-Score	Max F-Score	Min F-Score
Z Score	67.66	19.38	83.21	18.75
TF/IDF	80.74	11.50	93.36	52.27
Term Frequency	85.51	12.80	89.61	45.81

	Punctuation			
	Average F-Score	Stand. Dev. F-Score	Max F-Score	Min F-Score
Z Score	67.00	17.96	83.09	20.00
TF/IDF	81.90	23.75	100	16.67
Term Frequency	82.11	24.02	100	16.67

D.5 KNN Sentiment Frequency Results - Two Authors

Table 70: KNN Sentiment Frequency Results for Entire Corpus (500 Sentence) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	66.67	65.53	70.45	67.90	57.14	57.27	57.27	57.27
TF/IDF	65.08	64.39	65.80	65.09	61.90	60.61	65.31	62.87
Term Frequency	65.87	64.85	68.46	66.61	62.70	61.44	66.10	63.68
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	59.52	59.85	59.99	59.92				
TF/IDF	61.90	61.97	61.95	61.96				
Term Frequency	59.52	58.11	62.71	60.32				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	61.70	4.52	67.90	57.27				
TF/IDF	63.31	1.31	65.09	61.96				
Term Frequency	59.52	58.11	62.71	60.32				

Table 71: KNN Sentiment Frequency Results for Entire Corpus (1000 Sentence) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	74.60	74.55	74.55	74.55	66.67	66.52	66.58	66.55
TF/IDF	71.43	71.43	71.43	71.43	77.78	77.12	79.77	78.42
Term Frequency	73.02	72.88	72.97	72.92	74.60	74.39	74.64	74.51
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	55.56	55.91	56.02	55.96				
TF/IDF	71.43	70.76	72.78	71.76				
Term Frequency	71.43	70.45	75.00	72.65				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	65.69	7.61	74.55	55.96				
TF/IDF	73.87	3.22	78.42	71.43				
Term Frequency	73.36	0.82	74.51	72.65				

Table 72: KNN Sentiment Frequency Results for Entire Corpus (1500 Sentence) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	76.19	75.91	76.39	76.15	69.05	67.95	73.17	70.46
TF/IDF	78.57	78.86	79.06	78.96	71.43	71.59	71.59	71.59
Term Frequency	71.43	71.82	72.22	72.02	73.81	73.86	73.81	73.83
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	57.14	56.59	56.97	56.78				
TF/IDF	66.67	65.45	71.25	68.23				
Term Frequency	69.05	67.95	73.17	70.46				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	67.80	8.13	76.15	56.78				
TF/IDF	72.93	4.48	78.96	68.23				
Term Frequency	72.11	1.38	73.83	70.46				

Table 73: KNN Sentiment Frequency Results for Foundation Corpus [1980's Era] (500 Sentence) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	71.79	71.79	71.79	71.79	65.38	65.38	73.59	69.24
TF/IDF	67.95	67.95	68.06	68.00	64.10	64.10	65.58	64.83
Term Frequency	67.95	67.95	67.96	67.95	64.10	64.10	69.13	66.52
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	55.13	55.13	55.42	55.27				
TF/IDF	60.26	60.26	60.60	60.43				
Term Frequency	66.67	66.67	66.67	66.67				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	65.44	7.26	71.79	55.27				
TF/IDF	64.42	3.11	68.00	60.43				
Term Frequency	67.05	0.64	67.95	66.52				

Table 74: KNN Sentiment Frequency Results for Foundation Corpus [1980's Era] (1000 Sentence) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	75.68	75.73	75.73	75.73	72.97	72.51	75.61	74.03
TF/IDF	72.97	72.95	72.95	72.95	67.57	67.54	67.54	67.54
Term Frequency	75.68	75.58	75.74	75.66	70.27	70.32	70.32	70.32
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	70.27	69.88	71.79	70.82				
TF/IDF	72.97	73.25	74.09	73.67				
Term Frequency	78.38	78.51	78.68	78.59				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	73.53	2.03	75.73	70.82				
TF/IDF	71.39	2.74	73.67	67.54				
Term Frequency	74.86	3.43	78.59	70.32				

Table 75: KNN Sentiment Frequency Results for Foundation Corpus [1980's Era] (1500 Sentence) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	75.00	75.00	75.00	75.00	75.00	75.00	75.00	75.00
TF/IDF	58.33	58.33	59.38	58.85	50.00	50.00	50.00	50.00
Term Frequency	66.67	66.67	66.67	66.67	62.50	62.50	62.59	62.54
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	66.67	66.67	80.00	72.73				
TF/IDF	54.17	54.17	54.20	54.18				
Term Frequency	70.83	70.83	70.98	70.90				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	74.24	1.07	75.00	72.73				
TF/IDF	54.35	3.61	58.85	50.00				
Term Frequency	66.71	3.41	70.90	62.54				

Table 76: KNN Sentiment Frequency Results for Foundation Corpus [1950's Era] (500 Sentence) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	73.47	73.00	79.49	76.11	63.27	63.00	63.98	63.49
TF/IDF	67.35	67.50	67.86	67.68	63.27	63.25	63.25	63.25
Term Frequency	65.31	65.50	66.03	65.76	73.47	43.42	73.49	54.59
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	57.14	56.25	77.17	65.07				
TF/IDF	67.35	67.42	67.47	67.44				
Term Frequency	69.39	69.33	69.40	69.36				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	68.22	5.61	76.11	63.49				
TF/IDF	66.12	2.04	67.68	63.25				
Term Frequency	63.24	6.29	69.36	54.59				

Table 77: KNN Sentiment Frequency Results for Foundation Corpus [1950's Era] (1000 Sentence) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	75.00	75.00	75.71	75.35	75.00	75.00	75.00	75.00
TF/IDF	79.17	79.17	81.11	80.13	79.17	79.17	79.37	79.27
Term Frequency	79.17	79.17	85.29	82.12	79.17	79.17	79.37	79.27
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	62.50	62.50	68.95	65.57				
TF/IDF	66.67	66.67	72.22	69.33				
Term Frequency	70.83	70.83	75.21	72.95				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	71.97	4.53	75.35	65.57				
TF/IDF	76.24	4.90	80.13	69.33				
Term Frequency	78.11	3.83	82.12	72.95				

Table 78: KNN Sentiment Frequency Results for Foundation Corpus [1950's Era] (1500 Sentence) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	83.33	83.33	83.33	83.33	58.33	58.33	61.11	59.69
TF/IDF	66.67	66.67	68.75	67.69	50.00	50.00	50.00	50.00
Term Frequency	75.00	75.00	75.71	75.35	66.67	66.67	80.00	72.73
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	75.00	75.00	83.33	78.95				
TF/IDF	58.33	58.33	58.57	58.45				
Term Frequency	58.33	58.33	58.57	58.45				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	73.99	10.27	83.33	59.69				
TF/IDF	58.71	7.23	67.69	50.00				
Term Frequency	68.84	7.43	75.35	58.45				

Table 79: KNN Sentiment Frequency Results for Galactic Empire Corpus (500 Sentence) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	66.67	66.67	68.15	67.40	61.90	61.90	62.96	62.43
TF/IDF	71.43	71.43	75.07	73.20	69.05	69.05	69.09	69.07
Term Frequency	76.19	76.19	78.52	77.34	69.05	69.05	69.09	69.07
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	59.52	59.52	60.71	60.11				
TF/IDF	71.43	71.43	75.07	73.20				
Term Frequency	66.67	66.67	66.67	66.67				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	63.31	3.04	67.40	60.11				
TF/IDF	71.83	1.95	73.20	69.07				
Term Frequency	71.03	4.57	77.34	66.67				

Table 80: KNN Sentiment Frequency Results for Galactic Empire Corpus (1000 Sentence) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	61.11	61.11	62.50	61.80	66.67	66.67	67.53	67.10
TF/IDF	77.78	77.78	77.78	77.78	72.22	72.22	72.50	72.36
Term Frequency	72.22	72.22	72.50	72.36	66.67	66.67	66.67	66.67
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	72.22	72.22	72.50	72.36				
TF/IDF	72.22	72.22	72.50	72.36				
Term Frequency	66.67	66.67	67.53	67.10				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	67.08	4.31	72.36	61.80				
TF/IDF	74.17	2.56	77.78	72.36				
Term Frequency	68.71	2.59	72.36	66.67				

Table 81: KNN Sentiment Frequency Results for Galactic Empire Corpus (1500 Sentence) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00
TF/IDF	66.67	66.67	66.67	66.67	58.33	58.33	58.57	58.45
Term Frequency	58.33	58.33	58.57	58.45	58.33	58.33	58.57	58.45
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	66.67	66.67	80.00	72.73				
TF/IDF	58.33	58.33	58.57	58.45				
Term Frequency	75.00	75.00	75.71	75.35				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	57.58	10.71	72.73	50.00				
TF/IDF	61.19	3.88	66.67	58.45				
Term Frequency	64.08	7.97	75.35	58.45				

Table 82: KNN Sentiment Frequency Results for Robot Corpus (500 Sentence) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	94.64	94.64	94.70	94.67	60.71	60.71	65.91	63.20
TF/IDF	80.36	80.36	80.40	80.38	76.79	76.79	77.10	76.94
Term Frequency	80.36	80.36	80.40	80.38	76.79	76.79	77.10	76.94
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	51.79	51.79	75.45	61.42				
TF/IDF	73.21	73.21	73.48	73.34				
Term Frequency	75.00	75.00	75.13	75.06				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	73.10	15.27	94.67	61.42				
TF/IDF	76.89	2.87	80.38	73.34				
Term Frequency	77.46	2.20	80.38	75.06				

Table 83: KNN Sentiment Frequency Results for Robot Corpus (1000 Sentence) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	91.67	91.67	91.67	91.67	75.00	75.00	83.33	78.95
TF/IDF	87.50	87.50	90.00	88.73	83.33	83.33	84.29	83.81
Term Frequency	87.50	87.50	90.00	88.73	79.17	79.17	85.29	82.12
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	58.33	58.33	77.27	66.48				
TF/IDF	83.33	83.33	84.29	83.81				
Term Frequency	87.50	87.50	90.00	88.73				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	79.03	10.29	91.67	66.48				
TF/IDF	85.45	2.32	88.73	83.81				
Term Frequency	86.53	3.12	88.73	82.12				

Table 84: KNN Sentiment Frequency Results for Robot Corpus (1500 Sentence) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	93.75	93.75	94.44	94.09	93.75	93.75	94.44	94.09
TF/IDF	87.50	87.50	87.50	87.50	81.25	81.25	81.75	81.50
Term Frequency	87.50	87.50	87.50	87.50	81.25	81.25	81.75	81.50
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	56.25	56.25	76.67	64.89				
TF/IDF	81.25	81.25	81.75	81.50				
Term Frequency	75.00	75.00	76.67	75.83				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	84.36	13.77	94.09	64.89				
TF/IDF	83.50	2.83	87.50	81.50				
Term Frequency	81.61	4.77	87.50	75.83				

Table 85: Sentiment Frequency Statistics for KNN - Two Authors

	1 Decimal				2 Decimals			
	Average F-Score	Stand. Dev. F-Score	Max F-Score	Min F-Score	Average F-Score	Stand. Dev. F-Score	Max F-Score	Min F-Score
Z Score	75.70	11.58	94.67	50.00	68.43	10.08	94.09	50.00
TF/IDF	73.67	8.17	88.73	58.85	68.66	10.24	83.81	50.00
Term Frequency	73.99	8.10	88.73	58.45	70.18	7.94	82.12	54.59

	3 Decimals			
	Average F-Score	Stand. Dev. F-Score	Max F-Score	Min F-Score
Z Score	65.27	6.88	78.95	55.27
TF/IDF	68.54	8.24	83.81	54.18
Term Frequency	71.27	7.13	88.73	58.45

D.6 KNN POS Results - Multiple Authors

Table 86: KNN Sentiment Frequency Results for Entire Corpus(500 Sentences) - Multiple Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	57.94	37.75	45.16	41.12	53.97	22.85	19.31	20.93
TF/IDF	57.94	29.8	46.45	36.31	53.97	23.99	26.21	25.05
Term Frequency	56.35	30.43	52.4	38.50	55.56	27.9	38.13	32.22
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	53.17	18.06	17.2	17.62				
TF/IDF	55.56	30.18	26.67	28.32				
Term Frequency	53.97	25.13	24.68	24.90				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	26.56	10.39	41.12	17.62				
TF/IDF	29.89	4.73	36.31	25.05				
Term Frequency	31.88	5.56	38.50	24.90				

Table 87: KNN Sentiment Frequency Results for Entire Corpus(1000 Sentences) - Multiple Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	60.32	37.37	52.30	43.59	55.56	29.04	19.51	23.34
TF/IDF	68.25	48.99	64.33	55.62	63.49	42.93	49.11	45.81
Term Frequency	65.08	45.71	56.70	50.62	63.49	42.93	40.39	41.62
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	53.28	16.67	8.73	11.46				
TF/IDF	58.73	34.60	27.86	30.87				
Term Frequency	60.32	37.37	44.73	40.72				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	26.13	13.27	43.59	11.46				
TF/IDF	44.10	10.18	55.62	30.87				
Term Frequency	44.32	4.47	50.62	40.72				

Table 88: KNN Sentiment Frequency Results for Entire Corpus(1500 Sentences) - Multiple Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	59.52	32.58	37.78	34.99	57.14	31.82	37.08	34.25
TF/IDF	71.43	67.05	59.50	63.05	59.52	39.39	28.78	33.26
Term Frequency	66.67	48.48	45.66	47.03	59.52	32.58	18.63	23.70
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	52.38	16.67	8.73	11.46				
TF/IDF	57.14	35.23	24.98	29.23				
Term Frequency	61.90	43.56	30.08	35.59				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	26.90	10.92	34.99	11.46				
TF/IDF	41.85	15.08	63.05	29.23				
Term Frequency	35.44	9.52	47.03	23.70				

Table 89: KNN Sentiment Frequency Results for Foundation Corpus [1980's Era](500 Sentences) - Multiple Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	51.28	26.92	38.00	31.51	52.56	28.85	46.51	35.61
TF/IDF	53.85	30.77	54.84	39.42	50.00	25.00	12.66	16.81
Term Frequency	52.56	28.85	50.50	36.72	52.56	28.85	25.68	27.17
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	51.28	26.92	37.66	31.40				
TF/IDF	50.00	25.00	12.50	16.67				
Term Frequency	50.00	25.00	12.50	16.67				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	32.84	1.96	35.61	31.40				
TF/IDF	24.30	10.69	39.42	16.67				
Term Frequency	26.85	8.19	36.72	16.67				

Table 90: KNN Sentiment Frequency Results for Foundation Corpus [1980's Era](1000 Sentences) - Multiple Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	56.76	33.33	63.57	43.73	56.76	33.33	38.57	35.76
TF/IDF	56.76	41.89	35.71	38.55	51.35	25.00	12.84	16.97
Term Frequency	56.76	47.59	45.24	46.39	56.76	33.33	38.57	35.76
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	51.35	25.00	12.84	16.97				
TF/IDF	51.35	42.11	37.19	39.50				
Term Frequency	54.05	34.87	26.21	29.93				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	32.15	11.22	43.73	16.97				
TF/IDF	31.67	10.41	39.50	16.97				
Term Frequency	37.36	6.81	46.39	29.93				

Table 91: KNN Sentiment Frequency Results for Foundation Corpus [1980's Era](1500 Sentences) - Multiple Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	54.17	31.25	38.04	34.31	50.00	29.17	20.28	23.93
TF/IDF	54.17	31.25	38.04	34.31	50.00	25.00	12.50	16.67
Term Frequency	54.17	31.25	38.04	34.31	50.00	25.00	12.50	16.67
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	54.17	35.42	23.61	28.33				
TF/IDF	50.00	25.00	12.50	16.67				
Term Frequency	50.00	25.00	12.50	16.67				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	28.86	4.26	34.31	23.93				
TF/IDF	22.55	8.32	34.31	16.67				
Term Frequency	22.55	8.32	34.31	16.67				

Table 92: KNN Sentiment Frequency Results for Foundation Corpus [1950's Era](500 Sentences) - Multiple Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	53.06	34.50	54.06	42.12	55.10	37.62	51.60	43.51
TF/IDF	53.06	28.12	38.02	32.33	53.06	28.12	38.02	32.33
Term Frequency	55.10	31.25	38.89	34.65	55.10	31.25	30.25	30.74
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	51.02	25.00	12.76	16.90				
TF/IDF	51.02	25.00	12.76	16.90				
Term Frequency	53.06	28.12	38.02	32.33				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	34.18	12.23	43.51	16.90				
TF/IDF	27.18	7.28	32.33	16.90				
Term Frequency	32.57	1.61	34.65	30.74				

Table 93: KNN Sentiment Frequency Results for Foundation Corpus [1950's Era](1000 Sentences) - Multiple Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	70.83	60.42	71.43	65.47	58.33	37.50	63.64	47.19
TF/IDF	66.67	50.00	42.50	45.95	50.00	25.00	12.50	16.67
Term Frequency	62.50	43.75	39.29	41.40	50.00	33.33	31.73	32.51
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	54.17	31.25	38.04	34.31				
TF/IDF	54.17	31.25	26.79	28.85				
Term Frequency	54.17	35.42	24.47	28.94				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	48.99	12.78	65.47	34.31				
TF/IDF	30.49	12.01	45.95	16.67				
Term Frequency	34.28	5.24	41.40	28.94				

Table 94: KNN Sentiment Frequency Results for Foundation Corpus [1950's Era](1500 Sentences) - Multiple Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	50.00	25.00	16.67	20.00	66.67	58.33	65.62	61.76
TF/IDF	50.00	25.00	12.50	16.67	50.00	25.00	12.50	16.67
Term Frequency	50.00	25.00	12.50	16.67	50.00	25.00	12.50	16.67
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	66.67	50.00	54.17	52.00				
TF/IDF	50.00	25.00	12.50	16.67				
Term Frequency	50.00	25.00	12.50	16.67				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	44.59	17.84	61.76	20.00				
TF/IDF	16.67	0.00	16.67	16.67				
Term Frequency	16.67	0.00	16.67	16.67				

Table 95: Sentiment Frequency Statistics for KNN - Multiple Authors

	1 Decimal				2 Decimals			
	Average F-Score	Stand. Dev. F-Score	Max F-Score	Min F-Score	Average F-Score	Stand. Dev. F-Score	Max F-Score	Min F-Score
Z Score	39.65	11.59	65.47	20.00	36.25	12.39	61.76	20.93
TF/IDF	40.25	12.74	63.05	16.67	24.47	9.96	45.81	16.67
Term Frequency	38.48	9.44	50.62	16.67	28.56	7.92	41.62	16.67

	3 Decimals			
	Average F-Score	Stand. Dev. F-Score	Max F-Score	Min F-Score
Z Score	24.49	12.55	52.00	11.46
TF/IDF	24.85	7.90	39.50	16.67
Term Frequency	26.93	8.35	40.72	16.67

D.7 Naive Bayes POS Results - Two Authors

Table 96: Naive Bayes Sentiment Frequency Results for Entire Corpus(500 Sentences) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	54.76	55.45	55.93	55.69	59.52	57.95	64.29	60.96
TF/IDF	65.08	64.70	65.12	64.91	64.29	64.02	64.21	64.11
Term Frequency	71.43	71.06	71.67	71.36	67.46	67.12	67.52	67.32
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	59.52	59.47	59.46	59.46				
TF/IDF	54.76	54.62	54.63	54.62				
Term Frequency	51.59	51.06	51.11	51.08				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	58.70	2.22	60.96	55.69				
TF/IDF	61.22	4.67	64.91	54.62				
Term Frequency	63.26	8.76	71.36	51.08				

Table 97: Naive Bayes Sentiment Frequency Results for Entire Corpus(1000 Sentences) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	57.14	57.73	58.17	57.95	61.90	60.45	66.91	63.52
TF/IDF	73.02	73.03	72.98	73.00	73.02	72.58	73.53	73.05
Term Frequency	74.60	74.55	74.55	74.55	69.84	69.24	70.71	69.97
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	58.73	58.94	58.98	58.96				
TF/IDF	60.32	60.00	60.19	60.09				
Term Frequency	52.38	52.12	52.14	52.13				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	60.14	2.42	63.52	57.95				
TF/IDF	68.72	6.10	73.05	60.09				
Term Frequency	65.55	9.67	74.55	52.13				

Table 98: Naive Bayes Sentiment Frequency Results for Entire Corpus(1500 Sentences) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	59.52	60.91	66.16	63.43	64.29	63.18	67.01	65.04
TF/IDF	66.67	66.59	66.59	66.59	76.19	75.91	76.39	76.15
Term Frequency	66.67	66.59	66.59	66.59	80.95	80.91	80.91	80.91
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	59.52	60.23	61.11	60.67				
TF/IDF	59.52	59.77	59.84	59.80				
Term Frequency	57.14	56.82	56.94	56.88				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	63.04	1.81	65.04	60.67				
TF/IDF	67.51	6.70	76.15	59.80				
Term Frequency	68.13	9.87	80.91	56.88				

Table 99: Naive Bayes Sentiment Frequency Results for Foundation Corpus [1980's Era](500 Sentences) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	57.69	57.69	60.83	59.22	60.26	60.26	60.60	60.43
TF/IDF	69.23	69.23	69.70	69.46	60.26	60.26	60.43	60.34
Term Frequency	69.23	69.23	69.44	69.33	67.95	67.95	67.96	67.95
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	56.41	56.41	56.43	56.42				
TF/IDF	65.38	65.38	65.64	65.51				
Term Frequency	57.69	57.69	57.82	57.75				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	58.69	1.68	60.43	56.42				
TF/IDF	65.11	3.73	69.46	60.34				
Term Frequency	65.01	5.16	69.33	57.75				

Table 100: Naive Bayes Sentiment Frequency Results for Foundation Corpus [1980's Era](1000 Sentences) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	62.16	63.01	71.19	66.85	75.68	76.17	79.83	77.96
TF/IDF	62.16	61.99	62.20	62.09	67.57	67.40	67.71	67.55
Term Frequency	64.86	64.77	64.85	64.81	64.86	64.91	64.91	64.91
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	56.76	57.02	57.27	57.14				
TF/IDF	72.97	72.81	73.21	73.01				
Term Frequency	67.57	67.25	68.32	67.78				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	67.32	8.50	77.96	57.14				
TF/IDF	67.55	4.46	73.01	62.09				
Term Frequency	65.83	1.38	67.78	64.81				

Table 101: Naive Bayes Sentiment Frequency Results for Foundation Corpus [1980's Era](1500 Sentences) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	66.67	66.67	80.00	72.73	62.50	62.50	63.33	62.91
TF/IDF	75.00	75.00	75.71	75.35	62.50	62.50	62.59	62.54
Term Frequency	66.67	66.67	67.14	66.90	62.50	62.50	62.59	62.54
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	54.17	54.17	56.32	55.22				
TF/IDF	62.50	62.50	62.59	62.54				
Term Frequency	70.83	70.83	70.98	70.90				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	63.62	7.16	72.73	55.22				
TF/IDF	66.81	6.04	75.35	62.54				
Term Frequency	66.78	3.41	70.90	62.54				

Table 102: Naive Bayes Sentiment Frequency Results for Foundation Corpus [1950's Era](500 Sentences) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	67.35	67.75	70.88	69.28	61.22	61.75	65.88	63.75
TF/IDF	65.31	65.75	69.29	67.47	59.18	60.00	77.27	67.55
Term Frequency	69.39	69.67	71.15	70.40	61.22	61.67	64.29	62.95
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	59.18	59.67	62.39	61.00				
TF/IDF	57.14	57.50	58.52	58.01				
Term Frequency	71.43	71.75	73.99	72.85				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	64.68	3.44	69.28	61.00				
TF/IDF	64.34	4.48	67.55	58.01				
Term Frequency	68.74	4.21	72.85	62.95				

Table 103: Naive Bayes Sentiment Frequency Results for Foundation Corpus [1950's Era](1000 Sentences) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	70.83	70.83	70.98	70.90	70.83	70.83	70.98	70.90
TF/IDF	70.83	70.83	70.98	70.90	66.67	66.67	66.67	66.67
Term Frequency	70.83	70.83	70.98	70.90	70.83	70.83	70.98	70.90
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	58.33	58.33	61.11	59.69				
TF/IDF	75.00	75.00	75.71	75.35				
Term Frequency	83.33	83.33	84.29	83.81				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	67.17	5.29	70.90	59.69				
TF/IDF	70.98	3.55	75.35	66.67				
Term Frequency	75.21	6.08	83.81	70.90				

Table 104: Naive Bayes Sentiment Frequency Results for Foundation Corpus [1950's Era](1500 Sentences) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	58.33	58.33	58.57	58.45	50.00	50.00	50.00	50.00
TF/IDF	66.67	66.67	66.67	66.67	66.67	66.67	68.75	67.69
Term Frequency	75.00	75.00	75.71	75.35	66.67	66.67	68.75	67.69
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	66.67	66.67	80.00	72.73				
TF/IDF	75.00	75.00	83.33	78.95				
Term Frequency	75.00	75.00	75.71	75.35				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	60.39	9.38	72.73	50.00				
TF/IDF	71.10	5.56	78.95	66.67				
Term Frequency	72.80	3.61	75.35	67.69				

Table 105: Naive Bayes Sentiment Frequency Results for Galactic Empire Corpus (500 Sentences) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	69.05	69.05	69.09	69.07	59.52	59.52	65.44	62.34
TF/IDF	64.29	64.29	64.32	64.30	54.76	54.76	54.86	54.81
Term Frequency	66.67	66.67	66.67	66.67	50.00	50.00	50.00	50.00
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	61.90	61.90	62.01	61.95				
TF/IDF	57.14	57.14	57.41	57.27				
Term Frequency	50.00	50.00	50.00	50.00				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	64.45	3.27	69.07	61.95				
TF/IDF	58.80	4.02	64.30	54.81				
Term Frequency	55.56	7.86	66.67	50.00				

Table 106: Naive Bayes Sentiment Frequency Results for Galactic Empire Corpus (1000 Sentences) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	66.67	66.67	67.53	67.10	66.67	66.67	67.53	67.10
TF/IDF	61.11	61.11	62.50	61.80	55.56	55.56	55.84	55.70
Term Frequency	61.11	61.11	61.25	61.18	50.00	50.00	50.00	50.00
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	61.11	61.11	61.25	61.18				
TF/IDF	44.44	44.44	43.08	43.75				
Term Frequency	44.44	44.44	44.44	44.44				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	65.12	2.79	67.10	61.18				
TF/IDF	53.75	7.50	61.80	43.75				
Term Frequency	51.87	6.96	61.18	44.44				

Table 107: Naive Bayes Sentiment Frequency Results for Galactic Empire Corpus (1500 Sentences) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	58.33	58.33	58.57	58.45	66.67	66.67	68.75	67.69
TF/IDF	50.00	50.00	50.00	50.00	41.67	41.67	41.43	41.55
Term Frequency	66.67	66.67	66.67	66.67	41.67	41.67	41.43	41.55
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	58.33	58.33	58.57	58.45				
TF/IDF	58.33	58.33	58.57	58.45				
Term Frequency	33.33	33.33	33.33	33.33				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	61.53	4.36	67.69	58.45				
TF/IDF	50.00	6.90	58.45	41.55				
Term Frequency	47.18	14.18	66.67	33.33				

Table 108: Naive Bayes Sentiment Frequency Results for Robot Corpus (500 Sentences) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	76.79	76.79	84.15	80.30	80.36	80.36	80.71	80.53
TF/IDF	92.86	92.86	92.86	92.86	80.36	80.36	80.71	80.53
Term Frequency	92.86	92.86	92.86	92.86	89.29	89.29	90.10	89.69
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	55.36	55.36	59.93	57.55				
TF/IDF	80.36	80.36	80.71	80.53				
Term Frequency	64.29	64.29	64.97	64.63				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	72.80	10.78	80.53	57.55				
TF/IDF	84.64	5.81	92.86	80.53				
Term Frequency	82.39	12.63	92.86	64.63				

Table 109: Naive Bayes Sentiment Frequency Results for Robot Corpus (1000 Sentences) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	79.17	79.17	85.29	82.12	87.50	87.50	87.76	87.63
TF/IDF	95.83	95.83	96.15	95.99	87.50	87.50	90.00	88.73
Term Frequency	95.83	95.83	96.15	95.99	95.83	95.83	96.15	95.99
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	54.17	54.17	59.52	56.72				
TF/IDF	79.17	79.19	79.37	79.28				
Term Frequency	66.67	66.67	72.22	69.33				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	75.49	13.46	87.63	56.72				
TF/IDF	88.00	6.84	95.99	79.28				
Term Frequency	87.10	12.57	95.99	69.33				

Table 110: Naive Bayes Sentiment Frequency Results for Robot Corpus (1500 Sentences) - Two Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	93.75	93.75	94.44	94.09	87.50	87.50	90.00	88.73
TF/IDF	93.75	93.75	94.44	94.09	87.50	87.50	87.50	87.50
Term Frequency	93.75	93.75	94.44	94.09	81.25	81.25	81.75	81.50
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	56.25	56.25	76.67	64.89				
TF/IDF	81.25	81.25	81.75	81.50				
Term Frequency	75.00	75.00	75.00	75.00				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	82.57	12.69	94.09	64.89				
TF/IDF	87.70	5.14	94.09	81.50				
Term Frequency	83.53	7.93	94.09	75.00				

Table 111: Sentiment Frequency Statistics for Naive Bayes - Two Authors

	1 Decimal				2 Decimals			
	Average F-Score	Stand. Dev. F-Score	Max F-Score	Min F-Score	Average F-Score	Stand. Dev. F-Score	Max F-Score	Min F-Score
Z Score	68.38	10.32	94.09	55.69	68.63	10.36	88.73	50.00
TF/IDF	71.70	12.63	95.99	50.00	67.63	12.08	88.73	41.55
Term Frequency	73.85	10.82	95.99	61.18	68.26	14.21	95.99	41.55

	3 Decimals			
	Average F-Score	Stand. Dev. F-Score	Max F-Score	Min F-Score
Z Score	60.14	4.14	72.73	55.22
TF/IDF	65.91	11.07	81.50	43.75
Term Frequency	61.69	13.29	83.81	33.33

D.8 Naive Bayes POS Results - Multiple Authors

Table 112: Naive Bayes Sentiment Frequency Results for Entire Corpus(500 Sentences) - Multiple Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	32.54	44.44	30.28	36.02	39.68	18.31	22.59	20.23
TF/IDF	57.94	54.71	44.99	49.38	60.32	33.60	58.61	42.71
Term Frequency	61.11	44.44	46.97	45.67	52.38	16.67	8.73	11.46
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	52.38	16.67	8.73	11.46				
TF/IDF	52.38	16.67	8.83	11.54				
Term Frequency	60.32	32.83	59.88	42.41				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	22.57	10.16	36.02	11.46				
TF/IDF	34.54	16.49	49.38	11.54				
Term Frequency	33.18	15.42	45.67	11.46				

Table 113: Naive Bayes Sentiment Frequency Results for Entire Corpus(1000 Sentences) - Multiple Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	39.68	28.54	22.04	24.87	52.38	21.21	31.29	25.28
TF/IDF	58.73	39.14	49.63	43.77	52.38	16.67	8.87	11.58
Term Frequency	57.14	38.64	49.79	43.51	53.97	19.44	17.50	18.42
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	52.38	16.67	8.73	11.46				
TF/IDF	52.38	16.67	8.73	11.46				
Term Frequency	52.38	16.67	8.73	11.46				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	20.54	6.42	25.28	11.46				
TF/IDF	22.27	15.20	43.77	11.46				
Term Frequency	24.46	13.77	43.51	11.46				

Table 114: Naive Bayes Sentiment Frequency Results for Entire Corpus(1500 Sentences) - Multiple Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	59.52	42.80	64.05	51.31	52.38	16.67	8.73	11.46
TF/IDF	66.67	45.08	61.29	51.95	57.14	25.00	25.83	25.41
Term Frequency	66.67	48.48	66.67	56.14	57.14	25.00	25.83	25.41
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	52.38	16.67	8.73	11.46				
TF/IDF	52.38	16.67	8.73	11.46				
Term Frequency	52.38	16.67	8.73	11.46				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	24.74	18.79	51.31	11.46				
TF/IDF	29.61	16.79	51.95	11.46				
Term Frequency	31.00	18.66	56.14	11.46				

Table 115: Naive Bayes Sentiment Frequency Results for Foundation Corpus [1980's Era](500 Sentences) - Multiple Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	28.21	32.05	35.89	33.86	39.74	30.13	31.33	30.72
TF/IDF	48.72	35.90	36.53	36.21	51.28	32.05	50.18	39.12
Term Frequency	44.87	31.41	33.64	32.49	52.56	28.85	37.83	32.74
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	50.00	25.00	12.50	16.67				
TF/IDF	50.00	26.28	37.50	30.90				
Term Frequency	50.00	25.00	12.66	16.81				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	27.08	7.48	33.86	16.67				
TF/IDF	35.41	3.40	39.12	30.90				
Term Frequency	27.34	7.45	32.74	16.81				

Table 116: Naive Bayes Sentiment Frequency Results for Foundation Corpus [1980's Era](1000 Sentences) - Multiple Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	45.95	28.07	27.98	28.02	51.35	33.55	36.22	34.83
TF/IDF	54.05	37.72	52.31	43.83	51.35	25.00	13.97	17.92
Term Frequency	51.35	39.25	56.88	46.45	54.05	29.17	26.47	27.75
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	51.35	25.00	12.84	16.97				
TF/IDF	51.35	25.00	12.84	16.97				
Term Frequency	51.35	25.00	12.84	16.97				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	26.61	7.36	34.83	16.97				
TF/IDF	26.24	12.45	43.83	16.97				
Term Frequency	30.39	12.18	46.45	16.97				

Table 117: Naive Bayes Sentiment Frequency Results for Foundation Corpus [1980's Era](1500 Sentences) - Multiple Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	50.00	33.33	43.89	37.89	50.00	25.00	12.50	16.67
TF/IDF	54.17	31.25	26.14	28.47	54.17	31.25	38.64	34.55
Term Frequency	50.00	25.00	12.50	16.67	54.17	31.25	38.04	34.31
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	50.00	25.00	12.50	16.67				
TF/IDF	50.00	25.00	12.50	16.67				
Term Frequency	50.00	25.00	12.50	16.67				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	23.74	10.00	37.89	16.67				
TF/IDF	26.56	7.43	34.55	16.67				
Term Frequency	22.55	8.32	34.31	16.67				

Table 118: Naive Bayes Sentiment Frequency Results for Foundation Corpus [1950's Era](500 Sentences) - Multiple Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	46.94	35.75	44.94	39.82	42.86	27.38	23.57	25.33
TF/IDF	46.94	42.12	41.90	42.01	48.98	26.12	25.28	25.69
Term Frequency	46.94	33.62	28.91	31.09	53.06	28.12	38.02	32.33
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	51.02	25.00	12.76	16.90				
TF/IDF	34.69	27.62	30.00	28.76				
Term Frequency	53.06	30.25	46.67	36.71				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	27.35	9.47	39.82	16.90				
TF/IDF	32.15	7.08	42.01	25.69				
Term Frequency	33.37	2.41	36.71	31.09				

Table 119: Naive Bayes Sentiment Frequency Results for Foundation Corpus [1950's Era](1000 Sentences) - Multiple Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	54.17	35.42	48.61	40.98	50.00	25.00	12.50	16.67
TF/IDF	62.50	43.75	52.50	47.73	50.00	25.00	12.50	16.67
Term Frequency	62.50	43.75	64.29	52.07	50.00	25.00	12.50	16.67
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	50.00	25.00	12.50	16.67				
TF/IDF	50.00	25.00	12.50	16.67				
Term Frequency	50.00	25.00	12.50	16.67				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	24.77	11.46	40.98	16.67				
TF/IDF	27.02	14.64	47.73	16.67				
Term Frequency	28.47	16.69	52.07	16.67				

Table 120: Naive Bayes Sentiment Frequency Results for Foundation Corpus [1950's Era](1500 Sentences) - Multiple Authors

	1 Decimal Place				2 Decimal Places			
	Accuracy	Recall	Precision	F-Score	Accuracy	Recall	Precision	F-Score
Z Score	50.00	25.00	12.50	16.67	50.00	25.00	12.50	16.67
TF/IDF	50.00	25.00	12.50	16.67	50.00	25.00	12.50	16.67
Term Frequency	50.00	25.00	12.50	16.67	50.00	25.00	12.50	16.67
	3 Decimals							
	Accuracy	Recall	Precision	F-Score				
Z Score	50.00	25.00	12.50	16.67				
TF/IDF	50.00	25.00	12.50	16.67				
Term Frequency	50.00	25.00	12.50	16.67				
	Corpus Statistics							
	Average F-Score	Standard Deviation F-Score	Maximum F-Score	Minimum F-Score				
Z Score	16.67	0.00	16.67	16.67				
TF/IDF	16.67	0.00	16.67	16.67				
Term Frequency	16.67	0.00	16.67	16.67				

Table 121: Sentiment Frequency Statistics for Naive Bayes - Multiple Authors

	1 Decimal				2 Decimals			
	Average F-Score	Stand. Dev. F-Score	Max F-Score	Min F-Score	Average F-Score	Stand. Dev. F-Score	Max F-Score	Min F-Score
Z Score	34.38	9.54	51.31	16.67	21.98	7.17	34.83	11.46
TF/IDF	40.00	10.62	51.95	16.67	25.59	10.39	42.71	11.58
Term Frequency	37.86	13.65	56.14	16.67	23.97	7.92	34.31	11.46

	3 Decimals			
	Average F-Score	Stand. Dev. F-Score	Max F-Score	Min F-Score
Z Score	14.99	2.50	16.97	11.46
TF/IDF	17.90	6.80	30.90	11.46
Term Frequency	20.65	10.41	42.41	11.46

VITA

MICHAEL SCHNEIDER

Education: M.S. Computer and Information Science, East Tennessee State University,
Johnson City, Tennessee 2015

B.S. Computer and Information Science, East Tennessee State University
Johnson City, Tennessee 2014

Professional
Experience:

Graduate Assistant, East Tennessee State University,
Johnson City, Tennessee, 2013-2015

Software Developer Intern, IBM,
Raleigh, NC, Summer 2013

Software Developer Intern, CGI,
Abingdon, VA, Summer 2012