International Journal of Business, Human and Social Sciences
ISSN: 2517-9411
Vol:9, No:6, 2015

Effect of Personality Traits on Classification of Political Orientation

Vesile Evrim, Aliyu Awwal

Abstract**—**Today, there is a large number of political transcripts available on the Web to be mined and used for statistical analysis, and product recommendations. As the online political resources are used for various purposes, automatically determining the political orientation on these transcripts becomes crucial. The methodologies used by machine learning algorithms to do an automatic classification are based on different features that are classified under categories such as Linguistic, Personality etc. Considering the ideological differences between Liberals and Conservatives, in this paper, the effect of Personality traits on political orientation classification is studied. The experiments in this study were based on the correlation between LIWC features and the BIG Five Personality traits. Several experiments were conducted using Convote U.S. Congressional-Speech dataset with seven benchmark classification algorithms. The different methodologies were applied on several LIWC feature sets that constituted by 8 to 64 varying number of features that are correlated to five personality traits. As results of experiments, Neuroticism trait was obtained to be the most differentiating personality trait for classification of political orientation. At the same time, it was observed that the personality trait based classification methodology gives better and comparable results with the related work.

Keywords**—**Politics, personality traits, LIWC, machine learning.

I. Introduction

DUE to the intense use of Web environment, blogs, news and social networking sites have become the biggest communication medium. The availability of excessive amount of data about the user's behavior attracts companies and researchers to extract the valuable information from available resources by Opinion Mining techniques [1], [2] and later use with recommender systems to advertise the user-customized products [3].

The domain of politics has become one of the hot topics that is widely written and commented over internet. Especially during election times, following the opinions of people becomes even more important for statistical analysis and strategy determination. Therefore, it is important to automatically identify the hidden political affiliations in the documents by using classification algorithms [4], [5].

The politics and opinion mining are both broad research domains. Considering that Republicans and Democrats have different opinions and behaviors [6], [7] analyzing personality traits and its influence on leadership ability has gained importance over the years [8]. Therefore, How to characterize the opinion about politics? Become a big question.

Vesile Evrim is with the Computer Engineering Department, European University of Lefke, Mersin 10 Turkey (phone:+90-392-660-2000/2517; fax: +90-392-660-2503 ; e-mail: vevrim@ eul.edu.tr)

Aliyu Awwal is a M.S. student at Computer Engineering Department, European University of Lefke, Mersin 10 Turkey (e-mail:aliyu_eul@yahoo.com)

Considering the strong connection between personality and politics, this paper studies the effect of personality traits on political affiliation classification and is structured as follows. In Section II, background and related work on personality and politics is presented. The corpus and data pre-processing is explained in Section III. Next in Section IV, we talk about the methodology behind our study. Experiments are explained in Section V. In section VI, results are discussed and compared. Finally, in Section VII, we draw final conclusions and outline the future work.

II. Background and Related Work

A. Personality Traits

Personality is usually associated with five traits, Extraversion (E), Neuroticism (N), Agreeableness (A), Conscientiousness (C), and Openness (O), which are mostly tested in personality questionnaires. The model that uses these five traits and its facets is called Big Five [9] which has become a standard model in psychology, and intensively used among researchers.

Personality has an effect on the linguistics of people. For example, extraverts and conscientious people tend to be positive whereas neurotic and disagreeable people are associated with negativity [10]. Similarly, the outgoing extraverted individuals are considered successful while neurotics are not [11]. Yarkoni showed that agreeableness is negatively correlated with the use of "Anger" and "Swearing" words. On the other hand, people who score high on the Openness trait is tend to use more articles and propositions [12].

The significance of determining which linguistic features should be applied for experimentation is quite important. By the investigation of this, many researchers [13]-[15] have based their works on LIWC [16] software designed for analysis of texts. LIWC have different word classes over an extensive range of dimensions like "positive" or "negative emotions", "self-references", "causal words", as well as seventy other dimensions each with several dozens or hundreds of words.

Yarkoni found that there is a correlation between 66 LIWC (2001) categories and the Big Five personality traits [12], excluding the non-semantic and speech related features, such as non-fluencies. His findings replicated the previous studies that mostly analyzed only on a subset of LIWC categories due to a lack of insufficient data [13], [17]. Unexpectedly, Openness found to be negatively correlated with many of the

LIWC features and positively correlated with only 4 features due to the usage of more articles and prepositions with a tendency of using "function" words rather than "content" words [12]. Extraversion is strongly correlated to oral language and extraverts talk more in a less formal language [18] which leads the use of pronouns, verbs and interjections [19].

In addition to personality traits, Yarkoni also analyzed the correlation between the 6 facets of each Big Five personality traits and LIWC features. Although many facet-level correlation supports the trait level results, among the traits' facets considerable amount of heterogeneity was identified. For Neuroticism most of the heterogeneity was detected in "Self-Consciousness" with negative correlation to interpersonal interactions. Extraversion's facets, "Friendliness", "Gregariousness", and "Cheerfulness" all showed positive correlation with interpersonal interactions and positive effects, unlikely to other facets. Similarly the "Artistic Interest" and "Emotionality" facets of Openness have different correlation characteristics than the other four facets [12].

B. Personality and Politics

The Big Five personality traits are important in determining political ideologies and they influence socio-economic positions [20]. Obtaining the personality characteristics of Liberals and Conservatives are important to find out the understanding between two views [21]. The evidence shows that Liberals have more tendency to be Open and Agreeable [7], [22], [23]. On the other hand, conservatives have higher scores than liberals on Conscientiousness and Extraversion [6]. Based on the lack of proof, no relation was found between Neuroticism and Political Ideology.

Experiments have shown that personality effects many task oriented behaviors from leadership ability to motivation [8], [24], [25], which are important at characterizing politicians behavior. In order to estimate more detailed characteristics from text, that are crucial for politics, automated systems are designed to detect emotions, mood and dominance from text documents [26]. The studies have shown that mood and emotions of people are short term situations where personality is a more permanent characteristic that is reflected on behavior over time [27].

The individual differences in language usage affect the way information is conveyed and helps in the identification of the author's personality [28]. Therefore, the automatic recognition of personality traits from text has gained importance. However, there is little work on automatic recognition of personality traits on text [14], [29]. The detection of personality traits is important for Opinion Mining research, which is one of the backbones of determining political affiliations. The evidence shows that incorporating personality research into other task may improve the classification accuracy in opinion mining [29].

Pennacchiotti et al. used the Gradient Boosted Decision Tree Learning algorithm for detecting political affiliations on the Twitter user data (collected by We Follow and Twellow3) and found that topic-based linguistic features are promising in the classification of users' political orientation and ethnicity [4], [30]. Monroe identified and evaluated the linguistic differences between Democrats and Republicans in U.S. Senate speeches on a given topic like "Defense", "Taxes" or "Abortion". The relative utility of these approaches which are based on Bayesian shrinkage and regularization was illustrated [30].

Therefore with the aim of investigating the relationship between personality and political texts, in this paper, the automatic classification of political affiliations, via analysis of Big Five Personality traits and LIWC features are studied as explained in the following sections.

III. Corpus

In this study, Convote dataset (v1.1) was used. The dataset consists of U.S. Congressional-Speech was downloaded from http://www.cs.cornell.edu/home/llee/data/convote.html website. Each speech in the corpus was labeled with an identifier for its speaker and his recorded vote for the corresponding debate. The original dataset consists of three stages of tokenized speech-segment data [31]. For this study, the data in stage-three, consisting of speech-segments were used. Stage-three set comprises development, test and training sets. In this study, for comparison purposes, 1200 positive and 1200 negative documents of the stage-three training set were used, as in Alloty's experiments [32]. Each document in the stage-three, training-set was assigned as negative or positive; negative indicating Democrat and positive indicating Republican.

A. Data Pre-Processing and Feature Vector Construction

One of the aims of our study is to test the impact of personality traits over political ideologies. Building a new word-base for traits is beyond this study and falls into the research of Psychology domain. Therefore, 64 features of LIWC (2007), which covers most of the BIG Five Personality traits [16], were used to construct the feature set of this study. Each of the 64 dimensions of a feature vector represents a set of words. For example, the "Article" feature consists of words such as "a", "an", "the", and the "Anger" feature has words such as "kill", "shoot" "break" in their sets.

In processing the data, the documents in the dataset first went through stop-word and punctuation elimination. Although many of the words in LIWC features share common stems, the relationships between personality and stemmed words could be negatively affected in comparison to un-stemmed words [12]. The reason follows that LIWC features such as "present tense verbs", "past tense verbs" and "future tense verbs", are based on "tenses" and stemming would make it impossible to distinguish between such words. Therefore in this study, the common pre-processing step, stemming, was not applied and all words were left un-stemmed.

After document pre-processing, a vector, with 64 dimensions, were constructed for each document. The value for each dimension of a vector was obtained by calculating the term frequencies (TF) of the words for that feature. Thereafter, the excel sheet with a vector for each document was fed to the

RapidMiner Tool for experimentation, as explained in the following sections.

IV. Methodology

In order to measure the effect of Personality traits over political affiliations, the correlation between the LIWC and Big Five Personality traits were used. We based this study to the Yarkoni's correlation table between the LIWC (2001) and Big Five Personality traits [12]. In this study, we used LIWC (2007) which captures, on average, over 86% of the words people use in writing and speech. In the LIWC (2007) version, the rarely used categories of LIWC (2001) version; Optimism, Positive Feelings, Communication Verbs, Other References, Metaphysical, Sleeping, Grooming, School, Sports, Television, Up, Down, and the category of Unique Words have been removed from the set. Therefore, when Yarkoni's correlation table was used, the removed categories were omitted [33].

In this study, the top positively correlated (PC) and negatively correlated (NC) LIWC features, and also the absolute value of the top correlated LIWC features (AC) with BIG Five Traits were used by manually determined thresholds to test the effect of Personality traits over classifying the political orientation.

The seven classification algorithms namely, Decision Tree (Dec Tree), Rule Induction (Rule Ind), M5 Rule, SVM, SMO, ADTree, and J48 were tested using the RapidMiner [34] machine learning toolkit to classify each document. The task of validating performances, for the classification algorithms, was done through the X-Validation operator.

V. Experiments

In the experiments, for the comparison of related work, classification algorithms were applied on two sets of training/testing datasets:

- 80:20 Training/testing dataset with 5 Fold X-Validation
- 90:10 Training/testing dataset with 10 Fold X-Validation

The RapidMiner tool's performance validation operator produces performance results in terms of accuracy, as well as Average Precision and Average Recall, for Positive and Negative classes. Since the dataset used in this study are symmetric, accuracy was chosen as the main performance evaluation metric in the experiments.

A. Full Feature Classification

In this experiment 64 features of LIWC (2007) were tested using 7 classification algorithms of RapidMiner Tool on the Convote dataset. The experiments were conducted on both 80:20 and 90:10 training/testing ratios and the performance of the classification algorithms were measured by the accuracy as the results are provided in Table I.

From the results, it is observed that the "M5 Rule" classifier has the best performance, with 62.42% accuracy, on the 80:20 training/testing ratios of the dataset. Although the classification on 90:10 training/testing dataset were expected to give better results, only 4 classifiers improved their performances, and the "Rule Induction" and "M5 Rule" algorithms both had reduced accuracies with 90:10 training/testing ratios. The "Decision Tree" algorithm performed the worst because of an "unknown" average precision and 0% recall values on Liberal documents. In general, all classifiers had better precision value compared to the recall value. While the average precision values fell in the 55%-62% interval, recall values stayed in the range of 26%-56%.

TABLE I
CLASSIFICATION OF 64 LIWC FEATURES ON 80:20 AND 90:10 TRAINING/TESTING RATIOS

80:20 (5 Folds)						
Dec. Tree	Rule Ind	M5 Rule	SVM	SMO	ADTree	J48
50.00	55.00	62.42	57.71	57.88	58.38	57.67
90:10 (10 Folds)						
Dec. Tree	Rule Ind	M5 Rule	SVM	SMO	ADTree	J48
50.00	53.83	61.96	57.92	57.96	58.50	58.83

B. Personality Based Classification

The aim of this experiment is to test the effect of the top features that is highly correlated with all the personality traits over classifying the political affiliations. In this experiment, Yarkoni's correlation table was used and the average of the absolute correlation values of the LIWC features (AAC) over BIG Five traits were calculated as an example case as presented in Table II. Thereafter, all the 64 features were sorted based on the average correlation number.

TABLE II
EXAMPLE TO THE AVERAGE OF THE ABSOLUTE CORRELATION VALUE OF THE LIWC FEATURES OVER BIG FIVE TRAITS

LIWC Features	N	E	O	A	C	Average
Anger	0.13	0.03	0.03	0.23	0.19	0.122
Swear words	0.11	0.06	0.06	0.21	0.14	0.116

TABLE III
THE ACCURACY OF THE TOP 8, 15 AND 20 THE MOST CORRELATED LIWC FEATURES USING 6 CLASSIFIERS ON 80:20 AND 90:10 TRAINING/TESTING DATASET

80:20 (5 Fold)						
# Features	Rule Ind	M5 Rule	SVM	SMO	ADTree	J48
8	51.37	57.92	54.12	54.62	56.12	56.25
15	53.50	59.33	55.83	55.92	59.00	58.00
23	54.29	60.00	56.50	56.29	56.96	57.38
90:10 (10 Fold)						
# Features	Rule Ind	M5 Rule	SVM	SMO	ADTree	J48
8	51.17	57.50	54.25	55.00	56.67	56.54
15	54.92	59.71	55.79	55.50	58.67	57.04
23	55.38	60.04	55.75	55.92	58.50	57.71

In order to see the effect of the various numbers of features, three feature sets that were constructed by 8, 15 and 20 features were tested using the above mentioned 6 classifiers as presented in Table III, which excludes the "Decision Tree" classifier as it performed the worst on both 80:20 and 90:10 training/testing ratios.

The classification results shows that the accuracy of the algorithms increased as the number of features increased in the experiment sets. The differences in the performances between

8 to 15 feature sets were observed to be higher than the experiments between 15 to 23 feature sets. In fact, “ADTree” and “J48” classification algorithms were performed the best with 15 features on 80:20 training/testing ratios. The “M5 Rule” classifier outperformed all the classifiers in all feature sets, with the highest performance being 60% in 23 feature set. On the other hand, the “Rule Induction” algorithm had the worst accuracy 51.37% on the 8 feature set because of a very low (7%) recall value. No noticeable difference was observed in the experiments between 80:20 and 90:10 training/testing ratios.

C. Trait Based Classification

In this experiment, the absolute value of the top 10 correlated LIWC features were selected for each Personality trait category. Therefore, five different feature sets were constructed and tested with 6 Classification algorithms on 80:20 training/testing ratios of Convote dataset as presented in Table IV.

TABLE IV
ACCURACY OF THE CORRELATED LIWC FEATURES FOR 5 PERSONALITY TRAITS WITH 6 CLASSIFIERS USING 80:20 TRAINING/TESTING RATIOS OF CONVOTE DATASET

Traits	Rule Ind	M5 Rule	SVM	SMO	ADTree	J48
Neuroticism	53.58	61.29	57.75	57.83	59.83	58.92
Extraversion	51.17	53.79	52.21	52.04	53.33	54.00
Openness	52.79	56.75	54.29	53.79	56.46	57.04
Agreeableness	52.63	57.88	54.54	55.12	58.88	57.79
Conscientiousness	51.92	58.96	56.63	55.38	55.71	56.33

From the experiments it can be seen that the “M5 Rule” algorithm gives the best performance with 61.29% over the Neuroticism personality trait. The best results for the other traits were as followed: Conscientiousness 58.96%, Agreeableness 58.88%, Openness 57.04% and Extraversion 54%. The classifiers “M5 Rule”, “ADTree” and “J48” had the best performances over 5 Personality traits. The “J48” algorithm had the lowest precision and “M5 Rule” algorithm had the highest recall values among all the classifiers for all the traits.

D. Trait Based Classification Based on Positively and Negatively Correlated Features

In the previous experiments as reported in Section V A-C, features’ absolute correlation value was used in selecting the top rated LIWC features. Furthermore, in this experiment we were interested in testing the effect of the 10 most positively and 10 most negatively correlated LIWC features for each personality trait on political orientation classification. The experiments were conducted by both 80:20 and 90:10 training/testing ratios. However, since the accuracies of classifiers with both ratios did not show more than 0.5% variance, the results of the classifiers using the 80:20 training/testing ratios over Convote dataset is presented as in Table V.

The best results were obtained by the top 10 positively correlated features of Neuroticism using the “M5 Rule” that achieved 61.42% accuracy. The biggest difference between the positively and negatively correlated feature sets was obtained with the Neuroticism trait ~ 5% variations and the smallest difference were observed on Extraversion trait ~ 0.5%. The best results for positively and negatively correlated feature sets of Extraversion are performed by “ADTree” algorithm. While positively correlated feature set of Extraversion was performed 55% accuracy, negatively correlated set performed 55.5%. Both the negative and positive correlated sets have shown better results than the best performances obtained by the absolute correlation value of the LIWC features of Extraversion that was found to be 54% with “J48” classifier.

TABLE V
THE ACCURACY OF THE TOP 10 POSITIVELY AND NEGATIVELY CORRELATED LIWC FEATURES WITH FIVE PERSONALITY TRAITS USING 6 CLASSIFIERS

80:20 POSITIVELY CORRELATED FEATURES						
TRAITS	Rule Ind	M5 Rule	SVM	SMO	ADTree	J48
Neuroticism	53.58	61.42	57.58	57.42	58.62	58.88
Extraversion	50.75	52.75	51.67	51.42	55.00	53.83
Openness	53.96	58.21	54.46	54.75	58.46	57.33
Agreeableness	53.42	54.50	54.33	51.88	55.46	54.75
Conscientiousness	54.83	54.96	53.21	52.00	55.96	57.50
80:20 NEGATIVELY CORRELATED FEATURES						
TRAITS	Rule Ind	M5 Rule	SVM	SMO	ADTree	J48
Neuroticism	52.75	56.71	53.58	50.71	56.50	56.17
Extraversion	53.08	54.25	54.29	52.79	55.50	55.50
Openness	52.17	56.96	53.62	53.79	54.50	55.12
Agreeableness	52.46	58.29	55.88	55.46	56.42	56.25
Conscientiousness	52.17	59.21	56.83	55.42	56.42	55.58

VI. DISCUSSION AND COMPARISON OF THE EXPERIMENTAL RESULTS

In the experiments the “M5 Rule” algorithm had the best performance with the “Neuroticism” Personality trait for both the top 10 features of absolute correlation value, and the positive correlation features of LIWC with personality traits. Therefore, as opposed to the related research, where Liberals and Conservatives are said to have different tendencies to Openness, Agreeableness, Conscientiousness and Extraversion, and no relation is observed between Neuroticism and politics [6], [23], our experiments proved that Neuroticism is the most differentiating personality trait between the Liberals and Conservatives, while Extraversion was obtained to be the least differentiating personality trait.

When the classifiers were tested with the top features obtained by averaging absolute correlation value of the LIWC features over the five personality traits, it was observed that the classifiers performances were improved as the number of features were increased. Moreover, we found out that the success of the classifiers got even better when the top correlated features were selected based on each of the five personality trait categories separately.

Alloty used the Convote database to test the performance of 3 feature selection algorithms; “CPD”, “IG”, and “χ^2”, with “SVM” and “Naïve Bayes” algorithms [32]. The accuracies of the two classifiers on three feature selection algorithms obtained by Alloty on Convote dataset is presented in Table VI.

TABLE VI
ALLOTY'S ACCURACY PERFORMANCE EVALUATION [32]

# of Features	Classifier	CPD	IG	$\chi2$
15.549	SVM	**65.96**	**65.96**	**65.96**
11.662	SVM	65.50	65.18	62.90
4.332	SVM	65.50	61.93	62.82
3.334	SVM	65.50	60.03	62.25
1.555	SVM	65.50	59.30	60.45
15.549	Naïve Bayes	57.38	57.38	57.38
11.662	Naïve Bayes	57.90	57.17	57.38
4.332	Naïve Bayes	**57.90**	**57.90**	**61.46**
3.334	Naïve Bayes	57.90	57.40	60.04
1.555	Naïve Bayes	57.90	57.00	59.40

For the comparison purposes, in this study we have used the same dataset as Alloty. The best results of the classifiers on the top correlated Personality and LIWC features, as explained in Section V, are summarized in Table VII.

The overall best results in the experiments were obtained by the 64 features of LIWC (2007), achieving an accuracy of 62.42%. Among the 6 classifiers, the "M5 Rule" and "ADTree" classification algorithms had the best performances over all the feature sets. In general, the reduced feature space that considers the top correlated features of LIWC with all Big Five traits on average performed ~ 59%. On the other hand, the classifiers performance with the feature set for personality trait Neuroticism, was 61.42%, which was the best achievable among the traits for the top positively correlated features. Contrary, the Extraversion trait showed the worst performance 55.5% with the negatively correlated features among all the other traits.

TABLE VII
COMPARISON OF THE BEST RESULTS PERFORMED IN THE EXPERIMENTS

# Features	Kind of Features	Classifier	Accuracy
64	LIWC (2007) Full Set	M5 Rules	**62.42**
20	All Personality traits	M5 Rules	60.00
15	All Personality traits	M5 Rules	59.33
8	All Personality traits	M5 Rules	57.92
10	Neuroticisim-Positive Correlation	M5 Rules	61.42
10	Extraversion-Negative Correlation	ADTree	55.50
10	Openness-Positive Correlation	ADTree	58.46
10	Agreebalenss-Trait Avg Correlation	ADTree	58.88
10	Conscientiousness-Negative Correlation	M5 Rules	59.21

As presented in Table VI, Alloty obtained the best result with the "SVM" classifier over 15.549 features (65.96%). On the other hand, the best performance for "Naïve Bayes" algorithm was obtained over 4.332 features selected by "$\chi2$" feature selection algorithm that is recorded as 61.46%. Our best results for the experiments tested on the same dataset are presented in Table VII. The performance of the 64 LIWC features with the "M5 Rule" classifier measured as 62.42%, while the top 10 positively correlated features of Neuroticism gave 61.42%. Classifiers tested over the Personality traits were outperformed almost all the results obtained by the "Naive Bayes" algorithm in Alloty's work. They also outperformed the results of the "SVM" classifier with 1555 features, selected by IG and $\chi2$ algorithms in Alloy's experiments. Considering the reduced vector space in our study, the results indicate that the 64 LIWC features and also the features selected for personality traits, give comparable accuracies when they are combined with several classification algorithms.

VII. CONCLUSION AND FUTURE WORK

The evidence in this study shows that, when the LIWC features are correlated with the BIG Five Traits and tested over the "M5 Rule" and the "ADTree" algorithms, they have reasonable classification performances. In the Convote dataset, Neuroticism wqas found to be the most differentiating personality trait for the political texts. However, Convote dataset consist of audio transcripts which might possibly have different characteristics than written texts. Therefore, more experiments on different datasets are needed to generalize the conclusion. The experiments presented in this paper was used to test the effect of LIWC features, and the correlated LIWC features with the BIG Five personality traits on political orientation classification. As future work, it would be a good idea to check the effect of highly correlated LIWC features on the facets of personality traits in order to have a more fine grained analysis.

REFERENCES

[1] K. Michael, P. Sobkowicz, and G. Bouchard, "Opinion Mining in Social Media: Modeling, Simulating, andVisualizing Political Opinion Formation in the Web," 2012.
[2] B. Liu, M. Hu, and J. Cheng, "Opinion observer: analyzing and comparing opinions on the web," in WWW '05: Proceedings of the 14th international conference on World Wide Web, 2005, pp. 342–351.
[3] W. Yue and M. M. Tseng, "Customized products recommendation based on probabilistic relevance model," Journal of Intelligent Manufacturing, 2013, vol. 24(5), pp. 951-960.
[4] M. Pennacchiotti and A.M. Popescu, "A Machine Learning Approach to Twitter User Classification," ICWSM 11, 2011, pp. 281-288.
[5] M. Pennacchiotti and A.M. Popescu, "Democrats, republicans and starbucksafficionados: user classification in twitter," in Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, 2011.
[6] S. D. Gosling, P. J. Rentfrow, and W. B. Jr. Swann, "A very brief measure of the Big Five personality domains," Journal of Research in Personality, 2003, vol. 37, pp. 504–528.
[7] L. P. Naumann, S. Vazire, P. J. Rentfrow, and & S. D. Gosling, "Personality judgments based on physical appearance," Personality and Social Psychology Bulletin, 2009.
[8] R. Hogan, G. J. Curphy, and J. Hogan, "What we know about leadership: Effectiveness and personality," American Psychologist, 1994, vol. 49 (6), pp. 493–504.
[9] W. T. Norman, "Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality rating," Journal of Abnormal and Social Psychology, 1963, vol. 66, pp. 574–583.
[10] D. Watson and L. A. Clark, "On traits and temperament: General and specific factors of emotional experience and their relation to the five factor model," Journal of Personality, 1992, vol. 60 (2), pp. 441–76.
[11] R. E. Riggio, C. Salinas, and J. Tucker, "Personality and deception ability," Personality and Individual Differences, 1988, vol. 9 (1), pp. 189–191.
[12] T. Yarkoni, "Personality in 100,000 Words: A Large-Scale Analysis of Personality and Word Use among Bloggers," National Institute of Health Public Access, 2010, pp. 1-23.

[13] J. W. Pennebaker and L. A. King, "Linguistic styles: Language use as an individual difference," Journal of Personality and Social Psychology, 1999, vol. 77, pp. 1296–1312.
[14] F. Mairesse, M. A. Walker, M. R. Mehl, and R. K. Moore, "Using Linguistic Cues for the Automatic Recognition of Personality in conversation and text", Journal of Artificial Intelligence Research, 2007, pp. 457-500.
[15] S. Chris, "Predicting Dark Triad Personality traits From Twitter Usage and A Linguistic Analysis of Tweets," in 11th International Conference on Machine Learning and Applications (ICMLA), 2012, vol. 2.
[16] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Inquiry and Word Count: LIWC 2001," Mahwah, NJ, 2001.
[17] J. B. Hirsh and J. B. Peterson, "Personality and language use in self-narratives," Journal of Research in Personality, 2009.
[18] A. J. Gill, S. Nowson, and J. Oberlander, "What are they blogging about? Personality, topic and motivation in blogs," in Paper presented at the The Third International AAAI Conference on Weblogs and Social Media, San Jose, CA, 2009.
[19] F. Heylighen and J. M. Dewaele, "Variation in the contextuality of language: an empirical measure," Context in Context: Special issue of Foundations of Science, 2002, vol. 7 (3), pp. 293–340.
[20] S. Alan Gerber, A. Gregory Huber, David Doherty, and Conor M. Dowling, "Personality and Political Attitudes: Relationships across Issue Domains and Political Context," American Political Science Review, 2010, vol. 104(1), pp. 111-133.
[21] P. J. Rentfrow, J. T. Jost, S. D. Gosling, and J. Potter, "Statewide differences in personality predict voting patterns in U.S. Presidential elections, 1996–2004," in Social and psychological bases of ideology and system justification. New York: Oxford University Press, 2009.
[22] R. McCrae and P. Costa, "Toward a new generation of personality theories: Theoretical contexts for the five-factor model. ," in The five factor model of personality: Theoretical perspectives. New York: JS Wiggins, 1996, pp. 51-87.
[23] R. Carney Dana and etal, "The secret lives of liberals and conservatives: Personality profiles, interaction styles, and the things they leave behind," Political Psychology, 2008, vol. 29(6), pp. 807-840.
[24] A. Furnham and J. Mitchell, "Personality, needs, social skills and academic achievement: A longitudinal study," Personality and Individual Differences, 1991, vol. 12, pp. 1067–1073.
[25] M. Komarraju and S. J. Karau, "The relationship between the Big Five personality traits and academic motivation," Personality and Individual Differences, 2005, vol. 39, pp. 557–567.
[26] G. Mishne, "Experiments with mood classification in blog posts," in Proceedings of ACM SIGIR 2005 Workshop on Stylistic Analysis of Text for Information Access, 2005.
[27] K. R. Scherer, "Vocal communication of emotion: A review of research paradigms," Speech Communication, 2003, vol. 40 (1-2), pp. 227–256.
[28] E. Reiter and S. G. Sripada, "Contextual influences on near-synonym choice," in Proceedings of the International Natural Language Generation Conference, 2004, pp. 161– 170.
[29] J. Oberlander and S. Nowson, "Whose thumb is it anyway? Classifying author personality from weblog text," 2006.
[30] L. Monroe Burt, Michael P. Colaresi, and Kevin M. Quinn, "Fightin'words: Lexical Feature Selection And Evaluation For Identifying The Content Of Political Conflict," Political Analysis, 2008, vol. 16(4), pp. 372-403.
[31] M. Thomas, B. Pang, and L. Lee, "Get out the vote: Determining support or opposition from congressional floor-debate transcripts," in Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, Sydney, Australia, 2006, pp. 327–335.
[32] D. A. Allotey, "Sentiment analysis and classification of online reviews using categorical proportional difference," 2011.
[33] J. W. Pennebaker, C. K. Chung, M. Ireland, A. Gonzales, and R. J. Booth, "The Development and Psychometric Properties of LIWC2007," The University of Texas at Austin and The University of Auckland, New Zealand, Austin, Texas 78703 USA, 2007.
[34] RapidMiner Inc. "RapidMiner Documentation". (Online). http://docs.rapidminer.com/studio/operators/, 2014.