Predicting the Tolerance Level of Religious Discourse Through Computational Linguistics

Nicholas Venuti, Brian Sachtjen, Hope McIntyre, Chetan Mishra,
Matthew Hays, and Donald E. Brown
University of Virginia, nmv7de, bws7vs, hm7zg, cm5ky, mch9qs, deb@virginia.edu

Abstract - **Religious violence is one of the biggest and most complicated problems facing the world today. The number of incidents has been increasing in recent years and, unfortunately, scalable and accurate systems to predict which groups are likely to engage in such actions are not keeping pace. Additionally, this problem is compounded by lingual and cultural differences, which limit the effectiveness of understanding how tolerant or intolerant a group is without bias. To circumvent this challenge, recent studies indicate promise in the analysis of the performative character of discourse (how words are used) to estimate the tolerance level, rather than using the semantic or emotive character of text (what the words mean or imply). Using expert estimates of linguistic flexibility, a representation of the performative character of text, and thus also predictive of a text's tolerance level, this paper describes (a) new approaches to automating the quantification of the performative character of words and (b) the predictive efficacy of these approaches versus traditional semantic indicators of tolerance or intolerance. To implement the pipeline, a judgment identifier was developed along with multiple semantic density algorithms to extract the frequency of judgments and flexibility of keyword contexts, respectively. Test results show that text mining algorithms can accurately estimate the language flexibility of religious discourse. These results provide evidence that the performative characteristics of language better predict tolerance level than the semantic characteristics of language.**

Index Terms - Distributional vectors, Religious violence, Semantic density, Text mining

INTRODUCTION

Religious violence has long been one of the most destructive and divisive acts impacting society and, unfortunately, the number of casualties associated with these acts has been rising in recent years. For example, in 2014 alone, 12,737 deaths could be attributed to the operations of just two separate radical terrorist groups: Boko Haram with 6,664 deaths, and the Islamic State of Iraq and The Levant (ISIL), with 6,073 deaths [1]. Unfortunately, scalable and accurate systems to predict such actions are not keeping pace with the evolution of these groups [2]. In an effort to combat these actions, the University of Virginia/Global Covenant of Religions Ethnolinguistics Research Team (GCR-ERT) has been exploring ways of diagnosing tendencies to violence.

This is typically done by examining the affective or emotive attributes of keywords distributed by the semantic characteristics of these words. As words tend to carry multiple definitions, both within normative language and across cultural divides, the GCR-ERT found that it was insufficient to utilize solely semantic analysis as a method to estimate the level of tolerance displayed within religious discourse. Through their research, the GCR-ERT has found that by examining how things are said, rather than what is said, a more accurate diagnosis can be produced. The GCR-ERT has hypothesized that the performative character of discourse, the capacity for language to encapsulate an action or identity, is correlated to the level of intolerance within that discourse.

To quantify this performative character of discourse, the GCR-ERT have developed a 1 - 9 language rigidity scale which is used to classify texts as either rigid (scored a 1) or elastic (scored a 9). These rankings were assigned manually by analyzing a wide array of characteristics within religious discourse, mostly text. While encouraged by the initial results of their research, the GCR-ERT has been challenged by the resource-intensity of data collection and analysis by ethnographers and discourse analysts relying purely on manual processes. To obtain results of significance and breadth, the GCR-ERT required a methodology capable of accurately calculating the linguistic flexibility of discourses from a wide array of languages and religious groups.

To assist the GCR-ERT in their efforts, the University of Virginia Computational Linguistics Data Science (CLDS) team developed a text mining pipeline to (a) conduct preliminary research on automating the quantification of the performative character of words and (b) test the predictive efficacy of these implementations versus traditional semantic indicators of tolerance or intolerance. Through the manual process, the GCR-ERT discovered that certain characteristics of discourse, such as the frequency of judgments and the rigidity of keyword contexts, can be used to assist in their quantification. To implement the pipeline, the CLDS developed a judgment identifier along with multiple semantic density algorithms to replicate the frequency of judgments and rigidity of keyword contexts, respectively. The CLDS tested the effectiveness of these implementations against two traditional semantic indicators as baselines: a topic model and sentiment analysis. Using

978-1-5090-0970-1/16/$31.00

three classification models (SVM, neural networks, and random forest), the CLDS compared the traditional semantic signals against the performative signals to predict linguistic flexibility.

LITERATURE REVIEW

Because religious violence is a far-reaching and complex problem, many resources, such as intelligence analysts and NGOs, are dedicated to studying the actions and words of groups identified as high-risk. Most of these approaches, however, are not quantitative in nature. They are instead supported by the expertise of trained individuals familiar with the group's historic behavior and rhetoric instead of scientific techniques. Research on predicting religious violence using quantitative approaches is limited. One study conducted by a research group at New Mexico State University used Latent Semantic Analysis to track temporal shifts in language usage for Iranian leaders. While topical shifts were identified in this approach, few predictive capabilities were developed through this methodology [3]. When quantitative methodologies are used, they tend to be fixated on predicting specific incidences of violence [4]. Given these facts, we focused our research on the literature base surrounding methods which automatically detect semantic change.

According to Barsalou, "the conceptualization of an entity or set of entities can vary widely across individuals and occasions" [9]. Barsalou tested this hypothesis by providing subjects with a group of nouns such as bachelor, bird and chair and asking his subjects to provide definitions for those terms. Through his experiments he found that only "44% of the features in one subject's definition existed in another subject's definition" which was likely due to the beliefs and experiences individuals have in characterizing a concept [9]. These results assist in confirming the GRC-ERT's hypothesis that shifts in definitions correlate with the representational flexibility of a concept.

Studying semantic change through computation is a budding research area as the development of advanced computational resources and large data repositories has made the study more feasible. Results have established the feasibility of extracting word meaning from usage patterns [5][6][7][8]. While explicit word meanings were not directly studied, a mid-stage between 'usage patterns' and 'word meaning' is a mapping to a space which notes how words tend to be used, which are potential quantitative measures for the performative characteristics of words.

Sagi et al. attempted to detect semantic change by measuring the diversity of contexts in which a word was used [10]. They measured "broadening" (a word meaning becoming less restricted), "narrowing" (a word becoming more specific), and "pejoration" (a meaning becoming more negative). The notion of "diversity of contexts" directly correlates with the concept of performative characteristics of word [10]. Given this, we decided to directly test the efficacy of this method.

Boussidan and Ploux used a graph built from subsetted co-occurrence tables to create a map of lexical usages of words [11]. Finding the cliques in this graph leads to words grouped on a semantic plane where "drunk" and "stagger" are connected despite different definitions because they are used in very similar contexts [11]. While the work has interesting results, the requirement of computing cliques is computationally complex, limiting the scalability and accessibility of this method.

Cook and Stevenson attempted to detect "amelioration" (a word losing a negative meaning) and "pejoration" through a similar method [12]. This was done by computing the average pointwise mutual similarity between a word and a set of assumed pejorative or positive words [12]. Unfortunately, this method doesn't create an intermediate space which could be connected to performative characteristics.

DATA

The CLDS team obtained the data used to perform this analysis from online repositories that contained texts from different religious groups with a wide variety of affiliations and language rigidities, as described in Table I. We collected these texts from their respective online repositories using *BeautifulSoup* in Python and converted to .txt files prior to the pre-processing stage [13]. Following this step, the GCR-ERT manually reviewed and annotated the overall document sets with a language flexibility ranking to serve as ground truth.

TABLE I
DATA SOURCES

Group	Rank	Affiliation	~# of Doc.
Westboro Baptist Church	1	Baptist	419
Faithful Word Baptist Church	2	Baptist	228
Nouman Ali Khan	3	Sunni Muslim	88
Dorothy Day	4	Catholic	774
John Piper	4	Baptist	579
Steve Shepherd	4	Christian	728
Rabbinic texts	6	Jewish	166
Unitarian texts	7	Unitarian	276
Meher Baba	8	Spiritualist	27

We randomly split each group into a training set containing 70% of the documents and a testing set containing the remaining 30%. To balance the requirements for increased observations for modeling and the text size requirements for the semantic density algorithms, we randomly placed the documents into bins of 10 for each group. If any individual bin was smaller than half the targeted bin size (i.e. 5 documents), we discarded it so as to not introduce significant sample size disparities.

PRE-PROCESSING

We normalized and cleaned all the documents prior to analyzing bins of text, to ensure data continuity between each process. We extracted the raw text for each bin, from the documents and stored as a VCorpus using R's *tm* package [14]. After consolidating we normalized the text by removing punctuation, converting to lowercase, and removing numbers. We created processed tokens by tokenizing on whitespace, and stemming the tokens. We used these tokens as the inputs for the baseline topic model and sentiment analysis, as well as distributional scores, context vectors and network approaches for semantic density analysis.

We applied the Maxent POS Tagger from the *tm* package to the corpus in order to develop the list of keywords for the semantic density analysis [14]. Using this package, word counts were generated for each unique word - POS tag combination for each bin. The top ten most frequent adjectives and adverbs were selected for each bin and used to as the keywords. This method was suggested by the GRC-ERT based on their experience extracting keywords from religious discourse. These keywords were then fed into the various semantic density analyses.

BASELINE

In an effort to evaluate the effectiveness of the semantic-based signals for measuring linguistic rigidity, we developed two baseline features to serve as comparisons. These baseline features were meant to represent the standard approaches for evaluating the level of rigidity of the discourse: (1) topic modeling to compare thematic content between texts, and (2) measuring the sentiment of a group's texts.

I. Topic Modeling

We created a topic model using Latent Dirichlet Allocation (LDA) to find underlying thematic trends between texts in the corpus using the *topicmodels* package in R [15]. LDA is a text mining technique that creates a predefined number of topics based on word frequencies within a corpus of texts, and assigns topic probabilities to documents through a three-level hierarchical Bayesian model [16]. These topic probabilities can be used as features for predictive modeling on future unseen texts [16]. This methodology has been shown to be useful in text mining as it not only reduces the feature space from the number of words in the corporal dictionary to a much smaller, predefined number, but also provides flexibility within the feature space to allow a document to represent multiple topics [16].

To create this model, we built a document term matrix from all texts in the training bins. Terms with a sparsity less than 1% were removed to avoid overfitting. A 50-topic LDA model was created from this document term matrix. This model was used to generate predicted topic probabilities for all documents in the testing bins. These topic probabilities served as baseline features for predicting linguistic flexibility on the testing set. The gamma values from LDA as well as one-hot encoding were used in prediction, however, more accurate results were found through utilizing one-hot encoding. For this reason, we only included one-hot encoding, the results of which are shown below.

II. Sentiment

We created four sentiment metrics for each group of documents: average percent positive words per document, average percent negative words per document, percent positive documents, percent negative documents. The sentiment was calculated using a lexicon-based approach. A dictionary of approximately 6,800 positive and negative words, developed by Liu and Hu, was used [18]. These words were stemmed as discussed in the pre-processing steps above to ensure maximal matching between the sentiment word list and the pre-processed text. Average percent positive words was calculated by counting how many positive words appeared in the document, divided by the total word length of the document. The percent word matches for each document in the bin was averaged. The same procedure was followed for negative words.

Percent positive and negative documents was calculated by following the matching procedure outlined above and comparing the values of the number positive words against the number of negative words. The sentiment of each document was determined by the greater value of word matches (positive or negative). The percent positive documents was created by summing the number of documents with a positive sentiment and dividing by the total number of documents in the bin. The same was done for percent negative documents.

SIGNALS

We created four signals to model the GCR-ERT's hypotheses, three of which were related to semantic density and one of which was related to judgment quantification. These features were developed to ascertain the feasible lift from utilizing performative characteristics to predict tolerance.

I. Context Vectors

To capture the variations in semantic density surrounding keywords within religious discourse, we implemented an R-based context vector algorithm based on research by Sagi et al. The context vector-based approach relies upon the distributional vector, a mapping from word w_i to a vector $[x_{i1}, x_{i2}, \dots, x_{iN}]$. This was created by using the *wordspace* package in R [17]. This vector contains information about the specific contexts in which w_i appeared. Most minimally, x_{ij} measures the number of times w_j appeared within a window of w_i. A window can be k words or the same sentence, paragraph, document, etc. x_{ij} can be expanded to include document-level attributes as well. x_{ij} can be absolute counts or can be weighed according to any scheme. A context vector c_i represents information about the context in which a target term w_i appeared. For each of the l occurrences of this target term, the context vector is defined as (1).

$$c_{ij} = \sum_{k \in window(w_i)} [x_{k1}, x_{k2}, \dots, x_{kN}] \qquad (1)$$

A window here is defined in the same manner as for distributional vectors. A key statistic computable from the set of context vectors for w_i, average cosine distance, is defined by (2).

$$\sum_{k=1}^{l-1} \frac{\langle c_{ik}, c_{i(k+1)} \rangle}{||c_{ik}||*||c_{i(k+1)}||} \qquad (2)$$

Referred to as average semantic density, this metric was used in literature to detect changes in meanings of words by attempting to detect widening and narrowing in word contexts. In a similar manner, in our study this method was used to capture the relative flexibility of contexts surrounding keywords [10].

Directly computing average cosine similarity of n context vectors has a complexity of $\Theta(n^2)$. To reduce runtime the we instead estimated it using a Monte Carlo simulation with 1000 iterations.

II. Judgments Numbers

To quantify the judgments hypothesis, the CLDS team utilized a custom algorithm in conjunction with a Maxent POS tagger for the *tm* package to identify judgments in text through syntactic parsing [14]. The output of the algorithm is the average proportion of sentences that act as judgments. The algorithm split each document into sentence segments and word tokens. POS tagging was used on the word tokens to enable the identification of nouns, adjectives, and adverbs, the core components of a judgment. As depicted in Figure I our algorithm used the POS tagging results to flag a combination of a noun, "to be" verb, and adjective or adverb in a sentence.

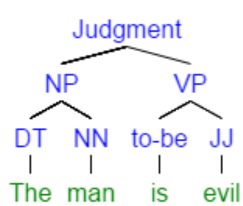

FIGURE I
EXAMPLE OF JUDGMENT SYNTACTIC PARSING

The number of flagged sentences divided by the total number of sentences was calculated for each document. This value was averaged for each document in the bin to generate the final judgment percentage signal.

III. Distributional Scores

The next custom approach relies upon distributional vectors to score the diversity of the set of words which occur around a keyword most frequently. The co-occurrences of all words with respect to the keywords were calculated using the same window as for distributional vectors. The top γ proportion of co-occurrences were subsetted. Using the distributional vector of these words, average cosine similarity was calculated with respect to the same procedure defined in the context vector section above. The distributional score for a bin was calculated as the average cosine similarity for all the identified keywords. $\gamma = 0.5$ was used. Given the smaller vector collections, average cosine similarity was directly calculated.

While the context vector approach seeks to quantify keyword usage diversity based on the context diversity, distributional scores seek to quantify keyword usage diversity based on diversity of the most co-occurring words. This was tried to address a potential weakness of the context vector-based approach: the potential for false diversity to be measured if a word was used in a different context but with the same usage. An example of this could be "building a house" versus "building software". While the contexts of the sentences are very different, the way in which the word is used is very similar.

IV. Network Quantification

Utilizing the distributional vector described in the prior section, we created a graph to estimate semantic density using the *igraph* package in R [20]. We did this by assigning the words within a corpus as nodes and weighting the edges based on the cosine similarity of the word pairs. An algorithm was constructed to extract properties of this graph. First the edges of this graph were removed if they were below a threshold radian χ. This graph was used to calculate subgraph centrality and eigenvector centrality. $\chi = 0.524$ was used.

Subgraph centrality measures the number of subgraphs a vertex participates in, weighing them inversely proportional to the subgraph size exponentiated [19]. Eigenvector centrality scores measure the influence of a network upon a node by looking at the number of influential nodes to which it is connected [19].

MODELING APPROACHES

We explored the predictive power of three modeling approaches. These three models were selected because they are able to handle the bounded continuous response variable, as well as their ability to handle complex input relations.

I. SVM

SVM has been shown to be one of the most powerful classification methodologies available, especially when dealing with noisy data in a high-dimensional space, and was selected to combat the noise present in this data. We modeled this using the *e1071* library in R with a radial kernel and a degree of three [21].

II. Random Forest

Random Forest is also a very robust classification methodology, particularly when there are many features from which to select. As there are a multitude of features, many of which exhibit multicollinearity, this method was selected to reduce this factor of the data. We modeled using the *randomForest* library in R with a tree count of 100 [22].

III. Neural Networks

Lastly, we tested an artificial neural networks modeling approach. Neural nets have been shown to have high prediction performance on complex classification problems, such as image recognition. As many of the signals show complex interactions, such as the context vector, this method was tested. The *neuralnet* library in R was used for this analysis with a threshold value of 0.1, a learning rate of 0.01, and a logistic activation function [23].

RESULTS

We calculated model performance given an allowed error margin of 1, as shown in (3) and (4), to account for the natural fluctuations in linguistic flexibility.

$$acc(bin, model) = \begin{cases} 1, & |\hat{y} - y| \leq 1 \\ 0, & otherwise \end{cases} \quad (3)$$

$$acc(model) = \frac{1}{|bins|} \sum_{bin \in bins} acc(bin, model) \quad (4)$$

Provided in Table II below are the outputs of this metric for each model.

TABLE II
MODEL RESULTS

Model Name	ANN	RF	SVM	Avg.
Topic Models	0.21	0.55	0.55	0.44
Total Sentiment	0.65	0.68	0.71	0.68
Coupled Sentiment – Word Percentage	0.66	0.64	0.70	0.67
Coupled Sentiment – Document Percentage	0.61	0.60	0.62	0.61
Full Model - Semantic	0.51	0.53	0.50	0.51
Distributional Scores	0.66	0.49	0.65	0.60
Context Vectors	0.66	0.49	0.65	0.60
Networks	0.66	0.71	0.76	0.71
Total Judgments	0.72	0.74	0.73	0.73
Percent Judgments	0.68	0.56	0.68	0.64
Number of Judgments	0.66	0.67	0.73	0.69
Full Model - Performative	0.78	0.84	0.80	0.80
Full Model	0.69	0.71	0.55	0.65
Optimal Model	0.80	0.86	0.84	0.83
Average	0.64	0.65	0.68	0.66

As seen above, the semantic models performed with varying degrees of success. Utilizing random forests and SVMs, the topic modeling signal was only able to achieve a prediction accuracy of 55%, with ANN exhibiting an extremely low 21% prediction accuracy. Conversely, sentiment performed much better, with both a coupled signal accuracy of 71% for SVM and 68% for RF. Individually, the word-based sentiment signal performed slightly better, with an average prediction accuracy of 67% versus 61%. Utilizing all semantic signals, the average prediction rate actually decreased to 51%, which was likely due to models overfitting the data.

Conversely, the performative models had similar accuracies to each other. Out of the three semantic density models, the network quantification was the most predictive with a mean prediction accuracy of 71%. The context vector and distributional score methods followed with average prediction accuracies of 60% The judgments signal also proved to be highly predictive with an average prediction rate of 73%. Individually, the number of judgments outstripped the percent judgments with average prediction accuracies of 69% and 64%, respectively. Utilizing all performative signals, there was an additional boost in performance, with an average prediction accuracy of 80%.

In an effort to ascertain the effectiveness of combining both semantic and performative models, we created a full model using all parameters. While more accurate than the full semantic model, it actually performed worse than the full model containing only performative features. Since topic modeling proved to be a poor indicator of language rigidity, and multi-collinearity was found between the two sentiment signals, an optimal model was created containing the judgments, semantic density and word-based sentiment signals. Using this model, there was an increase in average accuracy to 83%.

As seen in Figure II, this model did a good job of predicting across classes as most of the classes were accurately classified within their proper bin. However, classes near the edge, especially linguistic flexibility of 8 and 2, were less accurately classified. This is likely due to the heavily unbalanced classes at the edges and would likely be mitigated by more data in these classes.

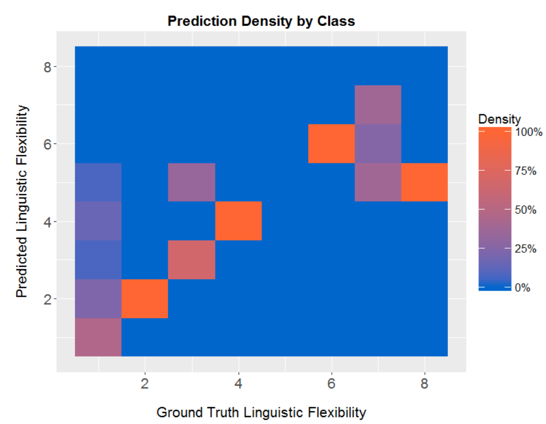

FIGURE II
HEAT MAP OF PREDICTED ACCURACY BY CLASS

CONCLUSIONS

As the number of incidents of religious violence continues to increase, the need for scalable and accurate prediction systems intensifies. Analysis of the semantic characteristics of language has shown promise in prior research and the manual evaluation of performative characteristics by the GCR-ERT, has produced good predictions of flexibility of keywords in religious discourse. The work in this paper has significantly extended these previous results and shown that automated systems can be developed to assist in modeling the language flexibility of religious discourse. Importantly, our results also show that performative characteristics of language yield better prediction signals than semantic characteristics of language.

While these results show promise, additional work is needed. As only base R parameters were used to develop the outputs, it would be ideal to tune both the model signals and parameters to develop a more robust classification system. Secondly, a wider array of documents needs to be analyzed with a greater variety of religious groups in each class. This would prevent the algorithms from over-fitting based on group specific tendencies and illustrate the performative characteristics of these group's discourse. It would also be advantageous to attempt a similar analysis on discourses written in different languages to ascertain if this is a global phenomenon versus an English-language phenomenon.

ACKNOWLEDGMENT

We would like to thank Professor Peter Ochs, Jonathan Teubner, Nauman Faizi, Syed Moulvi, and Patrick Derdall of the University Of Virginia's Religious Studies department for their critical insights. The work for this paper was partially supported by a grant from the U.S. Army Research Laboratory # W911NF-15-2-0044.

REFERENCES

[1] D. Searcey and M. Santora, "Boko Haram Tops ISIS in Ranking Terror Groups," *New York Times*, p A6, Nov. 14, 2015.

[2] J. White, "Strategic peacebuilding: addressing Abrahamic atrocities in the Middle East & Europe," presented at the *University of Virginia,* Charlottesville, VA, 2015.

[3] K. Hacker, D. Boje, V. Nisbett, A. Abdelali and N Henry, "Interpreting Iranian leaders' conflict framing by combining latent semantic analysis and pragmatist storytelling theory," presented at the *Political Communication Division of the National Communication Association annual conference,* Washington, DC, 2013.

[4] M. Yang, S. Wong, and J. Coid, "The efficacy of violence prediction: a meta-analytic comparison of nine risk assessment tools," *Psychological bulletin*, vol. 136, no. 5, p. 740, 2010.

[5] J. Bullinaria and J. Levy, "Extracting semantic representations from word co-occurrence statistics: A computational study," *Behavior research methods,* vol. 39, no. 3, pp. 510–526, 2007.

[6] M. Hearst, "Automatic acquisition of hyponyms from large text corpora," *Proc. 14th Conf. Computational linguistics,* vol. 2, 1992, pp. 539–545.

[7] H. Kozima, "Text segmentation based on similarity between words," *Proc. 31st Annual Meeting Association Computational Linguistics*, 1993, pp. 286–288.

[8] D. Lin, "Automatic retrieval and clustering of similar words," *Proc. 17th Int. Conf. Computational linguistics*, vol. 2, 1998, pp. 768–774.

[9] L. Barsalou, "Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols," *Theories of memory*, vol. 1, Lawrence Erlbaum Associates, 1993, pp. 29-31.

[10] E. Sagi, S. Kaufmann, and B. Clark, "Semantic density analysis: comparing word meaning across time and phonetic space," *Proc. Workshop Geometrical Models Natural Language Semantics*, Stroudsburg, PA, 2009, pp. 104–111.

[11] A. Boussidan and S. Ploux, "Using topic salience and connotational drifts to detect candidates to semantic change," *Proc. Ninth Int. Conf. Computational Semantics*, 2011, pp. 315–319.

[12] P. Cook and S. Stevenson, "Automatically identifying changes in the semantic orientation of words," *Proc. Language Resources and Evaluation Conf.,* 2010, pp.28-34.

[13] L. Richardson, *Beautiful soup documentation.* [Online]. Available: https://www.crummy.com/software/BeautifulSoup/bs4/doc/

[14] K. Hornik and I. Feinerer, *tm: Text Mining Package. R package version 0.6-1.* [Online]. Available: http://cran.r-project.org/web/packages/tm/.

[15] K. Hornik and B. Grün, *Package 'topicmodels'.* [Online]. Available: https://cran.r-project.org/web/packages/topicmodels/topicmodels.pdf

[16] D. Blei, A. Ng, and M. Jordan, "Latent Dirichlet Allocation," *The Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[17] S. Evert, "Distributional Semantics in R with the wordspace package*," Proc. 25th Int. Conf. Computational Linguistics: System Demonstrations*, Dublin, Ireland, 2014, pp. 110–114.

[18] M. Hu and B. Liu, *A list of positive and negative opinion words or sentiment words for English*. [Online]. Available: https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon.

[19] E. Estrada and J. Rodríguez-Velázquez, "Subgraph centrality in complex networks," *Phys. Rev. E*, vol. 71, no. 5, p. 56103, May 2005.

[20] G. Csardi and T. Nepusz, *Package 'igraph'.* [Online]. Available: https://cran.r-project.org/web/packages/igraph/igraph.pdf

[21] D. Meyer, E. Dimitriadou, A. Weingessel, F. Leisch, C. Chang, and C. Lin., *Package 'e1071'.* [Online]. Available: https://cran.r-project.org/web/packages/e1071/e1071.pdf

[22] A. Liaw and M. Wiener, *Package 'randomForest'.* [Online]. Available: https://cran.r-project.org/web/packages/randomForest/randomForest.pdf

[23] F. Günther and S. Fritsch, *Package 'neuralnet'.* [Online]. Available: https://cran.r-project.org/web/packages/neuralnet/neuralnet.pdf

AUTHOR INFORMATION

Nicholas Venuti, M.S. Student, Data Science Institute, University of Virginia.

Brian Sachtjen, M.S. Student, Data Science Institute, University of Virginia.

Hope McIntyre, M.S. Student, Data Science Institute, University of Virginia.

Chetan Mishra, B.A./B.S. Student, Applied Statistics & Systems Engineering, University of Virginia.

Matthew Hays, M.S. Student, Data Science Institute, University of Virginia.

Donald E. Brown, Director, Data Science Institute and William Stansfield Calcott Professor, University of Virginia