

Author Prop: Assisting the Creative Process with an Automated Intelligent Cognitive Prop for Writers

C. Roberts^{1,*}

¹Georgia Institute of Technology, kopant@gmail.com

Abstract

INTRODUCTION: Writing is a complex, metacognitive task which requires planning, execution, and evaluation of an evolving text simultaneously.

OBJECTIVES: This paper presents Author Prop, a web application for writing that aims to address the need for “cognitive props” during composition. The software particularly targets students, whose relative inexperience may lead to high levels of cognitive load and low engagement during writing.

METHODS: To provide a mental prop to writers, Author Prop provides topic-based decomposition of texts using intelligent highlighting. Furthermore, the software allows users to interact with topics (described as keyword lists) during periods of reflection, with the aim to ease and maintain creative thinking.

RESULTS: Of the 78 surveyed users, 42% said they would use the app for writing, and 70% said they enjoyed using the app.

CONCLUSION: Initial survey results from beta testing of the web application indicate encouraging reception of the software.

Keywords: writing, motivation, natural language processing, intelligent tutoring system, writing analytics, education, AI, software.

Received on 30 May 2019, accepted on 28 June 2019, published on 16 July 2019

Copyright © 2019 C. Roberts *et al.*, licensed to EAI. This is an open access article distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/3.0/>), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/eai.30-7-2019.162945

*Corresponding author. Email: kopant@gmail.com

1. Introduction

1.1. Student Engagement During Writing

Low engagement can be a problem when students participate in extended solo activities like writing. Students can have trouble remaining focused and engaged during the writing process. As well as being a solitary activity, writing is a complex task requiring extensive planning, awareness, and metacognition on the part of the writer [1, 2]. Students that write essays or articles can experience disconnectedness with their written material. Bereiter and Scardamalia’s Compare, Diagnose, and Operate model of writing

postulates that writers edit material by comparing their own mental model of what they planned to write with what they have actually written [1]. As the complexity of the text grows, writers may have trouble understanding whether or not what the text is communicating matches their intentions. They may also have trouble with editing or ordering sections of text to maintain coherence and relevance. This phenomenon can be summarized by saying that students experience high cognitive load when writing (cognitive load describes a situation where learners’ cognitive processing capabilities are overloaded during complex problem solving to the degree that they have no resources left for actually learning from the problem) [3]. Writing, which involves high levels of self-regulation, benefits from effective metacognitive strategies [2]. The proposed web application (“web app”), Author Prop, aims to provide writers with a

cognitive support that will lessen cognitive load through increased engagement and alternate ways of reflecting on or reconsidering the evolving state of document content during drafting.

1.2. Natural Language Processing Background

Natural language processing (NLP) is an area of artificial intelligence that focuses on taking structured or unstructured text or other language data, parsing it in a way that computers can understand, and conducting useful analyses on the data [4]. One sub-field of NLP is topic modelling, which typically works with written text data to extract the topics associated with a corpus of documents.

Topic models are probabilistic graphical models that postulate the occurrence of words in documents as a stochastic process governed by statistical distributions over unknown, “latent” topic variables. The values of these unknown topics determine which words are found in which document [5, 6]. Topic models are typically characterized by lists of keywords that characterize that topic. Fig. 1 shows a graphic illustrating topic modelling using Latent Dirichlet Allocation (LDA), one of the most widely used topic modelling algorithms.

2. Related Work

While NLP for student writing improvement is a relatively immature field, examples of related work exist. The largest area where NLP has found application is in software aimed at helping students improve essay structure and writing style. These products may be focused on a specific type of writing such as college academic or legal writing [7]; they are often designed to automatically detect and teach a predefined set of characteristics associated with “good”

essays. For instance, AcaWriter [8] automatically identifies patterns called “rhetorical moves” that indicate use of specific writing techniques such as use of emphasis and contrast in higher degree research writing; the tool offers feedback to students on essay rhetorical structure, such as the need for additional background information before presenting a research problem.

Fewer products use NLP with the aim to specifically improve student engagement and reflection (as opposed to improving essay structure); products that serve this niche typically aim to improve engagement and reflection in tandem with improving writing style. Few products take specific aim at the role of reflection in the writing process itself as a support for the cognitive load students may face while structuring their ideas during drafting.

Burstein et al [9] present an NLP-driven writing system that coaches students using feedback on essay qualities such as convincingness, clarity, and coherence. Their system flags student text with feedback and tips on techniques that can help improve essay quality. As part of the feedback, students are invited to review the coherence of topics automatically found in the essay. Topics, characterized by keywords identified either by the student directly or automatically by the system, are described using a cluster of semantically similar words from the student’s text. One of the authors’ explicit goals of their system is to “promote deep engagement and revision aimed at boosting the quality of students’ writing across genres and academic disciplines.”

Villalon and Calvo’s intent in developing Concept Map Miner [10] is much closer to Author Prop’s aims than other tools. Concept Map Miner automatically generates concept maps, semantically connected graphical structures similar to knowledge graphs, from student papers with the intent to engage students on a meta-cognitive level, triggering student reflection and understanding of draft papers. However concept maps as presented in Concept Map Miner end up being more low-level representations of paper concepts than

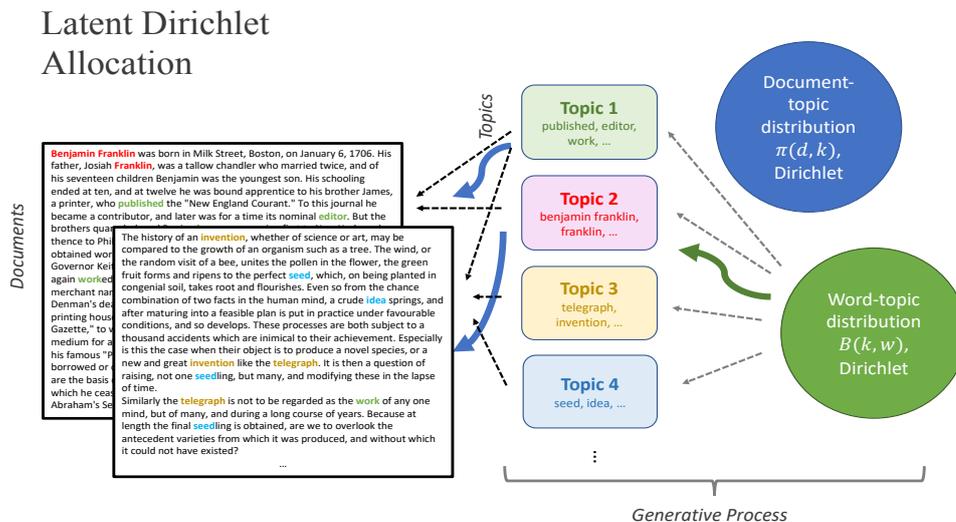


Figure 1. Representation of topic modelling using LDA

typical results from topic modelling, due to the reliance on parts of speech grammar to extract and connect concepts at the sentence level. In contrast, topic modelling frames evidence for concepts as groups of words or word phrases, allowing for more abstract interpretation on the part of student writers.

Whitelock et al's OpenEssayist system [11] aims to provide feedback to students on completed essay drafts to improve student reflection for understanding, motivation, and writing style. Feedback consists of feedback on essay structure (introduction, discussion, and conclusion) organized through automatic identification of essay keywords, interactive organization of sentences and key words, and overall visualization of word clouds and word dispersion through the parts of the essay. As opposed to using topic modelling, Whitelock et al use unsupervised graph-based ranking algorithms based on the TextRank approach from Mihalcea and Tarau [12] to extract keyword-oriented essay concepts. The OpenEssayist tool has been used in the UK's Open University system, and the authors have linked use of the tool to improved course overall grade [11]. Interestingly, the authors find that structure-based feedback of the kind provided by OpenEssayist can be as effective as content-based feedback in aiding student achievement [13].

While Author Prop is similar in intent to OpenEssayist and Concept Map Miner, it differs from these and other tools by:

- Focusing specifically on partial or in-progress student drafts as opposed to completed drafts
- Principally aiming to improve student cognitive load and engagement during drafting

- Being a general-purpose tool for any kind of writing, as opposed to a specific type of writing such as college academic writing

3. Solution

The Author Prop web app presents users with two tabs, Edit and Review.

The purpose of the Edit tab is for users to compare their entered text with a copy of their text highlighted according to the found topics. Text is highlighted with a different colour for each topic found in the text.

On the Edit tab, users are presented with a text area on the left, and an initially blank pane on the right. Placeholder text prompts users to either paste in existing text, or start writing in the text area. Once they have entered enough text to analyse, users can click the 'Analyse' button to trigger topic modelling. When the model is finished running, highlighted text populates the pane on the right-hand side.

The Edit tab was designed to display topic modelling results to users in a way that would not be distracting to ongoing editing. As such, alternate visual representations of topics were deferred to the Review tab. On the Edit tab, users can choose to ignore the highlighted results as they focus on writing; alternately they can quickly refer to model results by scanning through the colours in the mirror-image text on the right-hand side. Further text entered after a topic model is run will be highlighted according to the rules of the latest model. Users can click 'Analyse' at any point to update model results and highlighting.

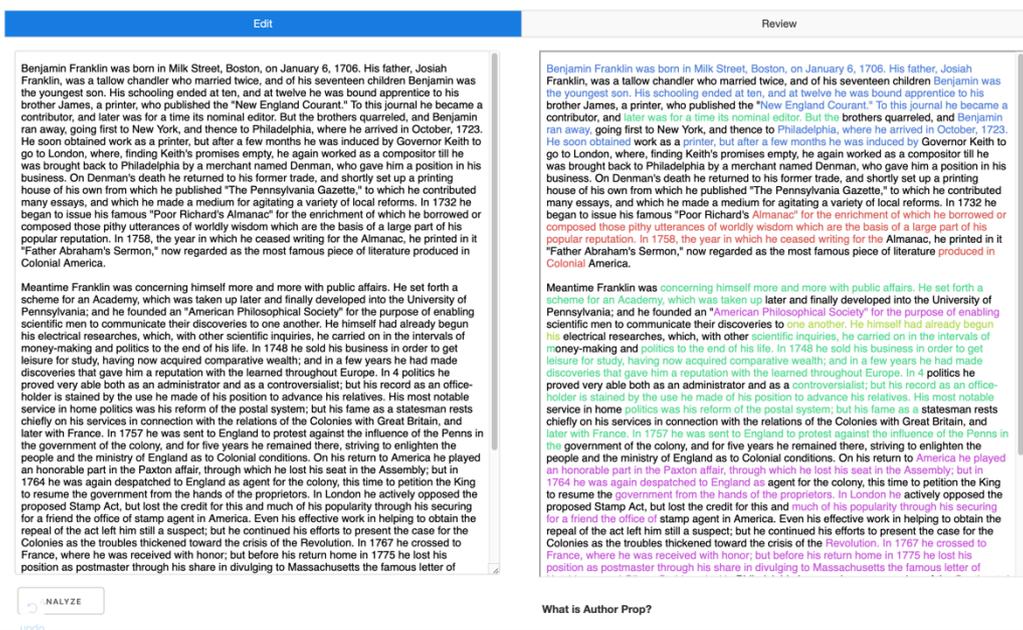


Figure 2. Author Prop's Edit Tab

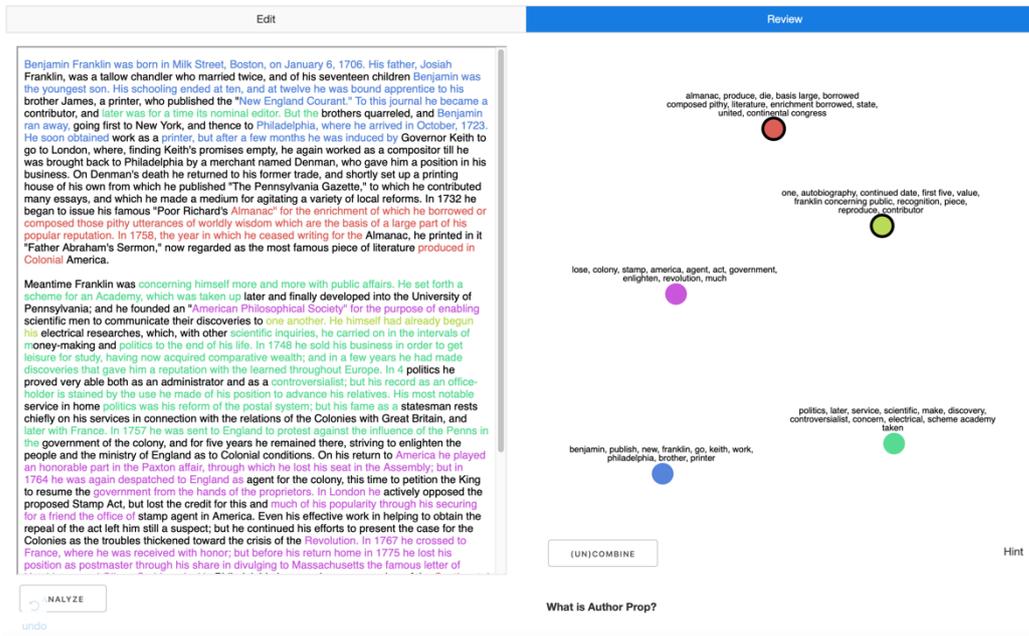


Figure 3. Author Prop’s Review Tab

The purpose of the Review tab is for users to reflect on their in-progress draft, and compare the topic model’s characterization of topics found in their paper with their own. Again, the tab is made up of two panes: on the left-hand side, users are presented with their highlighted text, matching the right-hand side of the Edit tab. On the right-hand pane, users are presented with a visualization of the individual topics found by the topic modelling algorithm. The visualization depicts topics as coloured bubbles associated with keyword lists. The topics are coloured to match the highlighted text shown at left and on the Edit pane. As a ‘Hint’ tooltip describes, users can ‘turn off’ topics, which greys out the topic bubble and adjusts the highlighted text for the loss of that topic. Users can also group/ungroup topics by dragging topics to select them and clicking the ‘(Un)Combine’ button. This combines two topics into a single bubble/colour, and aggregates their keyword lists. Topics are moveable: users can drag them around the right-hand pane area to associate them spatially.

After arriving at a highlighting they are satisfied with, users can then return to the Edit tab to return working with the updated highlighting.

Thus, users proceed in a cyclical fashion of editing and reflection, utilizing the Author Prop software as a cognitive prop.

4. Methodology

4.1. Natural Language Processing Methods

Text Pre-Processing

Author Prop follows a basic initial NLP pipeline of tokenizing, removing stopwords, and lemmatizing text [14, 15]. Tokenization refers to extracting words from documents by looking for whitespace or punctuation. Stopwords are very commonly occurring words like “a” and “the”; because such words occur very frequently, they can dominate found topics without adding value to results. Lemmatization refers to normalizing words across different inflectional forms such as tense or plurality. After this initial pre-processing, word combinations called n-grams are also created by looking at commonly occurring pairs and triples of words [14, 15].

The words describing a document are further filtered using the ‘term frequency-inverse document frequency’ (TF-IDF) metric [14]. TF-IDF provides a ranking of word importance in a specific document relative to the corpus as a whole, and is calculated as the ratio of the frequency with which that word occurs in the specific document relative to the frequency with which it occurs in the corpus.

Because the proposed tool is intended to work with single documents as opposed to a corpus of documents, Author Prop treats paragraphs as “sub-documents” comprising a ‘corpus’ in the larger text. Sub-documents are also subjected to minimum and maximum word length requirements, and are grouped consecutively or broken apart if necessary, to increase the stability of topic modelling results.

Topic Modelling Algorithms

The topic modelling algorithm used by Author Prop is Latent Dirichlet Allocation (LDA) [16]. Multiple models were tested, including LDA, hierarchical LDA (hLDA) [17], and Independent Components Analysis (ICA). The ICA approach is similar to Grant, Skillicorn, and Cordy [18]. Hyper-parameter testing suggested that hLDA actually

performs better than LDA for documents of longer length, for a certain range of minimum/maximum sub-document length; however, runtimes for the particular implementation of hLDA tested were significantly longer than LDA and deemed unacceptable for a web application.

Regardless of the algorithm used, the topic models tested take as input a document-word matrix of counts of words from the text pre-processing pipeline described in Section 4.1.1; one sub-document represents one row in the matrix. The output of the topic models, as mentioned in Section 1.2, are keyword lists characterizing topics found by the algorithm.

Hyper-Parameter Testing

Hyper-parameter testing was conducted to determine optimal combinations of text pre-processing values for minimum and maximum sub-document length, TF-IDF filter percentage, parts of speech to lemmatize, as well as topic modelling algorithm itself (as mentioned above). A grid search over the parameter space was conducted using a range of sample documents including at least five each from the Automated Student Assessment Prize dataset [19], a dataset of academic papers from the Neural Information Processing Systems conference [20], a dataset of New York Times articles [21], and Hans Christian Andersen fairy tales [22]. Hyper-parameters were selected based on average performance under the following metrics: distance from corpus, prominence within documents, and TC-PMI coherence [14]. The interested reader is referred to [14] for a discussion of these and other topic modelling metrics. Hyper-parameter testing results are shown in Appendix A: Web App NLP Hyper-parameter Testing Results.

4.2. Web App Interface

The web app utilizes the Plotly Dash[†] framework to integrate the python-based topic modelling pipeline with HTML and React JavaScript web frameworks. Dash utilizes the Flask web framework.

The python-based topic modelling pipeline utilizes the Natural Language Toolkit (NLTK)[‡] library for text pre-processing, as well as the lda python package[§] implementation of the LDA algorithm.

5. Results

The Author Prop web app is hosted at: <https://author-prop.herokuapp.com/>

Author Prop was beta tested via an eight-question survey, which involved testing the web app. Survey respondents were solicited through posting the survey on a class website (where participants get class participation points for completing the survey), through email to community writing

organizations, and by posting the survey on the neighbourhood-based social media site Nextdoor^{**} (hypothetically visible by over 23,000 potential users residing in zip codes 94121, 94129, and 94118). 78 responses were received for the survey; questions and aggregated responses are detailed in Appendix B: Survey Results. SurveyMonkey^{††} was used to generate descriptive statistics and graphs of the results. In addition, predictive models were fit to the responses for Question 6, “Was the app easy to use,” Question 7, “Did you enjoy using the app,” and Question 8, “Would you use the app for writing,” in order to generate inferential statistics about what types of writers are more receptive to the app.

5.1. Descriptive Statistics

Survey results show that the majority of users (60%) describe themselves as having above average (score of 4 or above on a scale of 1-5) writing experience, but the same population as a whole self-reports only slightly above average writing skill (38% with an average score of 3, and 45% with scores of 4 or above). Those with higher levels of writing experience are more likely to say that they enjoy writing, and to describe themselves as highly skilled.

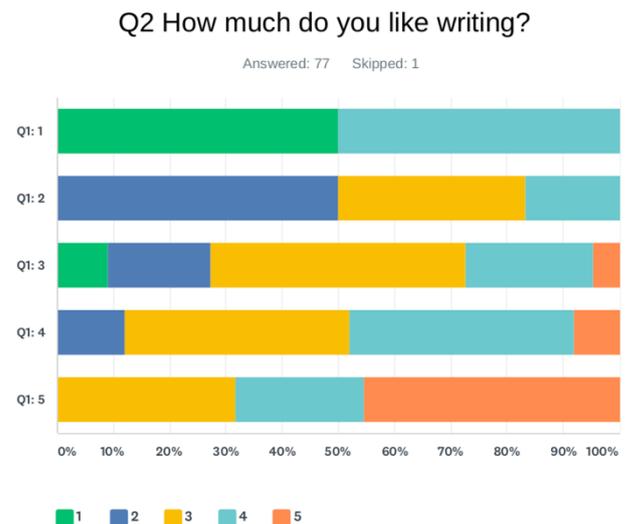


Figure 4. Question 2 split by scores to “Q1. How much experience do you have with writing overall, on a scale of 1-5, with 5 being the highest?”

[†] <https://plot.ly/dash/>

[‡] <https://www.nltk.org>

[§] <https://lda.readthedocs.io/en/latest/>

^{**} <http://nextdoor.com>

^{††} <http://www.surveymonkey.com>

Q3 How skilled do you think you are at writing?

Answered: 78 Skipped: 0

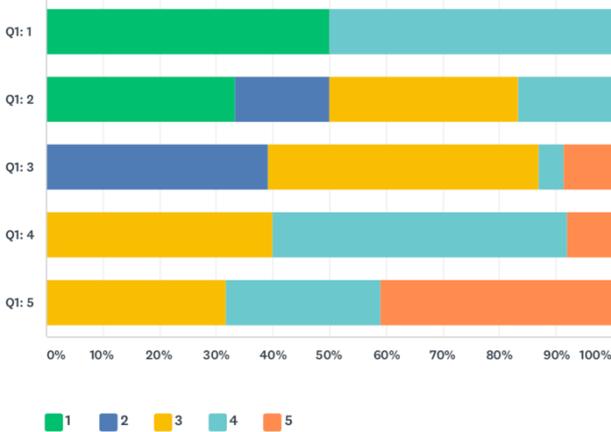


Figure 5. Question 3 split by scores to “Q1. How much experience do you have with writing overall, on a scale of 1-5, with 5 being the highest?”

Notable correlations between variables collected include the correlation between writing skill and writing experience at 0.58, the correlation between writing experience and writing enjoyment at 0.50, and the correlation between writing skill and writing enjoyment at 0.38.

In terms of reception, the majority of users (79%) enjoyed using the app (score of 3 or above on a scale of 1-5). Similarly, 71% thought the app was easy to use, and 39% responded positively when asked if they would use the app for writing. In terms of high correlations, the correlation between enjoying the app and finding the app easy to use was 0.67, the correlation between responding that one would use the app and one’s rating of app enjoyment is 0.54, and the correlation between responding one would use the app and finding the app easy to use is 0.29. Finding the app easy to use is evidently not the same thing as finding it useful for writing; however, one can rate lower ease of use scores and still find the app enjoyable and useful for writing.

One interesting subset of respondents is those users that prefer poetry to prose. While these respondents comprise a small sample (8 in total), 75% of them respond that they would use the app for writing (well over the overall average), even though they are fairly evenly split across the spectrum as to whether or not they enjoyed using the app (for the general population, the latter is strongly correlated with intention to use the app for writing or not). The explanatory models described in Section 5.2 confirm that preferring poetry to prose is in general a strong predictor of positive app reception, in particular intending to use the app for writing.

Q7 Did you enjoy using the app?

Answered: 76 Skipped: 1

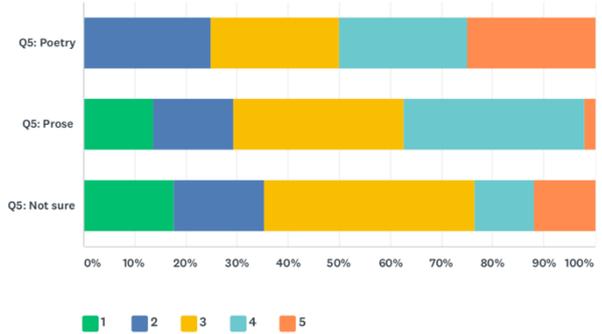


Figure 6. Question 7 split by scores to “Q5. Which do you prefer, poetry or prose?”

Q8 Would you use the app for writing?

Answered: 76 Skipped: 1

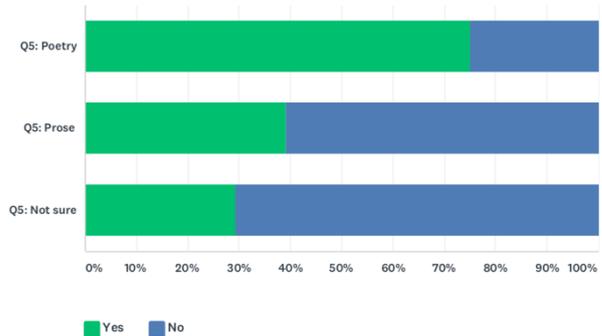


Figure 7. Question 8 split by scores to “Q5. Which do you prefer, poetry or prose?”

One hypothesis might be that Author Prop’s abstract visualization of document topics mirrors the abstract visualization associated with the creation and appreciation of poetry.

Overall, app reception is more positive among respondents that describe themselves as having higher levels of writing experience and skill.

Q7 Did you enjoy using the app?

Answered: 77 Skipped: 1

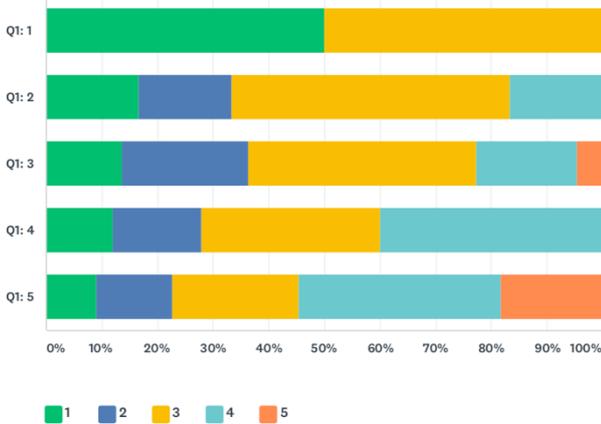


Figure 8. Question 7 split by scores to “Q1. How much experience do you have with writing overall, on a scale of 1-5, with 5 being the highest?”

Q7 Did you enjoy using the app?

Answered: 77 Skipped: 1

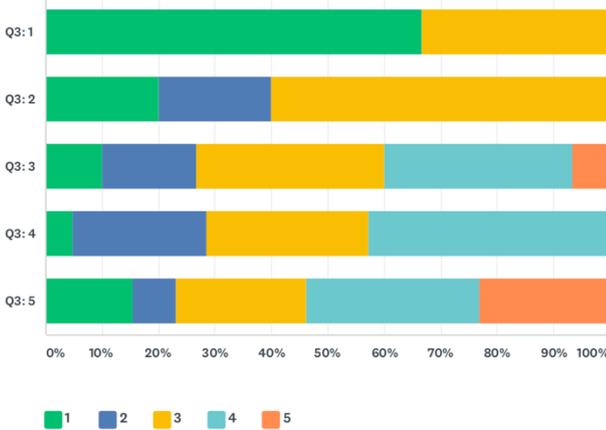


Figure 9. Question 7 split by scores to “Q3. How skilled do you think you are at writing?”

5.2. Explanatory Modelling and Inference

In order to enable objective and reproducible interpretations of the survey data results, and because of high correlations between variables noted above, models were used to facilitate explanatory inference.

Methodology

A series of models were fit to the survey data, to separately predict responses to Question 6, “Was the app easy to use,” Question 7, “Did you enjoy using the app,” and Question 8, “Would you use the app for writing.”

For each response variable, a set of model forms were tested, which included Logistic Regression, Random Forest,

Gaussian Naïve Bayes, and Gradient Boosted (tree-based) Classifier for classification (Question 8), and Linear Regression, Random Forest, and Gradient Boosted (tree-based) Regressor for regression (Questions 6 and 7). The Scikit-learn package [27] was used for modelling in python; the Statsmodels python package [28] was also used secondarily for Linear and Logistic Regression.

While learning curves with cross-validation were used to examine the sensitivity of the tested models to data size, due to small sample size and the explanatory (as opposed to predictive) purpose of the modelling, models were fit to the entire dataset for the purpose of explanatory inference [29]. For classification for Question 8, the predicted probability that someone would respond “Yes” was thresholded at the population sample mean for the purpose of classifying predicted probabilities as “Yes” or “No.”

In all cases, the Gradient Boosted Classifier/Regressor (GBM) outperformed the other model forms in terms of model validation metrics; ROCAUC was used as a metric for classification and MSE/MAE/explained variance score for regression. These models were selected for the purpose of explanatory inference.

To bolster confidence in variable-level inference, model validation metrics were examined across levels of individual variables to assess whether or not there were systematic patterns wherein certain variables were over or underpredicted; no significant evidence that would weaken the findings below was found.

Findings

For Question 6, “Was the app easy to use (scale of 1-5 with 5 highest)” the GBM achieved a MSE of 1.01, MAE of 0.75, and explained variance score of 35%. The most important features in this model using the tree-based feature importance metric were writing skill (0.28), writing enjoyment (0.20), writing experience (0.19), prefer poetry or prose (0.11), and the control variable for whether or not the survey began after initial improvements (0.11). As shown in the partial dependence plots in Figure 28 in Appendix C, people with high writing skill and either medium to low writing enjoyment or medium to low writing experience are more likely to find the app easy to use. We can confirm that this group, which represents 21% of the total sample, has an average ease of use rating of 3.78 compared to the overall average of 3.19. The standard deviation of the ease of use rating of the cohort is comparable to the population as a whole, 1.31 compared to 1.26.

For Question 7, “Did you enjoy using the app (scale of 1-5 with 5 highest)” the GBM achieved a MSE of 0.82, MAE of 0.75, and explained variance score of 34%. The most important features in this model using the tree-based feature importance metric were writing skill (0.27), writing enjoyment (0.20), writing experience (0.19), and the control variable for whether or not the survey began after the app was released to the Nextdoor website (0.12). As shown in the partial dependence plots in Figure 29 in Appendix C, people with either high writing skill or high writing experience or both are more likely to more highly rate enjoying the app. We can confirm that this group, which

represents 66% of the total sample, has an average enjoyment rating of 3.22 compared to the overall average of 3.00. The standard deviation of the enjoyment rating of the cohort is comparable to the population as a whole, 1.14 compared to 1.12.

For Question 8, “Would you use the app for writing (Yes/No),” the GBM achieved a ROCAUC of 0.82 on the training set. The true positive rate is 0.56, and the true negative rate is 0.78. The most important features in this model using the tree-based feature importance metric were writing skill (0.27), writing enjoyment (0.22), control for whether or not the survey began after initial improvements (0.16), and writing experience (0.14). As shown in the partial dependence plots in Figure 30 in Appendix C, people with either very high writing skill or the combination of medium writing enjoyment and high writing experience are more likely to say they would use the app for writing. We can confirm that this group, which represents 38% of the total sample, has an average “Yes” probability of 0.52 compared to the overall average of 0.42. The standard deviation of the “Yes” classification of the cohort is comparable to the population as a whole, 0.51 compared to 0.50.

6. Limitations

Survey results are somewhat biased by time; as survey responses were received, changes were made to the app UI to improve user experience. In particular, nearly half of users requested some form of explanation as to the purpose of the app (which was not explicitly stated on the website), and as to how they were expected to use the app. An additional popup, prompt text, and expanded tooltip were implemented to address these concerns. A control variable to indicate whether or not a survey response was begun after these initial improvements were made was introduced into the explanatory models to attempt to normalize for these effects.

Survey results are also biased by respondents, of which there are two main types: computer science graduate students, and members of the general population. If we infer that the general population responded most after the survey was initially released to them, we see that the general population respondents respond more positively to the app than the computer science graduate students. They also exhibit a self-selection bias of higher levels of writing experience and enjoyment. A control variable to indicate whether or not a survey response was begun after the survey was made available to the general population was introduced into the explanatory models to attempt to normalize for these effects.

How much experience do you have with writing overall, on a scale of 1-5, with 5 being the highest?

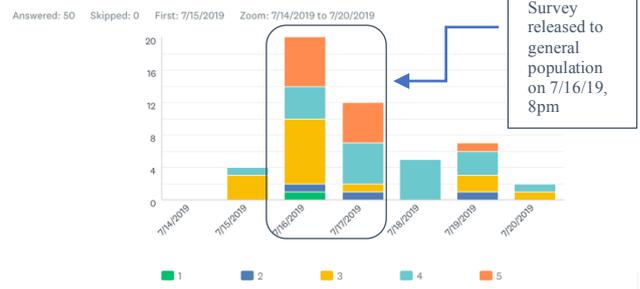


Figure 10. Time trend, Question 1. More respondents report higher levels of experience with writing on 7/16 – 7/17.

Was the app easy to use?

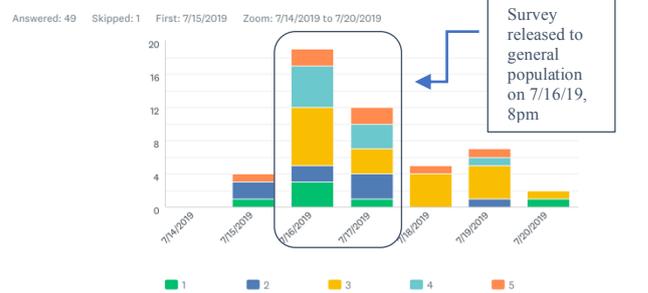


Figure 11. Time trend, Question 6. More respondents report the app being easy to use on 7/16 – 7/17.

Did you enjoy using the app?



Figure 12. Time trend, Question 7. More respondents report enjoying the app on 7/16 – 7/17.

7. Conclusion

Author Prop provides students with a tool that allows them to compare their mental maps of in-progress text drafts with topic model results from natural language processing. The web app provides modes for both editing and review/reflection, and allows users to refer to model results at their own pace. The review mode encourages users to think creatively by interacting with and informing alternate representations of their texts. The web app acts both as a cognitive prop and a tool to increase student engagement

during writing. Though survey results are limited, feedback has been encouraging on the app experience/concept, with refinement desired on instructions for using the app.

Reflecting on the survey results, it is interesting to note the effect that higher levels of self-identified writing skill and experience have on positive reception of the app. It is also interesting and somewhat surprising that the app seems to appeal more to those with medium to low levels of self-rated writing enjoyment, even though those that enjoy writing more tend to have higher levels of writing experience. These results may indicate a potential market for Author Prop; in the larger sense, it would be an area of future research to investigate how other tools aimed at improving the writing process have achieved similar or different success with related methods. It is interesting to reflect on what the results say about the nature of the writing process itself: how exactly do humans manage the cognition of writing, and what kinds of abstractions or visualizations are useful cognitive props? Through research with AI-based

tools, we may be able to provide evidence for theories about the cognition of writing.

8. Future Work

As noted by users, the app would benefit from a short animated introduction walking first-time users through how to use the app. Requests for better explanation of the web app's purpose and expected usage continued to be a prevalent theme in user-submitted comments even after adjustments were made to the app to explain functionality. Future work thus includes additional testing and research into intuitive user interfaces for the app, as well as better ways to guide first-time users through app usage. Further research into the app's reception is also desired.

Appendix A. Web App NLP Hyper-parameter Testing Results

Hyper-parameter and topic modeling algorithm grid search results; higher is better
Short documents (< 3,000 words)

Sub-doc min/max length; TF-IDF filter %; Lemmatized parts of speech*	Average of Distance from corpus dist			Average of Doc prominence: % largest			Average of TC-PMI coherence		
	hlda	ica	lda	hlda	ica	lda	hlda	ica	lda
100/200	0.69	0.69	0.72	0.70	1.00	0.42	6.38	10.12	13.62
20	0.69	0.69	0.72	0.70	1.00	0.42	6.51	10.12	13.62
J/N/N/R	0.68	0.69	0.72	0.70	1.00	0.42	7.35	10.12	13.62
V/N	0.68	0.69	0.72	0.70	1.00	0.42	6.54	10.12	13.62
40	0.69	0.69	0.72	0.70	1.00	0.42	5.64	10.12	13.62
J/N/N/R	0.68	0.69	0.72	0.70	1.00	0.42	6.62	10.12	13.62
V/N	0.68	0.69	0.72	0.70	1.00	0.42	6.54	10.12	13.62
60	0.69	0.69	0.72	0.70	1.00	0.42	6.89	10.12	13.62
J/N/N/R	0.69	0.69	0.72	0.70	1.00	0.42	6.43	10.12	13.62
V/N	0.69	0.69	0.72	0.70	1.00	0.42	6.03	10.12	13.62
100/400	0.65	0.64	0.68	1.00	1.00	0.60	9.53	15.52	9.14
20	0.63	0.64	0.68	1.00	1.00	0.60	9.50	15.52	9.14
J/N/N/R	0.65	0.64	0.68	1.00	1.00	0.60	11.99	15.52	9.14
V/N	0.61	0.64	0.68	1.00	1.00	0.60	9.15	15.52	9.14
40	0.64	0.64	0.68	1.00	1.00	0.60	7.35	15.52	9.14
J/N/N/R	0.66	0.64	0.68	1.00	1.00	0.60	10.34	15.52	9.14
V/N	0.69	0.64	0.68	1.00	1.00	0.60	7.50	15.52	9.14
60	0.65	0.64	0.68	1.00	1.00	0.60	17.48	15.52	9.14
J/N/N/R	0.63	0.64	0.68	1.00	1.00	0.60	6.04	15.52	9.14
V/N	0.65	0.64	0.68	1.00	1.00	0.60	8.75	15.52	9.14
50/200	0.70	0.71	0.73	0.90	1.04	0.54	7.07	8.40	14.75
20	0.70	0.71	0.73	0.90	1.04	0.54	6.58	8.40	14.75
J/N/N/R	0.70	0.71	0.73	0.90	1.04	0.54	7.21	8.40	14.75
V/N	0.70	0.71	0.73	0.90	1.04	0.54	5.62	8.40	14.75
40	0.70	0.71	0.73	0.90	1.04	0.54	6.92	8.40	14.75
J/N/N/R	0.70	0.71	0.73	0.90	1.04	0.54	7.52	8.40	14.75
V/N	0.70	0.71	0.73	0.90	1.04	0.54	7.57	8.40	14.75
60	0.70	0.71	0.73	0.90	1.04	0.54	7.12	8.40	14.75
J/N/N/R	0.70	0.71	0.73	0.90	1.04	0.54	7.12	8.40	14.75
V/N	0.70	0.71	0.73	0.90	1.04	0.54	6.49	8.40	14.75
50/400	0.72	0.70	0.74	1.22	1.13	0.73	3.06	3.79	11.38
20	0.72	0.70	0.74	1.22	1.13	0.73	3.06	3.79	11.38
J/N/N/R	0.72	0.70	0.74	1.22	1.13	0.73	2.14	3.79	11.38
V/N	0.72	0.70	0.74	1.22	1.13	0.73	3.80	3.79	11.38
40	0.72	0.70	0.74	1.22	1.13	0.73	1.82	3.79	11.38
J/N/N/R	0.72	0.70	0.74	1.22	1.13	0.73	1.25	3.79	11.38
V/N	0.72	0.70	0.74	1.22	1.13	0.73	3.43	3.79	11.38
60	0.72	0.70	0.74	1.22	1.13	0.73	0.79	3.79	11.38
J/N/N/R	0.72	0.70	0.74	1.22	1.13	0.73	0.75	3.79	11.38
V/N	0.72	0.70	0.74	1.22	1.13	0.73	1.49	3.79	11.38
J/N/N/R	0.72	0.70	0.74	1.22	1.13	0.73	0.89	3.79	11.38
V/N	0.72	0.70	0.74	1.22	1.13	0.73	-0.12	3.79	11.38

* J = adjective, V = verb, N = noun, R = adverb

Figure 13. Hyper-parameter testing results, short documents

Hyper-parameter and topic modeling algorithm grid search results; higher is better
Long documents (>= 3,000 words)

Sub-doc min/max length; TF-IDF filter %; Lemmatized parts of speech*	Average of Distance from corpus dist			Average of Doc prominence: % largest			Average of TC-PMI coherence		
	hlda	ica	lda	hlda	ica	lda	hlda	ica	lda
100/200	0.77	0.76	0.77	8.10	4.91	4.86	20.30	27.20	27.89
20	0.77	0.76	0.77	8.10	4.91	4.86	20.02	27.20	27.89
J/N/N/R	0.77	0.76	0.77	8.10	4.91	4.86	18.28	27.20	27.89
V/N	0.77	0.76	0.77	8.10	4.91	4.86	21.27	27.20	27.89
40	0.77	0.76	0.77	8.10	4.91	4.86	20.55	27.20	27.89
J/N/N/R	0.77	0.76	0.77	8.10	4.91	4.86	20.20	27.20	27.89
V/N	0.77	0.76	0.77	8.10	4.91	4.86	20.58	27.20	27.89
60	0.77	0.76	0.77	8.10	4.91	4.86	20.15	27.20	27.89
J/N/N/R	0.77	0.76	0.77	8.10	4.91	4.86	19.87	27.20	27.89
V/N	0.77	0.76	0.77	8.10	4.91	4.86	20.67	27.20	27.89
100/400	0.77	0.76	0.77	8.10	4.91	4.86	22.54	27.20	27.89
20	0.77	0.76	0.77	8.10	4.91	4.86	19.08	27.20	27.89
J/N/N/R	0.77	0.76	0.77	8.10	4.91	4.86	20.46	27.20	27.89
V/N	0.77	0.76	0.77	8.10	4.91	4.86	19.88	28.05	27.31
40	0.76	0.76	0.77	8.02	4.91	4.81	19.70	28.05	27.31
J/N/N/R	0.76	0.76	0.77	8.02	4.91	4.81	18.83	28.05	27.31
V/N	0.76	0.76	0.77	8.02	4.91	4.81	20.79	28.05	27.31
60	0.76	0.76	0.77	8.02	4.91	4.81	19.89	28.05	27.31
J/N/N/R	0.76	0.76	0.77	8.02	4.91	4.81	21.58	28.05	27.31
V/N	0.76	0.76	0.77	8.02	4.91	4.81	19.38	28.05	27.31
100/200	0.79	0.78	0.78	15.21	9.15	9.13	17.78	30.20	25.35
20	0.79	0.78	0.78	15.21	9.15	9.13	16.99	30.20	25.35
J/N/N/R	0.79	0.78	0.78	15.21	9.15	9.13	16.61	30.20	25.35
V/N	0.79	0.78	0.78	15.21	9.15	9.13	17.00	30.20	25.35
40	0.79	0.78	0.78	15.21	9.15	9.13	17.37	30.20	25.35
J/N/N/R	0.78	0.78	0.78	15.21	9.15	9.13	18.05	30.20	25.35
V/N	0.78	0.78	0.78	15.21	9.15	9.13	16.85	30.20	25.35
60	0.79	0.78	0.78	15.21	9.15	9.13	18.29	30.20	25.35
J/N/N/R	0.78	0.78	0.78	15.21	9.15	9.13	19.00	30.20	25.35
V/N	0.79	0.78	0.78	15.21	9.15	9.13	18.30	30.20	25.35
50/200	0.78	0.78	0.78	15.21	9.15	9.13	18.70	30.20	25.35
20	0.78	0.78	0.78	15.21	9.15	9.13	18.15	30.20	25.35
J/N/N/R	0.78	0.78	0.78	15.21	9.15	9.13	17.95	30.20	25.35
V/N	0.78	0.78	0.78	15.21	9.15	9.13	17.44	31.05	24.77
40	0.78	0.78	0.78	15.21	9.15	9.13	17.15	31.05	24.77
J/N/N/R	0.78	0.78	0.78	15.21	9.15	9.13	17.19	31.05	24.77
V/N	0.78	0.78	0.78	15.21	9.15	9.13	17.01	31.05	24.77
60	0.78	0.78	0.78	15.21	9.15	9.13	17.24	31.05	24.77
J/N/N/R	0.78	0.78	0.78	15.21	9.15	9.13	18.01	31.05	24.77
V/N	0.78	0.78	0.78	15.21	9.15	9.13	17.98	31.05	24.77
100/400	0.78	0.78	0.78	15.21	9.15	9.13	18.07	31.05	24.77
20	0.78	0.78	0.78	15.21	9.15	9.13	17.97	31.05	24.77
J/N/N/R	0.78	0.78	0.78	15.21	9.15	9.13	17.16	31.05	24.77
V/N	0.78	0.78	0.78	15.21	9.15	9.13	15.01	31.05	24.77
40	0.78	0.78	0.78	15.21	9.15	9.13	20.18	31.05	24.77
J/N/N/R	0.78	0.78	0.78	15.21	9.15	9.13	16.30	31.05	24.77
V/N	0.78	0.78	0.78	15.21	9.15	9.13			

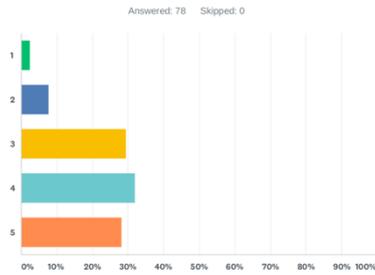
* J = adjective, V = verb, N = noun, R = adverb

Figure 14. Hyper-parameter testing results, long documents

Appendix B. Survey Results

B.1. Overall Summary by Question

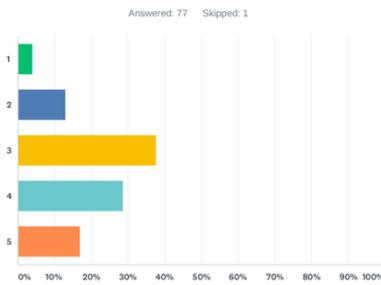
Q1 How much experience do you have with writing overall, on a scale of 1-5, with 5 being the highest?



ANSWER CHOICES	RESPONSES
1	2.56%
2	7.69%
3	29.49%
4	32.05%
5	28.21%
TOTAL	78

Figure 15. Overall Summary, Question 1

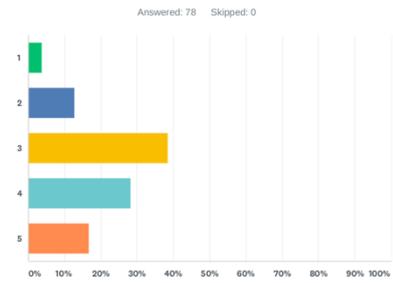
Q2 How much do you like writing?



ANSWER CHOICES	RESPONSES
1	3.90%
2	12.99%
3	37.66%
4	28.57%
5	16.88%
TOTAL	77

Figure 16. Overall Summary, Question 2

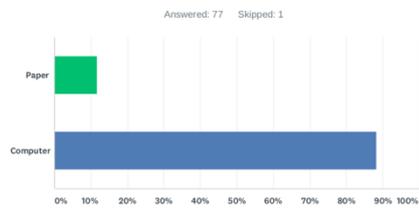
Q3 How skilled do you think you are at writing?



ANSWER CHOICES	RESPONSES
1	3.85%
2	12.82%
3	38.46%
4	28.21%
5	16.67%
TOTAL	78

Figure 17. Overall Summary, Question 3

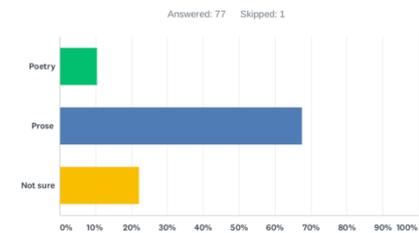
Q4 Do you prefer to write on paper or using a computer?



ANSWER CHOICES	RESPONSES
Paper	11.69%
Computer	88.31%
TOTAL	77

Figure 18. Overall Summary, Question 4

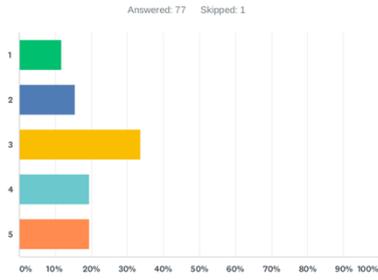
Q5 Which do you prefer, poetry or prose?



ANSWER CHOICES	RESPONSES
Poetry	10.39%
Prose	67.53%
Not sure	22.08%
TOTAL	77

Figure 19. Overall Summary, Question 5

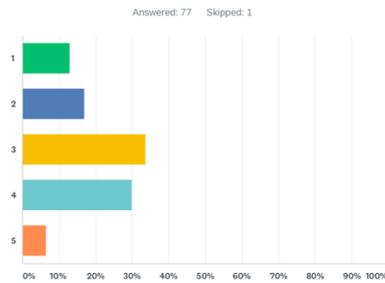
Q6 Was the app easy to use?



ANSWER CHOICES	RESPONSES	COUNT
1	11.69%	9
2	15.58%	12
3	33.77%	26
4	19.48%	15
5	19.48%	15
TOTAL		77

Figure 20. Overall Summary, Question 6

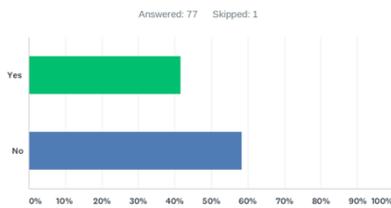
Q7 Did you enjoy using the app?



ANSWER CHOICES	RESPONSES	COUNT
1	12.99%	10
2	16.88%	13
3	33.77%	26
4	29.87%	23
5	6.49%	5
TOTAL		77

Figure 21. Overall Summary, Question 7

Q8 Would you use the app for writing?



ANSWER CHOICES	RESPONSES	COUNT
Yes	41.56%	32
No	58.44%	45
TOTAL		77

Figure 22. Overall Summary, Question 8

B.2. Additional Results, Time Trends in Survey Results

Results for Questions 1, 6, and 7 are shown in Section 6.

How much do you like writing?



Figure 23. Time trend, Question 2

How skilled do you think you are at writing?

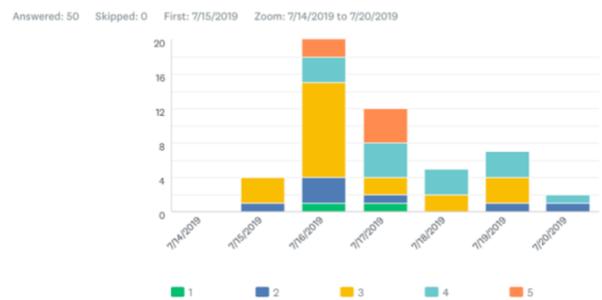


Figure 24. Time trend, Question 3

Do you prefer to write on paper or using a computer?



Figure 25. Time trend, Question 4

Which do you prefer, poetry or prose?

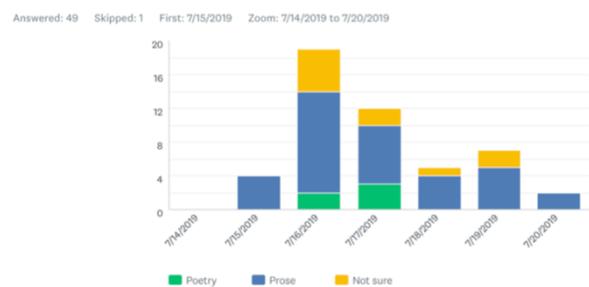


Figure 26. Time trend, Question 5

Would you use the app for writing?

Answered: 49 Skipped: 1 First: 7/15/2019 Zoom: 7/14/2019 to 7/20/2019

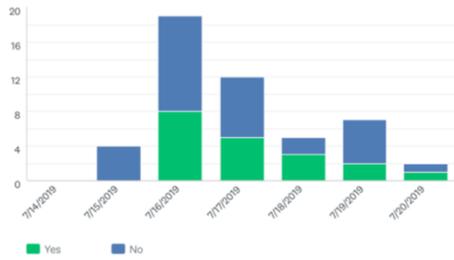


Figure 27. Time trend, Question 8

Appendix C. Explanatory Modelling

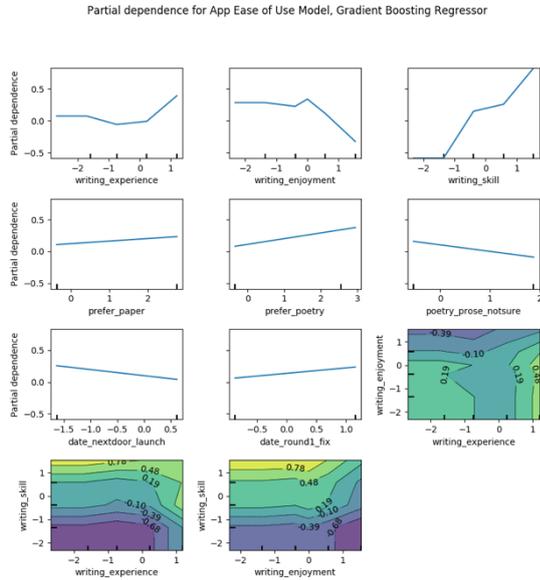


Figure 28. Partial dependence for Question 6 model

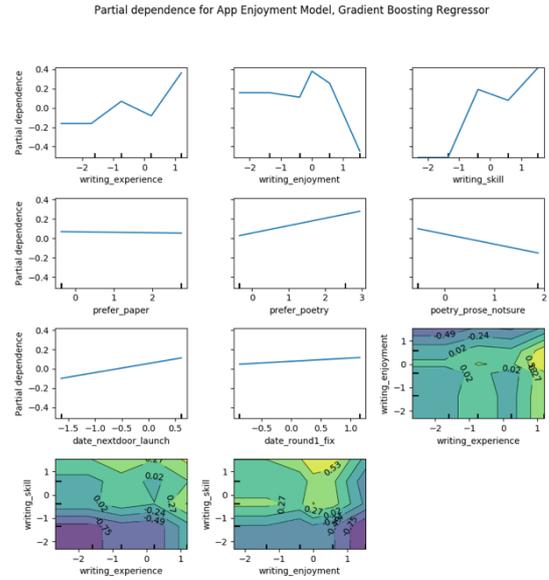


Figure 29. Partial dependence for Question 7 model

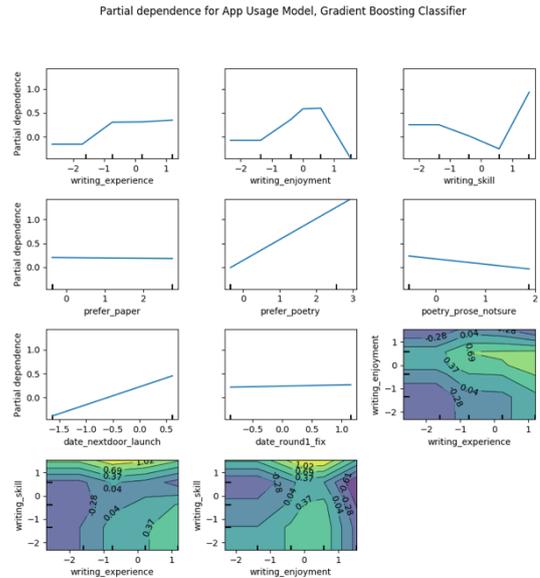


Figure 30. Partial dependence for Question 8 model

Acknowledgements.

The author would like to thank David Joyner at The Georgia Institute of Technology.

References

- [1] Becker, A. (2006) A Review of Writing Model Research Based on Cognitive Processes. In Horning, A., Becker, A. [ed.]. *Revision: History, Theory, and Practice* (West Lafayette, IN: Parlor Press), ch. 3.

- [2] Monem, R. (2010) Metacognitive Functions, Interest, and Student Engagement in the Writing Process: A Review of the Literature. In *Proceedings of the Ninth Annual College of Education and GSN Research Conference*, Miami, FL, (Miami, FL: Florida International University), 64-68.
- [3] Sweller, J. (1988) Cognitive load during problem solving: Effects on learning. *Cognitive Science* **12**(2): 257-285.
- [4] Cambria, E. and White, B. (2014) Jumping NLP Curves: A Review of Natural Language Processing Research. *IEEE Computational Intelligence Magazine* **9**(2): 48-57.
- [5] Blei, D. and Lafferty, A. (2009) Topic Models. In Srivastava, A., Sahami, M. [ed.] *Text Mining: Classification, Clustering, and Applications* (Boca Raton, FL: Chapman & Hall/CRC), ch. 4.
- [6] Blei, D. (2012) Probabilistic Topic Models. *Communications of the ACM* **55**(4): 77-84.
- [7] Knight, S., Shum, S., Ryan, P., Sandor, A., Wang, X. (2018) Designing Academic Writing Analytics for Civil Law Student Self-Assessment. *International Journal of Artificial Intelligence in Education* **28**(1): 1-28.
- [8] Abel, S., Kitto, K., Knight, S., Shum, S. (2018) Designing personalised, automated feedback to develop students' research writing skills. In *Proceedings ASCILITE 2018: 35th International Conference on Innovation, Practice and Research in the Use of Educational Technologies in Tertiary Education*, Victoria, Australia, 25-28 November 2018 (Melbourne).
- [9] Burstein, J., Klebanov, B., Elliot, N., Molloy, H. (2016) A Left Turn: Automated Feedback & Activity Generation for Student Writers. In *Proceedings of the 3rd Language Teaching, Learning & Technology Workshop*, San Francisco, CA, 6-7 September 2016, 6-13.
- [10] Villalon, J. and Calvo, R. (2011) Concept Maps as Cognitive Visualizations of Writing Assignments. *Educational Technology & Society* **14**(3): 16-27.
- [11] Whitelock, D., Twiner, A., Richardson, J., Field, D., Pulman, S. (2015) OpenEssayist: A supply and demand learning analytics tool for drafting academic essays. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge*, Poughkeepsie, NY, 16-20 March 2015 (New York, NY: ACM), 208-212.
- [12] Mihalcea, R., Tarau, P. (2004) TextRank: Bringing order into text. In *Proceedings of Empirical Methods in Natural Language Processing*, Barcelona, Spain, 25-26 July 2004 (Stroudsburg, PA: Association for Computational Linguistics), 404-411.
- [13] Whitelock, D., Twiner, A., Richardson, J., Field, D., Pulman, S. (2017) What types of essay feedback influence implementation: Structure alone or structure and content? In *Proceedings of Technology Enhanced Assessment (TEA 2016)*, Tallinn, Estonia, 5-6 October 2016 (Cham, Switzerland: Springer), 181-196.
- [14] Boyd-Graber, J., Minno, D., Newman, D. (2014) Care and Feeding of Topic Models: Problems, Diagnostics, and Improvements. In Airoldi, E., Blei, D., Erosheva, E., Fienberg, S. [ed.], *Handbook of Mixed Membership Models and Their Applications* (Boca Raton, FL: CRC Press), ch. 1.
- [15] Debortoli, S., Muller, O., Junglas, I., vom Brocke, J. (2016) Text Mining for Information Systems Researchers: An Annotated Topic Modeling Tutorial. *Communications of the Association for Information Systems* **39**(7): 1-30.
- [16] Blei, D., Ng, A., Jordan, M. (2003) Latent Dirichlet Allocation. *Journal of Machine Learning Research* **3**: 993-1022.
- [17] Blei, D., Jordan, M., Griffiths, T., Tenenbaum, J. (2003) Hierarchical topic models and the nested Chinese restaurant process. In *Proceedings of the 16th International Conference on Neural Information Processing Systems*, Barcelona, Spain, 5-10 December 2016 (Cambridge, MA: MIT Press), 17-24.
- [18] Grant, S., Skillicorn, D., Cordy, J. (2008) Topic Detection Using Independent Component Analysis. In *Proceedings of the 2008 Workshop on Link Analysis, Counterterrorism and Security, SIAM International Conference on Data Mining*, Atlanta, GA, 24-26 April 2008 (Philadelphia, PA: SIAM), 23-28.
- [19] The Hewlett Foundation: Automated Essay Scoring, <https://www.kaggle.com/c/asap-aes/data>, last accessed 2019/07/21.
- [20] NIPS Papers, <https://www.kaggle.com/benhamner/nips-papers>, last accessed 2019/07/21.
- [21] New York Times articles, <https://www.kaggle.com/nzalake52/new-york-times-articles/version/1>, last accessed 2019/07/21.
- [22] Andersen, H. (1999) *Andersen's Fairy Tales*. (Urbana, IL: Project Gutenberg).
- [23] Pedregosa et al. (2011) Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. **12**: 2825-2830.
- [24] Seabold, Skipper, and Josef Perktold. (2010) Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference*, Austin, TX, 28 June - 3 July 2010, 57-61.
- [25] Shmueli, G. (2010) To Explain or to Predict? *Statistical Science*. **25**:3, 289-310.