

# Natural Language Processing Used in Sentiment Analysis of Poetry: A Study of Six Common Techniques

Tanya Nair

Eastlake High School

## ABSTRACT

This paper explores and compares the accuracies of six different NLP model techniques when applied to analysis of English poetry. It was found that bigram-based and word embedding-based models were the most accurate in deriving emotions from the bodies of text in the corpus, with respective accuracies of 61.67% and 61.68%. The least accurate models were the unigram-based model and distributed dictionary-based model, constructed using the traditional approach rather than the new methodology, and had accuracies of 11.173% and 48.17% respectively. Most models passed the benchmark accuracies of 49.42% and 47.86%, the higher accuracy one being a word count model and lower one being a sentiment model. The need for newer methodologies that allow for higher dimensional levels of semantic analysis to be performed is also discussed in this paper, along with the potential impacts of this research to the field of Natural Language Processing.

## Introduction

Language, as a major form of communication among humans, is an essential part of our society and is just about second nature to us all. Natural Language Processing (NLP), commonly known as a sub-field of Computational Linguistics, consists of using Machine Learning models to manipulate natural language in different ways, whether that be through analysis, interpretation, or generation. NLP is an important field, with uses ranging from speech recognition to machine translation. It is this sheer versatility of applications that has allowed NLP to “come to play a central role in the multilingual information society” [1]. Poetry is an emotionally dense form of writing whose meaning is subjective and open to interpretation. Humans often struggle to grasp the meaning of certain poetic works, and making a machine learning model to automate this task of analysis is something difficult to achieve.

Mood recognition is a Natural Language Processing field that uses Machine Learning models to derive a sentiment from a body of text. This study focuses on applications of mood recognition to poetry and will aim to find the most accurate NLP framework to do so. Applications of mood recognition, however, range much further than conducting sentiment analysis of poetry. An example of this field being applied to everyday life is the recommendation systems in social media services, which gather data on its users, analyzing the reactions each post invokes. When it detects that users have relatively larger reactions to a certain post, it will then recommend that post to users with similar profiles of interests. Similarly, mood recognition can be used by law enforcement to monitor social media posts for threatening language, thereby improving public safety for both the local community and on a national level.

This paper will address the question: “Can Natural Language Processing models determine the overall emotion conveyed in poems accurately?”. I will conduct a comparative study of application of six Natural Language Processing models on poetry and evaluate which NLP technique is the most effective in recognizing the dense emotions that are conveyed in poetic works of the English language. From this study, I aim to gather

insights on what characteristics make certain NLP models more accurate at understanding emotion in poetry than others through experimentation and application of different NLP techniques. By making observations on improvements to be made to the models, I will uncover the strengths and weaknesses of each technique when applied to emotional analysis of poetry, while also gaining a deeper understanding of how machine learning models can understand more emotionally dense pieces of writing.

Human-like accuracy in machine analysis of prose is already quite the behemoth of a task, due to the need for “variability, ambiguity, and context-dependent interpretation of human languages” [2], but with the help of recent advances in technology, large strides have been made in the field. The field of Natural Language Processing has a comparatively lesser amount of literature on poetry analysis, likely due to the extra layers of subjectivity and difficulty the task of interpreting and analyzing poetry adds to the mix. A major focus of many studies has been to differentiate between poetry and prose, but to do so, there need to be multiple solid defining factors that a majority of poetic works follow. Marmik Pandya [3] of Northeastern University provides a working definition of poetry, that it is “a piece of natural language text that complies with the constraints of grammaticality, meaningfulness, and poeticness” (p. 5). These constraints are accommodating enough to ensure proper interpretation and expression of poetic works, while also offering a definite way to identify what makes a piece of writing a poem. When searching for poetry and Natural Language Processing in the current literature, one finds more studies focused on rhythmic analysis of poetry, such as analysis of Arabic poetry rhythm using NLP [4]. Analysis of poetic works in regard to their rhyme, meter, and word stress can help to distinguish between what is and is not poetry, however this style of analysis still does not derive the meaning behind the poem being analyzed. This is where affect analysis comes into play, by which machine learning models can analyze whether a poem’s mood is positive or negative. Researchers from Stanford University analyzed the style, affect, and imagery of poems in order to find the defining characteristics of good poetry [5], but other than this study, not much research has been done on higher dimensional forms of emotional analysis of poetry in the English language. I will delve deeper into the topic while answering my research question in this paper.

## Methodology

All models used in this study have been trained on the Poem Emotion Recognition Corpus (PERC) available for free on the Mendeley Data online database. This dataset was created by researchers following the publishing of their Natural Language Processing research paper providing an overview of the current literature in the field regarding emotional analysis of poems. In the paper [6], they proposed a categorical system of nine emotions that was outlined in the works of ancient sage Bharata Muni, contrary to the popular six-emotion system that is typically used to classify poetic works. I will use this nine-emotion classification system in the training and testing of my six base NLP models. Of this dataset, I will split the 717 entries of data into 70:30, with 70 percent of the data being used to train my models and the other 30 percent being used to test. In all models, each of the nine emotions are represented numerically by a mood code, an integer from 1 to 9. In the event that a model predicts a mood between two values on the integer scale, the value will be rounded to the nearest whole number. Due to technological limitations, all models will be run only once, so cross-validation will not be possible in this scenario.

Three of the six aforementioned models used in this study follow the popular bag-of-words approach, in which words from the unstructured text data of the corpus are extracted and encoded into a feature matrix. Depending on the values of each column in the matrix, with every column representing unique words across the corpus, and each value in the corresponding row being either 0 or 1 depending on the content of each document, the predicted emotion conveyed will be one of nine: anger, courage, fear, hate, joy, love, peace, sadness, or surprise.

The first two bag-of-words models are trained based on n-grams, with one model being unigram-based and the other bigram-based. To understand the strength and direction of association and improve upon the

unigram-based and bigram-based models after testing, I will use the Kendall accuracy function from the `glmnet` R package to find the measure of the accuracy of the predictions. The next step will be to find ways to minimize the difference between the prediction of the model and the actual value of the mood code available in the corpus.

The third model uses a more complex NLP strategy called topic modeling, another form of the bag-of-words technique, for identifying themes within a corpus. Topic modeling finds topics, or groups of words that frequently occur together across documents in the corpus. This is an unsupervised process, meaning that the model will find twenty designated topics and recurring patterns of words in the corpus without any human assistance. It is due to the reduced human intervention needed that topic modeling is widely known to be a good technique for sorting through and deriving meaning from large corpuses with hundreds of documents. As the generation of these twenty topics is unsupervised, they will not follow the nine static emotions that are defined in the training dataset, however slight alignments may occur.

The fourth model follows a more complex structural approach, word embeddings. Word embeddings help a machine learning model better understand the content of a body of text when compared to bag-of-words techniques due to their consideration of the context of a word when representing text data in a structured, numerical format. This technique groups words that occur in similar contexts closer together in embedding space, meaning that the more similar the words are in terms of their contexts, the closer together they will be. Word embeddings are also more efficient than bag-of-words techniques due to their use of dense vectors as opposed to sparse vectors. This could improve a machine learning model's performance when conducting emotional analyses of poetry. After training and testing the model, the next step will be to find ways to increase its accuracy past that of the benchmark models.

The fifth model uses a Natural Language Processing technique called dictionary-based text analysis. Dictionaries are a supervised method of training NLP models, requiring human-input during the preprocessing step, in which one has to map a set of words to certain emotions in the dictionary. Based on the frequency and emotional association of the dictionary defined words in any given document, the machine learning model can predict the mood code of said document. Prior research in the field [7] has acknowledged and provided a potential framework to counter the fact that dictionaries are domain-dependent, having mappings to emotions or sentiments that, when applied to models conducting text analysis in a different field, do not provide accurate predictions due to words in both domains being anatomically the same yet having different meanings based on context. The dictionary I used for this model was called `lexicon_loughran()` and from the package `textdata`.

The final model is trained based on distributed dictionaries, that enables measurement of similarity between dictionaries. This is like a combination of word embeddings and dictionaries.

## Results

Before sharing the accuracies of the NLP techniques, here are two benchmark models and their respective accuracies. The first benchmark model is word count based and ended up with an accuracy of 49.42%. The second model is a sentiment model and has an accuracy of 47.86%.

The unigram-based NLP model had a starting accuracy of 11.173%, but after clearing out relatively unimportant aspects of the data such as the rarest words in the corpus, the accuracy percentage increased to 57.41%. The bigram-based model presented a different story, with its accuracy score starting out at 61.67%. No improvements were made to the bigram-based model as it had already surpassed the benchmark accuracy score by almost 20%.

The word embedding model had an accuracy of 61.68%, the highest out of all models in this study. The dictionary-based model had an accuracy of 57.24%. The distributed dictionary based model had a starting accuracy of 48.17% when using the traditional approach, and an accuracy of 58.45% when using the new methodology.

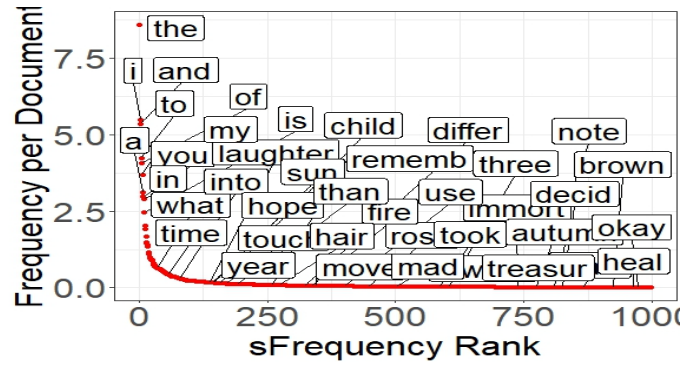


Figure 1. Frequency of various words per document in the corpus.

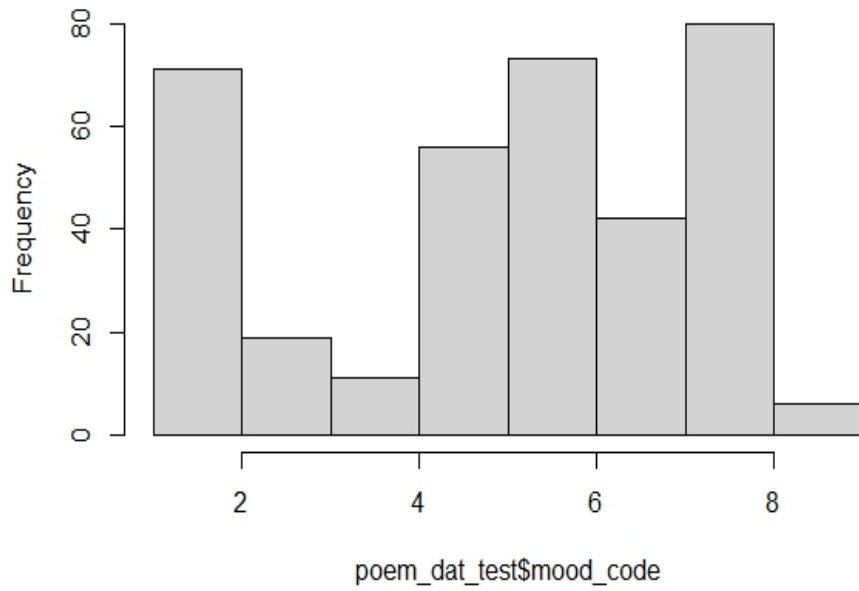


Figure 2. Histogram of predictions based on test data.

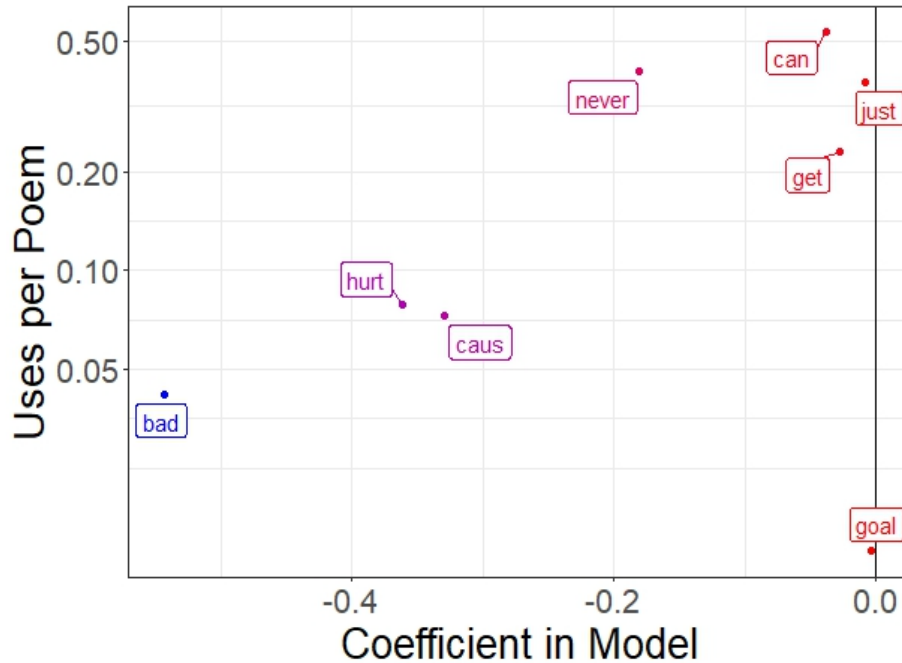


Figure 3. Kendall Rank Correlation Coefficient for important words in the corpus.

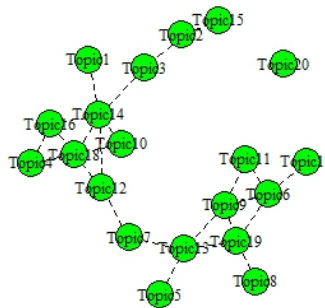


Figure 4. Correlation of generated topics.

## Discussion

When one is just beginning to learn a language, it is impossible to read and understand poetic works in said language right off the bat. One must first learn how to read the alphabet and derive meaning from the glyphs on paper. Then, they progress to learning how those alphabets can come together to make words, after which they explore the syntactic rules that give order and sense to the language. Only then can one progress on to learning about abstract concepts and figures of speech (that are often language- or culture-specific), something that will help them make sense of most of the poetry in that language. Although considerably expedited, this summary of a human being’s learning of language can also apply to the field of Natural Language Processing over time. We are currently in a transitional phase, computational linguists have made breakthroughs in the encoding of syntactic structures in machine learning models, and computers, for the most part, are able to derive basic meaning from bodies of text with reasonable accuracy.

These findings are important to the field of natural language processing because making advancements in emotion detection within somewhat convoluted, hard to analyze forms of text like poetry can help us find new avenues for analysis and improve upon our techniques when applied to less high-dimensional forms of text. The sentiment analysis methods outlined in this paper can be applied to many different situations aside from poetry.

## Conclusion

Natural Language Processing models can conduct emotional analysis of English poetic works, but their accuracy still is not yet at the level of humans. In this study, it was found that bigram and word embedding based models had the highest accuracies when applied to poetry, with percentages of 61.67% and 61.68% respectively. The lowest accuracy scores were held by the traditionally constructed distributed dictionaries at 48.17% and unigrams at 11.173%.

These accuracy scores are not great in the grand scheme of things. To improve them, better techniques could be devised, and better data could also be made publicly available. I was unable to cross-validate my findings in this study due to technological limitations. It took many minutes at a time to run each individual model, and cross validating all six models in the study would not be feasible due to the lack of processing power available. Along with this, I used a publicly available dataset to maintain a heavy focus on the NLP techniques in my paper, rather than spend much of my time manually creating a dataset of hundreds of poems.

As mentioned in the Discussion section, the more improvements that are made within the field of poetry analysis, the greater the impact will be on the field of NLP. Currently, the methods used for semantic analysis are meant to derive one of three sentiments – positive, negative, or neutral. Finding a new method that expands the list of sentiments we can extract from bodies of text will cause subsequent advancements in the field of sentiment analysis applied to both poetry and prose, as we will be able to derive more complex meanings from text.

As the field continues to progress, we will see improvements to NLP applications in our daily lives as well. From improved recommendation algorithms with the ability to understand your written review of a movie and recommend you movies that you may like based on the data, to improved anti-terrorist tracking systems for law enforcement to keep us safe, Natural Language Processing truly has a wide scale of applications, and the more developments we make in the field, the more we can improve our quality of life.

## Acknowledgments

Thank you for the guidance of mentor Seyed Mohammad Mahdi Seyed Ghafouri from the University of London in the development of this research paper and the Lumiere Research Scholar Program for this excellent opportunity.

## References

- [1] Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc."
- [2] Hirschberg, J. & Manning, C. D. (2015). Advances in natural language processing. *Science*, 349 ( 6245), 261-266. <https://doi.org/10.1126/science.aaa8685>

- [3] Pandya, Marmik. (2016). NLP based Poetry Analysis and Generation. <http://dx.doi.org/10.13140/RG.2.2.35878.73285>.
- [4] Ahmed, M.A., & Trausan-Matu, S. (2017). Using natural language processing for analyzing Arabic poetry rhythm. 2017 16th RoEduNet Conference: Networking in Education and Research (RoEduNet), 1-5. <https://doi.org/10.1109/ROEDUNET.2017.8123759>
- [5] Kao, J., & Jurafsky, D. (2012, June). A computational analysis of style, affect, and imagery in contemporary poetry. In Proceedings of the NAACL-HLT 2012 workshop on computational linguistics for literature (pp. 8-17).
- [6] P S, SREEJA & G S, Mahalakshmi. (2017). Emotion Models: A Review. International Journal of Control Theory and Applications. 10. 651-657.
- [7] Rice, D., & Zorn, C. (2021). Corpus-based dictionaries for sentiment analysis of specialized vocabularies. Political Science Research and Methods, 9(1), 20-35. doi:10.1017/psrm.2019.10