Examining Ambiguity in Natural Language Processing

Comparison of Tools to Detect and Evaluate Lexical Ambiguity

Laura Gould *l.gould@rutgers.edu*

Tarun Reddy *ta568@scarletmail.rutgers.edu*

Simran Singh *psimran.singh@rutgers.edu*

Masouma Tajik *mt1167@scarletmail.rutgers.edu*

ABSTRACT

Current natural language processing technology is not suited for detecting or remedying ambiguity in written text. This study tests whether existing tools related to NLP have the capacity to identify either lexical ambiguity or structural ambiguity in a provided dataset with limited context. Textblob and Pattern were used to test sentiment and subjectivity scores as two potential post-hoc tools, in addition to a series of logistic regression models trained on an ambiguous dataset. The research team determined the limitations of these three tools outweigh their potential benefits, and found that none of the techniques were useful for consistently identifying textual ambiguities.

Keywords

Lexical Ambiguity, Natural Language Processing, Sentiment Analysis, Subjectivity Analysis, Large Language Models

1 INTRODUCTION

Lexical ambiguity is a problem that has haunted natural language processing experts since the inception of NLP itself. Lexical ambiguity arises when a word has multiple meanings or senses. For example, the word "bank" can refer to a financial institution or the edge of a river. Syntactic ambiguity arises when a sentence can be parsed in multiple ways due to the structure or placement of words. For example, the sentence "I saw her duck" could mean either "I saw her (the person) duck (the bird)" or "I saw her (the bird) duck (the action). "Semantic ambiguity arises when a sentence has more than one interpretation due to the meaning of words. For example, the sentence "The chicken is ready to eat" could mean either "The chicken is cooked and ready to be eaten" or "The chicken is hungry and ready to eat." Pragmatic ambiguity arises when a sentence has more than one interpretation due to the context or speaker's intentions. For example, the sentence "Can you pass me the salt?" could mean either a request or a challenge depending on the context and tone. Referential ambiguity arises when a word or phrase refers to more than one possible antecedent. For example, the sentence "She saw her friend with a telescope" could refer to either the subject or the object having the telescope.

Ambiguous text can be particularly challenging in sentiment analysis, text classification, and machine translation, where the ability to accurately capture the meaning of a text is critical. In natural language processing, especially in particularly large data-scraping projects, researchers typically list lexical ambiguity as a limitation of their study. Current natural language processing technology struggles with complex forms of human communication like jokes and sarcasm, and even complex sentence structures which may present conflicting points of view in a single sentence. Lexical ambiguity is also present in equivocal words - words which may have multiple meanings based on varied contexts. Our project uses existing NLP techniques to determine which, if any, could be used as a post-hoc test to identify lexical and structural ambiguity in large natural language datasets for future research projects. For the purposes of this paper, we investigated one sentiment analysis tool, one subjectivity analysis tool, and one regression-based tool.

2 LITERATURE REVIEW

Ambiguity is a well-known challenge in natural language processing that poses a significant obstacle to accurate interpretation and understanding of text. Different types of ambiguity exist, and understanding them is crucial to addressing ambiguity in natural language processing tasks. In doing so, we can improve the accuracy of language processing systems and enhance our understanding of natural language. Researchers have proposed different metrics of ambiguity that can help to identify and quantify the level of ambiguity in a given text. In this literature review, we will explore and evaluate various metrics of ambiguity proposed in the literature, including domain-specific language models, dissimilarity scores, syntactic complexity, and cohesion and coherence scores.

Domain-Specific Language Models and Dissimilarity Scores

Domain-specific language models are a powerful tool for measuring ambiguity in technical domains. These models are typically trained on large amounts of text data from a specific domain and can capture the unique vocabulary and syntax of that domain. One way to use these models to measure ambiguity is to compare the use of specific terms across domains. For example, if a term has different meanings in different domains, it may be considered ambiguous. Gharibi et al. (2018) propose an approach for cross-domain ambiguity detection in requirements engineering that uses domain-specific language models and word embeddings to estimate potential ambiguity across domains. The authors use a combination of neural network models and probabilistic models to build domain-specific language models for different technical domains. They then use these models to compare the use of specific terms across domains, allowing them to identify potentially ambiguous terms.

Dissimilarity scores are another approach to measuring ambiguity. These scores are based on the idea that ambiguous words will have high variance in their similarity values across domains. Kageura and Umino (1996) propose a method for calculating ambiguity scores based on co-occurrence frequencies in large corpora. They argue that words that appear frequently in multiple contexts are more likely to be ambiguous than those that appear only in specific contexts. To calculate dissimilarity scores, researchers typically construct similarity lists for a set of words and then calculate the variance of similarity values for each word across domains. The dissimilarity score can then be calculated as a rank-weighted sum of these variance values. The higher the score, the more ambiguous the word is considered to be. Overall, domain-specific language models and dissimilarity scores are two powerful approaches to measuring ambiguity in technical domains. These techniques can be used to identify potentially ambiguous terms and improve communication between stakeholders in technical fields.

Subjectivity Scores

Subjectivity analysis is an important aspect of sentiment analysis that aims to identify the degree of subjectivity in a given text. Subjectivity scores are typically based on the presence of personal pronouns, such as 'I', 'my', 'our', and 'mine', and can be used to determine the degree of opinionatedness in a given text. Several studies have investigated the use of subjectivity scores in sentiment analysis. Liu et al. (2012) proposed a novel approach that combines both subjectivity and polarity scores to improve the accuracy of sentiment classification. They found that the use of both scores led to higher accuracy in classifying tweets as positive, negative, or neutral. Kim et al. (2006) analyzed customer reviews and found that subjectivity analysis can be used to identify the degree of opinionatedness in these reviews. By using subjectivity scores, they were able to distinguish between objective and subjective sentences and identify the overall sentiment expressed in the review. Pang et al. (2002) proposed a machine learning approach that uses both subjectivity and polarity features for sentiment classification. They found that the use of both features led to higher accuracy in classifying movie reviews as positive or negative.

Subjectivity scores can be particularly useful in addressing ambiguity in a statement. When a statement is ambiguous, it may be difficult to determine the sentiment expressed in the text. However, by analyzing the subjectivity score of the text, we can determine whether the text is more objective or subjective. This approach can lead to a better understanding of the sentiment expressed in a given text, which can be valuable in sentiment analysis tasks where it may be difficult to determine the polarity of a given text. In summary, subjectivity analysis is an important aspect of sentiment analysis that can be used to improve the accuracy of sentiment classification. The use of subjectivity scores, in combination with polarity scores, can provide a more nuanced understanding of the sentiments expressed in a given text. Moreover, subjectivity scores can help to address ambiguity in a statement and can be particularly useful in sentiment analysis tasks.

Syntactic Complexity

Syntactic complexity is another metric of ambiguity that refers to the complexity of the sentence structure in a given text. A sentence that is structurally complex can be more difficult to understand and may contain multiple interpretations, which can increase ambiguity. There are several different measures of syntactic complexity that have been proposed in the literature. One common approach is to use measures such as sentence length, number of clauses, and depth of embedding to capture the structural complexity of sentences (Hasselgård, 2010). For example, a sentence that is longer and contains more embedded clauses may be considered more syntactically complex than a shorter sentence with fewer clauses.

Another approach is to use treebank-based metrics such as phrase structure complexity (PSC) and dependency length minimization (DLM) (McDonald, 2006). PSC measures the degree of phrasal embedding in a sentence, while DLM measures the distance between dependent words in a sentence. These measures have been shown to correlate with reading difficulty and can be used to identify sentences that are likely to be more difficult to understand. Syntactic complexity has been studied in a variety of contexts, including second language acquisition, cognitive processing, and text simplification. For example, studies have found that learners of a second language struggle with more complex sentence structures and that text simplification can improve comprehension of complex sentences (Laufer, 2005; Xu et al., 2015). Overall, syntactic complexity is an important metric of ambiguity that can be used to identify sentences that are more difficult to understand and may contain multiple interpretations. By measuring syntactic complexity, researchers can gain insights into the cognitive processes involved in comprehension and develop strategies to improve text readability.

Cohesion and Coherence Scores

Cohesion and coherence scores have been used in natural language processing to measure the degree of ambiguity in a given text. Cohesion refers to the degree of connectivity between sentences or clauses in a text, while coherence refers to the degree of meaningfulness or logical connectedness of the text as a whole. Ambiguity can arise when a text has low cohesion or coherence, as the meaning of the text may be unclear or open to multiple interpretations. One approach to measuring cohesion and coherence is through the use of graph-based algorithms. For example, the LexRank algorithm, proposed by Erkan and Radev (2004), uses a graph-based representation of a text to identify the most important sentences based on their similarity to other sentences in the text. By identifying the most important sentences, the algorithm can improve the cohesion and coherence of the text, reducing ambiguity and improving readability. Another approach is to use machine learning techniques to predict the degree of coherence or ambiguity in a given text. For example, a study by Pitler and Nenkova (2008) used a support vector machine classifier to predict the coherence of a text based on a set of linguistic features, such as sentence length and the frequency of transitional phrases. The results of their study showed that the classifier was able to accurately predict the coherence of a text, indicating that machine learning techniques can be effective in measuring ambiguity.

Coherence and cohesion scores have also been used in the evaluation of machine translation systems, as ambiguity can be a common issue in machine translation. For example, the CoNLL 2016 Shared Task on Multilingual Parsing used a coherence scoring metric to evaluate the quality of machine translations, with the aim of identifying translations that were coherent and semantically meaningful (Zeman et al., 2016).Cohesion and coherence scores have been used in natural language processing to address ambiguity and measure the degree of meaningfulness and connectedness in a given text. Graph-based algorithms and machine learning techniques have been used to measure coherence and cohesion, with applications in improving readability, machine translation, and text evaluation.

In conclusion, ambiguity is a persistent challenge in natural language processing that continues to pose significant obstacles to the accurate interpretation and understanding of text. As such, it is critical to develop and evaluate various metrics to measure ambiguity, such as domain-specific language models, dissimilarity scores, syntactic complexity, and cohesion and coherence scores. The use of these metrics can provide valuable insights into the level of ambiguity present in a given text and inform the development of improved techniques for sentiment analysis, text classification, and machine translation. Moreover, by combining these metrics with other techniques, researchers can gain a deeper understanding of cross-domain ambiguity and enhance communication across a range of technical fields. Overall, ongoing research into ambiguity metrics promises to improve the accuracy and efficiency of natural language processing and facilitate more effective communication and collaboration between stakeholders.

3 DATASETS

In order to analyze lexical ambiguity in all the aforementioned ways, we needed two different datasets. The first dataset was created by hand and made of statements which are ambiguous to humans. This dataset was used to test the subjectivity and sentiment-based scores of ambiguous text. The second dataset was created by a user on GitHub, and was used to train the regression models.

Lexical Ambiguity Dataset

In order to test how sentiment analysis and subjectivity scores perform on ambiguous statements, the research team needed to create a dataset which was made up of entirely ambiguous statements. The English language has many types of ambiguity, but we chose to focus on two types which are particularly pressing in natural language processing matters - equivocal words and ambiguous structures. Equivocal word ambiguity is when the meaning of a phrase or sentence is unclear because one word has multiple meanings, several of which make sense in the given context. For example, upon hearing a sentence like “That was a quick duck!” it may not be immediately clear whether the speaker saw a well-trained athlete or a speedy bird. Structural ambiguity in language is when the meaning of a phrase or sentence is unclear because the order of the actual words themselves have multiple meanings. For example, when saying “the chicken is ready to eat,” it’s hard to tell whether the chicken has been prepared as a meal or whether it is hungry itself without additional information. The NLP team created a dataset of two hundred and sixty five of these kinds of statements, mixing in both lexical and structural ambiguities. Just over two-thirds of the dataset is made up of sentences with an equivocal word, and another third of sentences have an ambiguous structure. Around 3% of the sentences met both conditions.

Ambiguity Type	# of Statements	% of Total
Statements with an Equivocal Word	180 Statements	67.9%
Statements with an Ambiguous Structure	71 Statements	26.8%
Statements with an Equivocal Word and Ambiguous Structure	9 Statements	3.4%
All Statements	**265 Statements**	**100.0%**

Figure 1: Table of Ambiguity Breakdown in the Lexical Ambiguity Dataset

Each sentence was tagged by the creators to determine its ambiguity types and overall sentiment. For the purposes of this exercise, the team had to make a distinction between neutral sentiment and ambiguous sentiment. There were several examples in the dataset where a sentence's sentiment would change significantly based on the meaning or spoken tone. Both sentiment options are valid, but without the context of the sound or other sentences, the words as written had the ability to be either positive or negative, and were therefore marked as having "ambiguous" sentiment. This is distinct from "neutral sentiment", which has only cases where the sentiment is neither positive nor negative based on any interpretation of the original sentence created for the dataset.

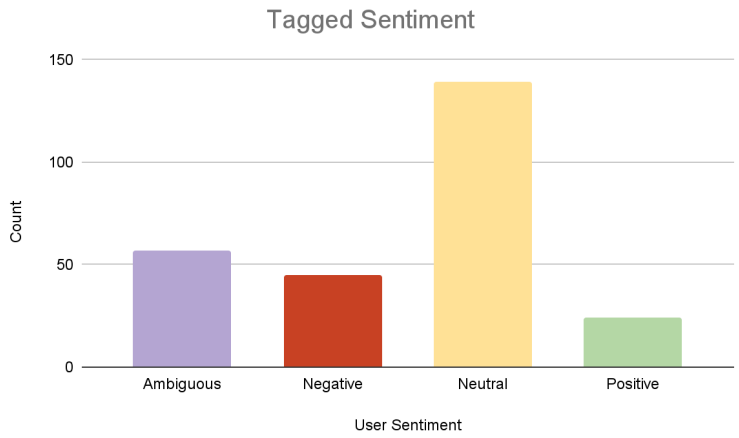


Figure 2: Total Number of Sentences by Sentiment Type

For all three ambiguity-type groups, most of the statements were tagged as having either neutral or ambiguous sentiment. Most of the sentences were simple statements, so neutrality makes sense. To break it down further, here is a cross-tabulation of the total number of sentences by both sentiment and ambiguity type. A plurality of sentences are neutral sentences with an equivocal word (about 33% of the whole dataset).

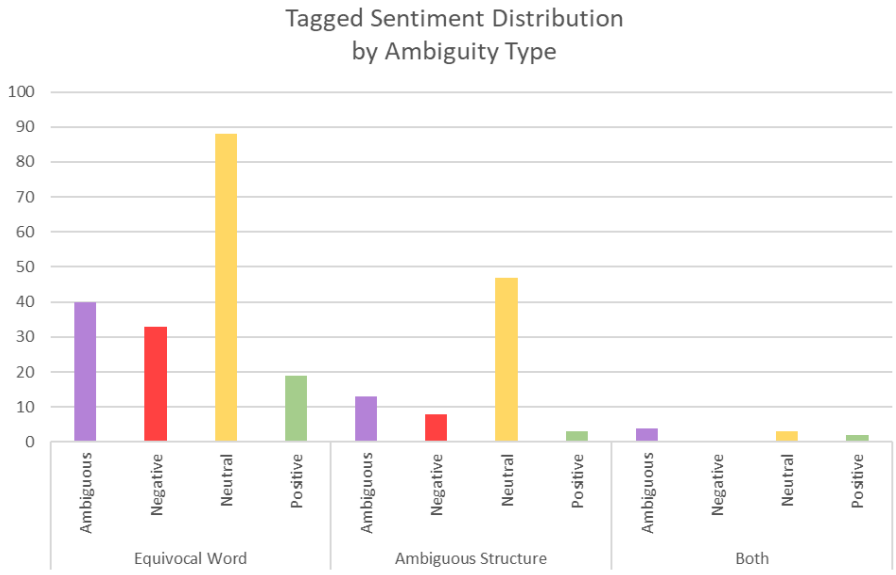


Figure 3: Total Number of Sentences by Sentiment Type

Training Dataset

The dataset at hand encompasses 5,988 distinct documents, all of which include at least one instance of 57 unique polysemic words. These words possess multiple meanings depending on the specific context in which they are used. Therefore, the meaning of a polysemic word can vary depending on the specific sense ID assigned to it. In this dataset, each word was used between 21 and 220 times across the nearly 6,000 sentences.

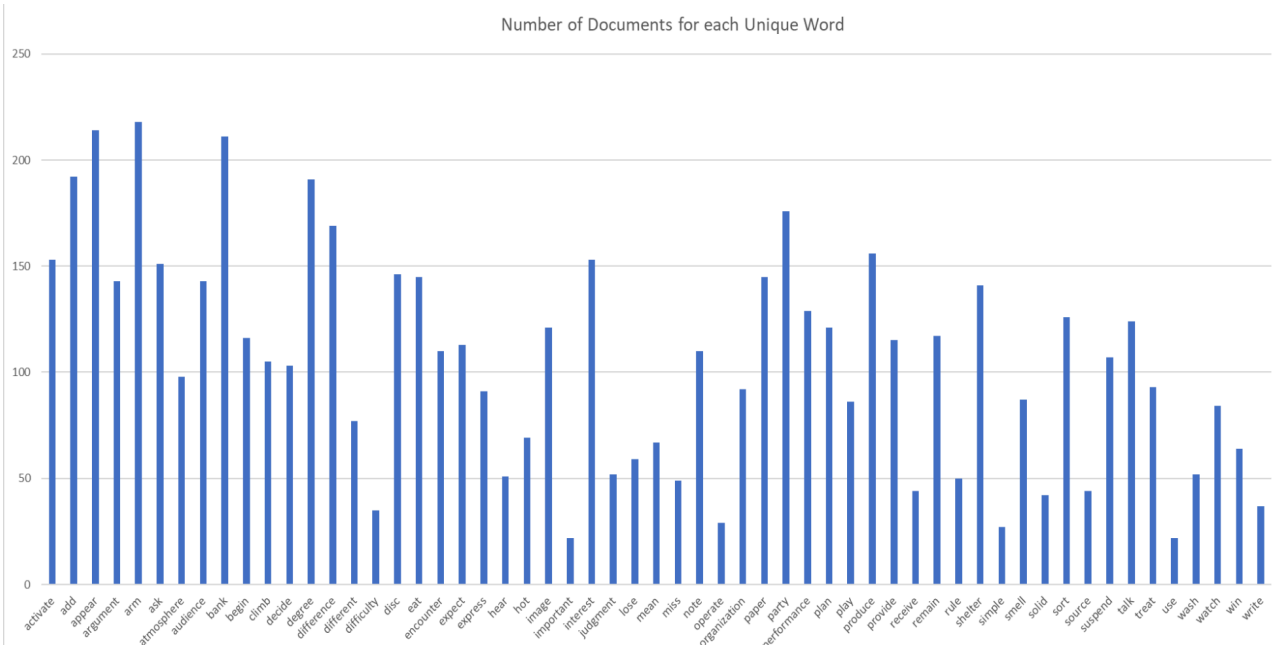


Figure 4: Total Number of Sentences for Each Word in the Training Data

In order to distinguish between the different meanings of a polysemic word, each context is identified and labeled with a unique “sense ID". For example, the word "bank" can refer to a financial institution or a riverbank depending on the specific context in which it is used. Therefore, in documents where "bank" is used to refer to a financial institution, it will be labeled with a sense ID of "1". Conversely, if "bank" is used in reference to a riverbank, it will be assigned a sense ID of "2". Similarly, if "bank" is used as a verb, a different sense ID will be assigned to it based on the specific context in which it is used. Therefore, the assignment of different sense IDs to polysemic words allows for the distinction of multiple contexts in which a word may be used, and facilitates the more accurate analysis of the data at hand. In this dataset, every word had between three and eleven sense IDs.

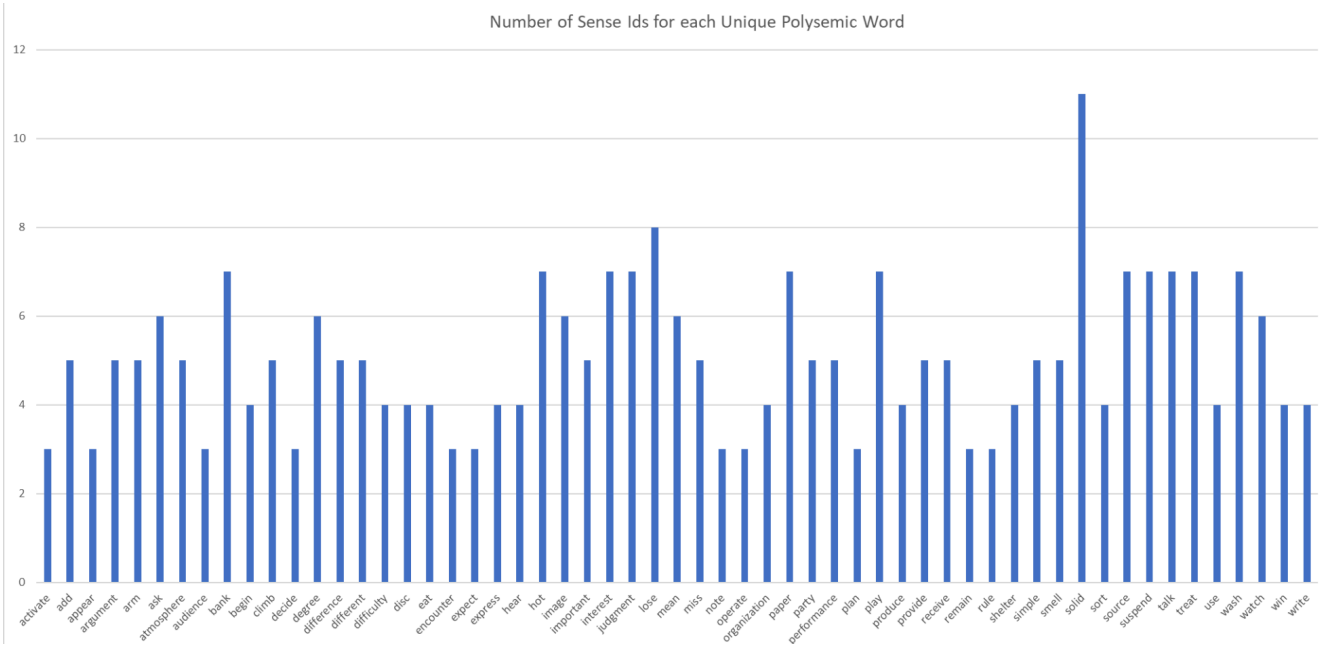


Figure 5: Total Number of Sense IDs for Each Word in the Training Data

4. METHOD

The methods for the research are shown in the diagram below. The top row (in red) indicates the portions of the methodology which relate to the subjectivity and sensitivity analysis conducted with the dataset the Rutgers team created. The bottom row (in green) indicates the portions of the methodology which relate to the regression analysis conducted with th

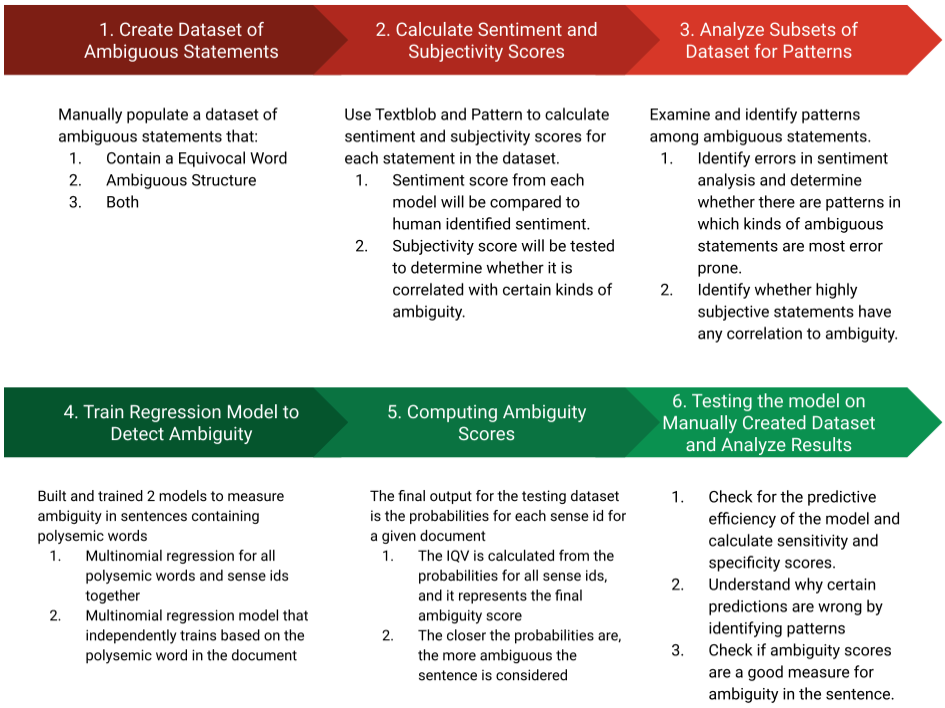


Figure 6: Outline of Study Processes

Sentiment and Subjectivity Methodology

First, the team manually created the dataset of 265 ambiguous statements. This is the dataset composed of equivocal words and ambiguous structures; it has already been described. Using this dataset, the team used Textblob and Pattern to calculate sentiment and subjectivity scores for each statement in the dataset. The sentiment score for each model was compared to the human-tagged sentiment score in the data.

The team looked at the data in subsets. The team was particularly interested in the cases where errors were present. For example: negative statements classified as positive, and positive statements classified as negative. The subsets were based on human-identified sentiment and compared the true sentiment to the sentiment scores generated by Textblob and Pattern to see if there were any trends in the data. The group also investigated whether polarity and subjectivity score in each model is prone to errors, or whether there is any suitable use for the task of identifying ambiguity, whether in sentiment, lexical, or structural. Based on this analysis, the team determined whether or not Textblob or Pattern can be used as post-hoc tests for identifying ambiguity in existing datasets.

Regression-Based Methodology

Second, a model was built to measure the ambiguity in sentences containing polysemic words. A text corpus of 5988 documents containing 57 unique polysemic words with different meanings represented by sense ids was used to train the model. In addition to example sentences and number of contexts for each unique equivocal word, the testing dataset generated the probabilities for each sense ID, or context of the equivocal word. The team calculated the Index of Qualitative Variation (IQV) for each equivocal word as a proxy measure for its ambiguity level - its ambiguity score. A word which has several meanings with similar likelihoods will have a high IQV, and thus be quite ambiguous. The closer the probabilities are, the more ambiguous the sentence is considered.

Two logistic regression models were trained. The first model is a multinomial regression for all polysemic words and sense ids together, and the second model is a multinomial regression model that independently trains based on the polysemic word in the document. The models considered only 10 words before and after the polysemic words for understanding context. (This choice was made because while context is necessary, too much context can actually reduce a model's efficiency.)

To check the predictive efficiency of the model, sensitivity and specificity scores were calculated. Sensitivity measures the proportion of actual positive cases that are correctly identified as positive, and specificity measures the proportion of actual

negative cases that are correctly identified as negative. The model was evaluated to ensure it had high sensitivity and specificity, which are essential to avoid false positives or negatives. In addition, the model was used to identify patterns in cases where the predictions were incorrect (similar to the methodology with the sensitivity and subjectivity scores). By identifying these patterns, it was possible to understand why certain predictions were wrong, and adjustments could be made to improve the model's accuracy in the future. Finally, the ambiguity scores for each sentence, calculated based on the IQVs of the equivocal words, were checked to understand whether they were a good measure of ambiguity in the sentence.

5. ANALYSIS and DISCUSSION

Sentiment and Subjectivity Analysis Results

In order to identify patterns among the ambiguous statements in the manually constructed dataset, various subsets were created based on the tagged variables.

Analysis of Statements with Ambiguous Sentiment

The data was subsetted to those statements which were tagged as having "ambiguous" sentiment. "Ambiguous" sentiment was given to statements that can be either positive or negative depending on the interpretation, of which there are potentially multiple. Unlike positive, negative, or neutral tagged statements, ambiguous tagged statements have no "correct" sentiment, so a proper sentiment analysis is not possible. However, Textblob and Pattern both utilize the score 0 to represent either purely neutral statements or potentially ambiguous ones.

For example, in the subset of ambiguous sentiment tagged statements, there were a total of 57 statements. Textblob gave 0 polarity scores to 100% of statements in the subset. Pattern gave 0 polarity scores to 54% of statements. According to documentation and user experience, Textblob and Pattern utilize 0 polarity when they are unable to determine a score, either due to positive and negative words averaging out or due to words that do not appear in the training datasets (Shah, 2020). Textblob and Pattern do differ to some extent, as Pattern assigns non-zero polarity scores to about 46% of the statements tagged with ambiguous sentiment. Of those statements which Pattern assigned non-zero polarity, all were also assigned non-zero subjectivity scores. This indicates that statements with zero polarity and subjectivity scores were unable to be scored by Pattern, and similarly Textblob, rather than them having truly zero polarity or being completely objective. It is possible that other statements tagged as neutral also exhibit this issue. Although, zero scores for these measures may be indicative of ambiguity it is not suited for the purpose of consistently identifying them.

Ambiguous sentiment statements which Pattern classified as positive or negative also received subjectivity scores. This indicates that opinion-like statements, typically with high polarity words, are more likely to be classified. However, given that these statements have ambiguous sentiment, it is dependent on the context whether they are correctly classified. For example, in Figure 7 (below), the first statement "Whenever you skip, she's happy," can have a positive or negative connotation:

1. Positive: "She's happy whenever you skip"
    a. Skip as in hopping around, which makes her happy
2. Negative: "She's happy whenever you skip"
    a. Skip as in your absence, which makes her happy

Pattern recognizes the word happy and assigns a high positive polarity to the statement, as expected. Similarly, the statement is classified as being very subjective. Pattern uses a bag-of-words approach to sentiment analysis, and this means that it recognizes the positive word in the statement and assigns a score accordingly. It is unable to detect the ambiguity in the sentence. However, in the case of this example, it is impossible to know without additional context around the statement, even for a human. Although Pattern would still be unsuited for detecting the ambiguity, a better more purpose-built model could be used to detect ambiguity. In conclusion, ambiguous sentiment adds an additional dimension to ambiguous statements which is difficult for NLP models and tools to contend with. When detecting sentiment, this type of ambiguity can not be properly handled with lexicon-based models. However, better models, i.e. the transformer, may be able to synthesize more context from a statement and predict the most likely meaning of a word, despite inherent ambiguity.

Statement	Ambiguity Type	User Sentiment	Textblob Polarity	Textblob Sentiment	Textblob Subjectivity	Pattern Polarity	Pattern Sentiment	Pattern Subjectivity
Whenever you skip, she's happy	Equivocal Word	Ambiguous	0.00	Neutral	0.00	0.80	Positive	1.00
He skips a lot, hope he's actually happy	Equivocal Word	Ambiguous	0.00	Neutral	0.00	0.80	Positive	1.00
Every scale is important in this line of work	Equivocal Word	Ambiguous	0.00	Neutral	0.00	0.40	Positive	1.00
He's an unbelievable worker.	Equivocal Word	Ambiguous	0.00	Neutral	0.00	-0.25	Negative	1.00
You look really hot	Equivocal Word	Ambiguous	0.00	Neutral	0.00	0.25	Positive	0.85
He ate the hot dog with relish	Equivocal Word	Ambiguous	0.00	Neutral	0.00	0.25	Positive	0.85
The wind was too strong to wind the sail	Equivocal Word	Ambiguous	0.00	Neutral	0.00	0.43	Positive	0.73
Skipping can be a good sign, but it can also be bad	Both	Ambiguous	0.00	Neutral	0.00	0.00	Positive	0.63
Someone needs to clean near the bank	Equivocal Word	Ambiguous	0.00	Neutral	0.00	0.23	Positive	0.55
Take your mother-in-law out back and shoot her	Equivocal Word	Ambiguous	0.00	Neutral	0.00	0.00	Neutral	0.00
Slay	Equivocal Word	Ambiguous	0.00	Neutral	0.00	0.00	Neutral	0.00
Break a leg	Equivocal Word	Ambiguous	0.00	Neutral	0.00	0.00	Neutral	0.00
The detective saw the man with the binoculars	Ambiguous Structure	Ambiguous	0.00	Neutral	0.00	0.00	Neutral	0.00
I heard her playing the guitar wearing a red dress	Ambiguous Structure	Ambiguous	0.00	Neutral	0.00	0.00	Neutral	0.00
I heard her sing with the headphones	Ambiguous Structure	Ambiguous	0.00	Neutral	0.00	0.00	Neutral	0.00
I fed her cat food.	Ambiguous Structure	Ambiguous	0.00	Neutral	0.00	0.00	Neutral	0.00
She left the bank	Equivocal Word	Ambiguous	0.00	Neutral	0.00	0.00	Neutral	0.00
They are hunting dogs	Ambiguous Structure	Ambiguous	0.00	Neutral	0.00	0.00	Neutral	0.00
She saw the bat flying around the room with her friend	Ambiguous Structure	Ambiguous	0.00	Neutral	0.00	0.00	Neutral	0.00
I read the book about the history of France in two days	Ambiguous Structure	Ambiguous	0.00	Neutral	0.00	0.00	Neutral	0.00

Figure 7: Screenshot of Output Demonstrating Textblob and Pattern Scores for Ambiguous Statements

Regression-Based Analysis Results

In this experiment, two different models were evaluated for their ability to predict the sense of polysemic words. The first model, which used a multinomial approach with 286 sense ids, demonstrated a moderate accuracy of 45.6% for the testing dataset, but only achieved a fit of 43.2% for the training dataset, indicating that there was still room for improvement in its performance.

The second model, which was an aggregate of 57 independent models, showed much better accuracy, especially for words with a higher number of documents. The accuracies of each individual word were recorded in a graph below, which provided insights into the specific words that were better predicted by the model.

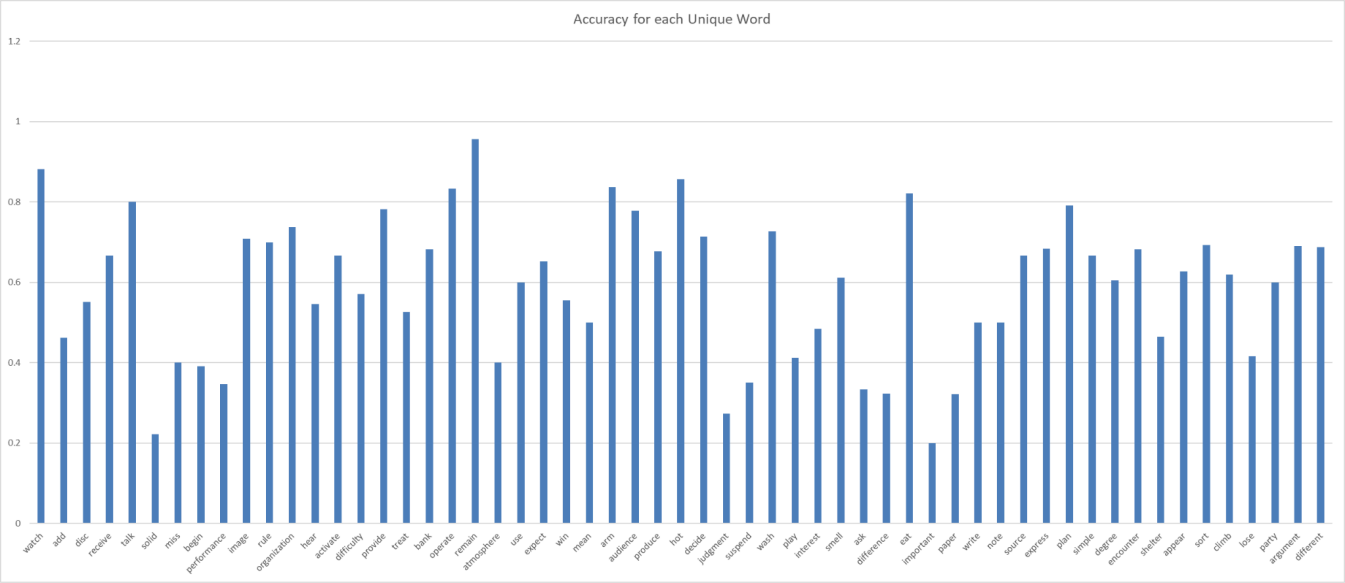


Figure 8: Percent Accuracy for Each Unique Word Represented in the Model

However, when the trained models were applied to the ambiguous dataset created by us, the results were mostly inaccurate. This was attributed to the high level of ambiguity in the statements, as well as the limited number of words available for the model to understand context. These findings suggest that while the second model demonstrated better overall performance, there are still limitations in its ability to handle new datasets with unique characteristics. The confusion matrix for each individual word is attached in the appendix. Following is a table of results:

Sentence	Word	Ambiguity Score
He sat by the bank	bank	[0.590]
On Wednesday, they made a watch	watch	[0.479]
The sewer expansion project is nearing complet...	eat	[0.314]
You look really hot	hot	[0.292]
She left the bank	bank	[0.660]
The wind was too strong to wind the sail	win	[0.862]
He ate the hot dog with relish	hot	[0.293]

Figure 9: Example Sentences and their Ambiguity Scores

This experiment highlights the importance of understanding the limitations of machine learning models and the need to continuously improve their performance. Future studies could explore ways to overcome these limitations, such as incorporating more contextual information or developing new models that can handle higher levels of ambiguity.

6. WEAKNESSES, FUTURE RESEARCH

The research conducted so far has not been without its limitations. In both parts of the analysis, the research was structured around very small sample sizes, particularly in the context of natural language processing work. On average, each unique word in the regression model training dataset had 105 documents and 5 sense IDs, which is quite small for model training. A larger dataset provides more examples for the model to learn from, reducing overfitting and increasing generalization to new data. However, too much data can lead to slower training times and increased computational costs.

Additionally, the research may have yielded more interesting and relevant results if the team had chosen to limit its work to a specific domain - for example, the context of fiction writing. As it is, the datasets are quite broad in context, and the ambiguous dataset is skewed in terms of having fewer ambiguous statements, and almost no statements which were both lexically and equivocally ambiguous in nature. If the team had more time to construct the datasets and train the model, results could have been fine tuned further.

In terms of the regression model, the team used Term Frequency-Inverse Document Frequency (TF-IDF) vectors to represent the importance of individual words. However, BERT embeddings may have been more appropriate since they capture the semantic relationships between words and can represent the meaning of a sentence as a whole. TF-IDF vectors may not be as effective as BERT embeddings for tasks that require understanding the context and meaning of the text, and as such an improved model would attempt to use BERT embeddings going forward.

In general, the regression-based model seems more promising for future attempts at identifying ambiguity in these datasets. However, the path forward is difficult. The current model focuses exclusively on lexical ambiguity resulting from polysemic words. Addressing other forms of ambiguity requires more complex algorithms, and combining multiple approaches to tackle all types of ambiguity is a particularly challenging task. Another challenge for future researchers is the switch from logistic regression to a neural network model. Neural network models are typically more efficient than logistic regression models in addressing ambiguity because they can learn complex nonlinear relationships between inputs and outputs. Unlike logistic regression models, which are limited to linear decision boundaries, neural networks can model more complex decision boundaries that can capture subtle variations in the data. They can also perform well on tasks where there is a high degree of ambiguity or uncertainty, as they can learn to make probabilistic predictions that take into account multiple sources of evidence.

7. CONCLUSION

This research study looked at three different potential tools for analyzing ambiguity in text. Based on the results, Textblob, Pattern, and even the trained regression model are weak post-hoc tests for identifying potentially ambiguous statements in large unstructured datasets. While subjectivity and polarity scores from sentiment analysis models may be altogether unsuited for detecting ambiguity, it is possible that a better-trained regression model will provide better results. Sense IDs provide a very valuable framework to train ambiguity detection models on, allowing for the machine to capture multiple possible meanings and identify the most likely one based on context. With a larger dataset and a more efficient model (perhaps via neural networks), perhaps future researchers will be able to solve the puzzle and identify and remedy ambiguity in text.

ACKNOWLEDGEMENTS

The research team would like to thank GitHub user niloofarMRZ (https://github.com/niloofarMRZ) for the contribution of their dataset which we used for our regression model training data. The team would also like to thank Professor Jim Samuel for his guidance and support throughout the semester.

REFERENCES

Barberá, P., Jost, J.T., Nagler,J., Tucker,J.A.& Bonneau,R.(2015). Tweeting from left to right: Is online political communication more than an echo chamber? Psychological Science ,26(10),1531–1542.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1), 1-8.

Erkan, G., & Radev, D. R. (2004). LexRank: Graph-based lexical centrality as salience in text summarization. Journal of artificial intelligence research, 22, 457-479.

Gharibi, W., et al. (2018). Cross-Domain Ambiguity Detection in Requirements Engineering.

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 168-177.

Kageura, K., & Umino, B. (1996). Methods for measuring lexical ambiguity. International Journal of Lexicography, 9(1), 1-26.

Kim, S. M., & Hovy, E. (2006). Extracting opinions, opinion holders, and topics expressed in online news media text. In Proceedings of the Workshop on Sentiment and Subjectivity in Text (pp. 1-8).

Liu, B., Zhang, L., & Tseng, B. (2012). Using multiple sources to improve the accuracy of sentiment classification. In Proceedings of the 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology (Vol. 1, pp. 205-212). IEEE.

McDonald, R. (2006). Discriminative sentence compression with soft syntactic evidence. In Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (pp. 377-384).

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10 (pp. 79-86).

Pitler, E., & Nenkova, A. (2008). Revisiting readability: A unified framework for predicting text quality. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (pp. 186-195).

Shah, P. (2020). *My absolute go-to for sentiment analysis - textblob.* Medium. Retrieved April 25, 2023, from https://towardsdatascience.com/my-absolute-go-to-for-sentiment-analysis-textblob-3ac3a11d524#:~:text=TextBlob%20is%20a%20simple%20library,classifying%20negative%20and%20positive%20words

Wiebe, J., Wilson, T., & Cardie, C. (2004). Annotating expressions of opinions and emotions in language. Language Resources and Evaluation Conference.

Xu, W., Callison-Burch, C., & Napoles, C. (2015). Problems in current text simplification research: New data can help. Transactions of the Association for Computational Linguistics, 3, 283-297.

Zeman, D., Dušek, O., Cinková, S., Hajic, J., Hajicová, E., Hladká, B., ... & Zabokrtský, Z. (2016). CoNLL 2016 shared task on multilingual parsing from raw text to universal dependencies. arXiv preprint arXiv:1606.07579.

APPENDIX 1 - LEXICAL AMBIGUITY DATASET

Attached is the dataset the team created and used for subjectivity and sensitivity score analysis.

Sentence	Equivocal Word	Ambiguous Structure	Sentiment
Duck!	Yes	No	Neutral
The chicken is ready to eat	No	Yes	Neutral
Shoot!	Yes	No	Ambiguous
He's a good shot	Yes	No	Positive
I'm dead	Yes	No	Ambiguous
let's eat grandma	No	Yes	Ambiguous
he sat by the bank	Yes	No	Neutral
the board keeps getting in the way	Yes	No	Negative
the bat hit me in the face	Yes	No	Negative
pass the bass please	Yes	No	Positive
she made a bow and left	Yes	No	Ambiguous
the bass is pumping	Yes	No	Ambiguous
that's the lead	Yes	No	Neutral
look at that crane!	Yes	No	Neutral
On Wednesday they made a watch	Yes	No	Neutral
I broke a nail	Yes	No	Negative
can you help me make these bows?	Yes	No	Neutral
his crown's been dented	No	No	Negative
my sibling is coming to the family reunion and they are very excited	No	Yes	Positive
what a big fan!	Yes	No	Positive
it was a tie	Yes	No	Neutral
this is a bad table - make it again	Yes	No	Negative
that mission broke the seal	Yes	No	Negative
he had a bad trip	Yes	No	Negative
that's a big plant	Yes	No	Neutral
he's got four diamonds	Yes	No	Positive
I hate jams	Yes	No	Negative
I don't like molds	Yes	No	Negative
Scales make me feel uncomfortable	Yes	No	Negative
his sole was flat and pale	Yes	No	Neutral
on the beach, she was greeted with waves	Yes	Yes	Neutral
there were bugs in the trunk	No	No	Negative
he saw her running down the block	No	Yes	Neutral
he likes to lie in the mornings	Yes	Yes	Ambiguous
their rolls didn't work out the way they wanted.	No	No	Negative
The dog is ready to eat	No	Yes	Ambiguous
The professor said on Friday she would give a test	No	Yes	Neutral
The burglar robbed the woman with the knife.	No	Yes	Negative
Visiting friends can be annoying.	No	No	Negative
Everyone at the game saw her duck.	No	Yes	Ambiguous
Seb got the bath ready for his son wearing a blue cape.	No	Yes	Neutral
Cowards run in my family	Yes	Yes	Ambiguous
I bought blue jeans and shirts	No	Yes	Neutral
Her pink dog pen was broken	Yes	No	Negative
We can add that to the case	Yes	Yes	Neutral

Sentence	Equivocal Word	Ambiguous Structure	Sentiment
My mouse died	Yes	No	Negative
Look at that spot	Yes	No	Neutral
The car is parked a yard away	Yes	No	Neutral
The teeth are worn down	Yes	No	Neutral
It's growing a new branch	Yes	No	Positive
Please do not disturb the set	Yes	No	Neutral
After they got a new head it worked much more smoothly	Yes	No	Positive
After digging around, they found some steel inside	Yes	No	Positive
She had a big bill	Yes	No	Neutral
After thinking about it, I gave her a ring	Yes	No	Neutral
Looking closely I saw a single tear	Yes	No	Negative
Every time we get in a boat, we end up rowing	Yes	No	Ambiguous
At church, all the couples were in a row	Yes	No	Ambiguous
New drills helped the army build tents more quickly	Yes	No	Positive
Small pieces were coming off the plane	Yes	No	Ambiguous
If you write down the address I'll deliver it	Yes	No	Neutral
What's the most important object of your life?	Yes	No	Neutral
She felt a brush against her hair	Yes	No	Neutral
The government is working on an experimental compound	Yes	No	Neutral
The archer took a bow after his performance	Yes	No	Neutral
The author's character was deeply flawed	Yes	No	Negative
The historian couldn't find a date	Yes	No	Negative
The politician went to the House on Tuesday	Yes	No	Neutral
I asked the king about his favorite subjects	Yes	No	Neutral
The master electrician reminded his apprentice to stay grounded	Yes	No	Neutral
The construction worker walked into a bar	Yes	No	Ambiguous
After the line broke, the sewer had to close for a day.	Yes	No	Negative
At recess, she decided to skip	Yes	No	Positive
That was a really rough wake	Yes	No	Negative
The sales rep told me to get down	Yes	No	Ambiguous
He ran by the bank	Yes	No	Neutral
She fell near the bank	Yes	No	Neutral
Next to the bank, there is a fish market	Yes	No	Neutral
There's a bank by the river	Yes	No	Neutral
That bank is going under	Yes	No	Negative
When the bank recovers, it'll be good for the community	Yes	No	Positive
The bank is always slippery	Yes	No	Neutral
Someone needs to clean near the bank	Yes	No	Ambiguous
Both banks are closed	Yes	No	Ambiguous
Both banks are open tomorrow, they're across the river from each other	Yes	No	Neutral
His jacket was down	Yes	No	Neutral
She had a lot of huge fans	Yes	No	Positive
Our next date is next month	Yes	No	Neutral
I hope I get another date	Yes	No	Neutral
She missed five dates in a row	Yes	No	Negative
Whenever a date comes up, he gets angry	Yes	No	Negative
I have a court date next week, I can't miss it	Yes	No	Ambiguous
You only get a date if you put in the effort	Yes	No	Neutral

Sentence	Equivocal Word	Ambiguous Structure	Sentiment
They decide who gets a date and who doesn't	Yes	No	Ambiguous
Whatever date you get, be grateful you got one	Yes	No	Ambiguous
Whenever you skip, she's happy	Yes	No	Ambiguous
Skipping every time shows that you don't care	Yes	No	Ambiguous
I've never skipped, I don't think I know how	Yes	No	Neutral
He skips a lot, hope he's actually happy	Yes	No	Ambiguous
Skipping can be a good sign, but it can also be bad	Yes	Yes	Ambiguous
He gave a her two shoes and shirts	No	Yes	Neutral
I purchased black pants and shoes	No	Yes	Neutral
It doesn't matter, he needs to get two new jackets and sweaters	No	Yes	Neutral
His duck was impressive	Yes	Yes	Positive
Her duck was so quick, I'm impressed	Yes	Yes	Positive
Another duck would be too much	Yes	No	Ambiguous
They added a plant to our town	Yes	No	Ambiguous
Plants are a large part of our economy	Yes	No	Ambiguous
Depending on what kind of plant it is, it can actually improve your health	Yes	No	Positive
He owns a lot of plants, but doesn't pay too much attention to them	Yes	No	Negative
They have people to make sure the plants are doing well	Yes	No	Neutral
I missed the wake	Yes	No	Negative
They jumped into the wake	Yes	No	Neutral
In the wake of his death, they were devastated	Yes	No	Negative
He owns three banks and plants	Yes	Yes	Ambiguous
Their company is in the plant business	Yes	No	Neutral
The shareholders approved the purchase of a new plant	Yes	No	Ambiguous
Every new drill is inspected	Yes	No	Neutral
The new drill increased safety by a lot	Yes	No	Neutral
He drilled it into him	Yes	No	Ambiguous
She did really well with the drills	Yes	No	Positive
No matter how many drills they complete, its still not enough	Yes	No	Negative
Too much drilling can be a bad thing	Yes	No	Negative
I don't think I've seen that drill	Yes	No	Negative
She spent so long looking for a match	Yes	No	Neutral
Matches aren't necessary, but are helpful	Yes	No	Neutral
She had her matches, and he had his	Yes	No	Neutral
When they got their match, they were so happy	Yes	No	Positive
The court is large and spacious	Yes	No	Positive
A court is not place to propose	Yes	No	Neutral
People who are ate there frequently praised it	No	Yes	Positive
Students who studied there frequently complained about it	No	Yes	Negative
Families who went to that doctor frequently spent a lot of time waiting	No	Yes	Negative
The workers who started there frequently made more money	No	Yes	Neutral
The new Italian pastry chef is great	Yes	No	Positive
Seven boys and girls were in the school at the time	No	Yes	Neutral
I don't think its an issue if the two men and women come to the party	No	Yes	Ambiguous
I don't think the drill is working well for the new recruits	Yes	No	Negative
She wants the new drill put together for the next meeting	Yes	No	Neutral
Every new spot on the map makes it harder to read	Yes	No	Ambiguous

Sentence	Equivocal Word	Ambiguous Structure	Sentiment
Every record you bring here is useful	Yes	No	Positive
His collection of records is quite impressive	Yes	No	Neutral
The seal on the docks is troubling	Yes	No	Negative
If you want it to work, we need better seals	Yes	No	Neutral
Every seal has to be a good seal	Yes	No	Neutral
My dad was a SEAL	Yes	No	Neutral
Every time I saw the wood, he was there	Yes	No	Ambiguous
If you saw the tree, you might have issues	Yes	No	Ambiguous
She asked for two rings and necklaces	Yes	Yes	Neutral
Give me a ring when you get the chance	Yes	No	Neutral
I don't believe a ring will be sufficient	Yes	No	Neutral
If you get rid of the scales, we won't be able to cook properly	Yes	No	Neutral
Every scale is important in this line of work	Yes	No	Ambiguous
I believe they had large scales but small bodies	Yes	No	Neutral
The next subject must be important	Yes	No	Neutral
I don't think you should change subjects my lord	Yes	No	Neutral
There hundreds of cells in there	Yes	No	Neutral
There are billions of cells in there	Yes	No	Neutral
Every cell is an important part of the whole system	Yes	No	Neutral
No matter how many cells are destroyed we continue to create new ones	Yes	No	Ambiguous
I saw the man walking along with a bat in his hand.	Yes	No	Neutral
John poked the man with the stick.	No	Yes	Negative
Sarah knew that she did well in the test, and so did Bob.	No	Yes	Positive
The teacher wrote an example of a letter on the board.	Yes	No	Neutral
Flying planes can be dangerous	Yes	No	Negative
“That was a big wave” said Joe and Simon	No	Yes	Neutral
Did you enjoy your trip?	Yes	No	Neutral
I don’t want anything to do with your club	Yes	No	Negative
Pete vowed to never lie again	Yes	No	Neutral
Tim and Lucy were part of different races	Yes	No	Neutral
We’ll have to look at the table to work out the answer	Yes	No	Positive
I saw someone across the street with binoculars	No	Yes	Neutral
Jon ran a bath for his son wearing a black t-shirt	No	Yes	Neutral
Call me a cab, please?	No	Yes	Neutral
The dog liked to guard the house and the postman could not make it to the door because he was barking viciously	No	Yes	Neutral
There is a bird in a cage that can talk	No	Yes	Neutral
I saw a man on a hill with a telescope	No	Yes	Neutral
Look at the dog with one eye	No	Yes	Neutral
I sent the bill to John.	Yes	No	
Let’s stop controlling people	No	Yes	Negative
Iraqi Head Seeks Arms	Yes	No	Ambiguous
There is an old man fishing on the bank	Yes	No	Neutral
Raj tried to reach his friend on the mobile, but he didn’t attend	No	Yes	Neutral
They are hunting bears with guns	No	Yes	Negative
The teacher graded the papers on the desk	No	Yes	Neutral
They're going to read a book about bats	Yes	No	Neutral

Sentence	Equivocal Word	Ambiguous Structure	Sentiment
How do you make a turtle fast? Take away his food bowl.	Yes	No	Neutral
I shot an elephant in my pajamas	No	Yes	Negative
It is too hot to eat	Yes	No	Negative
The duchess can’t bear children	Yes	No	Negative
The police shot the rioters with guns	Yes	No	Negative
The German teachers visited the British Museum yesterday	Yes	No	Neutral
Ralph took my picture.	Yes	No	Neutral
We will give you a ring tonight	Yes	No	Neutral
Jim took me to the court	Yes	No	Ambiguous
We were late but the coach was late too	Yes	No	Negative
Because of the cold, I had to put on a warm coat	Yes	No	Neutral
He stood watching the fireworks in the backyard	No	Yes	Neutral
We need more experienced drivers to drive the buses	Yes	No	Neutral
I repaired the car and returned the following day	No	Yes	Neutral
They fed her dog biscuits	No	Yes	Neutral
Tall boys and girls are needed to participate in the activities	No	Yes	Neutral
We gave the library books.	No	Yes	Neutral
The people who saw the movie frequently praised it.	No	Yes	Neutral
The protesters demanded more specific answers	No	Yes	Neutral
The girl who sat with Jane sometimes played the piano	No	Yes	Neutral
Suda is a very faithful and dependable clerk	Yes	No	Positive
Mother has baked potatoes for dinner	No	Yes	Neutral
Pat may not read the paper	Yes	No	Neutral
The hay farmer drank through a straw	Yes	No	Neutral
Ross was told what to do by the bank	Yes	No	Neutral
Transgenders appeal to pope	No	No	Neutral
Take your mother-in-law out back and shoot her	Yes	No	Ambiguous
The River Ravi flows in what state?	Yes	No	Neutral
She is looking for a match	Yes	No	Neutral
You know, somebody actually complimented me on my driving today. They left a little note on the windscreen; it said, 'Parking Fine.' So that was nice.	Yes	No	Neutral
Complaints about NBA referees growing ugly	No	Yes	Neutral
Squad helps dog bite victim	No	Yes	Neutral
Hospitals are sued by 7 foot doctors	No	Yes	Neutral
Come meet our new French pastry chef	No	Yes	Neutral
Reagan wins on budget, but no more lies ahead	Yes	No	Neutral
I saw a tall tree outside the house	Yes	No	Neutral
Farmer Bill dies in House	Yes	No	Neutral
TWO SOVIET SHIPS COLLIDE, ONE DIES	No	Yes	Neutral
KIDS MAKE NUTRITIOUS SNACKS	No	Yes	Neutral
War Dims Hope for Peace	No	Yes	Neutral
New Vaccine May Contain Rabies	Yes	No	Ambiguous
He put the dough in the bank.	Yes	No	Neutral
Ross baked the cake in the freezer	No	Yes	Neutral
Because you've got a lot riding on your tires	Yes	No	Neutral
A man like him is hard to find	Yes	No	Ambiguous
He's an unbelievable worker.	Yes	No	Ambiguous

Sentence	Equivocal Word	Ambiguous Structure	Sentiment
You would indeed be fortunate to get this person to work for you	Yes	No	Neutral
It seemed her career was just taking off	No	Yes	Neutral
There is nothing you can teach a man like him	No	Yes	Neutral
How can a man go eight days without sleeping?	Yes	No	Neutral
The sewer expansion project is nearing completion but City officials are holding their breath until it is officially finished	Yes	No	Neutral
The assembly passed and sent to the senate a bill requiring dog owners in New York City to clean up after their dogs, in penalty of $100 fine. The bill also applies to Buffalo.	Yes	No	Neutral
Two cars were reported stolen by the Groveton police yesterday	No	Yes	Neutral
We will sell gasoline to anyone in a glass container	No	Yes	Neutral
Dr. Benjamin Porter visited the school yesterday and lectured on "Destructive Pests." A large number were present.	No	Yes	Neutral
Gene Autry is better after being kicked by a horse	No	Yes	Neutral
Wanted. Man to take care of cow that does not smoke or drink.	Yes	No	Neutral
You look really hot	Yes	No	Ambiguous
Slay	Yes	No	Ambiguous
Break a leg	Yes	No	Ambiguous
I'm down	Yes	No	Ambiguous
The detective saw the man with the binoculars	No	Yes	Neutral
I heard her playing the guitar wearing a red dress	No	Yes	Neutral
I drove to the store with broken window	No	Yes	Negative
I heard her sing with the headphones	No	Yes	Neutral
I fed her cat food.	No	Yes	Neutral
She left the bank	Yes	No	Neutral
He picked up the bat	Yes	No	Neutral
The wind was too strong to wind the sail	Yes	No	Neutral
He ate the hot dog with relish	Yes	No	Ambiguous
They are hunting dogs	No	Yes	Ambiguous
I ate the sandwich on the bus	No	Yes	Neutral
She gave her cat to her sister wearing a pink hat	No	Yes	Neutral
They ate sandwiches with ham and cheese	No	Yes	Neutral
I convinced her children are important	No	Yes	Neutral
She saw the bat flying around the room with her friend	No	Yes	Ambiguous
After the accident, the police officer helped the woman with broken arm	No	Yes	Ambiguous
I read the book about the history of France in two days	No	Yes	Neutral
I shot an elephant in my pajamas	No	Yes	Negative
Mary gave the dog a bone in the kitchen.	No	Yes	Neutral
She cooked the soup with vegetables	No	Yes	Neutral
The band played the music in the park with the drums	No	Yes	Neutral
He read the book on the couch with the dog	No	Yes	Neutral
The woman drove the car into the garage with a broken door	No	Yes	Neutral
I saw the movie with the actor with my friend	No	Yes	Neutral
The lawyer examined the witness in the suit	No	Yes	Neutral
They painted the house with white spots	No	Yes	Neutral
John saw the man with the telescope while waiting for the bus	No	Yes	Neutral
The thief stole the painting in the dark with a flashlight	No	Yes	Negative
She cooked the chicken with vegetables in the oven	No	Yes	Neutral

Sentence	Equivocal Word	Ambiguous Structure	Sentiment
I heard the news about the fire on the radio in the morning	No	Yes	Negative
The doctor examined the patient in the gown	No	Yes	Neutral
The teacher gave the student pencil with eraser	No	Yes	Neutral
The child saw the man with the dog in the park with the leash	No	Yes	Neutral
The chef prepared the dish with the spicy sauce in the kitchen	No	Yes	Neutral
The construction worker built the house with the blueprints	No	Yes	Neutral
The firefighter put out the fire with the water hose	No	Yes	Neutral
The photographer took the picture of the model with the camera in the studio	No	Yes	Neutral
The cake was terrible with the frosting	No	Yes	Negative
The professor gave the lecture with the boring slides	No	Yes	Negative
She wrote the novel with the amazing characters	No	Yes	Positive
The band played the concert with the new instruments	No	Yes	Positive
The athlete won the championship with the record-breaking performance	No	Yes	Positive

APPENDIX 2 - AMBIGUOUS WORDS LIST

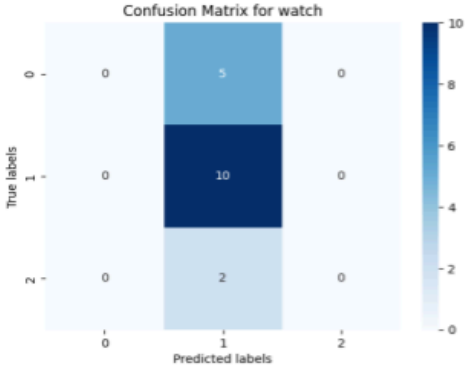
In developing the dataset of lexical ambiguity, the team found it helpful to create a list of equivocal words which may present a problem for NLP analysis tools. This made it easier to create sentences which were intentionally confusing.

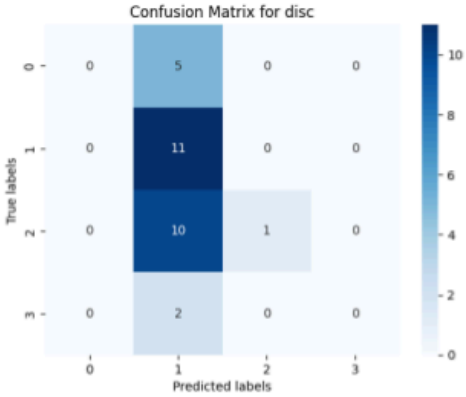
Homonyms	Definitions/Contexts
address	speech, where you live
agape	mouth, love
bank	river, investments
bar	pole, alcohol
bark	tree, dog
bass	fish, guitar, sound pitch
bat	animal, baseball
bill	beak, money
book	pages, reservation
bow	hair bow, arrows, curtsey, ship
branch	tree, subdivision
brush	light touch, painting, encounter
buckle	clasp, collapse
case	legal, container
cell	compartment, organism
character	fictional person, person's qualities
close	the door, nearby
club	social org, bat
compound	mixture, military base, to make worse
content	happy, subject matter
contest	competition, controversy
contract	smaller, agreement
crane	bird, construction equip
crown	top of head, royal headpiece
date	event time, romantic date
desert	barren land, abandon
diamond	stone, playing card
dove	bird, to dive
down	lower, bird fluff
drill	making holes, training exercise
duck	verb animal
dust	particles, vs removing them
extract	remove, something that has been removed

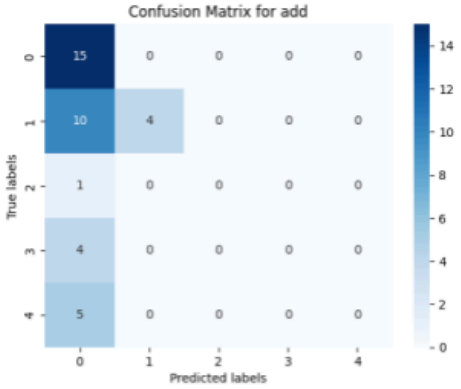
Homonyms	Definitions/Contexts
fair	carnival, justice, pleasant
fan	blowing air, admirer
foot	measurement, feet
ground	earth, electricity, "down to earth"/cool
head	highest part of body, person in charge, of a tool
House	legislative body, home
incorporate	form a corporation, include
invalid	weak person, something not recognized
jam	traffic, preserves
lead	metal, leader, past tense
lean	thin, rest against
leg	human, tripod, of a trip
letter	symbol, "dear so-and-so"
lie	rest, fib
light	heaviness, bright, pale color
March	month, walk
match	game, fire
minute	60s, tiny
mold	form, organism
moped	bike, gloomy
mouse	computer, animal
mouth	river connection, food hole
nail	wood, fingers
novel	book, new
object	physical thing, goal, "I disagree"
park	play, car
pen	writing, animal enclosure
permit	allow, license
pipe	musical, smoking, water
plane	wood, airplane
plant	industry, organism
polish	shoes, Polish people/language
pool	gather resources, swimming
present	right now, gift
record	account of event, act of recording it, vinyl
refuse	trash, nonconsent

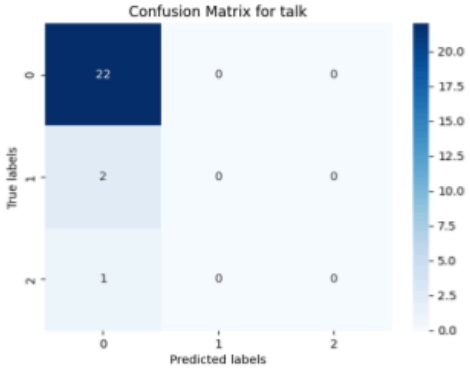
Homonyms	Definitions/Contexts
ring	finger, boxing, call
rock	music, geology
roll	cylinder, roll dice, baking, turn over
rose	flower, rise
row	line, rowing boat, argument
saw	see, wood
scale	fish weights
seal	animal, airtight, navy
set	belong together, in particular place, theater set
sewer	waste water, person who sews
sole	fish, feet
spot	stain, location
spring	season, coil
steel	metal, make strong
subject	place under, topic at hand, royal subject
table	work surface, rows and columns
tank	container, military equip
tear	drop, rip
teeth	saw, dentistry
tie	equal score, strip of cloth
tire	bored, wheel
train	transport, coach
trip	stumble, vacation
trunk	tree, storage
type	version, keyboard
wake	boat, funeral
watch	see, time keeper
wave	hello, ocean
wind	breeze, turning movement, curvy road
wound	hurt, wind
yard	lawn, 3ft

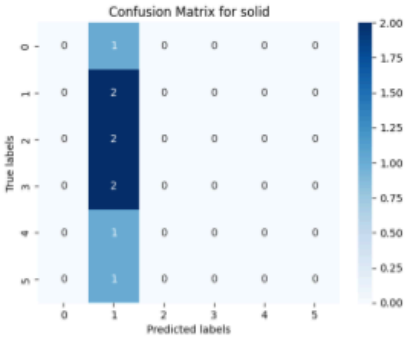
APPENDIX 3 - CONFUSION MATRIX FOR LOGISTIC REGRESSION MODEL

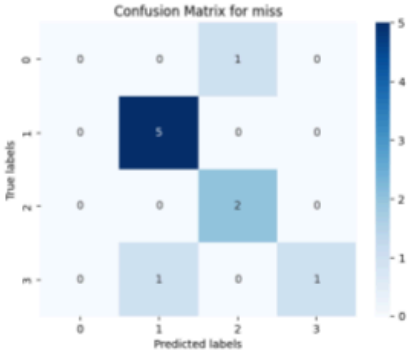


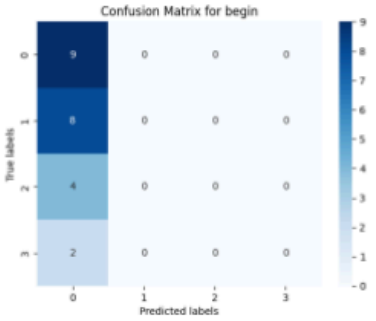


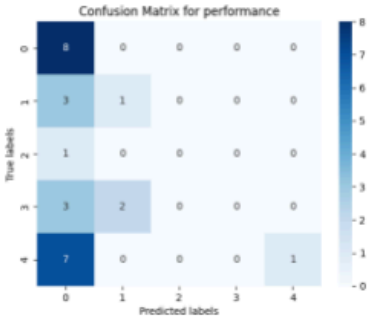


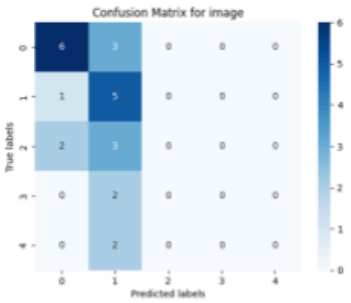


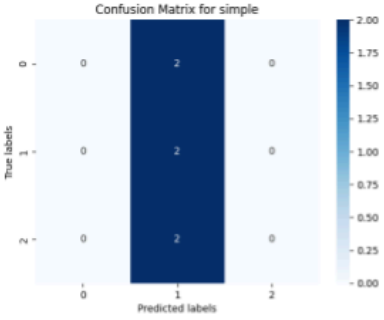


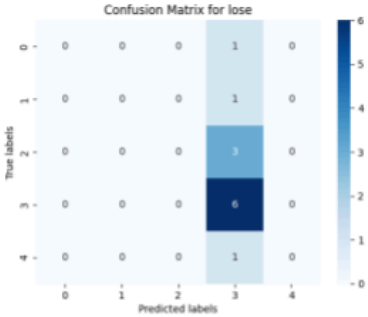


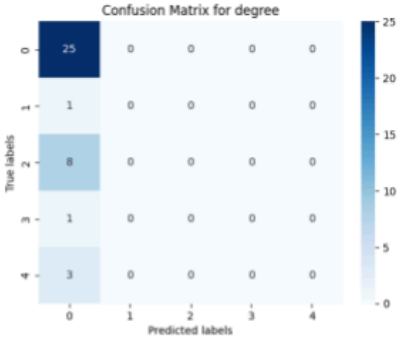


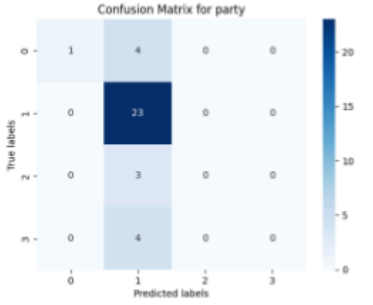


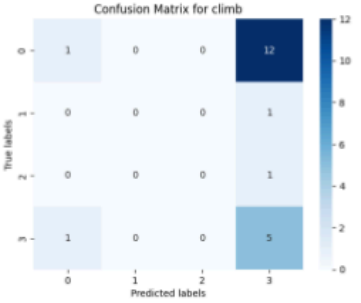


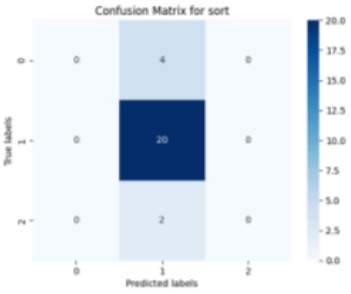

Confusion Matrix for simple
True labels
Predicted labels

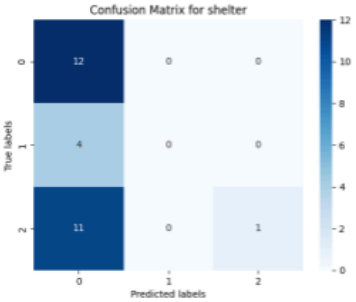

Confusion Matrix for lose
True labels
Predicted labels

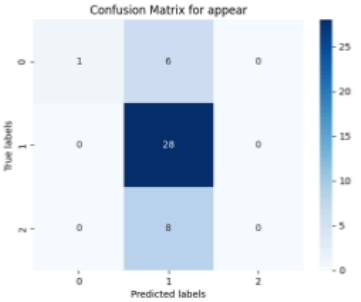

Confusion Matrix for degree
True labels
Predicted labels

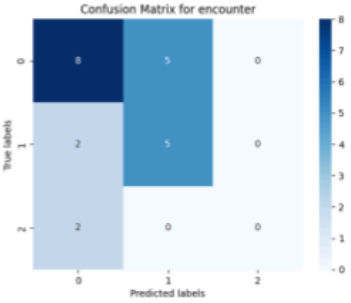

Confusion Matrix for party
True labels
Predicted labels

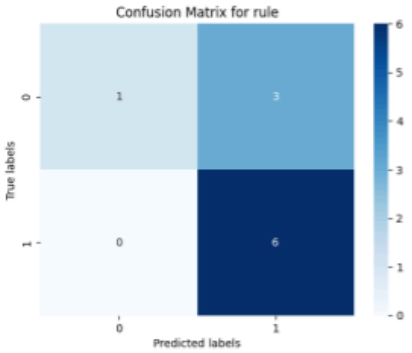

Confusion Matrix for climb
True labels
Predicted labels

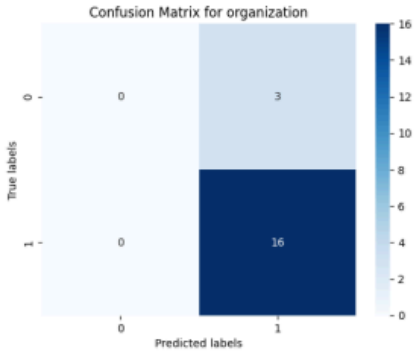

Confusion Matrix for sort
True labels
Predicted labels

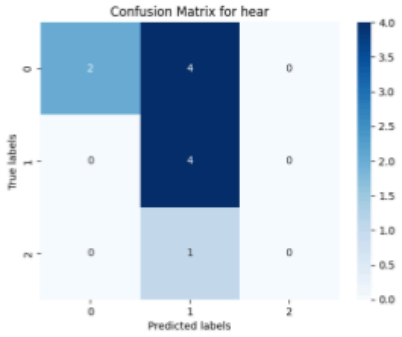

Confusion Matrix for shelter
True labels
Predicted labels

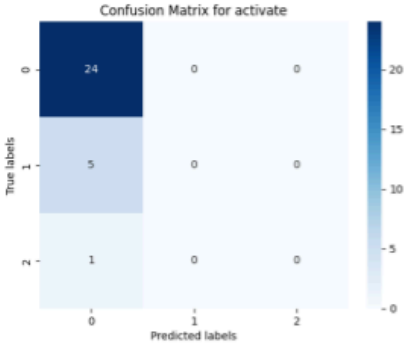

Confusion Matrix for appear
True labels
Predicted labels

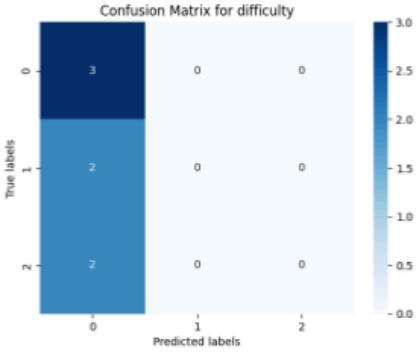

Confusion Matrix for encounter
True labels
Predicted labels

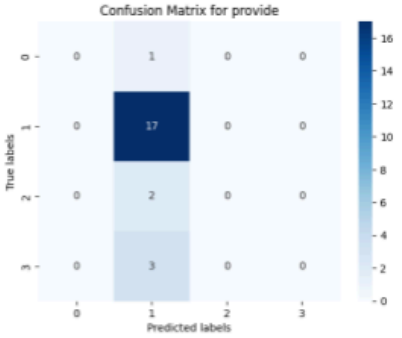

Confusion Matrix for rule
True labels
Predicted labels
1
3
0
6

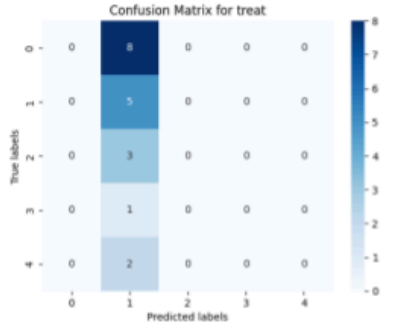

Confusion Matrix for organization
True labels
Predicted labels
0
3
0
16

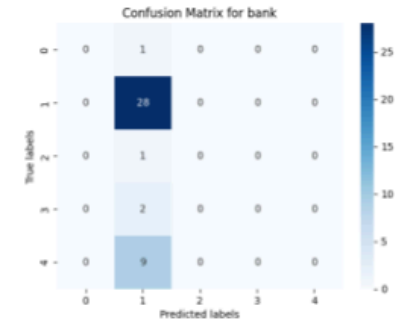

Confusion Matrix for hear
True labels
Predicted labels
2
4
0
0
4
0
0
1
0

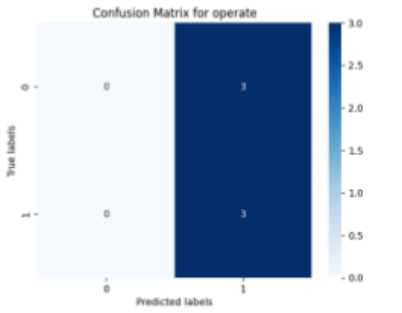

Confusion Matrix for activate
True labels
Predicted labels
24
0
0
5
0
0
1
0
0

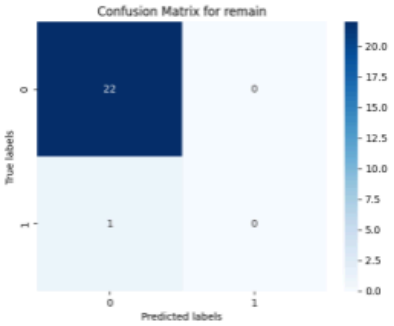

Confusion Matrix for difficulty
True labels
Predicted labels
3
0
0
2
0
0
2
0
0

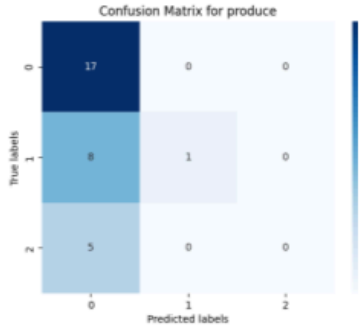

Confusion Matrix for provide
True labels
Predicted labels
0
1
0
0
0
17
0
0
0
2
0
0
0
3
0
0

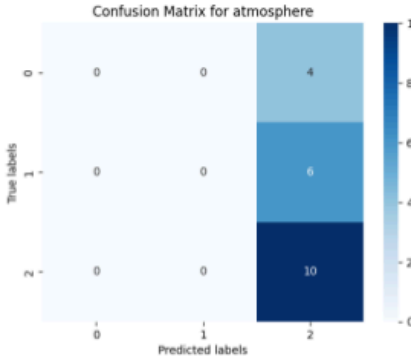

Confusion Matrix for treat
True labels
Predicted labels
0
8
0
0
0
0
5
0
0
0
0
3
0
0
0
0
1
0
0
0
0
2
0
0
0

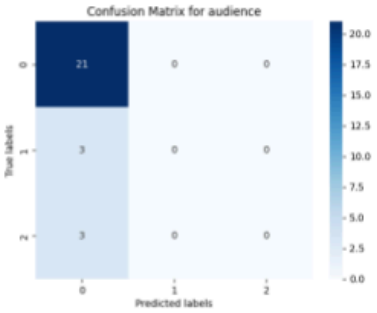

Confusion Matrix for bank
True labels
Predicted labels
0
1
0
0
0
0
28
0
0
0
0
1
0
0
0
0
2
0
0
0
0
9
0
0
0

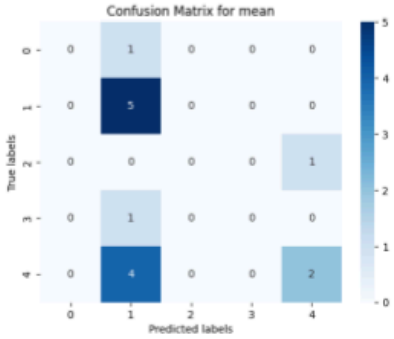

Confusion Matrix for operate
True labels
Predicted labels
0
3
0
3

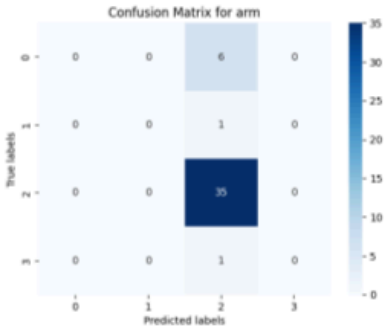

Confusion Matrix for remain
True labels
Predicted labels
22
0
1
0

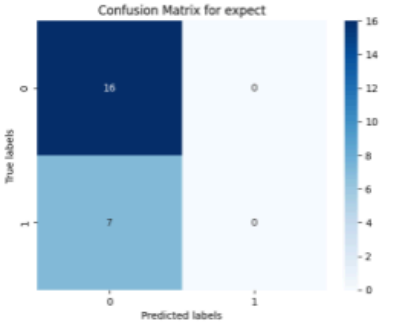

Confusion Matrix for produce
True labels
Predicted labels
17
0
0
8
1
0
5
0
0

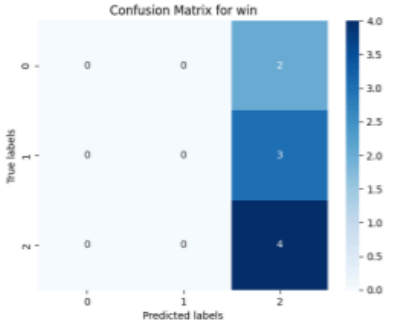

Confusion Matrix for atmosphere
True labels
Predicted labels
0
0
4
0
0
6
0
0
10

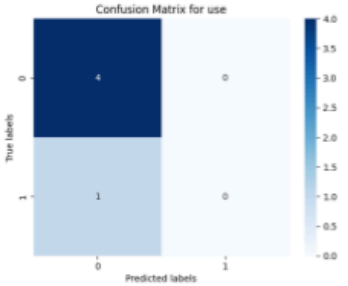

Confusion Matrix for audience
True labels
Predicted labels
21
0
0
3
0
0
3
0
0

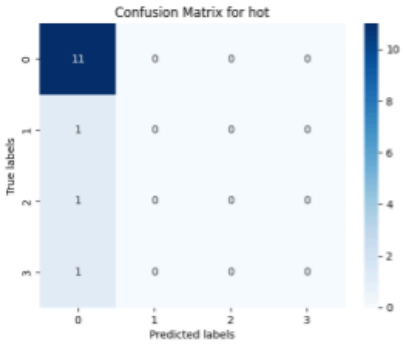

Confusion Matrix for mean
True labels
Predicted labels
0
1
0
0
0
0
5
0
0
0
0
0
0
1
0
1
0
0
0
0
4
0
0
2

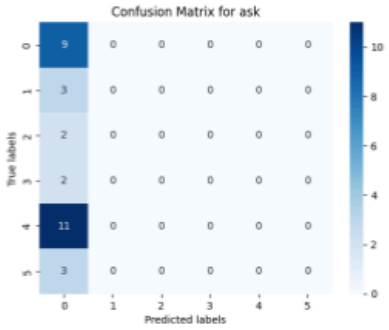

Confusion Matrix for arm
True labels
Predicted labels
0
0
6
0
0
0
1
0
0
0
35
0
0
0
1
0

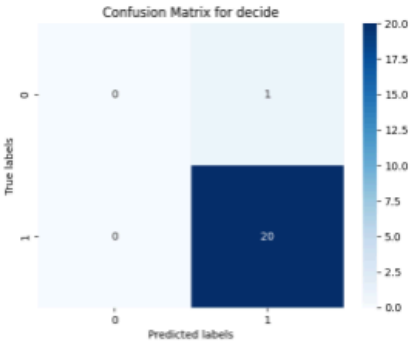

Confusion Matrix for expect
True labels
Predicted labels
16
0
7
0

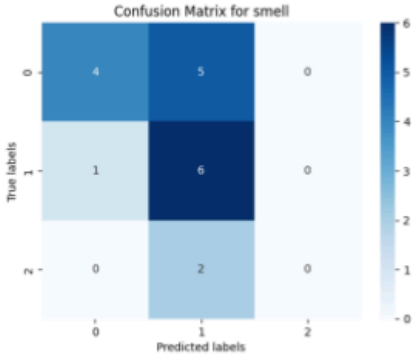

Confusion Matrix for win
True labels
Predicted labels
0
0
2
0
0
3
0
0
4

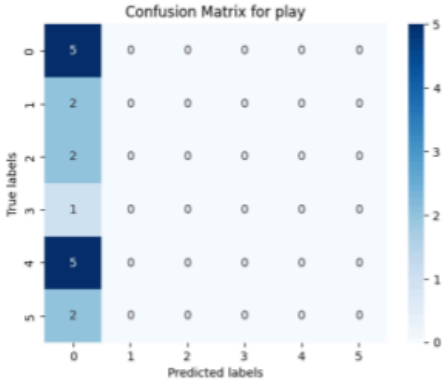

Confusion Matrix for use
True labels
Predicted labels
4
0
1
0

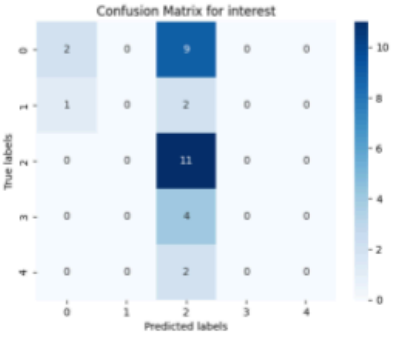

Confusion Matrix for hot
True labels
Predicted labels

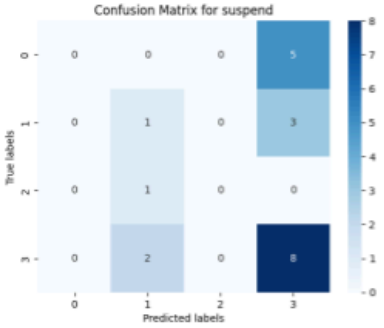

Confusion Matrix for ask
True labels
Predicted labels

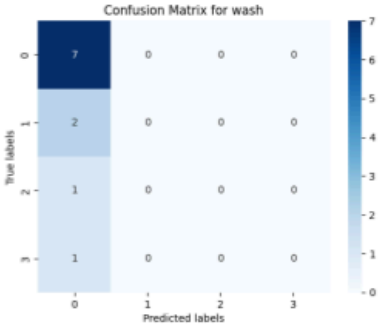

Confusion Matrix for decide
True labels
Predicted labels

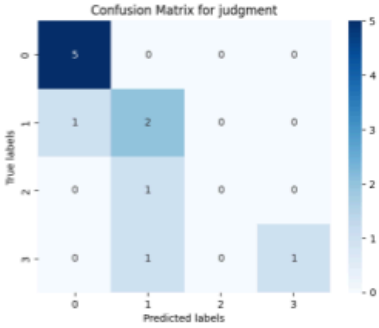

Confusion Matrix for smell
True labels
Predicted labels

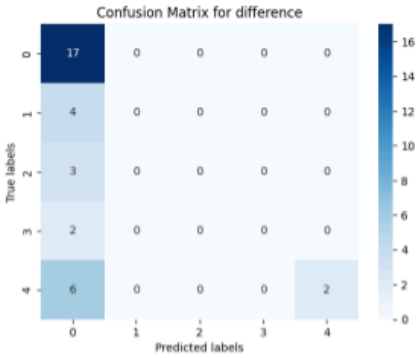

Confusion Matrix for play
True labels
Predicted labels

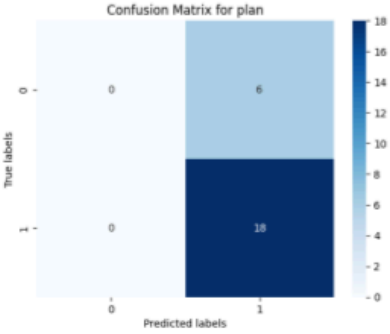

Confusion Matrix for interest
True labels
Predicted labels

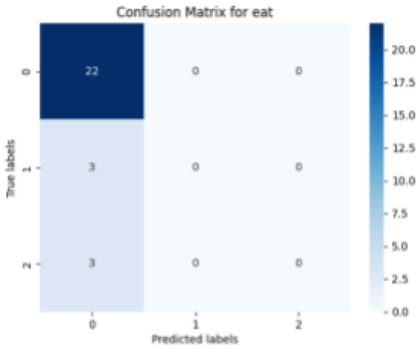

Confusion Matrix for suspend
True labels
Predicted labels

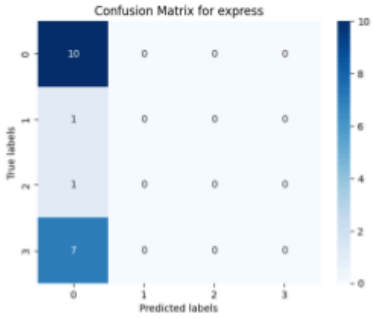

Confusion Matrix for wash
True labels
Predicted labels

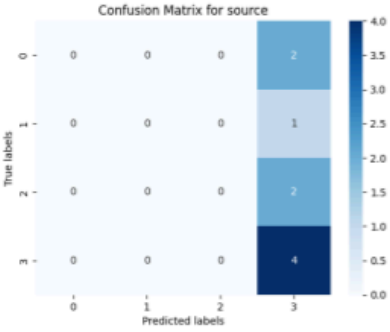

Confusion Matrix for judgment
True labels
Predicted labels

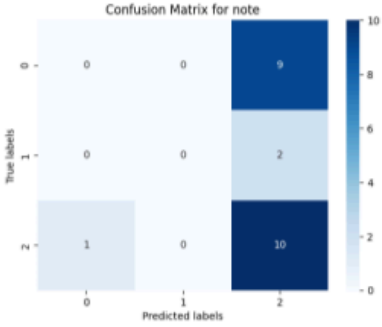

Confusion Matrix for difference
True labels
Predicted labels

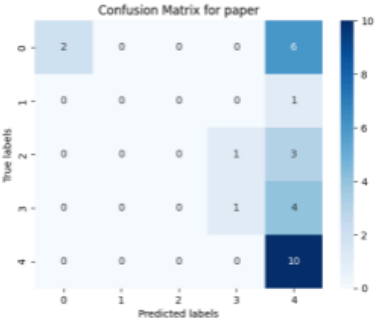

Confusion Matrix for plan
True labels
Predicted labels

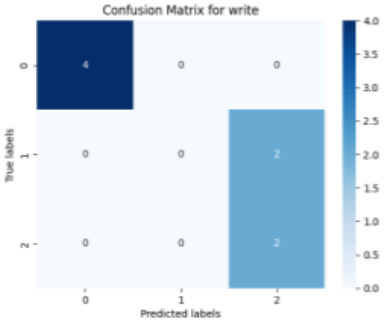

Confusion Matrix for eat
True labels
Predicted labels

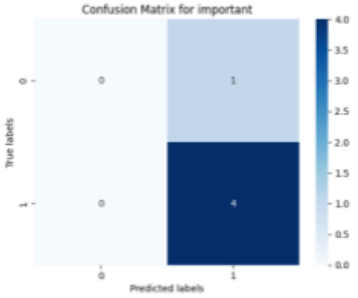

Confusion Matrix for express
True labels
Predicted labels

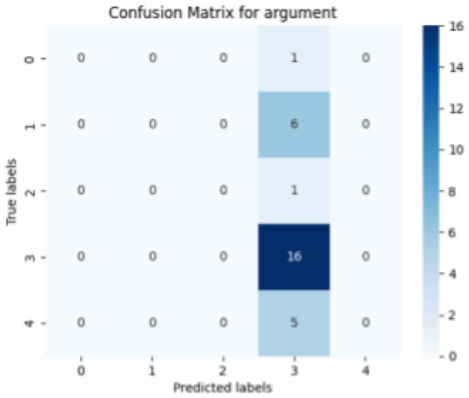

Confusion Matrix for source
True labels
Predicted labels

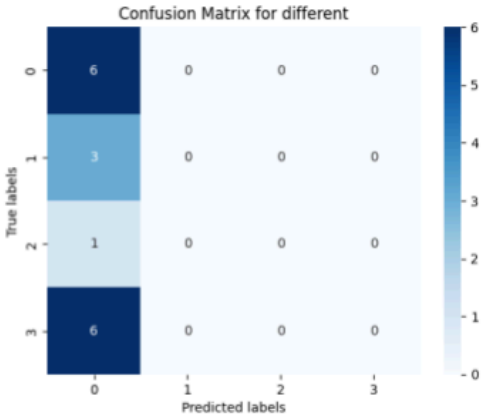

Confusion Matrix for note
True labels
Predicted labels


Confusion Matrix for paper
True labels
Predicted labels


Confusion Matrix for write
True labels
Predicted labels


Confusion Matrix for important
True labels
Predicted labels


Confusion Matrix for argument
True labels
Predicted labels
0 0 0 1 0
0 0 0 6 0
0 0 0 1 0
0 0 0 16 0
0 0 0 5 0
0 1 2 3 4
16 14 12 10 8 6 4 2 0


Confusion Matrix for different
True labels
Predicted labels
6 0 0 0
3 0 0 0
1 0 0 0
6 0 0 0
0 1 2 3
6 5 4 3 2 1 0