Carlos Orozco
Professor Maskharashvili
Linguistics 406

Assignment 3 Report

Baseline Table Results:

	Classifier	F1 Score	Precision	Recall
0	Naive_Bayes	0.358997	0.358997	0.358997
1	Decision_Tree	0.771593	0.771593	0.771593

Improved Table Results:

	Classifier	F1 Score	Precision	Recall
0	Naive_Bayes	0.380902	0.380902	0.380902
1	Decision_Tree	0.820714	0.820714	0.820714

Feature Engineering Table (With Improved System):

	Feature	F1 Difference	Precision Difference	Recall Difference	Classifier
0	Prev2	0.001708	0.001708	0.001708	Naive_Bayes
1	Prev1	0.001894	0.001894	0.001894	Naive_Bayes
2	Next1	0.008151	0.008151	0.008151	Naive_Bayes
3	Next2	0.000420	0.000420	0.000420	Naive_Bayes
4	Prev2	-0.014408	-0.014408	-0.014408	Decision_Tree
5	Prev1	-0.026005	-0.026005	-0.026005	Decision_Tree
6	Next1	-0.055613	-0.055613	-0.055613	Decision_Tree
7	Next2	-0.012861	-0.012861	-0.012861	Decision_Tree

Question 1:

Based on the results from both classifiers, in both the baseline and improved system, the Decision Tree model outperforms the Naïve Bayes model across all evaluation metrics. The precision measures how many of the predicted positive instances were positive. The decision tree achieves a higher precision in both the baseline and improved systems (0.7716 and 0.8207) compared to Naive Bayes (0.3590 and 0.3809). The recall evaluates how many of the actual positive instances were correctly identified. Like precision, the Decision Tree outperforms Naive Bayes by a large margin (0.7716 and 0.8207 vs. 0.3590 and 0.3809). For the F1 Score Decision Tree also significantly outperforms Naive Bayes on this metric (0.7716 and 0.8207 vs. 0.3590 and 0.3809).

Question 2:

For the improved model I decided to use a smaller context window and also consider the lemma of each word. For Naive Bayes model relies heavily on the frequency of words and their co-occurrences in the context window. Words that were seen in high frequency were most contributing to the performance. Contextual information or less frequent words did not contribute significantly to the performance. Since Naive Bayes assumes feature independence it does not capture the dependencies between the tokens' context or neighboring words. This likely hindered its performance. When it came to the Decision Tree model, specific token features, like the surrounding context, and their interactions likely had a major influence on the prediction. The Decision Tree can apply more complex rules, like combinations of nearby words or parts of speech, to capture more subtle patterns, which enhances accuracy. Features that were not common or redundant would have contributed less since they might not form useful splits in the tree.

Question 3:

The feature set with Naive Bayes, based on the context windows I used, worked reasonably well but did not provide a deep enough understanding of token dependencies. The simple feature set struggled to capture complex relationships between words and their POS Tags, where context and relationships between words are critical. The decision tree model benefited from the feature set, as it was able to partition the data based on the context windows. The tree could learn rules about how different combination of features predict POS Tags. In conclusion, the feature set was more effective for Decision Tree than for Naive Bayes, as the tree can capture more complex patterns and interactions between features.

Question 4:

After doing some research, I would consider using word embeddings to capture semantic relationships between words and more contextual information. This could help improve the decision tree model by providing more meaningful features, improving generalization and reducing sparsity found in the data. This would be much better than the bag-of-words approach that treats words as independent features.

I would also consider expanding the feature set to include N-Grams such as bigrams and trigrams instead of just individual words. This could help the model understand more complex patterns in token sequences. The model would understand syntactic structure more effectively, especially in a morphologically complex language like Polish