**Assignment2 Report**

**Implementation Details**

**Prior Calculation**

It is implemented in function **fit()** in **naive\_bayes.py** (line 30). It's done by counting the occurrences of each class in the training data and taking the logarithm of their probabilities.

**Likelihoods Calculation**

It is implemented in function **fit()** in **naive\_bayes.py** (line 43). The likelihood calculation is performed in the **fit** method. This calculation involves determining the frequency of each word in each class and applying Laplace smoothing.

**Feature Extraction**

It is implemented in function: **select\_words()** and **feature\_selection()** in **feature\_selection.py** (line 19). Feature extraction is done by selecting some useful lexical properties of the words. This is configurable and can be adjusted based on the text.

**Macro-F1 Score Calculation**

It is implemented in f unction: **macro\_f1\_score()** in **evaluation.py** (line 27).This function computes the F1 score for each class independently and then takes the average.

**Preprocessing Steps**

It is implemented in f unction: **preprocess()** in **utils.py**(line 25). This function converts the input text to lowercase first. Next, the function employs tokenization, breaking the text down into individual words or tokens. Following that, each token undergoes lemmatization. Subsequently, the function removes English stopwords and non-alphabetic characters (like punctuation), focusing on the most meaningful elements of the text. Finally, the preprocessed text is returned.

* 窗体顶端
* 窗体底端

**Feature selection**

My feature selection done based on the lexical properties of the words, including adjective, noun, verb, adverb, interrogative pronoun, etc. For example, adjectives are key, directly reflecting emotions with words like "wonderful" or "terrible." [1] However, relying solely on adjectives might overlook subtle sentiments expressed through other parts of text. Nouns, though not directly conveying sentiment, are crucial for identifying the main subjects and themes of a text. [2] Verbs and adverbs add depth by revealing actions and their intensities, contributing to a more comprehensive understanding of the text's emotional layers. Conjunctions can indicate the structure of arguments or the connection between sentiments. Numerals might be relevant in certain contexts like product reviews or factual descriptions. But based on the text and the context, this approach may be effective or ineffective. For instance, in highly contextual or idiomatic text, the chosen property might not capture all the subtleties.

I have tested it many times and noticed that only one or two cannot perform high scores, and for 3-class and 5-class, different lexical properties have different performance. Following are some outputs(f1 score) during the testing, which select different properties:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | adjective and noun | adjective, noun, verb, adverb and interrogative pronoun | adjective, noun, verb, adverb, conjunction and interrogative pronoun | adjective, noun, verb, adverb, numeral and interrogative pronoun |
| 3-class | 0.477344 | 0.517716 | 0.517737 | 0.522114 |
| 5-class | 0.285843 | 0.344783 | 0.345441 | 0.344738 |

This feature selection is implemented using NLTK, specifically the **pos\_tag** function. The flexibility to switch between 3-class and 5-class setting allows for tailored analysis, so people can select other lexical properties depending on the disparate text or setting. In 5-class I select adjective, noun, verb, adverb, interrogative pronoun and numerals. In 3-class, in addition to the words chosen earlier, I also choose foreign words, prepositions/subordinating conjunction, determiners, conjunction and particles.

**Results and Discussion**

|  |  |  |
| --- | --- | --- |
|  | all\_words | features |
| 3-class | 0.522169 | 0.525166 |
| 5-class | 0.350742 | 0.349075 |

图形用户界面, 应用程序

描述已自动生成图形用户界面, 应用程序

描述已自动生成The best 3-class model is the one using feature selection. The best 5-class model is the one without feature selection.

图表

描述已自动生成图表

描述已自动生成

Figure 1: 5-class all\_words

Figure 3: 3-class all\_words

Figure 2: 5-class features

Figure 4: 3-class features

In the context of movie reviews, nouns, verbs, adjectives, and adverbs offer significant clues about the film's content and the reviewer's emotions and perspectives. By extracting these parts as features, the model becomes more adept at capturing the essence and emotional nuances of the reviews. Additionally, the extraction of specific properties like interrogative pronouns, numerals, or interjections reveals particular sentiments, questions, or emphases in the reviews, aiding the model in gaining a deeper understanding of the text's semantic and emotional dimensions. The importance of various properties varies across different text. By selectively choosing features that align with the specific context, the model can be more effectively tailored to a given scenario.

Besides, I set the smoothing factor to 0.8 for 5-class and 0.5 for 3-class as it shows the best result.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 5-class all\_words | 5-class features | 3-class all\_words | 3-class features |
| 0.2 | 0.330022 | 0.326413 | 0.517815 | 0.513894 |
| 0.5 | 0.336403 | 0.336737 | 0.522169 | 0.525166 |
| 0.8 | 0.350742 | 0.349075 | 0.513071 | 0.522114 |

**Error analysis**

In the *3-class* setting, model using feature selection performs slightly better.

1. SentenceId: 4967 1 (neutral) 🡪 0 (negative)

- Original text: May be the most undeserving victim of critical overkill since Town and Country .

- Feature Selection: ['undeserving', 'victim', 'critical', 'overkill', 'town', 'country']

- Issue: The model failed to understand the overall neutral context of the statement, focusing instead on negative words like "victim" and "overkill".

2. SentenceId: 1207 1 (neutral) 🡪 0 (negative)

- Original Text: Absurdities and   
- Feature Selection: ['absurdity']

- Issue: The model incorrectly interpreted a text fragment with a single word "Absurdities" as negative, likely due to the lack of context.

3. SentenceId: 1243 0 (negative) 🡪 2 (positive)

- Original Text: Boasts eye-catching art direction but has a forcefully quirky tone that quickly wears out its limited welcome .   
- Feature Selection: ['boast', 'eyecatching', 'art', 'direction', 'ha', 'forcefully', 'quirky', 'tone', 'quickly', 'wear', 'limited', 'welcome']

- Issue: The model was misled by the initial positive phrase and failed to appropriately weigh the negative sentiment in the latter part of the sentence.

4. SentenceId: 8008 0 (negative) 🡪 1 (neutral)

- Original Text: There 's a little violence and lots of sex in a bid to hold our attention , but it grows monotonous after a while , as do Joan and Philip 's repetitive arguments , schemes and treachery . - Feature Selection: ['little', 'violence', 'lot', 'sex', 'bid', 'hold', 'attention', 'grows', 'monotonous', 'joan', 'philip', 'repetitive', 'argument', 'scheme', 'treachery']

- Issue: Complex sentences with a mix of sentiments or subtler negative cues seem to be very challenging.

In the *5-class* setting, model without feature selection performs slightly better.

1. SentenceId: 6235 2 (neutral) 🡪 1 (somewhat negative)

- Original text: The film is explosive , but a few of those sticks are wet .

- Feature Selection: ['film', 'explosive', 'stick', 'wet']

- Issue: The phrase contains opposing affective expressions ("explore" and "wet") resulting in a predictive sentiment that deviated from neutrality.

2. SentenceId: 4331 0 (negative) 🡪 1 (somewhat negative)

- Original text: All the Queen 's Men is a throwback war movie that fails on so many levels , it should pay reparations to viewers .

- Feature Selection: ['queen', 'men', 'throwback', 'war', 'movie', 'fails', 'many', 'level', 'pay', 'reparation', 'viewer']

- Issue: It may not fully understand the strong negative implications of the words "fail" and "reparations".

3. SentenceId: 2333 3 (somewhat positive) 🡪 0 (negative)

- Original text: Even at its worst , it 's not half-bad .

- Feature Selection: ['even', 'worst', 'halfbad']

- Issue: The absence of "not" in the phrase makes the word all negative.

4. SentenceId: 3602 4 (positive) 🡪 0 (negative)

- Original text: The overall effect is awe and affection -- and a strange urge to get on a board and , uh , shred , dude .

- Feature Selection: ['overall', 'effect', 'awe', 'affection', 'strange', 'urge', 'get', 'board', 'uh', 'shred', 'dude']

- Issue: The model failed to recognize positive emotions in words containing positive meanings such as 'awe' and 'affection'.

In general, it is necessary to enhance the model's ability to understand context, handle text fragments, balance mixed sentiments, and interpret neutral language more accurate. Also, some 5-class examples point to deficiencies in the model's ability to understand complex emotions, recognize essential words, and deal with emotional context. By enhancing training and understanding of these aspects, the accuracy and adaptability of the model should be significantly improved. However, achieving these improvements often requires a varied and well-structured training dataset that includes a wide range of linguistic nuances and complexities.

**Reference**

[1] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends® in Information Retrieval*, vol. 2, nos. 1–2, pp. 1-135, 2008.

[2] G. Weikum, "Foundations of statistical natural language processing," ACM SIGMOD Record, vol. 31, no. 3, pp. 37-38, 2002.