Sentiment Analysis of Movie Reviews
Supplementary materials

Yuna Lee

1. Results of Baseline model.

Our baseline consists of using TF-IDF representation combined with the NB classifier. This approach incorporates a commonly adopted preprocessing technique where stop words are removed, and lemmatisation is applied. Figure 1 and Table 1 show the confusion matrix and the performance metrics of this model. We also note that hyperparameters of this model were optimised through 10-fold cross validation with grid search, resulting in a smoothing parameter 'alpha' of 1 and 'fit_prior' set as True.

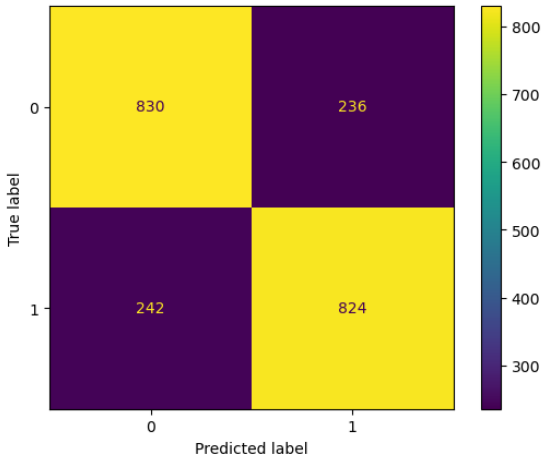

Figure 1 – Confusion matrix of the baseline model

Metrics	Baseline model
Accuracy	0.775797
Precision	0.775806
Recall	0.775797
F1-Score	0.775796

Table 1 – Performance metrics of baseline model.

2. Preprocessed text examples after performing Next Word Negation

For our first approach, we experimented with NWN. We display an example of how the NWN processes text, to help understand our findings.

Without NWN:

- **Raw text:**

"the uneven movie does have its charms and its funny moments but not quite enough of them."

- **Basic preprocessing**

"the uneven movie does have its charms and its funny moments but not quite enough of them"

- **Stop word removal**

"uneven movie charms funny moments quite enough"

- **Lemmatisation**

"uneven movie charm funny moment quite enough".

With NWN:

- **Basic preprocessing**

"the uneven movie does have its charms and its funny moments but not_quite enough of them"

- **Stop word removal**

"uneven movie charms funny moments not_quite enough"

- **Lemmatisation**

"uneven movie charm funny moment not_quite enough".

In this example, we observe that using Negation Word Marking (NWM), the negation word "not" is retained in the text. However, without NWM, "not" would be treated as a part of English stop words and removed during stop word removal. This would hinder the model's ability to identify negation in the text accurately.

The table below, which were also presented in the report confirms this observation: when NWM is implemented, performance improves significantly after stop words are removed. Conversely, without NWM, performance tends to decrease after stop word removal.

Model	Basic Preprocessing	Stop word removal	Lemmatisation
MNB	0.7814	0.7805	0.7758
MNB (with NMN)	0.7810	**0.7852**	0.7800
SVM	0.7697	0.7683	0.7744
SVM (with NMN)	0.7739	0.7725	0.7763
LR	0.7674	0.7613	0.7589
LR (with NMN)	0.7692	0.7678	0.7683

Table 2 – Accuracy scores for MNB, SVM, and LR models with and without NWN, with variations in pre-processing.

3. **Logistic Regression results with different combinations of N-grams and various pre-processing techniques.**

LR	Basic Preprocessing	Stop word removal	Lemmatisation
Unigram	0.7674	0.7598	0.7598
Bigram	0.7008	0.5994	0.6083
Uni + Bigram	**0.7683**	0.7655	0.7627
Uni + Bi + Trigram	0.7627	0.7641	0.7608

Table 3 – Accuracy scores for LR model with different combinations of N-grams and various pre-processing techniques.

Our analysis began with LR due to its computational efficiency, where we examined the accuracy scores of various N-gram combinations with different pre-processing methods. We see from Table 3 that the models consistently favoured Unigram + Bigram across all pre-processing methods, with basic pre-processing proving more effective than additional processing. Interestingly, Bigram extraction showed poor performance (0.6), particularly with added pre-processing steps. This indicates that the combination of Bigram extraction and additional pre-processing may not effectively capture sentiment in this dataset or task.