Assignment 2 Report

1. (3 marks) Develop Decision Tree models for training and testing: (a) with the 1% stopping criterion (the standard model), and (b) without the 1% stopping criterion.
(i) Show all metrics on the test set for scenario 1 comparing the two models (a) and (b), and explain any similarities and differences.

Evaluation	Accuracy	Micro-precision	Macro-precision	Micro-recall	Macro-recall	Micro-F1	Macro-F1
With 1%	0.638	0.638	0.302	0.638	0.302	0.254	0.144
Without 1%	0.534	0.534	0.256	0.534	0.237	0.144	0.122

When the stop criterion is used, the effect of the model is better. When the stop criterion is not used, the effect will be worse. Its function is to prune the decision tree, prevent the decision tree from overfitting, and improve the performance of the model.
(ii) Show all metrics on the test set for scenario 2 comparing the two models (a) and (b), and explain any similarities and differences.

Evaluation	Accuracy	Micro-precision	Macro-precision	Micro-recall	Macro-recall	Micro-F1	Macro-F1
With 1%	0.658	0.658	0.477	0.658	0.412	0.643	0.314
Without 1%	0.534	0.534	0.363	0.534	0.356	0.535	0.597

The situation in question 1 also occurs in question 2. It can be seen that even if the number of classifications is reduced, the overfitting of the decision tree will still occur.
(iii) Explain any differences in the results between scenarios 1 and 2.
The accuracy rates in Scenario 1 and Scenario 2 are the same, but the macro indicators are different. The macro indicators in Scenario 2 are better, and the number of classifications can improve the macro performance.

2. (3 marks) Develop BNB and MNB models from the training set using: (a) the whole vocabulary (standard models), and (b) the most frequent 1000 words from the vocabulary, as defined using scikit-learn Count Vectorizer, after preprocessing by removing "junk" characters.
(i) Show all metrics on the test set for scenario 1 comparing the corresponding models (a) and (b), and explain any similarities and differences.

model	vocabulary	Accuracy	Micro-precision	Macro-precision	Micro-recall	Macro-recall	Micro-F1	Macro-F1
BNB	1000	0.607	0.607	0.301	0.607	0.3	0.257	0.164
	whole	0.521	0.521	0.213	0.521	0.221	0.134	0.164
MNB	1000	0.586	0.586	0.364	0.586	0.326	0.246	0.133
	whole	0.528	0.528	0.264	0.528	0.219	0.167	0.294

For both the MNB and BNB models, using the first 1000 words performed better than using the full

word model. If all words are used, there will be a lot of invalid information that interferes with the judgment of the model and reduces the performance of the model.

(ii) Show all metrics on the test set for scenario 2 comparing the corresponding models (a) and (b), and explain any similarities and differences.

model	vocabulary	Accuracy	Micro-precision	Macro-precision	Micro-recall	Macro-recall	Micro-F1	Macro-F1
BNB	1000	0.669	0.669	0.346	0.669	0.345	0.246	0.142
	whole	0.548	0.548	0.253	0.548	0.246	0.127	0.153
MNB	1000	0.682	0.682	0.482	0.682	0.432	0.623	0.297
	whole	0.529	0.529	0.26	0.529	0.21	0.16	0.29

The situation of Scenario 2 is generally the same as that of Scenario 1, and the macro indicators are improved.

(iii) Explain any differences in the results between scenarios 1 and 2.

For BNB and MNB models, it is better to use the top 1000 most frequent words to represent the features of the samples than to use all the samples to represent the features of the model, regardless of the number of categories.

3. (3 marks) Evaluate the effect of preprocessing for the three standard models by comparing models developed with: (a) only the preprocessing described above (standard models), and (b) applying, in addition, Porter stemming using NLTK then English stop word removal using scikit-learn Count Vectorizer.

(i) Show all metrics on the test set for scenario 1 comparing the corresponding models (a) and (b), and explain any similarities and differences.

model	NLTK	Accuracy	Micro-precision	Macro-precision	Micro-recall	Macro-recall	Micro-F1	Macro-F1
DT	not	0.569	0.569	0.346	0.569	0.345	0.246	0.142
	use	0.631	0.631	0.248	0.631	0.248	0.12	0.145
BNB	not	0.582	0.582	0.482	0.582	0.432	0.623	0.297
	use	0.616	0.616	0.254	0.616	0.282	0.153	0.263
MNB	not	0.582	0.582	0.482	0.582	0.432	0.623	0.297
	use	0.612	0.612	0.266	0.612	0.211	0.166	0.29

After using NLTK, the effect of each model has been improved. NLTK can remove useless words in the model and enhance the expressive ability of features.

(ii) Show all metrics on the test set for scenario 2 comparing the corresponding models (a) and (b), and explain any similarities and differences.

model	NLTK	Accuracy	Micro-precision	Macro-precision	Micro-recall	Macro-recall	Micro-F1	Macro-F1
DT	not	0.592	0.592	0.356	0.592	0.342	0.177	0.169
	use	0.612	0.612	0.354	0.612	0.259	0.127	0.153
BNB	not	0.573	0.573	0.462	0.573	0.441	0.573	0.316
	use	0.623	0.623	0.243	0.623	0.282	0.146	0.255

MNB	not	0.563	0.563	0.492	0.563	0.492	0.563	0.222
	use	0.606	0.606	0.267	0.606	0.217	0.16	0.299

Even if the number of classes is reduced and the samples are balanced, the effect of using NLTK is still better than that without NLTK.

(iii) Explain any differences in the results between scenarios 1 and 2.

Whether it is scene 1 or scene 2, the effect of using NLTK is better than that without NLTK.

4. (3 marks) Evaluate the effect of converting all letters to lower case for the three standard models by comparing models with: (a) no conversion to lower case, and (b) all input text converted to lower case.

(i) Show all metrics on the test set for scenario 1 comparing the corresponding models (a) and (b), and explain any similarities and differences.

model	lower case	Accuracy	Micro-precision	Macro-precision	Micro-recall	Macro-recall	Micro-F1	Macro-F1
DT	not	0.618	0.618	0.346	0.618	0.323	0.18	0.153
	use	0.603	0.603	0.316	0.603	0.277	0.138	0.161
BNB	not	0.566	0.566	0.467	0.566	0.441	0.213	0.336
	use	0.503	0.503	0.316	0.503	0.272	0.166	0.235
MNB	not	0.563	0.563	0.492	0.563	0.492	0.563	0.222
	use	0.546	0.546	0.267	0.546	0.217	0.16	0.299

For these three models, after all words are converted to lowercase, the effect of the model will decrease, which proves that the capitalization of letters is also an important part of the model features.

(ii) Show all metrics on the test set for scenario 2 comparing the corresponding models (a) and (b), and explain any similarities and differences.

model	lower case	Accuracy	Micro-precision	Macro-precision	Micro-recall	Macro-recall	Micro-F1	Macro-F1
DT	not	0.618	0.618	0.346	0.618	0.323	0.18	0.153
	use	0.593	0.593	0.306	0.593	0.287	0.182	0.161
BNB	not	0.566	0.566	0.467	0.566	0.441	0.213	0.336
	use	0.513	0.512	0.366	0.512	0.272	0.166	0.235
MNB	not	0.563	0.563	0.492	0.563	0.492	0.563	0.222
	use	0.551	0.551	0.263	0.551	0.147	0.166	0.303

In Scenario 2, when all letters are converted to lowercase, the performance of the three models will also be degraded to varying degrees.

(iii) Explain any differences in the results between scenarios 1 and 2.

Whether it is scene 1 or scene 2, after converting all words to lowercase, the model will lose some information, resulting in a decrease in the effect of the model.

5. (5 marks) Describe your chosen “best” method for rating prediction. Give new experimental results for your method trained on the training set of 2000 reviews and tested on the test set of 500

reviews. Explain how this experimental evaluation justifies your choice of model, including settings and parameters, against a range of alternatives. Provide new experiments and justifications: do not just refer to previous answers.

scenario	Accuracy	Micro-precision	Macro-precision	Micro-recall	Macro-recall	Micro-F1	Macro-F1
1	0.683	0.683	0.316	0.683	0.348	0.246	0.216
2	0.684	0.684	0.416	0.684	0.654	0.222	0.246

Since the Linear SVC model has a good classification effect, I chose the Linear SVC model, and the effect is shown in the table above.