1. Develop Decision Tree models for training and testing: (a) with the 1% stopping criterion (the standard model), and (b) without the 1% stopping criterion.

Stop	With 1%		Without 1%	
scenario	1	2	1	2
accuracy	0.625	0.641	0.541	0.582
Micro- precision	0.625	0.641	0.541	0.582
Macro-precision	0.3	0.471	0.223	0.449
Micro-recall	0.625	0.641	0.541	0.582
Macro-recall	0.25	0.435	0.235	0.445
Micro-F1	0.12	0.641	0.114	0.582
Macro-F1	0.142	0.325	0.121	0.336

Table1 Decision Tree models stopping criterion

It can be seen from the above table that the model prediction result of scenario 2 is better than that of scenario 1, and with the 1% stopping criterion is better than without the 1% stopping criterion. It can be seen that the model has a better effect on tasks with few classifications, and A can effectively prevent the model from overfitting and improve the classification performance of the model.

2. 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.

model	BNB				MNB			
vocabulary	whole		1000		whole		1000	
scenario	1	2	1	2	1	2	1	2
accuracy	0.628	0.633	0.583	0.628	0.647	0.673	0.605	0.663
Micro- precision	0.628	0.633	0.583	0.628	0.647	0.673	0.605	0.663
Macro-precision	0.305	0.399	0.349	0.484	0.234	0.479	0.232	0.546
Micro-recall	0.628	0.633	0.583	0.628	0.647	0.673	0.605	0.663
Macro-recall	0.246	0.363	0.315	0.466	0.235	0.389	0.221	0.535
Micro-F1	0.116	0.633	0.184	0.628	0.113	0.673	0.605	0.663
Macro-F1	0.147	0.258	0.205	0.357	0.108	0.274	0.231	0.405

As can be seen from the above table, using all vocabulary to train the model is better than the model prediction effect obtained by training the model with the top 1000 vocabulary words. This proves that using all models to train the model does not overfit the model, but enhances the expressive ability of the model.

3. 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.

model	DT				BNB			
NLTK	using		not		using		not	
scenario	1	2	1	2	1	2	1	2
accuracy	0.625	0.634	0.637	0.644	0.583	0.649	0.625	0.63
Micro- precision	0.625	0.634	0.637	0.644	0.583	0.649	0.625	0.63

Macro-precision	0.293	0.458	0.637	0.492	0.335	0.482	0.226	0.561
Micro-recall	0.625	0.634	0.569	0.632	0.595	0.649	0.625	0.63
Macro-recall	0.257	0.418	0.324	0.446	0.287	0.205	0.256	0.525
Micro-F1	0.148	0.636	0.198	0.636	0.164	0.203	0.625	0.63
Macro-F1	0.126	0.315	0.216	0.346	0.178	0.274	0.234	0.415

model	DT				MNB			
NLTK	using		not		using		not	
scenario	1	2	1	2	1	2	1	2
accuracy	0.625	0.634	0.637	0.644	0.586	0.645	0.602	0.662
Micro- precision	0.625	0.634	0.637	0.644	0.586	0.645	0.602	0.662
Macro-precision	0.293	0.458	0.637	0.492	0.337	0.483	0.354	0.542
Micro-recall	0.625	0.634	0.569	0.632	0.595	0.645	0.602	0.662
Macro-recall	0.257	0.418	0.324	0.446	0.282	0.207	0.339	0.531
Micro-F1	0.148	0.636	0.198	0.636	0.164	0.205	0.261	0.662
Macro-F1	0.126	0.315	0.216	0.346	0.179	0.272	0.232	0.4

As can be seen from the above table, after using NLTK to remove some words, the prediction effect of the model does not increase but decreases. It shows that the words removed by NLTK can also represent the characteristics of the sample well. If removed, it will affect the accuracy of the model.

4. 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.

model	DT				BNB			
lower case	not		convert		not		convert	
scenario	1	2	1	2	1	2	1	2
accuracy	0.638	0.636	0.624	0.604	0.627	0.646	0.589	0.623
Micro- precision	0.638	0.636	0.624	0.604	0.627	0.646	0.589	0.623
Macro-precision	0.294	0.457	0.273	0.413	0.281	0.467	0.379	0.484
Micro-recall	0.628	0.636	0.624	0.604	0.627	0.646	0.585	0.623
Macro-recall	0.251	0.415	0.23	0.392	0.24	0.425	0.329	0.447
Micro-F1	0.144	0.636	0.102	0.604	0.155	0.646	0.236	0.623
Macro-F1	0.126	0.31	0.08	0.292	0.121	0.313	0.234	0.354

model	DT				MNB			
lower case	using		not		using		not	
scenario	1	2	1	2	1	2	1	2
accuracy	0.638	0.636	0.624	0.604	0.622	0.652	0.616	0.683
Micro- precision	0.638	0.636	0.624	0.604	0.622	0.652	0.616	0.6883
Macro-precision	0.294	0.457	0.273	0.413	0.293	0.415	0.345	0.554
Micro-recall	0.628	0.636	0.624	0.604	0.622	0.652	0.616	0.683
Macro-recall	0.251	0.415	0.23	0.392	0.258	0.426	0.356	0.554
Micro-F1	0.144	0.636	0.102	0.604	0.144	0.652	0.245	0.683
Macro-F1	0.126	0.31	0.08	0.292	0.165	0.356	0.254	0.448

It can be seen from the above table that after using lowercase, the predictive ability of the model decreases. In an English sentence, the first letter is generally capitalized, but after all of them are converted to lowercase, the model cannot learn the word position of the first letter of the sentence, and part of the position information is lost, which makes the model fit poorly.

5. 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.

model	Linear SVC	
scenario	1	2
accuracy	0.665	0.547
Micro- precision	0.665	0.547
Macro-precision	0.323	0.224
Micro-recall	0.665	0.546
Macro-recall	0.307	0.237
Micro-F1	0.221	0.114
Macro-F1	0.215	0.124

Comparing the performance of this model with the data in the previous data table, it can be found that the effect of this model is better than that of the previous model. It can be seen that the effect of SVC in dealing with classification problems is still very strong.