Day 70

DIY

Q1. Problem Statement: Sentiment Analysis

Write a Python program that reads the *mood_data.txt* (provided on LMS) file into a DataFrame. Then the following the given tasks, that has to be taken into consideration while constructing the solution.

Here dataset contains two columns where one is our target ("emotion" has 6 different categories) and another is the independent variable ("Text" contains data in form of sentences).

1. Load the *mobile mood_data.txt* data into a DataFrame
2. Generate tokens and remove punctuations, stop words and lower all rows
3. Join all the tokens as they were before and store them in a new column named "cleaned_text"
4. Now remove all single characters, extra space, and special characters and store processed data in a new column named "processed_text"
5. Create a final DataFrame containing dependent variable(emotion) and processed text
6. Extract independent variables (Xs) and dependent variables (Ys) into separate data objects
7. Generate tokens and do vectorization
8. Build a model with Multinomial Naive Bayes, Random Forest, Random Forest (Entropy), SVM and compare their accuracy

Dataset:

	Text	Emotion
0	i didnt feel humiliated	sadness
1	i can go from feeling so hopeless to so damned...	sadness
2	im grabbing a minute to post i feel greedy wrong	anger
3	i am ever feeling nostalgic about the fireplac...	love
4	i am feeling grouchy	anger

Sample Output:

3. Join all the tokens as they were before and store them in a new column named "cleaned_text"

	Text	Emotion	cleaned_text
0	i didnt feel humiliated	sadness	i didnt feel humiliated
1	i can go from feeling so hopeless to so damned...	sadness	i can go from feeling so hopeless to so damned...
2	im grabbing a minute to post i feel greedy wrong	anger	im grabbing a minute to post i feel greedy wrong
3	i am ever feeling nostalgic about the fireplac...	love	i am ever feeling nostalgic about the fireplac...
4	i am feeling grouchy	anger	i am feeling grouchy
...	...	...	...
15995	i just had a very brief time in the beanbag an...	sadness	i just had a very brief time in the beanbag an...
15996	i am now turning and i feel pathetic that i am...	sadness	i am now turning and i feel pathetic that i am...
15997	i feel strong and good overall	joy	i feel strong and good overall
15998	i feel like this was such a rude comment and i...	anger	i feel like this was such a rude comment and i...
15999	i know a lot but i feel so stupid because i ca...	sadness	i know a lot but i feel so stupid because i ca...

4. Now remove all single characters, extra space, and special characters and store processed data in a new column named “processed_text”

	Text	Emotion	cleaned_text	processed_text
0	i didnt feel humiliated	sadness	i didnt feel humiliated	i didnt feel humiliated
1	i can go from feeling so hopeless to so damned...	sadness	i can go from feeling so hopeless to so damned...	i can go from feeling so hopeless to so damned...
2	im grabbing a minute to post i feel greedy wrong	anger	im grabbing a minute to post i feel greedy wrong	im grabbing minute to post feel greedy wrong
3	i am ever feeling nostalgic about the fireplac...	love	i am ever feeling nostalgic about the fireplac...	i am ever feeling nostalgic about the fireplac...
4	i am feeling grouchy	anger	i am feeling grouchy	i am feeling grouchy
...	...	...	...	...
15995	i just had a very brief time in the beanbag an...	sadness	i just had a very brief time in the beanbag an...	i just had very brief time in the beanbag and ...
15996	i am now turning and i feel pathetic that i am...	sadness	i am now turning and i feel pathetic that i am...	i am now turning and feel pathetic that am sti...
15997	i feel strong and good overall	joy	i feel strong and good overall	i feel strong and good overall
15998	i feel like this was such a rude comment and i...	anger	i feel like this was such a rude comment and i...	i feel like this was such rude comment and im ...
15999	i know a lot but i feel so stupid because i ca...	sadness	i know a lot but i feel so stupid because i ca...	i know lot but feel so stupid because can not ...

8. Build a model with Multinomial Naive Bayes, Random Forest, Random Forest (Entropy), SVM and compare their accuracy

	Model	Accuracy
0	Multinomial Naive Bayes	0.740625
1	Random Forest(Gini)	0.828958
2	Random Forest(Entropy)	0.811458
3	SVC by SVM	0.810833