Task 3:
Machine Learning Model:

I concatenated the stories files into one dataframe, then I started to train different Machine Learning Algorithms to see how they will Perform on the data.

I used the following algorithms and the accuracies was:

1- For Naive Bayes: 81%

2- For Support Vector Machine: 84%

3- For Artificial Neural Network: 87%

what is the meaning of each accuracy metrics?

Precision:
Precision is a metric that measures how well the model predicts favorable outcomes. It is the proportion of accurate positive predictions to all instances of positive predictions in a given class. When a model predicts a positive class, it typically does so correctly because it has a low false positive rate or precision.

Recall:
The model's ability to recognize every positive case of a class is measured by recall, also known as sensitivity or true positive rate. It is the proportion of accurate forecasts to all actual instances of positive outcomes in that class. Recall is a measure of how well a model finds positive examples.

F1-score:
The harmonic mean of recall and precision is known as the F1-score. It offers a balanced measurement that takes into account both recall and precision. When precision and recall must be given equal weight, the F1-score is helpful. Its values vary from 0 to 1, with 1 representing the ideal compromise between recall and precision.

Accuracy:
Accuracy is a metric that measures how accurate the model's predictions are on the entire system. It is the proportion of accurate forecasts to all the predictions the model made. The model's accuracy gives a broad picture of how well it performs throughout the full dataset.

Some enhancements:
Try different NLP methods data preprocessing technique like one hot encoding. The reason one-hot encoding has become popular in machine learning algorithms is because it helps to avoid any ordinal association between categories and avoids the model from giving the encoded categories any unintentionally ordinal relevance. For algorithms like linear regression, support vector machines, and neural networks that do not directly deal with categorical data, it is very helpful.