SENTIMENT PREDICTOR FOR BRAND OWNERS.

Background

- A brand name or identity is now considered among the most valuable asset for a going concern. According to Wikipedia , A brand is a name, term, design, symbol or any other feature that distinguishes one seller's good or service from those of other sellers. Brands are used in business, marketing, and advertising for recognition and, importantly, to create and store value as brand equity for the object identified, to the benefit of the brand's customers, its owners and shareholders.
- Consequently, in is important to business owners or brand owners to know and understand how their current and potential perceive heir brand. This information would feed into their strategy on how to enhance, protect, course correct, where applicable on the status of their brand.

Problem Statement

- Brand perception by current and would customer is key to a business success. A negative brand perception, especially, in this current age of social media and interconnectedness, can quickly wipe out the value of a company within a short time. And conversely, a positive perception can quickly add value to a company. Therefore, being able to gauge how people feel and perceive about one's brand is a great asset.

Objective

- To develop sentiment prdiction model based upon Natural language multiclassification with features as customer reviews and social media views.

Data Understanding and Strategy.

- Our data is a CSV file of 9,093 records and 3 columns. The columns are tweet_texts, emotion_in_tweet_is_directed_at and is_there_an_emotion_directed_at_a brand_or_product. Essentially, the columns represent tweets of "customers" sentiments towards certain brands and/or products. We will apply data preparation techniques on out sentiment tweets in order to extract our features for modelling.

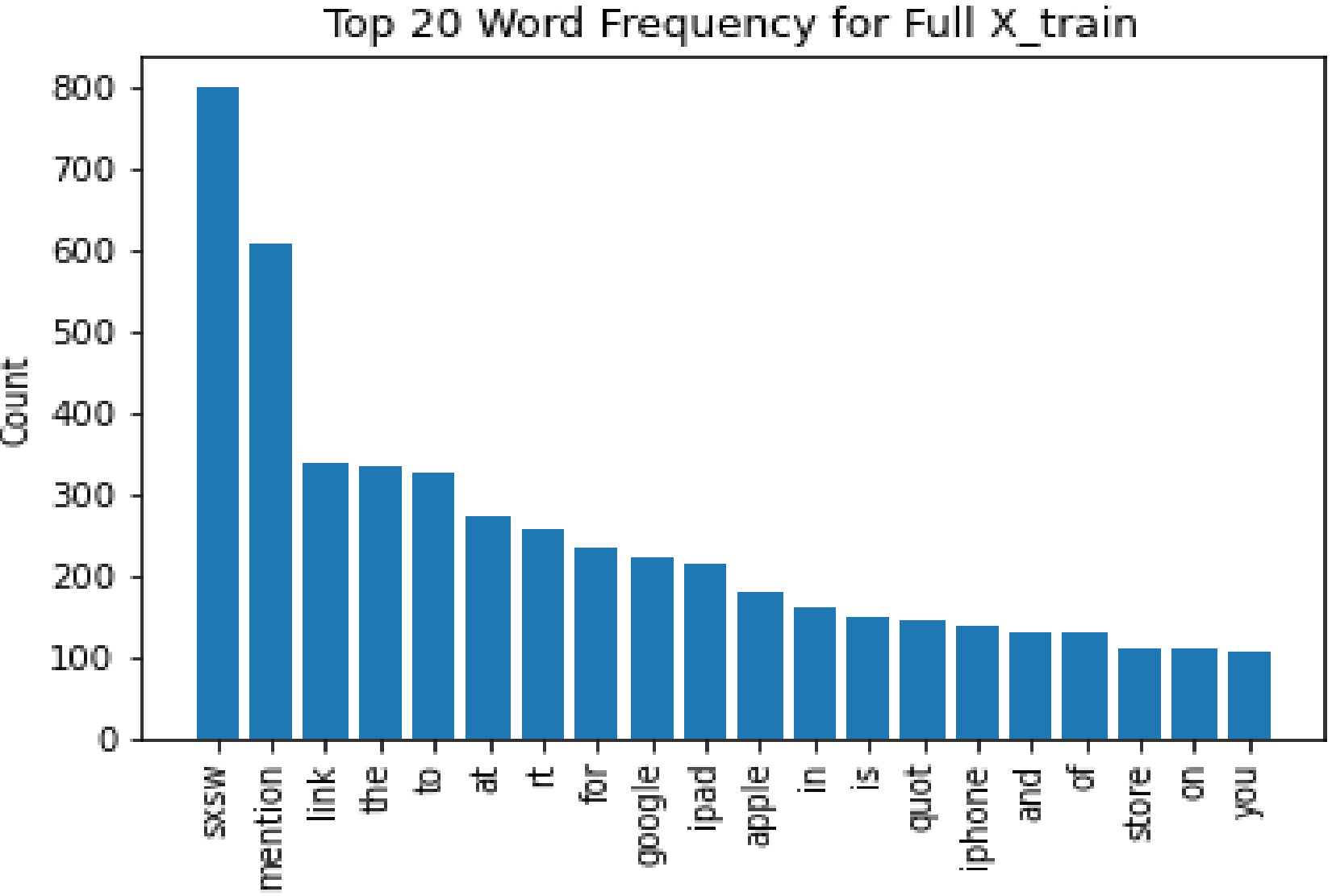
DATA PREPARATION.

Data Preparation Steps.

- The foolwoing are the data preparation steps we under took :

1. Assigning integers to the four sentiments under the TARGET column.
2. Sseparation of the data into test and train data sets using test_train_split.
3. Sstandardizing text by applying lower case function to the text from which we intend to extract our features.
4. tokenization of text by applying the tokenize library. This splits our text into individual tokens. This enables vectorization and modelling.
5. Elimination of stop words.

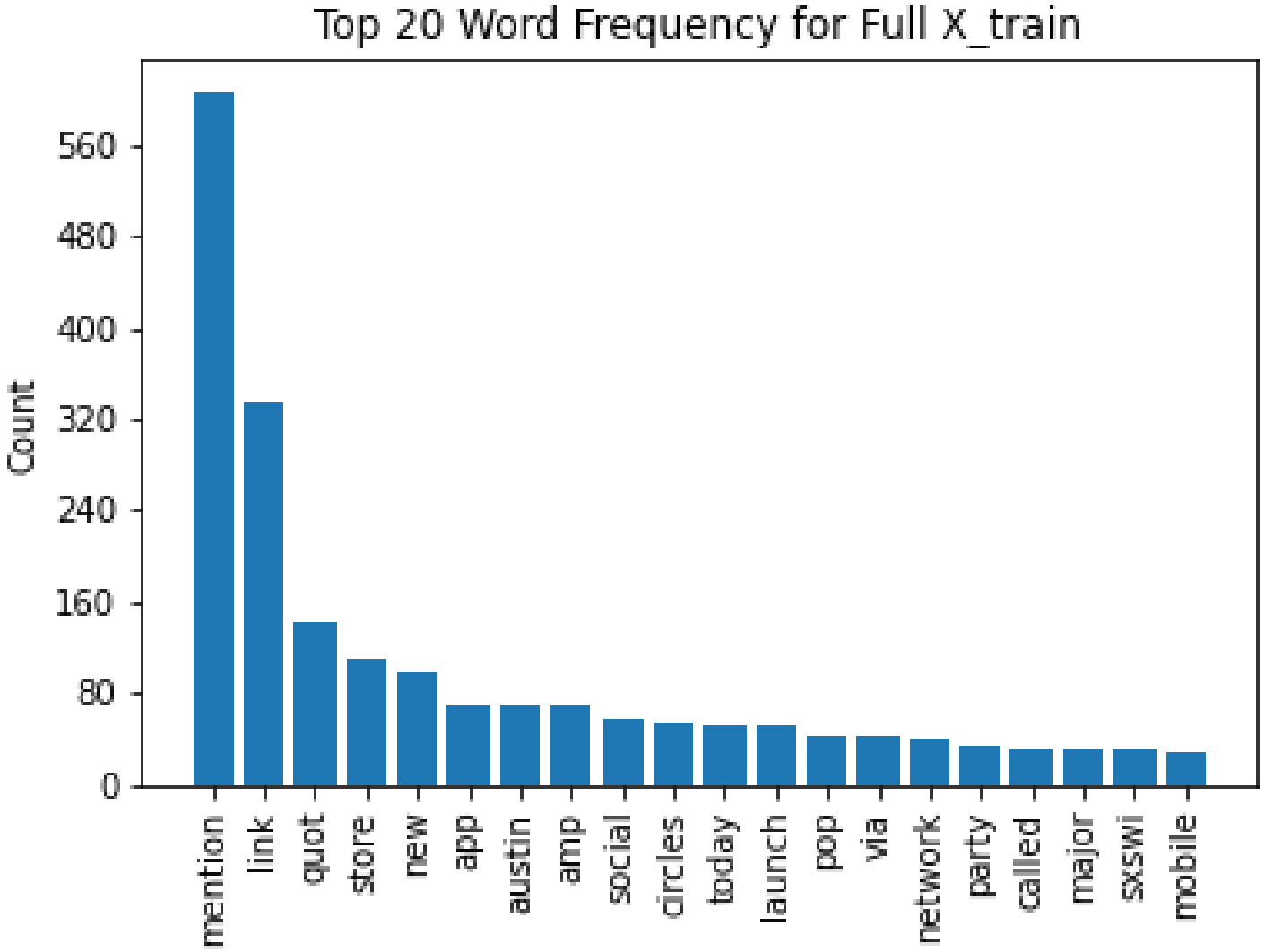
EDA BEFORE ELIMINATING STOP WORDS.



-Word frequency after tokenization and before elimination of stop words use in the base model.

This resulted in a model score of 61%.

EDA AFTER ELIMINATION OF STOP WORDS.



- Word frequency after elimination of stop words and used in the best model. This resulted in a model score of 64.8%

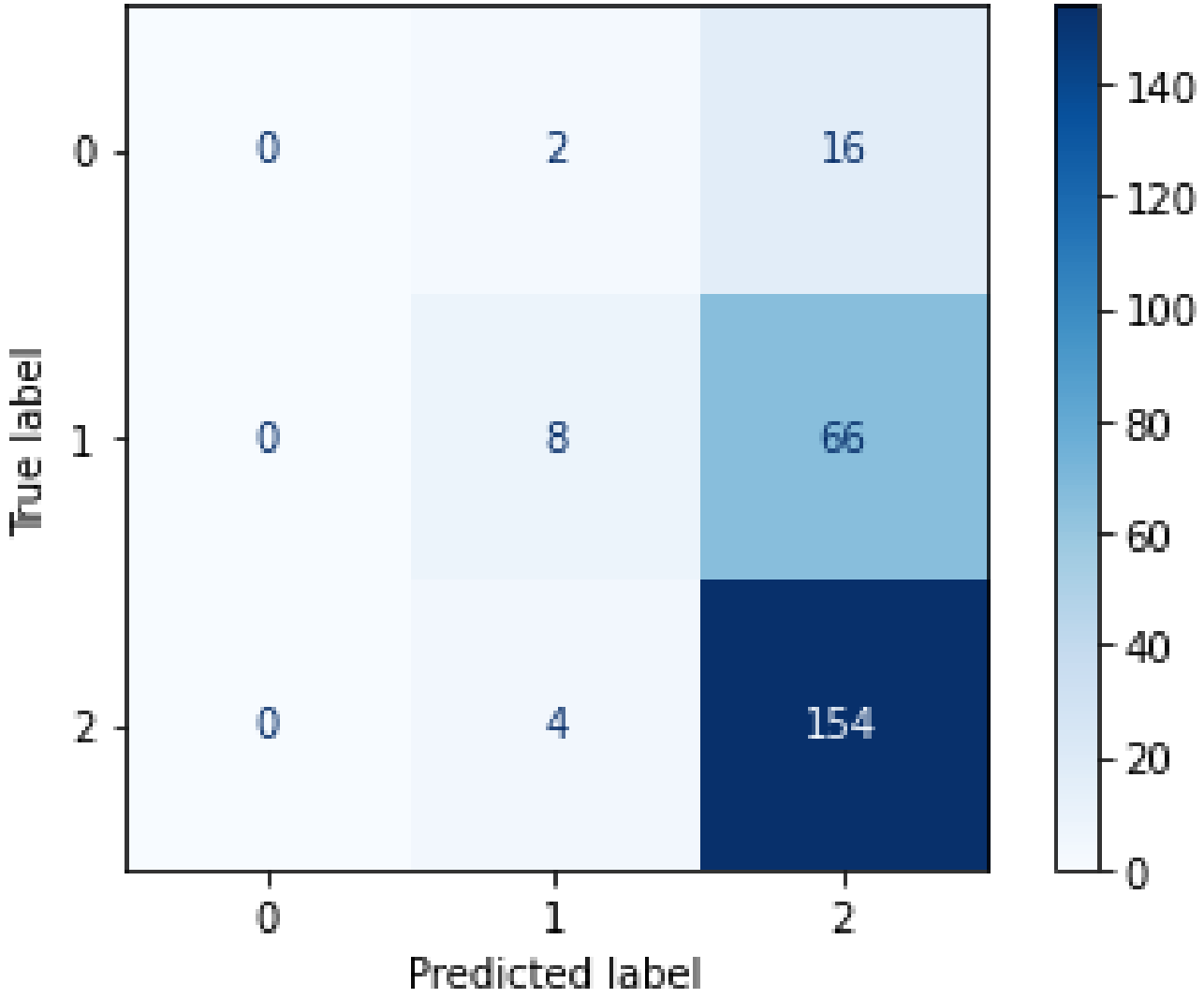
MODELS UTILISED.

We chose and used the following models:

- Multinomial Naive Bayes model because the model is suited for text classification problems and easy to use. In addition, it is easy to implement, efficient with large data sets, has low computational cost and works for both binary and multiclass classifications.

- The TfidVectorizer as our vectorizer because it enables us not to rely the raw frequencies of word occurrences through scaling down the impact of token that occur more frequently. Emphasis is placed on the importance of tokens.

MODEL RESULTS.



- The models has an accuracy score of 64.8%
- From the confusion matrix it is evident the model did not classify two labels quite well. The further steps would be to look into why these labels were not well classified by the model.

Recommendation.

1. The model to be considered to predict customer churn would be the Pruned decision tree which has a minimal difference between its train and test accuracy scores. Considering the confusion marix the model has correctly classified sentiment 2 (no emotion towards brand or product), has not correctly classified sentiment 0 (negative emotion)and sentiment 1 (positive emotion.)

2. We should now examine the misplaced sentiments 0 and 1 to determine what may have caused their misclassification. Is additional feature engineering required or preprocessing.