Graphical user interface, text, application

Description automatically generated

Importing the packages

Packages required to build the model are imported. Packages such as numpy ,pandas, pickle, seaborn, genism etc. are imported

Table

Description automatically generated with medium confidence

Subsetting the data frame

The data is subsetted which helps us to extract the required features.

A picture containing chart

Description automatically generated

Data visualization using bar graph

The dataset is visualized in bar graph for better understanding of the dataset.

Text

Description automatically generated

Data preprocessing and cleaning

Data preprocessing and cleaning takes place. Attributes that are not required are dropped off from the table

Table

Description automatically generated

The categories are grouped , in order to obtain ten unique categories. Categories are grouped in a way such that there is no category interception. For example, Categories related to debt are aggregated into debt collection column.

Text

Description automatically generated

, The baseline model is cross validated with metrics including accuracy, macro precision, macro recall, macro f1 ,weighted precision, weighted recall, weighted f1 score . These metric scores are analyzed across various models and the best suitable model is selected.

Table

Description automatically generated

Models like Linear SVC, Multinomial naïve bayes, SVM, Random Forest etc.… are trained with different word embeddings like glove 200d,300d etc.… and their respective metrics are calculated. GloVe is an unsupervised learning algorithm for obtaining vector representations for words. This is achieved by mapping words into a meaningful space where the distance between words is related to semantic similarity. Thus, logistic regression is the best suitable model with an accuracy of 82.93%

Graphical user interface, text, application, email

Description automatically generated

Fig 4.3.8 : Model Testing

In Figure 4.3.8, the proposed model correctly classifies the random complaints .

Graphical user interface, application

Description automatically generated

The model is connected through flask API, and frontend is designed using html\css. Consumer’s can enter their complaint in the text filed and the respective category of the complaint filed is displayed.