Dissertation on

"Financial Complaint Text Classification"

Submitted in partial fulfilment of the requirements for the award of degree of

Bachelor of Technology in
Computer Science & Engineering

UE18CS390B – Capstone Project Phase - 2

Submitted by:

Sameer Kulkarni	**PES1201801474**
Sarthak S	**PES1201801703**
Sanket S Kattimani	**PES1201801710**
Sumukha MK	**PES1201801995**

Under the guidance of

Prof. Ashwini M Joshi
PES University
Designation PES University

June - December 2021

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
FACULTY OF ENGINEERING
PES UNIVERSITY
(Established under Karnataka Act No. 16 of 2013)
100ft Ring Road, Bengaluru – 560 085, Karnataka, India

PES UNIVERSITY

(Established under Karnataka Act No. 16 of 2013)
100ft Ring Road, Bengaluru – 560 085, Karnataka, India

FACULTY OF ENGINEERING

CERTIFICATE

This is to certify that the dissertation entitled

'Financial Complaint Text Classification'

is a bonafide work carried out by

Sameer Kulkarni	**PES1201801474**
Sarthak S	**PES1201801703**
Sanket S Kattimani	**PES1201801710**
Sumukha MK	**PES1201801995**

in partial fulfilment for the completion of seventh semester Capstone Project Phase - 2 (UE18CS390B) in the Program of Study - Bachelor of Technology in Computer Science and Engineering under rules and regulations of PES University, Bengaluru during the period June 2021 – December 2021. It is certified that all corrections / suggestions indicated for internal assessment have been incorporated in the report. The dissertation has been approved as it satisfies the 6th semester academic requirements in respect of projectwork.

Signature	Signature	Signature
Prof. Ashwini M Joshi	Dr. Shylaja S S	Dr. B K Keshavan
	Chairperson	Dean of Faculty

External Viva

Name of the Examiners | **Signature with Date**

1.______________________________ ______________________________

2.______________________________ ______________________________

DECLARATION

We hereby declare that the Capstone Project Phase - 2 entitled **"Financial Complaint Text Classification"** has been carried out by us under the guidance of **Prof. Ashwini M Joshi** and submitted in partial fulfilment of the course requirements for the award of degree of **Bachelor of Technology** in **Computer Science and Engineering** of **PES University, Bengaluru** during the academic semester June – December 2021. The matter embodied in this report has not been submitted to any other university or institution for the award of any degree.

PES1201801474	**Sameer Kulkarni**	[illegible]
PES1201801703	**Sarthak S**	Sarthak
PES1201801710	**Sanket S Kattimani**	Kattimani
PES1201801995	**Sumukha MK**	Sumukha M.K

ACKNOWLEDGEMENT

We are grateful to our guide Prof. Ashwini M Joshi,Department of Computer Science and Engineering, PES University, for her support, helpfulness , lovingness and kindness for this UE18CS390B Capstone Project Phase 2.

We are thankful to the project coordinator - Prof. Silviya Nancy J ,for continuous support, management and organization of this project.

We are also thankful to Dr. Shylaja S S, Chairperson, Department of Computer Science and Engineering, PES University, for her kindness. We would also like to thank Dr. B.K. Keshavan, Dean of Faculty, PES University for his help.

We are grateful to Dr. M. R. Doreswamy, Chancellor, PES University, Prof. Jawahar Doreswamy, Pro Chancellor – PES University, Dr. Suryaprasad J, Vice-Chancellor, PES University for providing to me various opportunities and enlightenment every step of the way. Finally, this project could not have been completed without the continual support and encouragement we have received from my family and friends.

ABSTRACT

Financial Companies gets many complaints regarding their service and they have to handle those complaints quickly and effectively so that the customers who are already angry don’t get unhappy and continue to trust the company. The complaints these days are handled by customer support people who can be contacted through mail or through the company’s website.

Reading many complaints for the customer support people and tagging them to route to their respective departments becomes hectic and also takes time to handle the complaints. Also, now big financial companies have made their complaint handling system online but require the customers themselves to tag the department to which the complaint should be routed.

There are high chances of customers choosing the wrong department while tagging themselves due to lack of knowledge about the departments in financial companies and there are also chances of them choosing a random department since they are only interested in writing their complaint and not filling other details. Hence, an Automated Financial Complaint Classification System becomes important that automatically tags incoming complaints by using Artificial Intelligence.

This is a real world business problem that is solved by the use of AI and ML.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Financial Companies gets many complaints regarding their service and they have to handle those complaints quickly and effectively so that the customers who are already angry don't get unhappy and continue to trust the company. The complaints these days are handled by customer support people who can be contacted through mail or through the company's website.

Reading many complaints for the customer support people and tagging them to route to their respective departments becomes hectic and also takes time to handle the complaints. Also, now big financial companies have made their complaint handling system online but require the customers themselves to tag the department to which the complaint should be routed.

There are high chances of customers choosing the wrong department while tagging themselves due to lack of knowledge about the departments in financial companies and there are also chances of them choosing a random department since they are only interested in writing their complaint and not filling other details. Hence, an Automated Financial Complaint Classification System becomes important that automatically tags incoming complaints by using Artificial Intelligence.

This is a real world business problem that is solved by the use of AI and ML. Our project can help many financial companies in the following ways:

- Faster Response: Complaints are processed instantly and helps the companies to give faster response to customers hence saves time for both customer and bank employees.
- Workforce Reduction: No need of dedicated people who are assigned to manually read each complaint and then assign them to respective categories.

Financial Companies' perspective:

- Employees in banks come across hundreds of complaints daily via mails and social media and it becomes tedious task to read long,unstructured text and then categorize them as complaints pileup.
- This project will relieve the workload of workers to a great extent. Users'perspective:
- Most of the times, customers may get confused and can forget to specify the actual department of the financial company they are targeting to.
- The project helps users to target the proper department they are complaining to.

CHAPTER 2

PROBLEM STATEMENT

The problem that we are trying to solve is to automatically classify financial complaints that should be routed to their respective departments using Machine Learning and Deep Learning Algorithms. The project involves classifying long or short complaints to respective category. Bank complaint handling department can easily direct the complaint and help the customers as soon as possible. Users who are unsure about which exact department they need to post complaint to, can also utilize the platform and make use of it to be specific and precise while complaining. Hence the project targets banks and its customers who are unhappy with any of the banking operations.

This project involves the use of Natural Language Processing for processing the unstructured complaints text into vectors that computer can understand and later the use of ML and DL models for classification of complaints into respective categories.

Increasing the accuracy of the text classification model, handling outliers while classifying text, choosing the best model for prediction, building a friendly UI can be few of the challenges.

CHAPTER 3

LITERATURE SURVEY

3.1 Complaint Classification using Word2VecModel

- Authors:- Mohit , Dikshanth , Dinabandhu
- Year of Publication:-2018
- Dataset:- Financial Consumer Complaints Dataset

METHODOLOGY USED:

In this paper, the authors consider the complaints with maximum words of 750.For data pre-processing, they convert oral complaint to text and perform tokenization. They didn't remove Stop Words so that more contextual information can be preserved. For embedding layer, they used Word2Vec Model. Then they passed their representations from Word2Vec to GRU(Gated Recurrent Unit) Model sequentially. Next they passed the output of GRU Layer to MLP with single hidden layer and output layer equal to number of classes. For MLP they trained using standard backpropagation and for GRU they trained using backpropagation through time. They used Adam as an optimiser. They chose 60:40 split for training and testing and built the model using Keras library. They chose 4 epochs as optimal by plotting Loss vs Epochs and Accuracy vs Epochs graphs and 85% classification accuracy.

Since they implemented only unidirectional RNN, as a future work they recommend to use bidirectional RNN and other variations like Stacked RNN's.

MERITS

- RNN's are not able to keep track of long term dependencies due to vanishing gradients problem which are handled by GRU

DEMERITS

- Class Imbalance:- The authors does not describe how they handled class imbalance problem in their study. Hence, accuracy cannot be considered as a good evaluation metric.
- Uni-directional model:- The authors use an unidirectional RNN type model and recommends the use of bi-directional model which can handle context better.
- LSTM(3 gates) has a more complex structure than GRU(2 gates) and can handle context better.

POSSIBLE IMPROVEMENT

Bidirectional models which can handle context better than unidirectional models since they traverse input from both sides.

3.2 A Complaint Text Classification Model Based onCharacter-level Convolutional Network

- Authors:- Xuesong , Bin , Shuyang , Jinna
- Year of Publication:-2018
- Dataset:- Consumer Express Complaints :- 10000 training samples from each class were selected

METHODOLOGY USED:

Here, the authors first use Negative Elements Removal module to remove negative sentiment words and phrases present in complaints using Hownet sentiment lexicons. The main Model they use here is Character Level Convolutional Neural Networks. They use max pooling layer to prevent overfitting of the model. They considered complaints with maximum 600 characters in length. They used Relu activation for hidden layers and softmax activation function for output layer since its Multi-Class Classification task.

Batch size of 64 was taken and Adam optimiser was used. Cross entropy was the loss function implemented using Tensorflow. To compare the good performance of their model they implemented many ML models ,CNN based on words and Character level CNN with no NER and found that their model of character level CNN with NER has the best accuracy of 91.07 %.

MERITS

- They performed study on both English and Chinese language data.
- They compared traditional ML models with Deep learning models.
- They handled class imbalance problem by undersampling.

DEMERITS

- CNN’s are not capable of identifying context from sequential data like text.
- The authors considered only top three majority classes for their study which is not applicable for real life practical applications
- Character level CNN’s require more time to train compared to Word-Level CNN’S since they parse the input character by character.

POSSIBLE IMPROVEMENT

Word-level CNN with Auto Spell Check and Correct can be useful and requires less training time compared to Character level CNN.

.

3.3 Developing an LSTM based Classification Model of IndiHome Customer Feedback

- Authors:- Anditya, Dhamir, Dian, Suyanto, Irfan, Anis, Fedy, Rani
- Year of Publication:-2020
- Dataset:- User Feedback from ISP in Indonesia named IndiHome with total complaints of 1107 with 4 classes.

METHODOLOGY USED:

For Data Pre-processing they removed punctuations and numbers and converted text to lower case. Then they removed stop words and performed tokenization. For word vector representation, they used TF-IDF. Maximum limit of 1000 terms is used for their model. They used a ratio of 60:10:30 for training , validation and testing sets respectively. They used F1-score instead of accuracy to evaluate the model since the dataset was imbalanced. They then classified complaints using two models:- Naïve Bayes with TF- IDF and LSTM .They finally found that LSTM model gave an F1-score of 87.98% compared to Naïve Bayes model that gave an F1- score of 76.77%.

MERITS

- LSTM has a complex structure than GRU and hence can handle more long term dependencies than GRU does
- The authors considered F1 score as evaluation metric instead of accuracy since the dataset was imbalanced

DEMERITS

- The authors compare their deep learning model with just a single ML model.
- .They consider uni-directional LSTM model which can handle context from only one side i.e., left to right

POSSIBLE IMPROVEMENT

Bidirectional-LSTM model can be applied as it handles context better from sequential data like text.

3.4 Automatic Complaint Classification System Using Classifier Ensembles.

- Authors:- M Ali Fauzi
- Year of Publication:-2018
- Dataset:- 204 records from Sambhat system

METHODOLOGY USED:

Authors first applied text pre-processing steps . Bag of Words features were generated from this . Next , they trained machine learning models like Naive Bayes,MaximumEntropy,K N N ,Random Forest classifiers and SVM using Bag Of Words features.For the Naïve Bayes classifier, they used Gaussian, multinomial and Bernoulli NB. For SVM they used linear, polynomial, sigmoid, and RBF kernels .In the combination stage,hard and soft voting methods are used. In hard voting , document was assigned to the category which was predicted by major number of classifier and with the soft voting method, the average of different classifiers were used .

As result of research author got 80.7% accuracy using Multinominal naive bayes which was best among five individual classifiers.Also author got accuracy of 81.2% when they used ensemble method with three best classifiers.

MERITS

- Explains how different classifying methods work and compare their results.
- Ensemble methods avoids the use of bad classifier for text classification.

DEMERITS

- Better feature selection method like TF-IDF can be used along with of bag of words.
- Only 3 classifiers are used for ensemble method which gives better result for their dataset but may not be true for all cases

3.5 Active Learning SVM Classification Algorithm for Complaints Management Process Automatization.

- Authors:- Pavels Goncarovs.
- Year of Publication:-2019
- Dataset:- complains documents in Latvian areuse

METHODOLOGY USED:

The author made experiments to compare decision tree with SVM for complaints classification task.They did text preprocessing and represented data as bag of words. They selected terms with relative frequency greater than 3 for their study.

They used decision tree algorithm and Sequencial Minimal Optimization SVM.

Using only 20% of the data SVM performed far better than Decision tree. Results showed SVM got accuracy of 86% whereas decision tree produced irrelevant results.

MERITS

- Sequencial Minimal Optimization requires small amount of data.
- Training time reduced.

DEMERITS

- No information about dataset is given.
- Method may not be applicable for all type of classification problems.

3.6 Bank Customer Complaints Analysis Using Natural Language Processing and Data Mining

- Authors:- Chandana ,Neelashree,Nikitha ,Nisargapriya ,Vishwesh
- Year of Publication:-2020
- Dataset:- bank customer's complaints from Kaggle

METHODOLOGY USED:

This project is implemented using java programming language.Firstly they use LDA(*Linear Discriminant Analysis)* Algorithm which is an unsupervised method for dimensionality reduction for the process of text classification.

They later used t- S N E for data visualisation. They used sentence segmentation and later converted sentence segmentation into Tokenization.

MERITS

- They examine the complaint data to determine where the most complaints are being filed.
- Their model can help banks recognize the types of errors that need to be fixed, resulting in higher customer satisfaction.

DEMERITS

- For processing or analysing datasets, LDA takes longer. It's preferable if they combine the TF-IDF approach with a bag of terms.
- TF-IDF Accuracy is better than LDA.

3.7 Sentiment Classification of Indian Banks' Customer Complaints

- Authors:- Gutha , Vadlamani, Bheemidi , Harshal
- Year of Publication: 2019
- Dataset:- The customer complaints information of SBI, ICICI, AXIS and HDFC banksare gathered from the "complaintsboard.com".

METHODOLOGY USED:

TermFrequency(TF)is the word frequency and Inverse Document Frequency(IDF) is the find the weight of uncommon words. The words that infrequently appeared in the corpus have a high ID Fscore.Supportvectormachines(SVM)locatesthemaximumseparatinghyperplanes for the target classes.SVM utilizes a form of constrained optimization.Naive Bayes has the advantage of being built on a robust probabilistic foundation and is very robust. Decision tree models which take a discrete set of class names are called classification trees.Random Forest is a machine learning algorithm that can perform both classification and regression tasks.By combining multiple week learners, it produces a powerful ensemble learner.In most cases,Logistic Regression is used to solve a two-class problem. It uses regression to estimate the beta parameters for the individual inputs.

MERITS

- Information gathered from online complaints and feedback portals of banks was used to create a sentiment classification based on customer reviews from multiple banks.
- They used three different feature extraction techniques and found that RF and NB are statistically important.

DEMERITS

- On the results of the feature extraction methods, we can apply feature selection methods and use deep learning methods as classifiers.
- In addition, we plan to include more banks in the study and seek feedback from them in order to improve our results.

CHAPTER 4

Project Requirement Specification

4.1 Introduction

Financial complaint classifier application will mainly help bank employees who are under consumer complaint department to easily classify the complaints they receive to respective departments of the bank. Users who are unsure about which particular department they need to address to, can also make use of the application. So the end goal of this product is to provide intuitive and user friendly application which provides useful insights about the nature of the complaints.

4.1.1 Project Scope

The Bank employees,often receive lengthy, unstructured complaints. It takes lot of time and effort by these employees to know the department they need the complaint to be addressed to.So this application will organize these complaints without having to actually read the complaints. Also users who are uncertain as to which exact department of the bank they have to complain to can also make use of this application. The objective of this application is to instantly classify complaints, free up manpower. Limitations of this application will be regarding the language the user uses, application would not support all the languages the user uses.

4.1.2 Project Perspective and Features

This product will provide the employees and users with a friendly interface where people can type or copy paste the complaint. The results shown by the product would specify the classification levels of various departments present in the bank in terms of percentage. This product thus enables to instantly classify complaints. Product Features includes Displaying results in a user friendly manner will be an important feature of the application, The display result will contain all the departments along with the percentage. Spell Check feature: The application also checks the spelling of the complaint text and prompt users.

4.1.3 Operating Environment

The deployed web application will be operational by any system that has the popular updated web browsers,operating system, and a decent internet connection.

4.1.4 General Constraints, Assumptions and Dependencies

Constraint: Language constraint: English is the preferred language for the complaint text, not all languages will be supported

Assumptions:

- System is connected to internet
- System should be user friendly so that its easy for both employees and users
- System should be running 24 hours, throughout the year

4.1.5 Risks

Since users might share their banking information which is confidential , Risks include man in the middle attack,phishing.

4.2 Functional Requirements

Input : Complaint text will be given as input

Output: Classified results showing departments along with the percentage matching to that respective departments

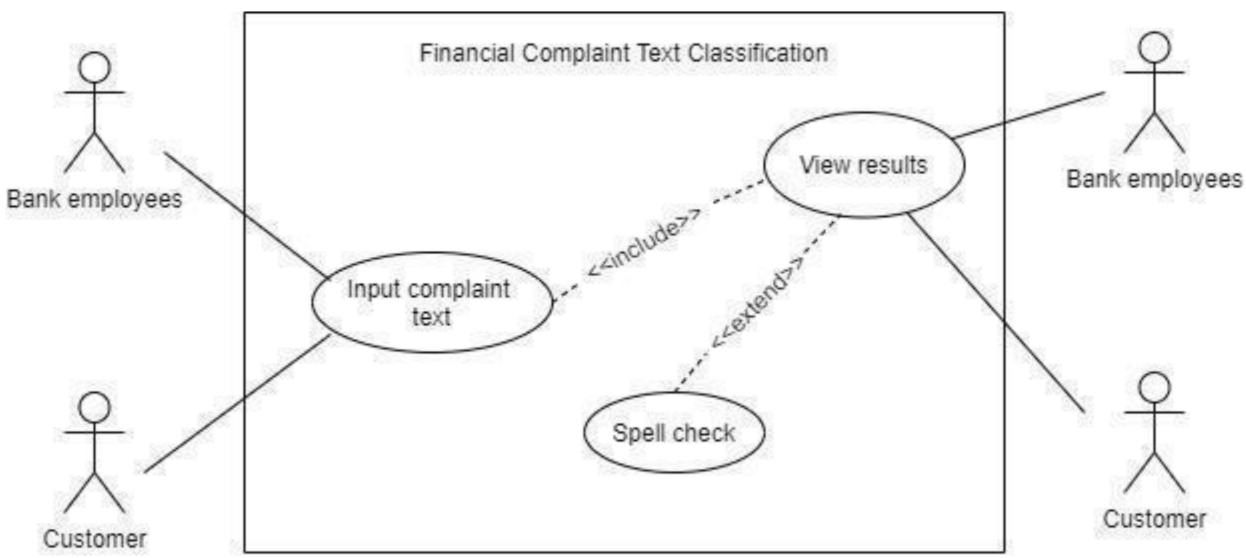


Fig: 4.2.1 Overall use case diagram

4.3 External Interface Requirements

4.3.1 User Interfaces

The application provides user friendly experience. It will be easy to use, the employees or the users will be given textbox where they can easily type or copy paste the complaint text. Submit button will be provided, the results will be shown in a very intuitive manner , hence enabling an easy to use platform. Spell check features will also be shown.

4.3.2 Hardware requirements

- Server : The Financial Text Classification application will run on a web server listening on port 80.

- Client :The web application will be displayed on client's monitor or laptop screen.The application will encourage users to use the mouse to interact with the components of the Website , Mouse will help users to activate buttons like submit and also helps to position the Cursor.The application also needs the requirement of a keyboard if users wish to type the complaint.

4.3.3 Software Requirements

- Serverside: Flask will be used for backend. It provides a development server and a debugger.

- Clientside: Any popular web browsers which supports JavaScript, HTML5, CSS Additional Tools: TensorFlow, scikit learn, etc

4.3.4 Communication Interfaces

Communication between the client and server is eased via https protocol.

4.4 Non Functional Requirements

4.4.1 Performance Requirements

The performance is mainly dependant on the internal working of machine learning model accuracy. The classification model chosen will provide accurate results while classifying. The application is reliable in terms of taking input and displaying results as soon as possible when users click the submit button. The application will be accessible in any devices with internet connection and browser.

4.4.2 Safety Requirements

Safety measures have to be taken to make sure that server won't face downtime while serving the web page. Maintenance team is desired for safeguarding the working of the team.

4.4.3 Security Requirements

Users tend to share their banking information like account number, credit card/debit card number while complaining , The website need to make sure that there is no man-in-the-middle attacks, so we use https instead of http

Appendix A: Definitions, Acronyms and Abbreviations

HTML5: hypertext markup language mainly acts as a foundation to the webpage; structuring is taken care her.

CSS: Cascading style sheet helps to add style to the web pages.
HTTPS: Hypertext transfer protocol secure (HTTPS) enables secure communication between client and server

port 80: By default, the port number for a Web server is 80 Flask : Flask is a web application framework written in Python.

CHAPTER 5

System Design

5.1 Introduction

Financial complaint classifier application will mainly help bank employees who are under consumer complaint department to easily classify the complaints they receive to respective departments of the bank. Users who are unsure about which particular department they need to address to, can also make use of the application. So the end goal of this product is to provide intuitive and user friendly application which provides useful insights about the nature of the complaints.

5.2 Current System

The Current System for Financial Complaints Classification uses various Machine Learning Algorithms. Recently Deep Learning Models like CNN and RNN are being tried for Complaints Classification.

5.3 Design Considerations

5.3.1 Design Goals

The Main Goal:- To classify financial complaints into respective categories for faster response and workforce reduction.

The proposed innovative approach is to implement Bi-Directional LSTM and Word Level CNN Model for our problem of classifying financial complaints along with implementing many MachineLearning, Ensemble and Deep Learning Models.

BENEFITS:-

- LSTM Models are better than RNN Models due to their ability to deal with long term dependencies.

- Since Bi-directional LSTM can handle input complaint from both sides, it handles context better than Vanilla LSTM Model.

- Since customers who complain are usually angry, there is high probability for spelling errors and correction of these errors as a pre-processing step can be useful.

Word level CNN trains faster than Character level CNN since it takes stream of words than characters.

DRAWBACK

- Bi-directional LSTM Model due to its complex structure can take large training time compared to Vanilla LSTM Model.

5.3.2 Architecture Choices

Deep Learning Models like CNN and RNN generally perform better than Machine Learning Models for text classification tasks because of their complex structure and the ability to learn feature extraction better compared to traditional Machine Learning Models.

RNN’s may be useful compared to CNN for text classification since CNN’s do not have the ability to understand context from sequential data like text.

LSTM and GRU are better than vanilla RNN because of their ability to handle long term dependencies from text data. RNN’s face this issue because of the Vanishing Gradients problem.

LSTM may handle context better compared to GRU Model since LSTM has a complex structure(3 Gates) than GRU Model(2 Gates). However, LSTM takes a huge training time compared to GRU Model due to its complex structure.

Bidirectional LSTM can handle context from text better than Vanilla LSTM model since it runs input from both left to right and right to left. However, Bidirectional Models take huge time for training.

5.3.3 Constraints, Assumptions and Dependencies

- The approach assumes that the customer enters his/her complaint only in English language.
- It also assumes that the complaint has no mixed language in its input.
- There is also a constraint that the user enters only financial complaints and not any other comments.

Dependency: The application is dependent on the maintenance team to look after the running of the server and ensure smooth functioning of the application.

Assumptions:

- System is connected to internet
- System should be user friendly so that its easy for both employees and users
- System should be running 24 hours, throughout the year Risks:
- Since users might share their banking information which is confidential,Risks include man in the middleattack, phishing.

5.4 High level System Design

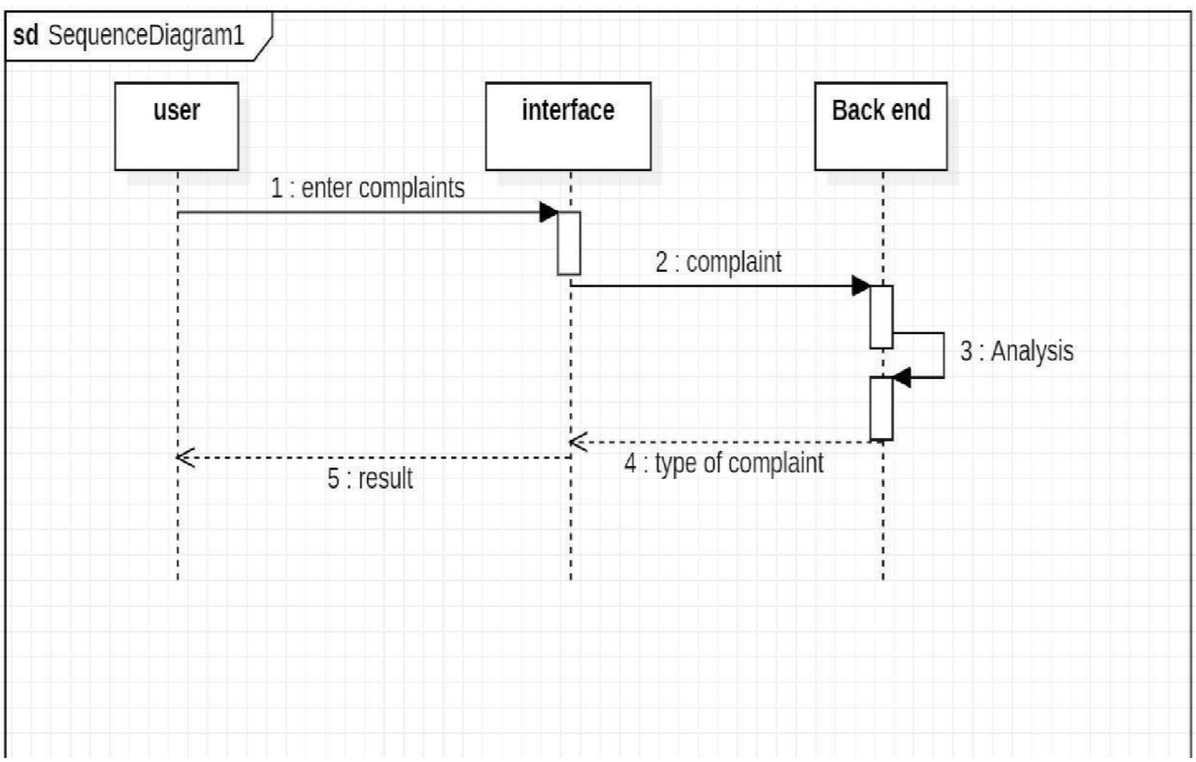


Fig 5.4.1

5.5 Design Description

5.5.1 Master Class Diagram

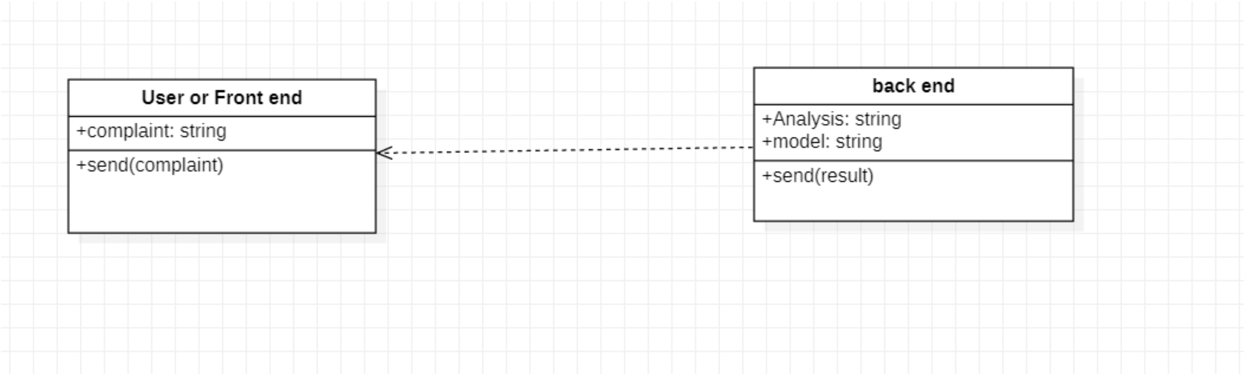


Fig 5.5.1

5.5.2 Reusability Considerations

- Scikit-learn library modules.
- NLTK modules.

5.6 Activity Diagram

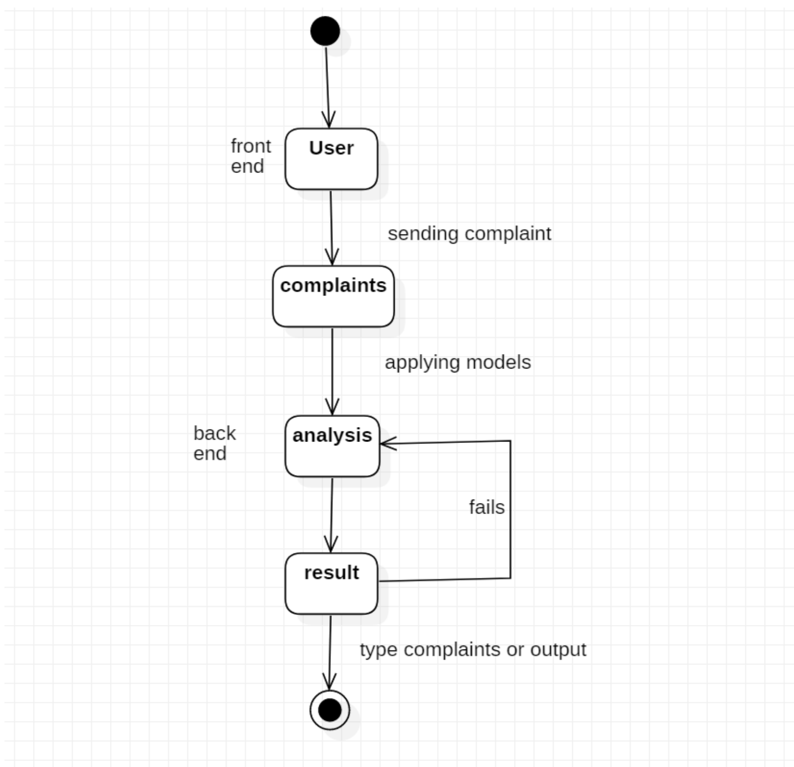


Fig 5.6.1

5.7 User Interface Design

There is only one webpage with which user will interact.There will be a text box in which user write his complaint and click the check button. After that the complaint will be processed by background algorithm and returns results in result section.

5.8 External Interfaces

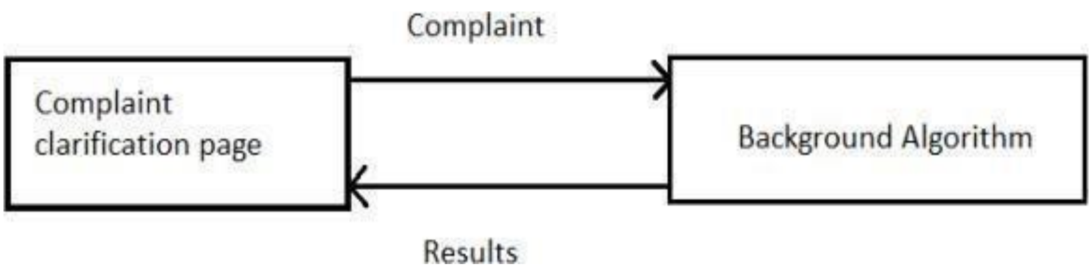
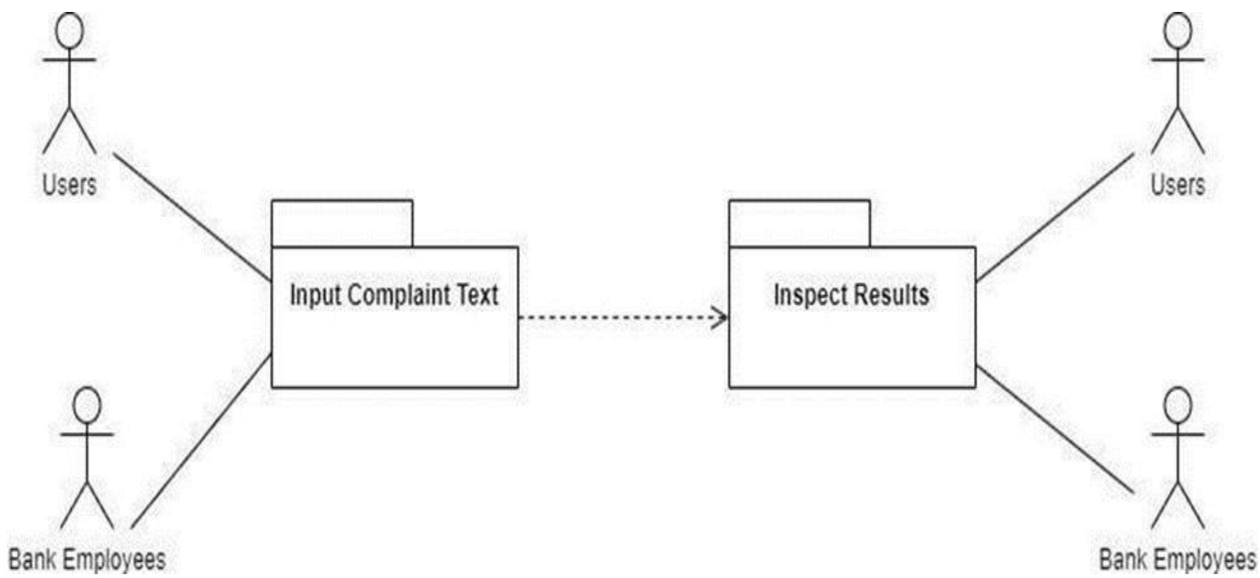
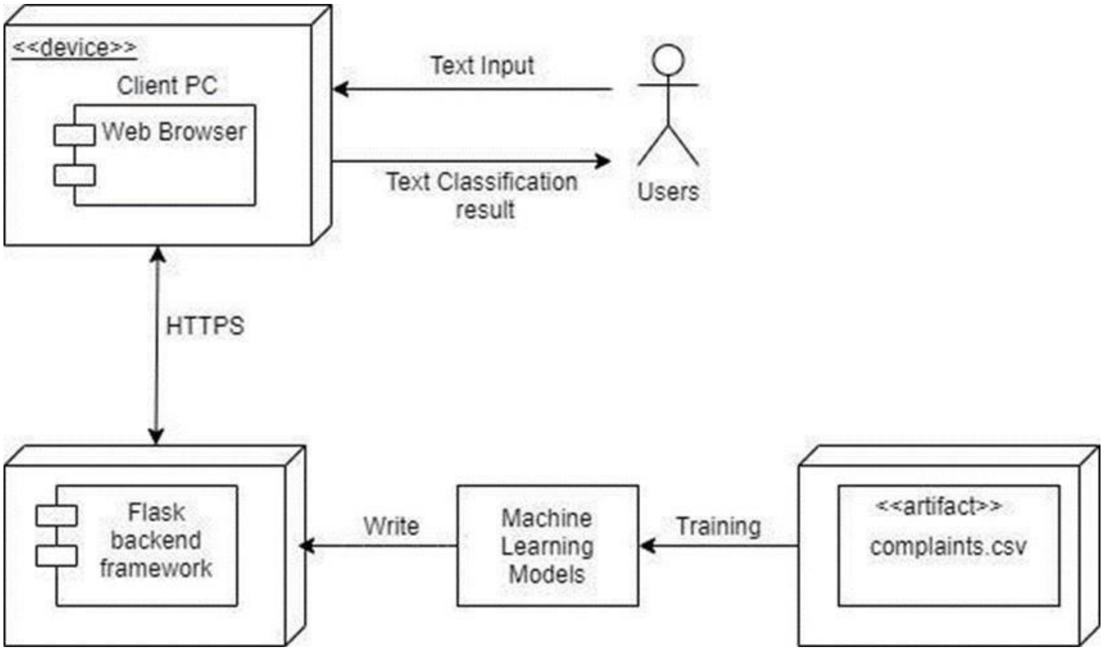


fig 5.8.1

5.9 Packaging and Deployment Diagrams



Packaging diagram : fig 5.9.1



Deployment Diagram : fig 5.9.2

5.10 Help

Instructions to the users :

- Connect to the internet and access the platform
- Enter the complaint text in the text box provided
- Click on Submit Button to view and inspect results

5.11 Design Details

5.11.1 Novelty

Implementing many models including bi-directional models and choosing the model with best accuracy for deployment purposes.

5.11.2 Innovativeness

Since complaints contain many spelling errors, applying spelling correct as a pre- processing task can be helpful. Since uni-directional LSTM can handle input through only one side, a bidirectional model LSTM which can handle context from both sides.

5.11.3 Performance

The performance is mainly dependant on the internal working of machine learning model accuracy. The classification model chosen will provide accurate results while classifying. The application is reliable in terms of taking input and displaying results as soon as possible when users click the submit button.The application will be accessible in any devices with internet connection and browser.

5.11.4 Safety

Safety measures have to be taken to make sure that server won't face downtime while serving the webpage.Maintenance team is desired for safeguarding the working of the team.

5.11.5 Security

Users tend to share their their banking information while complaining , The website need to make sure that there is no man-in-the-middle attacks, so we use https instead of http as the HTTPS protocol is secure over the internet

5.11.6 Maintenance

The system doesn't require any maintenance from customers/clients. Maintenance team handles the system during any failures or any need to improve the accuracy of the system using hyper parameter tuning.

System Design:

5.12 Process Chart

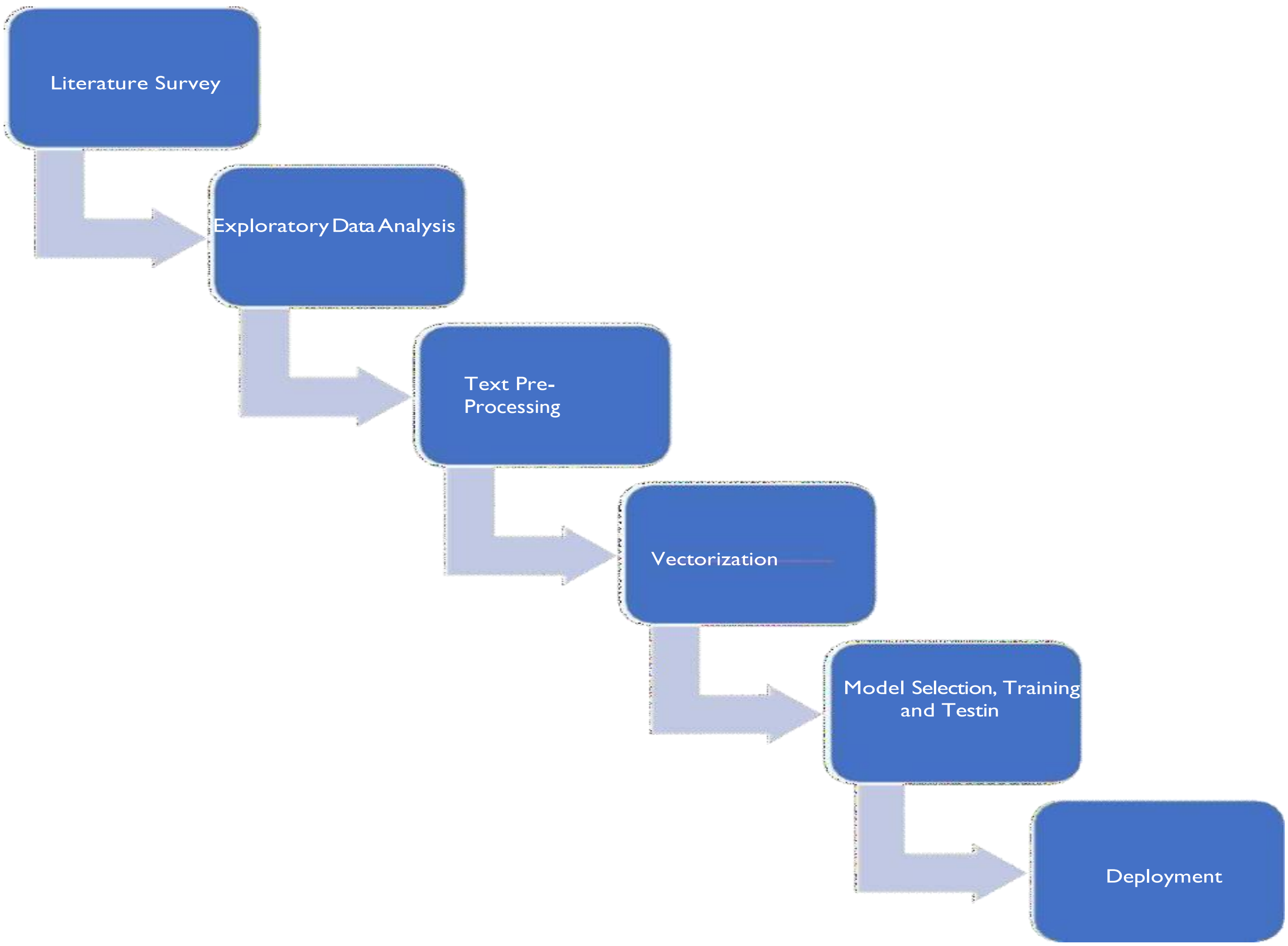


Fig 5.12

5.13 Design Diagram

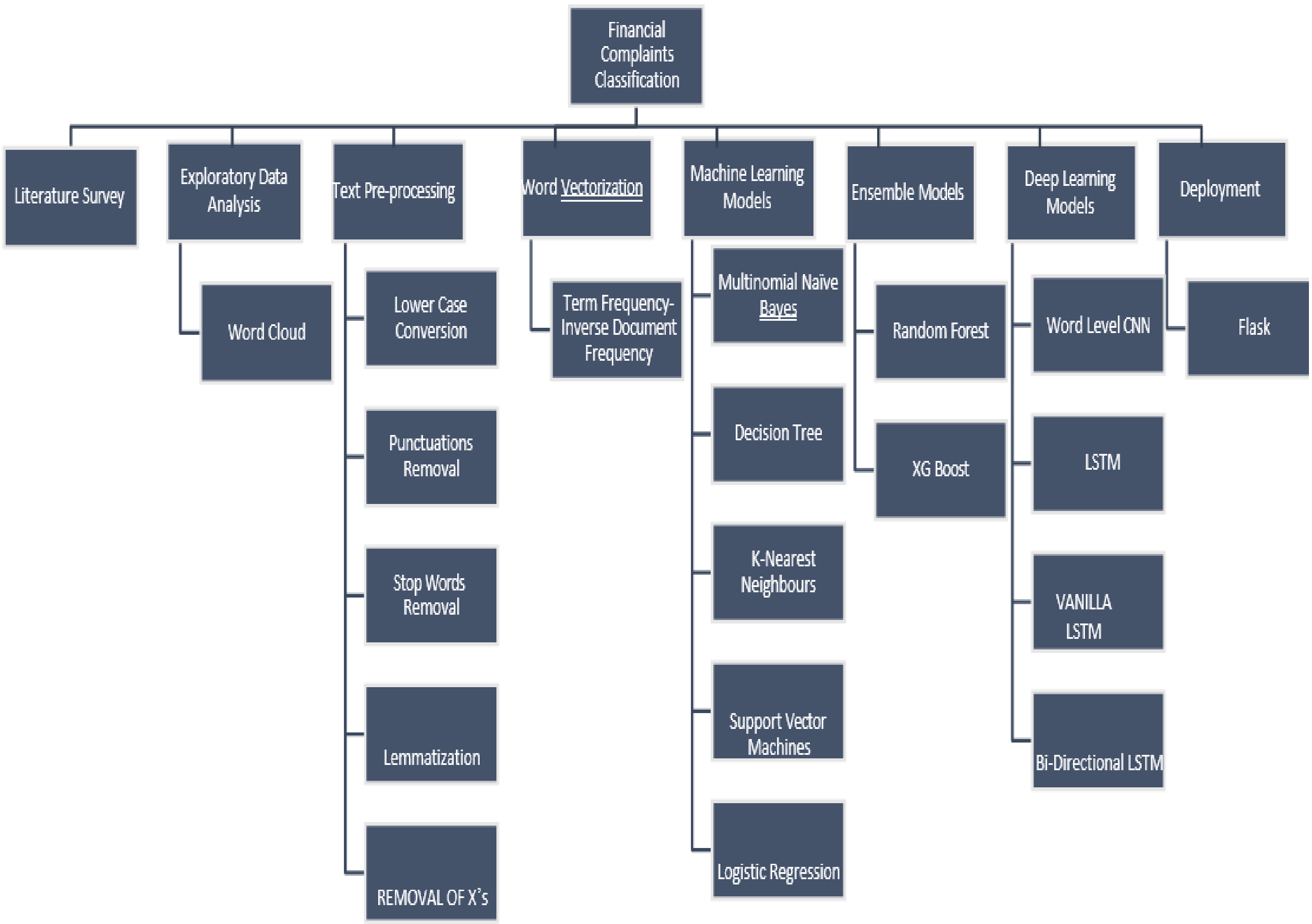


Fig 5.13

CHAPTER 6 Implementation

6.1 DATA:

For our analysis, we used dataset from Consumer Financial Protection Bureau government website. All the latest complaints are updated on their website weekly and the dataset is available for open research. We took latest complaints starting from the last year for our analysis. All the complaints description is present in “Consumer_complaint_narrative” column and the category of complaint is present in “Product” column. There are 9 categories in total in the Product column.

Since the dataset was highly imbalanced as shown in below figure, we handled class imbalance problem by undersampling majority classes. For increasing data for minority classes, we used Data Augmentation using the technique of Synonym Replacement by replacing random six words with its synonyms for each complaint. Using these techniques, 20000 rows from each class is taken for further analysis.

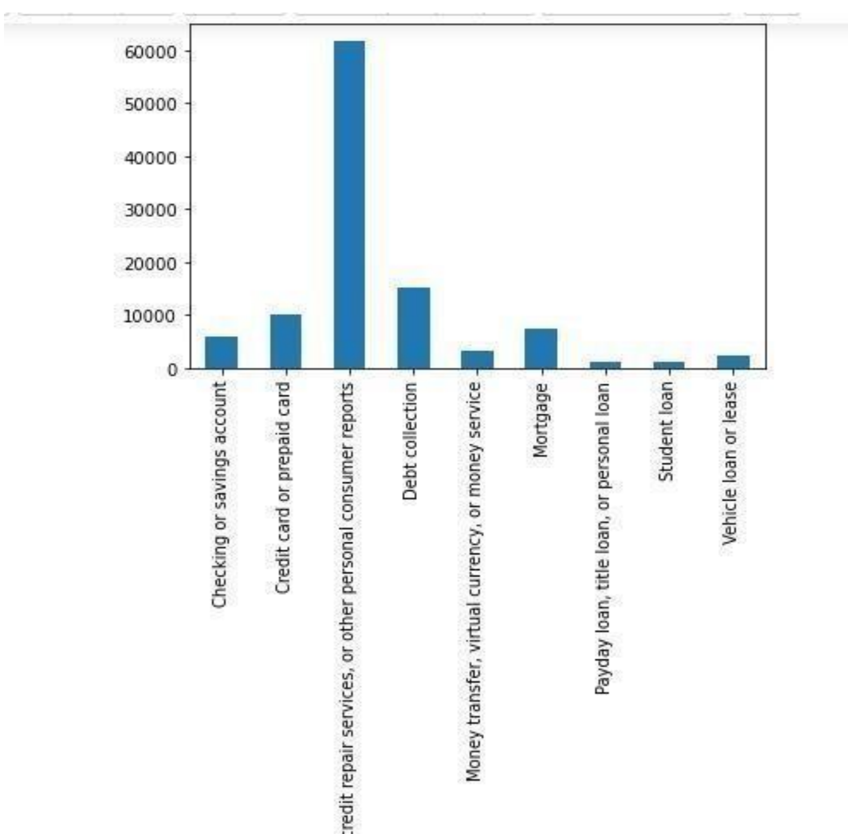


Fig 6.1.1 : Initial Class Distribution

6.2 DATA AUGMENTATION USING SYNONYM REPLACEMENT

```
Original text is:  someone in north Carolina has stolen my identity information and has purchased items including XXXX cell pho
nes thru XXXX on XXXX/XXXX/2015. A police report was filed as soon as I found out about it on XXXX/XXXX/2015. A investigation f
rom XXXX is under way thru there fraud department and our local police department.
-----------------------------------------------------------------------------------------------------------------------
```

Fig 6.2.1

```
-----------------------------------------------------------------------------------------------------------------------
This is ans: mortal in north Carolina has stolen my identity selective information and has purchased items including XXXX cell
phones thru XXXX on XXXX/XXXX/2015. A police report was charge as soon as unity line up out about it on XXXX/XXXX/2015. A inves
tigation from XXXX is under way thru there fraud section and our local police department.
-----------------------------------------------------------------------------------------------------------------------
```

Fig 6.2.2

6.3 EXPLORATORY DATA ANALYSIS:

We visualized our text data by finding the most frequent words present in each category by Word Cloud Visualisation.

Fig 6.3.1

6.4 TEXT PRE-PROCESSING:

We applied the following pre-processing steps to our text data:

1. Lower Case Conversion
2. Punctuations Removal
3. Digits Removal
4. Stop Words Removal
5. Lemmatization
6. Removal of confidential information which were represented by x's in complaints

6.5 MACHINE LEARNING MODEL BUILDING AND TESTING:

We used Term Frequency – Inverse Document Frequency technique for word vector representations. Then, we split the data into train and test data as 70% and 30% of data. Five Machine Learning algorithms are implemented and compared.

6.5.1 MUTINOMIAL NAÏVE BAYES

We achieved an accuracy of around 80 percent with Multinomial Naive Bayes Model with alpha value of one for smoothing.

	precision	recall	f1-score	support
Checking or savings account	0.71	0.83	0.76	6003
Credit card or prepaid card	0.74	0.77	0.76	6010
Credit reporting, credit repair services, or other personal consumer reports	0.75	0.74	0.75	5955
Debt collection	0.79	0.73	0.76	5922
Money transfer, virtual currency, or money service	0.86	0.78	0.81	6026
Mortgage	0.87	0.94	0.90	6077
Payday loan, title loan, or personal loan	0.81	0.69	0.74	5927
Student loan	0.88	0.90	0.89	6035
Vehicle loan or lease	0.81	0.81	0.81	6045
accuracy			0.80	54000
macro avg	0.80	0.80	0.80	54000

Fig 6.5.1

6.5.2 DECISION TREE

Decision Tree Model using Gini Impurity Index is built and we achieved an accuracy of 71.33 percent.

```
                                                                          precision    recall  f1-score   supp
ort

                                               Checking or savings account       0.65      0.63      0.64      6
003
                                               Credit card or prepaid card       0.66      0.65      0.66      6
010
Credit reporting, credit repair services, or other personal consumer reports       0.66      0.68      0.67      5
955
                                                           Debt collection       0.66      0.65      0.66      5
922
                                Money transfer, virtual currency, or money service       0.72      0.73      0.73      6
026
                                                                  Mortgage       0.84      0.82      0.83      6
077
                                      Payday loan, title loan, or personal loan       0.66      0.71      0.68      5
927
                                                              Student loan       0.85      0.81      0.83      6
035
                                                      Vehicle loan or lease       0.74      0.73      0.73      6
045

                                                                  accuracy                           0.71     54
000
                                                                 macro avg       0.71      0.71      0.71     54
000
                                                              weighted avg       0.71      0.71      0.71     54
```

Fig 6.5.2

6.5.3 LINEAR SVM

Linear SVM Model is built which intuitively is based on best fit line that separates the data points. Since this is a multi-class problem, One vs All method is being used to classify the complaints using Linear SVM. L2 penalty with squared hinge loss is used. C value of 1 is used. An accuracy of 83.51 percent is achieved using this model.

```
                                                                precision    recall  f1-score   supp
ort

                                         Checking or savings account       0.79      0.81      0.80      6
003
                                         Credit card or prepaid card       0.79      0.81      0.80      6
010
Credit reporting, credit repair services, or other personal consumer reports       0.80      0.77      0.78      5
955
                                                     Debt collection       0.79      0.79      0.79      5
922
                   Money transfer, virtual currency, or money service       0.85      0.83      0.84      6
026
                                                            Mortgage       0.91      0.94      0.92      6
077
                           Payday loan, title loan, or personal loan       0.81      0.81      0.81      5
927
                                                        Student loan       0.93      0.91      0.92      6
035
                                                 Vehicle loan or lease       0.84      0.85      0.85      6
045

                                                            accuracy                           0.84     54
000
                                                           macro avg       0.84      0.83      0.83     54
000
                                                        weighted avg       0.84      0.84      0.84     54
000
```

Fig 6.5.3

6.5.4 LOGISTIC REGRESSION

Logistic Regression which is based on logistic function is being built using one vs rest method since it is intuitively applicable for only binary classification problems. L2 regularizer is used with a C value of one. Accuracy obtained is 83.03 %.

	precision	recall	f1-score	support
Checking or savings account	0.78	0.82	0.80	6003
Credit card or prepaid card	0.79	0.81	0.80	6010
Credit reporting, credit repair services, or other personal consumer reports	0.79	0.77	0.78	5955
Debt collection	0.78	0.80	0.79	5922
Money transfer, virtual currency, or money service	0.84	0.83	0.84	6026
Mortgage	0.92	0.93	0.93	6077
Payday loan, title loan, or personal loan	0.78	0.79	0.78	5927
Student loan	0.93	0.90	0.91	6035
Vehicle loan or lease	0.84	0.83	0.84	6045
accuracy			0.83	54000
macro avg	0.83	0.83	0.83	54000
weighted avg	0.83	0.83	0.83	54000

Fig: 6.5.4

6.5.5 K NEAREST NEIGHBOURS

KNN model which is simple to understand and implement is the last ML Model we built. Since we need to explicitly specify the value of K for building the KNN Model, we used the technique of Elbow Method to find the optimal value of K as shown in the figure below. All K values taken are odd to avoid ties. Euclidean distance is used as distance metric.

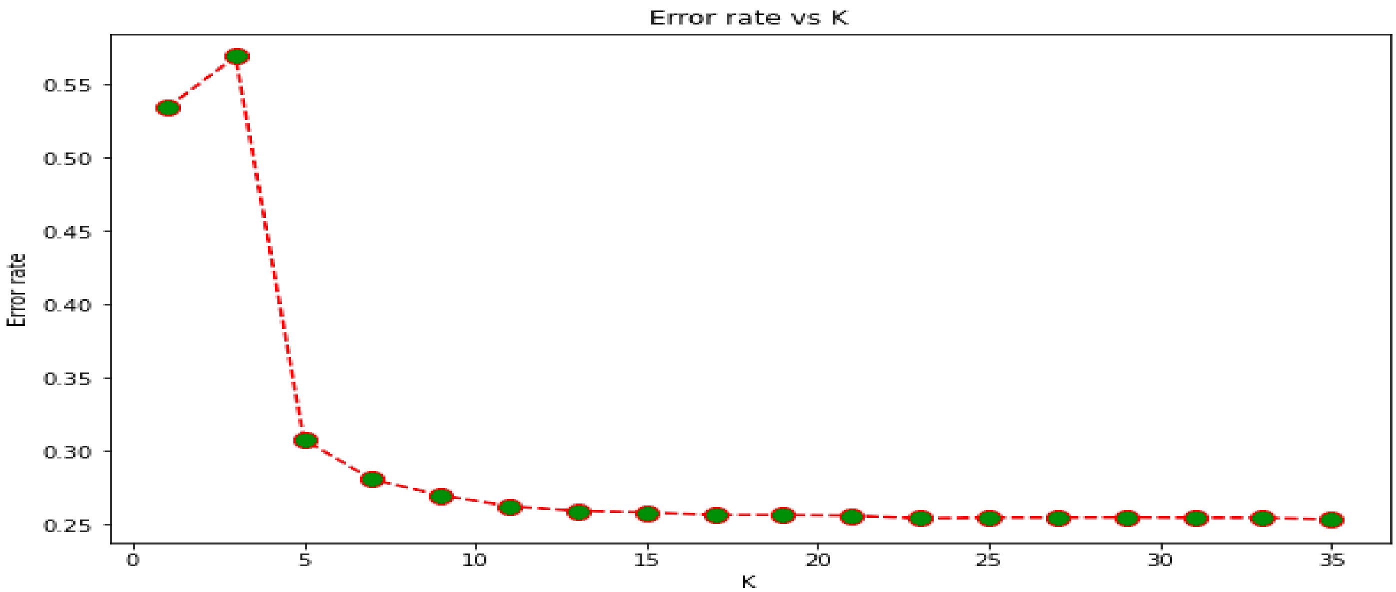


Fig: 6.5.5

Using elbow method from the above graph, K value of 11 is taken for building the model and we obtained an accuracy of 73.76 percent

	precision	recall	f1-score	support
Checking or savings account	0.66	0.74	0.70	6003
Credit card or prepaid card	0.70	0.73	0.71	6010
Credit reporting, credit repair services, or other personal consumer reports	0.62	0.82	0.71	5955
Debt collection	0.69	0.72	0.71	5922
Money transfer, virtual currency, or money service	0.80	0.75	0.77	6026
Mortgage	0.87	0.83	0.85	6077
Payday loan, title loan, or personal loan	0.71	0.58	0.64	5927
Student loan	0.84	0.80	0.82	6035
Vehicle loan or lease	0.80	0.67	0.73	6045
accuracy			0.74	54000
macro avg	0.74	0.74	0.74	54000
weighted avg	0.75	0.74	0.74	54000

Fig: 6.5.6

6.6 MACHINE LEARNING MODELS USING BIGRAMS

After implementing Machine Learning Models using Unigrams, we implemented these models using Bigrams which considers two words at a time for training. The results obtained are shown in the below table.

Sl.No.	Model	Unigram Accuracy	Bigram Accuracy
1	Multinomial Naïve Bayes	79.99%	82.94%
2	Decision Tree	71.33%	65.35%
3	Linear SVM	83.51%	85.35%
4	Logistic Regression	83.03%	81.71%
5	K-nearest Neighbours	73.76%	69.10%

Table 1

6.7 ENSEMBLE LEARNING MODELS

RANDOM FOREST AND XG BOOST

Two ensemble learning models namely Random Forest which is a bagging based model and XG Boost which is a boosting based model were built and the results obtained are shown below using 100 estimators and 500 estimators.

Sl.No.	Model	Accuracy with 100 estimators	Accuracy with 500 estimators
1	Random Forest	83.83%	84.78%
2	XG Boost	82.37%	84.68%

Table 2

Fig 6.7.1: Random Forest with 500 estimators

```
                                                        precision    recall  f1-score   supp
ort

                             Checking or savings account       0.79      0.81      0.80      6
003
                             Credit card or prepaid card       0.81      0.79      0.80      6
010
Credit reporting, credit repair services, or other personal consumer reports       0.77      0.81      0.79      5
955
                                         Debt collection       0.81      0.78      0.80      5
922
              Money transfer, virtual currency, or money service       0.85      0.87      0.86      6
026
                                                Mortgage       0.91      0.93      0.92      6
077
                       Payday loan, title loan, or personal loan       0.90      0.84      0.87      5
927
                                            Student loan       0.93      0.91      0.92      6
035
                                   Vehicle loan or lease       0.88      0.88      0.88      6
045

                                                accuracy                           0.85     54
000
                                               macro avg       0.85      0.85      0.85     54
000
                                            weighted avg       0.85      0.85      0.85     54
000
```

Fig 6.7.2: XG Boost with 500 estimators

```
                                                              precision    recall  f1-score   supp
ort

                                     Checking or savings account       0.80      0.82      0.81       6
003
                                     Credit card or prepaid card       0.81      0.83      0.82       6
010
Credit reporting, credit repair services, or other personal consumer reports       0.80      0.80      0.80       5
955
                                                 Debt collection       0.81      0.80      0.80       5
922
                      Money transfer, virtual currency, or money service       0.86      0.84      0.85       6
026
                                                        Mortgage       0.93      0.92      0.93       6
077
                           Payday loan, title loan, or personal loan       0.81      0.84      0.82       5
927
                                                    Student loan       0.94      0.91      0.92       6
035
                                          Vehicle loan or lease       0.86      0.87      0.86       6
045

                                                        accuracy                           0.85      54
000
                                                       macro avg       0.85      0.85      0.85      54
000
                                                    weighted avg       0.85      0.85      0.85      54
000
```

6.8 *DEEP LEARNING MODEL BUILDING AND TESTING*

Recently, deep learning models are becoming popular for researchers for many NLP tasks like sentiment analysis, machine translation, etc. We built a deep neural architecture (LSTM) for this phase.

6.8.1 *LONG SHORT TERM MEMEORY MODEL(LSTM)*

LSTM is a variant of RNN Model. It can handle long term dependencies better than vanilla RNN model and hence very useful for various NLP tasks.

Maximum of 400 words from each complaint is taken and embedding dimension of 100 is chosen. Train test ratio of 80:20 along with validation split of 10 percent is chosen. The number of epochs set is 9 along with a batch size of 64. Embedding layer is used between input and LSTM layer. Spatial

Dropout of 20 percent is used. Next, LSTM layer of 128 units is used with dropout and recurrent dropout percentage of 20 to avoid overfitting. Final layer consisted of a Dense Layer of 9 neurons since there are nine classes and softmax activation function is being used since it is multiclass text classification problem. Loss function used is categorical cross entropy along with Adam optimizer.

Early Stopping with a patience factor of 2 is used which means the model will stop when maximum of two epochs show no improvement in validation set. Min_delta is set to 0.0001 which means the minimum improvement should be greater than 0.0001 for it to be considered an improvement in Early Stopping.

The model stopped running after 7 epochs due to Early Stopping. Training accuracy of 92.11% and validation set accuracy of 84.02% is obtained after 7 epochs. Testing set accuracy obtained is 83.72 %.

Fig 6.8.1

```
Epoch 1/9
2025/2025 [==============================] - 1227s 604ms/step - loss: 1.6096 - accuracy: 0.3910 - val_loss: 0.7429
- val_accuracy: 0.7694
Epoch 2/9
2025/2025 [==============================] - 1214s 600ms/step - loss: 0.6972 - accuracy: 0.7813 - val_loss: 0.5606
- val_accuracy: 0.8251
Epoch 3/9
2025/2025 [==============================] - 1210s 598ms/step - loss: 0.4928 - accuracy: 0.8468 - val_loss: 0.5091
- val_accuracy: 0.8388
Epoch 4/9
2025/2025 [==============================] - 1208s 597ms/step - loss: 0.4044 - accuracy: 0.8728 - val_loss: 0.5092
- val_accuracy: 0.8404
Epoch 5/9
2025/2025 [==============================] - 1202s 593ms/step - loss: 0.3438 - accuracy: 0.8902 - val_loss: 0.4996
- val_accuracy: 0.8406
Epoch 6/9
2025/2025 [==============================] - 1209s 597ms/step - loss: 0.2949 - accuracy: 0.9060 - val_loss: 0.5193
- val_accuracy: 0.8398
Epoch 7/9
2025/2025 [==============================] - 1219s 602ms/step - loss: 0.2473 - accuracy: 0.9211 - val_loss: 0.5358
- val_accuracy: 0.8402
```

```
1125/1125 [==============================] - 98s 87ms/step - loss: 0.5450 - accuracy: 0.8373
Test Set Loss 0.5450457334518433
Test Set Accuracy 0.8372777700424194
```

Following are the loss and accuracy graphs for train and test data with epochs.

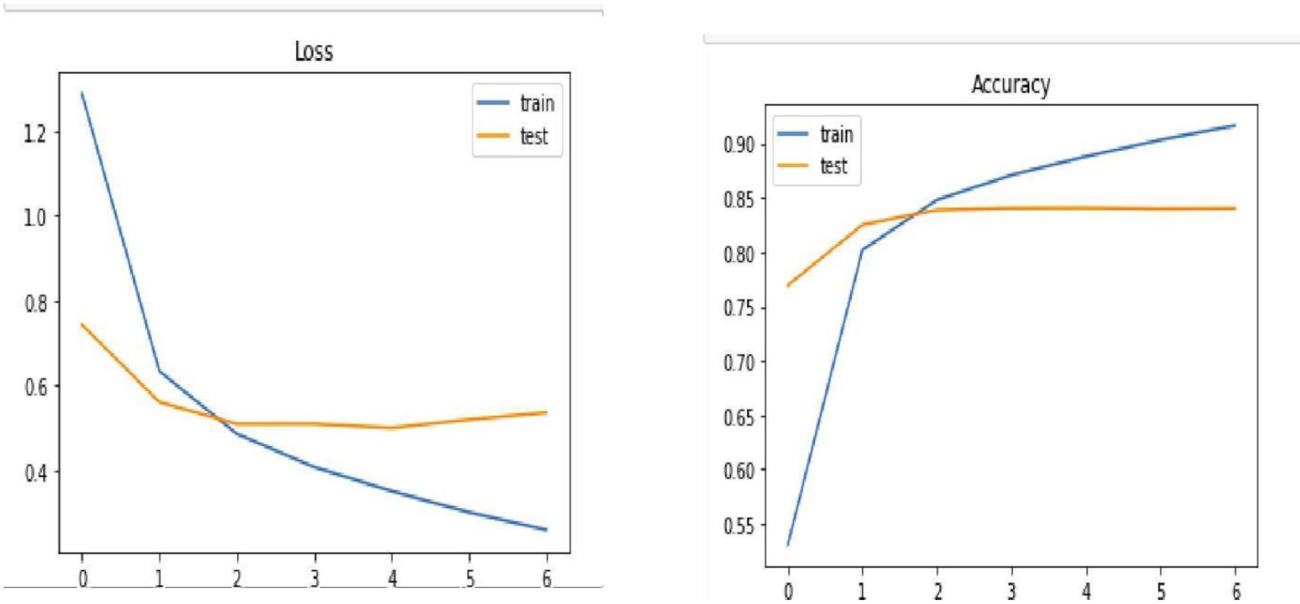


Fig 6.8.2

6.8.2 Bi-Directional LSTM

This model handles better dependencies than vanilla LSTM Model since it reads input from both left and right sides. Testing Accuracy obtained is 84.13%.

The below figure shows the training and testing results of Bi-directional LSTM.

```
Epoch 1/10
2025/2025 [==============================] - 11398s 6s/step - loss: 1.3383 - accuracy: 0.5165 - val_loss: 0.9436 -
val_accuracy: 0.6826
Epoch 2/10
2025/2025 [==============================] - 11393s 6s/step - loss: 0.7637 - accuracy: 0.7632 - val_loss: 0.6107 -
val_accuracy: 0.8118
Epoch 3/10
2025/2025 [==============================] - 11332s 6s/step - loss: 0.5167 - accuracy: 0.8420 - val_loss: 0.5267 -
val_accuracy: 0.8355
Epoch 4/10
2025/2025 [==============================] - 11394s 6s/step - loss: 0.4098 - accuracy: 0.8722 - val_loss: 0.5145 -
val_accuracy: 0.8354
Epoch 5/10
2025/2025 [==============================] - 11389s 6s/step - loss: 0.3341 - accuracy: 0.8953 - val_loss: 0.5198 -
val_accuracy: 0.8398
Epoch 6/10
2025/2025 [==============================] - 11392s 6s/step - loss: 0.2691 - accuracy: 0.9137 - val_loss: 0.5464 -
val_accuracy: 0.8388
```

```
In [34]: accu=model.evaluate(X_test,y_test)
         print('Test Set Loss',accu[0])
         print('Test Set Accuracy',accu[1])

1125/1125 [==============================] - 754s 664ms/step - loss: 0.5447 - accuracy: 0.8413
Test Set Loss 0.5447068810462952
Test Set Accuracy 0.8413055539131165
```

6.8.3 *Word Level Convolutional Neural Networks*

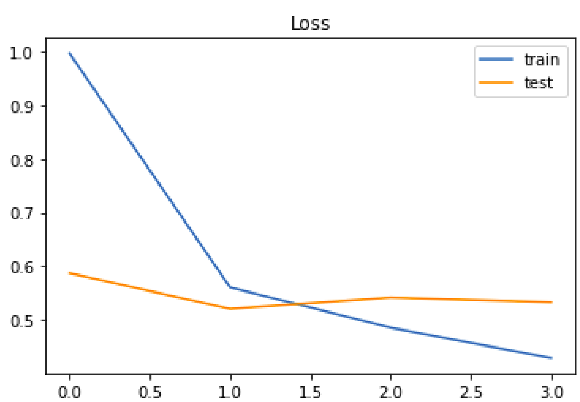
Two Convolutional layers with 128 filters each has been used with kernel size of 5. Activation function used in convolutional layers in relu. Convolutional layers are followed by Maxpooling layer with pool size of 5 along with dropout rate of 30%. The rest of the model is trained similar to LSTM Model. Testing accuracy achieved is 84.25%

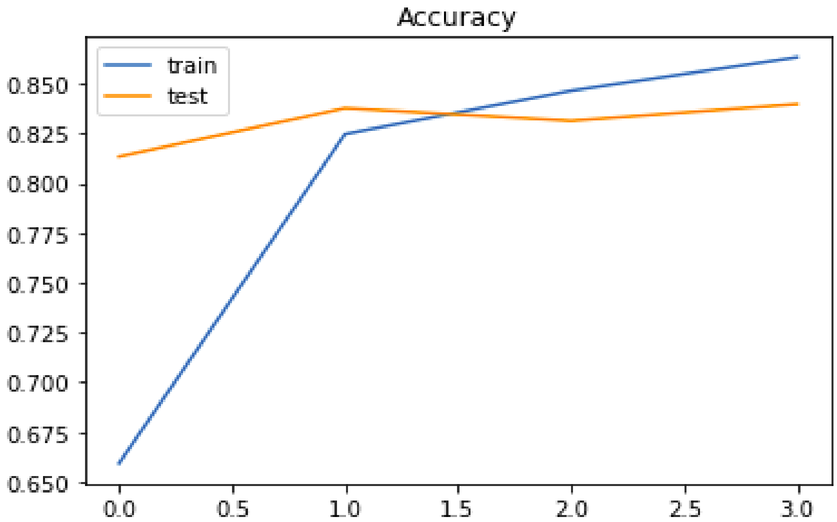
```
Epoch 1/10
2025/2025 [==============================] - 3793s 2s/step - loss: 1.3701 - accuracy: 0.5156 - val_loss: 0.5764 -
val_accuracy: 0.8181
Epoch 2/10
2025/2025 [==============================] - 3756s 2s/step - loss: 0.5408 - accuracy: 0.8305 - val_loss: 0.5399 -
val_accuracy: 0.8292
Epoch 3/10
2025/2025 [==============================] - 3735s 2s/step - loss: 0.4535 - accuracy: 0.8570 - val_loss: 0.5223 -
val_accuracy: 0.8345
Epoch 4/10
2025/2025 [==============================] - 3731s 2s/step - loss: 0.3913 - accuracy: 0.8747 - val_loss: 0.5148 -
val_accuracy: 0.8410
Epoch 5/10
2025/2025 [==============================] - 3748s 2s/step - loss: 0.3449 - accuracy: 0.8869 - val_loss: 0.5384 -
val_accuracy: 0.8371
Epoch 6/10
2025/2025 [==============================] - 3753s 2s/step - loss: 0.3050 - accuracy: 0.9016 - val_loss: 0.5436 -
val_accuracy: 0.8406
```

```
In [33]: accu=model.evaluate(X_test,y_test)
         print('Test Set Loss',accu[0])
         print('Test Set Accuracy',accu[1])
```

```
1125/1125 [==============================] - 216s 190ms/step - loss: 0.5487 - accuracy: 0.8426
Test Set Loss 0.5486834645271301
Test Set Accuracy 0.8425833582878113
```

The above figure 6.8.3 shows the training and testing results of CNN.
The below figure 6.8.4 shows the loss and accuracy graphs varying with epochs for CNN.





6.9 VOTING CLASSIFIER

We then implemented Hard Voting Classification of high accuracy models. The following models along with their respective accuracies are taken:-
1.Random Forest:-84.78%
2.XG Boost:-84.68%
3.Linear SVM:-83.51%
After implementing the following classifier gave an accuracy of 85.91%.

```
Out[47]: 0.8591666666666666
```

Bi-gram Linear SVM, Random Forest and XG Boost Hard Voting Classification model gave better accuracy but models were complex and take much time to process a complaint in real life.

To solve this issue we sampled 1,00,000 complaints from dateset and built a Voting classifier model as earlier but this time we used Multinomial Naive Bayes, Linear SVM and Logistic Regression as they were giving good accuracy as individual models and processing time also very quick.

Hard Voting Classification model gave accuracy of 82.46% which is a good accuracy and time taken is also very less.

6.10 USER INTERFACE

This user interface has been built using Flask.

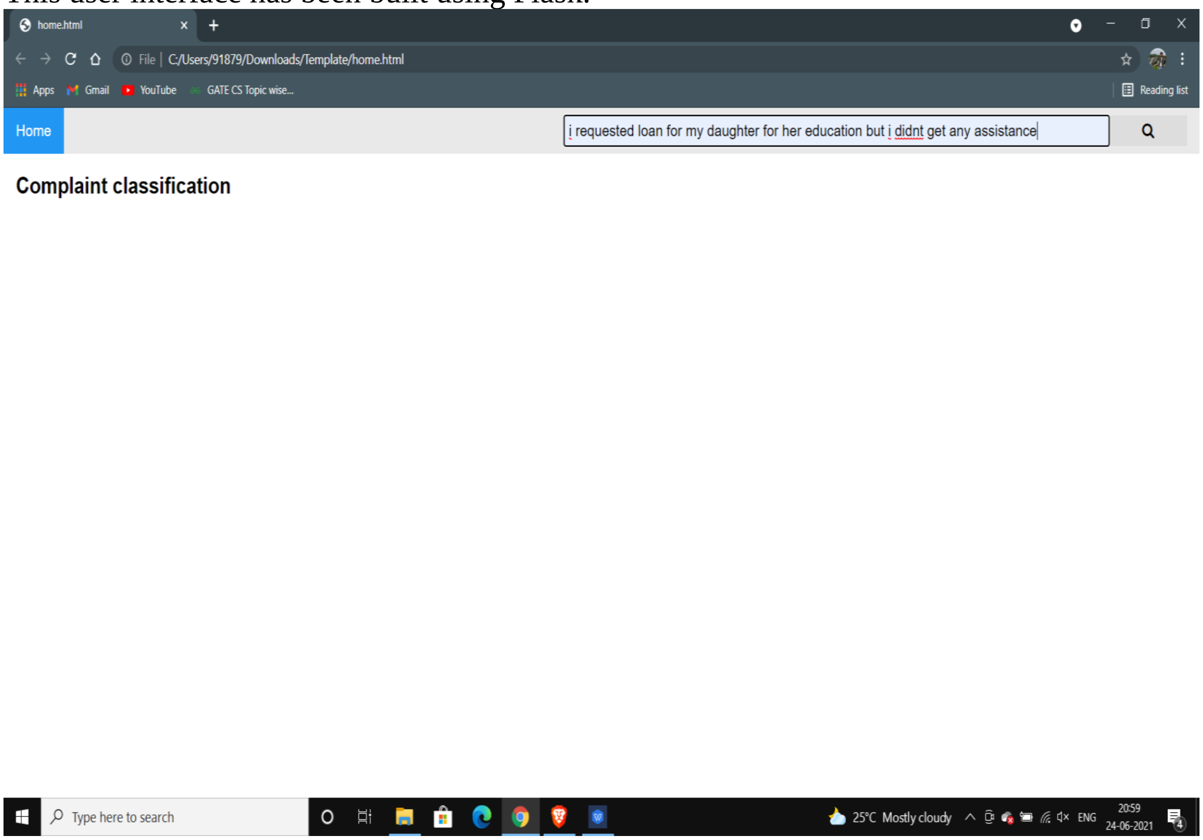


Fig 6.10

6.11 DATABASE

We used python sqlite3 database for project to store user complaints and results produced by classifier model.The database results can be used to check how the model is performing and to further improve the results.We also created a function to view the database

This is database view:

```
C:\Users\91879\Downloads\Template>python view.py
---------------------------------------------
I need education loan for my daughter
Student loan
---------------------------------------------
i am facing NEFT problem
Money transfer, virtual currency, or money service
---------------------------------------------
My car loan is pending
Vehicle loan or lease
---------------------------------------------
I forgot my ATM pin
Checking or savings account
---------------------------------------------
My credit card is not working.
Credit card or prepaid card
---------------------------------------------
```

Fig 6.11

7 CONCLUSION AND FUTURE WORK

Through our study we came to know that bi-grams performed better than uni-gram but increased model complexity. Hard voting method using Random forest, XG boost and linear SVM with bi-grams gave a higher accuracy of 85.9% but takes more time to build model and process result due to huge number of parameters. LSTM and SVM are more practical as they gave accuracy around 84% and took less time compare to voting method. Sampling of complaints and building a hard voting classifier (*uni-gram*) with linear SVM, NB and Logistic Regression gave acceptable accuracy of 82.4% and process complaints much faster in real time. Decision trees gave the least accuracy among all models. The choice of the model is based on user need based on speed and accuracy. The future work could be improvising speed while keeping accuracy as high possible and building fully working software for the finance companies.

8 REFERENCES

1. Mohit Rathore, Dikshanth Gupta, Dinabandhu Bhandari,"Complaint Classification using Word2Vec Model",International Journal of Engineering & Technology, 7 (4.5) (2018) 402-404 [1].
2. Xuesong Tong, Bin Wu, Shuyang Wang, Jinna Lv,"A Complaint Text Classification Model Based on Character- level Convolutional Network",978-1-5386-6565- 7/18/$31.00 (f)2018 IEEE [2].
3. Anditya, Dhamir, Dian, Suyanto, Irfan, Anis, Fedy, Rani,"Developing an LSTM based Classification Model of IndiHome Customer Feedbacks",2020 International Conference on Data Science and Its Applications [3].
4. M. Ali Fauzi,"Automatic Complaint Classification System Using Classifier Ensembles",Telfor Journal, Vol. 10, No. 2, 2018 [4].
5. Pavels Goncarovs,"Active Learning SVM Classification Algorithm for Complaints Management Process Automatization",978-1-7281-5709-2/19/$31.00 ©2019 IEEE [5].
6. Chandana C,Neelashree,Nikitha G N,Nisargapriya J,Vishwesh,"Bank Customer Complaints Analysis Using Natural Language Processing and Data Mining",International Journal of Progressive Researchin Science and Engineering Volume-1, Issue-3, June-2020 [6].
7. Gutha Jaya Krishna , Vadlamani Ravi, Bheemidi Vikram Reddy, Harshal Jaiswal,"Sentiment Classification of Indian Banks'Customer Complaints",978-1-7281-1895-6/19/$31.00©2019 IEEE [7].