Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225
ISSN: 2319-4847

Consumer Complaints Classification using Deep Learning & Word Embedding Models

Vadithya Vinod[1], Kalwakumtla Sathwika [2], Jarupla Pradeep Nayak [3], T.Bhargavi[4]

[1,2,3] UG Scholar, Department of IT, St. Martin's Engineering College, Secunderabad, Telangana,India – 500100

[4]Assistant Professor, Department of IT, St. Martin's Engineering College, Secunderabad, Telangana ,India– 500100

vadithyavinod267@gmail.com

Article Info

Received: 28-03-2025 ***Revised: 05 -04-2025*** ***Accepted: 16-04-2025*** ***Published:27/04/2025***

Abstract

The goal of Text classification is to categorize a document into predefined categories. Various supervised and unsupervised classifiers can be used to achieve this. In the research work proposed, SOTA (State of the Art) deep learning models and embedding techniques are used to classify consumers' complaints, which is in form of text, into 6 classes. Words and documents are represented as numeric vectors through embedding, enabling vector representations for related words. These representations are easily ingested by NLP algorithms. The classes represent departments where complaints are routed. Deep learning models like LSTM, Bi- LSTM, GRU and 1D CNN are used, along with word embedding techniques like Word2Vvec, Fasttext, Bert and Distilbert to represent text. The experimental results indicate that DistilBert and CNN achieved a 93% F-score.

Keywords— Transfer Learning, Text Classification, Word Embedding, Deep Learning

1. INTRODUCTION

The realm of Natural Language Processing (NLP) is situated at the intersection of computer science and computational linguistics, and is concerned with empowering machines to comprehend and produce human language. NLP leverages advanced algorithms and statistical models to extract meaningful insights from language data, and facilitates tasks that would typically necessitate human-level intelligence. By utilizing NLP, machines are capable of processing language in various formats, including text, audio, and video. NLP allows computers to process human language in text, audio, and video formats. NLP has numerous applications such as search engines, question answering, text summarization, sentiment analysis and machine translation [2].

Consumer Financial Protection Bureau (CFPB) is a US federal Organization which serves as negotiator When feud arises between financial institutions and consumers, Customers can send the organization a description of their dispute via an online form. The complaint is then manually tagged and send to the respective department. The current process takes 10-15 days, before it reaches to its right department. The proposed work introduces a model Which would take help of NLP techniques, Machine learning and Deep learning models to automatically classify a complaint without the need of human intervention.

The gathering and analysis of complaints is a crucial and required step in enhancing an organization's services [3]. The most common method for gathering complaints online is

through websites. The caller's complaint is recorded a web form and filed there. After analysis, the texts are given to several departments for potential resolution. In order to help the organization, resolve the problems as quickly as possible, it has become necessary to automatically classify the complaints into a number of established categories [4]. This would help in saving work and quick customer response and resolution of customer complaints. The problem can be approached through the methodology used in classical case of sentiment analysis [5]. The narratives are first converted to vector using several word embedding techniques and then a classifier is used to classify the complaint. State of the art word Embedding technique would help in Vectorizing the texts of complaints, so that they could be linearly separable in the hyperplane. Several Deep Learning classifiers could be employed further to predict the class of the complaint to which it originally belongs.

2. LITERATURE SURVEY

In [6], the authors conducted a survey on sentiment recognition methods, which is important in opinion mining and sentiment analysis. They suggested an MSCNN-BiGRU model for classifying text sentiment that extracts semantic features and incorporates text context data using multi-scale convolution kernels and BiGRU. The model was tested on Chinese and English e-commerce datasets and outperformed the combined CNN and RNN model by 3.5%. Recurrent neural networks, particularly LSTM, are commonly used in sentiment analysis due to their ability to extract features and handle long-term relationships. The model's performance was evaluated using various measures such as accuracy, F1 score, confusion matrix, and ROC curve.

In [7], authors have implemented BERT based transformer architecture. Research shows the benefits and drawbacks of the chosen models. The models presented include the Transformer-XL Cross-lingual Language Models (XLM), the Generative Pre-training (GPT) and its derivatives, and the Bidirectional Encoder Representations from Transformers (BERT). The study analyses recent studies in which researchers presented alternative BERT-based models, taking into account BERT's power and popularity in text-based emotion detection.

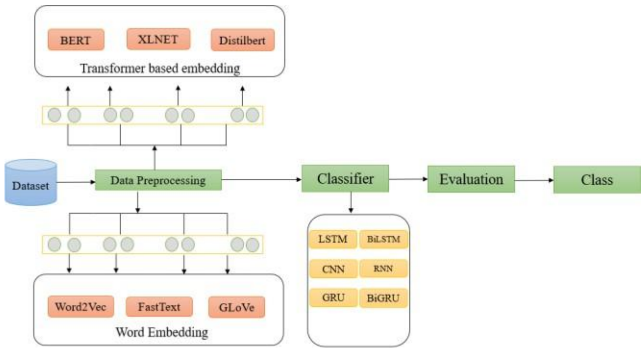
In [8], authors have conducted experiments related to sentiment analysis on Online reviews. The reviews from a car review

website. Rather than directly extracting the emotional inclination from comments, the application of a Conditional Random Forest model can be leveraged to identify emotional feature segments, which can subsequently be utilized to extract the emotional tendency. This approach can lead to an
increase in the classification accuracy. Given the differences between Chinese and English writing styles, the TF-IDF assigns greater weight to frequent characteristics, which improves classification differentiation. Additionally, in the majority of the comments, words indicating positive emotions are more likely to appear in phrases with positive emotions than negative emotions, and vice versa.

The authors go into great depth on BERT embedding model in [9]. A novel approach to multi-class text emotion identification has been devised, which employs an LSTM- CNN dual-channel system. Using a BERT model that has already been trained, the approach entails extracting embedding vectors from input phrases. This approach has been thoroughly researched and tested, and has shown promising results in accurately identifying emotions in text. BERT's ability to create high-quality sentences depends on two factors: its deep bidirectional language model and the label-segment embedding that provides label information. By leveraging these capabilities, BERT is capable of generating sentences that exhibit exceptional quality and fluency, which can be used for a variety of applications in natural language processing. A methodology has been developed to evaluate the system's proposed computation costs in terms of both a single value and the Big-O values of each of its constituent parts.

In [10], researchers have used NLP for text classification for Hindi sentence categorization. Fast text word vectors trained on a Hindi corpus and random word vectors were used to initialise the word vectors. CNN models perform better on the datasets utilized in this work than LSTM-based models. Due to the flexible word order in Hindi, LSTMs are no more effective than CNNs. The LASER multilingual model recorded richer sentence representations as compared to BERT. Lightweight models that just employed phrase encodings, however, typically underperformed models that were expressly trained on certain datasets.

suggests a novel attention-based BiLSTM text classification model that combined CNN with a gating mechanism (ABLG-CNN). ABLG-CNN employs Word2vec for training word vector representations. To determine keyword information, the attention mechanism calculates the word context vector. The context characteristics are captured by the bidirectional long short-term memory network (BiLSTM). Convolutional neural



network (CNN) collects topic relevant characteristics based on this. Long texts may contain phrases that refer to various topics, to create text fusion features that are optimal for classification, a gating mechanism is employed. This mechanism assigns weights to the output features of both the BiLSTM and CNN, thereby combining them into a single, cohesive feature vector. By using this technique, it is possible to achieve higher accuracy and better performance in various natural language processing tasks. Through experimental verification on two large text news datasets, ABLG-CNN has demonstrated its ability to capture both the contextual semantics and local phrase characteristics of text. This approach has proven to be highly effective in extracting meaningful insights from text data, and can be applied to a range of natural language processing tasks with great success.

Discusses the Glove & word2Vec Embedding in depth & employs the techniques of supervised learning models from Machine Learning domain. The dataset used is from Cancer patients & with the help of word embedding techniques we are converting the text of medical reports into vectors & later classifying the samples of patients into cancerous & non-cancerous classes. Log loss, Precision & recall are used as parameters to evaluate the performance of the model.

In [13], authors have discussed the Glove & word2Vec Embedding in depth & employs the techniques of supervised learning models from Deep Learning domain for detection of Cyber Bullying with the help of Text Classification. In-depth descriptions of several feature selection strategies, sentiment classification algorithms, and this research study presents deep learning methods for analyzing sentiment.. This study also addresses other sentiment categorization methods, including lexicon-based and machine learning-based approaches. For classifying sentiment, a variety of classification techniques like SVM, Decision Trees are employed. Deep learning models for sentiment analysis, including RNN, 1DCNN, and LSTM, are also covered in this study.

This research work encompasses all the techniques covered in this section and applies a combination of deep learning models along with word embeddings to classify consumer complaints. Since, very few research work has been done on this dataset, the following section discusses SOTA techniques applied on the dataset, which defines the novelty of the research work.

PROPOSED METHODOLOGY

The objective of this study is to accurately categorize consumer complaints into six distinct classes: Debt Collection, Mortgage, Credit Reporting, Savings Account, Student Loan, and Money Transfer. To achieve this, a comprehensive framework has been designed, as illustrated in Figure 1, which outlines the step-by-step process for classifying consumer complaints based on their textual content. The framework begins with the data collection and preprocessing stage, where raw textual data is gathered and cleaned to improve its quality. This step involves removing special characters, stopwords, and redundant information, as well as applying tokenization and text normalization to standardize the content. Techniques such as stemming or lemmatization may

International Journal of Application or Innovation in Engineering and Management
Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225
ISSN: 2319-4847
also be employed to reduce words to their root forms, ensuring consistency within the dataset. Following this, the processed text is fed into the word embedding layer, which converts the textual data into numerical representations that machine learning models can efficiently process. Word embeddings such as Word2Vec, GloVe, or BERT are often used to capture semantic relationships and contextual meanings within the text data, enhancing the model's ability to understand language patterns and improving classification accuracy.

Fig 1: Proposed Model block Diagram

A. Dataset

The Consumer Complaints Classification Dataset from the Consumer Financial Protection Bureau (CFPB) is a comprehensive dataset designed for text classification tasks, particularly in consumer complaint analysis. It comprises 75,000 observations in the training corpus and 8,000 observations in the testing corpus, making it suitable for robust model development and evaluation. This dataset is categorized into six distinct classes: Debt Collection, Mortgage, Credit Reporting, Savings Account, Student Loan, and Money Transfer. Each observation typically contains a detailed complaint description, a product or service category as the target label, and may include additional metadata such as the company name, date, and resolution status. The dataset offers valuable opportunities for various applications, including text classification, sentiment analysis, topic modeling, and fraud detection. By employing preprocessing techniques such as text cleaning and feature engineering methods like TF-IDF, Word2Vec, or BERT embeddings, practitioners can enhance model performance. Popular models such as Logistic Regression, Random Forest, XGBoost, or advanced deep learning architectures like LSTM and Transformer models are well-suited for achieving accurate classification results with this dataset. All the observations of the dataset fall under any one of the following 6 classes:

Debt Collection
Mortgage
Credit Reporting
Savings Account
Student Loan
Money Transfer

The unbalanced distribution of classes is seen in Figure 2.

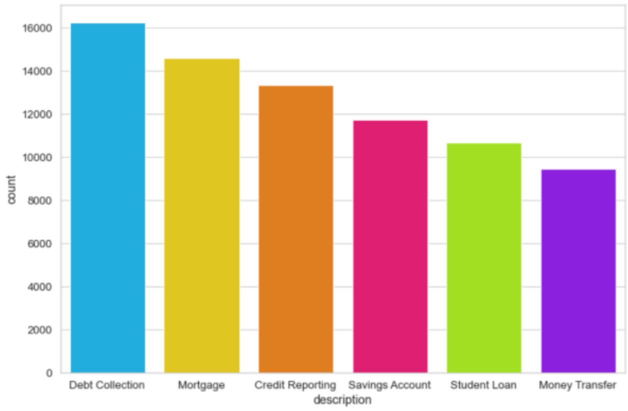


Fig 2. Class Distribution

To achieve the most precise outcomes, it is necessary to scrutinize the dataset using visualization approaches, such as assessing the frequency of word count and N-gram distribution, along with determining the maximum length of text. As demonstrated in Figure 3, for Complaint text, the maximum text length is 550 characters. These insights can assist in optimizing the model's training by adjusting the hyper parameters appropriately, leading to an improvement in accuracy.

International Journal of application or innovation in engineering and management

Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225
ISSN: 2319-4847

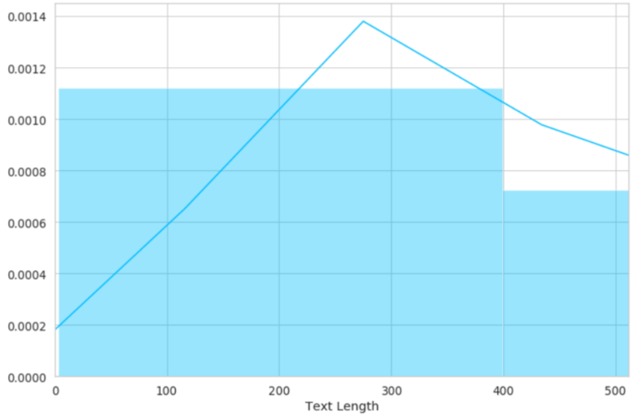


Fig 3. Length Distribution of Complaint Text

B. Implementation Environment

The programming language Python 3 was used to conduct this study on Google Colab. The implementation utilized hardware accelerators such as Graphical Processing Units (GPUs) and Tensorflow Processing Units (TPUs) that were provided by Colab.

Word Embeddings

To enhance the likeness among associated words, the word embedding method necessitates transforming text-based data into a vector depiction that captures the sense of the term. This study used six different word embedding approaches, including BERT, GLoVe, Word2Vec, FastText, DistilBert and XLNet to convert complaint text into real-valued vectors. [14].

In order to categorize the complaints, these word embeddings were subsequently fed into deep learning models. The word embeddings were acquired for this investigation using the Gensim, Hugging Face Transformer, and Ktrain TensorFlow lightweight wrapper packages.

Word2Vec is a well-known technique for producing word embeddings (NLP). It was introduced in 2013 by Tomas Mikolov and his team at Google. Word2vec learns to provide a word's vector representation using words from a huge corpus of literary works as input. The word2vec technique derives properties from the text for certain words, much as how CNNs analyze pictures. Word2vec generates vectors that represent words in the vector domain using these characteristics. The cosine similarity function, which highlights the semantic similarity between words, is used to pick these vectors.

Glove is an acronym that stands for Global Vectors for Word Representation. It creates embeddings based on the principle of co-occurrence of the words. These embeddings are created by putting comparable words in the same vector space. To put it another way, both positive and negative words will tend to cluster together

FastText embeddings was built by Facebook. Facebook expanded the Word2Vec concept by creating this pretrained model. The sub-structure or sub-words are used by the FastText embeddings to enhance the vector representation. The target word serves as the source of character n-grams in FastText embeddings, which are subsequently fed to the Word2Vec Skip-Gram model to produce the embeddings. FastText embeddings have the major benefit of being able to quickly produce embeddings for large amounts of data. This embedding require relatively little training time in order to build them. The FastText embeddings are implemented in the suggested system using the Python module Ktrain.

BERT Embedding supports the bi-directional encoder representations used by the Transformers. It's a pre- trained NLP model created by Google that anybody may use to train their own question-answering module for several hours on a single GPU. The company then showed the outcomes of 11 NLP jobs, including inquiries from the very competitive Stanford dataset. BERT, according to researchers, has effectively replaced all previous language models with its 93% accuracy when utilized as a pre-trained model in a deep neural network. BERT has only been pre-trained on 800 million words from a book corpus and 250 million words from Wikipedia, in contrast to other language models. BERT is superior to other standard LMs because it applies deep bidirectional context training of the sequence, which takes into account both left and right context while training, as opposed to other LM models like OpenAI GPT, which are unidirectional and limit each token's ability to attend to previous tokens in attention layers.

XLNet is a state-of-the-art pre-training method for natural language processing (NLP) tasks that outperforms previous models. It is based on the transformer architecture and uses an autoregressive model to generate text. Unlike other models, XLNet trains on all possible permutations of the input sequence,

making it more effective in handling bidirectional context. This allows it to capture long-range dependencies and context better than previous models. XLNet has been shown to achieve state-of-the-art results in various NLP tasks, including sentiment analysis, machine translation, and question answering.

DistilBERT is a smaller and faster version of BERT, a popular pre-training method for NLP tasks. It achieves this by distilling the knowledge from the larger BERT model into a smaller model while maintaining similar performance. Training the smaller model to imitate the bigger model's behaviour is a step in the distillation process. DistilBERT is trained using two techniques: distillation and parameter pruning. This allows it to have a smaller memory footprint and faster inference time than BERT. DistilBERT has been shown to be effective in various applications, including sentiment analysis, named entity recognition, and question answering.

Transfer Learning with word embedding Models

Transfer learning is an effective technique that leverages previously acquired knowledge from a related problem to address a new task, offering significant advantages in terms of efficiency and performance [15]. By utilizing a pre-trained model, this method helps bypass the often costly and time-consuming process of extensive data collection and model training from scratch, making it highly suitable for complex tasks requiring large datasets and computational resources [16]. In this study, several deep learning models were employed, including Gated Recurrent Units (GRU), Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), Bidirectional GRU (BiGRU), and Recurrent Neural Networks (RNN). Each of these models was used in combination with a word embedding layer as the input, which transformed the textual data into meaningful vector representations that capture semantic relationships between words [17][18]. The integration of these advanced classification models aims to effectively assign consumer grievances to the appropriate department class, ensuring that complaints are categorized correctly for efficient resolution. By leveraging transfer learning and robust deep learning architectures, this study enhances the accuracy and reliability of the complaint classification process, ultimately improving the consumer support system's responsiveness and effectiveness.

Model Configuration

The proposed system was developed and trained using a well-structured dataset, where the data was divided into training, validation, and test sets in a 60%, 20%, and 20% ratio, respectively, ensuring a balanced distribution for effective model training and evaluation. The maximum input sequence length for the pre-trained models was set to 300 characters, allowing the models to efficiently capture meaningful textual information while controlling computational overhead. For the Word2Vec model, a window size of 5 was chosen to define the context range for each target word, and a worker size of 4 was employed to enable parallel processing for improved training efficiency. In the case of the GloVe model, an embedding vector length of 300 dimensions was used to effectively capture semantic relationships within the text. Deep learning classifiers were trained over 15 epochs with a batch size of 32, ensuring stable learning convergence. A maximum pooling layer with a pool size of 2 was incorporated to reduce feature map dimensions while retaining key features. The LSTM and BiLSTM models were configured with 32 units each, while the GRU and BiGRU models employed 128 units, leveraging their enhanced efficiency in handling sequential data. The RNN model was designed with 100 units and included 300 convolutional filters to ensure robust feature extraction. All models utilized the Adam optimizer for adaptive learning rate adjustment, along with ReLU activation functions to introduce non-linearity for improved pattern learning. The sparse categorical cross-entropy loss function was applied to facilitate effective multi-class classification. Furthermore, the pre-trained transformer-based models were trained over 6 epochs with a dimension size of 768, a batch size of 6, and a learning rate of 1e-5, ensuring optimal fine-tuning for improved performance. For additional details on the experimental settings that were employed to achieve optimal results, refer to Table 1.

International Journal of application or innovation in engineering and management

Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225
ISSN: 2319-4847

Table 1. Details of Parameter tuning

Parameters	Values
Dataset split (train, test, validation)	60%-20%-20%
Max. sequence length	300
Embedding dimension	768
Learning rate	1e-5
Batch-size	32
Epochs	6, 15
Drop-out	0.5
Pool size	2
Filters	32
Kernel size	400
Optimizer	Adam
Loss function	Sparse categorical Cross entropy
Activation Function	ReLU
LSTM, BiLSTM Units	32
RNN	100
Window size	5
Worker size	4

Evaluation

To evaluate a classification task, its precision is measured by computing the number of correct predictions generated by the model. This metric determines the proportion of accurate predictions made by the model out of the total predictions made. Furthermore, the accuracy of a model can be improved by enhancing the features used to train it or by using more advanced algorithms to optimize its performance [19]. Further, in order to gauge the model's performance, a confusion matrix is created, which measures various performance metrics [20].

EXPERIMENTAL ANALYSIS

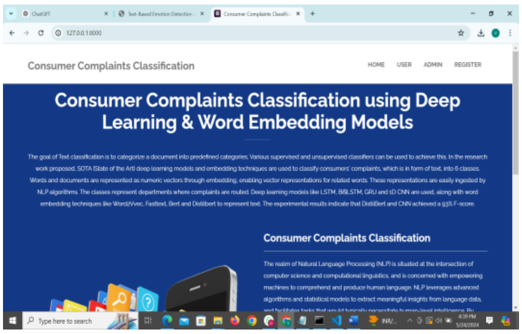


Fig 4.1: Home Page

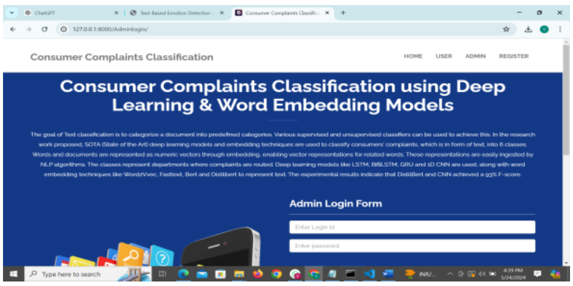


Fig 4.2: Admin Login Page

International Journal of application or innovation in engineering and management

Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225
ISSN: 2319-4847

Fig 4.3: Admin Home Page

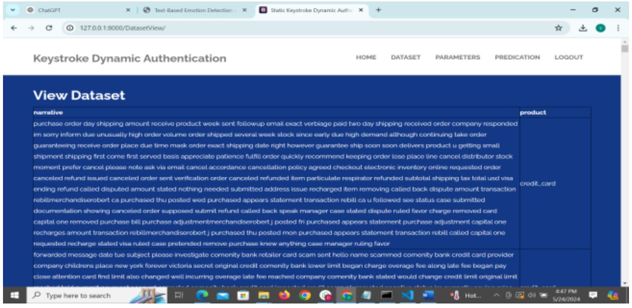


Fig 4.4: User List

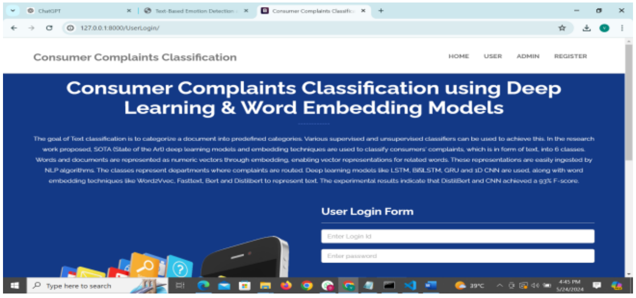


Fig 4.5: User Login

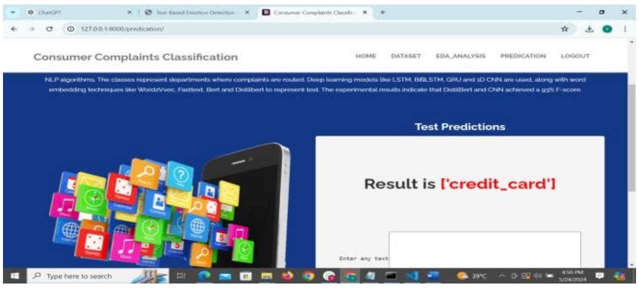
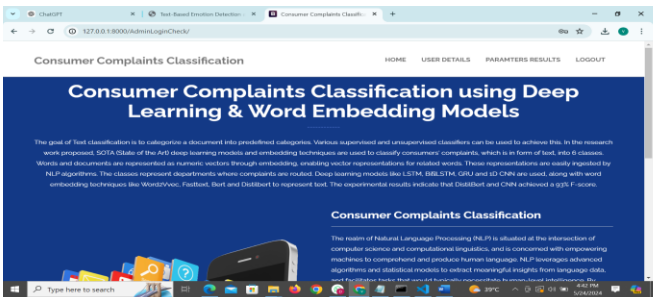


Fig 4.6: Data Set



International Journal of application or innovation in engineering and management

Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225
ISSN: 2319-4847

Fig 4.7: Predication

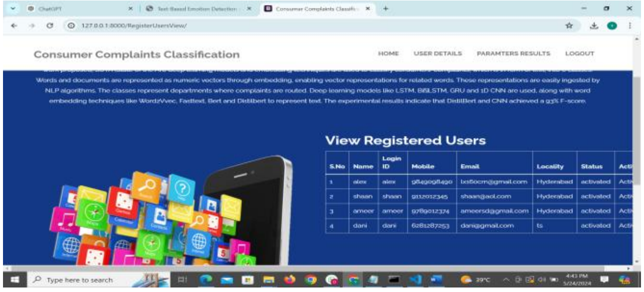
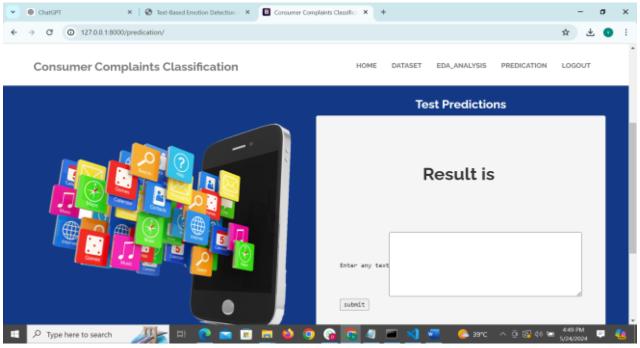


Fig 4.8: Result



CONCLUSION

This research paper focused on evaluating the effectiveness of different word embeddings and deep learning models for classifying consumer complaints in text form. The main goal was to develop a classification system capable of categorizing complaints into six distinct categories.

The results from the experiments indicated that the combination of Convolutional Neural Networks (CNN) and DistilBERT performed the best, achieving an accuracy rate of approximately 94.0% and an F-score of 93.0%. These results demonstrate the effectiveness of using CNN alongside pre-trained models like DistilBERT in handling text classification tasks, particularly in the context of consumer complaints.

The high accuracy and F-score show that this approach is well-suited for categorizing complaints efficiently, with minimal errors. It highlights the potential of deep learning techniques to improve consumer feedback management systems, providing businesses with a more reliable way to understand and address customer issues.

For future development, it is suggested to focus on enabling real-time complaint classification. This could significantly speed up the process of resolving consumer issues, providing quicker responses and enhancing customer satisfaction. Additionally, improvements in handling diverse, noisy, or unstructured complaint data could further enhance the model's performance.

In summary, the results from this study show that the CNN-DistilBERT combination is a strong candidate for consumer complaint classification. Future advancements in real-time processing and model refinement could lead to even more efficient systems for managing customer feedback.

REFERENCES

Manju Venugopalan and Deepa Gupta. An unsupervised hierarchical rule based model for aspect term extraction augmented with pruning strategies. Procedia Computer Science, 171:22–31, 2020.

Ashok Kumar, Erik Cambria, and Tina Esther Trueman. Transformerbased bidirectional encoder representations for emotion detection from text. In 2021 IEEE Symposium Series on Computational Intelligence (SSCI), pages 1–6. IEEE, 2021.

R. Y. Kim, "Using Online Reviews for Customer Sentiment Analysis," in *IEEE Engineering Management Review*, vol. 49, no. 4, pp. 162-168, 1 Fourthquarter,Dec. 2021.

Y. Zhou and S. Yang, "Roles of Review Numerical and Textual Characteristics on Review Helpfulness Across Three Different Types of Reviews," in *IEEE Access*, vol. 7, pp. 27769-27780, 2019.

Pansy Nandwani and Rupali Verma. A review on sentiment analysis and emotion detection from text. Social Network Analysis and Mining, 11(1):1–19, 2021.

İlkay Yelmen, Ali Güneş, Mehmet Zontul, Zafer Aslan, "Multi-Class Text Classification Based On Deep Neural Network And Word2vec", Journal of Aeronautics and Space Technologies, Vol. 15, No. 1, pp. 59- 65, Jan. 2022.

Mehta, Manan et al. “Hindi Text Classification: A Review.” 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N) (2021): 839-843.
Jianfeng Deng, Lianglun Cheng, Zhuowei Wang,Attention-based BiLSTM fused CNN with gating mechanism model for Chinese long text classification,Computer Speech & Language,Vol- ume 68,2021,101182,ISSN 0885-2308.
Veena P V, Anand Kumar M, Soman K P, “An Ef- fective way of word- level Language Identification for code mixed FAacebook comments using Word- Embedding via 35 Character-Embedding”, Interna- tional Conference on Advances in Computing.
Hickman, Louis & Thapa, Stuti & Tay, Louis & Cao, Mengyang & Srinivasan, Pad- mini. (2020). Text Preprocessing for Text Mining in Organizational Research: Review and Recommendations In-press at Organi- zational Research Methods. Organizational Research Methods. in press.
S Lalitha, Shikha Tripathi, and Deepa Gupta. Enhanced speech emotion detectionusing deep neural networks. International Journal of Speech Technology, 22(3):497– 510, 2019.
C. V. Amrutha, C. Jyotsna and J. Amudha, "Deep Learning Approach for Suspicious Activity Detection from Surveillance Video," 2020 2nd International Conference on Innovative Mechanisms for Industry Ap- plications (ICIMIA), Bangalore, India, 2020,pp. 335-339.
P. M. Kiritbhai and J. C, "An Automated University Gate Pass Monitoring System Using Deep Learning," 2021 International Conference on Communication, Control and Information Sciences (ICCISc), Idukki, In- dia, 2021, pp. 1-6.
Manju Venugopalan and Deepa Gupta. An enhanced guided lda model augmented with bert based semantic strength for aspect term extraction in sentiment analysis. Knowledg based Systems 246:108668, 2022.
Siddharth Karanchery and Suja Pal- aniswamy. Emotion recognition using one- shot learning for human-computer interac- tions. In 2021 International Conference on Communication, Control and Information Sciences (ICCISc), volume 1, pages 1–8,2021.
Wei Shiung Liew, Chu Kiong Loo, and Stefan Wermter. Emotion recognition using explainable genetically optimized fuzzy art ensembles. IEEE Access, 9:61513–61531, 2021.
M. Ohsaki, P. Wang, K. Matsuda, S. Kata- giri, H. Watanabe and A. Ralescu, "Confu- sion-Matrix-Based Kernel Logistic Regres- sion for Imbalanced Data Classification," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 9, pp. 1806-1819, 1 Sept. 2017