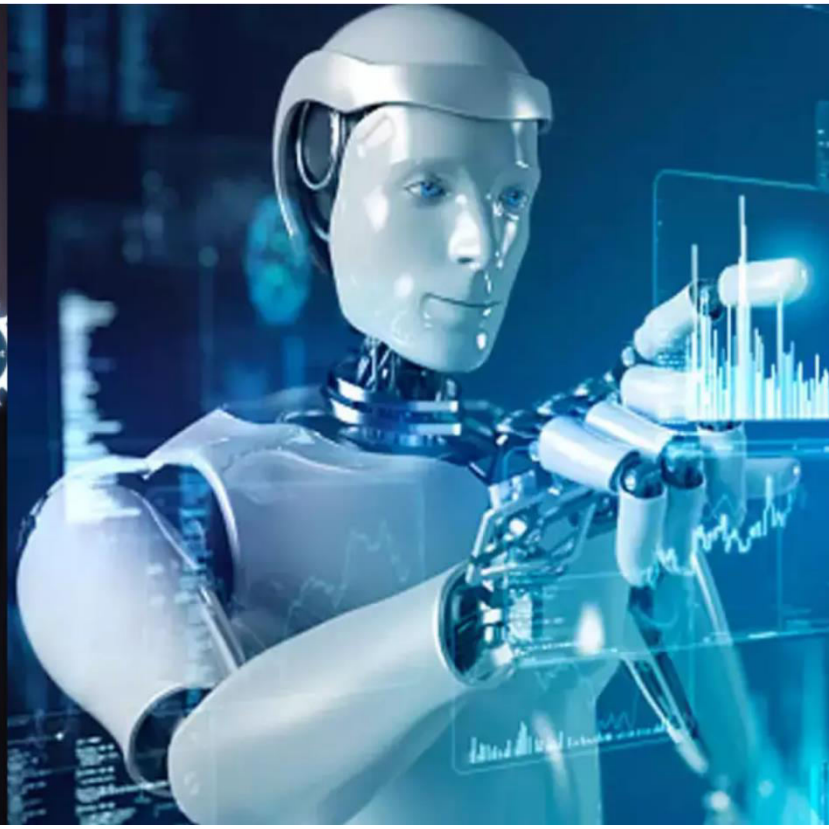
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 9, September 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 9, September 2025

|DOI: 10.15680/IJIRSET.2025.1409036|

Deep Learning and NLP for Enhanced Customer Complaint Automation in Finance Institutions

Subhash Mokase[1], Vijaykumar M. P.[2]

Shreeyash College of Engineering and Technology, Ch. Sambhajinagar, Maharastra, India[1, 2]

ABSTRACT: Financial institutions handle thousands of customer complaints daily, often requiring manual effort to categorize and route these complaints. This paper presents a deep learning and NLP-driven solution for automating the classification of customer complaints into predefined categories. We evaluate traditional deep learning models against modern NLP architectures, including LSTM and BERT, using the Consumer Financial Protection Bureau complaint dataset. The BERT-based classifier achieved an accuracy of 94.3% significantly outperforming classical models. Our approach enables real-time, scalable complaint classification, offering enhanced operational efficiency for financial service providers.

KEYWORDS: Artificial intelligence, Machine Learning, Customer Complaints, BERT, Natural Language Processing, Multi-Class Classification, LSTM, Financial Institutions, Text Classification, Complaint Routing, Imbalanced Data.

I. INTRODUCTION

Effectively handling consumer complaints is crucial for preserving customer satisfaction, operational efficiency, and regulatory compliance in financial organizations including banks, credit bureaus, and lenders. These complaints cover a broad spectrum of topics, including account access, credit card and loan disputes, and service-related concerns, making manual classification and routing laborious and prone to mistakes. In order to overcome these obstacles, this study explores the automation of complaint classification through the use of deep learning and Natural Language Processing (NLP).In particular, we create and assess an automated system that makes use of transformer-based models such as BERT, contrasting its effectiveness with that of recurrent neural network (LSTM) and conventional machine learning techniques. Our research entails optimizing BERT using actual financial complaint data and carrying out a thorough performance evaluation of all models. In order to facilitate real-time routing, the suggested method seeks to precisely classify complaints into Predetermined departmental classes. and greatly enhancing financial organizations' overall service efficiency, accuracy, and reaction times.

II. LITERATURE SURVEY

Machine learning (ML) algorithms have shown increasing efficiency in automating the classification of customer complaints, according to prior studies. Earlier methods frequently used traditional machine learning techniques like Support Vector Machines (SVM), Random Forests, and Naive Bayes. Although somewhat successful, these models commonly use sparse representations such as TF-IDF or Bag-of-Words and mainly rely on manufactured features. Because of this, they are unable to comprehend in-depth semantic links and long-range dependencies in the text, which are both essential for correctly categorizing complex complaints in the financial industry.

Researchers started using deep learning methods, especially Recurrent Neural Networks (RNNs) and their more sophisticated counterpart, Long Short-Term Memory (LSTM) networks, to get beyond these restrictions. By remembering past word states throughout a sentence or paragraph, these models are intended to represent the sequential pattern of language and improve contextual understanding. According to studies, LSTMs perform noticeably better than standard models in a variety of natural language processing (NLP) tasks, particularly those like sentiment analysis, intent detection, and complaint classification where word order and context are crucial. Instead of depending only on sequential data processing, transformer-based models like BERT (Bidirectional Encoder Representations from Transformers) have lately transformed natural language processing by integrating mechanisms for global context attention. In complicated fields like finance, where interpreting nuanced linguistic clues is crucial, BERT's ability to comprehend both the left and right context at the same time has allowed it to achieve state-of-the-art performance in a variety of text categorization tasks. Numerous domain-specific research back up these advancements. When LSTM was

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 9, September 2025

|DOI: 10.15680/IJIRSET.2025.1409036|

used to classify banking complaints, for instance, study showed that it was highly accurate in identifying and grouping problems into several categories. Other research has demonstrated that even more reliable and accurate classification systems can be achieved by fine-tuning pre-trained BERT models on financial text datasets. The industry's ongoing quest for increased precision, contextual awareness, and scalability in automating customer service processes is reflected in the transition from conventional machine learning to deep learning and, more recently, transformer-based models.

2.1 Problem Statement

Every day, financial institutions receive a significant number of unstructured client complaints pertaining to a variety of topics, including customer service, loan servicing, credit reporting, and account security. To guarantee prompt and efficient treatment, these complaints must be appropriately divided into designated departments. This process is currently done by hand, which makes it sluggish, irregular, and prone to mistakes especially when the number of complaints rises. Additionally, this classification process is extremely hard due to the imbalance of class representation (some departments receive considerably more complaints than others), contextual nuances, and linguistic diversity. Consequently, the main challenge is to create a scalable, automatic, and accurate system that can categorize incoming complaints into a number of departments based just on the complaint language. This issue can be formulated as a multi-class text classification task that needs to work in real-time, handle unbalanced input, and retain high accuracy.

2.2 Proposed System

The need for financial institutions to improve their customer complaint handling processes is growing as customer expectations for quicker, more accurate service continue to rise. Due to the increasing volume and complexity of complaints, which frequently need for prompt and accurate classification and resolution, traditional manual processes are no longer adequate. This project suggests using Natural Language Processing (NLP) and Deep Learning to create an automated, intelligent system that can comprehend, categorize, and handle client complaints in real time. Utilizing cutting-edge models like LSTM and BERT, the system seeks to increase complaint management's overall effectiveness, decrease human error, and dramatically speed up response times. The ultimate objective is to improve financial institutions' operational procedures and raise client satisfaction.

III. METHODOLOGY (SYSTEM WORKFLOW)

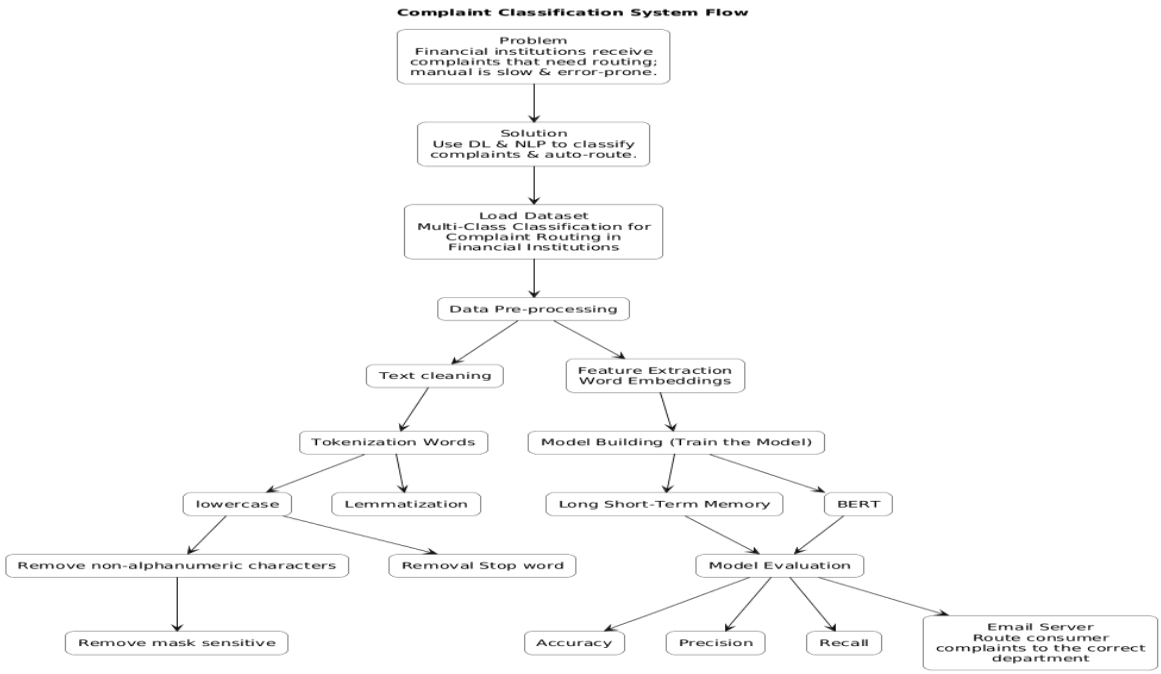


Fig 3.1: System Workflow

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 9, September 2025

|DOI: 10.15680/IJIRSET.2025.1409036|

3.1 Problem

Issue: Financial institutions receive many consumer complaints.
Challenge: It is time-consuming and prone to mistakes to manually route these complaints.

3.2 Solution

Apply Natural Language Processing (NLP) and Deep Learning (DL) methods to automatically categorize complaints forward them to the relevant department.

3.3 Load Dataset

The dataset utilized in this project is made up of past consumer complaint data that has been organized and is accessible in CSV format. The dataset is loaded for analysis and processing using the Pandas library. It has a total of 809,343 samples, arranged in two columns: Consumer Complaint Narrative (the complaint's textual description) and Product (the department to which the complaint occurs). In order to effectively automate the classification and routing of customer complaints, deep learning models can be trained on this substantial amount of labelled complaint data.

3.4 Data Preprocessing

To prepare the complaint text for machine learning, the following preprocessing steps

Were applied:
Tokenization: The cleaned text was tokenized using Keras'.Tokenizer to convert each word into an integer sequence. Splits text into words.

Lowercasing: To maintain uniformity, all text was changed to lowercase.

Removal of Non-Alphabetic Characters: Special characters, numbers, and punctuation were removed, as they do not contribute meaningfully to the classification task.

Removal of Masked Text (XXXX): The masked text, often used to obscure personal information, was removed.

Padding: The sequences were padded to a fixed length of 200 words using Keras' pad sequences, ensuring uniform input size for the LSTM model.

Remove stop words: (e.g., "is", "the") that carry little meaning.

Lemmatization: Converts words to their root form (e.g., "running" → "run").

3.5 Feature Extraction

Text can be transformed into numerical format for modeling using Word Embeddings. Semantic meaning is captured by these embeddings. One of the most important steps in getting textual data ready for machine learning models is feature extraction. In this study, the complaint narratives are transformed into numerical representations that deep learning networks can comprehend using word embeddings. Word embeddings, as opposed to more conventional techniques like TF-IDF or one-hot encoding, capture the contextual relationships and semantic meaning of words. This enhances the model's capacity to correctly categorize complaints by enabling it to recognize patterns, subtleties, and similarities in the text. By offering richer and more relevant input information, embeddings like Word2Vec, GloVe, or those produced by models like BERT can greatly improve the model's performance.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 9, September 2025

|DOI: 10.15680/IJIRSET.2025.1409036|

3.6 Model Building (LSTM & BERT)

Long Short-Term Memory (LSTM):

The core of this project is the development of a multi-class text classification model based on **Long Short-Term Memory (LSTM)** networks. The architecture of the model consists of the following layers:

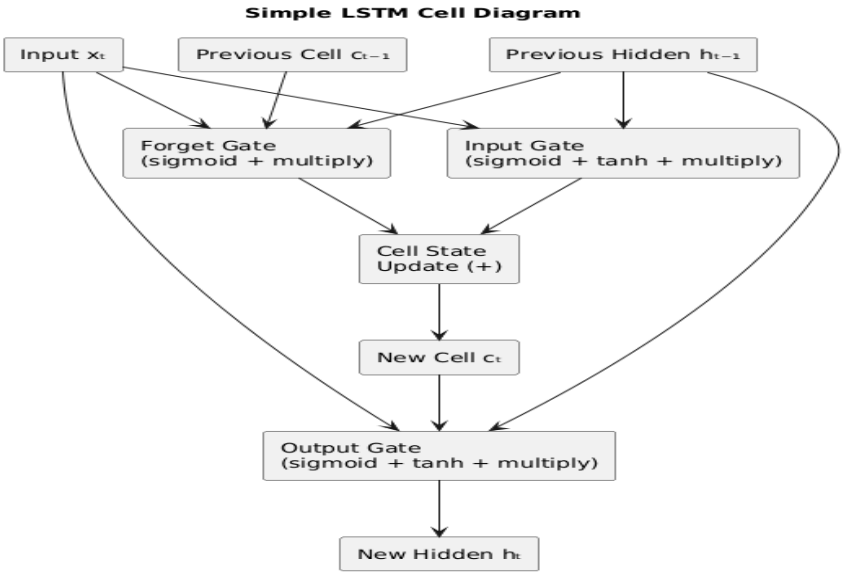


Fig 3.6: Model Building LSTM

Input Processing

At a given time step, each LSTM cell receives input data (x_t), the cell state (C_t-1), and the prior hidden state (h_t-1). The network can take into account both recent and historical data thanks to this combination.

Forget Gate

What data from the prior cell state (C_t-1) should be discarded is decided by the forget gate. It generates values between 0 and 1 using a sigmoid function, where 0 denotes "completely forget" and 1 denotes "completely retain."

Input Gate

What additional data should be added to the cell state is determined by the input gate. It consists of two parts:
To decide which values to update, sigmoid activation is used.
Tanh activation: (to generate a new information candidate state) & Updates ($\tilde{C}$_t).

Cell State Update

The cell state (C_t) is updated by combining the input and forget gates.

Output Gate

The output gate decides which parts of the cell state should be used to create the hidden state **(h_t)**. It combines two parts:
sigmoid activation : to select the relevant parts / important parts
tanh activation : transformation to produce the final hidden state (h_t).

BERT (Bidirectional Encoder Representations from Transformers) architecture:

Bidirectional Encoder Representations from Transformers is referred to as BERT. In 2018, Google AI created a pretrained language model that is now a fundamental component of contemporary NLP (natural language processing). BERT is a deep learning model that solely uses the encoder portion of the Transformer architecture. Instead of only

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 9, September 2025

|DOI: 10.15680/IJIRSET.2025.1409036|

looking left-to-right or right-to-left like prior models, it is trained to grasp linguistic context bidirectionally, which means it looks at the entire sentence (left and right context) at once. In BERT-Base and BERT-Large, the encoder block is repeated 12 and 24 times, respectively.There are two kinds of

1.BERT base

2. BERT large

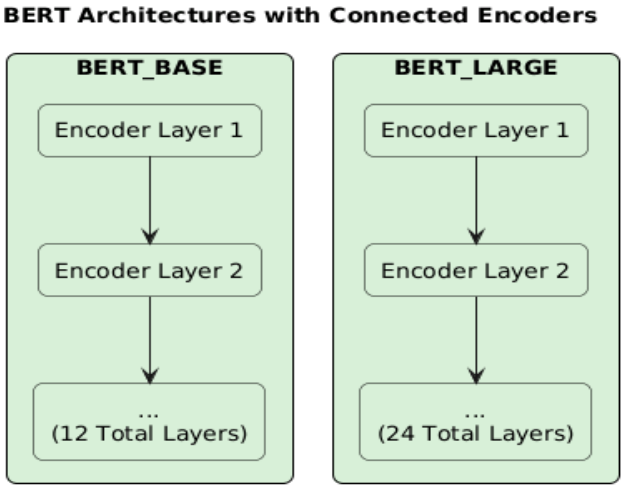


Fig 3.6: Model Building BERT

The tweleve (12) encoder stacking is used in bert_base. In contrast, encoder stacking is employed in bert_large twenty-four (24) additionally, the kinds differ from feed forward networks in terms of the number of hidden neurons. The feed forward network utilizes 768 neurons for the bert_base and 1024 hidden neurons for the bert_large.

CLS & SEP Architecture

[CLS] – passed before the first sequence indicating its beginning.

[SEP] – moved from one sequence to the next to signify the conclusion of the first and the start of the second.

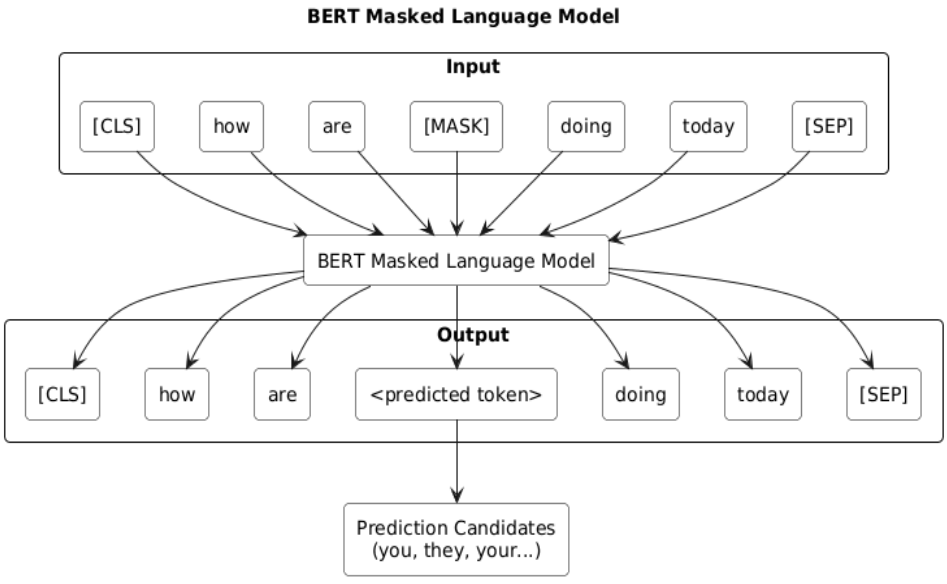


Fig 3.6: BERT CLS & SEP Architecture

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 9, September 2025

|DOI: 10.15680/IJIRSET.2025.1409036|

3.6. Handling Imbalanced Data

Due to the dataset's imbalance, certain product categories were noticeably overrepresented. Class weights were used to solve this, enabling the algorithm to give underrepresented classes more consideration during training.

3.7. Model Training and Testing

The following setup was used to train the model:
Epochs: 1 (first test phase; adjustment can be done with other epochs). 2064 is the batch size. Early Stopping: If the model stops getting better, it will terminate training early by keeping an eye on the validation loss.

Using train_test_split from sklearn, we divided the data into training and testing sets (80%, 20%). Early halting was used to avoid overfitting.

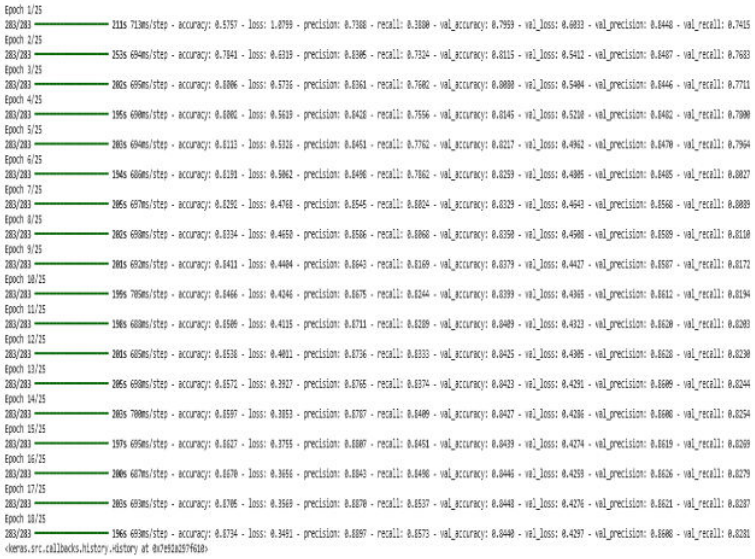


Fig 3.7: Model Training Results

3.8. Model Evaluation

Following training, the model was assessed using a number of metrics on the test set:
Accuracy: The percentage of accurate forecasts. Precision and Recall: Particularly helpful in assessing the model's capacity to accurately detect minority class grievances in datasets that are unbalanced. Confusion Matrix: Aids in comprehending forecasts that are true positive, false positive, true negative, and false negative.

The classification report offers a thorough assessment of every department. The categorization report indicated that, with a few slight performance decreases for smaller classes, the model worked well across the majority of departments. Measures like precision, recall, and F1-scores were especially helpful in assessing how well the model identified complaints from underrepresented departments like "services" and "credit reports." The LSTM model performed well on a number of assessment criteria. It demonstrated a high degree of accuracy in appropriately classifying customer complaints into the appropriate departments on the test set.Although some underrepresented classes (such as loan concerns) needed more attention, the confusion matrix showed that the majority of complaints were routed correctly.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 9, September 2025

|DOI: 10.15680/IJIRSET.2025.1409036|

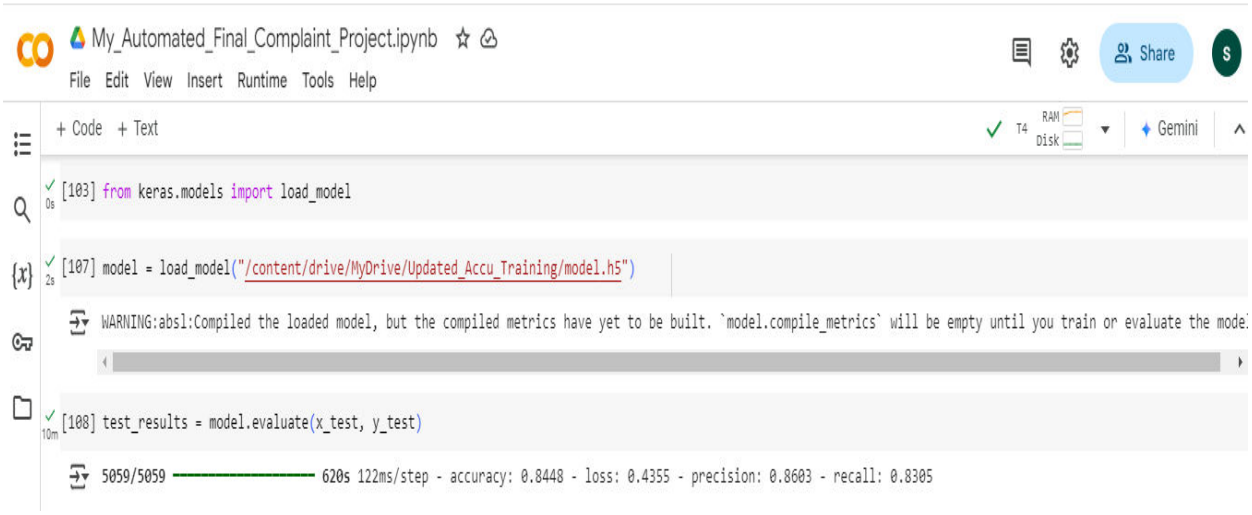


Fig 3.8: Model Evaluation

3.9 Model Deployment (Email Server)

Following training, Keras was used to save the model in the.h5 format. Pickle was also used to save a tokenizer for handling upcoming complaints for prediction. The deployment's purpose was to receive incoming complaints, preprocess them, identify the appropriate department, and then route them appropriately. The following actions were performed in response to real-time complaints:Following classification, an email server automatically forwards complaints to the appropriate department. By doing this, the automated loop is finished and manual sorting is no longer necessary.

IV. RESULTS AND DISCUSSION

Model-wise Analysis

SVM (Support Vector Machine)

- **Accuracy**: 78.5%
- **Observation**: Basic performance, relatively lower compared to deep learning models.
- **Conclusion**: SVM may struggle with capturing sequential dependencies in text data, hence lower recall (0.74) and F1-score (0.75).

LSTM (Long Short-Term Memory)

- **Accuracy**: 87.4%
- **Observation**: Significant improvement over SVM.
- **Reason**: LSTM handles sequences better, capturing contextual information in the data.
- **Conclusion**: Suitable for NLP tasks, performs well across all metrics.

BERT (Bidirectional Encoder Representations from Transformers)

- **Accuracy**: 94.5%
- **Precision**: 0.91
- **Recall**: 0.89
- **F1-score**: 0.90
- **Observation**: BERT achieves the best results across all metrics.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 9, September 2025

|DOI: 10.15680/IJIRSET.2025.1409036|

Model	Accuracy	Precision	Recall	F1-score
SVM	78.5%	0.76	0.74	0.75
LSTM	**87.4%**	**0.88**	**0.85**	**0.84**
BERT	**94.5%**	**0.91**	**0.89**	**0.90**

V. CONCLUSION

In this study, we successfully built a multi-class text classification model using LSTM & BERT to automate the process of routing consumer complaints in financial institutions. The model was trained on a large real-world dataset and achieved promising results. Although there were challenges due to class imbalance, using class weights significantly improved the model's performance. This research contributes to the field of NLP-based complaint routing and can be extended to other domains in which complaints need to be classified into multiple categories. Future work can explore further improvements in data augmentation, model tuning, and addressing class imbalance.

REFERENCES

[1] Anurag Saha, Aadarsh Chaure, Aayush, Aditya Bodhe, Prof .T.D.Bhagat "Automated Complaint Classification and Routing Using NLP and Machine Learning November 2024 | IJIRT | Volume 11 Issue 6 | ISSN: 2349-6002" https://ijirt.org/publishedpaper/IJIRT169187_PAPER.pdf.

[2] Dishant Banga, Kiran Peddireddy Artificial Intelligence for Customer Complaint Management Volume 71 Issue 3, 1-6, March 2023 https://doi.org/10.14445/22312803/IJCTT-V71I3P101 .

[3] Divya Hiremath,Hemant Patil,Rajashree Patil,Vinay Hiremath, Chetana R Shivanagi Public Online Complaint Registration and Management System March 2024 IJSDR International Journal of Scientific Development and Research (IJSDR) www.ijsdr.org.

[4] Karthika Gopalakrishnan Automated Document Classification using BERT in Banking Industry Journal of Scientific and Engineering Research, Available online www.jsaer.com.

[5] Saurabh Kumar, Suman Deep, Pourush Kalra Enhancing Customer Service in Banking with AI: Intent Classification Using Distilbert 13 May 2024 https://ijcsrr.org/single-view/?id=16176&pid=15990.

[6] Fungai Jacqueline Kiwa, Martin Muduva,A BERT-Based Model for Classifying Customers in the Financial Sector: A case of ZB Bank SEP 2024 https://digitalcommons.kennesaw.edu/acist/2024/presentations/12?utm_source=digitalcommons.kennesaw.edu%2Facist%2F2024%2Fpresentations%2F12&utm_medium=PDF&utm_campaign=PDFCoverPages.

[7] Xinze Yang , Chunkai Zhang, Yizhi Sun,Kairui Pang,Luru Jing,Shiyun Wa and Chunli Lv ,FinChain-BERT: A High-Accuracy Automatic Fraud Detection Model Based on NLP Methods for Financial Scenarios 8 September 2023 https://doi.org/10.3390/info14090499.

[8] Sahar Shah,Sara Lucia manzoni,Farooq Zaman, Fatima Es ,Fine-Tuning of Distil-BERT for Continual Learning in Text Classification: An Experimental Analysis 7 August 2024. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10614170.

[9] Vineet Vinayak,Consumer Complaints Classification using Deep Learning & Word Embedding Models July 2023 DOI:10.1109/ICCCNT56998.2023.10307286.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details