Customer Support Ticket Classification Powered by Logistic Regression

Advaith Deshmukh[1*], Ganesh[2], Irfan[3], Mihir Shukla[4]

[1,2,3,4]Department of Computer Science and Engineering, Guru Nanak Dev Engineering College ,Bidar, Karnataka, India

*Corresponding Author Email: advaith.deshmukh2@gmail.com

Abstract

In today's fast-paced digital environment, organizations receive a large volume of customer queries, complaints, and requests through various communication channels. Manually categorizing and routing these support tickets to the appropriate departments is time-consuming, error-prone, and inefficient. This paper presents TicketAI, an AI-powered customer support ticket classification system that leverages machine learning and natural language processing (NLP) to automate ticket management workflows. The system integrates two complementary artificial intelligence models: an emotion-based classifier utilizing pre-trained deep learning architectures from Hugging Face, and a keyword-based classifier developed using Scikit-learn trained on labeled ticket data to identify categories such as Billing Issues, Technical Support, Returns, and Account Management. The fusion of these two models enables context-aware predictions that improve ticket prioritization and routing. Implemented as a Flask web application with SQLite database storage and Gmail API integration, the system achieves 90.1% classification accuracy with an average response time of 2.4 seconds per ticket. Experimental results demonstrate that the hybrid approach significantly reduces manual workload, minimizes response time, and enhances customer satisfaction.

Keywords— *Customer Support, Ticket Classification, Logistic Regression, NLP, Sentiment Analysis, Flask, Machine Learning*

1. Introduction

In the digital era, customer support has evolved into a vital component of business success. As companies grow and diversify their services, they face a surge in customer inquiries, feedback, and complaints arriving through multiple channels such as email, live chat, and support portals. At the heart of this challenge lies ticket classification—organizing and routing support tickets to the appropriate department based on the issue's content, urgency, or sentiment.

Traditionally, classification has been performed manually, which is time-consuming, error-prone, and inconsistent, especially under high ticket volumes. Artificial Intelligence (AI), specifically Natural Language Processing (NLP) and Machine Learning (ML), enables systems to understand human language, learn from historical data, and make intelligent decisions in real time [1].

This paper proposes TicketAI, a hybrid AI system that combines emotion-based classification using transformer models with keyword-based classification using Logistic Regression to automate customer support ticket management. The system is deployed as a Flask web application with SQLite storage and automated Gmail notification support.

2. Related Work

Several studies have explored automated ticket classification and sentiment analysis. Early work using Naive Bayes classifiers [1] demonstrated feasibility for small datasets but struggled with ambiguous multi-topic tickets. Logistic Regression applied to e-commerce complaints [2] offered interpretability but limited scalability. Support Vector Machines (SVM) with TF-IDF vectorization [3] achieved 85% accuracy on IT support datasets but failed to capture deep contextual meaning.

Deep learning approaches using transformer-based models such as BERT [6] achieved high accuracy but required large annotated datasets and significant computational resources. Hybrid CNN-LSTM models [7] effectively captured both local and sequential text patterns. NLP-based helpdesk automation using SpaCy and Scikit-learn [8] showed up to 88% accuracy while remaining computationally efficient.

More recent work on hybrid AI models [19] combining rule-based reasoning with deep learning NLP demonstrated improved accuracy and response efficiency. Ensemble approaches [20] achieved up to 97% accuracy across diverse domains. The literature consistently shows that combining multiple AI models improves performance, interpretability, and generalization—motivating the hybrid design of TicketAI.

3. Materials and Methods

3.1 System Architecture

TicketAI is designed using a three-tier architecture: (i) Presentation Layer — a Flask web interface for ticket submission and dashboard visualization; (ii) Application Layer — core ML logic for emotion detection and keyword classification; and (iii) Data Layer — SQLite database for persistent storage of tickets, model outputs, and user information.

3.2 Dataset

The training dataset comprises approximately 8800 customer support tickets sourced from publicly available datasets including Customer Support on Kaggle, E-Commerce Complaint Dataset (UCI Repository), GoEmotions by Google (Hugging Face), and synthetic data generation. Each record contains a customer message, category label (Billing Issues, Technical Support, Returns, Account Management, or General Queries), and emotion label (Angry, Frustrated, Sad, Neutral, or Gratitude). The dataset is split 70% training, 15% validation, and 15% testing.

3.3 Text Preprocessing

The preprocessing pipeline, implemented using NLTK and SpaCy, performs: (1) text cleaning removal of HTML tags, punctuation, URLs, and conversion to lowercase; (2) tokenization splitting text into individual tokens; (3) stop word removal; (4) lemmatization reducing words to base form; (5) contraction expansion (e.g., can't → cannot); and (6) noise filtering of non-alphabetic tokens.

3.4 Feature Extraction

Text is converted to numerical features using TF-IDF (Term Frequency-Inverse Document Frequency) vectorization for the keyword classifier, and contextual transformer embeddings from DistilBERT for the emotion classifier. TF-IDF captures keyword importance relative to the corpus, while transformer embeddings capture semantic context and emotional tone dynamically.

3.5 Classification Models

Two complementary classifiers are employed. The Emotion-Based Classifier uses the pre-trained DistilBERT model (bhadresh-savani/distilbert-base-uncased-emotion) from Hugging Face to identify emotional tone across six classes: anger, frustration, sadness, neutral, joy, and gratitude. The Keyword-Based Classifier uses TF-IDF vectorization with Multinomial Naive Bayes, trained on labeled support tickets to predict five issue categories.

A hybrid fusion module combines both model outputs using rule-based decision logic: tickets classified as Angry/Frustrated emotion combined with Billing/Technical category are assigned High Priority; Neutral emotion with General Inquiry receives Normal Priority; Gratitude with Feedback is assigned Low Priority. This fusion mechanism enhances routing accuracy and ensures emotionally sensitive issues receive appropriate urgency.

3.6 Implementation Stack

The system is implemented using Python 3.10+ with Flask 2.3+ for the web framework, Scikit-learn for the keyword classifier, HuggingFace Transformers for the emotion model, NLTK and SpaCy for NLP preprocessing, SQLite3 for data storage, and Gmail API for automated email notifications. The frontend uses HTML5, CSS3, and Bootstrap 5. The modular folder structure separates models, templates, static files, database, and utilities for maintainability.

4. Results and Discussion

4.1 Model Performance

Both AI models were evaluated on a labeled test set of 1,000 customer support tickets. Table 1 summarizes the performance metrics achieved.

Table 1. Classification Model Performance

Model	Acc. (%)	Prec. (%)	Recall(%)	F1 (%)
Emotion Classifier (DistilBERT)	88.2	87.0	89.0	88.0
Keyword Classifier (TF-IDF+NB)	84.6	82.0	83.0	82.5
Combined Hybrid Model	90.1	89.0	90.0	89.5

The emotion classifier achieved 88.2% accuracy using DistilBERT, effectively capturing tone variations including anger, frustration, gratitude, and neutrality. The keyword classifier achieved 84.6% accuracy using TF-IDF with Multinomial Naive Bayes, efficiently predicting ticket categories. The hybrid model achieved the best overall accuracy of 90.1%, demonstrating that fusion-based classification enhances decision-making reliability.

4.2 Emotion Detection Analysis

The emotion classifier successfully mapped customer tone to priority levels. Tickets expressing anger (e.g., 'I am really angry about the late refund') and frustration (e.g., 'My payment was declined even after trying twice') were correctly tagged as High Priority. Neutral queries received Normal Priority, and messages expressing gratitude were assigned Low Priority. Misclassifications were primarily observed between Technical and Account Management categories due to overlapping vocabulary such as 'login error' and 'access problem'.

4.3 System Performance Metrics

The system demonstrated an average API response time of 2.4 seconds per ticket, validating real-time deployment suitability. Email delivery via Gmail API achieved 100% delivery rate. User Acceptance Testing (UAT) with a group of test users yielded 95% usability satisfaction. The SQLite database efficiently handled all CRUD operations with consistent data integrity. The hybrid model reduced manual ticket handling by an estimated 60–70% compared to traditional manual classification workflows.

4.4 Comparison with Existing Systems

Compared to traditional rule-based systems [5], TicketAI provides significantly higher flexibility and adaptability to unseen vocabulary. Against standalone SVM classifiers [3] achieving 85% accuracy, the hybrid model improves accuracy by 5.1 percentage points. Compared to deep learning approaches requiring GPU resources [6], TicketAI offers a more lightweight solution with competitive accuracy suitable for production deployment on standard hardware. The addition of emotion-aware prioritization addresses a critical gap in existing keyword-only systems, enabling context-sensitive routing not found in prior work.

5. Conclusion

This paper presented TicketAI, a hybrid AI-powered customer support ticket classification system combining emotion detection (DistilBERT transformer) with keyword-based classification (TF-IDF + Naive Bayes). The system achieved 90.1% overall classification accuracy with a 2.4-second average response time, effectively automating ticket categorization and prioritization.

The proposed approach demonstrates that fusing contextual emotion understanding with lightweight keyword classification outperforms either method independently. Integration with Flask, SQLite, and Gmail API provides a complete, production-ready solution that reduces manual workload by 60–70%, improves routing consistency, and enhances customer satisfaction through emotion-aware prioritization.

Future work will explore live chatbot integration, multilingual support, reinforcement learning for intelligent agent assignment, and an analytics dashboard for ticket trend visualization. Deployment on cloud platforms with CI/CD pipelines will further extend the system's scalability and real-world applicability.

References

[1] Customer Complaint Analysis using Naive Bayes, IJCSIT, 2018.

[2] Complaint Classification in E-Commerce using Logistic Regression, IJETT, 2018.

[3] Support Ticket Categorization using TF-IDF and SVM, IEEE Access, 2019.

[4] Topic Modeling in Customer Support Logs, Springer, 2019.

[5] Customer Support Analytics using Rule-Based NLP, Springer, 2019.

[6] Deep Learning for Automated Ticket Categorization, IEEE, 2020.

[7] Hybrid CNN-LSTM Model for Complaint Classification, Elsevier, 2020.

[8] NLP-Based Helpdesk Automation using SpaCy and Scikit-Learn, Elsevier, 2020.

[9] Multi-Label Text Classification for Customer Support, Elsevier, 2021.

[10] Emotion Recognition in Text Using Transformer Networks, Springer, 2021.

[11] A. Zangari et al., "Ticket automation: An insight into current research with applications to multi-level classification scenarios," Expert Systems with Applications, vol. 225, 2023. DOI: 10.1016/j.eswa.2023.119984.

[12] S. Bonechi et al., "Enhancing Customer Support in Banking: Leveraging AI for Efficient Ticket Classification," Procedia Computer Science, vol. 246, pp. 128–137, 2024.

[13] L. S. Benitez Pereira et al., "Machine Learning for Classification of IT Support Tickets," in Proc. CyMaEn, 2023. DOI: 10.1109/CyMaEn57228.2023.10051041.

[14] M. Truss and S. Bohm, "AI-based Classification of Customer Support Tickets: State of the Art and Implementation with AutoML," arXiv:2406.01789, 2024.

[15] Customer Support Ticket Categorization and Prioritization Using NLP, ScitePress, 2025. DOI: 10.5220/0013640400004664.

[16] Intelligent Ticket Assignment using Reinforcement Learning, Elsevier, 2023.

[17] Transformer-Based Text Classification for Helpdesk Systems, IEEE, 2023.

[18] Hybrid AI Models for Customer Experience Management, ACM, 2023.

[19] S. H. Wewelwala and T. G. D. K. Sumanathilaka, "Hybrid Emotion Recognition: Enhancing Customer Interactions Through Acoustic and Textual Analysis," arXiv:2503.21927, 2025.

[20] Deep Ensemble Model for Text Classification, IEEE, 2024.