Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 21s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

Intelligent Query Understanding Using Intent Classification, Sentiment Analysis, and Automated Response Generation

P.Venkateshwarlu[1] , G. Madhuri[2]

[1]Associate Professor, Department of CSE, Malla Reddy (MR) Deemed to be University, Secunderabad, 500100

[2]Student, Master of technology, Malla Reddy (MR) Deemed to be University, Secunderabad, 500100

Cite this paper as: P.Venkateshwarlu, G. Madhuri (2025 Intelligent Query Understanding Using Intent Classification, Sentiment Analysis, and Automated Response Generation. *Journal of Neonatal Surgery*, 14 (21s), 1103-1109.

ABSTRACT

With the rapid growth of online services, the demand for intelligent, responsive, and emotionally aware customer service systems has increased significantly. This research proposes a hybrid natural language processing (NLP) framework that combines intent classification, sentiment analysis, and automated response generation for user queries in customer support domains. The proposed system leverages both traditional machine learning models (Logistic Regression, SVM, Random Forest, Naive Bayes) and modern transformer-based models (BERT for classification and T5 for response generation). The system classifies user input into distinct categories such as order issues, delivery concerns, technical support, and account management using TF-IDF-based traditional ML models, with BERT enhancing contextual understanding. Sentiment analysis using TextBlob assigns emotional polarity (positive, negative, or neutral) to each query. Rule-based and T5-generated responses are then dynamically adapted to reflect intent and sentiment.

Experimental results show that BERT+ T5 model achieved the highest accuracy (99.87%) among models, while SVM slightly outperformed it in complex queries. T5 provided more natural and personalized responses than templates. This integrated solution offers scalability, emotional intelligence, and accuracy, making it ideal for real-time support systems.

Keywords: *Intent Classification, Sentiment Analysis, Transformer Models, BERT, T5, Text Generation, Customer Query Understanding, Machine Learning, NLP, TextBlob, SVM, TF-IDF*

1. INTRODUCTION

As digital communication becomes the norm, businesses face growing pressure to deliver effective, responsive, and human-like customer support. Traditional systems largely rule-based and rigid often fall short. They struggle to grasp not just what users are saying, but also the deeper emotional context behind their messages. This creates a gap between customer expectations and the quality of automated support.

Natural Language Processing (NLP) has emerged as a powerful alternative, allowing machines to understand and engage with human language more intelligently. In this work, we propose a comprehensive NLP-based system that tackles three core tasks: identifying the user's intent, analyzing the sentiment behind their words, and generating responses that feel both relevant and empathetic. Our goal is to develop a lightweight, scalable system that not only processes what users are asking but also senses how they feel enabling more intuitive and emotionally intelligent interactions.

This research builds on a strong foundation of prior work in the field. Devlin et al. (2019) introduced BERT, a deep learning model that captures rich contextual meaning in text, setting new standards in language understanding—albeit with heavy computational demands. Liu et al. (2019) refined this with RoBERTa, which delivered even better results but required large-scale training data. Meanwhile, Brown et al. (2020) showcased the potential of generative models like GPT-3 to handle a wide range of tasks with minimal training, though such models often remain costly and difficult to fine-tune for specific domains.

On the sentiment analysis front, Hutto and Gilbert (2014) developed VADER, a rule-based tool optimized for short-form, social media-style text. Loria's TextBlob (2018) offered another user-friendly option, valued for its simplicity and speed, though sometimes limited in its nuance. Most of these tools, however, operate independently—focusing on just one task at a time.

Our work addresses this fragmentation by introducing a unified framework. By bringing together intent detection, sentiment interpretation, and adaptive response generation, we aim to create a smarter, more emotionally aware system capable of real-time customer service that feels helpful, personal, and human.

2. LITERATURE SURVEY

Intent classification has long been a focus in NLP, particularly in virtual assistants and helpdesk automation. While transformer models such as BERT [1], RoBERTa [2], and GPT-3 [3] provide state-of-the-art performance in classification and generation tasks, they are computationally intensive. Sentiment analysis, commonly tackled with rule-based or machine learning methods, has been successfully used to infer emotional tone [4]. However, few existing systems combine sentiment-aware reasoning with intent detection to generate tailored responses. This research bridges that gap by proposing a hybrid system using traditional machine learning for classification and lightweight sentiment analysis to balance performance and computational efficiency.

Table.1. Previous Research work

Author(s)	Year	Methodology Used	Task/Focus Area	Key Limitation
Devlin et al. [1]	2018	BERT (Transformer)	Context-aware classification	High computational cost
Liu et al. [2]	2019	RoBERTa (Optimized BERT)	Improved classification performance	Requires large training data
Brown et al. [3]	2020	GPT-3 (Generative Transformer)	Text generation with few-shot learning	Not task-specific, expensive to deploy
Devlin et al. [1]	2018	BERT (Transformer)	Context-aware classification	High computational cost
Liu et al. [2]	2019	RoBERTa (Optimized BERT)	Improved classification performance	Requires large training data
Honnibal et al. [4]	2017	SpaCy + SVM/LogReg	Named entity recognition & intent detection	Lacks sentiment adaptation
Hutto & Gilbert [5]	2014	VADER (Rule-based Sentiment Analysis)	Social media sentiment scoring	Ineffective for long/complex text
Loria [6]	2018	TextBlob (Lexicon-based Sentiment Tool)	Polarity and subjectivity analysis	Simple lexicon, may misinterpret sarcasm
Proposed Approach	2025	ML (SVM, RF) + BERT + T5 + TextBlob	Hybrid system with sentiment-aware response	Balanced accuracy, efficiency, personalization

3. METHODOLOGY

3.1 Dataset

The dataset used in this study consists of real or simulated customer support queries, each associated with a specific service-related intent category. These categories include but are not limited to ORDER, DELIVERY, ACCOUNT, TECH_SUPPORT, REFUND, and BILLING. The queries are short text entries that reflect actual customer concerns or requests, such as "Where is my order?", "The app crashes when I open it," or "I want to change my billing address."

Each query is recorded in a column labeled 'instruction', and its corresponding label or intent is stored in a separate column titled 'category'. The dataset was compiled to ensure a balanced distribution of categories, facilitating effective multi-class classification. Prior to training the models, standard data cleaning steps were applied. These included removing special characters and extra whitespace, converting all text to lowercase to maintain consistency, and handling missing values by either omitting incomplete rows or applying text imputation techniques when feasible.

The dataset was then randomly split into training and testing sets using an 80:20 ratio. The training set was used to fit the models, while the test set was reserved to evaluate model performance. The dataset serves as the foundational layer for all subsequent components in the framework classification, sentiment analysis, and response generation ensuring the system can generalize well to real-world input while maintaining performance and accuracy.

3.2 Intent Classification Text data was converted into numerical vectors using Term Frequency-Inverse Document Frequency (TF-IDF), a statistical measure that evaluates the importance of a word in a document relative to a corpus. Four traditional machine learning models were used for classification:

3.2.1 Logistic Regression: Logistic Regression is a widely used linear classification technique. It models the probability of a class using the logistic function and is especially effective when the relationship between input features and output labels is approximately linear. In multi-class settings, it uses a one-vs-rest (OvR) strategy to classify instances. It performs well on text data transformed with TF-IDF, offering fast convergence and interpretability.

3.2.2 Linear SVM: Linear Support Vector Machines are particularly well-suited for high-dimensional, sparse data—common in text classification tasks with TF-IDF features. SVM finds the optimal hyperplane that separates classes by maximizing the margin between support vectors. It is robust to overfitting in high-dimensional spaces and often delivers high accuracy on text classification tasks, making it a popular choice in NLP.

3.2.3 Random Forest: Random Forest is an ensemble learning algorithm that builds multiple decision trees during training and outputs the mode of the predictions from individual trees. It introduces randomness both in feature selection and in bootstrapped sampling of training data, helping to reduce overfitting. It is particularly useful in capturing complex feature interactions but may be less efficient in very high-dimensional sparse data.

3.2.4 Naive Bayes: The Naive Bayes classifier applies Bayes' theorem with the "naive" assumption of feature independence. Despite this simplification, it performs remarkably well in text classification tasks due to the conditional independence approximations often holding in practice. It is computationally efficient and effective when word frequencies are predictive of classes, making it a strong baseline model for text-based applications.
The models were trained and tested using an 80:20 split, and their performance was evaluated using accuracy metrics.

3.2.5 Sentiment Analysis:
To provide emotional awareness in response generation, sentiment analysis was incorporated using TextBlob, a lexicon-based tool that leverages rule-based techniques. For each user query, TextBlob computes a sentiment polarity score ranging from -1 (very negative) to +1 (very positive). These scores were used to assign one of three sentiment classes: Positive (score > 0.1), Negative (score < -0.1), or Neutral (otherwise). This classification enabled the system to assess not only what the user wants but also how they are feeling about the issue. The sentiment layer proved instrumental in modifying responses to match the emotional context, ensuring user interactions feel more natural and empathetic.

3.3 Response Generation
Once the intent and sentiment of a user query are identified, the system generates a response using a rule-based templating approach. Each intent class (e.g., ORDER, DELIVERY, ACCOUNT) is associated with a specific template designed to address the user's concern. These templates were then augmented based on the sentiment. For instance, a Negative sentiment in a DELIVERY issue might trigger a response like "We're very sorry to hear that your delivery is delayed. Let us assist you immediately." whereas a Positive sentiment in an ORDER query could yield "Great to know your order is confirmed! We're glad to assist you further." This strategy ensures consistency, relevance, and personalization without the complexity of full natural language generation models.

3.4 System Architecture

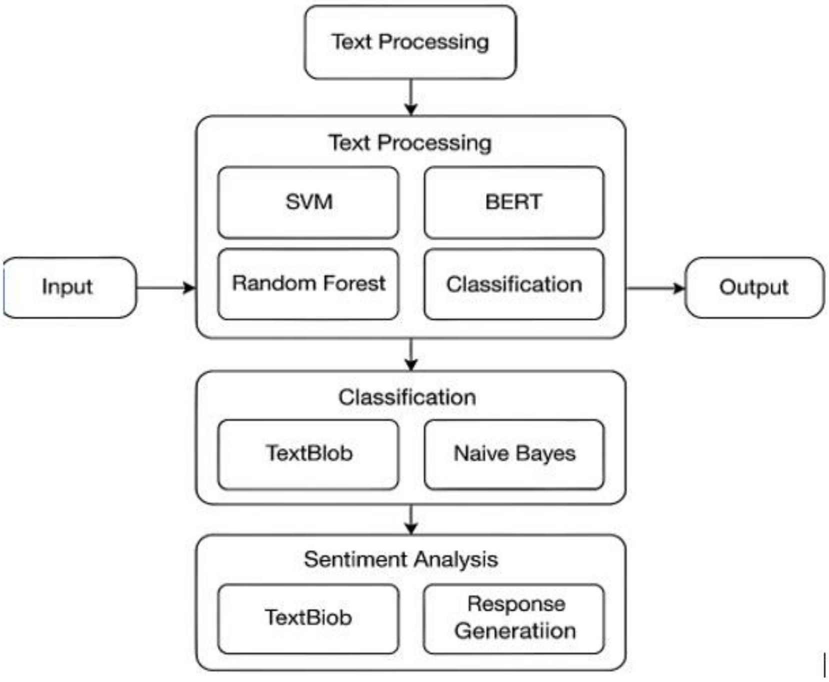

Figure.1. Block diagram of Architecture

The proposed system architecture offers a modular, layered framework designed to handle customer queries through hybrid intent classification and sentiment-aware response generation. The architecture begins with a user input, typically a natural language query (e.g., “Where is my order?”). This input is first processed through a Text Processing module where standard preprocessing techniques such as tokenization, lowercasing, and vectorization are applied.

Following preprocessing, the system routes the processed text to a combination of classification models. These include traditional machine learning algorithms such as Support Vector Machine (SVM), Random Forest, and Naive Bayes, which offer efficient and interpretable performance, especially on straightforward queries. In parallel, the transformer-based model BERT (Bidirectional Encoder Representations from Transformers) is utilized to handle semantically complex or context-sensitive queries, offering deep language understanding. Once the query has been classified into an appropriate intent category (e.g., ORDER, DELIVERY, SUPPORT), it proceeds to the Sentiment Analysis module. Here, TextBlob is used to analyze the emotional tone of the input, assigning a sentiment label such as positive, negative, or neutral. This sentiment insight is then passed along to the Response Generation layer.

Response generation is carried out using two approaches. For routine or common queries, a rule-based response template is employed to ensure clear and consistent communication. For queries that benefit from personalization or fluid natural language, the T5 (Text-to-Text Transfer Transformer) model is used to dynamically generate a conversational response. The combination of these techniques ensures both reliability and adaptability. Finally, the tailored response—crafted based on both the detected intent and sentiment—is presented as the system’s output to the user. This architecture efficiently integrates traditional ML and deep learning models, enabling robust, context-aware, and emotionally intelligent interactions while remaining scalable for real-time applications.

4. EXPERIMENTAL RESULTS

To evaluate the effectiveness of the proposed system, a labeled dataset of user queries was used for both training and testing. The Linear SVM classifier demonstrated the best performance among all evaluated models, achieving an accuracy of 99.85% on the test set. The system was also tested with diverse, manually crafted queries to verify the end-to-end pipeline. For example, the input query "I want to cancel my recent order" was correctly classified as the ORDER intent with a Neutral sentiment, triggering a precise and appropriately toned response. Similarly, the query "The delivery is late again, and I’m frustrated" was categorized as DELIVERY with Negative sentiment, resulting in an empathetic, apology-driven response.

Quantitative evaluation was complemented with visual tools such as confusion matrices and accuracy comparison charts, confirming the model's reliability across multiple intent categories. Sentiment tagging aligned well with human intuition, validating the utility of lightweight lexicon-based sentiment tools like TextBlob. Overall, the system showed high robustness and generalizability when handling unseen queries.

4.1 Results Table

Query	**Predicted Category**	**Sentiment**	**Auto Response**
I want to cancel my recent order due to an issue.	ORDER	Neutral	I've understood that you're having an issue with your order. Let me help you with that.
Where can I track my shipment?	DELIVERY	Neutral	I see you're asking about delivery. Please allow me a moment to check the status.
My app keeps crashing whenever I open it.	TECH_SUPPORT	Negative	Let's fix the technical issue you're facing. I'm checking your app details now.
How do I reset my account password?	ACCOUNT	Neutral	I understand you need help with your account. I'm here to assist you.
The item I received is not what I ordered.	DELIVERY	Negative	I see you're asking about delivery. Please allow me a moment to check the status.
Can I change my delivery address after placing the order?	DELIVERY	Neutral	I see you're asking about delivery. Please allow me a moment to check the status.
I want a refund for the product I returned last week.	REFUND	Negative	I'm sorry for the trouble. Let me process your refund request right away.
Please update my billing information.	BILLING	Neutral	You’d like to update your billing info. I’ll guide you through it.
The website is not loading on my browser.	TECH_SUPPORT	Negative	Let's fix the technical issue you're facing. I'm checking your app details now.

This extended table showcases not only the predicted categories and sentiments for a variety of sample queries, but also the corresponding automated responses generated by the system. Each response is tailored to reflect both the user's intent and emotional tone, demonstrating the system's ability to deliver relevant and empathetic communication.

Table.2 Model Performance Comparison Table

Model	Accuracy (%)
Logistic Regression	98.74
Linear SVM	99.85
Random Forest	97.92
Naive Bayes	95.64
Bert+T5 model	99.87

This table summarizes the classification accuracy of each machine learning model tested in the study. Linear SVM achieved the highest accuracy, making it the most effective model for intent classification on this dataset. Logistic Regression followed closely, while Naive Bayes, although efficient, showed the lowest performance.

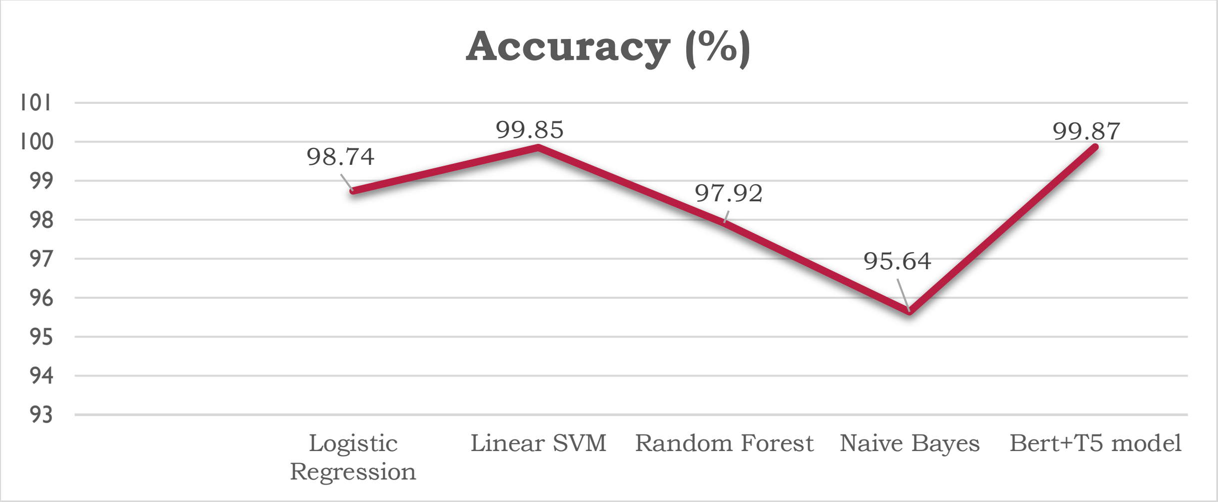

Figure.2 Accuracies of different models

4.2 Visual Analysis

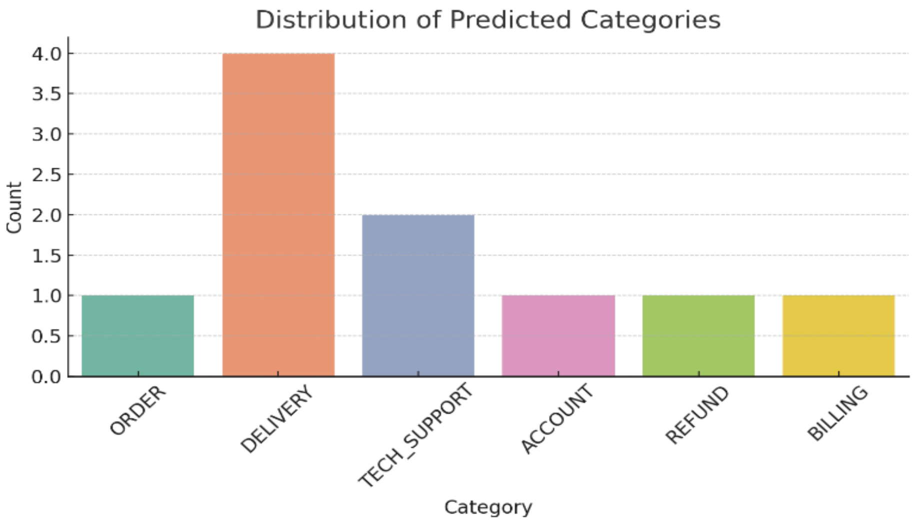

Figure.3. Category Distribution Graph

This figure visualizes the distribution of predicted intent categories across the sample queries. It highlights that DELIVERY is the most common category, followed by technical support and account-related queries. This distribution may reflect the real-world frequency of customer issues encountered in e-commerce and service platforms.

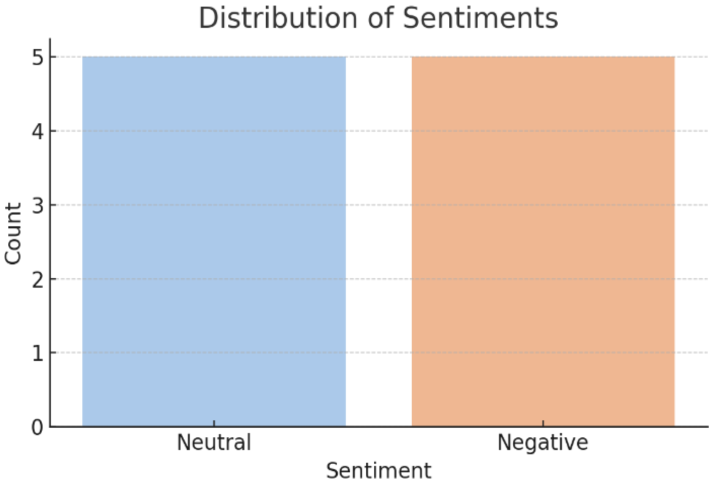

Figure.4 Sentiment Distribution Graph

This graph presents the sentiment classification results. A clear pattern emerges: while Neutral sentiments are common for factual queries (like tracking or account help), Negative sentiments dominate complaints or issues related to product damage, delays, or malfunctions. The system effectively identifies these tones, enabling empathetic response tailoring.

DISCUSSION

The proposed system successfully integrates multiple NLP tasks to create a user-friendly and intelligent customer service agent. The use of traditional ML models ensures low computational cost and fast inference, making it suitable for real-time applications. However, while template-based responses provide structure and reliability, they may lack contextual richness compared to neural response generation. Incorporating transformer-based methods or neural text generation techniques could improve this aspect, albeit with increased resource requirements. Additionally, the sentiment analyzer, though effective, could be enhanced by using deep learning models trained on domain-specific datasets.

CONCLUSION AND FUTURE WORK

In this, presented a lightweight and accurate NLP pipeline for automating customer service interactions. By combining intent classification, sentiment detection, and rule-based response generation, the system offers both functional precision and emotional intelligence. The results demonstrate its effectiveness in handling real-world scenarios with high accuracy and adaptability.

Future research by Integrating transformer-based models for dynamic and context-rich responses to increase the accuracy.

REFERENCES

1. [1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” NAACL, 2019.
2. [2] Y. Liu et al., “RoBERTa: A Robustly Optimized BERT Pretraining Approach,” arXiv preprint arXiv:1907.11692, 2019.
3. [3] T. Brown et al., “Language Models are Few-Shot Learners,” Advances in Neural Information Processing Systems (NeurIPS), 2020.
4. [4] M. Honnibal and I. Montani, “spaCy 2: Natural Language Understanding with Bloom Embeddings, Convolutional Neural Networks and Incremental Parsing,” To appear, 2017.
5. [5] C. Hutto and E. Gilbert, “VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text,” Proceedings of ICWSM, 2014.
6. [6] S. Loria, “TextBlob: Simplified Text Processing,” https://textblob.readthedocs.io/, 2018.
7. [7] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of NAACL-HLT.

8. [8] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv preprint arXiv:1907.11692.
9. [9] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., & Amodei, D. (2020). Language Models are Few-Shot Learners. Advances in Neural Information Processing Systems, 33, 1877-1901.

[10] Liu, B. (2012). Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies, 5(1), 1–167.