**Phase-2 Submission Template**

**Name:** -**D.Bala**

**Register Number:** - 421323205006

**Institution:** - Krishnasamy College of Engineering and Technology

**Department:** - IT (Information Technology)

**Date of Submission:** - 05-05-2025

**Github Repository Link:** https://github.com/asimahamed3115/Project.git

### **1. Problem Statement :**

This project addresses the challenge of enhancing customer support through an intelligent chatbot capable of providing automated assistance. Traditional customer service often struggles with delays, inconsistency, and high operational costs. Our solution leverages Natural Language Processing (NLP) and Machine Learning (ML) to build a smart chatbot that can resolve queries, offer information, and escalate complex issues, thereby improving customer satisfaction and operational efficiency.

### **2. Project Objectives :**

* Design and implement a chatbot that handles real-time customer queries efficiently.
* Integrate Natural Language Understanding (NLU) to ensure the chatbot comprehends diverse user intents.
* Incorporate Machine Learning techniques for continuous learning and improvement.
* Enable escalation to human agents for complex issues beyond the chatbot’s scope.
* Improve customer satisfaction and reduce response times.

**3. Flowchart of the Project Workflow :**

### *WhatsApp Image 2025-05-06 at 2.30.55 PM.jpeg*

### 

### **4. Data Description :**

The dataset used for training includes various customer interaction logs sourced from public datasets and simulated conversations.  
  
Features include:  
1. User Query

2. Intent Category  
3.Response Template  
4.Contextual Tags  
  
Number of Records: ~10,000 interactions  
Number of Features: 4

### **5. Data Preprocessing :**

* **Missing Values:** " Handled by removing incomplete entries."
* **Duplicate Records:** " Removed to maintain uniqueness."
* **Outliers:** " Not applicable as data is textual."
* **Data Types:** " Ensured all text data is properly encoded."
* **Categorical Encoding:** " Intent categories were encoded using Label Encoding."
* **Normalization/Standardization:** " Text normalization including lowercasing, stemming, and removal of stopwords.."
* **Transformation Explanation:** " Preprocessing ensured clean and consistent data, crucial for accurate intent classification and chatbot performance.."

### **6. Exploratory Data Analysis (EDA)** **:**

**Univariate Analysis:**

* " Frequency distribution of intents and common user queries.”

**Bivariate/Multivariate Analysis:**

* " Word clouds and intent-category relationships were visualized."
* **Relationship with Target Variable:** " Relationship with Target Variable: Analyzed how query length, keywords, and sentiment relate to intent accuracy."
* **Insights Summary:** " Frequent intents include basic troubleshooting, billing queries, and password resets. Sentiment analysis showed higher positivity for resolved queries."

### **7. Feature Engineering :**

* Created features such as query length, number of keywords, and sentiment polarity.
* Tokenized and vectorized queries using TF-IDF and word embeddings (Word2Vec/Glove).

**8. Model Building :**

* **Models Selected:**
  + " Logistic Regression (Baseline)"
  + " Random Forest Classifier"
  + “Deep Learning: Bidirectional LSTM for intent classification   “
* **Model Justification:**
  + " Classical ML models provide strong baselines; BiLSTM captures sequential information and improves intent recognition."
* **Data Splitting:**
  + " Training (80%) and Testing (20%)."
* **Initial Performance:**
  + " Logistic Regression Accuracy: 82%"
  + " Random Forest Accuracy: 85%"
  + “ BiLSTM Accuracy: 92%”

### **9. Visualization of Results & Model Insights :**

* Confusion matrix to evaluate intent prediction accuracy.
* Precision, Recall, F1-Score metrics for performance measurement.
* Error analysis showed confusion between similar intents like 'account issues' and 'login problems.'

### **10. Tools and Technologies Used :**

* **Programming Language:** Python
* **IDE/Notebook:** Google Colab
* **Libraries:** pandas, numpy, nltk, sklearn, keras, tensorflow, matplotlib, seaborn
* **Visualization:** Matplotlib, Seaborn
* **Models:** Linear Regression, Random Forest,BiLSTM

### **11. Team Members and Contributions :**

1. **M Asim Ahamed:** Model Deployment and Documentation
2. **M Madesh:** Data Collection and Preprocessing
3. **M Madhan:** Exploratory Data Analysis and Feature Engineering
4. **D Bala:** Model Building and Evaluation
5. **A Praveen Raj:** System Integration and Testing