**Phase-2 Submission Template**

**Student Name:** SUBAHARINI S

**Register Number:** 621523104159

**Institution:** Mahendra college of engineering

**Department:** Computer Science and Engineering

**Date of Submission:** 7/5/2025

**Github Repository Link:** https://github.com/Suba2118

### **1. Problem Statement**

*Efficient handling of customer queries*

*Scalable and always available support*

### **2. Project Objectives**

Objective: Automate customer support using a smart chatbotBenefits:

- Faster response time

- 24/7 availability

- Improved customer satisfaction

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

### Visual representation of workflow:

Data Collection -> Preprocessing -> EDA -> Feature Engineering

-> Model Building -> Evaluation -> Visualization & Deployment

### **4. Data Description**

*Sources: Chat logs, FAQs, customer feedback, ticketsFeatures:*

*- Text data (queries, issues)*

*- Metadata (customer ID, category, timestamps)*

*-Data Size: ~X records, avg. Y words per chat*

### **5. Data Preprocessing**

*Text cleaning: punctuation, HTML tags- Tokenization, Lemmatization*

*- Stop word removal*

*- Vectorization: TF-IDF, Word2Vec, BERT*

*- Encoding categorical data*

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

### *Top issue categories*

### *- Word clouds for complaints and queries*

### *- Distribution: Response time, satisfaction score*

### *- Customer demographics vs. ticket volume*

### **7. Feature Engineering**

### - Sentiment scores from messages

### - Chat length and frequency

### - Issue urgency detection

### - Response delay patterns

### **8.Model Building**

### - Intent Classification: SVM, BERT

- Response Generation: Retrieval-based (TF-IDF), Generative (GPT)

- Routing Prediction: Decision Trees, Logistic Regression

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

*- Intent Classifier Accuracy: XX%*

*- SHAP feature importance insights*

*- Chatbot vs. Human support metrics*

*- Confusion matrix & ROC curves*

### **10. Tools and Technologies Uses**

### *- Languages: Python*

### *- Libraries: scikit-learn, NLTK, spaCy, Transformers*

### *- Visualization: matplotlib, seaborn, Plotly*

### *- Deployment: Flask, Docker, Streamlit*

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

**VISHALINI S**

Contribution: Data cleaning

**SARANYA P**

Contribution: Exploratory data analysis (EDA)

**SANDHIYA S**

Contribution: Feature Engineering

**PREETHI G**

Contribution: Model development

**SUBAHARINI S**

Contribution: Documentation and Reporting