**Phase-1 Submission Template**

**Student Name:** DHARUN D

**Register Number:** 712523104015

**Institution:** PPG INSTITUTE OF TECHNOLOGY

**Department:** BE-CSE

**Date of Submission:** 26.04.2025

**1.Problem Statement**

*Traditional customer support systems often struggle with high response times, limited availability, and inconsistent service quality. These challenges lead to customer dissatisfaction, increased operational costs, and inefficiencies in handling repetitive queries. There is a growing need for an intelligent, automated solution that can provide instant, 24/7 assistance while maintaining a high level of accuracy and personalization.*

**2.Objectives of the Project**

* *Develop an AI-powered chatbot capable of understanding and responding to customer queries in real-time with high accuracy.*
* *Automate routine customer support tasks to reduce response time and workload on human agents.*

**3.Scope of the Project**

* *Development of a chatbot capable of handling frequently asked questions (FAQs) and basic support queries.*
* *Integration of the chatbot with platforms such as websites, mobile applications, or messaging services (e.g., WhatsApp, Facebook Messenger).*

**4.Data Sources**

* *Frequently Asked Questions (FAQs):  
  Pre-existing FAQ documents from the company’s website or help center will be used to train the chatbot on common customer queries.*
* *Historical Customer Support Logs:  
  Chat transcripts, email threads, and call center logs will provide real-world examples of customer interactions. These will help the chatbot understand context, intent, and common support scenarios.*
* *Datasource link:* [*https://www.kaggle.com/datasets/retailrocket/ecommerce-dataset*](https://www.kaggle.com/datasets/retailrocket/ecommerce-dataset)

**5.High-Level Methodology**

* *Data Collection:The data collection phase is critical to training, testing, and improving the chatbot. It involves gathering structured and unstructured data from relevant sources to enable the chatbot to understand customer queries and respond effectively.*
* *Data Cleaning: Data cleaning is a crucial step in preparing the collected data for training the chatbot. It ensures that the chatbot learns from high-quality, consistent, and relevant information, which directly affects its performance and accuracy.*
* *Exploratory Data Analysis (EDA): EDA helps you understand the structure, patterns, and key insights from your customer support data before feeding it into your chatbot model. It’s a foundational step in ensuring the model is trained on accurate, diverse, and relevant data.*
* *Feature Engineering : Feature engineering is the process of transforming raw data into meaningful inputs (features) that improve the performance of machine learning models. For a chatbot, especially one built using NLP and classification techniques, it plays a crucial role in understanding and generating human-like responses.*
* *Model Building: he goal is to build an intelligent chatbot that can understand customer queries and respond accurately. This typically involves two core components:*
  + *Intent Classification – To understand what the user wants.*
  + *Response Generation or Retrieval – To provide an accurate answer.*
  + *Model Evaluation: Model evaluation ensures that the chatbot performs accurately, responds appropriately to user inputs, and meets the goals of intent recognition and response generation. Evaluation depends on the model type: classification-based (for intent detection) or retrieval/generative-based (for response generation).*
  + *Visualization & Interpretation: Pie chart or bar chart of positive, neutral, and negative sentiments in user messages.Negative sentiment spikes around specific intents (e.g., refund\_request) indicate friction points.Helps prioritize features for chatbot tone, escalation triggers, or improvements in those flows.*

**6.Tools and Technologies**

* Programming Language: Python is the most widely used language for Natural Language Processing (NLP), Machine Learning, and AI chatbotdevelopment, due to its simplicity and strong ecosystem of libraries.
* Notebook/IDE: vectorizer = TfidfVectorizer()

X = vectorizer.fit\_transform(df['cleaned\_text'])

y = df['intent']

Libraries:

|  |  |
| --- | --- |
| LIBRARY | PURPOSE |
| Pandas | Dataloading,cleaning and manipulation |
| Numpy | Numerical operation and array handling |

* Optional Tools for Deployment :

|  |  |
| --- | --- |
| Tool | Purpose |
| Heroku | Beginner-friendly app hosting |
| Netlify | Static front-end hosting |

**7.Team Members and Roles**

|  |  |
| --- | --- |
| *NAME* | *ROLE* |
| *DHARUN D* | *LEADER* |
| *RAM PRASATH S* | *MEMBER* |
| *DHANAPRABHU R* | *MEMBER* |
| *UDHAYAKUMAR R* | *MEMBER* |
| *GOKUL JB* | *MEMBER* |
| *AGILESHWARAN S* | *MEMBER* |