Phase-2 Submission

Student Name: PRATHAP M

Register Number: 421223121036

Institution: KARPAGA VINAYAGA COLLEGE OF ENGINEERING AND TECHNOLOGY

Department: BIOMEDICAL ENGINEERING

Date of Submission15\05\2025

Github RepositoryLink:
https://github.com/rahamath10/shopping-chatbot-/tree/main
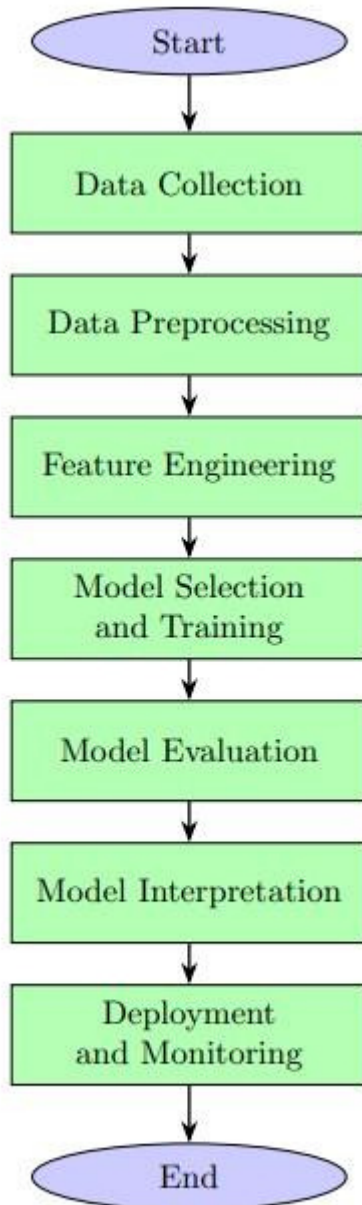
1. Problem Statement

Customer support is a critical part of the customer journey, yet many companies still rely on inefficient systems, leading to delayed responses and unsatisfactory service. The goal of this project is to develop an intelligent chatbot that can:
- *Understand customer shopping behavior,*
- *Provide personalized assistance, and*
- *Automate responses to common queries, using historical shopping data.*

2. Project Objectives

- Analyze shopping behavior and preferences from historical data.
- Preprocess and engineer features suitable for training ML/NLP models.
- Build a chatbot capable of:
- Answering product-related queries,
- Recommending items based on past purchases,
- Responding to FAQs on orders, payments, and shipping.
- Integrate a machine learning model for personalization.

3. Flowchart of the Project Workflow

```
              ┌─────────────┐
              │    Start    │
              └──────┬──────┘
                     ▼
           ┌───────────────────┐
           │  Data Collection  │
           └─────────┬─────────┘
                     ▼
           ┌───────────────────┐
           │ Data Preprocessing│
           └─────────┬─────────┘
                     ▼
           ┌───────────────────┐
           │Feature Engineering│
           └─────────┬─────────┘
                     ▼
           ┌───────────────────┐
           │  Model Selection  │
           │   and Training    │
           └─────────┬─────────┘
                     ▼
           ┌───────────────────┐
           │ Model Evaluation  │
           └─────────┬─────────┘
                     ▼
           ┌───────────────────┐
           │Model Interpretation│
           └─────────┬─────────┘
                     ▼
           ┌───────────────────┐
           │   Deployment      │
           │  and Monitoring   │
           └─────────┬─────────┘
                     ▼
              ┌─────────────┐
              │     End     │
              └─────────────┘
```

4. Data Description

- *Source: shopping_trends(e.g., from Kaggle or public APIs).*

- *Type: Text (Unstructured)*

  Rows: 3,900
  Columns: 19
  Key Features:

  - *Age, Gender, Location: Demographics*

  - *Item Purchased, Category, Size, Color, Season: Product information*

  - *Purchase Amount (USD), Review Rating: Purchase metrics*

  - *Subscription Status, Payment Method, Shipping Type: Behavioral data*

  - *Previous Purchases, Frequency of Purchases: Shopping frequency data*

5. Data Preprocessing

- *Removal of URLs, hashtags, mentions*

- *Lowercasing, tokenization, stopword removal*

- *Lemmatization/stemming*

- *Encoding target labels*

6. Exploratory Data Analysis (EDA)

○ *Word clouds for sentiment classes*

○ *Frequency plots*

○ *Sentiment distribution*

○ *Analysis of text length vs. sentiment*

○ *TF-IDF vectors*

○ *Word embeddings (GloVe, FastText)*

○ *Sentiment scores (TextBlob, VADER)*

○ *Part-of-speech tagging features*

7. Feature Engineering

Some ideas include:

- **Customer Value Score** = Purchase Amount **Review Rating** Frequency

- **Engagement Score** = Previous Purchases + Subscription Status + Promo Usage

- **Seasonal Buyer** = Based on dominant purchase Season

8. Model Building

Algorithms Used:

- Random Forest Classifier – To predict product categories. Chosen for its accuracy and ability to handle mixed data types.

- K-Nearest Neighbors (KNN) – For recommending products based on similar user profiles.

  Logistic Regression – To predict binary outcomes like subscription or

- promo code usage.

🧪 Evaluation Metrics:

- Accuracy – Overall correctness of the model.

- Precision & Recall – Measure relevance and coverage.

- F1 Score– Balance between precision and recall.

9. Visualization of Results & Model Insights

```
🤖 Hello! Welcome to our Smart Shopping Assistant.
Let's help you with your shopping needs. I'll ask a few quick questions.

👤 Age: 28
👤 Gender: Female
🛒 Purchase Amount: $75
⭐ Average Rating: 4.2
💟 Subscription: Yes
💳 Payment Method: Credit Card
📦 Previous Purchases: 12

🔍 Let me check some options for you...

🎉 Great news! You're likely to use a promo code. Here's one for you: **SAVE20**

🤖 Is there anything else I can help you with today?
```

10. Tools and Technologies Used

● *Language: Python*

● *IDE: Google Colab / Jupyter*

● *Libraries: pandas, sklearn, nltk, seaborn, matplotlib, transformers*

● *Visualization: Plotly, matplotlib*

11. Team Members and Contributions

PRATHAP.Mt — Data collection and cleaning

PREMKUMAR M — Eda and Feature Engineering

PRITHI NANDHA SEITHI P — Model building and Evaluation

PRIYADHARSHINI K — visualization and data preparation