Unified Mentor Pvt. Ltd.

Customer Satisfaction Prediction

---

1. Project Summary

This project focuses on the analysis of customer support tickets to uncover patterns and actionable insights that can enhance customer service strategies. It also involves building a machine learning model to predict customer satisfaction levels based on various features such as ticket type, priority, channel, and customer demographics. An interactive dashboard has been developed using Streamlit to visualize the results and enable user-friendly access to insights and predictions.

---

2. Project Objectives

- To analyze historical customer support ticket data for trends and insights.
- To identify relationships between ticket types, customer demographics, and satisfaction ratings.
- To build and evaluate a machine learning model to predict customer satisfaction.
- To visualize findings and predictions using a business-friendly Streamlit dashboard.
- To provide actionable recommendations to improve customer service operations.

---

3. Tools and Technologies Used

Programming Language:

- Python – Used for data preprocessing, analysis, visualization, model training, and deployment.

Libraries:

- Pandas – For data manipulation and handling structured datasets.
- NumPy – For numerical computations and efficient array operations.
- Seaborn – For creating attractive and informative statistical graphics.
- Matplotlib – For detailed and customizable data visualizations.
- Scikit-learn – For building and evaluating machine learning models.
- Joblib – For saving and loading the trained machine learning model.

Dashboard Framework:

- Streamlit – Used to develop an interactive web-based dashboard for presenting insights and visualizations.

Development Tools:

- PyCharm – Python IDE used for writing, testing, and debugging code.
- GitHub – Used for version control, collaboration, and project hosting.

---

5. Key Project Components and Activities

5.1 Data Collection and Cleaning

- Cleaned missing and inconsistent values from the dataset.
- Removed duplicates and irrelevant records.
- Formatted dates and derived additional fields such as age groups.

5.2 Exploratory Data Analysis (EDA)

- Analyzed ticket volume by type, customer demographics, and channels.
- Visualized key trends in satisfaction ratings and resolution times.

5.3 Feature Engineering

- Encoded categorical variables using Label Encoding.
- Derived additional features including resolution time and ticket age.
- Grouped customer age into logical segments.

5.4 Model Building

- Selected a Decision Tree Classifier for its interpretability.
- Used 70-30 train-test split for model evaluation.
- Saved the trained model using `joblib`.

5.5 Dashboard Development

- Created an interactive Streamlit application.
- Integrated model predictions and visual insights.
- Ensured clean and user-friendly UI layout with tabbed navigation.

7. Data Visualizations Used

The following visualizations were included to provide clear, data-driven insights:

- **Ticket Volume by Age Group** (Bar Chart)
- **Ticket Priority Distribution** (Pie Chart)
- **Ticket Type Frequency** (Bar Chart)
- **Satisfaction Ratings by Gender** (Bar Chart)
- **Ticket Trends Over Time** (Line Chart)
- **Resolution Time Distribution** (Histogram)
- **Ticket Channels by Satisfaction Level** (Stacked Bar Chart)
- **First Response Time by Ticket Type** (Box Plot)

These visualizations helped in uncovering patterns in customer behavior and service issues.

---

8. Streamlit Dashboard Overview

The dashboard includes the following:

- **Dataset Overview Tab**:
  - Displays dataset shape, basic statistics, and data preview.
- **Visual Insights Tab**:
  - Contains multiple interactive charts and plots generated from EDA.
  - Organized layout with appropriate chart sizing and titles for better readability.

---

9. Model Training and Evaluation

Model Used:

- **Decision Tree Classifier**

Features Considered:

- Customer Age, Gender
- Ticket Type, Ticket Channel, Ticket Priority
- Product Purchased, Resolution Time

Target Variable:

- **Customer Satisfaction Rating** (categorical)

---

10. Model Deployment

- The trained model was saved using `joblib` (`best_model.pkl`).
- The model was deployed locally via the Streamlit interface.
- The app loads the model automatically and provides real-time prediction on new inputs.

---

11. Results and Insights

- Most support tickets are raised by individuals aged 18–30.
- Female customers show slightly higher satisfaction than male customers.
- The most common ticket types are related to **Technical Issues**, **Cancellations**, and **Refunds**.
- Tickets with **High Priority** tend to result in faster resolutions and higher satisfaction.
- **Email and Phone Call** are the most frequently used support channels.
- Longer resolution times are generally correlated with lower satisfaction scores.

---

12. Conclusion

This project successfully combined data preprocessing, exploratory analysis, machine learning, and dashboard development into a comprehensive solution for analyzing and predicting customer satisfaction. The dashboard serves as a powerful tool for support teams to explore ticket data, understand customer behavior, and take proactive steps to improve service quality.

---

13. Future Recommendations

- Implement Natural Language Processing (NLP) to extract sentiment and topics from ticket descriptions.
- Experiment with ensemble models like Random Forest or XGBoost for improved predictive performance.
- Automate data ingestion and model retraining on a weekly/monthly basis.
- Deploy the model on a cloud platform with a backend API for real-time use by customer service platforms.
- Add role-based access and personalization in the dashboard.

Streamlit link - https://customersatisfactionprediction-d8k8ntysaymkt9upecvxhw.streamlit.app/