PREDICTING CUSTOMER PURCHASE LIKELIHOOD

Project Overview:

I developed a machine learning model to predict whether customers will purchase a product or service based on their demographic and behavioral data. This project involved various stages including data cleaning, feature engineering, exploratory data analysis, model selection, evaluation, and potentially deployment using Python and key data science libraries. The goal was to build an accurate and efficient predictive model to provide insights into the factors influencing purchase decisions.

Problem Statement

Understanding customer behavior and predicting purchase likelihood is crucial for targeted marketing and improving conversion rates. The objective of this project was to predict whether a customer will purchase a product or service based on their personal attributes and behavioral data.

Tools & Technologies

- **Python:** The primary programming language used.
- **Pandas & NumPy:** For data manipulation and analysis.
- **Scikit-Learn:** For building and evaluating machine learning models.
- **Matplotlib & Seaborn:** For data visualization and exploratory data analysis.
- **Flask:** For potential deployment as an interactive web application.

Data Preparation

- **Data Cleaning:** Handled missing values and inconsistencies in the dataset. For example, addressed missing values in categorical variables and imputed numerical data where necessary.
- **Feature Engineering:** Created new features such as:
    - **Age Bins:** Grouped ages into categories (e.g., young, adult, senior).
    - **Behavioral Scores:** Combined various behavioral metrics (e.g., frequency of purchases, time spent on website).
    - **Demographic Features:** Engineered features related to demographics (e.g., family size, income group).
- **Encoding Categorical Variables:** Converted categorical variables (e.g., job, marital status) into numerical format using techniques like one-hot encoding.

Exploratory Data Analysis (EDA)

- Visualized distributions and relationships between features using histograms, box plots, and heatmaps.

- Identified key factors affecting purchase decisions, such as demographic characteristics (age, income), behavioral patterns (purchase history, interaction frequency), and external factors (seasonality, marketing campaigns).

Model Development

- **Algorithm Selection:** Evaluated multiple machine learning algorithms including:
    - Logistic Regression
    - Decision Trees
    - Random Forest
    - Gradient Boosting
- **Model Tuning:** Used techniques like cross-validation and grid search to optimize hyperparameters and enhance model performance.
- **Evaluation Metrics:** Assessed models using metrics such as accuracy, precision, and possibly ROC curves to ensure robust performance.

Results

- **Performance:** The final model achieved an accuracy of XX%, demonstrating its effectiveness in predicting customer purchase behavior.
- **Key Insights:**
    - **Demographic Factors:** Certain demographics showed higher propensity to purchase.
    - **Behavioral Patterns:** Customers with specific behavioral patterns tended to be more likely to make a purchase.
    - **External Influences:** Seasonal variations or promotional activities significantly impacted purchase decisions.

Key Learnings

- **Data Preprocessing:** Importance of handling missing data and engineering relevant features to improve model accuracy.
- **Model Evaluation:** Understanding different evaluation metrics and selecting the best model based on performance trade-offs.
- **Potential Deployment:** Practical experience in deploying a machine learning model for real-world applications and creating an intuitive user interface.

GitHub Repository

Explore the complete project and code here:
https://github.com/sravanthi224/PRODIGY_DS_03.git