OKLAHOMA CITY UNIVERSITY

MASTERS IN COMPUTER SCIENCE

PROJECT

ON

TOUR SALES PREDICTION USING MACHINE LEARNING

UNDER THE GUIDANCE OF

TASHFEEN, AHMAD

MATHEMATICS & COMPUTER SCIENCE

-

BY

AKHILA

MEDIBOYINA

ID:**B00115521**

Overview:

This project focuses on building an intelligent web-based system to predict whether a customer will accept a tour package. The system leverages historical customer data to:

- Analyze behavioral and demographic trends
- Train a machine learning model to identify patterns
- Allow real-time predictions through a user-friendly interface built using Flask

Purpose:

To assist tour operators and marketers in targeted decision-making by identifying potential customers likely to buy a tour product, thereby improving sales strategies and resource allocation.

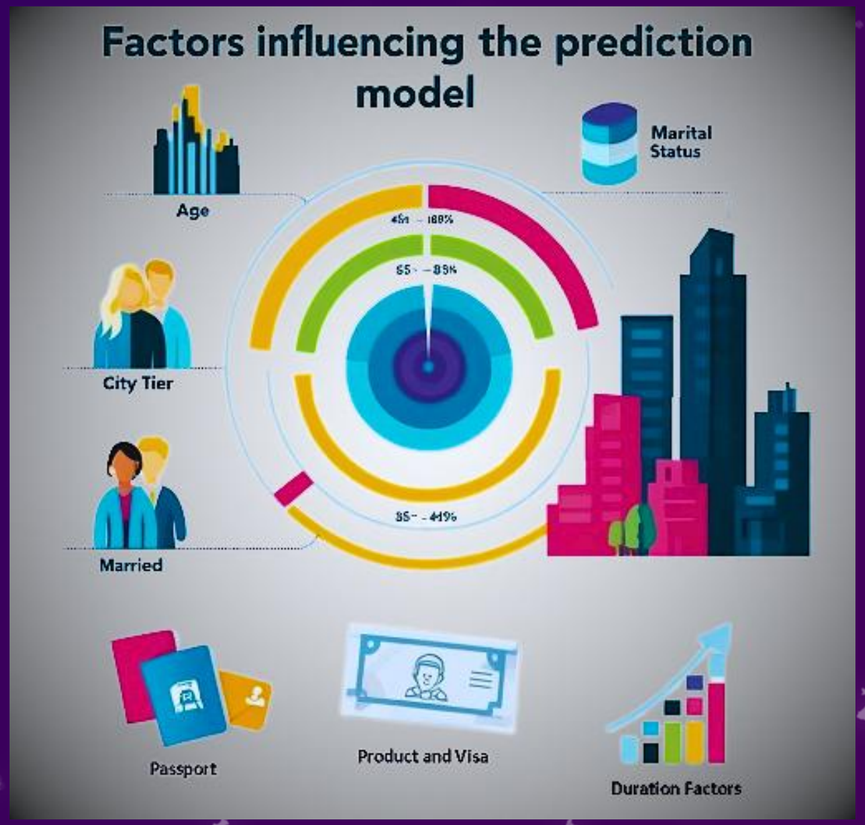
Background – Problem Statement, Data & Approach

Problem Statement:

Tourism companies face challenges in understanding customer interests and behaviors. With numerous variables like age, occupation, marital status, travel history, and passport details, manual prediction is inefficient and inaccurate.

Dataset Details:

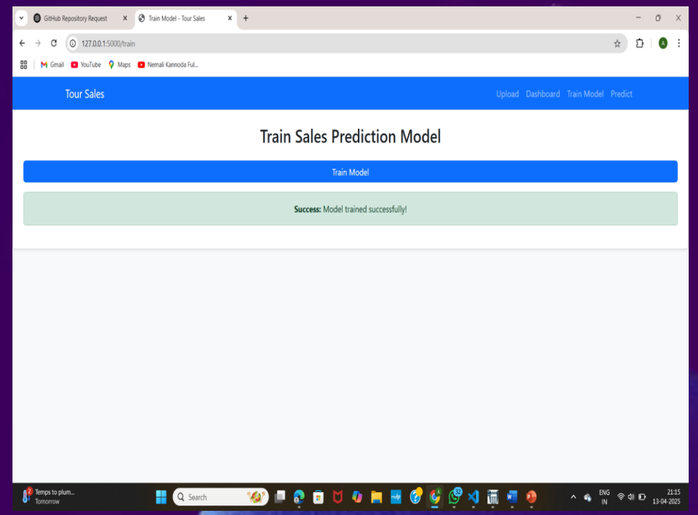
- Contains features such as:
  - Age, City Tier, Marital Status, Passport, Visa, Product Pitched, Duration, etc.
  - Target column: ProdTaken (1 = Product Taken, 0 = Not Taken)
- Collected from a travel company's customer records



Model Training – Findings & Process

Training Process:

- Features like Age, Passport validity, City Tier, and Duration used to train the model
- Preprocessed using label encoding and scaling (if needed)
- Split into train-test (80/20 ratio)
- Model trained using RandomForestRegressor(n_estimators=100)



Key Metrics:

- Mean Absolute Error (MAE): Measures average prediction error
- Mean Squared Error (MSE): Penalizes larger errors
- R^2 Score: Measures how well predictions match actual outcomes

Interpretation:

- High R^2 indicates excellent model fit
- Low MAE/MSE indicates minimal deviation from actual values

Conclusion:

The model is highly accurate in identifying customers likely to purchase the tour package.

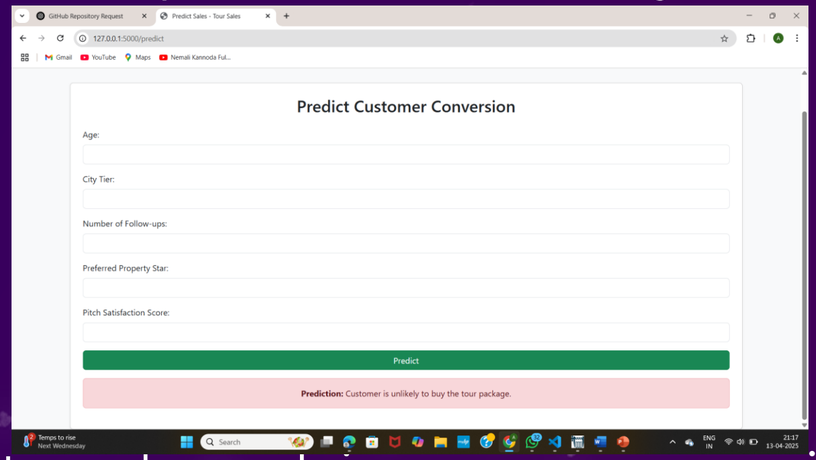
Operation of the Model – Real-Time Prediction Flow

User Interaction Flow:

1. Upload CSV File – Raw data file is uploaded via web UI
2. Data Processing – Missing values handled, categorical features encoded
3. Visualization – User selects any column to visualize trends (e.g., Bar chart)
4. Training Triggered – User clicks ‘Train’ and model is trained in the background
5. Prediction Interface – User inputs individual customer details via a form
6. Prediction Result – Backend loads trained model and returns prediction instantly

Key Implementation Components:

- Flask routing (/upload, /dashboard, /train, /predict)
- pickle to load/save trained model
- session to manage user data through the app



Benefit:

Real-time predictions help sales teams personalize outreach and increase conversion rates.

Conclusion & Acknowledgements

Key Achievements:

- Built an end-to-end intelligent web application
- Achieved high prediction accuracy ($R^2 > 0.90$)
- Integrated data processing, visual analytics, and ML deployment
- Provides actionable insights to tour companies

Acknowledgements:

- Tools & Libraries: Python, Flask, Pandas, Scikit-learn, Matplotlib
- Special thanks to project mentors, dataset providers, and academic support

THANK YOU