Car Budget Prediction using Machine Learning

Student Name : S.Sathvik Reddy
Roll Number : 225U1A05F6
Course Name & Code : Computer Science & CSE
Department Name : Computer Science
Guide Name : Dr.Madhu Sir
Institution Name : AVNIET

Date : 11-June-2025

Introduction

- This project predicts how much a customer might spend on purchasing a car using ML techniques.
- Motivation: To help automotive businesses better understand customer buying capacity and optimize targeting.
- Relevance: Useful in automotive sales, financial pre-approvals, and personalized marketing.

Problem Statement

- Problem: Estimating car purchase amount from customer demographics and financial data.
- Objective: Build a model that takes input like age, salary, net worth, etc., and predicts the purchase value.
- Scope: Focuses only on regression models using tabular customer data. Doesn’t include car features or brand preferences.

Dataset Description

- Source: Kaggle (Car Purchasing Dataset)
- Size: ~500 records in .csv format
- Key Features:
  - Input: Gender, Age, Annual Salary, Credit Card Debt, Net Worth
  - Output: Car Purchase Amount
- Preprocessing:
  - Dropped irrelevant columns (Name, Email)
  - Converted categorical variables
  - Scaled numerical features using StandardScaler

ML Model(s) Used

- Algorithm: Random Forest Regressor
- Why Random Forest?:
    - Handles non-linear data well
    - Less prone to overfitting
- Libraries/Tools:
    - pandas, numpy, scikit-learn, streamlit

Results & Evaluation

- Performance:

  - Mean Squared Error (MSE): 5.41 million

  - R^2 Score: 0.95 → Model explains 95% of the variance!

- Visualization: Model output vs actual purchase values
- Interpretation: Model performs very well on unseen data

Comparative Analysis

- Other models tested: Linear Regression (lower R^2), Decision Tree (less stable)
- Best Performer: Random Forest due to its ensemble nature and robustness
- Showcased better accuracy and generalization on test set

Conclusion

- Successfully built a car purchase predictor using financial and demographic data.
- High performance with $R^2 = 0.95$
- Could be extended with more features like car type, region, or historical purchases.
- App version created using Streamlit for real-time prediction.