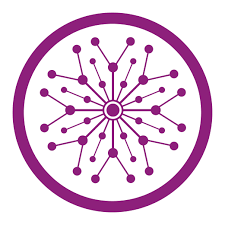
**

**The Superior University**

**Project Title**

Car Price Prediction

**Project Details**

1. Course: Artificial Intelligence Lab
2. Instructor: Sir Rasikh Ali
3. Semester: 3rd
4. Section: BSAI-3B

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Roll No** | **Email** | **Contact** |
| Muhammad Fasih Ud Din | BSAIM-F23-094 | su92-bsaim-f23-094@superior.edu.pk |  |
|  |  |  |  |

Contents

[1. Introduction 2](#_Toc184647264)

[2. Objectives 2](#_Toc184647265)

[3. System Requirements 2](#_Toc184647266)

[4. Methodology 3](#_Toc184647267)

[5. Challenges and Solutions 3](#_Toc184647268)

[6. Results: 3](#_Toc184647269)

[7. Conclusion 4](#_Toc184647270)

**Abstract**

This project focuses on predicting the selling price of cars based on their features, such as age, present price, kilometers driven, fuel type, seller type, transmission type, and ownership. Using the scikit-learn Linear Regression model, the project provides a data-driven approach to estimate car prices. The model was trained on a preprocessed dataset and achieved competitive accuracy, demonstrating the utility of regression techniques in predictive analytics.

# Introduction

Car price prediction is crucial for buyers and sellers to make informed decisions. This project uses Linear Regression, a fundamental machine learning technique, to build a predictive model for car prices. The dataset consists of historical data about cars, including their attributes and selling prices. By analyzing these features, the model predicts selling prices for new data inputs.

# Objectives

* Build a predictive model using Linear Regression.
* Preprocess the dataset, including feature engineering and encoding categorical variables.
* Evaluate the model using metrics like Mean Squared Error (MSE) and R-squared score.
* Demonstrate the relationship between car features and selling prices.

# System Requirements

Hardware Used:

* + Processor: Intel i5 8th Gen
  + RAM: 8 GB

Software:

* + Operating System: Windows 11
  + Python 3.8, Visual Studio Code or any other
* Libraries: Pandas,Numpy, Matplotlib, sns, sklearn.svc, metrices, linear regression.

# Methodology

1. **Data Understanding and Preprocessing**:
   * Analyzed dataset features: Year, Present Price, Kms Driven, Fuel Type, Seller Type, Transmission, and Owner.
   * Encoded categorical variables (Fuel\_Type, Seller\_Type, and Transmission) using one-hot encoding.
2. **Feature Selection**:
   * Selected relevant features ( Present\_Price, Kms\_Driven, Fuel\_Type, Seller\_Type, Transmission, and Owner) for the model.
3. **Model Training**:
   * Split the dataset into training and test sets (80%-20% split).
   * Trained the Linear Regression model on the training set.
4. **Evaluation**:
   * Assessed model performance using MSE and R-squared metrics on the test set.

# Challenges and Solutions

1. **Multicollinearity**:
   * Challenge: High correlation between certain features (e.g., Present\_Price and Fuel\_Type).
   * Solution: Analyzed correlation matrix to exclude redundant features.
2. **Categorical Encoding**:
   * Challenge: Ensuring proper handling of categorical variables.
   * Solution: Used one-hot encoding with drop\_first=True to avoid the dummy variable trap.
3. **Scaling**:
   * Challenge: Linear Regression is sensitive to feature magnitudes.
   * Solution: Standardized numerical features to improve model performance.

# Results:

**R-Squared Score: 0.8365766715026374**

# Conclusion

This project successfully predicts car selling prices using Linear Regression. The approach highlights the importance of preprocessing and feature engineering in machine learning projects.