# Car Price Predictor Platform

## MINI PROJECT – I SYNOPSIS



Department of Computer Science & Application

## Institute of Engineering & Technology

SUBMITTED TO: - SUBMITTED BY: -

Mr. Mandeep Singh Srasthi Shukla (201500706) (Technical Trainer) Harsh Mehrotra (201500270)

# Acknowledgement

It gives us a great sense of pleasure to present the synopsis of the B.Tech Mini project undertaken during B.Tech III Year. This project is going to be an acknowledgement to the inspiration, drive and technical assistance will be contributed to it by many individuals. We owe special debt of gratitude to Mr. Mandeep Singh, Technical Trainer , for providing us with an encouraging platform to develop this project, which thus helped us in shaping our abilities towards a constructive goal and for his constant support and guidance to our work.

His sincerity, thoroughness and perseverance has been a constant source of inspiration for us. We believe that he will shower us with all his extensively experienced ideas and insightful comments at different stages of the project & also taught us about the latest industry-oriented technologies. We also do not like miss the opportunity to acknowledge the contribution of all faculty members of the department for their kind guidance and co-operation.

Srasthi Shukla (201500706)

Harsh Mehrotra (201500270)

## ABSTRACT

The study described in this research report focused on variables which were posited to capture students’ experiences of the online tutoring service, e-Learning, and relationships with the students’ perceptions of their academic capabilities and academic performance. A theoretical model incorporating variables from the Technology Acceptance Model, the Theory of Planned Behaviour, and Social Cognitive Theory was developed and tested. A total of 506 undergraduate students from a university located in Sydney, Australia, completed an online survey. Data were analysed using confirmatory factor analysis (CFA) and structural equation modelling (SEM). The results suggested that the perceived usefulness of E-Learning had a direct positive relationship with academic self-efficacy, and an indirect positive association with the students’ academic grades through academic self-efficacy. There was a direct positive relationship between academic self-efficacy and students’ academic grades. The implications of these results and directions for future research are discussed in this report.

# Contents

Abstract Declaration Acknowledgement

1. Introduction
   1. Objective
   2. Motivation
   3. Problem Statement
2. Requirement
   1. Hardware Requirements
   2. Software Requirements
3. Project Description
   1. Future Work
4. Working
5. Implementation
6. References

# INTRODUCTION

Determining whether the listed price of a used car is a challenging task, due to the many factors that drive a used vehicle’s price on the market. The focus of this project is developing machine learning models that

can accurately predict the price of a used car based on its features, in order to make informed purchases. We implement and evaluate various learning methods on a dataset consisting of the sale prices of

different makes and models across cities in the United States. Our results show that Random Forest model and K-Means clustering with linear regression yield the best results, but are compute heavy.

Conventional linear regression also yielded satisfactory results, with the advantage of a significantly lower training time in comparison to the aforementioned methods.

**OBJECTIVE**

To develop an efficient and effective model which predicts the price of a used car according to user’s inputs. To achieve good accuracy.To develop a User Interface( UI ) which is user-friendly and takes input from the user and predicts the price.

**MOTIVATION**

Deciding whether a used car is worth the posted price when you see listings online can be difficult. Several factors, including mileage, make, model, year, etc. can influence the actual worth of a car. From the perspective of a seller, it is also a dilemma to price a used car appropriately[2-3]. Based on existing data, the aim is to use machine learning algorithms to develop models for predicting used car prices.

## SOFTWARE AND HARDWARE REQUIREMENTS

* **Hardware requirements**
* Operating system- Windows 7,8,10
* Processor- dual core 2.4 GHz (i5 or i7 series Intel processor  
  or equivalent AMD)
* RAM-4GB
* **Software Requirements**
* Python
* Pycharm
* PIP 2.7
* Jupyter Notebook
* Chrome

## PROJECT DESCRIPTION

To develop an efficient and effective model which predicts the price of a used car according to user’s inputs. To achieve good accuracy.To develop a User Interface( UI ) which is user-friendly and takes input from the user and predicts the price.

**FUTURE SCOPE**

In future this machine learning model may bind with various website which can provide real time data for price prediction. Also we may add large historical data of car price which can help to improve accuracy of the machine learning model. We can build an android app as user interface for interacting with user. For better performance, we plan to judiciously design deep learning network structures, use adaptive learning rates and train on clusters of data rather than the whole dataset.

## WORKING & IMPLEMTATION

There are two primary phases in the system:

1. Training phase: The system is trained by using the data in the data set and fits a model (line/curve) based on the algorithm chosen  
accordingly.

2. Testing phase: the system is provided with the inputs and is tested for its working. The accuracy is checked. And therefore, the data that is used to train the model or test it, has to be appropriate. The system is designed to detect and predict price of used car and hence  
appropriate algorithms must be used to do the two different tasks. Before the algorithms are selected for further use, different algorithms were compared for its accuracy. The well-suited one for the task was chosen.

The next step is to do Data Preprocessing which includes Data cleaning, Data reduction, Data Transformation. Then, using various machine learning algorithms we will predict the price. The algorithms involve Linear Regression, Ridge Regression and Lasso Regression. The best model which predicts the most accurate price is selected. After selection of the best model the predicted price is displayed tothe user according to user’s inputs. User can give inputthrough website to for used car price prediction to machinelearning model.

## Linear Regression

Linear Regression attempt to model the relationship between two variables by fitting a linear equation to observed data. The other is considered to be dependent variable. For Example: A modeler might want to relate weights of individuals to their heights using a linear regression model

Linear regression is useful for finding relationship between multiple continuous variables

There are multiple independent variables and single independent variable

y = m1X1+m2X2+......+b

m1, m2, m3 ....  slope

b  y intercept

X1, X2, X3 ......  independent variables

y  dependent variables

## Random Forest

Random Forest is an ensemble learning based regression model. It uses a model called decision tree, specifically as the name suggests, multiple decision trees to generate the ensemble model which collectively produces a prediction. The benefit of this model is that the trees are produced in parallel and are relatively uncorrelated, thus producing good results as each tree is not prone to individual errors of other trees. This uncorrelated behavior is partly ensured by the use of Bootstrap Aggregation or bagging providing the randomness required to produce robust and uncorrelated trees. This model was hence chosen to account for the large number of features in the dataset and compare a bagging technique with the following gradient boosting methods.

## REFERENCES;

* https://www.kaggle.com/jpayne/852k-used-car-listings
* N. Monburinon, P. Chertchom, T. Kaewkiriya, S. Rungpheung, S. Buya and P. Boonpou,
* "Prediction of prices for used car by using regression models," 2018 5th International Conference on Business and Industrial Research (ICBIR), Bangkok, 2018, pp. 115-119.
* Listiani M. 2009. Support Vector Regression Analysis for Price Prediction in a Car Leasing Application. Master Thesis. Hamburg University of Technology
* Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. ACM, 2016.
* Ke, Guolin, et al. "Lightgbm: A highly efficient gradient boosting decision tree." Advances in Neural Information Processing Systems. 2017.
* Fisher, Walter D. "On grouping for maximum homogeneity." Journal of the American statistical Association 53.284 (1958): 789-798.
* https://scikit-learn.org/stable/modules/classes.html: Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830

## Websites:

* [www.google.com](http://www.google.com/)
* [www.youtube.com](http://www.javawrench.com/)
* [www.Kaggle.com](http://www.Kaggle.com)
* [www.openml.](http://www.openml.)org
* [www.quora.com](http://www.quora.com)
* [www.github.com](http://www.github.com)
* [www.stackoverflow.com](http://www.stackoverflow.com)

## Faculty Guidelines:

Mr. Mandeep Singh (Technical Trainer in GLA University)

## GitHub Repository link:

https://github.com/Mehrotra01/Mini-Project1