

**Project Design Phase-I  
Proposed Solution Template**

S No.	Parameter	Description
1.	Problem Statement	The main aim of this project is to predict the price of used cars using the various Machine Learning (ML) models. This can enable the customers to make decisions based on different inputs or factors namely Brand or Type of the car one prefers like Ford, Hyundai, Model of the car namely Ford Figo, Hyundai Creta, Year of manufacturing like 2020, 2021, Type of fuel namely Petrol, Diesel, Price range or Budget, Type of transmission which the customer prefers like Automatic or Manual, Mileage to name a few characteristic features required by the customer. This project Car Price Prediction deals with providing the solution to these problems. Different techniques like multiple linear regression analysis, k-nearest neighbours, naïve bayes and decision trees have been used to make the predictions. The predictions are then evaluated and compared in order to find those which provide the best performances.
2.	Feasibility of Idea	New cars of a particular make, model, and year all have the same retail price, excluding optional features. This price is set by the manufacturer. Used car, however, are subject to supply-and-demand pricing. Further, used cars have additional attributes that factor into the price. These include the condition, milage, and repair history, which sets cars that may have shared a retail price apart.
3.	Novelty	The purpose of this thesis is to evaluate several different machine learning models for used car price prediction and draw conclusions about how they behave. This will deepen the knowledge of machine learning applied to car valuations and other similar price prediction problems.

4.	Social Impact / Customer Satisfaction	This work will focus on answering the research questions. They all entail a comparison of different ML algorithms for price prediction. This will be accomplished by sourcing and preparing a dataset on which all the algorithms can be trained on and compared fairly. The algorithms selected must therefore be similar enough for the same dataset to be used for all of them. This also means that no large optimization efforts on the dataset will be made to boost the performance, if these changes do not benefit the other models. Maximizing price prediction performance of any one algorithm in ways that do not offer better comparisons is outside the scope of this work.
5.	Business Model (Revenue Model)	A revenue model is a blueprint that shows how a startup business will earn revenue or gross income from its standard business operations, and how it will pay for operating costs and expenses.
6.	Scalability of the Solution	<p>Which of the models and parameters gives the best overall accuracy in making price predictions for used cars. The optimal parameters were determined in the process of implementing the models, and thus each model was implemented with the parameters that yielded the best performance by trial and error.</p> <p>All of the models approximated geometric appreciation, meaning that a constant percentage of value is lost every year independent of the age of the vehicle. Random Forest Regression had a significantly higher assessed average depreciation at approximately 13.8%, compared to the others with 9.7%. This is closer to the range of 15%-31% assessed by Karl Storchmann in his analysis of international depreciation rates</p>