Dr B R Ambedkar National Institute of Technology, Jalandhar

Minor Project -I

Report: AUTO-GAADI

Department of computer science and Engineering

Batch: 2020 -2024

Mentored by:

Mr. Rahul Agarwal

(Assistant Professor)

Presented by:

Group: 14

Md Iftear Ahmed (20103089)

Paras Gupta (20103107)

Smriti Chaudhary (20103140)

Table Of Contents

ACKNOWLEDGEMENT

*The success of our project, **AUTO-GAADI**, would not have been possible without the unwavering support and guidance from numerous individuals. We, Group 14, are grateful to have received assistance from a variety of sources throughout our project journey, and we would like to express our heartfelt appreciation to all those who have helped us.*

We extend our deepest gratitude to Professor Rajneesh rani, Head of the Department of Computer Science & Engineering, for his unwavering support and guidance. We are also grateful to the In-charge of Minor Project 3rd Year for providing us with a mentor and all other necessary support.

Our mentor, Mr. Rahul Agarwal, Assistant Professor, played a pivotal role in our project's success. he believed in our idea and offered invaluable suggestions whenever required, supporting us throughout the project.

In conclusion, we recognize that there were countless individuals who contributed to the success of our project. We are humbled by their support and grateful for the opportunity to work with such dedicated and supportive individuals.

Thank you.

[Group – 14]

DECLARATION

*As members of Group 14, we affirm that our project, entitled " **AUTO-GAADI** " submitted to the Department of Computer Science and Engineering, is the result of our own efforts under the excellent guidance of Mr. Rahul Agarwal. We confirm that the project has not been copied from any other source and has been solely created by our team.*

We will be solely responsible if some Plagiarism is found.

Thank You All.

Date: 12th May, 2023

[Group 14]

Plagiarism Check

*Group 14 has utilized the services of Turnitin to assess the level of plagiarism in our Project Report on tour project "**AUTO-GAADI**". The team expresses gratitude towards their mentor, Dr. Rajneesh Rani, for providing guidance throughout the process. The digital receipt generated by Turnitin confirms that the plagiarism in the report is below 10%.*

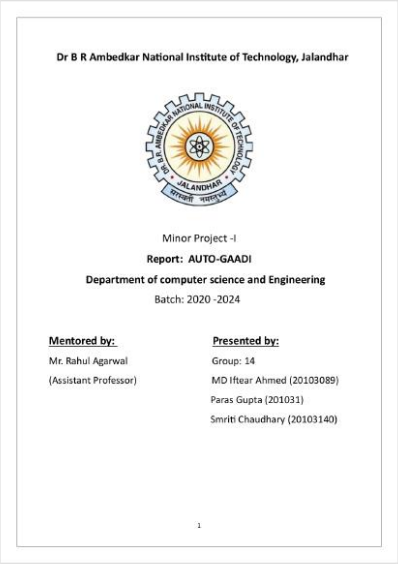
turnitin

Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	MD IFTEAR AHMED
Assignment title:	For Chatgpt or AI Subscription just contact by email hubxc...
Submission title:	Minor Project - 1
File name:	minor_project_-1,_group14.pdf
File size:	1.35M
Page count:	33
Word count:	5,866
Character count:	33,178
Submission date:	11-May-2023 02:35AM (UTC-0400)
Submission ID:	2090194783



Dr B R Ambedkar National Institute of Technology, Jalandhar

Minor Project -I

Report: AUTO-GAADI

Department of computer science and Engineering

Batch: 2020 -2024

Mentored by:
Mr. Rahul Agarwal
(Assistant Professor)

Presented by:
Group: 14
MD Iftear Ahmed (20103089)
Paras Gupta (201031)
Smriti Chaudhary (20103140)

1

INTRODUCTION - Building Up the Foundation

1.1 Background

The background of the "**AUTO-GAADI**" project would involve data analysis and machine learning techniques to predict the price of a car and analyze the trends in car sales.

To carry out this project, you would need access to a large dataset of car sales information, including factors such as car make, model, year, mileage, condition, location, and price. This data can be collected from various sources such as online car sales platforms, car dealerships, and public records.

The project would involve data cleaning and preparation to ensure that the data is accurate and consistent. This would include removing any missing or duplicate data, converting data into a standardized format, and identifying any outliers or anomalies.

Once the data is prepared, machine learning algorithms can be trained on the data to predictcar prices. This may involve using regression techniques such as random forest regression, which can identify the most important factors affecting car prices.

In addition to predicting car prices, the project may also involve analyzing car sales trends over time. This could involve creating visualizations of sales data, identifying seasonal patterns, and analyzing factors that affect sales, such as economic conditions or changes in consumer preferences.

Overall, the project would require a strong understanding of data analysis and machine learning techniques, as well as a background in programming and statistics.

1.2 Literature survey

To conduct a literature survey on the innovation of an **AUTO-GAADI**, we searched various sources available on the internet. Additionally, we explored the existing systems of several company to gain more insight into the topic. The results of our survey are presented below.

There is an online platform **TrueCar** that provides pricing transparency and market insights to help consumers and dealers make informed car buying and selling decisions. It has many features divided into four main modules namely- New Car, Used Car, sell your cars, Research (car Reviewing, car ranking, compare cars)

Another one is **CarGurus -** an online automotive marketplace that provides pricing insights, dealer reviews, and car valuation tools to help buyers and sellers make informed decisions.

CarGurus is a company based in Cambridge, Massachusetts that operates an automotive research and shopping website. Its platform helps users to compare local listings of both used and new cars and enables them to contact sellers.

The literature survey would identify gaps in the existing research and provide recommendations for future research. It would also evaluate the current trends and practices in the automotive industry related to car pricing and sales analysis, such as consumer preferences, environmental regulations, and technological advancements.

Overall, the literature survey would provide a comprehensive understanding of the current state of knowledge and research related to car price prediction and sales analysis. It would help to identify gaps in the existing research and provide insights and recommendations for our project.

1.3 Problem Statement and it's Necessity

The major problems estimated are:

1. **The automotive industry is highly competitive**

The automotive industry is a fiercely competitive space, with new models and brands constantly entering the market. To succeed in this industry, car dealerships need to have a deep understanding of the factors that impact car pricing and sales. The challenge, however, is that there are numerous factors that can influence these outcomes, including consumer demand, economic conditions, and competitor behavior. Therefore, car dealerships need an advanced data analytics solution that can help them predict car.

2. **Car price prediction is a crucial element in the car dealership's pricing strategy.**

It enables them to optimize the price of cars by analyzing the market demand and competitor pricing. Inaccurate pricing can lead to underpricing or overpricing, which can result in missed sales opportunities and lower profits. Furthermore, with the proliferation of online marketplaces, car buyers can easily compare prices, making it critical for dealerships to have a competitive and data-driven pricing strategy. By leveraging advanced data analytics techniques, such as machine learning algorithms and statistical modeling, car dealerships can predict car prices with greater accuracy and create optimal pricing strategies.

3. **Car sales analysis is equally critical for car dealerships to identify patterns in customer behavior and preferences.**

Analyzing customer data can help dealerships gain insights into the types of cars customers are interested in, the features they look for in a car, and their buying behavior. By analyzing this data, dealerships can tailor their marketing and sales strategies to maximize their revenue. For instance, they can develop targeted advertising campaigns, offer personalized recommendations, and improve their customer service. Ultimately, this can help dealerships stay ahead of the competition by providing a superior customer experience and increasing customer loyalty.

4. Traditional methods of determining car pricing and sales strategies

Traditional methods of determining car pricing and sales strategies rely on incomplete or outdated data, making it difficult for dealerships to make informed decisions. For instance, many dealerships use a manual process of collecting and analyzing data from various sources,such as customer feedback and sales reports. However, this process can be time-consuming and prone to errors, making it challenging to stay ahead of the competition. By using advanceddata analytics techniques, such as predictive modeling and machine learning algorithms, dealerships can gain deeper insights into car pricing and sales trends, allowing them to makemore informed decisions.

5. car dealerships must leverage advanced data analytics solutions to remain competitive

In today's increasingly digital and data-driven economy, car dealerships must leverage advanced data analytics solutions to remain competitive. By using these tools, they can gain valuable insights into market demand, pricing trends, and customer behavior. For instance, car dealerships can analyze data from social media platforms, such as Twitter and Facebook,to understand customer sentiment about a particular car model or brand. Additionally, they can leverage data from customer reviews to gain insights into the features that customers arelooking for in a car. By integrating these data sources, dealerships can develop a more comprehensive understanding of their customers, which can help them make data-driven decisions and stay ahead of the competition.

6. advanced data analytics solutions, car dealerships can gain a deeperunderstanding of market demand.

Car price prediction and sales analysis are critical for dealerships to optimize their operationsand provide better value to their customers. With these advanced data analytics solutions, car dealerships can gain a deeper understanding of market demand, competitor pricing, and customer behavior, enabling them to create more effective pricing and sales strategies. Ultimately, this can help dealerships increase their revenue, improve customer satisfaction, and build a loyal customer base. Furthermore, with the proliferation of digital technologies and data analytics, car dealerships that fail to embrace these tools risk falling behind their competitors and losing market share.

1.4 Feasibility- Technical and Non-Technical

To ensure the success of a project, it is essential to evaluate its feasibility. There are several types of feasibility that need to be considered, including

❖ Technical Feasibility:

- This is a web application; it can access easily trough mobile phone, Desktop, laptop, Tablet.
- It will run any kind of operating system like (Linux, Windows, Android)
- Internet connectivity is required for the system.
- It contains HTML, CSS, Bootstrap, Python Programming language.

❖ Economic Feasibility:

- The project's economic feasibility will be evaluated to determine its profitability and financial viability.
- The cost of development, implementation, and maintenance of the system will be calculated.
- The potential return on investment (ROI) will be analyzed.

❖ Operational Feasibility:

- The project's operational feasibility must be evaluated to ensure that it aligns with the existing operations, policies, and processes of the organization.
- The feasibility of implementing the system within the organization's existing structure and culture will be assessed.

❖ Legal Feasibility:

- The project's legal feasibility will be evaluated to ensure that it complies with relevant laws, regulations, and standards.
- The potential legal risks associated with the project will be analyzed, and measures will be taken to mitigate them.

PROPOSED SOLUTION

2.1 Identifying solution

The solution to the AUTO-GAADI Problem Statement involves building a web application that uses data analytics and machine learning techniques to provide accurate car price predictionand analysis of car sales trends. The following are the solutions to the problem statement:

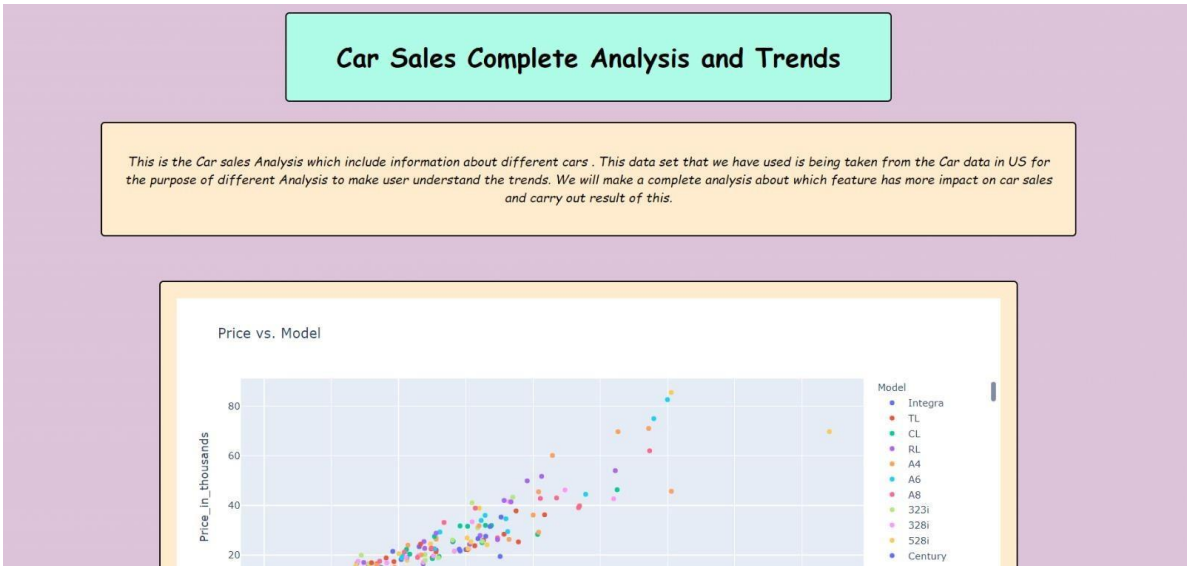
1. Taking a comprehensive dataset:

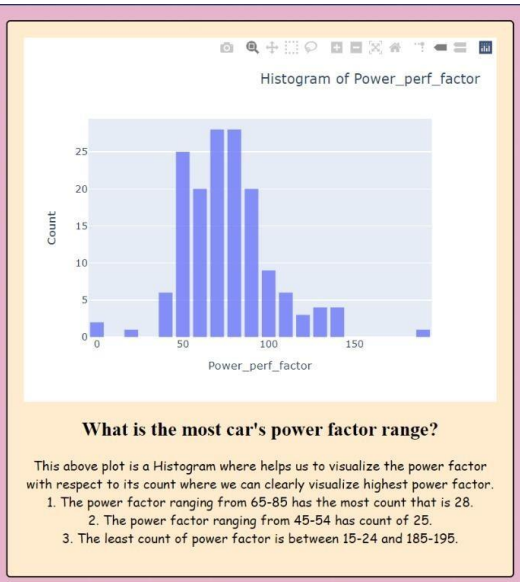
Building a comprehensive database is the first step towards building an effective AUTO-GAADI system. The database will need to include a wide range of information about car models, makes, and years to ensure that the predictions are accurate. The database will also need tobe updated continuously to reflect changes in the market, new car models, and other relevantfactors.

	A	B	C	D	E	F	G	H	I	J
1	car	price	body	mileage	engV	engType	registratio	year	model	drive
2	Ford	15500	crossover	68	2.5	Gas	yes	2010	Kuga	full
3	Mercedes-	20500	sedan	173	1.8	Gas	yes	2011	E-Class	rear
4	Mercedes-	35000	other	135	5.5	Petrol	yes	2008	CL 550	rear
5	Mercedes-	17800	van	162	1.8	Diesel	yes	2012	B 180	front
6	Mercedes-	33000	vagon	91	NA	Other	yes	2013	E-Class	
7	Nissan	16600	crossover	83	2	Petrol	yes	2013	X-Trail	full
8	Honda	6500	sedan	199	2	Petrol	yes	2003	Accord	front
9	Renault	10500	vagon	185	1.5	Diesel	yes	2011	Megane	front
10	Mercedes-	21500	sedan	146	1.8	Gas	yes	2012	E-Class	rear
11	Mercedes-	22700	sedan	125	2.2	Diesel	yes	2010	E-Class	rear

2. Developing a machine learning model:

Developing a machine learning model is the second step towards building an effective system.The machine learning model will be trained on historical sales data, market trends, and otherrelevant factors to make accurate predictions. The model will need to be tested extensively toensure that it is accurate and reliable.



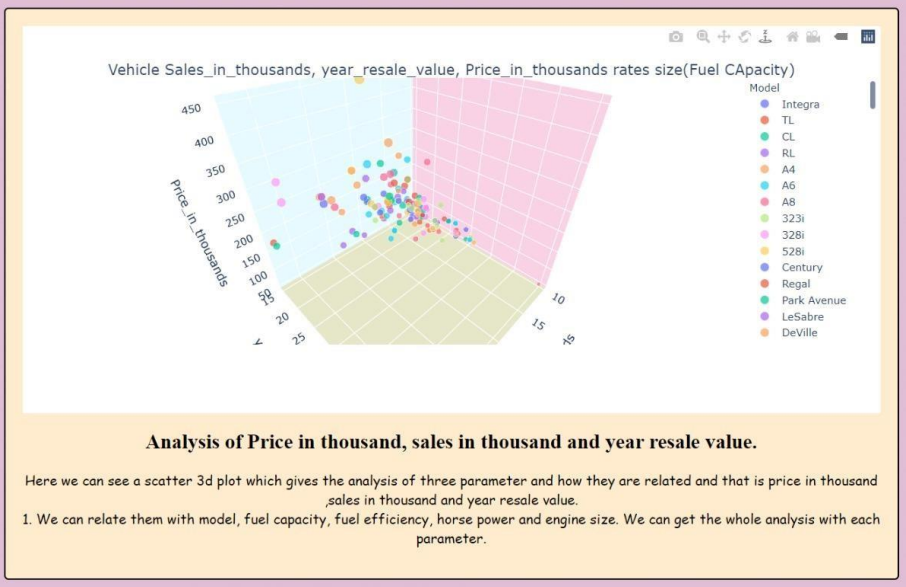


What is the most car's power factor range?

This above plot is a Histogram where helps us to visualize the power factor with respect to its count where we can clearly visualize highest power factor.
1. The power factor ranging from 65-85 has the most count that is 28.
2. The power factor ranging from 45-54 has count of 25.
3. The least count of power factor is between 15-24 and 185-195.

Which model has the most power factor?

The Histogram clearly states the model with highest power factor. Lets have a detailed analysis.
1. The highest Power factor is of Model Viper.
2. There are most models with power factor around 70 and 80.
3. The lowest power factor is of Town & country and CL.
4. The second lowerst power factor is of Metro.
5. The second highest power factor is of Corvette.



Analysis of Price in thousand, sales in thousand and year resale value.

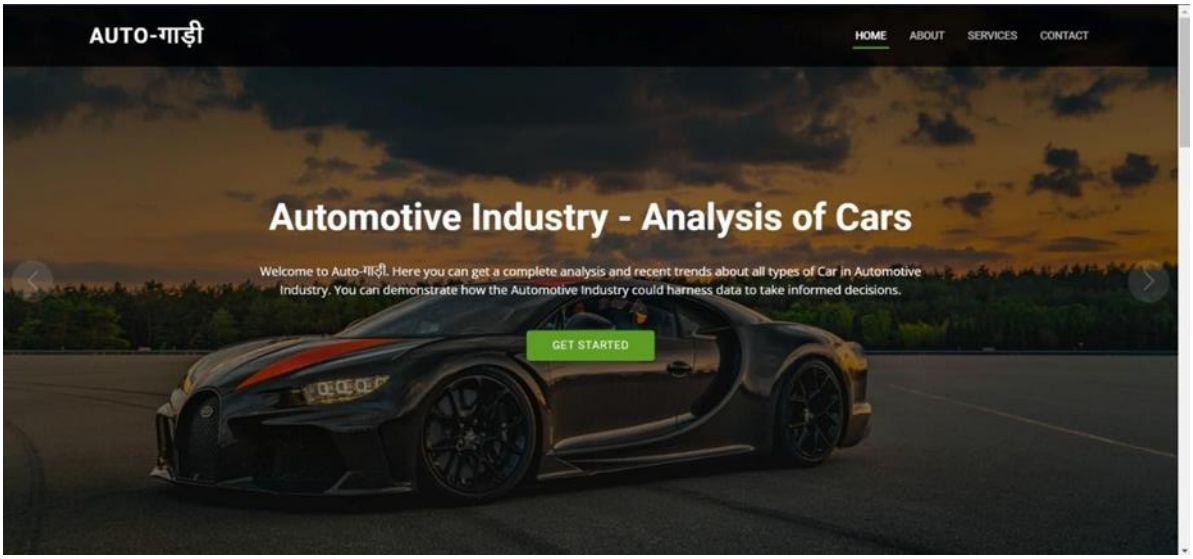
Here we can see a scatter 3d plot which gives the analysis of three parameter and how they are related and that is price in thousand ,sales in thousand and year resale value.
1. We can relate them with model, fuel capacity, fuel efficiency, horse power and engine size. We can get the whole analysis with each parameter.

3. Integration with other systems:

Integration with other systems is the third step towards building an effective system. The system will need to integrate with other systems such as car listing websites, dealer management systems, and customer relationship management systems to obtain real-time data and improve the accuracy of predictions. This integration will also allow the system to provide valuable insights into the car sales trends, including the average price of similar cars, demand and supply trends, and more.

4. User-friendly interface:

A user-friendly interface is essential for the success of the system. The interface should be easy to use and provide users with a clear understanding of the car sales trends and the pricesof similar cars. The interface should also allow users to input information about their cars andreceive an accurate price prediction. Additionally, the interface should provide users with valuable insights into the car sales trends, including the average price of similar cars, demandand supply trends, and more.



5. Integration with payment systems:

Integration with payment systems is also essential for the success of the system. The system should integrate with payment systems to enable users to complete the transaction seamlessly once a buyer is found. This will make it easier for users to sell their cars without having to deal with the hassle of payment processing.

6. Output:

2.2 consider some factors of solution of problem statement

- One important factor to consider is data quality. The accuracy and reliability of the predictions generated by the system will depend on the quality of the data used to train the machine learning model. Therefore, it's essential to ensure that the data is accurate, relevant, and up-to-date. This may involve verifying the data from multiple sources and using data cleansing techniques to remove any inaccuracies.

- Another factor to consider is the privacy and security of user data. The system will need to collect and store sensitive user data, such as personal and financial information. Therefore, it's essential to implement robust security measures to protectthis data from unauthorized access and data breaches. This may involve implementing encryption, firewalls, access controls, and regular security audits.

- Ensuring the system's scalability and flexibility is also crucial. The system should be able to handle a large volume of data and traffic, and it should be flexible enough to adapt to changing market trends and customer needs. This may involve using cloud-based infrastructure, implementing load balancing techniques, and designing the system with modularity and extensibility in mind.

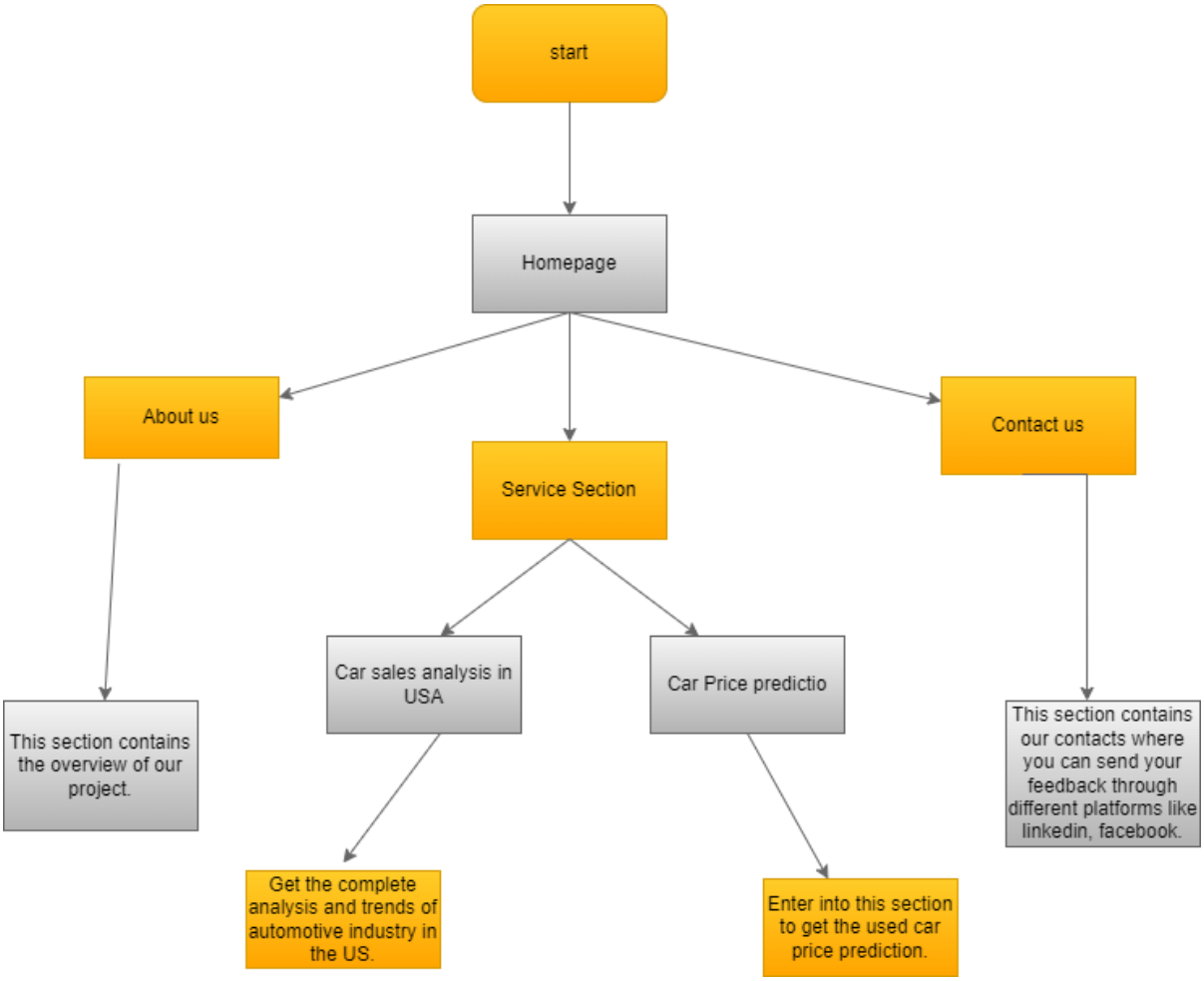
- Another important aspect to consider is regulatory compliance. The system should comply with various laws and regulations, such as data privacy regulations, consumer protection laws, and anti-money laundering laws. Failure to comply with these regulations could result in legal and financial consequences, such as fines and reputational damage.

Finally, the success of the system will depend on its adoption and usage by car sellersand buyers. Therefore, it's essential to develop effective marketing and user engagement strategies to promote the system's features and benefits. This may involve partnering with car dealerships, advertising on relevant websites and social media platforms, and providing incentives for users to use the system

TECHNICAL ANALYSIS

3.1 Flow chart

A flowchart is a graphical representation of a process or system that shows the sequence of steps involved in completing a task. In the context of the AUTO-GAADI system, a flowchart canhelp visualize the steps involved in predicting car prices and analyzing car sales. Here is a high-level overview of a possible flowchart for this system:



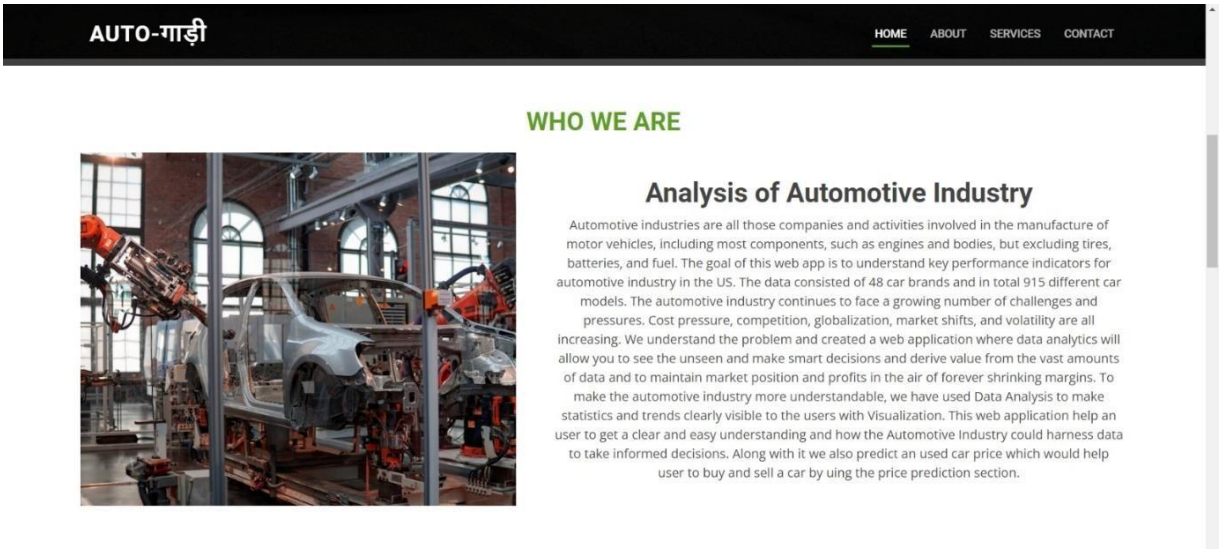
Explaining steps of flow chart:

Step1: This is the initial step for entering into web application.

Step2: After entering there are three sections divided under homepage section.

- About us
- Service section
- Contact us

Step3: About us section contains the overview of us, and our project.



Step4: service section divided into 2 steps again

- Car sells analysis in USA
- Car price prediction.

Step5: get the complete analysis and trends of automotive industry in the USA under car sells analysis section.

Step6: Enter into this section to get the used car price prediction under car price prediction section.

Step7: Contact us section contains our contacts where you can send your feedback through different online platform (link din, Instagram, Facebook)

3.2 Tech Stack Analysis:

We have utilized a diverse range of technology stacks to attain multiple solutions, and the selection of these technologies was based on several key criteria. These include:

1. **Ease of use and learnability**: We considered technologies that are user-friendly and have a short learning curve, making it easier for our team to work with them.
2. **Time to build:** We chose technologies that enable us to develop solutions quickly and efficiently, minimizing development time and allowing us to bring products to market faster.
3. **Efficiency:** We looked for technologies that are efficient in terms of performance, memory usage, and processing power, enabling us to create high-performing solutions.
4. **Security:** We prioritized technologies that have robust security features to ensure the protection of our solutions and the data they handle.

On the basis of the above mentioned criteria the Technologies used are:-

1.Web Designing- HTML, CSS, Bootstrap

Here, we have used HTML (Hypertext Markup Language) is a markup language to create the structure and content of web pages. It provides a way to define the different elements of a web page, such as headings, paragraphs, links, and images.

Next is CSS (Cascading Style Sheets) is a stylesheet language used to define the presentation and layout of web pages. It allows developers to control the appearance of HTML elements, such as the color, font, size, and positioning.

Last, we have used Bootstrap that is a popular front-end development framework that provides a set of pre-built CSS components, such as forms, buttons, and navigation menus. Itis designed to make it easier for developers to create responsive, mobile-first web pages andapplications.

2.Data Set

Took some data set for analysis to car price prediction and car sells analysis. It took from various online platform where has many data set about car related data set. Our project dataset based on USA market.

Mitsubish	12999	crossover	140	2.4	Gas	yes	2007	Outlander	full
Honda	30000	sedan	52	3.5	Gas	yes	2014	Accord	front
Porsche	26900	crossover	135	4.8	Petrol	yes	2008	Cayenne	full
Toyota	80000	crossover	23	4.5	Diesel	yes	2014	Land Cruis	full
BMW	8500	sedan	260	2.2	Gas	yes	2001	520	rear
Mercedes	7000	van	197	2.2	Diesel	yes	2002	Vito ïàññ.	front
Audi	36000	crossover	145	3	Petrol	yes	2010	Q7	full
Mercedes	9500	van	280	2.2	Diesel	yes	2005	Vito ïàññ.	rear
Toyota	103999	crossover	0	4.5	Diesel	yes	2016	Land Cruis	full
Toyota	1.00E+05	crossover	0	4.5	Diesel	yes	2016	Land Cruis	full

Manufacti	Model	Sales_in_t	__year_re	Vehicle_t	Price_in_t	Engine_si	Horsepow	Wheelbas	Width	Length
Acura	Integra	16.919	16.36	Passenger	21.5	1.8	140	101.2	67.3	172.4
Acura	TL	39.384	19.875	Passenger	28.4	3.2	225	108.1	70.3	192.9
Acura	CL	14.114	18.225	Passenger		3.2	225	106.9	70.6	192
Acura	RL	8.588	29.725	Passenger	42	3.5	210	114.6	71.4	196.6
Audi	A4	20.397	22.255	Passenger	23.99	1.8	150	102.6	68.2	178
Audi	A6	18.78	23.555	Passenger	33.95	2.8	200	108.7	76.1	192
Audi	A8	1.38	39	Passenger	62	4.2	310	113	74	198.2
BMW	323i	19.747		Passenger	26.99	2.5	170	107.3	68.4	176
BMW	328i	9.231	28.675	Passenger	33.4	2.8	193	107.3	68.5	176
BMW	528i	17.527	36.125	Passenger	38.9	2.8	193	111.4	70.9	188
Buick	Century	91.561	12.475	Passenger	21.975	3.1	175	109	72.7	194.6
Buick	Regal	39.35	13.74	Passenger	25.3	3.8	240	109	72.7	196.2
Buick	Park Aven	27.851	20.19	Passenger	31.965	3.8	205	113.8	74.7	206.8

3.Python

In our project we have used many libraries of python programming language for dataanalytics, data manipulation, data visualization.

- **NumPy:** This popular Python library that is used for numerical computing. It provides powerful tools and techniques for performing operations on arrays and matrices, which are often used in data analysis, scientific computing, and machine learning applications.

- **Pandas:** This used for data manipulation, analysis, and visualization. It provides easy-to-use data structures and data analysis tools that are widely used in data science, finance, social sciences, and other fields.
  Pandas provides two primary data structures: Series and Data Frame. A Series is a one-dimensional array-like object that can hold any data type, while a Data Frame is a two-dimensional table-like data structure consisting of rows and columns.

- **Pickle:** This is a Python library used for serializing and deserializing Python objects. It converts a Python object into a binary format that can be stored in a file or transmittedover a network, and then converts the binary data back into a Python object when needed.
  Pickle is commonly used for data persistence, caching, and transfer. For example, it canbe used to save the state of a Python object to a file, so that it can be loaded later without the need to recompute it. It can also be used to transfer data between different Python processes or machines.

- **Flask:** is a lightweight web framework for Python used to build web applications. It is designed to be easy to use and flexible, making it a popular choice for developers who want to build web applications quickly and easily.
  Flask provides a range of features and tools for web development, including routing, templates, sessions, and authentication. It also supports extensions that provide additional functionality, such as database integration and RESTful API development.

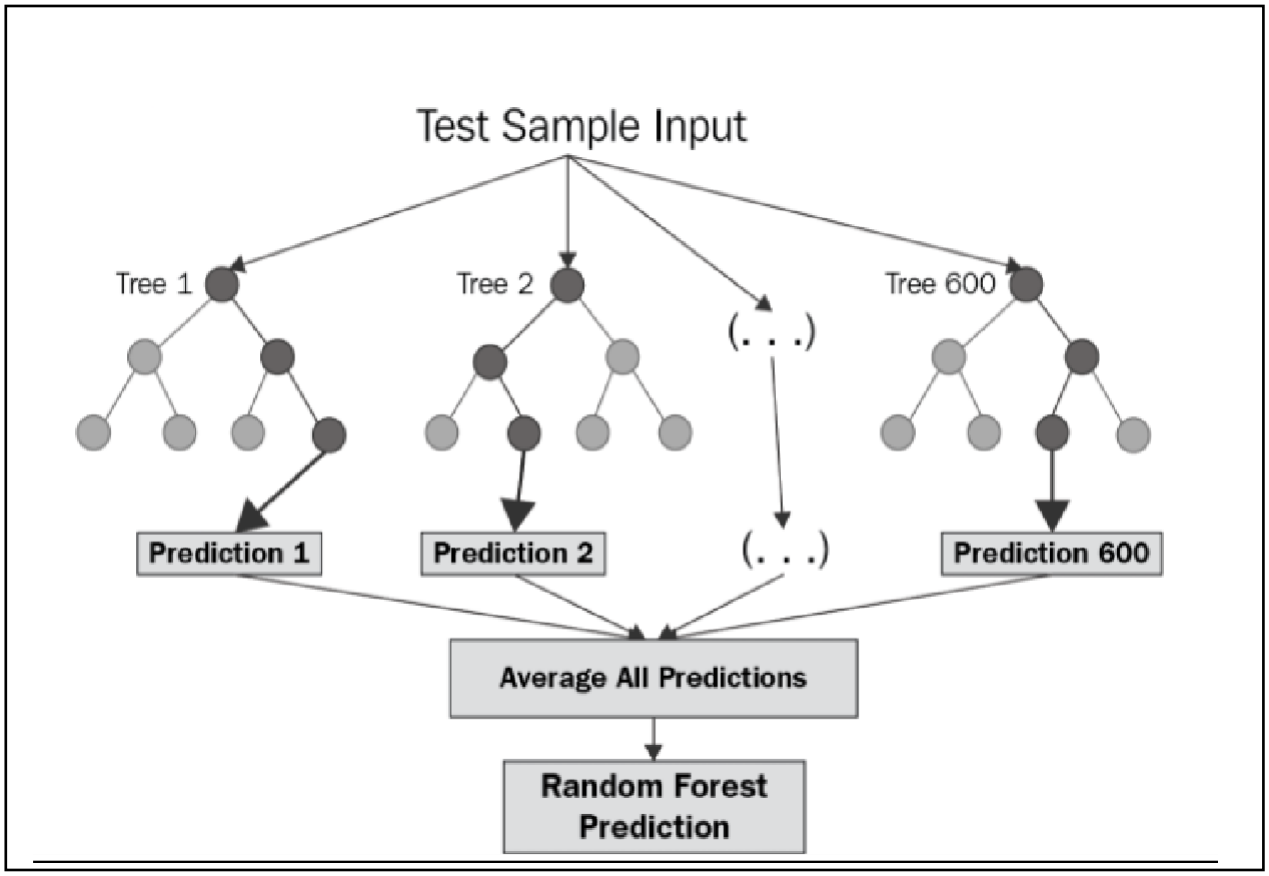
- **Plotly:** is a data visualization library for Python and other programming languages that provides a wide range of interactive visualization tools. It allows users to create interactive charts, plots, and graphs with just a few lines of code, making it a popular choice for data scientists and analysts who want to create engaging visualizations for their data.

4.Regression Algorithm

In our project, we have used Random Forest regression because of more accuracy. Random forest regression is an ensemble method that combines multiple decision tree models to improve the accuracy of the predictions. It is commonly used for high-dimensional data and data with complex relationships.

In Random Forest Regression, a set of decision trees is created using a random sample of the input features and a random subset of the training data. Each decision tree in the forest is trained independently, and the final prediction is made by averaging the predictions of all theindividual trees.

The main advantage of Random Forest Regression over a single decision tree is that it can handle a larger number of input features and avoids overfitting by reducing the variance of the predictions. It also provides an estimate of the feature importance, which can be useful for feature selection and understanding the



ECONOMIC ANALYSIS

1. **Implementation Costs:** Implementing the AUTO-GAADI system can involve significant upfront costs, such as acquiring and preprocessing data, developing and validating machine learning models, and designing and implementing a user interface. These costs need to be balanced against the potential benefits of the system.

2. **Maintenance Costs:** Once the AUTO-GAADI system is implemented, it will require ongoing maintenance and updates to ensure that it remains accurate and relevant. These costs need to be factored into the economic analysis.

3. **Competition:** The AUTO-GAADI system may face competition from other similar systems or traditional methods of car price estimation and sales analysis. This may impact the potential benefits and costs of implementing the system.

4. **Reduced Risk**: By using data-driven insights to guide pricing and sales decisions, the AUTO-GAADI system can help reduce the risk of overpricing or underpricing cars. This can result in more balanced supply and demand, and reduce the potential for price volatility.

5. **Improved Customer Satisfaction:** By providing accurate and transparent car price predictions and sales analysis insights, the AUTO-GAADI system can improve customer satisfaction. This can lead to increased sales and repeat business for car dealerships.

6. **Increased Efficiency:** By using machine learning algorithms to predict car prices and analyze car sales data, the AUTO-GAADI system can potentially increase the efficiency of the car sales process. This can reduce the time and resources required for manual price estimation and data analysis, leading to cost savings for car dealerships and consumers.

Comparison and Result

5.1 Comparison

Before starting this project, we have discussed and analysis many references book and research paper regarding our topic which is AUTO-GAADI based. Authors uses many Algorithms such as, Decision tree, Naïve bayes, support vector machine, Random Forest regression. Among all the algorithms, Random Forest regression more accuracy rather than others, therefore we have implemented here in our projects.

- According to "*The Framework of Car Price Prediction and Damage Detection Technique from Department of Software Engineering, Sir Syed University of Engineering & Technology, Karachi, Pakistan*". This research paper has been implemented many algorithms for better accuracy.

1. **Logistics Regression:** Logistic regression works by fitting a logistic function to the input features. The logistic function maps any real-valued input to a probability value between 0 and 1, representing the likelihood of the input belonging to a particular class.

   The logistic function used in logistic regression is called the sigmoid function, which has an S-shaped curve. The sigmoid function takes any real-valued input and outputsa value between 0 and 1. The output of the sigmoid function is then used to make a binary classification decision, where inputs with a probability above a certain threshold are classified as belonging to one class, and inputs with a probability belowthe threshold are classified as belonging to the other class.
   In this research paper the accuracy shows 0.77%.

2. **KNN Algorithm:** In KNN, the output variable is predicted by finding the k nearest data points in the training dataset and using their output values to predict the output for the new data point.

   The distance between the new data point and the training data points is calculated using a distance metric, such as Euclidean distance or Manhattan distance. The valueof k is a hyperparameter that must be specified before training the algorithm. The value of k determines the number of nearest neighbors that will be considered when making a prediction.

   In the case of classification tasks, the output variable is predicted by taking a majority vote among the k-nearest neighbors. In the case of regression tasks, the output

variable is predicted by taking the mean or median of the output values of the k- nearest neighbors.

KNN is a non-parametric algorithm, which means that it does not make any assumptions about the underlying distribution of the data. It is also considered a lazy learning algorithm, which means that it does not learn a model from the training databut instead memorizes the training data.

In this research paper the KNN Algorithm also 0.77%.

3. **Multiple Regression:** Multiple regression is a statistical technique used to model the relationship between multiple input variables and a single continuous output variable.It is an extension of simple linear regression, where only one input variable is used topredict the output variable.

   In multiple regression, a linear equation is fitted to the input variables, and the coefficients of the equation represent the relationship between the input variables and the output variable. The equation takes the form of y = b0 + b1x1 + b2x2 + ... + bnxn, where y is the output variable, b0 is the intercept, and b1, b2, ..., bn are the coefficients of the input variables x1, x2, ..., xn.

   The coefficients of the equation can be estimated using various techniques, such as the ordinary least squares (OLS) method or maximum likelihood estimation (MLE). Once the coefficients are estimated, the equation can be used to make predictions onnew data.

   In this research paper the Multiple Regression Algorithm accuracy shows also 0.76%.

4. **Random forest:** In Random Forest Regression, a random subset of the input variablesis used to construct each decision tree. This helps to reduce overfitting and improve the accuracy of the model. The output of each decision tree is then combined to makea prediction. The combination can be done by taking the mean or median of the outputs of the decision trees.

   Random Forest Regression is a powerful algorithm that can handle complex datasets and is relatively robust to outliers and noise. It is also able to handle missing data anddoes not require feature scaling.

   The accuracy of Random Forest Regression can be evaluated using various metrics, such as mean squared error (MSE), root mean squared error (RMSE), R-squared (R^2) and adjusted R-squared (R^2 adj).

MSE and RMSE are measures of the average squared difference between the predicted valuesand the actual values. The lower the MSE and RMSE values, the better the accuracy of the model.

R-squared is a measure of the proportion of variance in the output variable that is explained by the input variables. It ranges from 0 to 1, where 0 indicates that the model does not explainany of the variance in the output variable and 1 indicates that the model explains all of the variance. Higher R-squared values indicate better accuracy of the model.

According to "*Predicting the Price of Used Cars using Machine Learning Techniques, Computer Science and Engineering Department, University of Mauritius, Reduit, MAURITIUS".*

This research paper shows random forest regression got almost 0.81% accuracy of regardingour topic, therefore random forest implemented in our project.

5.2 Result

The result of the AUTO-GAADI project should be a fully functional system that accurately predicts car prices and provides insights into car sales data. The system should be easy to use, with a user-friendly interface that allows car dealerships and consumers to access the system's features and insights. The system should also be scalable and adaptable to different markets and geographic regions.

Ideally, the AUTO-GAADI system should provide tangible benefits such as cost savings, increased sales, and improved customer satisfaction. It should also help reduce the risk of overpricing or underpricing cars, leading to more balanced supply and demand in the market. Finally, the system should be able to keep up with changing market trends and customer preferences, ensuring its continuedrelevance and usefulness over time.

In addition to the functional system, the AUTO-GAADI project should also resultin valuable insights and analysis. The system should be able to identify trends and patterns in car sales data, helping dealerships and consumers to make informed decisions about pricing, inventory management, and sales strategies. The insights generated by the system should be based on accurate and reliable data, with a high degree of confidence and validity.

Moreover, the AUTO-GAADI system can potentially have broader economic and societal impacts. By providing more accurate and transparent pricing information, the system can help reduce information asymmetry in the car market. This can lead to a more efficient and competitive market, with greater consumer trust and confidence. Additionally, the system can potentially reducewaste in the car market by helping to match supply and demand more efficiently.

Overall, the result of the AUTO-GAADI project should be a valuable tool for car dealerships and consumers, providing accurate pricing information and insightful analysis. Thesystem should help to improve the efficiency,

5.3 Risk Analysis

The risks associated with the AUTO-GAADI project, we can consider each risk inmore detail:

1. **Data quality and availability:** To mitigate the risk of poor data quality, theproject team can implement data validation processes and quality control measures to ensure the accuracy and completeness of the data used. Thismay involve setting up data collection and cleaning processes, establishing data governance policies, and utilizing data verification techniques such as cross-checking and data profiling.

2. **Model accuracy:** To ensure model accuracy, the project team should perform rigorous testing and validation of the models used for car price prediction and sales analysis. This may involve conducting sensitivityanalysis, comparing model results to actual data, and using statistical techniques to evaluate model performance.

3. **Security and privacy:** To mitigate the risk of cyber-attacks and databreaches, the project team should implement appropriate security measures such as firewalls, encryption, access controls, and monitoring tools. They should also establish privacy policies and procedures to protect user data and comply with data protection regulations.

4. **User adoption:** To ensure user adoption, the project team should involvecar dealerships and consumers in the design and development of the system to ensure that it meets their needs and requirements. They shouldalso provide adequate training and support to users to facilitate the adoption of the system.

5. **Market changes:** To address the risk of market changes, the project team should continuously monitor the car market and update the system accordingly. This may involve integrating new data sources, updating models and algorithms, and implementing new features and functionalities to meet changing market needs.

6. **To manage costs:** the project team should develop a detailed budget and cost management plan that includes a comprehensive analysis of allproject costs, including development, maintenance, and operational costs. They should also consider cost-saving measures such as open- source software, cloud-based services, and outsourcing certain tasks.

Overall, by identifying and mitigating these risks, the AUTO-GAADI project can increase the likelihood of success and maximize the benefits to car dealerships,consumers, and the broader economy.

SOCIAL AND ENVIRONMENTAL IMPACT

The AUTO-GAADI project can have several social and environmental impacts,both positive and negative.

Positive impacts:

5.3.1 **Increased transparency:** The project can help increase transparency in thecar market by providing buyers and sellers with more accurate and reliableinformation about car prices and market trends.

5.3.2 **Improved decision-making:** The project can also help buyers and sellers make more informed decisions about buying and selling cars, potentially leading to better outcomes for both parties.

5.3.3 **Reduced carbon emissions**: By facilitating more efficient and effective carsales, the project could help reduce carbon emissions from thetransportation sector, thereby contributing to efforts to combat climate change.

Negative impacts:

5.3.1 **Job displacement:** The project may result in job displacement in the traditional car sales industry, as dealerships and other intermediaries mayno longer be needed.

5.3.2 **Data privacy concerns:** The collection and use of personal data in the project could raise concerns about data privacy and security, which mustbe addressed to ensure that user data is protected.

5.3.3 **Access barriers:** The project may create barriers to access for those who do not have access to technology or who lack the skills to use the systemeffectively, potentially exacerbating existing social and economic inequalities.

Conclusion

In conclusion, the AUTO-GAADI project is a promising initiative that has the potential to revolutionize the car sales industry by providing buyers and sellers with more accurate and reliable information about car prices and market trends. However, to ensure the success of the project and maximize its social and environmental impact, it is important to conduct a thorough feasibility study, perform rigorous risk analysis, and engage with stakeholders throughout the design and implementation process. By doing so, the project can be designedand implemented in a way that benefits all stakeholders and contributes to broader social andenvironmental goals, such as reducing carbon emissions, promoting access and inclusivity, andprotecting data privacy and security. Ultimately, the success of the project will depend on theability of the project team to navigate these challenges and develop a system that is effective,efficient, and equitable for all.

In addition to the feasibility study and risk analysis, it is important for the project team to consider the technical and logistical challenges associated with developing a AUTO-GAADI system. This may involve developing algorithms and machine learning models that can accurately predict car prices and market trends, as well as building a user-friendly platform that can handle large amounts of data and facilitate smooth transactions between buyers andsellers. The team must also consider factors such as data privacy and security, scalability, and compatibility with existing technologies and systems.

Furthermore, the project team must be aware of the potential ethical and legal implications of the AUTO-GAADI system. This may include issues related to data privacy, algorithmic bias, and the regulation of the car sales industry. To ensure that the system is designed and implemented in a way that is ethical and compliant with relevant laws and regulations, the project team may need to consult with legal experts and engage with industry stakeholders and regulatory bodies.

Overall, the AUTO-GAADI project has the potential to transform the car sales industry and contribute to broader social and environmental goals. However, to realize this potential, the project team must be diligent and thorough in their approach, taking into account the technical, logistical, ethical, and legal challenges associated with developing and implementing such a system. By doing so, they can build a system that is effective, efficient, and equitable for all stakeholders.

References:

- "*Predicting the Price of Used Cars using Machine Learning Techniques*", Sameer Chand Pudaruth, Computer Science and Engineering Department, University of Mauritius, Reduit, MAURITIUS
- "*US Auto Production and Price Prediction in the Context of Multiple Regression Analysis*", Qilin Li, Department of Economics, University of California, Irvine, the U.S.
- "*The Framework of Car Price Prediction and Damage Detection Technique*", Muhammad Kashif Shaikh, Department of Software Engineering, Sir Syed University of Engineering & Technology, Karachi, Pakistan.
- Web Development Tools: W3Schools.
- Stack Overflow
- Python tools: W3Schools.
- Youtube channels: Code with harry