Article

Are Used Cars More Sustainable? Price Prediction Based on Linear Regression

A'aeshah Alhakamy [1,2,*], Areej Alhowaity [3], Anwar Abdullah Alatawi [3] and Hadeel Alsaadi [3]

1   Department of Information Technology, Faculty of Computers and Information Technology, University of Tabuk, Tabuk 47512, Saudi Arabia
2   Researcher at Artificial Intelligence and Sensing Technologies (AIST) Research Center, University of Tabuk, Tabuk 47512, Saudi Arabia
3   Master of Artificial Intelligence, Department of Computer Science, Faculty of Computers and Information Technology, University of Tabuk, Tabuk 47512, Saudi Arabia
*   Correspondence: aalhakami@ut.edu.sa

Abstract: Currently, owning a car is a necessity, as it plays a significant role in human transportation for different purposes such as going to work and to the hospital. However, with the current economic challenges, buying expensive cars can be a burden. The car market has shifted toward more affordable used cars. Due to the increasing number of used cars being sold, the price of used cars has become a major issue that could affect our sustainable way of living. The objective of this research is to understand the impact of the problem and to find empirical solutions by implementing a variety of machine learning techniques and big data tools on the prices of used cars. Thus, we develop a linear regression model that can estimate used car prices based on various features to answer the following research questions: (R.Q.1) How significantly does an independent feature in the dataset affect the dependent variable (car price)? (R.Q.2) Is a linear regression model effective for prediction of used car prices? (R.Q.3) How does prediction of used car prices support sustainability? Finally, we present our results in the form of answers to these questions, including some limitations and future research.

Keywords: big data; machine learning; linear regression; prediction model; used car; car prices; sustainability

1. Introduction

Due to rapid technological development, many previously unpopular tools have become necessary for people. For instance, as people's lifestyles became more demanding, cars became more popular. The importance of cars cannot be underestimated because they play a huge role in our daily lives. They are also used to transport various items, such as food and furniture. It is expected that the need for automobiles and cars will increase as the number of people continues to increase [1]. The automotive industry is a vital part of the global economy. In most developed countries, around 70 percent of the vehicles manufactured are automobiles. This means that the automotive industry is a major contributor to economic development [2]. The locomotive industry is composed of a limited number of retailers and multinational companies. The manufacturers make cars; the retailers handle the trade of used and new vehicles [3]. It has been observed that the used car market has become more prominent in several countries. In fact, the market for used cars has become larger than that for new cars [2].

Most car owners buy and sell vehicles several times throughout their lives, but it is not always easy to determine the quality of a used car [4]. A study conducted on the value of used cars revealed that the average car owner might see its worth in an exaggerated manner due to various factors such as its attachment and transaction cost [5]. Being able to determine the right price for a used car is a good topic for research. In addition to the type of car a person is looking for, studies on the price of used cars also cover various factors

such as its characteristics, make, model, mileage, date of manufacture, number of cylinders, size, power, and distinctive features [6].

The process of producing predictions is usually performed by analyzing the observed data. After gathering the necessary information, it is then possible to come up with a prediction for the next occurrence. Unfortunately, most of the time, the data collected by these methods are too numerous and not easy to interpret. This is why manual analysis is often required [7]. Big data is a type of information that can be collected and used to perform various tasks, such as analyzing events and predicting the future. It can also help people with their problems and improve the efficiency of their businesses. According to studies, big data has been able to help many businesses grow [8].

One of the most common techniques used in forecasting and estimating is linear regression. In 1800, Gauss developed the least squares method, which is designed to fit an equation with its linear parameters. This type of statistical analysis is commonly used in both statistical and research studies. In both fields, linear regression is often used to show the relationship between a dependent factor and a random factor [9]. The main objective of this study is to analyze the changes in the car business from new to used cars and understand how it will impact sustainability. Citizens are not in favor of buying new cars due to the high cost of living, and they prefer to buy used cars due to their quality and lower price. In this study, we develop a model that uses linear regression to estimate the price of used cars based on their various features. This method will help car owners get the best possible price for their vehicle and for used car buyers to get a fair price.

In addition to the price prediction part of the research, we question whether used cars are a way to go green and to be more sustainability aware. Cars have received a lot of backlash due to their negative effects on the environment. However, there are some eco-friendly cars, though they are hard to buy [10]. During the pandemic last year, the media talked about how air pollution decreased after everyone was ordered to stay indoors [11]. Hybrid cars have been the talk of the town when it comes to discussing green automobiles, considering how much better these cars are than their gasoline counterparts. However, what if we told people that there is an alternative to buying a new hybrid? Well, it is actually a used car [12,13].

In general, we can say that there are four reasons why buying a used car is the best choice when it comes to going green: (1) older models are more fuel efficient, (2) people are not reducing unsustainable factory practices by buying a new hybrid car, (3) hybrid batteries are bad for the environment, and (4) such cars maintain the practice of reduce, reuse, and recycle [14]. To balance our thought process, we explore the effect of price prediction and the sustainability aspects of used cars to reach a conscious and aware decision when owning a car. Ultimately, our work addresses the following research questions: **(R.Q.1)** *How significantly does an independent feature in the dataset affect the dependent variable (car price)?* **(R.Q.2)** *Is a linear regression model effective for prediction of used car prices?* **(R.Q.3)** *How does prediction of used car prices support sustainability?*

Through the use of machine learning and big data analytical tools, we can help policymakers make informed decisions regarding the pricing of used cars and its influence on sustainable mobility. A sustainable transportation system should ensure that all its users are able to travel efficiently and effectively. It should also maintain a certain level of public health and economy. In response to increasing concerns about air pollution in urban areas, policymakers have started to introduce more eco-friendly transportation options. Although the various forms of transportation available are considered part of the environmental development of cities, they do not represent the entire solution. The decisions taken by policymakers include the introduction of car-sharing programs, recycling used cars, bus prioritization, and bike-sharing systems.

2. Related Work

This section of the study aims to discuss the various aspects of used cars and explores the research on this subject, focusing on (1) how machine learning can help predict their price, and (2) whether used cars are more sustainable than new cars.

2.1. Machine Learning and Old Cars

Many studies have been conducted on the prediction of car prices, performance, and behavior. Most of these studies have used advanced machine learning models [15–17]. Some of these have also used neural networks and deep learning models [18–20]. Although linear regression is commonly used in studies on the prediction of used car prices, it can also be used to analyze changes in the car business from new to used cars in different regions around the world. For instance, Safitri and Puteri used a linear regression model to predict the prices of used cars in Indonesia [9]. They wanted to identify the various factors that can influence the price of a used car. They performed a linear regression analysis to study the effects of different factors on a car's price. The variables taken into account in the analysis included the car's age, color, transmission type, price, and mileage. Analysis on the result revealed that the model was able to predict the car's age with 63.2 percent accuracy. The price-to-mileage analysis performed by the researchers showed a slightly lower accuracy of 33.3 percent. Further study revealed that the combination of car age and mileage could predict the used car's price with 63.6 percent accuracy.

To predict the prices of used cars, Monburinon et al. [21] used various programming languages to create a model with two phases. The first phase focuses on the linear regression model, and the second one focuses on the lasso regression model. The training phase was conducted after the data collected from the study were validated, and the testing phase was conducted to evaluate the model's performance. After the training phase, the model was evaluated against three regression models to determine which one gave the best accuracy when it came to predicting the prices of cars. Furthermore, Brahimi [22] provided a framework for the development of text mining techniques and their applications in the Arabic online classified market. They tested the proposed methods against two different types of ads: used cars and construction equipment. They performed a series of tests on the proposed models and their prices by implementing various statistical techniques, such as regression and k-nearest neighbor. After testing the prediction model, it was revealed that they were more accurate than the baseline methods.

Costa et al. [2] used the linear regression model to predict the prices of used cars. The 1870 cars included in the study were categorized as non-Maruti or Maruti. The results of the study revealed that the prices of used cars are inversely related to their mileage. This means that when the car's mileage is higher, its selling price goes down. It was also observed that the number of owners of the car is related to its mileage. The R-squared value of the linear regression model used in the study was almost a 100 percent. This shows that the model is compatible with the data collected by the study. Kiran et al. [23] used a linear regression model to analyze the used-car resale market. It was able to create a regression curve that represented the correlation between the various attributes of a vehicle and its price. After implementing the model, the authors found that the price of a car is influenced by the engine's power and the number of cylinders. The model was able to predict the car's price with a 90 percent accuracy with a 10.7 percent error rate.

Various methods can predict the price of a car based on its market value. Asghar et al. [24] proposed a method to help both the seller and buyer make an informed decision about buying or selling a vehicle. The model used a machine learning algorithm, known as linear regression, that outperformed other methods. The authors used a statistical test to get the optimal features of the model and the design value of P. After finding the RFE, they applied the statistical test to the VIF. The main objective of Arjun and Kamalraj [25] was to analyze the performance of the machine learning techniques used to predict the cost of used cars. The forecasts were based on the data collected from various sources, such as daily newspapers. The methods used for forecasting the prices of cars included linear

regression, k-closest neighbors, and logistic regression. The algorithms used for analyzing the results were also used to check their accuracy.

Reliable and accurate prediction is important in the field of used cars due to the various factors that affect their price. Samruddhi and Kumar [26] proposed a machine learning model that can analyze the used car market using KNN. The model was tested with different test ratios and trained on data collected from the Kaggle website. After analyzing the data, the proposed model was able to achieve an accuracy of 85 percent.

2.2. Sustainability and Used Cars

A green vehicle uses less harmful energy and reduces its impact on the environment. An eco-friendly vehicle is made from sustainable materials. A car should meet certain requirements to be sustainable. These involve its fuel consumption, carbon emissions, engine function, frequency of servicing, and seat materials. Other factors such as the vehicle's regenerative energy systems and the recyclability of worn components are also taken into account to determine its sustainability.

Cars in their current state cannot be sustainable due to how much effort goes into making them and how much energy they need to function. An evaluation of the benefit of a car should consider its various aspects, such as its social, economic, and ecological impact. If the focus is only on the technology, other factors such as the life cycle of the vehicle should also be considered. For instance, the effort required to produce a car should be taken into account during its in-use and end-of-life phases. Comprehensive analysis of a car is required to determine its sustainability. This process involves analyzing the various components of a vehicle's life cycle.

It is generally a better idea to keep an old car running as long as possible to ensure that it gets the best possible mileage. Doing so saves money and helps the environment by reducing the greenhouse gas emissions of new cars. In 2004, Toyota revealed that about 28 percent of a car's carbon dioxide emissions come from its manufacture and transportation [27]. The remaining emissions are released once the car is taken out of the factory and is driven. A study conducted by a Japanese university estimated that 12 percent of the car's pre-purchase emissions come from its driving [28,29]. As mentioned above, there are four main reasons why buying a used car can be sustainable: (1) older models are more fuel efficient, (2) people are not reducing unsustainable factory practice by buying a new hybrid car, (3) hybrid batteries are bad for the environment, and (4) used cars maintain the practice of reduce, reuse, and recycle.

2.2.1. Used Car Models Are More Fuel Efficient

Despite the supposedly better energy ratings of new cars, new models are not always the best choice when it comes to reducing greenhouse gas emissions. There are still plenty of cars from the 1990s that are more fuel efficient than modern cars. The waste materials are transformed into useful resources, such as raw materials and fuel. Fuel efficiency is important. The rising prices of gasoline will soon prevent the remaining raw materials from being used. Fossil fuels are not renewable resources [30]. Whether a person is planning on buying a newer used car or an older one, the rate at which that person will use these energy-producing substances depends on the vehicle's fuel efficiency. If a person is on a budget, buying used is an affordable way to reduce the impact on the environment. The options also expand when buying used, as newer cars typically do not fit into an individual's budget. For Instance, in Lebanon after a positive shock in gasoline prices, the sales of used cars increased by 2.69 percent in three months. Marrouch and Mourad [31] also found a unidirectional causality between the prices of gasoline and used car sales. Meanwhile, the results of the new car market show that an increase in gasoline prices does not affect the sales of new fuel-efficient cars. Instead, it shows that the most fuel-efficient cars are not preferred when prices are high.

Yang and Tang [32] assumed that the welfare and environmental programs in China are not affected by outside goods. However, these can also contribute to the accumulation of

greenhouse gases. For instance, if welfare programs are not in place, consumers might buy used cars instead of new ones. Although some consumers use public transportation to get around, they may also use other means of transportation that consume fuel. This suggests that we may not take into account the program's positive effects on the environment and social welfare. The demand for used vehicles will continue to decline even if the technology that allows drivers to sell their old gasoline cars is attractive because the prices of used vehicles will eventually rise until the new ones are cheaper. It is therefore important that the efficiency of gasoline vehicles is improved alongside the development of new business models and vehicle technologies [33,34].

2.2.2. Mitigate Unsustainable Factory Practices

Factories are unsustainable. Mass production is harmful to the environment because it requires the exploitation of resources to sustain itself. The latest hybrid cars are only as good as they are on the road. Their carbon footprint is still enormous when they are made. Factories can release toxic chemicals into the environment and contaminate groundwater reserves. Instead of supporting mass production, buyers should consider investing in used cars that are fuel efficient. These cars should be reliable vehicles that will not have issues. Before a buyer makes a purchase, he/she should speak with the dealer about the various expenses, such as the insurance and maintenance costs. Although it is an eco-friendly purchase, it should still be considered as a financial investment [35,36].

Due to the increasing number of cars being produced, many car factories have had to change their operations to improve air quality. This includes Volkswagen, which made numerous attempts to limit the harmful effects of car manufacturing by implementing programs that show the construction plan of producing new vehicles before they are made [37,38].

2.2.3. Hybrid Batteries and their Environmental Harm

There is a dark side to hybrid cars. Cars need batteries, and this can lead to an environmental crisis. The type of battery that a plug-in hybrid uses is one of the most important factors that affects its performance. Lead-acid batteries are crucial to a hybrid's functionality because they allow it to perform regenerative braking, which helps the vehicle save energy when it is running low on fuel. Unfortunately, the batteries used in electric cars are typically made from lithium and nickel, which have high toxicity levels [39,40] .

In one instance, NASA used a factory to test the batteries of rovers in an area declared a "dead zone". Unfortunately, there are no programs in place that ensure that these batteries are recycled. Not only does this increase the risk of contamination, but it also damages local communities [41,42].

The goal of electric vehicles is to reduce carbon emissions and to improve the environment by reducing transportation-related pollution. However, due to the large-scale production and use of these batteries, electric vehicles can cause environmental pollution and can be harmful to human health [43,44].

2.2.4. Reduce, Reuse, and Recycle

When it is time to get rid of your car, it is important that you bring it to the right people. Doing so will ensure that it does not contaminate the environment. There are many automotive recycling associations, such as the Canadian Automotive Recyclers Environmental Code, that are working to raise awareness about the proper procedures involved in car recycling The goal of these organization is to develop an industry-wide code that will help standardize the procedures involved in car recycling. This code will also protect the environment by preventing harmful materials from entering our water, air, and soil.

One of the most important reasons to recycle your car is that it can help reduce the amount of materials produced. This process can prevent harmful substances from being created and air pollutants from being released. Compared to the process of manufacturing

new steel, recycling uses less energy. It also helps reduce the need for mining for materials, which is a major contributor to air pollution. By recycling your car, you are contributing to the reduction of greenhouse gas emissions and improving the air you breathe. You are also helping keep animal habitats clean [45,46].

Apart from reducing the gases produced by the manufacturing of new materials, car recycling can help minimize the damage that vehicles do to the local environment. Having a professional car recycling company handle all of your vehicle's fluids and components can help ensure that they are properly disposed of. If you abandon your car, its leaking engine could cause significant damage to the surrounding environment. By recycling your vehicle, you can help reduce landfill use, which is very important, as car pollutants can contaminate soils. Professional disposal of vehicles helps keep forests and waterways clean [47,48].

You can give local businesses a boost by recycling your car because many companies cannot afford to buy new steel, and recycled materials allow them to continue making parts without the high price tag. In addition, by being able to produce cheaper components, many companies can lower their prices for their customers [49]. Not only is it better for the planet, but car recycling and reuse can also be advantageous for your wallet. Some services will gladly pay you for your old vehicle, and this option is great for those who have old cars taking up too much space in their yards [50,51].

An important aspect of a sustainable lifestyle is adopting the philosophy of recycling, prolonging usage, and reusing products. Buying a new car does not fit into this philosophy. In addition to being more energy efficient, buying used cars online also saves money [52,53]. Being conscious of the choices that we make can help us live a more eco-friendly lifestyle. For instance, buying a used car is one of the most important steps that we can take to reduce the impact on the environment [54–56].

3. Methodology

The goal of the current study was to analyze the used car resale market by implementing a linear regression model. This method was able to predict the prices of cars that will be sold by their owners and to analyze the previous patterns of the market by taking into account the dictated variables. The model was able to predict the price of a car based on its various factors such as its years of service, mileage, transmission type, and fuel type. It also took into account the number of previous owners and their experience.

To analyze the effects of an independent feature on the dependent variable, our assumptions indicate that the output of the linear regression model will be sufficient to answer the representative research questions below, in addition to the following assumptions: Assumption 1: there is a relationship between the independent variables and the dependent variable (car price). Assumption 2: the regression line has the same variance from all independent variable.

Then we asked the question, how does the dataset of independent features affect car prices? Big data tools are commonly used for processing data and for visualization. In this study, we used various tools to analyze the data and to develop a model that was able to predict the car prices. We used the PySpark framework to cluster the various data elements in the selected dataset. We also used the Spark MLlib to perform pipelines and to perform machine learning models. The data were collected using DataFrame from Spark SQL, which contains various types of data features and labels. In addition to an estimator, we needed a transformer and pipelines in order to use Spark for price prediction. All of the features have to be combined in one column.

After validating the model, we linked the price of old car prediction output to support the sustainability standard. We considered several aspects to support our argument to answer the following questions, where each of them represents a scene in our scenario for a complete picture.

3.1. Research Questions

We analyzed the dataset to answer the following research questions:

(R.Q.1) *How significantly does an independent feature in the dataset affect the dependent variable (car price)?*

(R.Q.2) *Is a linear regression model effective for the prediction of used car prices?*

(R.Q.3) *How does the prediction of used car prices support sustainability?*

3.2. Process Steps

For clarity, the steps taken to answer the previous questions are presented in this section by using the car market as an empirical setting to apply big data tools and a machine learning model.

1. **Apache Spark:** As a first step, we used Spark, which is an open-source data processing framework that runs on a variety of libraries. With the PySpark API, we can easily create pipelines and analyses that are capable of handling large amounts of data, such as ours. Pandas are typically run on a single machine, whereas PySpark is a multi-platform framework that can handle large datasets. Thus, PySpark is a better choice for machine learning projects because it can process operations several times faster than Pandas.

2. **Set and Load:** After building and setting up Pyspark packages, the dataset was unzipped and inflated in comma-separated values format. Then Spark Context and Session were created to load the data as Spark SQL. Spark context is an entry point to our Spark application. The Spark SQL module is a framework that can be used to process structured data. The DataFrames abstraction in Spark SQL can act as a distributed query engine. It can be used to read data from a Hive installation. The ability to process big data using a distributed data warehouse system known as Hive is a key component of the modern data-processing industry.

3. **Data Exploration:** The initial step in the process of data analysis is data exploration, which involves analyzing the data to determine its nature and characteristics. Data analysts use statistical techniques and data visualization to describe the data's size, accuracy, and quantity. Various types of data exploration techniques are commonly used by data scientists and include manual analysis and software solutions that allow them to explore and identify the relationships between various data elements. For car prices, this step is described in detail in Section 3.4.

4. **Spark MLlib:** The MLlib library is additionally compatible with Spark's APIs. It can be used to connect to various data sources, such as Hadoop. Its interoperation with the Python framework makes it easy to create and implement workflows designed to work with this type of data. The ability to run fast computations efficiently is one of the main factors that sets Spark apart from other platforms. Its ability to provide high-quality algorithms is also another reason why it is used in our work.

5. **Regression Model:** Regression analysis is a type of statistical procedure that predicts the future outcomes between a target and an independent variable. For instance, it can be used to identify the relationship between the price of a car and its physical features that could reflect on sustainability. A linear regression analysis is performed in a linear fashion and involves the use of a best-fit line between two variables. This model is explained further in Section 3.3.

6. **Prediction:** Prediction analysis is divided into three parts. First, we create a training dataset. Second, we fit a model to the training dataset. Third, we connect predictions with inputs to the model. To fit a model to a training dataset, we provide it with all of the necessary data. This allows the learning algorithm to find the mapping between the outputs and the inputs. A machine learning model can then connect the predictions to the model. The input should be described as an array of numbers, such as one row with two columns, and we can define this as a list of rows with each column having a given number. The model can be used in an application to directly relate the outputs and inputs of the prediction to the given data. This allows us to

perform more efficient analysis. In our work, the car price (dependent variable) is affected by all the independent features of a car (make, number of doors, number of engine cylinders, manufacturer, and transmission type). This is explored in Section 4.

7. **Evaluation:** The three main metrics that are used to evaluate a linear model are the mean absolute error (MAE), the mean squared error (MSE), and the root mean squared error (RMSE). The easiest to understand is the MAE, which represents the average error. Although the MAE is commonly used to interpret linear models, it is harder to interpret than the RMSE, which is a more popular metric. Square root constructions make it more interpretable. It is recommended to use the RMSE as the main metric to analyze a model. We evaluate our model with all three metrics but rely more on result of RMSE in our interpretation.

3.3. Linear Regression Model

One of the most common methods used in analyzing a relationship between two or more variables is regression analysis. It is the most common algorithm used in regression because of its attempt to fit a straight hyperplane to the dataset. This process can be performed on a variety of factors and can reveal the effects of a variable on a dependent. Regression can also be used to predict an unknown variable or situation.

This type of model is commonly used to analyze the effects of different factors on a given variable. In the well-known equation $X = mX + c$, X is the independent variable, Y is the dependent variable, and m is the slope/inclination of the straight line.

The linear regression equation is similar:

$$Y = a + m \times X + \varepsilon \tag{1}$$

where a is the intercept, m is the slope of the line or coefficient, and ϵ is the error term or residual. To perform a successful linear regression analysis, the model needs to get the best-fit line because the independent variable influences the dependent variable. For instance, if the width and height of people are related, we can predict the weight of a person based on his/her height. Linear regression is typically performed on a single input variable. In other cases, multiple input variables are used.

The dependent variable is y when implementing linear regression on the set of independent variables $x = (x_1, \ldots, x_r)$, where r is number of predictor. We assume a linear relationship between y and $x : y = \beta_0 + \beta_1 x_1 + \ldots + \beta_r x_r + \varepsilon$, which is the regression equation; $\beta_0, \beta_1, \ldots, \beta_r$ are the regression coefficients, and ε is the random error.

A linear regression function takes into account the predicted weights $b_0, b_1, \ldots b_r$ and the estimated coefficients. It then computes the expected regression function $f(x) = b_0 + b_1 x_1, \ldots + b_r x_r$ with the given inputs. This function should capture the various dependencies between the outputs and inputs.

The coefficient of determination, also known as R^2, is a measure of the variation in y that can be explained by the relationship between x and its dependence on other variables. A larger R^2 shows that the model can explain the variation with better accuracy.

The values of the actual and predicted responses coincide completely when $R^2 = 1$, which is the perfect fit.

$$R^2 = \sum (Y_p - \bar{Y})^2 / \sum (Y - \bar{Y})^2 \tag{2}$$

The standard error estimate is a representation of the distance between the predicted values of a linear regression model and the actual values of the data. It can be defined as the difference between the predicted values and the actual values.

$$\varepsilon = \bar{Y} - m\bar{X} \tag{3}$$

For model validation, the mean absolute error (MAE), the mean squared error (MSE), and the root mean squared error (RMSE) were used to evaluate the performance of a linear regression model with our dataset containing various data points using the following formulas:

The mean absolute error, or MAE, is a measure of the errors that occur when a model is evaluated. It can be calculated by comparing the actual values with the sample size.

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \tag{4}$$

The mean square error, or MSE, is a statistical measure that shows the difference between the predicted and observed values of a model. When a model has no errors, the MSE is zero. On the other hand, if a model has an increase in error, its value goes up.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y})^2 \tag{5}$$

The RMSE is a measure that shows the extent to which the residuals are spread out across the data points in the regression line. It is also used to measure the concentration of the data in the optimal fit.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y})^2}{n}} \tag{6}$$

3.4. Dataset

The dataset used for this analysis consists of a total of fifteen features that can be used to estimate the price of a car. These data were collected from the Kaggle platform for car features and the Cooper Union car dataset.

There are 11,914 rows in the dataset, which is significantly greater than the sixteen rows that were used in the previous analysis. The distribution of the rows is also different, with the numerical column being distributed over seven and the categorical column being nine; see Figure 1.

Make	Model	Year	Engine Fuel Type	Engine HP	Engine Cylinders	Transmission Type	Driven_Wheels	Number of Doors	Market Category	Vehicle Size	Vehicle Style	highway MPG	city mpg	Popularity	MSRP
BMW	1 Series M	2011	premium unleaded (required)	335	6	MANUAL	rear wheel drive	2	Factory Tuner,Luxury,High-Performance	Compact	Coupe	26	19	3916	46135
BMW	1 Series	2011	premium unleaded (required)	300	6	MANUAL	rear wheel drive	2	Luxury,Performance	Compact	Convertible	28	19	3916	40650
BMW	1 Series	2011	premium unleaded (required)	300	6	MANUAL	rear wheel drive	2	Luxury,High-Performance	Compact	Coupe	28	20	3916	36350
BMW	1 Series	2011	premium unleaded (required)	230	6	MANUAL	rear wheel drive	2	Luxury,Performance	Compact	Coupe	28	18	3916	29450
BMW	1 Series	2011	premium unleaded (required)	230	6	MANUAL	rear wheel drive	2	Luxury	Compact	Convertible	28	18	3916	34500
BMW	1 Series	2012	premium unleaded (required)	230	6	MANUAL	rear wheel drive	2	Luxury,Performance	Compact	Coupe	28	18	3916	31200
BMW	1 Series	2012	premium unleaded (required)	300	6	MANUAL	rear wheel drive	2	Luxury,Performance	Compact	Convertible	26	17	3916	44100
BMW	1 Series	2012	premium unleaded (required)	300	6	MANUAL	rear wheel drive	2	Luxury,High-Performance	Compact	Coupe	28	20	3916	39300
BMW	1 Series	2012	premium unleaded (required)	230	6	MANUAL	rear wheel drive	2	Luxury	Compact	Convertible	28	18	3916	36900
BMW	1 Series	2013	premium unleaded (required)	230	6	MANUAL	rear wheel drive	2	Luxury	Compact	Convertible	27	18	3916	37200
BMW	1 Series	2013	premium unleaded (required)	300	6	MANUAL	rear wheel drive	2	Luxury,High-Performance	Compact	Coupe	28	20	3916	39600
BMW	1 Series	2013	premium unleaded (required)	230	6	MANUAL	rear wheel drive	2	Luxury,Performance	Compact	Coupe	28	19	3916	31500
BMW	1 Series	2013	premium unleaded (required)	300	6	MANUAL	rear wheel drive	2	Luxury,Performance	Compact	Convertible	28	19	3916	44400
BMW	1 Series	2013	premium unleaded (required)	230	6	MANUAL	rear wheel drive	2	Luxury	Compact	Convertible	28	19	3916	37200
BMW	1 Series	2013	premium unleaded (required)	230	6	MANUAL	rear wheel drive	2	Luxury,Performance	Compact	Coupe	28	19	3916	31500
BMW	1 Series	2013	premium unleaded (required)	320	6	MANUAL	rear wheel drive	2	Luxury,High-Performance	Compact	Convertible	25	18	3916	48250
BMW	1 Series	2013	premium unleaded (required)	320	6	MANUAL	rear wheel drive	2	Luxury,High-Performance	Compact	Coupe	28	20	3916	43550
Audi	100	1992	regular unleaded	172	6	MANUAL	front wheel drive	4	Luxury	Midsize	Sedan	24	17	3105	2000
Audi	100	1992	regular unleaded	172	6	MANUAL	front wheel drive	4	Luxury	Midsize	Sedan	24	17	3105	2000
Audi	100	1992	regular unleaded	172	6	AUTOMATIC	all wheel drive	4	Luxury	Midsize	Wagon	20	16	3105	2000
Audi	100	1992	regular unleaded	172	6	MANUAL	front wheel drive	4	Luxury	Midsize	Sedan	24	17	3105	2000
Audi	100	1992	regular unleaded	172	6	MANUAL	all wheel drive	4	Luxury	Midsize	Sedan	21	16	3105	2000
Audi	100	1993	regular unleaded	172	6	MANUAL	front wheel drive	4	Luxury	Midsize	Sedan	24	17	3105	2000
Audi	100	1993	regular unleaded	172	6	AUTOMATIC	all wheel drive	4	Luxury	Midsize	Wagon	20	16	3105	2000

Figure 1. Detailed data about each used car in terms of 16 columns/features, among which are the make, model, year, engine fuel type, number of engine cylinders, transmission type, etc.

3.5. Data Preparation

Preparation of the data is usually carried out before it can be used. One of the most important steps that we must perform is data cleaning. This process involves removing duplicate and empty datapoints from the Kaggle dataset. In addition, all the null values have to be re-inserted. This process is very important because there are numerous null values in the dataset; see Table 1.

Table 1. The frequency of null values in the dataset varies between different features before removal of null values.

Feature	Null Frequency
Make	0
Model	0
Year	0
Engine Fuel Type	3
Engine HP	69
Engine Cylinders	30
Transmission Type	0
Driven Wheels	0
Number of Doors	6
Market Category	0
Vehicle Size	0
Vehicle Style	0
Highway MPG	0
City MPG	0
Popularity	0
MSRP	0

After the null values were removed, the data shape of the Kaggle dataset became: 11812, 16. Before implementation of the linear regression model, the data were split into two groups: 80 percent for training and 20 percent for testing.

3.6. Data Visualization

The distribution of the data with respect to various features is depicted in Figure 2. These features include the number of doors, number of engine cylinders, highway miles per gallon, and popularity. The distribution of the variables and features in our dataset is shown in Figure 3. Some of these factors influence the pricing of a car. For instance, the year of manufacture can greatly impact a car's price. The number of engine cylinders can also have a huge impact on the pricing of a car. For instance, if there are more engine cylinders, the price of a car goes up, and if there is less highway and city mileage, the car becomes more valuable.
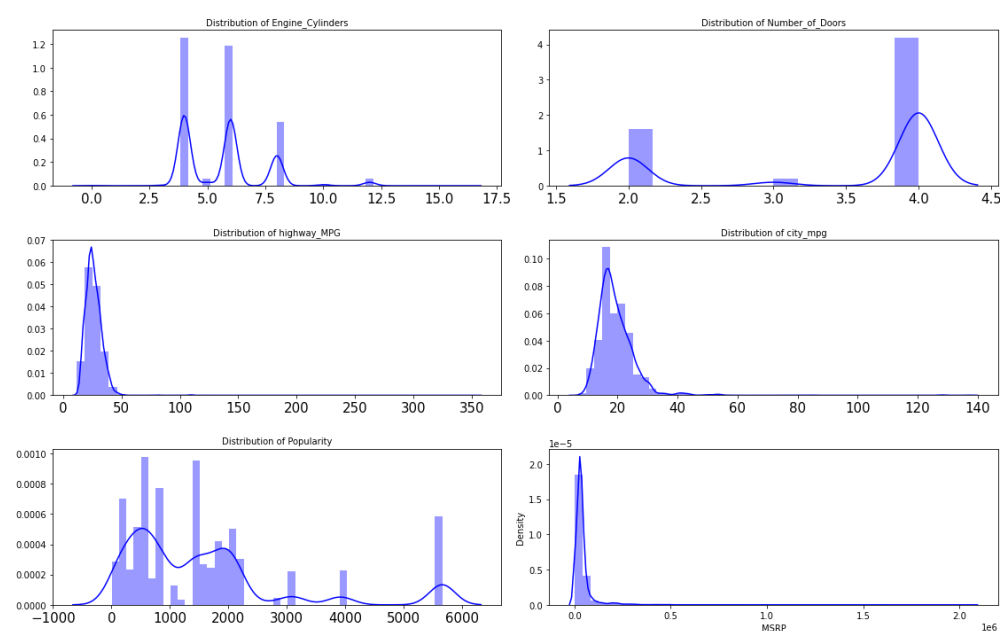
Figure 2. Distribution of various features from the dataset, such as number of engine cylinders, MSRP, and number of doors, showing the heterogeneity or homogeneity of acquired data.
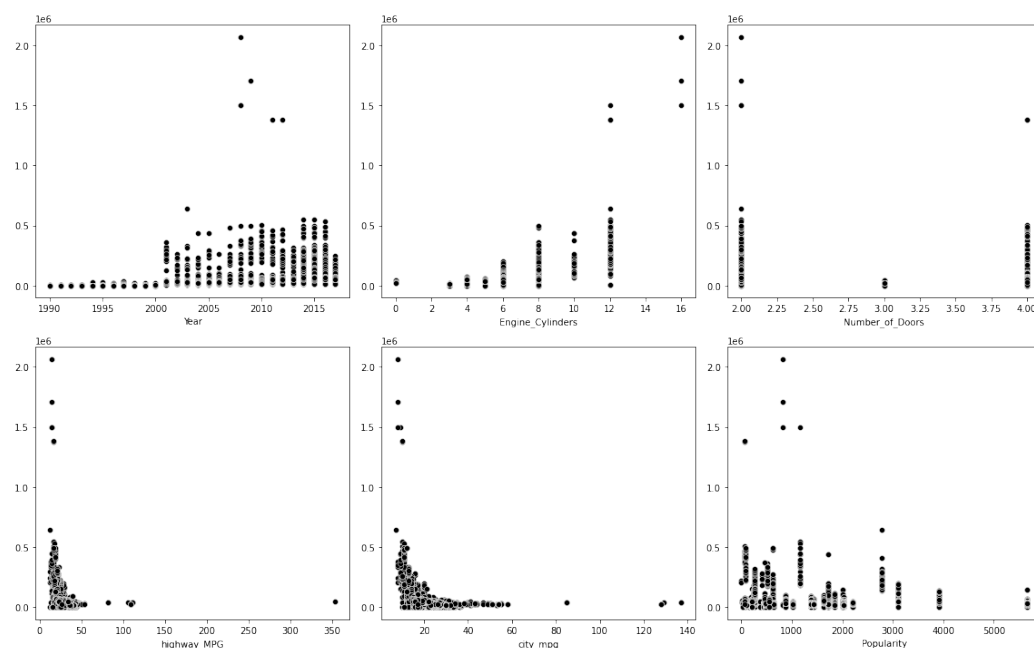
Figure 3. Dot plots of year, number of engine cylinders, number of doors, highway MPG, city, and popularity, showing the distribution of these factors in our given dataset.

The various factors that can affect the pricing of a used car are illustrated in Figure 4. The manufacturer of a car is also a significant factor that can affect the pricing of a used vehicle. For instance, if a certain model of car is more expensive than other models, then the manufacturer of that car can influence the prices of the used car market. A car's type of transmission can also have a huge impact on the price. For instance, if a car has semi-automatic transmission, then it costs more than vehicles that have manual or automatic transmissions; see Figure 5a. The price of a used car is affected by the drive wheels. Figure 5b shows that all-wheel-drive cars are more expensive than those with front or rear drive.
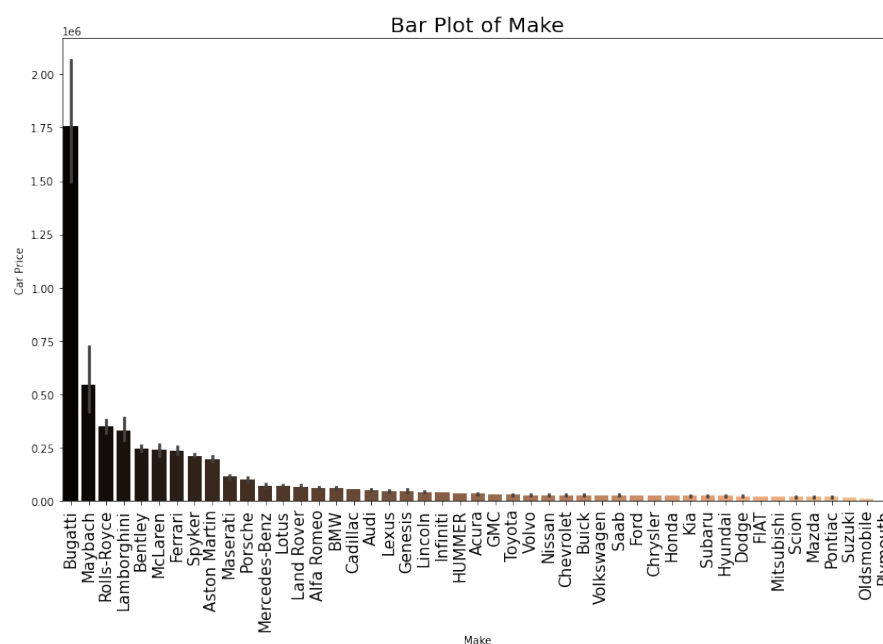
Figure 4. Bar graph representation of how the car make affects the price of the used car, i.e., the more popular the make, the more the car costs.

Our data have been readied for implementing the model, and our assumptions are held accountable: Assumption 1: for it to be applicable, we used corr() to examine the correlation between independent and dependent (price) variables; most of them are correlated and have a linear relationship, so this assumption is satisfied. Assumption 2: the variance around the regression line is the same for all independent variables, so this assumption is satisfied.
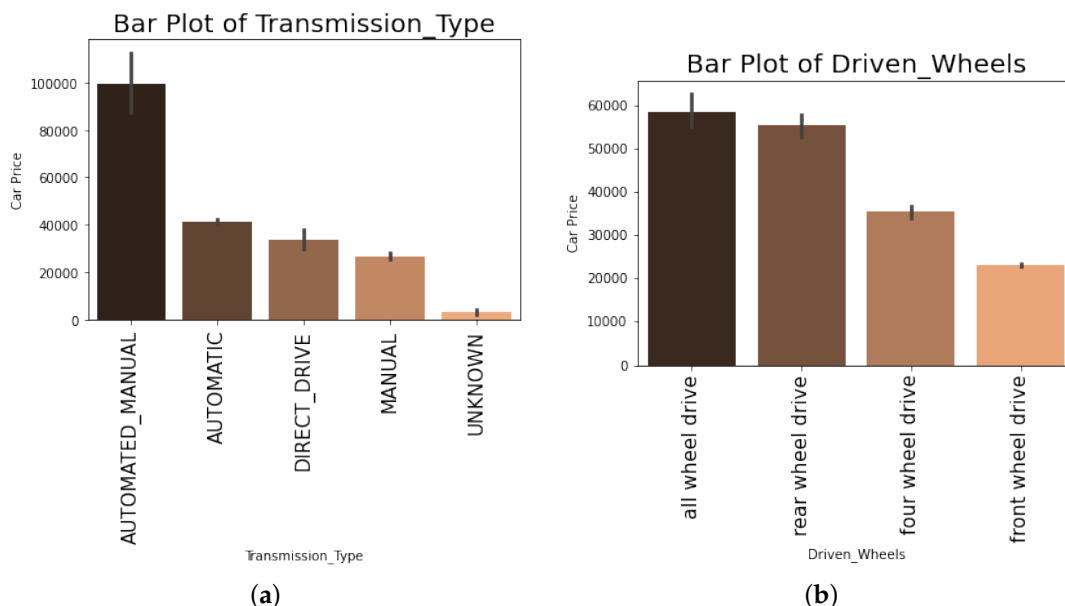

Figure 5. Bar plot representation that shows the effects of some features on the prices of used cars. (**a**) Car price variation based on transmission type, increasing when the type is automated-manual. (**b**) The effect of driven wheels on the car price, where the price increases when the car is all-wheel-drive.

4. Results

This section reports the results for each of the research questions.

4.1. (R.Q.1) How Significantly Does an Independent Feature in the Dataset Affect the Dependent Variable (Car Price)?

A car can be categorized based on various features such as make, model, number of engine cylinders, and the number of doors. According to data analysis, the make and model of the car can have a huge impact on the price of the vehicle. For instance, some cars have low prices because they were made by companies that are not well known. Popular and famous car brands have a huge advantage over these low-priced vehicles.

The prices of cars vary significantly depending on their type of transmission. However, the type of wheels does not seem to have a huge effect on the overall price of the vehicle. This suggests that the make and model of the car are some of the most important factors that can influence the price of a vehicle.

4.2. (R.Q.2) Is a Linear Regression Model Effective for Prediction of Used Car Prices?

The performance of a linear regression model is often evaluated to determine whether it is appropriate for predicting the price of used cars. There are various ways to achieve this, and the root mean square error (RMSE) is one of the most common metrics used in the evaluation.

The concept of the RMSE is that it shows the goodness of a model when compared to the actual results. For instance, in a case where the model predicts the price of a car, the value of the RMSE is used to compare the model's predicted results with the actual price of the car.

The residual value is referred to as "residual", and the root mean square is the collection of these values. The RMSE is then applied to the final assessment value by taking into account the square root operation. In this study, the model's RMSE was 1.43, which is very good accuracy. We used the PySpark big data tool for the evaluation.

We also evaluate the model using mean absolute error (MAE) and mean squared error (MSE) metrics; the results of each are represented in Table 2.

Table 2. The accuracy results of three primary metrics used to evaluate linear models: mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE).

Evaluation Matrix	Accuracy
MAE	1.86
MSE	2.04
RMSE	1.43

The results of the study revealed that the model's predicted price was close to the actual price of the car. The difference between the model's predicted and actual price was 1.43. Another metric used in the evaluation of a model's performance is the efficiency of determination, which is similar to the RMSE but shows the model's effectiveness in percentage terms. In the studies mentioned above, the R^2 value was calculated, but we decided to use the RMSE because it provides an overview of the model's overall performance.

5. Discussion

Through the use of linear regression, we have been able to predict the prices of cars and their good deals. This is considered one of the most important research topics related to the rise of artificial intelligence (AI) for sustainability [57]. The immense amount of data collected by the field have made vehicle value prediction challenging. In this section, we attempt to answer the remaining research question in connection with the previous ones, whereas Section 2.2 elaborates discussion on how used cars are related to sustainability.

Thus, the prediction model for used cars is crucial due to the various features that affect the value of a used car, which can be categorized into two parts: environmental factors and depreciation. While environmental factors include the vehicle's displacement, emission standards, and fuel consumption, depreciation is used determine the cost of a physical or tangible asset over its life. The latter focuses on the vehicle's usage period and miles. The longer the vehicle is used, the more likely it is that wear and tear will lower the value of the car. Because used car buyers are focused on fuel economy, small-displacement cars are more likely to retain their value [58,59]. Thus, using artificial intelligence for price predication in the form of linear regression can help us understand the impact of sustainability based on acquiring used cars.

(R.Q.3) How do Prediction of Used Car Prices Support Sustainability?

Due to the high cost of new cars and the inconvenient nature of bicycles, many people have shifted their focus to used vehicles. Pre-owned cars are often more sustainable than new ones. In terms of cost, second-hand cars are more enticing than new ones. Due to the rapid depreciation of a car, it can cost as little as half its original price after only five years of use by the original owner. In addition, buying a used car can be very beneficial as it allows people to get a highly reliable vehicle at a lower cost. One of the most important factors that people consider when it comes to buying a used car is the maintenance of its features [60].

Another issue that people should consider when it comes to buying a used car is the shortage of microchips. This shortage is a global issue that affects various tiny components. When the COVID-19 pandemic hit the world, many factories closed [10]. Due to the lack of orders, automakers were unable to make up for the lost production time. The sudden increase in the demand for consumer electronics caused the backlog to grow. When the

automotive industry started to recover, the demand for new vehicles outpaced the supply. Due to the shortage of microchips, people who are looking to buy a new car may have to consider other options. Most new cars require a variety of microchips to function properly. New cars also have their own set of environmental side effects. The exploitation of natural resources and the production of new vehicles have been a concern for a long time now, especially due to climate change. Even hybrid cars, which are more eco-friendly than regular vehicles, are hazardous during production.

According to estimates, a typical car produces about 4600 kg of CO_2 emissions annually, whereas the production of new electric cars can emit about 14,746 kg of CO_2 [62,63]. Extending the life cycle of older vehicles can help reduce harmful emissions.

6. Conclusions

Big data and machine learning tools were created to help people solve their daily problems. They became useful tools due to their ability to analyze and improve the efficiency of various tasks, such as predictions. Due to their ability to collect and analyze large amounts of data, they became easier to automate.

In this study, we used big data resources to analyze and improve the efficiency of a second-hand-car model. We then used a linear regression model to predict the price of cars. The data collected from Kaggle included various features such as the make and model of the car, its mileage, and its number of doors. We then used visualization tools to analyze the various factors that affect the price of a used car. The study's results revealed that the LR model could predict the car's price with a 1.43 RMSE value.

The dataset was analyzed to answer several questions: How does an independent feature affect the dependent variable (car price) in the dataset? Is a linear regression model good at predicting used car prices? How does it support sustainability? Thus, a linear regression model was developed to estimate used car prices using various features. It can then be used to analyze the effects of different factors on the dependent variable. For instance, the model can determine the degree to which an independent feature affects the car price and how the prediction of used car prices supports sustainability. In the future, we are eager to train and test various machine learning algorithms for better results and further insights.

Abbreviations

The following abbreviations are used in this manuscript:

MAE	Mean absolute error
MSE	Mean squared error
RMSE	Root mean square error
MLib	Apache Spark's scalable machine learning library
SQL	Structured query language
CSV	Comma-separated values
COVID-19	Coronavirus disease
CO_2	Carbon dioxide

References

1. Bilen, M. Predicting Used Car Prices with Heuristic Algorithms and Creating a New Dataset. *J. Multidiscip. Dev.* **2021**, *6*, 29–43.
2. Costa, L.; Souza, A.; Abhijith, K.; Varghese, D.M. Predicting True Value of Used Car using Multiple Linear Regression Model. *Int. J. Recent Technol. Eng.* **2020**, *8*, 42–45.
3. Chandak, A.; Ganorkar, P.; Sharma, S.; Bagmar, A.; Tiwari, S. Car Price Prediction Using Machine Learning. *Int. J. Comput. Sci. Eng.* **2019**, *7*, 444–450. [CrossRef]
4. Karakoç, M.M.; Çelik, G.; Varol, A. Car Price Prediction Using an Artificial Neural Network. *East. Anatol. J. Sci.* **2020**, *6*, 44–48.
5. Torbarina, M.; Jelenc, L.; Franulović, A.M.; Jukić, I. *Endowment Effect in the Used Cars Market*; 2021; SSRN 3957090. Available online: https://ssrn.com/abstract=3957090 (accessed on 1 March 2021)
6. Venkatasubbu, P.; Ganesh, M. Used Cars Price Prediction using Supervised Learning Techniques. *Int. J. Eng. Adv. Technol. (IJEAT)* **2019**, *9*. [CrossRef]
7. Rajesh, M. Price Prediction for Pre-Owned Cars Using Ensemble Machine Learning Techniques. *Recent Trends Intensive Comput.* **2021**, *39*, 178.
8. Chen, Y.; Li, C.; Xu, M. Business Analytics for Used Car Price Prediction with Statistical Models. In Proceedings of the 2021 3rd International Conference on Economic Management and Cultural Industry (ICEMCI 2021), online, 23 October 2021; pp. 542–547. [CrossRef]
9. Puteri, C.K.; Safitri, L.N. Analysis of linear regression on used car sales in Indonesia. *J. Phys. Conf. Ser.* **2020**, *1469*, 012143. [CrossRef]
10. Hosseini, S.E. An outlook on the global development of renewable and sustainable energy at the time of COVID-19. *Energy Res. Soc. Sci.* **2020**, *68*, 101633. [CrossRef]
11. Nelson, B. The positive effects of COVID-19. *BMJ* **2020**, *369*, m1785.
12. Anfinsen, M.; Lagesen, V.A.; Ryghaug, M. Green and gendered? Cultural perspectives on the road towards electric vehicles in Norway. *Transp. Res. Part D Transp. Environ.* **2019**, *71*, 37–46. [CrossRef]
13. Heffner, R.R.; Kurani, K.S.; Turrentine, T.S. Symbolism in California's early market for hybrid electric vehicles. *Transp. Res. Part D Transp. Environ.* **2007**, *12*, 396–413. [CrossRef]
14. Sathiya, V.; Chinnadurai, M.; Ramabalan, S.; Appolloni, A. Mobile robots and evolutionary optimization algorithms for green supply chain management in a used-car resale company. *Environ. Dev. Sustain.* **2021**, *23*, 9110–9138. [CrossRef]
15. Harahap, F.; Harahap, A.Y.N.; Ekadiansyah, E.; Sari, R.N.; Adawiyah, R.; Harahap, C.B. Implementation of Naïve Bayes Classification Method for Predicting Purchase. In Proceedings of the 2018 6th International Conference on Cyber and IT Service Management (CITSM), Parapat, Indonesia, 7–9 August 2018; pp. 1–5. [CrossRef]
16. Asim, M.; Khan, Z. Mobile price class prediction using machine learning techniques. *Int. J. Comput. Appl.* **2018**, *179*, 6–11. [CrossRef]
17. Pal, N.; Arora, P.; Kohli, P.; Sundararaman, D.; Palakurthy, S.S. How Much Is My Car Worth? A Methodology for Predicting Used Cars' Prices Using Random Forest. In *Future of Information and Communication Conference*; Arai, K., Kapoor, S., Bhatia, R., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 413–422.
18. Hao, H.; Zhang, Q.; Wang, Z.; Zhang, J. Forecasting the number of end-of-life vehicles using a hybrid model based on grey model and artificial neural network. *J. Clean. Prod.* **2018**, *202*, 684–696. [CrossRef]
19. Mozaffari, S.; Al-Jarrah, O.Y.; Dianati, M.; Jennings, P.; Mouzakitis, A. Deep Learning-Based Vehicle Behavior Prediction for Autonomous Driving Applications: A Review. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 33–47. [CrossRef]
20. Al-Mubayyed, O.M.; Abu-Nasser, B.S.; Abu-Naser, S.S. Predicting Overall Car Performance Using Artificial Neural Network. 2019. Available online: http://dstore.alazhar.edu.ps/xmlui/handle/123456789/127 (accessed on 1 March 2021).
21. Monburinon, N.; Chertchom, P.; Kaewkiriya, T.; Rungpheung, S.; Buya, S.; Boonpou, P. Prediction of prices for used car by using regression models. In Proceedings of the 2018 5th International Conference on Business and Industrial Research (ICBIR), Bangkok, Thailand, 17–18 May 2018; pp. 115–119. [CrossRef]
22. Brahimi, B. Arabic Text Mining for Used Cars and Equipments Price Prediction. *Comput. y Sist.* **2022**, *26*, 1015–1025. [CrossRef]

23. Kiran, S. Prediction of resale value of the car using linear regression algorithm. *Int. J. Innov. Sci. Res. Technol.* **2020**, *6*, 382–386.

24. Asghar, M.; Mehmood, K.; Yasin, S.; Khan, Z.M. Used Cars Price Prediction using Machine Learning with Optimal Features. *Pak. J. Eng. Technol.* **2021**, *4*, 113–119. [CrossRef]

25. Reddy, A.; Kamalraj, R. Old/Used Cars Price Prediction using Machine Learning Algorithms. *IITM J. Manag. IT* **2021**, *12*, 32–35.

26. Samruddhi, K.; Kumar, R.A. Used Car Price Prediction using K-Nearest Neighbor Based Model. *Int. J. Innov. Res. Appl. Sci. Eng. (IJIRASE)* **2020**, *4*, 629–632.

27. Charadsuksawat, A.; Laoonual, Y.; Chollacoop, N. Comparative Study of Hybrid Electric Vehicle and Conventional Vehicle Under New European Driving Cycle and Bangkok Driving Cycle. In Proceedings of the 2018 IEEE Transportation Electrification Conference and Expo, Asia-Pacific (ITEC Asia-Pacific), Bangkok, Thailand, 6–9 June 2018; pp. 1–6. [CrossRef]

28. Mulley, C.; Ho, C.; Balbontin, C.; Hensher, D.; Stevens, L.; Nelson, J.D.; Wright, S. Mobility as a service in community transport in Australia: Can it provide a sustainable future? *Transp. Res. Part A Policy Pract.* **2020**, *131*, 107–122. [CrossRef]

29. Gaton, B. There's no need to buy new: Buying a second-hand EV. *Renew Technol. Sustain. Future* **2022**, *158*, 35–38.

30. Martins, F.; Felgueiras, C.; Smitkova, M.; Caetano, N. Analysis of Fossil Fuel Energy Consumption and Environmental Impacts in European Countries. *Energies* **2019**, *12*, 964. [CrossRef]

31. Marrouch, W.; Mourad, J. Effect of gasoline prices on car fuel efficiency: Evidence from Lebanon. *Energy Policy* **2019**, *135*, 111001. [CrossRef]

32. Yang, Z.; Tang, M. Welfare analysis of government subsidy programs for fuel-efficient vehicles and new energy vehicles in China. *Environ. Resour. Econ.* **2019**, *74*, 911–937. [CrossRef]

33. Keith, D.R.; Houston, S.; Naumov, S. Vehicle fleet turnover and the future of fuel economy. *Environ. Res. Lett.* **2019**, *14*, 021001. [CrossRef]

34. Sidorenko, G.; Thunberg, J.; Sjöberg, K.; Vinel, A. Vehicle-to-Vehicle Communication for Safe and Fuel-Efficient Platooning. In Proceedings of the 2020 IEEE Intelligent Vehicles Symposium (IV), Las Vegas, NV, USA, 19 October–13 November 2020; pp. 795–802. [CrossRef]

35. Bueno-Suárez, C.; Coq-Huelva, D. Sustaining What Is Unsustainable: A Review of Urban Sprawl and Urban Socio-Environmental Policies in North America and Western Europe. *Sustainability* **2020**, *12*, 4445. [CrossRef]

36. Nieuwenhuis, P. Micro Factory Retailing: An Alternative, More Sustainable Automotive Business Model. *IEEE Eng. Manag. Rev.* **2018**, *46*, 39–46. [CrossRef]

37. Palla, M.; Bethani, H.; Mourtou, L.; Argyri, P. Planet Earth is screaming. Will you listen? *Open Sch. J. Open Sci.* **2020**, *3*. [CrossRef]

38. Ater, I.; Yoseph, N.S. The Impact of Environmental Fraud on the Used Car Market: Evidence from Dieselgate*. *J. Ind. Econ.* **2022**, *70*, 463–491.

39. Petrauskienė, K.; Galinis, A.; Kliaugaitė, D.; Dvarionienė, J. Comparative Environmental Life Cycle and Cost Assessment of Electric, Hybrid, and Conventional Vehicles in Lithuania. *Sustainability* **2021**, *13*, 957. [CrossRef]

40. Iwata, K.; Matsumoto, S. Use of hybrid vehicles in Japan: An analysis of used car market data. *Transp. Res. Part D Transp. Environ.* **2016**, *46*, 200–206. [CrossRef]

41. von Ehrenfried, M.D. Enabling Technology Advances. In *The Artemis Lunar Program: Returning People to the Moon*; Springer International Publishing: Cham, Switzerland, 2020; pp. 147–183. [CrossRef]

42. Kalita, H.; Thangavelautham, J. Exploration of extreme environments with currentand emerging robot systems. *Curr. Robot. Rep.* **2020**, *1*, 97–104. [CrossRef]

43. Shu, X.; Guo, Y.; Yang, W.; Wei, K.; Zhu, G. Life-cycle assessment of the environmental impact of the batteries used in pure electric passenger cars. *Energy Rep.* **2021**, *7*, 2302–2315. [CrossRef]

44. Zhao, J.; Xi, X.; Na, Q.; Wang, S.; Kadry, S.N.; Kumar, P.M. The technological innovation of hybrid and plug-in electric vehicles for environment carbon pollution control. *Environ. Impact Assess. Rev.* **2021**, *86*, 106506. [CrossRef]

45. Amatuni, L.; Ottelin, J.; Steubing, B.; Mogollón, J.M. Does car sharing reduce greenhouse gas emissions? Assessing the modal shift and lifetime shift rebound effects from a life cycle perspective. *J. Clean. Prod.* **2020**, *266*, 121869. [CrossRef]

46. Hertwich, E.G.; Ali, S.; Ciacci, L.; Fishman, T.; Heeren, N.; Masanet, E.; Asghari, F.N.; Olivetti, E.; Pauliuk, S.; Tu, Q.; et al. Material efficiency strategies to reducing greenhouse gas emissions associated with buildings, vehicles, and electronics—A review. *Environ. Res. Lett.* **2019**, *14*, 043004. [CrossRef]

47. Lin, H.T.; Nakajima, K.; Yamasue, E.; Ishihara, K.N. Recycling of End-of-Life Vehicles in Small Islands: The Case of Kinmen, Taiwan. *Sustainability* **2018**, *10*, 4377. [CrossRef]

48. Wang, J.; Wu, Q.; Liu, J.; Yang, H.; Yin, M.; Chen, S.; Guo, P.; Ren, J.; Luo, X.; Linghu, W.; et al. Vehicle emission and atmospheric pollution in China: Problems, progress, and prospects. *PeerJ* **2019**, *7*, e6932. [CrossRef]

49. Zhou, F.; Lim, M.K.; He, Y.; Lin, Y.; Chen, S. End-of-life vehicle (ELV) recycling management: Improving performance using an ISM approach. *J. Clean. Prod.* **2019**, *228*, 231–243. [CrossRef]

50. Beaudet, A.; Larouche, F.; Amouzegar, K.; Bouchard, P.; Zaghib, K. Key Challenges and Opportunities for Recycling Electric Vehicle Battery Materials. *Sustainability* **2020**, *12*, 5837. [CrossRef]

51. Khudyakova, T.; Shmidt, A.; Shmidt, S. Sustainable development of smart cities in the context of the implementation of the tire recycling program. *Entrep. Sustain. Issues* **2020**, *8*, 698. [CrossRef]

52. Glavič, P. Evolution and Current Challenges of Sustainable Consumption and Production. *Sustainability* **2021**, *13*, 9379. [CrossRef]

53. Pandey, R.U.; Surjan, A.; Kapshe, M. Exploring linkages between sustainable consumption and prevailing green practices in reuse and recycling of household waste: Case of Bhopal city in India. *J. Clean. Prod.* **2018**, *173*, 49–59. [CrossRef]

54. Kunamaneni, S.; Jassi, S.; Hoang, D. Promoting reuse behaviour: Challenges and strategies for repeat purchase, low-involvement products. *Sustain. Prod. Consum.* **2019**, *20*, 253–272. [CrossRef]

55. Martins, L.S.; Guimarães, L.F.; Botelho Junior, A.B.; Tenório, J.A.S.; Espinosa, D.C.R. Electric car battery: An overview on global demand, recycling and future approaches towards sustainability. *J. Environ. Manag.* **2021**, *295*, 113091. [CrossRef] [PubMed]

56. Fernando, Y.; Tseng, M.L.; Sroufe, R.; Abideen, A.Z.; Shaharudin, M.S.; Jose, R. Eco-innovation impacts on recycled product performance and competitiveness: Malaysian automotive industry. *Sustain. Prod. Consum.* **2021**, *28*, 1677–1686. [CrossRef]

57. Al-Turjman, F.; Hussain, A.A.; Alturjman, S.; Altrjman, C. Vehicle Price Classification and Prediction Using Machine Learning in the IoT Smart Manufacturing Era. *Sustainability* **2022**, *14*, 9147. [CrossRef]

58. Liu, E.; Li, J.; Zheng, A.; Liu, H.; Jiang, T. Research on the Prediction Model of the Used Car Price in View of the PSO-GRA-BP Neural Network. *Sustainability* **2022**, *14*, 8993. [CrossRef]

59. Murry, C.; Schneider, H.S. The economics of retail markets for new and used cars. In *Handbook on the Economics of Retailing and Distribution*; Edward Elgar Publishing: Cheltenham, UK, 2016; pp. 343–367. [CrossRef]

60. Wang, S.; Wang, J.; Li, J.; Wang, J.; Liang, L. Policy implications for promoting the adoption of electric vehicles: Do consumer's knowledge, perceived risk and financial incentive policy matter? *Transp. Res. Part A Policy Pract.* **2018**, *117*, 58–69. [CrossRef]

61. Zheng, T.; Ardolino, M.; Bacchetti, A.; Perona, M. The applications of Industry 4.0 technologies in manufacturing context: A systematic literature review. *Int. J. Prod. Res.* **2021**, *59*, 1922–1954.

62. Pavlovic, J.; Marotta, A.; Ciuffo, B. CO_2 emissions and energy demands of vehicles tested under the NEDC and the new WLTP type approval test procedures. *Appl. Energy* **2016**, *177*, 661–670. [CrossRef]

63. Carrera-Rodríguez, M.; Villegas-Alcaraz, J.F.; Salazar-Hernández, C.; Mendoza-Miranda, J.M.; Jiménez-Islas, H.; Segovia Hernández, J.G.; de Dios Ortíz-Alvarado, J.; Juarez-Rios, H. Monitoring of oil lubrication limits, fuel consumption, and excess CO_2 production on civilian vehicles in Mexico. *Energy* **2022**, *257*, 124765. [CrossRef]