
# King Fahd University of Petroleum and Minerals

*College of Computer Sciences and Engineering (CCSE)*

ST 399

Virtual Summer Training

# Final Report

August 2020



| | |
|---|---|
| Ye Myint Htun | 201668940 |
| Abdulaziz Alshehri | 201776310 |
| Abdullah Sheikh | 201641740 |

Dr. Yahya Osais

**Abstract**

This report studies the machine learning techniques needed to create a model that would assist individuals in estimating the value of their used vehicles. All the significant contributing features of a second-hand car would be utilized in estimating its value. The model is intended to be applied in the local markets to draw conclusions on how the market relates to its international counterpart.

# Contents

# List of Figures

# 1 Introduction

The number of vehicles registered in Saudi Arabia have been growing in recent years due to a thriving economy. Between 2014 and 2015, the number of registered vehicles grew by 360,000[1]. As the number of cars increase, so too does the scale of the used vehicle market to accommodate those that wish to exchange for a different vehicle without losing the full economic value of the vehicle they currently possess. The issue arises then, on how to best estimate the residual value of the vehicle they are selling. Since several features can affect the overall price, expert knowledge is required to estimate the resale value for the financial interests of the seller.

# 2 Problem Statement

Economic growth has led to a rise in the middle-class consumer group globally. As a result, it has become financially undemanding for many to own a car. With a steady growth in vehicle sales in recent years, the number of consumers

interested in selling their personal vehicle to exchange for a newer model or for monetary incentives has been on the rise. A problem then ensues on how to judge, with a great degree of accuracy, the monetary value of a used vehicle.

In some scenarios, individuals underestimate or overestimate the value of the vehicles they are advertising. Such an oversight would undermine the potential financial benefits of the completed sale. According to a survey conducted in 2017, 90% of customers did not have complete confidence in the information provided by the salesperson on the vehicle they were purchasing. In contrast, German manufacturers lost a potential revenue of $1 billion in the US markets due to a residual lease value miscalculation [2].

A Car Value Estimation (CVE) system would provide a reliable method to estimate the value of an automobile based on its features and condition. The system would use a vast, reliable, and relevant data set to achieve an accurate estimation of a vehicle's value. The developed and trained CVE could be later used on data from local markets to help draw conclusions on the similarities and differences between the local and international markets. These insights would provide valuable information to Saudi customers who would be seeking to make an informed judgement about their purchases.

# 3 Background

Machine learning is a useful method to provide systems with the capability of automated learning and improvement, by feeding on data without being programmed explicitly. Machine learning is typically categorized into supervised and unsupervised learning. The former utilizes labeled data and applies previous learning onto new sets of data to provide an output. The output would be contrasted against the true values as a measure of performance. In comparison, unsupervised machine learning is presented with raw unlabeled data. Where after, the model infers relations amongst data and computes hidden patterns.

Another subdivision of machine learning under supervision are the two types of learning algorithms. They are presented as either regression or classification

algorithms that borrow techniques found in statistics and probability theory to predict results for issues of a practical nature. Regression algorithms are used when involved with numerical values whilst classification algorithms are used when resolving categorical information.

A practical application of machine learning requires a model which uses the algorithms and data sets to make predictions. Different models include Artificial Neural Networks, Decision Trees, Support Vector Machines, Regression Analysis, and Bayesian Networks. Each model has its own advantages and disadvantages with some prioritizing high-dimensional accommodation or feature optimization. Regardless, each type of model requires updated information periodically to ensure optimal accuracy.

# 4 Literature Review

The following section will discuss the different methodologies and machine learning (ML) techniques that could be used to implement a CVE which would eliminate the need of hiring an expert and give individuals a marginally accurate estimate on the vehicle they desire to sell. Using multiple determining features of a car that would be variables in the ML models.

A study was conducted by Pudaruth [2] in Mauritius on using supervised machine learning to help predict the cost of used cars on the island. He gathered a data set from the local newspaper for only a month considering depreciation would affect the price of cars and hence the results. With an organized and labelled data set, he utilized different machine learning models to forecast the values of used vehicles. The algorithms he used included, multiple linear regression analysis, k-nearest neighbors (kNN), decision trees, and Naïve-Bayes [2]. He found that classifier algorithms did not produce a numerical output and had to be placed in value ranges. Whilst regressor algorithms were more suited for numeric prediction applications and provided higher accuracy [2] [3].

In Germany, another research was conducted on residual lease values using artificial neural networks (ANN) [4]. The study was based on a data set of

150,000 recorded transactions [4]. By employing an ANN, the research team was able to supplement large amounts of indeterminate data such as time depreciation and periodic changes in consumer trends [4]. The team fed the unstructured data into ANN allowing the model's hidden layers to process the information into quantified statistical analytics. The results obtained from the ANN averaged a 5.24% to 18.17% in accuracy of predictions [4].

Gegic et al. [3] attempted a different approach. They used a web scraper to extract a usable data set from their local online car retailer. The raw data was passed through a random forest classifier (RF) algorithm which sorted the vehicles into 3 categories based on the aftermarket price [3]. The labelled and organized data was then input into two different models. The first model was the Support Vector Machine (SVM) that allows some relations between variables as the dimensions increase. The second model was the ANN. Gegic et al. [3] created 3 price ranges corresponding to the 3 different price categories from the RF. The SVM and ANN would complete a secondary phase of classification to place the vehicle in its correct price point within its predetermined price ranges. Their results indicated that a phased method of classifications boosted accuracy as high as 92.38% [3].

Çelik and Osmanoğlu [5] conducted a prediction of prices on second–hand cars. Firstly, they acquired a data set from an online Turkish auctioning site called ikinciyeni.com. They obtained the data set by writing a script in Ruby programming language [5]. Their choice of model was the linear regression analysis. As a result of their work, they successfully obtained an 81.15% predictive accuracy within a 10% confidence interval [5].

Gültekin and Organ [6] used the artificial neural network approach to estimate second–hand car prices in Turkey's car market. They used Weka; an open-source machine learning software developed at the University of Waikato. In their research, they applied a multilayer perceptron regressor, a type of artificial neural network algorithm, to estimate second–hand car prices. Their data set was gathered from hurriyetoto.com, a site with many active car advertisements [6].

Lee et al. [7] were concerned about combining public and private data to maximize the utility of the available big data. They used a Random Forest algorithm (RF) to categorize and perform regression analysis on the collected data. In addition, a Deep Neural Network (DNN) was used for the building value analysis. The results indicated the RF scheme slightly outperformed the DNN. Furthermore, both schemes performed poorly (below 50%) when a model was trained using data obtained from one region and tested on another [7]. Consequently, the outcome confirms that machine learning models can be applied to predict the value of buildings with an accuracy exceeding 80% [7].

Lessmann and Voss [8], discussed the random forest regression (RF) approach for used car resale price forecasting and how to increase the accuracy of an estimated price. The study indicates that the diversity of forecasting methods available have varying degrees of accuracy when estimating the residual value [8]. The authors made three important contributions which are as follows. Firstly, they compared different methods of price forecasting for second–hand vehicles. Secondly, the paper discussed how private information, kept by sellers, about the car could affect the price accuracy [8]. Thirdly, the study examines the importance of the amount information required to update and manage different forecasting models.

# 5 Methodology

## 5.1 Data Gathering

All machine learning applications require a data set. Nine different websites that provide information on used cars for sale in Saudi Arabia were explored. The websites were cross-checked to use only the most suitable as sources for the data set. As seen in Figure [1], some of the needed features such as brand, model, year, color, and mileage were checked. Having chosen SaudiSale, KSAmotory, OpenSooq and Abisyara websites made a great combination to building the large data set. As seen from Figure [1], these four websites have many common

features creating a diverse combination for the data set. Some missing features such as car type, condition, and car options from the four websites were ignored. In addition, as SaudiSale did not provide a color feature for their cars, a random selection was given based on the three most common colors available.

| | Websites Check List | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Features | SaudiSale | KSA motory | OpenSooq | hatla2ee | fridaymarket | yalla motors | Haraj | syarah | abisayara | Quality Score |
| Car Make | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100% |
| Model | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100% |
| Year | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ | 89% |
| Car Type | ✗ | ✓ | ✗ | ✓ | ✗ | ✓ | ✗ | ✗ | ✗ | 33% |
| Trans | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ | ✓ | ✓ | 89% |
| Condition | ✗ | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ | ✗ | ✓ | 67% |
| Color | ✗ | ✓ | ✓ | ✓ | ✗ | ✓ | ✗ | ✓ | ✓ | 67% |
| Mileage | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ | ✗ | ✓ | ✓ | 78% |
| City | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 100% |
| Asking price | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ | 89% |
| real sold price | ✓ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | 11% |
| Sold Date | ✓ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | 11% |
| Car Options | ✗ | ✗ | ✓ | ✓ | ✗ | ✓ | ✗ | ✗ | ✗ | 33% |
| Quality Score | 69% | 77% | 77% | 85% | 38% | 85% | 38% | 62% | 69% | 67% |
| Comments | Suitable | Suitable | Suitable | not suitable to extract information | not suitable to extract information | few cars around 500 | not suitable to extract information | few cars around 500 and it's in arabic | Suitable | |

Figure 1: Websites check list.

## 5.2 Web Scraping Technique

Collecting the data set from different websites required knowledge on web scraping techniques. Web Scraping is tool that collects required data from the internet and exports it as an excel sheet. The two available methods of web scraping included either writing code or using available web scraping tools such as ParseHub and Octaparse. We initially used ParseHub, however, the free software limitations made it impossible to gather a decent data set. Octoparse, proved more useful with no limitations attached and only two-week window.

## 5.3 Cleaning The Data Set

The gathered data set needed a thorough cleaning to produce decent results. Consequently, rules were placed to eliminate records that did not uphold the criteria. The rules are as follows:

- The production year should be between 2000 and 2020.
- The price of the car should be above 10,000 and less than 200,000.

- Any records with null.
- Misinformation seen in the records.

In addition, capitalization differences were defined, and spelling errors were corrected for all records.

## 5.4 Final Data Set

Having cleaned the data set repeatedly, a final data set of 8467 cars from 27 different manufacturer with 514 different car models was created. Figure [2] indicates the top three car brands were Toyota, Hyundai, and Ford. The collected cars, distributed across 78 different cities in KSA, added variation to the prices as listed around the kingdom. Moreover, the top three car models were Ford Taurus, Hyundai Accent and Toyota Camry. Furthermore, Jeddah, Riyadh and Dammam contributed the most information on second-hand cars.

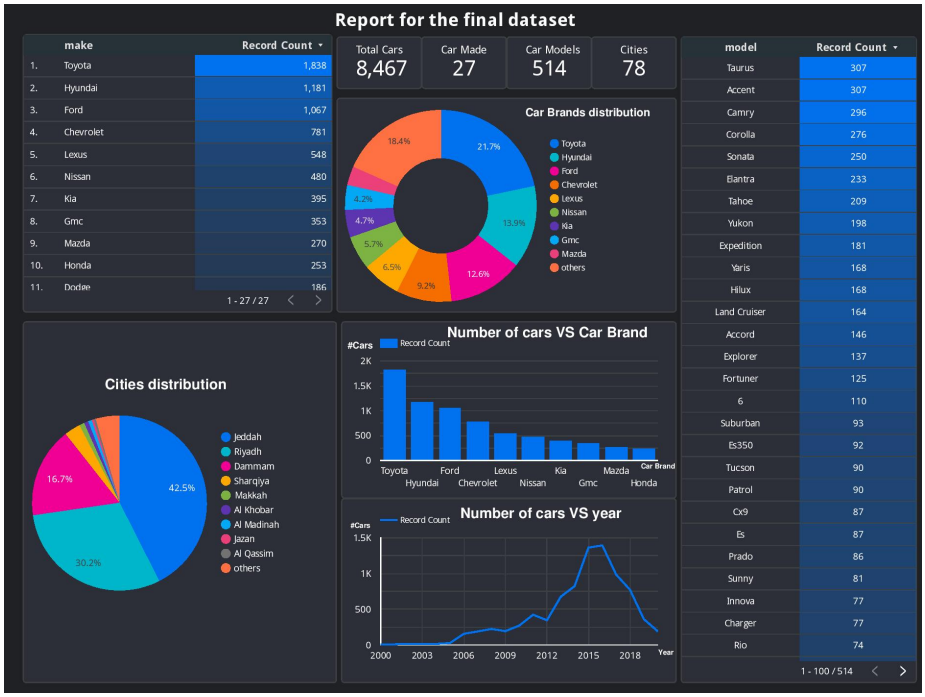


Figure 2: Final data set visualization.

### 5.5 Data Set Encoding

Machine learning algorithms can not deal with categorical values, so encoding is necessary. In our project we used One-Hot encoding at first, then replaced it with James-stein encoding.

- One-Hot Encoding: Replaces each categorical column with a number of numerical columns equal to the number of categories in that columns.

```
training time:  467.14059591293335 seconds
Training Set Mean Absolute Error: 7822.5129
Test Set Mean Absolute Error: 10635.3692
r2 score= 0.8409
MAPE= 19.5432
range accuracy( 6000 ) =47.93
```

Figure 3: Evaluation metrics for One-Hot encoding (showing training time. gradient boosting was used).

- James-Stein Encoding: Replaces each categorical value with the average value of the target (price) over that column.

```
training time:  66.06735301017761 seconds
Training Set Mean Absolute Error: 3349.1209
Test Set Mean Absolute Error: 9836.4174
r2 score= 0.8564
MAPE= 16.4716
range accuracy( 6000 ) =50.30
```

Figure 4: Evaluation metrics for James-Stein encoding (showing training time. gradient boosting was used).

## 5.6 Experiments

1. **Automatic Relevance Determination Regression**: Automatic relevance determination regression (ARD) is a linear regression model that helps reduce many irrelevant input features. Consequently, it infers only the relevant features that contribute towards the model to prevent over–fitting. The model conducts linear regression in an iterative manner, whereby it maximizes the evidence of certain features lacking contribution to the model. As the weight of the feature (sharing an inverse relation to evidence) approaches zero, the model removes the feature. Notable parameters of ARD regression include:

    - n_iter: The maximum number of iterations.
    - tol: Stopping the algorithm if the weight had converged to a specific value.
    - alpha_1: The precision of the distribution of the noise.
    - lambda_1: The precision of the distributions of the weights.
    - threshold_lambda: Threshold to remove weights from model.

    The model was initially used to accommodate the increased number of features observed when One–Hot encoding was employed to convert categorical data in numerical data. It was, however, discontinued when shifting encoding techniques from One–Hot encoding to James–Stein Encoding. The performance proved inferior to other machine learning algorithms to be discussed.

2. **Multi-Layer Perceptron Regression**: Multi-layer perceptron (MLP) regression is a feedforward artificial neural network (ANN) that contains many neurons and hidden layers. Each neuron (node) accepts multiple inputs that are adjusted by a weight. These weighted values are summed and then passed through a nonlinear activation function to map the value to the next neuron(s) in the next hidden layer. ANN uses backpropagation

to aid in adjusting the weights to minimize loss and provide an accurate prediction. Notable parameters of MLP regression include:

- hidden_layer_sizes: The number of hidden layers and neurons per hidden layer.
- activation: Nonlinear function used in hidden layer.
- solver: The method of weight optimization.
- learning_rate: Schedule to update weights in learning.
- tol: Stopping the algorithm when loss or score have not changed in consecutive iterations.
- max_iter: The maximum number of iterations.

ANN was initially tested in view of it being well regarded in literature. Further testing was halted, however, due to difficulties in hyperparameter optimization.

3. **Random Forest Regression**: Random Forest (RF) regression is a type of ensemble decision tree. The model creates many decision trees and trains each tree with a different random sample obtained from the testing data set. These trees may have high variance in relations to each other or whichever subset of data was used. Consequently, the final predictions are drawn from an averaging of these individual decision tree observations. Thereby reducing the variance and providing an optimal prediction. Notable parameters of RF regression include:

   - n_estimators: The number of trees in the forest.
   - max_depth: The maximum depth of each tree.
   - min_samples_leaf: The minimum number of samples required at a leaf node.
   - max_features: The number of features to consider when looking for the best split.

The earliest idea to implement a decision tree came as the RF regressor. Decision tree's robustness and very little need for pre-processing of that data made it very attractive. RF displayed promising results during experimentation and was only exceeded by Gradient Boosting regression by a small margin. The deciding factor was the inadequate results obtained when assigning a confidence interval using the dynamic range calculated using standard deviation.

4. **Gradient Boosting Regression**: Gradient Boosting (GB) regression is very similar to the RF regressor in also making use of decision trees. It creates a decision tree of a given data subset and generates a prediction and loss. The subsequent trees optimize the loss and generate new predictions until all values of the data set has been used. The process of continuous optimization along with the robustness of decision trees allow GB regression to be a powerful machine learning algorithm. Notable parameters of GB regression include:

    - loss: The loss function to be optimized.
    - learning_rate: The reduction in contribution of each tree towards the result.
    - n_estimators: The number of boosting stages to perform.
    - max_depth: The maximum depth of each tree.
    - min_samples_leaf: The minimum number of samples required at a leaf node.
    - max_features: The number of features to consider when looking for the best split.

    After testing many algorithms and different parameters, GB regression proved not only the best results but also a decent computation time. A static range for the confidence interval further provided satisfactory estimates for the predictions tested.

## 5.7 Optimizing Hyperparameters

In this project, two methods were used to optimize hyperparameters of the machine learning algorithms; randomized search, and grid search.

Randomized search randomly selects values for the parameters from a predefined set. This method is advantageous when no preexisting knowledge about the optimal ranges for the parameters' values are available.

```
model = ensemble.GradientBoostingRegressor()

param_grid = {
    'n_estimators': np.arange(500, 3000, 500),
    'max_depth': np.arange(5,10),
    'min_samples_leaf': np.arange(3,20),
    'learning_rate': np.arange(0.01,0.1),
    'max_features': np.arange(0.1,1),
    'loss': ['ls', 'lad', 'huber']
}
rs = RandomizedSearchCV(model, param_grid, n_jobs=-1, n_iter=300)
```

Figure 5: Example of randomized search to optimize hyperparameters.

Grid search is the most time intensive method of the two, as it tests every single combination of parameter values and return the best performers. This method, costly as it may be, is the most robust way to optimize hyperparameters and improve performance.

```
model = ensemble.GradientBoostingRegressor()

param_grid = {
    'n_estimators': [500, 1000, 3000],
    'max_depth': [4, 6],
    'min_samples_leaf': [3, 5, 9, 17],
    'learning_rate': [0.1, 0.05, 0.02, 0.01],
    'max_features': [1.0, 0.3, 0.1],
    'loss': ['ls', 'lad', 'huber']
}
bs = GridSearchCV(model, param_grid, n_jobs=-1)
```

Figure 6: Example of grid search to optimize hyperparameters.

### 5.8 Performance Evaluation Metrics

In order to improve the performance of the model, a method of gauging how erroneous it is was required. Therefore, four performance metrics were implemented for that purpose.

The first metric is Mean Absolute Error (MAE). This metric is a measure of errors between the true and predicted values. It is calculated using the following formula:

$$\mathrm{MAE} = \frac{1}{n}\sum_{j=1}^{n} |y_j - y_j|$$

Figure 7: Formula of MAE.

Where n is the number of cars in the test set, $x_i$ is the true price, x is the predicted price.

The second metric, R-squared, measures how much of the variance for a dependent variable is explained by the independent variable. Its formula is as follows:

$$\text{R-Squared} = 1 - \frac{\text{SS}_{\text{regression}}}{\text{SS}_{\text{total}}}$$

Figure 8: Formula of R-squared.

Where SS is the sum squared.

The third metric is Mean Absolute Percentage Error (MAPE). This metric is a measure of the model's accuracy. Its formula is as follows:

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{\left| y_i - \hat{y}_i \right|}{y_i} .100\%$$

Figure 9: Formula of MAPE

Where n is the number of cars in the test set, $y_i$ is the true price, $y_i$ prime is the predicted price.

Lastly, Range Accuracy percentage is a percentage of how many of the predictions fall within a set prediction range. The range and its implementation will be discussed in a later section of this report.

## 5.9 Visualization of Data

Using graphs and plots to visualize the data proved extremely valuable for the improvement of the model. Four different types of plots were used, each providing helpful insight.

- Histograms

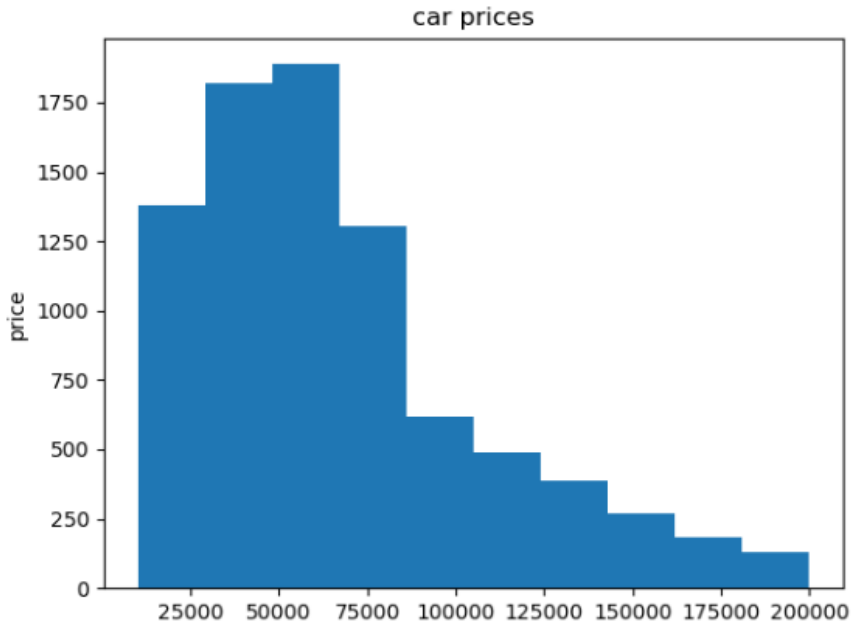


Figure 10: Histograms provide information about what range of prices is the most populated.

- Box plots

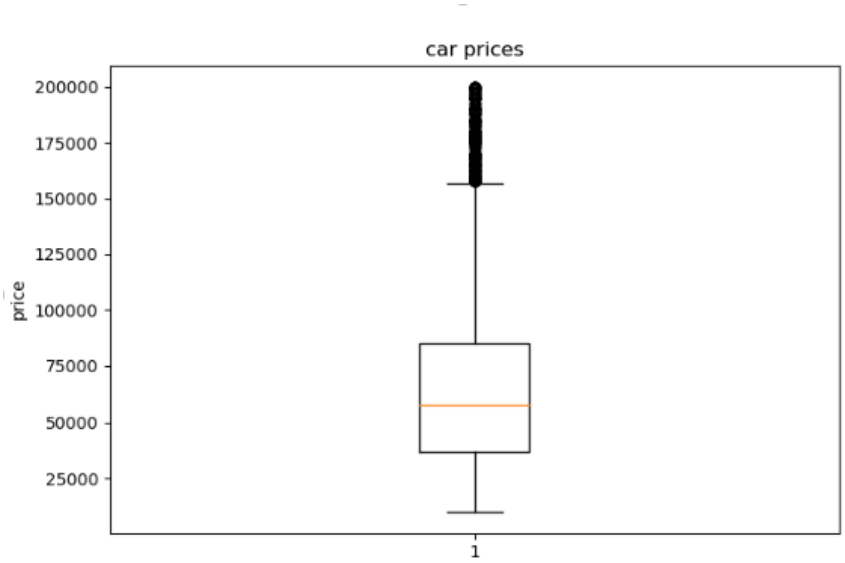


Figure 11: Box plots provide information about the existence of outliers.

- 'Predicted vs True' Scatter plot

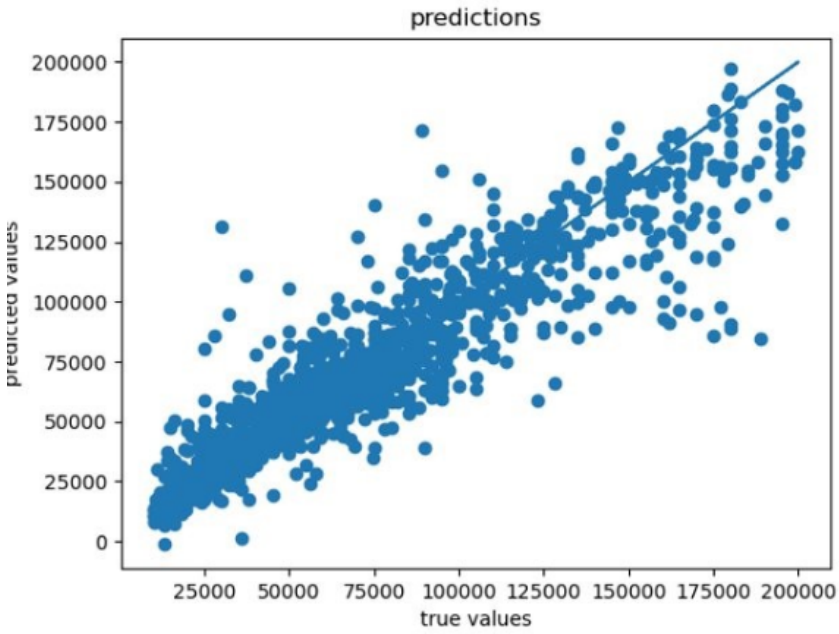


Figure 12: 'predicted vs true' scatter plot provides visual information about how well the model is performing.

- 'Error vs True' Scatter plot

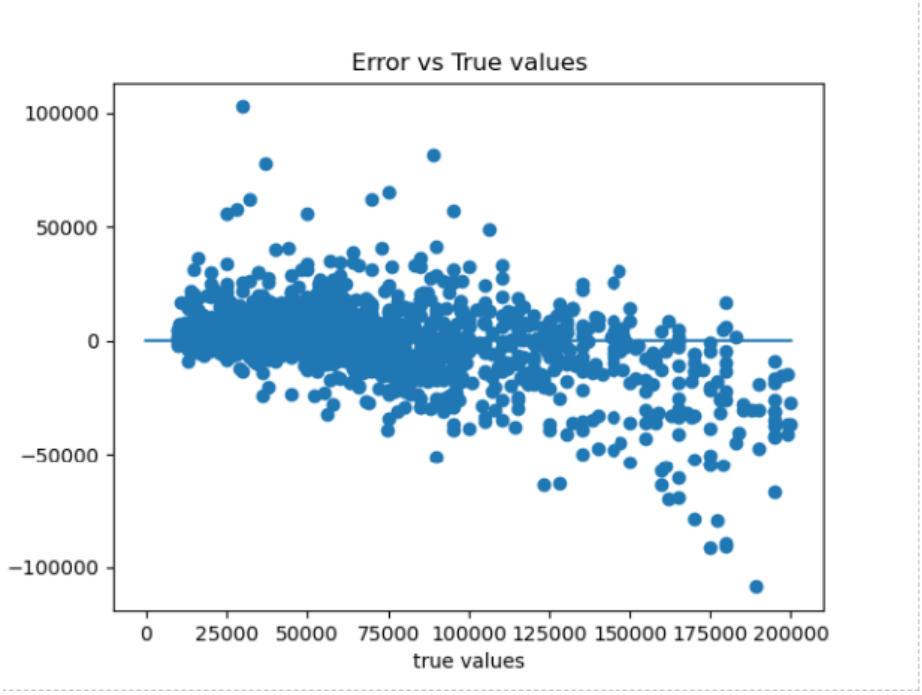


Figure 13: 'Error vs True' scatter plot provides information about where the model is underestimating or overestimating.

## 5.10 Trimming The Data

After inspecting some of the plots mentioned above, it became clear that the data was extremely skewed and needed to be trimmed. After trimming, excluding all cars with price higher than 200k, the error dropped considerably. Before trimming:

```
mean score for K-folds CV= 0.9316
Training Set Mean Absolute Error: 6987.6320
Test Set Mean Absolute Error: 21145.4327
r2 score= 0.9311
MAPE= 26.3037
```

Figure 14: Evaluation metrics before trimming the data (gradient boosting was used).

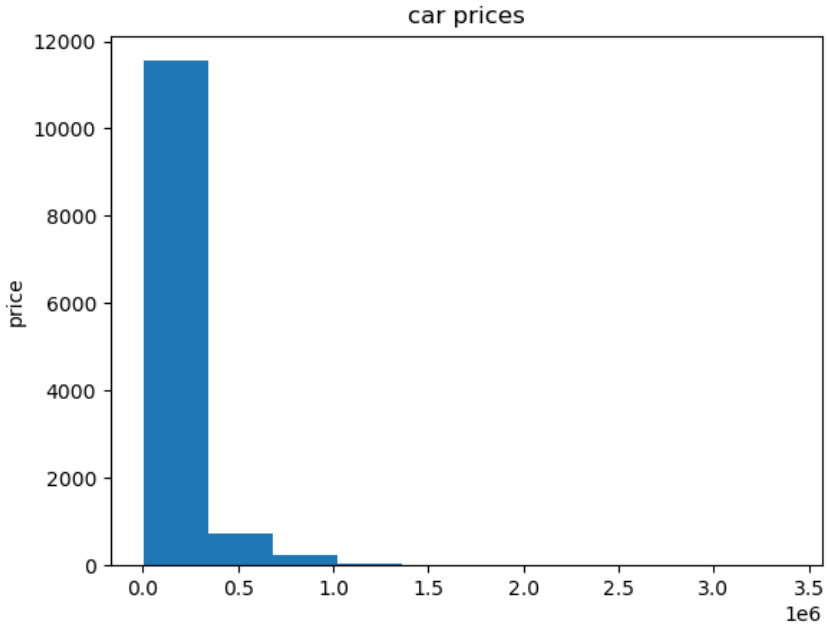


Figure 15: Histogram showing the data before trimming.

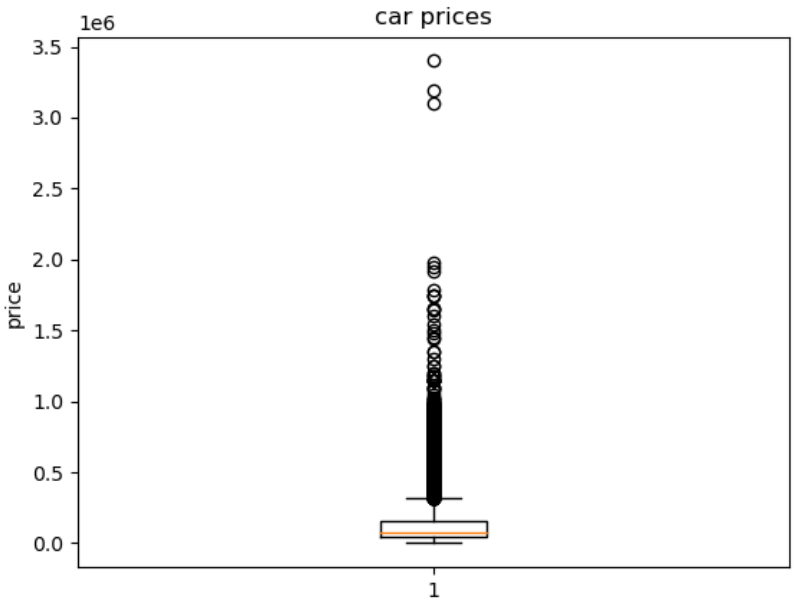


Figure 16: Box plot showing the data before trimming.

After trimming:

```
Training Set Mean Absolute Error: 3283.9847
Test Set Mean Absolute Error: 9869.0068
r2 score= 0.8568
MAPE= 16.5246
```

Figure 17: Evaluation metrics after trimming the data (gradient boosting was used).

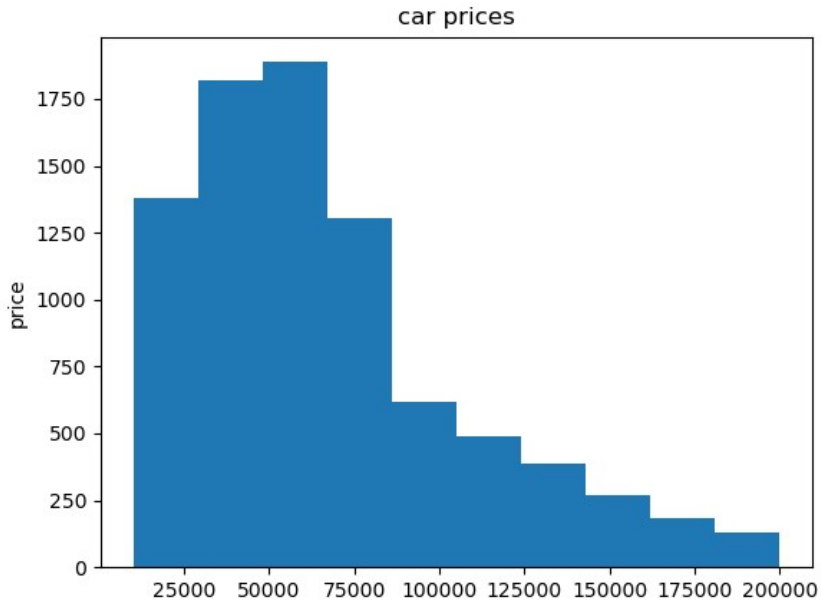


Figure 18: Histogram showing the data after trimming.

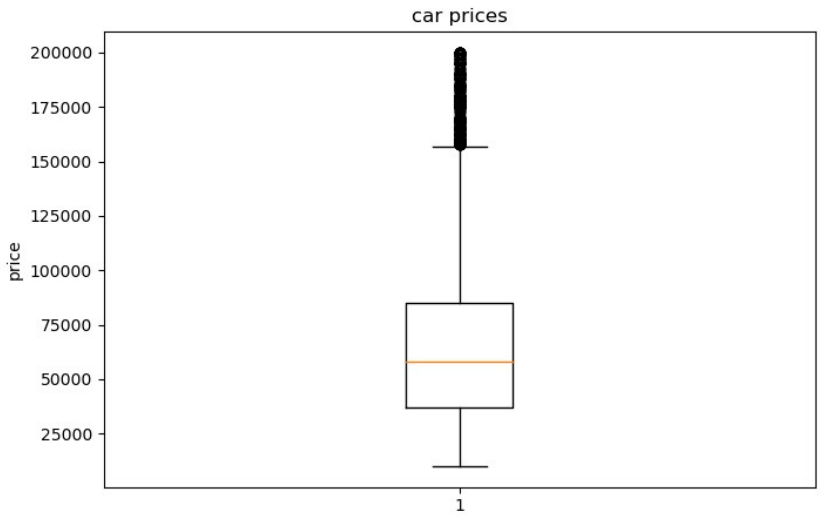


Figure 19: Box plot showing the data after trimming.

## 5.11 Estimation Range

Two types of ranges were implemented in this project, static and dynamic. Static range is a constant range applied to all predictions regardless of the predicted price, for example, if a car was estimated to be worth 150,000 riyals and the static range is 3500 riyals, then the estimation range would be [146,500 – 153,000] riyals. In contrast, the dynamic range is dependent on the predicted price so that each prediction made would have a unique estimation range. This was implemented by using the standard deviation of the set of predictions made by each individual decision tree for that particular car. However, one standard deviation proved too high for a feasible estimation range. Therefore, half a standard deviation was implemented as the dynamic range.

```
estimations = []
for i,x in enumerate(model.estimators_):
  estimations.append(model.estimators_[i].predict("The Car"))
arr = np.array(estimations)
rng_std = stdev(arr.ravel())
```

Figure 20: Code for implementing the dynamic range (one standard deviation).

```
Training Set Mean Absolute Error: 3283.9847
Test Set Mean Absolute Error: 9869.0068
r2 score= 0.8568
MAPE= 16.5246
range accuracy( 3500 ) =34.18
```

```
Training Set Mean Absolute Error: 3303.1295
Test Set Mean Absolute Error: 9856.3188
r2 score= 0.8557
MAPE= 16.4670
range accuracy( 5000 ) =45.40
```

```
Training Set Mean Absolute Error: 3268.3102
Test Set Mean Absolute Error: 9851.6658
r2 score= 0.8564
MAPE= 16.5315
range accuracy( 6000 ) =51.36
```

Figure 21: Different range accuracy percentages for different static range values (gradient boosting was used).

# 6 Web Application

Allowing the model to be used by the average consumer required making it easily accessible. For simplicity, a web application was developed. A web application runs on an online server and is accessible from anywhere through web browser. To develop a web application, a Front-End and Back-End system had to be developed. This will be explained in the next two paragraphs.

## 6.1 Front End

Front-End is the user interface. To develop the Front-End, a combination of different technologies such as HTML, CSS, and JavaScript was used. These technologies were provided by a very powerful web design tool called Bootstrap Studio. The tool provided a sizeable number of components for building the website. The web application was built using a combination of HTML, CSS, and JavaScript technologies with the use of Bootstrap studio tool.

## 6.2 Back End

Back-End is the layer that connects the user interface to the machine learning model. Flask was used as a Back-End for our application. Flask is a micro web framework based on Python. It assisted in creating get and post requests for the website as the connection between the Front-End and the machine learning model.

# 7 Conclusion

In this paper, a model was developed to aid the Saudi second-hand car market using machine learning methods. A data set, collected from multiple websites, had multiple outliers that reduced accuracy. As a result, the data set had to be trimmed, improving the predictions. The data set also had to be encoded to transform categorical data into numerical data that regression algorithms could understand. In addition, four different machine learning algorithms were

tested to decide which performed best. The results indicated Gradient Boosting regression to be the most suitable algorithm providing high accuracy. Results were analysed using statistical techniques such as mean absolute error and mean absolute percentage error. These methods assisted in understanding the results obtained from the multitude of tests conducted. A confidence interval was also added to further improve the results. The main limitation appeared in the data set and the confidence interval. The data, collected from multiple websites had many outliers, and lacked a wide variety of features available in a car. The confidence interval had to be static due to dynamic ranges varying greatly between predictions. Finally, the model was deployed as a web application to be accessible online for the consumers who desire assistance.

# 8 Appendices

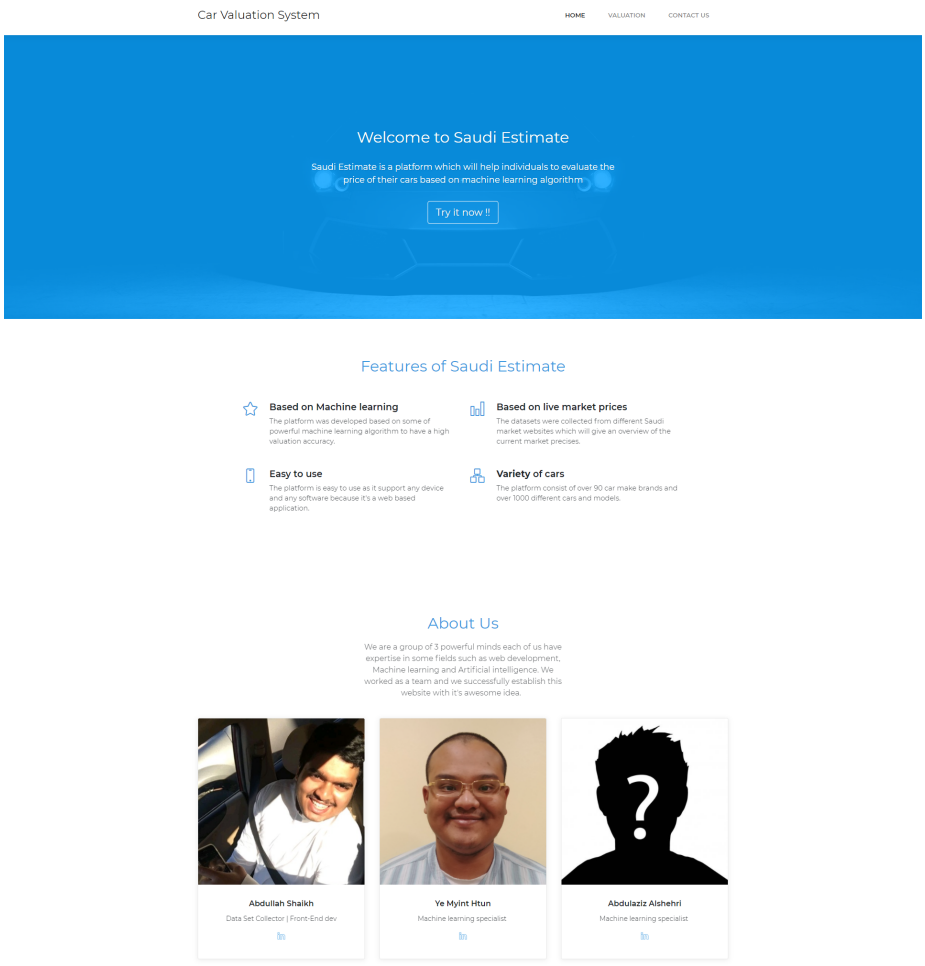


Figure 22: Index page of the website.

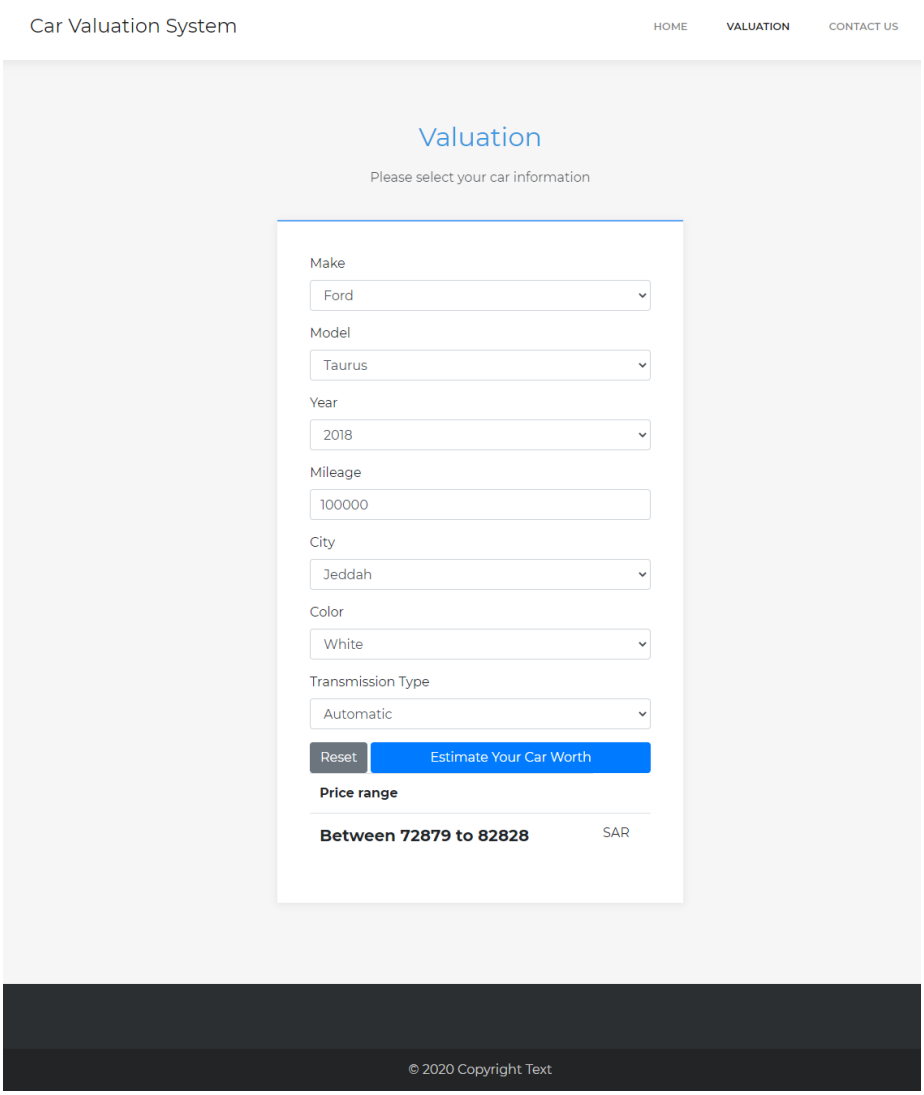


Figure 23: Sample run of the model.

This is a link for the code of the project: `https://github.com/userosq931655/SaudiEstimate-.git`.

# 9 Certificates

## 9.1 Abdulaziz Alshehri

Course Completion Certificate

knowledge hut

Business Skills Training

This is to certify that

**Abdulaziz Alshehri**

has successfully Completed the Business Skills Training

7 June - 11 June, 2020

UID:KH11-COURSECODE-CENTRO

SITE

Subramanyam M Reddy

Director, KnowledgeHut

Course Completion Certificate

knowledge hut

Data Science Fundamentals Training

This is to certify that

**Abdulaziz Alshehri**

has successfully Completed the Data Science Fundamentals Training

21 June - 25 June, 2020

UID:KH11-COURSECODE-CENTRO

SITE

Subramanyam M Reddy

Director, KnowledgeHut

## Course Completion Certificate

knowledge hut

### Blockchain Fundamentals Training

This is to certify that

**Abdulaziz Alshehri**

has successfully Completed the Blockchain Fundamentals Training

12 July - 16 July, 2020

UID:KH11-COURSECODE-CENTRO

SITE

Subramanyam M Reddy
Director, KnowledgeHut

## Course Completion Certificate

knowledge hut

### Cyber Security Training

This is to certify that

**Abdulaziz Alshehri**

has successfully Completed the Cyber Security Training

19 July - 23 July, 2020

UID:KH11-COURSECODE-CENTRO

SITE

Subramanyam M Reddy
Director, KnowledgeHut

LinkedIn LEARNING

Certificate of Completion

Congratulations, Abdulaziz Alshehri

## Machine Learning and AI Foundations: Value Estimations

Course completed on Jun 8, 2020 • 1 hour 4 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Tanya Staples

VP, Learning Content at LinkedIn

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: ASEc7F_yd1VVsCbyllV2CUS0zuEX

LinkedIn LEARNING

Certificate of Completion

Congratulations, Abdulaziz Alshehri

## Boosting Your Team's Productivity

Course completed on Jun 25, 2020

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Tanya Staples

VP, Learning Content at LinkedIn

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Project Management Institute

The PMI Registered Education Provider logo is a registered mark of the Project Management Institute, Inc.

Program: PMI® Registered Education Provider | Provider ID: #4101
Certificate No: Ac4XphYifiKZgVjsWLmiebFkrS2h
PDUs/ContactHours: 0.50 | Activity #: 100020003672

LinkedIn LEARNING

Certificate of Completion

Congratulations, Abdulaziz Alshehri

## Building High-Performance Teams

Course completed on Jun 25, 2020

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Tanya Staples

VP, Learning Content at LinkedIn

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Project Management Institute

The PMI Registered Education Provider logo is a registered mark of the Project Management Institute, Inc.

Program: PMI® Registered Education Provider | Provider ID: #4101
Certificate No: AaKzVUIZpvsnNk-TLDbGU-KnLJgq
PDUs/ContactHours: 2.25 | Activity #: 100020003044

LinkedIn LEARNING

Certificate of Completion

Congratulations, Abdulaziz Alshehri

## Building Your Team

Course completed on Jun 25, 2020 • 56 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Tanya Staples

VP, Learning Content at LinkedIn

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: AeRwgwzDFd_0jCo2-MSnWM0BSNpS

## 9.2 Ye Myint Htun

LinkedIn LEARNING

Certificate of Completion

Congratulations, Ye Myint Htun

COURSE COMPLETION

**Boosting Your Team's Productivity**

Course completed on Jun 21, 2020 • 39 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Tanya Staples

VP, Learning Content at LinkedIn

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: AbXTCg1yyWj_CcPV5F5usMIsXMfw

LinkedIn LEARNING

Certificate of Completion

Congratulations, Ye Myint Htun

COURSE COMPLETION

**Building High-Performance Teams**

Course completed on Jun 22, 2020 • 1 hour 50 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Tanya Staples

VP, Learning Content at LinkedIn

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: AdCkRoBCEt1p4uRA0dEHXt_PHTIo

LinkedIn LEARNING

Certificate of Completion

Congratulations, Ye Myint Htun

COURSE COMPLETION

## Building Your Team

Course completed on Jun 22, 2020 • 56 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Tanya Staples

VP, Learning Content at LinkedIn

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: AYmWjRuztYhHqcw0QNCPWb9JBc-9

LinkedIn LEARNING

Certificate of Completion

Congratulations, Ye Myint Htun

COURSE COMPLETION

## Deep Learning: Image Recognition

Course completed on Jun 12, 2020 • 1 hour 43 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Tanya Staples

VP, Learning Content at LinkedIn

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: AQr6H12UkZut9BxhT6AJLbd_9Q55


COURSE
COMPLETION
LinkedIn LEARNING
Certificate of Completion
Congratulations, Ye Myint Htun
Machine Learning and AI Foundations: Value Estimations
Course completed on Jun 9, 2020 • 1 hour 4 min
By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.
Tanya Staples
VP, Learning Content at LinkedIn
LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085
Certificate Id: AaVMQ7yj1yLKUbiqY3TMyAK5Zl5V


COURSE
COMPLETION
LinkedIn LEARNING
Certificate of Completion
Congratulations, Ye Myint Htun
Time Management Tips
Course completed on Jul 1, 2020 • 7 hours 51 min
By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.
Tanya Staples
VP, Learning Content at LinkedIn
LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085
Certificate Id: AR7GoEulqOSFYSzKRPSHZHZnYyOR

LinkedIn LEARNING

Certificate of Completion

Congratulations, Ye Myint Htun

COURSE COMPLETION

**Time Management: Working from Home**

Course completed on Jul 1, 2020 • 1 hour 25 min

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Tanya Staples

VP, Learning Content at LinkedIn

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Certificate Id: ATeUclgYdL6wBjN9kmGLX0LKqA9s

Course Completion Certificate

knowledge hut

Artificial Intelligence Fundamentals Training

This is to certify that

**Ye myint Htun**

has successfully Completed the Artificial Intelligence Fundamentals Training

28 June - 02 July, 2020

UID:KH11-COURSECODE-CENTRO

SITE

Subramanyam M Reddy

Director, KnowledgeHut

## Course Completion Certificate

knowledge hut

### Blockchain Fundamentals Training

This is to certify that

**Ye myint Htun**

has successfully Completed the Blockchain Fundamentals Training

12 July - 16 July, 2020

UID:KH11-COURSECODE-CENTRO

SITE

Subramanyam M Reddy
Director, KnowledgeHut

## Course Completion Certificate

knowledge hut

### Business Skills Training

This is to certify that

**Ye myint Htun**

has successfully Completed the Business Skills Training

7 June - 11 June, 2020

UID:KH11-COURSECODE-CENTRO

SITE

Subramanyam M Reddy
Director, KnowledgeHut


Course Completion Certificate
knowledge hut
Cloud Computing Fundamentals Training
This is to certify that
Ye myint Htun
has successfully Completed the Cloud Computing Fundamentals Training
14 June - 18 June, 2020
UID:KH11-COURSECODE-CENTRO
SITE
Subramanyam M Reddy
Director, KnowledgeHut


Course Completion Certificate
knowledge hut
Cyber Security Training
This is to certify that
Ye myint Htun
has successfully Completed the Cyber Security Training
19 July - 23 July, 2020
UID:KH11-COURSECODE-CENTRO
SITE
Subramanyam M Reddy
Director, KnowledgeHut

Course Completion Certificate

knowledge hut

Data Science Fundamentals Training

This is to certify that

**Ye myint Htun**

has successfully Completed the Data Science Fundamentals Training

21 June - 25 June, 2020

UID:KH11-COURSECODE-CENTRO

SITE

Subramanyam M Reddy

Director, KnowledgeHut

Course Completion Certificate

knowledge hut

IOT Fundamentals Training

This is to certify that

**Ye myint Htun**

has successfully Completed the IOT Fundamentals Training

05 July - 09 July, 2020

UID:KH11-COURSECODE-CENTRO

SITE

Subramanyam M Reddy

Director, KnowledgeHut

## 9.3 Abdullah Shaikh

Course Completion Certificate

knowledge hut

Artificial Intelligence Fundamentals Training

This is to certify that

**Abdullah Shaikh**

has successfully Completed the Artificial Intelligence Fundamentals Training

28 June - 02 July, 2020

UID:KH11-COURSECODE-CENTRO

SITE

Subramanyam M Reddy

Director, KnowledgeHut

Course Completion Certificate

knowledge hut

Blockchain Fundamentals Training

This is to certify that

**Abdullah Shaikh**

has successfully Completed the Blockchain Fundamentals Training

12 July - 16 July, 2020

UID:KH11-COURSECODE-CENTRO

SITE

Subramanyam M Reddy

Director, KnowledgeHut

## Course Completion Certificate

knowledge hut

Business Skills Training

This is to certify that

**Abdullah Shaikh**

has successfully Completed the Business Skills Training

7 June - 11 June, 2020

UID:KH11-COURSECODE-CENTRO

SITE

Subramanyam M Reddy
Director, KnowledgeHut

---

LinkedIn LEARNING

## Certificate of Completion

Congratulations, Abdullah Shaikh

COURSE COMPLETION

### Boosting Your Team's Productivity

Course completed on Jun 26, 2020

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Tanya Staples
VP, Learning Content at LinkedIn

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Project Management Institute

The PMI Registered Education Provider logo is a registered mark of the Project Management Institute, Inc.

Program: PMI® Registered Education Provider | Provider ID: #4101
Certificate No: AX-2A53mqkzORaSH958FDel9ZL5O
PDUs/ContactHours: 0.50 | Activity #: 100020003672

LinkedIn LEARNING

Certificate of Completion
Congratulations, Abdullah Shaikh

## Building High-Performance Teams

Course completed on Jun 26, 2020

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Tanya Staples
VP, Learning Content at LinkedIn

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Project Management Institute The PMI Registered Education Provider logo is a registered mark of the Project Management Institute, Inc.

Program: PMI® Registered Education Provider | Provider ID: #4101
Certificate No: AfIb3HQCbWjlX3flyw8S570omKyp
PDUs/ContactHours: 2.25 | Activity #: 100020003044

LinkedIn LEARNING

Certificate of Completion
Congratulations, Abdullah Shaikh

## Building Your Team

Course completed on Jun 26, 2020

By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.

Tanya Staples
VP, Learning Content at LinkedIn

LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085

Project Management Institute The PMI Registered Education Provider logo is a registered mark of the Project Management Institute, Inc.

Program: PMI® Registered Education Provider | Provider ID: #4101
Certificate No: ASaaDIrE0o1wZceQZThBW37UJLdo
PDUs/ContactHours: 0.75 | Activity #: 100020003185


COURSE
COMPLETION
LinkedIn LEARNING
Certificate of Completion
Congratulations, Abdullah Shaikh
Machine Learning and AI Foundations: Recommendations
Course completed on Jun 9, 2020 • 58 min
By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.
Tanya Staples
VP, Learning Content at LinkedIn
LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085
Certificate Id: AR4AIVq-78Nkr891CrNIVpLjZa6H


COURSE
COMPLETION
LinkedIn LEARNING
Certificate of Completion
Congratulations, Abdullah Shaikh
Machine Learning and AI Foundations: Value Estimations
Course completed on Jun 7, 2020 • 1 hour 4 min
By continuing to learn, you have expanded your perspective, sharpened your skills, and made yourself even more in demand.
Tanya Staples
VP, Learning Content at LinkedIn
LinkedIn Learning
1000 W Maude Ave
Sunnyvale, CA 94085
Certificate Id: ATHBPqAV15h53QAKRw_itNju7ueu

## Course Completion Certificate

knowledge hut

### Cloud Computing Fundamentals Training

This is to certify that

**Abdullah Shaikh**

has successfully Completed the Cloud Computing Fundamentals Training

14 June - 18 June, 2020

UID:KH11-COURSECODE-CENTRO

SITE

Subramanyam M Reddy
Director, KnowledgeHut

## Course Completion Certificate

knowledge hut

### Cyber Security Training

This is to certify that

**Abdullah Shaikh**

has successfully Completed the Cyber Security Training

19 July - 23 July, 2020

UID:KH11-COURSECODE-CENTRO

SITE

Subramanyam M Reddy
Director, KnowledgeHut

## Course Completion Certificate

knowledge hut

### Data Science Fundamentals Training

This is to certify that

**Abdullah Shaikh**

has successfully Completed the Data Science Fundamentals Training

21 June - 25 June, 2020

UID:KH11-COURSECODE-CENTRO

SITE

Subramanyam M Reddy
Director, KnowledgeHut

## Course Completion Certificate

knowledge hut

### IOT Fundamentals Training

This is to certify that

**Abdullah Shaikh**

has successfully Completed the IOT Fundamentals Training

05 July - 09 July, 2020

UID:KH11-COURSECODE-CENTRO

SITE

Subramanyam M Reddy
Director, KnowledgeHut


# References

[1] "Saudi arabia motor vehicle registered [2005 – 2020] [data & charts]," Jan 1970. [Online]. Available: https://www.ceicdata.com/en/indicator/saudi-arabia/motor-vehicle-registered

[2] S. Pudaruth, "Predicting the price of used cars using machine learning techniques," *International Journal of Information & Computation Technology*, vol. 4, pp. 753–764, 01 2014.

[3] E. Gegic, B. Isakovic, D. Kečo, Z. Mašetić, and J. Kevric, "Car price prediction using machine learning techniques," *TEM Journal*, vol. 8, pp. 113–118, 02 2019.

[4] C. Gleue, D. Eilers, H.-J. Mettenheim, and M. Breitner, "Decision support for the automotive industry: Forecasting residual values using artificial neural networks," *Business & Information Systems Engineering*, vol. 61, 02 2018.

[5] Ö. Çelik and U. Ö. Osmanoğlu, "Prediction of the prices of second-hand cars," 2019.

[6] S. Gültekin and A. Organ, "Price estimation of secondhand cars sold on the internet with artificial neural network method," *Journal of Internet Applications and Management*, vol. 11, pp. 49 – 61, 2020.

[7] W. Lee, N. Kim, Y.-H. Choi, Y. Kim, and B.-D. Lee, "Machine learning based prediction of the value of buildings," *KSII Transactions on Internet and Information Systems*, vol. 12, pp. 3966–3991, 08 2018.

[8] S. Lessmann and S. Voss, "Car resale price forecasting: The impact of regression method, private information, and heterogeneity on forecast accuracy," *International Journal of Forecasting*, vol. 33, pp. 864–877, 10 2017.