Second Hand Price Prediction for Tesla Vehicles

Sayed Erfan Arefin

Dhaka, Bangladesh

erfanjordison@gmail.com

Abstract—**The Tesla vehicles became very popular in the car industry as it was affordable in the consumer market and it left no carbo foot print. Due to large decline in the stock prices of Tesla Inc. at the beginning of 2019, Tesla owners started selling their vehicles in the used car market. This used car prices depended on attributes such as, model of the vehicle, year of production, miles driven and the battery used for the vehicle. Prices were different for a specific vehicle in different months. In this paper, it is discussed how a machine learning technique is being implemented in order to develop a second hand Tesla vehicle price prediction system. To reach this goal, different machine learning techniques such as decision trees, support vector machine (SVM), random forest and deep learning were investigated and finally was implemented with boosted decision tree regression. I future, it is intended to use more sophisticated algorithm for a better accuracy.**

Keywords—**SVM, Booted decision tree regression, Random forest, Second hand price prediction, Tesla vehicle**

I. INTRODUCTION

An American car manufacturing company Tesla Inc was founded on 2003. It is one of the companies that developed electric cars to a manufacturing standard and made it available for the consumer market [5]. In the beginning of 2019, the stock price of Tesla started to drop. This made Tesla vehicle owners feared and most of them started selling of their vehicle. A huge demand increased in the used car market for Tesla vehicles. But the price for Tesla vehicle depended on the stock price of the vehicle and it impacted largely in the used Tesla vehicle market.

Tesla vehicles highly depends on Tesla company's post sells services. In order to keep the vehicles usable in the market and also decline of the stock prices the accurate prediction of used car prices were very important for the new and existing customers and the company itself.

Predicting the price of a used Tesla vehicle became very important and also challenging. The whole process of making a second hand price prediction system consisted two sections. One section predicts the price of current tesla vehicles owned by a customer. It helps suggesting different driving habits which will minimize the price drop of the vehicle. Another section was to determine the price of Tesla vehicle by monitoring the prices of used cars from different websites that provided sells information of used Tesla vehicles. We will be focusing on how this second section worked. Become it is the most important part of the whole system.

The websites that provide this information includes truecar.com, autotrader.com, cars.com etc. We used a data scrapper to retrieve this information and feed it to our machine learning system in order to predict the price. The prices of a same vehicle were different in a range of three months (January, February and March of 2019). This means, the month of recorded price of a certain Tesla vehicle was also very important. The attributes that was found in the datasets were: Model, Year, Battery, Price, Miles, Exterior color, Interior type, Wheel type, Spoiler type etc. Since the prices were getting very saturated, four major attributes were focused to predict the price of Tesla vehicles in order to have an accurate result. These are: Model, Year, Battery, Price, Miles. In this paper how the datasets were used to implement to develop a prediction system using Microsoft Azure Machine Learning studio will be discussed. We will also discuss the test out of three different machine learning techniques that were primarily chosen to predict the price: Random forest, Boosted decision tree and K-nearest neighbor. Later the Boosted decision tree was selected as the main algorithm of the price prediction system, due to less errors.

II. BACKGROUND STUDY

Tesla shares faced their steepest price drop in 2019 when the company reported a large un expected loss. Around the same time their co-founder and Chief Technology Officer (CTO) JB Strubel departed from his executive ranks. The stock price dropped 13.6% and was $228.82 at the close. This was the largest drop since the share prices fell 13.9% on September 28, 2018. Later the stock price dropped more than 14% which is the steepest drop of price since 2013. [6]

Even though prices of stock were declining, the company increased their affords to prevent this. Some analysis suggested a similar decline of stock price was observed of a popular Tech company Netflix Inc. Later which recovered from such a situation. Eddie Yoon, founder of think and tank EddieWouldGrow, suggested Tesla's stock price in 2019 is very similar to Netflix's stock prices in 2011. [7]

It was very important for the company and also the customers to predict the price of second hand Tesla vehicles. The prices of a same vehicle were different in a range of three months (January, February and March of 2019). This means, as the time passed by prices of Tesla vehicle of a same type changed. Which is a very important factor in our machine learning technique to be used.

III. Dataset

This section includes dataset collection methods and daaset description. In sub-section III-A how the datasets were collected are discussed and in sub-section III-B data set description can be found. The attributes of the datasets and analysis of the dataset can be found in this sub-section.

A. Dataset collection

Tesla vehicle users sell their vehicles and this information can be retrieved from some well recognized websites. This includes truecar.com, autotrader.com, cars.com etc. There is another tool available online which is import.io. Using this tool data can be scrapped or extracted from the mentioned websites. 3 months of data of used cards were collected from cars.com, truecar.com and autotrader.com. Using import.io data was scrapped and datasets were collected in csv (Comma separated values) formatted files. Approximately total of 1600 entries were collected.

B. Description of Dataset

The dataset is labeled properly. There are 6 attributes of the dataset which are order independent and they are shown in table xyz. A sample of the dataset can be observed in table II. Attributes are described as follows.

- **Model:** This is the model of Tesla vehicle. There are four models available. They are, Model S, Model X, Roadstar 2dr and Model 3.
- **Year:** This indicates the manufacturing year.
- **Battery:** Battery indicates the battery model used for that specific vehicle. Battery is very important for a Tesla vehicle. Battery degradation, makes the battery to hold less charge which puts the vehicle in a bad condition. That also effects the price.
- **Price:** The price on which the vehicle was sold. The price is on US dollar.
- **Miles:** The total miles the vehicle has driven.

Rapid miner is a tool available for free, which can be used to pre process datasets, generate insights of the dataset and also run some machine learning techniques, such as deep learning, decision trees, support vector machine etc. [10]. Using this the following scatterplot and histogram were generated.
A scatterplot was plotted figure 2. In this graph, X-axis indicates the miles and Y-axis indicates price. The Models are marked with different colors.
It can be seen that, most of the Model 3 vehicles were sold before it reached a 20,000 miles drive by their owners and are sold at a price range of 40,000 USD to 70,000 USD. Average of Model 3 sold was 51,332.67 USD. For Model S, it can be observed that it has been sold in a vast range of driven miles, having large distribution in terms of mile, the average price it was sold is, 49,430.64 USD. Most of the Model X, were sold before it reached a 60,000 miles drive by the owners. The average price it was sold is, 83,475.01 USD which is much higher than any other models of Tesla vehicles. In case of Roadstar 2dr, the samples were very low.

TABLE I
THE DATA ATTRIBUTES

Serial	AttributeTitle	DataType
1	Model	String
2	Year	String
3	Battery	String
4	Price	Number
5	Miles	Number

TABLE II
Dataset Sample

Model	Year	Battery	Price	Miles
Model S	2013	Base	34200	36800
Model 3	2018	75	46995	2193
Model S	2018	75D	64900	1095
Model X	2016	P90D	84984	20680
Model S	2016	75D	58989	20303

Average price they were sold is, 51,493 USD.

A histogram was generated figure 1. In this graph, X-axis indicates the prices and Y-axis indicates frequency. The Models are marked with different colors. It can be observed that, the highest number of model sold was Model S, having a wide distribution in price. Compare to Model S, Model 3 and Model X are sold at a lower frequency. Price distribution of Model 3 is between 40,000 USD and 70,000 USD. Model X are sold over 60,000 USD. Roadstar 2dr had a minimum sale and is priced around 51,493 USD.
setminus The mean, median and mode of the whole dataset can be found in table III-B and model based average prices can be found in table IV.

IV. Implementation

As the dataset is labeled properly, it is considered to be used for supervised learning. In order to find out the best machine learning technique, different machine learning techniques were tested and based on the RMSE further decision was taken.
The following machine learning techniques were considered in the experiment.

A. Support Vector Machine (SVM)

Gegic, E., Isakovic, B. showed that in case of using datasets for Ford and Volkswagen the Support Vector Machine (SVM)

TABLE III
Dataset Mean, Median and Mode

Mean	Median	Mode
51041.84	46800	39999

TABLE IV
Average price based on models

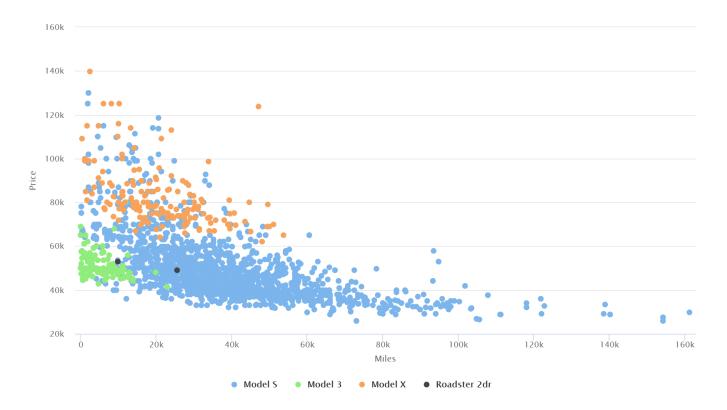| Models	Model 3	Model S	Model X	Roadstar 2dr
Average Price	51332.67	49430.64	83475.01	51493

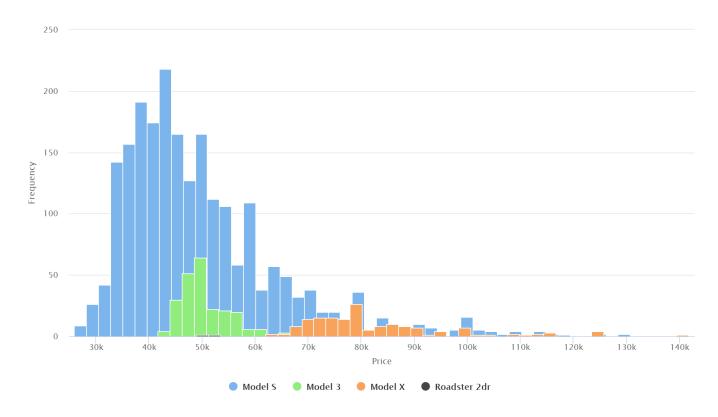Fig. 1. Boosted Decision Tree Implementation in Machine Learning Studio

Fig. 2. Scatter plot of the dataset: Price vs Frequency, colors indicating different models.

showed an accuracy of 48.23% which was higher than Artificial Neural Network algorithm [1]. In case of predictiong Tesla Prices SVM was also applied.

B. Deep Learning

Peerun, Saamiyah & Henna Chummun showed that a neural network which had one hidden layer with two nodes produced the smallest mean absolute error, among different machine learning techniques that were experimented. [4]. Though they also showed that SVM performed slightly well. They used data sets from Ford, Mazda, Opel, Renault, Toyota, Citreon prices in their paper.

Thus, in case of predicting the price of Tesla vehicles it will also be used in the experiment.

C. Random Forest

N. I. Nwulu showed that in case of oil price predictions, random forest performed well compare to decision tress, it had lower RMSE [2].

Pudaruth, Sameerchand also showed that in case of car price prediction with the dataset of Nissan and Toyota Random forest algorithm performed well compare to Decision Tree algorithm [3].

In Tesla price prediction Random forest algorithm was also used.

D. Decision Tree

Pudaruth, Sameerchand build decision trees with Nissan and Toyota cars prices as datasets [3]. They used decision tree to predict the prices and had a success rate of 84%. Thus, decision tree is also used in this experiment to predict Tesla Vehicle prices.

E. Boosted Decision Tree

In order to produce a prediction model, Grading boosting is a machine learning technique for regression and classification problems. This technique is mainly used for decision trees. The model is built in a stage-wise approach, similar to other boosting methods and it generalized by optimization of an arbitrary differentiable loss function.

Gradient boosted trees had performed well in the Rapid Miner experiment. It can be seen in the results from Rapid Miner containing the RMSE figure 5.

Boosted Decision Tree Regression means, the decision tree is formed based on boosting. This means, each tree is dependent on prior trees. Preceded tree residual is fitted in the current tree and this is how the algorithm learns.

Thus Boosted Decision Tree Regression was also considered for the experiment.

F. Error comparison

RMSE also known as Root Mean Square Error, is a good way to evaluate recommender systems. RMSE penalizes an algorithm more incase of wrong results. Thus, it was considered for deciding which Machine learning technique to be used for the prediction system. In table VII the RMSE results of different machine learning techniques are given, that was generated using the Tesla vehicle datasets.

It can be seen that Gradient Boosted Trees had a RMSE of 5849.8. But for Boosted Decision Tree Regression implementation on Machine Learning Studio, shows a RMSE of 4969.53 table V. Which is slightly better than Gradient Boosted Trees. Thus Boosted Decision Tree Regression was considered for the final implementation of the prediction system.

G. Final Implementation

As the boosted decision tree regression shows the lowest RMSE, it was decided to be used in the machine learning API. Machine Learning Studio is provided by Microsoft Azure [11]. In this platform models can be trained by uploading datasets, different machine learning techniques can be defined Later the trained data can be used to deploy a webservice to be consumed by other backend services. This has a modular interface.

In the implementation first, the datasets of different months were uploaded and combined into one dataset, then the columns discussed in section III (Model, Year, Battery, Price, Miles) were selected. There were some rows with missing data. Those were cleaned up before being used in the training. After that, Boosted Decision Tree Regression was executed and it provided a scored model. Web service input and web service output was also used but those had no purpose in this step, but need to be added for creating the web service later on. From the scored module the data can be converted to CSV and later was downloaded. A snapshot of the data is given in Table VI. The whole implementation can be visualized in figure 3. Modules used for implementing the Boosted Decision Tree Regression are given below with explanations.

- **Add Rows:** This module take two datasets as inputs and merges them. The columns of the datasets need to be same. The output contains the merged dataset.
- **Select Columns in Dataset:** This module helps select columns of the dataset provided in the input and outputs a dataset with the selected columns.
- **Clean Missing Data** This module cleans all the rows which have missing values.
- **Edit Metadata:** With this, types of the attributes can be changed or corrected in case of the attributes not being presented properly.
- **Split Data:** With this module the set can be split into two datasets based on the provided ratio. In this case the ratio was 75% and 25 %. The split was randomized. The first output gives the larger split and the second output provide the second half of the split.
- **Boosted Decision Tree Regression:** This module along with the Train Model module runs the Boosted Decision Tree Regression training. The algorithm is pre implemented with the module.
- **Train Model:** This takes in the module which defines the algorithm to use and the second input takes in the dataset that will be used for training.
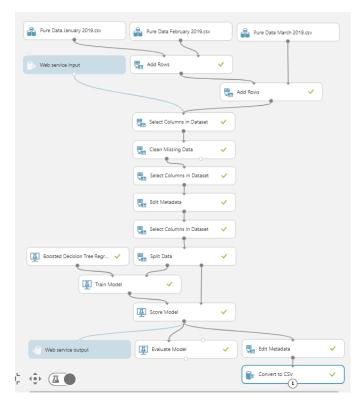
Fig. 3. Boosted Decision Tree Implementation in Machine Learning Studio

Fig. 4. Machine learning web-service with the trained model

This is only used for regression and classification problems. So, it is suitable for this implementation.

- **Score Model:** This generates a score or predicted value. In this case, a sample of the scored model is given in table VI.
- **Evaluate Model:** This generates Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Relative Absolute Error, Relative Squared Error, Co-efficient of Determination. In this case, the results from evaluate model are given in table V
- **Web service input:** This module takes input when deployed as web service.
- **Web service output:** This module provide response to web service input when deployed as web service.
- **Convert to CSV:** This converts the input into a CSV file, which can be downloaded for further reference.

It took 26 seconds to run the experiment. Because this is hosted in a cloud environment in this case, Microsoft Azure, it took very little time to run. After completing the experiment, it was turned into a webservice, where the Web service input and Web service output were used along with the trained model. This can be seen in figure 4.

This exposed a web service API, which later consumed by the back end. This API is discussed in section V, while testing.

V. TESTING

Testing the implemented API is a very important step. It will be tested if the price is predicted properly. Also, it
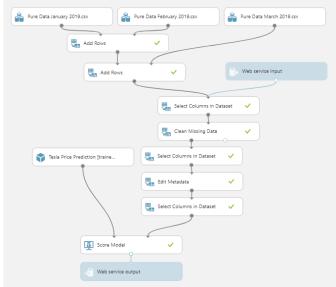

Fig. 5. Results of Rapid Miner: RMSE

TABLE V
ERRORS FROM EVALUATE MODEL

Error	Value
Root Mean Squared Error	4969.534921
Relative Absolute Error	0.297097
Relative Squared Error	0.098898
Coefficient of Determination	0.901102
Mean Absolute Error	3478.598936

Model	Year	Battery	Price	Miles	Scored Labels
Model S	2013	Base	34200	36800	36950.03516
Model X	2017	90D	88470	12229	88367.42969
Model 3	2018	75	46995	2193	52845.39063
Model X	2017	P100D	98500	33959	110876.7656
Model S	2014	Performance	44855	44980	44208.11328
Model X	2018	75D	84995	1379	83715.78906
Model X	2016	P90D	81500	25250	79317.74219
Model S	2016	90	59855	23010	52286.07031
Model S	2017	P100	98000	18871	104973.0547
Model S	2013	60	43995	15559	39556.19922
Model S	2015	70D	47995	47367	42450.09766
Model S	2016	90D	53995	55370	54899.95703
Model 3	2018	75	54900	1254	52441.74609
Model S	2012	Performance	36995	38030	42229.39844

Model	Root Mean Squared Error	Standard Deviation
Generalized Linear Model	6812.9	383.2
Deep Learning	6110.2	346.0
Decision Tree	9267.3	740.2
Random Forest	7557.1	542.7
Gradient Boosted Trees	5849.8	341.4
Support Vector Machine	10108.7	708.1

will be tested if the machine learning webservice is working properly.

In order to test if the price is predicted properly, the API is fed with Used vehicle prices randomly. A small piece of code that is used to run a loop. Each time of the loop, a request has been made to the API with the entries from the April dataset. The predicted price is listed with the corresponding entry. The predicted price is subtracted from the original price in order to find the error margin. It showed a RMSE to the RMSE found in the evaluate model errors shown in table V To run this piece of code, firstly it must me tested if the machine learning API is working properly. Web services in the Machine Learning Studio includes an input form to test out the webservice. It was tested and worked properly. The request form is shown in figure 6.

To test the API from another source, Postman will be used. Postman is a tool that helps create a Rest API request using all the necessary requirements such as providing the request with header, parameters, authorization tokens etc. [9]. To test out the web services API with Postman parameters, headers and authorization is populated with the values in table VIII, IX, X subsequently, which were retrieved from the web services API. It is also shown in figure 7, 8 subsequently. The request is finally populated with the following JSON (Java Script Object Notation) which included all the parameters required.

```
1   {
2       "Inputs": {
3           "input1":
4           [
5               {
```
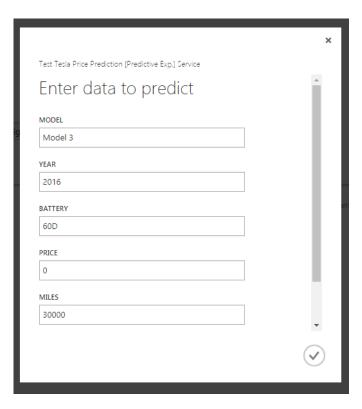

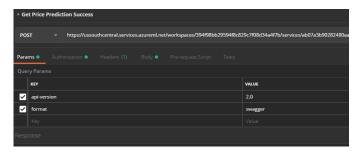Fig. 6. Web service input on Machine learning studio

Fig. 7. Testing with Postman: parameters for request.

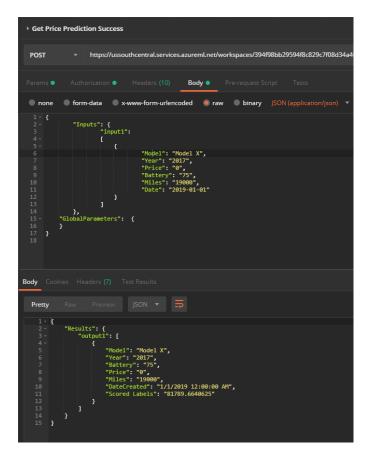Fig. 8. Testing with Postman: Headers for request.

TABLE VIII
POSTMAN PARAMETERS

Key	Value
api-version	2
format	swagger

TABLE IX
POSTMAN HEADERS

Key	Value
Content-Type	application/json

VI. THE FINAL SYSTEM

The final system was developed using the MERN stack. MERN stands for Mongo DB, Express Js, React Js and Node Js. The backend of the system was developed on Typescript. User needs to sign up in order to use the system and provide the VIN of the vehicle.

VIN is vehicle identification number. Which is unique for all the vehicles, similar to mac addresses that is being used in case of network devices. From VIN a vehicle model, year of production, manufacturer name etc can be extracted. Thus, it was used as an entry point to the system. With this, and the prediction web service, the system showed a predicted price for that specific vehicle.

Improvements on the final system is considered for future work.

VII. CONCLUSION

Predicting the price of a used vehicle can be challenging. In this paper, a price prediction system for Tesla vehicles was discussed. Different machine learning techniques were performed with the collected dataset in order find the lowest RMSE. Techniques considered were: Deep Learning, Support Vector Machine, Boosted Decision Tree Regression, Random Forest and Decision Trees. Boosted Decision Tree algorithm showed the lowest RMSE. It was later implemented in Machine Learning Studio and with the trained model a web service was deployed. Which was later used in the main back end of the prediction system. In the future it is intended to improve the accuracy of the prediction.

REFERENCES

[1] Gegic, E., Isakovic, B., Keco, D., Masetic, Z. and Kevric, J. (2019). Car Price Prediction using Machine Learning Techniques. TEM Journal, 8(1), pp.113-118.
[2] N. I. Nwulu, "A decision trees approach to oil price prediction," 2017 International Artificial Intelligence and Data Processing Symposium (IDAP), Malatya, 2017, pp. 1-5. doi: 10.1109/IDAP.2017.8090313
[3] Pudaruth, Sameerchand. (2014). Predicting the Price of Used Cars using Machine Learning Techniques. International Journal of Information & Computation Technology. 4. 753-764.

TABLE X
POSTMAN AUTHORIZATION

Key	Value
Bearer Token	{Retrieved from web service portal, kept secret}

Fig. 9. Testing with Postman: Request Body and Response

```
6               "Model": "Model X",
7               "Year": "2017",
8               "Price": "0",
9               "Battery": "75",
10              "Miles": "19000",
11              "Date": "2019-01-01"
12            }
13          ]
14        },
15      "GlobalParameters": {
16      }
17    }
```

The response from the API is shown below. It is received also as JSON.

```
1     {
2       "Results": {
3         "output1": [
4           {
5               "Model": "Model X",
6               "Year": "2017",
7               "Battery": "75",
8               "Price": "0",
9               "Miles": "19000",
10              "DateCreated": "1/1/2019 12:00:00AM",
11              "Scored Labels": "81789.6640625"
12            }
13          ]
14        }
15    }
```

This shows that the web service API is functioning properly. Then the error calculation step was executed.

[4] Peerun, Saamiyah & Henna Chummun, Nushrah & Pudaruth, Sameerchand. (2015). Predicting the Price of Second-hand Cars using Artificial Neural Networks.

[5] Wikipedia. (2019). Tesla, Inc.. [online] Available at: https://en.wikipedia.org/wiki/Tesla,_Inc. [Accessed 31 Aug. 2019].

[6] Ari Levy, L. (2019). Tesla suffers its worst day of the year after brutal earnings report and loss of technology chief. [online] CNBC. Available at: https://www.cnbc.com/2019/07/25/tesla-is-having-worst-day-of-2019-after-earnings-and-loss-of-cto.html [Accessed 2 Sep. 2019].

[7] Ari Levy, L. (2019). Tesla suffers its worst day of the year after brutal earnings report and loss of technology chief. [online] CNBC. Available at: https://www.cnbc.com/2019/07/25/tesla-is-having-worst-day-of-2019-after-earnings-and-loss-of-cto.html [Accessed 2 Sep. 2019].

[8] Ericson, G., Petersen, T., Lu, P., Takaki, J., Martens, J. and Sharkey, K. (2019). Boosted Decision Tree Regression - Azure Machine Learning Studio. [online] docs.microsoft.com. Available at: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/boosted-decision-tree-regression [Accessed 2 Sep. 2019].

[9] Postman. (2019). About Postman — Meet Our Founders, Learn About Our Values, & More. [online] Available at: https://www.getpostman.com/about-postman [Accessed 3 Sep. 2019].

[10] RapidMiner. (2019). Key Features of RapidMiner Studio — RapidMiner. [online] Available at: https://rapidminer.com/products/studio/feature-list/ [Accessed 1 Sep. 2019].

[11] Gilley, S., Ericson, G., Martens, J. and Rohm, W. (2019). What is - Azure Machine Learning Studio. [online] docs.microsoft.com. Available at: https://docs.microsoft.com/en-us/azure/machine-learning/studio/what-is-ml-studio [Accessed 30 Aug. 2019].