# Weekly Project Meeting Minutes

*The main purpose of the document is to capture all the work that has been done by the group over the course of one week and* ***not*** *to write down what was discussed in a single meeting. You should be meeting and/or working throughout the week*.

**Price Prediction of Used Cars**

**Time group spent on project:** 9 hours .

**Group Number**: 7

**Group members present (Name, ID):**

Delta Joseph(0735816)

Amith John Varkey(0735036)

Abdul Hannan Mansoor(0735818)

**Specific Activities from prior week:**

* We have worked on price prediction of used cars dataset from **Kaggle** to create a model on Artificial Neural Network (Keras) and XGBoost Regressor Model with k=5 parameter.
* Datasets(Kaggle and Automax) was splitted into .02 and 0.25 ratio

**Real Time dataset(AUTOMAX**)

* Null values was identified on real time dataset(AUTOMAX)
* Converted the categorical values into dummies using the get\_dummies function .
* Unwanted columns was dropped.
* We have created models like Linear Regression Model,K-Neighbours Regression,Decission Tree Regression ,Random Forest,Gradient Boosted Regression Model, XGBoost Regressor .
* We also did visualizations for every model.
* Completed project report writing till related work.

**Specific Output from prior week:**

1. Splitting the dataset(Kaggle) with the ratio **80%**(Training) and **20%**(Test).

|  |  |  |
| --- | --- | --- |
| **Model** | **MAE(%Test)** | **R2\_Score** |
| Linear Regression | -5037658137424.00 | 46026261912.8 |
| K-Neighbor Regression | 18.7 | 0.87 |
| Decision Tree Regression | 20.4 | 0.84 |
| Random Forest Model | 15.8 | 0.92 |
| Gradient Boosted Regression | 54.4 | 0.38 |
| Ridge Regression | 26.1 | 0.83 |
| Lasso Ridge Regression | 25.9 | 0.82 |
| Artificial Neural Network(Keras) | 15.4 | 0.91 |
| XGBoost Regressor(k=5) | 15.8 | 0.89 |

* Python link: <https://github.com/DAB-400-Capstone/Final-project/blob/master/Project%20.ipynb>

1. Splitting the dataset(Kaggle) with the ratio **75%**(Training) and **25%**(Test).

|  |  |  |
| --- | --- | --- |
| **Model** | **MAE(%Test)** | **R2\_Score** |
| Linear Regression | 47701943920.3 | -18146552 |
| K-Neighbor Regression | 19.2 | 0.84 |
| Decision Tree Regression | 20.03 | 0.82 |
| Random Forest Model | 16.1 | 0.89 |
| Gradient Boosted Regression | 51.0 | 0.43 |
| Ridge Regression | 25.8 | 0.83 |
| Lasso Ridge Regression | 25.6 | 0.81 |
| Artificial Neural Network(Keras) | 15.6 | 0.89 |
| XGBoost Regressor(k=5) | 15.2 | 0.90 |

* Python Link : <https://github.com/DAB-400-Capstone/Final-project/blob/master/Code_with_test_size%3D0.25.ipynb>

1. Splitting the dataset(**Automax**) with the ratio **80%**(Training) and **20%**(Test).

|  |  |  |
| --- | --- | --- |
| **Model** | **MAE(%Test)** | **R2\_Score** |
| Linear Regression | 67822346 | -79005870927 |
| K-Neighbor Regression | 16.3 | 0.54 |
| Decision Tree Regression | 16.4 | 0.49 |
| Random Forest Model | 12.0 | 0.73 |
| Gradient Boosted Regression | 18.2 | 0.44 |
| Artificial Neural Network(Keras) | 12.0 | 0.75 |
| XGBoost Regressor(k=5) | 11.8 | 0.76 |

Python Link: <https://github.com/DAB-400-Capstone/Final-project/blob/master/Automaxx.ipynb>.

* As we know , lowest the MAE value the better model we have .
* According results of model we can conclude XGBoost Regressor was the best model.

**Contributions from Delta Joseph**

* Worked on report writing
* Worked on all models for test size(0.2) on Kaggle dataset.
* Explained the concept of keras.
* Worked on data cleaning on “Automax” dataset.

**Contributions from Amith John Varkey**

* Worked on deep learning model
* Worked on report writing
* Worked on all models for test size(0.25) on Kaggle dataset.
* Worked on identifying the null values in Automax dataset and imputing the data accordingly

**Contributions from Abdul Hannan**

* Worked on XGboost Regressor in both dataset (Kaggle and Automax)
* Worked on report writing.
* Worked on model building in Automax dataset.
* Comparing the models.

**On Target:**

Indicate the current status of your project

\_\_\_\_\_ green: everything on track for completion by due date

\_\_\_\_\_ yellow: a small number of tasks are off track and completion by due date is at risk

\_\_\_\_\_ red: many tasks are off track and project will not be completed by due date

**Challenges/Disagreements:**

* Identify the columns which doesn’t help for model building.
* Identify the best feature corelated to target variable.
* Converting categorial values into dummy values.

**Planned Activities for coming week from all:**

**Delta Joseph**

* Will be working on the first 3 Chapters of the report.
* Work on Project Presentation.
* Will help in hyper parameter tuning.

**Amith John Varkey**

* Will be working on the Methods and conclusion part in report writing and will also start in making the final power point presentation.
* Work on hyper-parameter tuning in order to increase the accuracy.

**Abdul Hannan**

* Will work on the remaining part of the report and also will help in making the presentation.
* Will help in hyperparameter tuning.