**Description of the topic:**

The aim of this project is to predict the price of motor bikes according to the Bangladeshi market. This has been done with the help of regression of Machine Learning. We have made a dataset from scratch. After preprocessing and feature extraction we have trained different models of regression and then tested them with the help of our dataset.

**Brief description of the dataset:**

The name of our dataset is Bike\_Price\_Dataset. There are 8 attributes and 301 samples in the dataset. The data has been collected from four different bike selling websites of Bangladesh. Different features of bikes like mileage, engine capacity, top speed, bike type etc. are there in the dataset.

**Description Of The Used ML Models:**

1. **Linear Regression:** Itis a linear approach to find the relationship between a dependent variable and one or more independent variables. When there is only one independent variable then it is called simple linear regression.
2. **Lasso Regression:** Lasso regression is a type of [linear regression](https://www.statisticshowto.com/probability-and-statistics/regression-analysis/find-a-linear-regression-equation/) that uses [shrinkage](https://www.statisticshowto.com/shrinkage-estimator/). Shrinkage is where data values are shrunk towards a central point, like the [mean](https://www.statisticshowto.com/mean/).
3. **Extreme Gradient Boosting:** XGBoost is an implementation of Gradient Boosted decision trees. In this algorithm, decision trees are created in sequential form. Weights play an important role in XGBoost. Weights are assigned to all the independent variables which are then fed into the decision tree which predicts results. Weight of variables predicted wrong by the tree is increased and these variables are then fed to the second decision tree. These individual classifiers/predictors then ensemble to give a strong and more precise model. It can work on regression, classification, ranking, and user-defined prediction problems.
4. **Support vector machine:**In [machine learning](https://en.wikipedia.org/wiki/Machine_learning), support-vector machines (SVMs) are [supervised learning](https://en.wikipedia.org/wiki/Supervised_learning) models with associated learning [algorithms](https://en.wikipedia.org/wiki/Algorithm) that analyze data for [classification](https://en.wikipedia.org/wiki/Statistical_classification) and [regression analysis](https://en.wikipedia.org/wiki/Regression_analysis). In addition to performing [linear classification](https://en.wikipedia.org/wiki/Linear_classifier), SVMs can efficiently perform a non-linear classification using what is called the [kernel trick](https://en.wikipedia.org/wiki/Kernel_method#Mathematics:_the_kernel_trick), implicitly mapping their inputs into high-dimensional feature spaces.
5. **Bayesian regression:** The aim of Bayesian Linear Regression is not to find the single “best” value of the model parameters, but rather to determine the posterior distribution for the model parameters.Bayesian linear regression allows a fairly natural mechanism to survive insufficient data, or poor distributed data.
6. **Ada Boost:** Ada boost regressor is a meta-estimator that begins by fitting a regressor on the original dataset and then fits additional copies of the regressor on the same dataset but where the weights of instances are adjusted according to the error of the current prediction. As such, subsequent regressors focus more on difficult cases.

**Comparison of the performance scores of the models:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No.** | **ML Models** | **RMSE Score** | **MAE Score** | **MSE Score** | **R2 Score** |
| **1** | Linear Regression | 100656.9808 | 264.437 | 10131827788.345 | 0.8946 |
| **2** | SVM | 125150.98 | 247.328 | 15662767893.68 | 0.837089 |
| **3** | Lasso Regression | 100656.98082 | 264.4369 | 10131827789.734 | 0.8946 |
| **4** | Bayesian regression | 100809.373 | 263.339 | 10162529772.19 | 0.89429 |
| **5** | Extreme Gradient Boosting | 31832.6885 | 148.023 | 1013320058.582 | 0.9894 |
| **6** | AdaBoost Regressor | 64325.919 | 236.715 | 4137823902.635 | 0.9569 |

**Discussion:**

Here we successfully predict the motor bike price using different machine learning algorithms. The project uses a total of 6 machine learning algorithms where each has different performance values. In the end the model’s performance value are compared for better understanding of the models.

**Contribution:**

Sadman Sadik Hasan(180104110)

Dataset Collection, Data Pre Processing, Showing Co Relation, Applying ML algorithms (Extreme Gradient Boosting, AdaBoost Regressor),

Providing visual differences in evaluation metrics (RMSE Score, R2 Score, MAE Error) for different ML Algorithms.

Ayesha Afroze(180104118)

Dataset Collection, Data Pre Processing, Applying ML algorithms (LinearRegression, Bayesian regression)

Fabliha Nahid(180104116)

Dataset Collection, Data Pre Processing, Applying ML algorithms (SVM, Lasso Regression)

