52.5 %

Manufacturing Waste is Landfilled

Information retrieved from: https://www.epa.gov/facts-and-figures-about-materials-waste-and-recycling/national-overview-facts-and-figures-materials

Quality Predictions In The Mining Process

ML 3253 – Group # 21 - University Of Toronto – August 2019

The Team

Adam Gregg

Physics,
University Of Toronto

Nisarg Patel

Mechatronics Eng.
McMaster University

Omar Hamdy

Chemical Eng.
University Of Toronto

Khurram Shafiq

Computer Eng.
University Of Ottawa

AGENDA

1. Problem
2. Solution
3. Key Outcomes

Problem Statement

Predicting **% Silica (Impurity)** present in the Iron Ore concentrate

Background Process

Flotation Plant
[Physical Separation]

Iron Ore

% Silica

Other

DATASET

# of Instances	737,453
# of columns	24
Target Variable	% Silica Concentrate
Data Collection Period	20 seconds
Output Update Period	2 hours
Range	6 months
Missing Values	None
Problem Type	Supervised - Regression

Dataset retrieved from: https://www.kaggle.com/edumagalhaes/quality-prediction-in-a-mining-process

Data Preparation

Item	Approach # 1	Approach # 2
Sampling Period	20 seconds	1 hour
Outlier Detection & Removal	Z Score	Box-plots & IQR Score
Feature Selection	Manual	Semi-automatic
----------Output----------		
Total Features	6 (2 – engineered)	8 (2 – engineered)
Target Variable	1	1
# of Instances	346,000	1,817

ML Modelling Journey

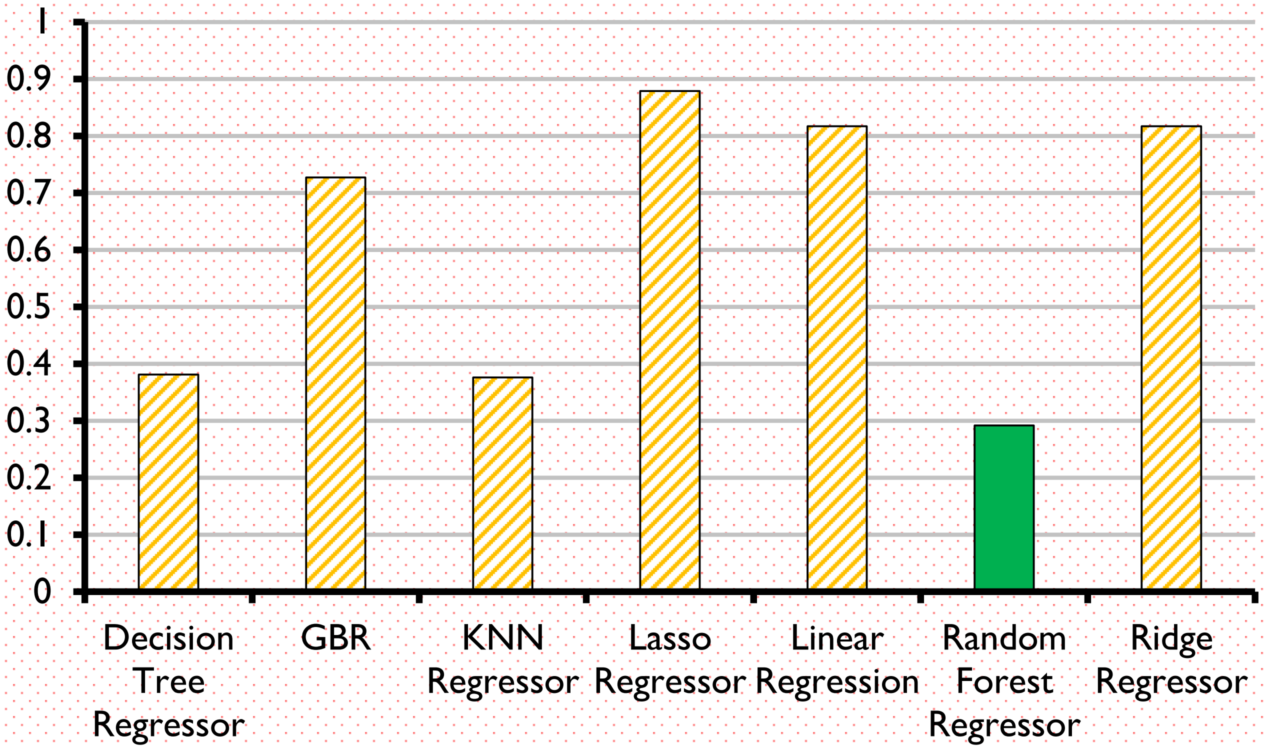
- Split Data
- Built Pipeline
- Scaling
- ML Model
- Train Model
- Evaluate Base Model
- Hyperparameter Tuning
- Re-evaluate Model
- Finalize

ML Modelling

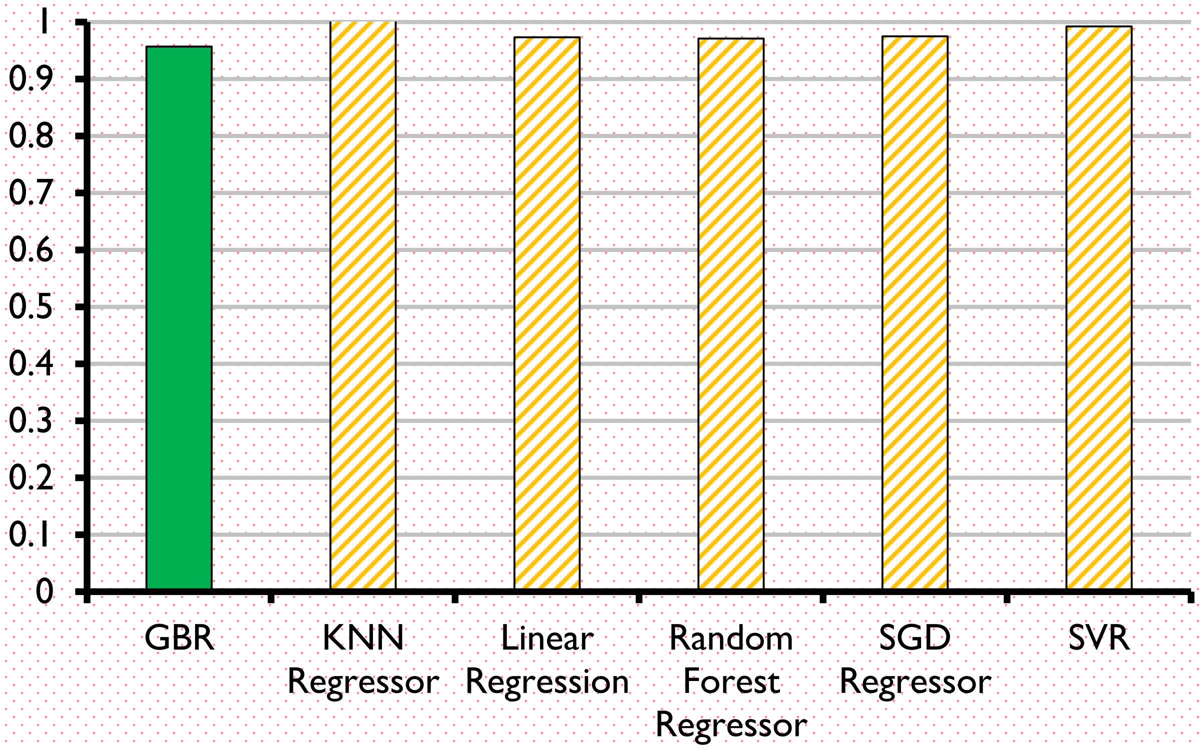
Item	Model # 1	Model # 2
Base Algorithms Used	LinearRegression,, KNeighborsRegressor, RandomForestRegressor, GradientBoostingRegressor, SGDRegressor	
Other Algorithms Used	Decision Tree Regressor, Ridge, Lasso	SVR
Scaling Data	Min_Max Scaler	Standard Scaler
Test Train Split	75% (train) - 25% (test)	70% (train) - 30% (test)
Cross Validation	K-folds 10 split	
Hyperparameter Tuning	Grid Search	
Scoring Methods	RMSE, R^2, Accuracy	

Results

Model # 1: RMSE Score



Model # 2: RMSE Score



Results

Before Tuning		
Scoring	**Model # 1 – Random Forest Regressor**	**Model # 2 – Gradient Boosting Regressor**
RMSE	0.282	0.913
R^2	0.897	0.221
Accuracy	93.64 %	62.24 %
After Tuning		
RMSE	0.282	0.907
R^2	0.908	0.232
Accuracy	93.80 %	62.89 %

Key Outcomes

Produced a generalized model that predicts % Silica (Impurity) present in the Iron Ore concentration with accuracy of:

93.80 %

Full Dataset

62.90 %

Resampled Dataset

Next Steps

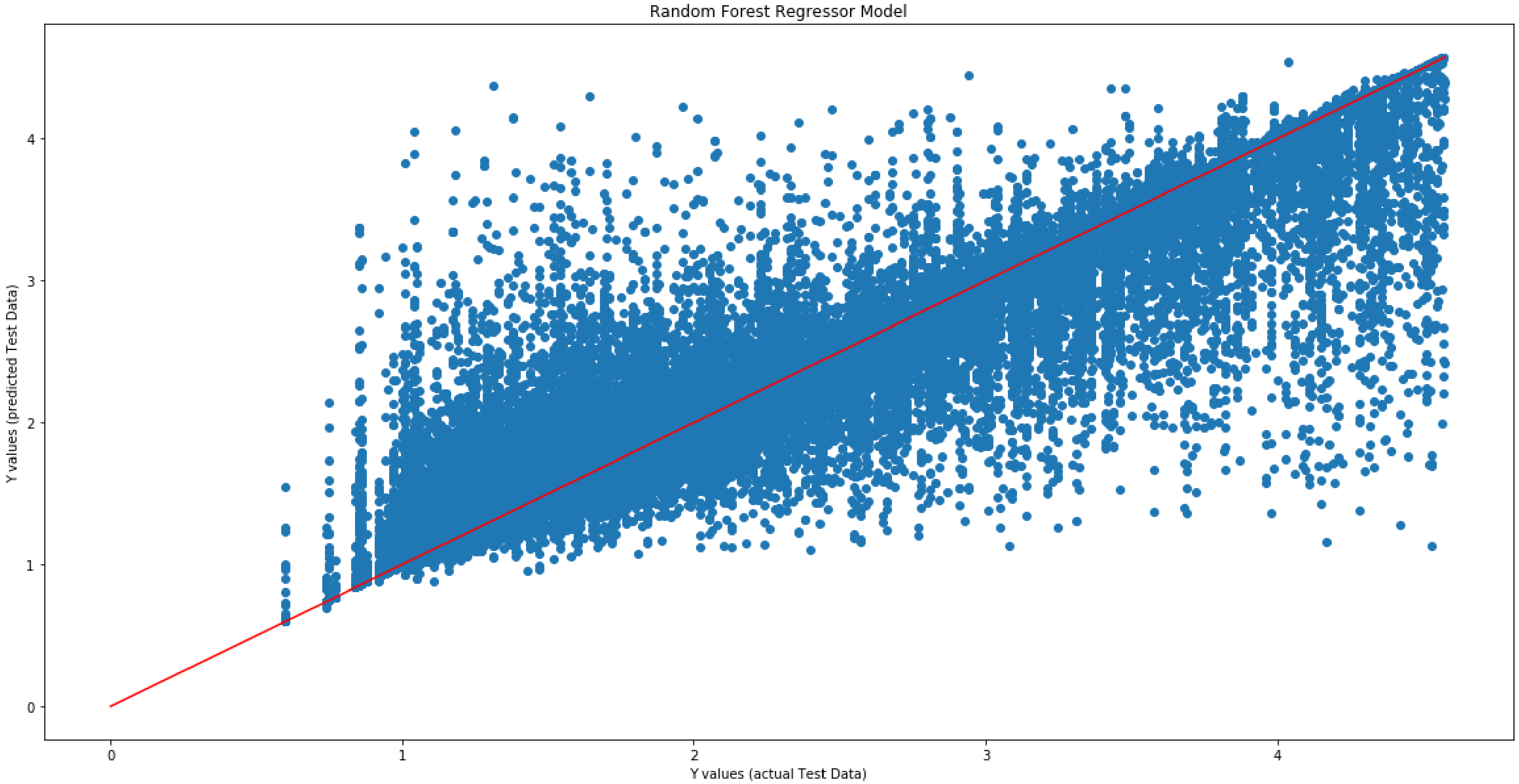
- Revisit and align on sampling period
- Consult with business to enhance the understanding of the data
- Perform in-depth hyper-parameter tuning on cloud computer (GCP, AWS, Azure)
- Experiment with different ML Packages / Libraries (Spark Mllib, TensorFlow)
- Use time series forecast to Predict Features (X) at a given time to Predict % Silica (Y)

Thank You

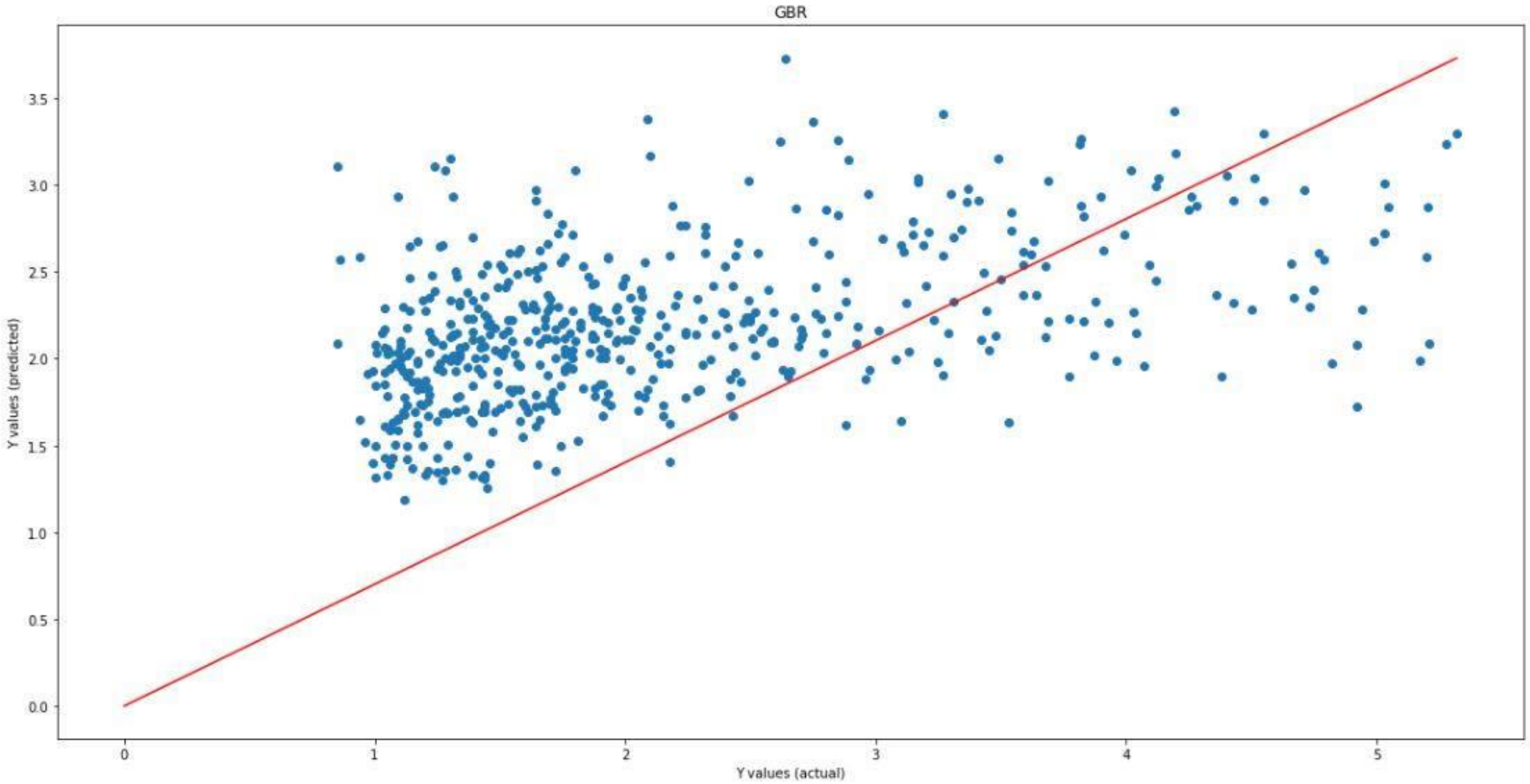
Scan me

https://github.com/nishp763/SCS-ML-3253---Final-Project.git

Appendix: Model # 1



Appendix: Model # 2



Appendix: Accuracy Function

```python
def evaluate(predictions, test_data):
    errors = abs(predictions - test_data)
    mape = 100 * np.mean(errors / test_data)
    accuracy = 100 - mape
    print('Model Performance')
    print('Average Error: {:0.4f}'.format(np.mean(errors)))
    print('Accuracy = {:0.2f}%'.format(accuracy))

    return accuracy
```

```python
base_accuracy = evaluate(y_pred, y_test)
```

```
Model Performance
Average Error: 0.1234
Accuracy = 93.64%
```