



DigiFab/ FONS Datathon 2021

Team 5: Balazs Striker, Katie Zeng, Marvin Alberts, Moe Thu Kha, Yixuan Wang



Data Preparation

All descriptors were used:

1. Drop all Columns with more than 60% missing Value
2. Replace all missing Values with the median of the column
3. Deal with Outliers:
 - a. Compute 5% and 95% quantile
 - b. Cap all data below or above to those value
4. Drop all Identifiers
5. Normalise Data



Method

1. Split Dataset into Train and Test set (80/20)
2. Test many algorithms on the Data => Random Forest won everywhere
3. Optimise Parameters of RandomForest
4. Calculate predictions on Test set



Adjustment in Task 4 & Bonus 2

- For both tasks the target showed high values
- Models work best on small values
- Thus the data was scaled down.
- Predictions were scaled up and submitted



Thank You



Results

	Mean Abs Error	Accuracy	R^2
Task 1	0.047		0.945
Task 2		0.842	
Task 3	0.242		0.862
Task 4	3.806		0.447
Bonus 1	0.019		0.469
Bonus 2	272.1		0.277

