NOV 7TH 2021

Heterogeneity Activity Recognizer Using Classification

Presented by: Ahad Almutairi & Shahad Alkaltham

Overview

- Classification probelm
- Timeline
- Model Selection
- Two Approaches
- Tools
- EDA
- Final Results
- Futrue Work

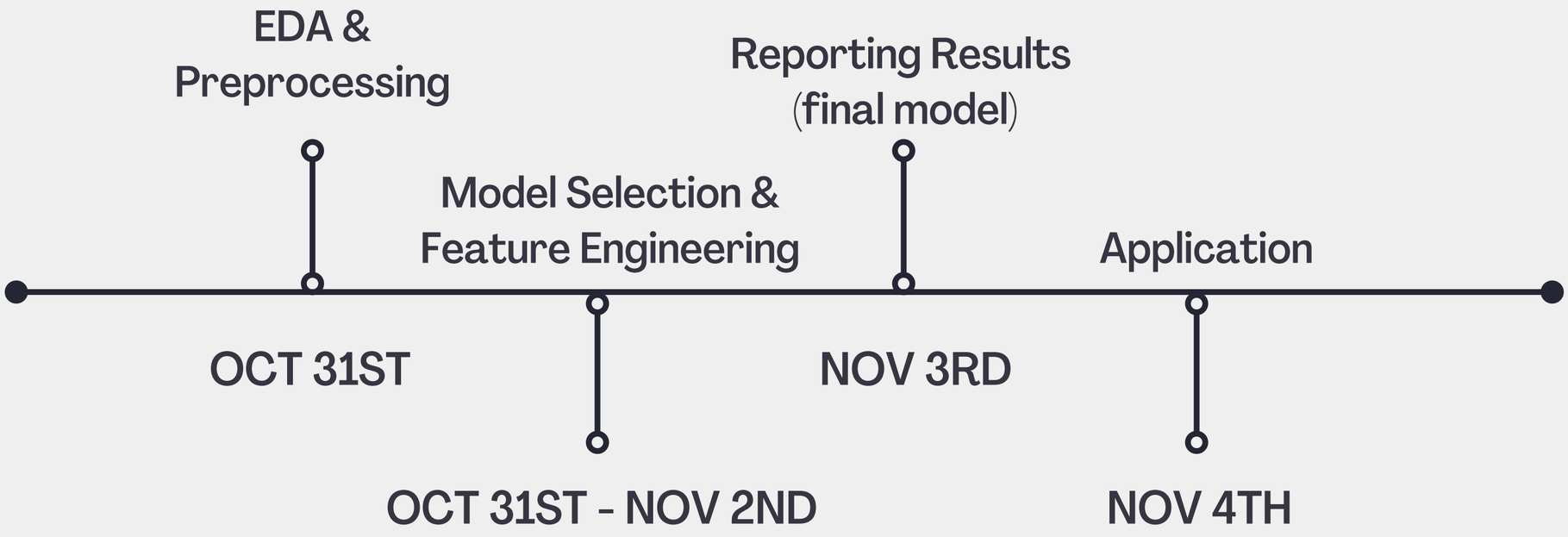
Heterogeneity Activity Recognizer

- What is the heterogeneity activity recognizer?
- Data contained 6 classes (UC Irvine Machine Learning Repository)
    - Stand
    - Sit
    - Bike
    - Walk
    - Stairs up
    - Stairs down

Stationary

Moving

Time Line


EDA &
Preprocessing
OCT 31ST
Model Selection &
Feature Engineering
OCT 31ST - NOV 2ND
Reporting Results
(final model)
NOV 3RD
Application
NOV 4TH

Tools

scikit learn

python™

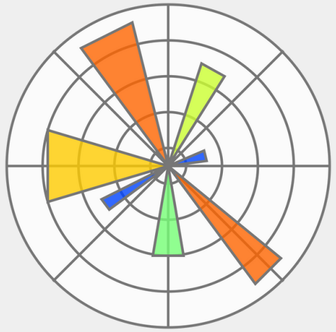
SQLite

NumPy

pandas

seaborn

Swift



MLXTEND

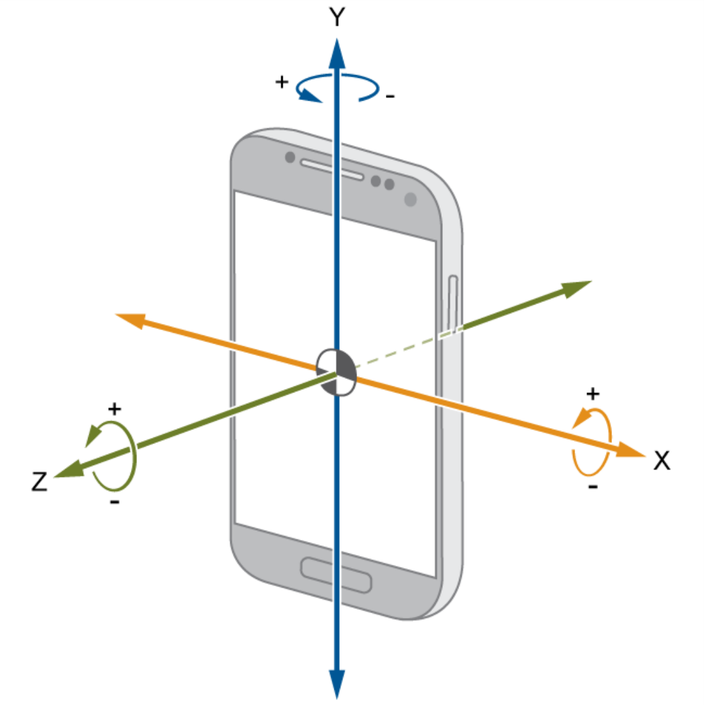
Exploratory Data Analysis

- **Data split into 4 csv files**

  Phones Accelerometer
  Phones gyroscope
  Watch Accelerometer
  Watch gyroscope

  MERGED

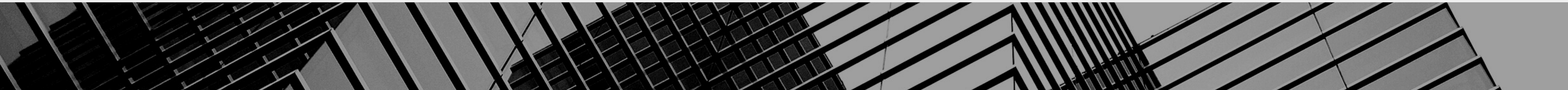
- **Contained 33M rows and 10 columns**



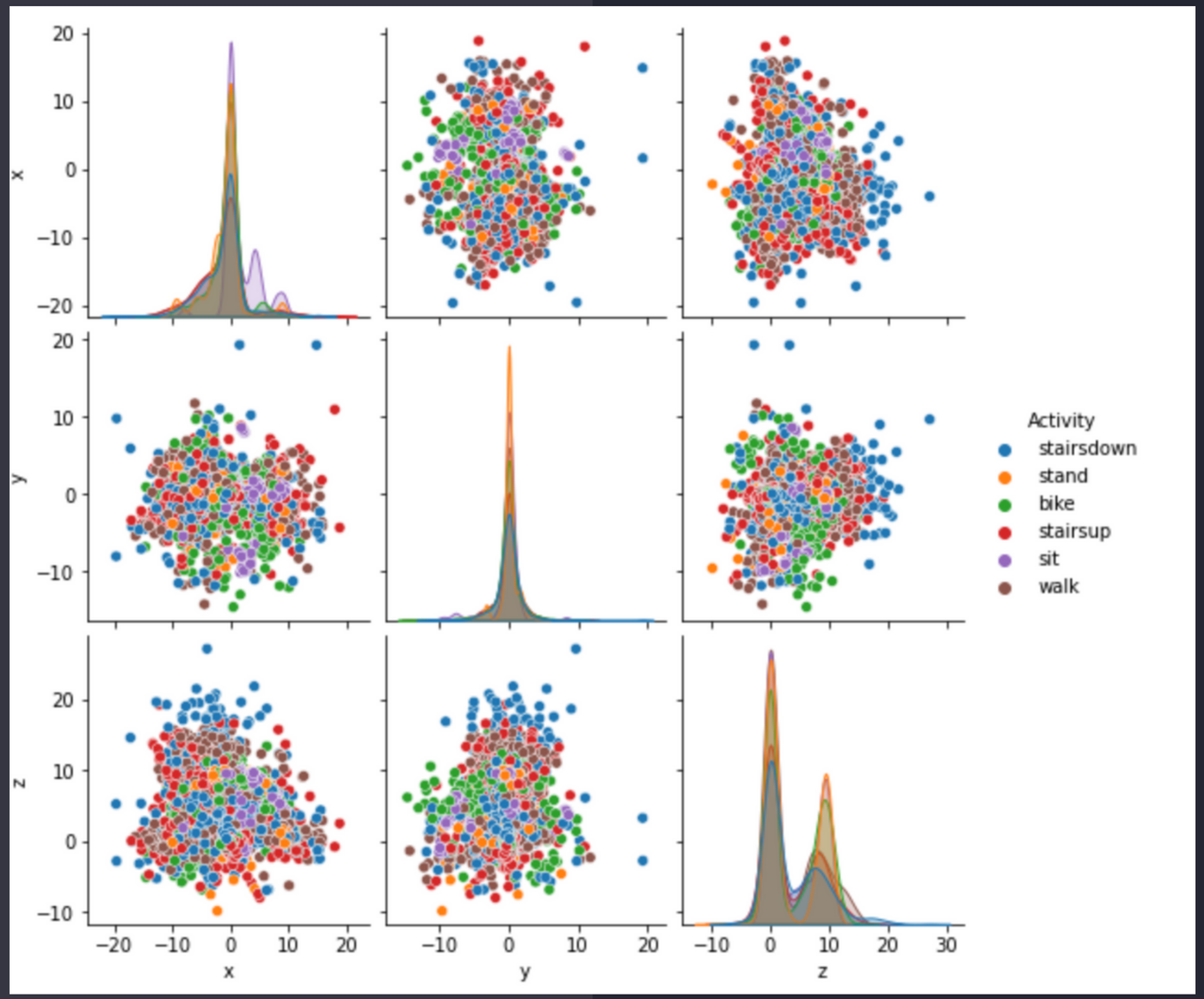
Exploratory Data Analysis

- Creating new columns (sensor)
- Handling nulls & duplicates
- Dummies (model, sensor)
- Dropping unneeded columns (index, creation time)
- Fixing timestamp issue
- Merging classes
- Taking a sample of 500,000

- Contains 500K rows and 19 columns



Issue with 1st Approach



Two Approaches

Multi-Class

- Classes:

Stand	0
Sit	1
Walk	2
Bike	3
Stairs up	4
Stairs down	5

- Classes are **balanced**

- Baseline (Logistic Regression):
  Accuracy Training: **0.2481**
  Accuracy Validation: **0.2512**
  F1: **0.2512**
  Recall: **0.2512**
  Precision: **0.2512**

Binary-Class

- Classes:

Stand	0
Sit	0
Walk	1
Bike	1
Stairs up	1
Stairs down	1

- Classes are **imbalanced**

- Baseline (Logistic Regression):
  Accuracy Training: **0.5936**
  Accuracy Validation: **0.5391**
  F1: **0.5579**
  Recall: **0.4399**
  Precision: **0.7625**

Model Selection

- Solving the imbalanced classes with oversampling
- Building 9 models
- Polynomial features with models
- ROC with different parameter
- Decision boundary for models
- Model comparison under ROC curve
- Fitting into stacking and voting
- Feature engineering

Model Selection

- **Imbalanced Data**

  Data before oversampling
  - Class 0: **137,482**
  - Class 1: **269,555**

  ×2

  Data after oversampling
  - Class 0: **274,964**
  - Class 1: **269,555**

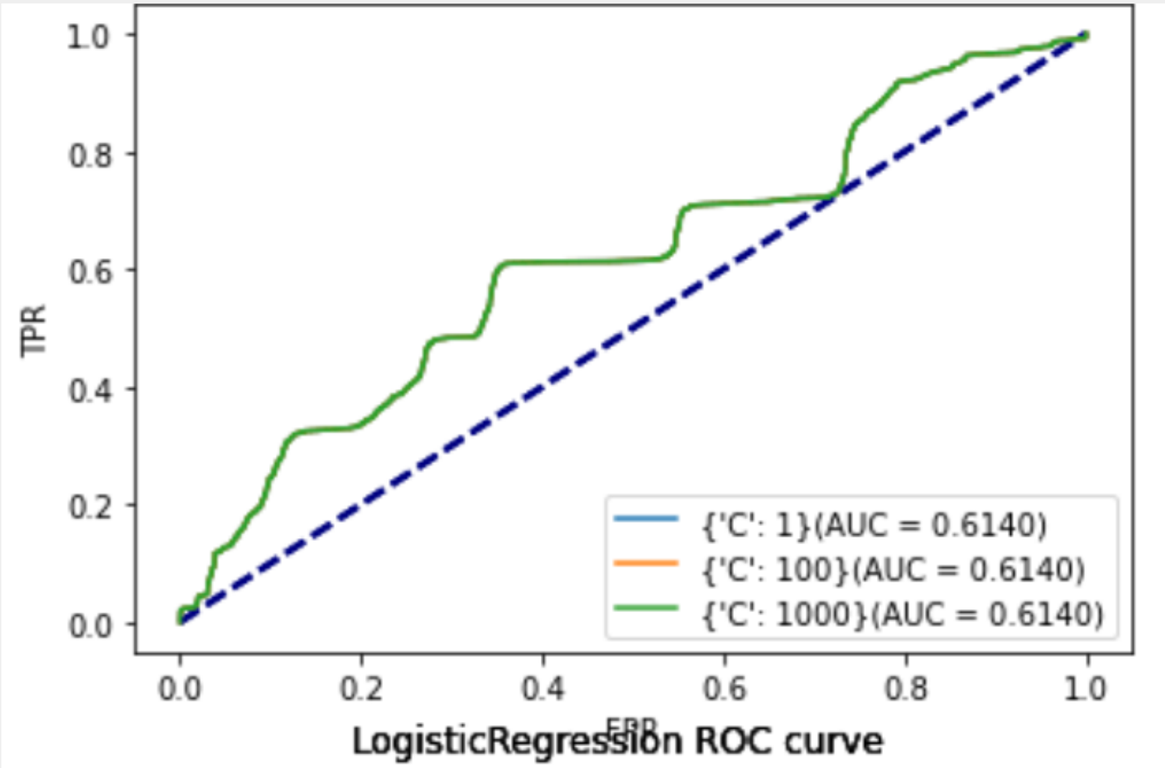
- **7 Models**

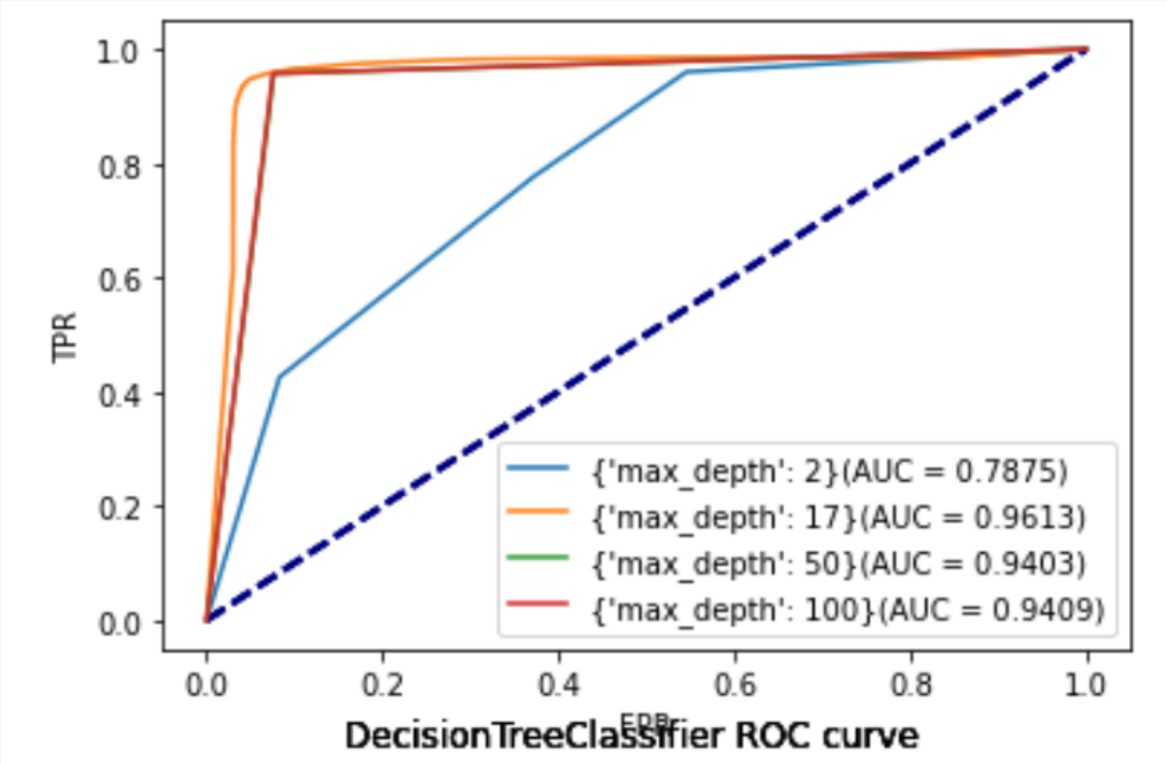
Model	Training Accuracy	Validation Accuracy	F1	Recall	Precision
Logistic Regression	0.5936	0.5391	0.5579	0.4399	0.7625
Decision Tree	0.9750	0.9492	0.9611	0.9487	0.9738
Random Forest	0.9991	0.9630	0.9720	0.9701	0.9739
Extra Trees	0.9991	0.9635	0.9725	0.9742	0.9709
XGB	0.9515	0.9457	0.9576	0.9365	0.9796
Gaussian	0.5733	0.5149	0.5240	0.4039	0.7460
Bernoulli	0.5740	0.6005	0.6846	0.6557	0.7161

- KNN Classifier

Model Selection

- ROC with different parameter





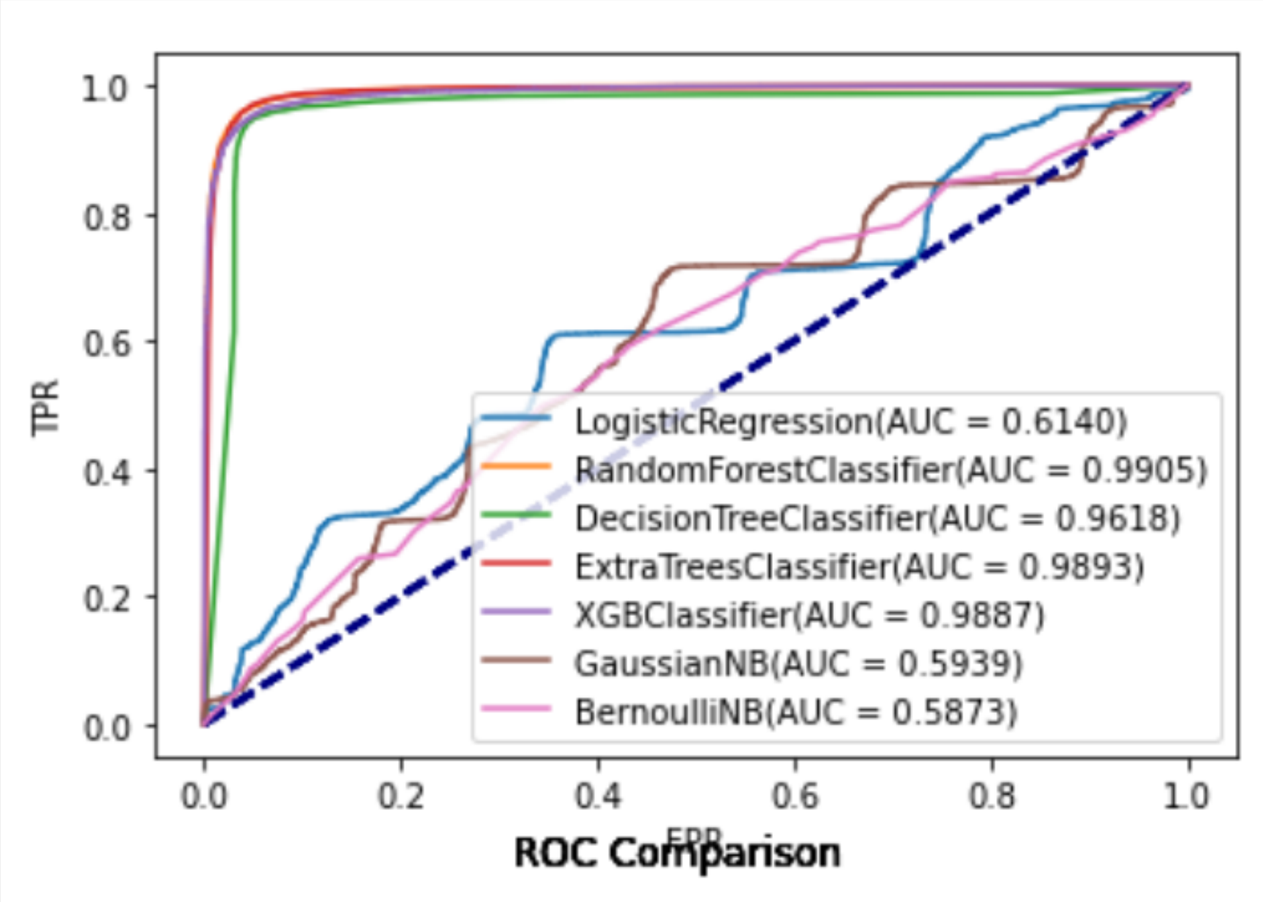
Model Selection

- Polynomial features with models

Model	Logistic Regresion	Decision Tree	Random Forest	Extra Trees	XGB
Training before	Tr:0.5936	Tr:0.9750	Tr:0.9991	Tr:0.9991	Tr:0.9515
Validation before	Val:0.5391	Val:0.9492	Val:0.9630	Val:0.9635	Val:0.9457
Difference	5.45%	2.58%	3.6%	3.56%	0.58%
Training after	Tr:0.6557	Tr:0.9668	Tr:0.9991	Tr:0.9991	-
Validation after	Val:0.6252	Val:0.9455	Val:0.9631	Val:0.9633	-
Difference	3.05%	2.13%	3.6%	3.58%	no overfit

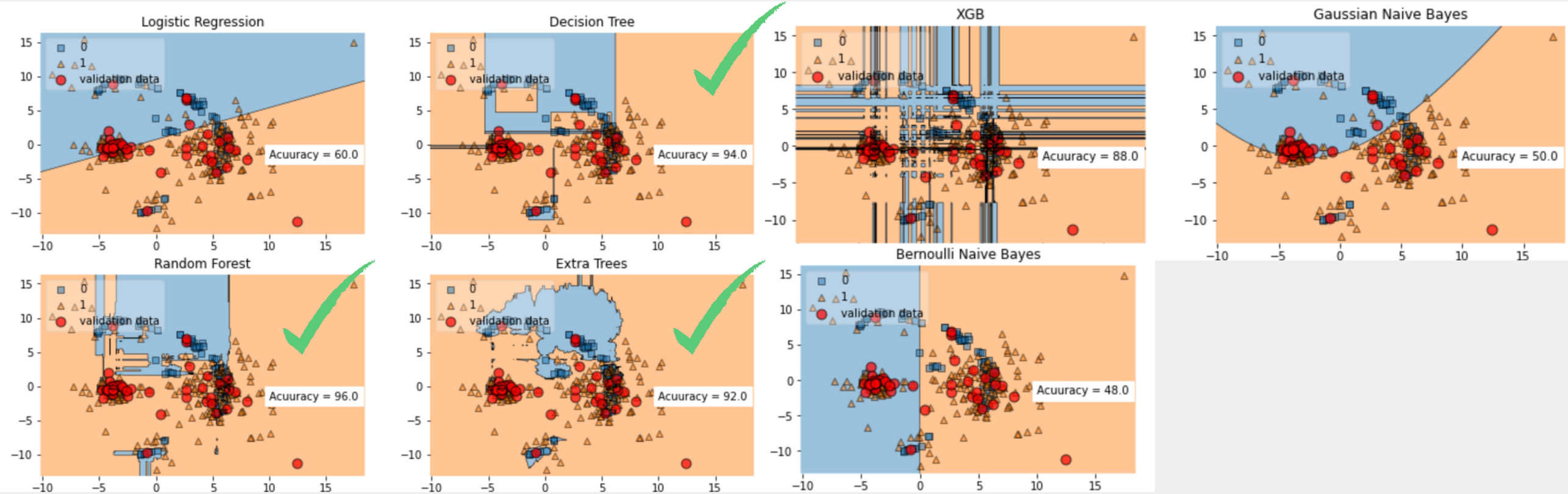
Model Selection

- ROC Curve



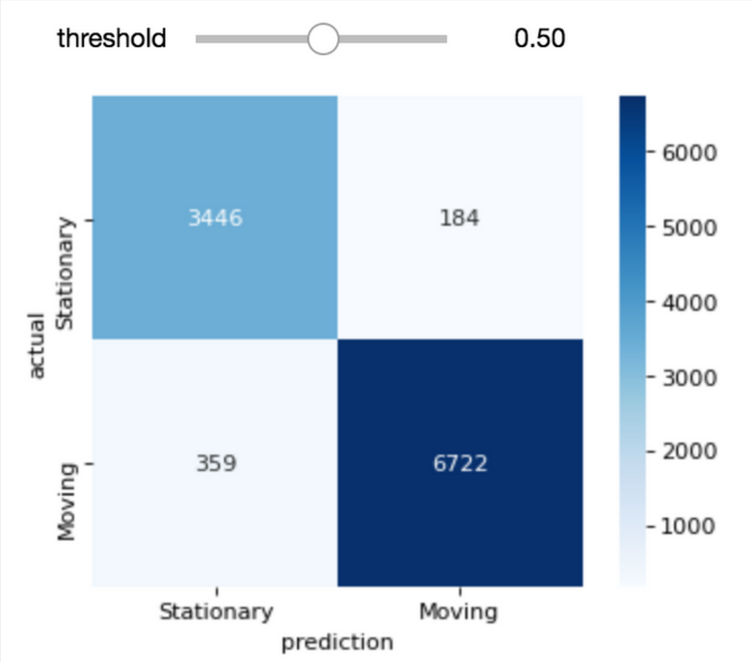
Model Selection

- Decision Boundaries

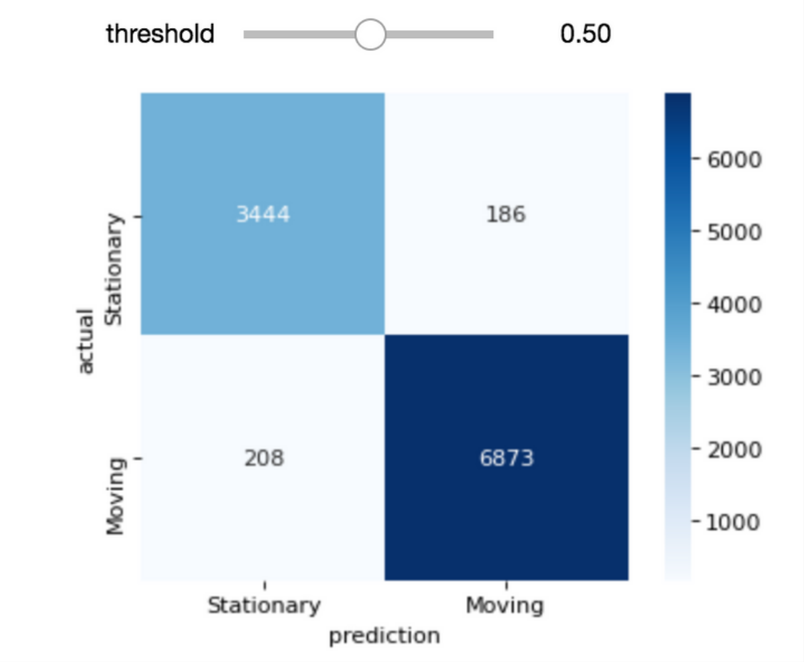


Model Selection

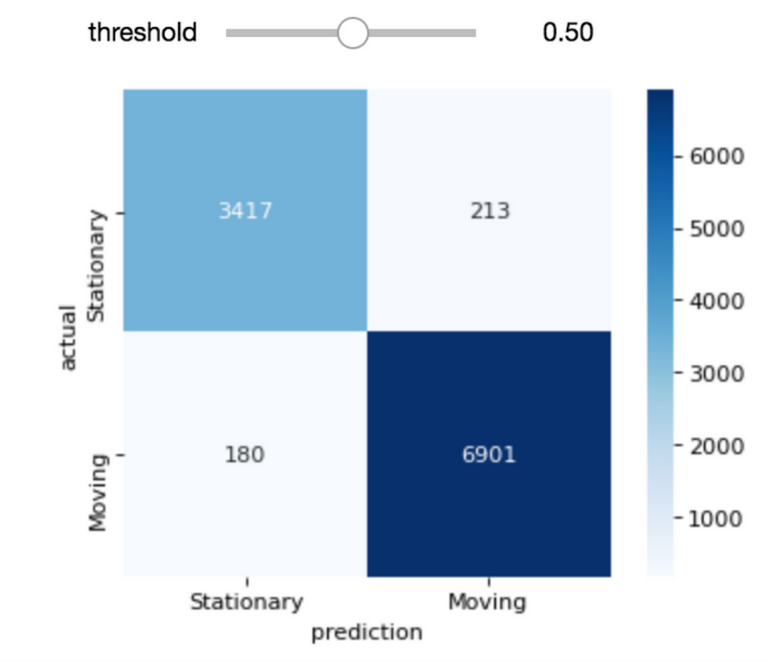
- Confusion Matrix



Decision Tree



Random Forest



Extra Trees

Model Selection

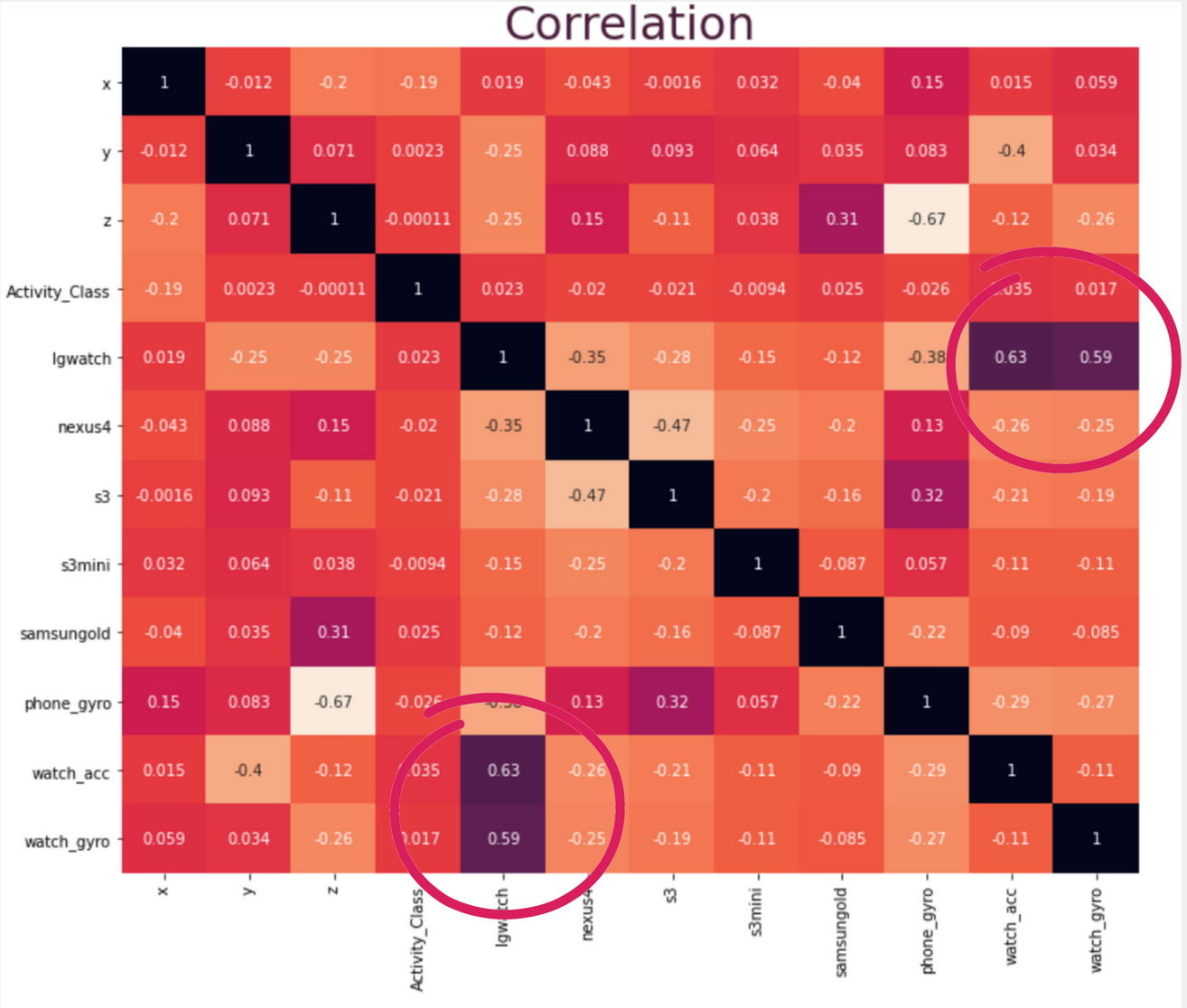
- Fitting into Stacking & Voting

Model	Training Accuracy	Validation Accuracy	F1	Recall	Precision
Voting	0.9991	0.9615	0.9709	0.9691	0.9726
Stacking	0.9991	0.9632	0.9723	0.9749	0.9697

Model Selection

- Feature Engineering on Stacking Model

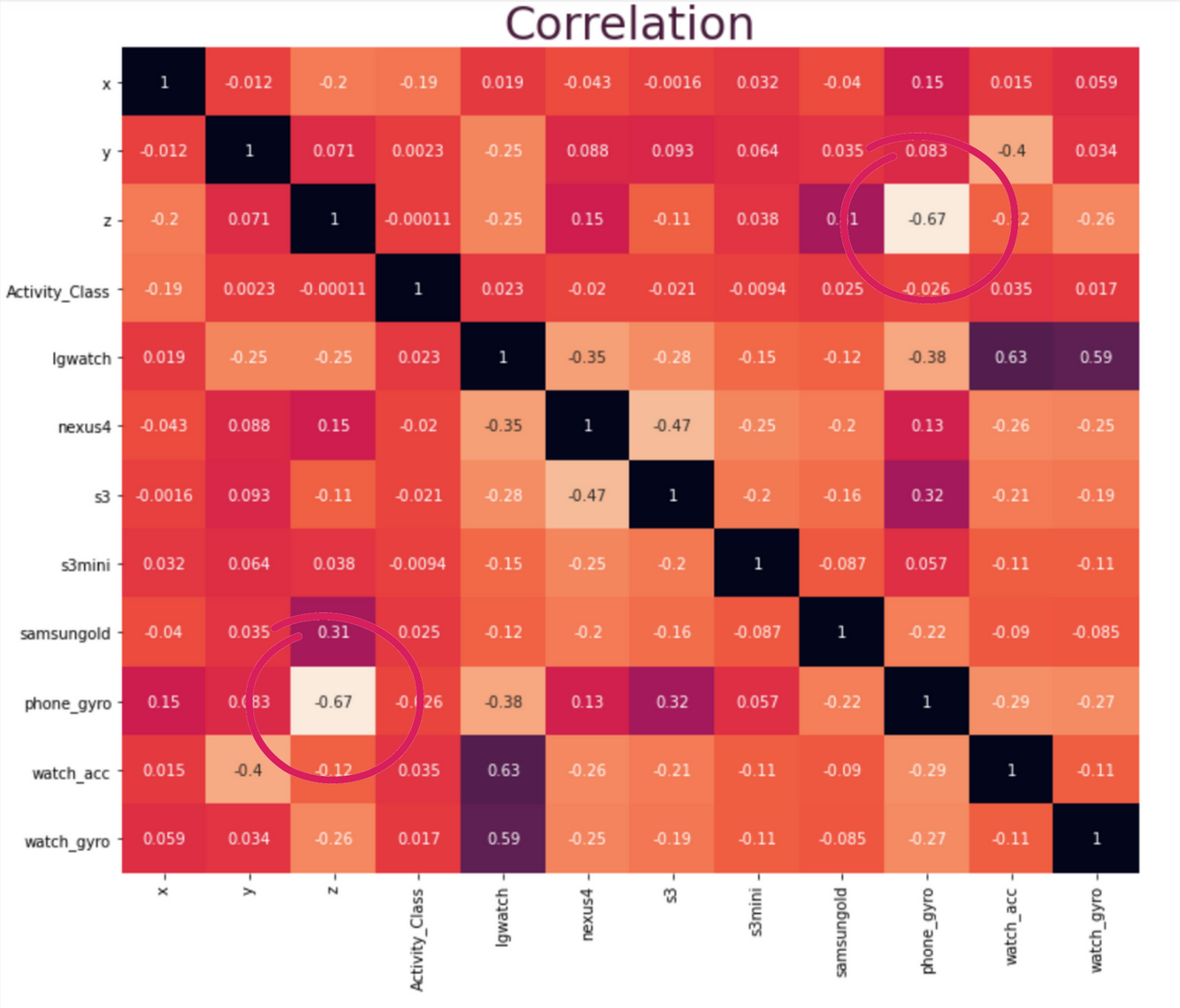
ORIGINAL	Train: 0.9991	Validation: 0.9632	STATUS
Attemp 1	Train: 0.9991	Validation: 0.9634	Approved
Attemp 2	Train: 0.9991	Validation: 0.9630	Failed
Attemp 3	Train: 0.9991	Validation: 0.9582	Failed
Attemp 4	Train: 0.9991	Validation: 0.9606	Failed
Attemp 5	Train: 0.9991	Validation: 0.8214	Failed



Model Selection

- Feature Engineering on Stacking Model

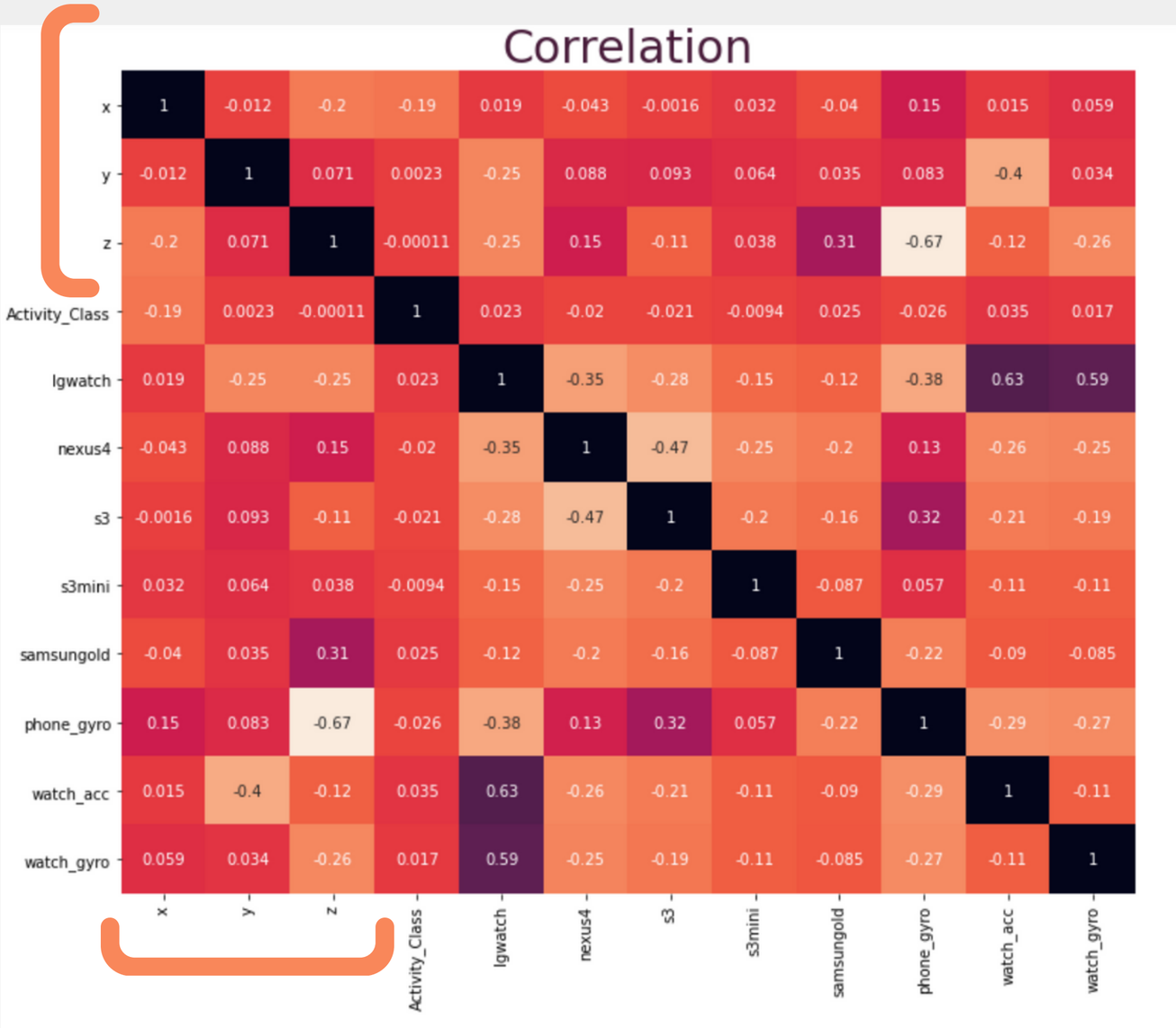
ORIGINAL	Train: 0.9991	Validation: 0.9632	STATUS
Attemp 1	Train: 0.9991	Validation: 0.9634	Approved
Attemp 2	Train: 0.9991	Validation: 0.9630	Failed
Attemp 3	Train: 0.9991	Validation: 0.9582	Failed
Attemp 4	Train: 0.9991	Validation: 0.9606	Failed
Attemp 5	Train: 0.9991	Validation: 0.8214	Failed



Model Selection

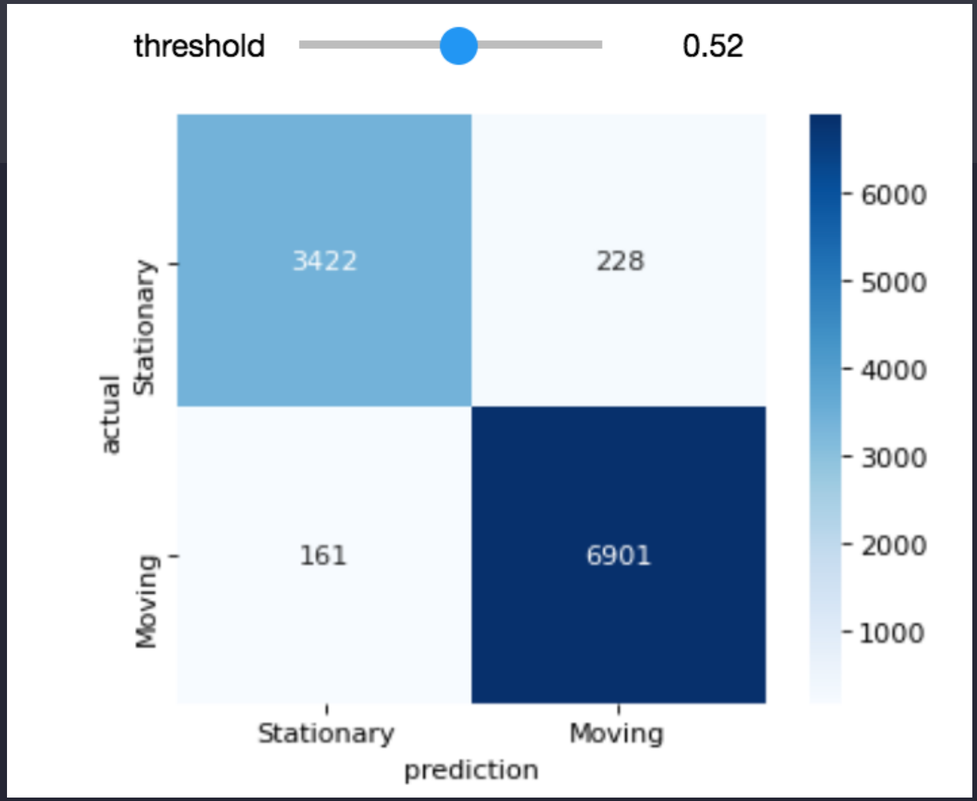
- Feature Engineering on Stacking Model

ORIGINAL	Train: 0.9991	Validation: 0.9632	STATUS
Attemp 1	Train: 0.9991	Validation: 0.9634	Approved
Attemp 2	Train: 0.9991	Validation: 0.9630	Failed
Attemp 3	Train: 0.9991	Validation: 0.9582	Failed
Attemp 4	Train: 0.9991	Validation: 0.9606	Failed
Attemp 5	Train: 0.9991	Validation: 0.8214	Failed



Reporting Final Results

Model	Whole Data Accuracy	Tesing Accuracy	F1	Recall	Precision
Stacking	0.9991	0.9636	0.9726	0.9772	0.9680



Future Work

- Support Apple devices
- Sending notifications to encourage exercises

Thank you for listening!

Any Questions?