



Team: Outliers

Project participant names – email ids

| First Name | Last Name | Email  |
|------------|-----------|--|
| Majid      | Sabbagh   | sabbagh.m@husky.neu.edu  |
| Samkeet    | Shah      | <a href="mailto:shah.sam@husky.neu.edu">shah.sam@husky.neu.edu</a>       |
| Shantanu   | Kawlekar  | <a href="mailto:kawlekar.s@husky.neu.edu">kawlekar.s@husky.neu.edu</a>   |
| Yixing     | Zhang     | <a href="mailto:zhang.yixin@husky.neu.edu">zhang.yixin@husky.neu.edu</a> |

#### Description

We have developed an Anomaly detection system using machine learning, to detect anomalies in power traces of computers, servers and later extended to household electric consumption.

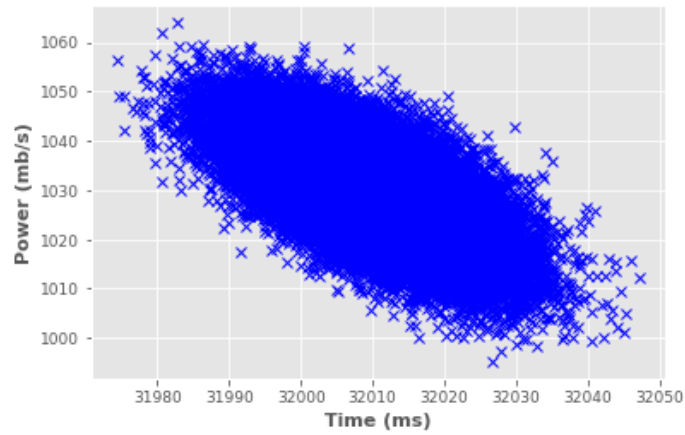
The motivation behind this system was, to detect anomalous power surges by a device which is integrated with Google (Firebase) Cloud messaging to notify users on their e-mail addresses and twitter and/or slack channels, of this anomalous power use.

The project was initially developed to notify system admins of organizations of anomalous, power traces usage which can be used by hackers to determine the general behavior and possibly learn the secret encryption keys of the machine.

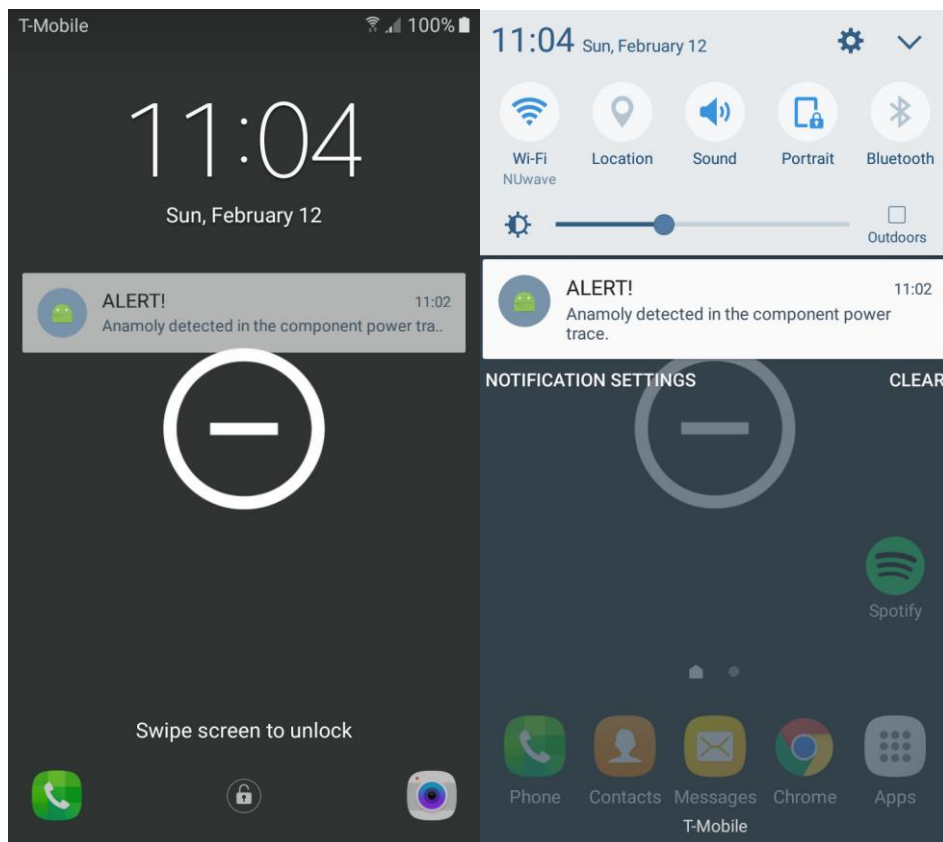
Hence we have use a masked and unmasked of power trace, where a masked version is undetectable by our machine learning model of any random behavior. Whereas the anomalous trace/plot clearly indicates outliers in the trace, which immediately alerts the user/admin.

### Snapshots of project

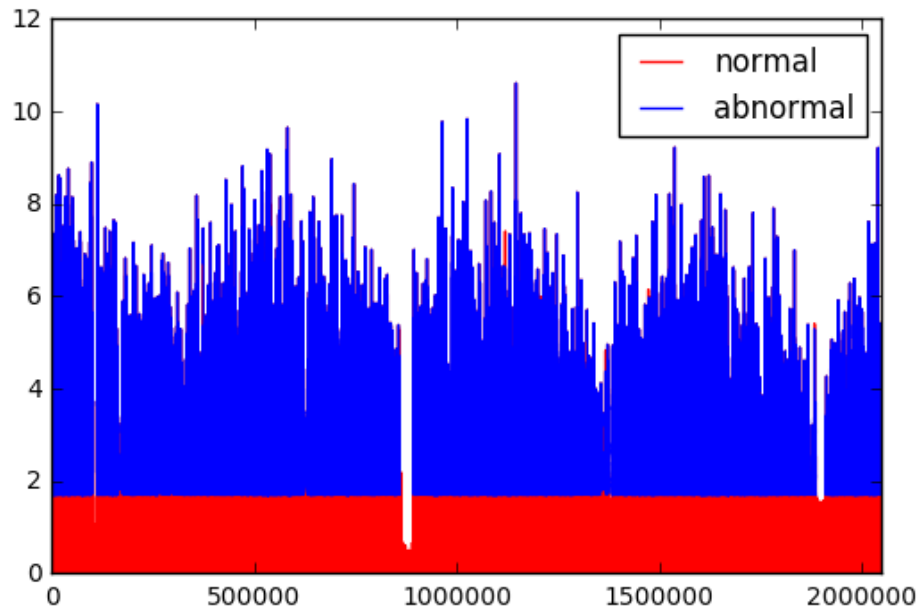
```
C:\Users\shahs\Anaconda3\lib\site-packages\matplotlib\font_und. Falling back to DejaVu Sans  
(prop.get_family(), self.defaultFamily[fonttext]))
```



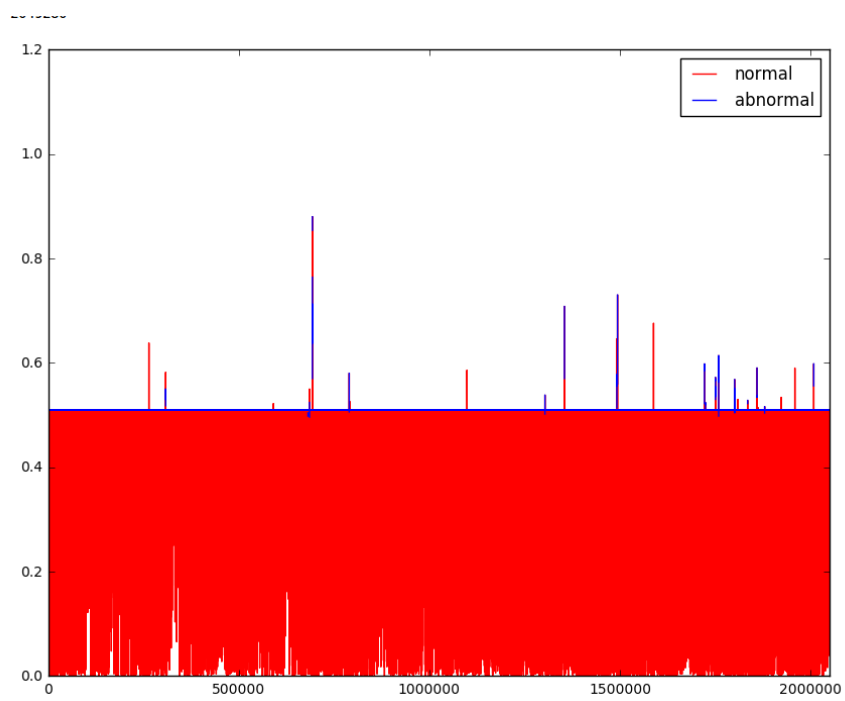
### Gaussian Invariate on Unmasked Data



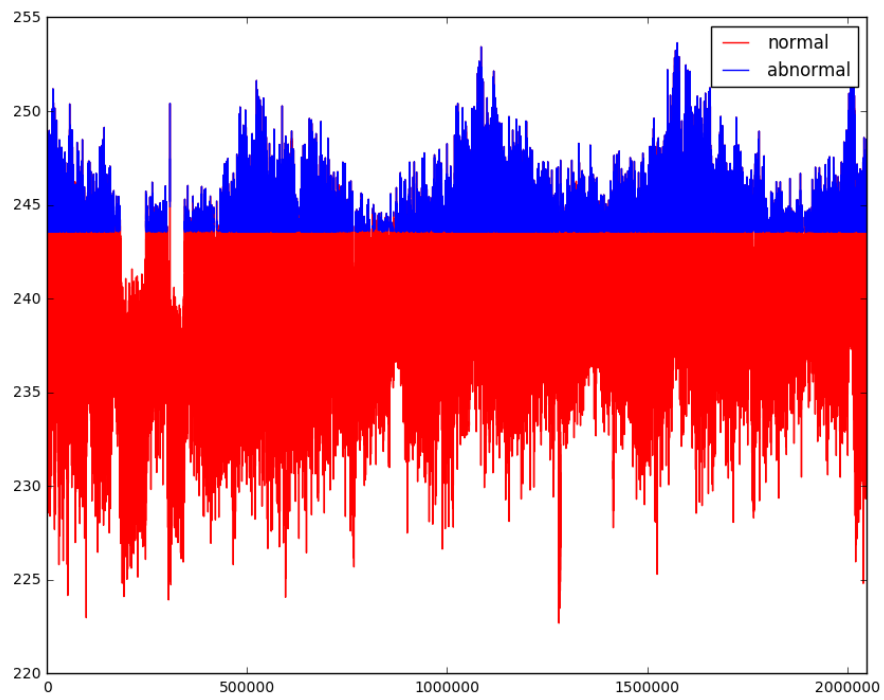
### Android Notification



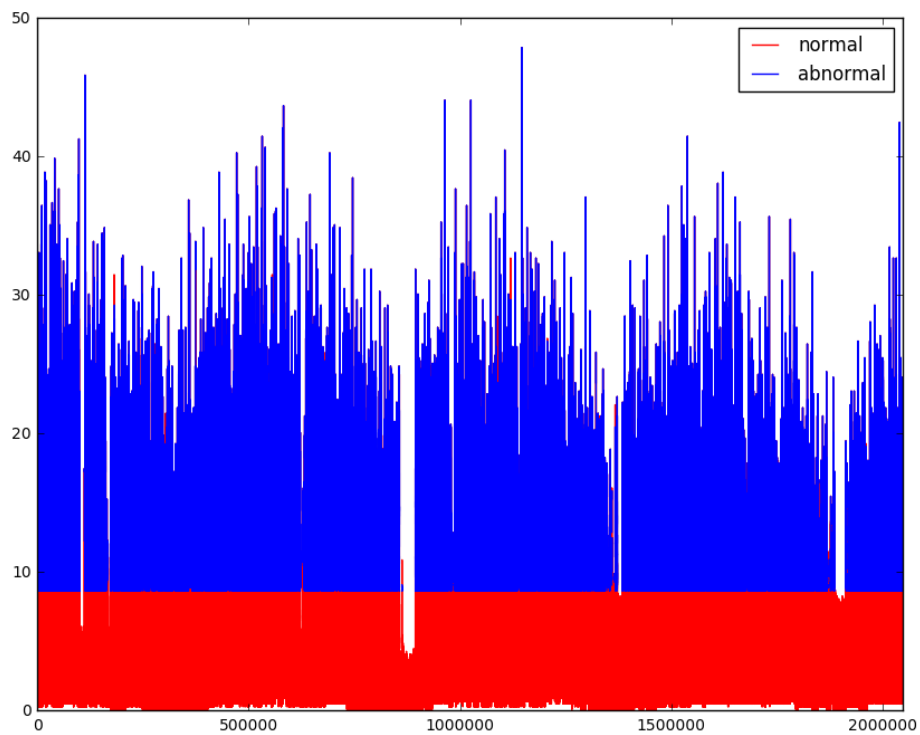
Plot of Global Active Power for a household



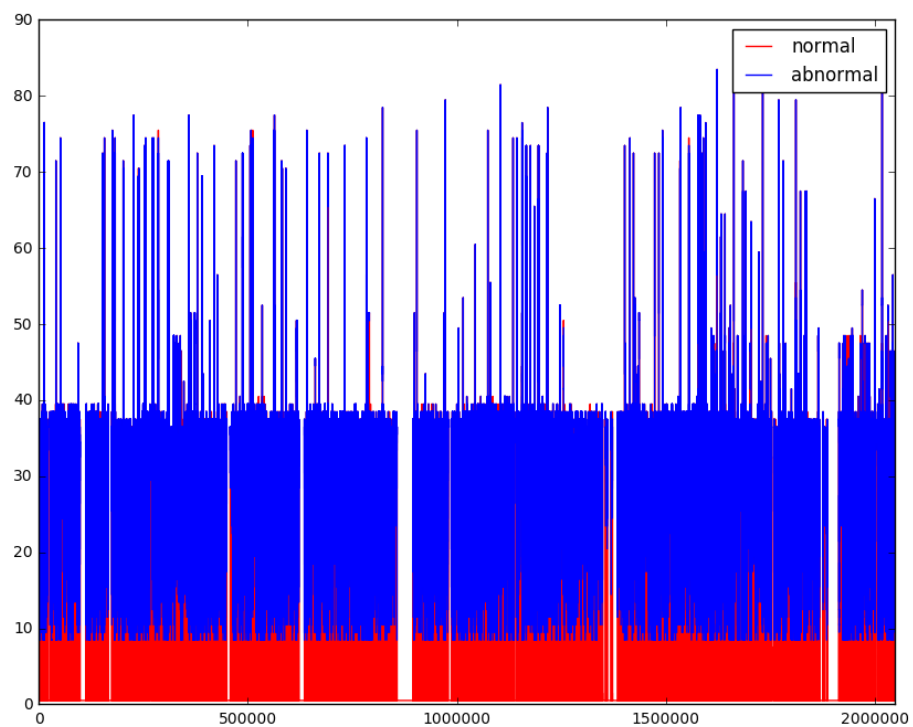
Plot of Global Reactive Power vs Time



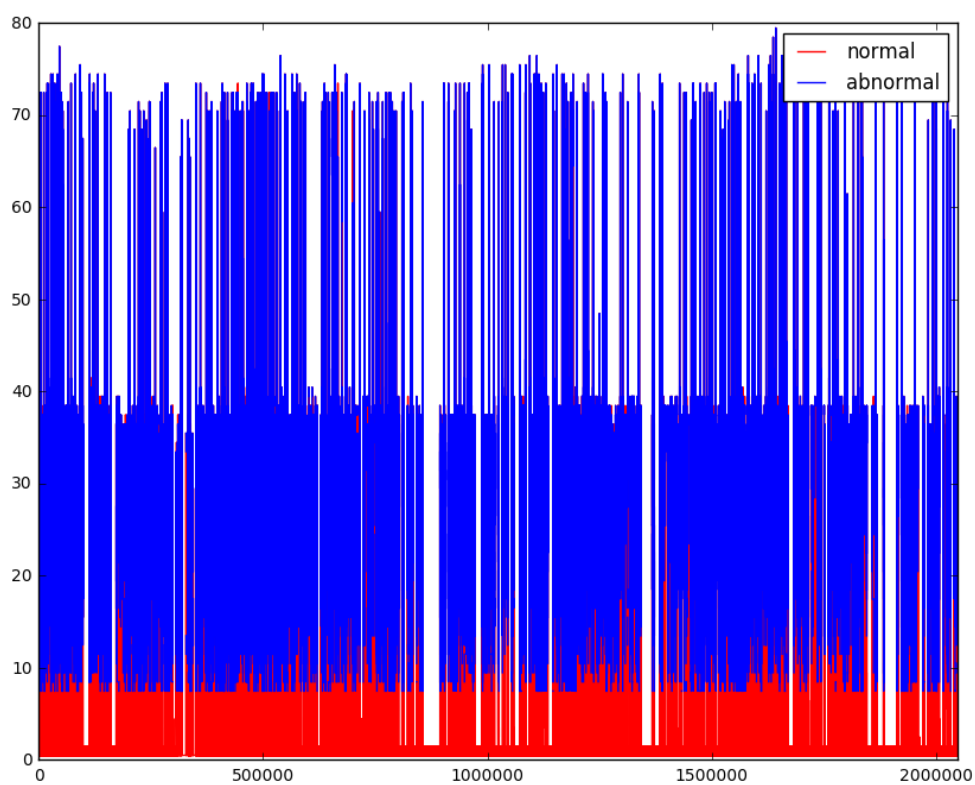
Voltage Deviation vs Time



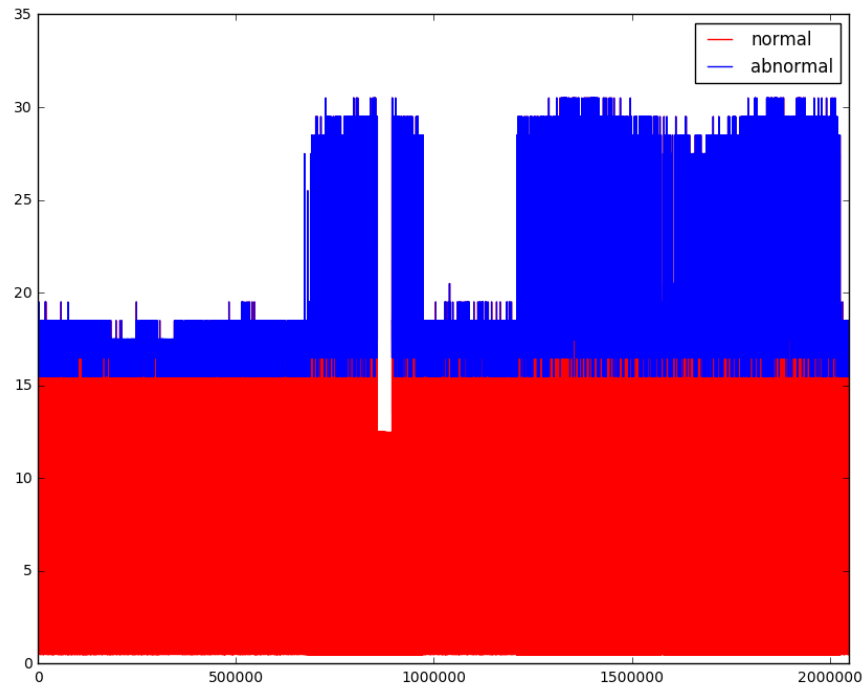
Global Intensity plot



Plot of Sub Meter 2



Plot of Sub Meter 3



Plot of Sub Meter 3 vs Time

#### Challenges faced during project

1. Handling of large datasets (100000 entries \* 3125 columns)
2. Cleaning the datasets to remove NULL or empty fields
3. Converting the input datasets into appropriate formats for applying the machine learning model
4. Integrating Google Firebase platform to notify users on a custom Android app.

Any additional information: