**IMPLEMENTATION:**

**MODULES:**

* User
* Admin
* Data Preprocessing
* Machine Learning Results

**MODULES DESCRIPTION:**

**User:**

The User can register the first. While registering he required a valid user email and mobile for further communications. Once the user register then admin can activate the user. Once admin activated the user then user can login into our system. User can upload the dataset based on our dataset column matched. For algorithm execution data must be in float format. Here we took hacking breach dataset. User can also add the new data for existing dataset based on our Django application. User can click the Classification in the web page so that the data calculated Accuracy, Precision, Recall and F1-Score based on the algorithms.

**Admin:**

Admin can login with his login details. Admin can activate the registered users. Once he activate then only the user can login into our system. Admin can view the overall data in the browser. Admin can click the Results in the web page so calculated ME, MSE, RMSE more based on the algorithms is displayed. All algorithms execution complete then admin can see the overall accuracy in web page.

**Data Pre-processing:**

Because we observed, as mentioned above, some days have multiple hacking breach incidents, one may suggest to treat such multiple incidents as a single “combined” incident (i.e., adding their number of breached records together) However, this method is not sound because the multiple incidents may happen to different victims that have different cyber systems. Given that the time resolution of the dataset is a day, multiple incidents that are reported on the same data may be reported at different points in time of the same day (e.g., 8pm vs. 10pm). As such, we propose generating small random time intervals to separate the incidents corresponding to the same day. Specifically, we randomly order the incidents corresponding to the same day, and then insert a small and random time interval in between two consecutive incidents (for the first interval, the starting point is midnight), while assuring that these incidents correspond to the same day (e.g., the two incidents on a two-incident day may be assigned at 8am and 1pm).

**Machine learning Results**:

Based on the split criterion, the cleansed data is split into 60% training and 40% test, then the dataset is subjected to machine learning regressor such as EDA (exploratory data analysis), ARMA Model. The ME(mean error),MSE(mean square error),RMSE(Root mean square error) of the regressor was calculated and displayed in my results. The regressor which bags up the highest mean square error could be determined as the best classifier.