**Project Title:** Comparative predictive analysis of social media comments for identifying depression levels using supervised machine learning algorithms.

**Project description:** Our aim is to investigate how Twitter user’s posts can help classify users according to mental health levels. We propose a system that uses Twitter data as a source of data and screening tool to classify the user using artificial intelligence according to the User comments on Twitter. We are creating a model that classify the user comments using four different classifiers: Logistic Regression, K-Nearest Neighbors, Random Forest and Support Vector Machine (SVM)

for predictive analysis, using supervised machine learning. The classifiers results are compared to provide the most accurate figures for prediction of depression levels of a group of datasets.

**Task Worked as a part of Team:**

|  |  |
| --- | --- |
| **Tasks** | **Mujtaba** |
| Collection of twitter datasets |  |
| Creating Bucket on S3 |  |
| Creating Cluster |  |
| Pre-processing of collected data |  |
| Sentiment Analysis |  |
| Implementing Logistic Regression |  |
| Implementing K-Nearest Neighbors |  |
| Implementing Random Forest |  |
| Implementing Support Vector Machine |  |
| Result Comparisons |  |
| Connecting Pycharm remotely to EC2 cluster |  |
| Installing Dependencies |  |
| Configuring EC2 cluster to connect to S3 |  |