# Major Project

- **Project Name:**

Data Science March Major Project

- **Project Description:**

**Problem statement:** Create a classification model to predict the sentiment either (1 or 0) based on Disaster tweets

**Context:** This dataset consists of a nearly 7000 disaster tweets (input text) and target (1 or 0) etc. for learning how to train Machine for sentiment analysis.

**Dataset:**

https://drive.google.com/file/d/11NIJ29_-X8GbuvSPp7TEg2ljmGikCTn0/view?usp=share_link

**Details of features:**

The columns are described as follows:

1. tweets: Product review
2. target: 1(Positive) or 0 (Negative)

Email: support@immensphere.com
Phone: 9019725366
Ground floor, 1537, 5th Main Rd, Rajiv Gandhi Nagar, Sector 7,
HSR Layout, Bengaluru, Karnataka 560102

**Steps to consider:**

1. Read the dataset
2. Remove handle null values (if any).
3. Preprocess the disaster tweets data based on the following parameter:
   a)Tokenizing words
   b)Convert words to lower case
   c)Removing Punctuations
   d)Removing Stop words
   e)Stemming or lemmatizing the words
4. Transform the words into vectors using
   a)Count Vectorizer
   OR
   b)TF-IDF Vectorizer
5. Select x(independent feature) as tweets after preprocessing and target as y(dependent feature).
6. Split data into training and test data.
7. Apply the following models on the training dataset and generate the predicted value for the test dataset
   a)Multinomial Naïve Bayes Classification
   b)Logistic Regression
   c)KNN Classification
8. Predict the target for test data
9. Compute Confusion matrix and classification report for each of these models
10. Report the model with the best accuracy.

Email: support@immensphere.com
Phone: 9019725366
Ground floor, 1537, 5th Main Rd, Rajiv Gandhi Nagar, Sector 7,
HSR Layout, Bengaluru, Karnataka 560102