**Project Scope: GLOBAL STOCK MARKET(FINANCIAL MARKETS) ANALYTICS**

**Objectives:**

1. Comparative Study of Global Markets – Returns, Volatility, Correlations.
2. Predicting **Nifty 50 open direction** based on global indicators.(you can decide the target market-I have considered India for which Nifty50 is the benchmark index)
3. Sentiment Analysis of Nifty 50 using X(twitter) data

**Phase 1 is all about creating master data:**

1. List down global market indices:
   1. Nifty 50
   2. Dow Jones Index
   3. Nasdaq
   4. Hang Seng
   5. Nikkei 225
   6. Dax
2. Check symbols for each of the index above which will help you to fetch data
3. Check R/Python syntax to download data for each index for Six years.(from 1st Jan 2018 till 31st December 2023-**OHLC Data**)
4. Add USA VIX data for Six years (Volatility Index- Fear Factor)
5. Calculate daily returns based on close prices of each index: ( Yt-Yt-1)/(Yt-1) \* 100
6. Check if daily returns follow Normal distribution
7. Rename variables in each file (Example: Nifty50\_Open,Nifty50\_Adjclose)
8. Merge all the above files with outer join (note that holidays are different globally)
9. Impute missing data using LOCF method
10. Create indicator variable for “Quarter”
11. Create indicator variable for “Month”
12. Create indicator variable for “Year”
13. Finalize Master Data which will be used for Phase 2

**Phase 2 is Exploratory Data Analysis**

**Questions:**

1. **Which index has given consistently good returns?**
2. **Which index was highly volatile?**
3. **How are global markets correlated during 6 years period and is the correlation structure similar in the recent year-2023?**
4. **Assuming primary target variable as “Nifty Opening Price Direction”, what are preliminary insights?**
5. Global Indices 5 years Performance Analytics:(

For each index (a to f above)

1. Box-Whisker Plot of daily returns by “YEAR”
2. Table of daily returns by “YEAR”- (n, mean and std. deviation)
3. Bar Plot of median daily return by “YEAR”
4. Heat Map by “YEAR” and “QUARTER” showing median/mean returns
5. Global Indices-Correlation Analysis
6. Correlation Matrix of 6 years daily returns (6 by 6 matrix)
7. Correlation Matrix of one year (2023) daily returns (6 by 6 matrix)
8. Pre-Post Covid Performance Analytics-time required for a market to come back

To pre-covid level

1. Nifty 50 Daily Movement- Pre Modeling Work
2. Define Nifty\_Open\_Dir=1 if Nifty 50 Open at t > Close at t-1

=0 other wise

1. Table of % of Nifty\_Open\_Dir=1 by year
2. Visualize/summarize global indices including VIX for 2 categories of Nifty\_Open\_Dir

Example: Box -whisker plot of lag ( DJI Returns) for 2 categories of Nifty\_Open\_Dir

\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Phase 3 is Predictive Modeling**

1. Define Nifty\_Dir\_Open =1/0 based on direction (dependent variable)

(already derived in phase 2)

1. Create data partition into train and test data sets (80/20)
2. Run Binary Logistic Regression with Nifty Direction as dependent variable and previous day global market returns (and VIX) as independent variables. You may add more features such as previous day High/Low ratio for Nifty 50 and/or DJI
3. Check multicollinearity and resolve if present
4. Check which variables are significant (revise the model if needed)
5. Obtain ROC curve and AUC for train data

7. Obtain threshold to balance sensitivity and specificity

Go to step 8 only if you are satisfied with model on train data

8.Obtain ROC curve and AUC for test data (compare with step 6)

9.Use above threshold to obtain sensitivity and specificity for test data

(compare with step 7)

10. Finalize the model

**Phase 4 is about using ML methods and compare with Binary Logistic Regression**

1. Create data partition into train and test data sets ( 80/20)

(already done for Logistic Regression)

1. Apply Naïve Bayes Method on train data with Nifty Direction as dependent variable and previous day global market returns (and VIX) as independent variables.
2. Obtain ROC curve and AUC for train data
3. Obtain Confusion Matrix for train data
4. Obtain ROC curve and AUC for test data
5. Obtain Confusion Matrix for test data
6. Repeat steps 2 to 6 for Decision Tree
7. Repeat steps 2 to 6 for Random Forest Method
8. Compare AUC for test data in case of 4/5 methods and finalize the method

**Phase 5 is about text mining/sentiment analysis**

1. Use the twitter(X) data about Nifty50
2. Perform text pre-processing
3. Obtain Word-Cloud
4. Perform Sentiment Analysis
5. Visualize sentiment scores

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*