Experiential Learning Assignment

Course No: MATH F432 (Applied Statistical Methods)
Max. Marks: 15
Last Date of Submission: 22 November, 2019

Problem 1: Data Analysis for Investments and Portfolio Management

Preparation files:

1. Visit www.nseindia.com
2. Locate "products" tab, and then under "capital markets" click on "indices"
3. On this page click on "+Historical Data"
4. Then, click on "search" belonging to "Historical Index Data"
5. Select Index :: Nifty 50 // Select a time period 01.03.2018 to 28.02.2019 -- Click "Get Data"
6. A data set will be displayed. Scroll down and click on "download file in csv format"
7. A file "data.csv" will be saved on your computer.
* This is the first data set for your assignment.
9. Go back to Step 2 above :: Now click on "+About indices"
10. Click on "Broad Market Indices"
11. Click on "Nifty 50 Index"
12. Download these three files and save them on your computer
(a) Methodology (.pdf)
(b) List of NIFTY 50 stocks (.csv)
(c) Fact Sheet of NIFTY 50 (.pdf)
*(a) and (c) are for our general information.
13. Open .csv file containing the list of NIFTY 50 stocks and find and note down the "symbol" of "Zee Entertainment Enterprises Ltd."
14. Go back to Step 2 above. This time click on 'equities' and then on "+Historical Data" and then on "search" under "Security-wise Price/Volume Archives"
15. Now get historical data for "all" for period 01.03.2018 to 28.02.2019 and save this as .csv file on your computer.
16. Now locate "live market" tab, go to "live analysis", and then click on "top ten gainers and losers"
17. Browse through other tabs and details to enrich your knowledge

Guidelines for report writing: Assume that you have to explain the importance, rationale, and methodology of stock exchange market to a layman who just have some theoretical exposure of Statistics. Therefore, prepare a detailed group report (limited to about 5000 words, or about 8 pages) in a lucid yet technical manner including tables/graphs that must include (but not limited to)

1. What are NSE and Nifty 50 Index? What is the index methodology? (explain basics of share market)
2. Why do we need to invest? Does it increase my stress level? Why shouldn't I deposit my money in a reputed bank, and get the nominal interest without any worry?
3. How can one be an effective investor (based on your analysis from the historical data and associated information)? Shall I invest some amount daily, or in a particular day of a month, or once in a month? Is the stock market operational daily, 24 hours? If not, at what time it becomes operational and when (e.g., early or late hours) should I invest? In which part of industry (IT, Telecom, …), should I invest? Shall I target to become a domestic or an international investor? Is it possible to determine the underlying best-fit probability distribution for the opening/closing prices? Using graphical tools, when you plot the open/close prices over time, do you see any trend/pattern? You are free to use any statistical software should there be any requirement.
4. What is a portfolio? What is portfolio management? Why do we need it?

Problem 2: Forecasting and Analysis of Renewable Energy

Preparation files and guidelines:

1. Visit www.mnre.gov.in and understand about various renewable energy resources.
2. Note that our aim is to analyse solar and wind energy for the study region. If possible, we may like to forecast solar/wind energy for next month.
3. Download 2000-2010 hourly data of Charanka Solar Park (Gujarat) from NALANDA.
4. Look at various terms such as DHI, DNI, GHI, dew point, temperature, pressure, relative humididty, wind speed, etc. Understand which terms are relevant to solar energy and which are to wind energy.
5. Perform several graphical plots or compute various descriptive statistics to understand the data, their correlation, etc.
6. Now let us concentrate on only two dataset: GHI and wind speed for the year 2000. Do GHI and wind speed seem to follow normal distributions? If not, identify the underlying probability distributions of these data set for the year 2000.
7. Now let us consider 2000-2010 wind-speed data. Plot them as a time series. Do you see any trend or seasonality in the time series data? Could you somehow decompose this time series into various components?
8. As our uhe ultimate aim is to forecast the wind speed for the next month (i.e., Jan 2011), how will you proceed for this?

Guidelines for report writing:

1. The total report should be of about eight pages.
2. Write about one page introduction for the renewable energy resources, highlighting the present status and future plan in India.
3. Please feel free to analyse the data by yourselves. Each step of your analysis must be supported with arguments and logic. If you perform any statistical test, make sure that the assumptions/conditions are met. In case you are unable to find any conclusion, justify it.
4. You are encouraged to use any statistical software of your choice for a stringent analysis.
5. You may need to read some book, journal papers or news articles of your choice to understand the inherent concepts.

Problem 3: Earthquake Interevent Time Distribution

Preparation files and guidelines:

1. Earthquake interevent time analysis helps up to assess earthquake hazards of a seismic region. Several probability distributions are used for this purpose. Before we proceed, let us first understand a few basic concepts about earthquakes. Visit https://earthquake.usgs.gov/education/ for details.
2. What is an earthquake? Is India prone to earthquakes? What do you mean by earthquake prediction? Is earthquake prediction same as earthquake forecasting? Find at https://www.usgs.gov/faqs/can-you-predict-earthquakes?qt-news_science_products=0#qt-news_science_products
3. Now let us collect previous earthquakes for the Himalaya and adjacent region. For this, visit http://www.isc.ac.uk/iscbulletin/search/catalogue/, in ISC Bulletin → CSV formatted catalogue → rectangular search region, latitude 24-34 deg N, longitude 75-95 deg E → time period 1900.01.01 to 2019.10.25 → additional parameters, depth 0-200 km, magnitude 6-10, magnitude type 'any', magnitude author ISC →output event catalogue. Thus you get a list of earthquakes that occurred in the region after 1900.
4. In order to remove the dependent events, such as foreshocks, aftershocks and seismic clusters, let us apply a dynamic window-based spatio-temporal filtering algorithm as:
   Search radius $r = \exp(-1.024 + 0.804M) \pm 15,$ and
   time window $t = \exp(-2.870 + 1.235M) \pm 60.$ If there is any event falling within the search radius and/or within the time window, we remove that event. In this way, we find a catalog of i.i.d events.
5. The interevent times of the declustered catalog can be obtained by subtracting the occurrence time from the next occurrence time. For example, if two events occurred on April 04, 1905 and August 20, 1908, then the difference of these dates will be the interevent time.
6. Having the list of interevent times, now we are ready for modeling using various probability distributions, such as exponential, gamma, Weibull and lognormal.
7. Use MLE and MoM parameter estimations to obtain the estimated parameters.
8. Now in order to prioritize the candidate probability distributions, let us apply two tests: AIC and K-S.
9. You may use MATLAB's inbuilt tool 'fitdist' or you can use any other software of your choice to summarize your results.
10. Now, having done the above steps, how can you forecast earthquakes?

Guidelines for report writing:

1. The total report should be of about eight pages.
2. Write about one page introduction on earthquakes and earthquake forecasting.
3. Please feel free to analyse the data by yourselves. Each step of your analysis must be supported with arguments and logic. When you perform any statistical test, make sure that the assumptions/conditions are met.
4. You are encouraged to use any statistical software of your choice for a stringent analysis.
5. You may need to read some book, journal papers or news articles of your choice to understand the inherent concepts.

Submission Process: As we discussed earlier, please form groups by yourselves (08 persons in a group, having at least three different branches such as B4, A2, A7, A3, etc.), and email me the group details (in excel sheet mentioning name, ID, with group leader's name in bold) by **05 November, 2019**. The group leader, on behalf of his/her group, should take this responsibility of submitting the group report in time. Make sure that you do not copy any paragraph/sentence directly from books/articles/online sources to avoid copywriting permission and to smoothly pass through plagiarism checker. **The last date of report submission (by e-mail) is 22 November, 2019. Delayed submission will yield a zero credit. Also note that individual report submission is strictly prohibited for this course.**