Mid-semester Test

March 01, 2024: 11:30 – 13:30

Marks: 20, Weightage: 20%

INSTRUCTIONS:

1. This test SHOULD be answered using Jupyter Notebook only.
2. The name of the Notebook SHOULD be your roll number.
3. The first cell of the Notebook SHOULD be a markup cell and it should list all the major decisions and steps you have taken to arrive at the required solution(s). Subsequently, include analysis and recommendations cells as and when required.
4. The code should flawlessly execute with your data file(s) kept in the same folder.
5. At the end of the test, you should upload your **data file** and the **Jupyter Notebook** (with code / illustrations / analysis / recommendations / etc.) to the **Midsem Test** submission point on Moodle.
6. **The submission point will close at 13:45 Hrs.**
7. Use of computers is allowed, with the following strict exceptions:
    a) **Use of MS-Team, WhatsApp, Google Docs, or any other collaboration software is NOT ALLOWED at any time through the duration of the test. Such applications should remain closed.**
8. Use of mobile phone is allowed, but **ONLY** for connecting to the internet, if required.
    a) **No other application (including calculators) should be used on the mobile phone during the period of the test.**
9. **MALPRACTICES / CHEATING: ANY MALPRACTICE OR VIOLATION OF ABOVE STATED RESTRICTIONS OR ANY CHEATING WILL RESULT IN THE SEVEREST POSSIBLE DISCIPLINARY ACTION – WHETHER DETECTED DURING THE TEST OR LATER / DURING EVALUATION.**

Read the following problem carefully and work out your solutions following all the best practices. You will be assessed as per the following scheme:

- Solution approach, thoroughness, completeness: 10 marks
- Correctness of analysis and recommendations: 5 marks
- Documentation of results – consolidation, quality, brevity: 5 marks

---

TopTex Textiles is an Indian company involved in the business of manufacture and sale of garments. Its customers span across age groups and gender. It has factories at two locations. It also operates its own sales channel, by way of company owned outlets (totally 136) located in several Class A, B, C cities and towns across India.

About a year back, the company operationalized online systems to track customer profiles and sales. The data related to *customers' first visit,* captured during the last 48 weeks, is now available and reproduced in the file **toptex.csv**.

The company is now seeking to open unique super stores at a few locations. These super stores will be targeted to provide enhanced engagement experiences to customers residing within 15-20 km radius. The company wants to make these stores unique, especially to first-time customers, and, also, super-efficient in utilizing their sales persons.

To this end, using the online data now available, the company is interested in creating effective *Customers' Purchase Value, and Engagement Time Prediction Models.* They are interested in engaging a competent person to create a prototype, and guide its development. *You are a candidate!*

The executive of the company, who is responsible for IT applications, has explained the requirement as follows:

> *The expectation is that when new customers walk in, they will be first ushered into a "meet and greet" area where an employee will offer them refreshments, have a chat with them, and get some basic data from them. This data will be fed into the model, which will then predict the overall value of the purchases that the*

customer is likely to make. These predictions will be used to assign an appropriately experienced sales person to attend to the customer.

The executive further stated the following:

We are also keen to predict the total duration of the customers' stay in the store. This is, again, to ensure our salespersons' efficient allocation to customers.

It is known that this executive has a good hold on Data Analysis and ML, and has an eye for, and very much likes details. The executive is a tough questioner, not given to easy convincing as far as analysis and recommendations are concerned. She asks probing questions, and likes to have well thought out and justified alternatives / recommendations to choose from. Talking of alternatives, the executive is a firm believer that most people goof up when it comes to the use of RF. If, for some reason, Linear Regression performs better than RF, you had it!

You are required to analyze the data that they have provided, work your way through to satisfy their requirements. Keep in mind the qualities of the executive you are dealing with. In addition to whatever you do, also be ready to answer pointed questions like - "Can you tell me what I should do to increase my sales?", "If I motivate people to spend more time in the stores, what kind of impact will that have on the sales?", etc.

Do your best to satisfy the executive, and get the assignment. It will help if you create logical, complete yet precise and well organized material for explanation and discussion. Good luck!

***

Remember:

- Your submissions should be uploaded to Moodle by 13:45 Hrs (March 1, 2024).
- Cheating, and use of any collaboration tools on the computer / phones is completely prohibited. Violations WILL attract disciplinary action.

Only in case of difficulty in submitting to Moodle by 13:40 Hrs:

- Using your IITB login, upload your submission to : https://tinyurl.com/2023-03-01-submission
- You can also use the following direct link:
    - https://docs.google.com/forms/d/e/1FAIpQLScb5GShkJmudEvfw2TZx6iIRF3XNv8zuAOqNp4kmgm2ktyumw/viewform
- **You will receive a confirmation email upon successful submission.**
- **If you are unable to submit even using this alternate method, immediately contact one of the invigilating TA / Instructor before 13:45 Hrs.**

Under no circumstances will submissions be accepted after 13:45 Hrs.

***