**Introduction to Machine Learning**

**Group project Deliverable #3: Unsupervised Machine Learning Development**

The purpose of this project is to provide students with the opportunity to use unsupervised machine learning algorithms to assist four or five industries in solving some of their strategic business problems. The solutions should cover the two types of ML algorithms:

1. Unsupervised
2. Forecasting

There are **two parts** to this project:

1. **Individual tasks**: Single model development (per industry) – completed by individual members.
2. **Group Activities**:
   1. Cross-model performance analysis - to be completed in groups (together)
   2. Cross-industry model analysis - to be completed in groups (together)

**Summary instructions:**

1. Individual tasks are to be completed first – by each member.
2. The group activities begin once the individual tasks are completed, reviewed, and validated by all team members.
3. The decks produced for individual tasks should be used to complete the group activities.
4. The project manager is responsible for creating the PowerPoint deck for the group activities.
5. The decks each member (individuals) produced, and the group activities deck should be used for the final presentation video.

**Industries**

1. Financial
2. Oil and Gas
3. Healthcare
4. Airlines
5. Motion pictures industry
6. Real Estate

**Part 1 – to be completed by individual member of the team.**

**Unsupervised ML/Forecasting**

**Team model assignments/responsibilities - each team member (except the project manager) should develop two models: one continuous and the other Categorical. The models might be for one or two industries depending on the available data.**

**Team member #1**

1. *Unsupervised ML* – Movie performance clustering
2. *Forecasting* - Oil price forecasting

**Team member #2**

1. *Unsupervised ML* – Credit card usage clustering
2. *Forecasting* – Walmart Weekly sales forecasting

**Team member #3**

1. *Unsupervised ML* – Bank IPO clustering
2. *Forecasting* – Energy consumption forecasting

**Team member #4**

1. *Unsupervised ML* – Cancer survival clustering
2. *Forecasting* – Electricity consumption forecasting

**Team member #5**

1. *Unsupervised ML* – Airline customer satisfaction clustering
2. *Forecasting* – US oil production forecasting

**Project Manager**: The last member will be the project manager for the group:

* Lead, facilitate, and coordinate team project activities and deliverables.
* Review the Python scripts and models developed by each member and provide feedback.
* Ensure project deliverables are completed on time and with quality.
* Provide weekly project updates to the class and should include the following:
  + Tasks completed the previous week.
  + Tasks planned for the current and next week's activities.
  + Risks, challenges, and issues
  + Overall project status – on track or not
* The deliverables from all members are collated and reviewed by all group members and approved for submission.
* **Work with the team to complete the group activities (below) and develop the PowerPoint slides for group activities**.
* The PM submits the final product on D2L.

**Additional Instructions**

1. Assess or evaluate the accuracy of the models using appropriate measures.
   1. Accuracy score
   2. Silhouette score.
2. Provide detailed/advanced analysis or explanation of how the model results help to solve business problems.
   1. **Clustering performance** – measure using silhouette coefficient score.
      1. The value of the silhouette coefﬁcient is between [-1, 1].
         1. The silhouette score of 1 means that the clusters are very dense and nicely separated.
         2. The score of 0 means that clusters are overlapping.
         3. The score of less than 0 means that data belonging to clusters may be wrong/incorrect.
      2. Determine the best way (or combination of variables) to cluster the dataset and produce the graph or graphs.
      3. Identify and analyze unique characteristics of each cluster or group.
   2. **Forecasting algorithms – Advanced analysis.**
      1. How good is the model?
         1. Explain or assess the model’s performance.
         2. Is the score good or not, why, and why not?
         3. How can you improve the model’s performance?
   3. Explain how the models you created helps solve the business problem(s)
   4. What would you do differently if you had to build/deploy this algorithm for a prospective employer?
   5. Document some of the weaknesses of the model.
   6. Provide suggestions on how to improve the model and the results.

**Part 2 – Group activities: to be completed in group (and coordinated by the PM)**

**Part 2A: Cross algorithms analysis: Unsupervised ML**

1. Compare clustering algorithm results across industries.
2. Which of the algorithms best fits or solves the business problems
3. Does the industry-specific context affect the variations in algorithms' performances?

**Cross algorithms analysis – Forecasting**

1. Compare forecasting algorithm results across industries.
2. Which of the algorithms best fits or solves the business problems
3. Does the industry-specific context affect variations in algorithms' performances?

**Part 2B: Cross-industry models analysis**

1. Compare the business problems across industries.
2. Determine and document their ease or difficulty of the models to solve the problem.
3. Compare the models created for each industry.
4. Determine the differences and whether the effectiveness or goodness of the model relates to industry effects.

**Project deliverables**

1. **A PowerPoint presentation deck:** should contain the group project results with different slides for what each student completed. Each student will present the results of his or her tasks. In essence, each member creates slides to present the results of their tasks and sends them to one person to combine into a single PowerPoint deck. During the presentation, each student goes over their slides and provides additional information about their tasks.
   1. Prepared and presented professionally.
   2. The Title page should contain the topic, group names, group members, and date.
   3. A slide with project and presentation objectives
   4. A slide contains the presentation agenda or table of contents.
   5. Should include graphs or plots and relevant pictures.
   6. Minimum 30 pages and 50 pages (Max)
   7. See additional instructions below.

**The PowerPoint deck and presentation must be professionally done and should conform to the following cadence:**

* + 1. Write before you design.
    2. Start with a title slide those piques interest.
    3. Stick to a simple design.
    4. Emphasize one point per slide.
    5. Use text sparingly (very important)
    6. Select images for impact.
    7. Practice your verbal presentation (prior to the presentation day)
    8. Run your presentation by other students.
    9. End with a persuasive call to action
    10. Explore the power of PowerPoint.
    11. Click and read the detailed instructions or points below

<https://www.monster.com/career-advice/article/powerpoint-hacks-1116>