**Introduction to Machine Learning**

**Group project Deliverable #2: Supervised Machine Learning Algorithms Development**

The purpose of this project is to provide students with the opportunity to use machine learning algorithms to assist four or five industries in solving some of their strategic business problems. This deliverable will cover **Supervised ML** algorithms.

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.

**Team Composition and Responsibilities**

* One of the team members should be a **project manager.**
* *An example*: Five or six-member team: 4 group members perform research and analysis for one of the four industries, and the fifth member, the project manager (PM)
* Each group member is assigned to an industry to develop two ML algorithms to solve two business problems.
* Each member builds their assigned models and shares the results with the team.
* A different member reviews the codes and model and provides feedback/suggestions.
* Original members review and incorporate the feedback.
* **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.
* The PM submits the final product on D2L.

**Industries**

The first task is for each team to select 4 or 5 industries based on team composition. A team with five members selects four industries and a project manager. A team with 6 members chose 5 sectors plus a project manager. Below are some of the suggested 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 members of the team.**

**Part 1A: Single model development - Supervised ML**

**(individual tasks: to be completed by each team member)**

1. Each student/team member selects an industry and develops two supervised ML algorithms to solve two business problems for each industry: 2 x 4 or 5 = 8 or 10 models in total (per group)
   1. One each from the categories below:
      1. **Continuous – linear regression**
         1. Housing Price predictions (y = price)
         2. Bank IPO performance predictions (y = Gross spread)
         3. Movie performance predictions (y = gross box office)
         4. Retail profit predictions (data – customer order analysis, y = profit)
         5. Airline revenue predictions (y = net sales)
      2. **Classification** - Logistic regression, decision tree, random forest, KNN, XGboost, etc.
         1. Airline customer satisfaction predictions (y = satisfaction)
         2. Fraud predictions (y = isFraud)
         3. Cancer survival predictions (y = patient\_status)
         4. Race car predictions (y = win)
         5. Bank IPO performance predictions (y = Day1perf)
2. Use one or multiple models (preferred option) for each algorithm (and determine which model is better).
3. Search or use the datasets from the following:
   1. Use datasets attached (recommended option) OR
   2. OR Search for dataset on Kaggle.com
   3. OR Find dataset from other sources.
4. You can use single or multiple datasets per industry.
5. Assess or evaluate the accuracy of the models using appropriate measures.
   1. Accuracy score
   2. Mean Squared errors.
   3. R-squared
   4. Others
6. Provide detailed analysis or explanation of how the model results help to solve the business problem.
   1. *Explain the results.*
      1. How good is the model?
         1. Explain the accuracy score or mean-square-error.
         2. Is the score good or not, why, and why not?
         3. How can you improve the model’s performance?
      2. Explain how the model you created helps solve the business problem(s)
      3. What would you do differently if you had to build/deploy this algorithm for a prospective employer?
7. Document some of the weaknesses of the model.
8. Provide suggestions on how to improve the model and the results.

**Part 2 – Group activities: to be completed in group (or together)**

**Part #2A: Cross-model performance analysis – Supervised ML (To be completed together or in group)**

* 1. Compare the performance of the 8 or 10 models.
  2. Evaluate and describe the model with the best performance vis-à-vis the business problems.
  3. Why does one model perform better than others?

**Part #3: Cross-industry models analysis (Group activities – to be completed by all team members)**

* 1. Compare the business problems across industries.
  2. Determine and document their ease or difficulty of developing 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.

**Practical exercises instructions**

1. Determine the best algorithms or model.
   1. Must use at least three models (one of them should be XGBoost)
2. Load dataset
3. Perform EDA
   1. Data analysis, cleansing, and wrangling
   2. Plots
   3. Check and Remove NA
4. Build model.
5. Test model
6. Evaluate model performance.
   1. Data
   2. Plot
   3. Continuous variable
      1. Mean squared error.
      2. Weighted mean errors.
   4. Categorical
      1. Accuracy score by models > determine the best model based on results.
      2. Confusion matrix
   5. Linear Regression
      1. P-value
      2. R-square
      3. Coefficient

**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>