**Milestone 1**

It is time to start using what you have learned throughout the first half of this course by developing an original data mining project. This week you will develop the project idea and do some data exploration/graphical analysis. You will continue working on and updating this project for the remainder of the term. The following link includes a sample project completed in Jupyter Notebook: Titanic Model Building Project Sample.

The first step is coming up with an idea – arguably one of the hardest steps! Identify an original business problem for your project that can be solved with an appropriate model. By a business problem, it is meant that you should work on a problem where there is a good reason to solve it. There should be some organization or company that would find the solution to the problem useful. There are lots of ideas available online through Kaggle and other sources, but your idea should have a unique spin on it. The second step is locating your data. This can come from a variety of sources, e.g., Kaggle, your job, a website, API, etc. Feel free to reach out to your instructor if you are not sure if your idea and data are suitable. You may need to adjust your idea on the availability of data.

Begin Milestone 1 with a 250-500-word narrative describing your original idea for the analysis/model building business problem. Clearly identify the problem you will address and the target for your model. Then, do a graphical analysis creating a minimum of four graphs. Label your graphs appropriately and explain/analyze the information provided by each graph. Your analysis should begin to answer the question(s) you are addressing. Write a short overview/conclusion of the insights gained from your graphical analysis.

As a reminder, Teams is a great place to discuss your project with your peers. Feel free to solicit feedback/input (without creating a group project!) and collaborate on your projects with your peers. Each milestone will build on top of each other, so make sure you do not fall behind.

**Predict Autism Spectrum Disorder (ASD)**

Autism spectrum disorder (ASD) is a developmental disability caused by differences in the brain. People with ASD often have problems with social communication and interaction, and restricted or repetitive behaviors or interests. People with ASD may also have different ways of learning, moving, or paying attention. It is important to note that some people without ASD might also have some of these symptoms. But for people with ASD, these characteristics can make life very challenging.

**Social Communication and Interaction Skills**

Social communication and interaction skills can be challenging for people with ASD. Examples of social communication and social interaction characteristics related to ASD can include:

* Avoids or does not keep eye contact
* Does not respond to name by 9 months of age
* Does not show facial expressions like happy, sad, angry, and surprised by 9 months of age
* Does not play simple interactive games like pat-a-cake by 12 months of age
* Uses few or no gestures by 12 months of age (for example, does not wave goodbye)
* Does not share interests with others by 15 months of age (for example, shows you an object that they like)
* Does not point to show you something interesting by 18 months of age
* Does not notice when others are hurt or upset by 24 months of age
* Does not notice other children and join them in play by 36 months of age
* Does not pretend to be something else, like a teacher or superhero, during play by 48 months of age
* Does not sing, dance, or act for you by 60 months of age

**Restricted or Repetitive Behaviors and Interests**

People with ASD have behaviors or interests that can seem unusual. These behaviors or interests set ASD apart from conditions defined by problems with social communication and interaction only.

Examples of restricted or repetitive behaviors and interests related to ASD can include:

* Lines up toys or other objects and gets upset when order is changed
* Repeats words or phrases over and over (called echolalia)
* Plays with toys the same way every time
* Is focused on parts of objects (for example, wheels)
* Gets upset by minor changes
* Has obsessive interests
* Must follow certain routines
* Flaps hands, rocks body, or spins self in circles
* Has unusual reactions to the way things sound, smell, taste, look, or feel

**Other Related Characteristics of ASD**

Most people with ASD have other related characteristics. These might include:

1. Delayed language skills
2. Delayed movement skills
3. Delayed cognitive or learning skills
4. Hyperactive, impulsive, and/or inattentive behavior
5. Epilepsy or seizure disorder
6. Unusual eating and sleeping habits
7. Gastrointestinal issues (for example, constipation)
8. Unusual mood or emotional reactions
9. Anxiety, stress, or excessive worry
10. Lack of fear or more fear than expected

**Milestone 2**

Now that you have created your idea, located data, and have started your graphical analysis, you will move on to the data preparation process of your project. After completing Milestone 2, your data should be ready for the model building/evaluation phase.

Here is a list of steps to consider performing in Milestone 2:

* Drop any features that are not useful for your model building and explain why they are not useful
* Perform any data extraction/selection steps
* Transform features if necessary
* Engineer new useful features
* Deal with missing data (do not just drop rows or columns without justifying this)
* Create dummy variables if necessary

Explain your process at each step. You can use any methods/tools you think are most appropriate. Do what makes the most sense for your data/problem. This will vary greatly among different projects. Be careful to avoid data snooping in these steps.

It is important to note that these milestones are meant to keep you on track for the final project submission. At any point, you can pivot or modify your project as needed based on what you discover. These milestones are not final versions; they are drafts of the many steps you need to complete along the way.

As a reminder, Teams is a great place to discuss your project with your peers. Feel free to solicit feedback/input (without creating a group project!) and collaborate on your projects with your peers.

Each milestone will build on top of each other, so make sure you do not fall behind. **Submit Milestones 1 & 2 together.** I recommend building your project milestones in a Jupyter Notebook, building upon one another. However, make sure it is clear where Milestone 1 ends and Milestone 2 begins.

**Milestone 3**

In Milestone 3, you will begin the process of selecting, building, and evaluating a model. You are required to train and evaluate at least one model in this milestone. Write step-by-step for performing each of these steps. You can use any methods/tools you think are most appropriate, but you should explain/justify why you are selecting the model(s) and evaluation metric(s) you choose. It is important to think about what type of model and metric makes sense for your problem. Again, do what makes the most sense for your project. Write a short overview/conclusion of the insights gained from your model building/evaluation.

It is important to note that these milestones are meant to keep you on track for the final project submission. At any point, you can pivot or modify your project as needed based on what you discover. These milestones are not final versions; they are drafts of the many steps you need to complete along the way.

As a reminder, Teams is a great place to discuss your project with your peers. Feel free to solicit feedback/input (without creating a group project!) and collaborate on your projects with your peers.

Each milestone will build on top of each other, so make sure you do not fall behind. **Submit Milestones 1-3 together.** I recommend building your project milestones in a Jupyter Notebook, building upon one another. However, make sure it is clear where each milestone begins and ends.

**Milestone 4**

You have made it to the final week of the course and the time has come to submit your final project! Using your own judgment and based on the feedback you have received, update your project accordingly. Add any new code and/or analysis to your content from Milestones 1-3. Clearly note what content has been added since Milestone 3. Include this as part of your final project submission.

The primary final submission for the term project is a minimum five-page project writeup, e.g., MS Word or PDF file, summarizing the details of your project. Your final submission should include the following:

* **Introduction**
  + Introduce the problem
  + Justify why it is important/useful to solve this problem
  + How would you pitch this problem to a group of stakeholders to gain buy-in to proceed?
  + Explain where you obtained your data
* **Organized and detailed summary of Milestones 1-3**
  + EDA; include any visuals you think are important to your project
  + Data preparation
  + Model building and evaluation
* **Conclusion**
  + What does the analysis/model building tell you?
  + Is this model ready to be deployed?
  + What are your recommendations?
  + What are some of the potential challenges or additional opportunities that still need to be explored?