Introduction to Business Data Analytics Final Project Grading Criteria

Group Project Proposal
Dataset Kaggle
Dataset ATH
*warning: drop last two columns *Naive_Bayes_Classifiers*
Column Descriptions

- **Business Understanding** (15 points)
    - Identify and motivate the business problem that you will address.

Business context
For businesses that offer subscriptions or memberships, it is typically more expensive to generate new customers than it is to retain current customers. For this reason, predicting customer churn is very valuable.

Business problem

Ambiguous classification of customers churn.

Business goals
Classification

Classification models help businesses run their marketing operation efficiently.Customers most likely to end service can be targeted with promotional offers to stay, thus raising revenue, while customers likely to maintain service do not need to be targeted with special offers, lowering costs.

    - Creativity? Precise statement of problem?

Ambiguous classification of customers churn.

- **Data Understanding** (15 points)
    - Identify and describe the data (and data sources) that will support data mining to address the business problem. Include those aspects of the data that we routinely talk about in class and/or in the homework assignments.

Our dataset comes from a bank experiencing high customer attrition. To address this issue, they need to know what customer attributes are predictive of churn, and thereby, which customers to target with retention offers.

- **Data Preparation** (15 points)
    - Specify how you will integrate these data to produce the format required for data mining.

    - Todri: "*The dataset is imbalanced; you might want to consider the strategy of rebalancing the target variable which is appropriate for imbalanced datasets. The model evaluation Python notebook of the class has an example of how to deal with imbalanced datasets..*"
    - Todri: *"It looks like it would be interesting to create new features (e.g., interaction effects, etc.) for this business problem. For instance, older male customers might be less likely to churn. To capture such patterns, you would have to interact with the corresponding features*."
- **Modeling** (15 points)
    - Specify the type of model built.
    - Discuss your choices for data mining algorithm: what are alternatives, and what are the pros and cons.
    - Discuss why and how this model should "solve" the business problem (i.e., improve along some dimension of interest to the firm).
- **Evaluation** (15 points)
    - Discuss how the result of the data mining is/should be evaluated.
    - How should a business case be developed to project expected improvement? ROI? If this is impossible/very difficult, why and is there a viable alternative?
    - Todri: "*For this context, it would be interesting if you estimate the expected value of classifiers to plot a profit curve. This is just a suggestion of course*.."
- **Deployment** (10 points)
    - Discuss how the result of the data mining will be deployed.
    - Discuss any issues the firm should be aware of regarding deployment.
    - Are there important ethical considerations?
    - Identify the risks associated with your proposed plan and how you would mitigate them
- **Overall** (15 points)
    - Presentation, communication.