Fall 2023 IST 557: Data Mining: Techniques and Applications

Instructor: Lu Lin

Individual Project I: Heart Attack Prediction (20 points + 1 bonus point)
Due Date: 11:59 PM, Friday, September 22, 2023

Goal: This project is to expose students to real-world Kaggle Data Mining competition for practicing basic skills on classification task. DO NOT CHEAT. You can only learn data mining and machine learning by getting your hands dirty. Enjoy it.

Logistics: To successfully complete the Kaggle competition, follow the following rules:

(1) ONLY create ONE account for the competition.

(2) Welcome to use ANY classification methods, e.g., logistic regression, Perceptron, K-Nearest Neighbors, SVM, Bagging, Boosting, Random Forest, etc. You can also apply ANY tricks/practices we mentioned in class, i.e. feature preprocessing, model selection.

(3) You can submit your prediction 10 times each day in maximum.

Kaggle Competition: The kaggle competition for this project can be accessed via this link: https://www.kaggle.com/t/7756b6874d924c86b28b81e8f93e30c4. This competition is created only for students enrolled in IST 557 to participate. **Please do not spread this link**.

Submission Checklist: You need to submit the following to **P1 on CANVAS**:

(1) Final code file (Jupyter notebook or python file) used to train your model and make predictions for the final submission.

(2) Brief project report in PDF (within 3 pages with proper fomatting) containing the following information:

- Kaggle account name, your name and PSU email
- Screenshot of your final rank in the leaderboard
- Brief description of your solution, such as data pre-processing, feature engineering, model building and selection, hyperparameter tuning, performance evaluation, any ideas and lessons learned

Grading Rubric: Total 20 + 1 bonus points consists of two parts:

- Performance (F1 Score) on the leaderboard **(10 points + 1 bonus point)**
  - F1 score $\in$ [**0.90**, 1.00): 10 points
  - F1 score $\in$ [**0.85**, 0.90): 8 points
  - F1 score $\in$ [**0.80**, 0.85): 6 points
  - F1 score $\in$ [**0.70**, 0.80): 4 points
  - F1 score $\in$ [**0.60**, 0.70): 2 points
  - F1 score $\in$ [**0.00**, 0.60): 0 points
  - **Students that achieve Top 3 highest F1 score are awarded 1 bonus point**
- Quality of report (or Jupyter notebook) **(10 points)**: it can be a brief document, as long as you summarize the practices you tried and things you observed.