University of Ottawa
Faculty of Engineering
School of Electrical Engineering and Computer Science

Assignment 2

Course	ELG7186 – AI for Cybersecurity Applications
Academic year	2023/2024
Semester	Fall
Instructor	Paula Branco
Announced	25 September 2023
Submission Deadline	**13 October 2023 11:59pm (GMT-4)**

NOTE: Strictly avoid copying your colleague's assignment. That would amount to plagiarism. Penalty in case plagiarism is detected: zero marks will be assigned for all parties whose assignment would be considered as plagiarized OR copies of each other.

Every student must submit the assignment **individually** on the Kaggle competition page provided

Assignment Overview

In this assignment, you must provide a solution for a network intrusion detection problem. The link to join the competition is the following: https://www.kaggle.com/t/e4cee33c5e47ae7c4a1bb3c082aa4d5e.

After joining the competition, you may access the assignment through the following link: https://www.kaggle.com/competitions/elg7186-assignment-2-is-this-an-intrusion/. In the assignment link, you will find a train set, a test set and a sampleSubmission file. Your goal is to train a model using the train set and obtain the predictions for the test set. After having obtained your model's predictions, you must build a submission file (the file sampleSubmission.csv shows an example of how this file should look like) and then you will upload your solution on the Kaggle competition page. Your score will be automatically calculated. Your main goal is to obtain the highest possible score for this problem! You can submit a maximum of 10 solutions per day and in the end, you will be able to select the 4 solutions that you want to be evaluated.

Instructions:

1. Open the Kaggle competition invitation link: https://www.kaggle.com/competitions/elg7186-assignment-2-is-this-an-intrusion/.
2. On the competition page, please join and register your team (yourself) in the format **Student Number - First & Last Name**, for example, **300300300 - Jane Doe**, so that it is easier to calculate your final grade.
3. Using the train data build a model to obtain predictions for the test set provided.
4. Format the solutions obtained according to the following guidelines: provide a csv file with a header containing two columns (ID, Class). The column ID should have all the values of the column ID in the test data. The column Class should contain your predicted class label. Check the sampleSubmission.csv to see a valid submission file.

5. Your score will be updated in the public leaderboard each time you upload a solution. The metric used for evaluating your solution is the F1-score (https://en.wikipedia.org/wiki/F1_score).
6. The public leaderboard uses roughly 50% of the test data to evaluate your solution. The final results (private leaderboard) will be based on the other 50%, so the final standings may be different. The final private leaderboard scores will be used for evaluating your assignment.
7. You can upload a **maximum of 10 solutions per day (GMT-4/Toronto) between September 25 and October 13.**
8. At the end of the competition you can select 4 solutions that you want for evaluation. If you don't select them, then the 4 solutions with the highest score in the public leaderboard will be used.
9. **You should have the code used for this assignment uploaded to Brightspace before the deadline as well.**
10. You can **publicly discuss** the assignment and ideas for solving it with your colleagues on Kaggle or Slack. However, **you are not allowed to share code or predictions**!

Assignment Evaluation:

1. The final evaluation will be made using the private leaderboard scores which will be revealed after the competition ends. Until then only the public leaderboard is available to observe the performance of the solutions.
2. The Kaggle competition is provided with a baseline (decisionTree-baseline.csv in the leaderboard). This baseline will match the score of 50% in this assignment. This means that all students having a model with a better score than the baseline will have a final assignment grade higher than 50%.
3. The maximum score of the assignment (100%) will be assigned to the best overall performing solution of the students.
4. A **sigmoid** function will be used to map the F1-scores into a 0-100% scale according to the rules described above.