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

Assignment 2

Course	ELG7186 – AI for Cybersecurity Applications
Academic year	2021/2022
Semester	Fall
Instructor	Paula Branco
Announced	27 September 2021
Submission Deadline	21 November 2021 11:59PM (EST)

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

Every student must submit the assignment **individually** on Brightspace

Assignment Overview

In this assignment, you will need to tackle two problems:

- **Network Intrusion Detection problem** in which you will perform binary class classification and predict whether there is an attack or not.
- **IOT Botnet Attack Detection problem** in which you will perform multiclass classification and predict is the case is normal or a specific type of attack.

For each one of these problems, you are expected to implement two predictive modeling solutions: one static solution and another solution that adapts through time.

For both the problems, you will have 2 sources of data:

- An Initial CSV file which you can use to train an initial model.
- A data stream (accessed from the Kafka Server) which will be used to evaluate the static solution and the solution that adapts with time.

You will need to implement, test, evaluate and compare the results obtained by the two solutions in each problem.

Instructions:

Problem 1 (Network Intrusion Detection):

- CSV file with the initial data provided on Brightspace: **cicids_static_data.csv**

- Implement two learning algorithms (at least one needs to adapt through time)

- Access Kafka Server (**refer to Kafka_Consumer.ipynb file on Brightspace**):

- Test the algorithms (choose adequate metrics, performance evaluation strategy, etc)

- Summarize, compare and discuss the results

- Read more about the data and its attributes here: https://www.unb.ca/cic/datasets/ids-2017.html

Problem 2 (IOT Botnet Attack Detection):

- CSV file with the initial data provided on Brightspace: **iot_static_data.csv**

- implement two learning algorithms (at least one needs to adapt through time)

- Access Kafka Server (**refer to Kafka_Consumer.ipynb file on Brightspace**):

- test the algorithms (choose adequate metrics, performance evaluation strategy, etc)

- summarize, compare and discuss the results

- Read more about the data and its attributes here:

https://archive.ics.uci.edu/ml/datasets/detection_of_IoT_botnet_attacks_N_BaIoT#

Important Note for both tasks:

- The messages that you will consume from the Kafka server contain the attributes as well as **the class label**. In a real-life scenario this wouldn’t be the case, you would have only the data without the class label (because this is what you are expect to predict). Here, **the label is provided just for the evaluation purposes.**
- You need to consume (read and evaluate) **100.000 instances for each task**. This should be your stopping criteria on the online task.

Deliverables:

(1) Source code used (should be clean and with comments)

(2) Report **(maximum 4 pages – 2 pages per problem)** summarizing the results of the experiments on the two problems

Submission:

- the report (pdf) and source code (zip) must be submitted on Brightspace

- deadline: 21 November 2021 11:59PM (EST)

Report guidelines:

In this report you should focus on briefly explaining the solutions you implemented, and describing the experiments carried out. The report should have two main sections: "Binary Problem" and "Multiclass Problem". Please include the following in each one of the mentioned sections:

- A subsection "Algorithms" describing the algorithms implemented for the problem. Be sure to add any necessary references. Provide only the overall idea of the algorithm (no pseudo code is necessary, no detailed explanation is required).
- A subsection "Experiments" containing:
    - a description of how you tested the algorithms (metrics selected, hyperparameters tuning, performance assessment setting, etc)
    - the results obtained (tables, plots, etc)
    - a discussion of the results (what do the plots/tables show us, the knowledge learned from the experiments, advantages and disadvantages of the solutions)