Texas Tech University Whitacre College of Engineering

CS 4331/5331 – Special Topics in Machine Learning and Information Security
Fall 2021
Assignment 3

Acknowledge your collaborators or source of solutions, if any. **Online submission is required.**

You are free to come up with your own assumptions, if every potential specification is not given to you. Just be reasonable and document your assumptions.

Your compliance with the "PROGRAMMING STYLE GUIDELINE" for CS 4331/5331 will affect your actual grade. All assignments will be checked for academic misconduct (cheating, plagiarism, collusion, falsifying academic records, misrepresenting facts, violations of published professional ethics/standards, and any act or attempted act designed to give unfair academic advantage to oneself or another student) defined by "OP 34.12: Grading Procedures, Including Academic Integrity" of TTU.

Objective: To implement the proposal submitted for Assignment 2 on an information security problem to solve with machine learning.

Tasks:

- Continue working with the team formed for assignment 2.
- Conduct the machine learning experiment proposed similarly to how the experiments are performed in chapters 9, 10, 11, 12, and 13 of Stamp with the experimental design, data collection and preparation, and experimental results. Some ideas to help include
    - Consider data in different formats, such as malware with only transpositions, malware with padding, malware with transpositions and padding, improved spam images, improved text spam, adding padding to encrypted messages, …
    - Consider adding features or attributes that are not original to a dataset and raising/lowering the dimensionality of the dataset
    - Avoid designing experiments where the results could be predicted beforehand; for example, it is well known that some models perform very well on certain problems such as Naïve Bayes performs well on text spam, performing the same experiment as on a blog site or research paper, using such a small amount of data that a model could not learn properly
    - Collect prior research results from research papers to compare against a model's results and implement more than one model for comparison purposes (larger teams should be implementing 2 or more models with several experiments)
- Format a report pdf document of not more than 5 pages as follows:
    1. Title
    2. Team member first and last names
    3. Problem to be solved
    4. Data collection and preparation
    5. Experimental design and changes during implementation
    6. Results
    7. Future Research
        - If you continued the research, what should be done next?
    8. References
        - 4 references minimum per team member from the journal and conference literature that gave you ideas and justification for the problem and machine learning models used
        - Only references cited in the report text should be included
- Write software, such as R scripts, to prepare the data, implement the model, and show the results

Learning Outcomes: Gain experience in implementing a machine learning approach to an information security problem.

Grading: 50 points

Due Date: 12/1/2021, 11:59pm (submitted on Blackboard)
The submission (report, scripts, and data) should be submitted by one team member on Blackboard.