CSIS

Coursework Specification

Assignment Details	
Module Name	Applied Machine Learning
Module Leader	Dr Paul Yoo
Set by	Dr Paul Yoo
Piece number	1 of 1
Assignment Type	Group Practical Project (with individual component)
Contribution	30%
Hand-in point	VLE – Turnitin
Due Date[1]	Sunday, 19 January 2020 – a copy must be uploaded to the VLE by *every* member of the team by 1600. See below for details. https://www.dcs.bbk.ac.uk/intranet/index.php/Coursework_Deadlines_Autumn_2019
Marks will be Returned on	1600 on Friday, 14 February 2020

[1] If you anticipate an issue with meeting the deadline due to exceptional circumstances, or have missed the deadline, please see http://www.bbk.ac.uk/management/current-students/mitigating-circumstances for details on the extension, deferral and suspension processes.

Assessment Requirements

IoT Intrusion Detection Competition using Machine Learning

Software to detect wireless network intrusions protects a computer system from various cyber attacks, including perhaps insiders. The task of intrusion detector learning is to build a predictive model (i.e. a machine-learning classifier) capable of distinguishing between "intrusive" traffic, called intrusions or attacks, and "good" normal traffic.

The Aegean WiFi Intrusion Dataset (AWID) project was prepared and managed by George Mason University and University of the Aegean. The objective was to survey and evaluate research in intrusion detection. A standard set of data to be audited, which includes real traces of both normal and intrusive 802.11 traffic with a wide variety of intrusions simulated in a physical lab which realistically emulates a typical SOHO infrastructure, was provided.

Details of the dataset can be found in the page below.

http://icsdweb.aegean.gr/awid/draft-Intrusion-Detection-in-802-11-Networks-Empirical-Evaluation-of-Threats-and-a-Public-Dataset.pdf

AWID dataset categorises the attacks according to the methodology of execution. Attacks that have similar patterns of expression fall under one of the groups:

a. injection attacks,
b. flooding attacks,
c. impersonation attacks,
d. passive attacks.

We will use the reduced CLS portion of the AWID dataset because it serves as a useful starting point for preliminary research, and affords a baseline against which the new built models can be compared with the state-of-the-art Deep Feature Extraction and Selection (D-FES) method[2], and other methods using the same reduced CLS portion[3,4].

Although flooding and injection attack signatures are also available within the AWID-CLS dataset, impersonation attacks were our focus as Hirte, Honeypot and EvilTwin

[2] ME Aminanto, R Choi, HC Tanuwidjaja, PD Yoo and K Kim (2018) Deep abstraction and weighted feature selection for Wi-Fi impersonation detection, IEEE Transactions on Information Forensics and Security, 13(3), 621–636.
[3] ME Aminanto and K Kim (2017) Detecting impersonation attack in WiFi networks using deep learning approach, Information Security Applications 17th International Workshop. Jeju Island, South Korea, 25-27 August 2016, 136–147.
[4] C Kolias, G Kambourakis, A Stavrou and S Gritzalis (2016a) Intrusion detection in 802.11 networks : empirical evaluation of threats and a public dataset, IEEE Communication Surveys and Tutorials, 18(1), 184–208.

impersonation attacks have previously been identified as the most severe threats to a network[5] and have been the focus of earlier research[6,7]. As a result, detecting impersonation attacks as the focus of our work will allow us to directly compare the performance of our model to others, a key weakness within the current body of machine-learning-based IoT IDS research.

A complete listing of the set of features defined for the connection records, the relevant papers, the dataset description is given in the link below.

http://icsdweb.aegean.gr/awid/

The dataset for your project is available on the module page of the VLE – please do not use the one in the above link.

Aim of the Assessment

The aim of this assessment is to provide a hands-on, practical, assessment of your machine learning skills for practical IoT intrusion detection application.

Description of Task to be Completed

Your task is to build a predictive model (i.e. a machine learning classifier) capable of distinguishing between “intrusive” traffic, called intrusions or attacks, and “good” normal traffic.

This is a group task with individual element, and you will work in a group of **5** students.

- **Team forming.** Find your partners as soon as possible, and *when a group is formed email the module leader (paul@dcs.bbk.ac.uk) with the details of your group members.* The module leader will update info on VLE so we know who has partners and who does not. Teaching assistants also has support for soliciting partners. If you are having trouble finding partners, ask the teaching staff, and we will try to find you a group in a fair way.

[5] C Kolias, G Kambourakis, A Stavrou and S Gritzalis (2016a) Intrusion detection in 802.11 networks : empirical evaluation of threats and a public dataset, IEEE Communication Surveys and Tutorials, 18(1), 184–208.

[6] ME Aminanto, R Choi, HC Tanuwidjaja, PD Yoo and K Kim (2018) Deep abstraction and weighted feature selection for Wi-Fi impersonation detection, IEEE Transactions on Information Forensics and Security, 13(3), 621–636.

[7] Parker L, Yoo P, Asyhari T, Chermak L, Jhi Y, Taha K. DEMISe: interpretable deep extraction and mutual information selection techniques for IoT intrusion detection. InARES'19 Proceedings of the 14th International Conference on Availability, Reliability and Security 2019 Aug 26. ACM.

- **Planning (group).** Machine learning projects are highly iterative; as you progress through the ML lifecycle, you'll find yourself iterating on a section until reaching a satisfactory level of performance, then proceeding forward to the next task (which may be circling back to an even earlier step). You need to plan carefully. You need to determine scope, resources, major tasks (and who is responsible for what) and schedule (e.g. Gantt chart). You may also need to discuss general model tradeoffs (accuracy vs speed).
- **Searching literature.** You are suggested to undertake a literature search. This is a search designed to identify existing research and information on IDS using AWID dataset. Their findings will be very helpful for your task (particularly in feature and algorithm selection tasks).
- **Downloading dataset.** The dataset for this project is available on VLE. You need to use `train_imperson_without4n7_balanced_data.csv` for training and *`test_imperson_without4n7_balanced_data.csv`* for testing. The first row of each dataset gives variable numbers (this may need to be removed). The original dataset has 154 input variables and 1 target variable however two of the input variables numbered 4 and 7 (*frame.time_epoch* and *frame.time_relative)* have been removed from both datasets as they provide temporal information which may cause unfair prediction. The training set has 97044 observations while testing set has 40158 observations.
- **Pre-processing (individual 1).** You are suggested to use suitable descriptive statistics and visualisation to better understand the data you have available. You need to consider various data pre-processing techniques such as data transformation, discretisation, cleaning, normalisation, standardisation, smoothing, feature construction and use them if necessary.
- **Selecting features (individual 2).** Consider various techniques within each of filter, wrapper and embedded methods. You may also need to consider some dimensionality techniques (e.g. PCA) or feature construction techniques (e.g. Autoencoder or GANs). Findings from literature review may also be helpful.
- **Exploring and selecting ML algorithms (individual 3).** Select candidate algorithms. Establish baselines for model performance and start with a simple model using initial data pipeline. Discuss the selection strategies for the candidate algorithms.
- **Refining algorithms (individual 4).** Finding the best configuration for these hyperparameters in such a high dimensional space is not a trivial challenge. Consider the model design components (e.g. no of layers, no of units per layer, loss

function, activations, optimisers, dropout layer etc) as well as the hyperparameters (e.g. learning rate, dropout rate, batch size etc). Perform model-specific optimisations and Iteratively debug model as complexity is added. Discuss the selection strategies for searching for the best configuration (e.g. trial and error, grid search, random search, Bayesian optimisation etc).

- **Evaluating model and analysing the results (individual 5).** Evaluate the classification performance (e.g. accuracy, detection rate, false alarm, type II error, MCC and TBM (time has taken to build model) and TTM (time has taken to test model) – go beyond these measures if necessary) of the selected models on the test data and interpret the results. You will also need to compare the chosen model's performance with the benchmarks.
- **Future work (group).** The future work section is a place for you to explain where you think the results can lead you. What do you think are the next steps to take? What other questions do your results raise? Do you think certain paths seem to be more promising than others? This lets people know what you're thinking of doing next and they may ask to collaborate if your future research area crosses over theirs.

Deliverables Required and Submission Information

You must produce a report of 4,000 words (±10%) which includes 2 groups components (i. Planning and ii. Future work) and 5 individual components (i. Pre-processing, ii. Selecting features, iii. Exploring and selecting ML algorithms iv. Refining algorithms and v. Evaluating model and analysing the results). The cover page must show who is responsible for each individual component and the wordcount of your report. There should be some substantial tables and figures (make a good use of appendix) that help to cram all your information into the word count.

The report must be presented using either Arial 10 point or Times New Roman 11 point font for the main body of the text and 1.5 line spacing. Pages must have a minimum of 2.54 (1 inch) margins, i.e. MS Word 'normal' margins. IEEE referencing must be used, for guidance see: https://ieeeauthorcenter.ieee.org/wp-content/uploads/IEEE-Reference-Guide.pdf

The front page must contain the module title, assignment type, assignment title, names of group members, student numbers. The page number should be in the footer. You are strongly recommended to upload it in PDF format where possible, especially if including tables or figures. The submitted file title must be in the following format:

Student Number SurnameInitial ModuleName Assignment type e.g. 654321 BloggsJ AML GPP. Your code must also be submitted along the report (email the module leader). It should be either *.ipynb* or *.py*.

Estimated Time to Complete

There will be time that is allocated for working on your group project in Weeks 7 and 8. However, it is your responsibility to allocate an appropriate amount of time to this piece of work and to form a group.

Marking Scheme

Marks will be awarded in the following areas	%Weighting	Marking Descriptors					
		Excellent 80-100%	Very Good 70-79%	Good 60-69%	Satisfactory 50-59%	Poor 40-49%	Very Poor 0-39%
Planning (Group)	30%	Particularly mature analysis and judgement in planning showing flexibility, and high awareness of contingencies, efficiency and monitoring. Full consideration has been given to analytic process and real depth of insight of each step is shown.	Very efficient plans show a high level of ability to plan with flexibility, contingencies, and monitoring. Clear justification for estimates. Selection and application of analytic process is very effective, showing insight and creativity.	Plans are based on sound estimates and show reasonable awareness of the need for flexibility, contingencies, and monitoring. Appreciation of principles. Evidence of consideration of alternatives and judgement in decision.	Adequate analysis in planning, with consideration given to some of the wider issues. Some justifications for estimates. Some weaknesses in the process and can apply these reasonably well with fair justification.	Just sufficient analysis in planning. The most important wider issues were addressed but in limited ways. No justifications for estimates. The process are weak and little evidence is provided but just workable. Justification for decisions is limited.	Basic plans, with little evidence of their use. Little or no attempt to make appropriate analytic process decisions.
Preprocessing (Individual 1)	40%	Strong justifications. Full consideration has been given to each of preprocessing steps and real depth of insight of each step is shown.	Very good justification for selection. Selection and application of preprocessing techniques is very effective, showing insight and creativity.	Appreciation of principles. Evidence of consideration of alternatives and judgement in decision.	Some weaknesses in the process and can apply these reasonably well with fair justification.	The process are weak and little evidence is provided but just workable. Justification for decisions is limited.	Little or no attempt to apply appropriate preprocessing techniques.
Selecting features (Individual 2)	40%	Strong justifications. Full consideration of various techniques in three different categories. Strong evidence of wide reading and research.	Very good justifications for selection. Application of feature selection techniques are effective showing insight and creativity. Evidence of wider reading and research.	Good justifications. Appreciation of principles. Evidence of consideration of alternatives and judgement in decision. Evidence of wider reading and research.	Some weaknesses in the process and can apply these reasonably well with fair justification.	The process are weak and little evidence is provided but just workable. Justification for decisions is limited.	Little or no attempt to apply appropriate feature selection techniques.

Marks will be awarded in the following areas	%Weighting	Marking Descriptors					
		Excellent 80-100%	Very Good 70-79%	Good 60-69%	Satisfactory 50-59%	Poor 40-49%	Very Poor 0-39%
Exploring and selecting ML algorithms (individual 3).	40%	Full consideration has been given to options for algorithms. Entirely appropriate choices made and expertly applied. Provision of baselines. Strong selection strategies.	Very good selection of algorithms and procedures - competently applied. Justification for selection is sound. Very good selection strategies. Evidence of wider reading.	Good judgement in selection and application of algorithm and procedures. Provision of baseline. Evidence of wider reading.	Reasonable selection and application of algorithm and procedures.	Poor but acceptable selection. Decisions may be based on student's convenience.	Appropriate analysis or judgement severely lacking or not provided.
Refining algorithms (individual 4).	40%	Real depth of insight is shown in final choices of both parameters and design, demonstrating unusual insight and very effective.	Selection of parameters and design is very effective, showing insight and creativity. Justification for selection is sound.	Selected parameters are appropriate to the objectives. Good justification for selection.	Reasonable justification for selection showing some awareness of appropriate principles.	Some of the selected parameters are appropriate, with limited justification.	An insufficient awareness of principles with very weak or no justification.
Evaluating model and analyzing the results (individual 5)	40%	Strong justification on evaluation methods. Selected models are fairly evaluated. All evaluation measures are correctly calculated. Evidence of wider reading and choice of extra measures. Quality of interpretation is very high and is based on several insightfully chosen sources.	Good justification on evaluation methods. Selected models are fairly evaluated. All evaluation measures are correctly calculated. Evidence of wider reading. Interpretation is very good and is based on comparison with some well-chosen sources, but some points could be developed.	Selected evaluation methods are appropriate to the objectives. Good justification for selection. All evaluations are provided. Interpretation is good, and based on comparison of some relevant sources, but the sources used could be extended.	Reasonable justification for selection showing some awareness of appropriate principles. Some calculations are incorrect or unfairly evaluated. Interpretation is reasonable and very few sources are used.	Some of the selected method are appropriate, with limited justification. Some calculations are incorrect or unfairly evaluated. Little interpretation on the experimental results. Lacks depth. Poor results.	An insufficient awareness of principles with very weak or no justification. Incorrect calculations. Unfair evaluation. No or trivial interpretation on the results.

Marks will be awarded in the following areas	%Weighting	Marking Descriptors					
		Excellent 80-100%	Very Good 70-79%	Good 60-69%	Satisfactory 50-59%	Poor 40-49%	Very Poor 0-39%
Future work (group)	15%	Very effectively addressing the questions. Strong evidence for future work.	Effectively addressing the questions. Good evidence for future work.	The questions are answered. Reasonable evidence for future work.	Some of the questions are answered. Limited evidence of future work.	Some of the questions are answered. Lack of evidence for future work.	No answer to the questions. No future work.
Report Documentation (group)	15%	Organisation of work is of a very high standard, likely to be highly stimulating, and at the limits of what may be expected at postgraduate level. Work is of a standard publishable in a refereed journal.	Documentation is very well ordered, concise and coherent. Excellent use of appendices and illustrations.	Organisation of work is likely to show few mistakes/limitations. Very good use of appendices and illustrations.	There is an overall structure evident but does not offer strong flow and progression. Appendices and illustrations mainly used well.	Structure of work is weak or inconsistent. Only the main points are logically organised/linked. Quite good use of appendices and illustrations.	Unstructured and/or incoherent. Illustrations are very poorly presented. Appendices are very poorly presented.