International Institute of Information Technology Hyderabad

System and Network Security (CS5470)

Lab Assignment 5: Develop a network intrusion detection system
Implement IDS with machine learning
Hard Deadline: **April 8, 2021 (23:55 P.M.)**
Total Marks: 100

Note1:- *Number beside the question denotes the number allocated to that question. Full marks is 100. You may either get 0 or full marks for each part of the question.No partial marks are there.Good luck*

Note2:- *It is strongly recommended that no student is allowed to copy programs from others. Hence, if there is any duplicate in the assignment, simply both the parties will be given zero marks without any compromise. Rest of assignments will not be evaluated further and assignment marks will not be considered towards final grading in the course. No assignment will be taken after deadline. Please upload in code along with a README file in the course moodle portal through a ZIP file (Groupnumber-Lab4.zip).*

IDS data set overview

- use the full.csv as your train set
- use the test.csv as your test set

feature name	*description*	*type*
duration	length (number of seconds) of the connection	continuous
protocol_type	type of the protocol, e.g. tcp, udp, etc.	discrete
service	network service on the destination, e.g., http, telnet, etc.	discrete
src_bytes	number of data bytes from source to destination	continuous
dst_bytes	number of data bytes from destination to source	continuous
flag	normal or error status of the connection	discrete
land	1 if connection is from/to the same host/port; 0 otherwise	discrete
wrong_fragment	number of ``wrong'' fragments	continuous
urgent	number of urgent packets	continuous

Table 1: Basic features of individual TCP connections.

feature name	*description*	*type*
hot	number of ``hot'' indicators	continuous
num_failed_logins	number of failed login attempts	continuous
logged_in	1 if successfully logged in; 0 otherwise	discrete
num_compromised	number of ``compromised'' conditions	continuous
root_shell	1 if root shell is obtained; 0 otherwise	discrete
su_attempted	1 if ``su root'' command attempted; 0 otherwise	discrete
num_root	number of ``root'' accesses	continuous
num_file_creations	number of file creation operations	continuous
num_shells	number of shell prompts	continuous
num_access_files	number of operations on access control files	continuous
num_outbound_cmds	number of outbound commands in an ftp session	continuous
is_hot_login	1 if the login belongs to the ``hot'' list; 0 otherwise	discrete
is_guest_login	1 if the login is a ``guest''login; 0 otherwise	discrete

Table 2: Content features within a connection suggested by domain knowledge.

feature name	*description*	*type*
count	number of connections to the same host as the current connection in the past two seconds	continuous
	Note: The following features refer to these same-host connections.	
serror_rate	% of connections that have ``SYN'' errors	continuous
rerror_rate	% of connections that have ``REJ'' errors	continuous
same_srv_rate	% of connections to the same service	continuous
diff_srv_rate	% of connections to different services	continuous
srv_count	number of connections to the same service as the current connection in the past two seconds	continuous
	Note: The following features refer to these same-service connections.	
srv_serror_rate	% of connections that have ``SYN'' errors	continuous
srv_rerror_rate	% of connections that have ``REJ'' errors	continuous
srv_diff_host_rate	% of connections to different hosts	continuous

Table 3: Traffic features computed using a two-second time window.

Questions

- use at least 3 different ML algorithm and compare their performance 20
- Performance metrics should include precision,recall,f1 score, accuracy. 30
- Now use your best algorithm to predict the label of test.csv 50
- submit full report in pdf(convert ipynb to pdf) , group contribution report along with predicted labels in a csv file named testLabel.csv which will contain one column i.e predicted label.
- your folder should contain 3 files - report.pdf , group-contri.pdf , testLabel.csv.
- please be sure that your submission follow the above points else you may get less marks as it will evaluated automatically

All the best!