MANIPAL INSTITUTE OF TECHNOLOGY
MANIPAL
(A constituent unit of MAHE, Manipal)

II SEMESTER M.TECH. (CSIS) (DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING) END SEMESTER EXAMINATIONS, MAY 2024

SUBJECT: AI AND ML TECHNIQUES IN CYBER SECURITY (CSE - 5416)

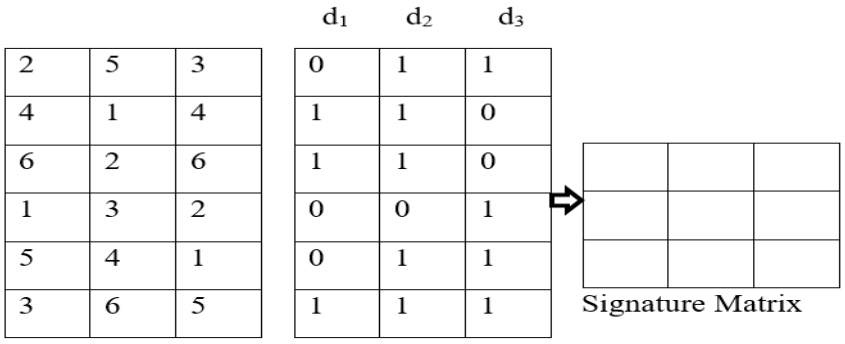
REVISED CREDIT SYSTEM

(XX/XX/2024)

Time: 9:30 am to 12:30 pm MAX.MARKS: 50

INSTRUCTIONS TO CANDIDATES:-

- Answer **ALL** the questions.
- Missing data may be suitable assumed.

		Marks
1A.	Spam detection is an example of pattern recognition because spam has a largely predictable set of characteristics, and an algorithm trained to recognize those characteristics as a pattern by which to classify emails. Is it possible that spam detection problem can also be an anomaly detection problem. Justify. Also Illustrate some other applications that clearly fall in the category of pattern recognition.	5M
1B.	Identify the algorithm which detects anomalies by fitting the SVM with data belonging to only a single class.	3M
1C.	List and explain the drawbacks of k-nearest neighbor model.	2M
2A.	Given the following shingling matrix and permutations for some documents (d_1,d_2,d_3):  1) Complete the corresponding signature matrix by Min-Hashing. 2) Compute the Jaccard similarities between documents. 3) To produce clusters of similar items, one must find groups of signatures that overlap in many places. Discuss the two ways to do this.	5M
2B.	Discuss the algorithm that divides data sets up into subgroups of high-density regions and the number of clusters is not operator defined but instead inferred from the data.	3M

2C.	Outline the number of techniques to address the feature selection problem.	2M
3A.	Generalize some considerations that data scientists use to improve data collection.	4M
3B.	Depict and discuss a typical malware attack flow.	4M
3C.	Using feature engineering, analyze and write the features to detect different kinds of attacks on web application.	2M
4A.	Suppose that you have concluded from your data that if more than 20 new accounts are created from the same IP address in the same hour, these accounts are certain to be fake. Scoring at account creation time, can count creation attempts per IP address in the past hour and block if the counter is greater than 20. For any attack, there will still be 20 fake accounts that got through and are free to send spam. If you score newly created accounts once per hour and take down any group of more than 20 from the same IP address, you will block all the spammers, giving them each one hour to wreak havoc. Clearly a robust approach combines instances of both techniques. Analyze and illustrate with valid answer.	4M
4B.	An anomaly detection system that raises too many false positive alerts to security operations personnel should take advantage of the correct labels given by human experts during the alert triaging phase to retrain and improve the model. Analyze the given example and describe and depict it with suitable justification.	3M
4C.	List and discuss the different approaches to speed up machine learning applications for performance bottlenecks in the program execution framework, find more efficient algorithms or using parallelism.	3M
5A.	Briefly outline the concept of model poisoning and evasion attack.	5M
5B.	"Even perfect learners can display vulnerabilities because the Bayes error rate might be a non-zero". Justify with a suitable answer.	3M
5C.	Discuss the working of botnets.	2M