**Mid-Term**
**Deadline: 20th March 2022, 11:59 PM**
**Total Marks: 40**

Q1. Discuss the working of Anti-virus software and state its limitations. **(5 marks)**

Q2. Discuss pros and cons of different types of malware detection methodologies. **(5 marks)**

Q3. Based on your experience, discuss some of the challenges in effectively using Machine learning based solutions for spam and malware detection. **(5 marks)**

Q4. Why it is important to examine the strings in malware detection. Discuss the process of examining malware strings for windows. **(5 marks)**

Q5. Preface: Spam email is unsolicited and unwanted junk email sent out in bulk to an indiscriminate recipient list. Typically, spam is sent for commercial purposes. It can be sent in massive volume by botnets, networks of infected computers – CISCO.

Spam detection is one of the oldest computer security challenges and the one which remained relevant throughout the history of E-mail communication and shall continue to do so in future as well. Machine Learning has shown great promise in addressing spam detection and is widely used for this purpose.

**Problem Statement**

1. Download the SPAMBASE data set using the following link: https://archive.ics.uci.edu/ml/datasets/spambase
2. Read the documentation of the database and devise a spam detection system using Naïve Bayes Classifier. You are free to use any Naïve Bayes model of your choice to achieve better results.
3. Report accuracy, precision, recall, confusion matrix and F-1 Score.
4. Discuss steps you have taken to improve your result. Justify, why your result is an acceptable solution for the given problem.

**(20 marks)**

**Note:**

**Submit the answer sheet and complete codes via moodle only.**

**Please answer all questions in your own words. Do not copy paste.**

**Plagiarism/cheating will lead to failing grades.**