**Microsoft Malware Prediction**

**Team Members(Team-2)**

1) Amit Subedi

2)Bhagavath Sai Darapureddy

3) Agarwal Raghav

4) Sai Srinivas Lakkoju

**Introduction**

**Our proposed project aims to predict the probability of Windows machines becoming infected by malware by analysing telemetry data, including various machine properties and infection records, from Microsoft's endpoint protection solution, Windows Defender.**

**Data Set Information**

The dataset is comprised of a multitude of rows, each representing a unique machine, identified by its "MachineIdentifier." The "HasDetections" column provides the ground truth, indicating whether malware was detected on the machine. The data is sourced from both heartbeat and threat reports collected by Windows Defender. The nature of this data offers a rich source of information for the development of our predictive model.

**SMART Questions**

1. What are the key features in the dataset that can be used to predict malware infections?
2. How does this project's outcome contribute to enhancing cybersecurity efforts?

**Dataset Source**

The "[Microsoft Malware Prediction](https://www.kaggle.com/competitions/microsoft-malware-prediction/overview)" dataset is a prominent data source available on Kaggle ([Link](https://www.kaggle.com/competitions/microsoft-malware-prediction/data)), a popular online platform for data science and machine learning enthusiasts. Kaggle hosts a wide range of datasets and competitions, including this one, which focuses on predicting the presence of malware in software based on various features and characteristics.

**Github Repository:**

[Saisrinivas783/Data\_mining\_Project\_Team\_2 (github.com)](https://github.com/Saisrinivas783/Data_mining_Project_Team_2)