Versione 1

21/06/2024

Relazione di Progetto:

ARCHITETTURA DATI

# Conseguenze dello sporcare un DataSet sui modelli predittivi

Presented by: cavallini Francesco

Matricola: 920835

UNIVERSITÀ MILANO BICOCCA

## Sommario:

[Conseguenze dello sporcare un DataSet sui modelli predittivi 0](#_Toc9733871)

[Sommario: 2](#_Toc743447526)

[1. Introduzione: goodware vs malaware 3](#_Toc454045560)

## Introduzione: goodware vs malaware

### 1.1. Premessa: Scelta del tema:

Nel panorama attuale, la minaccia informatica è in costante aumento e diversificazione. I malware, software dannosi progettati per danneggiare o rubare informazioni dai sistemi informatici, rappresentano una seria preoccupazione per individui e organizzazioni. La capacità di distinguere tra goodware, software legittimo e sicuro, e malware è fondamentale per la sicurezza informatica. Si decide quidni di proporre un progetto che rispecchia sia tematiche attuali che, banalmente, di puro interesse personale.

Per sviluppare il progetto si vuoule quindi leggere un dataset di applicazioni android. Si parte dunque da un dataset con le seguenti feature:

* colonna 1 214: Permission-based features   
  sono tutte feature binarie, dove 0=permesso non richiesto, 1=permesso richiesto
* colonna 215 241: API based features -> sono tutte feature binarie 0=api call non richiesta, 1=api call richiesta
* - colonna 242: label (target), dove le classi sono:
* - Malware
* - Goodware

Si vogliono sviluppare i seguenti 2 punti:

1. dimostrare che è possibile predirre correttamente se un applicazione è un malware o goodware in base ai permessi e le api call
2. verificare come cambiano le performance sporcando i dati all'interno del dataset