**The Problem of False Alarms**

**Background**

Cybersecurity monitoring of real-time transactions has always been a troublesome effort because of the very large volume of alarms created by security devices and applications. Many false alarms are part of this large volume.

Many alarms are ignored by security analysts during inspection due to time and volume constraints. This is partly due to a phenomenon called “alarm fatigue”.

This disregard of complete scrutiny of the alarm is many times intentional since there is not enough time to analyze, and act upon every alarm within the allotted security level agreement time span.

There have been organizations that have suffered negative consequences as a result of not identifying a real alarm that was among the many false positive alarms.

This challenge keeps increasing with the use of Big Data, given the faster velocity, larger volume, and variety of data streams, that need to be ingested, or disbursed, in real, or quasi real-time.

**How can we make the process of the manual verification of alarms more efficient?**

**Project Description**

Your customer, the deputy security director from the CISO (Chief Information Security Office) of your company, has asked you to come up with a recommendation to make the analysis of the alarms from a newly installed Data Loss Detection application more efficient. She is overwhelmed by the sheer size of alarms being generated as databases are suspiciously accessed 24 by 7, online and in batch mode.

Your customer is interested in improving the process to confirm any logged records that are true alarms and prioritize them ahead of the false ones.

She believes that if her security analysts had to focus on inspecting a confirmed and prioritized list of true alarms, they could do their inspection more thoroughly and be able to provide pertinent action on a timely basis.

You and your team will look to understand the nature of these output records from the DLD in order to verify them as true or false alarms.

In other words you will create an automated verification process of false alarms by training, validating and deploying a data mining model using a machine learning tool, such as RapidMiner, etc.

The data set containing the log of alarm records will be provided to your team for EDA in a separate Blackboard folder.

The project will require three team presentations of each phase of the project

1. EDA status
2. Training and validation status–
3. Final presentation. –

1 - **Exploratory Data Analysis phase.**

The team will analyze the data and fully understand it based on the EDA.

1. The team will use any software that can help in the EDA phase. Refer to examples in Session 5.
2. Each member will choose at least one data variable in the file and provide a detailed analysis of it in a separate slides.
3. The team will explore and recommend at least three possible machine learning models that could be used in the second and third phases to provide a solution based on the results of the EDA.
4. Each team will create a PowerPoint deck to present the EDA findings.
5. Each PowerPoint slide will have the name of the member or members that will be describing the work in the particular slide.
6. The presentation deck must have a slide for the introduction of the team members at the beginning of the deck and a summary slide at the end of the deck with the EDA findings and model recommendations.