**Abstract:**

In data center environments, hard disk drive (HDD) failures are a rare but costly occurrence. Therefore, HDD vendors are highly motivated to reduce the rate of failures as a cost saving measure. But currently, HDD manufacturers use Self-Monitoring and Reporting Technology (SMART) attributes collected during normal operations to predict failures. If a certain attribute considered critical to HDD health goes above its threshold value, the HDD is marked as likely to fail. Our project focuses on applying machine learning to improve prediction accuracy in hard disk drives. The goal of our project is, to analyze which of our subset of machine learning models is best suited towards predicting failure of HDDs. We analyze three different algorithms, Support Vector Machine (SVM), Naive Bayes and Random Forest, to see which has the highest accuracy, recall and precision when predicting HDD failures.

**Problem Statement**

Modern hard drives are reliable devices, yet failures can be costly to users and many would benefit from a warning of potential problems that would give them enough time to backup their data. The problem can be characterized as one of detecting rare events from a time series of noisy and non parametrically-distributed attributes. Hard drive manufacturers have been developing self-monitoring technology in their products, in an effort to predict failures early enough to allow users to backup their data. This Self-Monitoring and Reporting Technology (SMART) system uses attributes collected during normal operation (and during off-line tests) to set a failure prediction flag. SMART flag is a one-bit signal that can be read by operating systems and third-party software to warn users of impending drive failure.

**Proposed System**

Improving the performance of hard drive failure prediction will have many practical benefits. We compare machine learning methods applied to a difficult real-world problem: predicting computer hard-drive failure using attributes monitored internally by individual drives. This is achieved by training multiple classification modules with different prediction horizons to provide different performance trade-offs in terms of accuracy. We analyze three different algorithms, Support Vector Machine (SVM), Naive Bayes and Random Forest, to see which has the highest accuracy, recall and precision when predicting HDD failures.

**Scope**

Scope of the project listed following,

* Preparation of the best prediction algorithm to detect of hard drive failure prediction.
* We can reduce the amount of failures by taking the real time parameters.
* We can also build the hybrid classification algorithm for more accurace.

**SYSTEM CONFIGURATION:**

**Hardware requirements:**

Processer : Any Update Processer

Ram : Min 4 GB

Hard Disk : Min 100 GB

**Software requirements:**

Operating System : Windows family

Technology : Python 3.6

Front End : Pyqt5

IDE : PyCharm