**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.

**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