

# BALANCEAMENTO DE DADOS EM PROBLEMA DE PREDIÇÃO DE FALHA EM DISCOS RÍGIDOS

XXXVIII Encontro de Iniciação Científica

Davi Farias Martins Torres, Joao Paulo Pordeus Gomes

A capacidade de prever falhas em discos rígidos tornou-se um tópico bastante importante visto que falhas inesperadas podem acarretar em significativas perdas de dados e maior tempo de serviço fora do ar. Atualmente, a maioria dos discos rígidos está equipada com um sistema de monitoramento de falhas denominado Self-Monitoring, Analysis and Reporting Technology (SMART). Este sistema coleta diversos parâmetros relativos ao desempenho do equipamento. Apesar da grande popularidade do sistema SMART, estudos recentes mostram que somente 3% a 10% das falhas são identificadas previamente. Com isso muitos trabalhos foram desenvolvidos, mas nenhum deles trata o problema de desbalanceamento de classes através de balanceamento por sobreamostragem de dados. Este trabalho analisa o impacto causado por dados sintéticos gerados por redes generativas adversarias na tarefa de predição de falhas em disco rígido. Com o intuito de corrigir problemas gerados por desbalanceamento de dados. Para demonstração da efetividade do método o mesmo foi comparado com outros três métodos de sobreamostragem em um conjunto contendo dados de 2051 discos rígidos. Os resultados apresentados pela classificação com o método de balanceamento por redes generativas adversarias se mostraram promissores, o número de falhas de discos rígidos detectadas foi maior do que o obtido usando a classificação com o conjunto de dados original enquanto manteve uma acurácia próxima do mesmo.

Palavras-chave: Desbalanceamento de dados. Discos Rígidos. Detecção de falha. Dados sintéticos.