

The Data Science Institute's high-performance computing (HPC) cluster supports many students and researchers who submit jobs daily for data analysis, modeling, and machine learning. When resource usage spikes unexpectedly, these jobs can be delayed, which can waste resources like time and compute power. To improve the efficiency of the cluster's resource allocation, the HPC Cluster Team set out to predict whether usage will become extreme in the next hour.

Extreme usage was defined as exceeding the 85th percentile of historical usage data. The team created models that analyze past patterns in activity, time-based features (like day of the week), and job-related metrics such as memory usage. Several machine learning models were tested, including K-Nearest Neighbors, SVM, AdaBoost, Logistic Regression, and XGBoost.

XGBoost achieved the highest performance, with an F1 score of 0.885, accurately forecasting high-utilization periods. In parallel, the team also built interactive dashboard pages to interpret the cluster's log data from files to support deeper analysis, and explore usage patterns. These tools can offer cluster administrators better visibility and foresight, enabling faster job processing and more efficient resource use for the DSI community.

Model Performance Comparison (Test Set)

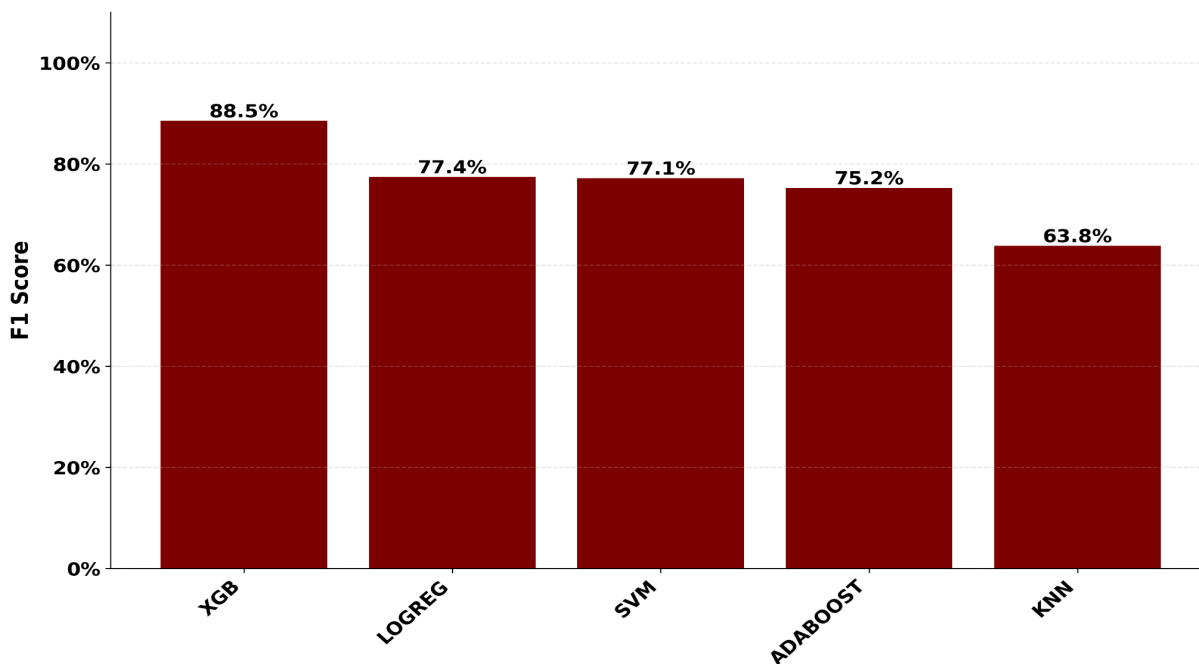


Figure 1: F1 Score Comparison Across Models