CAPPA INTERNSHIP

SUMMER 2024

PROJECT: FloDX

NAME: Ashik Rahman

ID: R00212471

ONE-DRIVE:

TABLE OF CONTENTS

1. Introduction
2. Task 1:

* Sensor Variance
* Finding the Optimal Wavelength

1. Task 2

* Design and Creation of Xgboost Model
* Initial Model design and testing
* Observing variance within data set

1. Task 3

* Model testing

1. **Introduction**

The monitoring of bacterial growth is a fundamental process in microbiological research and various industrial applications. However, traditional methods for bacterial growth assessment are often laborious, time-consuming, and prone to contamination due to frequent sampling. These methods, typically involving spectrophotometric analysis, not only require significant manual intervention but also generate substantial waste, presenting both environmental and economic challenges.

FloDx, an innovative ultraportable wireless device, offers a groundbreaking solution to these issues. Designed for continuous monitoring of bacterial growth, FloDx eliminates the need for repetitive sampling, thereby reducing the risk of contamination and minimizing waste. The device's efficiency translates to significant savings in both time and cost, potentially reducing labour by up to 10,000 hours and cutting expenses by $2,000,000 annually—a tenfold reduction compared to conventional spectrophotometer-based methods.

This report details the development and implementation of a sophisticated growth model algorithm for FloDx, undertaken as part of an internship project. The primary objective is to enhance the predictive accuracy of bacterial growth monitoring, ensuring reliable and precise data output. Additionally, this research explores further optimization strategies for the prediction factors, aiming to maximize the operational efficiency and overall performance of the FloDx system.

1. **A black device with red caps

   Description automatically generatedFundamentals Of Xgboost**

The development of advanced machine learning models has significantly enhanced the capabilities of predictive algorithms, particularly in the field of artificial intelligence. XGBoost represents a major advancement over traditional models by addressing several limitations inherent in earlier methods. The primary motivation behind XGBoost's creation was the need for a scalable and efficient boosting model capable of handling large datasets with greater speed and accuracy.

At the core of XGBoost is the decision tree, a foundational algorithm used for classification and regression tasks. Decision trees split data into smaller subsets based on input features, forming a tree-like structure where each branch represents a decision rule, and each leaf represents an outcome. Although decision trees are useful, they are often prone to overfitting, where the model performs well on training data but poorly on unseen data.

To overcome these limitations, ensemble methods like Random Forest and AdaBoost were introduced. Random Forest builds multiple decision trees using different subsets of data, aggregating their predictions to improve accuracy and reduce overfitting. AdaBoost, on the other hand, combines weak learners into a strong learner by focusing on the errors of previous models, though it uses simpler tree structures known as stumps.

Gradient Boosting further refined these concepts by using gradient descent to optimize model performance. It builds trees sequentially, where each tree corrects the errors of its predecessors. Despite its improvements, Gradient Boosting models can still suffer from overfitting if not properly tuned.

XGBoost enhances the Gradient Boosting framework by incorporating features like regularization and advanced tree pruning techniques to combat overfitting and improve model performance. These innovations make XGBoost a powerful and flexible tool for various machine learning tasks, offering significant advantages in terms of accuracy and computational efficiency. This paper explores the implementation of XGBoost in the context of bacterial growth prediction, aiming to leverage its capabilities to advance monitoring technologies and improve predictive accuracy.