

# Midterm Project Report

Huanyu Chen, Yifei Liu, Longyu Zhang

## Abstract

This study examines the predictive performance of various data science models in estimating recovery time from COVID-19 based on demographic and health-related factors. The dataset includes variables such as age, BMI, and smoking status. The techniques used include LASSO regression, PLS, MARS, and GAM. The results show that LASSO and PLS models had similar predictive accuracies and MARS outperformed GAM. The importance of considering both linear and nonlinear relationships when modeling recovery time is underscored by the findings. This has implications for clinical decision-making and resource allocation in COVID-19 management.

## 1 Exploratory Analysis and Data Visualization

## 2 Model Training

## 3 Result

## 4 Conclusion

## 5 Appendix