Forecasting ICU Admissions for COVID-19 Patients in Brazil: A Data-Driven Approach

Introduction

This project uses machine learning to predict ICU admissions for COVID-19 patients, improving resource allocation and healthcare services in Brazil.

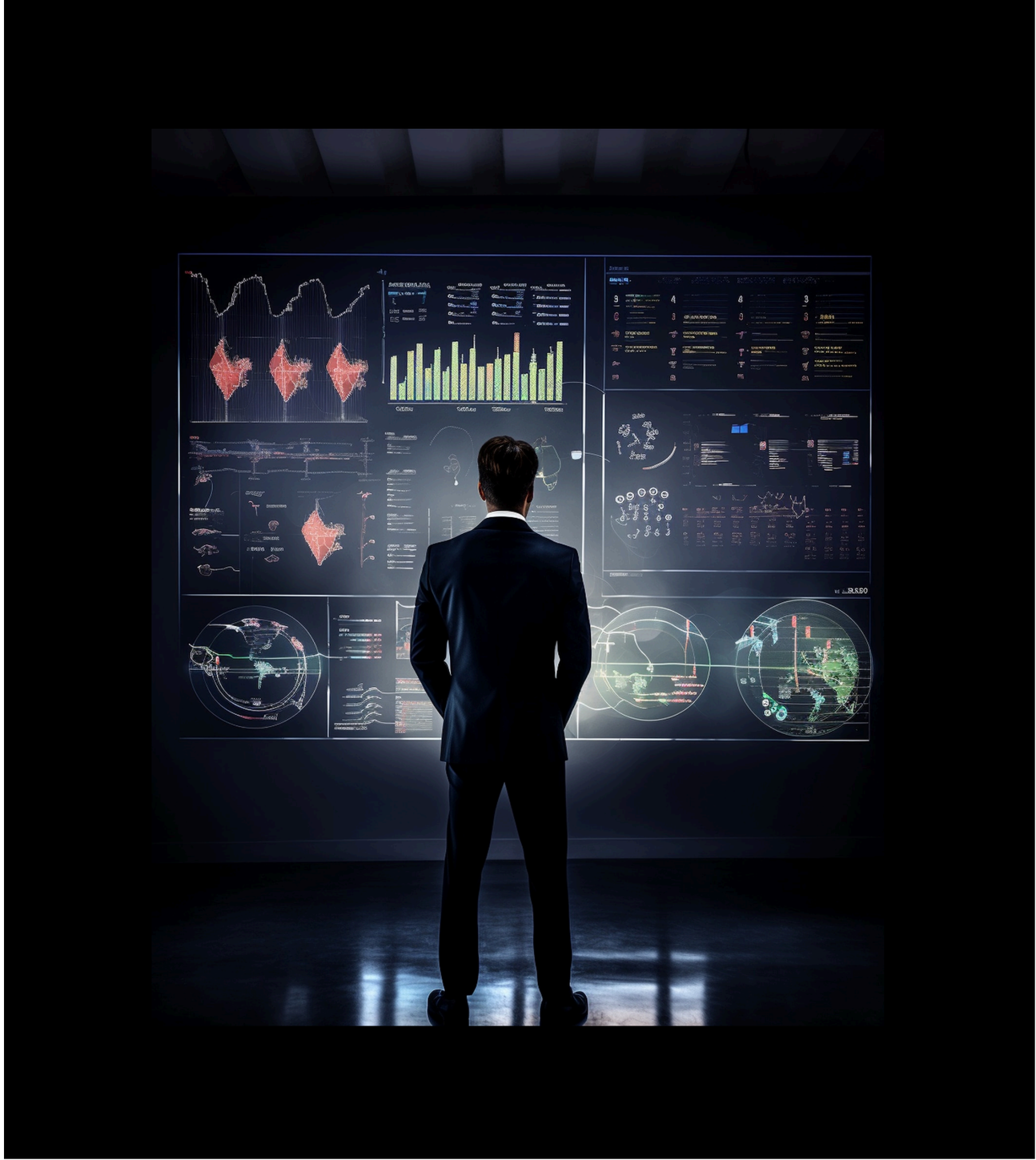
Data Sources

the data used in this project was obtained from the Sírio-Libanês Hospital's Kaggle dataset.

click here to access the dataset

Models

1. Logistic regression
2. Random Forest Classifier
3. Decision Tree Classifier



Results and Insights

Logistic Regression
Accuracy before optimization: 80%
Accuracy after optimization: 81%

Class 0 (non-ICU patients)
precision: 82%
Recall: 93%

class 1 (ICU patients)
precision: 71%
recall: 47%

Results and Insights Cont...

Random Forest
Accuracy: 81%

Class 0 (non-ICU patients)
precision: 89%
Recall: 85%

class 1 (ICU patients)
precision: 65%
recall: 72%

Results and Insights Cont...

Decision Tree
Accuracy before optimization: 75%
Accuracy after optimization: 80%

Class 0 (non-ICU patients)
precision: 82%
Recall: 85%

class 1 (ICU patients)
precision: 56%
recall: 50%

Conclusion

we now have working models to help the healthcare system use its resources better. Random Forrest proves to work the best with 81% accuracy.

Recommendations for future work

1. class balancing techniques such as SMOTE
2. Feature Engineering
3. Ensemble Techniques
4. Adding time stamps to the window column to better integrate it into the models

Thanks!

Author: Ousman Njie