Predictive Analysis of Oxygen Concentrator Performance

PURDUE UNIVERSITY | The Data Mine

Max Chen, Kestan Kamei, Irfan Firosh, Rohan Sangameswaran , Tanay Sahasrabudhe, Sanjana Tatavarthi, Billy Chen, Harshvardhan Sharma

Inogen

Introduction

High Level Overview:
Our corporate mentors have assigned us the task of analyzing a large dataset from Portable Oxygen Concentrators (POC) to develop a predictive model for future metrics.

About Inogen:
Inogen is a medical technology company that specializes in Portable Oxygen Concentrators (POC) for home use. Their products improve quality of life for people around the world dealing with chronic respiratory conditions.

Specific Goal:
Our goal is to use accurate and efficient ML models to analyze data and predict future oxygen concentration levels of POC's to allow Inogen to predict when they need to repair or replace devices, greatly increasing cost efficiency as well as ensuring that the patient always has medical grade oxygen.

What is a POC:
An Inogen Portable Oxygen Concentrator is a compact, mobile device that supplies purified oxygen from the air, designed for convenient, on-the-go oxygen therapy.

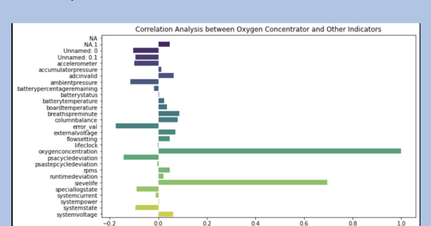
Methodology

Random Forest: Scikit learn
LSTM: TensorFlow and Keras
Programming language: Python
Data manipulation and visualization: Pandas and Matplotlib

Data Cleaning

Why did we clean the data?
- NA values that were inconsistent, could possibly hinder accuracy of the models
- Abnormal data beyond acceptable range

Steps:
- Removed NA values in the database
- Filtered data between appropriate ranges as specified by the handout
- Attempted to find most common causes of POC errors

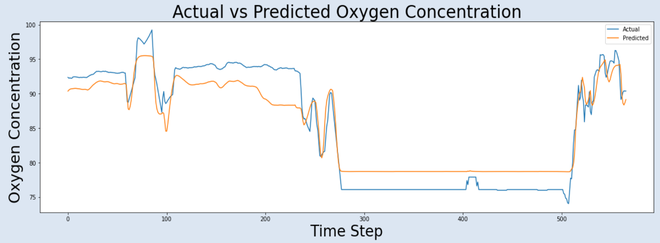


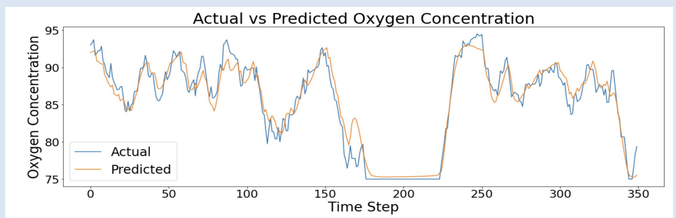
LSTM

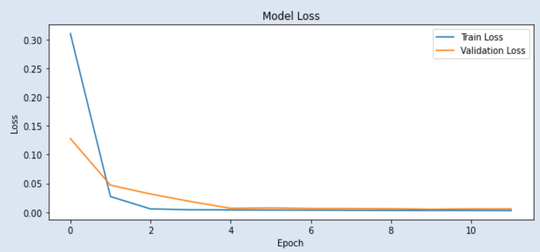
Introduction on LSTM
- A Recurrent Neural Network designed for long-term sequence data dependencies.

Strengths of the Method:
- Captures long-term time-series patterns.
- Uses past information for better predictions.
- Predicts equipment failures from historical data.
- Manages complex, interrelated variables.

Implementation Highlights:
- Divided data for O_2 concentration prediction.
- Applied rolling mean to minimize variability.
- Explored different LSTM architectures.
- Tested on validation data.







Example model loss graph

Results:
- Achieved high R^2 score of 94% for one machine
- Achieved high R^2 score of 92.09% for one machine
- Predicted O_2 concentration values closely following trend of actual values

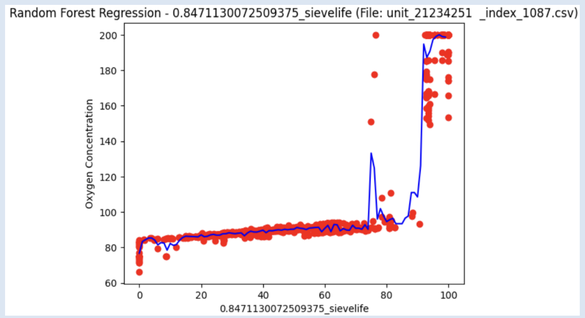
Impact:
- Predicting O_2 concentration can help Inogen estimate when a machine will stop producing medical-grade oxygen.
- Once fully implemented, it will streamline maintenance, reducing costs for the company and enhancing customer convenience.

Random Forest

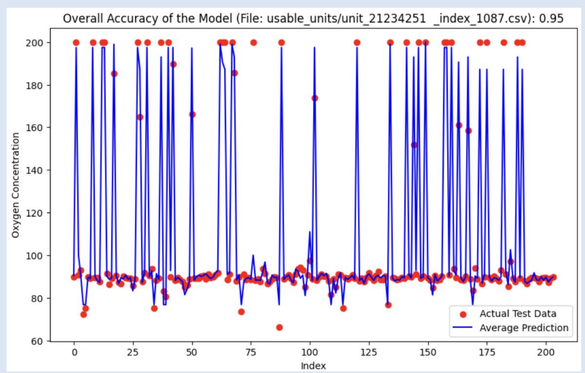
Introduction on Random Forest
- Utilizes collective decision-making of multiple decision trees for enhanced accuracy and stability (less overfitting).

Strengths of the Method:
- Excels in handling complex, multifaceted datasets.
- Offers validation mechanism through error estimation.

Implementation Highlights:
- Aimed to predict oxygen levels using diverse input features.
- Applied `RandomForestRegressor` from scikit-learn library
- Delineated data into 80% for training to capture patterns, 20% for testing to validate predictions.
- Determined model accuracy using R^2 score



Plotting Oxygen Concentration vs Sieve Life



Plotting Oxygen Concentration vs Index (using highly correlated variables)

Results
- Average overall accuracy (R^2) across different datasets: 81%
- This high accuracy is essential for Inogen's maintenance scheduling and product longevity assurance.

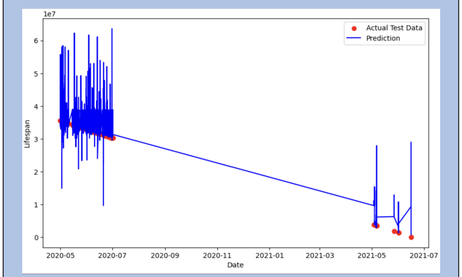
Impact
- Useful for finding relations between different types of collected data to provide a better understanding patient machine use.
- However, random forest proved to be less useful for prediction of future values, as seen in the visualization in the future plans section.

Conclusion

- With our models, Inogen can now anticipate when components in a POC may fail.
- Knowing the time of failure can speed up the maintenance process of a POC
- Leveraging these predictions to perform maintenance will allow patients to have necessary medical grade oxygen more consistently

Future Plans

In the future, we plan to:
- Tune and refine Random Forest models for more well-defined results.
- Decreasing mean absolute error in the following visualization for the following model, a random forest model intended to predict device maintenance periods.



- Tune hyper parameters on the LSTM model increase accuracy scores and streamline results.
- Develop the VAR model further and check its effectiveness on the dataset compared with other models.

Acknowledgements

Data Mine Staff:
- Nicolas Lenfestey
- David Glass
- Joshua Robert
- Maggie Betz

Inogen Mentors:
- Norbert Leinfellner
- Horst Pichler

TAs:
- Stanley Kim
- Anant Mishra