EC Utbildning Data Science

October 2024

Predictive Maintenance Project Overview

Predictive Maintenance Project

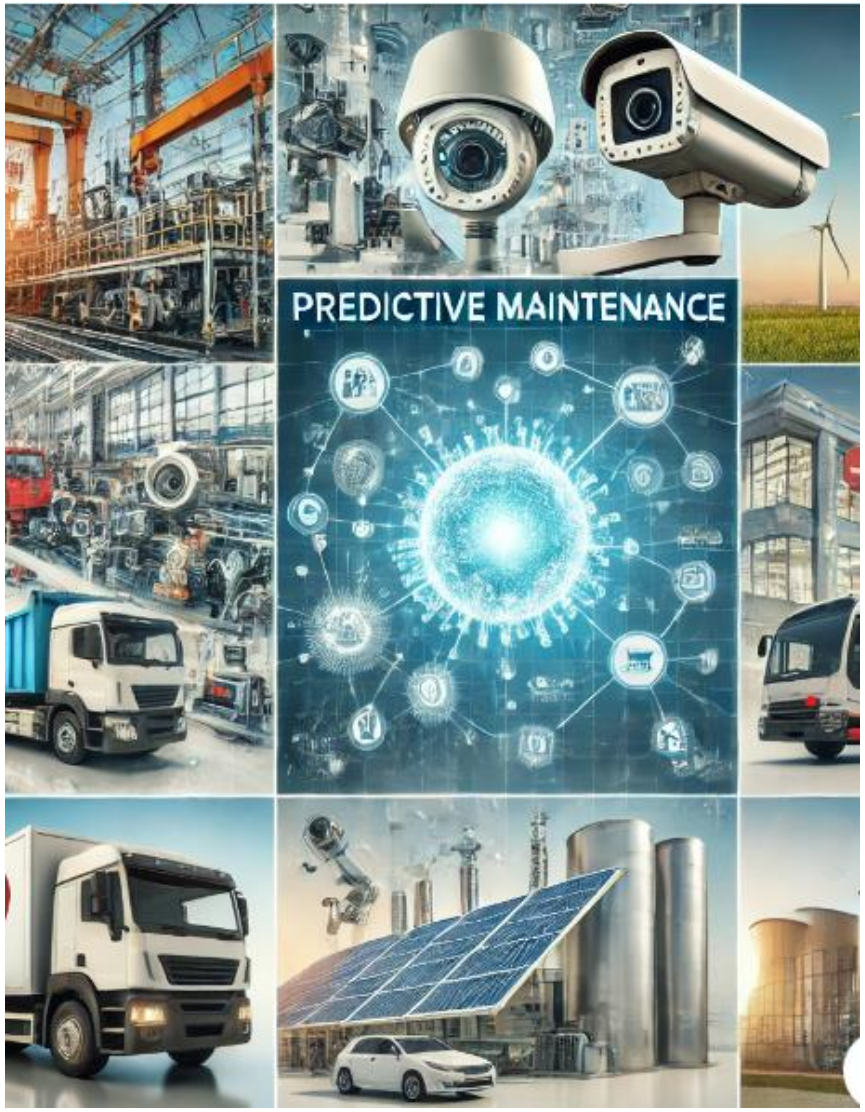
OVERVIEW

This project aims to develop a predictive maintenance system using the Predictive Maintenance: Time-Series Forecasting dataset from Kaggle, progressing from data storage in SQL to machine learning model development and visualization.

Predictive Maintenance and Its Impact Across Industries

- **Manufacturing**
- **Security**
- **Healthcare**

- **Cost Efficiency**: Reduces maintenance costs by 10–30% annually.
- **Operational Uptime**: Cuts unplanned downtime by up to 50%.
- **Annual Savings**: Can save companies **€1M–€5M** per facility.



Role of Time Series Analysis

Forecasting: Predict future outcomes to enable proactive planning.

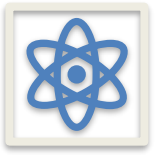
Anomaly Detection: Identify unusual patterns to respond quickly (e.g., machine spikes)

Collaborative Model Development

Define Goals:
Failure or RUL:

Data Prep:
Modules

Agile EDA:
Understand &
Redefine



Model Build:
From Linear
Light to
Nuclear Neural

Parallel push to
Deployment
Deadline

Agile Methodology & Scrum Roles

Scrum Master
Robert Shaw

Product Owners
Missi Hansson & Jakob Rask

Dev Team:
Jakob Rask,
Missy Hansson
& Robert Shaw

Tools

- Kanban Board: Trello for task management.
- Version Control: Git for code teamwork and tracking.

Scrum-based Timeline

Sprint 1: Setup & Data Acquisition – Set up the project and pull initial data.
Deliverable: Dataset ready for analysis.

Sprint 2: Data Exploration & Feature Engineering – Explore data and create features.
Deliverable: Initial features for model.

Sprint 3: Model Building & Evaluation – Build and test the model.
Deliverable: Prototype model with performance metrics.

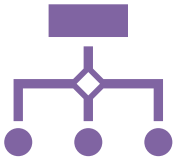
Sprint 4: Model Refinement & Dashboard – Refine model and build dashboard.
Deliverable: Improved model and working dashboard.

Sprint 5: Final Refinement & Presentation – Finalize model and prepare presentation.
Deliverable: Complete project ready for delivery.

Technical Implementation

Data Handling: SQL integration for structured data storage.

ML Models: Random Forest, Neural Networks, ARIMA.



Version Control: Git workflow with branches and pull requests for code management.

Which model and target

Remaining Useful Life (RUL) Prediction Using LSTM

What is RUL?

- **RUL**: Predicts time remaining before a machine fails, allowing proactive maintenance and cost savings.

Why LSTM?

- **Captures Long-Term Dependencies**: Ideal for sequential data like RUL.
- **Handles Nonlinear Patterns**: More effective than ARIMA for complex time-series data

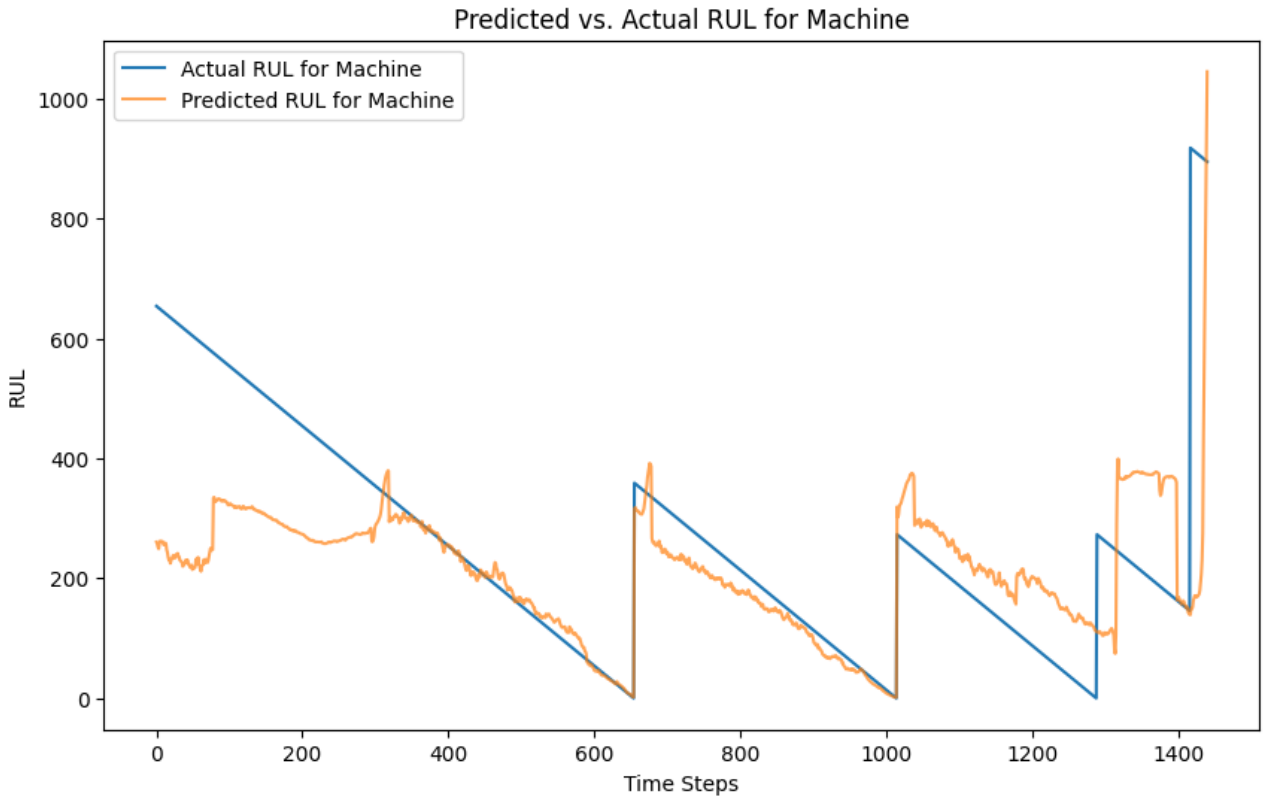
Handling RUL & Missing Data

Challenge:

Missing RUL values from lack of recent failure data, leading to inaccuracies.

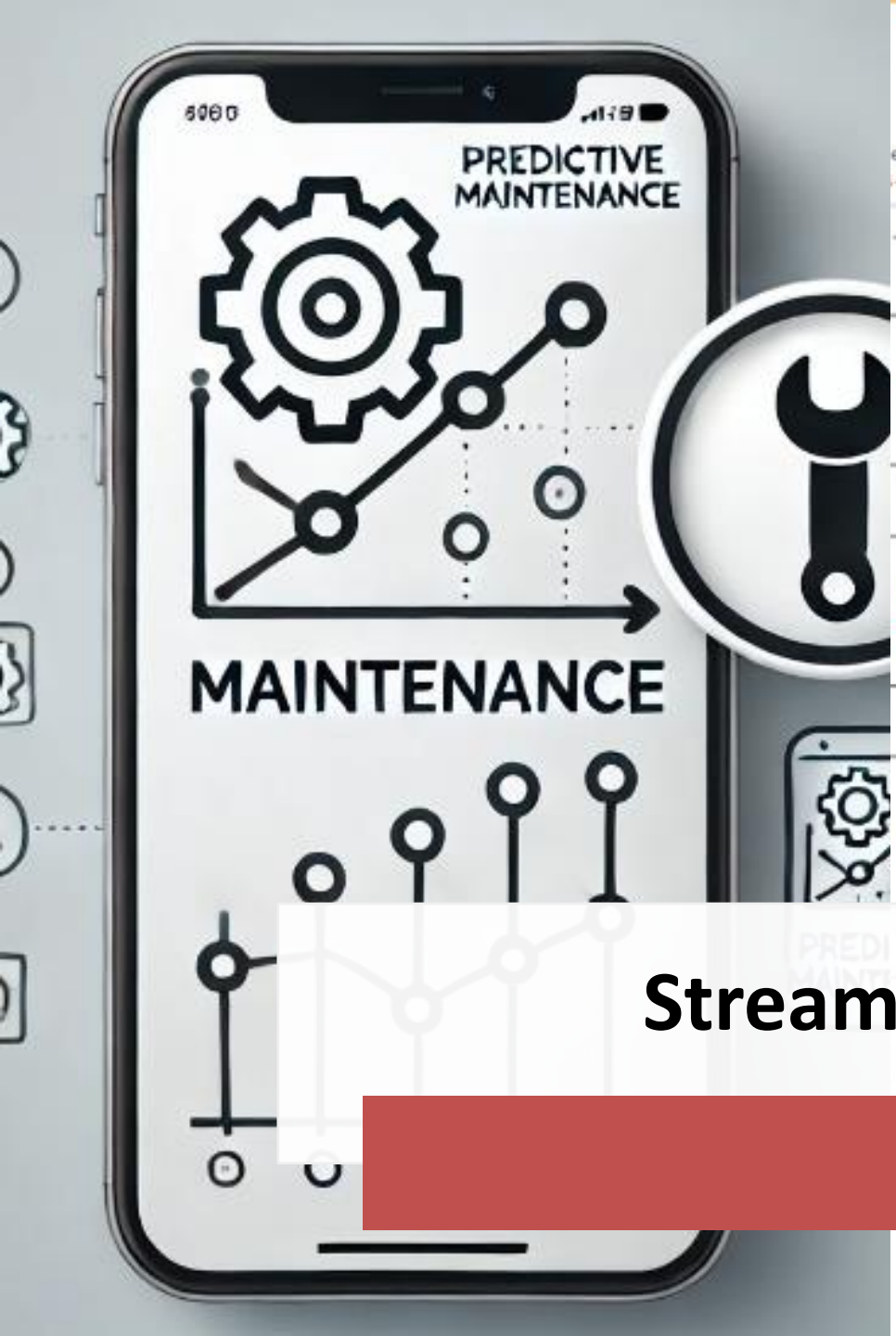
Solution:

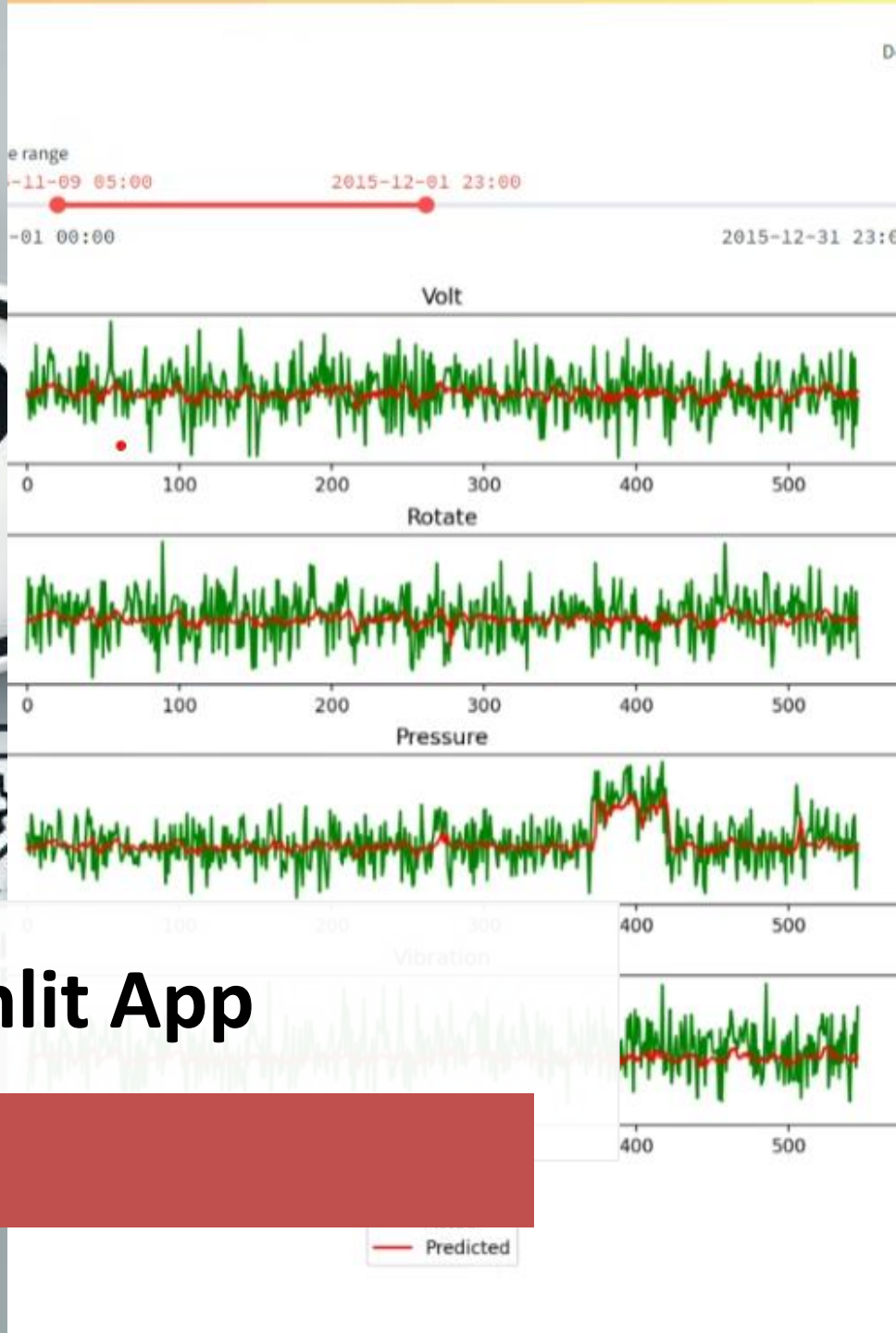
- **Machine-Specific Means**: Filled NaNs with machine-specific mean RULs, using a countdown.
- **Similarity-Based Imputation**: Used average RUL from similar machines (age/model) for machines without failure data.



Predicitve Power!!

We managed to create a magical model ready for our streamlit app





Streamlit App