Patient Mobility Forecasting using Time-Series & Clinical Data

Machine Learning Intern Assignment

Liberdat B.V.

This project forecasts daily patient step counts by integrating wearable mobility data with longitudinal clinical information to predict activity patterns and understand health factor impacts.

Problem Statement

Why It Matters

Patient activity levels are critical indicators of health status and recovery progress. However, raw step data from wearables is high-frequency, noisy, and difficult to interpret without context.

Clinical events–such as therapy sessions, medications, and side effects–strongly influence daily mobility patterns.

Primary Goal

Forecast daily step counts for the next 365 days

Secondary Goal

Explain impact of health factors on mobility

Data Overview

Time-Series Step Data

High-frequency mobility measurements captured from mobile devices and wearable sensors throughout the day

Clinical Data

Longitudinal health records including patient demographics, therapy schedules, medication administration, and documented side effects

Unified Timeline

Both datasets merged on a standardized daily timeline for comprehensive analysis

Data Preprocessing

Cleaning & Preparation

Raw wearable data required extensive preprocessing to create a reliable foundation for forecasting models.

We established a continuous daily timeline to ensure no gaps in the analysis period.

01

Timestamp Conversion

Converted all timestamps to standardized datetime format

02

Timezone Standardization

Unified all time zones to ensure accurate temporal alignment

03

Daily Aggregation

Aggregated high-frequency step measurements into daily totals

04

Timeline Creation

Generated continuous daily timeline spanning the entire study period

Feature Engineering

Therapy Indicators

Binary flags indicating therapy status (on/off) for each day, capturing treatment schedules and their potential impact on patient activity

Side Effect Metrics

Quantified side effect counts and severity scores, providing insights into how adverse events affect mobility patterns

Calendar Features

Temporal attributes including day of week and week of year to capture cyclical patterns in patient behavior

Lag Features

Historical step counts at t-1, t-7, and t-30 day intervals to incorporate recent activity trends into predictions

Modeling Approach

Two-Model Strategy

We implemented both a statistical baseline and an advanced machine learning model to compare performance and ensure robust predictions.

Time-based splitting preserved temporal ordering, preventing data leakage and ensuring realistic evaluation.

Baseline Model

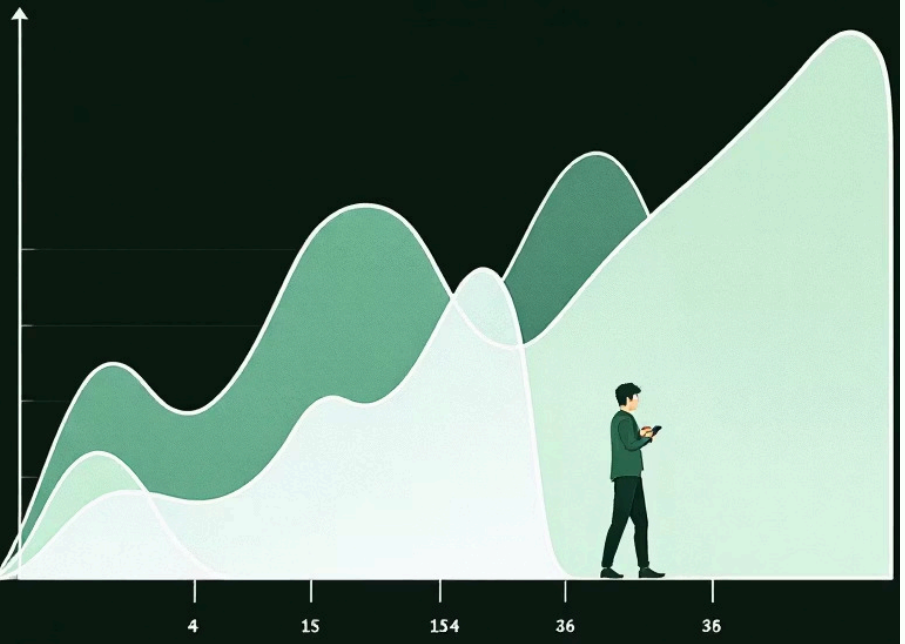
SARIMA

Univariate time-series forecasting capturing seasonal patterns

ML Model

Explainable Boosting Machine (EBM)

Multivariate approach with built-in interpretability



Model Evaluation

RMSE	MAE	↓
Root Mean Square Error	**Mean Absolute Error**	**ML Improvement**
Primary metric measuring prediction accuracy	Secondary metric for average deviation	Significant accuracy gains over baseline

The Explainable Boosting Machine demonstrated superior performance compared to the SARIMA baseline, achieving lower error rates across both RMSE and MAE metrics. The ML model's ability to incorporate clinical features alongside time-series patterns enabled more accurate forecasts.

Explainability Analysis

Understanding Predictions

Model interpretability is crucial for clinical adoption and trust.

EBM's global explanations revealed which factors drive mobility predictions.

Lagged Steps

Historical activity was the strongest predictor of future mobility

Therapy Impact

Active therapy periods correlated with increased patient activity

Side Effects

Documented side effects showed clear negative impact on mobility

365-Day Forecast Results

1

Daily Predictions

Generated step count forecasts for each of the next 365 days with confidence intervals

2

Trend Decomposition

Separated baseline activity trends from clinical event impacts for clearer interpretation

3

System Integration

Formatted output for seamless integration with downstream clinical decision support systems

Scalability & Conclusion

Scaling to 100,000 Patients

Big Data Infrastructure

Apache Spark + S3 for distributed processing

Batch Pipelines

Automated prediction workflows

1

2

3

Global Modeling

Single model serving multiple patient cohorts

Key Takeaway

This project demonstrates an **end-to-end interpretable and scalable forecasting system** that aggregates wearable data, integrates clinical events, engineers time-series features, and delivers accurate 365-day mobility predictions while explaining the impact of health factors.