Supporting Document

Time Series Forecasting for Patient Mobility

1. Top Three Challenges in Data Engineering Pipeline

Challenge 1: Temporal Alignment of Clinical Events

Problem: Clini cal events (therapies, side effects, diagnoses) have variable start/end dates that must be mapped to each daily timestep in the step count timeline.

Solution:- Created date range masks for each clinical event- Applied binary f lags for active periods- Aggregated overlapping events (e.g., multiple concur rent therapies)

Impact: Successfully merged 80K+ step count intervals with clinical data across a continuous daily timeline.

Challenge 2: Handling Missing Data and Timeline Gaps

Problem: Step count data had irregular intervals and potential missing days.

Solution:- Aggregated granular intervals to daily totals- Created continuous date range using pd.date_range()- Filled missing days with 0 (assuming no activity)- Validated no gaps in final timeline

Impact: Ensured robust time series without gaps, critical for lag features and rolling averages.

Challenge 3: Feature Engineering at Scale

Problem: Creating lag fea tures and rolling averages for 40+ features while managing memory and com putation.

Solution:- Vectorized operations using pandas- Efficient rolling window cal culations-Dropped redundant features early- Validated feature distributions

Impact: Generated 40+ meaningful features while maintaining reasonable com putation time.

2. Chosen Modeling Approach and Justification

Baseline Model: Prophet Choice: Facebook Prophet for univariate time series forecasting P

Justification:- Seasonality: Automatically detects weekly and yearly pat terns in step counts- Simplicity: Provides baseline without feature engineer ing- Interpretability: Clear trend and seasonal decomposition- Robustness: Handles missing data and outliers well

Results: Established performance benchmark for comparison.

Advanced Model: Explainable Boosting Machine (EBM) Choice: EBM from InterpretML library for multivariate forecasting Justification:

1. Explainability: Healthcare applications require inter pretable models- EBM provides global feature importance- Shows exact contribution of each clinical variable- Critical for clinical decision support

2. Performance: Gradient boosting accuracy with GAM interpretability

   - Captures non-linear relationships
   - Handles mixed feature types (continuous, categorical)
   - Competitive with XGBoost while being fully explainable

3. Clinical Context: Incorporates therapies, side effects, and events

   - Quantifies impact of clinical interventions
   - Identifies which factors drive mobility changes
   - Enables personalized predictions

Alternative Considered: XGBoost with SHAP- Rejected because EBM pro vides native interpretability- SHAP adds computational overhead- EBM's glass box nature preferred for healthcare

3. Key Learnings from Explainability Phase

Learning 1: Lag Features Dominate Predictions Finding: Steps from previous days (t-1, t-7, t-30) were consistently the top predictors.

Implication:- Patient mobility is highly autocorrelated- Recent behavior is the strongest predictor of future behavior- Clinical interventions have secondary (but significant) impact

Action: Prioritize lag features in future models; consider LSTM/RNN for se quence modeling.

Learning 2: Clinical Features Provide Meaningful Signal Finding: Active therapy count and side effect intensity ranked in top 10 features.

Implication:- Clinical context improves predictions beyond pure time series- Therapies and side effects measurably affect mobility- Model can quantify treatment impact on patient

activity Action: Expand clinical feature engineering; collect more granular therapy data.

Learning 3: Temporal Patterns Matter Finding: Day of week and week of year showed significant importance.

Implication:- Weekly routines strongly influence step counts- Seasonal vari ations exist in mobility patterns- Weekend vs. weekday behavior differs signifi cantly

Action: Consider separate models for weekdays vs. weekends; account for holi days.

Learning 4: Explainability Builds Trust Finding: Being able to explain "why" a prediction was made is as valuable as the prediction itself.

Implication:- Healthcare stakeholders need interpretable models- Black-box models (even if accurate) face adoption barriers-

Explainability enables model debugging and improvement

Action: Always prioritize interpretable models in healthcare applications; use EBM or similar glass-box approaches.

Conclusion

This project demonstrated that combining time series forecasting with clinical features yields superior predictions while maintaining full interpretability. The key success factors were:

1. Robust data engineering to align temporal and clinical data

2. Thoughtful feature engineering capturing both historical patterns and clinical context

3. Explainable modeling using EBM to quantify feature impacts

The resulting model provides actionable 365-day forecasts with clear explana tions of driving factors, making it suitable for clinical decision support.