PROJECT PRESENTATION

Title: Predictive Analytics for Network KPIs: Finding the Best ML Model
Presenter: Akila S | Date: 22/05/2025

OBJECTIVE

To analyze 5G KPI data (Avg. UE Number & DL PRB Utilization), forecast future trends, and identify the best ML model for predicting downtrends.

Key Goals:

- Forecast KPI behavior
- Identify when both UE and PRB utilization drop
- Compare several ML models (Random Forest, SVR, Prophet, etc.)
- Recommend the best model for forecasting
- Use machine learning models to identify trends and make predictions.
- Outline my approach and results.

Identifying Drop Patterns in DRB Utilization & Avg UE

- **DRB Utilization:** Measures downlink physical resource block usage.
- **Avg UE Number**: Represents average number of active users per interval.
- By analyzing temporal patterns (hour of day, day of week), we identify low-traffic periods.

Concept

- Both DRB Utilization and Avg UE Number decline during off-peak hours, and this can be accurately forecasted using time-based ML models.

Key Insight:

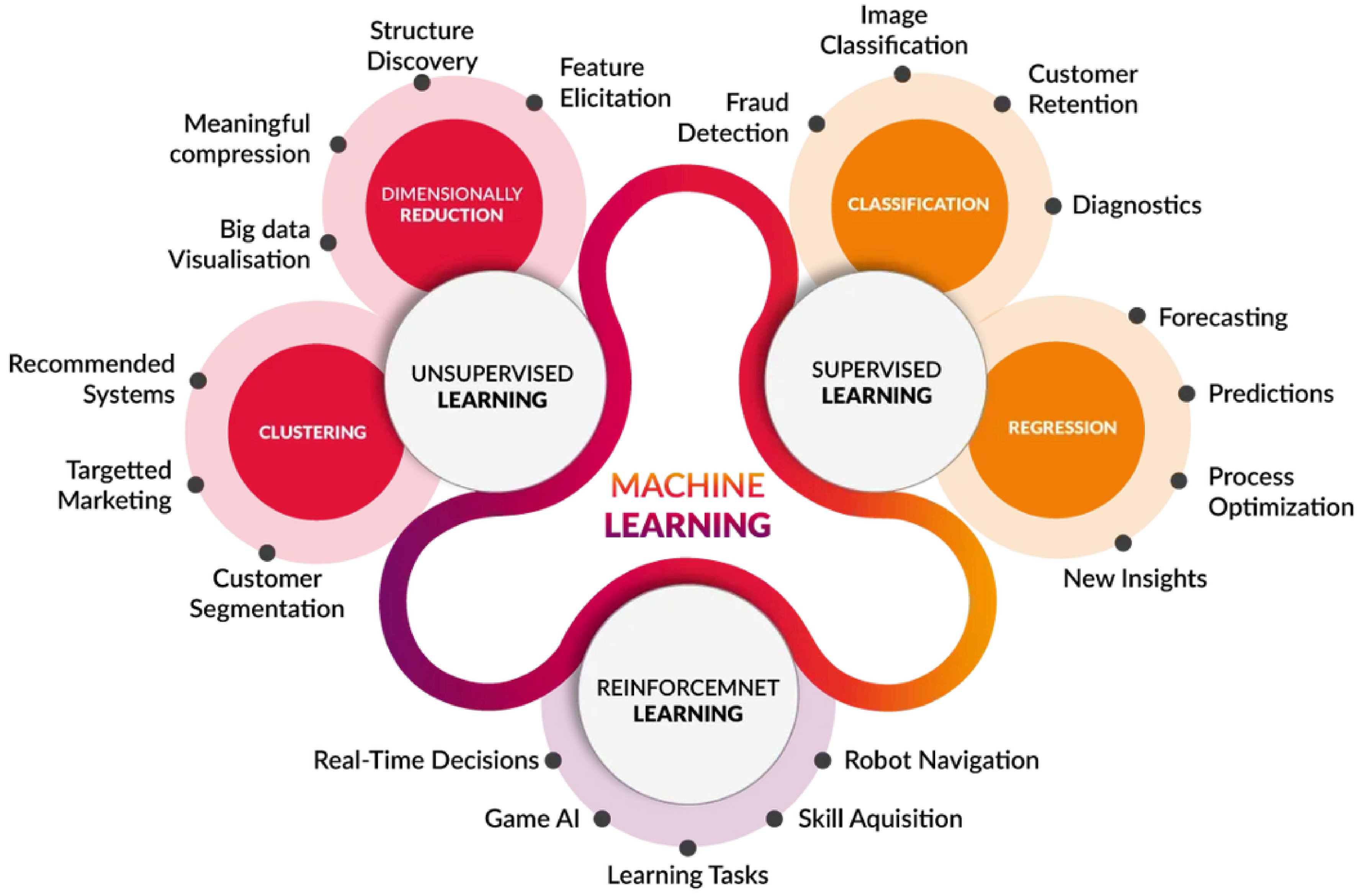
- Usage consistently drops during late-night hours (12 AM – 6 AM).
- Drop patterns are location-specific, depending on gNB and NCI.

Why It Matters:

- Suitable for multivariate prediction using time, gNB, and NCI as features.
- Enables network resource optimization and cost-effective planning.

📌 Algorithms

- Random Forest Regressor
- Gradient Boosting Regressor
- Long Short-Term Memory
- XGBoost
- Linear Regression
- Decision Tree Regressor
- Support Vector Regression (SVR)
- Prophet
- ARIMA / SARIMA
- Weighted Average Regression (WAR)



ML Algorithm Comparison for KPI Forecasting

Algorithm	Time Series Support	Multivariate Support	Long-Term Forecasting	Handles Nonlinearity	Interpretability
Random Forest Regressor	❌ Needs lag features	✅ With manual features	⚠️ Limited	✅ Good	✅ High
Gradient Boosting Regressor	❌ Needs lag features	✅ With manual features	⚠️ Limited	✅ Good	✅ Moderate
LSTM (Long Short-Term Mem	✅ Native	✅ Native	✅ Excellent	✅ Excellent	❌ Low
XGBoost	❌ Needs lag features	✅ With manual features	⚠️ Limited	✅ Excellent	✅ Moderate
Linear Regression	❌ Needs lag features	✅ With manual features	❌ Poor	❌ Poor	✅ High
Decision Tree Regressor	❌ Needs lag features	✅ With manual features	❌ Poor	✅ Good	✅ High
Support Vector Regression	❌ Needs lag features	✅ With manual features	❌ Poor	✅ Good (with kernels)	⚠️ Moderate
Prophet	✅ Native (Univariate)	❌ One KPI only	✅ Good	✅ Trend/Seasonality	✅ High
ARIMA / SARIMA	✅ Native (Univariate)	❌ One KPI only	✅ Good (if stationary)	❌ Linear only	✅ High
Weighted Average Regressior	❌ Not time-aware	⚠️ Limited	❌ Poor	❌ Poor	✅ High

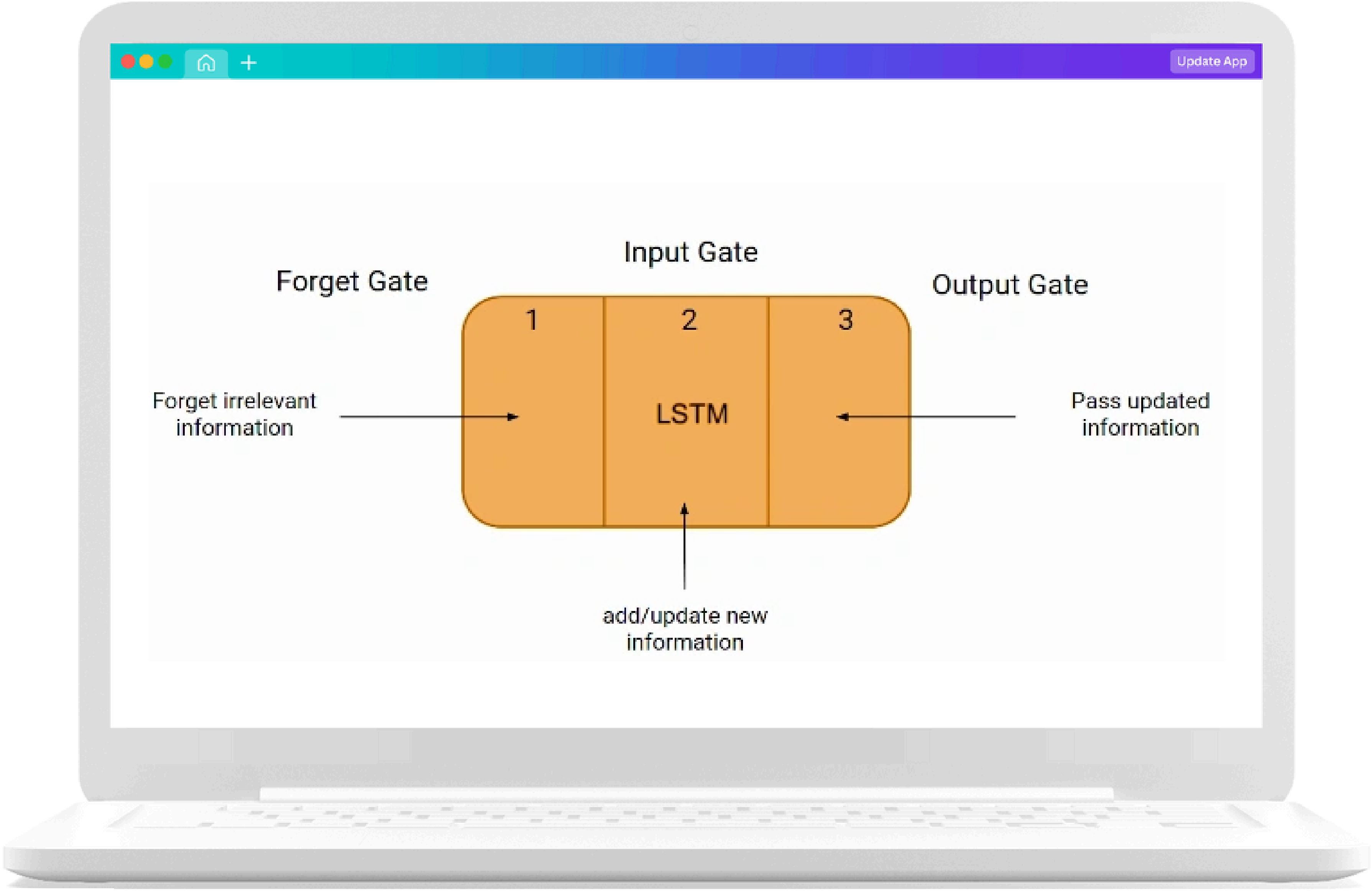
BEST THREE ALGORITHMS

Aspect	LSTM	LightGBM	Prophet
Multivariate KPI Support	✅ Natively supports multivariate time series fore	⚠️ Possible via feature engineering	❌ Only univariate
Sequential Dependency Handling	✅ Excellent (captures long-term and short-term	⚠️ Needs engineered lag/rolling features	⚠️ Limited to trend/seasonality components
15-min / 30-min Interval Support	✅ Fully supported via adjustable sequence wind	✅ Supported with fine-grained feature engineeri	⚠️ Not natively designed for sub-hour granulariti
6-month Ahead Forecasting	✅ Can predict multi-step ahead at any interval (i	⚠️ Doable, but error compounds in long horizons	⚠️ Limited, works best for daily/weekly/monthly
Non-Linear Relationship Handling	✅ Excellent (deep learning strength)	✅ Good (via boosted trees)	⚠️ Limited (piecewise linear trends only)
Interpretability	⚠️ Low (black-box, hard to explain predictions)	✅ Medium (feature importance available)	✅ High (easy to explain via trend & seasonality c
Training Speed & Complexity	⚠️ Slower, needs careful tuning (benefits from GI	✅ Fast and efficient	✅ Very fast, minimal config

BEST ALGORITHM

LSTM is the Best Choice

- Captures Temporal Patterns
  Learns long-term dependencies in time-series data (ideal for telecom KPIs).
- Supports Multivariate Forecasting
  Predicts multiple KPIs like Avg_UE_Number & DL_Prb_Utilization together.
- Handles Non-linear Trends
  Effectively models complex traffic behavior and utilization patterns.
- Superior Accuracy (Low Error Metrics)
- Flexible with Time Windows
  Performs well across 30-min, 1-hour, or custom intervals.



Superior Accuracy (Low Error Metrics)

Metric	LSTM	LightGBM	Prophet
MAE	✅ **8.12**	9.75	10.41
RMSE	✅ **10.03**	12.29	13.54

How to Approach with These Data

1. Data Preparation
   - Merge datasets on Timestamp, NCI, gNB
   - Convert Timestamp to datetime, create features (hour, day)
2. Exploratory Data Analysis (EDA)
   - Visualize usage trends over time
   - Identify peak and off-peak periods
3. Feature Selection
   - Inputs: Hour, Day, Avg_UE_Number, gNB, NCI
   - Target: DL_Prb_Utilization (or vice versa)
4. Model Building & Evaluation
   - Train/Test split (80/20)
   - Use models like Random Forest, XGBoost
   - Evaluate using MAE, RMSE, R^2 Score
5. Prediction & Insights
   - Predict high-utilization periods
   - Export results for deployment

Conclusion:

A structured ML workflow enables accurate predictions and actionable insights.

PROS AND CONS

LSTM – Pros

- Remembers long-term patterns

  Designed to retain information over long time steps.
- Handles sequential/time-series data

  Ideal for forecasting, NLP, and sensor data.
- Reduces vanishing gradient

  Uses gates to preserve gradients during training.
- Supports multivariate inputs

  Can learn from multiple KPIs/features simultaneously.
- Flexible output structure

  Supports many-to-one, one-to-many, and many-to-many forecasting.

LSTM – Cons

- Slower training

More computationally expensive than basic models.

- Complex architecture

Requires more tuning and understanding.

- Needs more data

Performs best with large datasets.

- Harder to interpret

Acts as a black-box model, unlike linear models.

Thank you!