Smart Energy Consumption Monitoring and Prediction using Big Data Techniques

Supervisor – Dr. Samir Patel

Kavya Patel (22BCP406) | Dev Pandya (22BCP409)

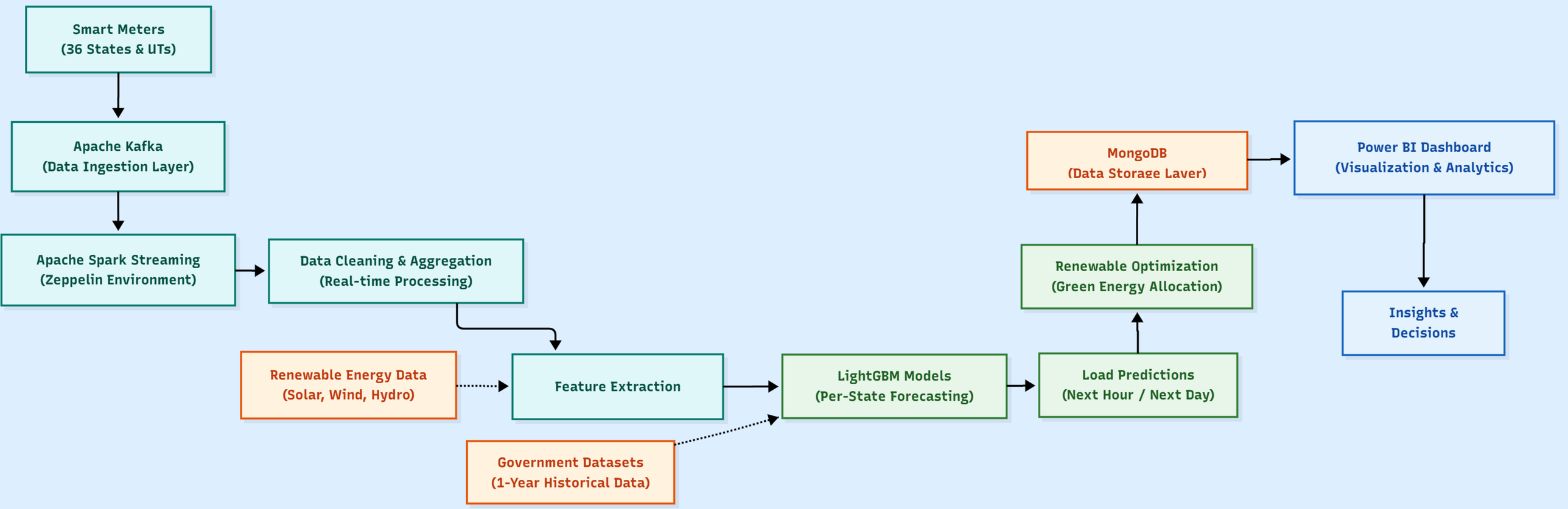
Brief Motivation

- India's rapid urban growth has led to rising and unpredictable electricity demand, making efficient grid management essential.
- Traditional systems cannot handle large, real-time smart meter data, causing imbalance and wastage.
- Our project uses Big Data tools like Kafka and Spark with LightGBM models to forecast energy demand and integrate renewable sources.
- This enables smarter decisions, reduces wastage, and supports India's shift toward sustainable, data-driven energy management.

Objective

- Design a scalable Big Data pipeline using Apache Kafka and Spark for real-time energy data collection and processing from 36 states and union territories.
- Integrate renewable energy data (solar, wind, hydro) into predictions to optimize green energy usage and reduce wastage.
- Implement real-time analytics dashboards in Power BI for visualizing load patterns, predictions, and renewable utilization.
- Enable data-driven decision-making for efficient grid management, load balancing, and sustainable energy planning.

Methodology / Design Component



All services Kafka, Zookeeper, MongoDB, and Zeppelin are containerized using Docker and Docker Compose for easy orchestration and scalability.

The Zeppelin notebook acts as the Spark interpreter, processing streaming data and running LightGBM models in real time.

Development is done on WSL Ubuntu, ensuring smooth integration between local scripts and containerized services for efficient big data processing and visualization.

Expected Results

Model-Level Results

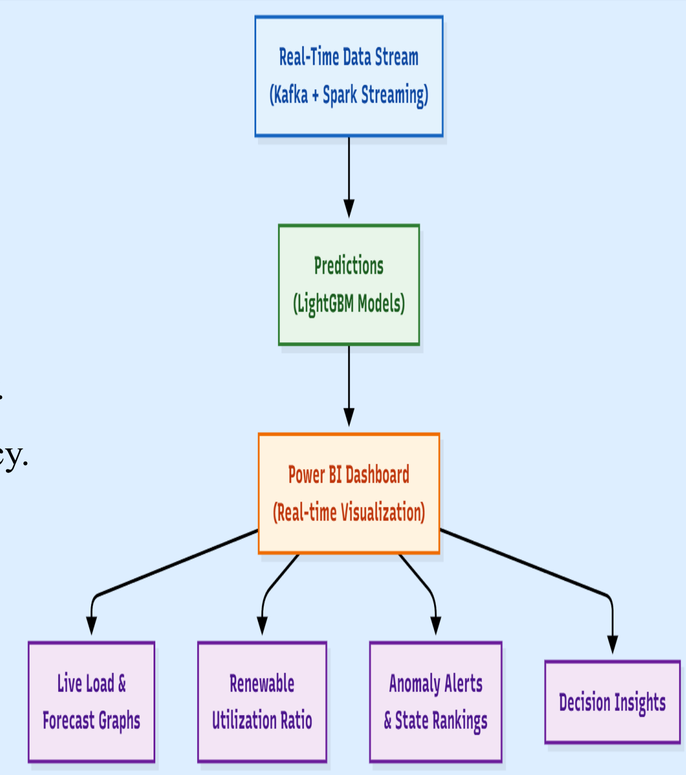
- LightGBM achieves >90% forecasting accuracy with an average RMSE of ~70 MW.
- Outperforms ARIMA and Prophet in speed and accuracy on large-scale data.

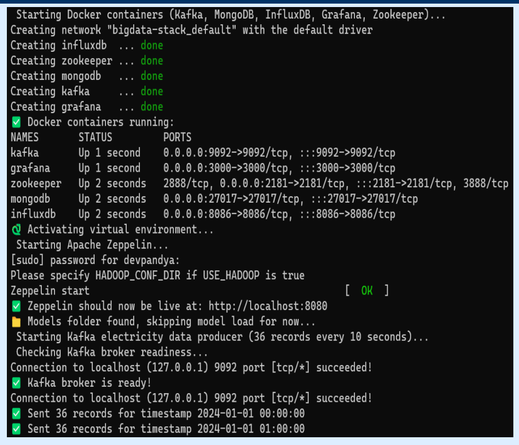
System-Level Results

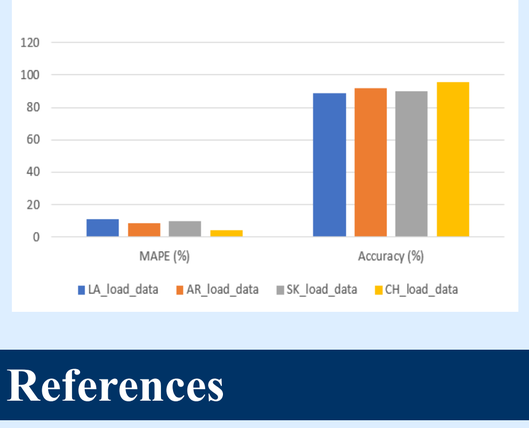
- Kafka–Spark pipeline processes live data from 36 states with <3 seconds latency.
- Power BI dashboard displays real-time load, predictions, and renewable efficiency.
- Renewable optimization module improves green energy utilization.

Practical Impact

- Enables real-time energy management through data-driven forecasting.
- Provides scalable, fault-tolerant architecture deployable on cloud or local grids.
- Supports India's smart grid and clean energy transition with actionable analytics.







Outcomes

- Accurate energy demand forecasting with over 90% prediction accuracy using LightGBM.
- Real-time monitoring and visualization of load and renewable usage through Power BI dashboards.
- Optimized renewable energy integration, improving solar and wind utilization.
- Data-driven decision support for efficient grid management and policy planning.
- Enhanced grid stability and sustainability, reducing energy wastage and ensuring balanced power distribution.

Bibliography/ References

- "A transfer learning-based hybrid model with LightGBM for smart grid short-term energy load prediction" (Simaiya et al., 2024)
- "Electricity Load and Peak Forecasting: Feature Engineering, Probabilistic LightGBM and Temporal Hierarchies" (Rubattu, Maroni, Corani, 2023)
- "Integrating Wind and Solar in the Indian Power System: An Assessment with a Unit Commitment and Dispatch Model"
- Dataset Resources -
  - **POSOCO/NLDC Reports: Collected state-wise daily demand and peak load data for model training.**
  - **CEA Dashboard: Provided monthly and annual installed capacity and supply position statistics.**
  - **National Power Portal (NPP): Offered region-wise renewable generation data for solar, wind, and hydro.**
  - **data.gov.in: Supplied machine-readable datasets on power demand, generation, and availability gaps.**