

Big Data-Based Energy Consumption Prediction for Smart Grids

Varun Kumar D , Mrs. Jyothika K R

P.G. Student, Department of Master of Computer Applications, BIET, Davangere-577004.

varunkumar.d66@gmail.com, Davangere, Karnataka, India

Assistant Professor & HOD Department of Master of Computer Applications, BIET, Davangere-577004.

jyothikakrkushi@gmail.com, Davangere, Karnataka, India

ABSTRACT

The rapid deployment of smart meters and IoT devices in modern power systems has led to an exponential growth of energy consumption data. Traditional statistical forecasting techniques struggle to manage the volume, velocity, and variety of such data. This research proposes a Big Data-driven predictive analytics framework for Smart Grids using Hadoop, Apache Spark, and Deep Learning models such as Long Short-Term Memory (LSTM). The system supports real-time anomaly detection, scalable distributed processing, and improved forecasting accuracy, enabling proactive grid management and operational efficiency.

I. INTRODUCTION

Smart Grid represent the modernization of traditional power systems by integrating digital communication and intelligent monitoring systems. With the deployment of Advanced Metering Infrastructure (AMI), utilities collect high-frequency data from millions of consumers.

However, conventional forecasting methods such as linear regression and ARIMA struggle to capture complex consumption patterns influenced by weather, human behavior, industrial cycles, and seasonal variations.

This research proposes a distributed Big Data architecture combined with deep learning techniques to provide accurate, scalable, and real-time energy forecasting for smart grid environments.

on:

- Similar Day Method
- Linear Regression
- ARIMA models
- RDBMS-based storage systems

These systems suffer from:

- Poor scalability
- Low prediction accuracy
- High latency in decision-making

- Inability to process real-time streaming data

Recent studies highlight the effectiveness of:

- Hadoop ecosystem for large-scale data storage
- Spark Streaming for real-time analytics
- Deep Learning models such as LSTM and GRU for time-series forecasting

However, integration of all these technologies into a unified Smart Grid prediction framework remains limited.

III. EXISTING SYSTEM

Traditional systems rely on relational databases and batch-processing mechanisms. They are reactive rather than predictive and experience significant latency in decision-making. Additionally, they lack horizontal scalability and real-time anomaly detection capabilities.

IV. PROPOSED SYSTEM

The proposed system uses:

- **Hadoop (HDFS)** for distributed storage
- **Apache Kafka** for data ingestion
- **Apache Spark** for processing
- **LSTM & Random Forest** for prediction
- **Flask API** for deployment
- **Web Dashboard** for visualization

Key Features:

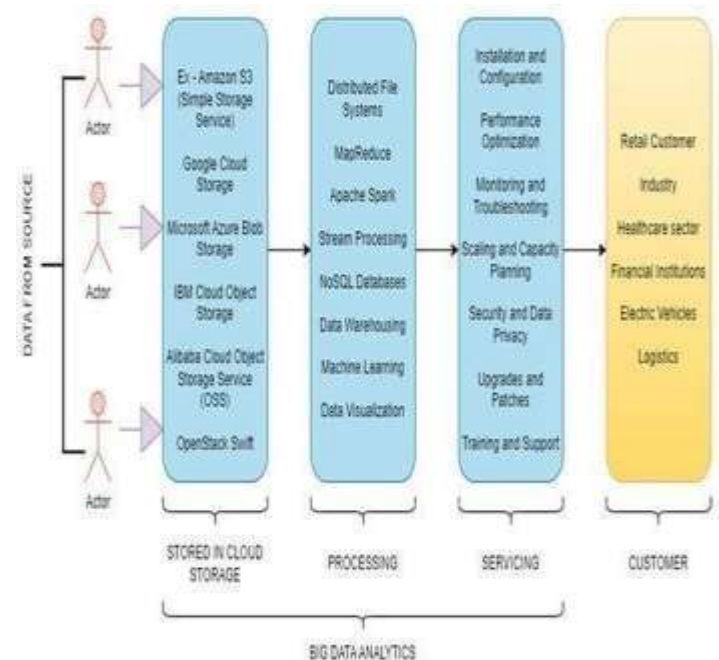
- High-velocity data ingestion

- Distributed storage
- Automated ETL pipeline
- Real-time anomaly detection
- Interactive dashboard
- Horizontal scalability

Advantages:

- Higher forecasting accuracy
- Real-time operational intelligence
- Reduced maintenance cost
- Better renewable integration
- Demand-response optimization

V. SYSTEM ARCHITECTURE



The system consists of:

1. Data Sources (Smart meters, Weather APIs)
2. Data Ingestion Layer (Kafka)
3. Distributed Storage (HDFS)
4. Processing Layer (Spark)
5. ML Model Training (LSTM, RF)
6. Prediction API
7. Visualization Dashboard

VI. METHODOLOGY

1. Data Collection

- Smart meter data
- Weather data
- SCADA logs

2. Data Preprocessing

- Missing value imputation
- Timestamp standardization
- Data normalization
- Feature engineering (Lag features, Day of Week, Hour of Day)

3. Model Development

- LSTM neural network
- Random Forest Regressor
- Training (70%), Validation (15%), Testing (15%)

4. Evaluation Metrics

- MAE
- RMSE
- R^2 Score

5. Deployment

- Model serialization
- Flask API
- Web-based dashboard visualization

VII. RESULTS AND DISCUSSION

The experimental results demonstrate improved forecasting accuracy compared to traditional methods. The LSTM model effectively captures long-term dependencies and seasonal patterns. Real-time anomaly detection successfully identifies abnormal consumption deviations exceeding predefined thresholds.

- **High-volume data ingestion:** The dashboard (1,000+ filtered data points) reflects how smart grids collect continuous streams from smart meters, IoT sensors, and substations—forming the foundation for big data analytics.
- **Time-series forecasting:** The energy consumption trend line illustrates historical patterns used by models (like ARIMA, LSTM, Prophet) to predict future demand at hourly, daily, or monthly levels.
- **Real-time analytics:** Metrics such as *Avg Grid Load (21.2%)* and *Peak Demand (3.5 kWh)* show how streaming analytics helps utilities respond instantly to fluctuations.
- **Load balancing & demand response:** Predictive models use consumption trends to shift or reduce load during peak hours, preventing overloads and reducing costs.
- **Cost optimization:** The *Total Cost (INR 262.89)* indicates how prediction helps

estimate future billing and enables dynamic pricing strategies based on demand forecasts.

- **Anomaly detection:** Sudden spikes or drops in the trend graph can be flagged using big data techniques (e.g., clustering, outlier detection) to detect faults, theft, or abnormal usage.
- **Renewable energy integration:** Prediction models combine consumption data with weather and solar/wind forecasts to balance supply from intermittent renewable sources.
- **Scalability with big data tools:** Technologies like distributed storage and processing frameworks enable handling massive datasets across regions in real time.
- **User behavior insights:** Consumption patterns over dates (Nov–Dec) help segment users and personalize energy-saving recommendations.
- **AI-driven forecasting:** The “AI Forecast” feature suggests use of machine learning models to improve accuracy by learning complex nonlinear patterns in energy usage.
- **Grid health monitoring:** The *Grid Health indicator* (Avg Load %) reflects predictive maintenance and system stability analysis powered by big data.
- **Peak demand prediction:** Identifying when peaks occur (e.g., 12:30 am) allows proactive infrastructure planning and avoids blackouts.

VIII. CONCLUSION

This project successfully demonstrates the transformation of traditional energy forecasting into a Big Data-driven intelligent system. By integrating distributed storage, parallel processing, and deep learning models, the proposed system provides:

- Accurate energy load forecasting

- Real-time anomaly detection
- Scalable infrastructure
- Optimized grid management

The system contributes toward building autonomous and self-healing Smart Grids.

X. FUTURE ENHANCEMENTS

Future work includes integration of edge computing for localized predictions, blockchain-based peer-to-peer energy trading mechanisms, electric vehicle load forecasting, and reinforcement learning for autonomous grid control.

REFERENCES

1. Al-Fuqaha A. Guizani M. Mohammadi M. Aledhari M. & Ayyash M. (2015). Internet of Things: A Survey on Enabling Technologies Protocols and Applications. IEEE Communications Surveys & Tutorials 17(4) 2347-2376.
2. Hochreiter S. & Schmidhuber J. (1997). Long Short-Term Memory. Neural Computation 9(8) 1735 1780. (Foundational paper for the LSTM algorithm used).
3. Chen M. Mao S. & Liu Y. (2014). Big Data: A Survey. Mobile Networks and Applications 19(2) 171-209.
4. ang X. Misra S. Xue G. & Yang D. (2012). Smart Grid — The New and Improved Power Grid: A Survey. IEEE Communications Surveys & Tutorials 14(4) 944-980.
5. White T. (2015). Hadoop: The Definitive Guide. 4th Edition. O'Reilly Media. (Reference for the Big Data architecture implementation).