FINAL CONSOLIDATED PROJECT REPORT

Energy Consumption Prediction, Smart Grid Analytics, and Interactive Educational Platform

---

1. Executive Summary

This project integrates **machine learning analytics** with an **interactive Smart Grid educational system** to address the growing need for intelligent energy management and public awareness. Using historical U.S. state-level energy consumption data, the study develops two high-performance ML models:

1. **Task A:** Classification of monthly consumption levels (LOW / MED / HIGH)
2. **Task B:** Prediction of the **minimum energy source** used per state-month producer record

The Smart Grid prototype website provides an accessible platform to explain Smart Grid architecture, AMI, benefits, and project workflow. The solution is suitable for **educational demonstrations, analytics dashboards, and smart grid research**.

---

2. Project Objectives

2.1 Machine Learning Objectives

- Develop a classification model to categorize consumption levels.
- Identify the minimum energy source from multiple fuel types.
- Build deployable ML pipelines suitable for integration with a front-end dashboard.

2.2 Smart Grid Educational System Objectives

- Create an interactive, accessible website explaining Smart Grid concepts.
- Include AMI visualization, architecture diagrams, and educational elements.
- Provide a workplan that guides ML integration into the Smart Grid system.

---

3. Dataset Description

The dataset used is **consumptionstatemon1.xlsx**, consisting of **12,423 records** and the following columns:
YEAR, MONTH, STATE, TYPE_OF_PRODUCER, ENERGY_SOURCE, CONSUMPTION.

It provides detailed monthly energy consumption at state-level granularity across various fuel sources.

---

4. Data Preprocessing

A structured preprocessing pipeline was created.

4.1 Cleaning

- Standardized column names
- Removed missing or invalid rows
- Final dataset shape: **12423 × 6**

4.2 Encoding & Target Generation

- OneHotEncoder for categorical features
- LabelEncoder for classification tasks
- Created **CONSUMPTION_CLASS** using quantile-based thresholds:

    - LOW
    - MED
    - HIGH

4.3 Dataset Balance

The distribution is naturally balanced:

- HIGH: 4229
- LOW: 4112
- MED: 4082
  → No need for SMOTE.

---

5. Feature Engineering

5.1 Task A: Consumption Classification

- One-hot encoded vectors created **64 features**.
- Training / Testing split: **9317 / 3106 records**.

5.2 Task B: Minimum Source Prediction

- Converted dataset into a pivot table with shape **4462 × 8**, representing:
    - Coal
    - Natural Gas
    - Petroleum
    - Other Gases
    - NONE

- Target: **MIN_SOURCE**
- Final model input features: **64 encoded features**.

---

6. Machine Learning Models

Four ML models were evaluated:

- Logistic Regression
- Decision Tree
- Random Forest
- SVM with PCA

---

7. Results — Task A: Consumption Classification

7.1 Model Performances

Logistic Regression

- Accuracy: **73%**
- Poor non-linear learning

Decision Tree

- Accuracy: **89.89%**
- Strong interpretability

Random Forest — Best Model

- Accuracy: **91.21%**

- Robust to noise, strong generalization

Class	Precision	Recall	F1
HIGH	0.94	0.94	0.94
LOW	0.93	0.93	0.93
MED	0.87	0.87	0.87

Selected Model Parameters:

- n_estimators: 200
- max_depth: None
- min_samples_split: 5

Conclusion:
Random Forest achieved the highest and most consistent accuracy for Task A.

---

8. Results — Task B: Minimum Source Prediction

Random Forest again delivered exceptional performance.

8.1 Accuracy

- **99.82%** (≈ perfect prediction)

8.2 Class Performance

Energy Source	Precision	Recall	F1
Coal	1.00	1.00	1.00
NONE	1.00	1.00	1.00

Natural Gas	1.00	1.00	1.00
Other Gases	1.00	1.00	1.00
Petroleum	1.00	0.95	0.97

Conclusion:
This near-perfect performance makes the model suitable for real-time smart grid optimization systems.

---

9. Exported ML Artifacts

The following serialized files were generated:

Task A Artifacts

- taskA_randomforest.pkl
- taskA_preprocessor.pkl

Task B Artifacts

- taskB_randomforest.pkl
- taskB_preprocessor.pkl

These are fully deployable and include preprocessing pipelines.

PART II — SMART GRID EDUCATIONAL SYSTEM

---

10. Purpose of the Website

The Smart Grid website serves as an **educational and interactive platform** to help users visualize and understand modern electricity systems.

Features include:

- Animated Smart Grid architecture
- AMI system visualization
- Stepwise project workplan
- Benefits and future vision
- Responsive UI for accessibility

---

11. Website Structure & Features

11.1 Hero Section

- Introduces Smart Grid concepts clearly
- Buttons to navigate to Architecture and AMI sections

11.2 About Section

Explains:

- Smart Grid core idea
- Key technologies: AMI, DER, analytics
- Impacts: reliability, sustainability, consumer empowerment

11.3 AMI System Section

Highlights:

- Smart meter
- Data communication path
- Utility analytics system
- Animated data flow representing secure transfer

11.4 Interactive Architecture

An SVG-based interactive diagram showing:

- Generation
- Transmission
- Substation
- Distribution
- DER
- Smart Meter

Includes hover tooltips, keyboard navigation, click-to-pin details.

11.5 Workplan Section

Aligned with the Smart Grid project plan:

Phase	Deliverables
Phase 1	Literature review, dataset collection
Phase 2	UX and diagram design
Phase 3	Prototype build

Phase 4	Demo materials and presentation

11.6 Benefits & Vision

- Enhanced reliability
- Renewable energy integration
- Customer empowerment
- Future focus: microgrids, EVs, decentralized markets

11.7 Contact Section

Includes feedback form (non-functional in prototype).

---

12. Workplan Summary (From Smart Grid Workplan PDF)

Completed Tasks

- Literature review
- Consumption dataset preparation
- AMI simulation creation
- Interactive UI prototyping

New Objectives

- ML-based energy prediction
- Grid stability classification

- Integrated analytics dashboard
- UI/UX optimization

Timeline (Aug 25–Sep 30, 2025)

- Aug 25–31: Data & ML model development
- Sep 1–10: Dashboard & interactivity
- Sep 11–20: UI/UX testing
- Sep 21–30: Documentation and final review

---

13. Integrated Insights

Based on combined analysis:

13.1 Machine Learning

- Random Forest is superior for both classification and regression-type tasks.
- Structured, categorical-heavy datasets benefit from tree-based methods.
- Energy consumption patterns are predictable with high accuracy.

13.2 Smart Grid System

- Demonstrating Smart Grid visually enhances public understanding.
- AMI integration showcases real-time data flow capabilities.
- Educational UI supports training, awareness, and stakeholder engagement.

13.3 Combined System Potential

By merging ML analytics with the Smart Grid website, a **real-time energy dashboard** can be created.

---

14. Conclusion

This project successfully integrates:

✓ **High-accuracy machine learning models** for consumption classification and minimum source prediction
✓ **Interactive Smart Grid educational website** explaining modern energy systems
✓ **A clear project plan** aligning ML analytics with Smart Grid visualization tools

Final Selected Models

Task	Model	Accuracy
Consumption Classification	Random Forest	**91.21%**
Minimum Source Prediction	Random Forest	**99.82%**

The system serves as a holistic solution for **energy analytics, Smart Grid education, and scalable future development**.

---

15. Future Enhancements

Machine Learning

- Integrate time-series forecasting (LSTM, Prophet)
- Add SHAP/LIME explainability

- Deploy via Flask/FastAPI

Smart Grid Platform

- Add real AMI dataset streaming
- Add EV and microgrid simulation
- Build a full analytics dashboard