**Powering Sustainability: Forecasting Electricity Prices and Renewable Energy Trends with Machine Learning**

A data-driven analysis of electricity generation, pricing dynamics, and renewable energy evolution in Turkey (2018–2023)

**1. Summary**

This project leverages machine learning techniques to model and forecast electricity prices while uncovering key insights into the integration and performance of renewable energy sources in Turkey's energy mix from 2018 to 2023.

Using a rich dataset of hourly electricity generation and consumption, we:

* Performed comprehensive exploratory data analysis (EDA)
* Engineered time-based and energy-specific features
* Applied regression models including Linear Regression and Random Forest to predict electricity prices (TRY/MWh)
* Evaluated model performance using R², MAE, and RMSE metrics

Special emphasis was placed on identifying trends in solar, wind, and hydro generation, their contributions to the grid, and how they impact pricing and sustainability.

The results demonstrate the strength of predictive modeling in supporting data-driven energy policy and grid management strategies, especially in the transition toward a cleaner and more resilient energy system.

**2. Analysis Questions**

|  |  |
| --- | --- |
| Category | Questions |
| Consumption & Demand | What are the hourly, daily, and monthly patterns in electricity consumption? |
| How does electricity demand differ between weekdays and weekends? |
| Are there significant seasonal trends in electricity usage (e.g., summer vs. winter)? |
| Which year had the highest average electricity consumption, and why? |
| Generation & Sources | What percentage of total electricity generation comes from renewable vs. fossil sources? |
| How has the reliance on each energy source changed from 2018 to 2023? |
| Which renewable source (solar, wind, hydro) has shown the most growth? |
| What is the variability of solar and wind generation throughout the day and year? |
| Demand-Supply Balance | How often does electricity generation exceed or fall short of consumption? |
| During which time periods (hour, month, season) do imbalances most frequently occur? |
| How dependent is the system on imported electricity during peak hours? |
| Pricing & Economics | How do electricity prices correlate with demand levels? |
| Which energy sources are most associated with high pricing periods? |
| Is there a seasonal or annual trend in electricity prices? |
| How do currency fluctuations affect local pricing in TRY? |
| Variable Relationships | What are the strongest predictors of electricity price per MWh? |
| Are there any surprising negative or positive correlations between variables? |
| How do combinations of energy sources impact overall system efficiency and price stability? |

**3. Dataset Description**

**Structure:**

* **Timeframe:** January 2018 – December 2023 (hourly records).
* **Total Records:** 52,584
* **Total Columns:** 22
  + **Time:** Timestamp of each record.
  + **Consumption & Generation:**
    - consumption\_MWh.
    - total\_generation\_MWh.
  + **Energy Sources:**
    - Fossil-based: natural\_gas, coal\_imported, fuel\_oil, asphaltite\_coal, hard\_coal, lignite, LNG, naphtha.
    - Renewable: hydro\_dam, hydro\_river, solar, wind, biomass, geothermal, waste\_heat.
    - Others: International (Imports).
  + **Electricity Prices:**
    - TRY/MWh, USD/MWh, EUR/MWh.

**4. Objectives**

* Forecast electricity prices (TRY/MWh).
* Analyze contributions of renewable vs. fossil sources.
* Study temporal patterns in energy consumption and generation.
* Support strategic decisions for renewable integration.

**5. Methodology**

**Data Preparation**

* Clean and parse datetime values.
* Engineer new features:
* hour, day\_of\_week, month, season.
* % share of each generation source.
* Handle outliers and scaling for modeling.

**Exploratory Data Analysis (EDA)**

* **Heatmaps** of average usage by hour/week/month.
* **Scatterplots** for price correlation.
* **Boxplots** for seasonal variations in pricing and generation.

**Modeling Techniques**

* **Regression Models:**
  + Linear Regression.
  + Random Forest Regression.

**6. Model Performance**

|  |  |  |  |
| --- | --- | --- | --- |
| Model | R² Score | MAE (TRY) | RMSE (TRY) |
| Linear Regression | 0.793 | 124.58 | 172.93 |
| Random Forest | **0.976** | **27.34** | **59.47** |

**7. Tools**

|  |  |
| --- | --- |
| Tool | Purpose |
| Python (Pandas, NumPy) | Data handling and transformation |
| Matplotlib, Seaborn, Plotly | Visualization |
| Scikit-learn | |  | | --- | | Regression modeling and clustering |  |  | | --- | |  | |
| Streamlit | Reporting and dashboards |
| Colab Notebook | |  | | --- | | Workflow and documentation |  |  | | --- | |  | |

**8. Conclusion**

This project outlines a clear and practical approach to analyzing electricity generation, consumption, and pricing data between 2018 and 2023. By leveraging a rich, hourly dataset and applying descriptive analysis, regression modeling, and clustering techniques, the study aims to deliver meaningful insights into how energy is produced, consumed, and priced over time.

Key questions around demand patterns, source dependency, supply balance, and price dynamics will be addressed using classical statistical tools. The analysis will not only help identify inefficiencies in the current energy mix but also highlight which sources drive pricing volatility and when the system relies heavily on imports or fossil fuels.

The results of this project will serve as a valuable resource for energy planners, policymakers, and sustainability stakeholders, enabling data-driven decisions that enhance reliability, reduce costs, and support the shift toward a cleaner and more balanced energy future.

**9. Data Source**

<https://www.kaggle.com/code/pythonafroz/eda-on-energy-deficit/notebook>