**Project Overview**

**Authors: Christopher Watson, Trevor Mcgirr, Bin Lu**

**Business Name: London Electricity**

**Publication Date: January 22, 2024**

**Github Project Link: https://github.com/Arcpolar/Smart\_Grid**

**Asana Board Link:**

**Project Background:**

1-2 paragraphs describing, at a high level, the problem you are trying to solve. This should provide the reader context for the technical solution they will review. This should be a quick elevator pitch for your project. Be sure to answer the following:

* What is the model’s objective?
* What type of Machine Learning problem will you be solving?

Traditional electricity utilities are currently going through digitization with the smart grid. With

electrification rapidly happening in many industries (i.e., the automotive industry moving to

electric vehicles), demand and consumption have been pushed to a new record high. The

fundamental problem that must be solved is balancing generational and consumption

demand. Smart grids can provide a lot of demand and consumption data

sample points for the utility to analyze, estimate, and predict demand.

Advanced Metering Infrastructure (AMI), as a fundamental component of the smart grid, records and transmits the customer consumption and demand as a load profile back to utilities. These load profile data were used for billing only at the beginning stage of the smart meter. After that, more power quality data (voltage, current, and harmonic) was measured and returned to the AMI system. Predictive regression machine learning models can be created to utilize these load profile data and provide load forecasts and predictions. These load forecasts and load predictions are critical information to control the grid and balance the load to minimize the brownout.

**Technical Background:**

1-2 paragraphs describing the technical details of the problem you are trying to solve. This should help the reader understand the project constraints. Be sure to answer the following:

* How will you evaluate your model?
* What is your data source?
  + How will you need to prepare your data?
  + How will you explore your data?
  + What do you hypothesize your main features will be?
* What type of model do you want to use?

AMI system include 5,567 London households' energy consumption readings between November 2011 and February 2014. This data set includes not only the energy consumption of these households but also weather and bank holidays within the same period. Energy consumption is affected by weather (temperature) as colder temperatures trigger heater usage, and hot weather brings air conditioning loads in. The holiday schedule will also play an essential factor in energy consumption. When people stay home, the household consumption and load profile are different from when people are away from home. All this information could create a group of load profiles to indicate customer power consumption behaviour. It may predict future load based on historical consumption, weather, holiday schedule, and load profile behaviour.

Primary Goal of the project: Create a customer load prediction regression model based on customer load profile clusters, weather information, and bank holiday schedules (Supervised learning). This model should be evaluated with test data and calculated with the mean squared error. Algorithms like linear regression, decision tree regression, or Artificial Neural Network regression model.

**Goals vs Non-Goals:**

Write a bulleted list of (3-5 points each) of goals and non-goals. This should help the reader understand the context that would factor into solution selections and trade-offs. Goals will help the reader understand what a successful outcome looks like. Non-goals will help limit the scope of your project and prevent scope creep.

Goals:

• All metering consumption information will be processed automatically

• The supervised learning regression model provides a prediction of the energy consumption (kWh) of London for one month.

• The supervised learning regression model provides a feature importance list.

Non-Goal:

• The energy consumption prediction model is not designed for the individual customers.

• The energy consumption prediction model will not make predictions for more than a month.

• The energy consumption prediction model will not take in any detailed customer information.

**Solution Overview**

1-2 paragraphs summarizing your ML System. Include a system architecture diagram containing the components you use to store data, pre-process data, engineer features, build/train/debug your model, and deploy your model. Also add notes on what you will monitor and what you will test prior to releasing a new model.

**Data Sources:**

What is your data source?

“Smart meters in London” from Kaggle.com ( <https://www.kaggle.com/datasets/jeanmidev/smart-meters-in-london> )

What is your data volume?

Details of the dataset

1. 5,567 households’ energy consumption data in kWh that is measured every 30 min from

November 2011 to February 2014, including device ID, date and time, and kWh

consumption.

2. 5,567 household information, including device ID, block, and tariff.

3. UK bank holiday schedule

4. hourly weather data from November 11, 2011, to February 15, 2014. Data include

visibility, wind bearing, temperature, dew point, pressure, apparent temperature, wind

speed, principle, icon, humidity, and summary.

Why did you select this data set?

The electricity industry has not commonly utilized machine learning or machine learning operations internally; this data set is the most detailed smart meter data available to the public.

Any risks (bias, sensitive features, etc)?

**Data Engineering:**

How will you store this data?

Data will be stored on the Amazon S3 data lake bucket.

What data pre-processing do you need to do before you feed it into your ML system?

Missing or NaN data: Missing data needs to be filled or dropped depending on the size and availability.

Feature Engineering and Feature Combine: different features must combine. Time Stamp, kWh,

Weekday, Temperature, Hours, and Holidays will all need to merge into one data frame for processing.

**Training Data:**

How will you split your data into training, test and validation?

Training data selection is from January 2013 to October 2013, and test and validation data selection is from November 2013 to February 2014.

Will you use any data labelling techniques?

Nope, this dataset is the time sequence data set; prediction of the load of kWh has been given with a time stamp.

**Feature Engineering:**

Which fields from your data will you use or exclude?

Only temperature will be used in the weather data. And households information will not be used for this project. Data will be using a hourly meter dataset, and UK bank holiday will be used as a full dataset.

Which fields will be combined or bucketed?

Time Stamp, kWh, Weekdays, Temperature, Hours, and Holidays must merge into one data frame for processing. All the meter consumption needs to be aggregated for the system load.

What other data transformations will you apply to your data?

Hours, Weekdays, and Weekends need to be extracted from the time stamp of the dataset.

**Model Training & Evaluation:**

How will you train your model?

What algorithm will you use?

This project will use LightBGM. And all 5,567 household meters data have been used for

calculating the mean of the data.

What parameters will you use?

LightBGM will

How will you evaluate your model?

**Model Deployment:**

What instance size will you use?

Will your model function as a batch or real time model? Why?

**Model Monitoring:**

How will you monitor your model?

How will you monitor your infrastructure?

How will you monitor your data?

**Model CI/CD:**

What checkpoints will your CI/CD pipeline contain?

What tests will your CI/CD pipeline contain?

**Security Checklist, Privacy and Other Risks:**

Will this store or process Personal Health Information (PHI)?

No

Will this store or process Personal Identifiable Information (PII)?

No

Will user behavior be tracked and stored?

No, not for every user, the behavior is for whole system

Will this store or process credit card information?

No

If you answered yes to any of the above questions, please justify.

What S3 buckets will this application read from or write to?

What data bias should be considered?

Will your model have potential for bias along sensitive features (race, ethnicity, gender, age, religion, disability, sexual orientation, or other personal attributes)

Are there any ethical concerts with the data or business problems that should be addressed?

**Future Enhancements:**

Provide at least 3 ways you would improve your ML system if you had more time or additional resources.–