

Forecasting hourly power demand of Peru with Amazon Sagemaker

Machine Learning Engineer Nanodegree
Capstone Proposal

April 10th, 2023

Elaborated by:

Javier Castillo Peña

Domain Background

In Peru, electricity bill for free users includes charges for generation, transmission, distribution, and other regulated fees such as subsidies. Free users are large consumers (with a maximum annual demand over 200 kW) which contract directly with a power generator the price of their electricity, while the prices for transmission and distribution are established by the national regulator.

Transmission charges may represent about 20% of the monthly electricity bill of a free user. This toll fee for using the transmission system is determined by a regulated price multiplied by the user's power demand (in kW), also known as “coincident peak demand”, at the time of the monthly national electric system's peak demand measured in 15-minute periods. This methodology by the national regulator aims to incentivize the reduction of consumption during peak hours from 5:00 pm to 11:00 pm.

In order to optimize their energy bill, free users should improve their consumption behavior by reducing their “coincident peak demand”. Besides, other “peak shaving” strategies include using locally produced electricity or energy battery storage to reduce demand from the electricity grid. In any case, predicting national power demand can help to apply those strategies to reduce “coincident peak demand” and the respective transmission charges.

Although national demand data is published by the independent system operator (COES), no open-source tool to predict power demand is available in the market. The aim of this project is to build a machine learning model with Amazon Sagemaker to forecast Peru's power demand to help consumers to optimize their “coincident peak demand” and reduce their electricity bill.

Problem Statement

The problem to be solved in this project is to forecast hourly power's demand of the Peruvian national electric system (SEIN) using time series historical data. By predicting power demand, a free user (large consumer) can make better decisions such as optimizing their “coincident peak demand” at the time of the monthly national electric system's peak demand from 5:00 pm to 11:00 pm and therefore reduce their transmission charges that can represent about 20% of their electricity bill. To solve the problem, a machine learning model will be built to forecast one-day ahead hourly power demand using historical data. Based on the characteristics of this forecasting problem, the performance of the model will be measure using the RMSE metric.

Datasets and Inputs

The data of power demand of the national electric system (SEIN) is publicly available by the independent system operator (COES), however, some tasks for extraction and transformation will be required in order to prepare the required dataset. The available data of power demand include the following:

1. **Monthly national demand** registered in 15-minute periods and reported the first days of each month. This data from meters is used to determine the monthly peak demand.¹
2. **Daily national demand** registered in 30-minute periods and reported the next day of operation. This data from the SCADA systems is close but has some differences to the monthly data from meters.²
3. **Near real-time national demand** registered and reported in 30-minute periods. Data from the SCADA systems.³

To forecast one-day ahead hourly power demand, the daily national demand data (second item) will be used. The original data is published in Excel reports for each day of operation; therefore, some scripts will be developed to download, extract, and transform the data into a single csv file. The final dataset includes the following fields:

datetime: date and time of instances.

sein_demand: power demand of the national system (SEIN) in MW.

workday: boolean field to identify the instance as a working day.

on_peak: boolean field to identify the instance as demand on peak hours from 5:00 pm to 11:00 pm.

Solution Statement

To solve the problem, a machine learning model will be built to forecast one-day ahead hourly power demand using historical data. This project will use Amazon Sagemaker platform and the AutoGluon package which performs advanced data processing, deep learning, and multi-layer model ensemble methods.⁴ The performance of the model will be measure using the RMSE metric.

Benchmark Model

A simple naïve model will be used as a benchmark model where the most recent known value is used as the predicted value at the same time the next day. In this manner, the previous day power demand at a specific time is used as the predicted value for the next day demand at the same time. In order to compare

¹ <https://www.coes.org.pe/Portal/portalinformacion/demanda?indicador=maxima>

² <https://www.coes.org.pe/Portal/PostOperacion/Reportes/leod>

³ <https://www.coes.org.pe/Portal/portalinformacion/demanda>

⁴ [How AutoGluon-Tabular works - Amazon SageMaker](#)

the performance of the benchmark model and the machine learning model, the RMSE will be calculated on the same test dataset.

Evaluation Metrics

In this project, based on the characteristics of this forecasting problem, the performance of the model will be measured using the RMSE metric. The objective is to build a machine learning model with the lowest RMSE that helps free users to make decisions on reducing “coincident peak demand” from 5:00 pm to 11:00 pm.

Project Design

The project will be developed according to the following steps:

1. Data extraction, transformation, and load.
2. Exploratory data analysis (EDA).
3. Build a benchmark model.
4. Model training on raw data.
5. Feature engineering.
6. Model training with new features.
7. Deployment on an Amazon Sagemaker endpoint