Project Scoping – Boston Public Library's Electricity Consumption Prediction based on Weather Conditions

Group 3 - Team Members:

- Team Member 1: Debanjan Saha
- Team Member 2: Ajay Bana
- Team Member 3: Akhil Krishna Nair
- Team Member 4: Sai Venkat Madamanchi
- Team Member 5: Siddharth Banyal
- Team Member 6: Venkatesh Gopinath Bogem

Table of Contents

1. INTRODUCTION **3**

2. DATASET INFORMATION **3**

2.1 DATASET INTRODUCTION: 3
2.2 DATA CARD: 4
2.3 DATA SOURCES: 5
2.4 DATA RIGHTS AND PRIVACY: 5

3. DATA PLANNING AND SPLITS **6**

4. GITHUB REPOSITORY **6**

5. PROJECT SCOPE **6**

5.1 PROBLEMS 6
5.2 CURRENT SOLUTIONS 7
5.2.1 Smart Building Management Systems (BMS) *7*
5.2.2 IoT (Internet of Things) Sensors *7*
5.2.3 Renewable Energy Integration *7*
5.3 PROPOSED SOLUTIONS 7
5.3.1 Predictive Analytics and Monitoring Software *7*

6. CURRENT APPROACH FLOW CHART AND BOTTLENECK DETECTION **8**

7. METRICS, OBJECTIVES, AND BUSINESS GOALS **8**

7.1 BUSINESS GOALS 8
7.2 OBJECTIVES 9
7.3 SUCCESS METRICS 10

8. FAILURE ANALYSIS **11**

9. DEPLOYMENT INFRASTRUCTURE **12**

9.1 INFRASTRUCTURE COMPONENTS: 13
9.1.1 GCP GKE Cluster: *13*
9.1.2 Docker Containers: *13*
9.1.3 MLFlow for Model Tracking: *13*
9.1.3 Airflow for Orchestration: *14*
9.2 DEPLOYMENT PROCESS: 14
9.2.1 CI/CD Pipeline: *14*
9.2.2 Kubernetes Deployments: *14*
9.2.3 MLFlow Integration: *14*

10. MONITORING PLAN **14**

10.1 MONITORING COMPONENTS: 14
10.1.1 Model Performance Metrics: *14*
10.1.2 Resource Utilization: *14*
10.1.3 Data Quality Checks: *15*
10.1.4 MLFlow Tracking: *15*
10.1.5 Log Management: *15*
10.2 VISUALIZATION AND REPORTING: 15
10.2.1 Grafana Dashboards: *15*
10.2.2 Kibana Visualizations: *15*

11. SUCCESS AND ACCEPTANCE CRITERIA **15**

12. TIMELINE PLANNING **15**

13. ADDITIONAL INFORMATION **16**

13.1 REFERENCES 16

1. Introduction

Electricity demand forecasting is crucial for efficient resource management in the energy sector. This project aims to leverage machine learning models to predict electricity demand based on weather conditions. By understanding the intricate relationship between weather patterns and demand fluctuations, we intend to provide energy providers with valuable insights to enhance operational efficiency.

In the dynamic landscape of the energy sector, accurately predicting electricity demand is paramount for optimizing resource allocation. Our endeavor revolves around harnessing the power of cutting-edge machine learning techniques to forecast electricity demand with precision, particularly honing in on the synchronization between weather dynamics and consumption patterns.

This project is not just about crunching numbers; it's about deciphering the intricate interplay between atmospheric conditions and energy usage.We will collect historical electricity usage data as well as weather data from multiple sources. This rich dataset consisting of temperature, humidity, wind patterns, precipitation, and other meteorological factors along with actual electricity demand will be used to train machine learning algorithms. We will experiment with different ML approaches like linear regression, random forests, and neural networks to determine the best model. Through iterative training and validation, we expect to develop an accurate demand forecasting system that can predict load requirements for the next day, next week, and even next month.

By delving deep into historical data and employing sophisticated algorithms, we strive to unravel the hidden patterns that govern demand fluctuations. Through this process, we aim to equip energy providers with actionable insights, empowering them to fine-tune their operations and meet the evolving needs of consumers effectively. Overall this project has the potential for significant cost savings and sustainable energy utilization through data-driven demand modeling capabilities. We intend to deploy the models in a cloud-based architecture so energy companies across different geographies can leverage these capabilities.

Ultimately, our mission transcends mere forecasting; it's about driving tangible improvements in operational efficiency and sustainability within the energy ecosystem. By staying at the forefront of innovation and leveraging the vast potential of machine learning, we pave the way for a future where energy resources are utilized intelligently and responsibly.

2. Dataset Information

2.1 Dataset Introduction:

The comprehensive dataset has historical electricity demand data and corresponding weather conditions, serving as the foundational cornerstone for in-depth analysis. The primary objective is to discern the intricate relationship between weather variables and electricity consumption, ultimately leading to the development of a robust forecasting model.

Purpose: The dataset is selected to dissect and quantify the impact of diverse weather conditions on electricity demand and consumption patterns. The main goal is to construct a machine-learning model that predicts electricity demand based on prevailing weather conditions.

Relevance: This dataset's significance lies in its ability to empower energy providers with invaluable insights, enabling them to enhance resource planning, minimize wastage, and elevate overall grid stability to new heights.

2.2 Data Card:

The data sources are selected to ensure broad temporal and spatial coverage, capturing variations in electricity demand and weather conditions across different seasons, and times of day. This coverage enhances the dataset's applicability for a wide range of research and forecasting applications.

Size: The dataset has an extensive temporal span, encompassing approximately 6 years of hourly data. This temporal breadth ensures a thorough examination of electricity demand fluctuations and corresponding weather conditions over an extended period.

Format: Structured in CSV (Comma-Separated Values) format, the dataset facilitates seamless manipulation and analysis using a multitude of data processing tools and programming languages.

Data Types: The dataset exhibits data types, including numerical representations of electricity demand, date-time values for precise timestamps, and categorical data categorizing different weather conditions.

This structured Data Card provides a detailed breakdown of the key features and attributes present in both the electricity demand and weather prediction datasets. It outlines the sources, format, and privacy considerations, fostering transparency and accessibility for users.

Electricity Demand Data:
- **Usage ID**: An identifier for each electricity usage entry.
- **Start Time:** Timestamp indicating the beginning of the usage period.
- **End Time**: Timestamp indicating the end of the usage period.
- **Usage (kW):** The electricity demand measured in kilowatts during the specified period.

Weather Predictions Data
- **Number**: An index or identifier column for each weather prediction entry.
- **Name**: Descriptive name for the weather prediction entry.
- **Start Time**: Timestamp marking the start of the weather prediction period.
- **End Time**: Timestamp marking the end of the weather prediction period.
- **Is Daytime**: Boolean indicating whether it is daytime during the prediction.
- **Temperature**: Temperature during the prediction period.
- **Temperature Unit**: Unit of temperature measurement (e.g., Fahrenheit).
- **Temperature Trend**: Trend in temperature during the prediction period.
- **Probability of Precipitation:** Likelihood of precipitation during the prediction period.
- **Dewpoint:** Dewpoint temperature during the prediction period.
- **Relative Humidity:** Humidity level during the prediction period.
- **Wind Speed:** Wind speed during the prediction period.
- **Wind Direction:** Wind direction during the prediction period.
- **Icon:** Icon representing the weather condition.
- **Short Forecast:** Brief description of the weather forecast.

2.3 Data Sources:

Electricity demand data is curated from the Data Boston API. This entity is recognized for capturing and recording electricity consumption patterns. The data collection methodology includes real-time monitoring and reporting, ensuring the dataset's fidelity in reflecting the actual demand on the grid.

Simultaneously, weather data is sourced from the live National Weather Service (NWS API). This source is chosen for its reliability and comprehensive coverage of meteorological variables crucial for understanding the dynamic interplay between weather conditions and electricity demand. The weather data is derived from advanced meteorological instruments, satellite observations, and ground-based stations, providing a multi-dimensional view of atmospheric conditions.

2.4 Data Rights and Privacy:

Data Compliance: The dataset aligns with GDPR, exemplifying adherence to the highest standards of data protection and privacy.

Privacy Considerations: Prioritizing privacy, the dataset is anonymization, safeguarding PII information. By meticulously removing personally identifiable details, the dataset ensures the utmost privacy for consumers while still facilitating profound insights into electricity demand patterns.

3. Data Planning and Splits

Our data pre-processing involves handling missing values, scaling numerical features, and encoding categorical variables. Given the temporal nature of the data, a time-based split is adopted to maintain chronological order, ensuring the model is trained on historical data and tested on future periods.

In light of the temporal dynamics inherent in our data, a thoughtful time-based split strategy is embraced. This split methodology prioritizes maintaining chronological integrity, thereby fortifying the model's ability to glean insights from historical trends and generalize effectively to future periods. By adhering to this temporal framework, we lay a robust foundation for training our models on past data while rigorously evaluating their performance on unseen future data, thus fostering a comprehensive understanding of electricity demand patterns across time.

We will use the first few years of hourly electricity demand data paired with corresponding weather predictions to train our models. The later few years will be held out as a test set to evaluate demand forecasting accuracy on future data. Within the training data, we will further split off the few most recent months to serve as our validation set for hyperparameter tuning and model selection.

This rigorous time-based splitting approach ensures our models learn from a sizable portion of historical data, while still being evaluated on unseen future data. The chronological order also realistically mimics the practical use case of using past electricity demand patterns and weather conditions to forecast future load requirements.

4. GitHub Repository

- GitHub Repository Link: https://github.com/debanjansaha-git/mlops-group3
- **Folder Structure:**

```
.
├── data                 # Data Collected (alternatively `dataset`)
├── docs                 # Documentation files (alternatively `doc`)
├── environments_setup      # Setup files for GCP (alternatively `setup` or `build`)
├── images                 # Images used (alternatively `assets`)
├── models                 # Models used
├── src                 # Source files (alternatively `lib` or `app`)
├── utils                 # Tools and utilities
├── LICENSE
└── README.md
```

5. Project Scope

5.1 Problems

Public spaces, particularly libraries, present unique challenges in efficient energy management. Public spaces, being frequented by diverse user groups, require meticulous attention to energy consumption. Libraries, as essential community hubs, operate for extended hours, making energy efficiency crucial. Traditional management approaches often struggle to meet the dynamic demands of such spaces, necessitating a more sophisticated solution. Maintaining these spaces demands a structured approach to effectively address energy consumption.

The primary objective of this study is to develop and implement a forecast system tailored to the specific energy needs of libraries. By leveraging predictive analytics and advanced modeling, we aim to enhance the energy management infrastructure, ensuring optimal resource utilization while maintaining a comfortable and conducive environment for library users.

5.2 Current Solutions

Here are some current solutions that were being explored or implemented for energy management in public spaces:

5.2.1 Smart Building Management Systems (BMS)

BMS integrates various building systems, including HVAC, lighting, and security, to enhance overall energy efficiency. These systems often incorporate sensors, real-time monitoring, and automation to optimise energy usage based on occupancy, environmental conditions, and other factors. Smart lighting systems with occupancy sensors can adjust lighting levels based on the number of people in a given area. This helps to minimize energy wastage in unoccupied spaces

5.2.2 IoT (Internet of Things) Sensors

IoT sensors can be deployed throughout a library to collect real-time data on occupancy, temperature, and lighting conditions. This data can then be analysed to make informed decisions about energy consumption and resource allocation.

5.2.3 Renewable Energy Integration

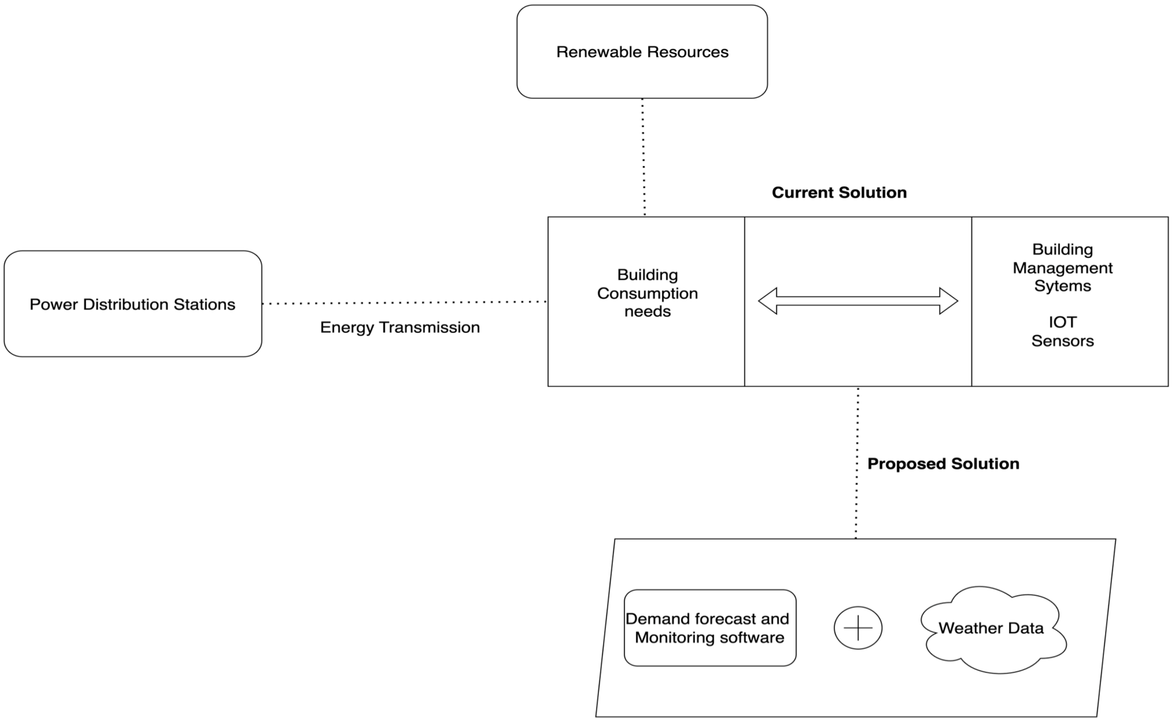
Libraries and other public spaces are increasingly incorporating renewable energy sources, such as solar panels or wind turbines, to supplement their energy needs.

5.3 Proposed Solutions

5.3.1 Predictive Analytics and Monitoring Software

Predictive analytics tools use historical data and machine learning algorithms to forecast future energy usage patterns. By analysing past trends and considering external factors (such as weather forecasts or special events), these tools can provide insights for proactive energy management. Developing software platforms with dashboards and reports helps organizations monitor, analyse, and optimise energy usage. These systems often allow facility managers to track energy consumption trends and identify areas for improvement.

6. Current Approach Flow Chart and Bottleneck Detection



While the existing energy management system efficiently leverages sensor technology, management tools, and renewable resources to optimize energy utilization, there is a significant opportunity to enhance this optimization further. By integrating real-time energy consumption monitoring with external weather condition data, and utilizing this information for predictive demand forecasting, a more refined, data-driven decision-making process can be established. The development and implementation of bespoke software solutions and the requisite infrastructure to support this advanced analytics capability would mark a transformative advancement in energy management practices.

7. Metrics, Objectives, and Business Goals

7.1 Business Goals

Forecasting energy demand for the Boston Public Library (BPL) based on historical energy demand data and weather data can bring several business benefits to the institution, including, but not limited to the following:

7.1.1 *Cost Optimization*

By accurately forecasting energy demand, the Boston Public Library can optimize its energy consumption, leading to potential cost savings. The library can adjust heating, cooling, and lighting systems based on predicted demand, reducing unnecessary energy usage during off-peak hours.

7.1.2 *Efficient Maintenance Scheduling*

With reliable forecasts, the BPL can better allocate resources to meet energy demand efficiently. This includes scheduling maintenance activities during periods of lower energy consumption and ensuring sufficient staffing levels during peak demand times.

7.1.3 *Environmental Sustainability*

By reducing energy waste and optimizing consumption, the BPL can decrease its carbon footprint and contribute to environmental sustainability efforts. This aligns with the library's commitment to green initiatives and demonstrates responsible stewardship of resources. The government has recently launched many campaigns such as Go Boston 2030 (2017), Carbon Free Boston (2019), Zero-Waste Boston (2019) all of which cater to the same mission.

7.1.4 *Enhanced User Experience*

Consistent temperature and lighting levels contribute to a comfortable and inviting environment for library visitors. By forecasting energy demand and optimizing HVAC and lighting systems accordingly, the BPL can ensure a pleasant experience for patrons, leading to increased satisfaction and repeat visits. Increasing crowds and worsening weather conditions can be accounted for and a smooth experience can be provided.

7.1.5 *Data-Driven Insights*

Implementing an AI solution for energy demand forecasting generates valuable data insights over time. By analysing trends and patterns in energy consumption alongside weather conditions, the BPL gains a deeper understanding of its operational dynamics, which can inform future strategies and investments.

7.2 Objectives

This project's primary goal is to create a predictive model for a commercial building's (*Boston Public Library* in this case) *energy consumption* using a variety of machine learning techniques in a cloud-based machine learning platform. The accuracy of the methodology used to predict energy consumption is the main focus of this project. Since it is essential to lower the energy consumption of all types of buildings, from residential to industrial, advances in machine learning research have a significant impact on the subject of smart building energy management.

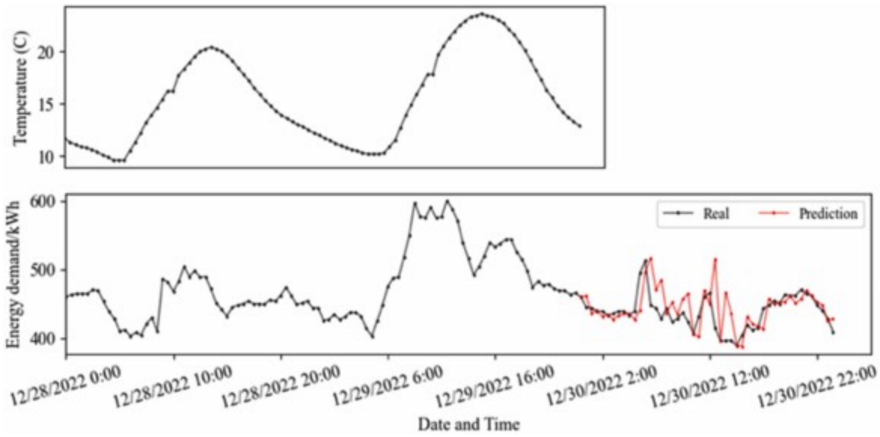


Fig. [9] Goals of prediction: Energy Demand prediction accurate to actual consumption

7.3 Success Metrics

This project is focused on energy forecasting using weather data, incorporating Continuous Training (CT), alongside CI/CD, Continuous Monitoring (CM), and dynamic dashboards for real-time metrics, the success criteria can be streamlined as follows:

1. Automated CI/CD and Continuous Training (CT) Workflow:
   - Efficient automation of data ingestion, model retraining, evaluation, and deployment processes to adapt to new weather data and consumption patterns.
   - Seamless integration and deployment of updates with minimal manual effort, ensuring the model stays current with the latest data and algorithms.

2. Continuous Monitoring (CM) and Dashboards:
   - Effective real-time monitoring of model performance (e.g., forecasting accuracy) and operational metrics (e.g., latency, throughput).
   - Interactive dashboards that provide insights into model health, data quality, and the impact of weather on energy consumption.
   - Automated alerts for model drift, data anomalies, or performance degradation, prompting timely adjustments.

3. Model and Data Management:
   - Robust version control for models and datasets, enabling traceability and quick rollback if needed.
   - High-quality data ingestion and preprocessing to ensure accurate and reliable forecasts.

4. Scalability and Efficiency:
   - Scalable architecture to handle varying volumes of weather data and energy consumption records.
   - Optimized resource management, balancing computational costs with forecast accuracy and timeliness.

5. Adaptability and Continuous Improvement:

- Flexibility to incorporate new data sources, weather patterns, or consumption trends.
- Commitment to iterative improvement through regular feedback loops and model updates.

Success in this context is defined not just by technical robustness but also by the model's ability to deliver actionable insights, drive operational efficiencies, and adapt to evolving data landscapes and business needs in the energy sector.

8. Failure Analysis

To address potential risks, a comprehensive failure analysis strategy is essential.

Failures in data pipelines can significantly impact data analysis, business intelligence, and decision-making processes. Here are some common examples of failures in pipelines:

- Missing or incomplete data due to extraction errors or source system availability issues. Transformation errors, leading to inaccurate analytics and business intelligence insights.
- Slow processing due to inefficient code or inadequate hardware resources, causing delays in data availability.
- Inability to handle increased data volumes or new data sources, leading to system overload or significant performance degradation.
- Hardcoded or inflexible designs that make it difficult to adapt to changing data processing requirements.
- Failures in external services or data sources that the pipeline depends on, cause data ingestion issues.
- Inconsistent data across different stages of the pipeline or when integrating data from multiple sources.

Building software solutions to forecast energy consumption using weather data involves complex data processing and modeling steps. Several potential failures can occur during this process, affecting the accuracy and reliability of the forecasts. Here are some common issues:

- Including features that have little to no predictive power can reduce model performance.
- Failing to include critical weather parameters or derived features that significantly influence energy consumption.
- Building a model that is too complex for the available data can lead to overfitting, where the model performs well on training data but poorly on unseen data.
- Conversely, a model that is too simple may not capture the underlying relationships between weather conditions and energy consumption, resulting in underfitting.

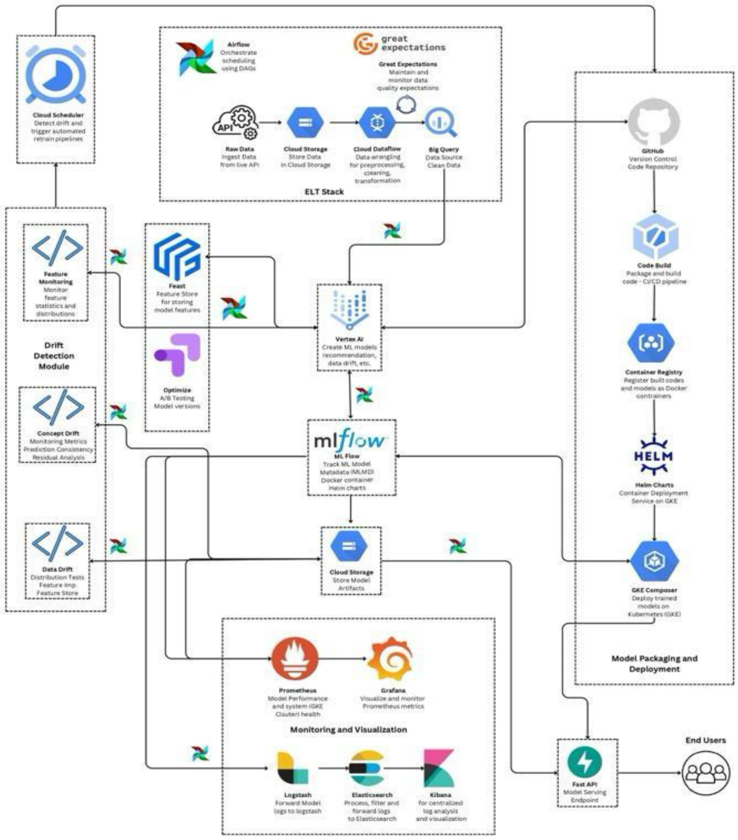
- Failing to account for seasonal variations, time dependencies, and lag effects between weather conditions and energy usage.
- Using models that cannot adapt to changing patterns over time, such as new consumption habits or weather trends.
- Failure to regularly update the model with new data or retrain it to adapt to changing patterns.

Effective monitoring, robust error handling, and regular maintenance are essential to mitigate these failures. Implementing best practices in data pipeline design, such as ensuring data quality, scalability, security, and fault tolerance, can help in preventing these issues and maintaining reliable data processing systems.

In summary, a proactive approach involving ongoing monitoring, well-defined triggers, and thorough documentation serves as a foundation for robust failure analysis. This strategy enables quick identification and mitigation of issues, contributing to the overall effectiveness of machine learning systems and their adaptability to changing conditions.

9. Deployment Infrastructure

Our deployment infrastructure will be hosted on Google Cloud Platform (GCP) using Google Kubernetes Engine (GKE) for efficient container orchestration. This choice provides scalability, ease of management, and integration with various GCP services.



End-to-End MLOps Workflow on GCP
System Design Architecture
Concept: Debanjan Saha

9.1 Infrastructure Components:

9.1.1 GCP GKE Cluster:

- GKE will serve as the foundation for hosting our machine learning model containers. It allows for automated scaling, management, and Helm charts for Kubernetes orchestration, ensuring a robust and resilient deployment.

9.1.2 Docker Containers:

- The machine learning model, along with its dependencies, will be containerized using Docker. This ensures consistent deployment across different environments, facilitating reproducibility.

9.1.3 MLFlow for Model Tracking:

- MLFlow will be integrated into our MLOps pipeline for comprehensive model tracking and management. It provides capabilities for tracking experiments, packaging code, and sharing and deploying models.

9.1.3 Airflow for Orchestration:

- Most of our non-kubernetes components will be orchestrated using Apache Airflow as it allows the creation of DAGs which facilitate seamless scheduling periodic tasks like executing data flows, data pre-processing, running experiments, executing custom tasks, and much more.

9.2 Deployment Process:

9.2.1 CI/CD Pipeline:

- A continuous integration and continuous deployment (CI/CD) pipeline will be established to automate the deployment process using Code Build. This pipeline will include steps for testing, building Docker images, deploying to GKE, and managing MLFlow experiments.

9.2.2 Kubernetes Deployments:

- Helm charts containing Kubernetes manifests will define the deployment specifications for our machine learning model and accompanying services. These manifests will be version-controlled and applied to the GKE cluster as part of the CI/CD process.

9.2.3 MLFlow Integration:

- MLFlow server components will be deployed as part of the GKE cluster. MLFlow Tracking will be integrated to log and organize experiments, parameters, metrics, and artifacts. MLFlow Models will enable easy model versioning and deployment.

10. Monitoring Plan

A robust monitoring plan is essential for ensuring the continuous health, performance, and reliability of our electricity demand forecasting model. The monitoring plan encompasses various aspects of the MLOps pipeline, including model performance, system metrics, and data quality. The integration of Prometheus, Grafana, and the ELK stack will play a pivotal role in capturing and visualizing these metrics.

10.1 Monitoring Components:

10.1.1 Model Performance Metrics:

- **Metrics Tracked:** Mean Absolute Error (MAE), Root Mean Squared Error (RMSE).
- **Monitoring Frequency:** Real-time monitoring, updated every hour.
- **Alerts:** Trigger alerts if MAE or RMSE deviates significantly from the baseline or exceeds a predefined threshold.

10.1.2 Resource Utilization:

- **Metrics Tracked:** CPU and memory usage of the deployed model containers.
- **Monitoring Frequency:** Real-time monitoring with Prometheus.

- **Alerts:** Notify if resource utilization approaches predefined limits to prevent performance degradation.

10.1.3 Data Quality Checks:
- **Metrics Tracked:** Missing values, outliers, and distribution shifts in incoming data.
- **Monitoring Frequency:** Daily batch checks and real-time streaming checks.
- **Alerts:** Flag anomalies in the data distribution or significant data quality issues.

10.1.4 MLFlow Tracking:
- **Metrics Tracked:** Experiment metrics, model versions, and deployment artifacts.
- **Monitoring Frequency:** Continuous tracking with every model update.
- **Alerts:** Notify if there are discrepancies in logged metrics or issues with model versions.

10.1.5 Log Management:
- **Logs Tracked:** Deployment logs, application logs, and error logs.
- **Monitoring Frequency:** Real-time log streaming with the ELK stack.
- **Alerts:** Alert on critical errors or unusual patterns in logs that may indicate issues.

10.2 Visualization and Reporting:

10.2.1 Grafana Dashboards:
- Customized Grafana dashboards will provide a visual representation of model performance, resource utilization, and other critical metrics. These dashboards will enable the operations team to quickly identify trends and potential issues.

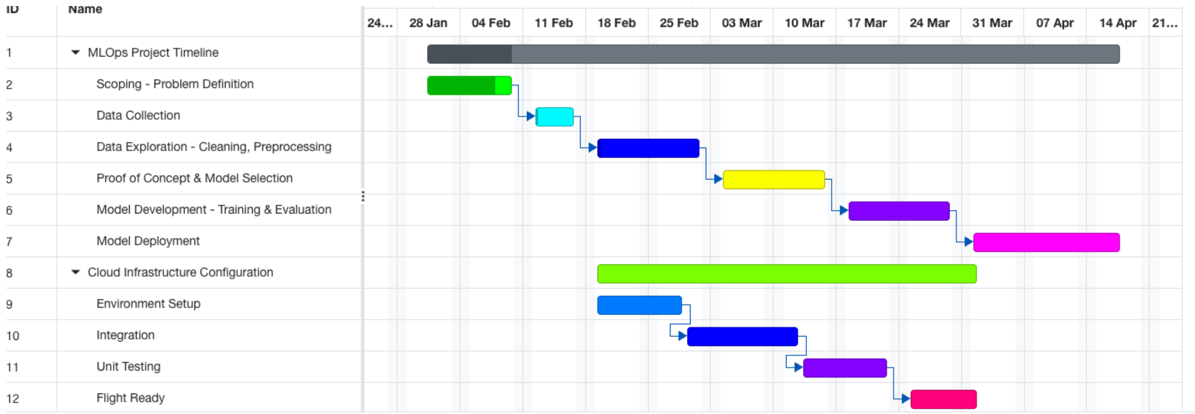
10.2.2 Kibana Visualizations:
- Kibana will be used to create visualizations for log data, allowing for efficient analysis of log patterns and facilitating troubleshooting.

11. Success and Acceptance Criteria

Success defined by achieving accuracy goals. Acceptance criteria include stakeholder approval and successful deployment.

12. Timeline Planning

We have planned a phase-by-phase execution of this project starting from scoping to data collection, creating a proof-of-concept (POC), model building, testing, evaluating, to final deployment and monitoring. While some of the critical components of our architecture are being developed we also plan to bring up our cloud development environment in Google Cloud Platform. Below is a gantt chart showing our development activities in phase by phase manner.



- Data Collection and Preprocessing: 3 weeks
- Model Development and Training: 4 weeks
- Testing and Validation: 3 weeks
- Deployment and Monitoring: 3 weeks

13. Additional Information

13.1 References

1. Zhao, H., & Magoulès, F. (2012). A review on the prediction of building energy consumption. *Renewable & Sustainable Energy Reviews, 16*, 3586-3592.
2. Heidarinejad, M., Cedeño-Laurent, J.G., Wentz, J.R., Rekstad, N.M., Spengler, J.D., & Srebric, J. (2017). Actual building energy use patterns and their implications for predictive modeling. *Energy Conversion and Management, 144*, 164-180.
3. Kim, Y., Heidarinejad, M., Dahlhausen, M., & Srebric, J. (2017). Building energy model calibration with schedules derived from electricity use data. *Applied Energy, 190*, 997-1007.
4. Ruijin, Z., Yiqun, P., Zhizhong, H., & Qiujian, W. (2013). Building Energy Use Prediction Using Time Series Analysis. *2013 IEEE 6th International Conference on Service-Oriented Computing and Applications*, 309-313.
5. Global Protocol for Community-Scale Greenhouse Gas Emission Inventories (GHG). November 14, 2017, https://ghgprotocol.org/sites/default/files/standards/GPC_Full_MASTER_RW_v7.pdf
6. City of Boston, Boston Greenhouse Gas Inventory Methodology, August 2023, https://docs.google.com/document/d/1DoN9etK_emn3C2glSAXJCFGKl1Gwru99jR6GwrRtBYE
7. NOAA Local Climatological Data Publication, August 2021, https://www.ncdc.noaa.gov/IPS/lcd/lcd.html

8. Mel Keytingan M. Shapi, Nor Azuana Ramli, Lilik J. Awalin. Energy consumption prediction by using machine learning for smart building: Case study in Malaysia https://www.sciencedirect.com/science/article/pii/S266616592030034X#coi0010
9. Ammar Kamoona, Hui Song, Kian Keshavarzian, Kedem Levy, Mahdi Jalili, Richardt Wilkinson, Xinghuo Yu, Brendan McGrath, Lasantha Meegahapola. Machine learning based energy demand prediction, Energy Reports https://www.sciencedirect.com/science/article/pii/S2352484723013975