

Designing Building Energy Prediction

Internship Report

Submitted for the partial fulfilment of the degree of

Bachelor of Technology

In

Mathematics & Computing

Submitted By

Akash Gaur

0901MC20101

UNDER THE SUPERVISION AND GUIDANCE OF

Dr. Jitendra Kumar

Associate Professor

Department of Engineering Mathematics & Computing



माधव प्रौद्योगिकी एवं विज्ञान संस्थान, ग्वालियर (म.प्र.), सार
MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR (M.P.), INDIA

Deemed to be University

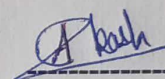
NAAC ACCREDITED WITH A++ GRADE

January-June 2024

DECLARATION BY THE CANDIDATE

I hereby declare that the work entitled "**Building Energy Prediction**" is my work, conducted under the supervision of **Dr. Jitendra Kumar, Associate Professor**, during the session Jan-May 2024. The report submitted by me is a record of bonafide work carried out by me.

I further declare that the work reported in this report has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.



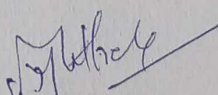
Date: 17-May-2024

Akash Gaur

Place: Gwalior

0901MC20101

This is to certify that the above statement made by the candidates is correct to the best of my knowledge and belief.



Guided By:

Dr. Jitendra Kumar

Associate Professor

Department of Engineering Mathematics & Computing
MITS, Gwalior

Departmental Project Coordinator

Dr D K Jain
Professor

Department of Engineering Mathematics &
Computing, MITS, Gwalior

Approved by HoD

Dr. Vikas P Shinde
Professor & Head


Department of Engineering Mathematics &
Computing, MITS, Gwalior

PLAGIARISM CHECK CERTIFICATE

This is to certify that I/we, a student of B.Tech. in **Engineering Mathematics & Computing** have checked my complete report entitled "**Building Energy Prediction**" for similarity/plagiarism using the "Turnitin" software available in the institute.

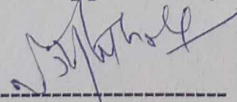
This is to certify that the similarity in my report is found to be 1.3% which is within the specified limit (20%).

The full plagiarism report along with the summary is enclosed.



Akash Gaur
0901MC20101

Checked & Approved By:



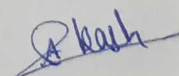
Dr. Jitendra Kumar
Associate Professor
Department of Engineering Mathematics & Computing
MITS, Gwalior

ACKNOWLEDGEMENT

Completing the semester-long internship has been a defining moment for my career, and I want to express my appreciation to Madhav Institute of Technology and Science for enabling me to pursue my disciplinary or interdisciplinary project as a mandatory part of the curriculum. The Flexible Curriculum Scheme, has given me the flexibility to explore my interests and develop my skills. I would like to thank Dr. R.K., the Director of the institute, for his support throughout this journey.

I would sincerely like to thank my department, the Department of Engineering of Mathematics and Computing, for allowing me to explore this internship. I humbly thank Dr Vikas P Shinde, Professor and Head, Department of Engineering Mathematics and Computing, for his continued support during the course of this engagement, which eased the process and formalities involved.

I am sincerely thankful to my faculty mentors. I am grateful to the guidance of Dr. Jitendra Kumar, Associate Professor of Engineering Mathematics and Computing, MITS, Gwalior for his continued support and guidance throughout the internship. I would also like to express my gratitude to the department's faculty and staff for their invaluable support and efforts.



Akash Gaur
0901MC20101

EXECUTIVE SUMMARY (FOR INTERNSHIP)

The efficient management of energy consumption in buildings is crucial for reducing environmental impact and optimizing resource utilization. In this project, we explore the application of machine learning techniques for predicting building energy consumption. Leveraging a dataset containing various features such as building characteristics, weather data, and historical energy usage, our goal is to develop accurate predictive models that can forecast future energy consumption patterns.

The project begins with exploratory data analysis to gain insights into the relationships between different variables and their impact on energy consumption. Feature engineering techniques are employed to extract relevant information from the dataset, including time-series analysis of historical energy usage and weather patterns.

We then implement and evaluate several machine learning algorithms, including linear regression, random forest, and gradient boosting, to predict energy consumption. Model performance is assessed using metrics such as Root Mean Squared Error (RMSE), and coefficient of determination (R-squared).

Our results demonstrate the effectiveness of machine learning approaches in accurately predicting building energy consumption. The developed models provide valuable insights into the factors influencing energy usage and can be utilized for energy-efficient building management and policy-making initiatives.

Overall, this project highlights the importance of data-driven approaches in addressing energy challenges and underscores the potential of machine learning in optimizing energy efficiency in buildings.

ACKNOWLEDGEMENT

Completing the semester-long internship has been a defining moment for my career, and I want to express my appreciation to Madhav Institute of Technology and Science for enabling me to pursue my disciplinary or interdisciplinary project as a mandatory part of the curriculum. The Flexible Curriculum Scheme, has given me the flexibility to explore my interests and develop my skills. I would like to thank Dr. R.K., the Director of the institute, for his support throughout this journey.

I would sincerely like to thank my department, the Department of Engineering of Mathematics and Computing, for allowing me to explore this internship. I humbly thank Dr Vikas P Shinde, Professor and Head, Department of Engineering Mathematics and Computing, for his continued support during the course of this engagement, which eased the process and formalities involved.

I am sincerely thankful to my faculty mentors. I am grateful to the guidance of Dr. Jitendra Kumar, Associate Professor of Engineering Mathematics and Computing, MITS, Gwalior for his continued support and guidance throughout the internship. I would also like to express my gratitude to the department's faculty and staff for their invaluable support and efforts.

Akash Gaur
0901MC20101

CERTIFICATE OF INTERNSHIP



+91-8277-111444
contact@kloudspot.com
www.kloudspot.com

Date: May 10th, 2024

CERTIFICATE OF INTERNSHIP

This is to certify that **Mr. Akash Gaur**(ID: 0901MC201001, Mathematics and Computing, Madhav Institute of Technology & Science, Gwalior) is engaged as an intern in our organization from Feb 1st 2024 to May, 31st 2024. Post completion of his internship, he will be joining us as Software Engineer, as per our offer letter.

During this period Akash, has met all expectations of the project and his conduct was found to be satisfactory.

We take this opportunity to thank and wish him best in future.

Sincerely,
For KLOUDSPOT INDIA PRIVATE LIMITED,

Neetu Goyal
HR and Finance Manager



KLOUDSPOT INDIA PRIVATE LIMITED (A 100% subsidiary of Kloudspot Inc. USA)
Registered Office: FC052, Founders Cube, Plot No. 77, JBR Tech Park, 6th Rd, Whitefield EPIP, Bangalore, KA 560066
CIN: U62099KA2023FTC172996

CONTENT

Table of Contents	
Declaration by the Candidate.....	i
Plagiarism Check Certificate	ii
Executive Summary (For Internship).....	i
Acknowledgement	iv
Certificate of Internship	iii
Content.....	vi
List of Figures	vii
List of Tables	viii
Chapter 1: Introduction	1
Chapter 2: Literature Survey.....	2
Chapter 3: Company Profile	4
Chapter 4: Problem Formulation	5
Chapter 5: Proposed Methodology	6
Chapter 6: Dataset Detail	10
Chapter 7: Exploratory Data Analysis (EDA)	12
Chapter 8: Pycaret Result Analysis.....	15
Chapter 9: ML Training Result.....	16
Chapter 10: ANN Training Result	22
Chapter 11: Daily Model Training Result	24
References.....	25
Turnitin Plagiarism Report	26
Annexure-1	27
Learning Outcomes.....	27

Annexure-2	28
Daily Diary.....	28
Annexure-3a.....	29
MPR-1.....	29
Annexure-3b	30
MPR-2.....	30
Annexure-3c.....	31
MPR-3.....	31
Annexure-4 The End.....	32

LIST OF FIGURES

Fig 1: Feature Summary

Fig 2: Site Geographical Location

Fig 3: Correlation Heatmap

Fig 4: Meter Reading Distribution

Fig 5: Meter Reading Pie chart

Fig 6: Site Wise Energy Consumption

Fig 7: Distribution Of Building by Primary Use

Fig 8: Energy Trend Through the Day

Fig 9: Energy Trend Through the year

Fig 10: Distribution Of Weather Features

Fig 11: Trends in Weather Features

Fig 12: Model Training Result without Weather Features for Electricity

Fig 13: Model Training Result without Weather Features for Water

Fig 14: Model Training Result with Weather Features for Electricity

Fig 15: Model Training Result with Weather Features for Water

Fig 16: ANN Training Summary

Fig 17: ANN Model Training Result for Electricity

Fig 18: ANN Model Training Result for Water

Fig 19: Daily Model Training Result for Electricity

Fig 20: Daily Model Training Result for Water

LIST OF TABLES

Table 1: Categorical Features Description

Table 2: Dataset Summary

Table 3: Pycaret Parameters

Table 4: Pycaret Result

Table 5: ANN Model Summary

Table 6: Daily Diary

CHAPTER 1: INTRODUCTION

Energy consumption in buildings constitutes a significant portion of overall energy usage, making it a key focus area for sustainability efforts and cost optimization. The ability to accurately predict building energy consumption can enable better resource management, reduced environmental impact, and improved energy efficiency. In this project, we delve into the realm of building energy prediction using machine learning techniques.

The project aims to develop predictive models that can forecast building energy consumption based on various factors such as building characteristics, weather conditions, occupancy patterns, and historical energy usage data. By leveraging a diverse dataset encompassing these features, our objective is to harness the power of data-driven approaches to optimize energy management strategies.

Through exploratory data analysis, feature engineering, and machine learning model development, we seek to uncover insights into the complex relationships between different variables and their influence on energy consumption patterns. By understanding these relationships, we aim to create robust predictive models capable of accurately forecasting future energy usage trends.

The outcomes of this project have the potential to revolutionize energy management practices in buildings, offering stakeholders valuable insights and tools for making informed decisions regarding energy efficiency initiatives, resource allocation, and sustainability goals. By harnessing the predictive power of machine learning, we endeavor to pave the way towards a more sustainable and energy-efficient future for buildings and communities.

CHAPTER 2: LITERATURE SURVEY

The prediction of building energy consumption has garnered significant attention in recent years, driven by the increasing emphasis on sustainability, energy efficiency, and environmental conservation. A review of the existing literature reveals a wealth of research focusing on various aspects of building energy prediction, ranging from modeling techniques to data sources and applications.

Modelling Techniques:

Several modelling techniques have been explored for building energy prediction, including traditional statistical methods and machine learning algorithms. Linear regression models have been commonly used for their simplicity and interpretability, but they may not capture the complex nonlinear relationships inherent in building energy data. In contrast, machine learning algorithms such as random forests, gradient boosting, and neural networks offer greater flexibility and predictive power, allowing for more accurate and robust energy consumption forecasts.

References:

1. Hong, T., & Lin, H. (2013). A review of building energy modeling for control and operation. *Renewable and Sustainable Energy Reviews*, 22, 34-45.
2. Zhao, L., Magouleès, F., & Sicard, P. (2019). A review of building energy consumption prediction models: Data-driven and hybrid approaches. *Renewable and Sustainable Energy Reviews*, 104, 275-291.

Data Sources:

Building energy prediction models rely on a diverse range of data sources, including building characteristics, weather data, occupancy patterns, and historical energy usage. The availability and quality of these data sources significantly impact the accuracy and reliability of energy consumption forecasts. Integration of advanced sensing technologies, IoT devices, and building management systems has enabled real-time data collection and enhanced model performance.

References:

1. Kumar, S., & Wahab, M. A. (2020). A review of data-driven and machine learning approaches for building energy consumption prediction. *Applied Energy*, 268, 115074.
2. Afshari, S., Mahyari, A., & Sayyaadi, H. (2018). Review of prediction models for building energy consumption. *Renewable and Sustainable Energy Reviews*, 82, 110-124.

Applications:

Predictive models for building energy consumption have diverse applications across various sectors, including commercial buildings, residential complexes, educational institutions, and healthcare facilities. These models are utilized for energy management, demand response, optimization of HVAC systems, and decision-making support for building operators and policymakers.

References:

1. Jin, X., & Jin, Y. (2016). Review of building energy-use prediction models. *Renewable and Sustainable Energy Reviews*, 56, 368-379.
2. Tsanas, A., Xifara, A., & Karamanos, K. (2020). A review of building energy consumption prediction studies: From statistical models to machine learning algorithms. *Energy and Buildings*, 225, 110365.

In summary, the literature review underscores the importance of building energy prediction as a critical component of energy management and sustainability efforts. By leveraging advanced modelling techniques and integrating diverse data sources, predictive models hold immense potential for optimizing energy usage, reducing costs, and mitigating environmental impact in buildings.

CHAPTER 3: COMPANY PROFILE



Ground Floor, IndiQube, Gamma 154, 172, Outer Ring Rd, Kadubeesanahalli, Marathahalli, Bengaluru, Karnataka 560103

Kloudspot provides a platform for transforming physical locations with data. Enhancing people's work experiences and customer connections is the company's aim. The goal of Kloudspot's platform is to establish secure, interesting, and cooperative work environments.

Location-based services, actionable insights, and real-time data and analytics are a few of Kloudspot's capabilities. Additionally, Kloudspot may be linked with already-in-use corporate software.

The post lists a number of advantages of utilizing Kloudspot, such as higher income, lower costs, and better staff engagement. Numerous sectors, including retail, healthcare, and transportation, use Kloudspot.

The platform developed by Kloudspot collects data from multiple sources, including Bluetooth signals, Wi-Fi, temperature and humidity sensors, and even cameras that are connected. They evaluate this data using AI, offering real-time insights on people's move about and communicate in a physical area.

These observations are useful in addition to being instructive. Companies can use this data to tailor employee and consumer experiences. Consider a retail establishment that uses Kloudspot to provide customized promotions to a consumer's phone depending on where they are in the store.

CHAPTER 4: PROBLEM FORMULATION

Building energy consumption represents a significant portion of overall energy usage and is influenced by various factors such as building characteristics, weather conditions, occupancy patterns, and operational practices. The ability to accurately predict building energy consumption is crucial for optimizing energy management strategies, reducing environmental impact, and achieving sustainability goals. However, building energy prediction poses several challenges, including the complexity of energy usage patterns, the dynamic nature of building environments, and the availability of diverse and heterogeneous data sources.

In this project, our primary objective is to develop predictive models that can forecast building energy consumption with high accuracy and reliability. Specifically, we aim to address the following key aspects:

- 1. Data Collection and Preprocessing:** We will gather a comprehensive dataset containing information on building characteristics, weather data, historical energy usage, and other relevant factors. This dataset will be subjected to thorough preprocessing steps, including data cleaning, normalization, and feature engineering, to ensure its suitability for model training.
- 2. Model Development:** We will explore various machine learning algorithms, including linear regression, random forest, gradient boosting, and neural networks, to develop predictive models for building energy consumption. These models will be trained using historical energy usage data and other relevant features to learn the underlying patterns and relationships.
- 3. Evaluation and Validation:** The performance of the developed models will be evaluated using appropriate metrics such as Root Mean Squared Error (RMSE) and coefficient of determination (R-squared). We will conduct rigorous validation experiments to assess the generalization ability of the models across different time periods and building types.
- 4. Insights and Interpretation:** We will analyse the insights gained from the predictive models to understand the key factors influencing building energy consumption. This analysis will provide valuable information for optimizing energy management strategies, identifying areas for efficiency improvements, and informing decision-making processes.

Building Energy Prediction

OVERVIEW

Building energy prediction and performance analytics offer significant opportunities to save energy, reduce carbon emissions, and lower operating costs for building owners. Machine learning and prediction techniques are essential in identifying energy-saving opportunities and assessing the risks and benefits of implementing changes.

GOALS

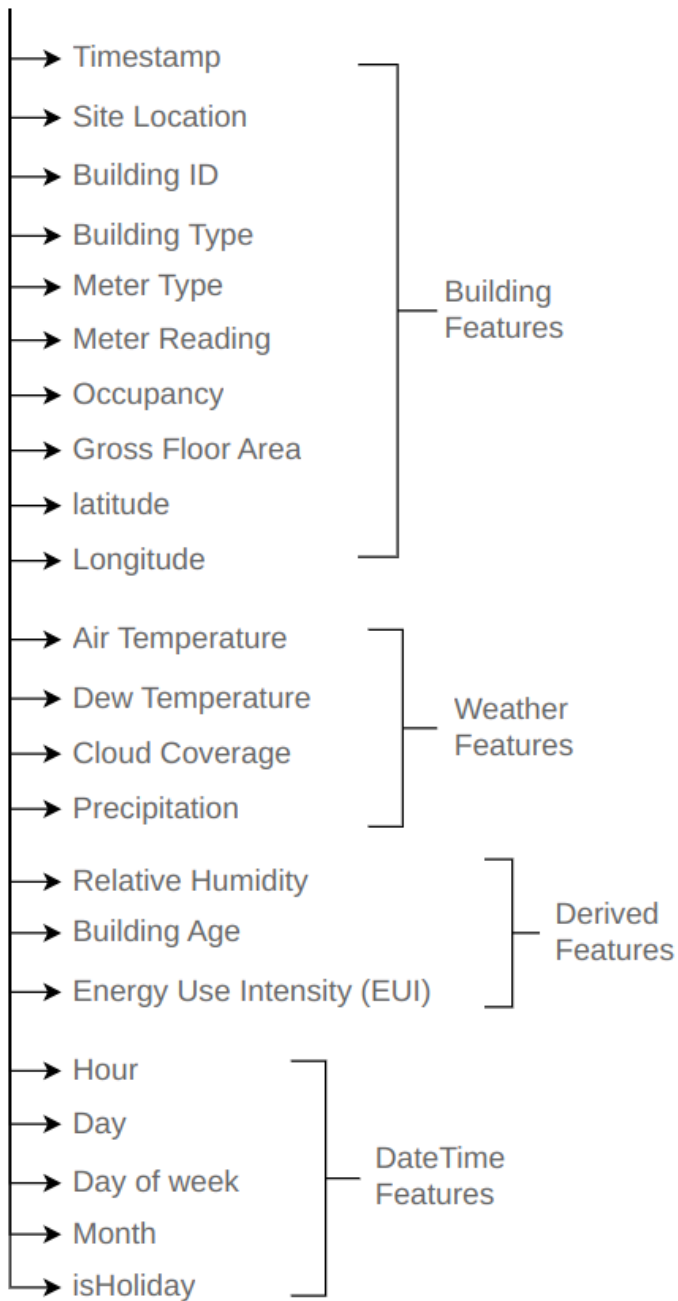
1. Identifying the essential features for building energy prediction involves distinguishing between hard-coded and optional features. Hard-coded features are those crucial for accurate predictions, while optional features may provide additional insights but are not strictly necessary
2. Collecting and preparing data for prediction entails gathering relevant information such as building characteristics, weather data, and historical energy usage.
3. Applying various machine learning models for building energy prediction involves testing different ml algorithms
4. Finally, analysing and presenting the predictions or results involves interpreting the model outputs and identifying patterns or trends. These insights help stakeholders to understand the implications of the predictions and make informed decisions about energy efficiency measures.

HARD CORE & OPTIONAL FEATURES

Hard Core Features are essential features that directly impact energy consumption. They are crucial for accurate energy predictions and are typically included in all models.

Optional Features provide additional insights but may not be essential for basic energy predictions. Including these features can enhance the accuracy and granularity of predictions but may require more data and computational resources

Hard Core Feature



Optional Feature

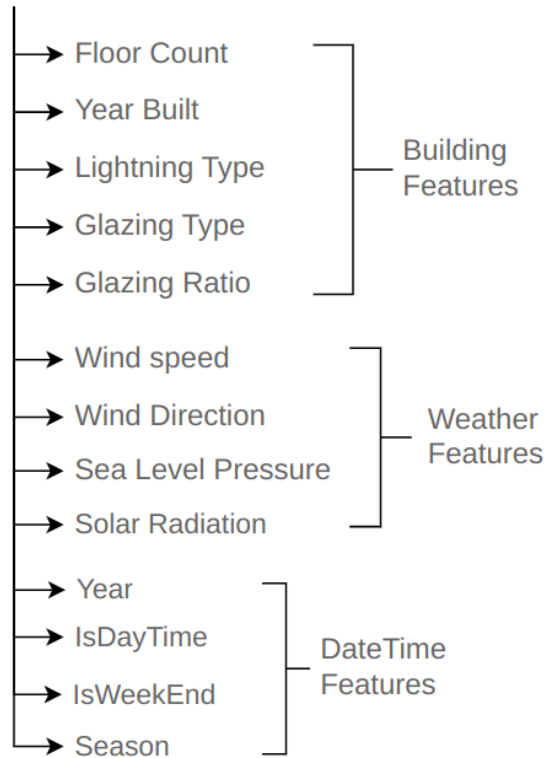


Fig 1: Feature Summary

DETAILS OF FEATURES

1. **Timestamp:** Date and time when the data was recorded.
2. **Site Location:** Physical location or site of the building.
3. **Building ID:** Unique identifier for each building in the dataset.
4. **Building Type:** Classification or type of the building (e.g., residential, office).
5. **Meter Type:** Type of meter used for measuring (e.g., electricity, water).
6. **Meter Reading:** Recorded measurement from the meter at a specific time.
7. **Occupancy:** Number of individuals occupying the building.
8. **Gross Floor Area:** Total floor area of building in square feet.
9. **Latitude:** Latitude of building location to city level.
10. **Longitude:** Longitude of building location to city level.
11. **Floor count:** Number of floors corresponding to building.
12. **Building Built:** Year corresponding to when building was constructed, in the format YYYY.
13. **Building Age:** Length of time since the building was constructed.
14. **EUI:** Energy Use Intensity of a building. (kWh/year/m²)
15. **Lighting Type:** Type of lighting system used in the building.
16. **Glazing Type:** Type of glazing material or glass used in windows.
17. **Glazing Ratio:** Ratio of window area to wall area.
18. **Air Temperature:** The temperature of the air in degrees Celsius (°C)
19. **Dew Temperature:** The dew point (the temperature to which a given parcel of air must be cooled at constant pressure and water vapor content in order for saturation to occur) in degrees Celsius (°C).
20. **Cloud Coverage:** Portion of the sky covered in clouds, in oktas.
21. **Precipitation:** The depth of rain or snow precipitation that is measured over a one-hour accumulation period (mm).
22. **Relative Humidity:** The amount of moisture presents in the air relative to the maximum amount it can hold at a given temperature

23. Wind speed: The rate of horizontal travel of air past a fixed point (m/s).

24. Wind Direction: The angle, measured in a clockwise direction, between true north and the direction from which the wind is blowing (degrees).

25. Sea Level Pressure: The air pressure relative to Mean Sea Level (MSL) (hPa).

26. Solar Radiation: The intensity of radiant energy emitted by the sun.

	Site_id	Building Type	Meter Type	Time Zone	Lighting Type	Glazing Type
1	Orlando FL	residential	chilledwater	Europe/Dublin	CFL	Single-pane
2	Cardiff UK	education	electricity	Europe/London	LED	Double-pane
3	Salt Lake	office	gas	US/Central	Fluorescent	Triple-pane
4	Dublin Ireland	entertainment	hotwater	US/Eastern	Incandescent	Low-emissivity coating
5	London Site1	retail	irrigation	US/Mountain	Halogen	Insulated
6	London Site2	parking	solar	US/Pacific	Solar-powered LED	Laminated
7	London Site3	public services	steam		other	Tempered
8	Tempe AZ	storage	water			Reflective
9	Washington DC	food				Vacuum insulated
10	Berkeley CA	religion				
11	Princeton NJ	healthcare				
12	Austin TX	utility				
13	Ottawa Site1	science				
14	Ottawa Site2	other				
15	Saint Paul					
16	Cornell NY					

Table 1: Categorical Features Description

CHAPTER 6: DATASET DETAIL

The entire dataset contains 3,053 energy meters from 1,636 buildings with a range of two full years (2016 and 2017) at an hourly frequency. These meters were collected from 19 sites across North America and Europe, with one or more meters per building measuring whole building electrical, heating and cooling water, steam, and solar energy as well as water and irrigation meters. The entire dataset after preprocessing has 4,34,90,402 rows and 21 columns.

A sample of a single site (Salt Lake) is taken from the entire dataset for prediction. This site consists of 10 buildings belonging to 5 different types: Residential, Office, Educational, Entertainment, and Other. Two-meter types, electricity, and water, are used for prediction. The sample dataset contains 3,31,834 records and 21 features.

S. No.	Site	Kaggle	Actual Site Name	Location	Buildings	Meters
1	Panther	0	Univ. of Central Florida (UCF)	Orlando, FL	136	299
2	Robin	1	Univ. College London (UCL)	London, UK	52	67
3	Fox	2	Arizona State Univ. (ASU)	Tempe, AZ	137	306
4	Rat	3	Washington DC - City Buildings	Washington DC	305	305
5	Bear	4	Univ. of California - Berkeley	Berkeley, CA	92	92
6	Lamb	5	Cardiff - City Buildings	Cardiff, UK	147	265
7	Eagle	6	Anonymous	N/A	47	106
8	Moose	7	Ottawa - City Buildings	Ottawa, Ontario	15	43
9	Gator	8	Anonymous	N/A	74	74
10	Bull	9	Univ. of Texas - Austin	Austin, TX	124	308
11	Bobcat	10	Anonymous	Salt Lake	36	116
12	Crow	11	Carleton Univ.	Ottawa, Ontario	5	15
13	Wolf	12	Univ. College Dublin (UCD)	Dublin, Ireland	36	66
14	Hog	13	Anonymous	Anonymous	163	336
15	Peacock	14	Princeton University	Princeton, NJ	106	298
16	Cockatoo	15	Cornell University	Cornell, NY	124	282
17	Shrew	—	UK Parliment	London, UK	9	13
18	Swan	—	Anonymous	N/A	21	55
19	Mouse	—	Ormand Street Hospital	London, UK	7	7

Table 2: Dataset Summary

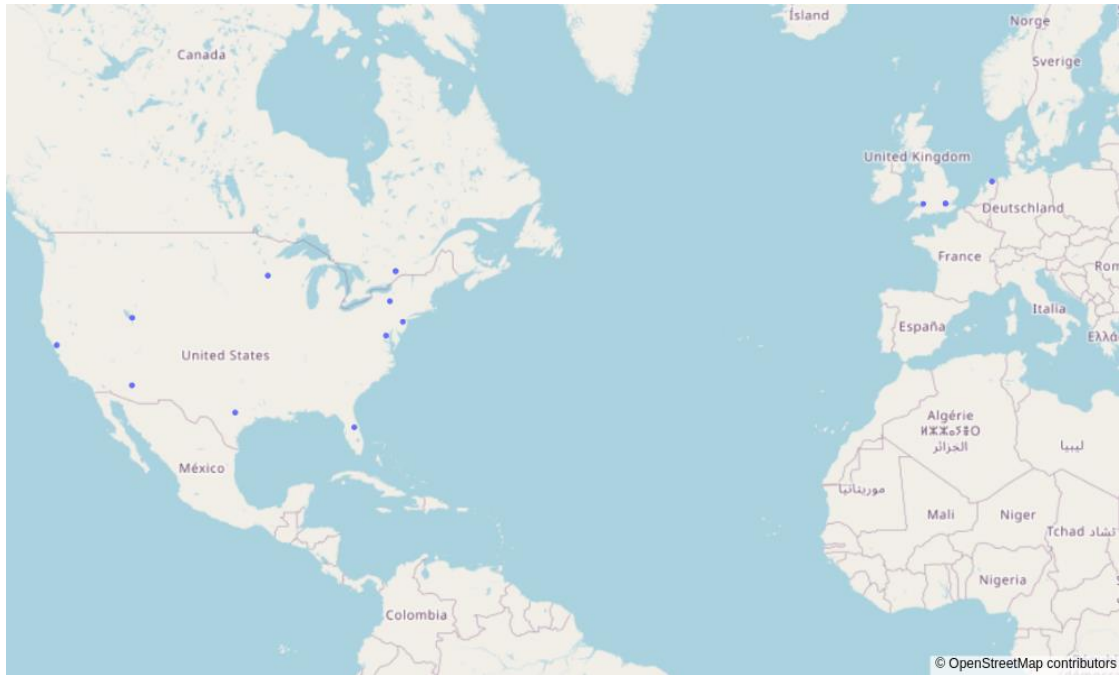


Fig 2: Site Geographical Location

Correlation with Target Variable

The correlation heatmap reveals the relationships between various features and the target variable, `meter_reading`. Features like `meter` (0.18061), `occupants` (0.17648), and `sqft` (0.13337) have the highest positive correlations with meter readings, indicating these factors significantly impact energy consumption. `Primary_use` (0.08414) and `airTemperature` (0.07485) also show positive correlations but are less influential.

Conversely, `building_id` (-0.14771) and `relative_humidity` (-0.06815) have negative correlations, meaning certain buildings and higher humidity levels tend to lower meter readings. Features such as `timestamp`, `hour`, `day`, `month`, `precipDepth1HR`, `windDirection`, and `windSpeed` show near-zero correlations, indicating they have minimal impact on meter readings. This analysis highlights the most relevant factors for predicting energy usage, aiding in the development of effective predictive models.

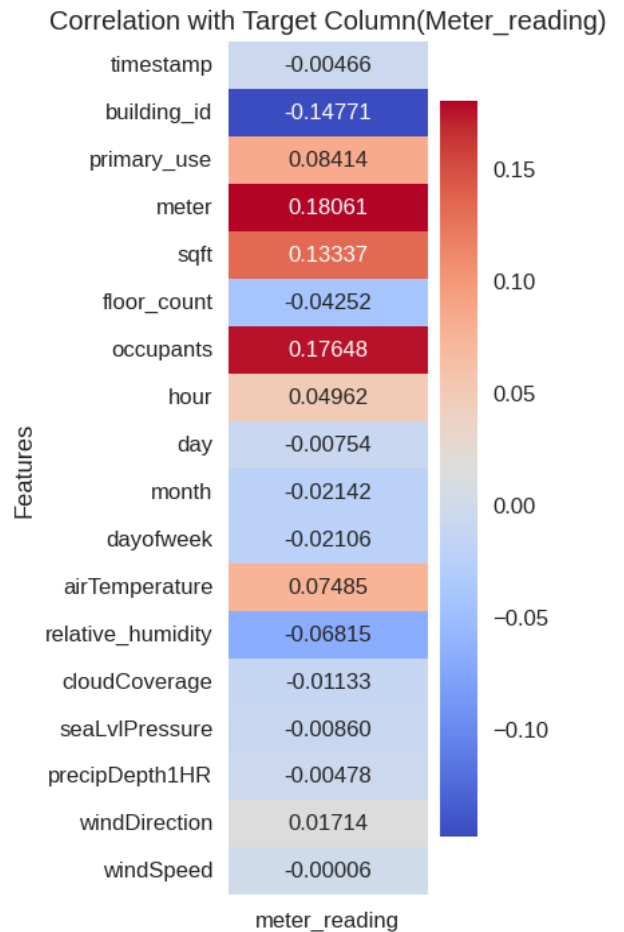


Fig 3: Correlation Heatmap

CHAPTER 7: EXPLORATORY DATA ANALYSIS (EDA)

In the exploratory data analysis (EDA) for our building energy prediction project, we examined the correlations between various features and the target variable, `meter_reading`. The analysis revealed that `meter` type, number of `occupants`, and building `sqft` are the most influential factors, showing significant positive correlations with energy consumption. As we see in the Fig 6 the meter reading is highly skewed so we that log meter reading for our model training. These insights will guide the development of more accurate predictive models for building energy usage.

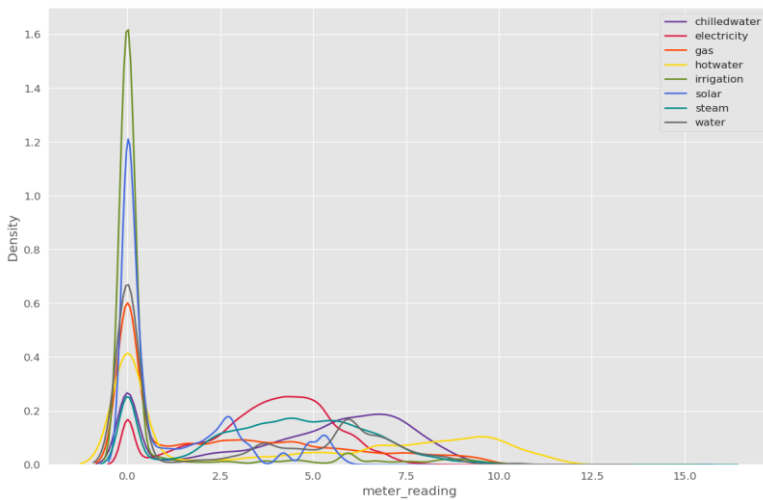


Fig 4: Meter Reading Distribution

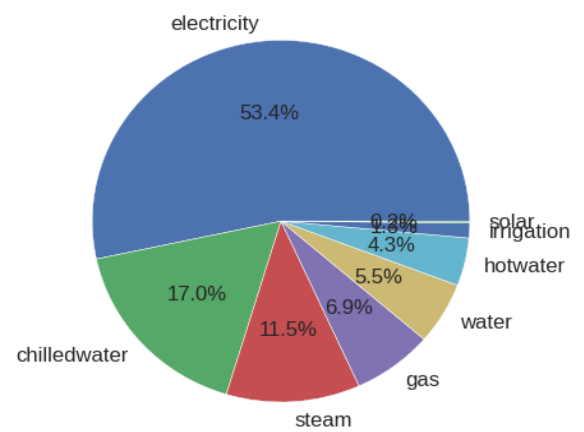


Fig 5: Meter Reading Pie chart

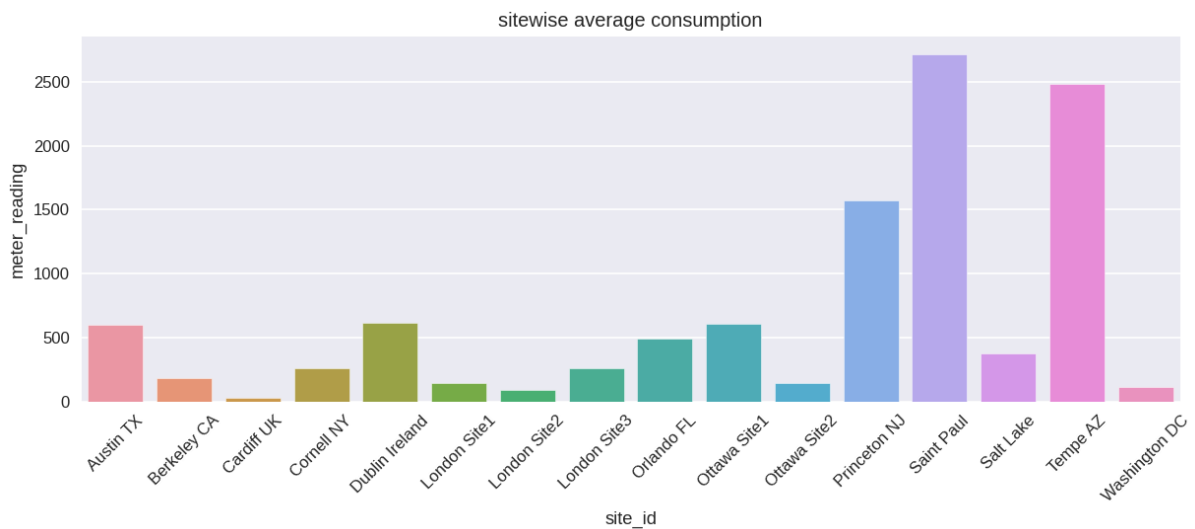


Fig 6: Site Wise Energy Consumption

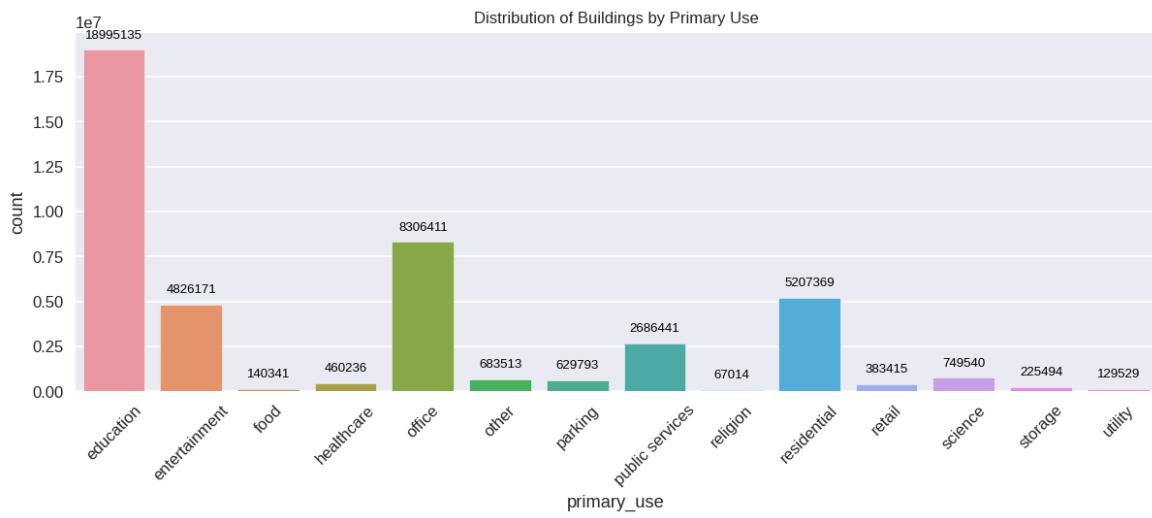


Fig 7: Distribution Of Building by Primary Use

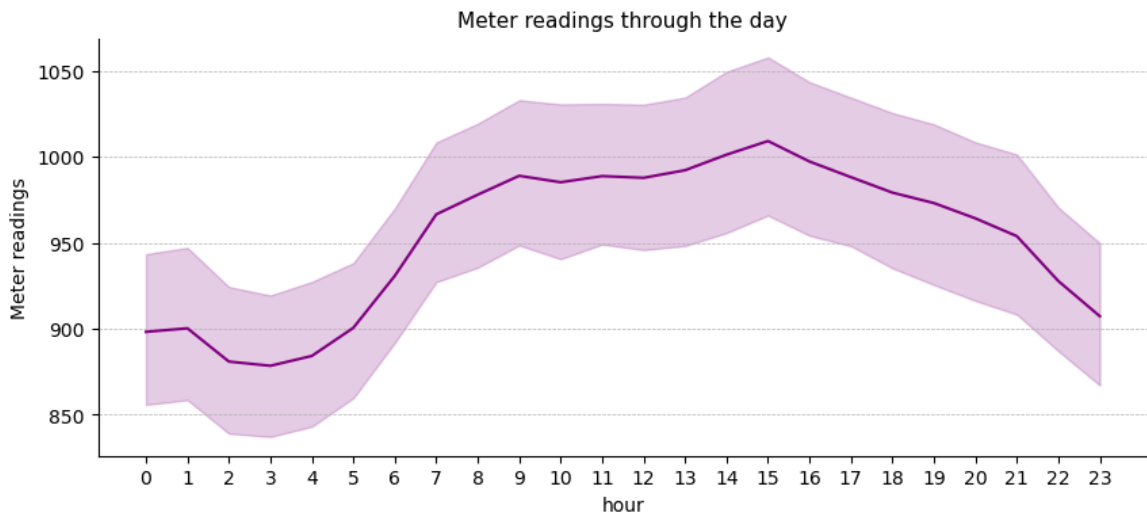


Fig 8: Energy Trend Through the Day

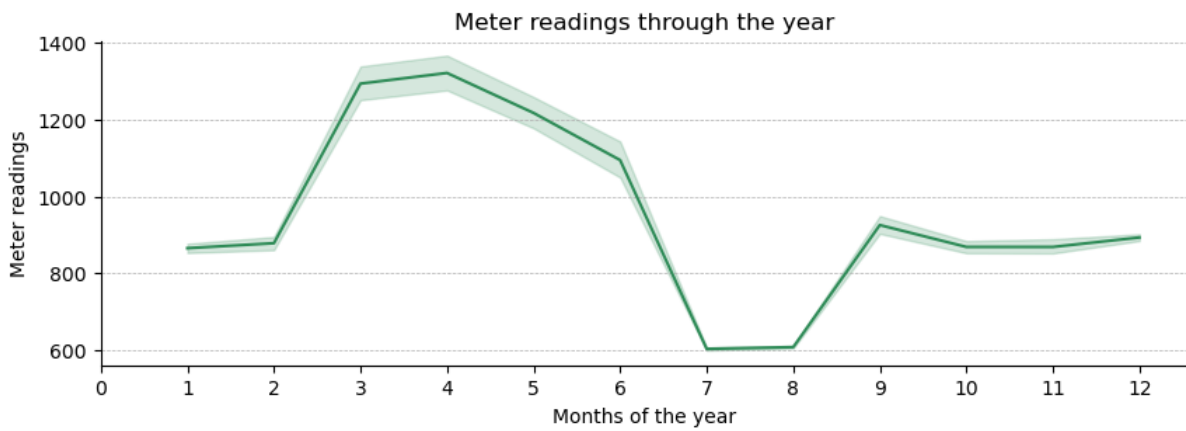


Fig 9: Energy Trend Through the year

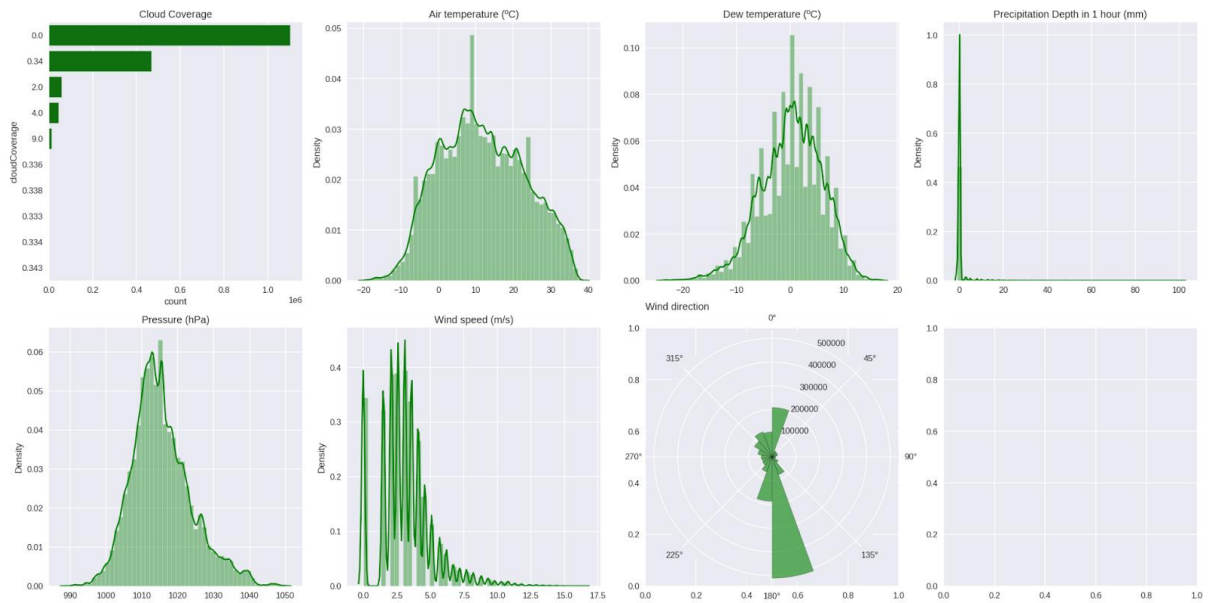


Fig 10: Distribution Of Weather Features

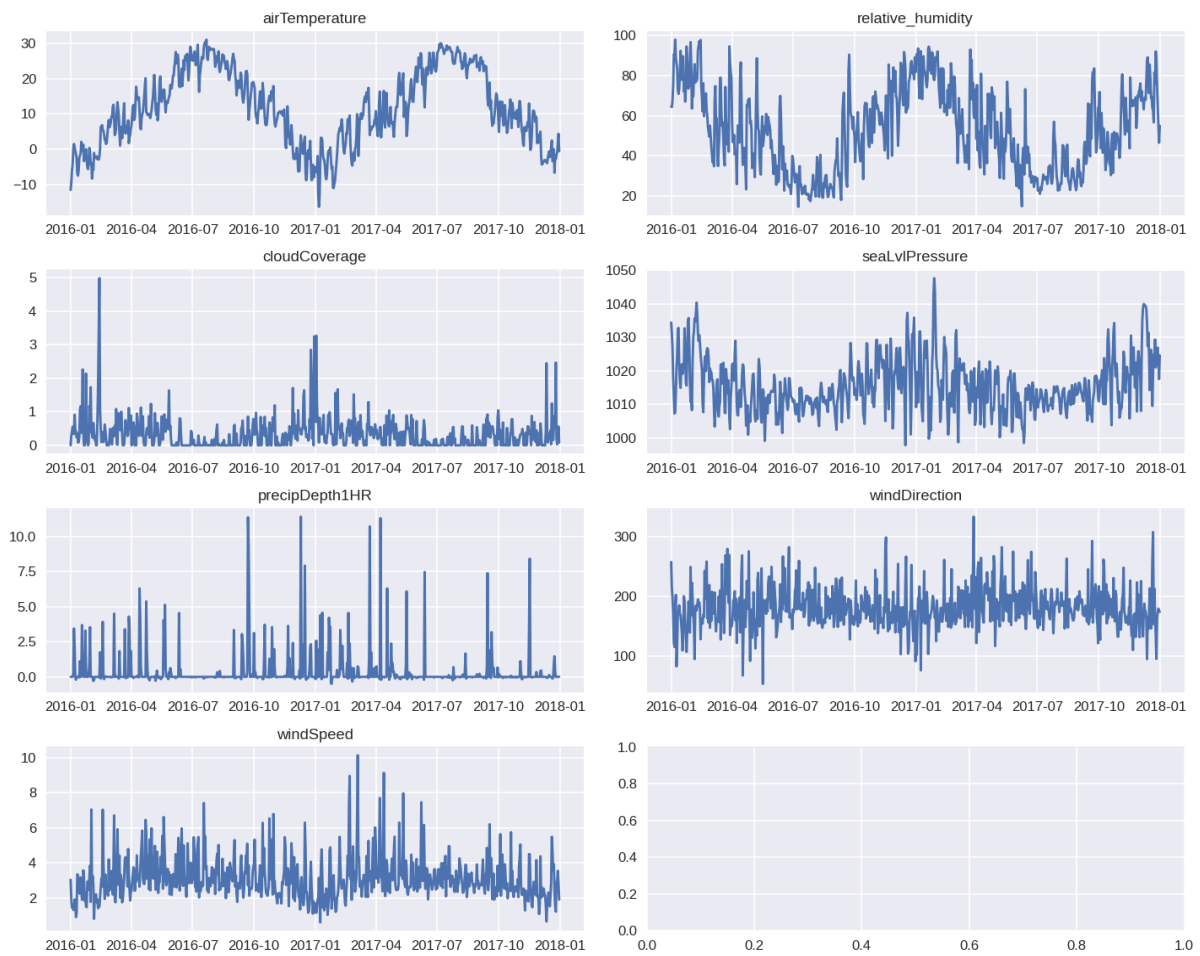


Fig 11: Trends in Weather Features

CHAPTER 8: PYCARET RESULT ANALYSIS

	Description	Value
0	Session id	5649
1	Target	meter_reading
2	Target type	Regression
3	Original data shape	(331834, 11)
4	Transformed data shape	(331834, 13)
5	Transformed train set shape	(232283, 13)
6	Transformed test set shape	(99551, 13)
7	Numeric features	9
8	Date features	1
9	Preprocess	True
10	Imputation type	simple
11	Numeric imputation	mean
12	Categorical imputation	mode
13	Fold Generator	KFold
14	Fold Number	2
15	CPU Jobs	-1
16	Use GPU	False
17	Log Experiment	False
18	Experiment Name	reg-default-name
19	USI	740c

Table 3: Pycaret Parameters

The Pycaret analysis reveals valuable insights into the predictive models for our project. We conducted regression analysis on the dataset with a target variable of "meter reading". After training the data, we will identify the best-performing model that demonstrates high performance.

The Random Forest exhibits the lowest Mean Absolute Error (MAE) and highest R-squared (R2) values, indicating better performance compared to other models.

Ensemble methods such as Gradient Boosting and AdaBoost Regressors show moderate performance but may benefit from further optimization or feature engineering.

Based on these results, we recommend considering Random Forest for our predictive modelling tasks, owing to their superior performance metrics.

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
rf	Random Forest Regressor	108.4571	147616.2278	384.2086	0.6295	2.0852	0.5063	35.0700
et	Extra Trees Regressor	106.7494	152649.4887	390.7038	0.6169	2.0064	0.4893	22.4900
lightgbm	Light Gradient Boosting Machine	128.2970	159327.1853	399.1581	0.6001	2.4501	9.0738	2.2800
xgboost	Extreme Gradient Boosting	137.7266	185653.6719	430.8754	0.5341	2.4756	12.2764	2.0600
knn	K Neighbors Regressor	128.8314	202746.4219	450.2737	0.4912	2.1463	0.7926	1.7700
dt	Decision Tree Regressor	122.9288	242998.9021	492.9492	0.3902	2.0209	0.5200	3.2300
gbr	Gradient Boosting Regressor	177.5836	270163.9248	519.7730	0.3220	2.7358	21.5874	52.4600
lr	Linear Regression	256.5672	364914.3593	604.0814	0.0842	3.3017	105.5956	0.4400
lasso	Lasso Regression	255.6671	364931.2534	604.0954	0.0842	3.3002	105.5690	0.5200
ridge	Ridge Regression	256.5667	364914.3632	604.0814	0.0842	3.3017	105.5953	0.4000
lar	Least Angle Regression	256.5672	364914.3593	604.0814	0.0842	3.3017	105.5956	0.4000

Table 4: Pycaret Result

CHAPTER 9: ML TRAINING RESULT

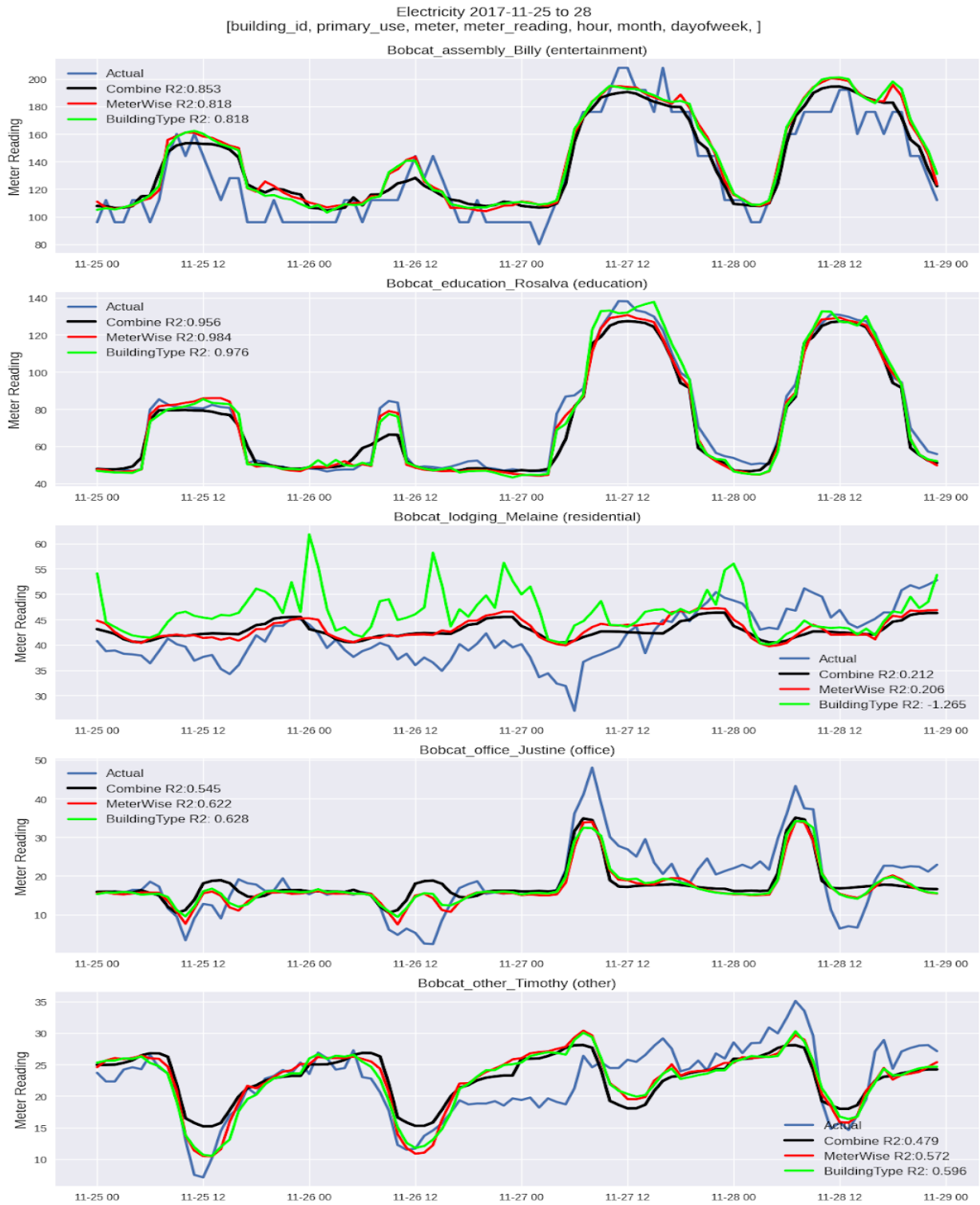


Fig 12: Model Training Result without Weather Features for Electricity

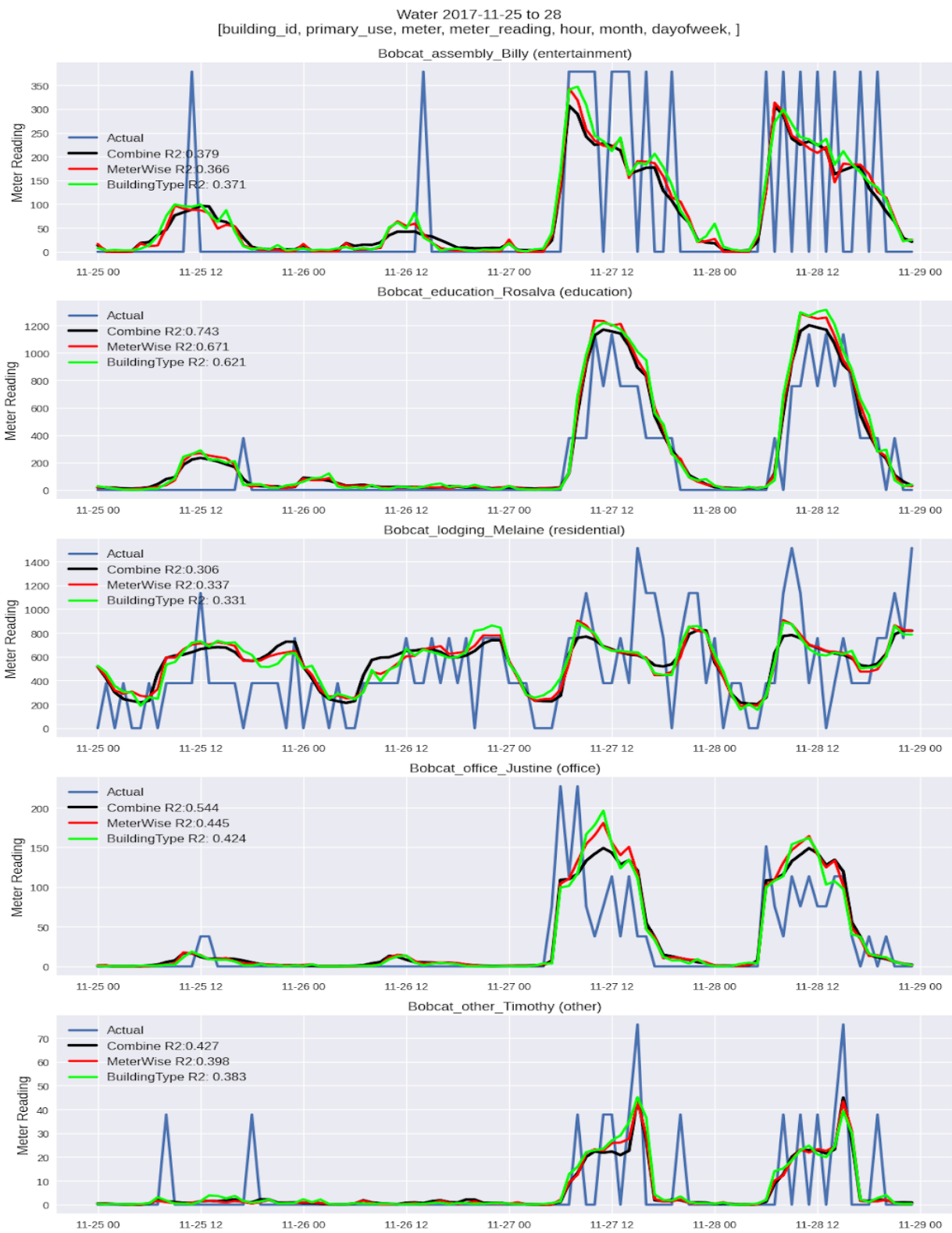


Fig 13: Model Training Result without Weather Features for Water

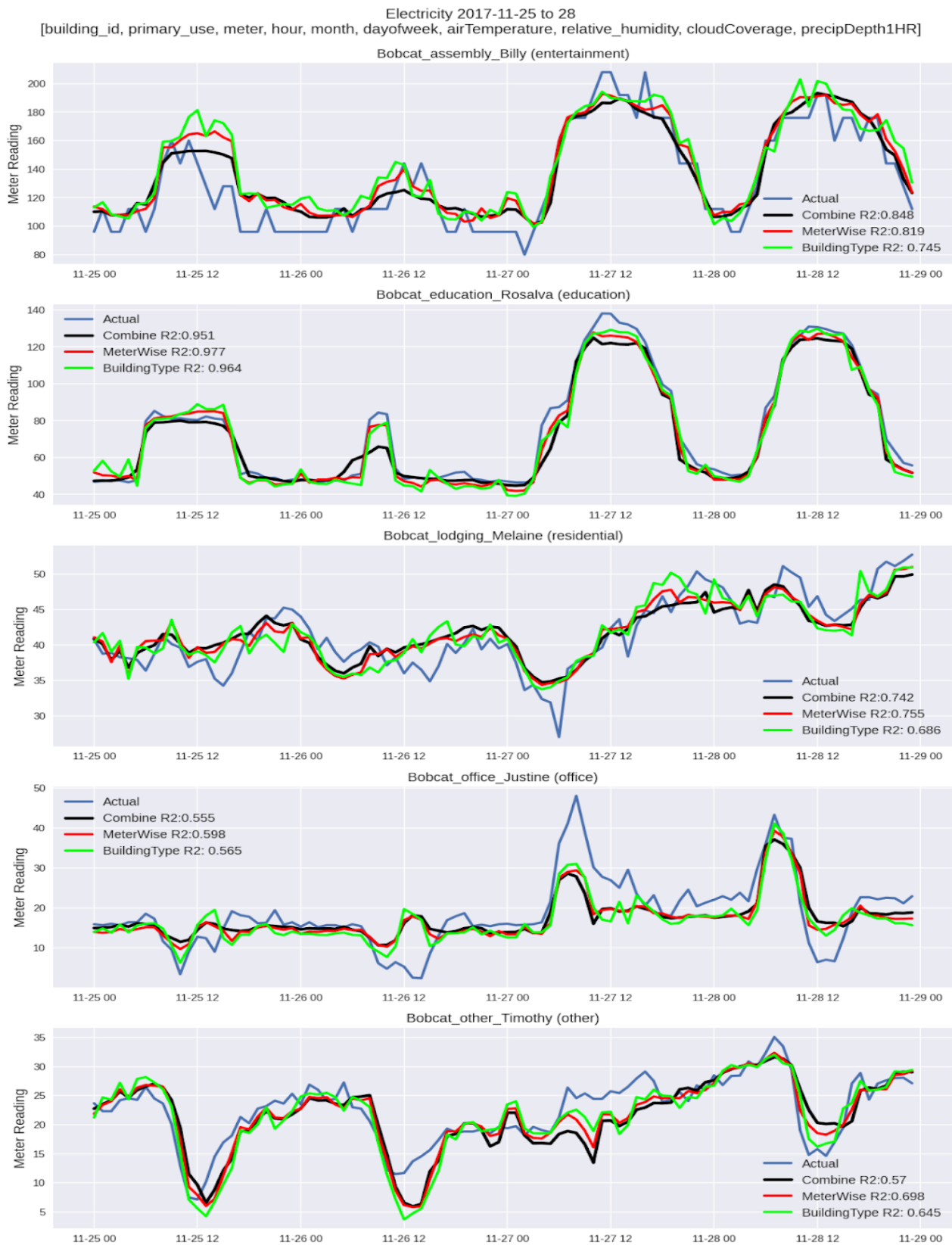


Fig 14: Model Training Result with Weather Features for Electricity

Water 2017-11-25 to 28
 [building_id, primary_use, meter, hour, month, dayofweek, airTemperature, relative_humidity, cloudCoverage, precipDepth1HR]



Fig 15: Model Training Result with Weather Features for Water

CHAPTER 10: ANN MODEL RESULT

Model: "sequential"

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 64)	704
dense_1 (Dense)	(None, 128)	8,320
dense_2 (Dense)	(None, 256)	33,024
dense_3 (Dense)	(None, 128)	32,896
dense_4 (Dense)	(None, 64)	8,256
dense_5 (Dense)	(None, 1)	65

Total params: 83,265 (325.25 KB)

Trainable params: 83,265 (325.25 KB)

Non-trainable params: 0 (0.00 B)

Table 5: ANN Model Summary



Fig 16: ANN Training Summary

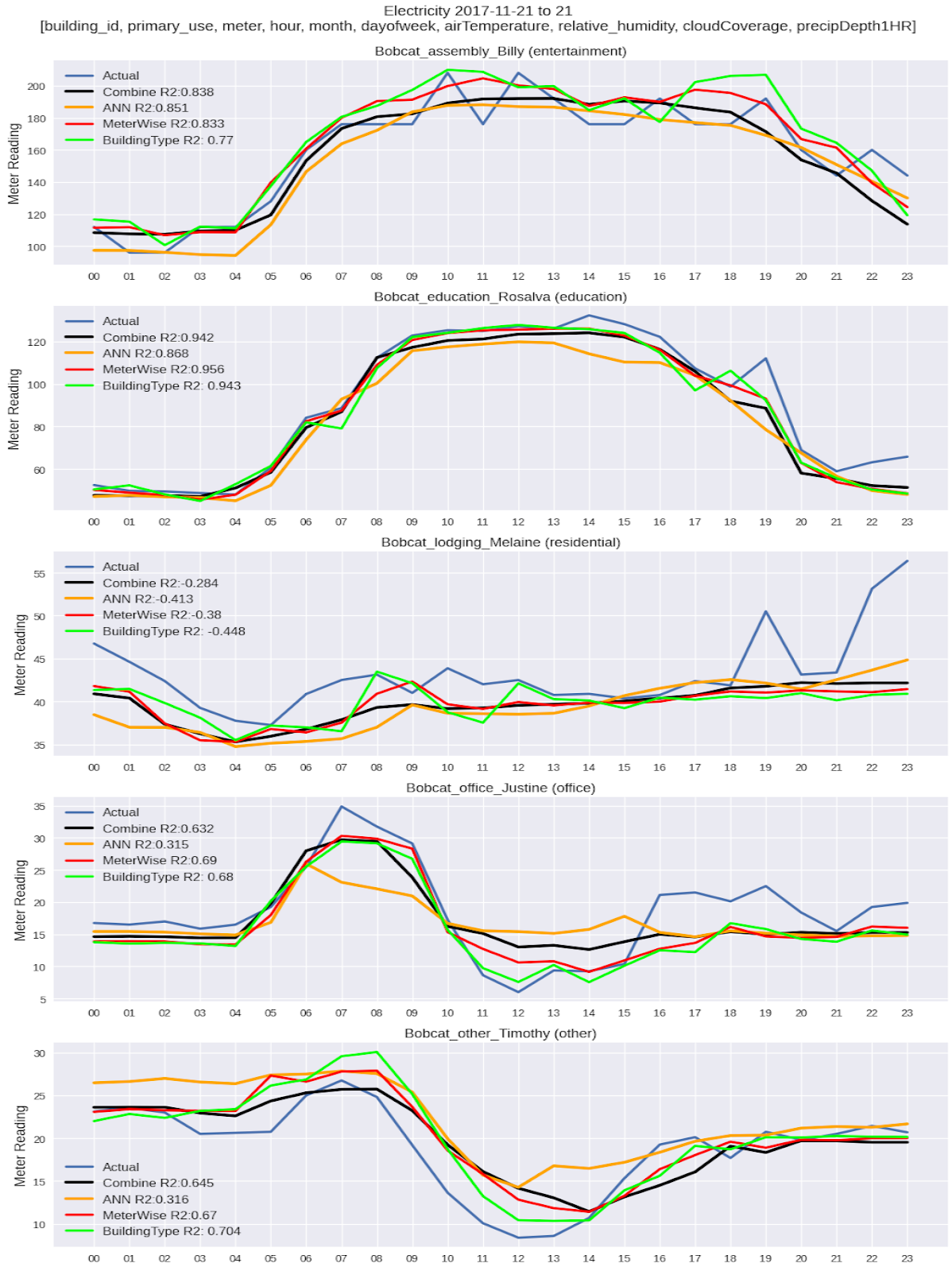


Fig 17: ANN Model Training Result for Electricity

Water 2017-11-21



Fig 18: ANN Model Training Result for Water

CHAPTER 11: DAILY MODEL TRAINING RESULT

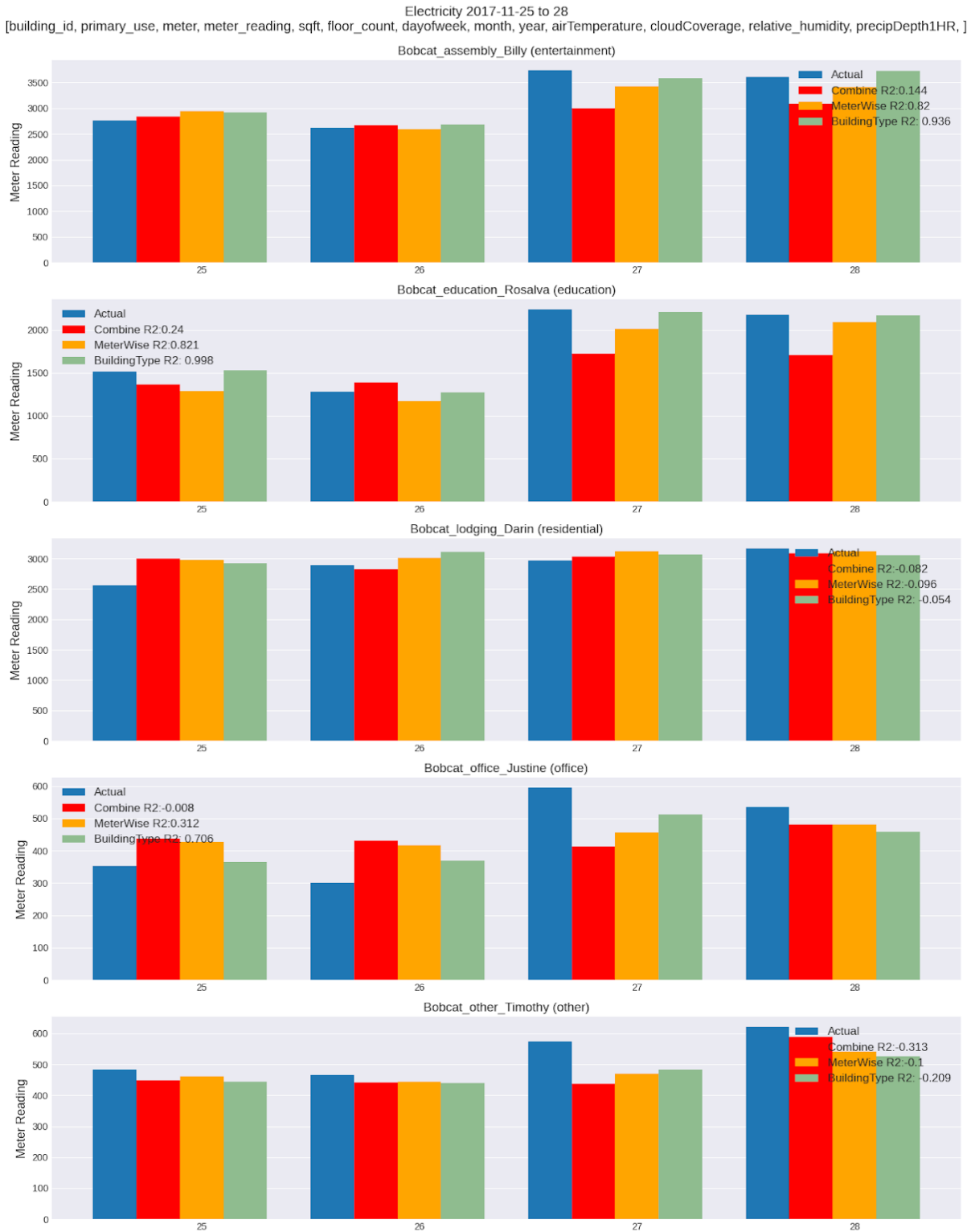


Fig 19: Daily Model Training Result for Electricity

Water 2017-11-25 to 28
 [building_id, primary_use, meter, meter_reading, sqft, floor_count, dayofweek, month, year, airTemperature, cloudCoverage, relative_humidity, precipDepth1HR,]

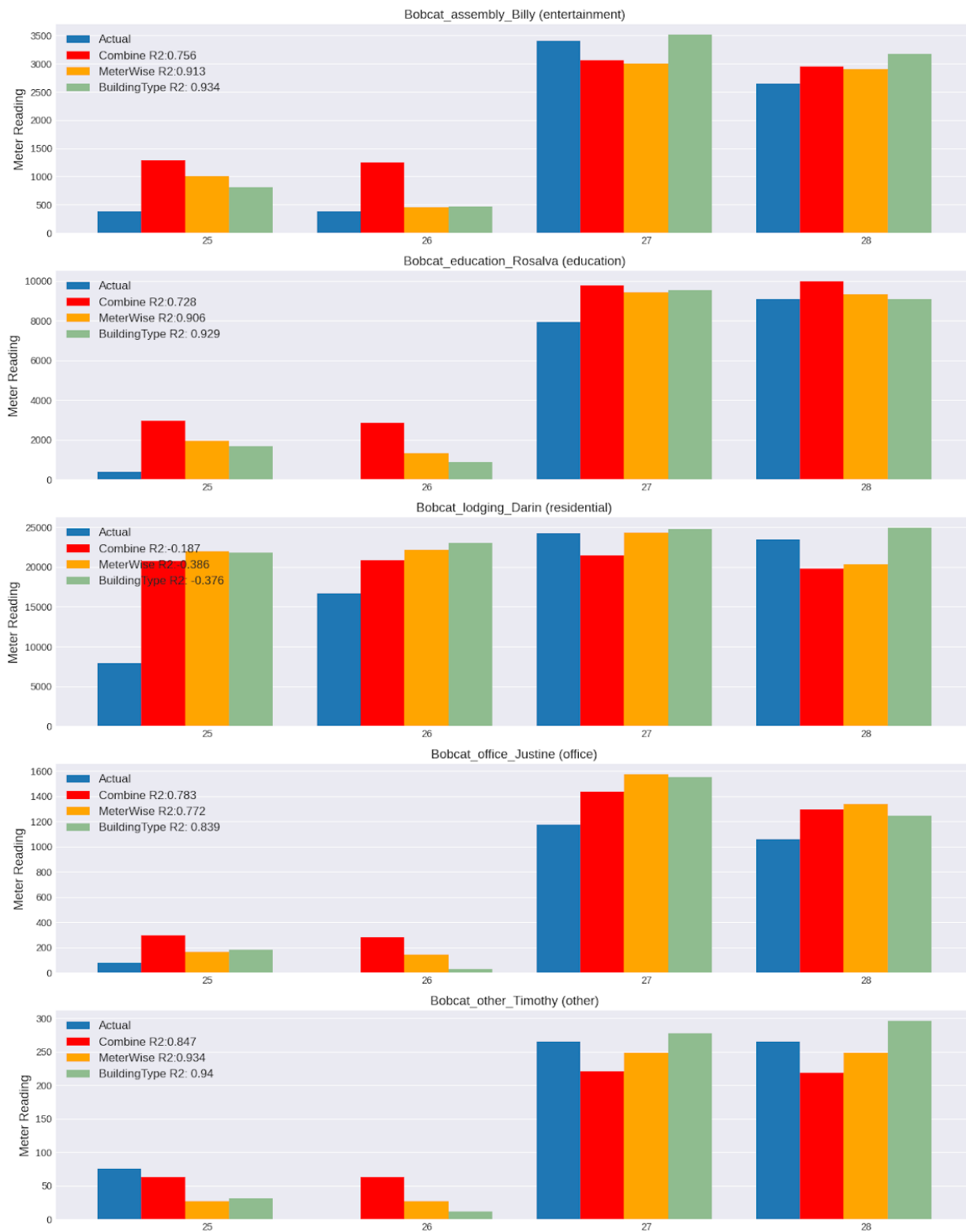


Fig 20: Daily Model Training Result for Water

REFERENCES

Machine learning for estimation of building energy paper

https://www.researchgate.net/publication/328032571_Machine_learning_for_estimation_of_building_energy_consumption_and_performance_a_review

ASHRAE: Great Energy Predictor

<https://www.kaggle.com/code/caesarlupum/ashrae-start-here-a-gentle-introduction/notebook#Model-in-%E2%9A%A1%F0%9F%94%8CASHRAE:-Lgbm-Simple-FE>

Building Energy

<https://www.kaggle.com/code/sayansh001/energy-efficient-buildings-detailed-eda>

ML Model

<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/>

Zero Energy Buildings

https://www.researchgate.net/publication/237429041_Zero_Energy_Buildings_A_Critical_Look_at_the_Definition

TURNITIN PLAGIARISM REPORT

Similarity Report

PAPER NAME

Akash Gaur 0901MC201 (1).pdf

WORD COUNT

3646 Words

CHARACTER COUNT

27016 Characters

PAGE COUNT

41 Pages

FILE SIZE

5.3MB

SUBMISSION DATE

May 17, 2024 1:56 PM GMT+5:30

REPORT DATE

May 17, 2024 1:56 PM GMT+5:30

● 13% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 3% Internet database
- 0% Publications database
- Crossref database
- Crossref Posted Content database
- 13% Submitted Works database

● Excluded from Similarity Report

- Bibliographic material

Summary

LEARNING OUTCOMES

During my internship focused on building energy prediction using machine learning, I achieved several key learning outcomes:

1. **Data Preprocessing and Cleaning:** I gained hands-on experience in preparing large datasets for analysis, including handling missing values, normalizing data, and encoding categorical variables.
2. **Exploratory Data Analysis (EDA):** I developed skills in conducting comprehensive EDA to uncover patterns and relationships within the data, such as identifying key features that influence energy consumption.
3. **Feature Engineering:** I learned how to create new features and select the most relevant ones to improve model performance, based on insights from the correlation analysis and domain knowledge.
4. **Model Selection and Evaluation:** I became proficient in selecting appropriate machine learning algorithms for regression tasks, including linear regression, decision trees, and ensemble methods. I also learned how to evaluate model performance using metrics such as RMSE, MAE, and R^2 .
5. **Hyperparameter Tuning:** I practiced optimizing model performance by tuning hyperparameters using techniques like grid search and random search, enhancing the accuracy and efficiency of the predictive models.
6. **Visualization and Reporting:** I improved my ability to visualize data and model results using tools like Matplotlib and Seaborn, and effectively communicated findings and insights through clear, concise reports and presentations.
7. **Collaboration and Communication:** I experienced the collaborative nature of data science projects, working with team members to share insights, troubleshoot issues, and integrate feedback into the modelling process.

These outcomes have equipped me with a solid foundation in machine learning and data analysis, preparing me for future challenges in the field of predictive modelling and beyond.

ANNEXURE-2**DAILY DIARY**

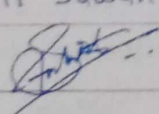

Week	Duration Start date – End date (DD/MM/YY) - (DD/MM/YY)	Progress of Internship/Project
Week 1	01/01/24 - 07/01/24	Introduction and Orientation Introduction to the team and project Overview of project goals and timelines. Set up development environment and tools.
Week 2	08/01/24 - 14/01/24	Data Collection and Initial Analysis Collect initial datasets for analysis Perform data cleaning and preprocessing. Begin exploratory data analysis (EDA).
Week 3	15/01/24 - 21/01/24	Deep Dive into EDA Conduct detailed exploratory data analysis. Identify key features and correlations. Document findings and initial insights.
Week 4	22/01/24 - 28/01/24	Feature Engineering Create new features based on EDA insights. Select relevant features for modelling. Begin preliminary feature selection.
Week 5	29/01/24 - 04/02/24	Model Selection Research and select appropriate machine learning models. Set up baseline models for comparison. Evaluate initial model performance.
Week 6	05/02/24 - 11/02/24	Model Training Train multiple machine learning models. Implement cross-validation techniques. Analyse and compare model performance metrics.

Week 7	12/02/24 - 18/02/24	<p>Hyperparameter Tuning</p> <p>Conduct hyperparameter tuning for selected models.</p> <p>Use grid search and random search techniques.</p> <p>Optimize model parameters.</p>
Week 8	19/02/24 - 25/02/24	<p>Model Evaluation</p> <p>Evaluate tuned models on validation datasets.</p> <p>Compare performance using RMSE, MAE, and R^2 metrics.</p> <p>Select the best performing model.</p>
Week 9	26/02/24 - 03/03/24	<p>Deployment Preparation</p> <p>Prepare the final model for deployment.</p> <p>Document the model pipeline and workflow.</p> <p>Develop scripts for model deployment.</p>
Week 10	04/03/24 - 10/03/24	<p>Model Deployment</p> <p>Deploy the model in a test environment.</p> <p>Monitor model performance and outputs.</p> <p>Debug and refine deployment scripts as needed.</p>
Week 11	11/03/24 - 17/03/24	<p>Performance Monitoring</p> <p>Continuously monitor deployed model.</p> <p>Collect performance data and user feedback.</p> <p>Make necessary adjustments and improvements.</p>
Week 12	18/03/24 - 24/03/24	<p>Final Adjustments and Documentation</p> <p>Implement final adjustments based on feedback.</p> <p>Thoroughly document the project process and outcomes.</p> <p>Prepare final report.</p>
Week 13	25/03/24 - 31/03/24	<p>Presentation and Wrap-Up</p> <p>Prepare presentation of the project.</p> <p>Present findings and outcomes to the team.</p> <p>Reflect on learning outcomes and experiences.</p>

MPR-1

FORMAT

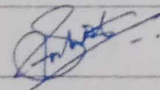
MONTHLY REPORT OF PROGRESS (MRP) FROM INDUSTRY MENTOR

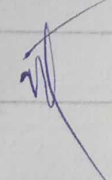
Name of student	AKASH GAUR		Department	MAC	
Industry/Organization	Kloudspot		Date/Duration	08/08/24 09/08/24	
Criterion	Poor	Average	Good	Very Good	Excellent
Punctuality/Timely completion of assigned work					✓
Learning capacity/Knowledge upgradation				✓	
Performance/Quality of work				✓	
Behaviour/Discipline/Team work					✓
Sincerity/Hard work					✓
Comment on nature of work done/Area/Topic	<p>→ Data Collection & Exploration on Time Series Prediction related data.</p> <p>→ Research Analysis on few research papers on Energy Index prediction.</p>				
<u>OVERALL GRADE (Any one)</u>	<u>POOR/AVERAGE/GOOD/VERY GOOD/EXCELLENT</u>				
<u>Name of Industry Mentor</u>	Ankit Sawant				
<u>Signature of Industry Mentor</u>					
Receiving Date		Name of Faculty Mentor	Dr. J K Mutheti	Sign	

MPR-2

FORMAT

MONTHLY REPORT OF PROGRESS (MRP) FROM INDUSTRY MENTOR

Name of student	AKASH GAUR		Department	MAC	
Industry/Organization	kloudspot		Date/Duration	01/02/24 - 29/02/24	
Criterion	Poor	Average	Good	Very Good	Excellent
Punctuality/Timely completion of assigned work					✓
Learning capacity/Knowledge upgradation				✓	
Performance/Quality of work				✓	
Behaviour/Discipline/Team work					✓
Sincerity/Hard work					✓
Comment on nature of work done/Area/Topic	<p>→ Data Collection & Exploration on Time Series Prediction related data.</p> <p>→ Research Analysis on few research papers on Energy Index prediction.</p>				
<u>OVERALL GRADE (Any one)</u>	<u>POOR/AVERAGE/GOOD/VERY GOOD/EXCELLENT</u>				
<u>Name of Industry Mentor</u>	Ankit Sawant				
<u>Signature of Industry Mentor</u>					

Receiving Date	Name of Faculty Mentor	Dr. J K Muthale	Sign	
----------------	------------------------	-----------------	------	---


MPR-3

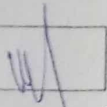
MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR

(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV Bhopal)

NAAC Accredited with A++ Grade

MONTHLY PROGRESS REPORT (MPR) FROM INDUSTRY/COMPANY MENTOR

Name of Student	Akash Gaur		Enrollment No.	0901ML201001	
Department	MAC		Sem.	8 (session Jan-June 24)	
Industry/Organization	Kloudspot		Date/Duration	DD/MM/YY - DD/MM/YY 01/03/24 - 31/03/24	
Criterion	Poor	Average	Good	Very Good	Excellent
Punctuality/Timely completion of assigned work					✓
Learning capacity/ Knowledge upgradation					✓
Performance/ Quality of work					✓
behavior/Discipline/Team work					✓
Sincerity/Hard work					✓
Comment on nature of work done/Area/Topic	Energy Index Prediction using ML models				
<u>OVERALL GRADE</u> (Anyone)	<u>POOR/AVERAGE/GOOD/VERY GOOD/EXCELLENT</u> ✓				
<u>Name of Industry/ Company Mentor</u>	ANKIT SAWANT, KLOUDSPOT, INDIA				
<u>Signature of Industry/ Company Mentor</u>					

Receiving Date		Name of Faculty Mentor	Dr JK Muthale	Sign	
----------------	--	------------------------	---------------	------	---

THE END