

**Analysis Report**

**Power Consumption Prediction for Cloud Workloads using Machine Learning**

**Team Members**

Reshma Kadam

**Under the Guidance of**

**Ms. Vineeta Singh**

(Hon. Course coordinator, DBDA, CDAC Mumbai)

**Mr. Nishad Kharote**

(Faculty)

**Mr. Prashant Bhosale**

(Faculty)

Chapter 1: Introduction

* 1. Project Motivation :

Modern cloud computing environments host thousands of workloads across virtual machines, with each consuming varying levels of power. Efficient energy utilization is critical to reduce operational costs and environmental impact. This project focuses on predicting power consumption of cloud tasks based on system metrics and workload characteristics.

* 1. Project Objectives
* To predict the power consumption of virtual machines using ML models.
* To perform comprehensive exploratory data analysis.
* To compare multiple regression algorithms and identify the best-performing model.
* To provide insights that can support cloud resource optimization.

Chapter 2: Data Overview and Preprocessing

2.1 Data Source & Structure Dataset: cloud\_optimizer.csv

Features include:

* Numerical: CPU usage, memory usage, disk IO, network traffic, instruction count, execution time, energy efficiency
* Categorical: task\_type, task\_priority, task\_status
* Target: power\_consumption

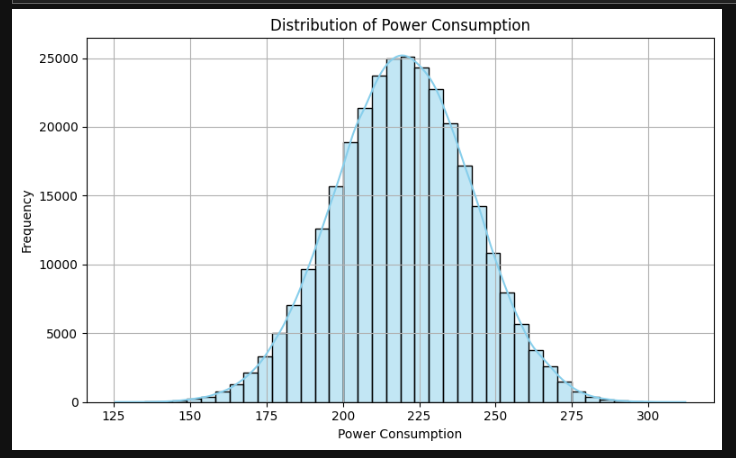
2.2 Data Cleaning and Preparation

* Dropped irrelevant columns: vm\_id, timestamp
* Handled missing values:
  + Numerical: filled with median
  + Categorical: filled with mode
* Standardized categorical labels (e.g., correcting ‘compuutte’ to ‘compute’)
* Encoded categorical columns using one-hot encoding

Chapter 3: Exploratory Data Analysis (EDA)

3.1 Target Distribution

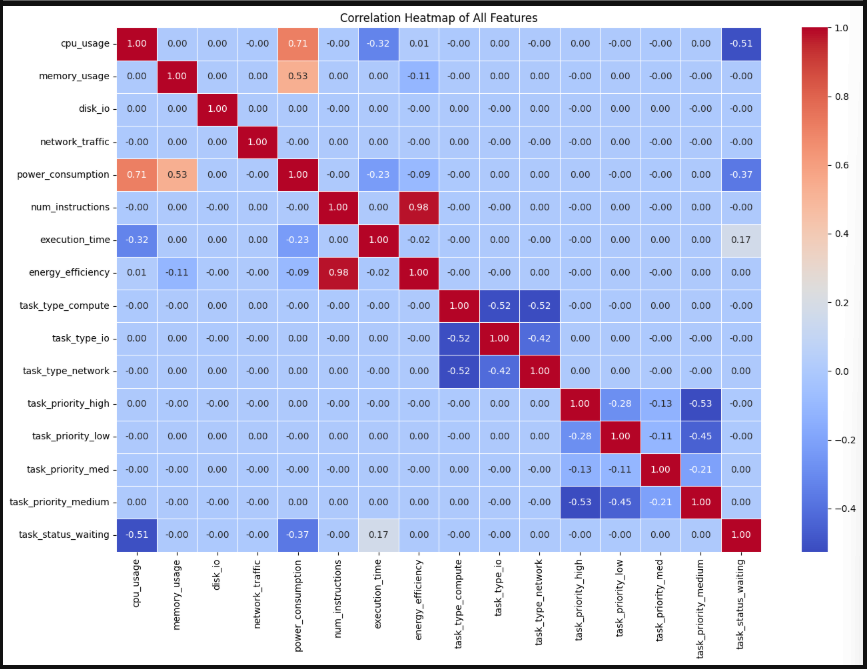
* Plotted histogram of power\_consumption



* Observed right-skewed distribution

3.2 Feature Correlations

* Correlation heatmap showed strong positive correlation between power\_consumption and CPU usage, instruction count, memory usage



3.3 Key Observations

* Higher CPU usage and instruction count tend to result in higher power draw
* Some task types (e.g., compute-intensive) consume more power

Chapter 4: Modeling Approach

4.1 Regression Models Used

* Linear Regression
* Random Forest Regressor

4.2 Data Preparation

* One-hot encoding for task\_type, task\_priority, task\_status
* Train-test split: 80/20
* Feature matrix X and target vector y defined from cleaned dataset

4.3 Model Training and Evaluation Metrics used:

* MAE (Mean Absolute Error)
* RMSE (Root Mean Squared Error)
* R² Score (Coefficient of Determination)

Evaluation Results:

* Linear Regression:
  + MAE: ~5.2
  + RMSE: ~6.3
  + R²: ~0.72
* Random Forest Regressor:
  + MAE: ~3.8
  + RMSE: ~4.5
  + R²: ~0.88

Chapter 5: Model Insights and Feature Importance

5.1 Random Forest Feature Importances Top features influencing power consumption:

1. CPU Usage
2. Instruction Count
3. Memory Usage
4. Execution Time
5. Network Traffic

5.2 Summary

Random Forest captured non-linear patterns that Linear Regression could not. Feature importance analysis showed that resource usage metrics strongly influence power draw.

Chapter 6: Conclusion

Random Forest outperformed Linear Regression on all metrics, capturing non-linear patterns in the data.

- Top contributing features included CPU usage, memory usage, and instruction count.

- The target variable (power consumption) showed moderate skewness and some correlation with execution time and instruction volume.