BATCH NUMBER: 47

You Choose, We Do It

St. JOSEPH'S COLLEGE OF ENGINEERING

(An Autonomous Institution)

St. Joseph's Group of Institutions

OMR, Chennai - 119

COMPARATIVE ANALYSIS OF LOAD FORECASTING ALGORITHMS USING MACHINE LEARNING

KARTHIKEYAN S (312320105061)

PRAVEEN KUMAR J (312320105092)

PROJECT GUIDE: Mr. H.PRASAD M.E., (Ph.D.), Assistant Professor

AGENDA

1. Literature Survey
2. Objective of project
3. Scope and Novelty
4. Block Diagram
5. Algorithms description
6. Output Results
7. Conclusion
8. Future Scope
9. Reference

LITERATURE SURVEY

- Electricity holds immense importance in modern life, serving as the lifeblood of technological, social, and economic progress. From powering our homes and lighting our streets to fueling industries and driving innovations, electricity is integral to nearly every aspect of our daily existence.

- Estimating the load is essential for planning energy production in order to reliably supply enough electricity to meet needs at any moment. Due to the increase in electricity demand, the management of the maintenance schedule, the selection of suitable and affordable generators, and plant planning is a difficult procedure for the electricity supply [1],[3],[8].

LITERATURE SURVEY

- Electrical load forecasting, a critical component of energy management, which can be broadly classified into three distinct categories based on the forecast period: short-term, medium-term, and long-term [2],[3].
- **Short-term** load forecasting, spanning hours to a few days, focuses on predicting immediate electricity demand fluctuations [5],[7].
- **Medium-term** load forecasting extends the horizon to several weeks, months, or even a year, aiding in mid-range planning and resource allocation. It considers factors like seasonal variations and specific events affecting demand.
- **Long-term** load forecasting, covering periods beyond a year, is essential for strategic planning, infrastructure development, and policy formulation.

SCOPE & NOVELTY OF THE PROJECT

SCOPE:

The scope of this project extends to a broad exploration and techniques aimed at enhancing the accuracy and reliability of load predictions using machine learning algorithms in load forecasting, with a particular analysis and practical implementation. Each algorithm will be scrutinized based on its ability to capture and extrapolate complex load patterns, adapt to changing load dynamics, and provide interpretable insights into the underlying factors influencing energy consumption. The study will not only address the technical aspects of algorithmic performance but also deliver into the scalability and computational efficiency of the models, ensuring that the findings are pertinent to real-world energy management systems. Furthermore, the research will consider the implications of model interpretability and transparency, acknowledging the significance of comprehensible insights for decision-making processes in the energy sector.

SCOPE & NOVELTY OF THE PROJECT

NOVELTY:

- ❑ Integration of Exogenous Factors For increase in efficiency of prediction/forecasting
- ❑ Hybrid Models of ARIMA/SARIMAX models
- ❑ Compare with different ML algorithms for best results
- ❑ Real-world Dataset Evaluation
- ❑ Advancement in Load Forecasting Techniques

ALGORITHMS USED

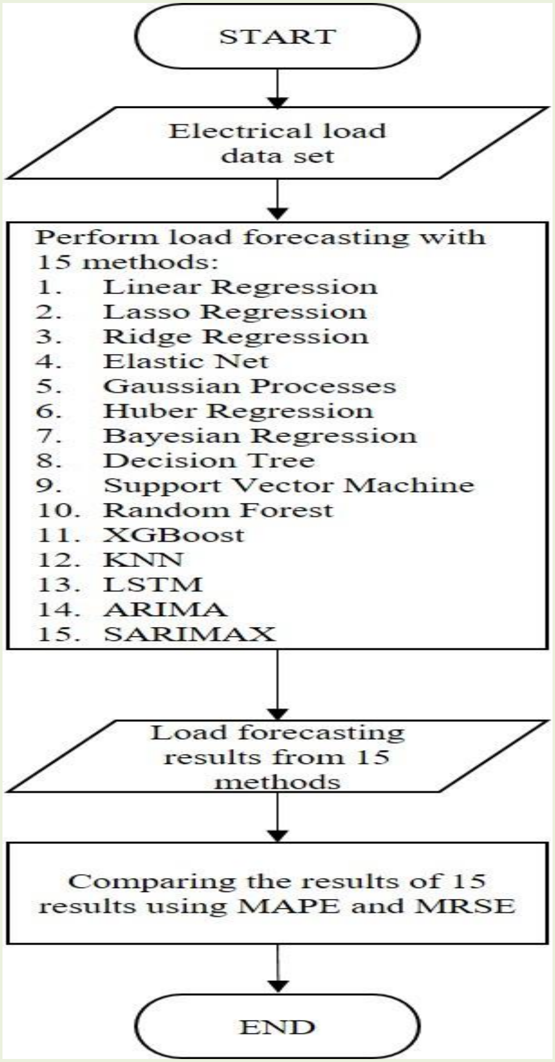
- Linear Regression
- Lasso Regression
- Ridge Regression
- Elastic Net
- Huber Regression
- Bayesian Regression
- Decision Tree
- Support Vector Machine (SVM)
- Random Forest
- Gradient Boosting/ XGBoost
- K-Nearest Neighbor (KNN)
- Long-Short Term Memory (LSTM)
- Autoregressive Integrated Moving Average (ARIMA)
- Seasonal and Exogenous Auto Regressive Integrated Moving Average (SARIMAX)

PROCEDURE

STEP BY STEP APPROACHES:

1. Data Collection and Preprocessing
2. Algorithm Selection and Integration
3. Data Exploration
4. Algorithm Implementation
5. Parameter Tuning
6. Model Training and Evaluation
7. Comparative Analysis
8. Documentation and Reporting

BLOCK DIAGRAM



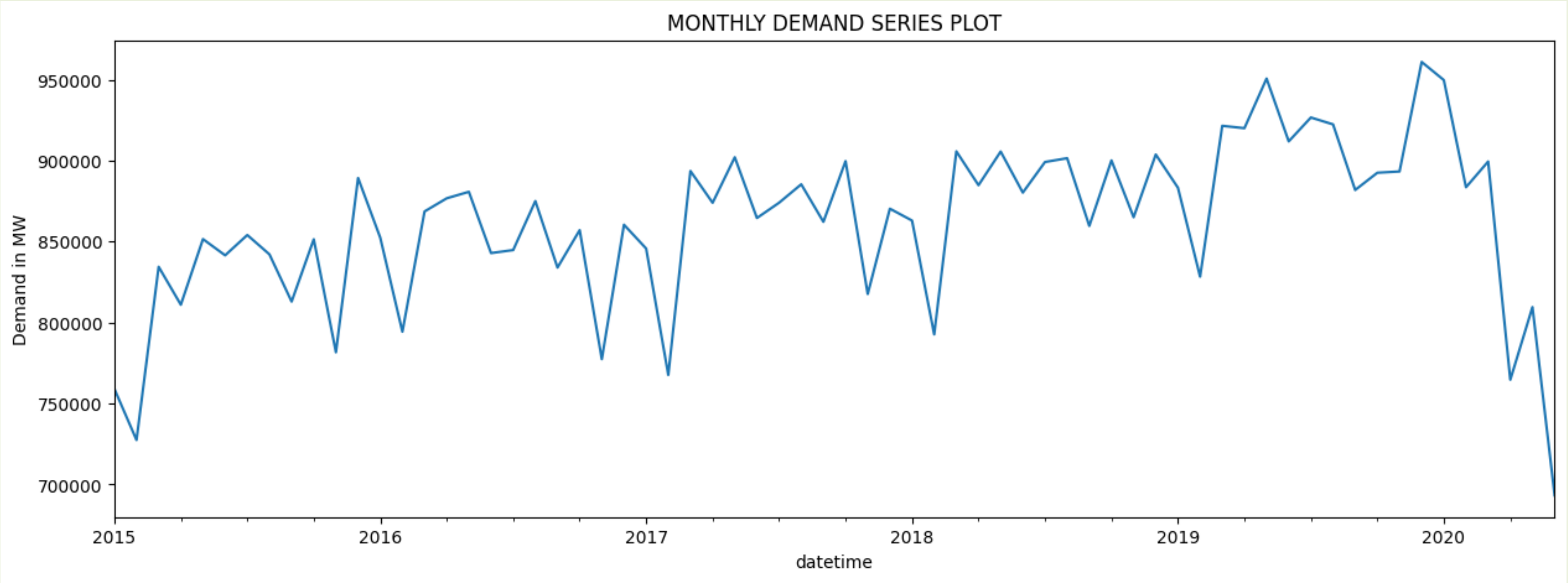
INPUT DATA DESCRIPTION

Parameters	Type	Description
Date ID	Numerical	-
Date	‘yyyy-mm-dd’	2015-01-03 to 2020-06-27
Time	‘hh:mm:ss’	01:00:00 – 23:00:00
Day of Week	Categorical	{0,1,2,3,4,5,6}
Holiday	Categorical	{0, 1}
Load	Numerical	Mega Watts

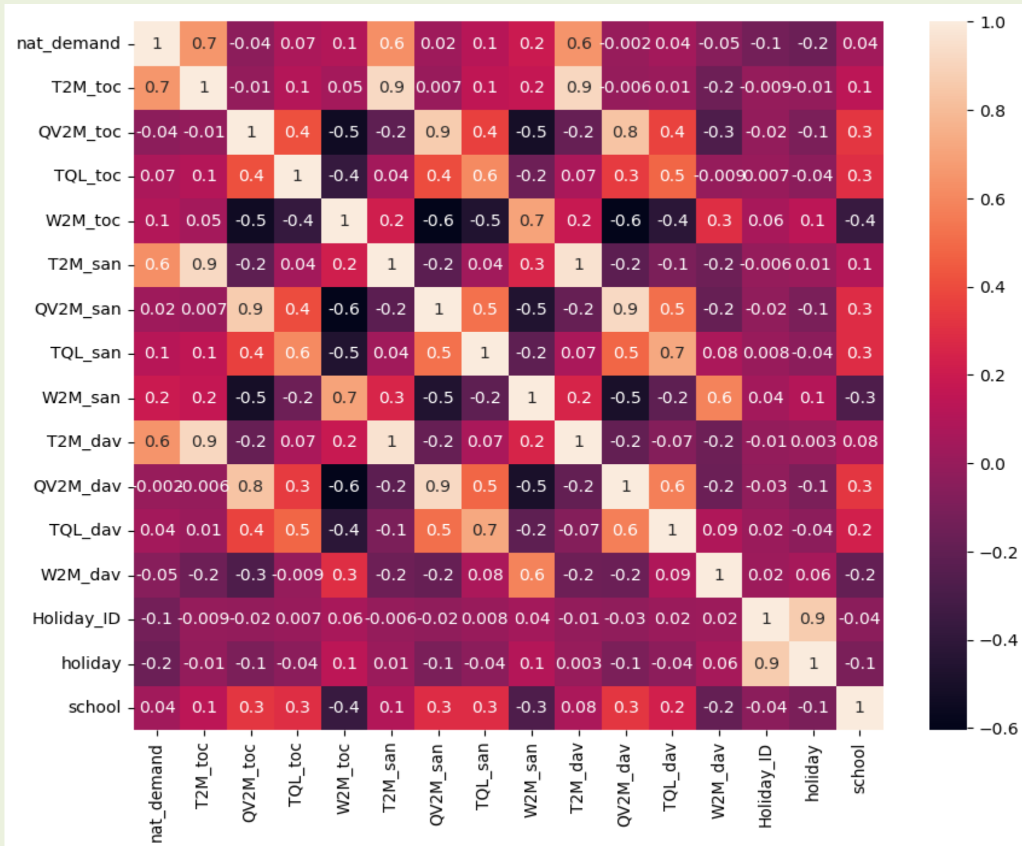
REAL TIME DATA SET

datetime	nat_demand	T2M_toc	QV2M_toc	TQL_toc	W2M_toc	T2M_san	QV2M_san	TQL_san	W2M_san	T2M_dav	QV2M_dav	TQL_dav	W2M_dav	Holiday_ID	holiday	school
2015-01-03 01:00:00	970.3450	25.865259	0.018576	0.016174	21.850546	23.482446	0.017272	0.001855	10.328949	22.662134	0.016562	0.096100	5.364148	0	0	0
2015-01-03 02:00:00	912.1755	25.899255	0.018653	0.016418	22.166944	23.399255	0.017265	0.001327	10.681517	22.578943	0.016509	0.087646	5.572471	0	0	0
2015-01-03 03:00:00	900.2688	25.937280	0.018768	0.015480	22.454911	23.343530	0.017211	0.001428	10.874924	22.531030	0.016479	0.078735	5.871184	0	0	0
2015-01-03 04:00:00	889.9538	25.957544	0.018890	0.016273	22.110481	23.238794	0.017128	0.002599	10.518620	22.512231	0.016487	0.068390	5.883621	0	0	0
2015-01-03 05:00:00	893.6865	25.973840	0.018981	0.017281	21.186089	23.075403	0.017059	0.001729	9.733589	22.481653	0.016456	0.064362	5.611724	0	0	0

INPUT GRAPH


MONTHLY DEMAND SERIES PLOT
Demand in MW
950000
900000
850000
800000
750000
700000
2015
2016
2017
2018
2019
2020
datetime

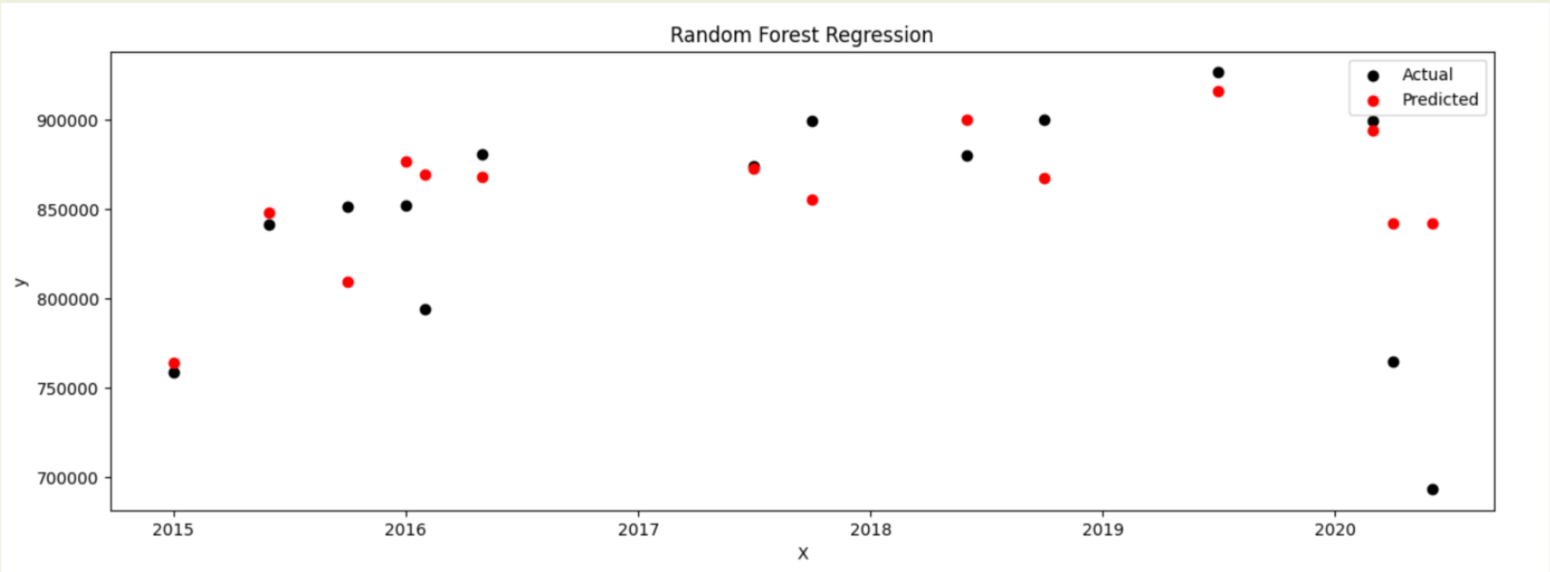
HEATMAP DIAGRAM



Predicted values

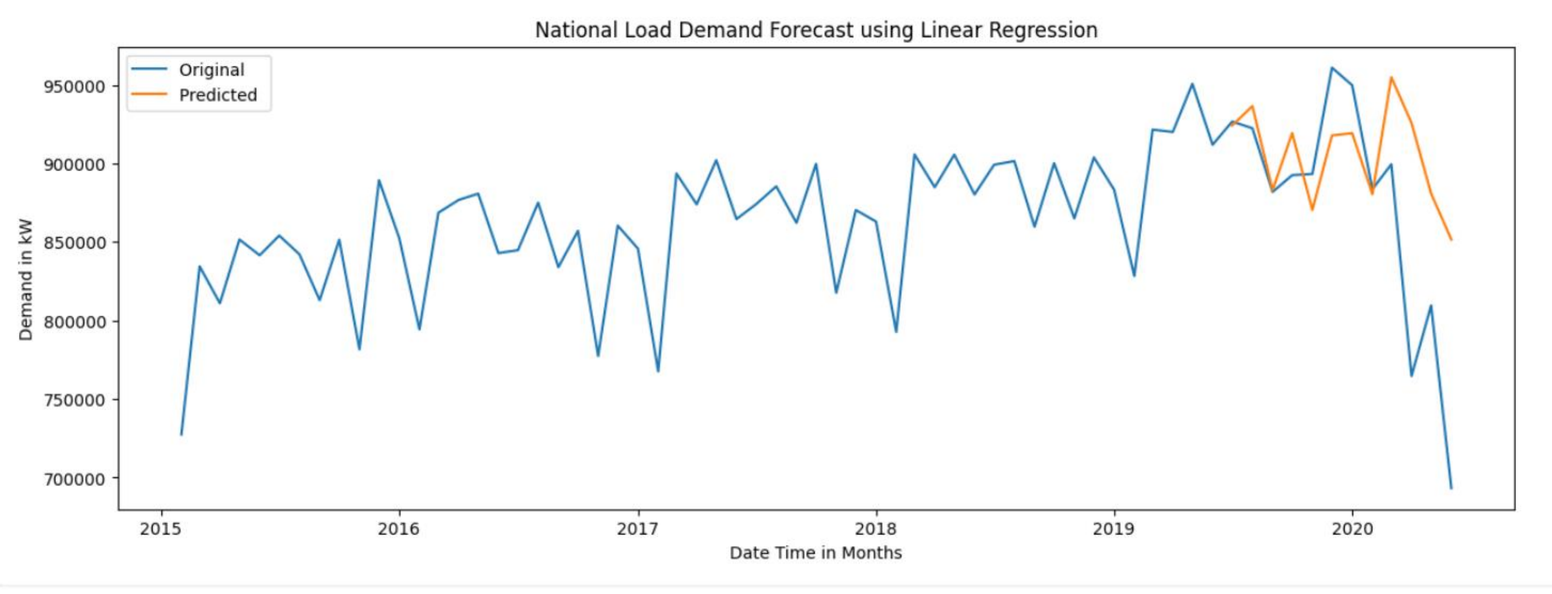
DATE	TEST DATA	PREDICTED VALUES
2018-06-30	880275.513500	832943.531199
2018-07-31	899181.589600	864517.484130
2018-08-31	901531.617850	855958.947029
2018-09-30	859703.530100	833249.130336
2018-10-31	900163.013300	850059.042015
2018-11-30	865062.682600	803958.630069
2018-12-31	903805.973500	877417.721637

RANDOM FOREST OUTPUT



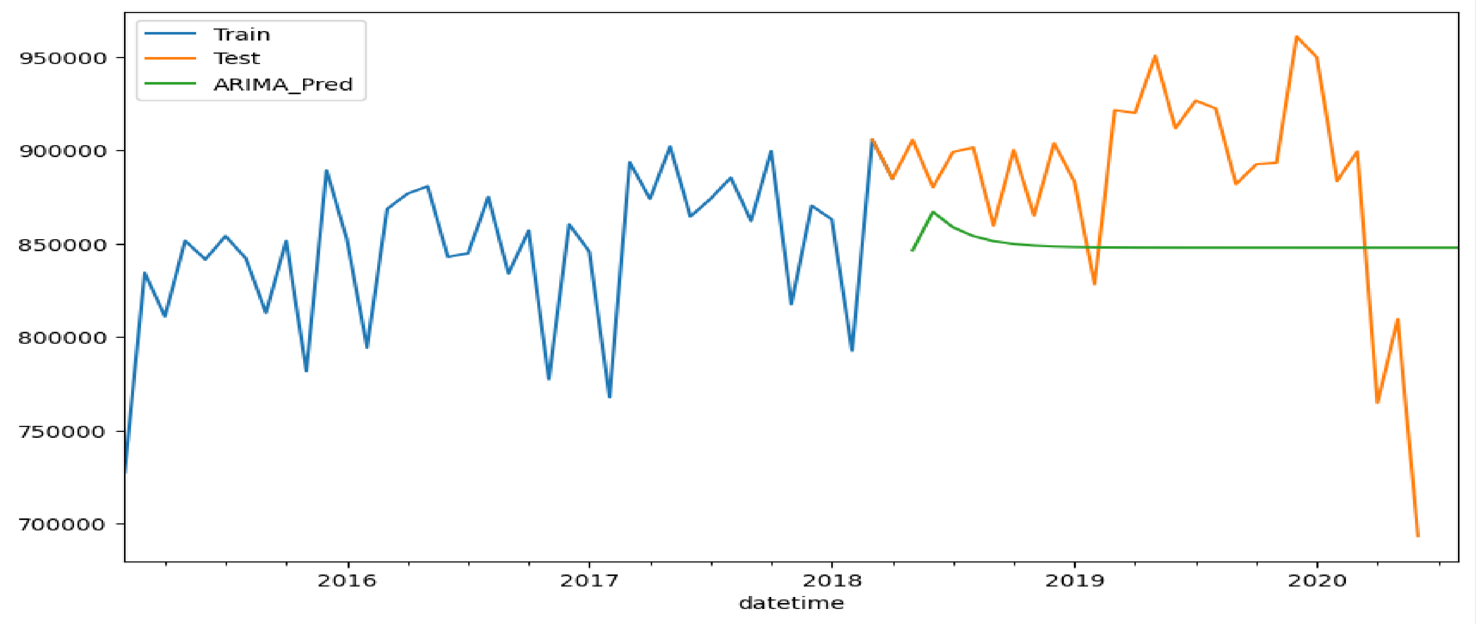
MODEL	Random Forest
Root Mean Squared Error (RMSE)	134.95257812661671
R-squared (R2) Score	0.47335130993430474

LSTM LINEAR REGRESSION OUTPUT



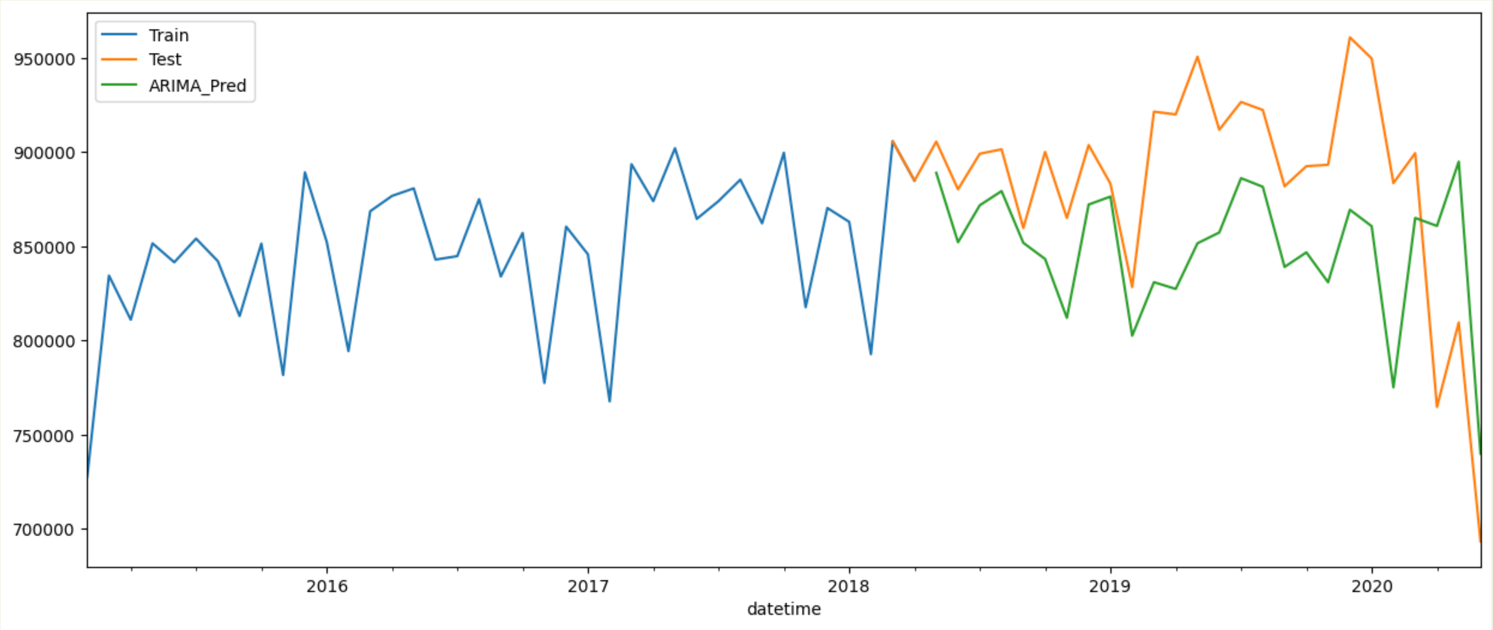
MODEL	Linear Regression
Root Mean Squared Error (RMSE)	170.75527121552568
R-squared (R2) Score	0.1568460569635063

ARIMA without exo variables OUTPUT



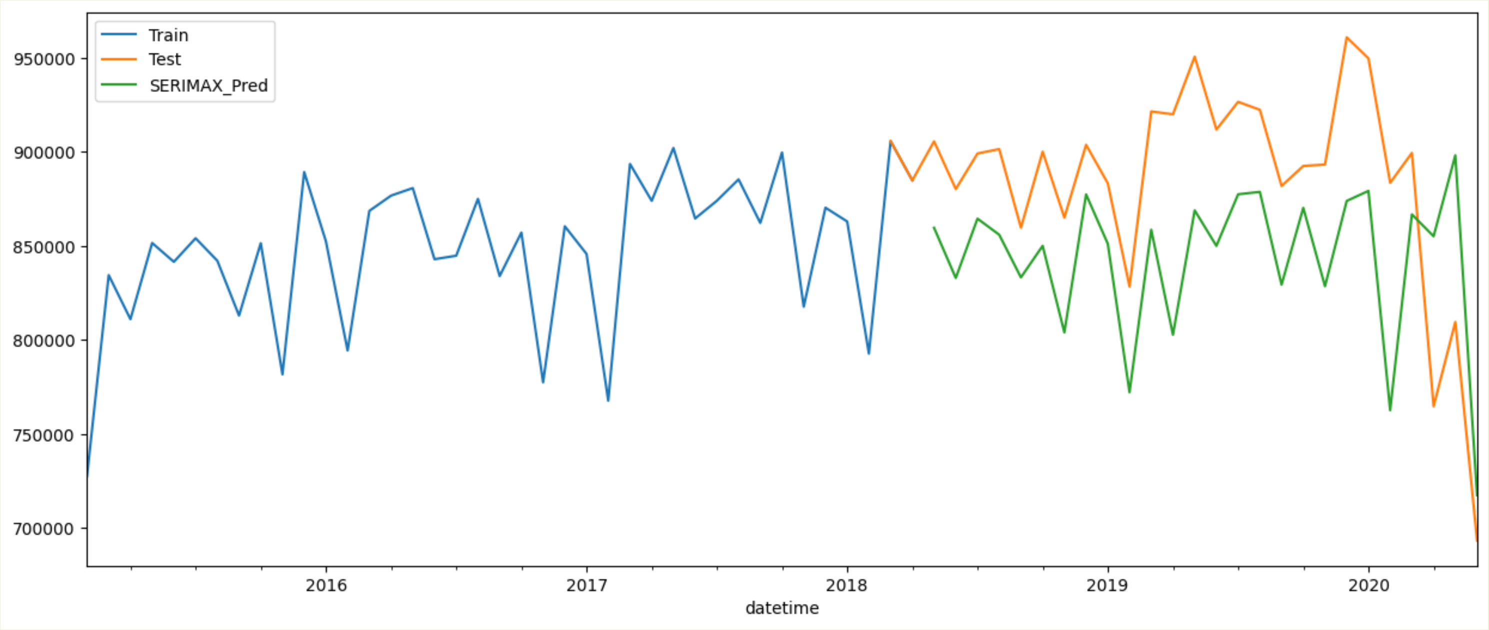
MODEL	ARIMA
Root Mean Squared Error (RMSE)	12745.584882387127
R-squared (R2) Score	-1.7426633795099633

ARIMA with exo variables OUTPUT



MODEL	ARIMA with exogenous variables
Root Mean Squared Error (RMSE)	8515.136350323046
R-squared (R2) Score	-0.22415435440763765

ARIMA/SARIMAX with exo variables RESULTS



MODEL	SARIMAX with exogenous variables
Root Mean Squared Error (RMSE)	8320.40974073744
R-squared (R2) Score	-0.16880590239907822

OUTPUT RESULTS

MODEL	ELASTIC NET
Root Mean Squared Error (RMSE)	181.0937999437399
R-squared (R2) Score	0.05165621935602005
RMSE for Baseline Model (mean prediction)	189.43165503806821
Performance Percentage	4.401510979066692

MODEL	BAYESIAN REGRESSION
Root Mean Squared Error (RMSE)	174.20693697072895
R-squared (R2) Score	0.12241432333123692
RMSE for Baseline Model (mean prediction)	189.43165503806821
Performance Percentage	8.037050652532018

OUTPUT RESULTS

MODEL	RIDGE REGRESSION
Root Mean Squared Error (RMSE)	174.10043071271087
R-squared (R2) Score	0.1234870681471878
RMSE for Baseline Model (mean prediction)	189.43165503806821
Performance Percentage	8.09327476037538

MODEL	HUBER REGRESSION
Root Mean Squared Error (RMSE)	171.223942675962
R-squared (R2) Score	0.1522113011840347
RMSE for Baseline Model (mean prediction)	189.43165503806821
Performance Percentage	9.61175805513983

OUTPUT RESULTS

MODEL	LINEAR REGRESSION
Root Mean Squared Error (RMSE)	170.75527121552568
R-squared (R2) Score	0.1568460569635063
RMSE for Baseline Model (mean prediction)	189.43165503806821
Performance Percentage	9.859167317515816

MODEL	LASSO REGRESSION
Root Mean Squared Error (RMSE)	170.60616179784125
R-squared (R2) Score	0.158317956484766
RMSE for Baseline Model (mean prediction)	189.43165503806821
Performance Percentage	9.93788141503794

OUTPUT RESULTS

MODEL	SVM
Root Mean Squared Error (RMSE)	165.65802749486647
R-squared (R2) Score	0.20643298360725137
RMSE for Baseline Model (mean prediction)	189.43165503806821
Performance Percentage	12.549976158115806

MODEL	KNN
Root Mean Squared Error (RMSE)	144.4713124990982
R-squared (R2) Score	0.3964380215238891
RMSE for Baseline Model (mean prediction)	189.43165503806821
Performance Percentage	23.734334438420433

OUTPUT RESULTS

MODEL	DECISION TREE
Root Mean Squared Error (RMSE)	143.20956572140287
R-squared (R2) Score	0.4069344569179547
RMSE for Baseline Model (mean prediction)	189.43165503806821
Performance Percentage	24.400404096863614

MODEL	ARIMA
Root Mean Squared Error (RMSE)	12745.584882387127
R-squared (R2) Score	-1.7426633795099633
RMSE for Baseline Model (mean prediction)	16233.529103447227
Performance Percentage	21.48605025335758

BEST OUTPUT RESULTS

MODEL	RANDOM FOREST
Root Mean Squared Error (RMSE)	134.95257812661671
R-squared (R2) Score	0.47335130993430474
RMSE for Baseline Model (mean prediction)	189.43165503806821
Performance Percentage	28.759225537307042

MODEL	GRADIENT BOOSTING
Root Mean Squared Error (RMSE)	127.28027133252687
R-squared (R2) Score	0.5315310386286948
RMSE for Baseline Model (mean prediction)	189.43165503806821
Performance Percentage	32.809396979111746

BEST OUTPUT RESULTS

MODEL	ARIMA with exogenous variables
Root Mean Squared Error (RMSE)	8515.136350323046
R-squared (R2) Score	-0.22415435440763765
RMSE for Baseline Model (mean prediction)	16233.529103447227
Performance Percentage	47.54599387440136

MODEL	SARIMAX with exogenous variables
Root Mean Squared Error (RMSE)	8320.40974073744
R-squared (R2) Score	-0.16880590239907822
RMSE for Baseline Model (mean prediction)	16233.529103447227
Performance Percentage	48.74552731130669

INFERENCE

- Through careful testing and thorough examination in my project, it became clear that the SARIMAX (Seasonal Auto Regressive Integrated Moving Average with exogenous Regressors) algorithm is the best for predicting load.
- After trying out different machine learning methods, SARIMAX consistently showed better results than the others. It's really good at understanding complex time patterns and considering outside factors, which helps it make more accurate predictions.
- This discovery not only confirms SARIMAX as the top choice for load prediction but also emphasizes the importance of following a structured approach and knowing the specific field well when using machine learning. By using SARIMAX in my project, I've not only improved load forecasting but also laid the groundwork for smarter energy management decisions.

CONCLUSION

- In conclusion, the project "Comparative Analysis of Load Forecasting Algorithms Using Machine Learning" stands as a testament to the relentless pursuit of excellence in energy management systems.
- Through an exhaustive examination of diverse machine learning and statistical algorithms, coupled with meticulous data preprocessing and feature engineering, the project has provided invaluable insights into the realm of load forecasting.
- The comparative analysis conducted has illuminated the relative strengths and weaknesses of each algorithm, offering a roadmap for energy practitioners and decision-makers to navigate the intricacies of load prediction with confidence and clarity. Moreover, the integration of exogenous variables has underscored the importance of considering external factors in enhancing forecasting accuracy, paving the way for more robust and adaptive energy management strategies. As the project draws to a close, it the trajectory of energy management towards a more sustainable and resilient future.

REFERENCES

[1]. Agus Setiawan; Zainal Arifin; Budi Sudiarto; Fauzan Hanif Jufri; Qasthalani Haramaini; Iwa Garniwa, “Comparison of Medium-Term Load Forecasting Methods (Splitted Linear Regression and Artificial Neural Networks) in Electricity Systems Located in Tropical Regions” 2022 3rd International Conference on Clean and Green Energy Engineering (CGEE)

[2]. Pande Popovski; Goran Veljanovski; Mitko Kostov; Metodija Atanasovski, “Optimizing Short Term Load Forecast: A study on Machine Learning Model Accuracy and Predictor Selection” 2022 57th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST)

[3]. M. Abdullah Al Amin; Md. Ashraful Hoque, “Comparison of ARIMA and SVM for Short-term Load Forecasting” 2019 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON)

REFERENCES

[4]. K. W. Yu; C. H. Hsu; S. M. Yang, “A Model Integrating ARIMA and ANN with Seasonal and Periodic Characteristics for Forecasting Electricity Load Dynamics in a State” 2019 IEEE 6th International Conference on Energy Smart Systems (ESS)

[5]. Rajat Sethi; Jan Kleissl, “Comparison of Short-Term Load Forecasting Techniques” 2020 IEEE Conference on Technologies for Sustainability (SusTech)

[6]. Mansi Bhatnagar; Vivek Dwivedi; Divyanshu Singh; Gregor Rozinaj, “Comprehensive Electric load forecasting using ensemble machine learning methods” 2022 29th International Conference on Systems, Signals and Image Processing (IWSSIP)

REFERENCES

[7]. Jinjin Zhang; Tao Wang; Junyong Wu; Hainan Zhu; Dong Lan; Fengshuo Li, “Short-term Load Forecasting Method Based on Artificial Intelligence Highway Neural Network” 2021 IEEE 5th Conference on Energy Internet and Energy System Integration (EI2)

[8]. Mansi Bhatnagar; Vivek Dwivedi; Divyanshu Singh; Gregor Rozinaj, “Comprehensive Electric load forecasting using ensemble machine learning methods” 2022 29th International Conference on Systems, Signals and Image Processing (IWSSIP)

[9]. Akanksha Jain; S.C. Gupta, “Peak load Forecasting using Machine Learning Algorithms” 2023 IEEE Renewable Energy and Sustainable E-Mobility Conference (RESEM).

REFERENCES

[10]. M. Abdullah Al Amin; Md. Ashraful Hoque, "Comparison of ARIMA and SVM for Short-term Load Forecasting" 2019 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON)

[11]. Rizwan A Khan; C. L. Dewangan; S. C. Srivastava; S. Chakrabarti, "Short Term Load Forecasting using SVM Models" 2018 IEEE 8th Power India International Conference (PIICON)

[12]. Stefan Ungureanu; Vasile Ţopa; Andrei Cziker, "Integrating the industrial consumer into smart grid by load curve forecasting using machine learning" 2019 8th International Conference on Modern Power Systems (MPS)

REFERENCES

[13]. Can Wang; Thomas Bäck; Holger H. Hoos; Mitra Baratchi; Steffen Limmer; Markus Olhofer, “Automated Machine Learning for Short-term Electric Load Forecasting” 2019 IEEE Symposium Series on Computational Intelligence (SSCI)

[14]. Denis Sidorov; Qing Tao; Ildar Muftahov; Aleksei Zhukov; Dmitriy Karamov; Aliona Dreglea; Fang Liu, “Energy balancing using charge/discharge storages control and load forecasts in a renewable-energy-based grids” 2019 Chinese Control Conference (CCC)

[15]. Lingxiao Wang; Shiwen Mao; Bogdan Wilamowski, “Short-Term Load Forecasting with LSTM Based Ensemble Learning” 2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)

REFERENCES

[16]. Ihsan A. S. Abu Amra; Ashraf Y. A. Maghari, “Forecasting Groundwater Production and Rain Amounts Using ARIMA-Hybrid ARIMA: Case Study of Deir El-Balah City in GAZA” 2018 International Conference on Promising Electronic Technologies (ICPET)

[17]. Stylianos I. Vagropoulos; G. I. Chouliaras; E. G. Kardakos; C. K. Simoglou; A. G. Bakirtzis, “Comparison of SARIMAX, SARIMA, modified SARIMA and ANN-based models for short-term PV generation forecasting” 2016 IEEE International Energy Conference (ENERGYCON)

[18]. Akshita Gupta; Arun Kumar, “Mid Term Daily Load Forecasting using ARIMA, Wavelet-ARIMA and Machine Learning” 2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe)