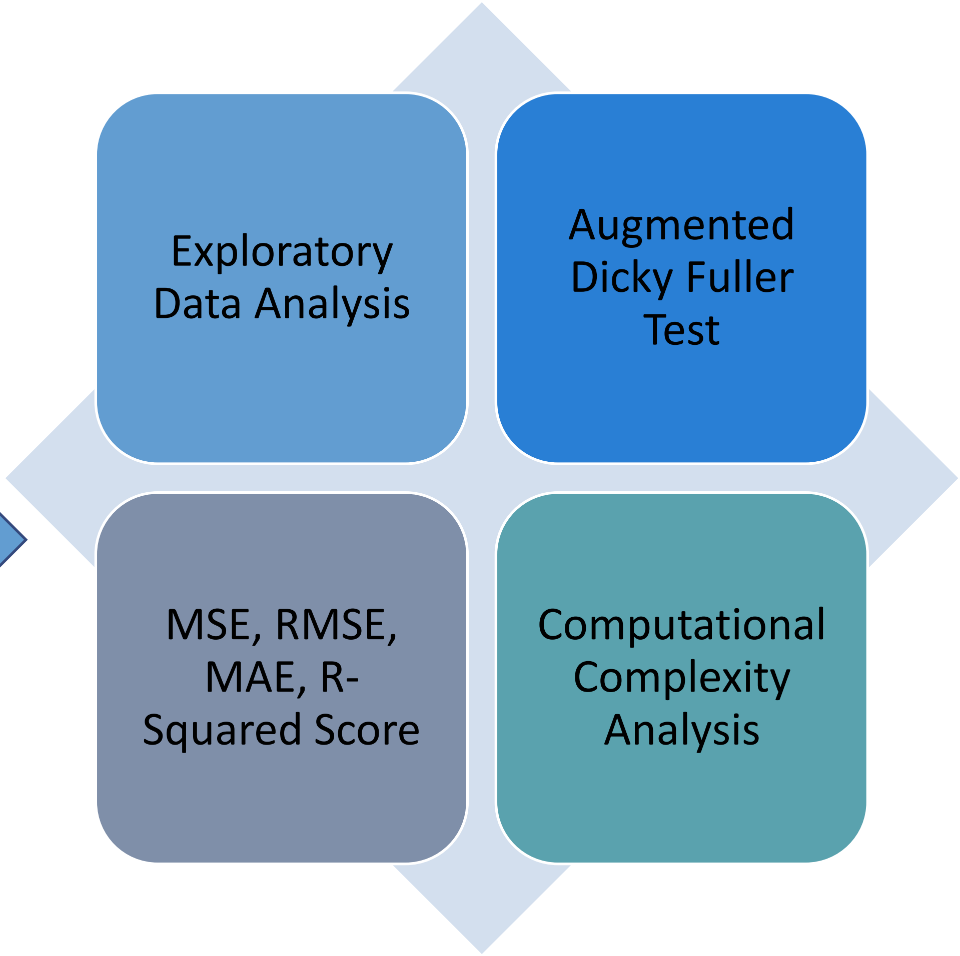
MACHINE LEARNING BASED MULTIVARIATE TIME SERIES FORE-CASTING OF ISO NEW ENGLAND ELECTRICITY DEMAND

Breakdown

- What has been Updated
- Introduction
- Research Objectives
- Machine Learning Algorithms Used
- Research Methodology
- Results Interpretation
- Conclusion
- Questions and Answering Session



What has been revised

Level of achievement of project	Poor	The approach and results produced need to be explained in greater detail and linked to the application much more.	I will focus on the improvement of chapter 4 and 5, so the results can be explained and linked to the applications
Identification and Analysis of Problem	Marginal Fail	The student sets out a reasonable set of aims for the project, but the results do not meet these aims.	I will improve the results interpretation so it can meet the criteria
Related Work	Marginal Fail	The Related work section contains a lot of information about forecasting generally and applications to various areas, but only a small amount is relevant to predicting energy usage.	done
Design and Justification	Clear Fail	The overall approach is broadly acceptable, but, in Chapter 4, there needs to be more detail on why the respective plots were chosen, what they find and how it relates to the overall problem. Just doing a set of exploratory data analsys plots, with little justification for why these ones were chosen, is not apprpriate. More significantly, beyond some vague statements at the start of Chapter 5, on how Machine Learning is supposedly wonderful, there is no discussion of the application (energy use) nor attempt to justify the approaches taken.	I will update the chapter 4 and 5, so their results can be clearly explained and their impact on the energy forecasting or what insights they provide will be discussed.
Critical Evaluation	Clear Fail	In Chapter 4, the results from the Exploratory data analysis is done in very vague terms. There are statements that "I can can spot discrepancies, mistakes, or anomalies by visualizing the data", but this is not done, and a majority of the detail is about what histograms, box-and-whisker plots etc are generally. In addition, statements like "on Thursday, the average recorded electricity demand was higher" should be backed up with evidence (statictial analysis) given the differences are so small. (This also appears to be contradicted in Figured 12, which shows Thursday slightly lower.) Chapter 5 similarly has a lack of detail on the application of the results.	I will modify the chapter 4 in a way so its is presented with critical evaluation. Moreover, in chapter 5 I will link it to the application along with the description of other datasets and how those are connected to this. Moreover, I will include the explanation of taken models, stationary concepts and energy prediction examples.
References	Pass	[Quality and quantity of referencing.]	
Presentation	Marginal Fail	The presentation is acceptable in Chapters 1-4, but the layout of Chapter 5 is not appropriate for the aim of the work.	I will improve the captions and labels. Along with improving the layout of chapter 5.
Overall Mark [0-100]	Clear Fail	The overall aim of the prioject is acceptable, but the stduent needs to show more evidence of a understanding of the results in terms of the application, and a mucg deeper understanding of the Machine Learning approach and output.	I will improve the indicated issues and problems.

Introduction

- Significance of Electric Load Forecasting (ELF) in Energy Market
- Compromised Forecasting Since Decades
  - Example: Under developing has no proper data set
  - Example: Lack of input parameters i.e., Humidity, Temperature etc.
- Compromised Energy Trends
- Need of effective prediction tools

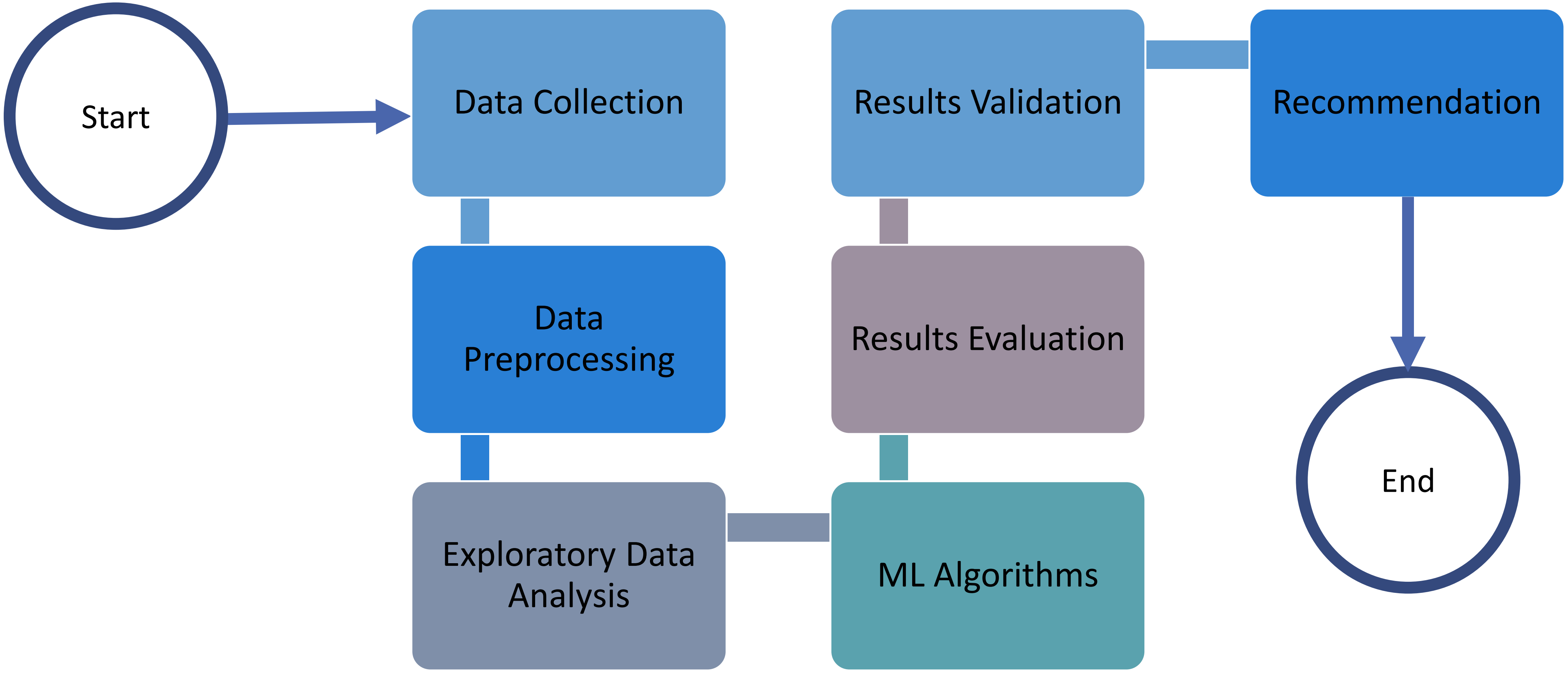
Research Objectives

- To conduct a thorough exploratory data analysis to learn more about the traits, trends, and connections found in the ME Zone electricity demand dataset.
- To successfully explain the results and offer insightful context for the patterns of electricity demand, use narrative and visualizations.
- To utilize the Random Forest, XGBoost, CatBoost, and Prophet algorithms to anticipate power consumption using multivariate time series forecasting approaches.
- To use measures like RMSE (Root Mean Square Error), MAPE (Mean Absolute Percentage Error), R2 Score, MAE (Mean Absolute Error), and computational complexity analysis to assess and compare the performance of the suggested algorithms.

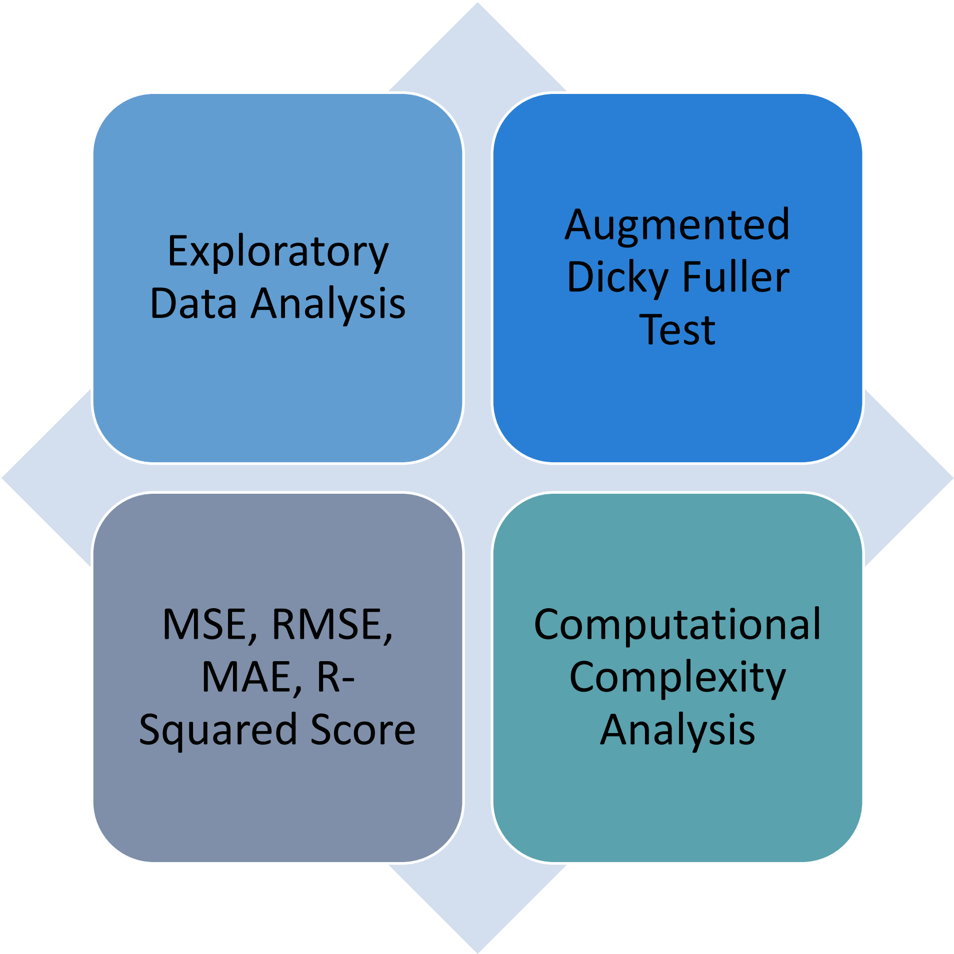
Machine Learning Algorithms Used

Algorithm	Random Forest	CatBoost	XGBoost	Prophet
Type	Ensemble (Bagging)	Ensemble (Boosting)	Ensemble (Boosting)	Time Series Forecasting
Architecture	Collection of decision trees	Collection of decision trees	Collection of decision trees	Trend, seasonality, holiday components
Categorical Handling	No special handling	Automatic handling	No	N/A
Tree Construction	Recursive feature selection (information gain/Gini)	Specialized algorithm, ordered boosting	Gradient-based optimization	N/A
Parallelism	Can be parallelized	Supports multicore & GPU	Can be parallelized	N/A
Feature Importance	Yes	Yes	Yes	No
Missing Data Handling	Yes	Yes	Yes	No

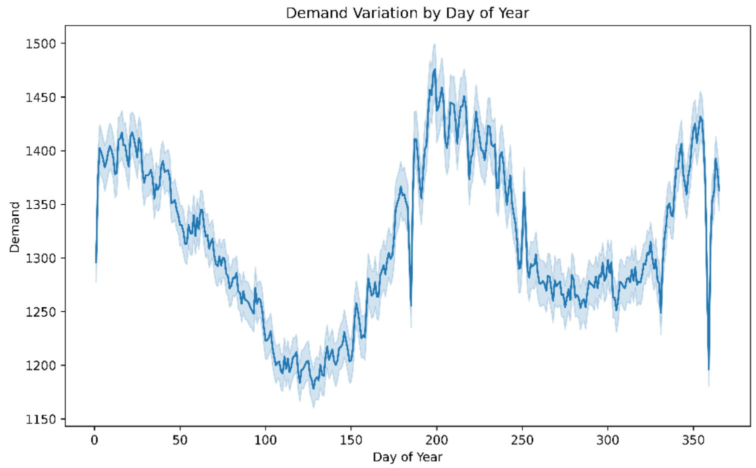
Research Methodology

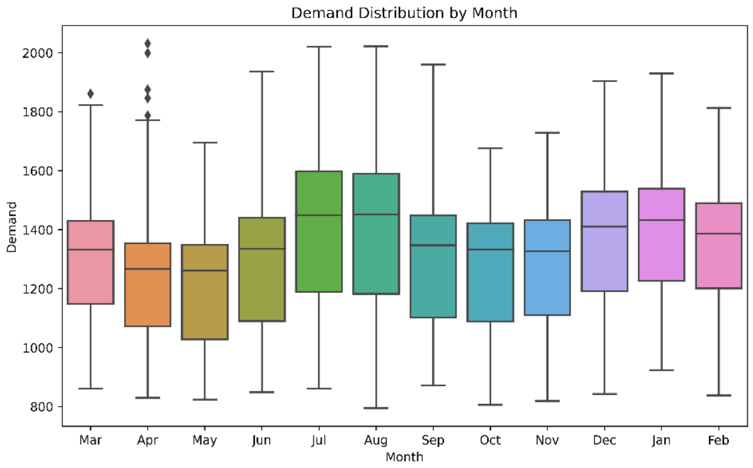


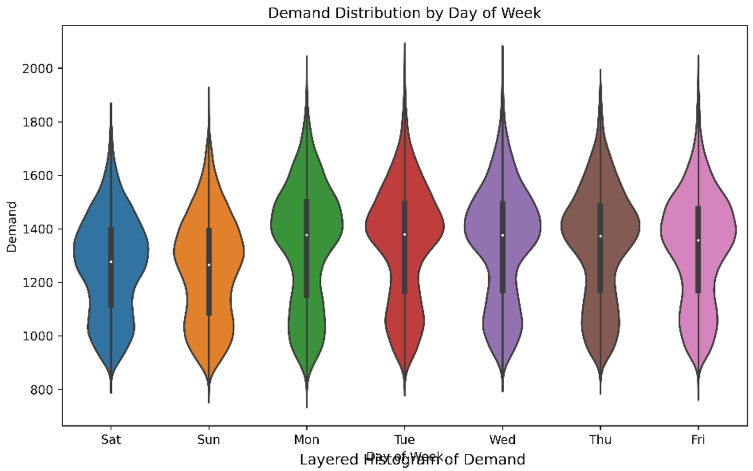
Results Interpretation

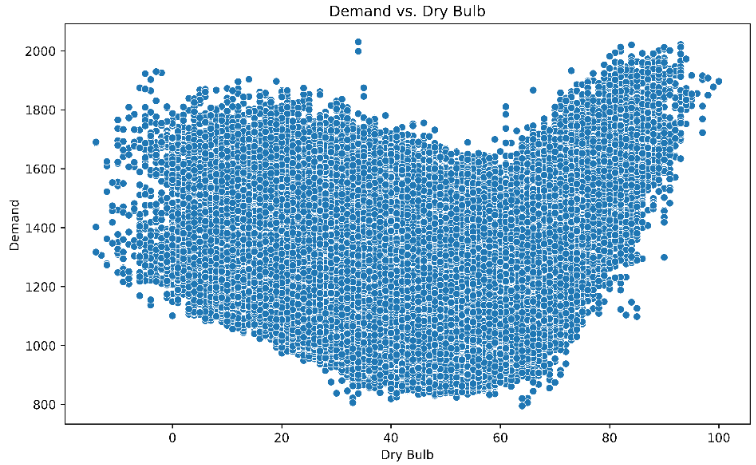


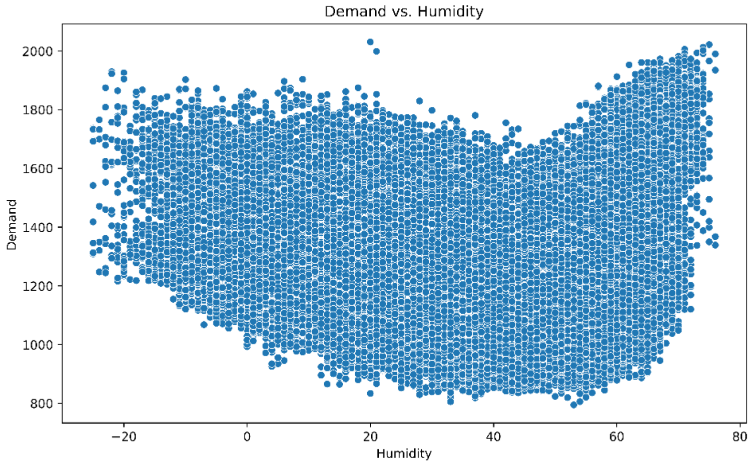
Exploratory Data Analysis

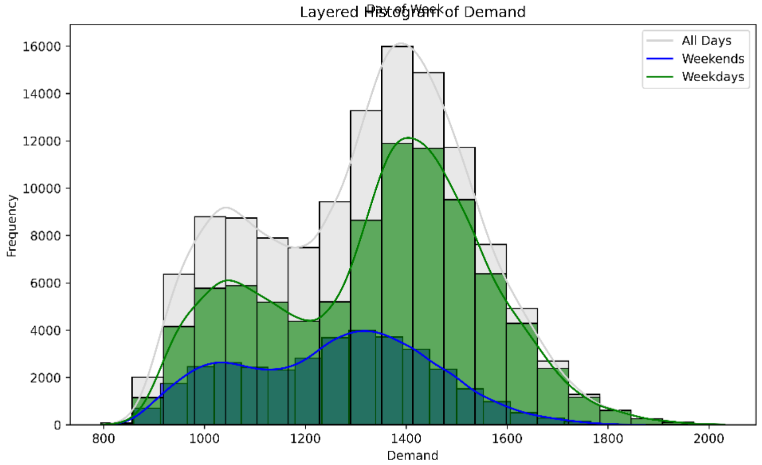












Augmented Dicky Fuller Test

Main Dataset

Attributes		Feature Variable	Target Variable 1	Target Variable 2
Dataset 1	ADF Value	-19.6346	-9.8057	-11.6349
	p-value	0.0000	0.0000	0.0000
	Critical Value 1%	-3.4304	-3.4304	-3.4304
	Critical Value 5%	-2.8616	-2.8616	-2.8616

Feature Variable: Demand
Target Variable 1: Temperature
Target Variable 2: Humidity

For Validation

Attributes		Feature Variable	Target Variable 1	Target Variable 2
Dataset 2	**ADF Value**	**-1.0177**	**-0.9152**	**-0.9817**
	p-value	0.7467	0.7829	0.7599
	Critical Value 1%	-3.4395	-3.4394	-3.4393
	Critical Value 5%	-2.8656	-2.8655	-2.8655
Dataset 3	**ADF Value**	**-9.5118**	**-1.4996**	**-0.9447**
	p-value	0.0000	0.5337	0.7729
	Critical Value 1%	-3.4486	-3.4487	-3.4490
	Critical Value 5%	-2.8696	-2.8696	-2.8698
Dataset 4	**ADF Value**	**-8.6222**	**-2.1543**	**-1.9100**
	p-value	0.0000	0.2232	0.3274
	Critical Value 1%	-3.4349	-3.4349	-3.4349
	Critical Value 5%	-2.8635	-2.8635	-2.8635
Dataset 5	**ADF Value**	**-0.8491**	**-0.7669**	**-0.8363**
	p-value	0.8043	0.8287	0.8082
	Critical Value 1%	-3.4350	-3.4350	-3.4350
	Critical Value 5%	-2.8636	-2.8636	-2.8636

Model Evaluations

Main Dataset

Attributes		Random Forest	CatBoost	XGBoost	Prophet
Dataset 1	MSE	4614.47	3658.15	**3633.99**	4415.99
	RMSE	67.92	60.48	**45.53**	68.89
	MAE	50.77	45.89	**60.28**	50.78
	R^2 Score	0.88	0.90	**0.90**	0.89

For Validation

Attributes		Random Forest	CatBoost	XGBoost	Prophet
Dataset 2	MSE	**19540.62**	40768.26	21611.41	19440.32
	RMSE	**139.78**	201.91	129.74	145.98
	MAE	**121.49**	189.71	147.00	122.07
	R^2 Score	**-3.08**	-7.53	-3.52	-2.89
Dataset 3	MSE	5851.95	**5615.39**	6827.75	5721.68
	RMSE	76.49	**74.93**	56.78	74.98
	MAE	53.33	**51.91**	82.63	53.01
	R^2 Score	0.35	**0.38**	0.25	0.35
Dataset 4	MSE	1149.27	1158.70	**1128.25**	1152.25
	RMSE	33.90	34.03	**18.25**	34.56
	MAE	18.41	17.78	**35.75**	18.32
	R^2 Score	0.31	0.30	**0.23**	0.31
Dataset 5	MSE	638.76	729.58	**627.73**	635.12
	RMSE	25.27	27.01	**17.62**	24.12
	MAE	17.76	19.54	**25.05**	18.45
	R^2 Score	-0.09	-0.24	**-0.07**	-0.09

Computational Complexity Analysis

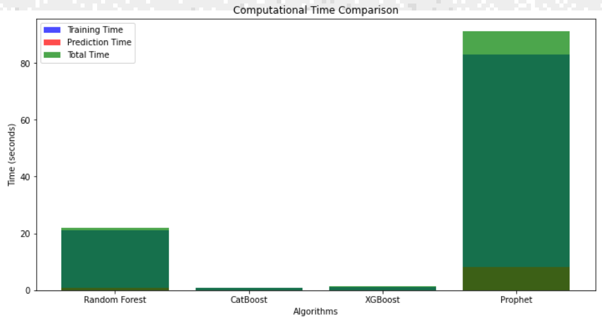
Main Dataset

		Random Forest	CatBoost	XGBoost	Prophet
Dataset 1	Training	20.9662	**0.9093**	1.2517	82.8203
	Prediction	0.9036	**0.0040**	0.0160	8.2750
	Total Time	21.8698	**0.9132**	1.2676	91.0954

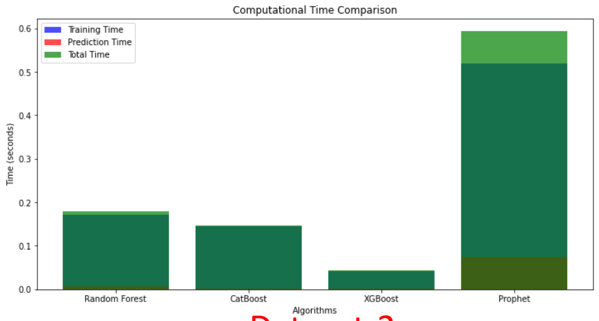
For Validation

		Random Forest	CatBoost	XGBoost	Prophet
Dataset 2	Training	0.1706	0.1452	0.0418	0.5186
	Prediction	0.0080	0.0021	**0.0020**	0.0748
	Total Time	0.1785	0.1472	**0.0438**	0.5934
Dataset 3	Training	0.1466	0.1305	**0.0349**	0.0841
	Prediction	0.0080	0.0020	**0.0020**	0.0489
	Total Time	0.1546	0.1325	**0.0369**	0.1330
Dataset 4	Training	0.4259	0.1616	**0.0878**	0.2762
	Prediction	0.0150	0.0010	**0.0030**	0.1255
	Total Time	0.4408	0.1626	**0.0908**	0.4017
Dataset 5	Training	0.2952	0.1476	**0.0788**	1.0472
	Prediction	0.0100	0.0010	**0.0020**	0.1137
	Total Time	0.3052	0.1486	**0.0808**	1.1609

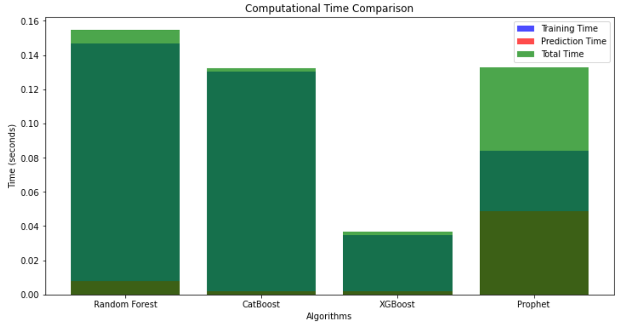
Computational Complexity Analysis (Cont.)



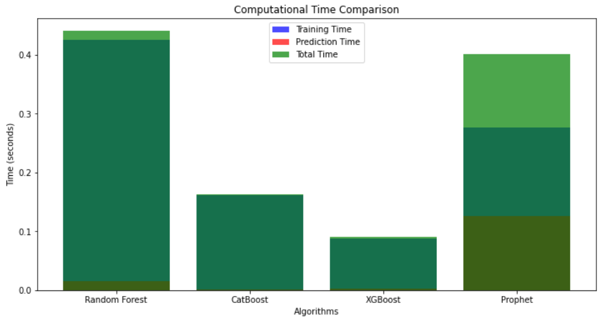
Dataset-1



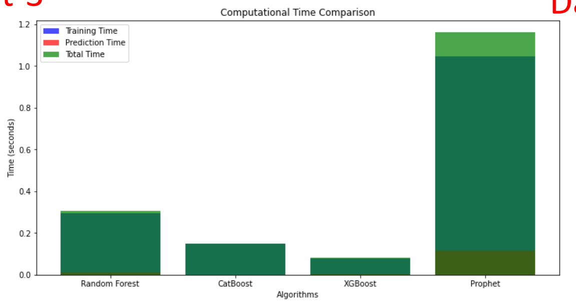
Dataset-2



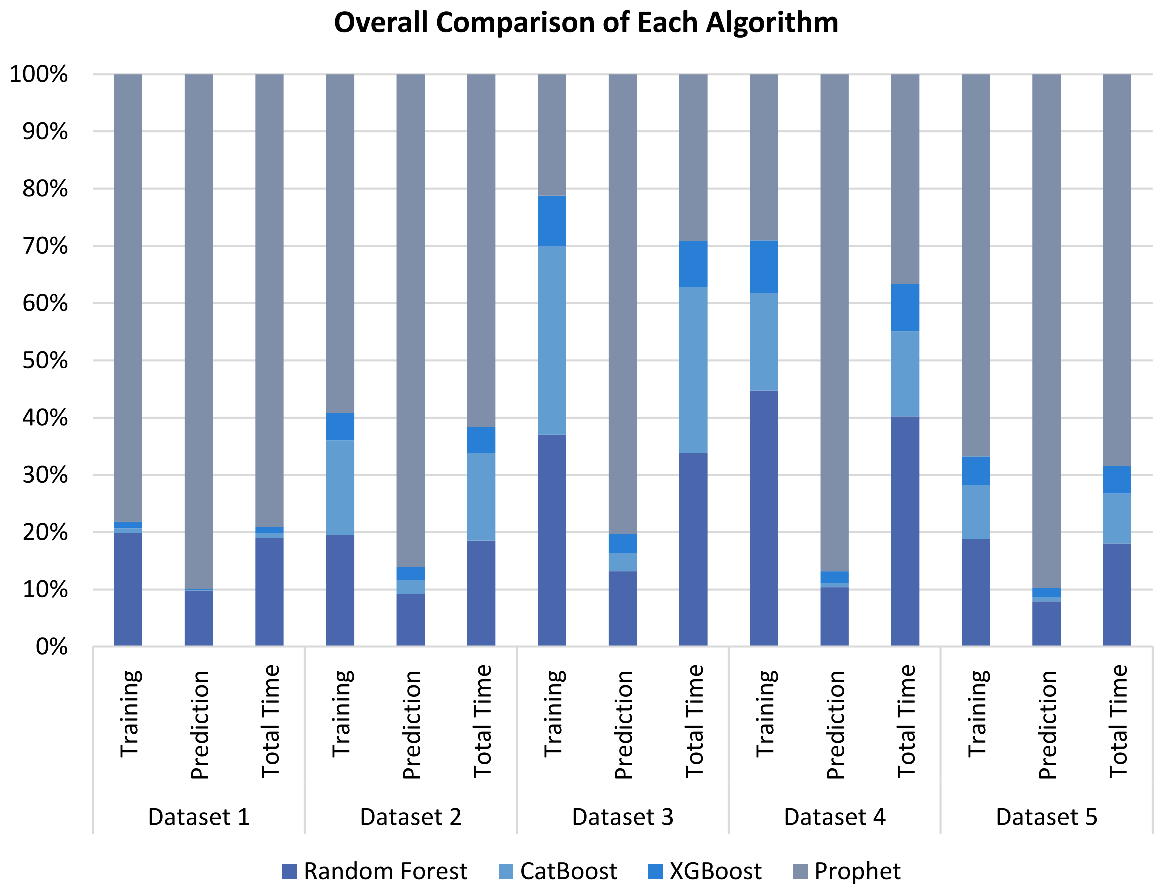
Dataset-3



Dataset-4

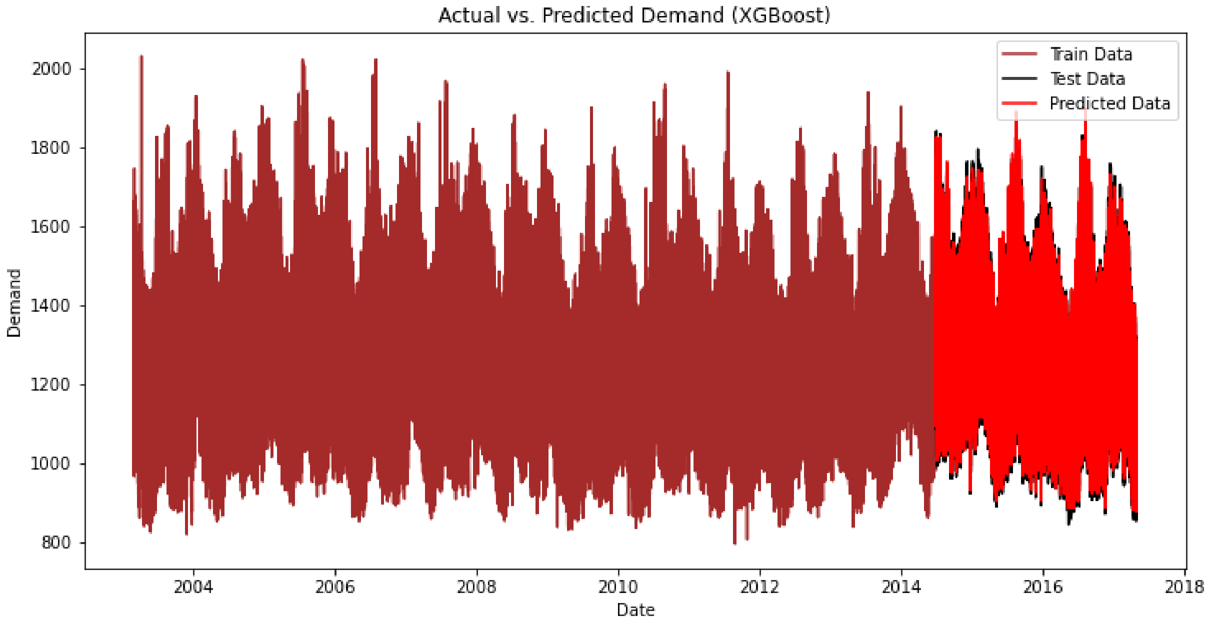


Dataset-5



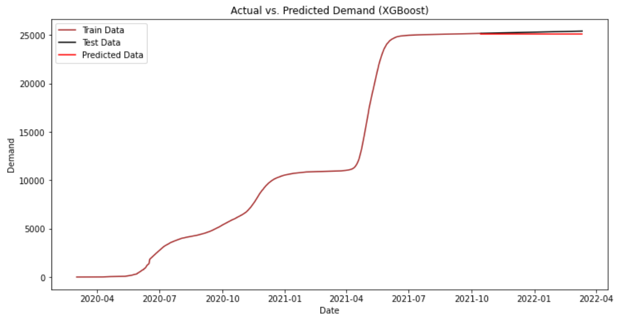
Best Performing Algorithm

Main Dataset

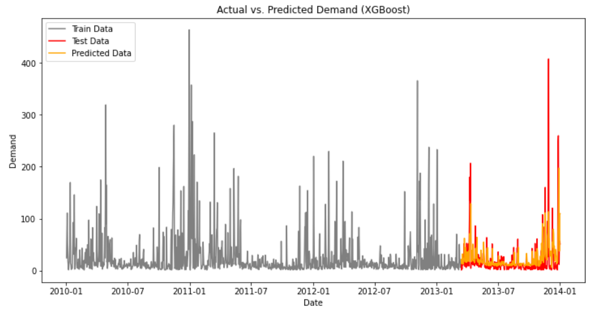


Dataset-1

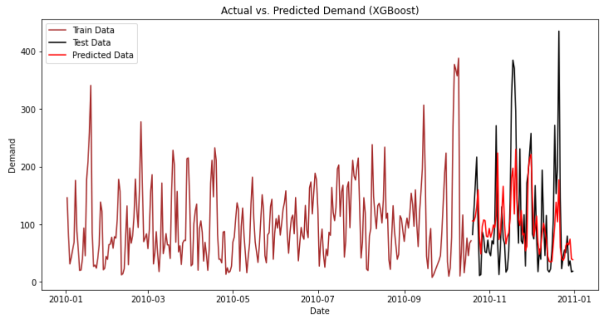
For Validation



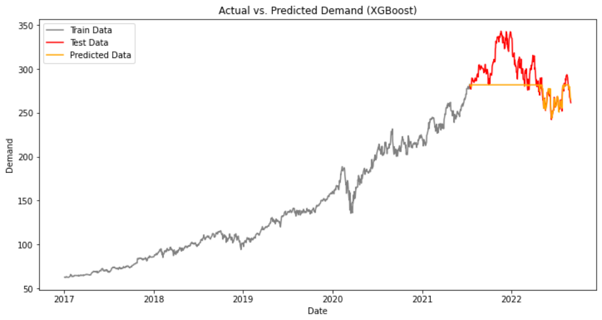
Dataset-2



Dataset-3



Dataset-4



Dataset-5

Conclusion

- Electricity Consumption patterns are dependent on the weather parameters and consumer behaviors.
- Machine learning solution are easy, time efficient and require less-computing powers.
- Among decision tree-based architectures XGBoost provides effective results as compared to others.
- Effective predictions can be obtained through machine learning algorithms.
- Accuracy can be enhanced by using anomaly detection models.

THANK YOU