Available online www.ejaet.com

European Journal of Advances in Engineering and Technology, 2021, 8(10):83-87

Research Article ISSN: 2394 - 658X

A Comprehensive Exploration of Regression Techniques for Building Energy Prediction

Vibhu Sharma

Energy Engineer, Energy Group Grumman/Butkus AssociatesEvanston IL, USA
**sharmav9@icloud.com*

ABSTRACT

This paper presents a detailed investigation into various regression techniques applied to predict building energy consumption. The dataset utilized in this study encompasses diverse attributes related to buildings, including size, location, usage type, construction materials, and energy consumption patterns, alongside weather-related data such as temperature, humidity, and precipitation. After preprocessing, which includes loading, inspecting, and imputing missing data, the dataset undergoes evaluation using regression techniques including Ridge Regression, Bayesian Ridge, Linear Regression, Orthogonal Matching Pursuit, Elastic Net, Huber Regression, Lasso Regression, and Passive Aggressive Regression. Evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-squared (R2), Root Mean Squared Logarithmic Error (RMSLE), and Mean Absolute Percentage Error (MAPE) are employed to assess each model's performance. The results offer valuable insights into predictive accuracy, computational efficiency, and robustness, aiding in the selection of the most suitable regression technique for building energy prediction tasks.

Key words: Building energy prediction, Regression techniques, Ridge Regression, Bayesian Ridge, Linear Regression, Orthogonal Matching Pursuit, Elastic Net, Huber Regression, Lasso Regression, Passive Aggressive Regression, Data preprocessing, Model evaluation, Computational efficiency, Predictive accuracy, Robustness, Sustainability

---

INTRODUCTION

Building energy consumption prediction is crucial for optimizing energy usage, reducing costs, and mitigating environmental impacts. Regression analysis plays a pivotal role in this context, facilitating insights into the relationships between various predictors and energy consumption patterns. [1] Following are the different regression techniques:

Ridge Regression

Ridge Regression, a regularized version of Linear Regression, aims to mitigate overfitting by penalizing coefficients. It introduces a regularization term that shrinks coefficients, particularly those with high multicollinearity, thereby achieving a balance between fitting the training data and reducing model complexity. The strength of regularization is controlled by a tuning parameter, facilitating flexibility in model optimization. [2][3]

Bayesian Ridge Regression

Bayesian Ridge Regression is a probabilistic technique that incorporates prior knowledge about coefficient distributions. By employing Bayesian inference, it estimates posterior distributions of coefficients, rendering predictions more robust in the face of uncertainties. This principled framework accounts for uncertainties in both model parameters and predictions, offering a comprehensive approach to regression analysis. [4]

Linear Regression

Linear Regression, a fundamental technique in regression analysis, models predictor-target variable relationships linearly. It estimates coefficients that minimize the sum of squared errors, assuming linearity and additivity in the relationship between predictors and the target variable. Despite its simplicity, Linear Regression remains a powerful tool in predictive modeling, particularly when the relationship between predictors and the target variable is predominantly linear. [4]

Orthogonal Matching Pursuit

Orthogonal Matching Pursuit is a sparse regression method that aims to select a subset of relevant predictors to represent the target variable efficiently. It iteratively updates coefficient estimates by selecting predictors with the highest correlation with the residual, thereby identifying critical features for prediction. Orthogonal Matching Pursuit excels in handling high-dimensional data and is particularly useful in scenarios where feature selection is crucial for model interpretability and efficiency. [5]

Elastic Net Regression

Elastic Net Regression combines the strengths of Ridge and Lasso Regression by introducing both L1 and L2 penalties. This hybrid approach balances the sparsity-inducing property of Lasso and the regularization effect of Ridge Regression, effectively addressing multicollinearity and selecting pertinent predictors. By tuning the mixing parameter, Elastic Net Regression provides a flexible framework for model optimization, accommodating various data characteristics and modeling objectives. [6]

Huber Regression

Huber Regression is a robust regression technique that minimizes a combination of squared and absolute errors. By introducing a tuning parameter δ, Huber Regression adapts the threshold between quadratic and absolute loss functions, making it less susceptible to outliers while maintaining computational efficiency. This robust approach is particularly beneficial in scenarios where the data exhibit heteroscedasticity or contain influential outliers that may distort model estimates. [7]

Lasso Regression

Lasso Regression, leveraging L1 regularization, encourages sparsity in coefficient estimates, facilitating feature selection. Controlled by the regularization strength parameter λ, Lasso Regression balances model complexity and prediction accuracy, effectively shrinking less important coefficients to zero. This feature selection capability makes Lasso Regression particularly useful in scenarios where interpretability and parsimony are essential considerations in model development.[7]

Passive Aggressive Regression

Passive Aggressive Regression is a type of online learning algorithm suitable for large-scale regression tasks. By updating model parameters incrementally, Passive Aggressive Regression adapts to changing data distributions efficiently, making it suitable for streaming data scenarios or situations where retraining the model from scratch is impractical. This adaptive approach ensures computational efficiency and memory-friendliness, enabling real-time updates to model predictions as new data becomes available. [8][6]

METHODOLOGY

Data Preprocessing

The dataset undergoes preprocessing, including loading, inspecting, and imputing missing data, to ensure its suitability for analysis. This process optimizes memory usage and enhances computational efficiency, preparing the dataset for regression model evaluation.

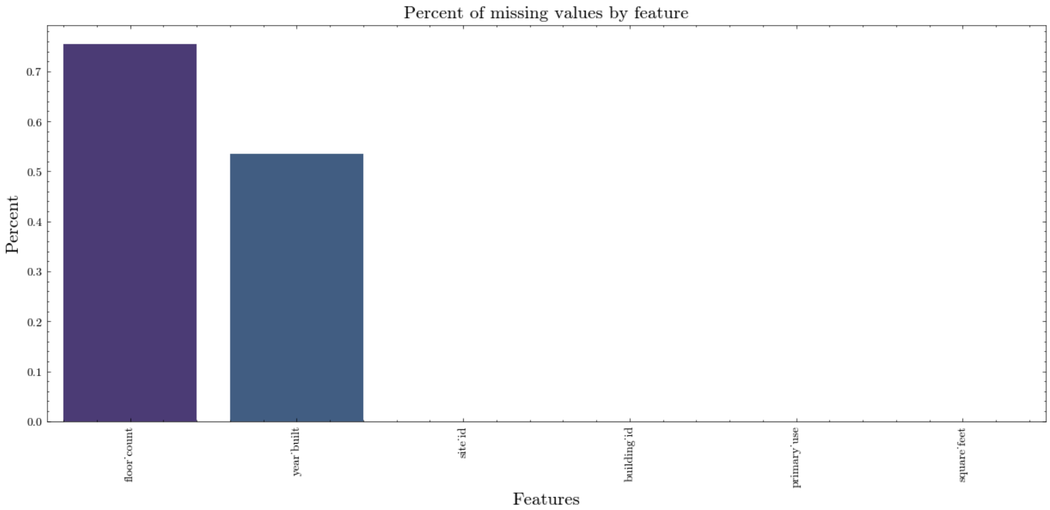

Figure 1: Missing Data in Building Metadata Dataset

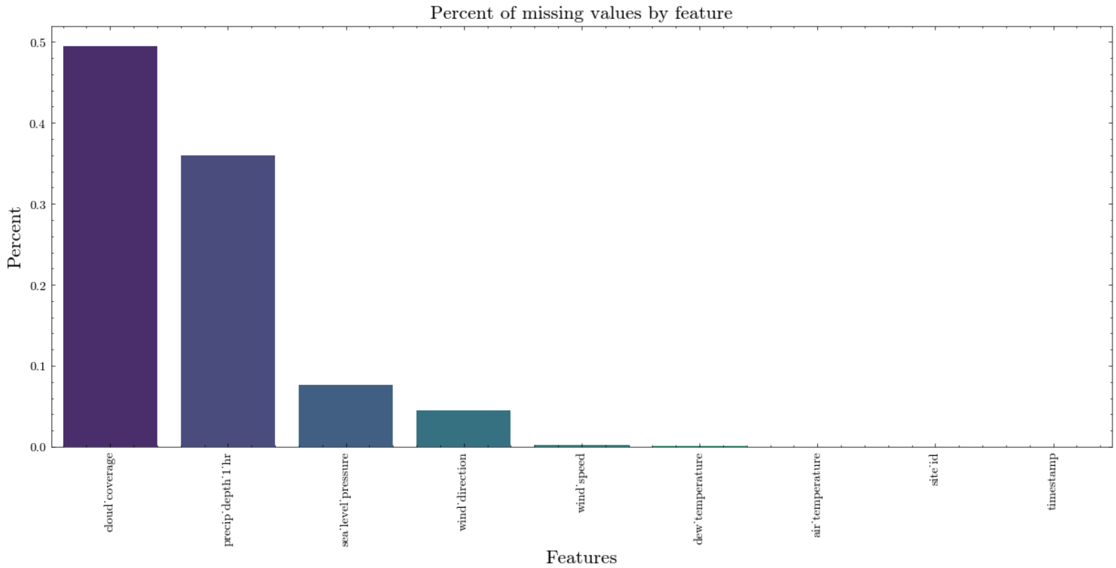

Figure 2: Missing Data in Weather (Training) Dataset

Figure 3: Hourly & Daily Mean Plot of Various Meter from Weather Dataset

Regression Model Evaluation

Various regression techniques are applied to the preprocessed dataset, and each model is trained and evaluated using standard metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-squared (R^2), Root Mean Squared Logarithmic Error (RMSLE), and Mean Absolute Percentage Error (MAPE). Computational efficiency is also assessed in terms of training time (TT) and memory usage to understand each model's performance and scalability.

RESULTS ANALYSIS

Performance metrics obtained from each regression model are analyzed to identify the most suitable technique for predicting building energy consumption. Factors such as predictive accuracy, computational efficiency, and robustness to outliers are considered during the analysis. Additionally, visualization techniques such as scatter plots and residual plots are utilized to gain further insights into each model's performance and identify potential areas for improvement.

RESULTS

The results offer valuable insights into predictive accuracy, computational efficiency, and robustness, aiding in the selection of the most suitable regression technique for building energy prediction tasks.

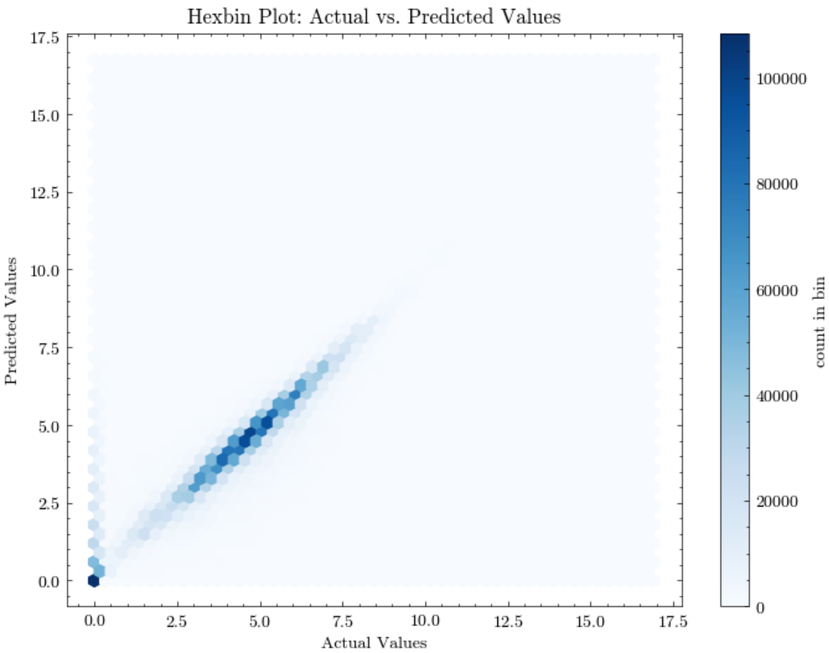

Figure 4: Hourly & Daily Mean Plot of Various Meter from Weather Dataset

Table 1: Summary of Regressions

Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT(Sec)
Ridge Regressiion	1.3189	3.3581	1.8325	0.2710	0.5615	4.9009	107.3067
Bayesian Ridge	1.3189	3.3581	1.8325	0.2710	0.5615	4.9009	147.4733
Linear Regression	1.3189	3.3583	1.8326	0.2709	0.5616	4.9005	143.7400
Orthogonal Matching Pursuit	1.3974	3.6621	1.9137	0.2050	0.5789	6.0528	96.9667
Elastic Net	1.5296	3.9842	1.9959	0.1352	0.5931	6.5091	108.7467
Huber Regressor	1.4737	3.9842	1.9960	0.1351	0.6059	7.0570	537.0633
Lasso Regression	1.6525	4.4140	2.1010	0.0417	0.6086	6.7712	126.2733
Lasso Least Angle Regression	1.6525	4.4140	2.1010	0.0417	0.6086	6.7712	96.5267
Passive Aggressive Regressor	2.1208	8.4533	2.7623	-0.8355	0.7108	4.9804	127.1600
Least Angle Regression	9587.9	468856129.3	12594.9	-101701614.4	4.7	16232.1	3791.8

CONCLUSION

This study provides valuable insights into the efficacy of different regression techniques for building energy prediction tasks. Understanding their strengths and limitations empowers researchers and practitioners to make informed decisions when developing predictive models for enhancing energy efficiency and sustainability in buildings. Future research endeavors may explore advanced regression techniques, ensemble methods, and hybrid approaches to further enhance predictive accuracy and model robustness in building energy prediction tasks.

REFERENCES

[1]. Y. Geng, W. Ji, B. Lin, J. Hong, and Y. Zhu, "Building Energy Performance diagnosis using energy bills and weather data," Energy and Buildings, vol. 172, pp. 181–191, Aug. 2018. doi:10.1016/j.enbuild.2018.04.047

[2]. S. C. Keat, B. B. Chun, L. H. San, and M. Z. M. Jafri, "Multiple regression analysis in modelling of carbon dioxide emissions by energy consumption use in Malaysia," AIP Conference Proceedings, Jan. 2015, doi: 10.1063/1.4915185.

[3]. J. S. Hygh, J. F. DeCarolis, D. B. Hill, and S. R. Ranjithan, "Multivariate regression as an energy assessment tool in early building design," Building and Environment, vol. 57, pp. 165–175, Nov. 2012, doi: 10.1016/j.buildenv.2012.04.021.

[4]. S. Asadi, S. S. Amiri, and M. Mottahedi, "On the development of multi-linear regression analysis to assess energy consumption in the early stages of building design," Energy and Buildings, vol. 85, pp. 246–255, Dec. 2014, doi: 10.1016/j.enbuild.2014.07.096.

[5]. D. Hsu, "Identifying key variables and interactions in statistical models of building energy consumption using regularization," Energy, vol. 83, pp. 144–155, Apr. 2015, doi: 10.1016/j.energy.2015.02.008.

[6]. Ghareeb, W. Wang, and K. P. Hallinan, "Data-driven modelling for building energy prediction using regression-based analysis," Dec. 2019, doi: https://doi.org/10.1145/3368691.3368701.

[7]. R. K. Jain, T. Damoulas, and C. E. Kontokosta, "Towards Data-Driven Energy Consumption Forecasting of Multi-Family Residential Buildings: Feature Selection via The Lasso," Computing in Civil and Building Engineering (2014), Jun. 2014, doi: 10.1061/9780784413616.208.

[8]. H. Naganathan, W. K. Chong, and N. Ye, "Learning Energy Consumption and Demand Models through Data Mining for Reverse Engineering," Procedia Engineering, vol. 118, pp. 1319–1324, Jan. 2015, doi: 10.1016/j.proeng.2015.11.392.