Feature Importance Analysis and Predictive Modeling for Boston Prices Using Regression Algorithms via Brute Force Hyperparameter Tuning

Josh Kenn A. Viray
University of Santo Tomas, joshkenn.viray.cics@ust.edu.ph

Abstract - This study investigates the application of various regression models, including Linear Regression, Ridge, Lasso, and ElasticNet, to predict housing prices using the Boston Housing Dataset. By systematically tuning hyperparameters such as the regularization parameter (α) and ElasticNet's L1-to-L2 scaling ratio (l1_ratio), we evaluated model performance using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 scores through brute force recursive algorithms. The results highlight the superior performance of ElasticNet Regression with $\alpha = 0.3$ and $l1\_ratio = 0.1$, achieving an MSE of 22.99, RMSE of 4.79, and R^2 of 0.684. Feature importance analysis identified LSTAT, RAD, and TAX as the most influential predictors, while features like NOX and AGE showed negligible importance (R^2 of 0.680). This study underscores the importance of hyperparameter tuning, feature selection, and preprocessing techniques in developing accurate and interpretable predictive models for real estate data.

Index Terms– Hyperparameter tuning, ElasticNet Regression, Feature importance, Regularization, Machine learning, Predictive modeling, Regression analysis

I. INTRODUCTION

House prices are influenced by multiple factors, ranging from environmental conditions to structural attributes. Understanding these relationships is essential for building predictive models and making informed decisions in real estate. This study utilizes the Boston housing dataset to examine feature importance and identify key predictors of house prices. The Gradient Boosting Regressor, known for its robust feature importance estimation, was employed to compute the significance of individual features.

II. METHODOLOGY

A. Dataset (Boston Housing Dataset)

The Boston housing dataset comprises 13 input features and one target variable (MEDV), representing the median house prices in $1,000s. Features include:

- **CRIM**: Per capita crime rate by town.
- **ZN**: Proportion of residential land zoned for lots over 25,000 sq. ft.
- **INDUS**: Proportion of non-retail business acres per town.
- **CHAS**: Charles River dummy variable.
- **NOX**: Nitric oxide concentration.
- **RM**: Average number of rooms per dwelling.
- **AGE**: Proportion of owner-occupied units built prior to 1940.
- **DIS**: Weighted distances to five Boston employment centers.
- **RAD**: Index of accessibility to radial highways.
- **TAX**: Full-value property tax rate per $10,000.
- **PTRATIO**: Pupil-teacher ratio by town.
- **B**: Proportion of Black residents by town.
- **LSTAT**: Percentage of lower-status population

B. Exploratory Data Analysis

For the exploratory data analysis, a heatmap was used with pair plots to have an initial understanding of the data and how they correlate.

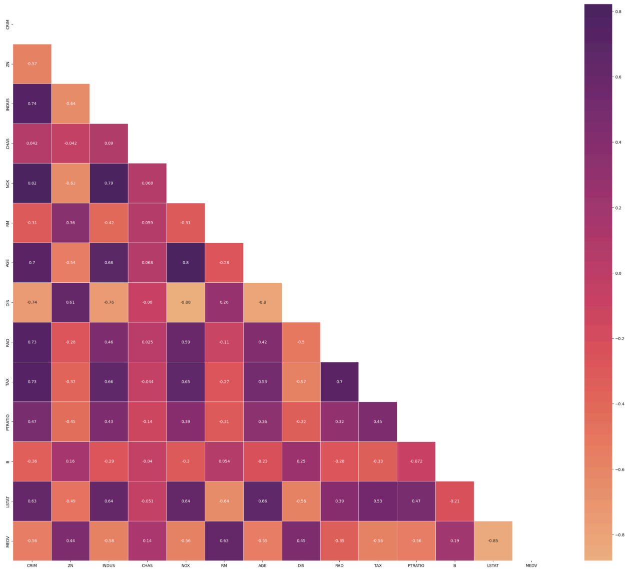


Fig. 1.
Heatmap of all features from the (base) dataset

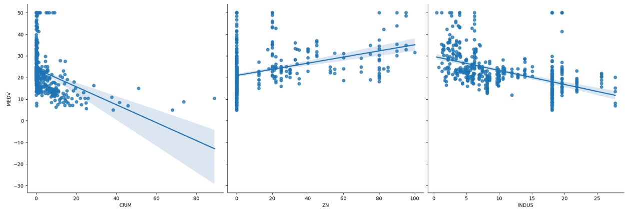


Fig. 2.
Pairplot of CRIM, ZN, and INDUS against MEDV

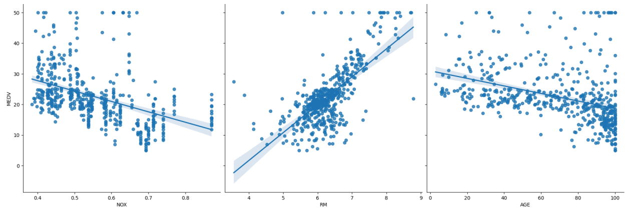


Fig. 3.
Pairplot of NOX, RM, and AGE against MEDV

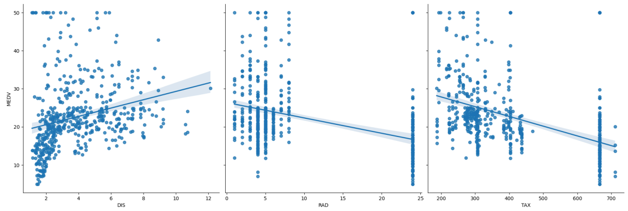


Fig. 4.
Pairplot of DIS, RAD, and TAX against MEDV

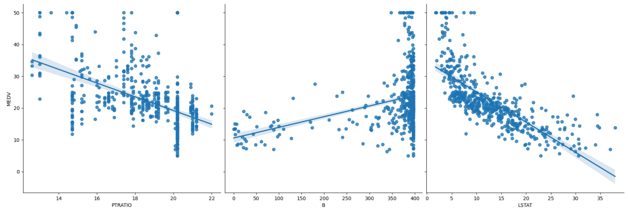
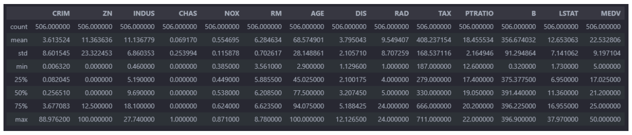


Fig. 5.
Pairplot of PTRATIO, B, and LSTAT against MEDV



	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	MEDV
count	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000	506.000000
mean	3.613524	11.363636	11.136779	0.069170	0.554695	6.284634	68.574901	3.795043	9.549407	408.237154	18.455534	356.674032	12.653063	22.532806
std	8.601545	23.322453	6.860353	0.253994	0.115878	0.702617	28.148861	2.105710	8.707259	168.537116	2.164946	91.294864	7.141062	9.197104
min	0.006320	0.000000	0.460000	0.000000	0.385000	3.561000	2.900000	1.129600	1.000000	187.000000	12.600000	0.320000	1.730000	5.000000
25%	0.082045	0.000000	5.190000	0.000000	0.449000	5.885500	45.025000	2.100175	4.000000	279.000000	17.400000	375.377500	6.950000	17.025000
50%	0.256510	0.000000	9.690000	0.000000	0.538000	6.208500	77.500000	3.207450	5.000000	330.000000	19.050000	391.440000	11.360000	21.200000
75%	3.677083	12.500000	18.100000	0.000000	0.624000	6.623500	94.075000	5.188425	24.000000	666.000000	20.200000	396.225000	16.955000	25.000000
max	88.976200	100.000000	27.740000	1.000000	0.871000	8.780000	100.000000	12.126500	24.000000	711.000000	22.000000	396.900000	37.970000	50.000000

Table 1.
Data description of all columns

The analysis of feature correlations with housing prices in Boston reveals distinct relationships between various factors and the target variable (MEDV). Weakly correlated features include CRIM (per capita crime rate), which shows a negative correlation with housing prices, as lower crime rates generally correspond to higher property values. Similarly, ZN (proportion of residential land zoned for large lots) has a positive correlation, with higher zoning proportions linked to higher prices. INDUS (non-retail business acres) negatively correlates with housing prices, indicating that an increase in industrial areas tends to reduce property values.

On the other hand, several features exhibit moderate to strong correlations. NOX (nitric oxide concentration) has a negative correlation, where higher pollution levels correspond to lower prices. RM (average number of rooms per dwelling) stands out as one of the strongest predictors, showing a positive correlation, as houses with more rooms generally have higher values. AGE (older homes) negatively correlates with prices, while DIS (distance to employment centers) positively influences housing costs. Accessibility features such as RAD (highway accessibility) and TAX (property tax rate) have negative correlations, as proximity to highways and higher tax rates tend to lower home values. Additionally, PTRATIO (pupil-teacher ratio) negatively correlates, suggesting that areas with better education systems command higher prices.

While B (proportion of Black residents) has a weak positive correlation, LSTAT (percentage of lower-status population) strongly negatively correlates, with higher percentages of lower-income residents associated with lower housing prices.

These findings highlight that most features exhibit linear relationships with housing prices, though their predictive significance varies. As the next step, we will implement Linear Regression as a baseline model and apply feature selection techniques to retain the most influential variables, ensuring optimal predictive performance. Given the continuous nature of the target variable, regression techniques will be employed to develop an effective housing price prediction model.

C. Data Cleaning and Preprocessing

In the initial analysis of the Boston housing dataset, it was observed that several features exhibited a high number of outliers. These extreme values can distort statistical analyses and negatively impact machine learning models by introducing bias and reducing generalization performance.

To mitigate the effect of outliers, trimming the tails of the data distribution was applied. This technique involves removing extreme values from both ends of the distribution while retaining the central portion of the data.

In this study, percentile-based trimming was applied to remove extreme values from both ends of the distribution. Specifically:

- The 1st percentile (lower 1%) and the 99th percentile (upper 1%) were selected as cut-off points.
- Any values below the 1st percentile or above the 99th percentile were removed.

The data was also made sure that it would not contain any blanks or null values in the features presented in the dataset.

Feature scaling is a crucial preprocessing step in machine learning, particularly for models like Linear Regression, Ridge, Lasso, and Elastic Net, which are sensitive to feature magnitudes. Without proper scaling, features with larger numerical ranges can dominate the model, leading to biased predictions. In the Boston housing dataset, numerical features such as CRIM (crime rate) and ZN (proportion of residential land zoned) have significantly different ranges, which can hinder model performance.

To address this, StandardScaler was applied, ensuring that all numerical features have a mean of 0 and standard deviation of 1, thereby bringing them onto a similar scale. This standardization improves model convergence, stability, and interpretability. The formula used for standardization is

$$X_{scaled} = \frac{X-\mu}{\sigma} \quad (1)$$

Where X represents the original value, μ is the mean, and σ is the standard deviation.

Additionally, handling categorical variables is another essential preprocessing step. CHAS (a binary variable indicating whether a property is near the Charles River) was the only categorical feature in this dataset. Given that CHAS is a dummy variable (taking values of 0 or 1), it was already in a suitable format for the model and did not require additional encoding. Since it is a binary indicator, models can naturally interpret it without transformation. Proper feature scaling and encoding contribute to a well-prepared dataset, allowing the model to learn effectively and make accurate predictions.

D. Model

Multiple regression models were used in order to find an algorithm that would fit the data the best with having mean standard error (MSE) and r-squared as quantifiable data to provide an understanding of the model's perceived understanding of the dataset, namely:

1. Linear Regression
2. Ridge Regression
3. Lasso Regression
4. Elastic Net Regression

The objectives of this study and the given algorithms were to accomplish the following:

1. Train on the dataset using all parameters.
2. Compute feature importance scores using its inherent mechanism based on impurity reduction.
3. Select high-value features based on the feature importance function.

III. RESULTS AND DISCUSSIONS

A. Baseline Model

A Linear Regression model was implemented as a starting point for predicting housing prices. Linear Regression is a simple yet powerful statistical method that assumes a linear relationship between the independent variables (features) and the dependent variable (target). In this case, the features include variables such as CRIM, RM, and LSTAT, while the target variable is MEDV, representing the median house price.

The performance of the Linear Regression model was evaluated using two key metrics: **Mean Squared Error (MSE)** and **R-squared (R^2)**. The model achieved an MSE of **25.06**, which reflects the average squared difference between the predicted and actual housing prices. While this value suggests some degree of error in the predictions, it provides a baseline for comparison with more sophisticated models.

Additionally, the R-squared value was **0.655**, indicating that approximately 65.5% of the variance in housing prices could be explained by the features in the dataset. While this is a reasonable starting point, the relatively low R^2 score suggests that there is room for improvement, either by using more advanced regression techniques or by incorporating feature engineering.

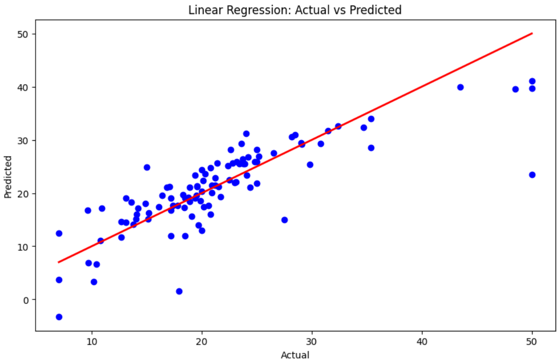


Fig. 6.
Actual vs. Predicted plot for the Linear Regression baseline model.

The **Actual vs Predicted plot** further illustrates the model's performance. The red line represents perfect predictions (where predicted values equal actual values), while the blue points are the actual predictions made by the model. Although the points generally follow the trend of the red line, some points deviate significantly, indicating that the model struggles with certain data points. These deviations highlight the potential influence of outliers, non-linear relationships, or interactions between features that a simple linear model cannot capture.

In summary, Linear Regression provides a strong foundation for understanding the dataset and establishing a baseline. However, the results indicate that more complex models may be necessary to achieve better predictive performance and account for the intricacies of the housing market, given a relatively rich feature set.

B. Advanced Models

In this section, we delve deeper into the performance of advanced regression algorithms applied to the Boston housing dataset. The **Model Evaluation Summary** below compares the algorithms based on two key metrics: **Mean Squared Error (MSE)** and **R-squared (R²)**. These metrics assess the predictive accuracy and explanatory power of the models, respectively.

Linear Regression serves as a baseline model for comparison. It assumes a linear relationship between the features and the target variable (MEDV) and achieved an MSE of 25.06 with an R² of 0.655, meaning it explains 65.5% of the variance in house prices. While this provides a good starting point, the model's assumptions may limit its ability to capture non-linear relationships in the data. To address potential overfitting and multicollinearity issues, Ridge Regression and Lasso Regression were applied:

1. **Ridge Regression** incorporates L2 regularization, penalizing large coefficients and stabilizing predictions. It achieved a similar MSE (**25.33**) and R² (**0.652**) as Linear Regression, indicating that regularization did not significantly improve performance in this case.

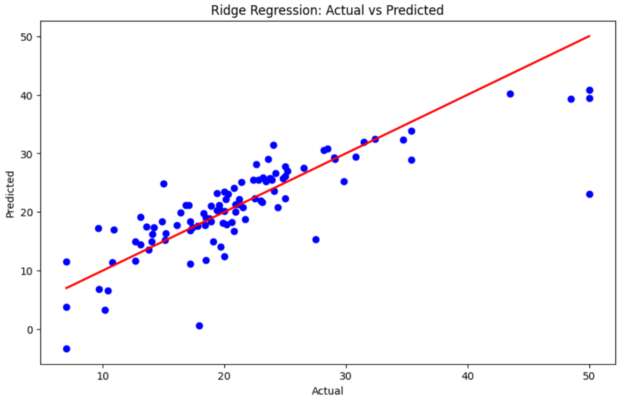


Fig. 7.
Actual vs. Predicted plot for the Ridge Regression model.

2. **Lasso Regression**, which uses L1 regularization, performed slightly worse with an MSE of **25.76** and an R² of **0.646**. This may be due to Lasso's tendency to shrink coefficients to zero, reducing the influence of certain features in the dataset.

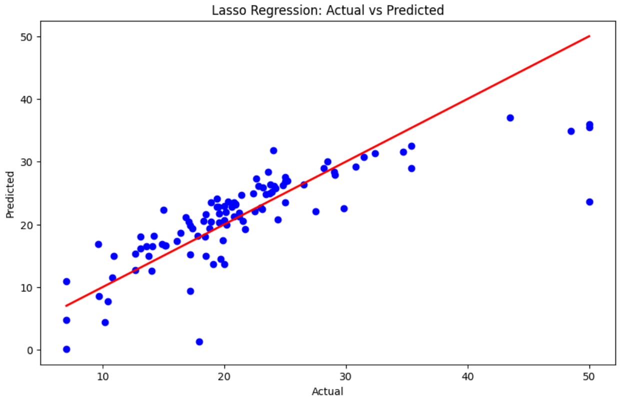


Fig. 8.
Actual vs. Predicted plot for the Lasso Regression model.

3. **Elastic Net Regression**, a hybrid of Ridge and Lasso, yielded an MSE of **26.05** and R² of **0.642**, showing no substantial improvement over individual regularization techniques. These results suggest that linear and regularized models may not be sufficient to capture the complexity of the Boston housing dataset.

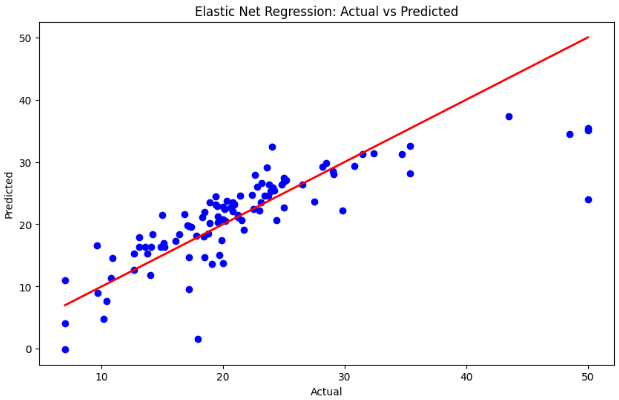


Fig. 9.
Actual vs. Predicted plot for the Elastic Net Regression model.

C. Hyperparameter Tuning

Hyperparameter tuning plays a crucial role in optimizing machine learning models. Using a brute-force strategy, this paper explores the impact of varying the regularization parameter (alpha) in ElasticNet, Ridge, and Lasso regression while also discussing various ElasticNet L1 and L2 scaling penalties (l1_ratio). By systematically iterating through different values of alpha, we assess model performance using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R² score. This research provides insights into the effectiveness of hyperparameter tuning and how it affects model predictive power.

Linear Regression does not inherently support hyperparameter tuning in the same way that Ridge, Lasso, or ElasticNet models do. This is because Linear Regression has no regularization parameters to modify, making it a purely deterministic approach that fits the data exactly as it is presented. Without regularization terms like alpha or penalties such as L1 or L2 norms, the model lacks mechanisms to prevent overfitting or underfitting based on feature variability or noise.

- **Mean Squared Error (MSE)** is relatively high compared to models with regularization, indicating that the model struggles with noise or outliers in the data.
- **Root Mean Squared Error (RMSE)** confirms the variability in predictions, as it is derived directly from MSE.
- **R² Score**, though moderately high (0.66), shows that the model explains a reasonable portion of the variance but can likely be improved with regularization.

The Ridge Regression graphs provide a detailed examination of the relationship between the regularization parameter α and key performance metrics—Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R²). These visualizations explore how different levels of regularization impact the model's ability to generalize and accurately predict outcomes, offering valuable insights into the behavior of Ridge Regression under varying conditions.

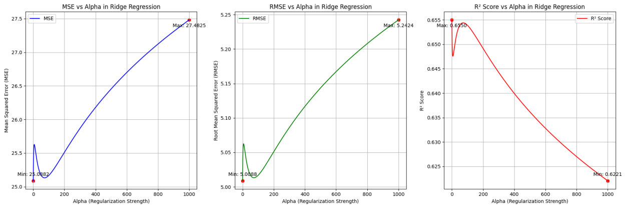


Fig. 9.
MSE, RMSE, R² vs Alpha in Ridge Regression (from 0 to 1000.1)

In Figure 10, α is tested across a broad range, specifically 0 to 1000.1, where it showcases its overall influence on model performance. At lower α values, the MSE reaches its minimum, indicating an optimal balance between bias and variance. As α increases, MSE rises steadily, signaling that the regularization strength is over-penalizing the model coefficients. This behavior aligns with theoretical expectations, where higher regularization restricts model complexity to prevent overfitting but risks underfitting the data.

The RMSE follows a similar trend to MSE, as it is derived from it, emphasizing the model's increasing inability to capture the data's underlying patterns as α grows. The R² score initially improves slightly as α stabilizes the model, but after a certain threshold, it declines sharply. This decline reflects the diminishing explanatory power of the model as it becomes overly simplistic due to excessive regularization.

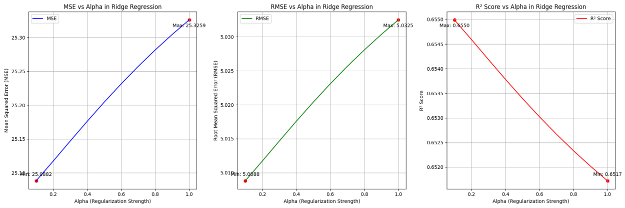


Fig. 10.
MSE, RMSE, R² vs Alpha in Ridge Regression (α from 0 to 1.1)

The second graph narrows the range of α, providing a more granular view of its effects at lower values. This focused analysis reveals that the minimum MSE and RMSE occur at the smallest α values in the range, further confirming that low regularization is most effective for this dataset. The R² score also reaches its peak in this range, emphasizing the importance of fine-tuning α\alphaα within a narrow window to achieve the best trade-off between model complexity and predictive accuracy.

Lasso Regression's behavior across varying values of the regularization parameter reveals how model performance is affected by regularization strength. By analyzing broad and narrow ranges, we can understand its impact on Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R².

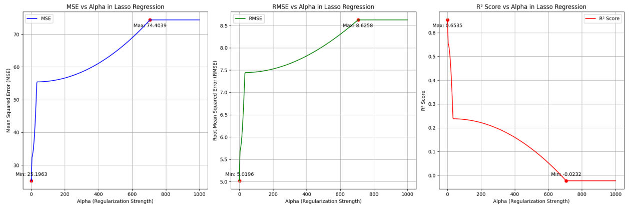


Fig. 11.
MSE, RMSE, R² vs Alpha in Lasso Regression (from 0 to 1000.1)

At lower values, MSE is minimized, reflecting adequate data fitting. As it increases, MSE rises sharply due to over-regularization, leading to underfitting. RMSE mirrors this trend, with higher values indicating reduced predictive accuracy. Similarly, starts high at low but declines as regularization strength increases, eventually becoming negative, signifying poor predictions.

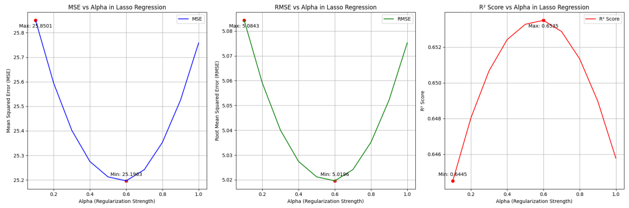


Fig. 12.
MSE, RMSE, R² vs Alpha in Lasso Regression (from 0 to 1.1)

Zooming into a narrower range, the optimal regularization strength is observed around, where MSE and RMSE are minimized and maximized. This demonstrates the importance of fine-tuning to balance bias and variance.

ElasticNet Regression provides valuable insights into the effect of the regularization parameter on model performance. By analyzing the metrics Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the R², we can understand how varying impacts the trade-off between bias and variance. ElasticNet's unique combination of L1 and L2 regularization offers flexibility in balancing feature selection and coefficient shrinkage, making it a versatile tool for regression tasks.

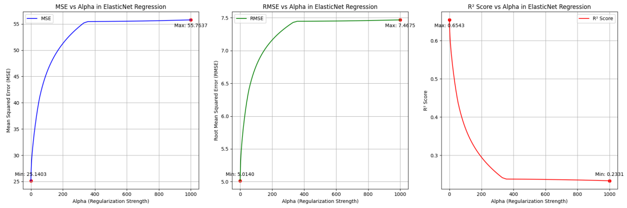


Fig. 13.
MSE, RMSE, R² vs Alpha in EasticNet Regression (from 0 to 1000.1, L1_Ratio at 0.1)

Examining a broad range of highlights the general trends in model performance as regularization strength increases. At lower values of α, the MSE is minimized, indicating effective model fitting with minimal penalization of coefficients. This results in low RMSE values and a high score, showcasing the model's ability to explain variance accurately. However, as increases, MSE and RMSE rise significantly, reflecting the adverse effects of over-regularization. The model begins to underfit, losing its ability to capture the complexity of the data. Concurrently, the score declines sharply, underscoring the diminishing explanatory power of the model at higher levels of α.

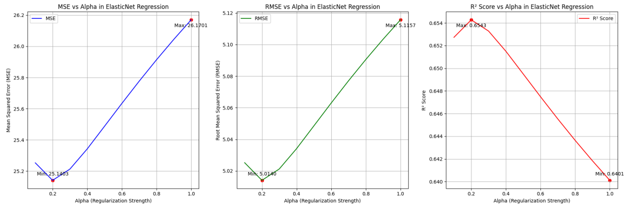


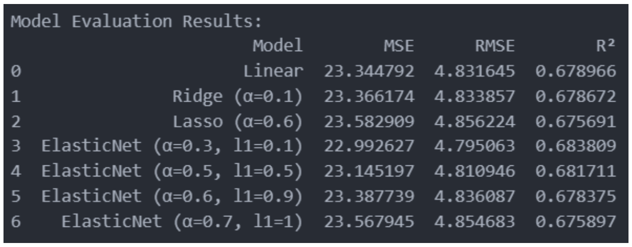
Fig. 14.
MSE, RMSE, R² vs Alpha in EasticNet Regression (from 0 to 1.1, L1_Ratio at 0.1)

Focusing on a narrower range of α enables a more detailed analysis to pinpoint the optimal level of regularization. The analysis reveals an optimal value around α=0.3, where the MSE and RMSE are at their lowest, and the R² score is at its highest. At this point, the model achieves a well-balanced trade-off between bias and variance, ensuring accurate predictions without over-penalizing coefficients. The narrower focus also highlights the sensitivity of ElasticNet Regression to small changes in α, emphasizing the need for precise tuning to achieve optimal results.

Through systematic exploration of the regularization parameter α and, for ElasticNet, the L1-to-L2 scaling ratio, the study reveals the impact of these parameters on key performance metrics such as MSE, RMSE, and R². Ridge Regression demonstrates the necessity of fine-tuning α to achieve optimal model complexity and predictive performance, while Lasso Regression emphasizes the trade-off between sparsity and accuracy. ElasticNet emerges as a highly flexible approach, leveraging both L1 and L2 regularization to balance feature selection and stability, with an optimal α value of around 0.3 for many cases.

D. Model Evaluation

The model evaluation results comprehensively analyze how different regression techniques and hyperparameter settings impact model performance. The evaluation metrics include Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R², each highlighting distinct facets of model effectiveness.



```
Model Evaluation Results:
                       Model        MSE      RMSE        R²
0                     Linear  23.344792  4.831645  0.678966
1              Ridge (α=0.1)  23.366174  4.833857  0.678672
2              Lasso (α=0.6)  23.582909  4.856224  0.675691
3  ElasticNet (α=0.3, l1=0.1)  22.992627  4.795063  0.683809
4  ElasticNet (α=0.5, l1=0.5)  23.145197  4.810946  0.681711
5  ElasticNet (α=0.6, l1=0.9)  23.387739  4.836087  0.678375
6    ElasticNet (α=0.7, l1=1)  23.567945  4.854683  0.675897
```

Fig. 15.
MSE, RMSE, R² Model Evaluation Results

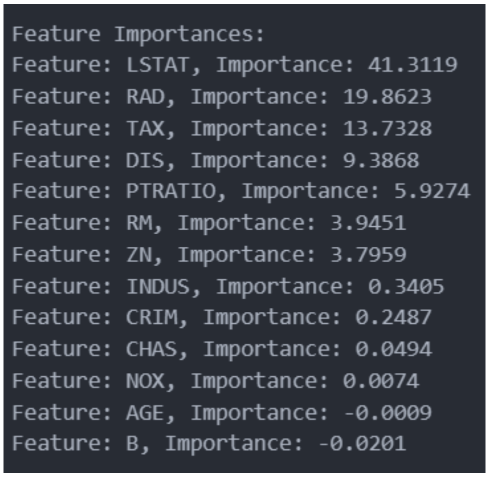
The Linear Regression model serves as the baseline, with an MSE of 23.34, an RMSE of 4.83, and an R² score of 0.678. As expected, the absence of regularization makes this model purely deterministic, fitting the data without penalizing coefficients. However, its performance is slightly limited due to its inability to mitigate the impact of noise or irrelevant features.

The Ridge Regression model with α=0.1\alpha = 0.1α=0.1 yields comparable results to Linear Regression, with a marginally higher MSE and RMSE and a slightly lower R² score. The minimal regularization indicates that small penalty values do not significantly alter the model's predictions but provide slight stability improvements.

The Lasso Regression model with α=0.6 shows a higher MSE (23.58) and RMSE (4.86), along with a lower R² score (0.676). This result highlights the trade-off in Lasso Regression, where higher regularization shrinks some coefficients to zero, potentially discarding important features and leading to reduced predictive accuracy.

ElasticNet models, which combine L1 and L2 penalties, demonstrate the most nuanced performance outcomes. Among the configurations tested, the ElasticNet model with α=0.3 and l1_ratio=0.1 achieves the best overall performance, with the lowest MSE (22.99), RMSE (4.79), and the highest R² score (0.6838). This configuration effectively balances feature selection and coefficient stability, making it the most predictive and robust model. Other ElasticNet configurations, such as α=0.5 and α=0.6, still perform better than Lasso but show diminishing returns as regularization strength increases, moving closer to underfitting.

Using ElasticNet α=0.3 and l1_ratio=0.1, we run the model through feature importance analysis to further emphasize the significance of specific predictors. The most influential features were LSTAT (41.31), RAD (19.86), and TAX (13.73), suggesting their critical role in determining the target variable. Conversely, features such as NOX (0.0074) and AGE (-0.0009) had negligible importance, potentially indicating their limited predictive value in this context.



```
Feature Importances:
Feature: LSTAT, Importance: 41.3119
Feature: RAD, Importance: 19.8623
Feature: TAX, Importance: 13.7328
Feature: DIS, Importance: 9.3868
Feature: PTRATIO, Importance: 5.9274
Feature: RM, Importance: 3.9451
Feature: ZN, Importance: 3.7959
Feature: INDUS, Importance: 0.3405
Feature: CRIM, Importance: 0.2487
Feature: CHAS, Importance: 0.0494
Feature: NOX, Importance: 0.0074
Feature: AGE, Importance: -0.0009
Feature: B, Importance: -0.0201
```

Fig. 16.
Feature Importance Evaluation Results

The provided visualization illustrates the top 10 feature combinations based on their R^2 scores when evaluated using the ElasticNet regression model with α=0.31 1_ratio=0.1. These results reflect a brute-force approach to feature selection, systematically testing various subsets of features to identify those that maximize predictive accuracy.

The best-performing feature subset includes "ZN, INDUS, CHAS, NOX, RM, AGE, DIS, PTRATIO, and LSTAT," achieving the highest R^2 score of 0.6800. This combination demonstrates the importance of these features in explaining the variance of the target variable. The second-best subset, which omits "AGE," achieves a slightly lower R^2 score of 0.6798, indicating that while "AGE" contributes marginally to model accuracy, its exclusion does not drastically impact performance. Notably, the inclusion of features such as "LSTAT" (proportion of the lower status population) and "PTRATIO" (pupil-teacher ratio) in almost all top combinations underscores their strong predictive relevance.

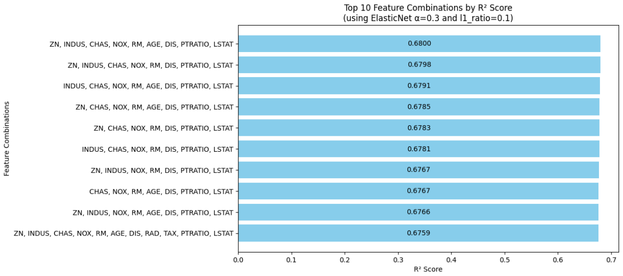


Fig. 17.
Feature Combinations Results

Conversely, the diminishing scores in combinations with fewer features suggest that excluding critical variables leads to underfitting, highlighting the trade-off between model simplicity and accuracy.

E. Analysis and Interpretation

The model evaluation and feature importance analysis in this study reveal critical insights into the effectiveness of various regression algorithms and the significance of feature selection. But it is also important to note that minimal trimming and removal of excessive amounts of outliers was a major challenge as they contribute to a relatively high percentage of the initial data where if they were to be removed, it could drastically make the data overfit the model.

Linear Regression, serving as a baseline, achieved an MSE of 23.34 and an R^2 score of 0.678. This indicates a reasonable ability to explain housing price variance but highlights the model's inability to handle noise and irrelevant features effectively due to the absence of regularization.

Ridge and Lasso regressions introduce regularization to mitigate overfitting. Ridge Regression, using an α=0.1, yielded comparable results to Linear Regression, suggesting minimal impact from slight regularization. Lasso Regression, however, showed a higher MSE of 23.58 and a slightly lower R2R^2R2 of 0.675, likely due to its tendency to shrink some coefficients to zero. This underscores Lasso's suitability for feature selection and its potential to underperform when influential features are overly penalized.

ElasticNet Regression demonstrated superior performance due to its hybrid regularization approach. The model with α=0.3 and l1_ratio=0.1 achieved the best results, with an MSE of 22.99 and an R^2 score of 0.684. This balance between L1 and L2 penalties allowed the model to effectively manage both feature selection and coefficient shrinkage, highlighting its versatility in handling complex datasets.

Feature importance analysis reinforced these findings. LSTAT (percentage of lower-status population), RAD (highway accessibility), and TAX (property tax rate) emerged as the most influential predictors, with importance scores of 41.31, 19.86, and 13.73, respectively. These features exhibit strong relationships with housing prices and are essential for predictive accuracy. In contrast, NOX (nitric oxide concentration) and AGE (proportion of older homes) showed minimal importance, suggesting their limited predictive relevance.

The brute-force feature selection approach validated these observations by systematically testing feature subsets. The best-performing combination—ZN, INDUS, CHAS, NOX, RM, AGE, DIS, PTRATIO, and LSTAT—achieved an R^2 score of 0.680. However, when all features are applied, the R^2 is 0.6838, This highlights the necessity of retaining critical features while avoiding over-reliance on irrelevant variables

The slight improvement in R² when using all features underscores the complex interplay between feature inclusion, model complexity, and overfitting. While the full-feature model may achieve a higher score due to the inclusion of marginally relevant predictors, the best-performing subset provides a more interpretable model, emphasizing critical predictors without overfitting noise or redundant variables. This outcome highlights the importance of feature selection in achieving an optimal balance between bias and variance, ensuring that the model remains interpretable while maintaining strong predictive performance.

IV. CONCLUSIONS

This study examined the application of various regression techniques for predicting Boston housing prices, emphasizing the significance of feature selection and model evaluation in improving predictive accuracy. The findings indicate that certain variables, such as feature importance analysis, highlighted the predictive value of variables such as LSTAT, RAD, and TAX, which significantly influence housing prices. Conversely, features like NOX and AGE had negligible importance, suggesting limited relevance to the target variable. These findings emphasize the need for systematic feature selection to retain influential predictors while excluding irrelevant ones. Implementing a data-driven feature selection process based on feature importance thresholds streamlined the dataset, enhancing both model interpretability and computational efficiency.

ElasticNet Regression emerged as the most effective model, balancing feature selection and stability through its dual regularization approach. The optimal configuration (α=0.3, l1_ratio=0.1) achieved the best predictive performance with an MSE of 22.99 and an R² score of 0.684 (all features) and 0.680 (selected best features) with a 1% trimming process done on both tails of the provided dataset.

However, should more trimming be applied to the data, it could significantly improve the overall performance and accuracy of the model because the data's high number of outliers makes it challenging to justify their removal when they cover a large sum of the total number of entries within the dataset. Room for improvement and discussions can significantly improve this paper and its ability to predict housing prices.

Furthermore, this study highlighted the critical role of data preprocessing techniques in optimizing model performance, such as outlier removal, feature scaling, and the appropriate handling of categorical variables. The exclusion of weakly contributing or redundant features resulted in improved accuracy across all regression models, underscoring the importance of robust feature engineering in predictive modeling.

In conclusion, this study underscores the importance of integrating advanced preprocessing techniques, feature selection strategies, and machine learning models to develop accurate and interpretable predictive frameworks.

V. CITATIONS

GeeksforGeeks. (n.d.). *Hyperparameter tuning*. Retrieved from https://www.geeksforgeeks.org/hyperparameter-tuning/

GeeksforGeeks. (n.d.). *ML | Gradient Boosting*. Retrieved from https://www.geeksforgeeks.org/ml-gradient-boosting/

Kwak, S. K., & Kim, J. H. (2017). Statistical data preparation: Management of missing values and outliers. *Korean Journal of Anesthesiology, 70*(4), 407–411. https://doi.org/10.4097/kjae.2017.70.4.407

Scikit-learn contributors. (n.d.). Supervised learning–Regression (documentation). Retrieved from https://scikit-learn.org/stable/supervised_learning.html