REPORT

Introduction:

Using the California housing dataset, we examine the effectiveness of various regression techniques in this research. The goal is to comprehend how various regression models behave and assess each model's performance using a range of measures.

Regression Algorithms:

1. Linear Regression
2. RANSAC Regression
3. Ridge Regression
4. Lasso Regression
5. ElasticNet Regression
6. Ridge Regression with Polynomial

Dataset:

We made use of the California housing dataset, which includes information on population, average occupancy, median income, median age of housing, average number of rooms and bedrooms, latitude, and longitude. The median home value for California districts is the goal variable.

Methodology:

- loaded the dataset and used an 80-20 split ratio to divide it into training and testing sets.
- To guarantee that every feature has a mean of 0 and a variance of 1, the features were scaled using StandardScaler.
- The regression model was trained using the training set of data.
- evaluated the model's performance using the R-squared (R2) score and Mean Squared Error (MSE) on both the training and testing sets.
- measured the computing efficiency by keeping track of the training time.

Results:

\---

Linear Regression:

Training MSE: 0.5179
Testing MSE: 0.5559
Training R2: 0.6126
Testing R2: 0.5758
Training Time: 0.0049 seconds

RANSAC Regression with Lasso:
Training MSE: 0.9587
Testing MSE: 0.9482
Training R2: 0.2828
Testing R2: 0.2764
Training Time: 0.1631 seconds

Ridge Regression:
Training MSE: 0.5179
Testing MSE: 0.5559
Training R2: 0.6126
Testing R2: 0.5758
Training Time: 0.0078 seconds

Lasso Regression:
Training MSE: 0.5192
Testing MSE: 0.5545
Training R2: 0.6125
Testing R2: 0.5769
Training Time: 0.3835 seconds

ElasticNet Regression:
Training MSE: 0.5192
Testing MSE: 0.5546
Training R2: 0.6125
Testing R2: 0.5768
Training Time: 1.7080 seconds

Non-linear Regression (Ridge Regression with Polynomial Features)
Training MSE: 0.4207
Testing MSE: 0.4641
Training R2: 0.6853
Testing R2: 0.6458
Training Time: 0.0900 seconds

Results in a table:

Regression Algorithm	Training MSE	Testing MSE	Training R2	Testing R2	Training Time (seconds)
Linear Regression	0.5179	0.5559	0.6126	0.5758	0.0049
RANSAC Regression	0.9587	0.9482	0.2828	0.2764	0.1631
Ridge Regression	0.5179	0.5559	0.6126	0.5758	0.0078
Lasso Regression	0.5192	0.5545	0.6125	0.5769	0.3835
ElasticNet Regression	0.5192	0.5546	0.6125	0.5768	1.7080
Non-linear Regression (Ridge Regression with Polynomial Features)	0.4207	0.4641	0.6853	0.6458	0.0167

Regression Algorithm Analysis:

Linear Regression:

Linear Regression is a basic regression algorithm that models the relationship between the independent variables and the target variable. Here are the results and analysis:
Training MSE: 0.5179
Testing MSE: 0.5559
Training R2: 0.6126
Testing R2: 0.5758
Training Time: 0.0049 seconds

Analysis:

Linear Regression performs reasonably well with low MSE and high R2 scores on both training and testing sets. However, there's a slight indication of potential overfitting, as the testing MSE is slightly higher than the training MSE.

RANSACRegressor:

RANSACRegressor is a robust regression algorithm that fits a regression model to a subset of the data, iteratively refining the model by reducing the influence of outliers. Here are the results and analysis:
Training MSE: 0.9587
Testing MSE: 0.9482
Training R2: 0.2828
Testing R2: 0.2764
Training Time: 0.1631 seconds

Analysis: RANSACRegressor shows poorer performance compared to Linear Regression, with higher MSE and lower R2 scores. This algorithm may not be suitable for this dataset due to its inferior performance.

Ridge Regression:

Ridge Regression is a regularized regression algorithm that adds a penalty term to the loss function to prevent overfitting. Here are the results and analysis:
Training MSE: 0.5179
Testing MSE: 0.5559
Training R2: 0.6126
Testing R2: 0.5758
Training Time: 0.0078 seconds
Analysis: Ridge Regression performs similarly to Linear Regression with comparable MSE and R2 scores. It effectively mitigates overfitting and maintains good generalization performance.

Lasso Regression:

Lasso Regression is another regularized regression algorithm that uses L1 regularization to penalize the absolute size of coefficients. Here are the results and analysis:
Training MSE: 0.5192
Testing MSE: 0.5545
Training R2: 0.6125
Testing R2: 0.5769
Training Time: 0.3835 seconds
Analysis: Lasso Regression exhibits performance similar to Ridge Regression, with slightly higher MSE but comparable R2 scores. It effectively reduces overfitting and maintains good generalization performance.

ElasticNet Regression:

ElasticNet Regression is a hybrid of Lasso and Ridge Regression, combining L1 and L2 regularization. Here are the results and analysis:
Training MSE: 0.5192
Testing MSE: 0.5546
Training R2: 0.6125
Testing R2: 0.5768
Training Time: 1.7080 seconds
Analysis: ElasticNet Regression provides similar performance to Lasso Regression, effectively balancing between L1 and L2 regularization to prevent overfitting.

Non-linear Regression (Ridge Regression with Polynomial Features)

Ridge Regression with Polynomial Features introduces non-linear transformations to the features, allowing for capturing non-linear relationships. Here are the results and analysis:
Training MSE: 0.4207
Testing MSE: 0.4641
Training R2: 0.6853
Testing R2: 0.6458
Training Time: 0.0167 seconds

Analysis: Non-linear Regression demonstrates superior performance compared to linear models, with lower MSE and higher R2 scores on both training and testing sets. However, there's a slight indication of potential overfitting, as the testing MSE is higher than the training MSE.

Conclusion:

Ridge Regression with Polynomial Features works better than other regression techniques, according to the analysis, showing better generalization ability with a lower testing MSE and a higher testing R2 score. Ridge Regression with Polynomial Features is the best choice for this dataset, even though Linear Regression also works well. Potential overfitting, however, indicates that more regularization is necessary, especially in non-linear regression. On the other hand, ElasticNet, Lasso, and RANSAC regressions perform worse and might not be the best options. In conclusion, ElasticNet, Lasso, and Linear regressions may show some little overfitting, while Ridge and Non-linear regressions provide the best fits with low MSE and high R2 values. The RANSACRegressor performs worse and might not be the best choice for this dataset.