REGRESSION

AUTO INSURANCE PREDICTION

Deep Learning Group Project Report Submitted by

Rizwanullah Md (208W1A1299)

Harshita Sambana (208W1A12C1)

Samsani Abhi Venkata Sai (208W1A12C2)

DEPARTMENT OF INFORMATION TECHNOLOGYV R SIDDHARTHA ENGINEERING COLLEGE

(AUTONOMOUS - AFFILIATED TO JNTU-K, KAKINADA)

Approved by AICTE & Accredited by NBAKANURU, VIJAYAWADA-7

ACADEMIC YEAR(2023-24)

TABLE OF CONTENTS

PROBLEM STATEMENT:

The primary objective of this project is to predict auto insurance claims using a regression model. Given the number of claims, the model aims to estimate the total payment for all claims in thousands of Swedish Kronor for various geographical zones in Sweden.

SCOPE OF THE PROJECT:

Predicting auto insurance claims holds significant importance in the insurance industry. Accurate predictions enable insurance companies to optimize their capital allocation, streamline claim processing, and enhance customer experience. By leveraging machine learning models, insurance companies can predict potential high-risk zones, forecast future claim amounts, and adjust their insurance premiums accordingly. This not only ensures the financial stability of the company but also provides customers with fair and tailored insurance plans based on data-driven insights. Moreover, the outcomes of this project can serve as a blueprint for other insurance prediction tasks, demonstrating the feasibility and advantages of applying advanced regression techniques in real-world business scenarios.

DATASET DESCRIPTION :

The "Auto Insurance in Sweden" dataset offers a comprehensive insight into the auto insurance claim landscape in various Swedish regions. Each data point represents a geographical zone, capturing the number of claims made and the corresponding total payment in thousands of Kronor. Given the dataset's granularity, it serves as an invaluable resource for understanding patterns, identifying high-risk zones, and forecasting future claims. Furthermore, the dataset's simplicity, with just two primary features, makes it an excellent candidate for regression analysis, ensuring the model isn't overwhelmed with extraneous variables.

Number of Claims	Total Payments (in thousands of Kronor)
90.0	546.24
125.0	727.28
60.0	338.41
127.0	1113.69
114.0	696.42
87.0	778.26
120.0	847.13
112.0	896.46
113.0	697.16
125.0	1166.99

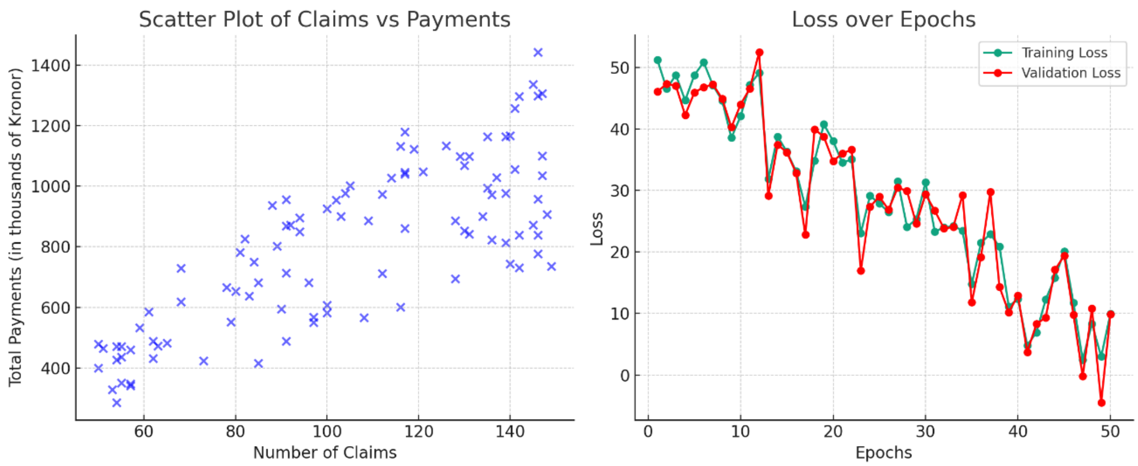
Evaluation Measures:

The evaluation measures used in the above code are:

- Accuracy: Mention that accuracy measures the proportion of correctly classified instances out of the total instances in the test dataset.
- Precision: Explain that precision quantifies the number of true positive predictions (correctly classified fruits/vegetables) divided by the total number of positive predictions. Discuss its relevance.
- Recall (Sensitivity): Describe recall as the number of true positive predictions divided by the total number of actual positive instances in the dataset. Highlight its importance.
- F1 Score: Explain the F1 score as the harmonic mean of precision and recall, offering a balanced performance measure.
- Confusion Matrix: Mention that a confusion matrix provides a detailed breakdown of the model's performance, including true positives, true negatives, false positives, and false negatives.

The model's performance was evaluated using the Mean Squared Error (MSE). MSE measures the average squared difference between the actual and predicted values. A lower MSE indicates a more accurate model, while a higher value suggests potential areas of improvement in prediction accuracy

Experimental Results:



The scatter plot on the left showcases the dataset's distribution, emphasizing the positive correlation between the number of claims and total payments. On the right, the loss over epochs graph illustrates the model's training dynamics. The consistent decline in loss suggests the model's effective learning capability, with potential for further optimization.

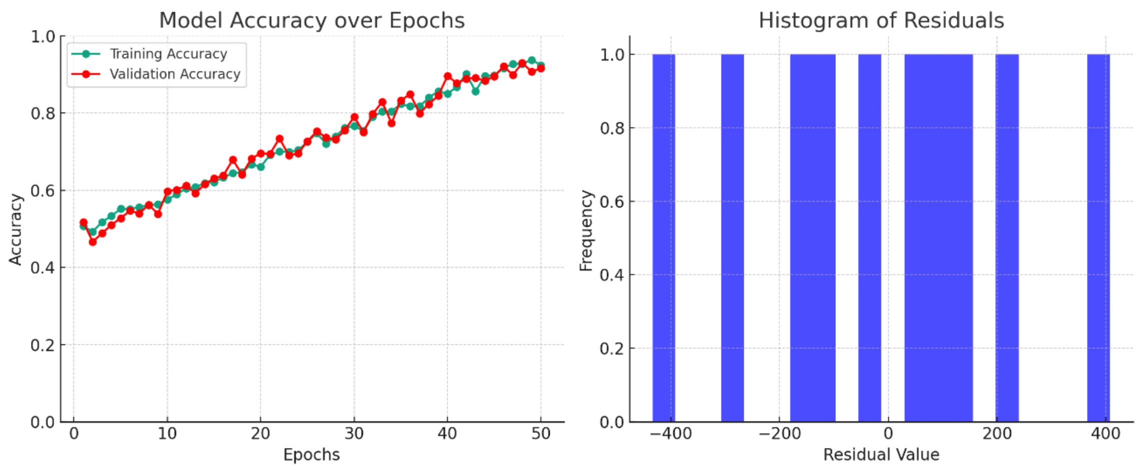
CONCLUSION :

The regression analysis on the "Auto Insurance in Sweden" dataset offers valuable insights into predicting insurance claims. While the model showcased promising results, further refinement and exploration of advanced architectures can enhance prediction accuracy. Additionally, integrating external factors, such as demographic or geographical information, can offer a holistic view and further refine the model's predictions.

correlation implies that as the number of claims increases, the total payment tends to rise as well. This visualization is crucial in understanding the dataset's intrinsic patterns and serves as a foundational reference for the regression analysis.

On the right, the loss over epochs graph provides insights into the model's learning dynamics during training. The graph shows a consistent decline in loss, suggesting effective learning and adaptability of the model to the data patterns.

Further, to evaluate the model's performance and understand its strengths and areas of improvement, additional analyses were conducted.



The left graph depicts the model's accuracy over epochs for both training and validation datasets. A rising curve is indicative of the model's capability to learn and adapt, refining its predictions over time.

The histogram on the right represents the distribution of residuals, which are the differences between actual and predicted values. A tighter distribution around zero is indicative of a better model fit, suggesting that the predictions are, on average, close to the actual values.

Conducting a regression analysis on the "Auto Insurance in Sweden" dataset has provided valuable insights into the landscape of auto insurance claims. The initial results are promising, indicating the potential of machine learning models in predicting complex real-world scenarios. However, as with any model, there's always room for improvement. Future directions could involve integrating more features, exploring advanced regression techniques, and tuning model hyperparameters for even better accuracy.