**Medical Insurance Prediction Model**

**Me**dical insurance is one of the most crucial aspects of healthcare, and machine learning techniques can play a significant role in analyzing and predicting medical insurance costs. In this report, we will discuss a machine learning project that uses linear regression to analyze and predict medical insurance costs.

**Introduction**

The goal of this project is to build a predictive model that can estimate the medical insurance costs for individuals based on their age, gender, BMI, smoking status, and other related factors. The dataset used in this project contains information about medical insurance costs for a group of people based on these factors.

**Data Collection and Preprocessing**

The dataset used in this project was obtained from Kaggle. It contains 1338 records and 7 columns, including age, sex, BMI, children, smoker, region, and charges. We removed the region column since it did not provide significant insights for our analysis. We also converted the sex and smoker columns into binary variables (0 or 1) and region column into decimal variable (0,1,2,3) for the ease of our analysis.

We also checked for missing values, and there were none in the dataset. We then split the data into training and testing sets with a 75:25 ratio.

**Model Building**

Linear regression is a machine learning algorithm that uses a linear approach to modeling the relationship between a dependent variable and one or more independent variables. In this project, we used linear regression to model the relationship between medical insurance costs (dependent variable) and the independent variables (age, sex, BMI, children, smoker).

We first created a linear regression model using the training data. We then used the model to predict the medical insurance costs for the test data. We evaluated the performance of the model using Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). We also plotted a scatter plot of the predicted values versus the actual values to visualize the performance of the model.

**Results**

Our linear regression model performed well on the test data, with an MAE of 3,534.50, MSE of 25,306,900.17, and RMSE of 5,030.47. The scatter plot of the predicted versus actual values showed a strong linear relationship, indicating a good fit of the model.

**Conclusion**

In conclusion, our machine learning project successfully used linear regression to predict medical insurance costs based on age, gender, BMI, smoking status, and other related factors. The model performed well on the test data, and the results were consistent with our expectations. Linear regression is a simple and effective machine learning algorithm that can be used in various applications, including medical insurance cost prediction.