Predicting Insurance premium charged from customer based on their age, BMI and smoking status

I have chosen a dataset with age, sex, BMI, number of children, smoking status, region and premium charged from customers. In this dataset premium charged from customers as our target variable. The model built can be used to predict the insurance premium to be charged from a customer basis the given set of features.

The target variable consists of continuous numerical values and the target variable in our dataset is defined this is a ‘Supervised Regression problem’.

I have built 5 model in total using K- Nearest Neighbours’ Regressor, Linear Regression, Decision Tree Regressor, Lasso Regression, Ridge Regression.

K-NN Regressor using the concepts of similarity to predict the target variable.

Linear Regression uses a linear equation of 1st degree as it’s mapping function to predict target variable.

Lasso and Ridge regularize the gap between train and test performance by adding a term to the cost function.

Decision tree is a greedy algorithm that splits the dataset which gives minimum MSE for the split at each depth and uses the leaf node to make predictions.

I have used Mean Absolute Percentage Error as the metric for my model performance.

I am also attaching the Jupyter notebook with the python code demo to build my models.

Thank You,

Ksheetij Raj Sinha