Machine Learning Mini Project

HEALTH INSURANCE PREMIUM PREDICTION

- Rishi Raj
- Sahaj Gupta
- Vishal Kumar Prajapati
- Yash Agrawal

Objective

To give people an estimate of how much they need based on their individual health situation. After that, customers can work with any health insurance carrier and its plans and perks while keeping the projected cost from our study in mind. I am considering variables as age, sex, BMI, number of children, smoking habits and living region to predict the premium.

Dataset

For training and testing the model, I used the public data set available in Kaggle, “Insurance Premium Prediction”

URL: https://www.kaggle.com/noordeen/insurance-premium-prediction

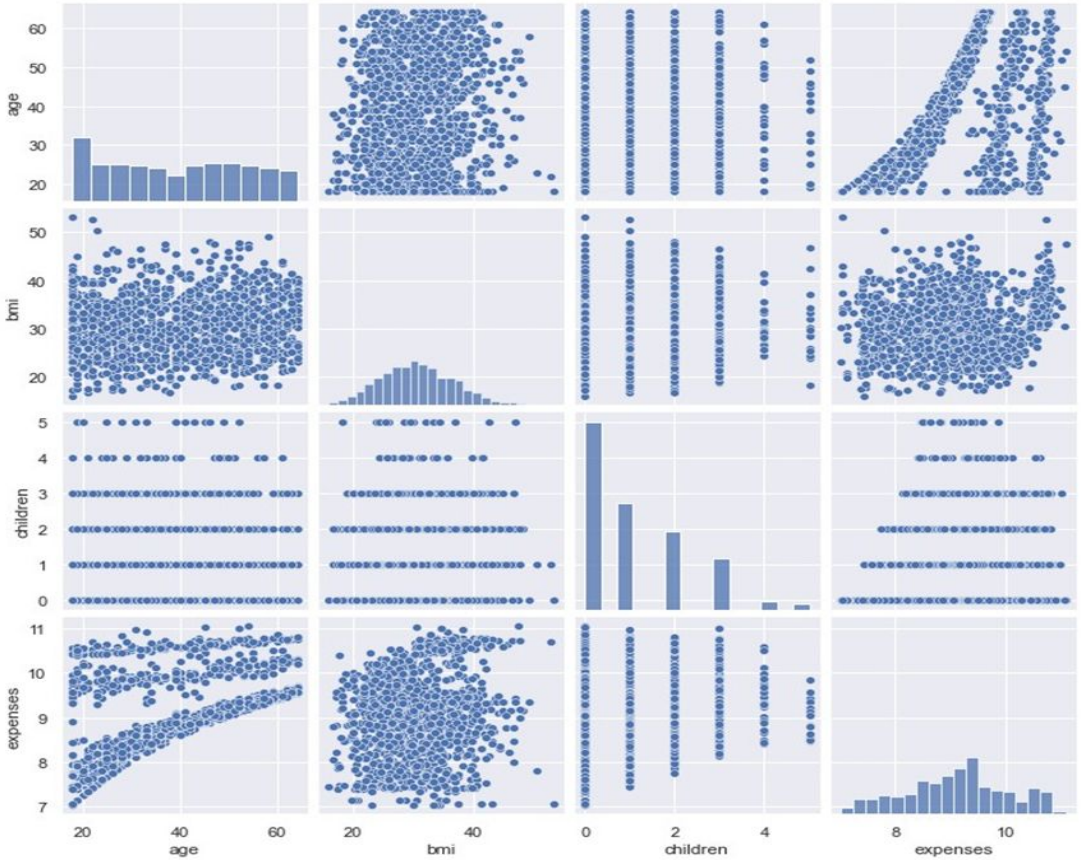
Dataset cont.

- The given data set has 7 features and each one is quantitative
- in nature.
- This data set doesn’t have any missing values.
- The shape of the data set is 1718 rows × 7 columns.

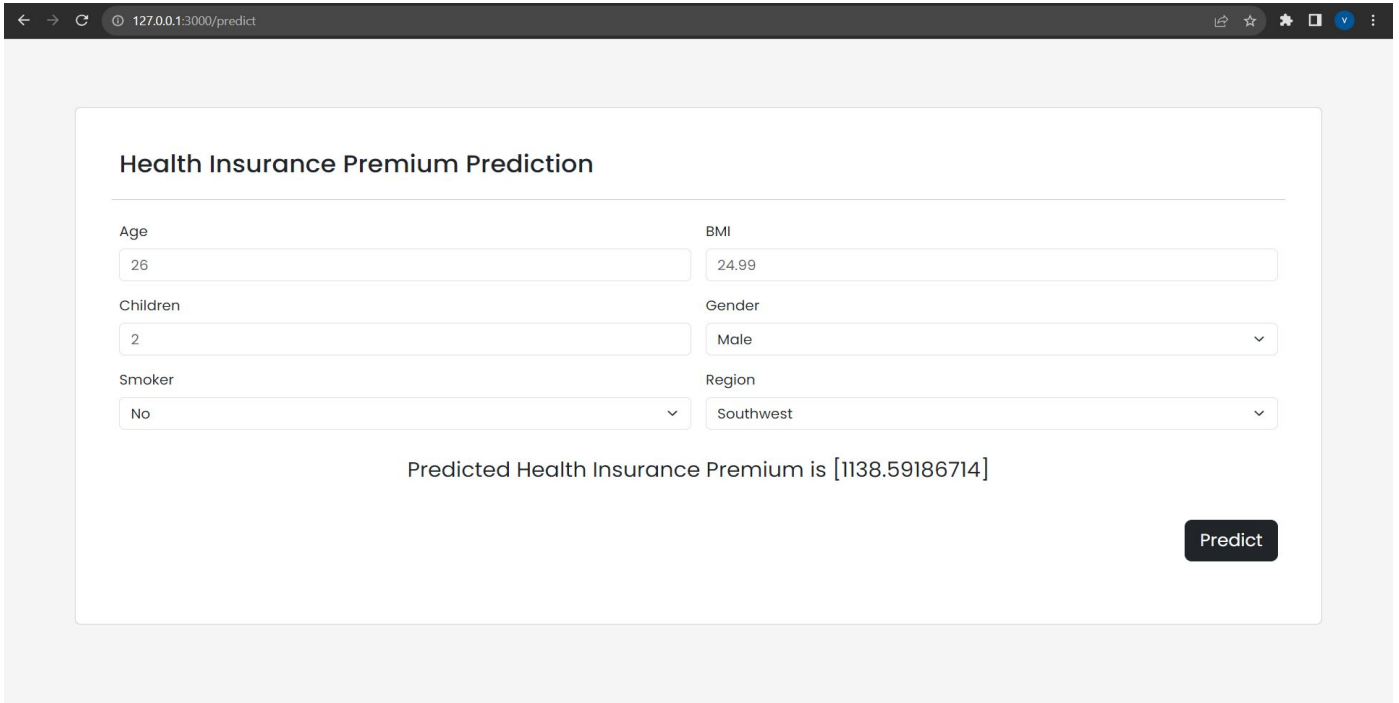
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1718 entries, 0 to 1717
Data columns (total 7 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   age       1718 non-null   int64
 1   sex       1718 non-null   object
 2   bmi       1718 non-null   float64
 3   children  1718 non-null   object
 4   smoker    1718 non-null   object
 5   region    1718 non-null   object
 6   charges   1718 non-null   float64
dtypes: float64(2), int64(1), object(4)
memory usage: 94.1+ KB
```

Data Visualisation

- Age values are constrained between 18-64 years.
- BMI values varies from 16 to 53.1
- Age & BMI have spread data points.
- With age expenses are also increasing following linear positive relationship.
- With number of children expenses also follow linear positive relationship



Web App


127.0.0.1:3000/predict
Health Insurance Premium Prediction
Age
26
BMI
24.99
Children
2
Gender
Male
Smoker
No
Region
Southwest
Predicted Health Insurance Premium is [1138.59186714]
Predict

Conclusion

Models & Accuracy

- Liner Linear Regression : 0.73
- Polynomial Regression : 0.77
- Linear Regression using Lasso : 0.73
- Linear Regression using Ridge : 0.73
- Decision Tree Regression : 0.71
- MLP : 0.79
- KNN : 0.76
- SVM : 0.79

Thank You 🥲