



Insurance Charges Prediction

Machine Learning-Regression

Sowmiya R A

Assignment

Client's requirement is to predict insurance charges based on parameters such as age,sex,BMI,smoker or not, number of children using Artificial Intelligence.

Basic Findings

1. **Domain Selection:** Categorized it under Machine Learning (as they deal with numerical values)
2. **Learning Selection:** Supervised Learning(as our input and output are known)
3. **Algorithm Selection:** Regression Algorithm(as output is a numerical value).

Let us check with the following algorithms to find the best model:

- i)Multiple Linear Algorithm
- ii)Support vector Machine
- iii)Decision Tree
- iv)Random Forest

Specifications

The provided dataset has five input columns(age,sex,BMI, number of children and smoking status) and one output column(charges). It has 1338 rows of data to find the best model. The data need to be preprocessed as it has two columns with categorical data which is to be converted into nominal data. Both the columns are of ordinal types. So Label Encoder should be used for conversion.

Machine Learning Algorithm

This algorithm provides the R score value of **0.78**

Support Vector Machine

Let us find the best model from the following hyper tuning parameters:

C value	linear	rbf	poly	sigmoid
10	0.46	-0.03	0.038	0.039
50	0.60	0.14	0.41	0.40
100	0.62	0.32	0.61	0.52
200	0.63	0.47	0.75	0.54
500	0.76	0.66	0.82	0.44
1000	0.76	0.81	0.85	0.28
2000	0.74	0.85	0.86	-0.59
3000	0.74	0.86	0.85	-2.12

The best Model from this algorithm has an r score value of **0.86**

Decision Tree

Let us find the best model from this algorithm by using the following hyper tuning parameters:

criterion	max_features	splitter	R value
mse	auto	best	0.69
mse	auto	random	0.66
mse	sqrt	best	0.72
mse	sqrt	random	0.66
mse	log2	best	0.78
mse	log2	random	0.60
mae	auto	best	0.67
mae	auto	random	0.69
mae	sqrt	best	0.71
mae	sqrt	random	0.73

mae	log2	best	0.65
mae	log2	random	0.72
friedman_mse	auto	best	0.68
friedman_mse	auto	random	0.67
friedman_mse	sqrt	best	0.70
friedman_mse	sqrt	random	0.63
friedman_mse	log2	best	0.71
friedman_mse	log2	random	0.72

The best model in this algorithm gives an r score value of **0.78**

Random Forest

Let us find the best model from this algorithm by using the following hyper tuning parameters:

critierion	splitter	n=10	n=50	n=100
mse	auto	0.83	0.84	0.85
mse	sqrt	0.85	0.87	0.87
mse	log2	0.85	0.87	0.87
mae	auto	0.83	0.85	0.85
mae	sqrt	0.85	0.87	0.87
mae	log2	0.85	0.87	0.87

The best model in this algorithm gives an r score value of **0.87**

Conclusion:

By comparing the above algorithms **Random Forest** provides the best model for the given dataset with an **r score value of 0.87**. Hence the model is **finalized** for the **prediction** of charges.