

MEDICAL INSURANCE PREDICTION

Introduction

Overview

The objective of proposed work is to consider the effects of smoking, BMI, gender and region to determine how much these factors can account for our increase or decrease in insurance premium. An illustration of how healthcare providers can take advantage of machine learning is being able to predict hospital re-admission for chronically ill patients.

Linear Regression is used in this project to understand how much smoking increases their premium by predicting how much customers will have to pay within seconds.

Machine learning is when a computer has been taught to recognize patterns by providing it with data and an algorithm to help understand that data

LITERATURE SURVEY

Existing methods:

The company have their parameters and charges which are made up of collected data and their charges vary from one company to other. Here they use clients daily habits and other medical records to calculate charges and other parameters, but many of them don't have any solution or integrated software which can predict the charges.

Proposed Solution:

Logistic regression is used as outcome always depend on sum of inputs and parameters.

Solution that I suggest for the project is Linear Regression as it predict the charges for given input parameters. This will not only help the companies but also the clients to determine the charges for existing parameters.

Hardware/Software designing

Software requirements :

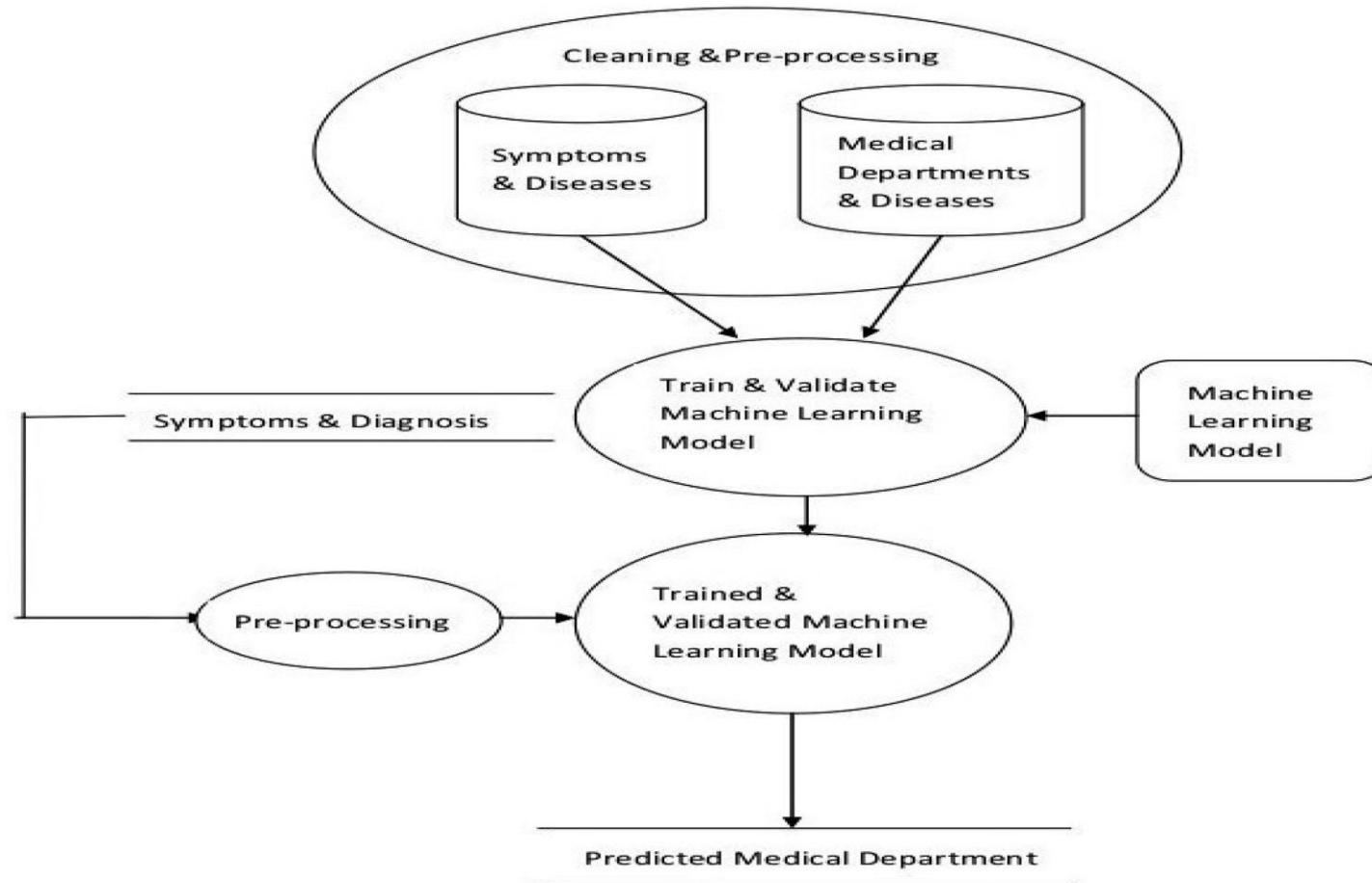
- Java JDK10
- Eclipse IDE

Hardware requirements :

- Processor – i3,i5
- Min Hard Disk – 4 GB
- Min Memory – 4GB RAM

THEORETICAL ANALYSIS

BLOCK DIAGRAM FOR MEDICAL INSURANCE PREDICTION



EXPERIMENTAL INVESTIGATIONS

Analysis is using in this project to make sure which method is best to know or predict the medical insurance.

RESULT

java/org/ml/Medical1.java - Eclipse IDE

Search Project Run Window Help

```
1 package org.ml;
2
3 import java.io.IOException;
4 import java.util.Arrays;
5
6 import tech.tablesaw.api.Table;
7 import tech.tablesaw.plotly.Plot;
8 import tech.tablesaw.plotly.components.Figure;
9 import tech.tablesaw.plotly.components.Layout;
10 import tech.tablesaw.plotly.traces.BoxTrace;
11 //import tech.tablesaw.plotly.traces.HistogramTrace;
12 import tech.tablesaw.plotly.traces.ScatterTrace;
13 import weka.classifiers.Classifier;
14 import weka.classifiers.Evaluation;
15 import weka.core.Instance;
16 import weka.core.Instances;
17 import weka.core.converters.ConverterUtils.DataSource;
18
19 public class Medical1 {
20
21     public static Instances getInstances (String filename)
22     {
23
24         DataSource source;
25         Instances dataset = null;
26         try {
27             source = new DataSource(filename);
28             dataset = source.getDataSet();
29             dataset.setClassIndex(dataset.numAttributes()-1);
30
31         } catch (Exception e) {
32             // TODO Auto-generated catch block
33             e.printStackTrace();
34
35         }
36
37         return dataset;
38     }
39 }
```

Problems @ Javadoc Declaration Console Internal Web Browser

<terminated> Medical1 [Java Application] C:\Users\hp\.p2\pool\plugins\org.eclipse.justi.openjdk.hotspot.jre.full

Structure of dataset2.csv

Index	Column Name	Column Type
0	age	INTEGER
1	bmi	DOUBLE
2	children	INTEGER
3	smoker=no	INTEGER
4	charges	DOUBLE

dataset2.csv

Summary	age	bmi	children
Count	1338	1338	1338
sum	52459	41027.624999999985	1466
Mean	39.20702541106125	30.663396860986524	1.0956651718983534
Min	18	15.96	0
Max	64	53.13	5
Range	46	37.17	5
Variance	197.40138665754355	37.18788360977321	1.4523222572902084
Std. Dev	14.049960379216147	6.098186911679012	1.2051233369619097

669

** Logistic Regression Evaluation with Datasets **

Correctly Classified Instances	669	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0		
Root mean squared error	0		
Relative absolute error	0	%	

ra/org.ml/Medical1.java - Eclipse IDE

Search Project Run Window Help



org.ml/pom.xml Medical1.java

```
1 package org.ml;
2
3 import java.io.IOException;
4 import java.util.Arrays;
5
6 import tech.tablesaw.api.Table;
7 import tech.tablesaw.plotly.Plot;
8 import tech.tablesaw.plotly.components.Figure;
9 import tech.tablesaw.plotly.components.Layout;
10 import tech.tablesaw.plotly.traces.BoxTrace;
11 //import tech.tablesaw.plotly.traces.HistogramTrace;
12 import tech.tablesaw.plotly.traces.ScatterTrace;
13 import weka.classifiers.Classifier;
14 import weka.classifiers.Evaluation;
15 import weka.core.Instance;
16 import weka.core.Instances;
17 import weka.core.converters.ConverterUtils.DataSource;
18
19 public class Medical1 {
20
21     public static Instances getInstances (String filename)
22     {
23
24         DataSource source;
25         Instances dataset = null;
26         try {
27             source = new DataSource(filename);
28             dataset = source.getDataSet();
29             dataset.setClassIndex(dataset.numAttributes()-1);
30
31         } catch (Exception e) {
32             // TODO Auto-generated catch block
33             e.printStackTrace();
34
35         }
36
37         return dataset;
38     }
39 }
```

Problems Javadoc Declaration Console Internal Web Browser

<terminated> Medical1 [Java Application] C:\Users\hp\.p2\pool\plugins\org.eclipse.justj.openjdk.hotspot.jre.full

children	7.9136
smoker=no	114.6069
charges	-0.0003
Intercept	-110.958

Odds Ratios...

Variable	Class
=====	
age	2.2208
bmi	1.2654
children	2734.0832
smoker=no	5.931202965433636E49
charges	0.9997

Confusion matrix:

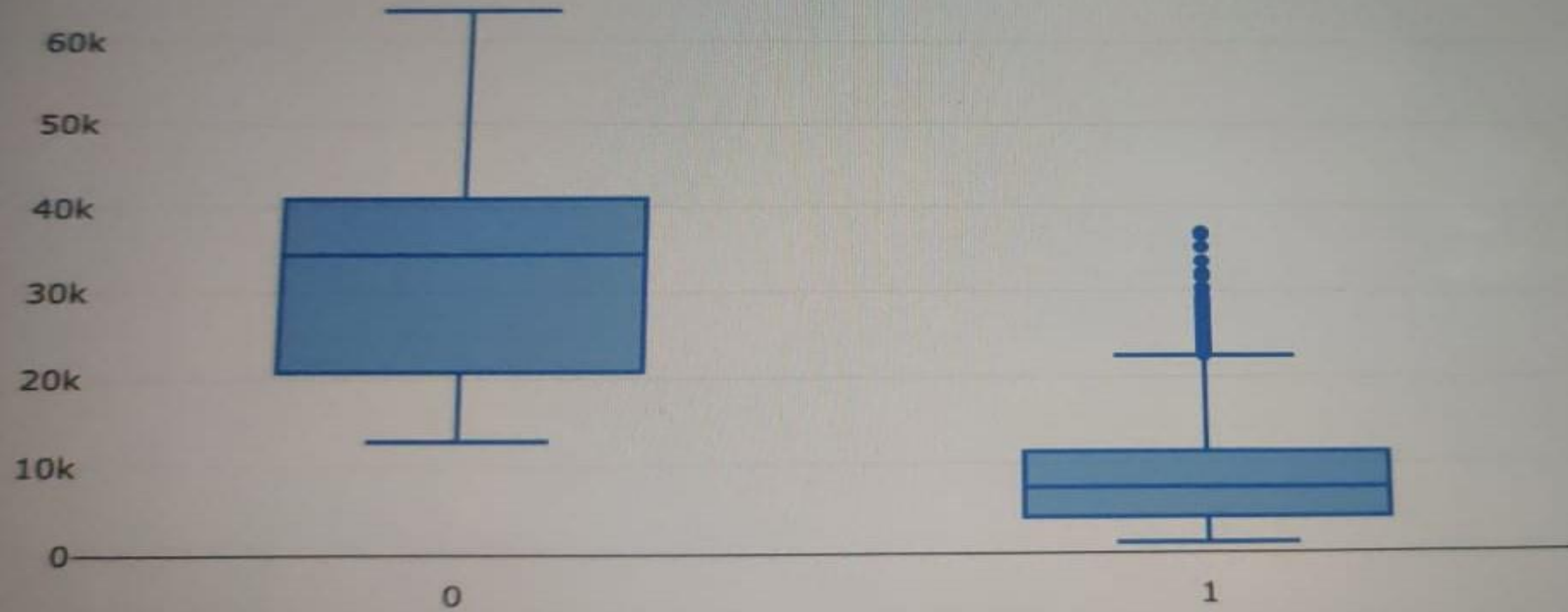
[535.0, 0.0]
[0.0, 134.0]

Area under the curve

1.0

[Correct, Incorrect, Kappa, Total cost, Average cost, KB relative, KB information,
Recall :1.0
Precision:1.0
F1 score:1.0
Accuracy:1.0

age distribution



Advantages and Disadvantages :

Advantages :

- Linear regression performs exceptionally well for linearly separable data.
- When we know the relationship between the independent and dependent variable have a linear relationship, this 'Linear Regression' is the best to use because of its less complexity compared to the other algorithms.
- Logistic regression is easier to implement, interpret, and very efficient to train.

Disadvantages :

- The assumption of linearity between dependent and independent variables
- Linear regression technique outliers can have huge effects on the regression and boundaries are linear in this technique.

CONCLUSION

In this project, we have explored the linear regression model and applied it to various parameters. By using this model, the customers can understand how much smoking increases their premium by predicting how much they will have to pay within seconds.

References:

Dataset for Medical Insurance Prediction

<https://www.kaggle.com/mirichoi0218/insurance>