Medical Insurance Prediction

INTRODUCTION:

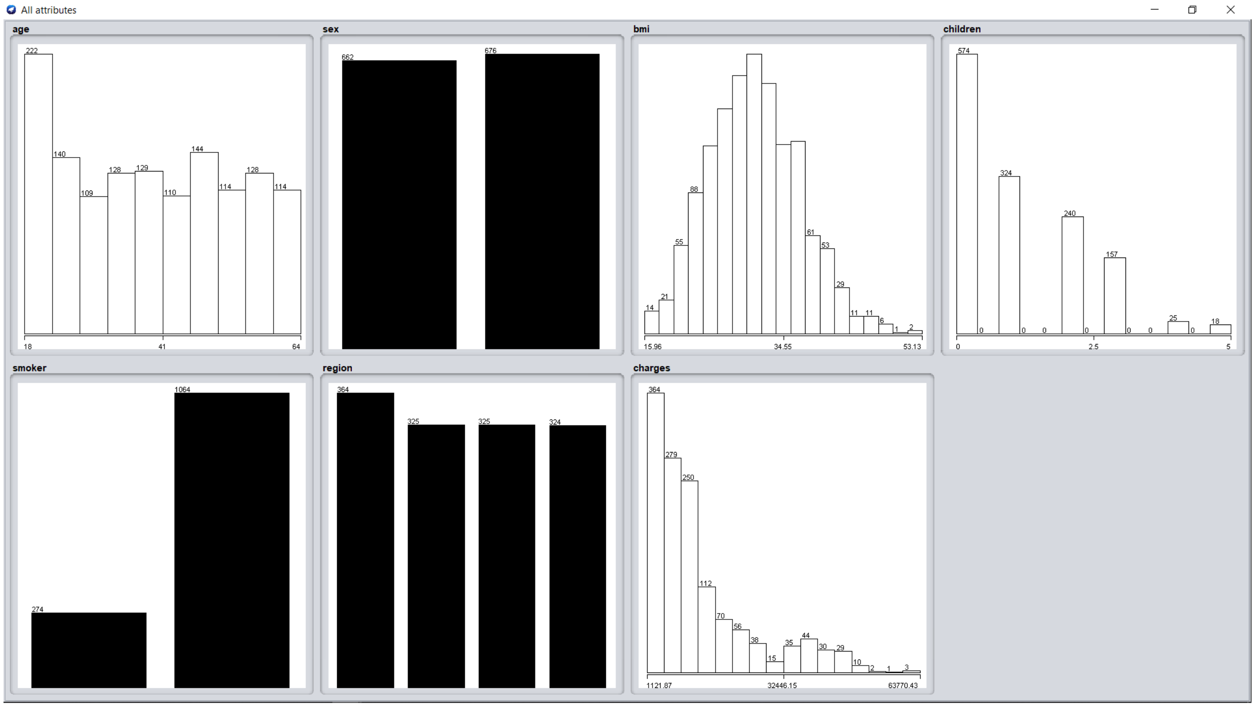
Overview:

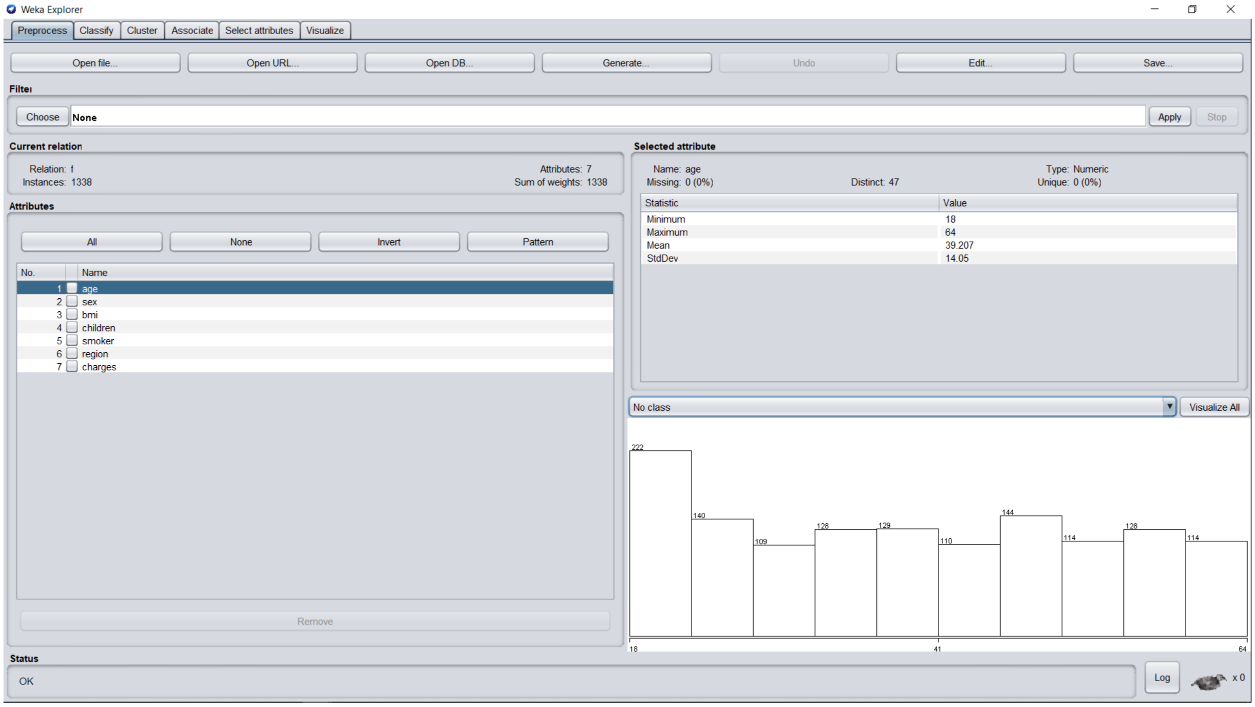
A health insurance company can only make money if it collects more than it spends on the medical care of its beneficiaries. On the other hand, even though some conditions are more prevalent for certain segments of the population, medical costs are difficult to predict since most money comes from rare conditions of the patients.

Purpose:

to predict insurance costs based on people's data, including age, Body Mass Index, smoking or not, etc.

A health insurance company can only make money if it collects more than it spends on the medical care of its beneficiaries. On the other hand, even though some conditions are more prevalent for certain segments of the population, medical costs are difficult to predict since most money comes from rare conditions of the patients. The objective of this article is to accurately predict insurance costs based on people’s data, including age, Body Mass Index, smoking or not, etc. Additionally, we will also determine what the most important variable influencing insurance costs is. These estimates could be used to create actuarial tables that set the price of yearly premiums higher or lower according to the expected treatment costs. This is a regression problem.





Filteı

None

Current relation

Relation: f
Instances: 1338

Attributes: 7
Sum of weights: 1338

Attributes

No.		Name
1	□	age
2	□	sex
3	□	bmi
4	□	children
5	□	smoker
6	□	region
7	□	charges

Selected attribute

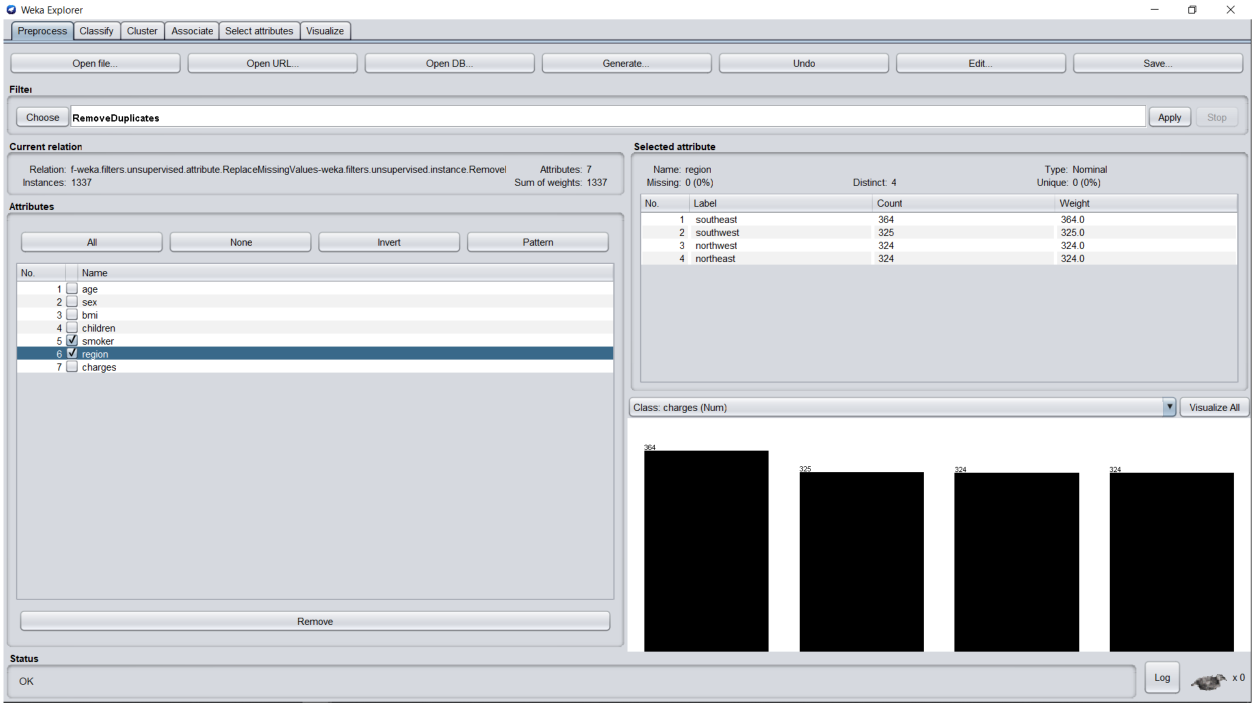
Name: age
Missing: 0 (0%)
Distinct: 47
Type: Numeric
Unique: 0 (0%)

Statistic	Value
Minimum	18
Maximum	64
Mean	39.207
StdDev	14.05

No class

Status

OK



Filter: RemoveDuplicates

Current relation

Relation: f-weka.filters.unsupervised.attribute.ReplaceMissingValues-weka.filters.unsupervised.instance.Removel Attributes: 7
Instances: 1337 Sum of weights: 1337

Attributes

No.		Name
1	□	age
2	□	sex
3	□	bmi
4	□	children
5	☑	smoker
6	☑	region
7	□	charges

Selected attribute

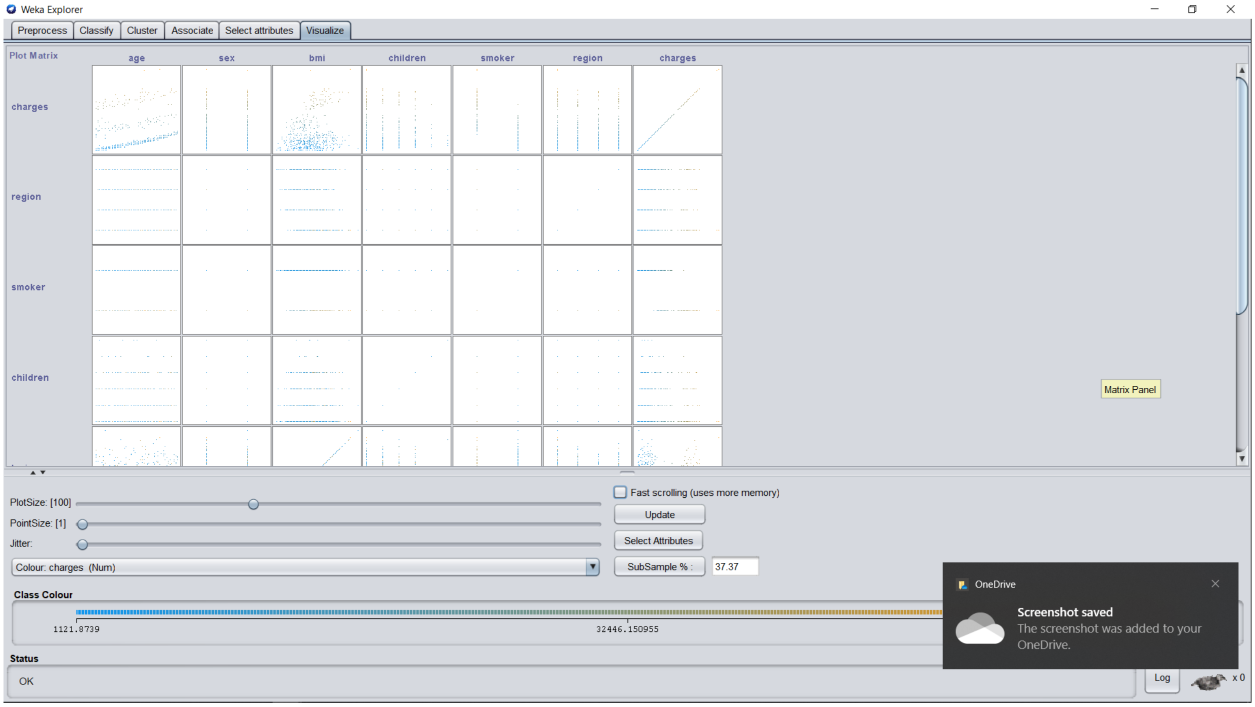
Name: region Type: Nominal
Missing: 0 (0%) Distinct: 4 Unique: 0 (0%)

No.	Label	Count	Weight
1	southeast	364	364.0
2	southwest	325	325.0
3	northwest	324	324.0
4	northeast	324	324.0

Class: charges (Num)

Status

OK



PlotSize: [100]

PointSize: [1]

Jitter:

□ Fast scrolling (uses more memory)

SubSample % : 37.37

Colour: charges (Num)

Class Colour

1121.8739 32446.150955

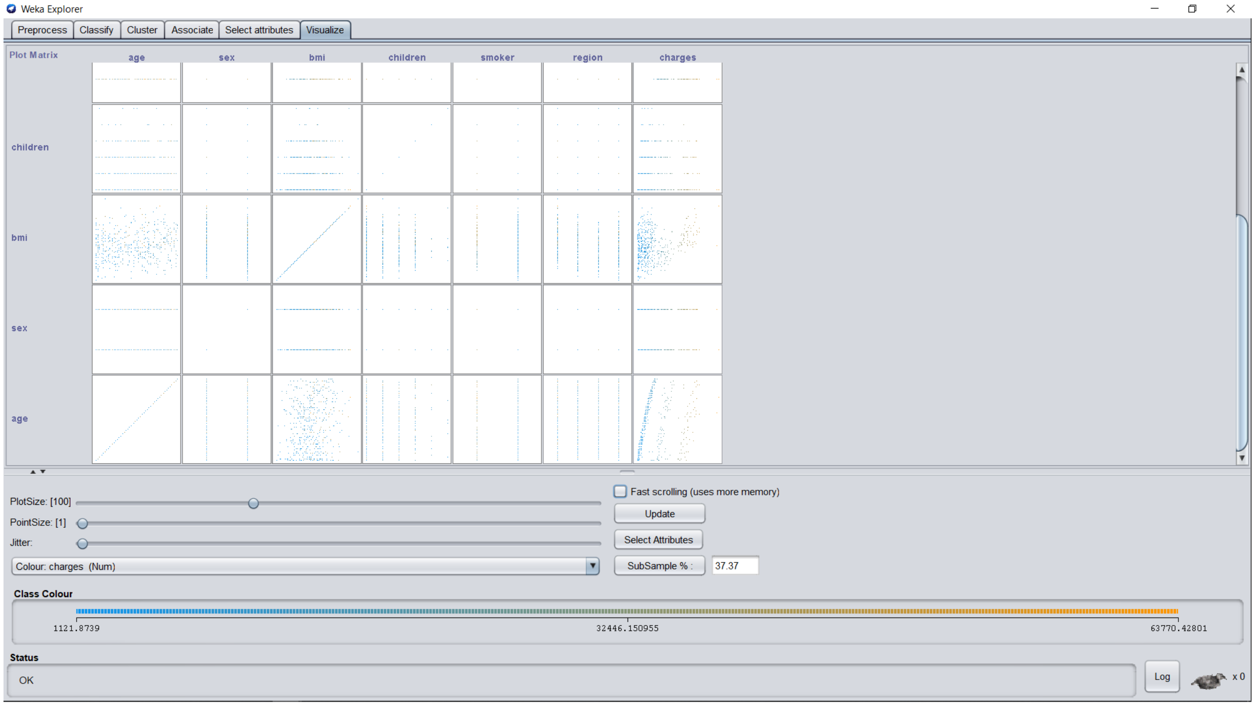
Status

OK

OneDrive

Screenshot saved

The screenshot was added to your OneDrive.



PlotSize: [100]

PointSize: [1]

Jitter:

Colour: charges (Num)

☐ Fast scrolling (uses more memory)

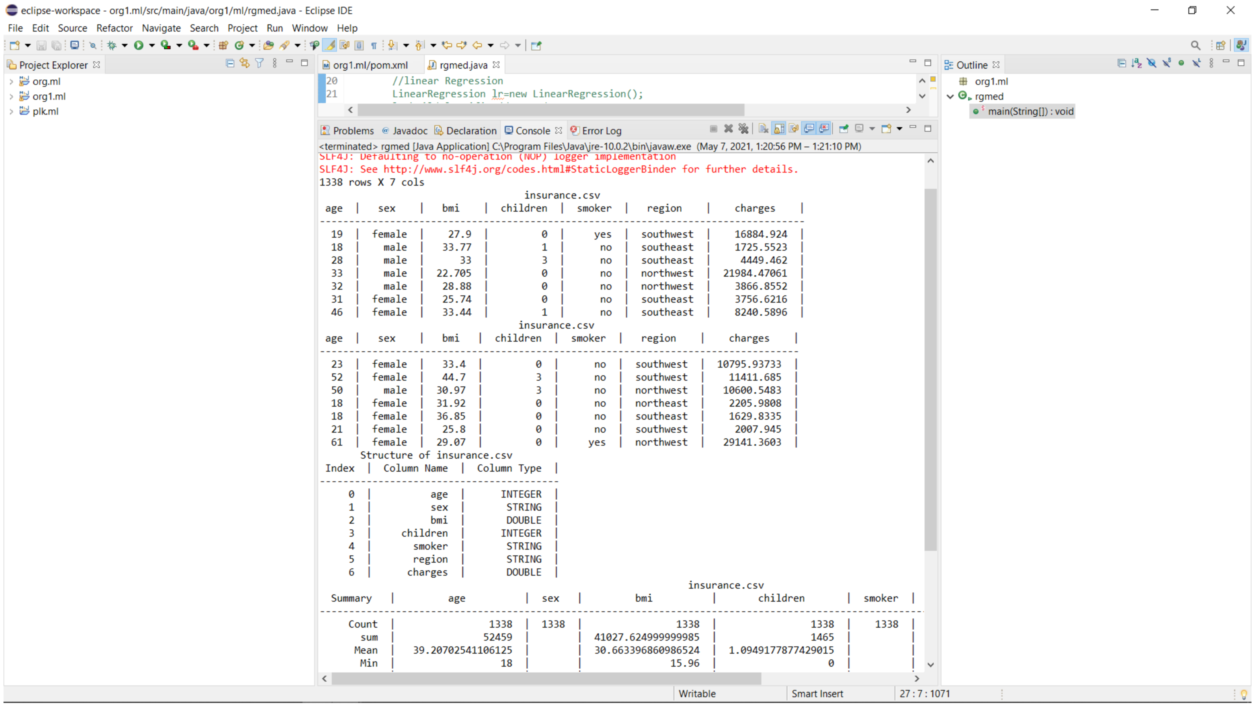
SubSample % : 37.37

Class Colour

1121.8739 32446.150955 63770.42801

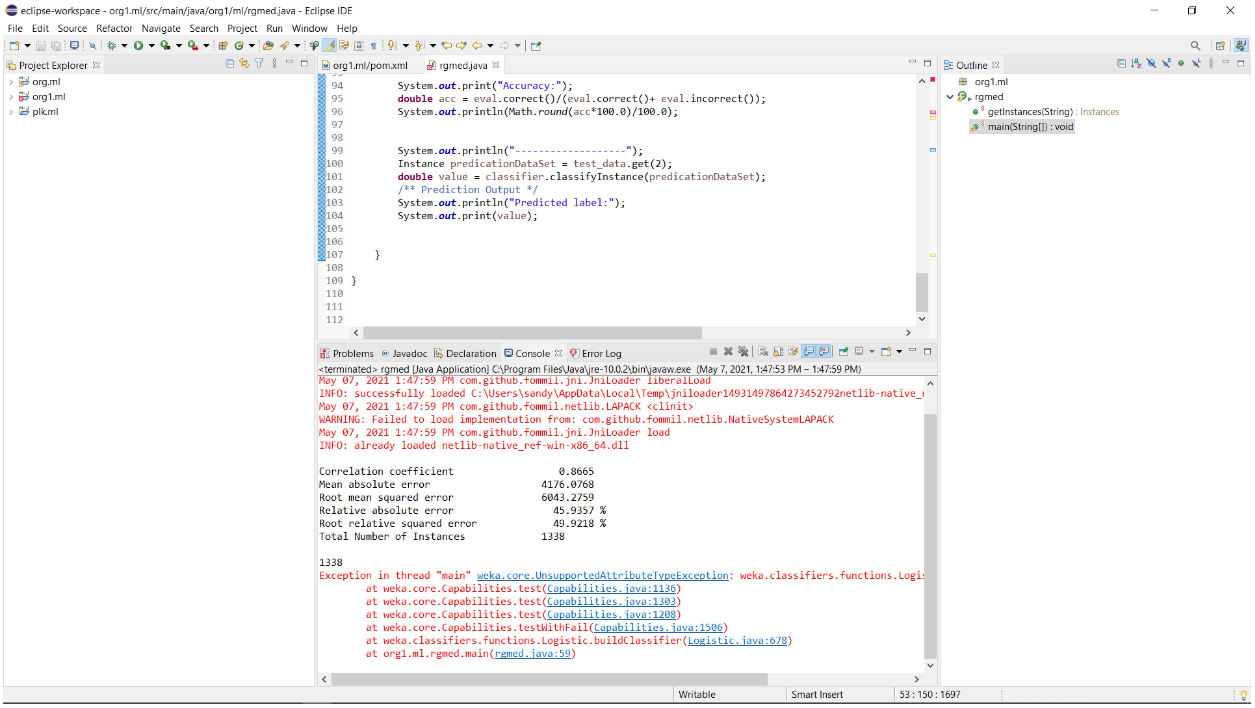
Status

OK



```
        //linear Regression
        LinearRegression lr=new LinearRegression();
```

```
<terminated> rgmed [Java Application] C:\Program Files\Java\jre-10.0.2\bin\javaw.exe  (May 7, 2021, 1:20:56 PM – 1:21:10 PM)
SLF4J: Defaulting to no-operation (NOP) logger implementation
SLF4J: See http://www.slf4j.org/codes.html#StaticLoggerBinder for further details.
1338 rows X 7 cols
                      insurance.csv
 age  |   sex    |  bmi   | children |  smoker  |   region    |   charges    |
-------------------------------------------------------------------------------
  19  |  female  |  27.9  |        0 |      yes |  southwest  |   16884.924  |
  18  |    male  |  33.77 |        1 |       no |  southeast  |   1725.5523  |
  28  |    male  |     33 |        3 |       no |  southeast  |    4449.462  |
  33  |    male  | 22.705 |        0 |       no |  northwest  | 21984.47061  |
  32  |    male  |  28.88 |        0 |       no |  northwest  |   3866.8552  |
  31  |  female  |  25.74 |        0 |       no |  southeast  |   3756.6216  |
  46  |  female  |  33.44 |        1 |       no |  southeast  |   8240.5896  |
                     insurance.csv
 age  |   sex    |  bmi  | children |  smoker  |   region    |   charges    |
------------------------------------------------------------------------------
  23  |  female  |  33.4 |        0 |       no |  southwest  | 10795.93733  |
  52  |  female  |  44.7 |        3 |       no |  southwest  |   11411.685  |
  50  |    male  | 30.97 |        3 |       no |  northwest  |  10600.5483  |
  18  |  female  | 31.92 |        0 |       no |  northeast  |   2205.9808  |
  18  |  female  | 36.85 |        0 |       no |  southeast  |   1629.8335  |
  21  |  female  |  25.8 |        0 |       no |  southwest  |    2007.945  |
  61  |  female  | 29.07 |        0 |      yes |  northwest  |  29141.3603  |
      Structure of insurance.csv
 Index  |  Column Name  |  Column Type  |
-----------------------------------------
     0  |          age  |      INTEGER  |
     1  |          sex  |       STRING  |
     2  |          bmi  |       DOUBLE  |
     3  |     children  |      INTEGER  |
     4  |       smoker  |       STRING  |
     5  |       region  |       STRING  |
     6  |      charges  |       DOUBLE  |
                                                          insurance.csv
  Summary  |        age         |  sex  |        bmi         |      children      |  smoker  |
----------------------------------------------------------------------------------------------
    Count  |              1338  |  1338 |              1338  |              1338  |    1338  |
      sum  |             52459  |       | 41027.624999999985 |              1465  |          |
     Mean  |  39.20702541106125 |       | 30.663396860986524 | 1.0949177877429015 |          |
      Min  |                18  |       |             15.96  |                 0  |          |
```



```java
        System.out.print("Accuracy:");
        double acc = eval.correct()/(eval.correct()+ eval.incorrect());
        System.out.println(Math.round(acc*100.0)/100.0);


        System.out.println("------------------");
        Instance predicationDataSet = test_data.get(2);
        double value = classifier.classifyInstance(predicationDataSet);
        /** Prediction Output */
        System.out.println("Predicted label:");
        System.out.print(value);


    }

}
```

```
<terminated> rgmed [Java Application] C:\Program Files\Java\jre-10.0.2\bin\javaw.exe  (May 7, 2021, 1:47:53 PM – 1:47:59 PM)
May 07, 2021 1:47:59 PM com.github.fommil.jni.JniLoader liberalLoad
INFO: successfully loaded C:\Users\sandy\AppData\Local\Temp\jniloader14931497864273452792netlib-native_r
May 07, 2021 1:47:59 PM com.github.fommil.netlib.LAPACK <clinit>
WARNING: Failed to load implementation from: com.github.fommil.netlib.NativeSystemLAPACK
May 07, 2021 1:47:59 PM com.github.fommil.jni.JniLoader load
INFO: already loaded netlib-native_ref-win-x86_64.dll

Correlation coefficient                  0.8665
Mean absolute error                   4176.0768
Root mean squared error               6043.2759
Relative absolute error                 45.9357 %
Root relative squared error             49.9218 %
Total Number of Instances             1338

1338
Exception in thread "main" weka.core.UnsupportedAttributeTypeException: weka.classifiers.functions.Logi
        at weka.core.Capabilities.test(Capabilities.java:1136)
        at weka.core.Capabilities.test(Capabilities.java:1303)
        at weka.core.Capabilities.test(Capabilities.java:1208)
        at weka.core.Capabilities.testWithFail(Capabilities.java:1506)
        at weka.classifiers.functions.Logistic.buildClassifier(Logistic.java:678)
        at org1.ml.rgmed.main(rgmed.java:59)
```