

Cycle Sheet-I

Prashanth.S 19MID0020

I) Determine, How many instances and attributes are contained in the data-set???

```
print("Number of attributes : ",len(df.columns))
print("Number of instances : ",len(df))
```

Number of attributes : 9
Number of instances : 768

Current relation

Relation: wheat seeds	Attributes: 8
Instances: 210	Sum of weights: 210

2) Give the class label and number of instances in each class. Which class is dominant in the dataset?

```
classes = df['Outcome'].nunique()
print("Number of classes in the data-set : ",classes)
print("\n")

for i in range(classes):
    print("Number of instances in Class-{} : {}".format(i,len(df[df['Outcome'] == i])))
```

Number of classes in the data-set : 2

Number of instances in Class-0 : 500
Number of instances in Class-1 : 268

Class-0 is the dominant class.

CYCLE SHEET-I

Prashanth.S 19MID0020

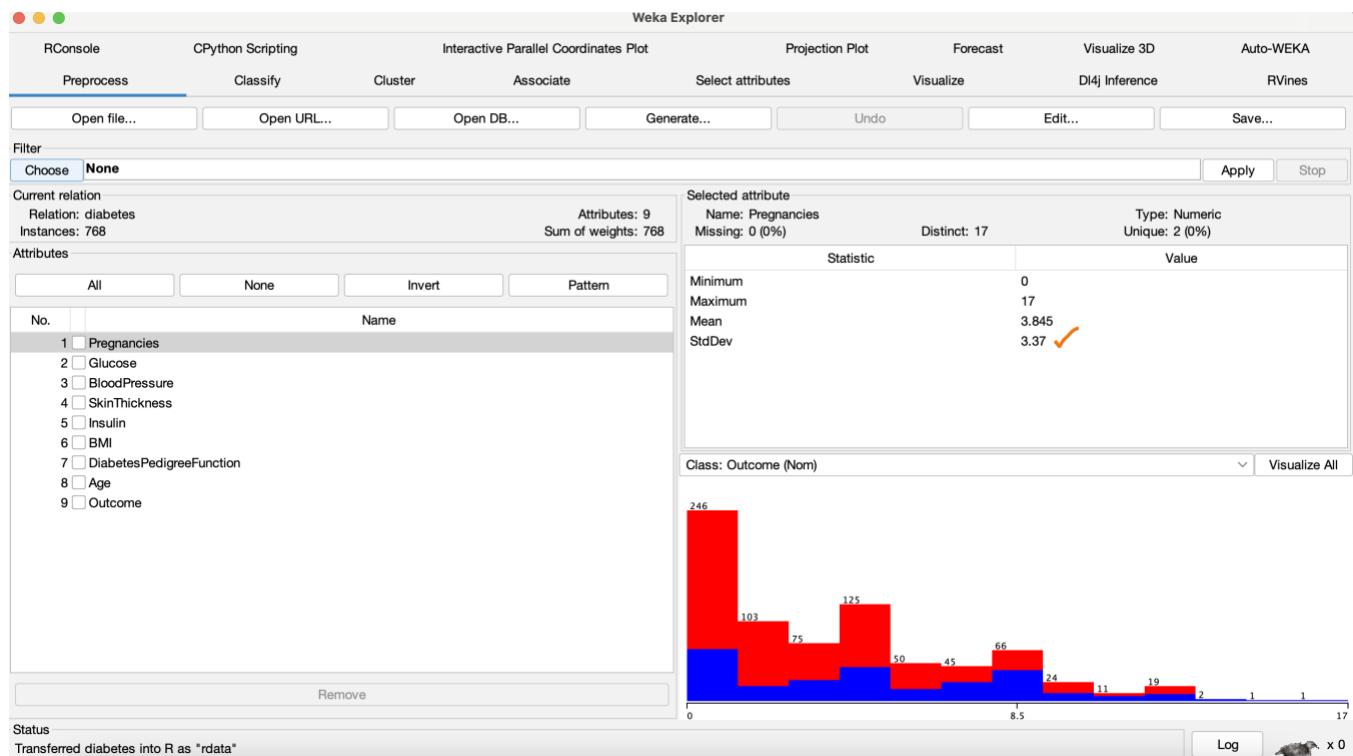
3) Which attribute has a sparse set of attributes from the mean? And Why?

```
list1 = []
for i in df.columns:
    list1.append((i,np.std(df[i])))
    print("{} -> {:.2f}".format(i,np.std(df[i])))
```

Pregnancies --> 3.37
Glucose --> 31.95
BloodPressure --> 19.34
SkinThickness --> 15.94
Insulin --> 115.17
BMI --> 7.88
DiabetesPedigreeFunction --> 0.33
Age --> 11.75
Outcome --> 0.48

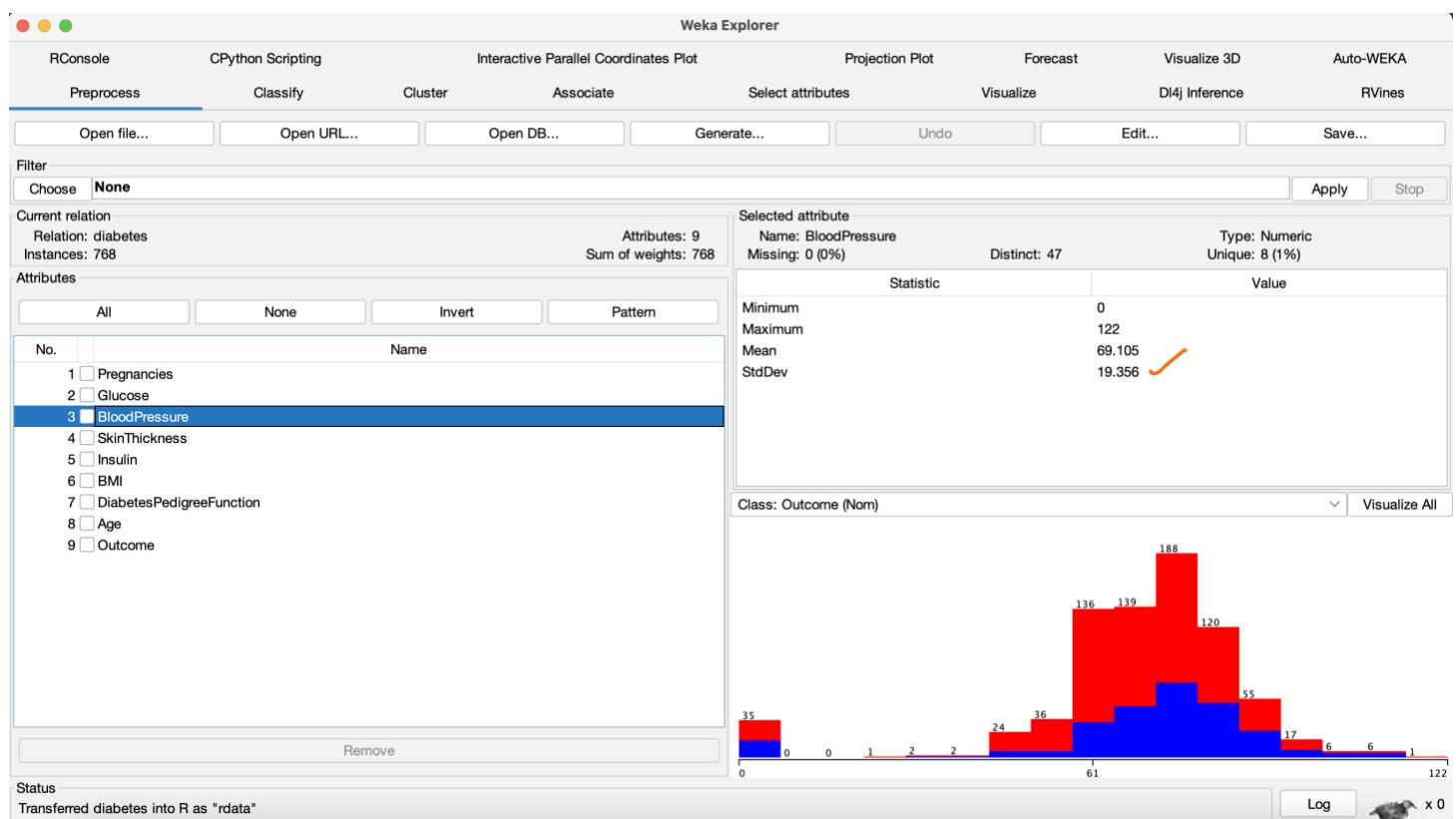
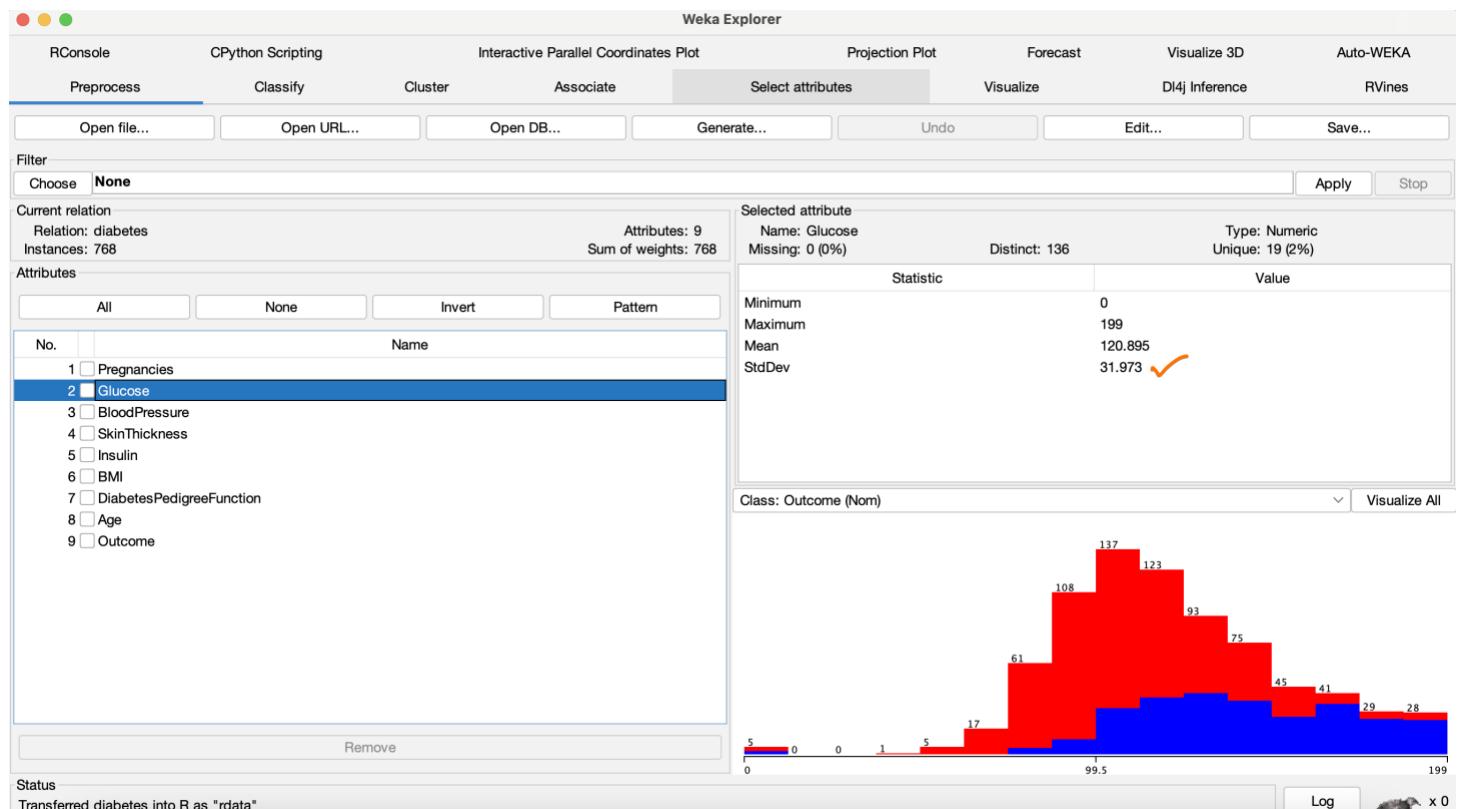
```
print("Maximum spread : ",max(list1,key=lambda item:item[1]))
print("Minimum spread : ",min(list1,key=lambda item:item[1]))
```

Maximum spread : ('Insulin', 115.1689492646728)
Minimum spread : ('DiabetesPedigreeFunction', 0.331112816028629)



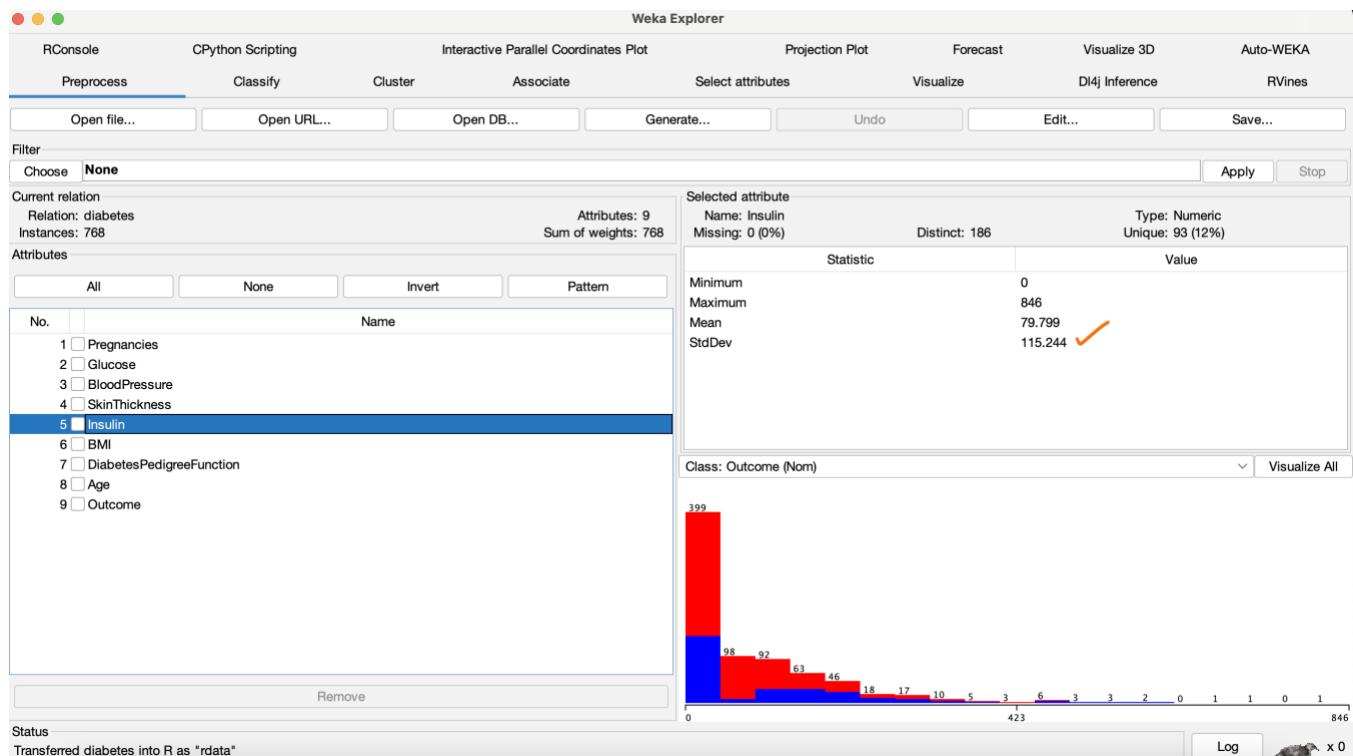
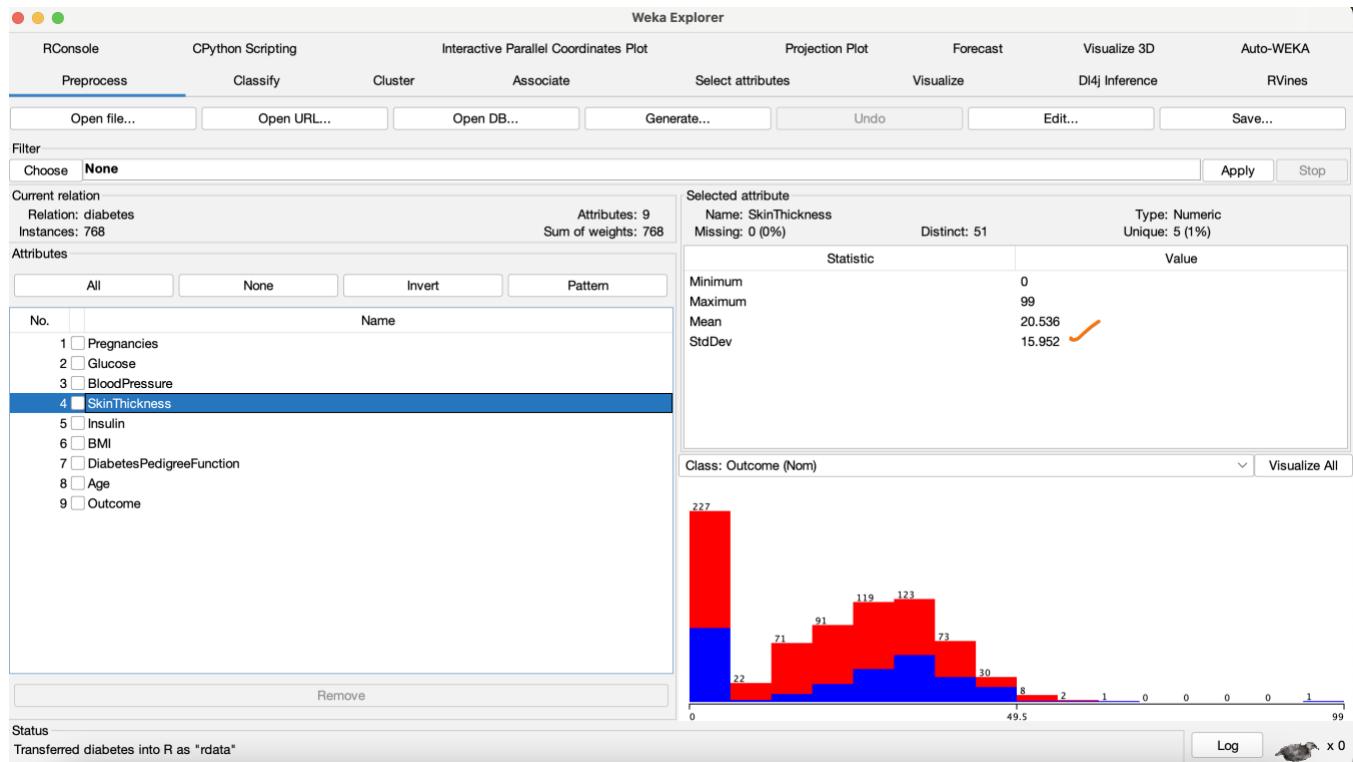
CYCLE SHEET-I

Prashanth.S 19MID0020



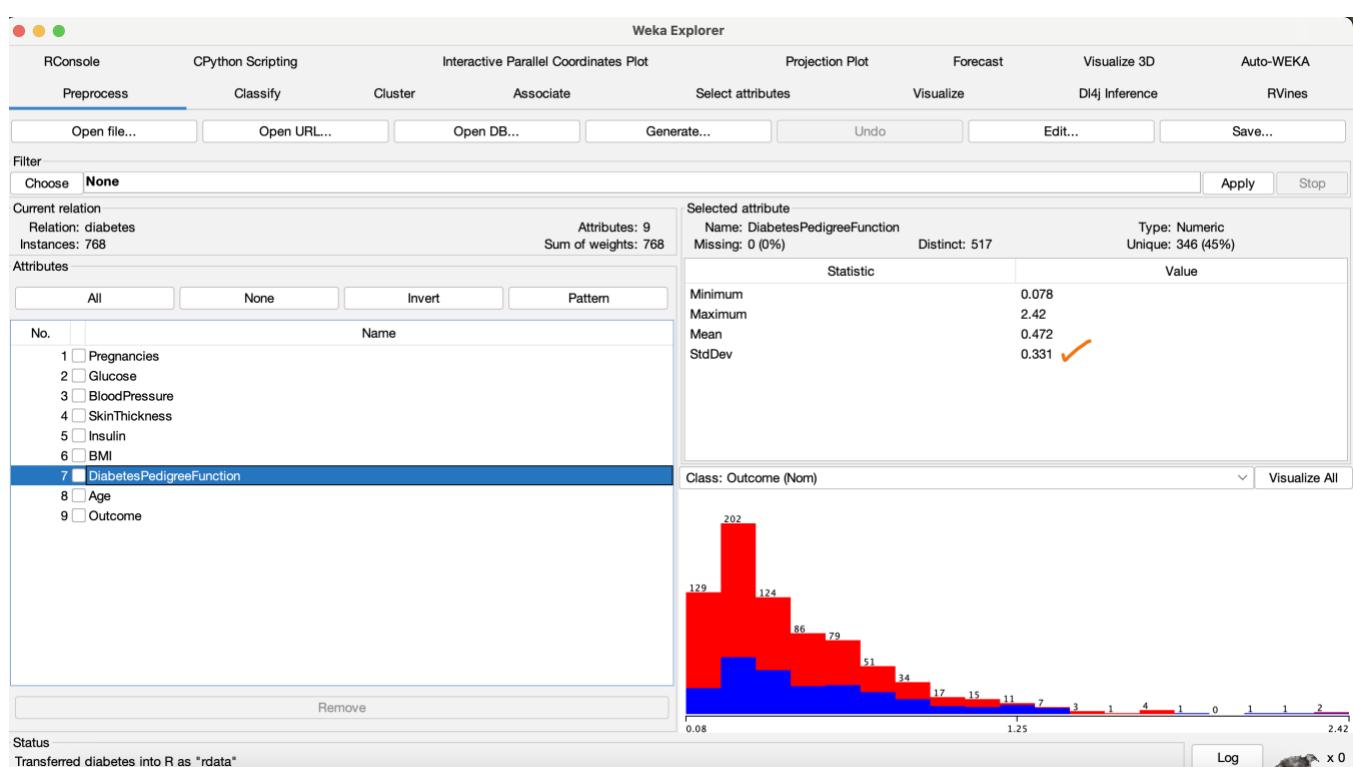
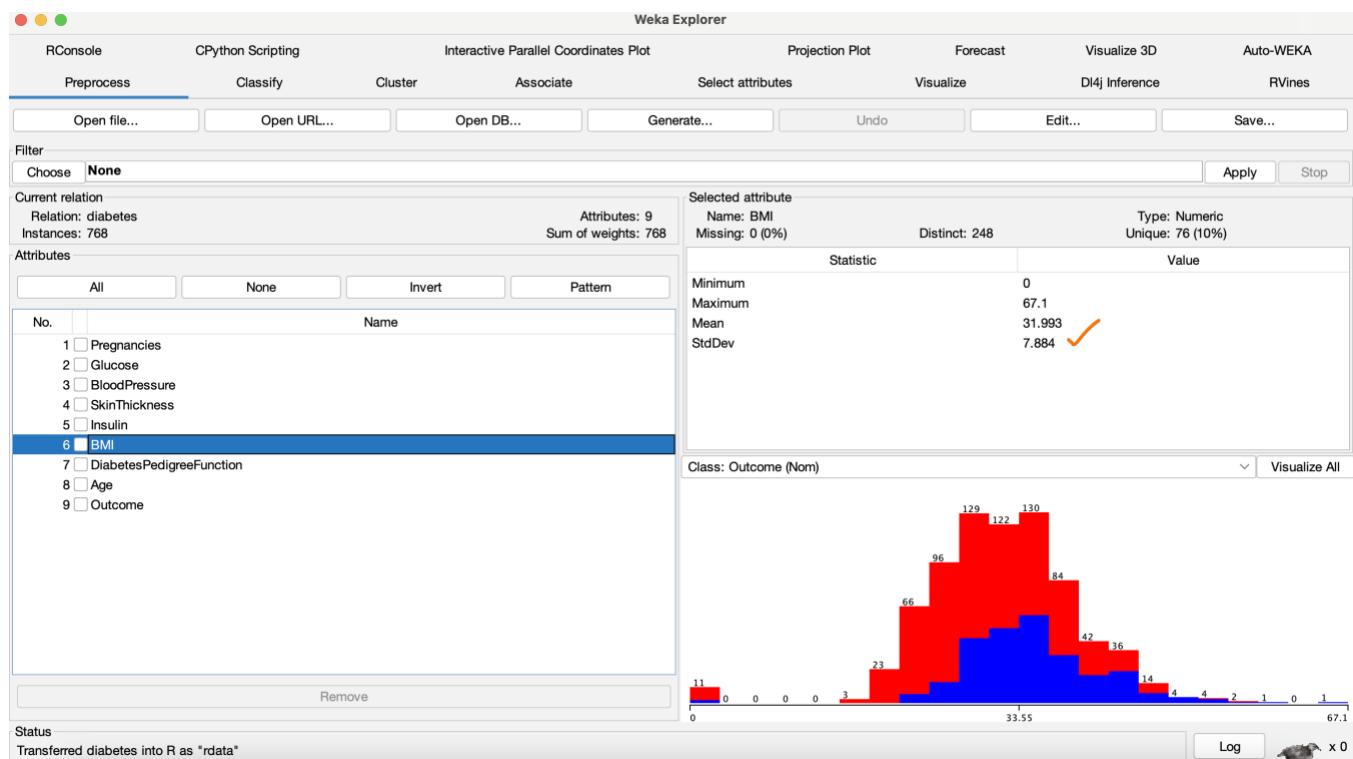
CYCLE SHEET-I

Prashanth.S 19MID0020



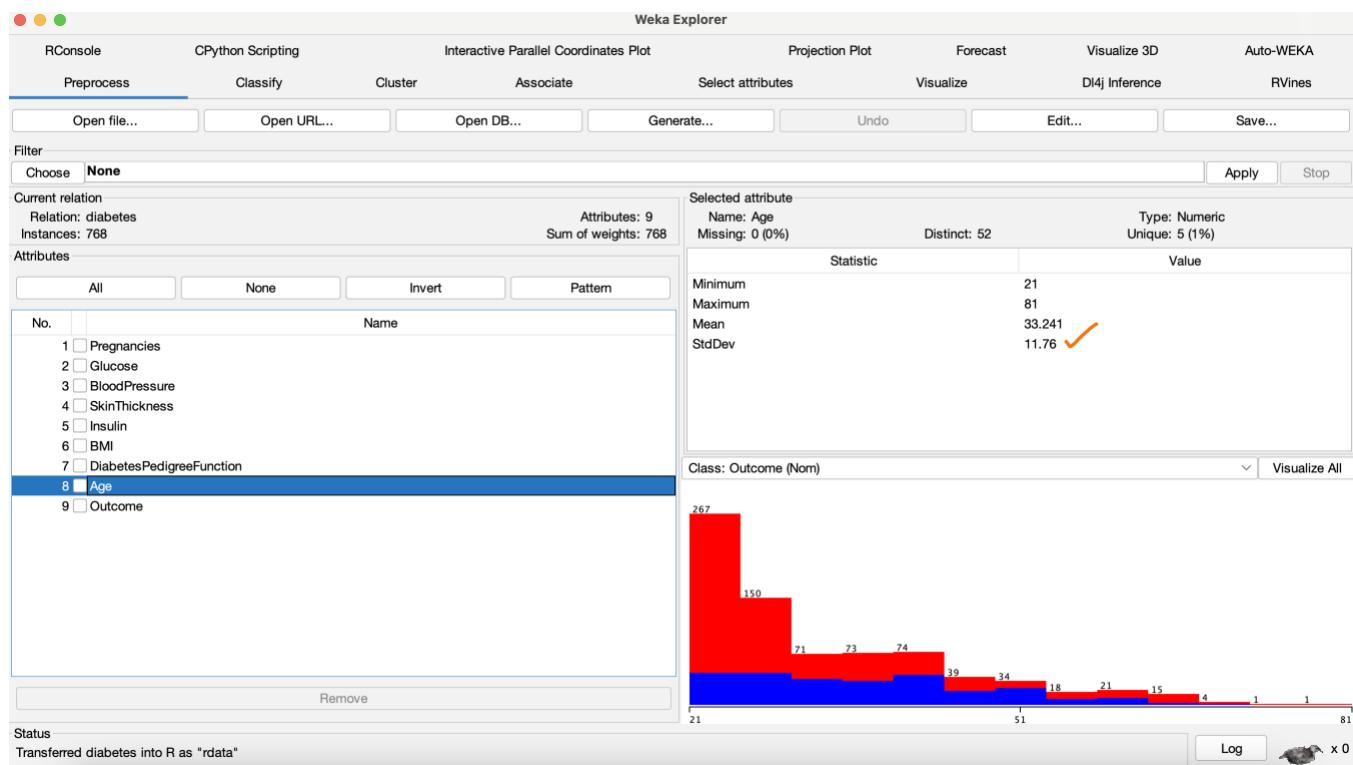
CYCLE SHEET-I

Prashanth.S 19MID0020



CYCLE SHEET-I

Prashanth.S 19MID0020



Maximum Spread → Insulin attribute

Minimum Spread → DiabetesPedigreeFunction

CYCLE SHEET-I

Prashanth.S 19MID0020

4) Find how many outliers in the Dataset? What we should do on the outlier?

Without applying the outliers

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The 'Classifier' dropdown is set to 'J48 - C 0.25 - M 2'. The 'Test options' section shows 'Percentage split' selected with 70% chosen. The 'Classifier output' pane displays the following text:

```

    === Run information ===
Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: diabetes
Instances: 768
Attributes: 9
Pregnancies
Glucose
BloodPressure
SkinThickness
Insulin
BMI
DiabetesPedigreeFunction
Age
Outcome
Test mode: split 70.0% train, remainder test
    === Classifier model (full training set) ===
J48 pruned tree
-----
```

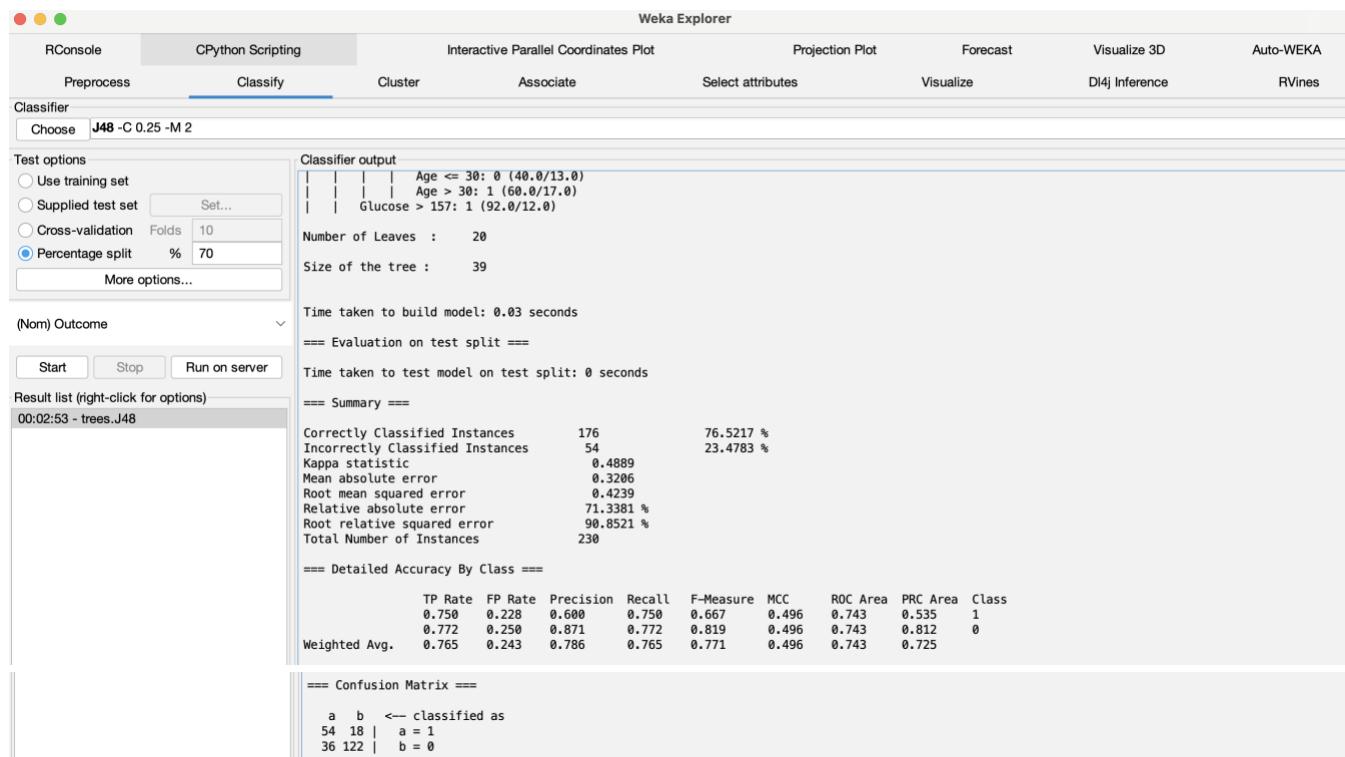
The second screenshot shows the same interface after applying outliers. The 'Classifier output' pane now displays the decision tree structure with outliers included:

```

    Glucose <= 127
    |   BMI <= 26.4: 0 (132.0/3.0)
    |   BMI > 26.4
    |       Age <= 28: 0 (180.0/22.0)
    |       Age > 28
    |           Glucose <= 99: 0 (55.0/10.0)
    |           Glucose > 99
    |               DiabetesPedigreeFunction <= 0.56: 0 (84.0/34.0)
    |               DiabetesPedigreeFunction > 0.56
    |                   Pregnancies <= 6
    |                       Age <= 30: 1 (4.0)
    |                           Age > 30
    |                               Age <= 34: 0 (7.0/1.0)
    |                                   Age > 34
    |                                       BMI <= 33.1: 1 (6.0)
    |                                           BMI > 33.1: 0 (4.0/1.0)
    |                                               Pregnancies > 6: 1 (13.0)
    Glucose > 127
    |   BMI <= 29.9
    |       Glucose <= 145: 0 (41.0/6.0)
    |       Glucose > 145
    |           Age <= 25: 0 (4.0)
    |           Age > 25
    |               Age <= 61
    |                   BMI <= 27.1: 1 (12.0/1.0)
    |                   BMI > 27.1
    |                       BloodPressure <= 82
    |                           DiabetesPedigreeFunction <= 0.396: 1 (8.0/1.0)
    |                           DiabetesPedigreeFunction > 0.396: 0 (3.0)
    |                               BloodPressure > 82: 0 (4.0)
    |                                   Age > 61: 0 (4.0)
    BMI > 29.9
    |   Glucose <= 157
    |       BloodPressure <= 61: 1 (15.0/1.0)
    |           BloodPressure > 61
```

CYCLE SHEET-I

Prashanth.S 19MID0020



Finding the outliers via inter-quartile range

```

length = 0
outlier_factor = 3
for i in df.columns:

    print((colored("\n{}-Feature".format(i), attrs=[bold])))

    q1_y = df[i].quantile(0.25)
    print("1st quantile : ",q1_y)

    q3_y = df[i].quantile(0.75)
    print("3rd quantile : ",q3_y)

    IQR = q3_y - q1_y
    print("Inter-Quartile Range : ",IQR)

    lower_limit = q1_y - outlier_factor * IQR
    upper_limit = q3_y + outlier_factor * IQR

    print("Lower Limit : ",lower_limit)
    print("Upper Limit : ",upper_limit)

    length += len(df[np.logical_or( df[i] < lower_limit , df[i] > upper_limit )])
    attribute_name = i + ' outlier'
    df[attribute_name] = np.logical_or( df[i] < lower_limit , df[i] > upper_limit)

```

CYCLE SHEET-I

Prashanth.S 19MID0020

Pregnancies-Feature

1st quantile : 1.0
3rd quantile : 6.0
Inter-Quartile Range : 5.0
Lower Limit : -14.0
Upper Limit : 21.0

Glucose-Feature

1st quantile : 99.0
3rd quantile : 140.25
Inter-Quartile Range : 41.25
Lower Limit : -24.75
Upper Limit : 264.0

Insulin-Feature

1st quantile : 0.0
3rd quantile : 127.25
Inter-Quartile Range : 127.25
Lower Limit : -381.75
Upper Limit : 509.0

BMI-Feature

1st quantile : 27.3
3rd quantile : 36.6
Inter-Quartile Range : 9.3
Lower Limit : -0.6000000000000014
Upper Limit : 64.5

Outcome-Feature

1st quantile : 0.0
3rd quantile : 1.0
Inter-Quartile Range : 1.0
Lower Limit : -3.0
Upper Limit : 4.0

BloodPressure-Feature

1st quantile : 62.0
3rd quantile : 80.0
Inter-Quartile Range : 18.0
Lower Limit : 8.0
Upper Limit : 134.0

SkinThickness-Feature

1st quantile : 0.0
3rd quantile : 32.0
Inter-Quartile Range : 32.0
Lower Limit : -96.0
Upper Limit : 128.0

DiabetesPedigreeFunction-Feature

1st quantile : 0.24375
3rd quantile : 0.62625
Inter-Quartile Range : 0.3824999999999995
Lower Limit : -0.9037499999999999
Upper Limit : 1.77375

Age-Feature

1st quantile : 24.0
3rd quantile : 41.0
Inter-Quartile Range : 17.0
Lower Limit : -27.0
Upper Limit : 92.0

```
: print("Number of outliers : ",length)
print("Percentage of Outliers : ",length/len(df) * 100)
```

Number of outliers : 51
Percentage of Outliers : 6.640625

There are 51 outliers in the data-set via normal solving

CYCLE SHEET-I

Prashanth.S 19MID0020

weka.filters.unsupervised.attribute.InterquartileRange

About

A filter for detecting outliers and extreme values based on interquartile ranges.

attributeIndices: 1-8

debug: False

detectionPerAttribute: False

doNotCheckCapabilities: False

extremeValuesAsOutliers: False

extremeValuesFactor: 6.0

outlierFactor: 3.0

outputOffsetMultiplier: False

More **Capabilities**

Information

SYNOPSIS
A filter for detecting outliers and extreme values based on interquartile ranges. The filter skips the class attribute.

Outliers:
 $Q3 + OF \cdot IQR < x \leq Q3 + EVF \cdot IQR$
 or
 $Q1 - EVF \cdot IQR \leq x < Q1 - OF \cdot IQR$

Extreme values:
 $x > Q3 + EVF \cdot IQR$
 or
 $x < Q1 - EVF \cdot IQR$

Key:
 Q1 = 25% quartile
 Q3 = 75% quartile
 IQR = Interquartile Range, difference between Q1 and Q3
 OF = Outlier Factor

RConsole **CPython Scripting**

Preprocess **Classify** **Cluster** **Associate**

Open file... **Open URL...** **Open DB...** **Generate...** **Undo** **Edit...** **Save...**

Filter Choose **InterquartileRange -R 1-8 -O 3.0 -E 6.0** **Apply** **Stop**

Current relation
 Relation: diabetes-weka.filters.unsupervised.attribute.InterquartileRange-R1-8-O3... Attributes: 11
 Instances: 768 Sum of weights: 768

Attributes

No.	Name
1	Pregnancies
2	Glucose
3	BloodPressure
4	SkinThickness
5	Insulin
6	BMI
7	DiabetesPedigreeFunction
8	Age
9	Outcome
10	<input checked="" type="checkbox"/> Outlier
11	<input type="checkbox"/> ExtremeValue

Remove

Interactive Parallel Coordinates Plot

Projection Plot

Forecast

Visualize 3D

Auto-WEKA

Select attributes

Visualize

DI4j Inference

RVines

Selected attribute
 Name: Outlier
 Missing: 0 (0%) Distinct: 2 Type: Nominal
 Unique: 0 (0%)

No.	Label	Count	Weight
1	no	719	719
2	yes	49	49

Class: ExtremeValue (Nom) **Visualize All**

There are 49 outliers in the data-set via Weka tool.

CYCLE SHEET-I

Prashanth.S 19MID0020

Removing the found outliers

```
print(colored("1st Quantile", attrs=['bold']))  
q1 = df.quantile(0.25) ## calculates the 1st quantile of df  
print(q1)
```

1st Quantile

```
Pregnancies      1.00000  
Glucose         99.00000  
BloodPressure   62.00000  
SkinThickness   0.00000  
Insulin          0.00000  
BMI              27.30000  
DiabetesPedigreeFunction 0.24375  
Age              24.00000  
Outcome          0.00000  
Name: 0.25, dtype: float64
```

```
print(colored("3rd Quantile", attrs=['bold']))  
q3 = df.quantile(0.75) ## calculates the 3rd quantile of df  
q3
```

3rd Quantile

```
Pregnancies      6.00000  
Glucose         140.25000  
BloodPressure   80.00000  
SkinThickness   32.00000  
Insulin          127.25000  
BMI              36.60000  
DiabetesPedigreeFunction 0.62625  
Age              41.00000  
Outcome          1.00000  
Name: 0.75, dtype: float64
```

```
print(colored("IQR", attrs=['bold']))  
IQR = q3 - q1  
print(IQR)
```

IQR

```
Pregnancies      5.0000  
Glucose          41.2500  
BloodPressure    18.0000  
SkinThickness    32.0000  
Insulin          127.2500  
BMI              9.3000  
DiabetesPedigreeFunction 0.3825  
Age              17.0000  
Outcome          1.0000  
dtype: float64
```

```
print("Before Removing the Outliers : ",len(df))  
df = df[~((df < (q1 - 1.5 * IQR)) | (df > (q3 + 1.5 * IQR))).any(axis=1)]  
print("After Removing the Outliers : ",len(df))
```

Before Removing the Outliers : 768
After Removing the Outliers : 639

CYCLE SHEET-I

Prashanth.S 19MID0020

The screenshot shows the WEKA interface with the 'Preprocess' tab selected. A 'Filter' dialog is open, displaying the 'RemoveWithValues' filter configuration. The 'Selected attribute' is set to 'Outlier' (Nominal, Unique: 0%). The 'No.' column shows two entries: '1 no' with a count of 719 and '2 yes' with a count of 49. The 'Attributes' section shows the filter's capabilities: Class -- Binary class, Date class, Empty nominal class, Missing class values, No class, Nominal class, Numeric class, Relational class, String class, Unary class; Attributes -- Binary attributes, Date attributes, Empty nominal attributes, Missing values, Nominal attributes, Numeric attributes, Relational attributes, String attributes, Unary attributes; Interfaces -- StreamableFilter, UnsupervisedFilter, WeightedAttributesHandler, WeightedInstancesHandler.

Filter Dialog (Top):

No.	Label	Count	Weight
1	no	719	719
2	yes	49	49

Filter Dialog (Bottom):

Choose RemoveWithValues -S 0.0 -C 10 -L last

Current relation: Relation: diabetes-weka.filter weka.filters.unsupervised.instance.RemoveWithValues

Attributes:

- All
- No.:
 - 1 Pregnancies
 - 2 Glucose
 - 3 BloodPressure
 - 4 SkinThickness
 - 5 Insulin
 - 6 BMI
 - 7 DiabetesPedigreeFunction
 - 8 Age
 - 9 Outcome
 - 10 Outlier
 - 11 ExtremeValue

attributeIndex: 10

nominalIndices: last

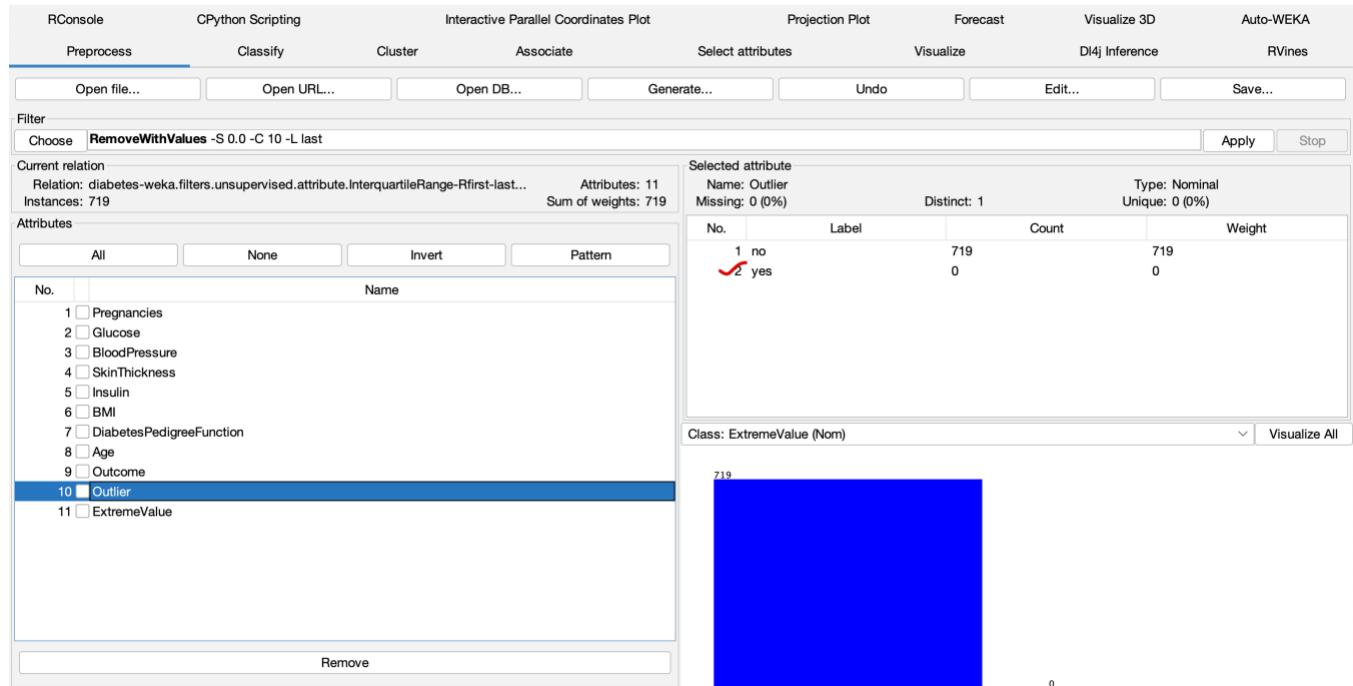
Result View:

49

CYCLE SHEET-I

Prashanth.S 19MID0020

After removing the outliers

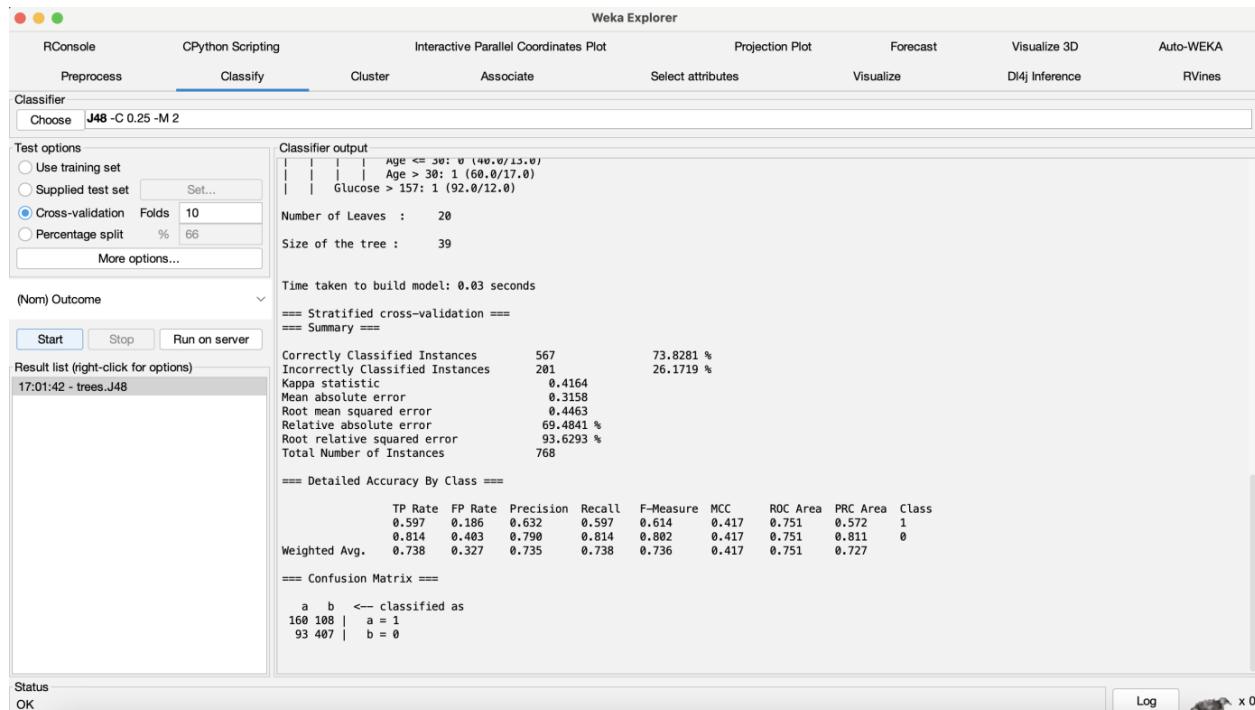


CYCLE SHEET-I

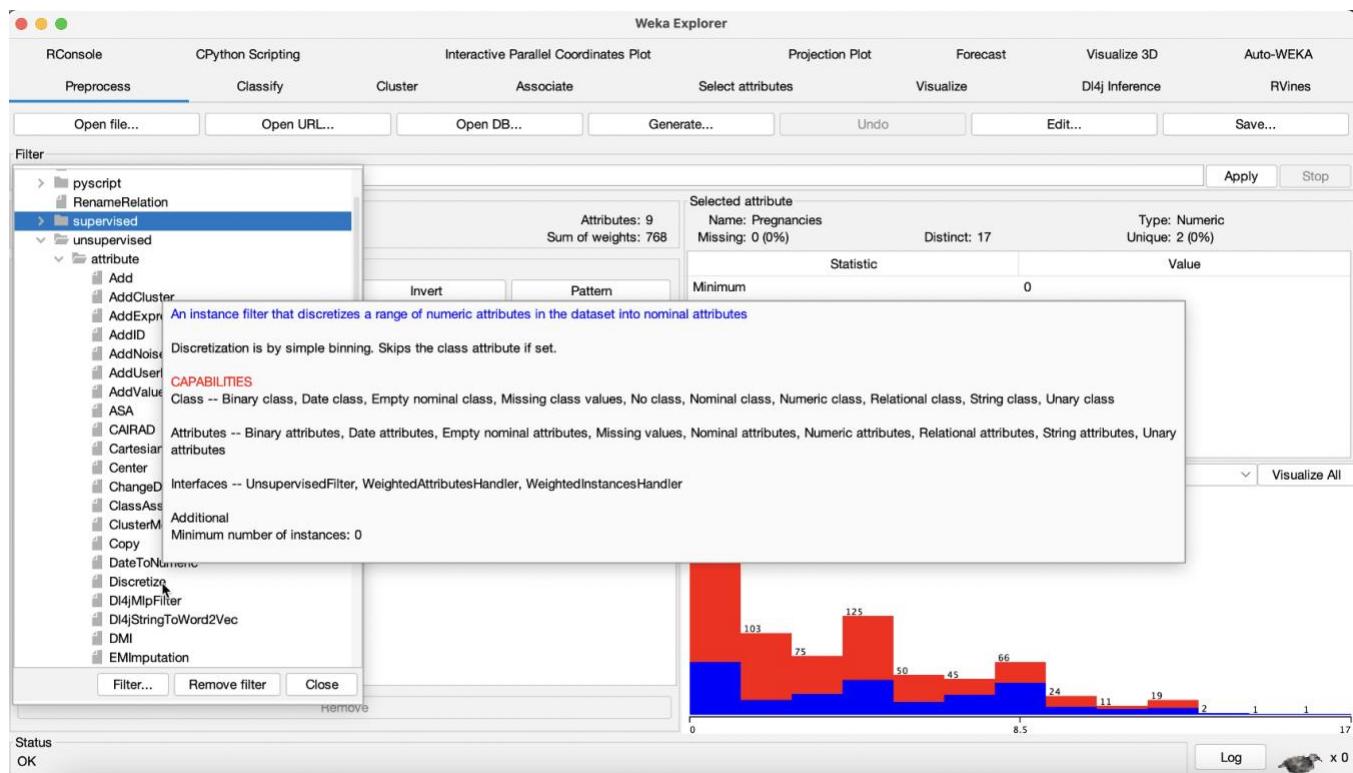
Prashanth.S 19MID0020

5) Create 3 Bin for first 2 attributes. Draw the Bar chart and the count for each?

Before Discretization



Discretization

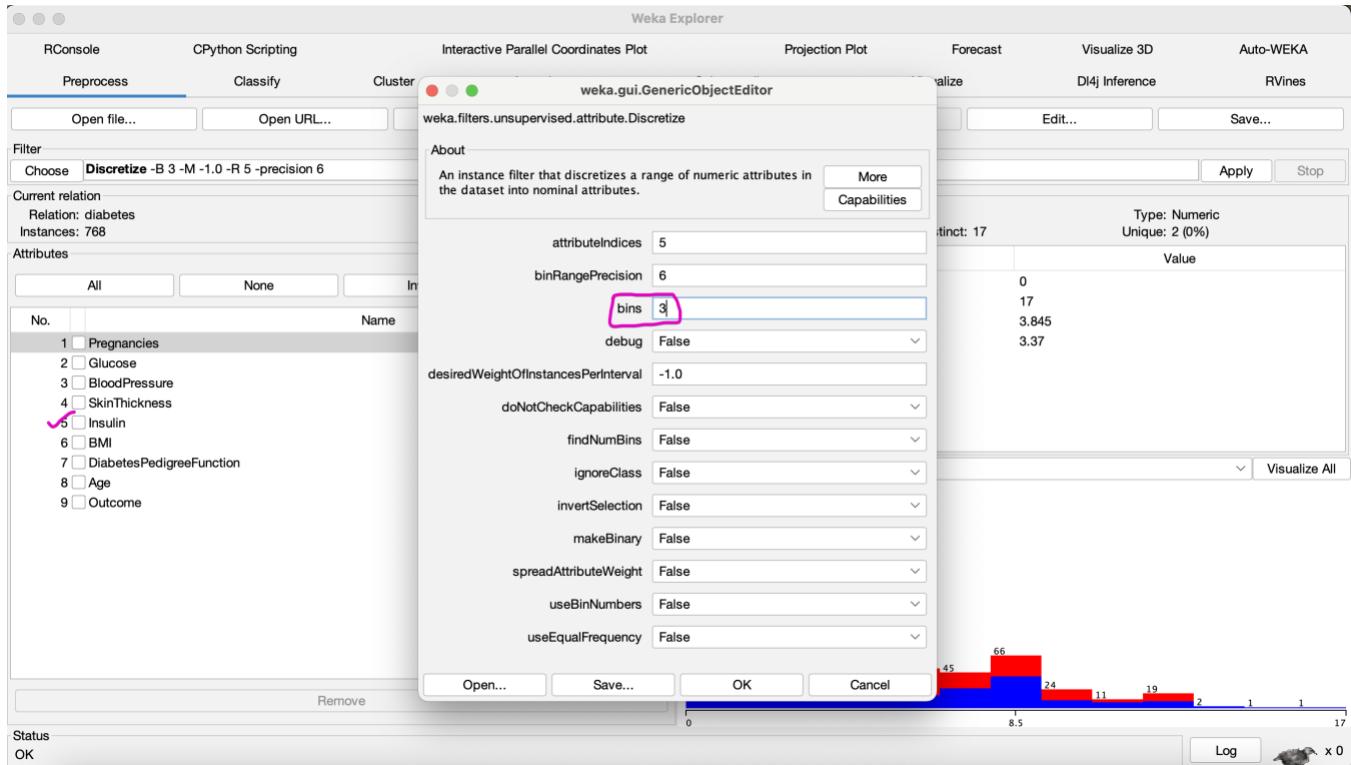


CYCLE SHEET-I

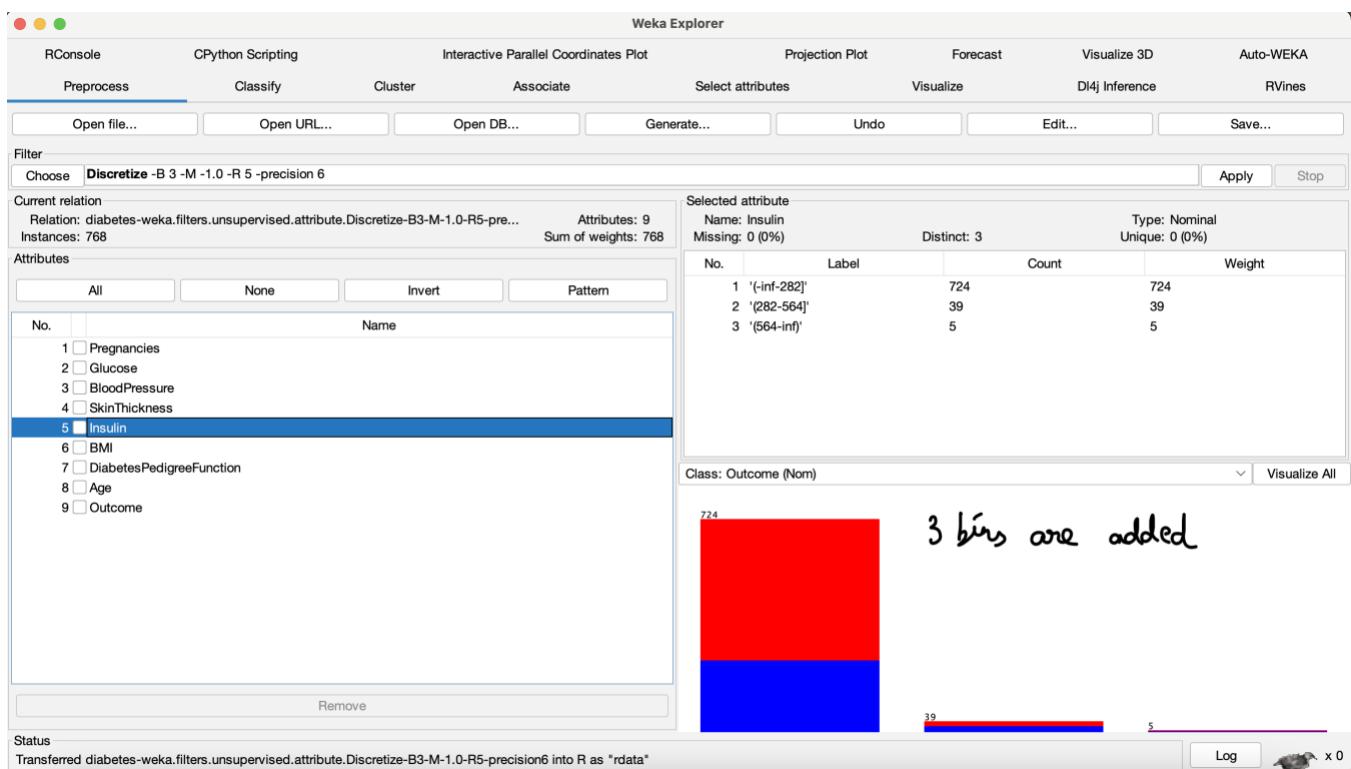
Prashanth.S 19MID0020

After Discretization

Attribute-1 → Insulin



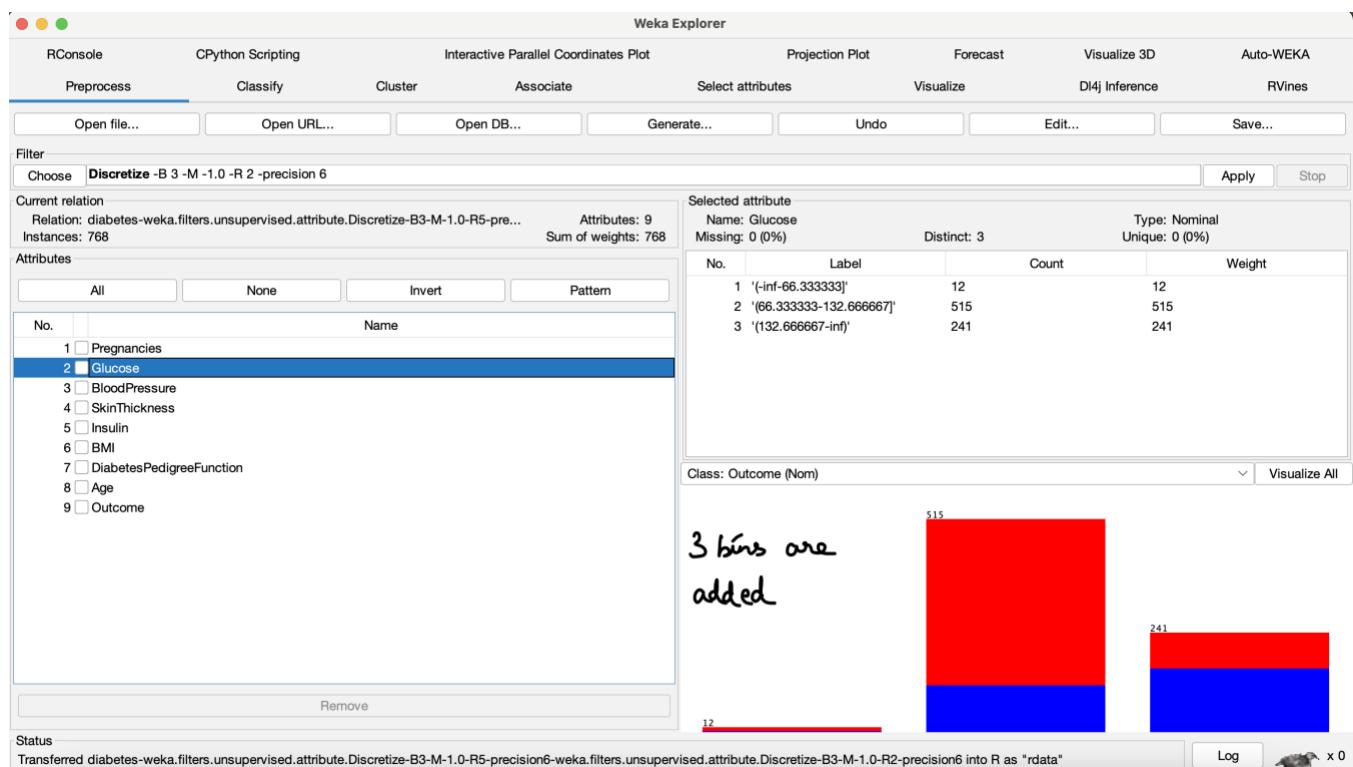
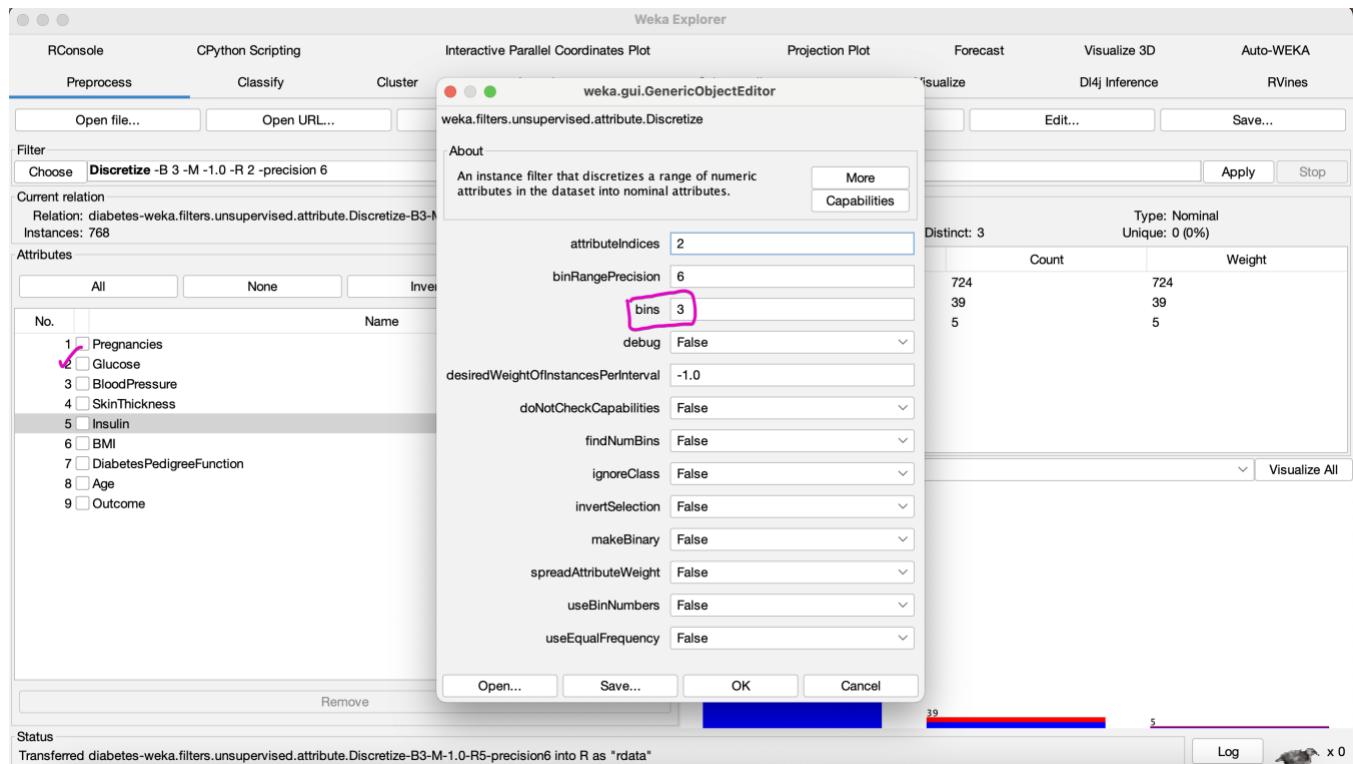
3 bins are added



CYCLE SHEET-I

Prashanth.S 19MID0020

Attribute-2 → Glucose



CYCLE SHEET-I

Prashanth.S 19MID0020

The screenshot shows the Weka Explorer interface with the following details:

- Classifier:** Choose **J48 - C 0.25 -M 2**
- Test options:**
 - Use training set
 - Supplied test set
 - Cross-validation Folds **10**
 - Percentage split % **66**
- Classifier output:**
 - | BMI <= 29.9: 0 (62.0/22.0)
 - | BMI > 29.9: 1 (179.0/47.0)
- Number of Leaves :** 12
- Size of the tree :** 22
- Time taken to build model:** 0.01 seconds
- (Nom) Outcome** dropdown menu with buttons: Start, Stop, Run on server.
- Result list (right-click for options):**
 - 17:01:42 - trees.J48
 - 17:11:11 - trees.J48The second item, "17:11:11 - trees.J48", is highlighted with a blue selection bar.
- Classifier output (continued):**
 - Correctly Classified Instances 568 73.9583 %
 - Incorrectly Classified Instances 200 26.0417 %
 - Kappa statistic 0.4137
 - Mean absolute error 0.3297
 - Root mean squared error 0.4351
 - Relative absolute error 72.5445 %
 - Root relative squared error 91.2802 %
 - Total Number of Instances 768
- Detailed Accuracy By Class:**

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.578	0.174	0.640	0.578	0.608	0.415	0.759	0.588	0.834	1
0.826	0.422	0.785	0.826	0.805	0.415	0.759	0.759	0.748	0
Weighted Avg.	0.740	0.335	0.735	0.740	0.736	0.415	0.759	0.748	
- Confusion Matrix:**

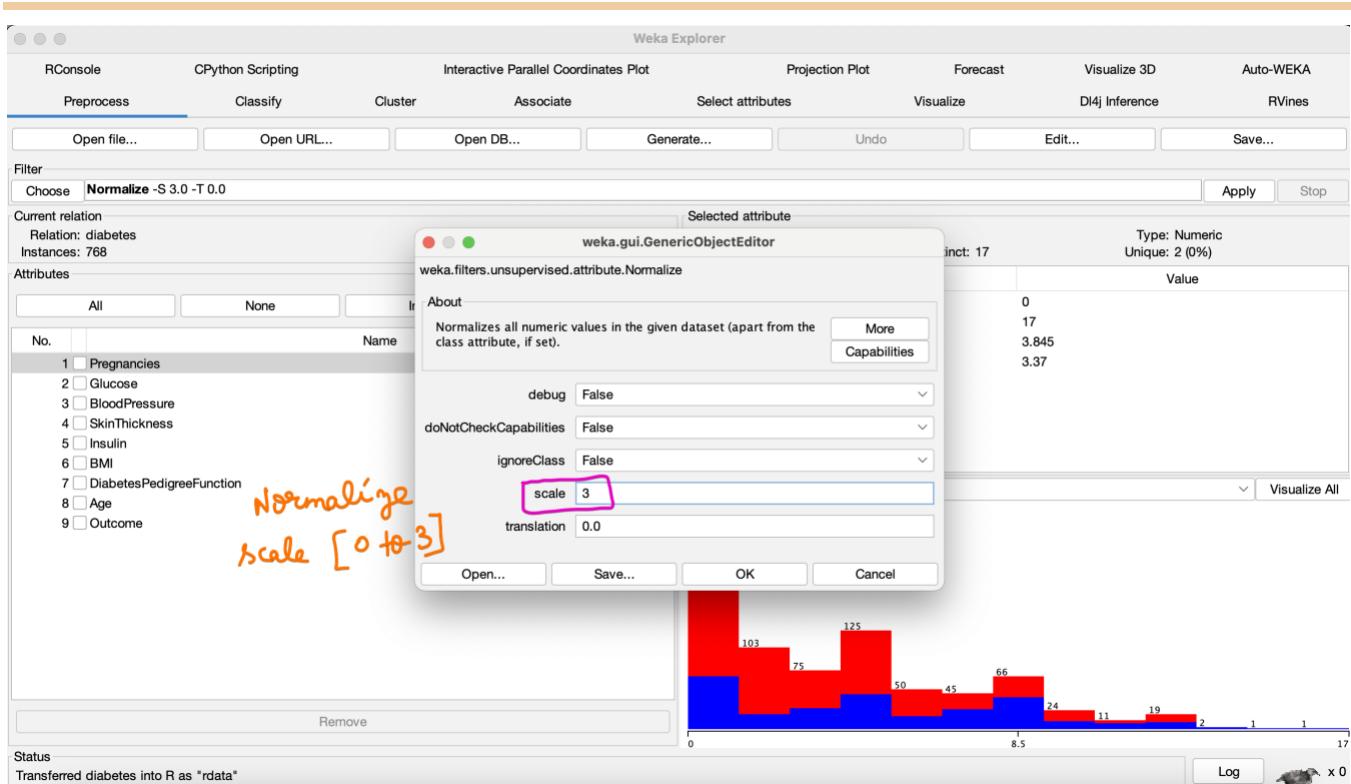
	a	b	<- classified as
155	113		a = 1
87	413		b = 0
- Status:** OK
- Log:** A small icon representing the log file, showing 0 entries.

Both the accuracies before and after Discretization remains the same.

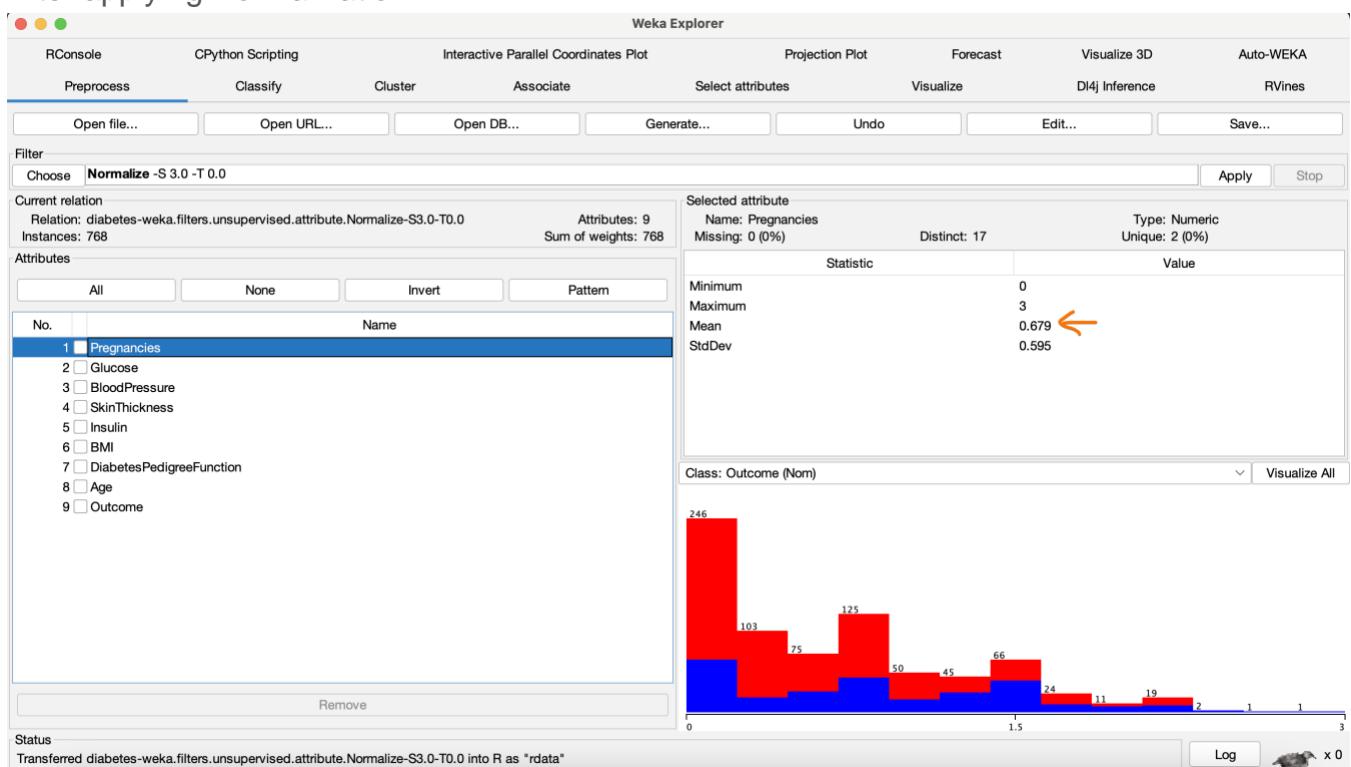
CYCLE SHEET-I

Prashanth.S 19MID0020

6) Apply the Normalize for range [0, 3]. Give mean value for each related attributes.

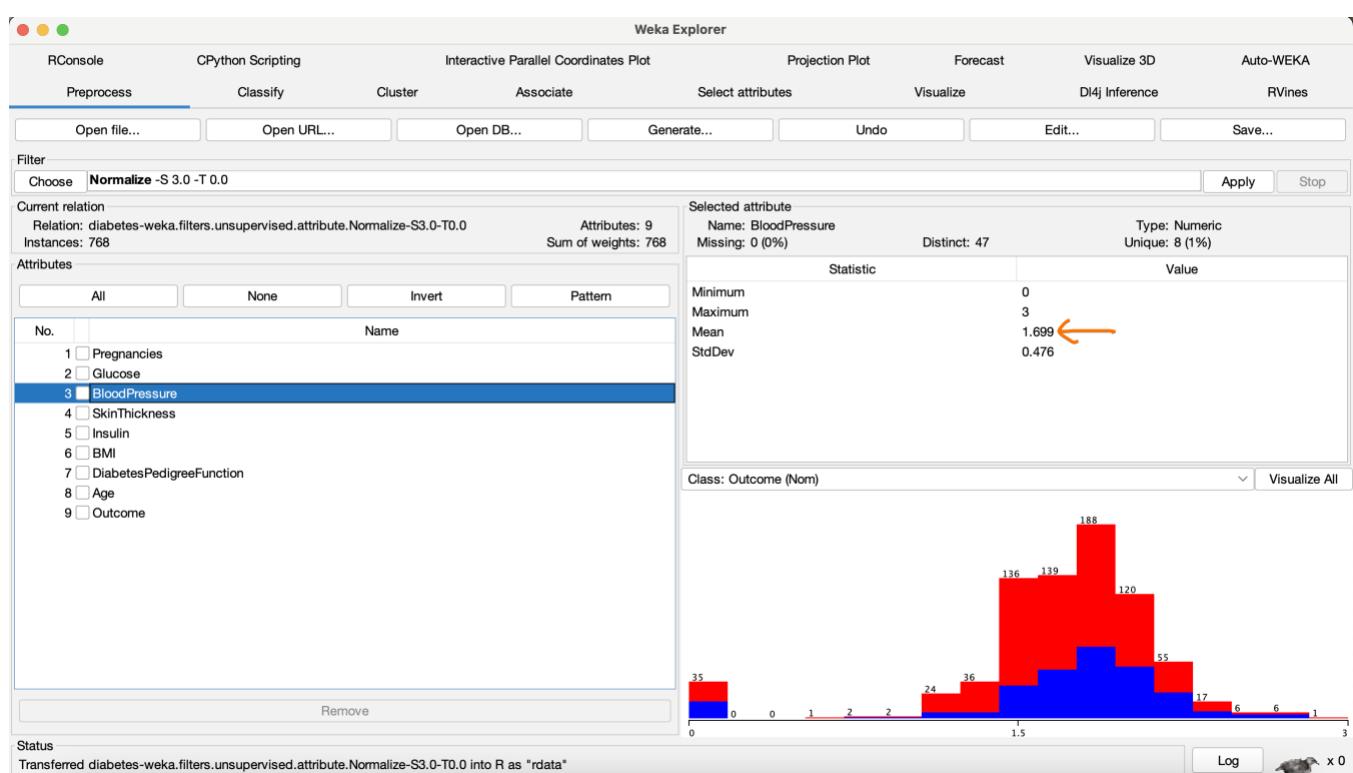
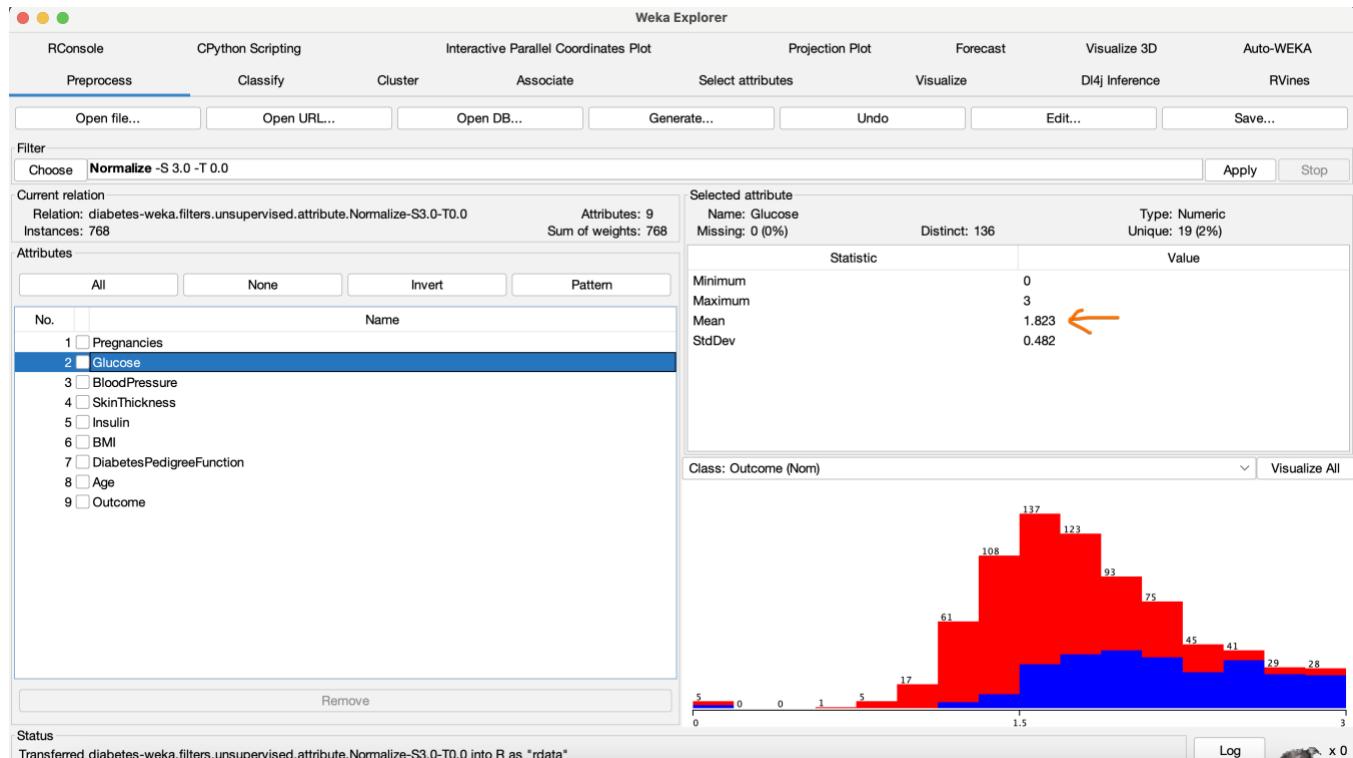


After applying Normalization



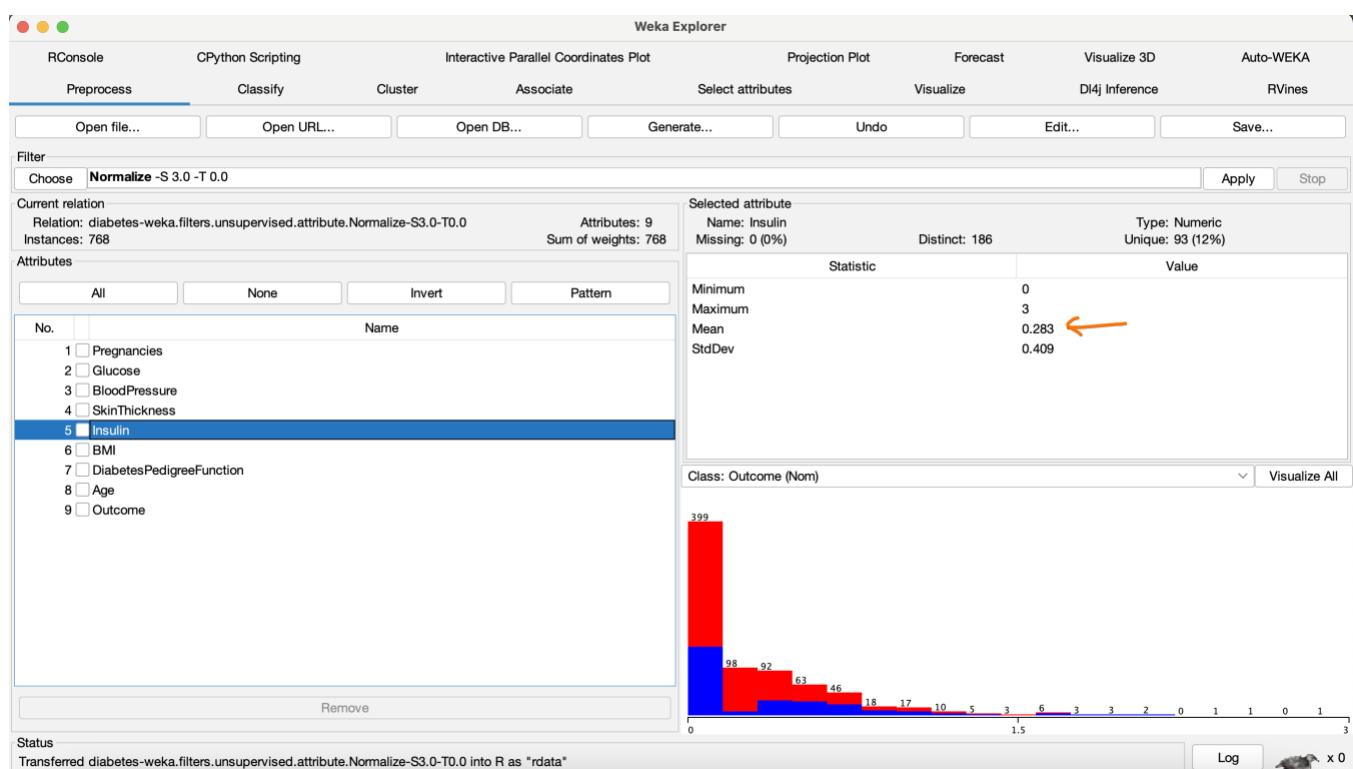
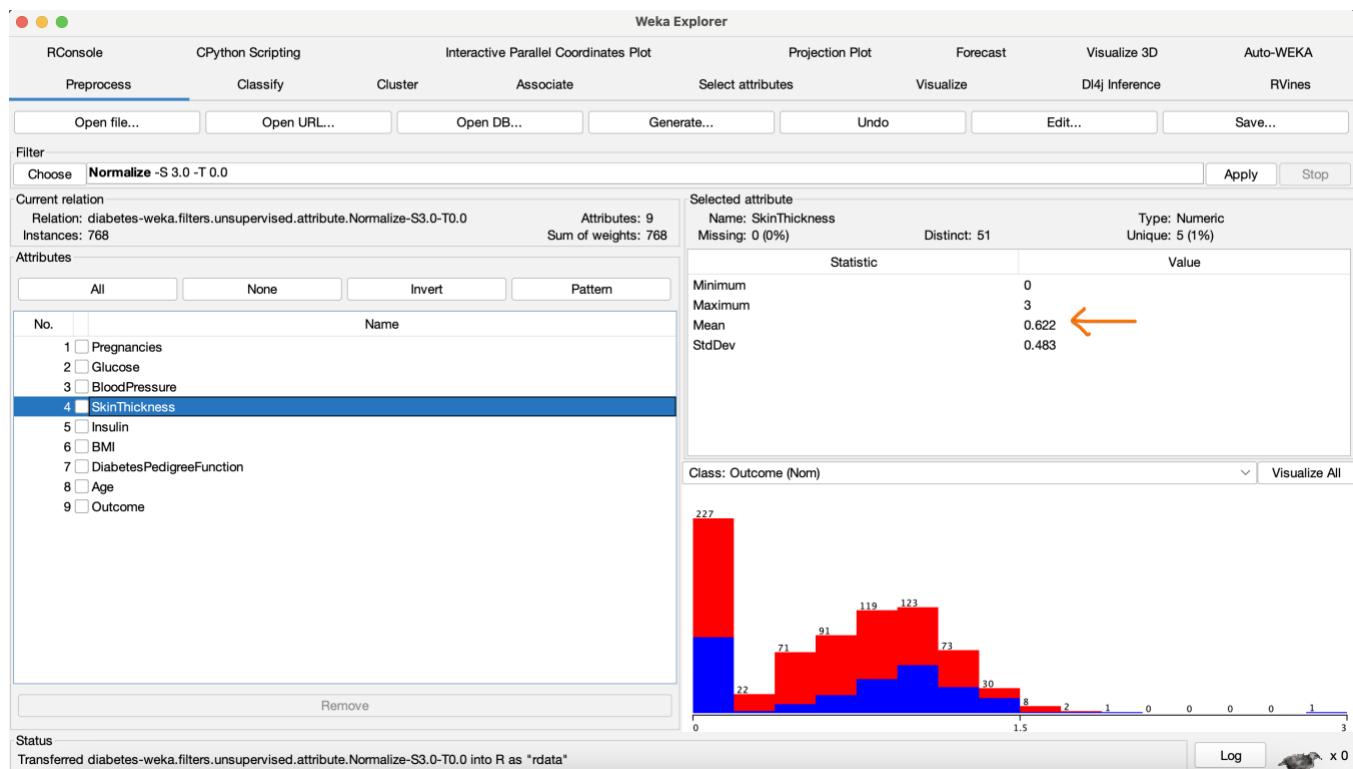
CYCLE SHEET-I

Prashanth.S 19MID0020



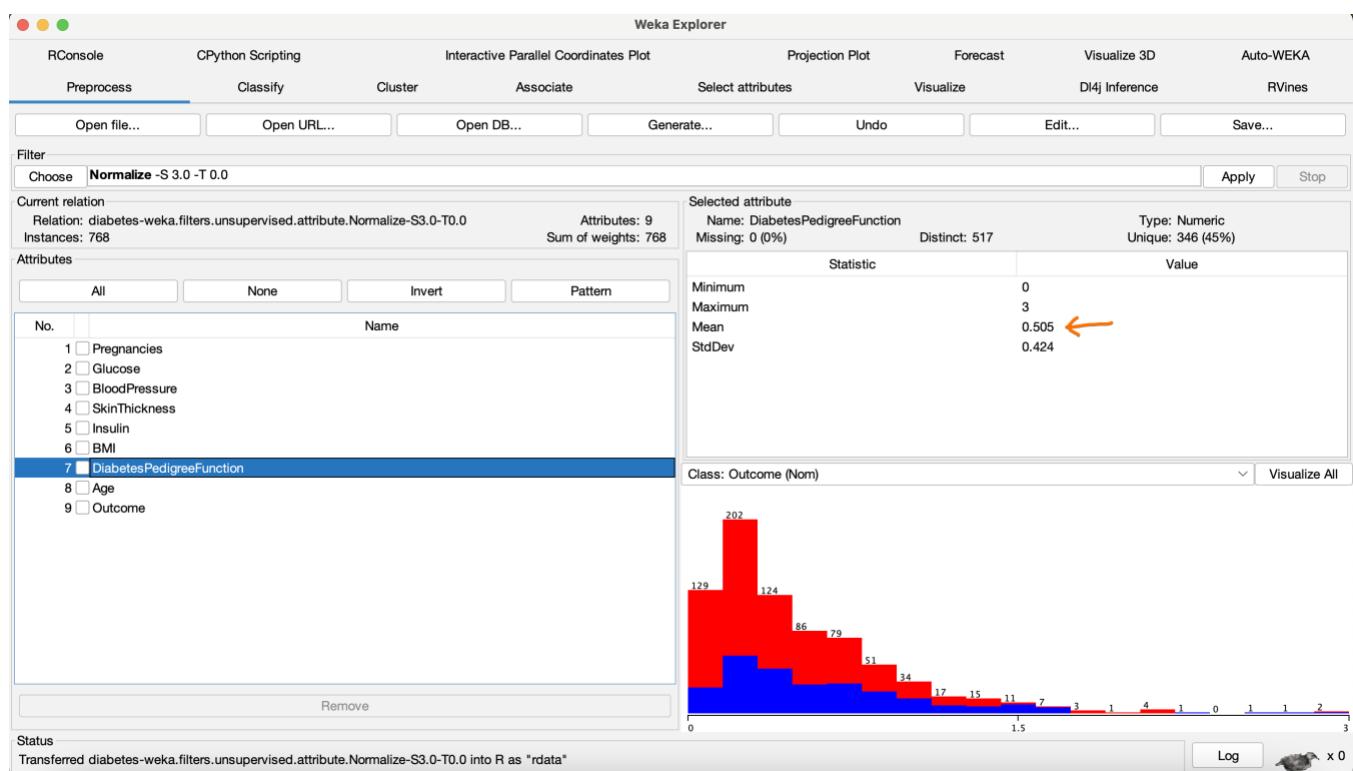
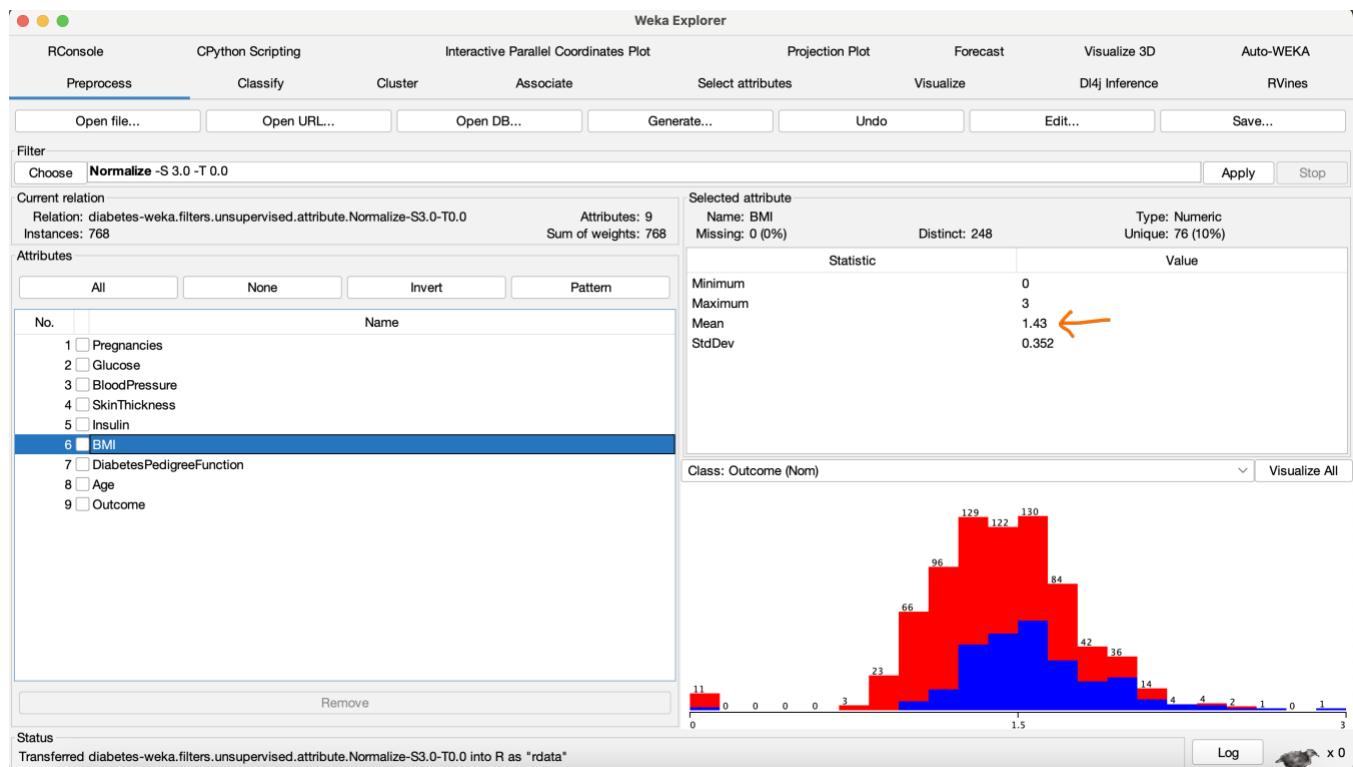
CYCLE SHEET-I

Prashanth.S 19MID0020



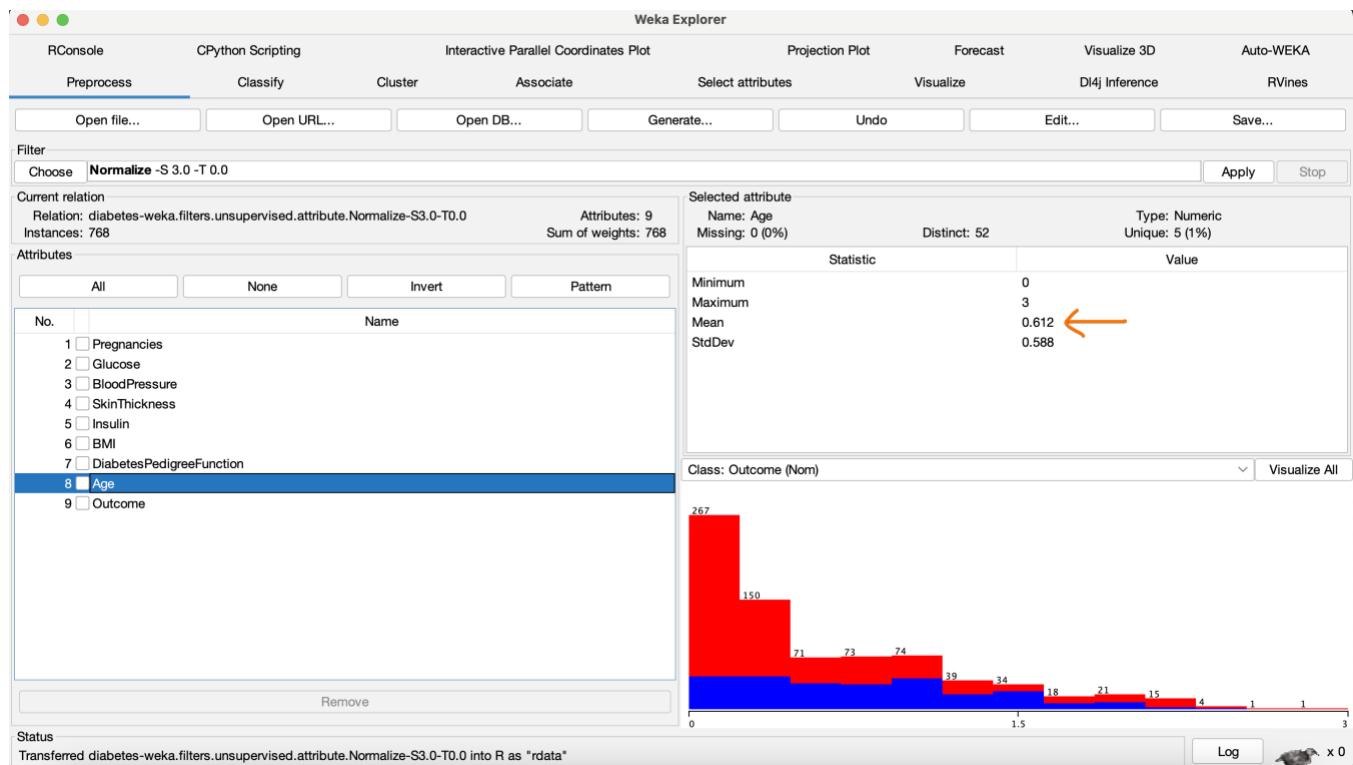
CYCLE SHEET-I

Prashanth.S 19MID0020



CYCLE SHEET-I

Prashanth.S 19MID0020



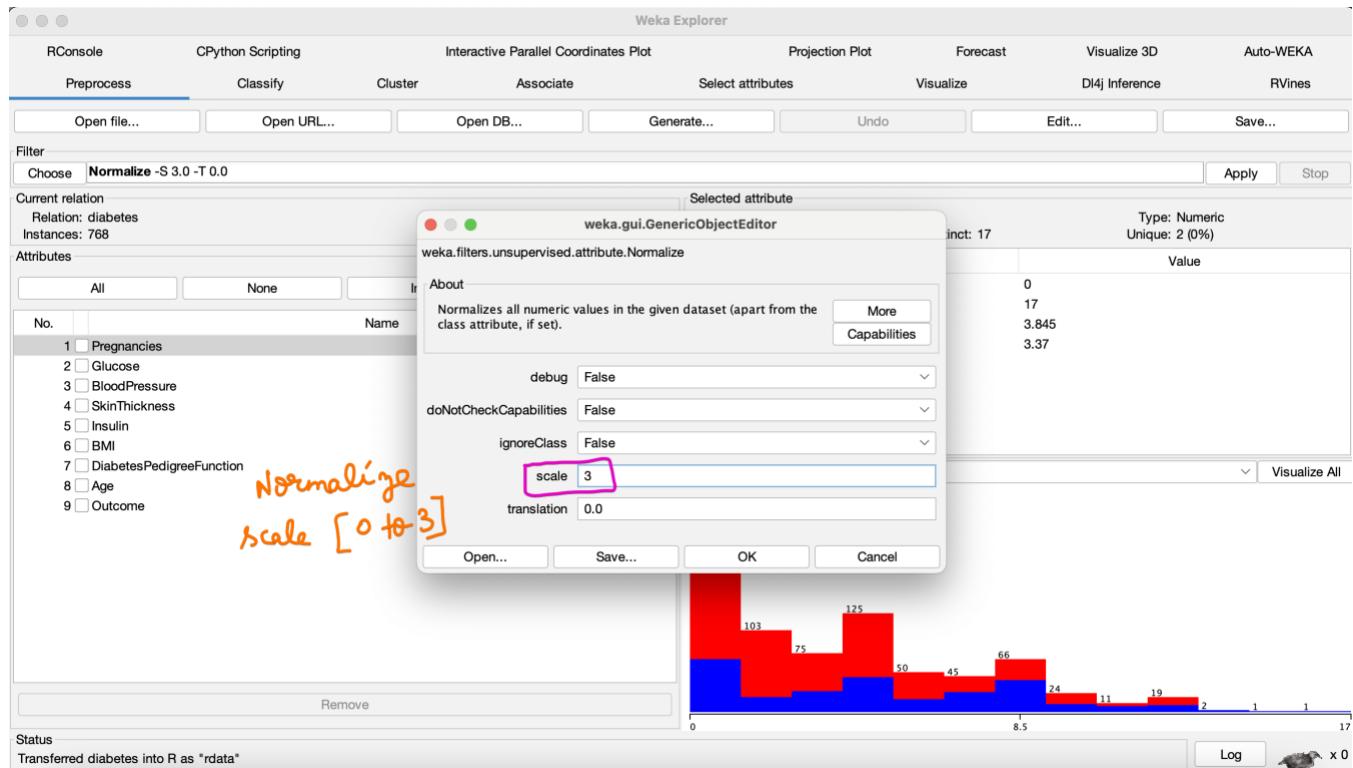
CYCLE SHEET-I

Prashanth.S 19MID0020

7) Discretise and normalize the dataset?

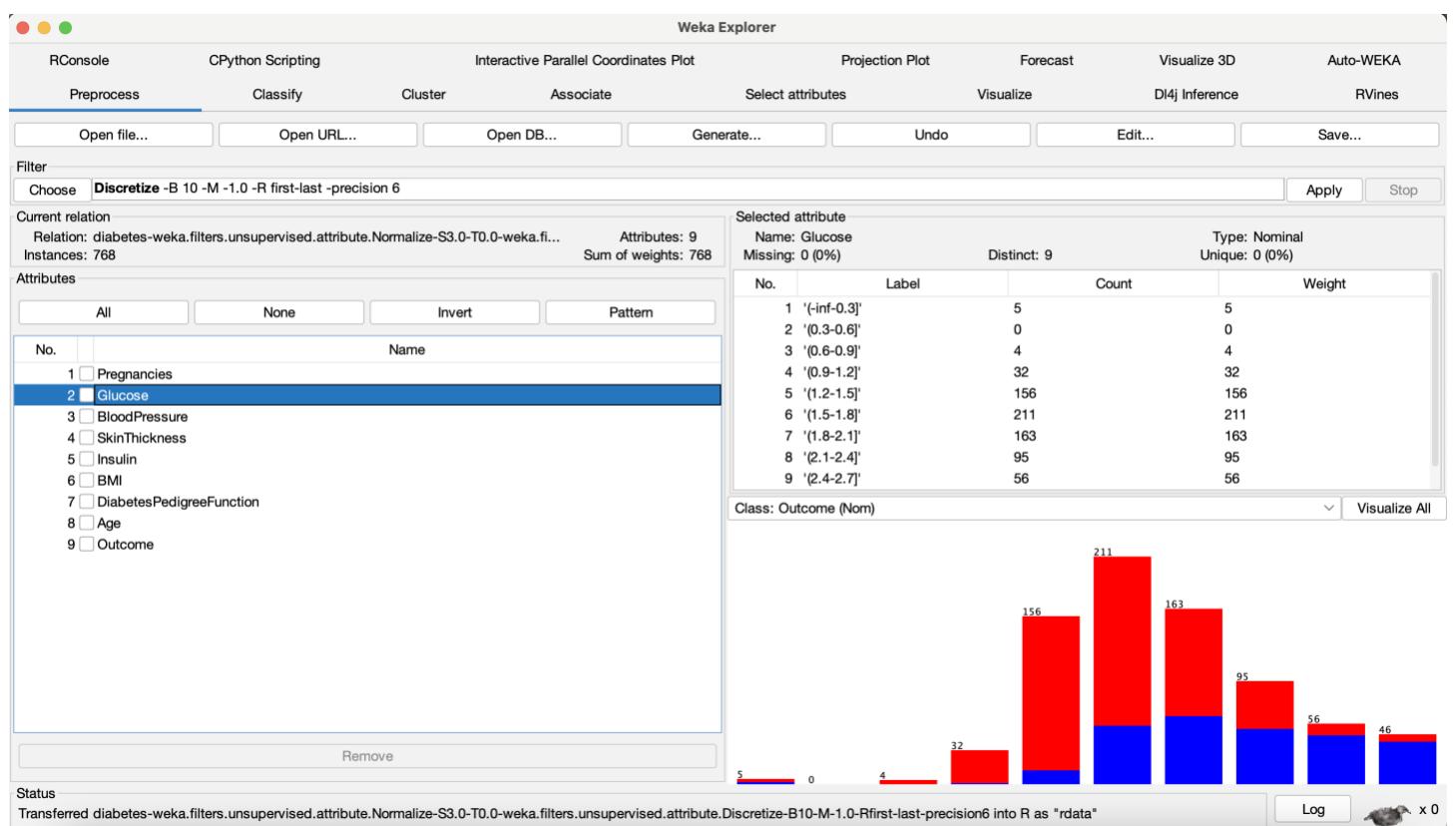
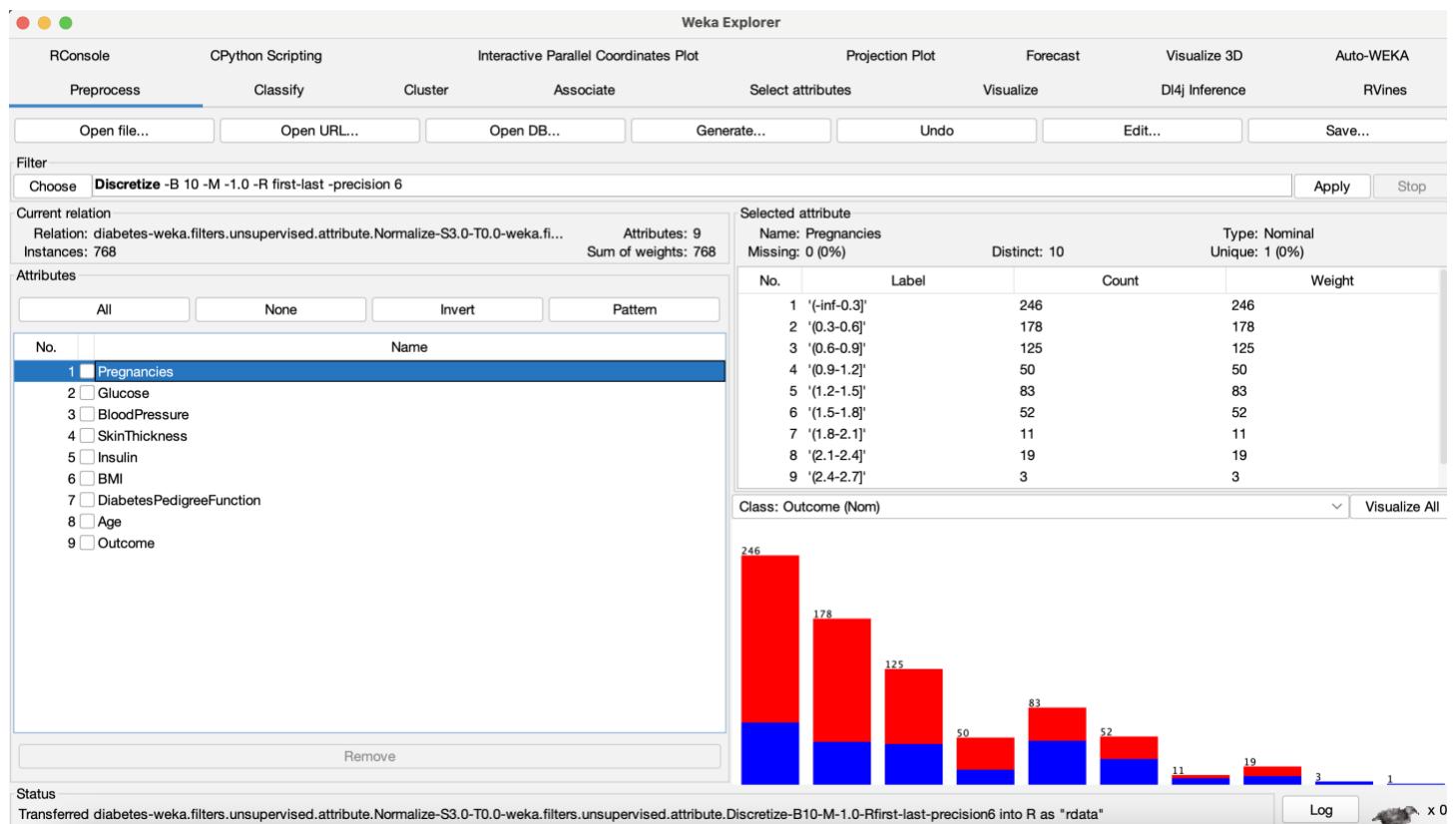
There is no possibility of Discrete sizing and Normalizing the data-set.

So we will first Normalize and Discretise the data-set.



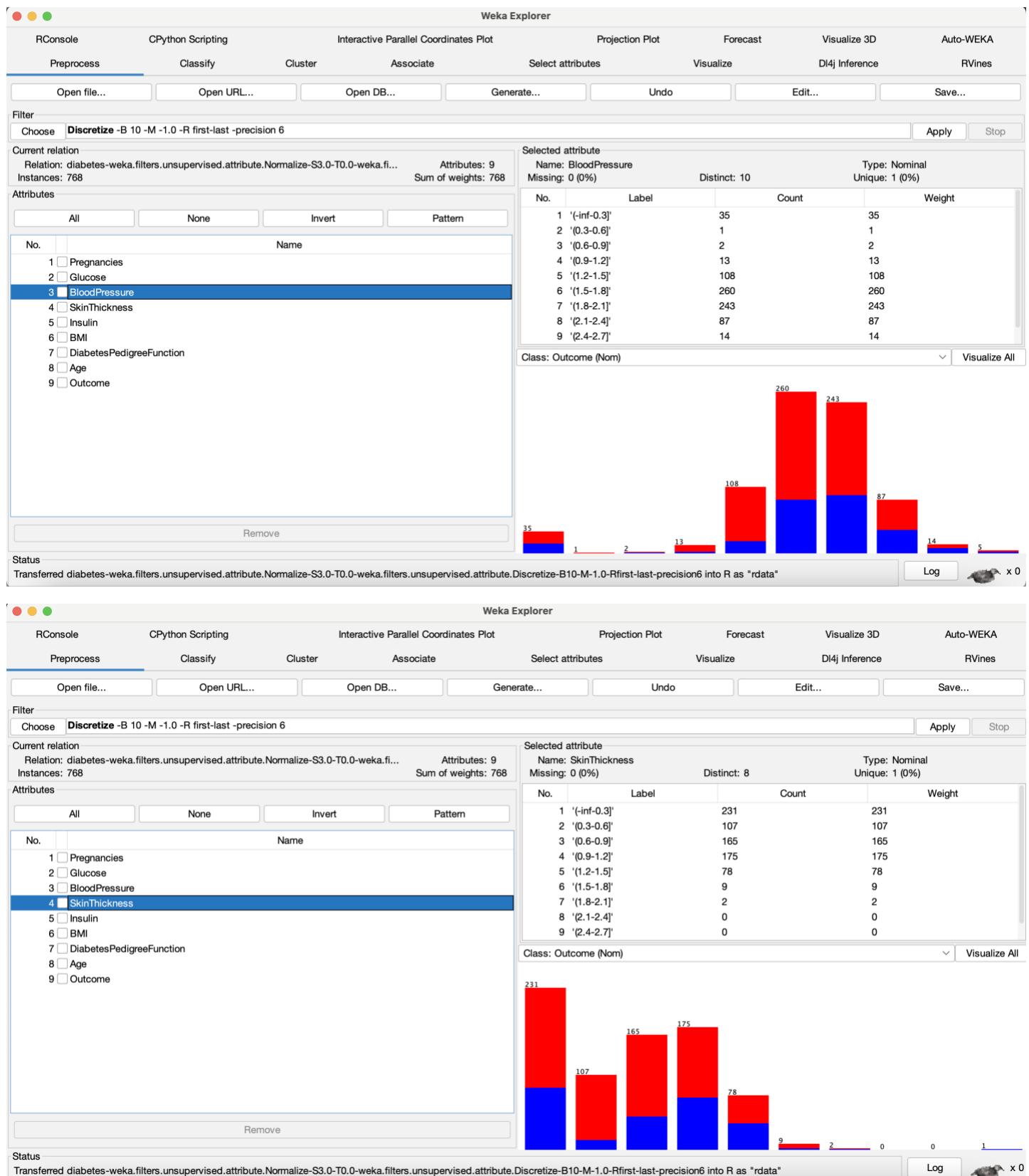
CYCLE SHEET-I

Prashanth.S 19MID0020



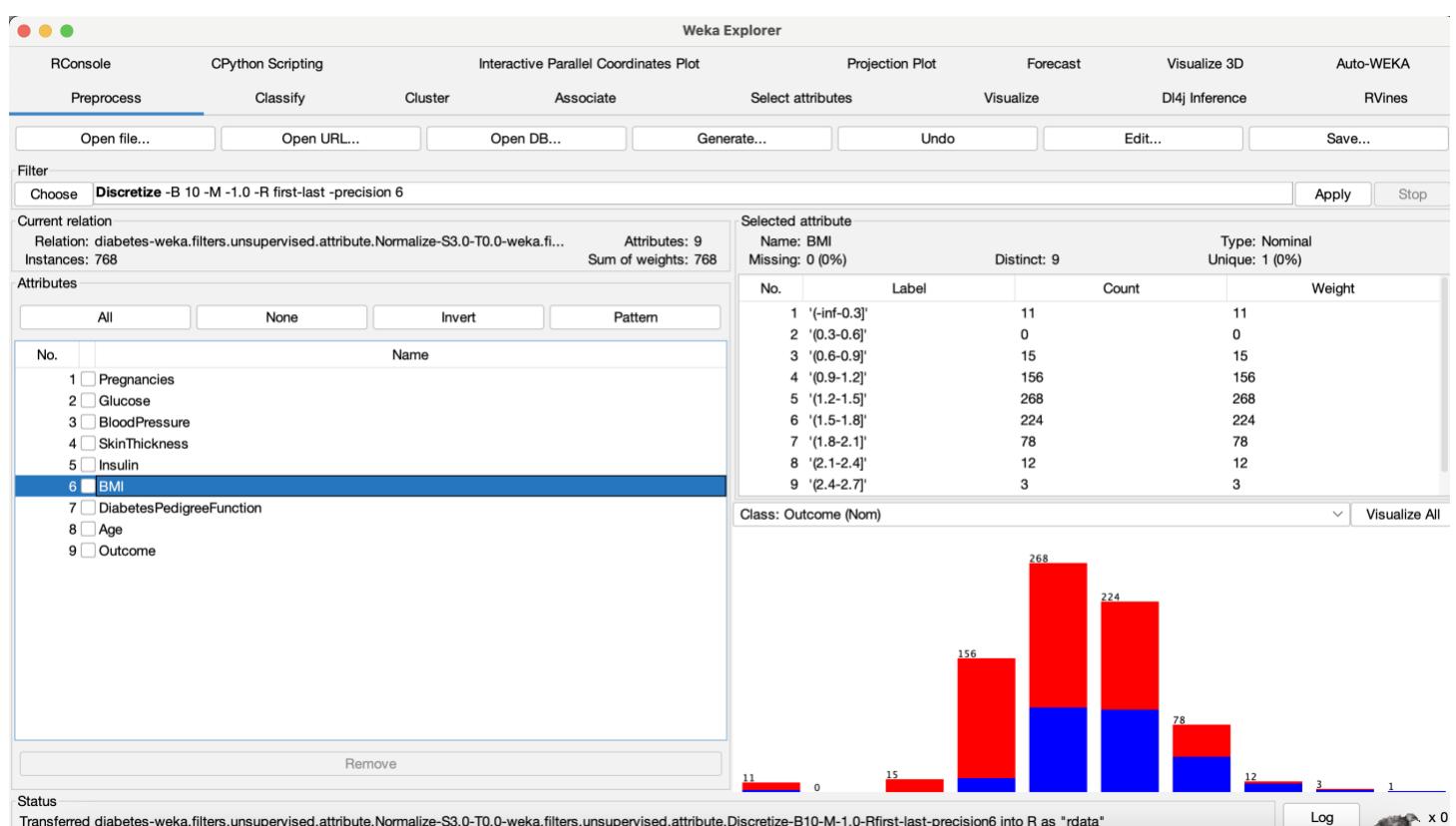
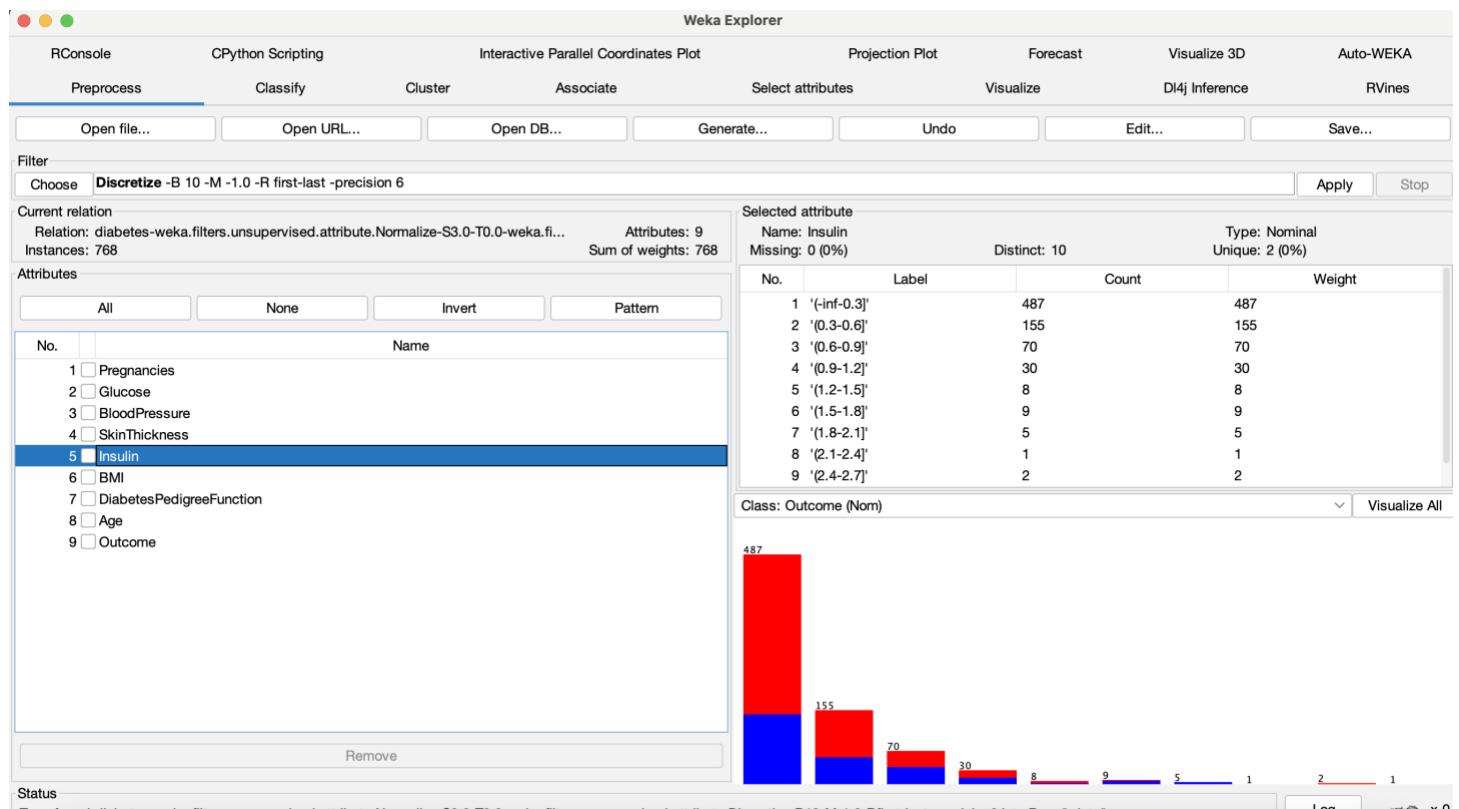
CYCLE SHEET-I

Prashanth.S 19MID0020



CYCLE SHEET-I

Prashanth.S 19MID0020



CYCLE SHEET-I

Prashanth.S 19MID0020

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize Forecast Projection Plot DI4j Inference RVines

Open file... Open URL... Open DB... Generate... Undo Edit... Save... Filter Choose Discretize -B 10 -M 1.0 -R first-last -precision 6 Apply Stop

Current relation Relation: diabetes-weka.filters.unsupervised.attribute.Normalize-S3.0-T0.0-weka.filters.unsupervised.attribute.Discretize-B10-M-1.0-Rfirst-last-precision6 Attributes: 9 Instances: 768 Sum of weights: 768

Attributes

No.	Name
1	Pregnancies
2	Glucose
3	BloodPressure
4	SkinThickness
5	Insulin
6	BMI
7	DiabetesPedigreeFunction
8	Age
9	Outcome

Remove

Status Transferred diabetes-weka.filters.unsupervised.attribute.Normalize-S3.0-T0.0-weka.filters.unsupervised.attribute.Discretize-B10-M-1.0-Rfirst-last-precision6 into R as "rdata"

Selected attribute Name: DiabetesPedigreeFunction Missing: 0 (0%) Distinct: 10 Unique: 1 (0%) Type: Nominal

No.	Label	Count	Weight
1	'(-inf-0.3]'	318	318
2	'(0.3-0.6]'	206	206
3	'(0.6-0.9]'	136	136
4	'(0.9-1.2]'	58	58
5	'(1.2-1.5]'	25	25
6	'(1.5-1.8]'	15	15
7	'(1.8-2.1]'	3	3
8	'(2.1-2.4]'	3	3
9	'(2.4-2.7]'	1	1

Class: Outcome (Nom) Visualize All

Age Group	Count
'(-inf-0.3]'	318
'(0.3-0.6]'	206
'(0.6-0.9]'	136
'(0.9-1.2]'	58
'(1.2-1.5]'	25
'(1.5-1.8]'	15
'(1.8-2.1]'	3
'(2.1-2.4]'	3
'(2.4-2.7]'	1

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize Forecast Projection Plot DI4j Inference RVines

Open file... Open URL... Open DB... Generate... Undo Edit... Save... Filter Choose Discretize -B 10 -M 1.0 -R first-last -precision 6 Apply Stop

Current relation Relation: diabetes-weka.filters.unsupervised.attribute.Normalize-S3.0-T0.0-weka.filters.unsupervised.attribute.Discretize-B10-M-1.0-Rfirst-last-precision6 Attributes: 9 Instances: 768 Sum of weights: 768

Attributes

No.	Name
1	Pregnancies
2	Glucose
3	BloodPressure
4	SkinThickness
5	Insulin
6	BMI
7	DiabetesPedigreeFunction
8	Age
9	Outcome

Remove

Status Transferred diabetes-weka.filters.unsupervised.attribute.Normalize-S3.0-T0.0-weka.filters.unsupervised.attribute.Discretize-B10-M-1.0-Rfirst-last-precision6 into R as "rdata"

Selected attribute Name: Age Missing: 0 (0%) Distinct: 10 Unique: 1 (0%) Type: Nominal

No.	Label	Count	Weight
1	'(-inf-0.3]'	300	300
2	'(0.3-0.6]'	157	157
3	'(0.6-0.9]'	104	104
4	'(0.9-1.2]'	74	74
5	'(1.2-1.5]'	60	60
6	'(1.5-1.8]'	31	31
7	'(1.8-2.1]'	25	25
8	'(2.1-2.4]'	12	12
9	'(2.4-2.7]'	4	4

Class: Outcome (Nom) Visualize All

Age Group	Count
'(-inf-0.3]'	300
'(0.3-0.6]'	157
'(0.6-0.9]'	104
'(0.9-1.2]'	74
'(1.2-1.5]'	60
'(1.5-1.8]'	31
'(1.8-2.1]'	25
'(2.1-2.4]'	12
'(2.4-2.7]'	4

CYCLE SHEET-I

Prashanth.S 19MID0020

After Normalization and Discretization

```
17:40:09 - trees.J48

== Run information ==

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: diabetes-weka.filters.unsupervised.attribute.Normalize-S3.0-T0.0-weka.filters.unsupervised.attribute.Discretize-B10-M-1.0-Rfirst-last-precision6
Instances: 768
Attributes: 9
Pregnancies
Glucose
BloodPressure
SkinThickness
Insulin
BMI
DiabetesPedigreeFunction
Age
Outcome
Test mode: 10-fold cross-validation

== Classifier model (full training set) ==

J48 pruned tree
-----
Glucose = '(-inf-0.3]': 0 (5.0/2.0)
Glucose = '(0.3-0.6]': 0 (0.0)
Glucose = '(0.6-0.9]': 0 (4.0)
Glucose = '(0.9-1.2]': 0 (32.0/1.0)
Glucose = '(1.2-1.5]': 0 (156.0/13.0)
Glucose = '(1.5-1.8]': 0 (211.0/54.0)
Glucose = '(1.8-2.1]'
| BMI = '(-inf-0.3]': 1 (2.0/1.0)
| BMI = '(0.3-0.6]': 0 (0.0)
| BMI = '(0.6-0.9]': 0 (1.0)
| BMI = '(0.9-1.2]': 0 (29.0/3.0)
| BMI = '(1.2-1.5]':
|   | BloodPressure = '(-inf-0.3]': 1 (2.0)
|   | BloodPressure = '(0.3-0.6]': 0 (0.0)
|   | BloodPressure = '(0.6-0.9]': 0 (0.0)
|   | BloodPressure = '(0.9-1.2]': 0 (1.0)
|   | BloodPressure = '(1.2-1.5]': 0 (9.0/3.0)
|   | BloodPressure = '(1.5-1.8]':
|   |   | SkinThickness = '(-inf-0.3]': 1 (8.0/3.0)
|   |   | SkinThickness = '(0.3-0.6]': 1 (2.0)
|   |   | SkinThickness = '(0.6-0.9]': 1 (5.0/1.0)
|   |   | SkinThickness = '(0.9-1.2]': 0 (3.0)
|   |   | SkinThickness = '(1.2-1.5]': 0 (1.0)
|   |   | SkinThickness = '(1.5-1.8]': 1 (0.0)
|   |   | SkinThickness = '(1.8-2.1]': 1 (0.0)
|   |   | SkinThickness = '(2.1-2.4]': 1 (0.0)
|   |   | SkinThickness = '(2.4-2.7]': 1 (0.0)
|   |   | SkinThickness = '(2.4-2.7]': 1 (0.0)
|   |   | SkinThickness = '(2.7-inf)': 1 (0.0)
|   | BloodPressure = '(1.8-2.1]': 0 (22.0/6.0)
|   | BloodPressure = '(2.1-2.4]': 0 (7.0/2.0)
|   | BloodPressure = '(2.4-2.7]': 1 (2.0)
|   | BloodPressure = '(2.7-inf)': 0 (0.0)
| BMI = '(1.5-1.8]': 0 (51.0/20.0)
| BMI = '(1.8-2.1]': 1 (13.0/4.0)
| BMI = '(2.1-2.4]': 1 (3.0)
| BMI = '(2.4-2.7]': 0 (1.0)
| BMI = '(2.7-inf)': 1 (1.0)
Glucose = '(2.1-2.4]'
| BloodPressure = '(-inf-0.3]': 1 (4.0)
| BloodPressure = '(0.3-0.6]': 1 (0.0)
| BloodPressure = '(0.6-0.9]': 1 (0.0)
| BloodPressure = '(0.9-1.2]': 1 (0.0)
| BloodPressure = '(1.2-1.5]': 0 (8.0/3.0)
| BloodPressure = '(1.5-1.8]':
|   | DiabetesPedigreeFunction = '(-inf-0.3]': 0 (8.0/1.0)
|   | DiabetesPedigreeFunction = '(0.3-0.6]': 1 (8.0/1.0)
|   | DiabetesPedigreeFunction = '(0.6-0.9]': 1 (4.0/2.0)
|   | DiabetesPedigreeFunction = '(0.9-1.2]': 1 (0.0)
|   | DiabetesPedigreeFunction = '(1.2-1.5]': 1 (0.0)
|   | DiabetesPedigreeFunction = '(1.5-1.8]': 1 (0.0)
|   | DiabetesPedigreeFunction = '(1.8-2.1]': 1 (0.0)
|   | DiabetesPedigreeFunction = '(2.1-2.4]': 1 (0.0)
|   | DiabetesPedigreeFunction = '(2.4-2.7]': 1 (0.0)
|   | DiabetesPedigreeFunction = '(2.7-inf)': 1 (0.0)
| BloodPressure = '(1.8-2.1]': 0 (46.0/20.0)
| BloodPressure = '(2.1-2.4]': 1 (16.0/3.0)
| BloodPressure = '(2.4-2.7]': 1 (0.0)
| BloodPressure = '(2.7-inf)': 1 (1.0)
Glucose = '(2.4-2.7]': 1 (56.0/11.0)
Glucose = '(2.7-inf)': 1 (46.0/7.0)

Number of Leaves : 55
Size of the tree : 61

Time taken to build model: 0 seconds
```

CYCLE SHEET-I

Prashanth.S 19MID0020

```
== Stratified cross-validation ==
== Summary ==

Correctly Classified Instances      561          73.0469 %
Incorrectly Classified Instances   207          26.9531 %
Kappa statistic                      0.3512
Mean absolute error                  0.341
Root mean squared error              0.4348
Relative absolute error              75.0371 %
Root relative squared error         91.2259 %
Total Number of Instances           768

== Detailed Accuracy By Class ==

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC  ROC Area  PRC Area  Class
0  0.429    0.108    0.680     0.429   0.526    0.369   0.741    0.576     1
1  0.892    0.571    0.745     0.892   0.812    0.369   0.741    0.810     0
Weighted Avg.                     0.730    0.409    0.722     0.730   0.712    0.369   0.741    0.729

== Confusion Matrix ==

 a   b   <-- classified as
115 153 |   a = 1
 54 446 |   b = 0
```

8) Apply J48 with default setting and determine the accuracy before and after data pre-processing.

Before Discretization

The screenshot shows the Weka interface with the J48 classifier selected. The 'Classify' tab is active. The 'Classifier output' pane displays the decision tree structure and statistics:

```

Classifier output
=====
|   |
|   | Age <= 30: 0 (40.0/15.0)
|   | Age > 30: 1 (60.0/17.0)
|   |
|   | Glucose > 157: 1 (92.0/12.0)

Number of Leaves : 20
Size of the tree : 39

Time taken to build model: 0.03 seconds

```

The 'Summary' section shows the following performance metrics:

```

== Stratified cross-validation ==
== Summary ==

Correctly Classified Instances      567          73.8281 %
Incorrectly Classified Instances   201          26.1719 %
Kappa statistic                      0.4164
Mean absolute error                  0.3158
Root mean squared error              0.4463
Relative absolute error              69.4841 %
Root relative squared error         93.6293 %
Total Number of Instances           768

```

The 'Detailed Accuracy By Class' section provides the following data:

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.597	0.186	0.632	0.597	0.614	0.417	0.751	0.572	1
1	0.814	0.483	0.790	0.814	0.802	0.417	0.751	0.811	0
Weighted Avg.	0.738	0.327	0.735	0.738	0.736	0.417	0.751	0.727	

The 'Confusion Matrix' is as follows:

```

a   b   <-- classified as
160 108 |   a = 1
 93 407 |   b = 0

```

The status bar at the bottom indicates 'OK'.

CYCLE SHEET-I

Prashanth.S 19MID0020

After Discretization

```
== Stratified cross-validation ==
== Summary ==
Correctly Classified Instances      561          73.0469 %
Incorrectly Classified Instances    207          26.9531 %
Kappa statistic                      0.3512
Mean absolute error                  0.341
Root mean squared error              0.4348
Relative absolute error              75.0371 %
Root relative squared error         91.2259 %
Total Number of Instances            768

== Detailed Accuracy By Class ==
      TP Rate   FP Rate   Precision   Recall   F-Measure   MCC   ROC Area   PRC Area   Class
           0.429     0.108     0.680     0.429     0.526     0.369     0.741     0.576       1
           0.892     0.571     0.745     0.892     0.812     0.369     0.741     0.810       0
Weighted Avg.      0.730     0.409     0.722     0.730     0.712     0.369     0.741     0.729

== Confusion Matrix ==
      a   b  <-- classified as
115 153 |  a = 1
 54 446 |  b = 0
```

CYCLE SHEET-I

Prashanth.S 19MID0020

Classifiers with the data-sets

academicrecords.arff file

```
academicrecords.arff
1 @relation academicrecords
2
3 @attribute Coursecode {CSI2006, CSI3002, CSI3003,CSI3005,CSI3010,CSI3017,HUM1002,ESP1001,STS3022}
4
5 @attribute CAT1 numeric
6 @attribute CAT2 numeric
7 @attribute FAT Integer
8 @attribute GRADE {S,A,B,C,D,F}
9
10 @data
11 CSI2006, 80, 70, 90, A
12 CSI3002, 60, 50, 89, C
13 CSI3003, 90, 20, 54, F
14 CSI3005, 80, 65, 70, C
15 CSI3010, 89, 93, 70, B
16 CSI3017, 40, 50, 52, F
17 HUM1002, 80, 85, 92, S
18 ESP1001, 60, 90, 89, A
19 STS3022, 70, 70, 85, B
20
```

Viewer					
No.	1: Coursecode Nominal	2: CAT1 Numeric	3: CAT2 Numeric	4: FAT Numeric	5: GRADE Nominal
1	CSI2006	80.0	70.0	90.0	A
2	CSI3002	60.0	50.0	89.0	C
3	CSI3003	90.0	20.0	54.0	F
4	CSI3005	80.0	65.0	70.0	C
5	CSI3010	89.0	93.0	70.0	B
6	CSI3017	40.0	50.0	52.0	F
7	HUM1002	80.0	85.0	92.0	S
8	ESP1001	60.0	90.0	89.0	A
9	STS3022	70.0	70.0	85.0	B

Successfully imported academicrecords.arff file

CYCLE SHEET-I

Prashanth.S 19MID0020

Decision Tree

Weka Explorer

RConsole	CPython Scripting	Interactive Parallel Coordinates Plot	Projection Plot	Forecast	Visualize 3D	Auto-WEKA
Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	Di4j Inference
Classifier						
<pre>> timeseries -> trees -> J48</pre>						
<pre>Time taken to test model on test split: 0 seconds Summary === Classified Instances 0 0 % Correctly Classified Instances 2 100 % Pct. of correctly classified instances 0.3333 Relative absolute error 0.3333 Root relative squared error 0.5501 Absolute error 113.0435 % Root mean squared error 137.2159 % Number of Instances 2 Detailed Accuracy By Class ===</pre>						
<pre>Class for generating a pruned or unpruned C4 J48 JC JC LA LM NB NB Op PC Ra 5 decision tree. For more information, see Ross Quinlan (1993). C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo, CA.</pre>						
CAPABILITIES <p>Class -- Binary class, Missing class values, Nominal class</p> <p>Attributes -- Binary attributes, Date attributes, Empty nominal attributes, Missing values, Nominal attributes, Numeric attributes, Unary attributes</p> <p>Interfaces -- Drawable, PartitionGenerator, Sourcable, WeightedInstancesHandler</p>						
<p>Additional</p> <p>Minimum number of instances: 0</p>						

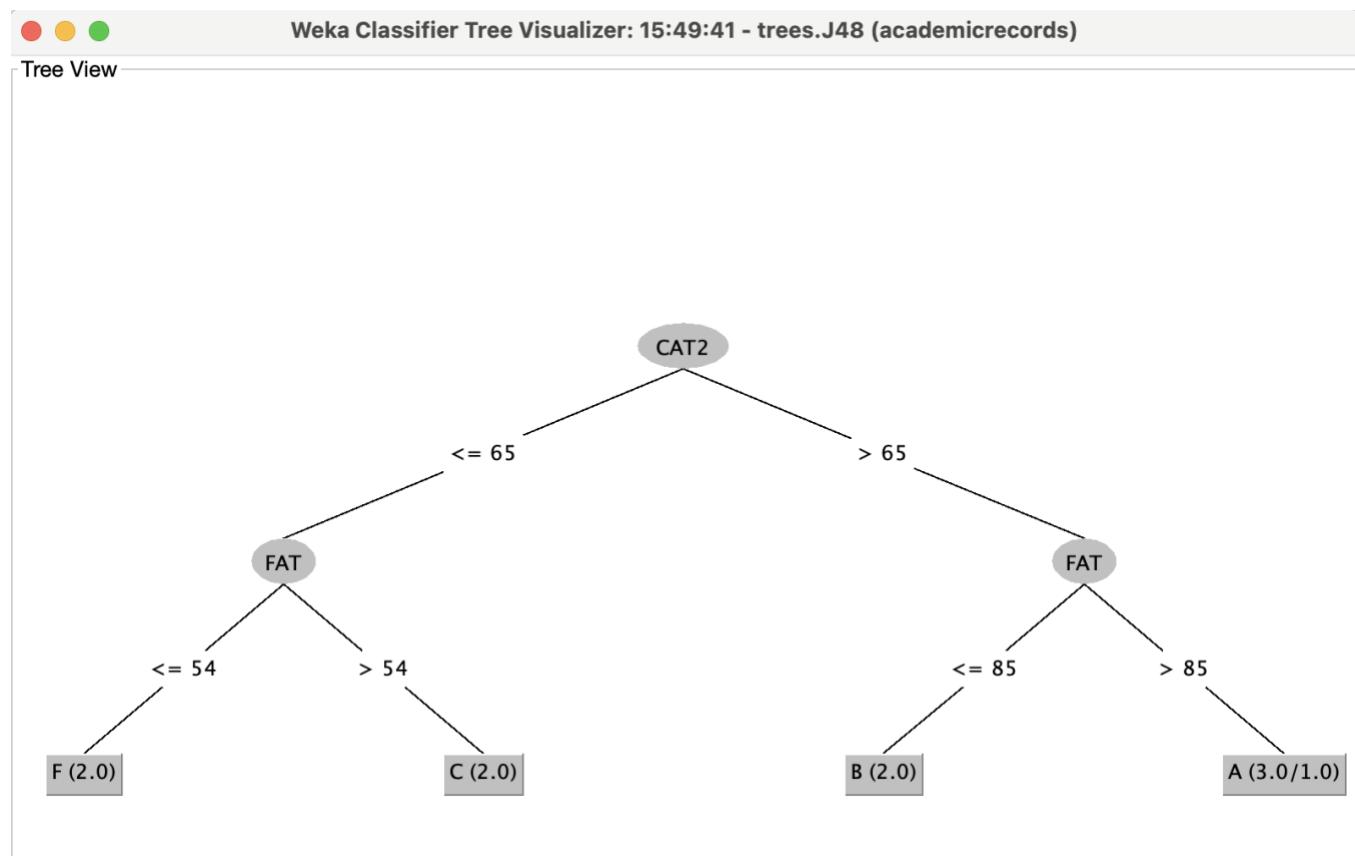
Status OK

Weka Explorer

RConsole	CPython Scripting	Interactive Parallel Coordinates Plot	Projection Plot	Forecast	Visualize 3D	Auto-WEKA																																																																																																																							
Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	Di4j Inference																																																																																																																							
Classifier																																																																																																																													
<input type="button" value="Choose"/> J48-C 0.25 -M 2																																																																																																																													
<p>Test options</p> <p><input type="radio"/> Use training set</p> <p><input type="radio"/> Supplied test set <input type="button" value="Set..."/></p> <p><input checked="" type="radio"/> Cross-validation Folds 2</p> <p><input type="radio"/> Percentage split % 80</p> <p><input type="button" value="More options..."/></p> <p>(Nom) GRADE</p> <p><input type="button" value="Start"/> <input type="button" value="Stop"/> <input type="button" value="Run on server"/></p> <p>Result list (right-click for options)</p> <p>15:49:41 - trees.J48</p>																																																																																																																													
<p>Classifier output</p> <p>Time taken to build model: 0 seconds</p> <p>==== Stratified cross-validation ==== ==== Summary ==== </p> <table border="1"> <thead> <tr> <th></th> <th>Correctly Classified Instances</th> <th>1</th> <th>11.1111 %</th> </tr> </thead> <tbody> <tr> <td>Incorrectly Classified Instances</td> <td>8</td> <td>88.8889 %</td> </tr> <tr> <td>Kappa statistic</td> <td>-0.0746</td> <td></td> </tr> <tr> <td>Mean absolute error</td> <td>0.2778</td> <td></td> </tr> <tr> <td>Root mean squared error</td> <td>0.441</td> <td></td> </tr> <tr> <td>Relative absolute error</td> <td>101.7263 %</td> <td></td> </tr> <tr> <td>Root relative squared error</td> <td>119.6069 %</td> <td></td> </tr> <tr> <td>Total Number of Instances</td> <td>9</td> <td></td> </tr> </tbody> </table> <p>==== Detailed Accuracy By Class ==== </p> <table border="1"> <thead> <tr> <th></th> <th>TP Rate</th> <th>FP Rate</th> <th>Precision</th> <th>Recall</th> <th>F-Measure</th> <th>MCC</th> <th>ROC Area</th> <th>PRC Area</th> <th>Class</th> </tr> </thead> <tbody> <tr> <td>0.000</td> <td>0.500</td> <td>0.000</td> <td>0.000</td> <td>0.000</td> <td>-0.316</td> <td>-0.060</td> <td>0.250</td> <td>0.111</td> <td>S</td> </tr> <tr> <td>0.500</td> <td>0.571</td> <td>0.200</td> <td>0.500</td> <td>0.286</td> <td>-0.060</td> <td>0.464</td> <td>0.211</td> <td>0.211</td> <td>A</td> </tr> <tr> <td>0.000</td> <td>0.000</td> <td>?</td> <td>0.000</td> <td>?</td> <td>?</td> <td>?</td> <td>0.536</td> <td>0.236</td> <td>B</td> </tr> <tr> <td>0.000</td> <td>0.000</td> <td>?</td> <td>0.000</td> <td>?</td> <td>?</td> <td>?</td> <td>0.464</td> <td>0.211</td> <td>C</td> </tr> <tr> <td>?</td> <td>0.000</td> <td>?</td> <td>?</td> <td>?</td> <td>?</td> <td>?</td> <td>?</td> <td>?</td> <td>D</td> </tr> <tr> <td>0.000</td> <td>0.000</td> <td>?</td> <td>0.000</td> <td>?</td> <td>?</td> <td>?</td> <td>0.500</td> <td>0.222</td> <td>F</td> </tr> <tr> <td>Weighted Avg.</td> <td>0.111</td> <td>0.183</td> <td>?</td> <td>0.111</td> <td>?</td> <td>?</td> <td>0.464</td> <td>0.208</td> <td></td> </tr> </tbody> </table> <p>==== Confusion Matrix ==== </p> <table border="1"> <thead> <tr> <th>a b c d e f</th> <th>--- classified as</th> </tr> </thead> <tbody> <tr> <td>0 1 0 0 0 0</td> <td> a = S</td> </tr> <tr> <td>1 1 0 0 0 0</td> <td> b = A</td> </tr> <tr> <td>1 1 0 0 0 0</td> <td> c = B</td> </tr> <tr> <td>1 1 0 0 0 0</td> <td> d = C</td> </tr> <tr> <td>0 0 0 0 0 0</td> <td> e = D</td> </tr> <tr> <td>1 1 0 0 0 0</td> <td> f = F</td> </tr> </tbody> </table>								Correctly Classified Instances	1	11.1111 %	Incorrectly Classified Instances	8	88.8889 %	Kappa statistic	-0.0746		Mean absolute error	0.2778		Root mean squared error	0.441		Relative absolute error	101.7263 %		Root relative squared error	119.6069 %		Total Number of Instances	9			TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class	0.000	0.500	0.000	0.000	0.000	-0.316	-0.060	0.250	0.111	S	0.500	0.571	0.200	0.500	0.286	-0.060	0.464	0.211	0.211	A	0.000	0.000	?	0.000	?	?	?	0.536	0.236	B	0.000	0.000	?	0.000	?	?	?	0.464	0.211	C	?	0.000	?	?	?	?	?	?	?	D	0.000	0.000	?	0.000	?	?	?	0.500	0.222	F	Weighted Avg.	0.111	0.183	?	0.111	?	?	0.464	0.208		a b c d e f	--- classified as	0 1 0 0 0 0	a = S	1 1 0 0 0 0	b = A	1 1 0 0 0 0	c = B	1 1 0 0 0 0	d = C	0 0 0 0 0 0	e = D	1 1 0 0 0 0	f = F
	Correctly Classified Instances	1	11.1111 %																																																																																																																										
Incorrectly Classified Instances	8	88.8889 %																																																																																																																											
Kappa statistic	-0.0746																																																																																																																												
Mean absolute error	0.2778																																																																																																																												
Root mean squared error	0.441																																																																																																																												
Relative absolute error	101.7263 %																																																																																																																												
Root relative squared error	119.6069 %																																																																																																																												
Total Number of Instances	9																																																																																																																												
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class																																																																																																																				
0.000	0.500	0.000	0.000	0.000	-0.316	-0.060	0.250	0.111	S																																																																																																																				
0.500	0.571	0.200	0.500	0.286	-0.060	0.464	0.211	0.211	A																																																																																																																				
0.000	0.000	?	0.000	?	?	?	0.536	0.236	B																																																																																																																				
0.000	0.000	?	0.000	?	?	?	0.464	0.211	C																																																																																																																				
?	0.000	?	?	?	?	?	?	?	D																																																																																																																				
0.000	0.000	?	0.000	?	?	?	0.500	0.222	F																																																																																																																				
Weighted Avg.	0.111	0.183	?	0.111	?	?	0.464	0.208																																																																																																																					
a b c d e f	--- classified as																																																																																																																												
0 1 0 0 0 0	a = S																																																																																																																												
1 1 0 0 0 0	b = A																																																																																																																												
1 1 0 0 0 0	c = B																																																																																																																												
1 1 0 0 0 0	d = C																																																																																																																												
0 0 0 0 0 0	e = D																																																																																																																												
1 1 0 0 0 0	f = F																																																																																																																												

CYCLE SHEET-I

Prashanth.S 19MID0020



CYCLE SHEET-I

Prashanth.S 19MID0020

Naïve Bayes

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. In the 'Classifier' tree, 'NaiveBayes' is selected. A tooltip provides detailed information about the classifier, mentioning it uses J48 as the estimator with a precision of 0.25, and it splits the data at 60% for training. It also notes that numeric estimator precision values are chosen based on training data analysis.

Class for a Naïve Bayes classifier using estimator classes

Numeric estimator precision values are chosen based on analysis of the training data. For this reason, the classifier is not an UpdateableClassifier (which in typical usage are initialized with zero training instances) -- if you need the UpdateableClassifier functionality, use the NaiveBayesUpdateable classifier. The NaiveBayesUpdateable classifier will use a default precision of 0.1 for numeric attributes when buildClassifier is called with zero training instances.

For more information on Naïve Bayes classifiers, see George H. John, Pat Langley: Estimating Continuous Distributions in Bayesian Classifiers. In: Eleventh Conference on Uncertainty in Artificial Intelligence, San Mateo, 338-345, 1995.

CAPABILITIES

Class -- Binary class, Missing class values, Nominal class

Attributes -- Binary attributes, Empty nominal attributes, Missing values, Nominal attributes, Numeric attributes, Unary attributes

Interfaces -- WeightedAttributesHandler, WeightedInstancesHandler

Additional
Minimum number of instances: 0

The screenshot shows the Weka interface with the 'Classify' tab selected. The 'Classifier' dropdown is set to 'Choose' and 'NaiveBayes'. The 'Test options' section shows 'Cross-validation' selected with 2 folds. The 'Result list' pane shows the command '15:50:42 - bayes.NaiveBayes' being run. The 'Classifier output' pane displays the results of the classification, including:

- Time taken to build model: 0 seconds
- Stratified cross-validation summary
- Classification statistics: 33.3333 % correctly classified instances, 66.6667 % incorrectly classified instances.
- Detailed accuracy by class table:

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.000	0.250	0.000	0.000	0.000	-0.189	0.000	0.111	0.268	S
0.500	0.429	0.250	0.500	0.333	0.060	0.429	0.292	0.489	A
0.000	0.143	0.000	0.000	0.000	-0.189	0.000	0.174	0.248	B
0.000	0.000	?	0.000	?	?	0.429	0.292	0.489	C
?	0.000	?	?	?	?	?	?	?	D
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	F
Weighted Avg.	0.333	0.155	?	0.333	?	0.413	0.397		

- Confusion matrix table:

a b c d e f	<-- classified as						
0 1 0 0 0 0		a = S					
1 1 0 0 0 0		b = A					
1 1 0 0 0 0		c = B					
0 1 1 0 0 0		d = C					
0 0 0 0 0 0		e = D					
0 0 0 0 0 2		f = F					

CYCLE SHEET-I

Prashanth.S 19MID0020

K-Nearest Neighbor

Weka Explorer

RConsole	Cython Scripting	Preprocess	Classify	Cluster	Associate	Select attributes	Projection Plot	Forecast	Visualize 3D	Di4j Inference	Auto-WEKA
			Classify								
Classifier											
weka											
classifiers											
> bayes											
> clojure											
> functions											
> keras											
lazy											
> AnalogicalModeling											
> IB1											
IBk											
> IB											
K-nearest neighbours classifier											
> KNN											
> LDA											
> LV											
> RBF											
> RBFNB											
> meta											
> mi											
> misc											
> mirel											
> pycosat											
> rules											
> scripts											
> sklean											
> times											
> trees											

Output

```
Time taken to build model: 0 seconds
Stratified cross-validation ===
IBk ==

    Classified Instances      1      11.111 %
    Newly Classified Instances 8      88.889 %
    Predictive accuracy        -0.0746
    Percentage Error          0.2778
    Root Absolute Error       0.441
```

Can select appropriate value of K based on cross-validation. Can also do distance weighting.

For more information, see

D. Aha, D. Kibler (1991). Instance-based learning algorithms. Machine Learning. 6:37-66.

CAPABILITIES

Class -- Binary class, Date class, Missing class values, Nominal class, Numeric class

Attributes -- Binary attributes, Date attributes, Empty nominal attributes, Missing values, Nominal attributes, Numeric attributes, Unary attributes

Interfaces -- UpdateableClassifier, WeightedInstancesHandler

Additional

Minimum number of instances: 0

Close Class
S S
A A
B B
C C
D D
F F

Weka Explorer

RConsole	Cython Scripting	Preprocess	Classify	Cluster	Associate	Select attributes	Projection Plot	Forecast	Visualize 3D	Di4j Inference	Auto-WEKA
			Classify								
Classifier											
Choose: IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last"											
Test options											
<input type="radio"/> Use training set											
<input type="radio"/> Supplied test set <input type="button" value="Set..."/>											
<input checked="" type="radio"/> Cross-validation <input type="button" value="Folds"/> 2											
<input type="radio"/> Percentage split % 80											
<input type="button" value="More options..."/>											
(Nom) GRADE											
<input type="button" value="Start"/> <input type="button" value="Stop"/> <input type="button" value="Run on server"/>											
Result list (right-click for options)											
15:49:41 - trees.J48											
15:50:42 - bayes.NaiveBayes											
15:54:38 - lazy.IBk											

Classifier output

```
Time taken to build model: 0 seconds
Stratified cross-validation ===
Summary ===

    Correctly Classified Instances      2      22.222 %
    Incorrectly Classified Instances   7      77.7778 %
    Kappa statistic                   0.0308
    Mean absolute error              0.2697
    Root mean squared error          0.3949
    Relative absolute error          98.767 %
    Root relative squared error     107.1007 %
    Total Number of Instances        9
```

Detailed Accuracy By Class

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.000	0.250	0.000	0.000	0.000	-0.189	0.500	0.143	S	
0.500	0.429	0.250	0.500	0.333	0.060	0.500	0.236	A	
0.000	0.143	0.000	0.000	0.000	-0.189	0.357	0.194	B	
0.000	0.143	0.000	0.000	0.000	-0.189	0.429	0.211	C	
?	0.000	?	?	?	?	?	?	D	
0.500	0.000	1.000	0.500	0.667	0.661	0.857	0.667	F	
Weighted Avg.	0.222	0.187	0.278	0.222	0.222	0.055	0.532	0.307	

Confusion Matrix

```
a b c d e f  <- classified as
0 1 0 0 0 0 | a = S
1 1 0 0 0 0 | b = A
1 1 0 0 0 0 | c = B
0 1 1 0 0 0 | d = C
0 0 0 0 0 0 | e = D
0 0 0 1 0 1 | f = F
```

CYCLE SHEET-I

Prashanth.S 19MID0020

Accuracy Report

Classification Models Accuracy	
Decision Tree	11.11%
Naïve Bayes	33.33%
K-Nearest Neighbor	22.22%

All these test were under-taken at Cross folds=2 and with 80% training data-set.

Linear Regression

Since our Target attribute is Grade and it is Nominal, removing the Grade attribute and considering the FAT marks which is numeric as Target attribute.

The screenshot shows the Weka Explorer interface. The 'Classify' tab is selected. In the left panel, under the 'Classifier' category, the 'LinearRegression' class is highlighted. The right panel displays the configuration for the 'LinearRegression' classifier. The command entered is:

```
weka.classifiers.functions.LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4  
academicrecords-weka.filters.unsupervised.attribute.Remove-R5
```

The output pane shows the execution results:

```
: 9  
$: 4  
Coursecode  
CAT1  
CAT2  
FAT  
: 2-fold cross-validation  
ifier model (full training set) ===  
  
gression Model
```

A detailed description of the classifier is provided in a modal window:

- Class for using linear regression for prediction**
- Uses the Akaike criterion for model selection, and is able to deal with weighted instances.
- CAPABILITIES**
 - Class -- Date class, Missing class values, Numeric class
 - Attributes -- Binary attributes, Date attributes, Empty nominal attributes, Missing values, Nominal attributes, Numeric attributes, Unary attributes
 - Interfaces -- WeightedInstancesHandler
- Additional**
 - Minimum number of instances: 1

At the bottom, performance metrics are shown:

Relative absolute error	98.9522 %
Root relative squared error	95.2147 %

Status: OK

CYCLE SHEET-I

Prashanth.S 19MID0020

Weka Explorer

RConsole CPython Scripting Interactive Parallel Coordinates Plot Projection Plot Forecast Visualize 3D Auto-WEKA

Preprocess Classify Cluster Associate Select attributes Visualize Dl4j Inference RVines

Classifier Choose **LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4**

Test options

Use training set
 Supplied test set Set...
 Cross-validation Folds 2
 Percentage split % 80
More options...

(Num) FAT

Start Stop Run on server

Result list (right-click for options)

15:49:41 - trees.J48
15:50:42 - bayes.NaiveBayes
15:54:38 - lazy.IBk
16:05:49 - lazy.IBk
16:52:57 - functions.LinearRegression

Classifier output

```
Scheme: weka.classifiers.functions.LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4
Relation: academicrecords-weka.filters.unsupervised.attribute.Remove-R5
Instances: 9
Attributes: 4
Coursecode
CAT1
CAT2
FAT
Test mode: 2-fold cross-validation
== Classifier model (full training set) ==
Linear Regression Model
FAT =
+
76.7778
Time taken to build model: 0.05 seconds
== Cross-validation ==
== Summary ==
Correlation coefficient -0.4378
Kendall's tau -0.4002
Mean absolute logarithmic error 0.2136
Mean absolute percentage error 0.227
Root mean square logarithmic error 0.259
Root mean square percentage error 0.29
Spearman's rho -0.5126
Mean absolute error 15.9148
Root mean squared error 18.9736
Relative absolute error 98.9522 %
Root relative squared error 95.2147 %
```

Status OK Log x 0