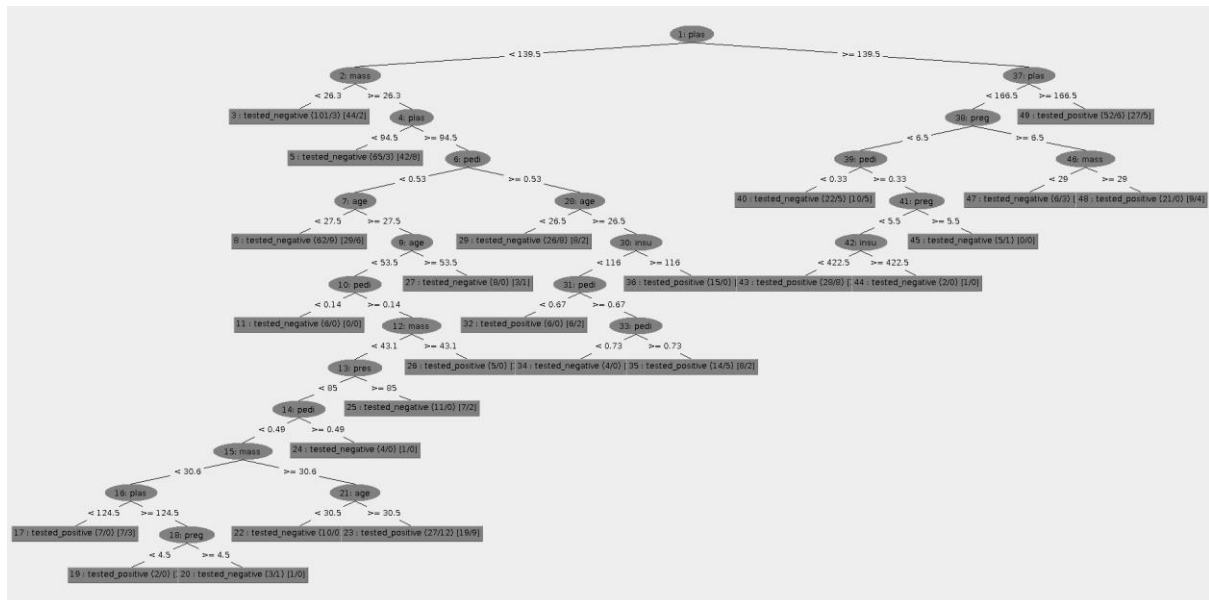


Roll No: 106120043

Name: I Viswanath

Considering the diabetes.arff dataset,  
REP Tree with 10 cross-validation folds



=== Run information ===

Scheme:weka.classifiers.trees.REPTree -M 2 -V 0.001 -N 3 -S 1 -L -1

Relation: pima\_diabetes

Instances: 768

Attributes: 9

preg

plas

pres

skin

insu

mass

pedi

age

class

Test mode:10-fold cross-validation

=== Classifier model (full training set) ===

REPTree

=====

plas < 139.5

| mass < 26.3 : tested\_negative (101/3) [44/2]

| mass >= 26.3

| | plas < 94.5 : tested\_negative (65/3) [42/8]

| | plas >= 94.5

| | | pedi < 0.53

| | | | age < 27.5 : tested\_negative (62/9) [29/6]

| | | | age >= 27.5

| | | | | age < 53.5

| | | | | | pedi < 0.14 : tested\_negative (6/0) [0/0]

| | | | | | pedi >= 0.14

| | | | | | | mass < 43.1

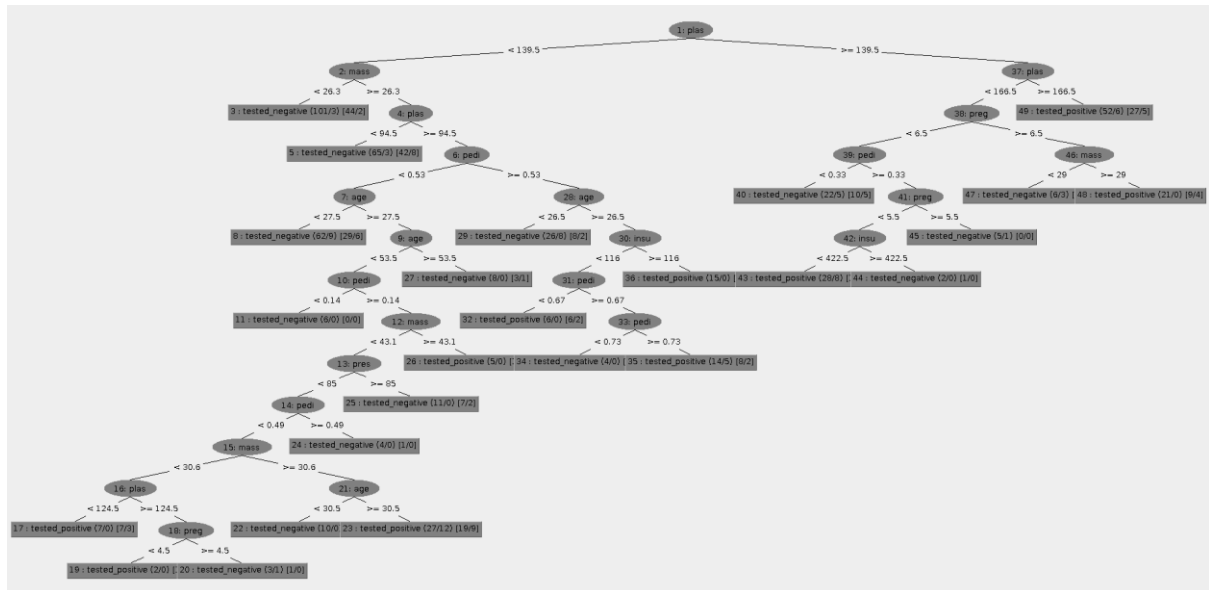
| | | | | | | pres < 85



=== Confusion Matrix ===

a b <-- classified as  
 423 77 | a = tested\_negative  
 113 155 | b = tested\_positive

REP Tree with 20 cross-validation folds



=== Run information ===

Scheme:weka.classifiers.trees.REPTree -M 2 -V 0.001 -N 3 -S 1 -L -1

Relation: pima\_diabetes

Instances: 768

Attributes: 9

preg  
 plas  
 pres  
 skin  
 insu  
 mass  
 pedi  
 age  
 class

Test mode:20-fold cross-validation

=== Classifier model (full training set) ===

REPTree

=====

```

plas < 139.5
| mass < 26.3 : tested_negative (101/3) [44/2]
| mass >= 26.3
| | plas < 94.5 : tested_negative (65/3) [42/8]
| | plas >= 94.5
| | | pedi < 0.53
| | | | age < 27.5 : tested_negative (62/9) [29/6]
| | | | age >= 27.5
| | | | | age < 53.5
| | | | | | pedi < 0.14 : tested_negative (6/0) [0/0]
| | | | | | pedi >= 0.14
| | | | | | | mass < 43.1
| | | | | | | | pres < 85
| | | | | | | | | pedi < 0.49

```



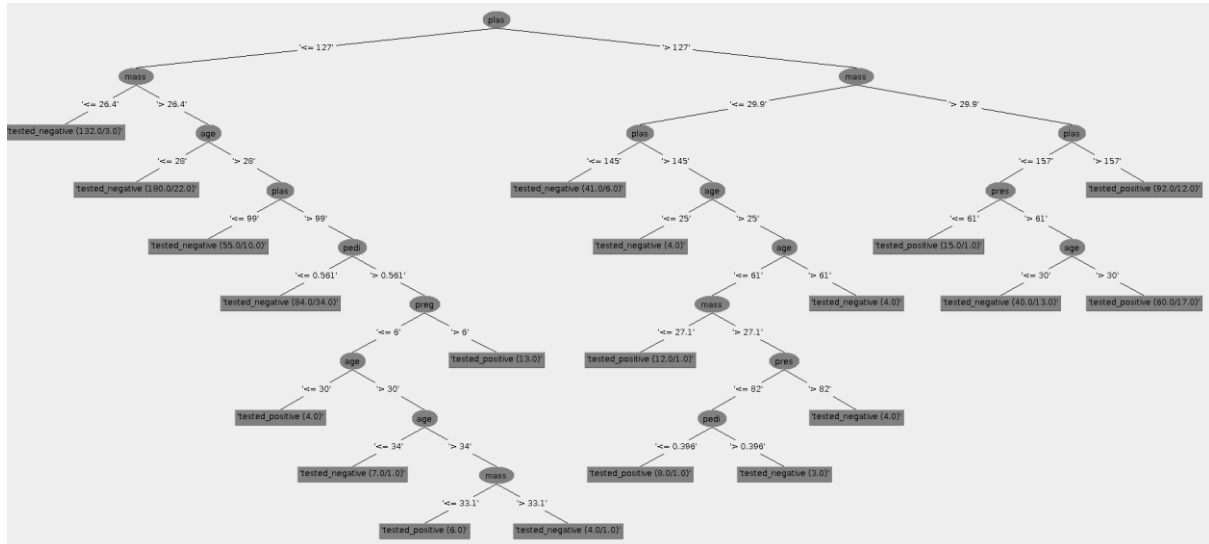
```

a b <-- classified as
411 89 | a = tested_negative
123 145 | b = tested_positive

```

**Both REP Trees produce the same decision tree as the algorithm is the same. The accuracy of classification is better in 10-fold REP tree (75.26%) compared to 20-fold REP tree (73.39%).**

J48 with 10 cross-validation folds



=== Run information ===

Scheme:weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: pima\_diabetes

Instances: 768

Attributes: 9

```

    preg
    plas
    pres
    skin
    insu
    mass
    pedi
    age
    class

```

Test mode:10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree

-----

```

plas <= 127
| mass <= 26.4: tested_negative (132.0/3.0)
| mass > 26.4
| | age <= 28: tested_negative (180.0/22.0)
| | age > 28
| | | plas <= 99: tested_negative (55.0/10.0)
| | | plas > 99
| | | | pedi <= 0.561: tested_negative (84.0/34.0)
| | | | pedi > 0.561
| | | | | preg <= 6
| | | | | age <= 30: tested_positive (4.0)
| | | | | age > 30

```

```

| | | | | | | age <= 34: tested_negative (7.0/1.0)
| | | | | | | age > 34
| | | | | | | mass <= 33.1: tested_positive (6.0)
| | | | | | | mass > 33.1: tested_negative (4.0/1.0)
| | | | | | | preg > 6: tested_positive (13.0)
plas > 127
| mass <= 29.9
| | plas <= 145: tested_negative (41.0/6.0)
| | plas > 145
| | | age <= 25: tested_negative (4.0)
| | | age > 25
| | | | age <= 61
| | | | | mass <= 27.1: tested_positive (12.0/1.0)
| | | | | mass > 27.1
| | | | | | pres <= 82
| | | | | | | pedi <= 0.396: tested_positive (8.0/1.0)
| | | | | | | pedi > 0.396: tested_negative (3.0)
| | | | | | | pres > 82: tested_negative (4.0)
| | | | | | | age > 61: tested_negative (4.0)
| mass > 29.9
| | plas <= 157
| | | pres <= 61: tested_positive (15.0/1.0)
| | | pres > 61
| | | | age <= 30: tested_negative (40.0/13.0)
| | | | age > 30: tested_positive (60.0/17.0)
| | plas > 157: tested_positive (92.0/12.0)

```

Number of Leaves: 20

Size of the tree: 39

Time taken to build model: 0.02 seconds

=== 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	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.814	0.403	0.79	0.814	0.802	0.751	tested_negative
	0.597	0.186	0.632	0.597	0.614	0.751	tested_positive
Weighted Avg.	0.738	0.327	0.735	0.738	0.736	0.751	

=== Confusion Matrix ===

```

a b <-- classified as
407 93 | a = tested_negative
108 160 | b = tested_positive

```

**The J48 tree produces a much simpler tree compared to REP Tree but the accuracy is 73.83% which is lower than the REP Tree with same number of cross-validation folds. The**

accuracy of classification of REP Tree is usually expected to be higher than that of J48, as REP Tree uses reduced-error pruning in addition to the C4.5 algorithm used by J48 Tree.

---