**DEMOGRAPHICS**

**Group\_14**

**Section\_29**

**TEAM MEMBERS**

Bethi, Bharath Reddy 999990725

Roja suryuadhevara 999990511

Venkateswara rao alluri 999990691

Deepthi kota 999990506

Priyanka pappuri 999990648

Contents

[Introduction: 3](#_Toc449890139)

[Dataset: 3](#_Toc449890140)

[Preprocessing : 4](#_Toc449890141)

[Dataset with arff format: 5](#_Toc449890142)

[Selection of the Key Attributes: 6](#_Toc449890143)

[Assumptions: 6](#_Toc449890144)

[Application of Classify Algorithms: 7](#_Toc449890145)

[M5Rules: 8](#_Toc449890146)

[M5Rules Result: 10](#_Toc449890147)

[Improvements introduced by M5Rules: 12](#_Toc449890148)

[Decision Table: 13](#_Toc449890149)

[Classifier Errors: 14](#_Toc449890150)

[Clustering Algorithms: 22](#_Toc449890151)

[Research questions: 27](#_Toc449890152)

[Project Management: 28](#_Toc449890153)

[Conclusion: 29](#_Toc449890154)

[Reference: 30](#_Toc449890155)

# 

# Introduction:

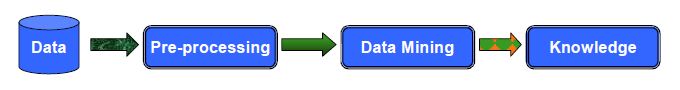
Demographics is an important field of study with important implications for planning and resource management. Changing lifestyle patterns and human migration can change the infrastructure needs of communities, but infrastructure projects by their nature require large investments in both time and capital. We applied data mining techniques to the socio-economic data from the '90 Census, law enforcement data from the 1990 Law Enforcement Management and Admin Stats survey. We predicted that there is a strong correlation between population and mass transit usage. We also predicted that there is a strong correlation between income and population. In other words, income and mass transit usage are both higher in heavily populated areas.

# Dataset:

For finding an appropriate dataset we went through several websites and finally found this dataset on the link. [http://archive.ics.uci.edu/ml/datasets/Communities+and+Crime+Unnormalized#](http://archive.ics.uci.edu/ml/datasets/Communities+and+Crime+Unnormalized).

This dataset is a collection of several attributes based on the US census which predict many interesting things. This is a very large dataset with 147 attributes and 2215 instances. The creator of this dataset is Michael Redmond; Computer Science; La Salle University; Philadelphia, PA, 19141, USA.

# Preprocessing :

****

We converted the dataset to arff format so that our data mining program can read it. In our program we put comma between each pair of attributes, finally our dataset is as shown below

@relation Demographics

@attribute pctUrban numeric

@attribute perCapInc numeric

@attribute pctPoverty numeric

@attribute pctHousOccup numeric

@attribute popDensity numeric

@attribute pctUsePubTrans numeric

@data

pctUrban,0,100,70.46530926,44.08027532,0.086294263,100,100,0

perCapInc,5237,63302,15603.5246,6281.558523,-0.315255977,14101,11252,0

pctPoverty,0.64,58,11.62053725,8.600352277,0.505349223,9.33,3.26,0

pctHousOccup,37.47,99,92.93397291,5.04073584,-0.256836226,94.21,95.38,0

popDensity,10,44229.9,2783.835034,2828.993341,0.256966815,2027.3,3217.7,0

pctUsePubTrans,0,54.33,3.041124153,4.91291686,0.190478991,1.22,0,0

# Dataset with arff format:

At this step our dataset is ready to be given to data mining software for which we chose Weka.

**Attributes and Relations:** We identified thatsome further pre-processing of the dataset must be done for better understanding of the correlation between population, mass transit usage and income. Using Weka's Unsupervised Attribute Remove Filter we removed the unwanted attributes. So the source dataset consists of 16 attributes. Many variables are included so that algorithms that select or learn weights for attributes could be tested. We have to identify the attributes that are not predictive, and would get in the way of some algorithms. However, clearly unrelated attributes were not included. Attributes were picked if there was any plausible connection between them. The attributes included in the dataset are as follows:

**racepctblack:** percentage of population that is african american (numeric - decimal)

**racePctWhite:** percentage of population that is caucasian (numeric - decimal)

**racePctAsian:** percentage of population that is of asian heritage (numeric - decimal)

**racePctHisp:** percentage of population that is of hispanic heritage (numeric - decimal)

**pctUrban:** percentage of people living in areas classified as urban (numeric - decimal)

**pctWWage:** percentage of households with wage or salary income in 1989 (numeric - decimal)

**perCapInc:** per capita income (numeric - decimal)

**PctPopUnderPov**: percentage of people under the poverty level (numeric - decimal)

**PersPerFam**: mean number of people per family (numeric - decimal)

**PersPerOccupHous**: mean persons per household (numeric - decimal)

**PctHousOccup**: percent of housing occupied (numeric - decimal)

**PopDens**: population density in persons per square mile (numeric - decimal)

**PctUsePubTrans**: percent of people using public transit for commuting (numeric - decimal)

# Selection of the Key Attributes:

On closely examining the dataset we identified 3 key attributes. Using these attributes we can predict that income and mass transit usage are both higher in heavily populated areas. The key attributes are as follows:

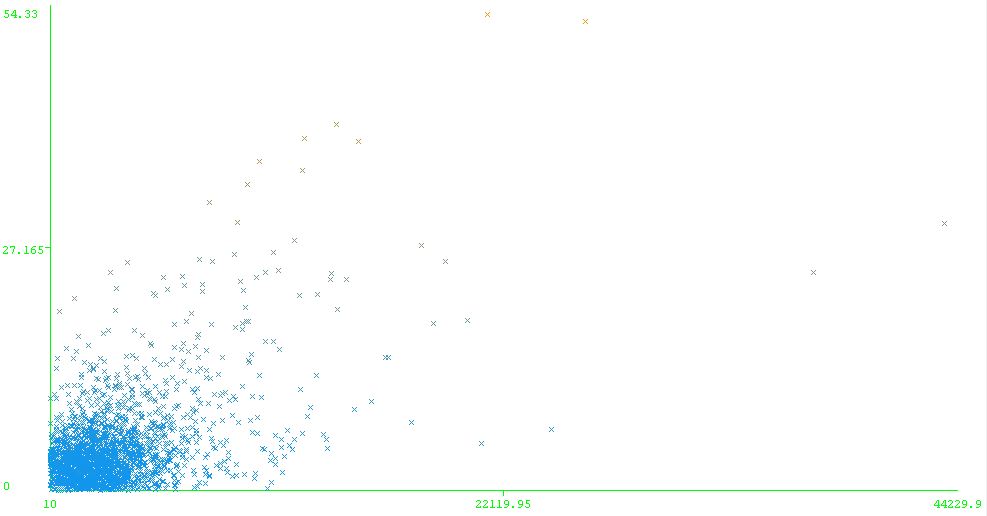
**PctUsePubTrans**: percent of people using public transit for commuting (numeric - decimal)

**perCapInc:** per capita income (numeric - decimal)

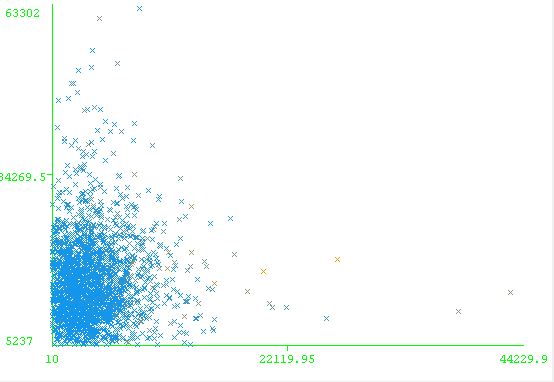
**PopDens**: population density in persons per square mile (numeric - decimal)

# Assumptions:

* We are predicting that the Population Density influences the percentage of people using public transit for commuting in a directly proportional manner taking popDens on the X axis and pctUsePubTrans on Y axis.



* We are predicting that the Population Density influences the per capita income in a directly proportional manner taking popDens on the X axis and perCapInc on Y axis.



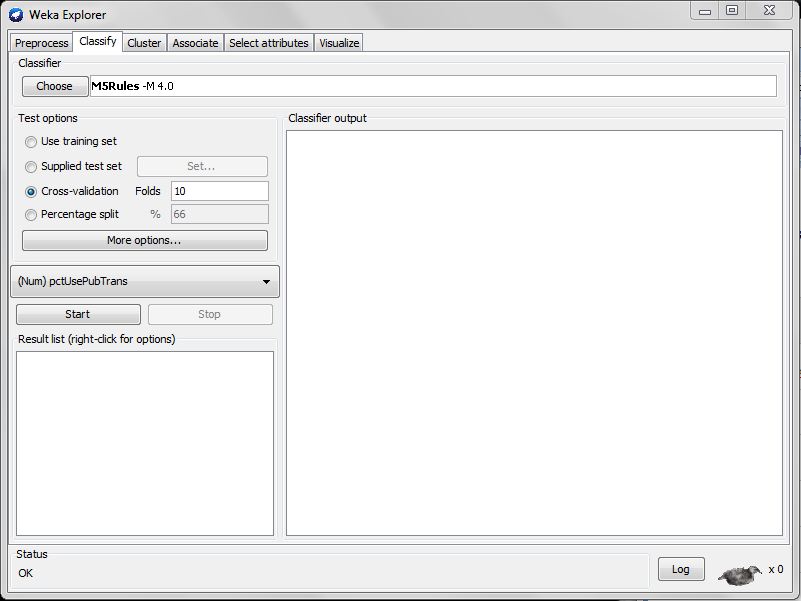
# Application of Classify Algorithms:

We applied different rules for classification such as DecisionTable and M5Rules.

* M5Rules: It generates a decision list for regression problems using separate-and-conquer strategy. In each iteration it builds a model tree using M5 and makes the “best” leaf into a rule. The number to the left of a variable is its coefficient in the linear model.
* Decision Table: It’s a class for building and using a simple decision table majority classifier.

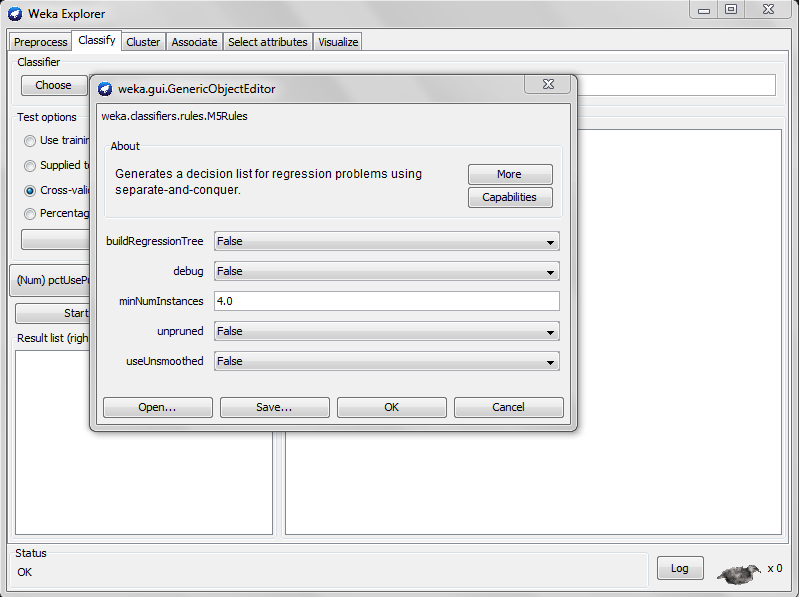
# M5Rules:

How we applied these algorithms to our data set: First we loaded our dataset to the Weka software and chose which type of algorithm we wanted to apply. According to this we selected Classify. Then there will be an option called ‘Choose’ where we found the algorithms. From those we applied two algorithms Decision Table and M5Rules and saw the results for our dataset. For the application of M5Rules we chose Cross Validation(10 folds) in the Test options.

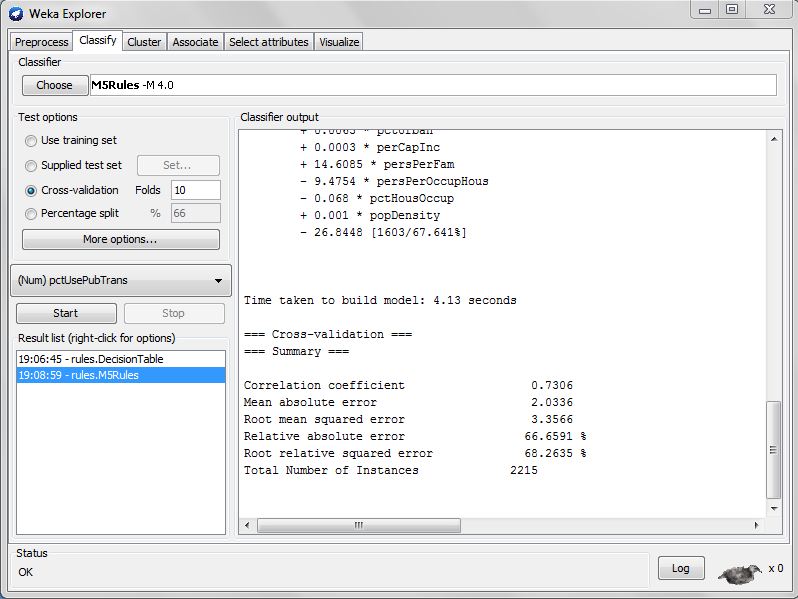


The table below describes the options available for M5Rules.

|  |  |
| --- | --- |
| **Option** | **Description** |
| buildRegressionTree | Whether to generate a regression tree/rule instead of a model tree/rule. |
| debug | If set to true, classifier may output additional info to the console. |
| minNumInstances | The minimum number of instances to allow at a leaf node. |
| unpruned | Whether unpruned tree/rules are to be generated. |
| useUnsmoothed | Whether to use unsmoothed predictions. |



# M5Rules Result:



M5 pruned model rules

(using smoothed linear models) :

Number of Rules : 2

Rule: 1

IF

popDensity <= 3767.25

pctAsian <= 1.555

pctUrban > 68.185

THEN

pctUsePubTrans =

0.0418 \* pctBlack

+ 0.0016 \* pctWhite

+ 0.0123 \* pctAsian

- 0.0001 \* pctHisp

- 0.0219 \* pctUrban

- 0.0227 \* pctWwage

+ 0.0001 \* perCapInc

- 0.087 \* pctPoverty

+ 11.1133 \* persPerFam

- 7.203 \* persPerOccupHous

- 0.0008 \* pctHousOccup

+ 0.0004 \* popDensity

- 11.9643 [612/38.15%]

Rule: 2

pctUsePubTrans =

0.1251 \* pctBlack

+ 0.0761 \* pctWhite

+ 0.0643 \* pctAsian

- 0.0224 \* pctHisp

+ 0.0063 \* pctUrban

+ 0.0003 \* perCapInc

+ 14.6085 \* persPerFam

- 9.4754 \* persPerOccupHous

- 0.068 \* pctHousOccup

+ 0.001 \* popDensity

- 26.8448 [1603/67.641%]

M5Rules uses 2 rules to explain this data. These rules predict linear models which make the algorithm much more ﬂexible. M5 builds a tree by splitting the data based on the values of predictive attributes. M5 chooses attributes that minimize intra-subset variation in the class values of instances that go down each branch. After constructing a tree, M5 computes a linear model for each node; the tree is then pruned back from the leaves, so long as the expected estimated error decreases. The expected error for each node is calculated by averaging the absolute diﬀerence between the predicted value and the actual class value of each training example that reaches the node. The best leaf (according to some heuristic) is made into a rule and the tree is discarded. All instances covered by the rule are removed from the dataset. The process is applied recursively to the remaining instances and terminates when all instances are covered by one or more rules. This is the basic separate-and conquer strategy for learning rules. However, instead of building a single rule, as it is done usually, we build a full model tree at each stage, and make its “best” leaf into a rule.

When we applied this algorithm to our dataset it took 1.59 seconds to build this model and the correlation coefficient is less 0.73 and Relative absolute error and Root relative squared error are 66.6% and 68.2%.

# Improvements introduced by M5Rules:

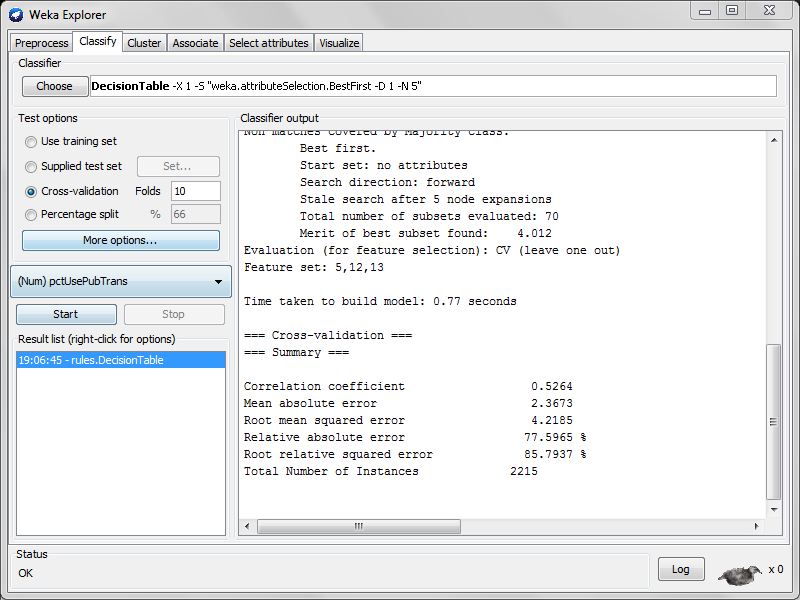
Handles discrete attributes

Handles missing values

Reduces the tree size

# Decision Table:

In a similar manner to M5Rules we also applied the Decision Table algorithm to our dataset.The results that we obtained are clearly shown below

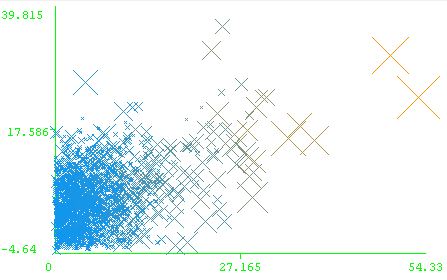


This algorithm did not give us the accurate results as M5Rules. Decision Table’s correlation coefficient is 0.52 which is very low when compared to M5Rules. Even the Relative Absolute Error and Root relative squared error are 77.59% and 85.79%.

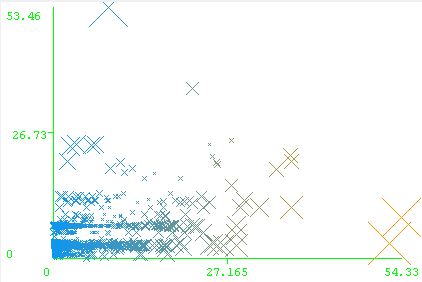
# Classifier Errors:

Both the Classifier Errors graphs have Percentage of Public Transit Usage on the X axis and Predicted Percentage of Public Transit Usage on the Y axis. Comparatively it is clearly understood that M5Rules yields the best results.

M5Rules



Decision Table:

****

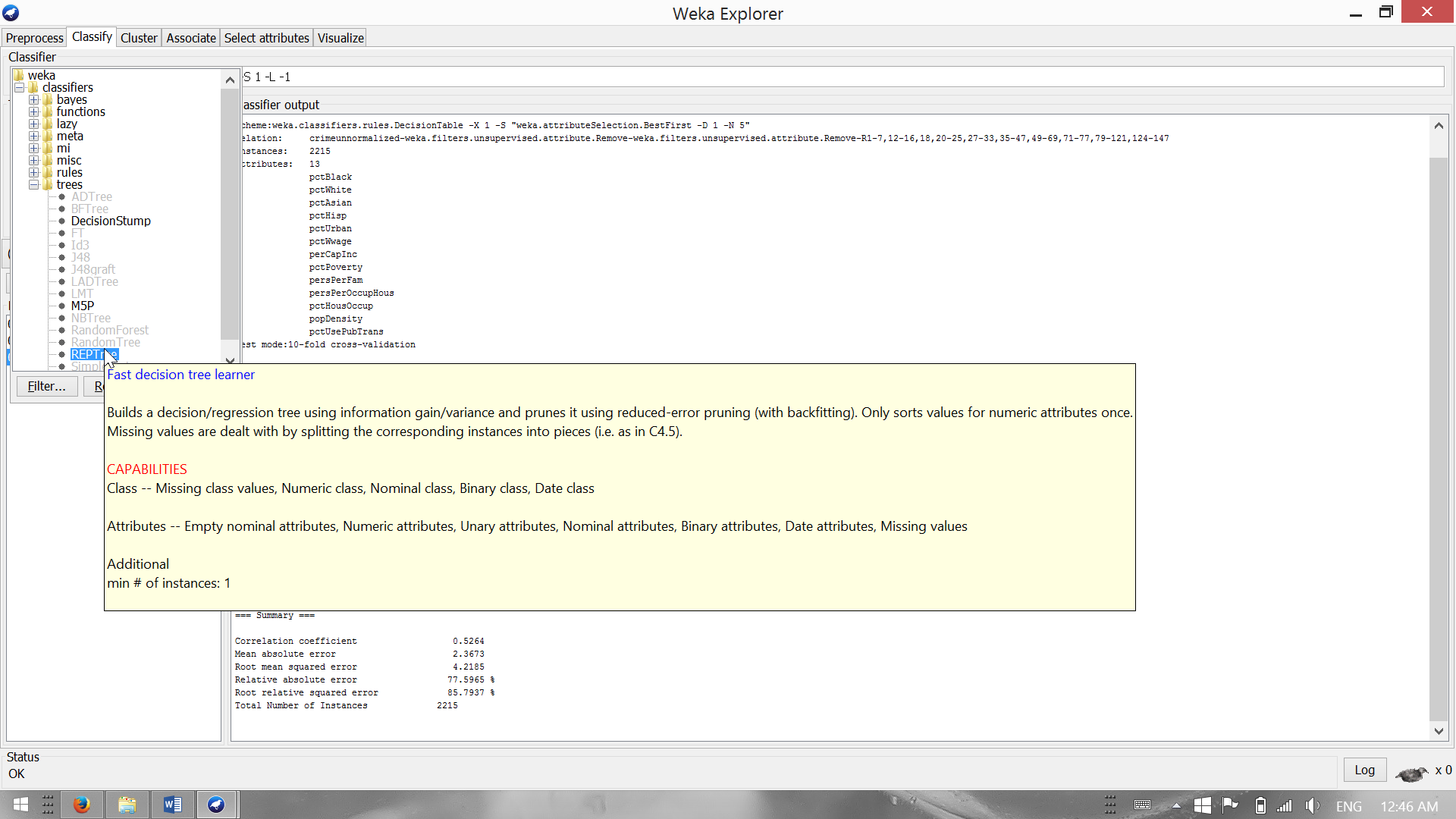
**Classify algorithms:**

We applied different trees for classification such as REP Tree and Decision Stump to classify our training data set.

1. REP tree
2. Decision stump

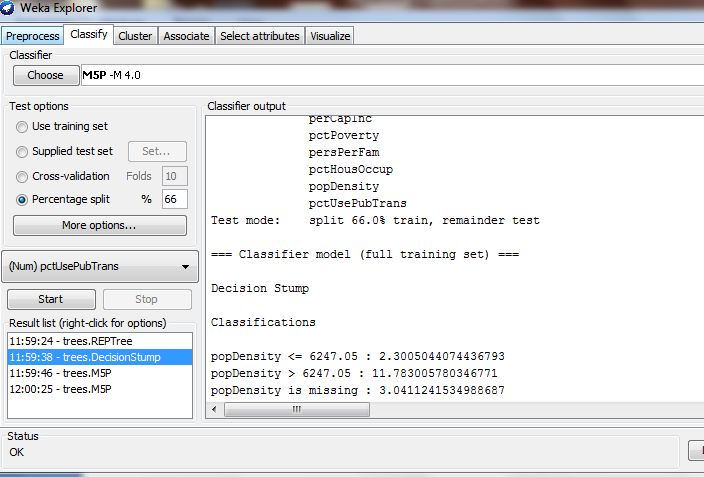
How we applied these algorithms to our data set:

First we loaded our dataset to the Weka software and chose which type of algorithm we wanted to apply. According to this we selected Classify. Then there will be an option called ‘Choose’ where we found the algorithms. From those we applied two algorithms REP tree and decision stump and saw the results for our dataset.



After that we selected which set of data should be applied to our algorithm. We selected Percentage split option from Test options and gave 66%. This will divide our dataset into 66% training and remaining as test dataset.

The next important issue is selecting the key attribute which we want to predict and see how other attributes are varying with respect to this attribute. So we selected Mass usage transmit (pctusePubTrans) from the drop list where we will find all the attributes from our dataset. After that, clicking the start button below the dropdown list will yield the results of our algorithms.



REP tree:

REPTree algorithm is a fast decision tree learner. It builds a decision/regression tree using information gain/variance and prunes it using reduced-error pruning (with back-fitting).

The algorithm only sorts values for numeric attributes once. Missing values are dealt with by splitting the corresponding instances into pieces.

Result:

**popDensity < 6248.9**

| perCapInc < 20565

| | popDensity < 2634.5

| | | pctBlack < 65.96

| | | | pctUrban < 77.75

| | | | | pctWwage < 87.74 : 0.53 (354/0.41) [172/1.03]

| | | | | pctWwage >= 87.74 : 1.98 (23/8.95) [13/11.98]

| | | | pctUrban >= 77.75

| | | | | pctAsian < 2.92

| | | | | | pctWhite < 54.41

| | | | | | | pctAsian < 0.51 : 1.58 (6/1.69) [4/0.81]

| | | | | | | pctAsian >= 0.51 : 3.83 (11/1.45) [2/7.89]

| | | | | | pctWhite >= 54.41

| | | | | | | pctAsian < 0.72 : 0.84 (105/1.08) [61/1.08]

| | | | | | | pctAsian >= 0.72

| | | | | | | | pctHousOccup < 97.44

| | | | | | | | | popDensity < 1785.6 : 1.23 (152/1.32) [74/2.07]

| | | | | | | | | popDensity >= 1785.6 : 1.78 (86/1.53) [41/2.98]

| | | | | | | | pctHousOccup >= 97.44 : 2.74 (13/4.69) [9/3.92]

| | | | | pctAsian >= 2.92 : 2.74 (61/7.02) [26/4.13]

| | | pctBlack >= 65.96 : 7.86 (6/45.37) [0/0]

| | popDensity >= 2634.5

| | | pctBlack < 49.63

| | | | perCapInc < 16572

| | | | | pctBlack < 19.06

| | | | | | pctUrban < 49.88 : 0.88 (51/1.1) [28/0.77]

| | | | | | pctUrban >= 49.88 : 2.66 (171/3.56) [84/7.78]

| | | | | pctBlack >= 19.06

| | | | | | perCapInc < 13533.5

| | | | | | | popDensity < 4137.35 : 2.69 (19/2.72) [3/1.41]

| | | | | | | popDensity >= 4137.35

| | | | | | | | perCapInc < 11048.5 : 3.73 (4/0.44) [1/0.06]

| | | | | | | | perCapInc >= 11048.5 : 8.24 (4/1.99) [2/17.37]

| | | | | | perCapInc >= 13533.5

| | | | | | | persPerFam < 3.27

| | | | | | | | pctBlack < 34.81 : 7.74 (4/4.94) [5/21.48]

| | | | | | | | pctBlack >= 34.81 : 13 (2/0.16) [1/0.63]

| | | | | | | persPerFam >= 3.27 : 3.61 (5/2.82) [0/0]

| | | | perCapInc >= 16572 : 4.41 (81/15.31) [44/14.94]

| | | pctBlack >= 49.63 : 10.91 (8/42.81) [6/75.12]

| perCapInc >= 20565

| | pctAsian < 2.74 : 3.89 (90/17.69) [49/23.26]

| | pctAsian >= 2.74

| | | pctWwage < 85.37 : 8.3 (60/40.35) [30/38.49]

| | | pctWwage >= 85.37 : 4.32 (46/15.16) [26/13.55]

**popDensity >= 6248.9**

| popDensity < 11324.55

| | pctHisp < 5.73

| | | popDensity < 8535.05

| | | | pctBlack < 1.03 : 4.87 (6/37.57) [2/18.31]

| | | | pctBlack >= 1.03 : 12.86 (20/31.38) [7/70.6]

| | | popDensity >= 8535.05 : 18.95 (8/69.72) [4/64.37]

| | pctHisp >= 5.73

| | | pctWhite < 82.93

| | | | pctBlack < 18.24

| | | | | pctPoverty < 18.07

| | | | | | pctWwage < 78.8

| | | | | | | pctBlack < 2.99 : 4.32 (4/0.93) [1/4.31]

| | | | | | | pctBlack >= 2.99 : 10.7 (3/0.33) [3/140.33]

| | | | | | pctWwage >= 78.8

| | | | | | | pctHisp < 54.35

| | | | | | | | persPerFam < 3.03 : 6.28 (3/0.7) [1/119.03]

| | | | | | | | persPerFam >= 3.03 : 2.14 (9/0.58) [7/2.76]

| | | | | | | pctHisp >= 54.35 : 4.35 (3/0.62) [2/0.76]

| | | | | pctPoverty >= 18.07

| | | | | | pctAsian < 5.58 : 5.09 (3/1.47) [1/8.64]

| | | | | | pctAsian >= 5.58

| | | | | | | pctAsian < 10.75 : 9.05 (2/0.25) [1/7.9]

| | | | | | | pctAsian >= 10.75 : 6.04 (2/0.18) [1/0]

| | | | pctBlack >= 18.24 : 11.36 (16/20.04) [10/54.06]

| | | pctWhite >= 82.93

| | | | pctHisp < 20.71 : 7.53 (9/17.08) [3/34.08]

| | | | pctHisp >= 20.71 : 19.32 (2/4.2) [0/0]

| **popDensity >=** 11324.55 : 19.41 (24/162.69) [15/169.83]

Time taken to build model: 0.22 seconds

=== Evaluation on test split ===

=== Summary ===

Correlation coefficient 0.7177

Mean absolute error 2.097

Root mean squared error 3.6123

Relative absolute error 67.4654 %

Root relative squared error 69.828 %

Total Number of Instances 753

When we applied this algorithm it took 0.22 seconds to build this model. The correlation coefficient is more than 0.7 which is near to one and the accuracy is more than 97%. The Relative absolute error and Root relative squared error are 67 and 69%.

As we stated in the predictions that Population density and per-captia income are strongly related with the Mass usage transit, when we applied our dataset to this algorithm it automatically showed the population density and per-capita at the top most nodes which are our key attributes. We also proved from this algorithm that as the population density increases the mass transit usage also increases.

The decision tree algorithm identified population density as the most significant attribute. A population density greater than or equal to 11324.55 leads to a prediction of 19.41 public transportation usage with no other attributes considered. For a population density less than 11324.55 but greater than or equal to 6248.9 leads to a prediction greater than the mean in all cases. However racial demographics become significant. Generally speaking, the highest values for public transportation usage in this population density range are for communities with a low percentage Hispanic and a high percentage black.

For population densities less than 6248.9, public transportation usage is generally quite low. However, there are two important exceptions. Communities with higher mean per capita income have higher public transportation usage and communities that have a high percentage black have higher public transportation usage.

Decision Stump:

Decision stumps are basically decision trees with a single layer. As opposed to a tree which has multiple layers, a stump basically stops after the first split.

Decision stumps are usually used in population segmentation for large data. Occasionally, they are also used to help make simple yes/no decision model for smaller data with little data.

Result:

When we applied this algorithm to our dataset it took 0.09 seconds to build this model but the correlation coefficient is less 0.533 and Relative absolute error and Root relative squared error is more than 80%. It took 753 instances for giving these results.

This algorithm did not give us the accurate results as REP tree. The Classifications it gave is just the value for mass transit usage when popDensity<= 6247.05 as 2.3005 and popDensity > 6247.05 as 11.783.

Classifications

popDensity <= 6247.05 : 2.3005044074436793

popDensity > 6247.05 : 11.783005780346771

popDensity is missing : 3.0411241534988687

Time taken to build model: 0.09 seconds

=== Evaluation on test split ===

=== Summary ===

Correlation coefficient 0.5333

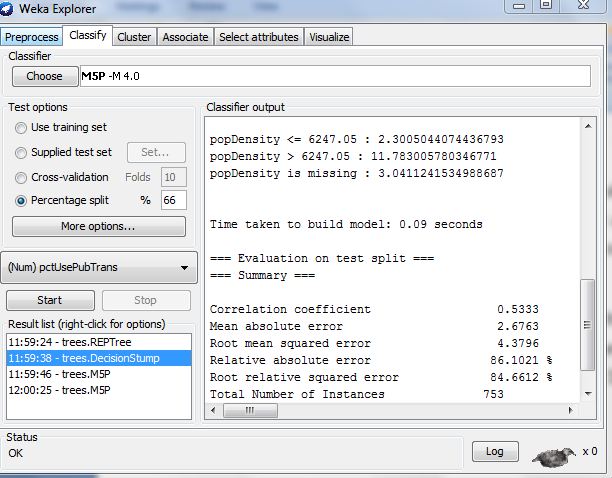
Mean absolute error 2.6763

Root mean squared error 4.3796

Relative absolute error 86.1021 %

Root relative squared error 84.6612 %

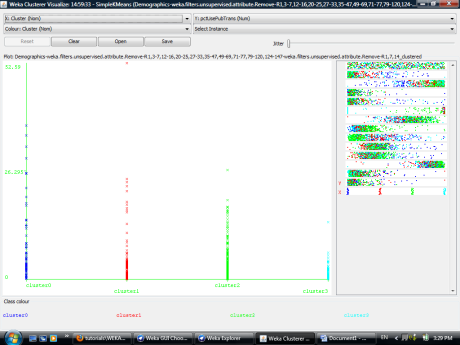
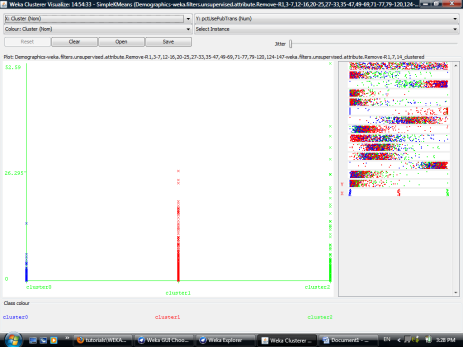
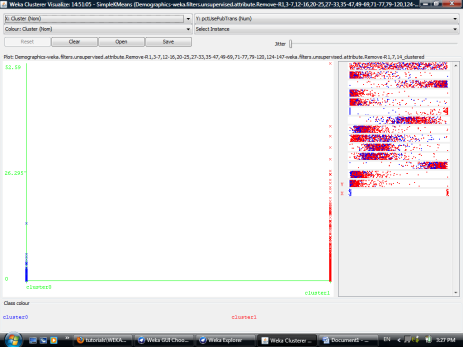
Total Number of Instances 753



# Clustering Algorithms:

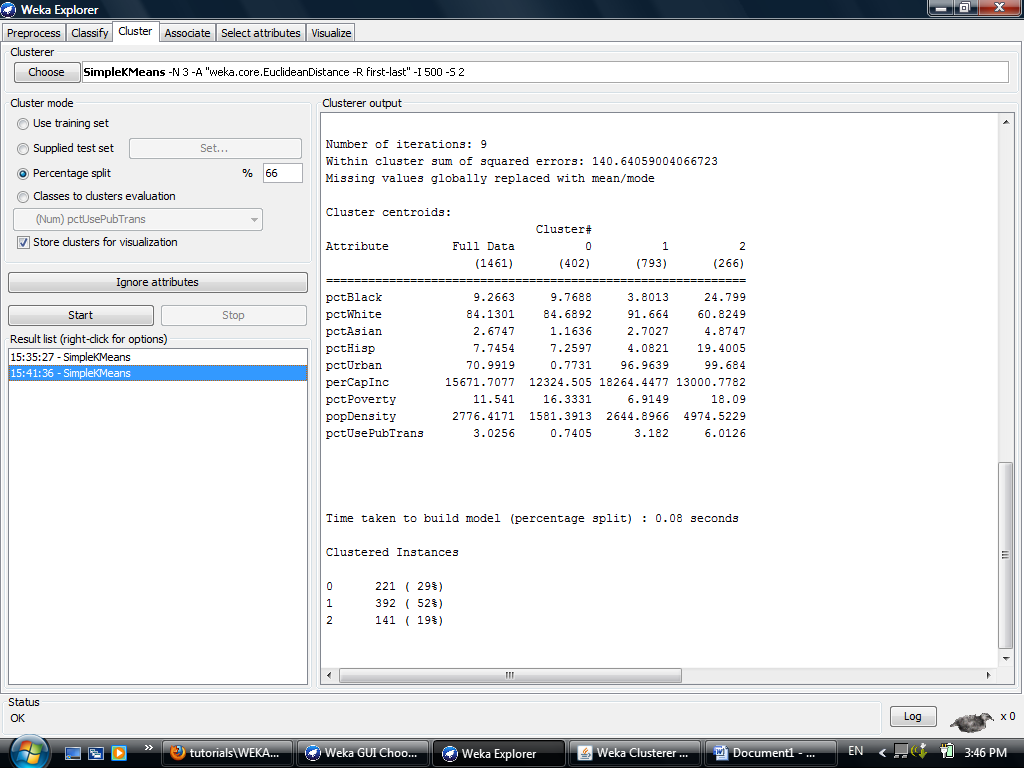
We attempted a number of clustering algorithms to find useful patterns in the data. We attempted hierarchical clustering but it did not work well for this problem. At a low level of acuity and pruning, the data did not form into useful groups. At a higher level, there was too much individual variation in the data to be helpful. We found the SimpleKMeans and EM clustering algorithms to be the most effective for this problem.

We used a SimpleKMeans clustering algorithm to identify data clusters. We initially examined values of K between 2 and 6. Since the primary focus of our research question was to examine percentage public transportation usage as a dependent variable, as a means to select the best value of K to analyze we visualized the degree to which this variable was dependent on the cluster number for each value of K.



Regardless of the value of K, in each group some values of public transportation usage were close to zero. However a K value of 3 showed the most clarity with respect to this variable, with ranges having a max value low, medium, and high.

Next, to simplify the analysis we ignored a few attributes that seemed to have similar centers for all clusters and ran the algorithm again. We then examined the differences between the three clusters, and identified the defining characteristics of each.



Cluster 0 (29%): Rural Non-Commuters

This group has the lowest public transportation usage. It also has an extremely low urban identification and population density. Incomes are the lowest for any group and poverty levels are moderately high. This group lives in rural areas and generally does not use public transportation.

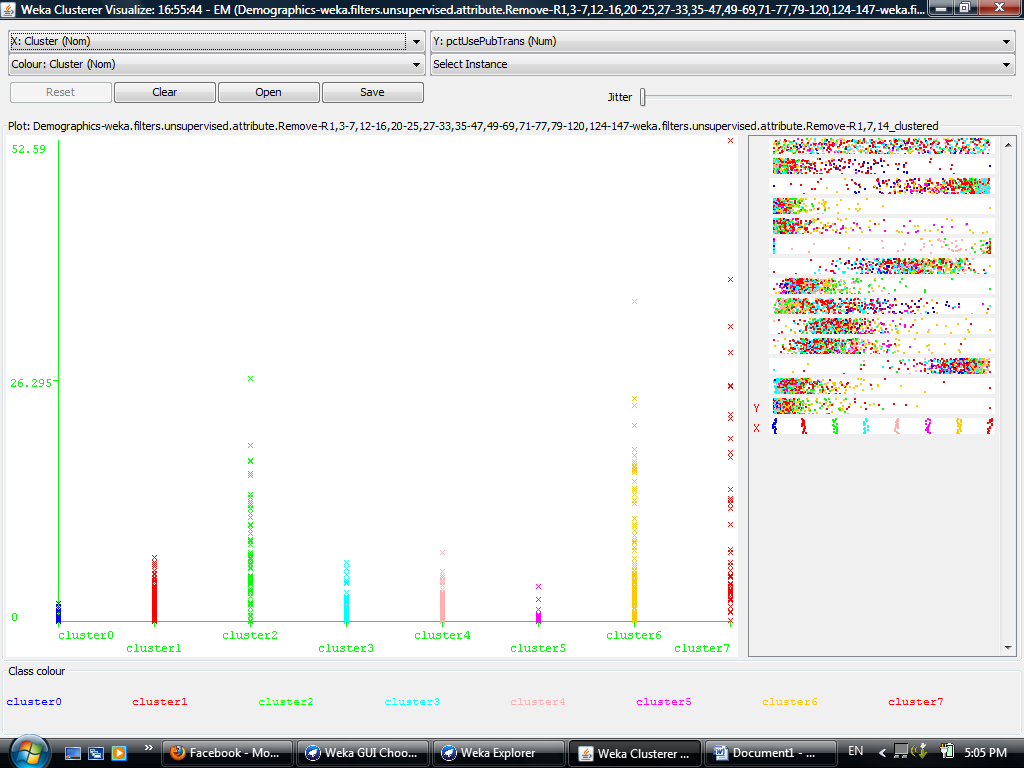
Cluster 1 (52%): Suburban Commuters

This group has moderate to high public transportation usage. The cluster center is about ninety-seven percent urban, but the population density shows that it is only moderately so. This group is approximately ninety-two percent white, and has the lowest percent black and Hispanic of any cluster. The per capita income is significantly higher and the poverty level is significantly lower for this group than the other clusters.

Cluster 2 (19%): Urban Commuters

This group has the highest public transportation usage. This cluster is ninety-nine percent urban, with the highest population density. This group has the lowest percent white and the highest percentages Asian, black, and Hispanic. This group also has the highest poverty level. This group generally lives in large cities and commutes to work.

Examining the same attributes, the EM (expectation maximizing) algorithm automatically selected eight clusters. Examining the clusters for prediction of transportation usage, we first looked at the visualization of the data.



Cluster 7 has virtually all data points above the mean, and maximizes the variable. Cluster 0 and cluster 5 are excellent predictors of low public transportation usage. Clusters 1, 4, and 5 have moderate to low usage, while clusters 2 and 6 have moderate to high usage.

Examining the data in detail, we can describe the clusters as follows. We have ordered the clusters according to their value as predictors of public transportation usage.

Cluster 0 (5%): extremely impoverished non-commuters

This group has the highest poverty rate and the lowest per capita income. It is twenty-one percent urban but has a low population density, indicating that the urban areas are likely blight areas. This cluster centers on sixty-three percent white and thirty-five percent black.

Cluster 5 (5%): rural, impoverished non-commuters

This cluster is very similar to cluster 0, but has a much lower urban percentage and a much higher percentage Hispanic, around thirty-two percent.

Cluster 3 (20%): rural lower middle class non-commuters

This group is similar to Cluster 5, but has a slightly higher per capita income, and is ninety-three percent white.

Cluster 4 (15%): suburban lower middle class non-commuters

This cluster has a fairly high percentage urban, but a fairly low population density. This suggests suburban areas. The group is ninety-six percent white and has a somewhat higher per capita income than the previous groups.

Cluster 1 (25%): semi-urban lower middle class non-commuters

This group is approximately one-hundred percent urban, but the population density is not extremely high. The cluster has a higher poverty percentage and lower per capita income than cluster 4. The group is eighty-seven percent white and nine percent black.

Cluster 2 (9%): semi-urban upper middle class commuters

This cluster is similar in demographic characteristics to cluster 1, but it has the lowest poverty percentage of any cluster. It also has the highest per capita income of any group. This group has a relatively high public transportation usage.

Cluster 6 (15%): urban middle class commuters

This group is urban, with an extremely high population density. It has a mid-range per capita income and moderate poverty rate. It is approximately seventy-seven percent white and sixteen percent Hispanic. This group has a very high public transportation usage.

Cluster 7 (6%): urban lower middle class commuters

This cluster is urban, with a high population density. It has a low per capita income and a relatively high poverty rate. It is approximately forty-six percent white and forty-six percent black. This group has the highest public transportation usage.

More detailed and significant information was gathered using the EM clustering algorithm, but the results generally agree. The most important indicators for public transportation usage are population density and urban identification, with the greatest usage needed by the nineteen to twenty-four percent of the population living in large population centers. The next greatest indicator is per capita income. The clusters with the lowest income have the lowest public transportation usage, while semi-urban areas with high per capita income have a high usage.

# Research questions:

Is there a definable population density threshold which sees a large increase in commuting times and mass transit usage?

Does the data not conform well to predictable patterns based on human planning and decision making?

Can we predict where mass transit infrastructure will be most and least needed based on demographic trends?

What is the relationship between commuting times and mass transit usage? Does mass transit decrease commuting times (it is clear that it has a positive ecological impact)?

# Project Management:

| **Team member** | **Roles and skills** | **Contributions** |
| --- | --- | --- |
| Bethi Bharath Reddy | code development | Updated the project proposal, analyzed the available traffic data and working on coding part. |
| Venkateswara rao alluri | Works on testing | Cleaning of data and eliminated some of the bad data |
| Roja suryuadhevara | Collected data | Progressing demographics |
| Deepthi kota | Data analysis | Analysing data required, reporting accurate data |
| Priyanka pappuri | Installation and setup | Finished installing and setting up the environment required to test our project outcomes. |

# Conclusion:

* We applied different rules for classification such as DecisionTable, M5Rules and identified M5Rules as the best rule set. We applied different trees for classification such as REP Tree and Decision Stump to classify the dataset and identified REP Tree as the best. We also applied the Cluster Algorithms such as SimpleKMeans and EM clustering and identified EM Clustering as the best clustering algorithm. Using classification rules, decision trees, and clustering we were able to confirm our hypothesis. We were also able to identify the specific numeric thresholds for target groups and clustered demographics for target populations.70% of the population has very low usage of public transportation. Infrastructure improvement should be targeted to the 30% with high usage.

# Reference:

* For finding an appropriate dataset we went through several websites and finally found this dataset on the link. [http://archive.ics.uci.edu/ml/datasets/Communities+and+Crime+Unnormalized#](http://archive.ics.uci.edu/ml/datasets/Communities+and+Crime+Unnormalized).
* The creator of this dataset is Michael Redmond; Computer Science; La Salle University; Philadelphia, PA, 19141, USA.