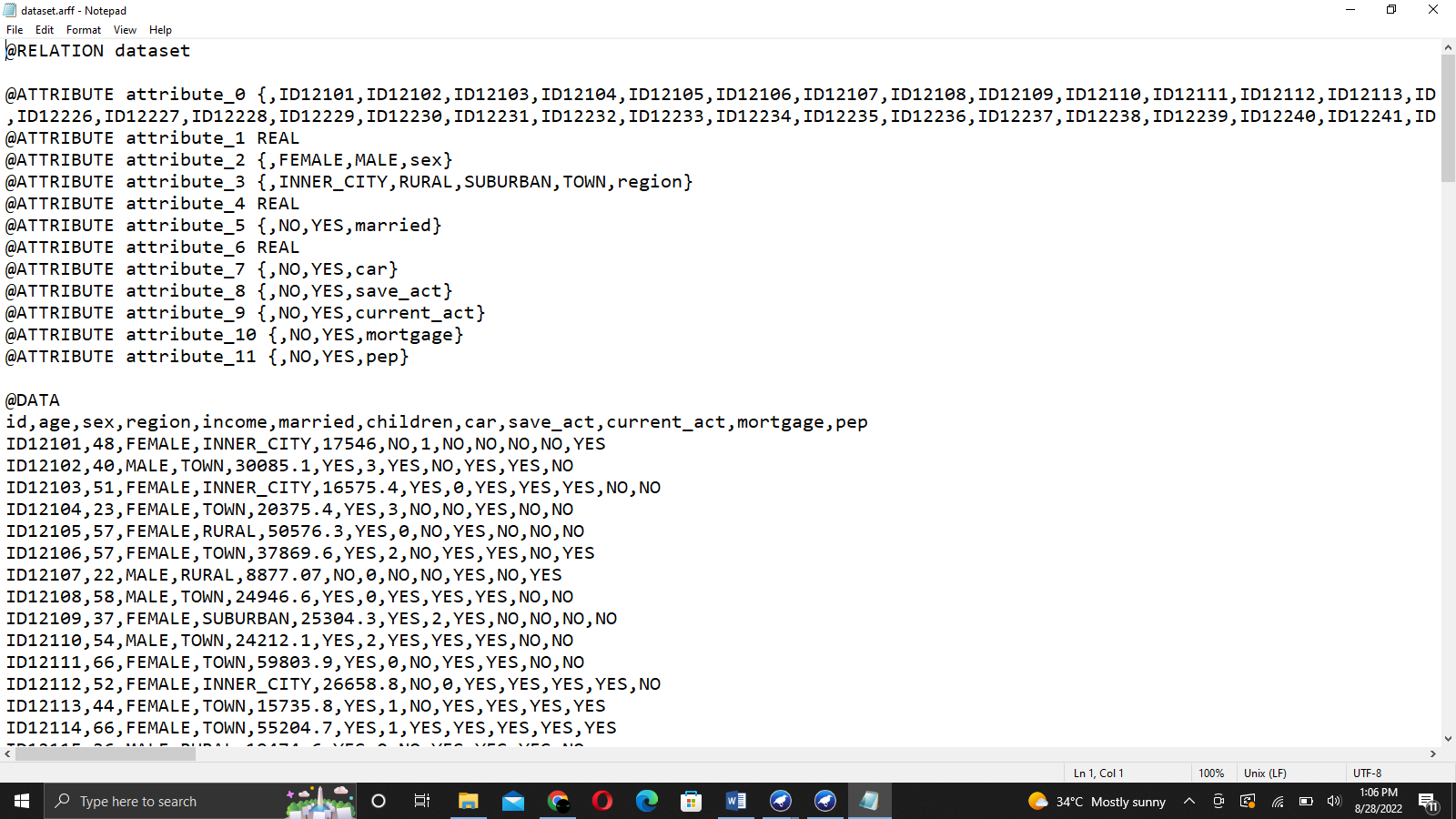
**ARFF for weka Report**

**Task 1:**

1. **Screenshot of corrected arff file**



1. **Which attribute in the dataset do you think is useless and did not provide useful**

Children attribute is useless because it didn’t provide useful information for prediction.

1. **How many attributes the dataset has?**

The dataset has total of 12 attributes.

1. **How many instances the dataset has?**

The dataset has 150 instances.

1. **What is the class attribute in the data.arff dataset?**

Id is the class attribute in dataset.arff

1. **What proportion of customers who has a mortgage and live in Inner City?**

14.67%

1. **What proportion of customers who has a mortgage and not living in Inner City?**

16.67%

1. **What proportion of customers have a mortgage, and their income is between**

**$1000 and $10000?**

0.0%

1. **How many customers are married and have no mortgage?**

44.67%

1. **How many customers have not owned a car and have a mortgage?**

15.34%

**Task 2:**

* **The attributes are:**
  + Id
  + Age
  + Sex
  + Region
  + Income
  + Married
  + Children
  + Car
  + Save\_act
  + Current\_act
  + Mortgage
  + Pep
* **Real value and nominal attributes**

Genuine esteemed credits are numeric traits containing just genuine qualities. These are quantifiable amounts. These qualities can be span scaled, for example, temperature or proportion scaled like mean, or middle.

Ostensible traits address names or some portrayal of things. There is no organization in such traits and they address some classification. For instance, variety.

The board underneath the ongoing connection shows the name of characteristics.

In the right board, the chose characteristic measurements are shown. Select the characteristic "checking\_status".

It shows:

Name of the trait

Missing: Any missing upsides of the trait in the dataset. 0% for this situation.

Unmistakable: The quality has 4 particular qualities.

Type: The trait is of the ostensible sort that is, it takes no numeric worth.

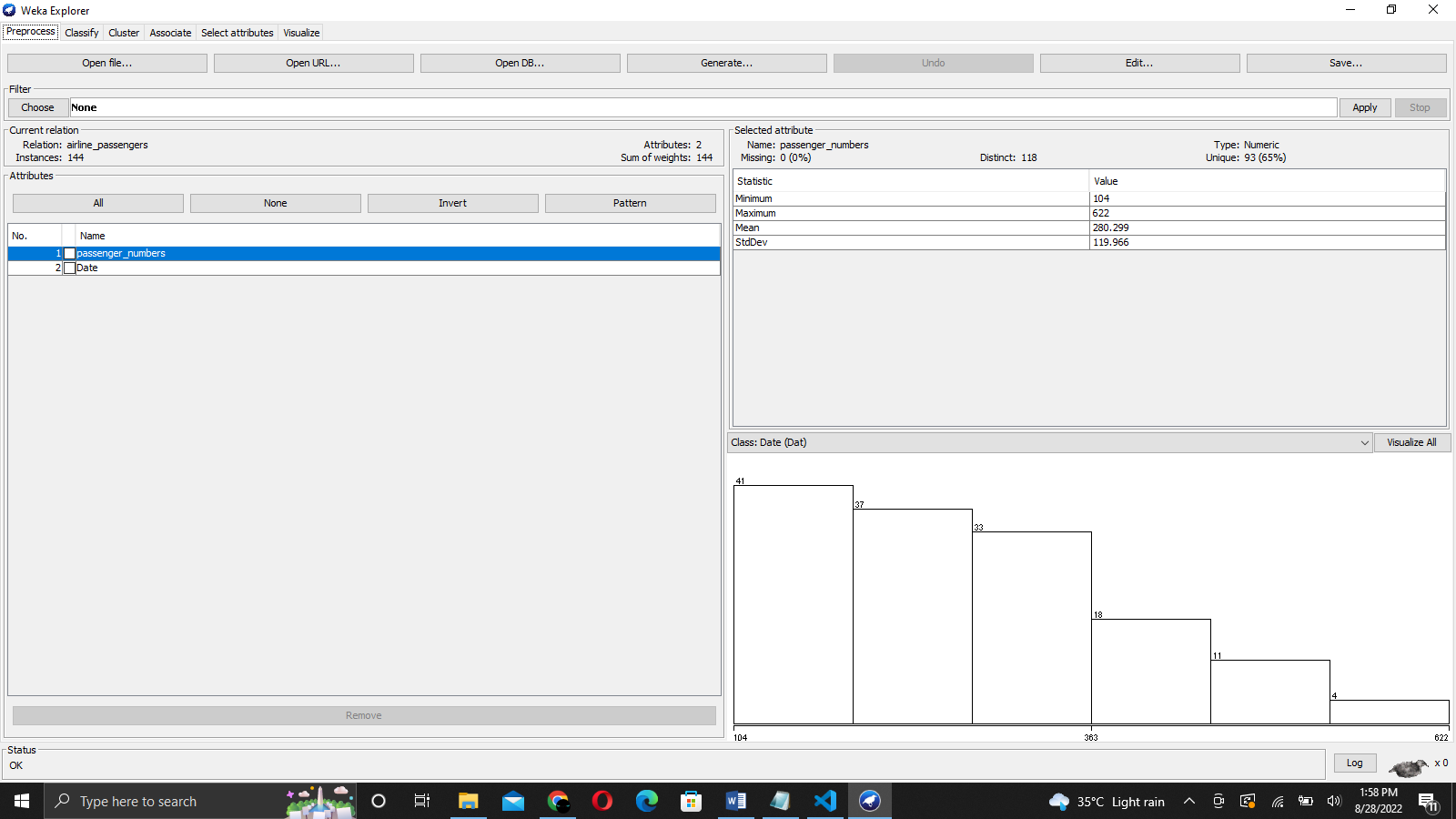
Count: Among the 1000 cases, the count of each unmistakable class name is written in the count section.

Histogram: It will show the result class mark for the characteristic. The class mark in this dataset is either positive or negative. There are 700 examples of good (set apart in blue) and 300 cases of awful (set apart in red).

For the mark < 0, the occurrences for good or awful are practically a similar in number.

For mark, 0<= X<200, the examples with choice great are more than cases with terrible.

Also, for name >= 200, the maximum cases happen for good and no checking mark has more occurrences with choice great.



Name: This is the Name of the property.

Type: The kind of the property is numeric.

Missing worth: The property has no missing worth.

Unmistakable: It has 33 particular qualities in 1000 examples. It implies in 1000 cases it has 33 unmistakable qualities.

Novel: It has 5 extraordinary qualities that don't coordinate with one another.

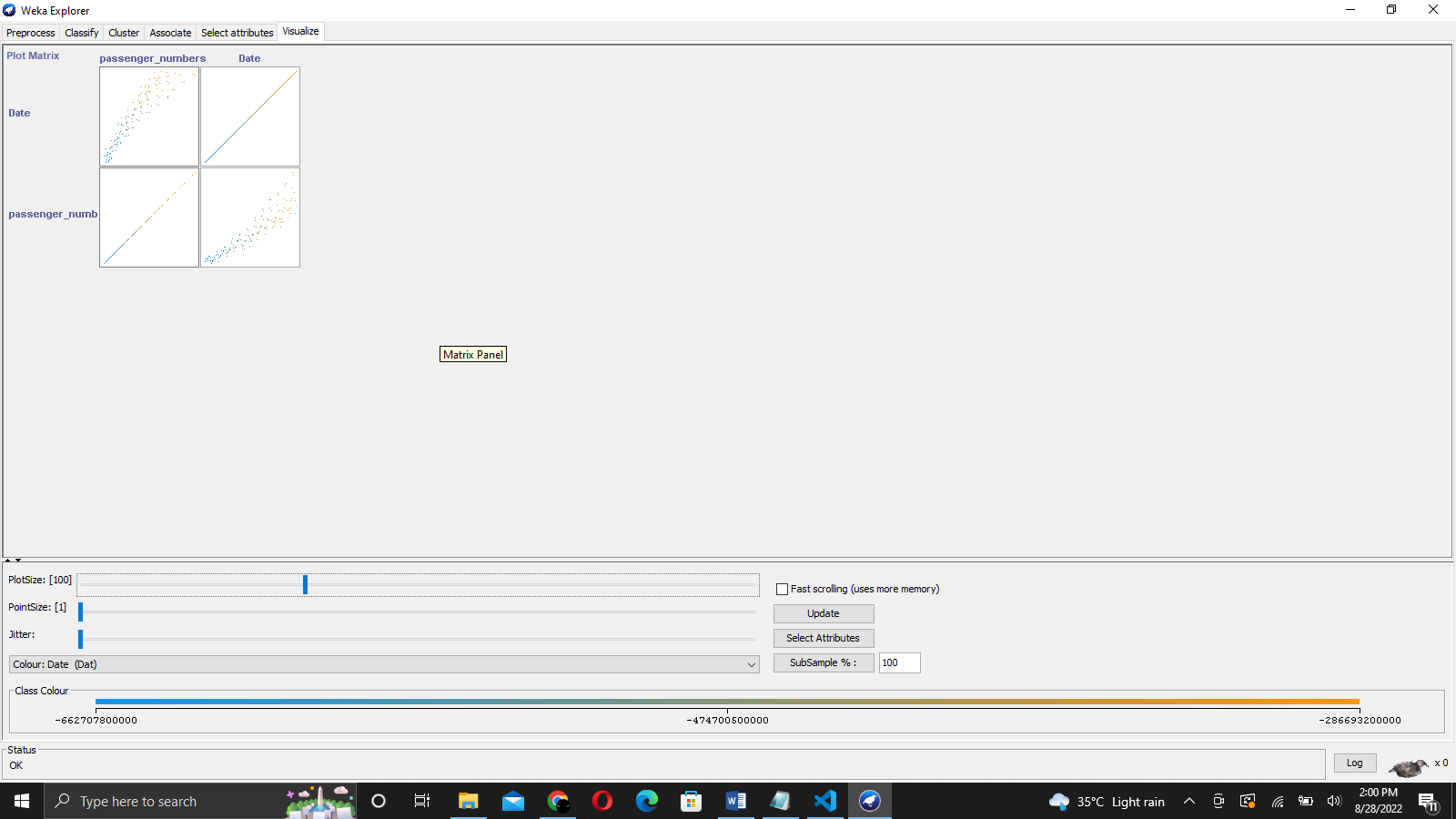
Least worth: The min worth of the property is 4.

Most extreme Value: The maximum worth of the property is 72.

Mean: Mean is adding every one of the qualities separated by cases.

Standard Deviation: Stddeviation of trait span.

Histogram: The histogram portrays the length of 4 units, the maximum cases happen for a decent class. As the span increments to 38 units, the quantity of occasions lessens for good class marks. The span arrives at 72 units which has just a single occasion which characterizes the choice as terrible.



WEKA channels have numerous functionalities to change the quality upsides of the dataset to make it appropriate for the calculations. For instance, the numeric change of characteristics.

Sifting the ostensible and genuine esteemed credits from the dataset is one more instance of utilizing WEKA channels.

**Output**

The genuine esteemed and ostensible qualities credits in the dataset are distinguished. Perception with the class mark is found as histograms.

**Decision Tree**

A choice Tree is the grouping strategy that comprises of three parts root hub, branch (edge or connection), and leaf hub. Root addresses the test condition for various traits, the branch addresses generally potential results that can be there in the test, and leaf hubs contain the name of the class to which it has a place. The root hub is at the beginning of the tree which is additionally called the highest point of the tree.

**J48 Classifier**

It is a calculation to produce a choice tree that is created by C4.5 (an expansion of ID3). It is otherwise called a measurable classifier. For choice tree order, we really want an information base.

Steps include:

#1) Open WEKA traveler.

#2) Select weather.nominal.arff document from the "pick record" under the preprocess tab choice.

Pick dataset

#3) Go to the "Arrange" tab for grouping the unclassified information. Click on the "Pick" button. From this, select "trees - > J48". Allow us likewise to have a brief glance at different choices in the Choose button:

Bayes: It is a thickness assessment for mathematical properties.

Meta: It is a multi-reaction straight relapse.

Capabilities: It is calculated relapse.

Sluggish: It sets the mix entropy consequently.

Rule: It is a standard student.

Trees: Trees characterize the information.

Characterize tab

#4) Click on Start Button. The classifier result will be seen on the Right-hand board. It shows the run data in the board as:

Plot: The order calculation utilized.

Occasions: Number of information lines in the dataset.

Credits: The dataset has 5 ascribes.

The quantity of leaves and the size of the tree depicts the choice tree.

Time is taken to fabricate the model: Time for the result.

Full characterization of the J48 pruned with the characteristics and number of occasions.

Characterized yield data

Envision tree

#5) To envision the tree, right-click on the outcome and select imagine the tree.

Choice tree

Yield:

The result is as a choice tree. The principal characteristic is "standpoint".

On the off chance that the viewpoint is radiant, the tree further examines the dampness. On the off chance that stickiness is high, the class mark play= "yes".

Assuming the viewpoint is cloudy, the class name, play is "yes". The quantity of occurrences which comply with the characterization is 4.

On the off chance that viewpoint is blustery, further grouping happens to examine the trait "breezy". In the event that windy=true, the play = "no". The quantity of occasions which comply with the arrangement for outlook= blustery and windy=true is 2.

**Result**

WEKA offers an extensive variety of test datasets to apply AI calculations. The clients can perform AI errands like characterization, relapse, quality choice, and relationship on these example datasets, and can likewise gain proficiency with the instrument utilizing them.

WEKA traveler is utilized for carrying out a few roles, beginning from preprocessing. Preprocessing accepts input as a .arff document, processes the information, and gives a result that can be utilized by other PC programs. In WEKA the result of preprocessing gives the characteristics present in the dataset which can be additionally utilized for factual examination and correlation with class marks.

WEKA additionally offers numerous order calculations for choice trees. J48 is one of the well known grouping calculations which yields a choice tree. Utilizing the Classify tab the client can picture the choice tree. Assuming that the choice tree is excessively populated, tree pruning can be applied from the Preprocess tab by eliminating the characteristics which are not needed and beginning the arrangement interaction once more.

**Comparison**

The Naive Bayes calculation is a basic probabilistic classifier that computes a bunch of probabilities by counting the recurrence and blends of values in a given informational collection. The calculation utilizes Bayes hypothesis what's more, expects all credits to be free given the worth of the class variable. This contingent freedom supposition seldom turns out as expected in genuine world applications, thus the portrayal as Guileless yet the calculation will in general perform well and advance quickly in different managed characterization issues.

Genuine positive (TP): If the result from a forecast is p and the genuine worth is likewise p, then it is known as a genuine positive. Bogus positive (FP): However in the event that the genuine esteem is n then it is supposed to be a misleading positive. Accuracy and review: Precision is the portion of recovered occasions that are significant, while review is the negligible part of pertinent cases that are recovered. Both accuracy and review are subsequently founded on a comprehension and proportion of significance. Accuracy should be visible as a proportion of precision or quality, while review is a proportion of fulfillment or then again amount. Review is only the valid positive rate for the class.

Two classifiers, gullible bayes calculation what's more, J48 choice tree calculation are utilized for examination. Examination is made on precision, responsiveness and particularity utilizing genuine positive and misleading positive in disarray grid created by the particular calculations. Likewise we can utilize the right furthermore, wrong occasions that give us a generally effective technique for order by utilizing the disarray framework.

J48 classifier is a straightforward C4.5 choice tree for order. It makes a double tree. The choice tree approach is most helpful in arrangement issue. With this method, a tree is built to model the order cycle. When the tree is constructed, it is applied to each tuple in the data set and brings about order for that tuple.

The exhibition of characterization calculation is generally analyzed by assessing the precision of the characterization. Anyway since characterization is frequently a fluffy issue, the right response might rely upon the client. Customary calculations assessment draws near for example, deciding the reality above can be utilized however these methodologies are generally optional. Figuring out which better best is relies upon the understanding of the issue by clients. Characterization exactness is normally determined by deciding the level of tuples set in a right class. This overlooks the way that there likewise may be an expense related with a wrong task to some unacceptable class. It would be ideal for this maybe to likewise decide.

This demonstrates that the, J48 is a basic classifier procedure to go with a choice tree. Productive outcome has been taken from dataset utilizing weka apparatus in the analyze. Guileless Bayes classifier additionally appearing great outcomes. The tests results displayed in the review are about arrangement precision and cost examination. J48 gives more order exactness for class contract in bank dataset having two qualities Yes and negative. However here in this model, cost examination esteemed same for both the classifier, with orientation quality, we can demonstrate that J48 is cost productive than the Naïve Bayes classifier.

**Summary**

The J48 is a tree-based calculation, which is utilized to find the manner in which the property vector performs for various occurrences. Additionally, based on the preparation occasions, the classes for the recently delivered information are found. This calculation created the principles for the result variable expectation. With the guide of tree grouping calculation, the right appropriation of the information is effectively sensible. J48 is a development of ID3. The extra elements of J48 incorporate ascertaining the missing qualities, choice trees pruning, nonstop characteristic worth reaches, and so forth. The J48 is an open source Java execution of the C4.5 calculation. It presents a bunch of choices connected with tree pruning. On account of expected over fitting, pruning can be utilized as an accuracy device.

For different calculations, the arrangement is accomplished recursively until each and every leaf is unadulterated. For this situation, the arrangement of the information ought to be essentially as exact as could really be expected. The J48 creates the guidelines from which a particular personality of the information is delivered.

Naive bayes (NB) is an order calculation that delivers a probability of a particular arrangement of clarifications connected with a particular class, which vary because of the upsides of the class name variable. The NB classifier has been acknowledged as an essential probabilistic classifier, which depends on clear free standards of Bayesian hypothesis.

In AI and information mining, characterization is an essential issue. In a characterization, the idea of this calculation is to build a classifier with class marks. The NB approach is a managed characterization calculation which utilizes the hypothesis of Bayes.

**Conclusion**

Contract trait has been decided arbitrarily for bank informational index. J48 is applied on the informational index and the disarray framework is produced for class orientation having two potential qualities for example YES or NO.

We have performed arrangement utilizing Naïve Bayes calculation and J48 choice tree calculation on dataset.arff dataset in weka apparatus. Weka instrument give inbuilt calculations to guileless Bayes and J48. genuine up-sides for class a='YES' is 33 while bogus up-sides is 72 though, for class b='NO', genuine up-sides is 170 and bogus up-sides is 25 for example corner to corner components of grid 33+170 =203 addresses the right occurrences ordered and other components 25+72 = 97 addresses the wrong occurrences.

Genuine positive rate = corner to corner component/amount of applicable line

Bogus positive rate = non-corner to corner component/amount of applicable line

Consequently,

TP rate for class a = 33/(33+72) = 0.314

FP rate for class a = 25/(25+170) = 0.128

TP rate for class b = 170/(25+170) = 0.871

FP rate for class b = 72/(33+72) = 0.685

Normal TP rate = 0.677

Normal FP rate = 0.491

Here same, Mortgage trait has been decided for dataset index. Guileless Bayes is applied on the informational index also, the disarray network is created for class orientation having two potential qualities.