# CHAPTER 5

# BULK ARRIVAL REDUCTION AND CLASSIFICATION OF AUTHENTICATED REQUIREMENTS

## 5.1 SERVICE CLASSIFICATION AND FILTERING

Generally in online service products data retrieval is based on information. The information in context depends on what kind of data to be taken for service execution. During service execution, data gets frequently varied, based on the customer options and their needs. Customers have various options to choose their performed services based on service offer, service quality and service availability. For the selection of services, requirements are significant and the performed service should be available at the required time. In the present marketing business environment, multiple services carry out vast amount of transactions. Similarly also they receive bulk requests from the customers. Due to the bulk arrival of transactions, service providers most frequently receive partial credential information only. To avoid such bulk arrival of requests, data is classified as separate packages. The formation of information under classification acts as center of attention for multiple factors. When considering these factors, it improves the cohesiveness in group information. Classification of data takes place under groups namely topic based information, structured based information and user environment information.

### 5.1.1 Topic Based Information

The classification of the grouped information is specifically oriented on the subject. The collected content of the abstracted data should represent with some disciplines. For example: engineering, medicine and agricultural data are categorized with respect to their domain. In olden days, information retrieval analysis was analyzed with respect to domain related contents. The chosen content are matched with existing datasets from that abstracted data are selected. From the abstracted content, the classification has been done with the matched sub component. While matching topic based composition, the abstracted data gets connected with other matched data. Such a way of matching, the collected group was meaningful and produced the valued output. Considering, the travel booking services contain only the relevant content of booking travel tickets. It also has the sub composition of hotel booking, vehicle arrangements and tourism place connectivity that are connected together. With respect to the data composition, the individual services are established and choreographed.

Similar to orchestration, the choreography has centralized source that to be connected with multiple sub sources. The orchestrate services do not have centralized grouping to control sub services. In travel booking service, if a customer wants to book a flight ticket it may either accessed by the individual customer or accessed by the agent service. If it accessed with the single user the full control stated with the customer. Otherwise it service reply sent to the agent not to the requesting customers. The first condition is said to be of orchestrated service and second condition is considered as choreographed service. These services are coordinated together to produce valuable services to the customers.

### 5.1.2 Structured Based Information

The collective groups of services are used for many web applications. During formation of these applications, the intended services should be framed in a structured way. The structural hierarchy is based on the ontological frameworks. In ontology, the semantic web represents every component, each component are having valuable meaning. For example, the ‘yahoo’ group of services framed with the base of ontology. The main group and sub groups are categorized into more than 10,000 categories. The scheme was developed based on the ontological scheme and the 10,000 category groups are worked with the team works. As the design of structural yahoo groups, it will contain up to date information by adding and updating easily for making the logical evaluation. The structured information is formulated based on certain rules there are timings, service response and proper message transmission. The design of hierarchical classification was mainly focused to timing constraint. With the assigned period of time, the team has to collect the information. After collecting the information the team discussion are carried out that conduct review for the unwanted information and threats. From the finalized requirements, the work was divided with the number of groups and starts the initialization processing. In allocation of works, the time duration was assigned with each team and monitored by the team leader. Finally, the finalized goal was collected from individual groups and was designed based on that information the structure.

The next formation of information was based on the service response. The multi source collected data is presented to the team leader. The multiple information are collected from multiple sources. The data are gathered with an acknowledgement or service response from the accessed services. If the acknowledgment is not received then the service requests are discarded. In third, the proper message transmission also comes as part of the service response. The link or bond between source data component to destination component are clearly represented and the messages are transmitted with the stated rules. Association, aggregation, generalization, uni-directional and bi-directional are often used to represent the linkages.

### 5.1.3 User Environment Information

Online applications are mostly viewed and accessed by end users. These types of users are persons who access service information from current source or remote source. Current source is an application that is accessed with the tightly coupled network and remote source is the loosely coupled network. With respect to the networks, the end user accessing mode gets insipidly varied. Either it is of tightly coupled or loosely coupled network the user accessing is based on the Application Programming Interface (API) interface. Two types of interfaces generally defined for web services they are admin API and service based API. The admin API contains information about deploy, configure, control and update multiple services in the distributed network. The role of admin interface is to allow users to access the required services with a valid authentication user Id. Only after a valid authentication ID is entered, the applied services are accessed through admin API. For service based API, the selected services are assigned with an Id. These Id values are used, whenever the admin requires any particular service. If any particular service requires updation, deletion or creation of new information then the necessary operation is carried out easily with this valid Id. Also, the updated or deleted or newly created contents using an individual Id is automatically stored in admin API. If any authentication Id gets lost, then the contents are accessed from the source admin itself.

With the authenticated user environment information services are accessed authentically and are collected from both local and remote resources. With this information, classification factors improve the reliability, framed structure and end user accessibility. By discussing the above factors, the proposed HOAG system is in the position to reduce the vast transaction using Classifier Filter method. Also, this chapter deals service classification, with the proposed Efficient Trim down (ETD) algorithm. The proposed ETD algorithm was intended with Empirical estimation, Risk analysis and Complexity of the transaction data. Also, the proposed ETD accuracy was matched with the existing Collective Group classifiers of (J48, Random Tree, Random Forest and AD Tree). Collective Group classifiers are matched with ETD classifier and the accurate results are passed back to the service providers.

## 5.2 DESIGN ARCHITECTURE OF CLASSIFIED FILTER

The Figure 5.1 represents layer 2 technique called Classified Filter and its processing components. The sub components framed inside are Multi Classifier Mixture (MCM), Collective Group-Efficient Trim Down (CG-ETD) classifiers, Accuracy Analyzer and Request Recognizer.

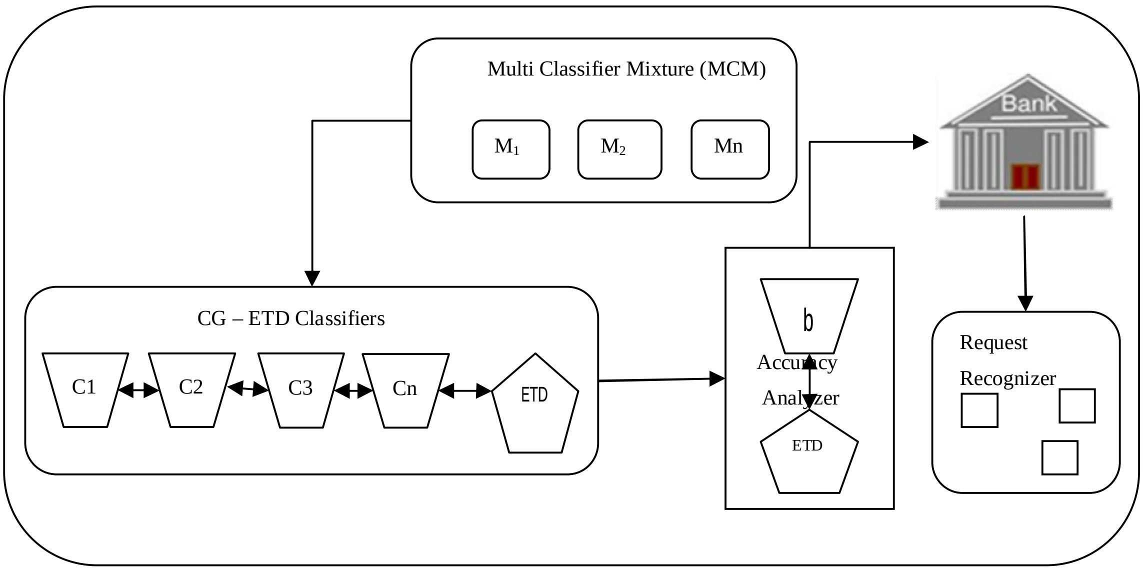


**Figure 5.1 Layer 2 Technique of Classified Filter**

In first layer secure information is executed by the Enforcer algorithm and moves on to the second layer, where the first sub component MCM receives the bulk arrival and allocates the bulk information to each individual MCM sub packages. The sub packages are next moved to CG-ETD classifiers. Here the bulk transactions are processed using CG and ETD classifiers. Accuracy rates are measured by Accuracy Analyzer and the results are carried or sent to the service provider services through bank progression.

### 5.2.1 Multi Classifier Mixture

In Multi Classifier Mixture (MCM), bulk information is received initially are packaged. The design of this package is such that it receives all the generated data from layer 1 and are distributed to its small sub packages. Each package contains information such as service Id, password, bill no and card number. Each small sub packages are named as $M_1$, $M_2 \ldots M_n$. The proposed algorithm for MCM mixture is given below. Figure 5.2 displays the MCM algorithm with functional steps.

| | |
|---|---|
| Step 1: | Check for satisfaction of Requirements. |
| Step 2: | If the authenticated data (objweka.PreAuthenticate = true) is true then the credential data passed to the classifiers. |
| Step 3: | Service the clients to purchase items, with the credential Information's of shopcode, Billno, cardno and amount. Here the, credential information results are concatenated with the classifiers using Weka classifier using the following code.<br>CentralizedBank.weka.WekaClassifierService objweka = new CentralizedBank.weka.WekaClassifierService();<br>objweka.PreAuthenticate = true;<br>objweka.Credentials= System.Net.CredentialCache.<br>DefaultCredentials; |
| Step 4: | The bulk of credential data is<br>(objweka.getallResult1(transaction Id, Service Id, password, amount))<br>retrieved from the customers and passed to the classifiers of CGP and ETD. |

**Figure 5.2 MCM Algorithm**

### 5.2.2 CG -ETD Classifiers

All the credential information that is received from the MCM layer is next fed into CG-ETD classifiers. These classifiers are descriptively known as Collective Group (CG) Classifiers and Efficient Trim Down (ETD) Classifiers. Collective Group (CG) indicates the existing classifiers of J48, Random Tree, Random Forest and AD Tree. The existing classifiers are compared to test the accuracy measure with the proposed ETD classifier. The proposed and existing classifiers are discussed with their functionality and algorithm.

#### 5.2.2.1 Collective Group Classifiers

The existing Collective Group (CG) Classifiers was considered in order to compare then along with the proposed classifiers shown in Figure 5.3

##### 5.2.2.1.1 J48

This is the first classifier that receives the training data set as input. Two nodes are used for assigning information namely 1) vector node and 2) base node. The base node is a node which is considered to be having the highest normalized base and the remaining data are fixed as vector nodes. Also based on the prediction vector, nodes are assigned in the base node. For layer 2 transaction, password, amount, transaction Id and Service Id are included in training data set and the highest normalized base among these parameters is identified. Thus, the training dataset is classified using J48 classifier.

#### 5.2.2.1.2 Random Tree

In random tree classifier, the training or sample data is allotted using decision trees. The trees have multiple sub branches from which nodes are agreed randomly. The random allocation based on the branch allocation. The filtered nodes are assigned using an acceptability parameter.

#### 5.2.2.1.3 Random Forest

Random Forest classifier has fixed classifier variables. The sample training inputs are matched with these classifier variables. During matching, classification takes place only when if the sample training data is less than classifier data. If not, the sample data get discarded and are not assigned with the nodes. In such way nodes are assigned and discarded using this classifier.

#### 5.2.2.1.4 AD Tree

AD Tree classifier has two set of nodes. They are decision node and precision node. Decision node allocates the incoming retrieved samples. Precision node allocates the instance of the decision node. Based on node fixing, instance classes are defined for the main class. The received inputs are settled with their nodes in a formal way and are classified using accurate results. To make the comparison the training data is again input to the proposed Efficient Trim Down (ETD) classifier. Figure 5.3 displays the Collective Group classifier algorithm.

**CGP1: (C1) J48 Classifier:**

1. Read the training Data.
2. Augment the vector nodes with the dataset of shopecode, bill no, Card no and amount.
3. Identify the highest normalized base.
4. Base recursions are defined and add the sub nodes to the base using the following code.

```
String ret= WriteFile(shopecode,billno,CardNumber,amount);
String[] args={"CLASSIFIER","weka.classifiers.trees.J48","U",
"FILTER","weka.filters.unsupervised.instance.Randomize","DATASET",
"iris.arff"
```

**CGP2: (C2) Random Tree:**

1. Construct multiple Decision Trees.
2. Randomly allocate the parameters of shopecode, billno,Card no, amount into the tree.
3. Filtered nodes are identified and tree class distributions are recorded with the following code.

```
String ret= WriteFile(shopecode,billno,CardNumber,amount);
String[]
args={"CLASSIFIER","weka.classifiers.trees.RandomTree",
"FILTER","weka.filters.unsupervised.instance.Randomize","DAT
ASET","iris.arff"};
```

**CGP3: (C3) Random Forest:**

1. Read the training data set and define the classifier variables.
2. Check the input variables shope code, billno,Cardno, amount with the classifier variable.
3. Input variables should less than the classifier variables.
4. Random allocation provide for the Input variables.
5. Pruning is not permissible. And the following code describes the classification.

```
String ret= WriteFile(shopecode,billno,CardNumber,amount);
String[]
args={"CLASSIFIER","weka.classifiers.trees.RandomForest",

"FILTER","weka.filters.unsupervised.instance.Randomize","DATASET",
"iris .arff"};
```

**CGP4: (C4) AD Tree:**

1. Define the Decision nodes and precision nodes.
2. Assign the inputs shopecode, billno, cardno, amount to decision nodes.
3. Calculate the instance based on the traverse of the prediction nodes.

Code depicts the processing of classification.

```
String ret= WriteFile(shopecode,billno,CardNumber,amount);
String[]args={"CLASSIFIER","weka.classifiers.trees.ADTree",
"FILTER","weka.filters.unsupervised.instance.Randomize","DATASET","iris.arff"}
```

**Figure 5.3 Collective Group (CG) Classifier Algorithm**

### 5.2.2.2 Efficient Trim Down Classifier

In this Collective Group classifiers, the accurate results are matched with the ETD classifier. This is the proposed classifier, it evaluates the filtered data based on the Empirical estimation, Complexity and Risk analysis. ETD classifier uses Efficient Accuracy (EA).

The formula for Efficient Accuracy (EA) is as follows.

$$\boxed{\text{Efficient Accuracy (EA)} = \text{Remp}() + (h(\log(2m*a/h + 1) - \log(n/4))/(m*a))^{\wedge}(1/2)}$$

where…

| | | |
|---|---|---|
| Remp | - | is the Empirical estimation of risk factors |
| m | - | number of records input from the client requests |
| h | - | Complexity of classification |
| Logarithm (log) | - | used to define the accuracy of the collected data. For example, Probability distribution a value 2 has 1.414 and 3 have 1.752. |
| a | - | is the stated Error rate of input attributes, where the number of attributes are directly proportional to the error risks. When the number of attributes increases, the error rate and the accuracy are also increased. |

At first, the Empirical estimation is calculated based on the data analysed. Here the comparison happens with previously analyzed sample data and with the filtered input data. The estimated value is further calculated with data complexity and risk analysis. After making estimation, the complexity of the data calculated by the logarithmic method of h(log2m*a/h+1) – log(n/4), where ‘h’ represents the complexity of classification and logarithm is used to calculate the number of received records of ‘m’ and ‘n’ for attributes present in the records. Logarithm is used to reduce the unmanaged value of input parameters. For example, probability distribution of value 2 is 1.414, 3 is 1.752 such that different range of arrived input data is calculated with the probability.

Finally, the estimated Empirical estimation and complexity measure of data are divided with the total count of data records. It is represented as (m*a)) ^ (1/2). When the classifier receives data beyond bounds the scheming process stops automatically. The following Figure 5.4 demonstrates the procedural steps of ETD Classification algorithm.

R1: Define the inputs shoppecode, billno, CardNumber, amount with the value of n.
R2: Calculate the Empirical estimation based on the risk factors.
R3: Match the predicted data with input data.
R4: Estimation of risk factors defined with
    Ramp( ) ->{ Emperical Estimation for risk factors}.
R5: Complication of the data is analyzed with the complexity of “h”.
R6: Identify the Error rate and match the input data with the measured Probabilities.
    It is defined as (log(2m*a/h + 1)
Where a-> is the stated error rate of input attributes which is directly proportional to the error risks.
R7: Estimate the probability estimation with respect to the input variables and assume the probability ratio to produce accurate error rate of log(n/4).
R8: Finalize the total count of input records with m/a.
R9: Calculate the logarithm probability with parameter of ½.
R10: Finally, find the better accuracy using Efficient Accuracy (EA).

**Figure 5.4 ETD Classification Algorithm**

### 5.2.3 Accuracy Analyzer

CG-ET classifiers are set of Collective Group classifiers, where each classifier has its own accurate results. Similarly, ETD classification also does filtering and classifies the results based on the Empirical estimation, Complexity and Risk analysis. All the classifier results are fed to the Accuracy Analyzer. The best results are compared and produced in the Analyzer, where it chooses the accurate results among the CG classifiers and judge their results along with ETD classifier’s result. In the test experiment, the proposed ETD classifier outputs more accurate results. Finally the results are passed on to bank progression.

### 5.2.4 Bank Progression

Customer data are frequently executed using bank progression. Based upon the user account details, transactions are done or carried out either with nationalized bank or internationalized bank. Next, the generated data is verified during bank progression and moves on to the Request Recognizer. Finally, the huge amount of transactions is sequentially executed and sent to RR as a small package.

#### 5.2.4.1 Request Recognizer

The major functionality of Request Recognizer (RR) shown in Figure 5.5 is to reduce the bulk arrival of data and transmit only the authenticated data to the preferred service provider. Also RR contains multiple bulkers which receives and redirects the arrival of filtered data to the preferred service providers. Figure 5.5 displays Request Recognizer algorithm with its functionalities.

Step1: Stream the accuracy and validation results from CG-ETD Classifiers.
Step2: Compare the results in the Accuracy Analyzer (AA).
Step3: Pass the finalized accuracy results to Request Recognizer.
Step4: Bulk data get split in the RR with different bulker.
Step5: If Accuracy is not matched, invalid reply is sent from the bank processing.
Step6: Otherwise, finalized data pass to the customer in a Queue way.
The processing is described with the code.

```
if(obj.CommandQuery("Insert into
Tbl_TransactionDetails(AccNo,Amount,Date,ChequeNo) values(" +
Accountno + "," + Amount + ",'" + DateTime.Now.ToShortDateString()
+ "','" + ChequeNo + "')"))
{ obj.CommandQuery("Insertinto Tbl_DepostDetails(AccNo,Date,Amount)
values(" + dsOrgFrom.Tables[0].Rows[0][0].ToString() + ",'" +
DateTime.Now.ToShortDateString() + "'," + TxtAmount.Text.Trim() + ")");
obj.CommandQuery("Update Tbl_CheckDetails set Status=1 where CheckNo='"
+
ChequeNo + "' and AccountNo_FK=" + Accountno + "");
Result = "Amount CreditSucessfully"; }
else
{ Result = "Error Occured";}}
else { Result = "Sorry Insufficent Balance";}}}}
else { Result = "Cheque Bounces";}}
else{
Result = "Invalid Cheque Account Number"; }}
else{ Result = "Invalid Credit Account No";}
```

**Figure 5.5 Request Recognizer Algorithm**

## 5.3 SAMPLE VALUE ANALYSIS FOR LAYER 2

As viewed in the Analyzer Security Originator, the generated credential information is filtered and passed on to the Classified Filter. This filter reduces the source using the proposed Efficient Trim Down (ETD) classifier. The classified filter layer executes with the example:

- In the initial processing step collect the credential information of customer service Id ‘HP200’,Transaction Id ‘2325’, password ‘12#77gg’and purchase amount ‘8000’ is assigned as the value of ‘n’. Such as n1= ‘HP200’ , n2= ‘2325’, n3= ‘12#77gg’ and n4=‘8000’ produce the count value of n=count(n1,n2,n3,n4) = ‘4’.

- ‘m’ is the number of records to be input, so as for one customer ‘ m = 1’.
- $R_{emp()}$ is the Empirical estimation of the total risk factors when compared with the existing predicted results. With the comparison $R_{emp()}$ is found to be ‘3’thus it produces three risk factors than compared with the previous survey.
- ‘a’ – is the stated error rate of attributes. The count of attributes is directly proportional to the error rate. So the value of ‘a’ is 4 (service Id, transaction Id password, amount).
- ‘Log’- is to identify the logarithm and is used to measure the accuracy of data.
- ‘h’ is the complexity to be measured for the value ‘n’. It is found to be of ‘4’.
- Using all above values, the Efficient Accuracy (EA) is calculated as ..

EA = 4 + ( 4 (log ( 2 * 1 * 4/4 + 1)) – log (( 4/4 ) / (1/4) ) ^ (1/2)

= 4 + (4 (log(8/4+1))-log((1) / (0.25)) ^ 0.5

= 4 + (4 (log (8/5))-log(4) ^ 0.5

= 4 + (4(log (1.6))-0.6020 ^ 0.5

= 4+ (4(0.204))-0.6020 ^ 0.5

=4+ 0.816-0.6020 ^ 0.5

= 4+0.214 ^ 0.5

= 4.214 ^ 0.5

EA = 4.2 accuracy

- EA’s 4.2 accuracy measure was used to compare the accuracy measure of existing Collective Group classifiers (J48, Random Tree, Random Forest and AD Tree).
- Among the four collective classifiers, Random Tree produces 3.4 accuracy measures.
- The Accuracy Analyzer (AA) is used to the compare random tree classifier value of 3.4 and ETD’s accuracy measure value 4.2 accuracy measures and passes the credentials for bank progression.
- The Figure 5.6 displays the sample value execution of the proposed ETD classification.

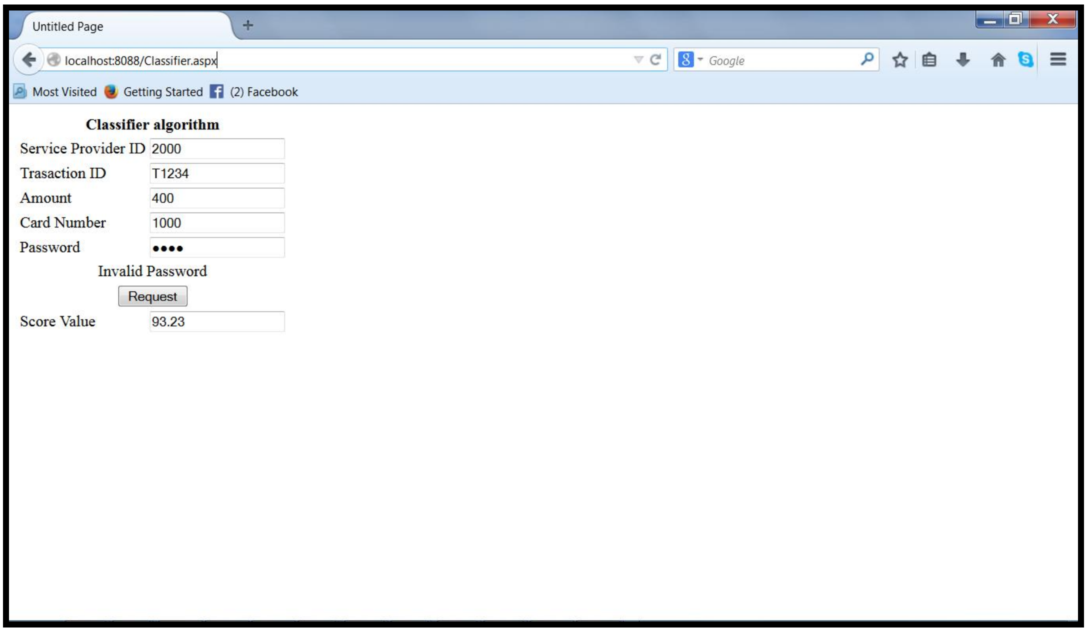


**Figure 5.6 Proposed Efficient Trim Down (ETD) Execution**

## 5.4 EXPERIMENTAL TESTING OF LAYER 2

Layer 2 was tested with the concept of comparative classification for the proposed and existing classifiers. During this testing, 10,000 filtered samples were taken and two types of tests were experimented to evaluate Stratified Cross Validation and Accuracy Test Analysis. Stratified Cross Validation is used to classify the proposed ETD and CG classifiers with different measures. Accuracy Test Analysis measures the accuracy of the transaction data with respect to usage.

### 5.4.1 Stratified Cross Validation

The validation starts with 10,000 testing samples collected from Layer 1. The samples are first classified using Collective Group classifiers (J48, Random Tree, Random Forest and AD Tree) and then they are parallely executed with the proposed ETD classifier. The resultant test validation analysis is displayed in table 5.1.

During validation, the input sample instances are categorized as either correctly classified instance or incorrectly classified instance. Correct classification instances are stated with correct authenticated user data and incorrectly classified instance are defined with unauthenticated data that does not match with the source data. From the source of correct instance classification, J48 is 50.1292% data, Random Forest is 86.0465% data, Random Tree is 88.889% data and AD Tree is 41.3437% data. When comparing with these classifiers the proposed ETD classifier is 93.7% of correctly classified instance. Likewise for incorrectly classified instances, J48 is 49.8708% data, Random Forest is 13.9535% data, Random Tree is 11.111% data and AD Tree is 58.6563% data. The proposed ETD classifier is 63.25% data when compared with the other classifiers.

**Table 5.1 Stratified Cross Validation of CG-ETD classifiers**

| Verification | CLASSIIFIERS | | | | |
|---|---|---|---|---|---|
| | **J48** | **Random Forest** | **Random Tree** | **AD Tree** | **ETD** |
| Correctly classified Instance | 50.1292 % | 86.0465 % | 88.8889 % | 41.3437 % | 93.7% |
| Incorrectly classified Instance | 49.8708 % | 13.9535 % | 11.1111 % | 58.6563 % | 63.25% |
| Kappa Statistic | 0.128 | 0.7209 | 0.7778 | 0.173 | 0.106 |
| Mean Absolute Error | 0.5 | 0.3279 | 0.1111 | 0.5105 | 0.0678 |
| Root mean squared error | 0.5 | 0.4149 | 0.3333 | 0.5121 | 0.2568 |
| Relative absolute error | 85 % | 65.5785 % | 22.2212 % | 92.0929 % | 11.65% |
| Root relative squared error | 88 % | 82.9827 % | 66.6633 % | 86.4052 % | 60.452% |
| Total No of Instance | 10000 | 10000 | 10000 | 10000 | 10000 |

During validation process, failure range analysis was measured using kappa statistic, Mean absolute error, Root mean squared error, Relative absolute error and Root relative squared errors. These error measures are used to measure the different categorical items and estimate the difference between estimated values and true values. From Table 5.1 it shows that kappa static errors for CG classifiers, J48 is 0.128, Random Forest is 0.7209, Random Tree is 0.778 and AD Tree is 0.173 with the comparison the proposed ETD classifier is 0.106 errors. The validation analysis shows that the proposed ETD classifier produces fewer amounts of errors than the existing CG classifiers.

Similarly, in testing of mean absolute error for CG classifiers, J48 is 0.5 errors, Random Forest is 0.3279 errors, Random Tree is 0.1111 errors and AD Tree is 0.5105 errors. Compared to the CG classifiers, ETD classifier is of fewer amounts of 0.0678 errors. The root mean squared error CG classifiers J48 is 0.5 errors, Random Tree is 0.4149 errors, Random Forest is 0.3333 errors and AD Tree is 0.5121 errors and comparatively the proposed ETD classifier is 0.2568 errors. For the Relative absolute error for J48 is 85% errors, Random Forest is 65.5785% errors, Random Tree is 22.2212% errors and AD Tree is 92.0929 % errors and the proposed ETD is 11.65% errors. For the root relative squared error testing CG classifiers J48 is 88% errors, Random Forest is 82.9827% errors, Random Tree is 66.6633% errors and AD Tree is 86.4052% errors and the proposed ETD classifier is 60.452% errors. From all the above analysis, it can be seen that the proposed ETD classifier produce fewer amounts of errors and has the best analysis results, than the existing CG classifiers.

### 5.4.2 Accuracy Measure Class

Based on the Stratified classification represents error measures and comparative classified data analysis for CG classifiers and ETD classifier. Accuracy Measure Analysis was based on whether the user account is male or female. According to the user account details, the accuracy measure was collected as per the gender test.

**Table 5.2 Accurate Measure Results for Male and Female Dataset**

| | **Male** | | | | | **Female** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Accuracy** | **J48** | **Random Forest** | **Random Tree** | **AD Tree** | **ETD** | **J48** | **Random Forest** | **Random Tree** | **AD Tree** | **ETD** |
| TP Rate | 0.237 | 0.881 | 0.897 | 0.397 | 0.932 | 0.7321 | 0.839 | 0.881 | 0.43 | 0.903 |
| FP Rate | 0.46 | 0.161 | 0.119 | 0.57 | 0.183 | 0.322 | 0.119 | 0.103 | 0.603 | 0.130 |
| Precision | 0.501 | 0.847 | 0.883 | 0.412 | 0.927 | 0.633 | 0.876 | 0.895 | 0.415 | 0.901 |
| Recall | 0.432 | 0.881 | 0.897 | 0.397 | 0.952 | 0.754 | 0.839 | 0.881 | 0.43 | 0.940 |
| F-Measure | 0.668 | 0.864 | 0.89 | 0.404 | 0.913 | 0.821 | 0.857 | 0.888 | 0.422 | 0.913 |

Table 5.2 Displays the Accuracy Calculated with respect to True Positive rate, False Positive rate, Precision, Recall and F-Measure.

#### 5.4.2.1 Male Data Set Analysis

In male data set analysis, TP rate estimation for J48 classifier is 0.237, Random Forest is 0.881, Random Tree is 0.897 and AD Tree is 0.397. Comparatively ETD classifier produces high accuracy with the value of 0.932 than the CG classifiers. Secondly FP rate measure for J48 is 0.46, Random Forest is 0.161, Random Tree is 0.119 and AD Tree is 0.57. When compared with ETD, it produces 0.183 measures than the existing CG classifiers. Similarly, Precision, Recall and F-Measure was estimated for both male and female dataset.

Precision rate is defined as the retrieved instances that produce relevant results from the given sample instances. Precision rate measured for CG classifiers of J48 is 0.501, Random Forest is 0.847, Random Tree is 0.883 and AD Tree is 0.412. Compared with CG classifiers, the proposed ETD has the highest accurate measure of 0.927. Recall is the authenticated information that is filtered from the sample data analysis. In Recall measure, CG classifiers J48 filter 0.432 data, Random Forest filter 0.881 data, Random Tree produce 0.897 data, AD Tree produce 0.397 data. Compared with CG classifiers, ETD has an accuracy rate of 0.952. In F-measure, the failure rate measure was calculated with the combination results of precision and recall. F-measure for J48 is 0.668, Random Forest is 0.864, Random Tree is 0.89, AD Tree is 0.404. Compared with CG classifiers, ETD classifier has an accuracy rate of 0.913.

#### 5.4.2.2 Female Data Set Analysis

As the male data set analysis, the female dataset is tested for their measures of TP rate, FP rate, Precision, Recall and F-measure. In TP rate analysis, CG classifier of J48 is 0.7321, Random Forest is 0.839, Random Tree is 0.881 and AD Tree has 0.43. When compared with CG, the proposed ETD classifier has a higher accuracy rate with the value of 0.903. In FP rate analysis, J48 is 0.322 error rate, Random Forest is 0.119, Random Tree is 0.103 and AD is 0.603. The comparative analysis of CG with ETD produces a lesser error value of 0.130. Precision Measure for J48 is 0.633, Random Forest is 0.876, Random Tree is 0.895 and AD Tree is 0.415. Compared with CG the ETD classifier has an accuracy value of 0.901.

The recall measure value for J48 is 0.754, Random Forest is 0.839, Random Tree is 0.881 and AD Tree is 0.43. In comparison, ETD an accuracy value of 0.940. Finally, F-measure for J48 is 0.821, Random Forest is 0.857, Random Tree is 0.888 and AD Tree is 0.422. The proposed ETD algorithm has an F-Measure value of 0.913.

## 5.5 PERFORMANCE ANALYSIS OF CG AND ETD CLASSIFIERS FOR MALE, FEMALE DATASET

The accuracy rate ratio for both male and female is represented graphically. Figure 5.7 shows how the frequency ranges from CG and ETD classifiers with respect to male users. The graph represents different measures such as with TP rate, FP rate, Precision, Recall and F-measure. The classifiers are represented using multiple colors. Dark blue represents J48, red Random Forest, green Random Tree, violet curve for AD Tree and light blue ETD.

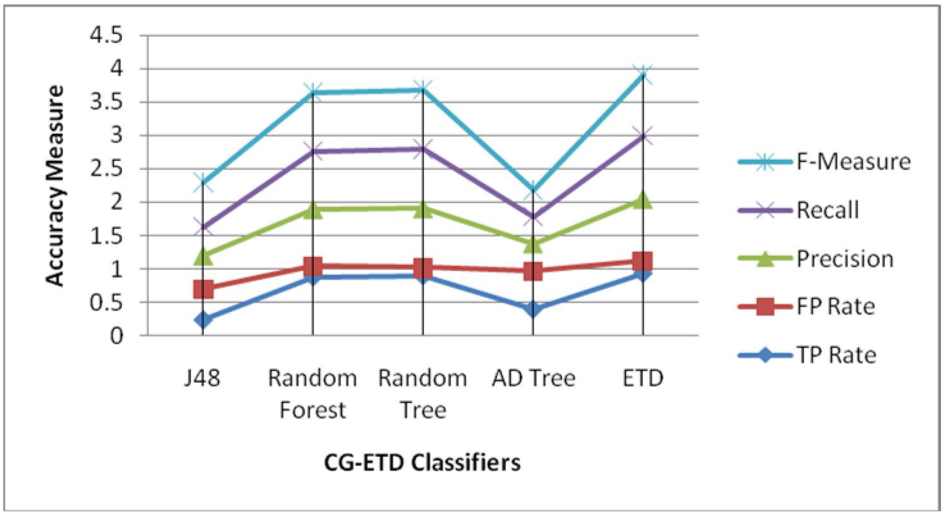


**Figure 5.7 Accuracy Measure Comparison for Male Dataset**

Compared to CG classifier ETD classifier occurs with the sequence curve. Also, ETD classifier has an accurate cervical representation than CG classifiers. The decreased deep curve of ETD classifier denotes the decreased error range, when compared with CG classifiers.

Figure 5.8 displays the accuracy rates for both CG classifiers and ETD classifier with respect to female users. CG classifiers are represented with different color namely blue for J48, Random Tree for red curve, Random Forest for green, AD Tree for violet and ETD classifier for light blue curve. When compared with CG classifiers, ETD classifier produces a low bending curve. When compared with existing classifiers, the proposed ETD classifier has a low deep curve which represents fewer errors. With the overall comparison the proposed ETD classifies the bulk data accurately than with the existing CG classifiers.

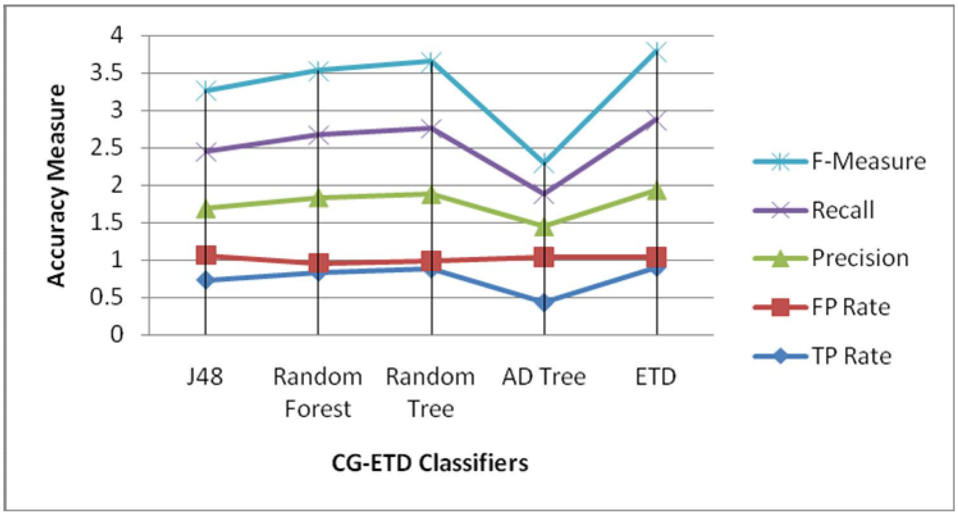


**Figure 5.8 Accuracy Measure Comparison for Female Dataset**

## 5.6 PICTORIAL VIEW REPRESENTATION

Pictorial view represents the evaluation of the existing Collective Group and proposed ETD classification. Figure 5.9 shows the Card Authentication process.

### 5.6.1 Card Authentication

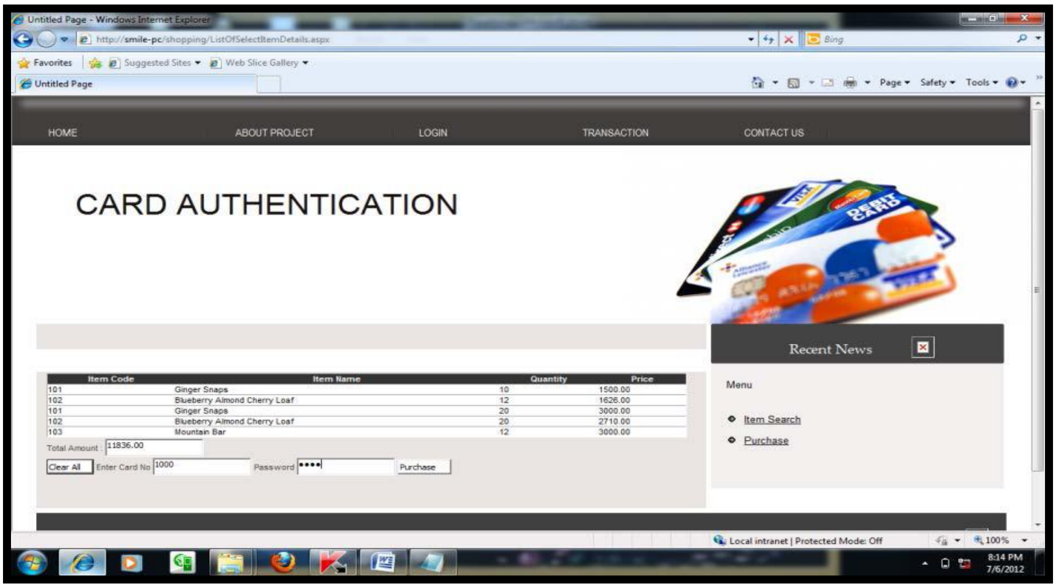


**Figure 5.9 Card Authentication Details from Bank Database**

### 5.6.2 Client Details in SQL Database

Figure 5.10 shows the client details which are stored in the SQL database

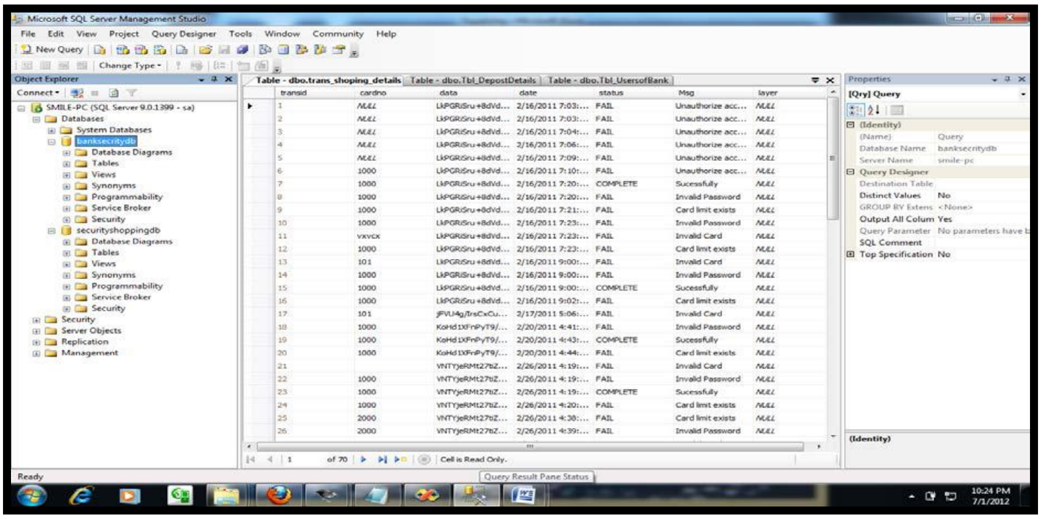

**Figure 5.10 Client Details in SQL Database**

### 5.6.3 J48 Classifier

Figure 5.11 shows the output of J48 classifier

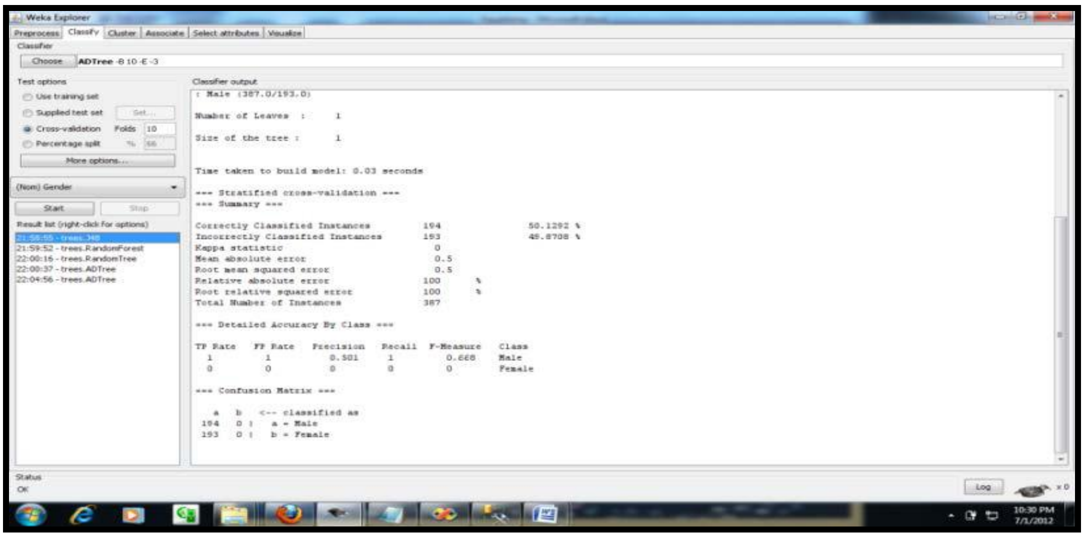

**Figure 5.11 Sample Data Execution in J48 Classifier**

### 5.6.4 Random Forest

Figure 5.11 shows the output of J48 classifier

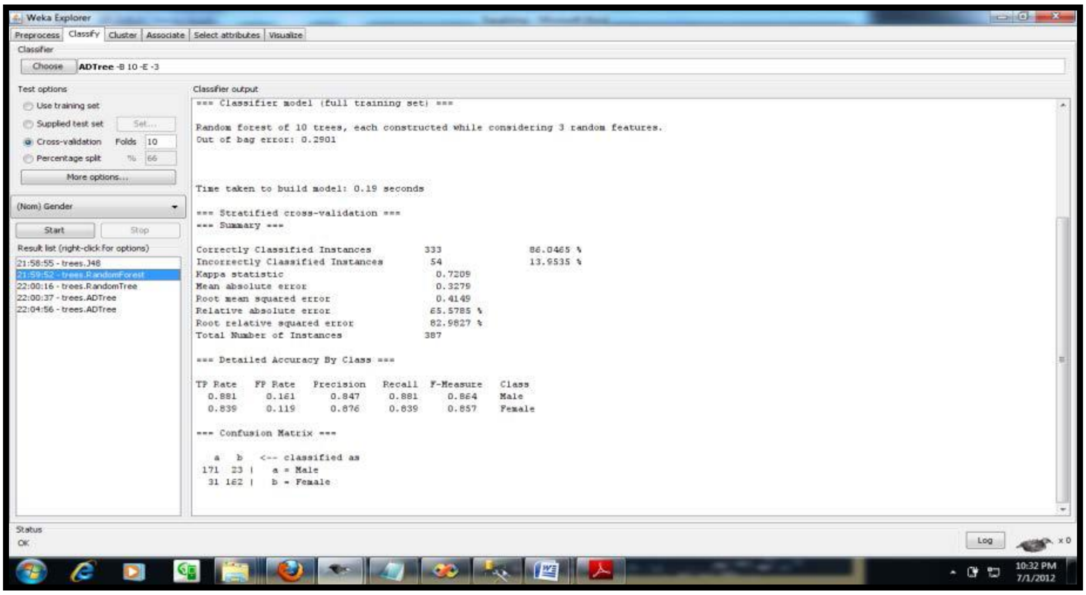


**Figure 5.12 Sample Data Execution in Random Forest Classifier**

### 5.6.5 Random Tree

Figure 5.12 shows the output of Random Forest algorithm

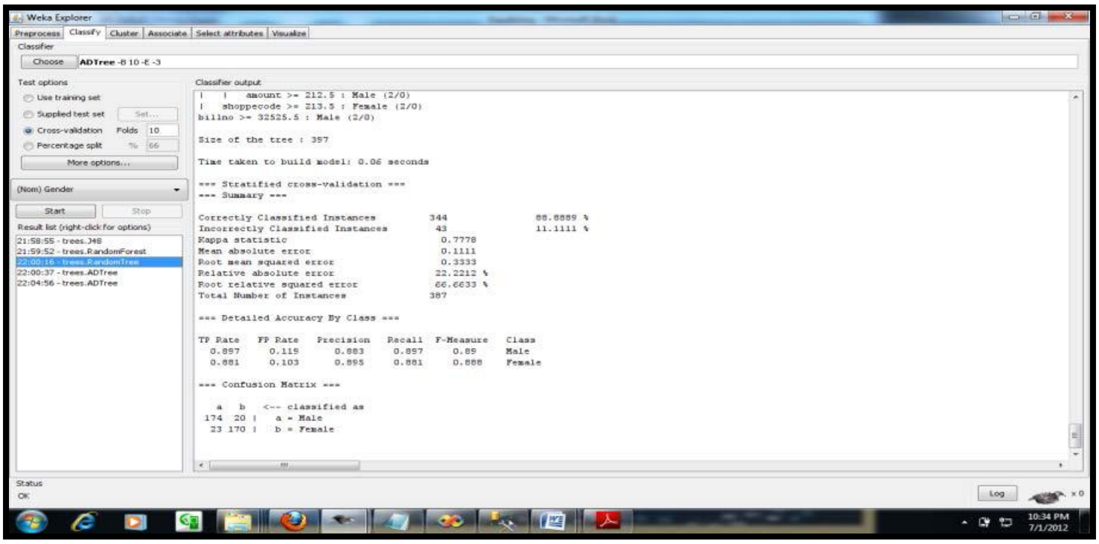


**Figure 5.13 Sample Data Execution in Random Tree Classifier**

### 5.6.6 AD Tree

Figure 5.14 shows the output of AD Tree algorithm.

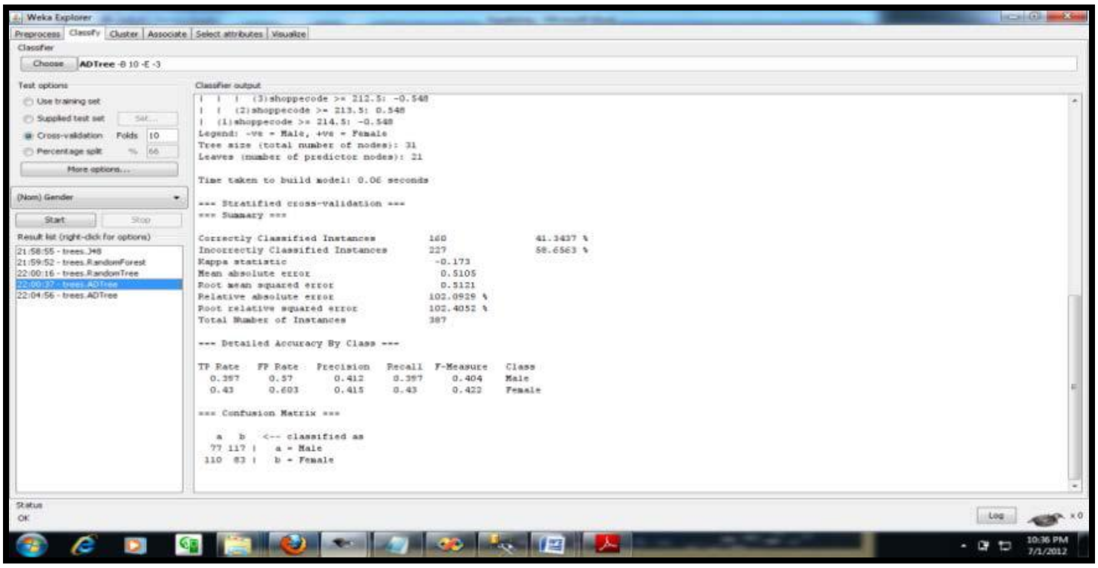


**Figure 5.14 Sample Data Execution in AD Tree Classifier**

### 5.6.7 Proposed Efficient Trim Down Classification

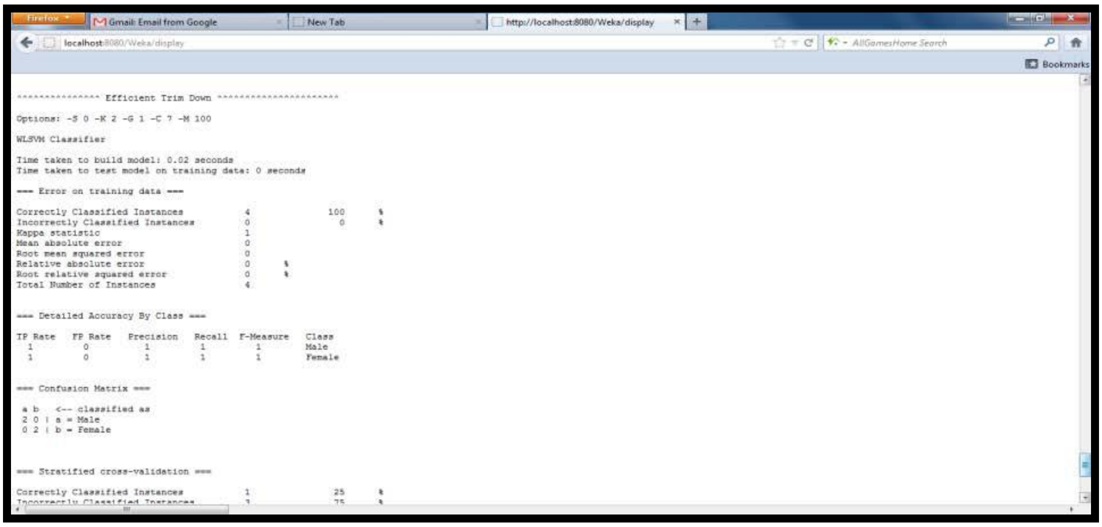


**Figure 5.15 Sample Data Execution in Proposed ETD Classification Algorithm**

Figure 5.15 shows the output of the ETD classification algorithm. Figure 5.11 to 5.15 it is observed that the classification accuracy is more in ETD classifier.

## 5.7 SUMMARY

This chapter discusses about the bulk reduction of credential information from layer 1.The Analyzer Security Originator evaluates and produces authenticated credentials for layer 2. Efficient Trim Down (ETD) classifier classifies the input data based on the Empirical estimation, Risk analysis and Complexity. Also, the bulk data is equally distributed to the CG classifiers using the proposed ETD classifier.

Existing classifiers mostly extract input data, classify the input and allocate the data to corresponding nodes. Whereas, the proposed Efficient Trim Down (ETD) classifier first analyzes the filtered data, then it carries out its processing with error factors. During execution, un-authenticated data if found, is discarded and resent to the customers. Only authenticated data is executed and passed on to layer 3 executions. This chapter focuses on new classification ETD algorithm and reduces the vast amount of credential information in a predetermined period of time.