An Automatic Weight Calculation Based Associative Classifier

Prepared By

Swati Gupta

12MCEC38

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2014

An Automatic Weight Calculation Based Associative Classifier

Major Project

Submitted in partial fulfillment of the requirements

For the degree of

Master of Technology in Computer Science and Engineering

Prepared By

Swati Gupta

(12MCEC38)

Guided By

Prof. Sapan H. Mankad

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2014

Certificate

This is to certify that the Major Project Report entitled ""**An Automatic Weight Calculation Based Associative Classifier"** submitted by **Swati Gupta (Roll No: 12MCEC38)**, towards the partial fulfillment of the requirements for the degree of Master of Technology in Computer Science and Engineering of Nirma University, Ahmedabad is the record of work carried out by her under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination. The results embodied in this major project, to the best of my knowledge, haven't been submitted to any other university or institution for award of any degree or diploma.

Prof. Sapan H. Mankad
Guide & Assistant Professor,
CSE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Prof. Vijay Ukani
Associate Professor
Coordinator M.Tech - CSE
CSE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Dr. Sanjay Garg
Professor and Head,
CSE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Dr K Kotecha
Director,
Institute of Technology,
Nirma University, Ahmedabad

Undertaking for Originality of the Work

I, **Swati Gupta**, Roll. No. **12MCEC38**, give undertaking that the Major Project entitled "**An Automatic Weight Calculation Based Associative Classifier**" submitted by me, towards the partial fulfillment of the requirements for the degree of Master of Technology in **Computer Science & Engineering** of Nirma University, Ahmedabad, is the original work carried out by me and I give assurance that no attempt of plagiarism has been made. I understand that in the event of any similarity found subsequently with any published work or any dissertation work elsewhere; it will result in severe disciplinary action.

_____

Signature of Student

Date:

Place:

Endorsed by

Prof. Sapan H. Mankad

(Signature of Guide)

Acknowledgements

It provides for me tremendous delight in communicating much appreciated and significant appreciation to **Prof. Sapan H. Mankad**, Assistant Professor, CSE Department, Institute of Technology, Nirma University, Ahmedabad for his important direction and consistent consolation all around this work. The gratefulness and consistent help he has conferred has been an extraordinary inspiration to me in arriving at a higher objective. His direction has activated and nourished my learned development that I will profit from, for quite a while to come.

My deepest appreciation is reached out to **Prof. Vijay Ukani**, PG CSE - Coordinator, Division of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad for an outstanding backing and consistent consolation all around the Major Project.

It provides for me a tremendous delight to thank **Dr. Sanjay Garg**, Hon'ble Head of Computer Science and Engineering Department, Institute of Technology, Nirma University, Ahmedabad for his kind help and giving fundamental framework and solid exploration nature.

An exceptional appreciation is communicated wholeheartedly to Dr K Kotecha, Hon'ble Director, Foundation of Technology, Nirma University, Ahmedabad for the unmentionable inspiration he has stretched out all around course of this work.

I might additionally thank the Institution, all working parts of Computer Engineering Office, Nirma University, Ahmedabad for their exceptional consideration and recommendations towards the task work.

The gifts of God and relatives make the route for consummation of Project. I am really thankful to them.

- Swati Gupta

12MCEC38

Abstract

Late studies in grouping have proposed methods for exploiting the association data mining standard. These studies have performed broad examinations to show their strategies to be both proficient and faultless. Be that as it may, existing studies in this standard it is possible that don't give any hypothetical avocation behind their methodologies or expect inreliance between a few parameters. Associative classification is another order methodology incorporating acquaintanceship mining and characterization. It turns into a noteworthy apparatus for information disclosure and information mining. This new approach coordinates association mining and classification into a solitary framework. Association mining, or example disclosure, points to uncover enlightening information from database, while arrangement concentrates on building an arrangement model for classifying new information. All around, both cooperation design finding and arrangement principle mining are fundamental to commonsense information mining applications. Impressive exertions have been made to incorporate these two strategies into one framework. Cooperation principle mining is a standout amongst the most essential and generally inquired about techniques of information digging for graphic assignment, at first utilized for business crate dissection. It discovers all the guidelines existing in the transactional database that fulfill some base support and least trust stipulations. Grouping utilizing Association tenet mining is an alternate real Predictive investigation procedure that expects to uncover a little set of principle in the database that structures a precise classifier.

In this dissertation, another system for figuring of weights in weighted associative order by presenting the idea of Value Frequency (VF) and Inverse Class Recurrence (ICF) in order of social database is proposed. It likewise gives the programmed figuring of weights of every thing in the social database for forecasts. The VFICF is much like the TFIDF idea utilized as a part of report order. It utilized the significance of a quality which is extraordinary in different classes yet visit in a specific class.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Classification and association standard mining are two vital information mining methods. Any classification strategy utilizes a set of characteristics or parameters to describe every object, where these characteristics ought to be important to the current workload. We consider here routines for directed classification, implying that a human master has both decided into what classes an item may be sorted and additionally has given a set of example objects with known classes. This set of known items is known as the preparation situated in light of the fact that it is utilized by the classification projects to figure out how to classify objects. There are two stages to constructing a classifier. In the preparation stage, the preparation set is utilized to choose how the parameters should be weighted and joined together with a specific end goal to divided the different classes of articles. In the application stage, the weights decided in the preparation set are connected to a situated of questions that don't have known classes with a specific end goal to focus what their classes are liable to be. Association guideline mining discovers all rules in the database that fulfill some base backing and least confidence constraints [2]. In association principle mining, the target is not decided ahead of time for mining, while there is one and only one decided beforehand focus in classification, i.e., the class label [11]. Association mining plans to find expressive information from database along these lines known as example disclosure, while classification concentrates on building a model for ordering the data [11]. In down to earth information mining application both association design revelation and classification are essential. Hence, one can get extraordinary investment funds and conveniences if these two mining systems can by one means or another be integrated [12]. Such a coordinated system is known as association classification. This

schema could be fabricated proficiently and without misfortune of execution.

The integration is done by concentrating on an extraordinary subset of association rules whose right- hand-side are confined to the classification class property. We allude to this subset of rules as the class association rules (Cars) [1]. Another classification approach known as association classification coordinates association mining and classification into a solitary system. Association mining, or example disclosure, means to run across unmistakable learning from database, while classification concentrates on building a classification model for categorizing new information. Generally, both association design revelation and classification tenet mining are crucial to down to earth information mining applications. Considerable endeavors have been made to coordinate these two systems into one framework.

By concentrating on a constrained subset of association rules, i.e. those rules where the consequent of the tenet is confined to the classification class trait, it is conceivable to manufacture more exact classifiers. We are attempting to use association guideline revelation techniques to construct classification frameworks. It has as of recently been demonstrated that the Associative Classifiers are performing admirably than traditional classifiers methodologies, for example, decision tree and guideline induction. A mixed bag of methods is produced through the integration of association rules and classification referred to as affiliated classifications, for example, Classification Based on Associations (CBA), Mining Car Association Rules (MCAR), and Classification based on Multiple Association Rules (CMAR). Yet at the same time there is extent of change. We are concentrating on enhancing the consistency of existing Associative classifiers.

Chapter 2

Associative Classification

Associative Classification(AC) mining is a guaranteeing approach in information mining that uses the affiliation principle disclosure procedures to build arrangement frameworks. Associative Classification (AC) is an extension of a bigger range of experimental study known as information mining. Air conditioning coordinates two known information mining undertakings, cooperation tenet disclosure and grouping, to fabricate a model (classifier) with the end goal of forecast. Arrangement and acquaintanceship guideline disclosure are comparable undertakings in information mining, with the exemption that the principle point of characterization is the expectation of class names, while cooperation standard revelation portrays correspondences between things in a transactional database.

Association rule mining discovers all principles in the database that fulfill some base support and least trust constraints [2]. For cooperation principle mining, the focus of mining is not foreordained, while for arrangement principle mining there is unparalleled one decided target, i.e., the class. Both characterization tenet mining and affiliation standard mining are imperative to pragmatic provisions. Along these lines, extraordinary investment funds and comforts to the client could come about if the two mining procedures can by one means or another be coordinated. In this dissertation, we propose such a coordinated schema, called affiliated characterization. The reconciliation is carried out by concentrating on an uncommon subset of affiliation runs whose right-hand-side are confined to the characterization class quality. We allude to this subset of leads as the class acquaintanceship tenets. Another characterization approach known as associative order coordinates companionship mining and characterization into a solitary framework.

Association mining, intends to uncover elucidating learning from database, while arrangement concentrates on building an order model for classifying new information. Impressive deliberations have been made to incorporate these two strategies into one framework. A normal affiliated arrangement framework is built in two stages:

1. uncovering all the occasion cooperations (in which the recurrence of events is huge as per a few tests)

2. creating arrangement principles from the cooperation examples to manufacture a classifier.

In the first stage, the taking in target is to uncover the cooperation designs intrinsic in a database (additionally alluded to as learning disclosure). In the second stage, the errand is to select a little set of important affiliation examples uncovered to develop a classifier given the anticipating property.

Chapter 3

Literature Survey

Numerous classifiers based on association rules have been proposed in the writing. The fundamental thought is to create standards with a solitary thing in the subsequent and to select guidelines with the characterized target trait happening at the ensuing. These tenets are known as Classification Association Rules. The forecast technique works by selecting leads whose forerunner blankets another occurrence (case) to be characterized. At that point, a request is forced on these standards as indicated by a measure, normally lead quality (e.g. certainty, lift). The best manage is decided to flame and the new case forecast is the ensuing of this tenet. This strategy is known as Bestrule expectation.

Another methodology focused around both positive and negative principles was additionally introduced[13]. The "interestingness" of the tenets was focused around the association coefficient that measures the quality of the direct relationship between a couple of variables. As information is not generally static in nature, it changes with time, so receiving sequential measurement to this will give more reasonable approach and yields much better comes about as the intention is to give the example or relationship among the things in time domain[14]. As the record has a place with one and only of the set brings about sharp limit issue which offers ascent to the idea of Fuzzy Association Rules (Far)[3][9]. Weighted cooperative classifiers which are focused around the not at all like characteristics weights are then introduced[15]. Each characteristic fluctuates regarding importance[4][5]. Weights might additionally change with the proficiencies of predicting[6].

Chapter 4

The Proposed Approach

Associative Classification is another grouping methodology incorporating acquaintanceship mining and grouping. It turns into a critical instrument for information disclosure and information mining.

The association classifier has positive qualities, quick preparing, great classification exactness, and phenomenal elucidation. A cooperative classifier is a classifier utilizing grouping decides that are transformed through a continuous example mining methodology from a preparing information accumulation. This procedure is the same one utilized as a part of conventional information digging for substantial log information of transactional database.

Plenty of research has been done in the area of associative classification but still there is scope of improving the accuracy and efficiency of associative classifier in many ways. We proposed the algorithm Value Frequency based Associative Classifier (VFAC) for:

- weighted associative classification
- automatic calculation of weights of each item
- introducing the concept of VF(Value Frequency) and ICF(Inverse Class Frequency) in relational database for the calculation of weight evolved from the concept of TFIDF in document classification

Term Frequency

- Term frequency(TF) is the easiest measure to weight each one term in a content.
- In this strategy, each one term is accepted to have criticalness corresponding to the number of times it happens in a content.
- Term recurrence of each one statement in a report is a weight which relies on upon the dispersion of each one statement in archives.
- It communicates the imperativeness of the statement in the archive.
- The weight of a term t in a content d is given by W(d, t) = TF(d, t); where TF(d, t) is the term recurrence of the term t in the content d.
- It is calculated as:
  [8]

$$TF(t_i, d_j) = \begin{cases} 0 & if\ n_{ij} = 0 \\ \frac{n_{ij}}{\sum_{k=1}^{m} n_{kj}} & if\ n_{ij} > 0 \end{cases} \tag{4.1}$$

- where n_{ij} is the number of times i^{th} term appeared in j^{th} document;
  k=1 to m;
  m is the number of terms in that document

Inverse Document Frequency

- While term recurrence concerns term event inside a content, Inverse Document Frequency (IDF) concerns term event over an accumulation of texts.[8]
- The natural significance of IDF is that terms which infrequently happen over an accumulation of writings are profitable.
- The vitality of each one term is thought to be conversely corresponding to the number of messages that hold the term.
- The IDF factor of a term t is given by:[8]

$$IDF(t_i) = log\frac{1 + |D|}{|D_{ti}|} \tag{4.2}$$

- where |D| is the number of documents;
  $|D_{ti}|$ is the number of documents containing i^{th} term.

TF/IDF

- TF/IDF is a system which utilizes both TF and IDF to focus the weight a term.[7]
- TF/IDF plan is exceptionally well known in content grouping field and very nearly the various weighting plans are variants of this scheme
- TF/IDF can be calculated using formula 4.3.

$$TF(t_i, d_j) * IDF(t_i) \tag{4.3}$$

4.1 Introduction to VF and ICF

Value Frequency(VF)

- In this method, each item in relational database is assumed to have importance proportional to the number of times it occurs in a class.
- It can be calculated as

$$VF(v_i, c_j) = \begin{cases} 0 & if\ n_{ij} = 0 \\ \frac{n_{ij}}{\sum_{k=1}^{m} n_{kj}} & if\ n_{ij} > 0 \end{cases} \tag{4.4}$$

- where VF(v_i,c_j) is the Value Frequency of i^{th} attribute value for class attribute c_j;
  n_{ij} is the number of times i^{th} attribute value appeared in tuples containing j^{th} class in class attribute;
  k=1 to m;
  m is the total number of attribute values present in tuples containing j^{th} class in class attribute.
- For example:

Record ID	Age	Smoke	Hypertension	BMI	Heart disease
1	42	YES	YES	40	YES
2	62	YES	NO	28	NO
3	55	NO	YES	40	YES
4	62	YES	YES	50	YES
5	45	NO	YES	30	NO

Table 4.1: Example of relational database

- In the given example calculation of VF($Hypertension_{YES}$, $Heartdisease_{YES}$) is as follows:

  n_{ij} is $n_{Hypertension_{YES},Heartdisease_{YES}} = 3$

  $\sum_{k=1}^{m} n_{kj} = 3$

  Therefore,

  VF($Hypertension_{YES}$, $Heartdisease_{YES}$)=3/3=1

- Similarly, VF($Hypertension_{NO}$, $Heartdisease_{YES}$)= 0

  VF($Hypertension_{YES}$, $Heartdisease_{NO}$)= 1/2=0.5

  VF($Hypertension_{NO}$, $Heartdisease_{NO}$)=1/2=0.5

Inverse Class Frequency(ICF)

- The importance of each value is assumed to be inversely proportional to the number of items that contain the value
- It basically finds out how rarely certain value is present in other classes.
- It can be calculated as:

$$ICF(v_i) = log\frac{1 + |C|}{|C_{vi}|} \tag{4.5}$$

- where |C| is the number of classes;

$|C_{vi}|$ is the number of classes containing i_{th} value.

- For the same above example, the calculation of ICF (v_i) is as follows:
  ICF($Hypertension_{YES}$)= log $\frac{1+|C|}{|C_{Hypertension_{YES}}|}$= log $\frac{1+2}{2}$= log(1.5)= 0.176
  ICF($Hypertension_{NO}$)= log $\frac{1+|C|}{|C_{Hypertension_{NO}}|}$= log $\frac{1+2}{1}$= log(3)= 0.477

VF/ICF

- It is used for assigning weights to each value in relation.
- It can be calculated as:
  VF(v_i,c_j)∗ICF(v_i)
- For the above example, the calculation of weight by VF/ICF is as follows:
  VF($Hypertension_{YES}$, $Heartdisease_{YES}$) ∗ ICF($Hypertension_{YES}$)= 1∗0.176=0.176
  VF($Hypertension_{YES}$, $Heartdisease_{NO}$) ∗ ICF($Hypertension_{YES}$)= 0.5 ∗ 0.176= 0.088
  VF($Hypertension_{NO}$, $Heartdisease_{YES}$) ∗ ICF($Hypertension_{NO}$)= 0*0.477=0
  VF($Hypertension_{NO}$, $Heartdisease_{NO}$) ∗ ICF($Hypertension_{NO}$)= 0.5∗0.477=0.2385

Interpretation of results:

As we can see in the given example, when Hypertension=No then class Heart disease is always No so the weight of Hypertension=No for class Heart disease=No must be high. According to the calculation proposed VF*ICF is also giving it high weight and as there are no chances when Hypertension=No and Heart disease=Yes so it is also giving it zero weight.

And also whenever Hypertension=Yes then mostly Heart disease=Yes and rarely present for class Heart disease=No so its weight for Heart disease=Yes must be higher than that of Heart disease=No. According to above calculation VF*ICF is giving higher weight to Hypertension=Yes and Heart disease=Yes than that of Hypertension=Yes and Heart disease=No.

Hence, it is theoretically proved that VF*ICF is giving right weights to each value.

4.2 The Proposed Algorithm

The flowchart of the proposed algorithm is shown in figure 4.1.

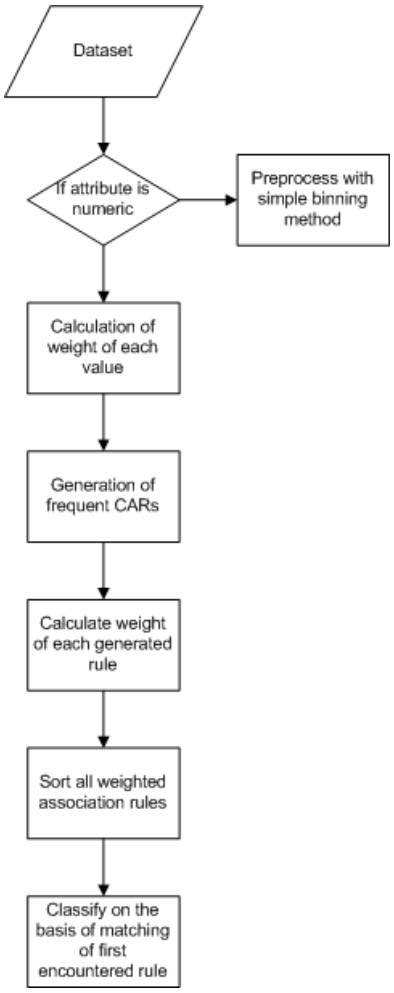

Figure 4.1: Flowchart of the proposed algorithm

The steps of proposed algorithm Value Frequency based Associative Classifier(VFAC) are as follows:

1. **Pre-processing of data :**
   Discretization of data must be done so as to make it suitable for applying to apriori algorithm.

2. **Calculation of weights of each distinct values** in the relation using introduced VF/ICF method.

3. **Generation** of frequent Class associative rules.

4. **Calculation of weights of each rule:**

   suppose a rule is such that:

   $\{v_1, v_2, v_3, v_4\} \rightarrow c_j$

   where $\{v_1, v_2, v_3, v_4\}$ is the condition set and C is the class; Then, the weight of this rule is= $\sum_{i=1}^{n} W(v_i)$

   where W(v_i) is the weight of value v_i= VF(v_i,c_j)$*$ICF(v_i);

   n is the number of values in the condition set

5. **Sort the weighted associative rules** according to decreasing precedence of each rule.

6. **Classify** on the basis of matching of first encountered weighted rule. If no rule is matched then the default class will be classified.

The precedence of each rule can be defined as:

- Given two rules, r_i and r_j, r_i precedes r_j or r_i has a higher precedence than r_j if:
  - The confidence of r_i is greater than that of r_j[1]
  - Or their confidences are equal, but the support of r_i is greater than that of r_j[1]
  - Or both the confidences and supports of r_i and r_j are the same but the weight of r_i is greater than that of r_j

Chapter 5

Results and findings

5.1 Data set used

Implementation is done with real ordinal data **SWD(Social Workers Decisions)**[10] which contains real-world assessments of qualified social workers regarding the risk facing children if they stayed with their families at home and is composed by ten features and four classes. The original donor of this dataset is: Arie Ben David, MIS, Dept. of Technology Management Holon Academic Inst. of Technology and the owner is Yoav Ganzah, Business Administration School, Tel Aviv Univerity. The data set is in the Arff format. The brief description about the data set is given below:

- Number of Instances: 100
- Number of Attributes: 10 inputs and 1 output
- All input attributes contains numeric values and it has 4 classes
- There are no missing values

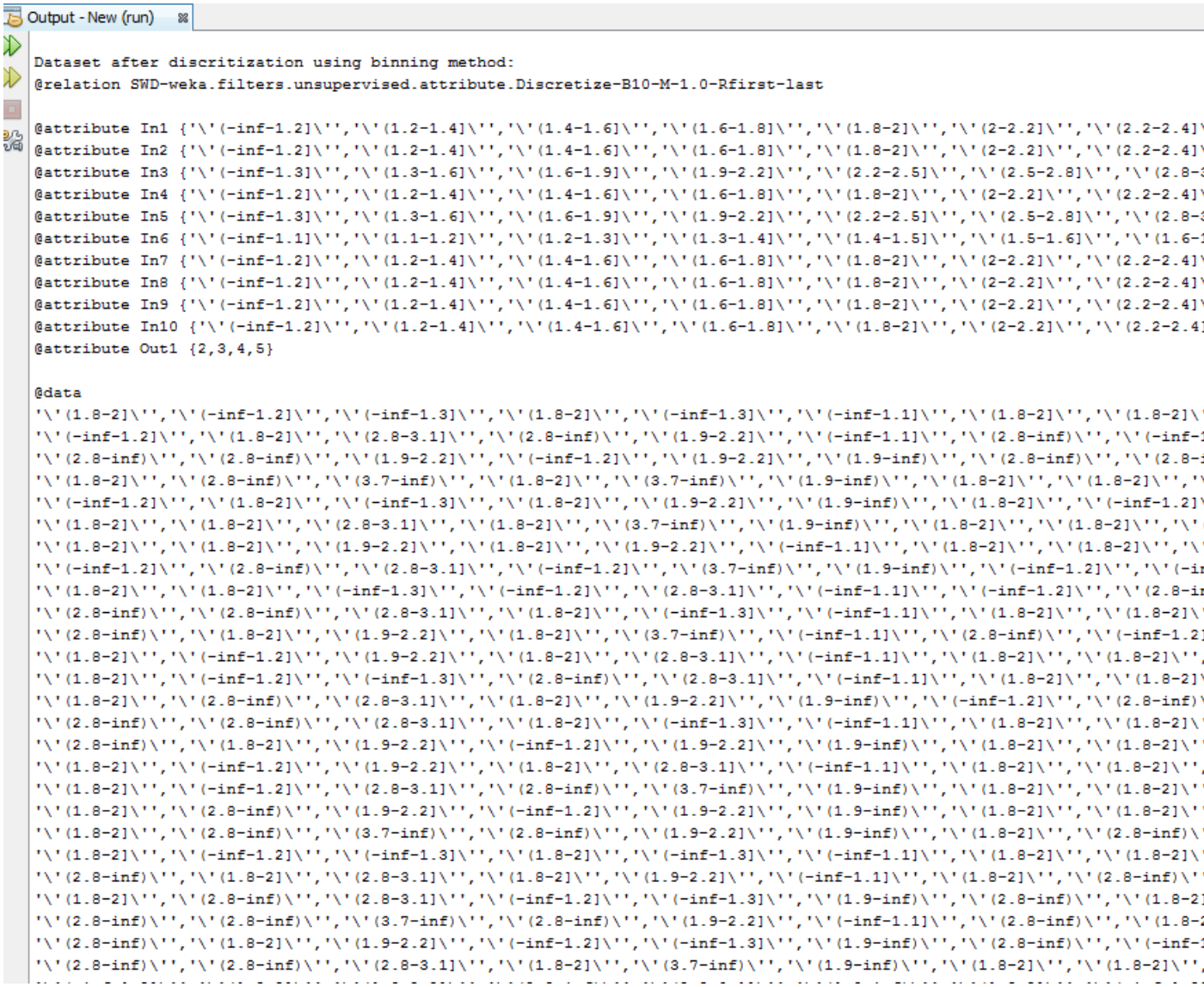

Figure 5.1: Result of discretization of dataset

Tools used

Netbeans IDE 7.3.1 tool is used for the implementation of the project. We have used Weka API in java for the implementation.

Results:

The proposed algorithm is implemented in following order. We:

1. **Pre-process** the standard dataset i.e. the data set is discretized using the simple binning method. The entire numeric attribute is discritized and the range of its value is divided into 10 bins as shown in figure 5.1.

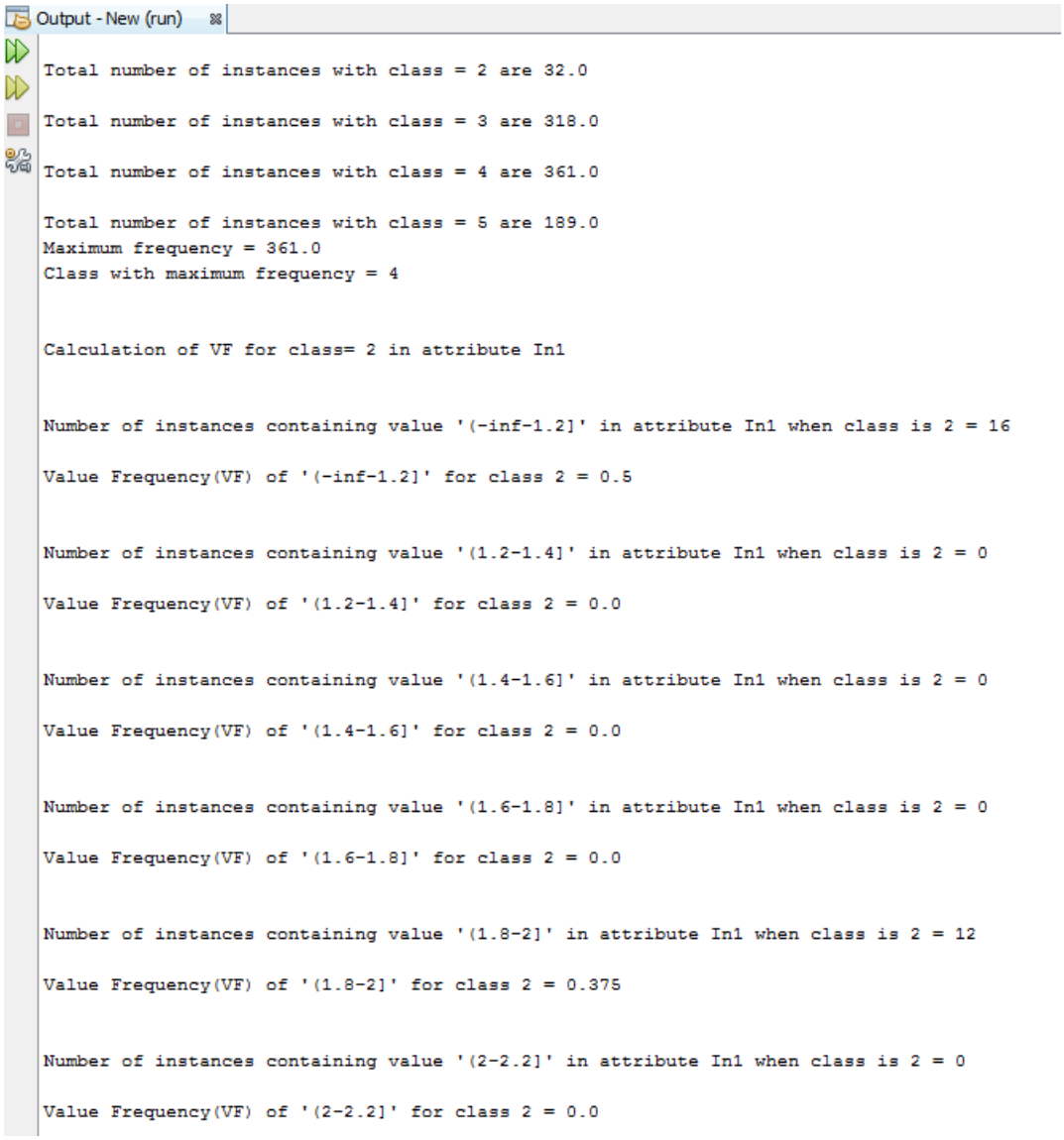

Figure 5.2: Result for calculation of VF for each value of dataset

2. Calculate the Value Frequency(VF) of each value. For this, first total number of instances for both classes was calculated. Then, the Value Frequency of each value is calculated using the equation 4.3. The output of the same is shown in figure 5.2.

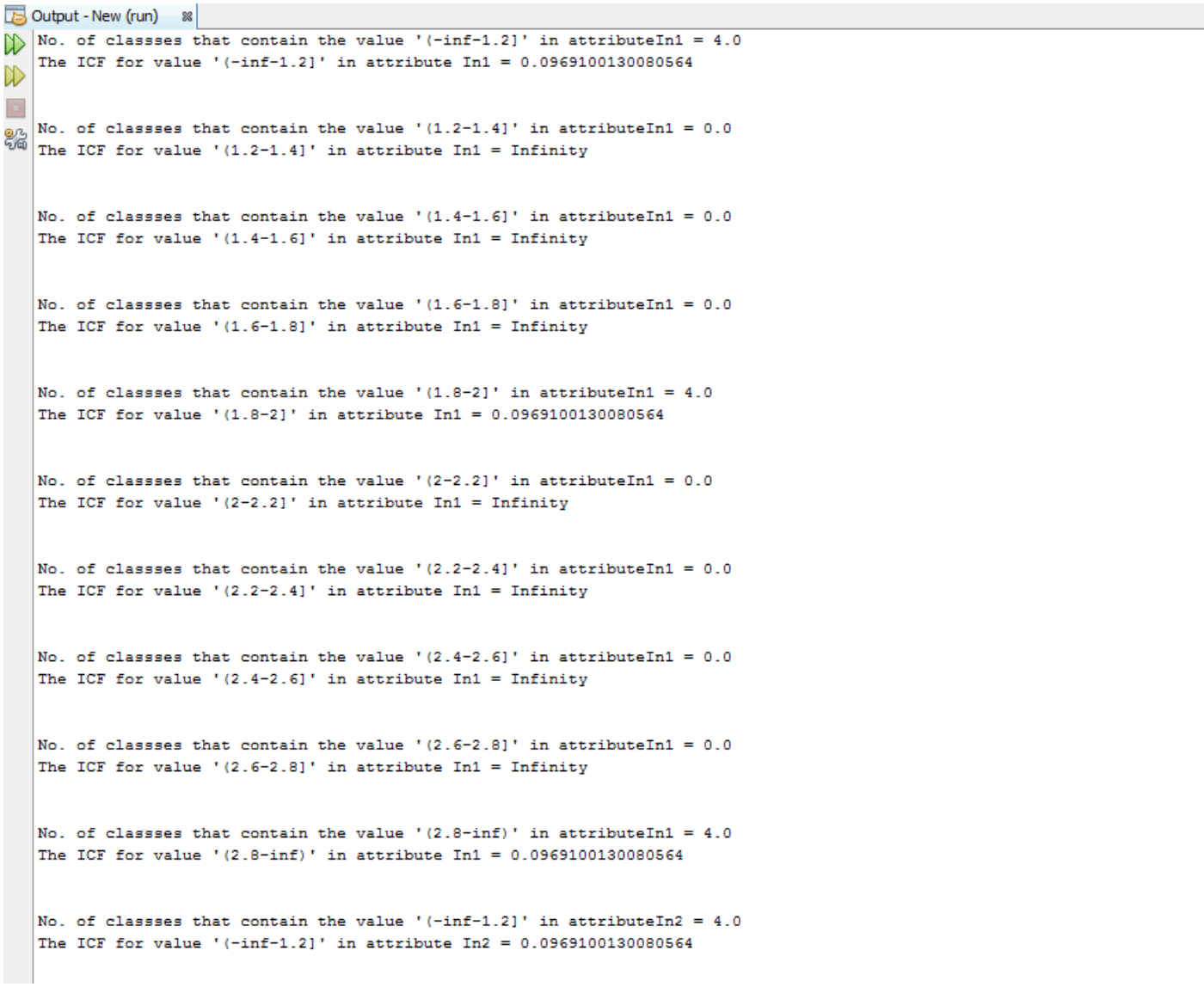

Figure 5.3: Result for calculation of ICF for each value of dataset

3. Calculate the ICF(Inverse Class Frequency) of each value for all classes. This is calculated using the equation 4.4. The output of the same is shown in figure 5.3.

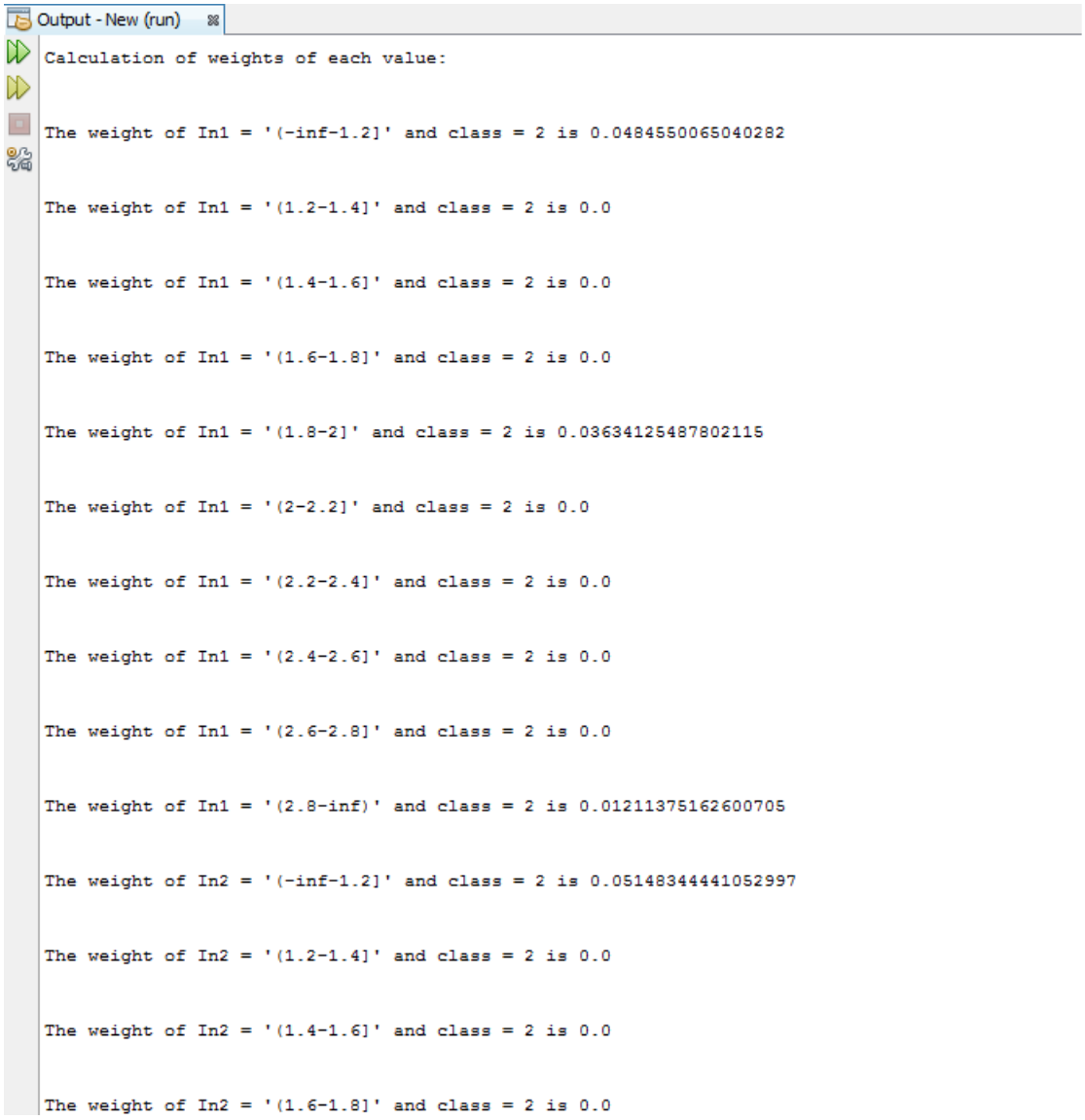

Figure 5.4: Result for calculation of weights for each value of dataset

4. **Calculate the weights** of each value. This is done by using VF/ICF. The output of the same is shown in figure 5.4.

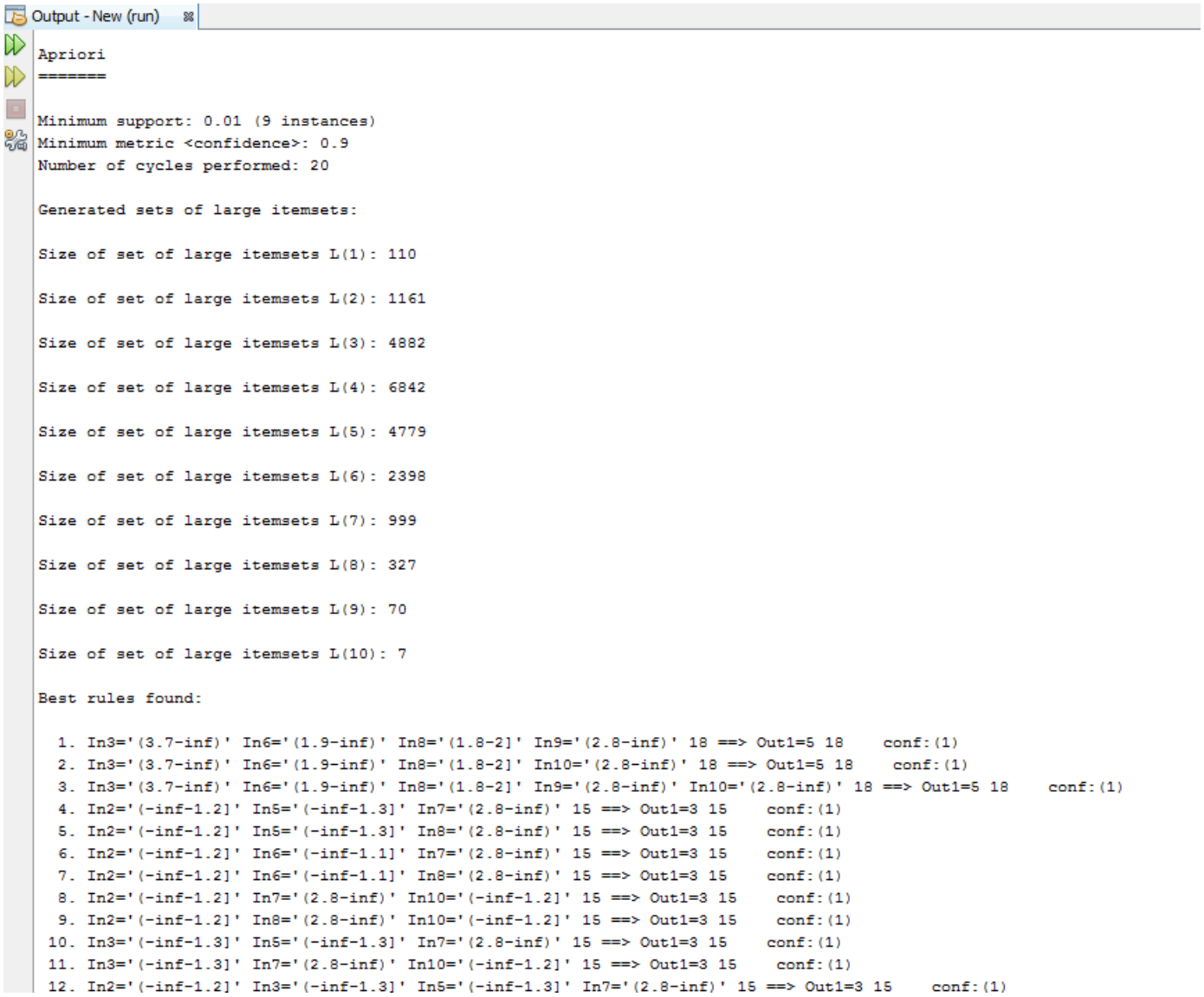

Figure 5.5: Result for generation of association rules

5. **Generate association rules** which are shown in figure 5.5. Generation of association rules is done by using Apiori algorithm.

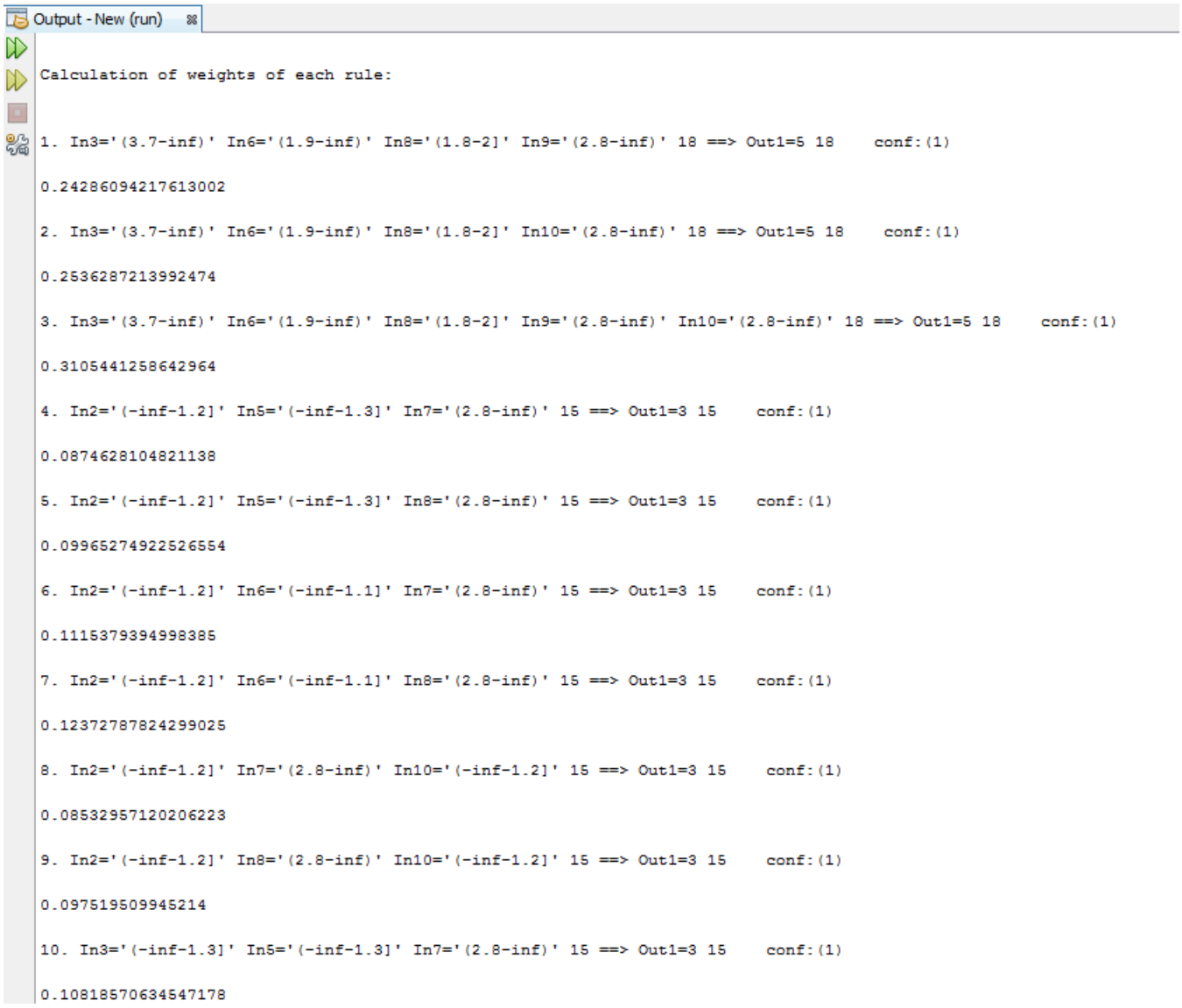

Figure 5.6: Result for calculation of weights of each association rule

6. **Calculate weights of each association rule** which is shown in figure 5.6. For this, weights of each item in antecedents with respected to class in consequent of the same rule.

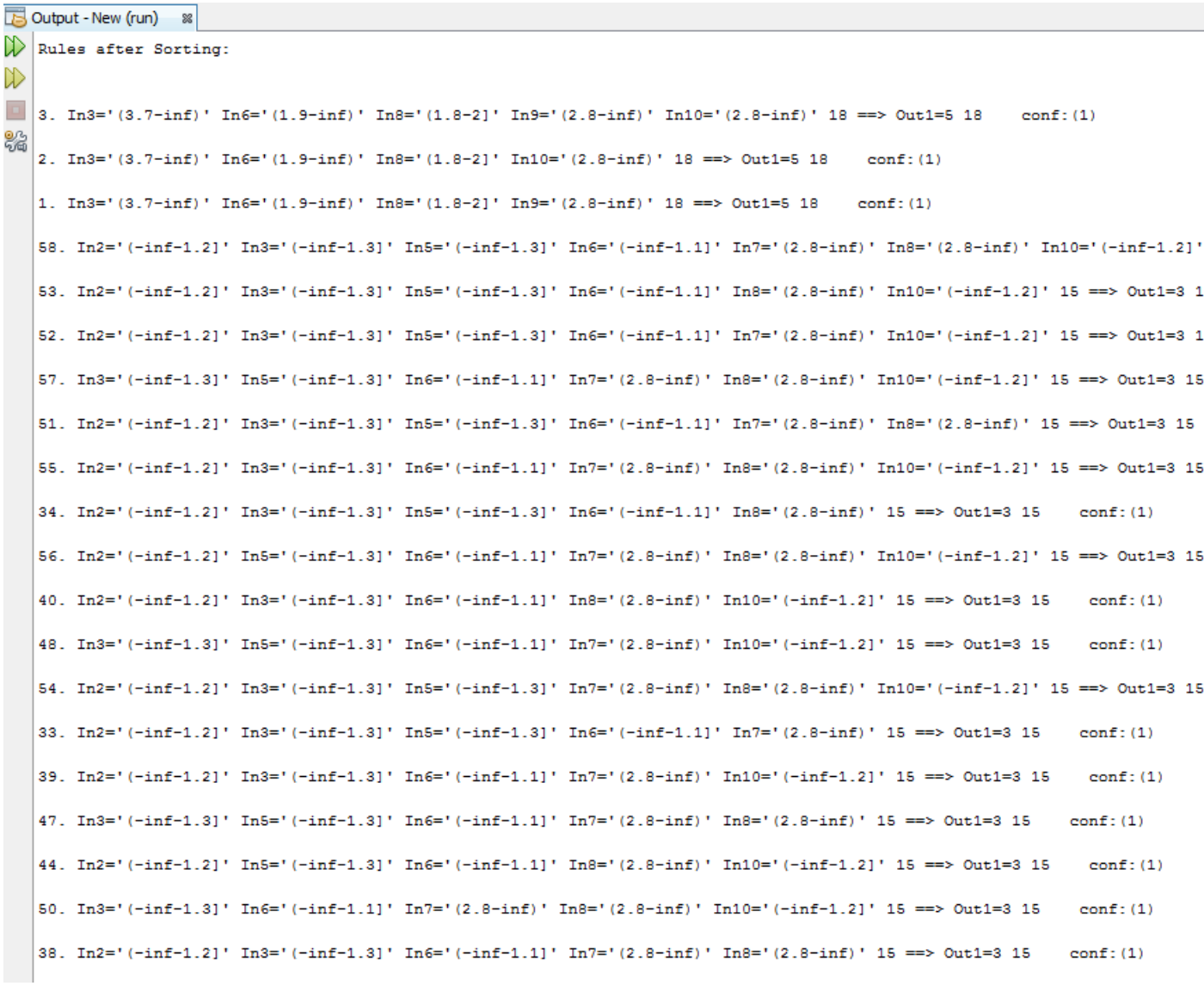

Figure 5.7: Result for sorting of association rules according to their precedence

7. **Sort the association rules** according to decreasing precedence of each rule. The output of the same is shown in figure 5.7.

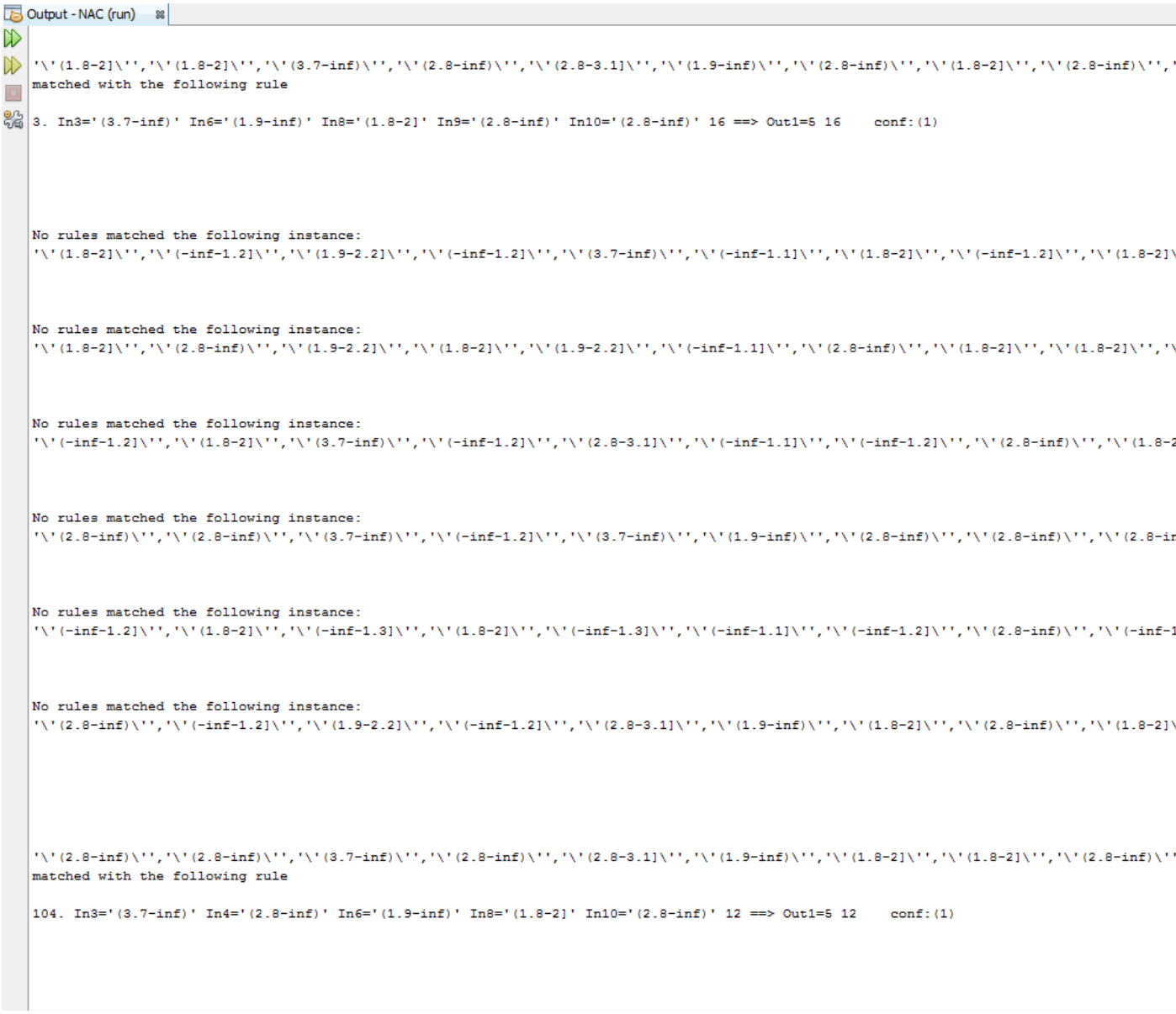

Figure 5.8: Result for classification on the basis of matching of rule

8. **Classify** on the basis of matching of first encountered weighted rule. The output of the same is shown in figure 5.8.

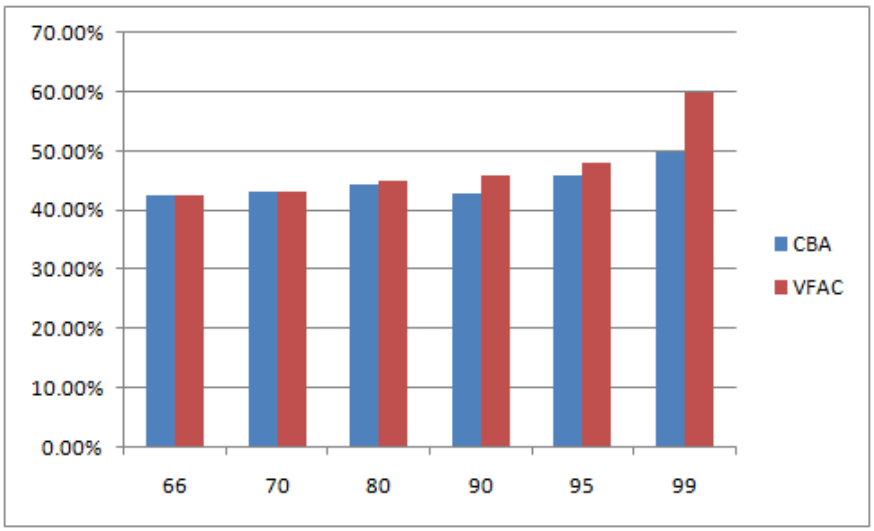

Figure 5.9: Comparison of accuracy of CBA and proposed algorithm(VFAC)

5.2 Accuracy:

The accuracy of our algorithm is measured by two methods:

1. **H**oldout Method: A certain amount of data (66%, 70%, 80%, 90%, 95% and 99%) was reserved for training then the testing has been performed on the rest of the data. The accuracy of the proposed classifier is compared with the existing classical algorithm CBA(Classification Based on Association Rules). The comparison is shown in the figure 5.9. The X-axis shows the amount of data in % used for training and Y-axis shows the accuracy. It is observed that the proposed classifier works better as the training data is increasing.

2. **K**-fold Cross Validation: The accuracy of the proposed algorithm(VFAC) and CBA is calculated on the basis of 10-fold cross validation. Not much variation is observed in the accuracy of both algorithm. The accuracy of proposed algorithm(VFAC) is 46.3% whereas the accuracy of CBA is 46.2%.

5.3 An Experiment:

As per suggestion recieved in a review, an experiment was carried out to test the time taken by the proposed classifier with dataset of multiple class beacause when we are calculating weight with VF*ICF then we need to calculate VF for each value and ICF for all classes then VF*ICF will needed to be calculated for number of values * number

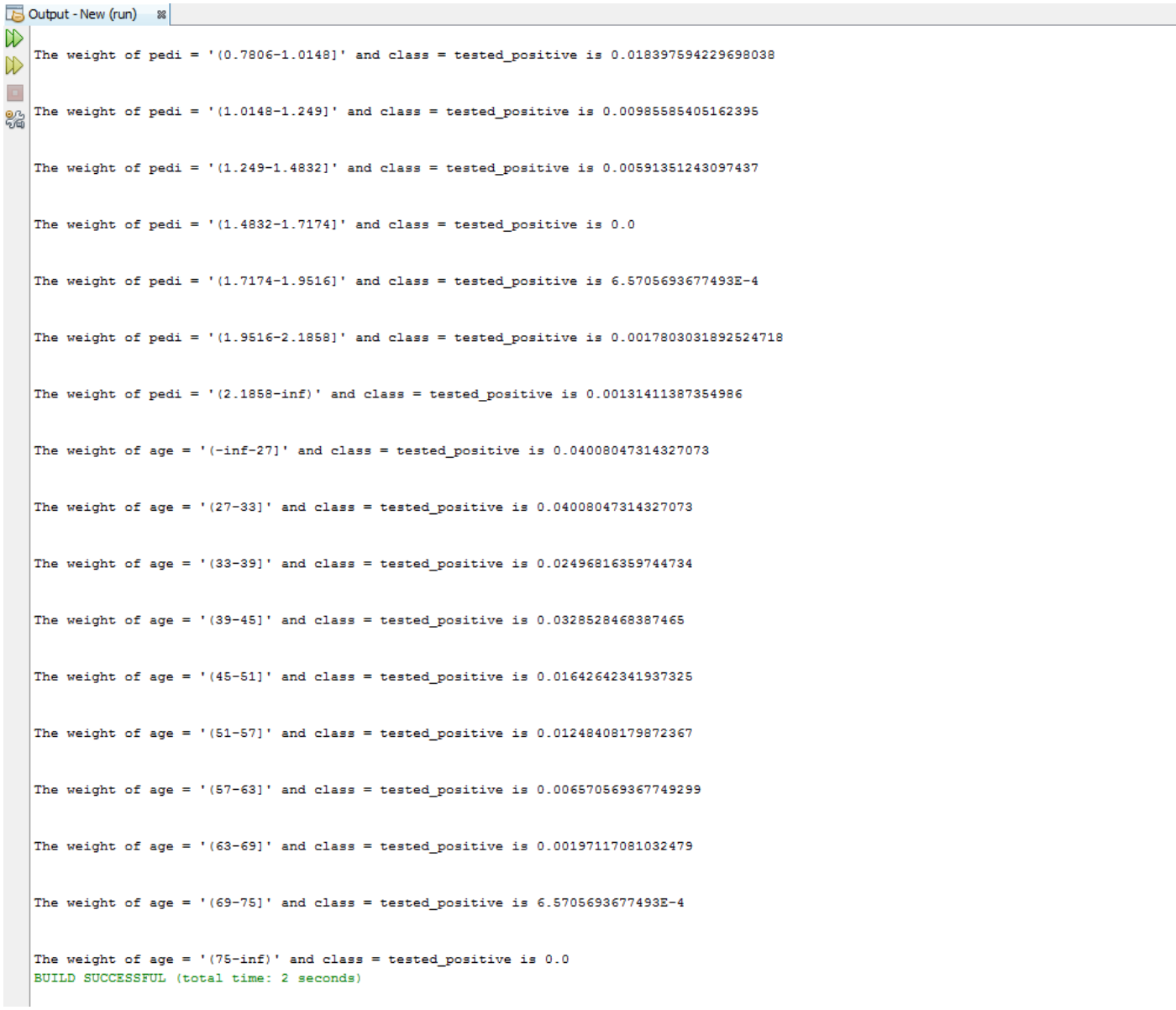

Figure 5.10: Result of calculation of weights on binary class dataset

of classes. So, it may take large amount of time. So, an experiment was carried out to calculate weights with both datasets, one with binary class and other with multiple classes and we found out that both datasets have taken same time for the calculation. The result with binary class dataset is shown in figure 5.10 and the result with multiple class dataset is shown in figure 5.11.

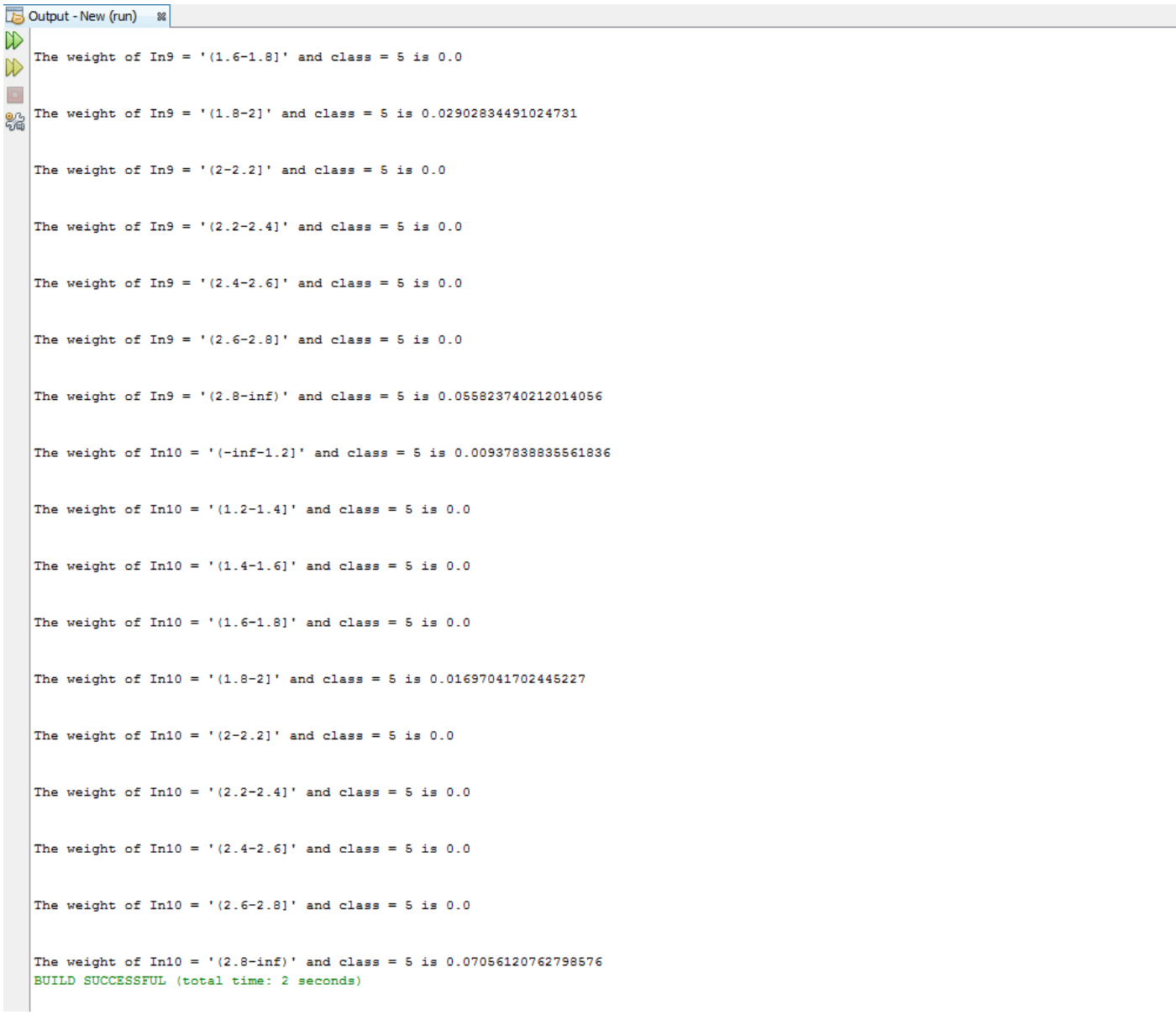

Figure 5.11: Result of calculation of weights on multiple class dataset

Chapter 6

Conclusion and Future Work

In this thesis, Value Frequency based Associative Classifier(VFAC) is introduced which used the concept of VF/ICF in relational database for calculating weights of each value to implement a modified weighted associative classifier in which weights are calculated automatically. VF/ICF concepts in relation database are similar to the TF/IDF concept in Document representation. Much accurate results are observed from the proposed algorithm when compared to the CBA algorithm.

Implementation has been done with binary class dataset as well as multiple class dataset and it is found that it is not taking more time with multiple class. So, the proposed algorithm is not limited to binary class dataset. And the proposed algorithm can work with any number of classes.

The method of calculating weights proposed in this dissertation can be integrated with any type of weighted associative classifier to improve the accuracy of classification.

References

[1] Bing Liu, Wynne Hsu, Yiming Ma-Integrating Classification and Association Rule Mining,Appeared in KDD-98, New York, Aug 27-31.

[2] Yanmin Sun, Andrew K. C. Wong, Fellow, IEEE Yang Wang, Member, IEEE-An Overview of Associative Classifiers

[3] Zuoliang Chen, Guoqing Chen - Building an Associative Classifier Based On Fuzzy Association Rules, International Journal of Computational Intelligence Systems, Vol.1, No. 3 (August, 2008), 262 - 273

[4] Sunita Soni, O.P.Vyas-Using Associative Classifiers for Predictive Analysis in Health Care Data Mining - International Journal of Computer Applications (0975% 8887) Volume 4 No.5, July 2010

[5] Jyoti Soni, Uzma Ansari, Dipesh Sharma, Sunita Soni- Intelligent and Effective Heart Disease Prediction System using Weighted Associative Classifiers-Jyoti Soni et al. /International Journal on Computer Science and Engineering (IJCSE)

[6] Mrs. Suwarna Gothane, Dr. G.R.Bamnote- An Automated Weighted Support Approach Based Associative Classification With Analytical Study For Health Disease Prediction-International Journal of Engineering Research and Applications (IJERA), Vol. 2, Issue 5, September- October 2012.

[7] V. Srividhya, R. Anitha- "Evaluating Preprocessing Techniques in Text Categor ization", International Journal of Computer Science and Application Issue 2010.

[8] Tokunaga, Takenobu Iwayama, Makoto - Text categorization based on weighted inverse document frequency, TITCS 94-TR0001 March 1994.

[9] Joel Pinho Lucas, Anne Laurent, Maria N. Moreno and Magueloenne Teisseire- A Fuzzy Associative Classification Approach For Recommender Systems,hal-00797379, version 1 - 6 Mar 2013

[10] www.tunedit/repo/Data

[11] Dr. V. Vaithiyanathan, K. Rajeswari, Rahul Pitale and Kapil Tajane, Performance Improvement Using Integration of Association Rule Mining and Classification Techniques, International Journal of Scientific & Engineering Research, Volume 4, Issue 5, May-2013 1891 ISSN 2229-5518.

[12] N. Aditya Sundar, P. Pushpa Latha and M. Rama Chandra, Performance Analysis of Classification Data Mining Techniques Over Heart Disease Database, [IJESAT] International Journal of Engineering Science & Advanced Technology Volume-2, Issue-3, 470 - 478

[13] Antonie, M., & Zaiane, O. (2004). An associative classifier based on positive and negative rules. Proceedings of the 9th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery (pp. 64-69). Paris, France.

[14] Aydin, M. Karakose, and E. Akin, The Prediction Algorithm Based on Fuzzy Logic Using Time Series Data Mining Method, World Academy of Science, Engineering and Technology 27 2009

[15] S. soni, O.P. Vyas, J. pillai, Associative Classifier Using Weighted Association Rule, Symposium 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC 2009) page(s):1492-1496