International Research Journal in Global Engineering and Sciences. (IRJGES)
Vol. 1, No. 1, May, 2016 | ISSN : 2456-172X

Implementation of Association Rule Mining for Bridge Datasets Using Weka

[1]Dr. M. Thangamani & [2]Ms.V.Prasanna
[1]Assistant Professor, Kongu Engineering College, India
[2]Research Scholar, Kongu Engineering College, India

Abstract: Data mining playing vital information in extracting useful information from large amount of data set. Apriori algorithm generate useful rule by finding frequent itemset from huge data set. In this paper can apply the Apriori Algorithm to generate rules for the given data set (bridge) using Waikato Environment for Knowledge Analysis tool. Bridge dataset is taken from UCI machine learning repository. These articles explore and visualize the apriori technique in data mining concept.

Keywords: Data mining, Apriori technicque, UCI machine

1. INTRODUCTION

The data mining represents mining the knowledge from large data. Topics such as knowledge discovery, query language, decision tree induction, classification and prediction, cluster analysis, and how to mine the Web are functions of data mining. Manual analyses are time consuming in the real world. In this situation, WEKA can use for automating the task.

Weka is a collection of machine learning algorithms for data mining tasks. Classification was performed using WEKA in data mining research. WEKA is a data mining workbench that allows comparison between many different machine learning algorithms. In addition, it also has functionality for feature selection, data pre-processing and data visualization [1]. The algorithms can either be applied directly to a dataset or called from Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules and visualization. Well-suited for developing new machine learning schemes. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.

2. RELATED WORK

The more associations between accident factors and accident severity were illustrated when applying Apriori algorithm [2]. The predictive Apriori algorithm could derive more number of rules that could be useful when studying the effect of each individual factor to accident severity. These results can help the decision makers in the traffic accident department to take actions based on various hidden patterns from the data. The swarm based techniques to extract association rules for student performance prediction as a multi-objective classification problem is analysis by [3]. In this algorithm takes a low convergence time and it used a few number of parameters. Honeybee Colony Optimization and Particle Swarm Optimization are the two used metaheuristics to extract association rules. These are used in this investigation and WEKA, Rapidminer and KEEL tools are used for comparing the technique. Various type of analysis is carried out using association rules [4-6] in data mining through WEKA environments.

3. EXPERIMENTS DESIGN

Implementation of Association Rule Mining is carried out in Bridge datasets using Weka tool.

3.1 Dataset description

Association rule works only with nominal type and the data values are discrete in nature.
Data set Characteristics: Multivariate
Number of Instances:108
Number of Attributes: 13
Attribute Characteristics: Categorical, Integer

3.2 Attributes description

Table.1 shows the list of attributes in bridge dataset. It also represents the data type for each attributes. Bridge datasets attributes are viewed by viewer in the WEKA explorer panel. It is illustrated in Fig. 1

Table.1 List of attributes

Attribute	**Possible Values**	**Data type**
Id		Nominal
River	A,M,O	Nominal
Location	1 to 52	**Numeric**
Erected	1818-1986; Crafts, Emerging, Mature, Modern	**Numeric**
Purpose	Walk, Aqueduct, RR, Highway	Nominal
Length	804-4558; Short, Medium, Long	**Numeric**
Lanes	1,2,4,6	**Numeric**
Clear-G	N, G	Nominal
T-OR-D	Through, Deck	Nominal
Material	Wood Iron, Steel	Nominal
Span	Short, Medium, Steel	Nominal
REL-L	S, S-F, F	Nominal
Type	Wood, Suspen, Simple-T, Arch, Cantilev, Cont-T	Nominal

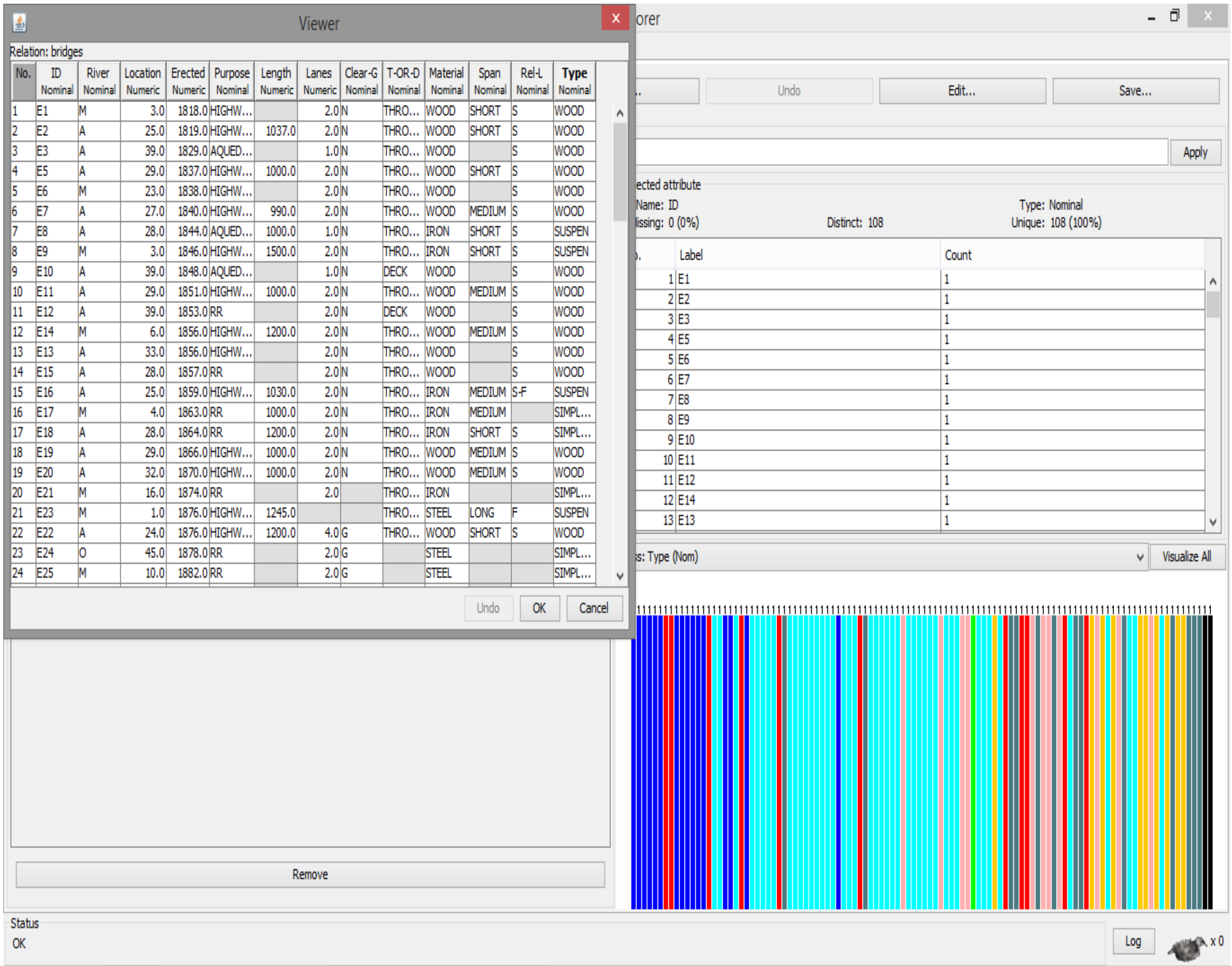


Fig.1 Weka Database Viewer and front panel

4. IMPLEMENTATION STEPS

Since Apriori algorithm works with only nominal data, the data set is preprocessed. Save the intermediate files after each step. The preprocessing WEKA is shown in Fig.2 and Fig.3. The Fig.4 represents the pure data after preprocessing.

The following preprocessing methods are applied:

- **Removing the attribute:**
  - Remove the attribute id, since it uniquely identifies the tuples. It is done by selecting the remove attribute filter.
  - Remove the attribute location, since it does not play a vital role in generating the rules.

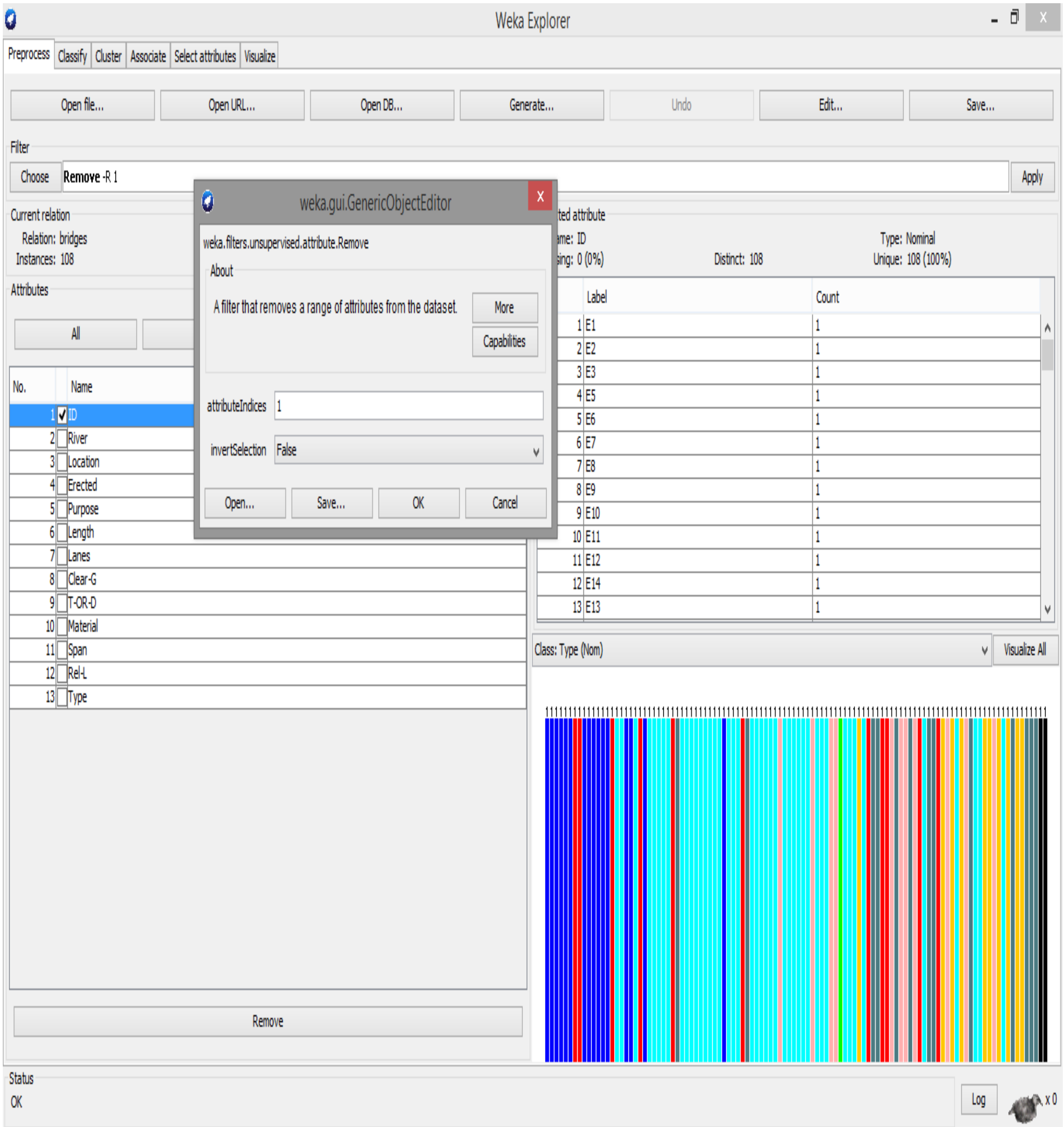


Fig.2 Preprocessing Weka

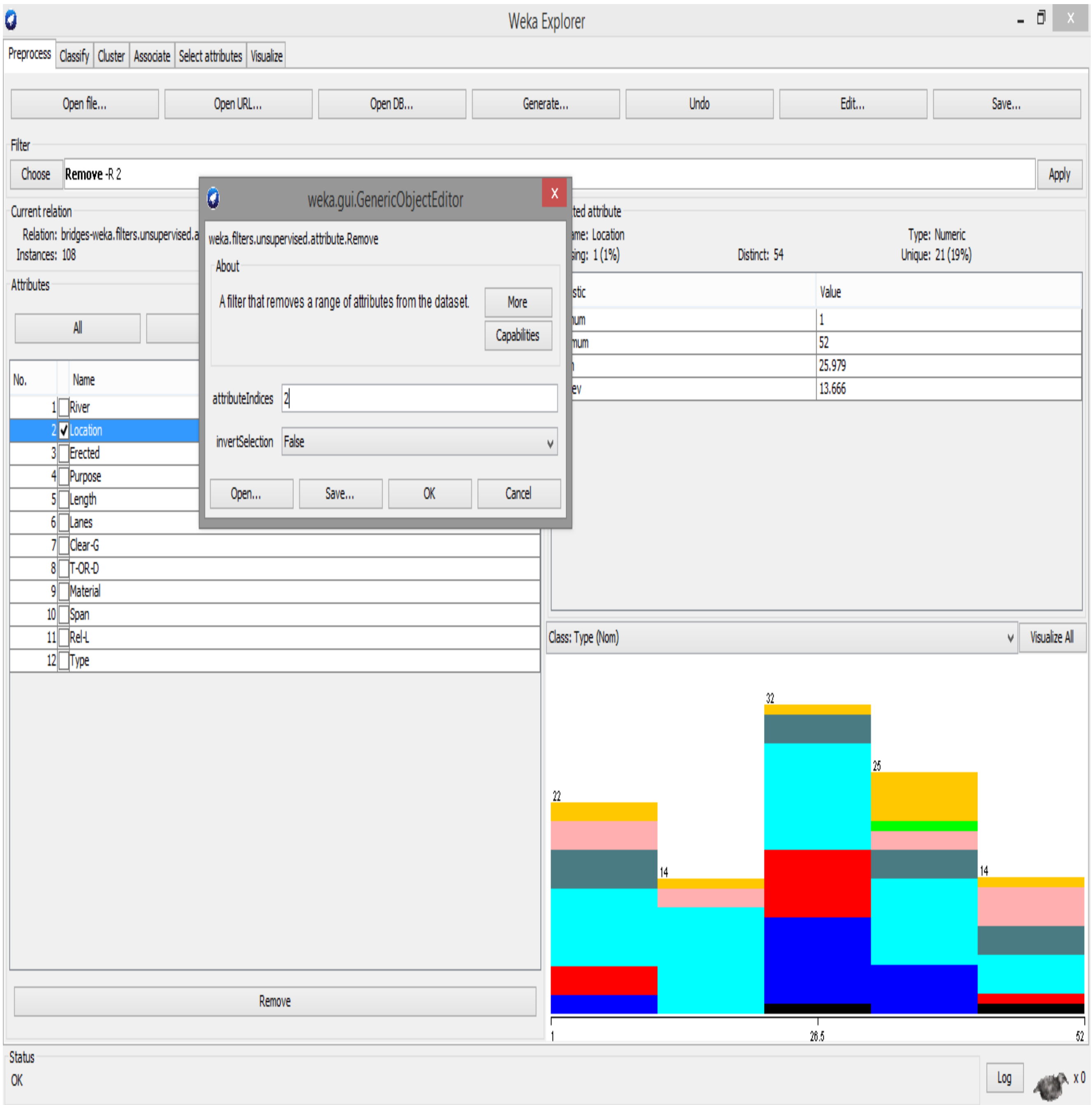


Fig.3 Unwanted attribute removing in Preprocessing Weka

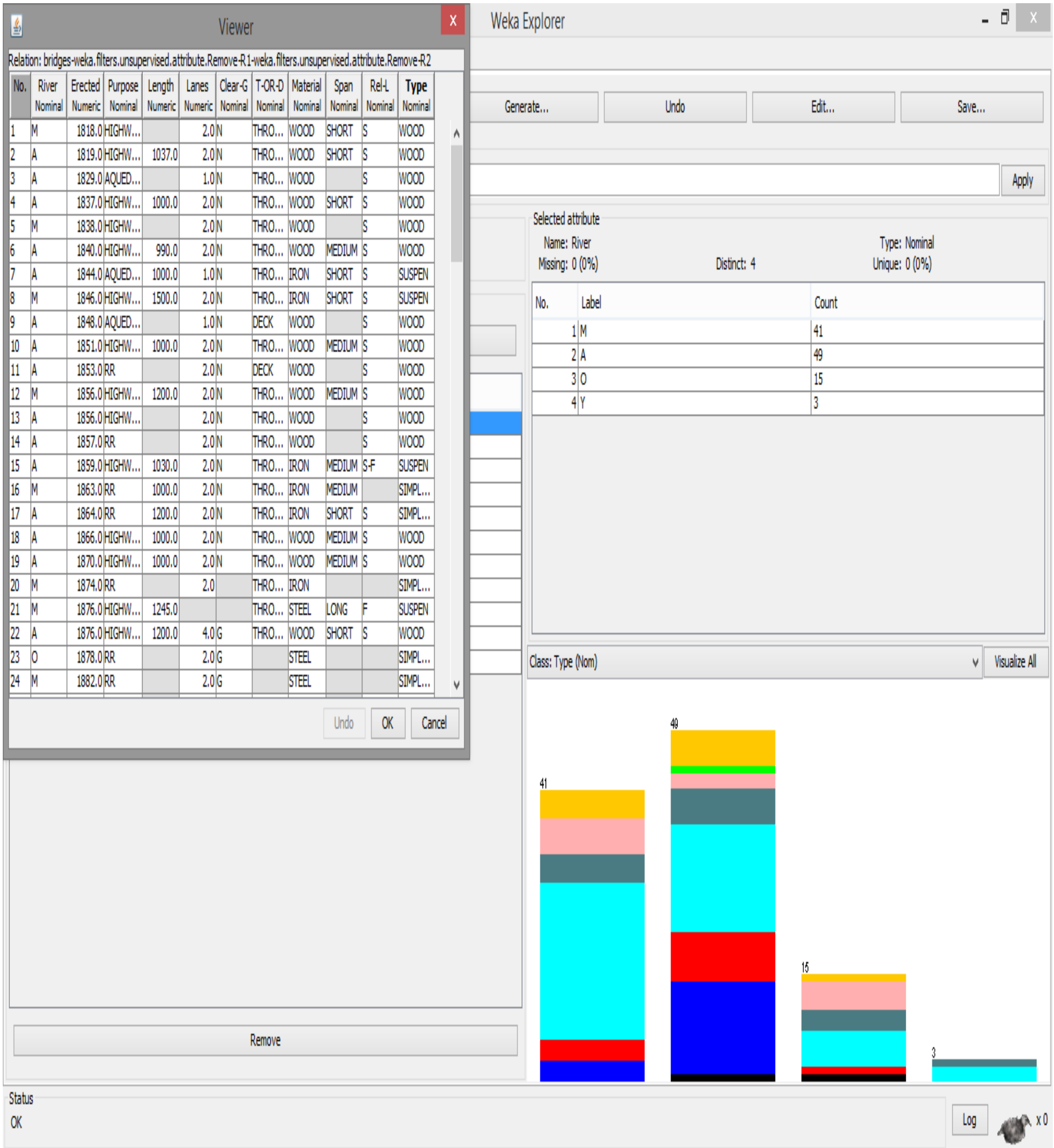


Fig.4 After preprocessing

Discretization: Association rule mining can be applied on categorical data, so the three numeric attributes erected, length and lanes in the data set are discretized and it shown in Fig.5. The Fig.6 represents the how to modify the normalized value for discretization.

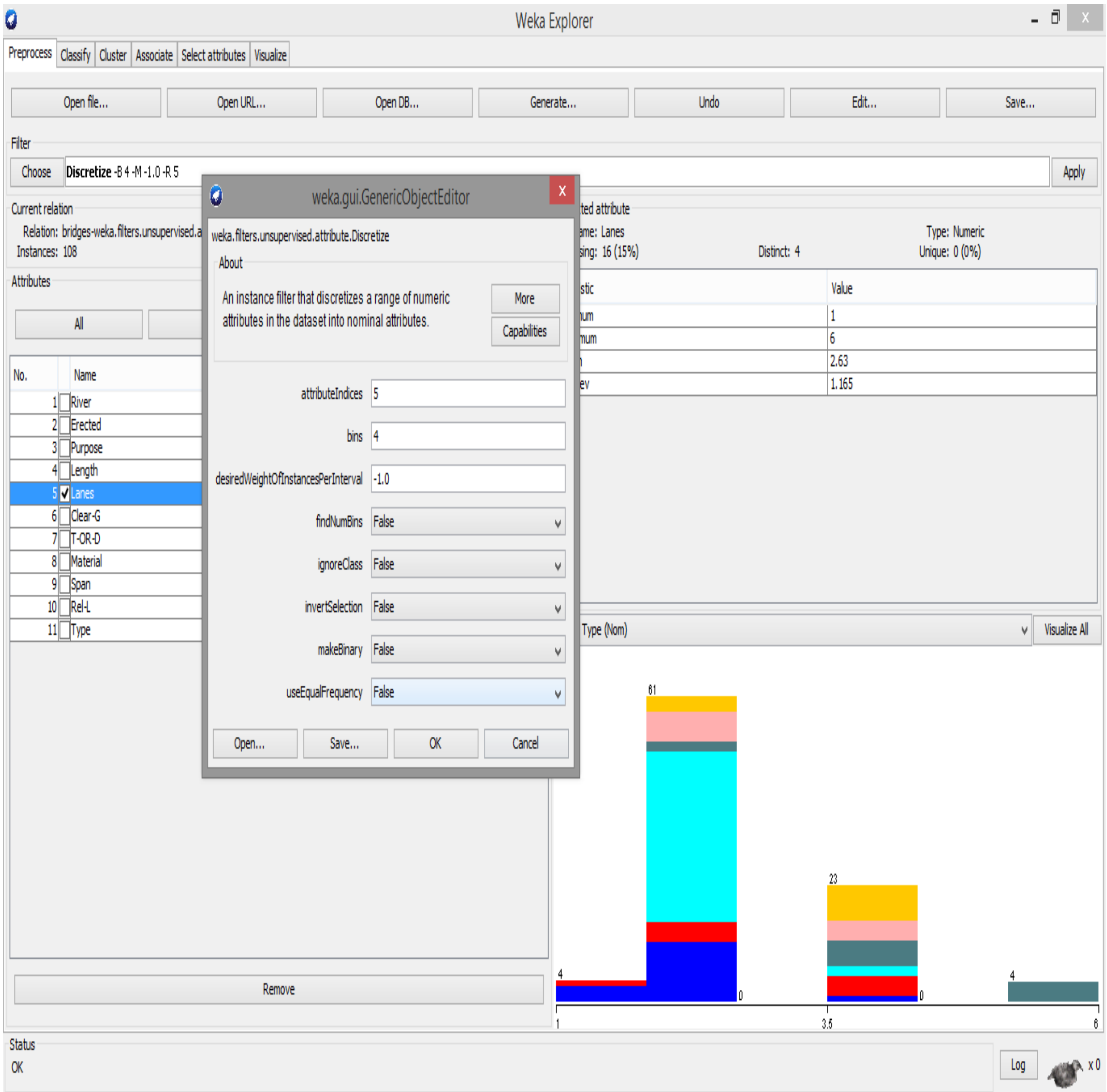


Fig.5 Discretization in Bridge datasets

The input file with the above changes is shown below Fig.6.

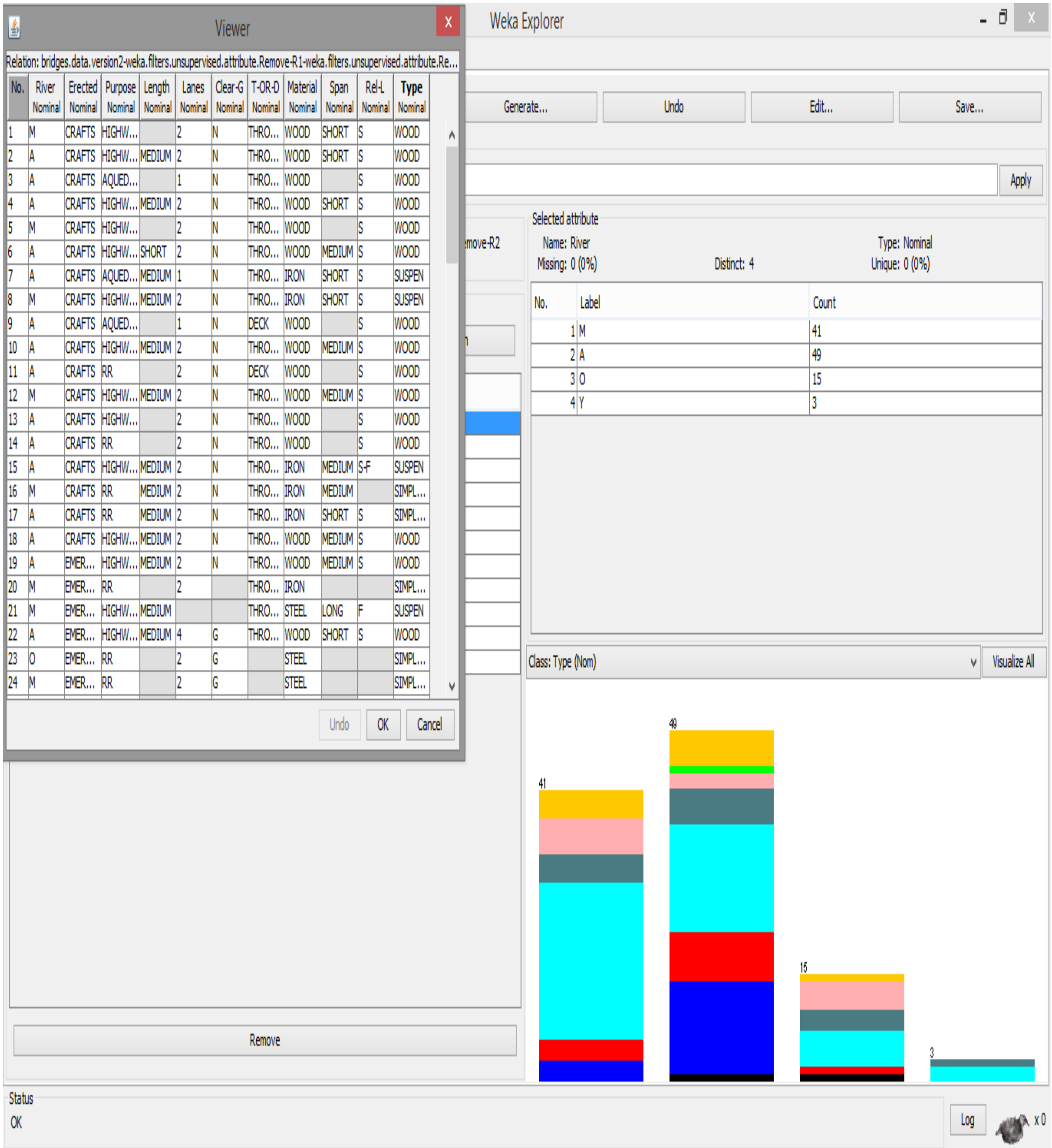
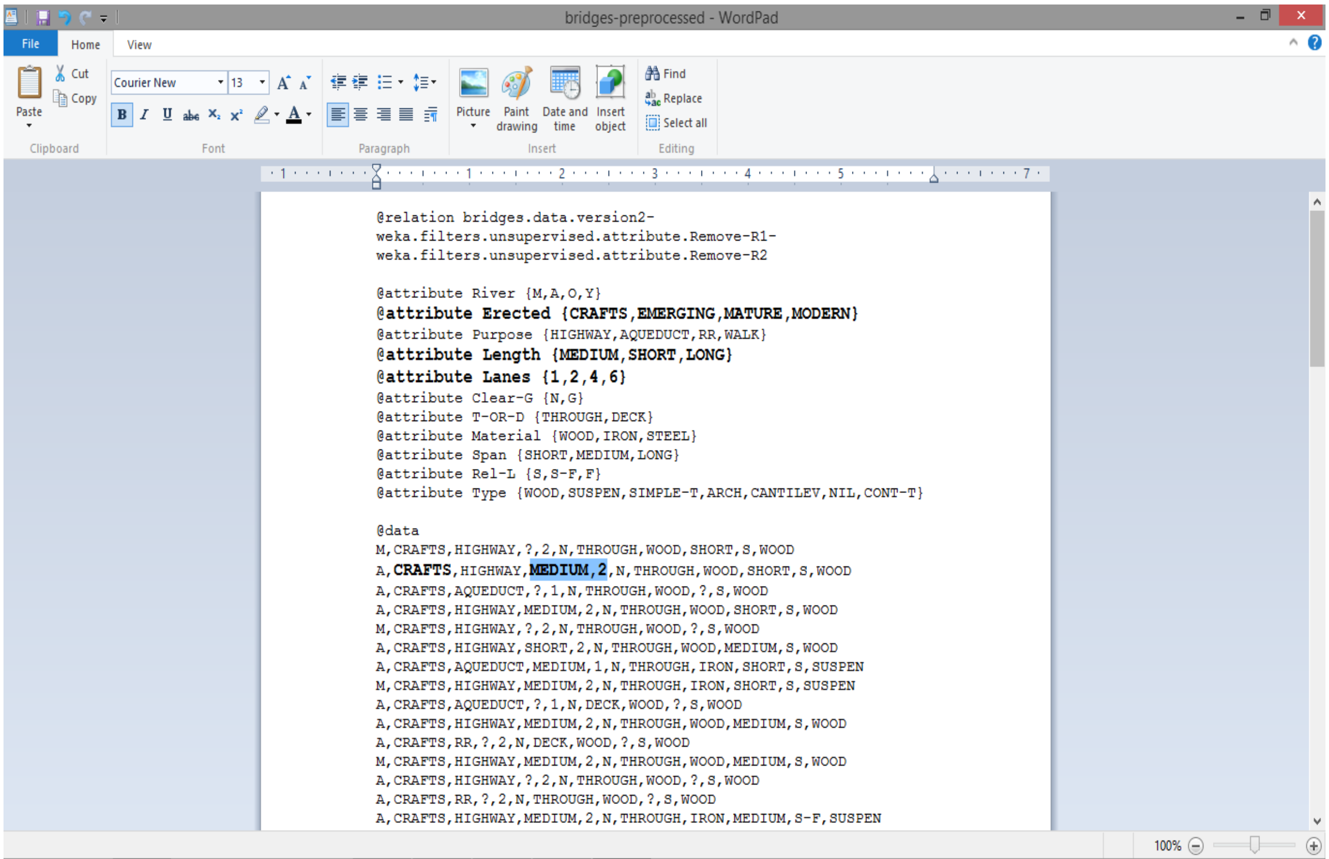


Fig.6 After Discretization on Bridge datasets

The following Fig.7 depict the labels assigned for the attributes and the changes in the instances (one instance highlighted):




```
@relation bridges.data.version2-
weka.filters.unsupervised.attribute.Remove-R1-
weka.filters.unsupervised.attribute.Remove-R2

@attribute River {M,A,O,Y}
@attribute Erected {CRAFTS,EMERGING,MATURE,MODERN}
@attribute Purpose {HIGHWAY,AQUEDUCT,RR,WALK}
@attribute Length {MEDIUM,SHORT,LONG}
@attribute Lanes {1,2,4,6}
@attribute Clear-G {N,G}
@attribute T-OR-D {THROUGH,DECK}
@attribute Material {WOOD,IRON,STEEL}
@attribute Span {SHORT,MEDIUM,LONG}
@attribute Rel-L {S,S-F,F}
@attribute Type {WOOD,SUSPEN,SIMPLE-T,ARCH,CANTILEV,NIL,CONT-T}

@data
M,CRAFTS,HIGHWAY,?,2,N,THROUGH,WOOD,SHORT,S,WOOD
A,CRAFTS,HIGHWAY,MEDIUM,2,N,THROUGH,WOOD,SHORT,S,WOOD
A,CRAFTS,AQUEDUCT,?,1,N,THROUGH,WOOD,?,S,WOOD
A,CRAFTS,HIGHWAY,MEDIUM,2,N,THROUGH,WOOD,SHORT,S,WOOD
M,CRAFTS,HIGHWAY,?,2,N,THROUGH,WOOD,?,S,WOOD
A,CRAFTS,HIGHWAY,SHORT,2,N,THROUGH,WOOD,MEDIUM,S,WOOD
A,CRAFTS,AQUEDUCT,MEDIUM,1,N,THROUGH,IRON,SHORT,S,SUSPEN
M,CRAFTS,HIGHWAY,MEDIUM,2,N,THROUGH,IRON,SHORT,S,SUSPEN
A,CRAFTS,AQUEDUCT,?,1,N,DECK,WOOD,?,S,WOOD
A,CRAFTS,HIGHWAY,MEDIUM,2,N,THROUGH,WOOD,MEDIUM,S,WOOD
A,CRAFTS,RR,?,2,N,DECK,WOOD,?,S,WOOD
M,CRAFTS,HIGHWAY,MEDIUM,2,N,THROUGH,WOOD,MEDIUM,S,WOOD
A,CRAFTS,HIGHWAY,?,2,N,THROUGH,WOOD,?,S,WOOD
A,CRAFTS,RR,?,2,N,THROUGH,WOOD,?,S,WOOD
A,CRAFTS,HIGHWAY,MEDIUM,2,N,THROUGH,IRON,MEDIUM,S-F,SUSPEN
```

Fig.7 Labels assigned for the attributes and the changes in the instances

Apriori Algorithm Implementation in Weka:

The preprocessed data file is used for Association rule mining (Apriori Algorithm) and the following rules are generated by setting the necessary measures such as support and confidence is shown in Fig.8 and Fig.9.

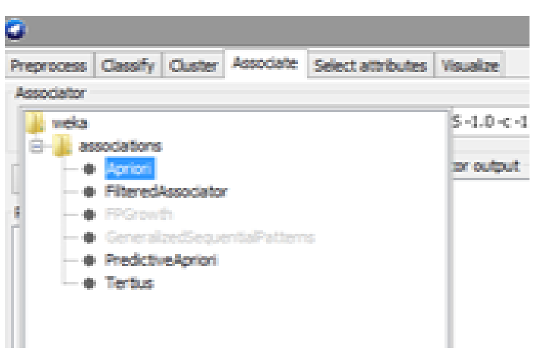


Fig.8 Apriori Algorithm Implementation in Weka

Minimum Support and Confidence threshold:
The following Fig.9 shows the parameters set

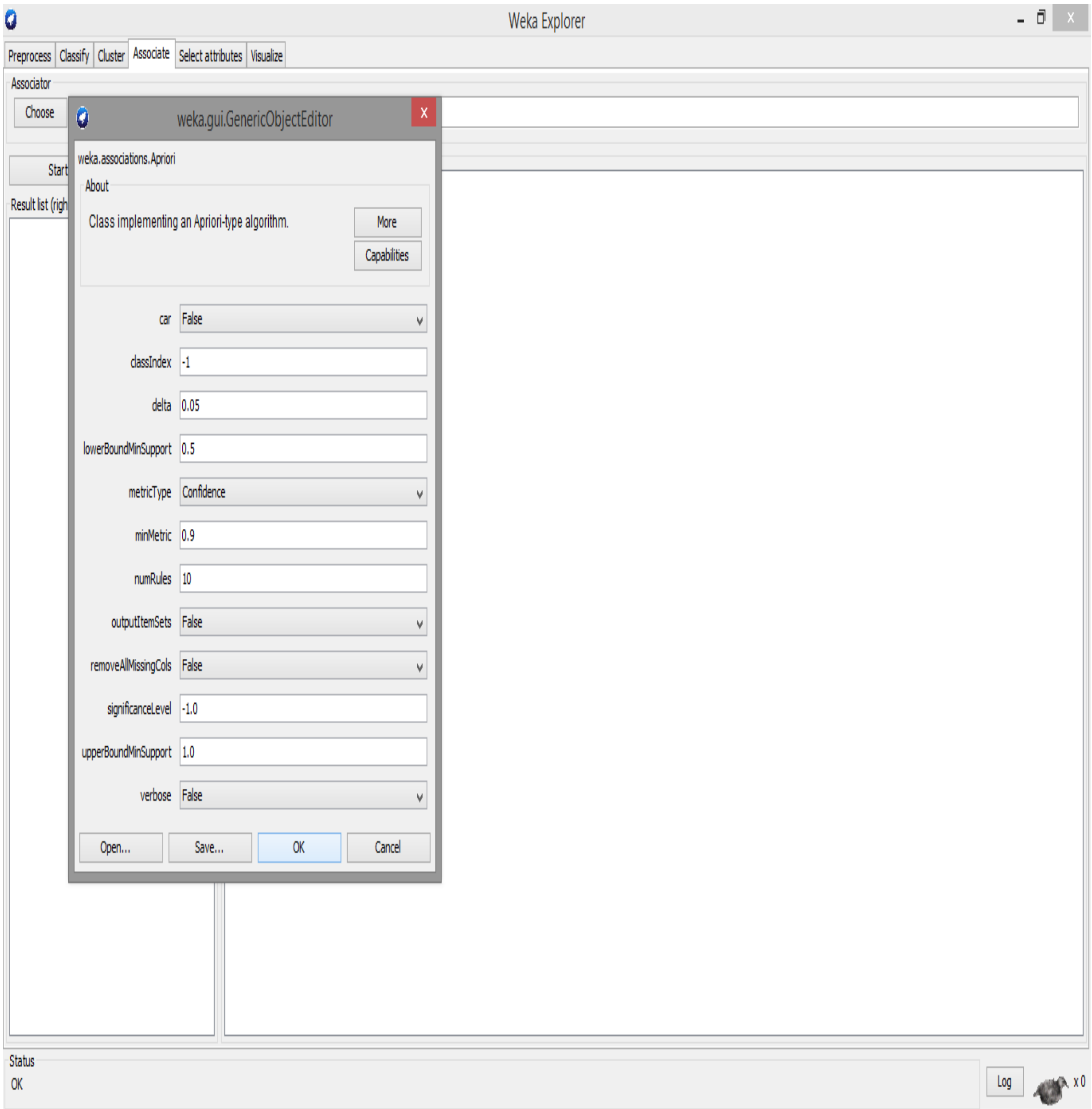


Fig. 9 Minimum Support and Confidence threshold

Output-Rules Generated:

The screen shot shows the rules generated by applying Apriori Algorithm for association rule mining is shown in Fig.10.

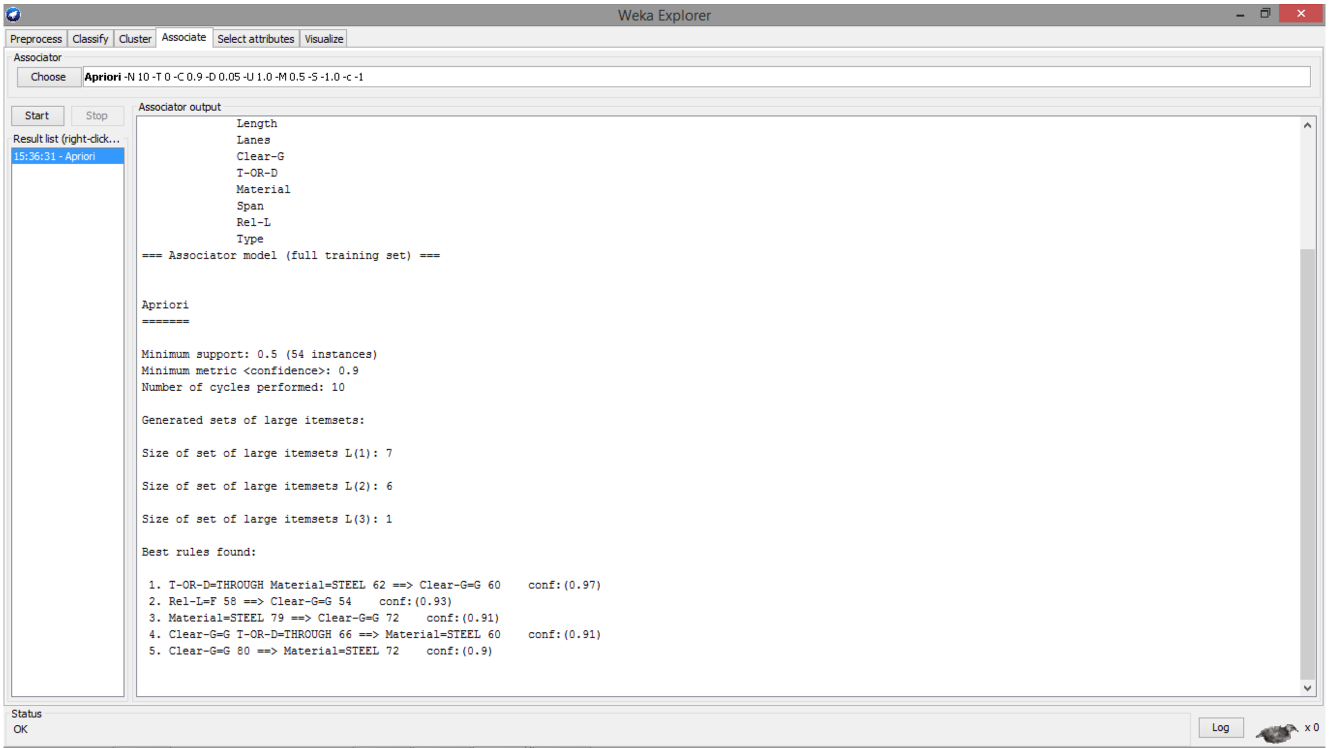


Fig.10. Output rule generated

=== Run information ===
Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.5 -S -1.0 -c -1
Relation: bridges.data.version2-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove-R2
Instances: 108
Attributes: 11

River
Erected
Purpose
Length
Lanes
Clear-G
T-OR-D
Material
Span
Rel-L
Type

=== Associator model (full training set) ===
Apriori
=======
Minimum support: 0.5 (54 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 10
Generated sets of large itemsets:
Size of set of large itemsets L(1): 7
Size of set of large itemsets L(2): 6
Size of set of large itemsets L(3): 1

Best rules found:
1. T-OR-D=THROUGH Material=STEEL 62 ==> Clear-G=G 60 conf:(0.97)
2. Rel-L=F 58 ==> Clear-G=G 54 conf:(0.93)
3. Material=STEEL 79 ==> Clear-G=G 72 conf:(0.91)
4. Clear-G=G T-OR-D=THROUGH 66 ==> Material=STEEL 60 conf:(0.91)
5. Clear-G=G 80 ==> Material=STEEL 72 conf:(0.9)

5. CONCLUSION AND FUTURE DIRECTION

The above rules infer that Most of the THROUGH bridges are constructed using the Material STEEL. If Bridges built on Clear Ground and are THROUGH bridges then the Material used to build such bridges is STEEL.

References

1. Donn Morrison, Ruili Wang, Liyanage C. De Silva, Ensemble methods for spoken emotion recognition in call-centres, Speech Communication, Elsevier, Vol. 49, pp.98-112, 2007
2. Amira A. El Tayeb, Vikas Pareek, Abdelaziz Araar, Applying Association Rules Mining Algorithms for Traffic Accidents in Dubai , International Journal of Soft Computing and Engineering, Vol. 5, No.4, pp. 1-12, 2015
3. Roghayeh Saneifar and Mohammad Saniee Abadeh, Jan Zizka et al., Association Rule Discovery for Student Performance Prediction Using Metaheuristic Algorithms, pp. 115–123, 2015, DOI : 10.5121/csit.2015.51510
4. Amit Dipchandji Kasliwal, Dr. Girish S. Katkar,"Web Usage mining for Predicting User Access Behaviour", International Journal of Computer Science and Information Technologies, Vol.6, No.1, pp-201-204, 2015.
5. Rahul Neve 1, K.P Adhiya ", Comparative Study of Web Mining Algorithms for Web Page Prediction in Recommendation System" , international Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, No.1, pp.969-976, 2013
6. Dhruva Mistry, Kirti Sharma, Samip A.Patel, Recommend Websites through Weblog Files using Association Rule, International Journal of Computer Applications, Vol.126, No.2, pp.16-19, 2015

Author(s) Biography

Thangamani completed her B.E., from Government College of Technology, Coimbatore, India. She completed her M.E in Computer Science and Engineering from Anna University and PhD in Information and Communication Engineering from the renowned Anna University, Chennai, India in the year 2013.

Dr. M. Thangamani possesses nearly 23 years of experience in research, teaching, consulting and practical application development to solve real-world business problems using analytics. Her research expertise covers Medical data mining, machine learning, cloud computing, big data, fuzzy, soft computing, ontology development, web services and open source software. She has published nearly 70 articles in refereed and indexed journals, books and book chapters and presented over 67 papers in national and international conferences in above field. She has delivered more than 60 Guest Lectures in reputed engineering colleges and reputed industries on various topics. She has got best paper awards from various education related social activities in India and Abroad. She has organized many self-supporting and government sponsored national conference and Workshop in the field of data mining, big data and cloud computing. She continues to actively serve the academic and research communities and presently guiding Ph.D Scholars under Anna University. She is also seasonal reviewer in IEEE Transaction on Fuzzy System, international journal of advances in Fuzzy System and Applied mathematics and information journals. She has organizing chair and keynote speaker in international conferences in India and countries like California, Dubai, Malaysia, Singapore, Thailand and China. She has received many awards for academic activities. She is on the editorial board and reviewing committee of leading research journals, which includes her nomination as the Editor in chief to International Scientific Global journal for Engineering, Science and Applied Research (ISGJESAR) & International Research Journal in Global Engineering and Sciences (IRJGES) and on the program committee of top international data mining and soft computing conferences in various countries. She has Life Membership in ISTE, Member in CSI, International Association of Engineers and Computer Scientists in China, IAENG, IRES, Athens Institute for Education and Research and Life member in Analytical Society of India. She is currently working as Assistant Professor at Kongu Engineering College at Perundurai, Erode District.

Mrs. V. Prasanna completed her Master's degree in Software Engineering from Anna University of Technology, Coimbatore, Tamil Nadu, India. Currently, she is full time research scholar in Kongu Engineering College under the Anna University, Chennai. She has nearly 7 years of academic experience and one year industry experience. She has presented papers in four National and International conferences. She has published 5 papers in reputed international journals. Her research interests include Data Mining, Image retrieval, Big data and Cloud Computing.