wjert, 2020, Vol. 6, Issue 1, 249-255. Original Article ISSN 2454-695X

World Journal of Engineering Research and Technology

WJERT

www.wjert.org

SJIF Impact Factor: 5.924

PERFORMANCE ANALYSIS OF EDTT AND CONJUNCTIVE RULE ALGORITHM IN CLOUD COMPUTING ENVIRONMENTS

Kuldip Kaur*[1] and Er. Harwinder Kaur[2]

[1]Student. M.Tech CSE Soldier Institute of Engg. & Technology, Near NIT, Jalandhar (Punjab).

[2]Assistant Professor, Department of Computer Sc. & Engg. St. Soldier Institute of Engg. & Technology, Near NIT, Jalandhar (Punjab).

Article Received on 01/12/2019 **Article Revised on 22/12/2019** **Article Accepted on 12/01/2020**

***Corresponding Author**
Kuldip Kaur
Student. M.Tech CSE Soldier Institute of Engg. & Technology, Near NIT, Jalandhar (Punjab).

ABSTRACT

AI is the information revelation process by investigating the huge volumes of information from different viewpoints and abridging it into helpful data. Because of the significance of removing information/data from the enormous information vaults, AI has become a fundamental segment in different fields of human life. One of the special advantages - not yet completely acknowledged - of the developing distributed computing pattern is additionally one of the most disputable parts of the present on-request benefits. The usage of AI systems through Cloud figuring will enable the clients to recover significant data from for all intents and purposes coordinated information distribution center that decreases the expenses of framework and capacity.

INTRODUCTION

Man-made intelligence is the data disclosure process by exploring the tremendous volumes of data from various perspectives and abbreviating it into supportive information. In view of the criticalness of expelling data/information from the colossal data vaults, AI has become a basic fragment in various fields of human life. One of the exceptional favorable circumstances - not yet totally recognized - of the creating circulated figuring design is furthermore one of the most debatable pieces of the present on-demand benefits. The utilization of AI frameworks through Cloud figuring will empower the customers to recoup

noteworthy information from in every practical sense facilitated data dissemination focus that diminishes the costs of system and limit.

Cloud Computing is a recently developed technology but its popularity is becoming larger day by day. You are probably using some of its forms without even realising it. From small start-ups to global corporations, from government agencies to non-profit organisations, we are embracing Cloud Computing for all sorts of reasons.

Utilising cloud services has many perks. Whether you are an individual doing online banking or your organisation is running an application that share photos to millions of mobile users, Cloud Computing can satisfy all in providing low cost IT resources.

The multi-faceted Cloud Computing

Cloud service providers, such as Amazon Web Services (AWS), offer on-demand cloud services platform for data storage, applications, and other IT resources via the internet (a.k.a. the cloud) with a flexible pricing, suitable for even small and medium-sized businesses.

The concept of Cloud Computing is you can access your files anywhere, anytime as long as you are connected to the internet. Email, mobile banking, online shopping, skyping are just a few of the many applications that Cloud Computing is capable of.

1. Cloud database

Your business is in need of a database but is on a tight budget or lack the adequate expertise to build one on-site. In such case, cloud database is a better alternative.

Cloud databases give IT personnel a powerful database that just works without the need for a physical infrastructure. Your cloud service provider will not only support but also take care of all the maintenance and operation of the database, your sole responsibility is handling your own data.

2. Test and development

Testing and development testing are crucial steps to ensure your application can run smoothly, error-free, and usable. In order to successfully test your application, you need a simulated environment that mimics the actual business operations to validate the results.

This is where Cloud Computing comes in to ease your pain of building your own test environments. There are various ready-made environments, tailored to your specific needs at your fingertips.

3. Website hosting

Hosting your website on the cloud might be necessary if the current one could not meet up with the constant growth of your business. If you have built a steady website, you would know that hosting takes up a large portion of IT resources.

4. Big data analytics

Every data that we encounter today, such as the stack of paperwork, resumes, or digital data like your Facebook messages, is categorised under a big umbrella term called "big data".

5. File storage and sharing

This is one of the most basic forms of Cloud Computing. Files are stored in the cloud which makes sharing, retrieving and archiving extremely easy. Google Drive, Dropbox, Shutterstock are the most popular examples of this service.

6. Backup and disaster recovery

Documents, files and data should be backed up frequently, but not many of us follow through with the routine. Today, we still manually dispatch data using portable devices and storage facilities which are bottom line time consuming and not very cost effective.

7. Business applications

There are more than a handful of cloud-based applications that have intuitive interfaces, are easy to use and tailored for a specific industry.

Machine Learning Techniques

EDTT (Enhanced Decision Tree Technique EDTT): Enhanced Decision Tree Technique, like its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction. The fundamental concept behind random forest is a simple but powerful one — the wisdom of crowds. In data science speak, the reason that the EDTT works so well is: A large number of relatively uncorrelated models (trees) operating as a committee will outperform any of the individual constituent models.

Conjunctive Rule Algorithm: Conjunctive Rule algorithm implements a single conjunctive rule learner that can predict for numeric and nominal class values. Conjunctive rule uses the relation of logical AND to link stimulus attributes. The rule involves 'AND' ing the antecedents together and the consequent (class value) for the classification. In this case, the consequent is the distribution of the available classes (or mean for a numeric value) in the dataset. If this rule does not enclose the test instance, then the default class distributions/value of data that is not enclosed by the rule in the training data is used to predict it. An antecedent is selected by this learner by calculating the Information Gain of each antecedent and the generated rule is pruned using Reduced Error Pruning (REP) or simple pre-pruning depending on the number of antecedents. The weighted mean of the entropies of both the data covered and not covered by the rule is the Information of one antecedent used for classification. Single conjunctive rule learner is one of the machine learning algorithms and is normally known as inductive Learning. The goal of rule induction is generally to induce a set of rules from data that captures all generalizable knowledge within that data, and at the same time being as small as possible. Classification in rule-induction classifiers is typically based on the firing of a rule on a test instance, triggered by matching feature values at the left-hand side of the rule. Rules can be of various normal forms, and are typically ordered; with ordered rules, the first rule that fires determines the classification outcome and halts the classification process. Uncovered test instances are assigned the default class value (or distribution) of the uncovered training instances. The information gain (nominal class) or variance reduction (numeric class) of each antecedent is computed, and rules are pruned using reduced-error pruning.

Dataset Used

Name	Class	Semester	Name of the Institute	Reason	Cloud Services
JOGINDER PAL	BSC	3	St. Soldier	Google Drive	File storage
SOMNATH	BCA	4	CT Inst.	Web Hosting	Development
RAM PRAKASH	B.Tech	5	LKC	online money transfers	Finance
NAND SINGH	MBA	4	St. Soldier	ERP	Backup
BAL RAJ	Pharmacy	6	LKC	Facebook whatsapp	Big data analyticsÿ
RAMPAT THAKUR	B.Tech	2	DAV	Office Applications Email	Business Applications
SOHAN SINGH	HMCT	5	St. Soldier	Google Drive	File storage

RESULT AND DISCUSSION

Dataset Used

Table 1: Shows the attributes of Cloud dataset with instances.

Name	Class	Semester	Name of the Institute	Reason	Cloud Services
JOGINDER PAL	BSC	3	St. Soldier	Google Drive	File storage
SOMNATH	BCA	4	CT Inst.	Web Hosting	Development
RAM PRAKASH	B.Tech	5	LKC	online money transfers	Finance
NAND SINGH	MBA	4	St. Soldier	ERP	Backup
BAL RAJ	Pharmacy	6	LKC	Facebook whatsapp	Big data analytics
RAMPAT THAKUR	B.Tech	2	DAV	Office Applications Email	Business Applications
SOHAN SINGH	HMCT	5	St. Soldier	Google Drive	File storage

```
Correctly Classified Instances         120               21.2389 %
Incorrectly Classified Instances       445               78.7611 %
Kappa statistic                          0
Mean absolute error                      0.2755
Root mean squared error                  0.3712
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              565

=== Detailed Accuracy By Class ===

TP Rate   FP Rate   Precision   Recall  F-Measure   Class
  0          0          0          0          0        File storage
  0          0          0          0          0        Development
  1          1          0.212      1          0.35     Finance
  0          0          0          0          0        Backup
  0          0          0          0          0        Big data analyticsÿ
  0          0          0          0          0        Business Applications

=== Confusion Matrix ===

   a   b   c   d   e   f   <-- classified as
   0   0  70   0   0   0 |   a = File storage
   0   0  89   0   0   0 |   b = Development
   0   0 120   0   0   0 |   c = Finance
   0   0 106   0   0   0 |   d = Backup
   0   0 109   0   0   0 |   e = Big data analyticsÿ
   0   0  71   0   0   0 |   f = Business Applications
```

Figure 1: Shows the parameters of Conjunctive Rule Algorithm with CCI, ICI, Error Rate and Accuracy on cloud dataset.

```
Correctly Classified Instances         229               40.531  %
Incorrectly Classified Instances       336               59.469  %
Kappa statistic                          0.2594
Mean absolute error                      0.2082
Root mean squared error                  0.3227
Relative absolute error                 75.5575 %
Root relative squared error             86.9332 %
Total Number of Instances              565

=== Detailed Accuracy By Class ===

TP Rate   FP Rate   Precision   Recall  F-Measure   Class
  0          0          0         0         0        File storage
  0          0          0         0         0        Development
  1          0          1         1         1        Finance
  0          0          0         0         0        Backup
  1          0.737      0.245     1         0.394    Big data analyticsÿ
  0          0          0         0         0        Business Applications

=== Confusion Matrix ===

   a   b   c   d   e   f   <-- classified as
   0   0   0   0  70   0 |   a = File storage
   0   0   0   0  89   0 |   b = Development
   0   0 120   0   0   0 |   c = Finance
   0   0   0   0 106   0 |   d = Backup
   0   0   0   0 109   0 |   e = Big data analyticsÿ
   0   0   0   0  71   0 |   f = Business Applications
```

Figure 2: Shows the parameters of EDTT with CCI, ICI, Error Rate and Accuracy on cloud dataset.

CONCLUSION

In this research, using EDTT built a analytical model of cloud environment, using the model after pruning algorithm the model was optimized by using the model after pruning algorithm and to evaluate Correctly classified instances, Incorrect classified instances and accuracy ratio of the classification method, it achieved good the excavation results.

In future course, we will review the various classification algorithms and significance of Prediction approach in designing of efficient classification algorithms for data mining. Selection of data and methods for data mining is an important task in this process and needs the knowledge of the domain.

REFERENCES

1. Ping Dong, Vice-processor, Junjun Dong, Engineer, Tiansheng Huang, Graduate student, "Application of Data Warehouse Technique in Educational Decision Support System", IEEE, 2006.
2. Ning Fang and Jingui Lu, "Work in Progress - A Decision Tree Approach to Predicting Student Performance in A High-Enrollment, High-Impact, and Core Engineering", 39th ASEE/IEEE Frontiers in Education Conference, October 18 - 21, San Antonio, 2009.
3. Pratiyush Guleria, Niveditta Thakur, Manu Sood, "Predicting Student Performance Using Decision Tree Classifiers and Information Gain", International Conference on Parallel, Distributed and Grid Computing, IEEE, 2014.
4. Tomas Hasbun, Alexandra Araya and Jorge Villalon," Extracurricular activities as dropout prediction factors in higher education using decision trees", IEEE 16th International Conference on Advanced Learning Technologies, 2016.
5. Hanjun Jin, Tianzhen Wu, "Application of Visual Data Mining in Higher-education Evaluation System", First International Workshop on Education Technology and Computer Science, 2009.
6. xiaojian.long, "Application of decision tree in student achievement Evaluation", International Conference on Computer Science and Electronics Engineering, 2012.
7. Sajan Mathew and Dr. John T. Abraham, "Application Of Data Mining In Higher Secondary Directorate Of Kerala", IEEE, 2016.
8. Ms.Tismy Devasia Ms.Vinushree T P, Mr.Vinayak Hegde, "Prediction of Students Performance using Educational Data Mining", International Journal of Innovative Research in Science, Engineering and Technology, March 2016; 3(3).
9. Fezile Matsebula, "A Big Data Architecture For Learning Analytics In Higher Education", IEEE Africon Proceedings, 2017.