Scholastic Seed Inc.

www.gjeis.com

global Journal of ENTERPRISE INFORMATION SYSTEM

Divers Principles, Algorithm & An Approach To Data Mining: A Comparative View

– Sachin Kumar*

Research Executive, Scholastic Seed Inc., New Delhi,India

ksack123@outlook.com *https://orcid.org/0000-0002-1540-5989*

– Prerita Talwar

Bachelor of Computer Applications, Amity University, (AIIT), Noida, India

prerita.talwar@gmail.com *https://orcid.org/0000-0002-1491-1732*

ARTICLE HISTORY

Paper Nomenclature: View Point
Paper Code: GJEISV11N4OD2019VP2
Submission Online: 04-Oct-2020
Manuscript Acknowledged: 05-Oct-2020
Originality Check: 07-Feb-2020
Originality Test Ratio: 20%
Peer Reviewers Comment: 12-Nov-2020
Blind Reviewers Remarks: 14-Feb-2020
Author Revert: 18-Feb-2020
Camera-Ready-Copy: 20-Dec-2020
Editorial Board Citation: 31-Feb-2020
Published Online First: 10-Feb-2020

ABSTRACT This paper proposes associate rule for data mining referred to as Ant-Miner (ant- colony-based information miner). The goal of Ant-Miner is to extract classification rules from the information. The rule is impressed by each kind of analysis on the behavior of real Emmet colonies and a few processing ideas additionally as principles. Classification is also an important topic in data mining analysis. Given a bunch of knowledge records, every of that belongs to a minimum of 1 amongst many predefined categories, the categorification downside cares with the invention of classification rules which is ready to allow records with unknown class membership to be properly classified. Several algorithms are developed to mine giant information sets for classification models which they're effective. However, once it involves deciding the possibility of every classification created, several of them aren't designed with such purpose in mind.

KEYWORDS ~~Green Banking~~ | ~~Sustainable Development~~ | ~~Rural India~~ | ~~Awareness~~ | ~~Financial Corporate Social Responsibility~~ | ~~Rural Banking~~

Introduction

In recent years, Infobahn has become the first suggests that for data dissemination. It's obtaining used for business, recreation, or academic functions, and,thus,

itsquality resulted in serious traffic on the net. Since the net capability is not keeping pace, the net impact of this growth was a large increase at intervals the user-perceived latency, that is, the time between once consumer problems an invitation for a document then the time the response arrives. Potential sources of latency square measure Infobahn servers' serious load, network congestion, low information

measure, information measure underutilization, and propagation delay a clear answer would be to extend the information measure. This doesn't appear a viable answer since the Web's infrastructure (Internet) can't be simply modified, while not a vital economic price. Aside from this price, higher information measures would ease users to create a lot of subtle and "heavy" documents, "choking" once more the network. Moreover, propagation delay can't be reduced on the far side a particular purpose since it depends on the physical distance between the act endpoints. The primary answer that was investigated toward the reduction of latency was the caching of internet documents at varied points at intervals the network (client, proxy, and server) Caching capitalizes on the temporalsection.

Effective consumer and proxy caches cut back the consumer perceived latency, the server load, then the amount of traveling packets, so increasing the offered information measure. Many caching policies square measure planned throughout the previous

years, particularly for proxy servers. Yet, the advantages reaped thanks to caching square measure typically restricted once internet resources tend to vary oft

***Corresponding Author** (Sachin Kumar)

https://doi.org/10.18311/gjeis/2019
Volume-11 | Issue-4 | Oct - Dec, 2019 | Online ISSN : 0975-1432 | Print ISSN : 0975-153X
Frequency : Quarterly, Published Since : 2009

Objective

In our last tutorial, we tend to study data mining Techniques. Today, we'll learn process Algorithms. we'll try to cowl all types of Algorithms in knowledge Mining: methodology primarily based Approach, Machine Learning-Based Approach, Neural Network, and Classification Algorithms in data processing, ID3 algorithmic program, C4.5 algorithmic program, K Nearest Neighbors algorithmic program, Naïve Thomas Bayes algorithmic program, SVM algorithmic program, ANN algorithmic program, forty-eight call Trees, Support Vector Machines, and Sense Clusters.

What is data mining Algorithms?

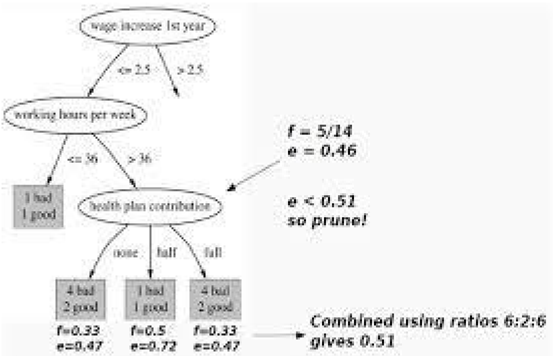
There is too several data mining Algorithms gift. We'll discuss each of them one by one.

These are the examples, wherever the data analysis task is Classification Algorithms in knowledge Mining-

A loan officer desires to analyses the data to understand that the client is risky or that are safe. A selling manager at a corporation should analyses a client with a given profile, the

WHO can purchase a replacement laptop.

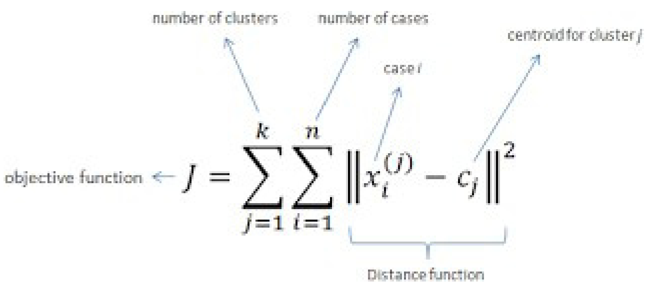
Types Algorithms in data mining 1. C4.5:



Source; - https://octaviansima.wordpress.com/2011/03/25/decision-trees-c4-5/

C4.5 is an associate degree algorithmic rule that is accustomed to generate a classifier at intervals the sort of an alternative tree and has been developed by Ross Quinlan. And to try and do to identical, C4.5 is given a bunch of information that represents things that have already been classified. C4.5 that is typically mentioned as an applied math classifier is Associate in nursing extension of Quinlan's ID3 algorithmic rule. The choice trees that area unit generated by C4.5 area unit typically any used for classification. The C4.5 algorithmic rule has conjointly been delineating as "a landmark call tree program that is in all probabilitythe machine learning workhorse most usually utilized in apply to date" by the authors of the rail machine learning code.

k-means:



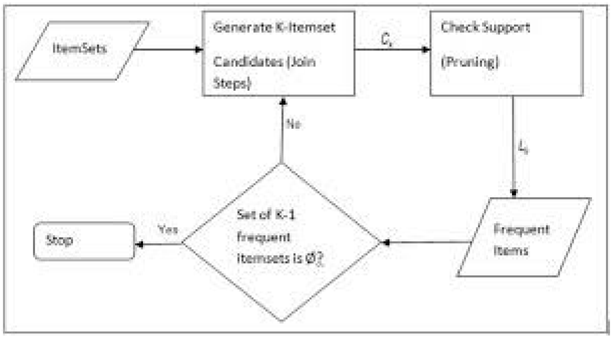
Source: - https://www.saedsayad.com/clustering_kmeans.htm

k-means cluster that is conjointly cited as the nearest center of mass classifier or The Rocha algorithmic program might even be a way of vector quantization, that is significantly widespread for cluster analysis in process .k-means is employed to create k teams from a gaggle of objects with care that the members of a gaggle area unit additional similar. It is an accepted widespread cluster analysis technique used for exploring a dataset.

Support vectormachines:

When it involves machine learning, support vector machines that square measure cited as support vector networks square measure supervised learning models that accompany associated learning algorithms that then analysesknowledge that is used for the analysis of regression and classification. associate degree SVM model is made that is an illustration of the examples as points in the house, that square measure any mapped thus as that the samples of the separate classes square measure then divided by a clear gap that has to be compelled to be as wide asdoable.

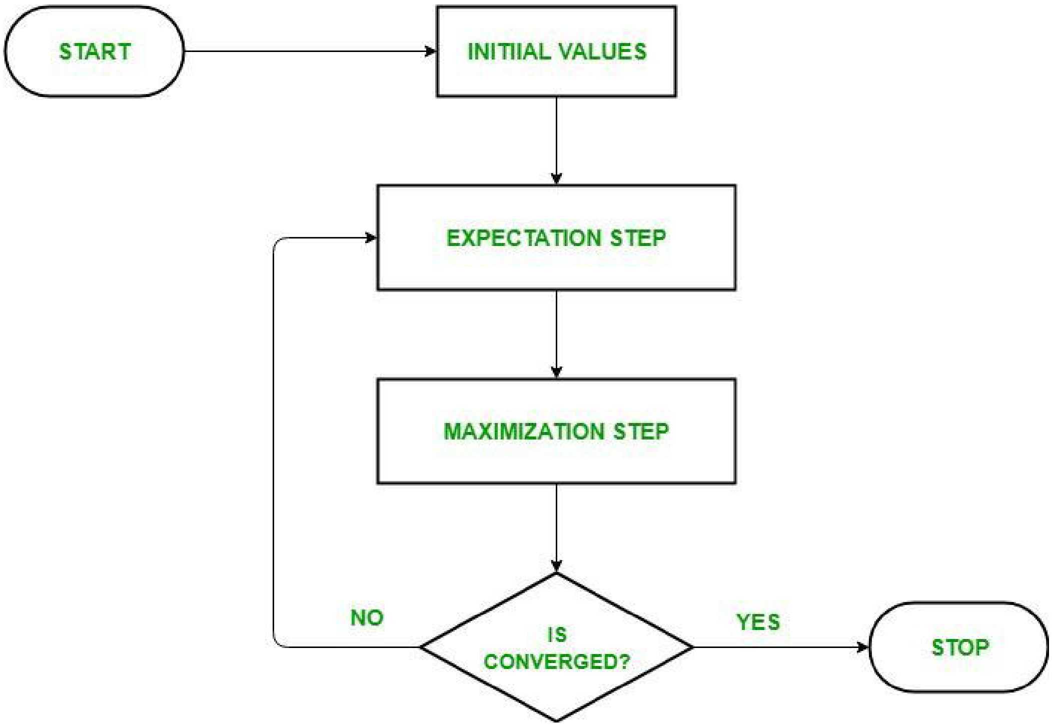
Apriority:



Source: - https://dwgeek.com/mining-frequent-itemsets-apriori-algorithm.html/

Apriori is an associate degree rule that is used for frequent itemset mining and association rule learning overall transactional databases. The rule is preceded by the identification of the individual things that are frequent among the information then extending them to larger itemset as long as sufficiently those item sets seem usually enough among the information. These frequent itemset that is determined by Apriority are usually used for the determination of association rules that then highlight generaltrends.

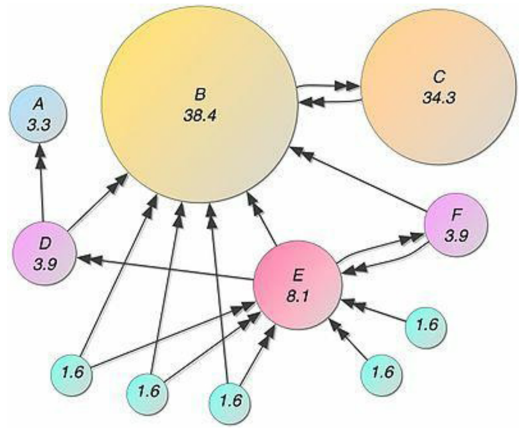
EM(Expectation-Maximization):



Source: -https://www.geeksforgeeks.org/ml-expectation-maximization-algorithm/

An expectation-maximization (EM) formula, once it involves statistics is an Associate in nursing unvarying technique that is accustomed to notice most a posteriori (MAP) ormost probability estimates of parameters in applied mathematics models that depend upon unobserved latent variables.

PageRank (PR):



PageRank (PR) that was named once Larry Page United Nations agency is one altogether the founders of Google is associate degree formula that is utilized by Google Search to rank the websites in their program results. PageRank, that's the first formula that was utilized by the corporate isn't the only formula that is obtaining utilized by Google to order program results, however it's the known manner of mensuration the importance of web site pages.

AdaBoost:

Adaptive Boosting or AdaBoost that has been developed by Yoav Freund and Henry M. Robert Schapiro is additionally a machine learning meta-algorithm that won the founders the 2003 Kurt Gödel Prize for constant. The algorithmic rule is usually used in composition with several different types of learning algorithms to spice up performance. AdaBoost is sensitive to screeching knowledge additionally as outliers.

kNN:

The k-nearest neighbors algorithmic rule (k-NN) is additionally a kind of lazy learning or instance-based learning and is taken into account as a non-parametric methodology that is used for classification and regression.\In each the mentioned cases, the input consists of the k nighest coaching examples among the feature area then the output depends on whether or not the algorithmic rule is obtaining used for classification or regression. This kNNalgorithmic rule is taken into account and is besides among the only of all machine learning algorithms.

NaiveBayes:

When it involves machine learning, Naive Thomas Bayes classifiers that square measure thought-about to be extremely ascendible square measure celebrated to be a family of simple probabilistic classifiers that square measure supported the appliance oftheorem with the assistance of sturdy independence assumptions between the options.

CART:

CART is an algorithmic program that stands for classification and regression trees. It's a selection tree learning technique that either outputs classification or regression trees and equally like C4.5, CART is, also a classifier. Several of the explanations that a user would use C4.5 for conjointly apply to that of CART since each of them square measure call tree learning techniques and options like straightforward interpretation and clarification applied to CART also.

Data Mining Implementation Process

Business understanding:

In this part, industrial enterprise and data-mining goals area unit established.

- First, you'd favour to acknowledge enterprise and client objectives. You'd favour to stipulate what your client desires (which repeatedly even they are doing not perceive themselves)
- Take inventory of the contemporary data processing situation. Offer some thought to resources, assumption, constraints, and alternative vast factors into yourassessment.
- Victimization industrial enterprise objectives and fashionable situation outline your records mining goals.
- Associate in nursing honest facts mining format is very special and should be developed to accomplish each business and statistics mining goals.

Data understanding:

In this part, saneness check on facts is run to check whether or not it's attractive for the information mining goals.

- First, data is gathered from several records sources accessible at intervals theorganization.
- These records sources might comprehend quite one database, flat filer or facts cubes. There square measures issues like object matching and schema integration which might arise throughout knowledge Integration method. It's a quite difficult and tough technique as statistics from quite range of sources not progressing to healthy simply. As Associate in nursing example, table A consists of Associate in nursing entity named custom whereas Associate in nursing other table B carries an entity named Cust-id.
- Thus, it's pretty laborious to form certain that every of those given Objects raise identical worth or not. Here, information ought to be accustomed minimize mistakes at intervals the records integration method.
- Next, the step is to look for homes of obtained knowledge. A right because of discover the information is to reply the facts mining queries (decided in enterprise phase) the utilization of the question, reporting, and visualizationtools.
- Supported the outcomes of question, the information quality ought to be determined. Missing statistics if any ought to beno inheritable.

Data preparation:

In this section, knowledge is formed production prepared.

- The data preparation method consumes concerning ninetieth of the time of the project. The data from extraordinary sources ought to be hand-picked, cleaned, remodeled, formatted, anonym zed, and designed (if required).
- Data cleansing may even be manner a technique} to "clean" the records by way of smoothing screaming data and filling in lacking values.
- For example, for a client demographics profile, age records are missing. The statistics is incomplete and should be stuffed. In some cases, there might be records outliers. As associate degree example, age options a fee three hundred. Knowledge may wish to be inconsistent.
- As associate degree example, determine of the shopper is outstanding in distinct tables. Data transformation operations alternate the records to make it useful in data mining. Following transformation area unit typically applied

Data transformation:

Data transformation operations would create a contribution to the success of the mining method.

- Smoothing: It helps to eliminate noise from the information.
- Aggregation: outline or aggregation operations square measure applied to the information. I.e., the weekly financial gain info is aggregative to calculate the monthly and every year total.
- Generalization: throughout this step, Low-level info is changed by victimization higher- level concepts with the assistance of plan hierarchies. For instance, the city is replaced by the victimization of the county.
- Normalization: normalization administered once the attribute statistics square measure scaled up o scaled down.
- Example: information has to be compelled to fall at intervals the vary -2.0 to 2.0 post- normalization.

- Attribute construction: these attributes square measure developed and blanketed the given set of attributes helpful for statistics mining.
- The results of this method may even be a final record set that is in a position to be used in modeling.

Modeling

In this section, mathematical fashions square measure accustomed to decide info patterns.

- Supported the enterprise objectives, applicable modeling strategies have to be compelled to be chosen for the readydataset.
- Produce a scenario to require a glance at taking a look at the exceptional and validity of the model.
- Run the mannequin on the organizeddataset.
- Results ought to be assessed via all stakeholders to make positive that mannequin will meet info mining objectives.

Evaluation:

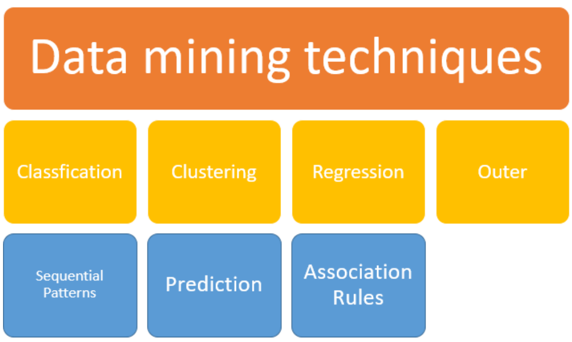
In this section, patterns known square measure evaluated con to the industrial enterprise objectives.

- Results generated by the records mining mannequin have to be compelled to be evaluated con to the business enterpriseobjectives.
- Gaining enterprise perception is an associate unvaried method. While understanding, new business enterprise wants may be raised because of infomining.
- A go or no-go call is taken to travel the model among the readyingsection.

Deployment:

In the readying part, you ship your statistics mining discoveries to daily business operations.

- The experience or records ascertained at intervals the course of information mining systems needs to be created easy to grasp for non-technicalstakeholders.
- A singular readying set up, for shipping, maintenance, and observation of facts mining discoveries ismade.
- A final assignment file is made with directions discovered and key experiences at intervals the course of the project. This helps to strengthen the Organization's business enterprise policy.



Classification:

This analysis is employed to retrieve important and relevant statistics concerning knowledge, and information. This records mining technique helps to classify knowledge in one all

- Told a part of speech.

Clustering:

- Clustering analysis is an associate degree data mining technique to
- Urge data that are like each alternative. This method helps to acknowledge the variations and similarities between the information.

Regression:

- Regression analysis is that the data mining methodology of distinctive and analyzing
- The affiliation between variables. It's accustomed to notice the chance of a singular variable, given the presence of alternative variables.

AssociationRules:

- This records mining approach helps to urge the affiliation between 2 or larger things. It discovers a hidden sample inside the information set.

Outerdetection:

- This kind of records mining methodology refers to the observation of information gadgets inside the dataset that do not match a foretold sample or foretold behavior. This method are typically used in associate
- Degrees passing quite domains, like an intrusion, detection, fraud or fault detection, etc. Outer detection is also explicit as Outlier Analysis or Outlier mining.

ConsecutivePatterns:

This statistics mining technique helps to urge or discover similar patterns or tendencies in dealing with knowledge uncalled-for to say the amount.

Prediction:

The prediction has used a mix of the numerous data mining strategies like Trends, consecutive patterns, clustering, classification, etc. It analyses past activities or instances throughout the right sequence for predicting a future event.

Comprehensive List of tools for Data Mining

The 6 companies has been taken into consideration in a comparative manner

S. No	About	Logo	Modus-Operandi	Used For
1	RapidMiner is a data science software platform developed by the company of the same name that provides an integrated environment for data preparation, machine learning, deep learning, text mining, and predictive analytics		RapidMiner is additionally a liberal to use process tool. it's used For information preparation, pc learning, and model preparation. It presents a Variety of merchandise to create new facts mining methods and prognostic setup analysis. Features: • enable several information administration was yes • interface orexecution • Integrates with in-house databases • Interactive, shareable dashboards • massive information prognostic analytics • Remote analysis process • information filtering, joining, merging, and aggregating • Build, educate and validate prognosticmodels • Reports and precipitatednotifications	RapidMiner is a data science software platform developed by the company ofthe same name that provides an integrated environment for data preparation, machine learning, deep learning, text mining, and predictive analytics.
2	R is also a language and surroundings for statistical computing and graphics. it is a GNU assignment that is paying homage to the S language and surroundings which was once as soon as developed at Bell Laboratories by John Chambers and colleagues. R is often considered as an incredible implementation of S. R presents a considerable type of statistical and graphical techniques, and is extremely extensionble.		R might even be a language for applied Math computing and graphics. It also used for big information analysis. It presents a broad vary of applied math tests. Features: • Effective statistics managing and storage facility, • It provides a gaggle of operators for calculations on arrays, especially,matrices, • It provides a coherent, inherent series of large facts tools for infoanalysis • It provides graphicalamenities • For information analysis that show either on- screen or on hardcopy. • It offers graphicalfacilities • For info analysis that show each on-screen or on hardcopy.	R may additionally be a free software program environment for statistical computing and graphics written in C++. R Studio is IDE specifically design need for the R language. It is one in each of the foremost tools used to do records mining responsibilities and comes with big community help as nicely as packaged with plenty of libraries inbuilt specific for data mining
3	KNIME is that the magnificent in titration platform for data analytics and reporting developed via was y of KNIME. com AG. It operates on the questioning of the Modular information on pipeline. KNIME constitutes of range Desktop learning a ND statistics price sing factors ember added together.		KNIME is open to provide a software system program for making statisticsscience Applications and services. This processing device helps you to acknowledge info and to layout facts science workflows. Features: • Helps you to create AN stop to quit facts scienceworkflows • permits you to the mixture, sort, filter, and be an element of facts either on your close machine, in-database or In assigned vast knowledge environments. • Build laptop mastering models for classification, regression, dimensiondiscount	Primarily used for information pre- processing — i.e. data extraction, transformation, and loading, Knime must be A prime quality device with GUI that suggests the community of fact nodes. Popular amongst monetary information analysts, it's modular documents pipelining, leveraging computer learning, and statistics mining ideas liberally for constructing enterprise talent Reports.

4	Statistical Analysis System (SAS) is a product of the SAS Institute developed for analytics & information ma engagement. SAS can mine data, alter it, control document nets from particular sources And overall performance states tidal analysis. It gives a graphical UI for non-technical users	sas THE POWER TO KNOW.	Statistical Analysis System may even be a product of SAS. it had been developed for analytics And knowledge management. It presents a graphical UI for not technical users. Features: • SAS processing tools assist you in analyzing hugeknowledge • It is an excellent device for processing, text mining &improvement. • SAS offers adistributed • Memory process design that is a variety of ascendible. • on discount	SAS information mine r permits customers to lookup Giant statistics and derives correct percept ion to structure well-timed decisions. SAS carries a dispense d memory processing structure which Is tremendously scalar le. it is Precisely acceptable of r facts mining, text mi nine & optimization.
5	Teradata is commonly called the Teradata database. It's an enterprise facts warehouse that includes information on administration tools together with data processing software re. It may be used for enterprise analytics.	TERADATA	Teradata is additionally a vastly parallel Open process device for developing large-scale statistics reposition applications. Teradata will run on Unix/Linux/Windows server platform. Features: • Teradata Optimizer will manage up to sixty- four joins throughout anissue. • Timandra records options associate the occasional whole price of possession. It's convenient to line up, maintain, andhandle. • It helps SQL to possess interaction with the facts saved in tables. It provides itsextension. • It helps you to distribute the data to the disks robotically with no manualintervention. • Teradata offers load & dump utilities to cross statistics into TeradataSystem.	Teradata is employed to very own a perception of or animation records like sales, product placement, patron prep erects, etc. it can moreover different irate between 'hot' & 'cold' data, which potential that it puts much less often used data in An exceptionally slug is storage section.
6	Sisense is a very beneficial an d first-class suited BI software program when it entails reporting functions inner the organization. it is developed by using the ability of the agency of equal Discover 'Sisense'. it's a superb performance to manipulate and approach facts for the little scale/large scale organizations.	SISENSE	Sisense is the different quality facts processing tool. It immediately analyzes and visualizes both massive and disparate datasets. it's an ideal gadget for growing dashboards with a massive kind of visualizations. Features: • Allows to assemble interactive dashboards with no techskills • Create one mannequin of the truth with seamless data • Unify unrelated facts into one centralized place • East drag-and-drop character interface • Allows to set off right of entry to dashboards even within the cell device • Eye-grabbingvisualization • Identifies vital metrics the use of filtering and calculations • Handles massive scale statistics at one commodityserver	Sisense is a very beneficial and first-class suited BI software program when it entails reporting functions inner the organization. it is developed by using the ability of the agency of equal discover 'Sisense'. it's a superb performance to manipulate and app roach facts for the little scale/large scale organizations.

Data is truly priceless. But it is no longer a cake stroll to lookup it as increased matters come at a bigger cost. With the exponential increase in data, there requires a technique to extract meaningful files as conclude to useful insights. Data mining is that the approach where the invention of patterns among giant units of understanding to seriously trade it into incredible records is performed. This method makes use of unique algorithms, statistical analysis, synthetic intelligence, and database structures to juice out The statistics from large datasets and convert them into an understandable form. Above are the 6 comprehensive data mining tools which are widely used in the big data industry.

Reference

- https://www.solver.com/frontline-systems-releases-solver-sdk-platform-v2016-support-r-python-and-rason-languages
- https://www.solver.com/blog/whats-new-solver-sdk-v2016-r-and-python-support
- https://www.semanticscholar.org/paper/A-Comparative-Analysis-of-Data-Mining-Tools-in-Christa-Madhuri/24c5041d044a590cef9528c9c655a31c49051db0
- https://www.businesswire.com/news/home/20130429005524/en/New-SAS%C2%AE-High-
- Performance-Analytics-products-deliver-performancehttps://paralleldesk.com/job-details/online-analytics
- https://books.google.se/books?id=IROhDQAAQBAJ&pg=PA347&lpg=PA347&dq=%22the+data+mining+method+of%22&source=bl&ots=A8RQrXmEJm&sig=ACfU3U35hn6nU5VDiEja0NDJMjC88ZqAog&hl
- =sv&sa=X&ved=2ahUKEwiW0IKgltbnAhUsAxAIHRz_D7MQ6AEwDnoECAgQAQhttps://insidebigdata.com/2013/05/02/sas-adds-to-its-big-data-analytics-product-lineup/
- https://ieeexplore.ieee.org/abstract/document/7867310
- https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118029145.app1
- https://in.wisdomjobs.com/jobs/software-developer-jobs-in-bengaluru-ibm-india-pvt-dot-limited-4-dot-0-6217-revi-openings-23170647
- https://books.google.com/books?id=kPm4DwAAQBAJ&pg=PT585&lpg=PT585&dq=%22data+mining
- +text+mining+optimization%22&source=bl&ots=EDtnrSD-fD&sig=ACfU3U2nNNx5ZKW-9rCN7udwdZlU3PQyRQ&hl=en&sa=X&ved=2ahUKEwjThpOgltbnAhWR_J4KHZ-TyAo4Q6AEwA3oECAsQAQ
- https://books.google.se/books?id=mxncBQAAQBAJ&pg=PA155&lpg=PA155&dq=%22the+data+mining+method+of%22&source=bl&ots=Y34SmETpv2&sig=ACfU3U3c9FFO-vCUXUzLyFulasnwecIrqg&hl=sv&sa=X&ved=2ahUKEwiW0IKgltbnAhUsAxAIHRz_D7MQ6AEwD3oECAcQAQ
- https://ieeexplore.ieee.org/document/1255389https://towardsdatascience.com/data-mining-tools-f701645e0f4c
- https://books.google.ch/books?id=f9vRpOidzHsC&pg=PA236&lpg=PA236&dq=%22of+tools+for+data+mining%22&source=bl&ots=6r-CNavCGy&sig=ACfU3U2Y45P11zD-h6mTaQcYO3TuM48P4A&hl=de&sa=X&ved=2ahUKEwjx4fWfltbnAhVOd6wKHR7ABl4Q6AEwE3oECAsQAQ
- https://in.wisdomjobs.com/jobs/data-scientist-jobs-in-bengaluru-educational-initiatives-private-limited-openings-5021601
- https://ieeexplore.ieee.org/document/4771925/https://www.coursehero.com/file/19574605/ba-sg/
- https://www.informationweek.com/software/information-management/sas-gets-hip-to-hadoop-for-big-data/d/d-id/1106844?print=yes
- https://www.ijert.org/data-mining-tools-an-analytical-approach
- https://www.computer.org/csdl/journal/tk/2018/08/08259370/13rRUxASuvL
- https://www.nsf.gov/awardsearch/showAward?AWD_ID=1758807&HistoricalAwards=false
- https://theswissbay.ch/pdf/Gentoomen%20Library/Data%20Mining/ISR.Encyclopedia.Of.Data.Warehousing.And.Mining.2nd.Edition.Sep.2008.eBook-ELOHiM.pdf
- https://stackoverflow.com/a/16233417https://www.museumsandtheweb.com/paper_keywords/metadata_enrichment.html
- https://wikihub.berkeley.edu/display/istrit/Framework+for+workflows+with+task+dependencies+on+HPC
- https://www.adaptivecomputing.com/tag/big-data/feed/https://link.springer.com/article/10.1023/A:1022497517599https://www.adaptivecomputing.com/blog-hpc/nfl-hpc-big-data/https://www.nextchapter.pub/books/crow-of-thornshttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6696378/
- https://cran.r-project.org/web/packages/ClusterR/vignettes/the_clusterR_package.htmlhttps://www.museumsandtheweb.com/mw2011/session/linked_data.html
- https://books.google.com.au/books?id=vMJNDwAAQBAJ&pg=PA132&lpg=PA132&dq=%22in+this+phase+data+is%22&source=bl&ots=n_g04kEUZZ&sig=ACfU3U2FVvykyYdsm_chqaSJSBPSkAuo-GQ&hl=en&sa=X&ved=2ahUKEwjCiPm7ltbnAhXcITQIHd5RANwQ6AEwA3oECAcQAQ
- https://blog.saleslayer.com/data-lifecycle-management-what-you-should-know
- https://books.google.com.au/books?id=Fp-MdU1q2AkC&pg=PA112&lpg=PA112&dq=%22in+this+phase+data+is%22&source=bl&ots=WdqYuK9GOb&sig=ACfU3U1H1OVoUKUh1yHGXUiVMxOwhuBasw&hl=en&sa=X&ved=2ahUKEwjCiPm7ltbnAhXcITQIHd5RANwQ6AEwBXoECAwQAQ
- https://books.google.com.au/books?id=PKLTDAAAQBAJ&pg=PT193&lpg=PT193&dq=%22in+this+phase+data+is%22&source=bl&ots=vqZgeA33sY&sig=ACfU3U0YYWFzlxkAzBztwFli_hwpAY9n1Q&hl=en&sa=X&ved=2ahUKEwjCiPm7ltbnAhXcITQIHd5RANwQ6AEwBHoECAkQAQ
- https://link.springer.com/content/pdf/10.1023/A:1022497517599.pdfhttps://www.cs.rit.edu/~amt/dataminingFall04/DMFall04Schedule.html
- https://books.google.com.au/books?id=oLqeBQAAQBAJ&pg=PA1153&lpg=PA1153&dq=%22in+this+phase+data+is%22&source=bl&ots=xnh6zIsvZl&sig=ACfU3U3DyCnGmzU_LhP3jY1nSKjYXa79ag&hl=en&sa=X&ved=2ahUKEwjCiPm7ltbnAhXcITQIHd5RANwQ6AEwAnoECAYQAQ
- https://www.guru99.com/digital-forensics.htmlhttps://study.com/academy/lesson/nursing-process-purpose-and-steps.html
- https://books.google.be/books?id=5qb9sCXdBQYC&pg=PA261&lpg=PA261&dq=%22and+a+few+data+mining%22&source=bl&ots=Hu3mL0bile&sig=ACfU3U1q_0oim9YuF-gryS93lnvRwfcH7A&hl=nl&sa=X&ved=2ahUKEwjovKK9ltbnAhVDSxoKHeLCA20Q6AEwAHoECAoQAQ
- https://books.google.at/books?id=or8WODqdsdYC&pg=PA259&lpg=PA259&dq=%22an+algorithm+for+data+mining%22&source=bl&ots=g_zphd8D4F&sig=ACfU3U30pg6lWJqfvOjA_jOoW3ihLMx_2w&hl=de&sa=X&ved=2ahUKEwjqhIi-ltbnAhWkNn0KHSqwCMEQ6AEwCXoECAsQAQ

- https://books.google.com.au/books?id=fmBeAgAAQBAJ&pg=PA52&lpg=PA52&dq=%22in+this+phase+data+is%22&source=bl&ots=NZ3C8I-Z48&sig=ACfU3U28_W8OQ7AMTItpA0mHq4BhWKv-
- _w&hl=en&sa=X&ved=2ahUKEwjCiPm7ltbnAhXcITQIHd5RANwQ6AEwBnoECAgQAQhttps://www.grin.com/document/441758
- https://www.semanticscholar.org/paper/MineTool-3-DM-2-%253A-An-Algorithm-for-Data-Mining-of-3-Sipes-Karimabadi/c124be594f2725345b65f3fa3c6548fa4494859d
- https://ieeexplore.ieee.org/document/1027744
- https://www.academia.edu/34507683/An_Algorithm_for_Data_Mining_On_Fuzzy_Weighted_Association_Rules
- https://www.freelancer.com/u/masterpythonhttps://whatis.techtarget.com/definition/data-life-cycle
- https://www.reddit.com/r/OSU/comments/3wt7ny/any_osu_courses_that_deal_with_ibm_watsonbig_data/
- https://run.unl.pt/bitstream/10362/31971/1/TEGI0400.pdf
- https://www.researchgate.net/publication/220451882_Sampling_and_Subsampling_for_Cluster_Analysis_in_Data_Mining_With_Applications_to_Sky_Survey_Data
- https://www.researchgate.net/publication/50946165_A_Review_on_Data_mining_from_Past_to_the_Future
- https://www.ijcaonline.org/archives/volume15/number7/1961-2623
- https://www.researchgate.net/publication/316531075_A_frequency-domain_adaptive_filter_FDAF_prediction_error_method_and_ARLS_for_speech_echo_cancellation
- https://wenku.baidu.com/view/d4c698bcf121dd36a32d82e0.htmlhttps://bsbi.org/ScottishNsl2012.pdf
- https://www.researchgate.net/publication/263203644_Large_Scale_and_Big_Data_Processing_and_Management
- https://www.researchgate.net/figure/Estimated-MSE-and-misalignment-for-the-cXMNL-NLMS-algorithm_fig13_224218410
- https://research-portal.uws.ac.uk/files/912138/ASOC_GAF-NN_Revision24June2015Final.pdf
- https://www.researchgate.net/publication/281817338_A_Survey_Data_Classification_Approaches_and_Tools
- https://link.springer.com/chapter/10.1007/978-1-4302-5990-9_1
- https://rd.springer.com/chapter/10.1007%2F978-1-4302-5990-9_1
- https://www.researchgate.net/publication/29467751_Top_10_algorithms_in_data_mining
- https://www.softwaretestinghelp.com/data-mining-techniques/

GJEIS Prevent Plagiarism in Publication
The Editorial Board had used the turnitin plagiarism [http://www.turnitin.com] tool to check the originality and further affixed the similarity index which is 20% in this case (See Annexure-I). Thus the reviewers and editors are of view to find it suitable to publish in this Volume-11, Issue-4, Oct-Dec, 2019

Annexure 1

Submission Date	Submission Id	Word Count	Character Count
21-Oct-2019	1203455879 (turnitin)	7037	43571

URKUND

Urkund Analysis Result

Analysed Document: 4.1 RT-1 peggy chowdhary 17-02-20.pdf (D64021025)
Submitted: 2/17/2020 5:16:00 PM
Submitted By: editorial.scholastic.seed@gmail.com
Significance: 7 %

Sources included in the report:

Nguyen_Anh-Tuan_MAS-BI_MAS-1DA18_P.pdf (D40509296)
Medicinal Counterfeiting Paper.docx (D56980400)
EE_TU49_Manal_Al Hosawi.docx (D40583315)
Literature review_Liu_Lundin_Wareborn 1101.docx (D58209667)
https://www.ncbi.nlm.nih.gov/books/NBK202520/
https://www.ice.gov/factsheets/ipr-pangea
https://www.ncbi.nlm.nih.gov/books/NBK202531/
https://www.coursehero.com/file/p19c7jc3/An-international-study-by-INTERPOL-in-2014-identified-two-well-established/
https://cmr.asm.org/content/28/2/443
https://www.cbp.gov/sites/default/files/assets/documents/2019-Aug/IPR_Annual-Report-FY-2018.pdf
2d012325-621b-4443-8967-bc39eecc8eb9
https://www.fraserinstitute.org/sites/default/files/pharmaceutical-counterfeiting-endangering-public-health-society-and-the-economy.pdf
https://www.todaysgeriatricmedicine.com/archive/ND16p10.shtml
https://www.reajetus.com/wp-content/uploads/2016/04/Pharma-Crime-Sub-Directorate.pdf
https://enact-africa.s3.amazonaws.com/site/uploads/2018-11-12-counterfeit-medicines-policy-brief.pdf

Instances where selected sources appear: 28

Reviewers Comment

Reviewer's comment 1:
The data mining models that are widely used to extract valuable knowledge from huge amounts of data. The criteria used to evaluate the classifiers are mostly accuracy, computational complexity, robustness, scalability, integration, comprehensibility, stability, and interestingness.

Reviewer's comment 2:
This study compares the classification of algorithm accuracies, speed (CPU time consumed) and robustness for various datasets and their implementation techniques

Reviewer's comment 3:
In this paper reviewed the utility and application of data mining technique in the field of privacy. Preservation. Privacy preservation is technique for hiding of information and secured the information during transmission.

Citation

Sachin kumar & Prerita Talwar
"Divers Principles, Algorithm & an Approach to Data Mining: A Comparative View"
Volume-11, Issue-4, Oct-Dec, 2019. (www.gjeis.com)

https://doi.org/10.18311/gjeis/2019
Volume-11, Issue-4, Oct-Dec, 2019
Online ISSN : 0975-1432, **Print ISSN** : 0975-153X
Frequency : Quarterly, Published Since : 2009

Google Citations: Since 2009
H-Index = 96
i10-Index: 964

Source: https://scholar.google.co.in/citations?user=S47TtNkAAAAJ&hl=en

Conflict of Interest: Author of a Paper had no conflict neither financially nor academically.

EDITORIAL BOARD EXCERPT At the time of submission, the paper had.20 % of plagiarism which is an accepted percentage as per the norms and standards of the journal for the publication. As per the editorial board's observations and blind reviewers' remarks the paper had some minor revisions which were communicated on timely basis to the authors (Sachin & Prerita) and accordingly all the corrections had been incorporated as and when directed and required to do so. The comments related to the manuscript are related to the theme "Divers, Principles, Algorithm & an Approach to Data Mining" both subject-wise and research-wise. Data mining via utilizing specific algorithms, statistical analysis, and artificial intelligence & database systems aims to discover the patterns among large volumes of data to extract information from it and to convert it into more refined and understandable structure for future use. The algorithm in data mining (or machine learning) is a set of heuristics and calculations that creates a model from data. To create a model, the algorithm first analyses the data you provide, looking for specific types of patterns or trends. Overall the paper promises to provide a strong base for the further study in the area. After comprehensive reviews and editorials boards remarks the manuscript has been decided to categories and publish under the "View Point" category.