A Survey Analysis on Data Mining Techniques: ANN & VSM

.PROF.KALAM NARREN
[1, 2, 3] Department of CSE, Vignan's Institute of Management and Technology for Women, Hyderabad, India

ABSTRACT:

The knowledge or data extraction from some huge collection of data that is preserved in many data collections, which are heterogeneous, is termed as Data mining. The message relaying in this procedure is sometimes direct and sometimes indirect. The research in the current study is to provide a survey on various aspects of data mining in other fields. The aspects are the Artificial Neural-Networks(ANN), Vector-Space Model (VMS)and Spatial-Image Mining (SIM)along with the artificial neural-networks. This paper furnishes the data regarding the aspects of data mining in the above mentioned themes. The current research visualizes the SIM with the relative implementations and computations utilizing VSM and artificial neural-networks.The comparison between different algorithms like BM-Cilin, VSM and VSM-Cilin with respected to different datasets or different types of areas like Entertainment, Military, Sports and Political. Among all our method got better results the

Keywords: Data-mining, VSM, RST, SIM, Neural-networks.

I. INTRODUCTION:

The current enhancements in the IT have resulted in an outrageous growth of data in the last decade in many areas. The process of storing and processing the present data and decision making have been on research for quite some time for enhancements. The methodology in which the extraction of patterns and some data is done from a huge set of data is termed as data mining. It's a logical method which is preferred to compute on data sets to find some useful information. The datasets are very essential for the determination of the patterns and abide their optimum. Requirementsand accuracy of the system developed.The term "data mining" is often used interchangeably with KDD. The term confusion is understandable, but "Knowledge Discovery of Databases" is meant to encompass the overall process of discovering useful knowledge from data. Meanwhile "data mining" refers to the fourth step in the KDD process. This is commonly thought of the "core step" which applies algorithms to extract patterns from the data. It parallels the "modeling" phase of other data science processes.

While KDD variants can range from 5 to 7 steps, many influential and authoritative voices on the matter regard KDD as the following 5-steps process:Selection,Pre-processing,Transformation:Data Mining:Interpretation/Evaluation:The below figure-1 shows An Outline of the Steps of the KDD Process.

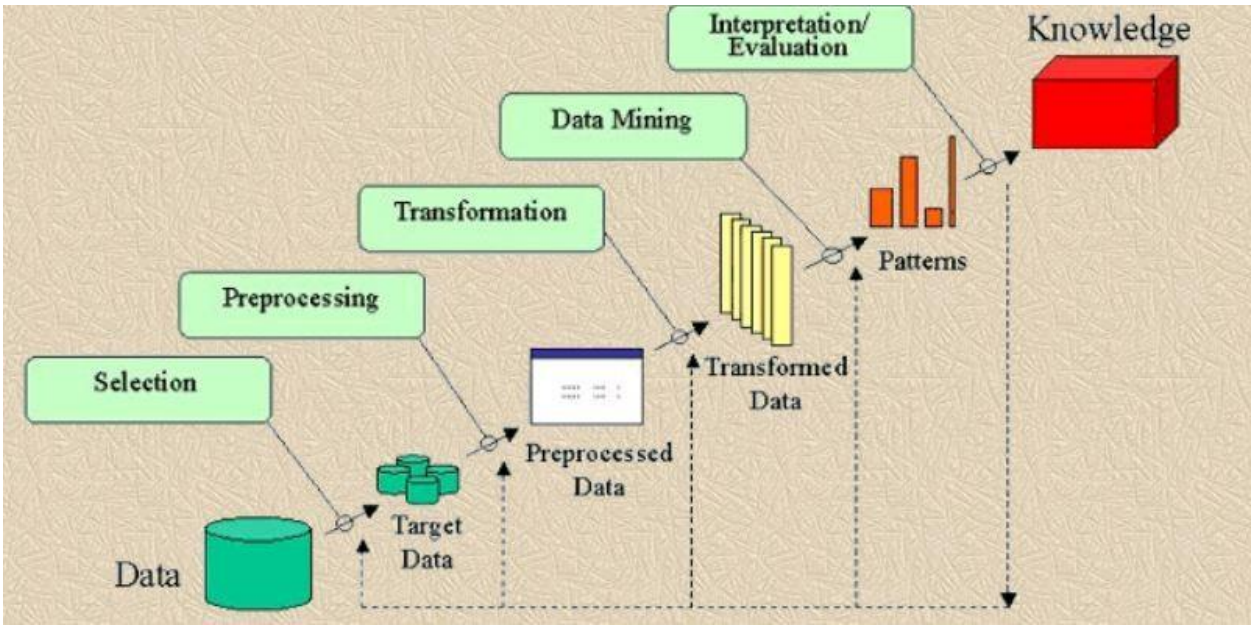

FIGURE-1: An Outline of the Steps of the KDD Process

Unknown pattern discovery is the theme of this methodology to be put forth. After the discovery of patterns, we can research more about their use cases and functionalities on how the decisions are made and how it can help for the business development.

To find anything we need a sequence of steps as a procedure to follow. They are

1. Exploration
2. Identification of patterns
3. Deployment

For the start step, the presented data is preserved and transformation of data from one form to another takes place which gives chance for the developer to identify and determine the data nature and variables of the problem.For the next step, after the exploration of data, refinement of data along with definition of the data identification of patterns takes place and the developer will have to choose a pattern, which gives the optimum prediction and computation.

The research in the current study is to provide a survey on various aspects of data mining in other fields. The aspects are the Artificial Neural-Networks(ANN), Vector-Space Model (VMS)and Spatial-Image Mining (SIM)along with the artificial neural-networks. This paper furnishes the data regarding the aspects of data mining in the above-mentioned themes. The current research visualizes the SIM with the relative implementations and computations utilizing VSM and artificial neural-networks. The comparison between different algorithms like BM-Cilin, VSM and VSM-Cilin with respected to different datasets or different types of areas like Entertainment, Military, Sports and Political. Among all our method got better results the

The organization of this paper as follows, in the next section explain about the relatedwork and literaturesurvey. In section-III, explain about the methodologies like ANN,SIM and VSM. Next comparison in section-IV Finally conclusion of this paper in section-V.

II. RELATED WORK

Many researchers have developed certain measure for the implementation of datasets and their pattern identification along that many have done various implementations of knowledge discovery in plenty of aspects. The mining algorithms and methodologies are very essential for the identification of new possibilities for new researches and many more patterns to compute towards accuracy with respect to speed and time. E-catalogue, word matching, product matching are some of the implementations done by many researchers in various fields and aspects.

2.1 Literature Survey

There are many works on these aspects but some surveyed for the beneficiary of this paper. The study called data mining concepts along with techniques, which is furnished by Morgan_Kauffman as a second edition, and it is written by Jiawci_Han in collaboration with Micheline_Kamber. There are other influences like the emerging fields in the study of knowledge discovery by Gary_Parker are also a good reference of literature. Stepwise data mining guide presents the crisp-DM_1.0 which can be accessed from the internet. Student performance evaluation by the use of data mining methodologies by the professor Rimmy_Chuchra_Er which has been published in the international journal of CSMR. The methodologies and applications which are distinct using Data mining which is a survey of literature is illustrated by the Aakanksha_Bhatnagar along with Shweta_P_Jadye in collaboration with M_M_Nagar is also one of the published research work. Educational information mining for the performance of the students is also a published paper in an international journal by the Brijesh_K_Baradwaj.

III METHODOLOGIES

In section, we explain about the methodology like ANN, SIM and VSM.

3.1.1 Vector-Space Model (VMS)

The process in which all the existing facts or data is represented in the form of identifiers to vectors. It is primarily used in the filtration mechanisms of data along with the indexing of data represented or computed and ranking is done based on their relevancies.Conversion of the similarities in the document to some angle b/w vectors where for each word taken a dimension as every document is noted as a space vector while the words are assumed as independents, which therefore minimizes the complexity relationships between the words and the computations. The below figure-2 shows that working of the VMS. Where d and q are the parameters of this model.

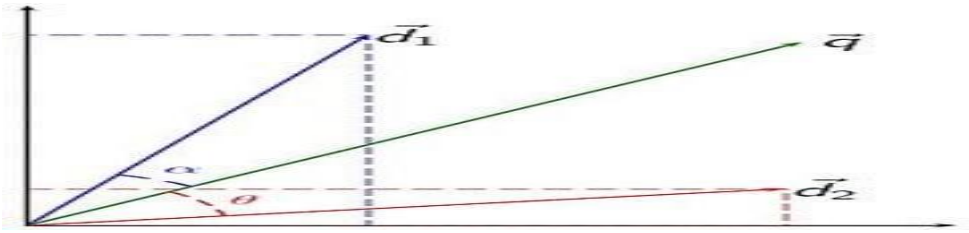

FIGURE-2: The working of the VMS

All vectors beneath hobby via way of using the current methodology the entitywise positive, a cosine value for 0 for the fee, this way the question and corresponding document vector to it are orthogonal and have no skillful representation. In addition statistics we have to observe the similarities in the cosine values.

3.1.2 Artificial Neural-Networks(ANN)

Artificial neural network simulate the functions of the neural network of the human brain in a simplified manner. ANN act as the pattern recognized problem. A neural network consists of three layers. The first layer is the input layer. It contains the input neurons that send information to the hidden layer. The hidden layer performs the computations on input data and transfers the output to the output layer. It includes weight, activation function, and cost function.The below figure-3 shows that the Architecture of the ANN.

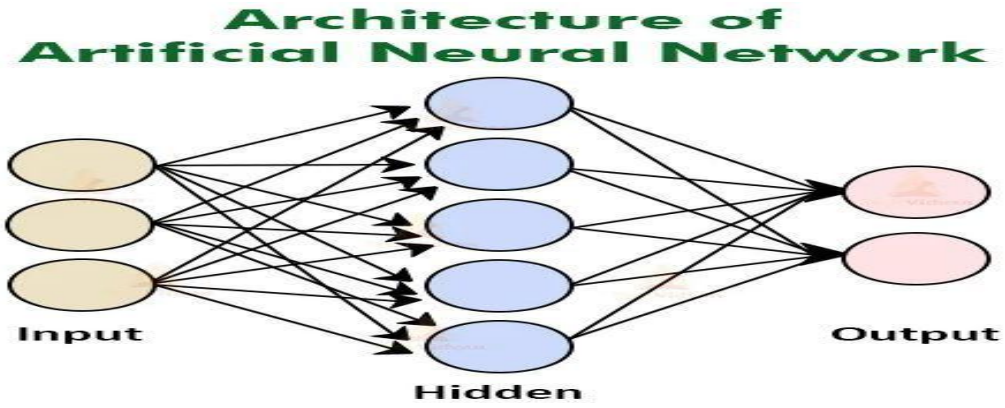

FIGURE-3: The Architecture of the ANN.

Aswe, all know that the computer models are the critical contributors in understanding the functionality of the brain through artificial means of experimentations and one such means is the artificial neurons developed which are also computational models for the development in cognitive neural science. Their root feature is that they can easily handle the complexity we present to them and produce results which we can analyze by means of comparison with the actual results. It follows the basic principle of generating the model computation then test run the computation and calculate the computational results of the computer model. We plan and construct a computational approach such that the model follows a less complex methodology and

can retrieve maximum computations and get the massive base of information. The following are the different types of Artificial Neural Network namely Feedforward Network and Feedback network shown below figure-4&5 respectively.

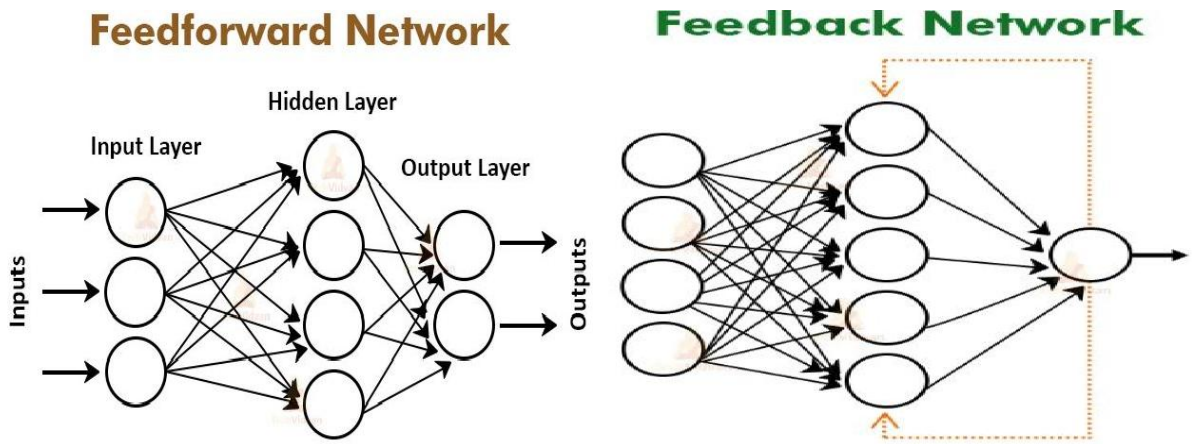

FIGURE-4: Types of ANN

3.1.3 Spatial-Image Mining (SIM)

From an image, if there is an extraction of facts then it is called as the spatial-image mining. Usage of given image mining is to enable the transformation of a set of images into data and it also proves hidden undertaking by datasets of mage with the use if methodologies regarding data mining. The spatial mining strategies explored is beneficial for unique software and on precise or particular pictures, mining strategies targets now not handiest to reduce the time of searching however additionally to be greater accurate in retrieving precise photo, reading, predicting. Those techniques may be efficiently utilized by combining one among a kind set of policies and thoughts. Crucial purpose of any mining strategies isn't always fine sincerely mining it want to moreover be green and powerful. Mining method performance is measured in terms of its standard overall performance – cut price time in transforming the picture to records set .Reduction in the period of the transformed picture facts and the time to mine the majority statistics. Mining approach effectiveness is the percentage of accuracy in identifying, retrieving, searching photos.

IV.COMPARISONS

On the analyzation and computation of the both VSM and ANN, we can state that the image conversions with the use of vector identifiers and the network of data representation both are valid ways for the implementation of the space image mining. However, the implementation of the VSM is considerably easy on comparison with the neural-networks.The below figure-5 shows that the comparison between different algorithms like BM-Cilin , VSM and VSM-Cilin with respected to different datasets or different types of areas like Entertainment , Military , Sports and Political . Among all our method got better results.This research shows us that the

image transmigration to datasets of facts is not an easy task. As the advancements are preceding in the field of data mining and image mining are being progressing rapidly the new implementations are to be speculated and proposed for a definition of a new set of research to take place and go on for further improvements.

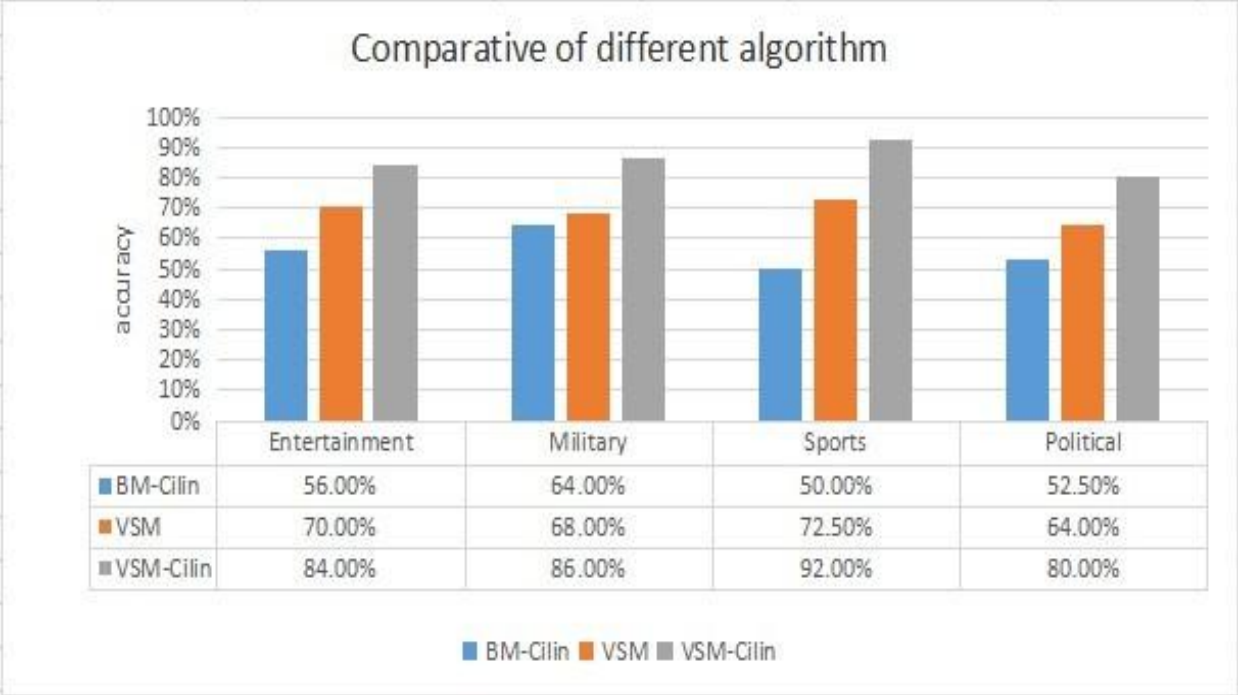

FIGURE-5: Comparison of Different Algorithms

V. CONCLUSION:

The implementation of the VSM and ANN together will allow a special compatibility, which further can be classified or optimized in the future along with the cost efficiency optimizations. The image fact extraction procedure can be most simply computed when we integrate the VSM and ANN into the space image mining for the extraction of facts and the optimum extraction of data from the image sets to form a new dataset.

REFERENCES:

A., Buyya, R. (eds) Innovations in Computer Science and Engineering. Lecture Notes in Networks and Systems, vol 171. Springer, Singapore. https://doi.org/10.1007/978-981-33-4543-0_79

Aakanksha Bhatnagar, Shweta P. Jadye, Madan Mohan Nagar" Data Mining Techniques & Distinct Applications: A Literature Review" International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 9.

Brijesh Kumar Baradwaj, Saurabh Pal "Mining Educational Data to Analyze Students Performance" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No.

Crisp-DM 1.0 Step by step Data Mining guide from http://www.crisp-dm.org/CRISPWP-0800.pdf.

Dr. Gary Parker, vol 7, Data Mining: Modules in emerging fields, CD-ROM.

Er. Rimmy Chuchra "Use of Data Mining Techniques for the Evaluation of Student Performance:A Case Study" International Journal of Computer Science and Management Research Vol 1 Issue 3 Oct.

Han J and Kamber M. "Data Mining, Concepts and Techniques", Morgan Kaufmann.

Kumar, A. (2012). ARTIFICIAL NEURAL NETWORKS FOR DATA MINING.

Rao, M.V.V., Chaparala, A. A Novel Feature-Based SHM Assessment and Predication Approach for Robust Evaluation of Damage Data Diagnosis Systems. Wireless Pers Commun (2022). https://doi.org/10.1007/s11277-022-09518-z

Sirisati, R.S., Vishnu Vardhana Rao, M., Dilli Babu, S., Narayana, M.V. (2021). An Energy-Efficient PSO-Based Cloud Scheduling Strategy. In: Saini, H.S., Sayal, R., Govardhan,

V. V. R. M AND A. CHAPARALA, "data mining technique for structural strength monitoring system methodologies," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-6, doi: 10.1109/ICCCI50826.2021.9402640.

Vishnu Vardhana Rao, M., Chaparala, A. (2021). Structural Strength Monitoring System Practices Using Machine Learning. In: Singh Mer, K.K., Semwal, V.B., Bijalwan, V., Crespo, R.G. (eds) Proceedings of Integrated Intelligence Enable Networks and Computing. Algorithms for Intelligent Systems. Springer, Singapore. https://doi.org/10.1007/978-981-33-6307-6_26