iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** XII **Month of publication:** December 2022

DOI: https://doi.org/10.22214/ijraset.2022.47833

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue XII Dec 2022- Available at www.ijraset.com

Data Mining Application: Rainfall Predictions

Siddhartha Shukla[1], Sandeep Kumar[2], Navdeep Sharma[3], Dr. Manjot Kaur Bhatia[4]
[1, 2, 3]*Student,* [4]*Faculty, Jims Sector 5*

Abstract: ***Data mining is method or process of extracting (Implicit previously unknown and potentially useful) pattern or information from large amount of Data. It is used to extract relevance knowledge from raw data. Some data mining methods and algorithms or some organizations used this because to enhance their businesses and they found required result. In 1936 Alan Turing introduced this idea first time. And other name of Data Mining is Knowledge Discovery in Database (KDD), because of unknown and potentially important data stored in database. Rainfall is the prime input for wedding seasons or any occasion in India . It is designed for information about the rainfall season on occasion or by taking the previous 10 years data. This is very helpful for primary sector workers for crop planning .The Daily rainfall data for a period of 10 years is used to understand usual rainfall, deficit rainfall, Excess rainfall and Seasonal rainfall This analysis will provide useful facts for water resources planner and formula available is used to evaluate return period of monthly , seasonally and annual rainfall.***
Keywords: Data mining, Analysis, algorithm, extract, KDD, potentially database.

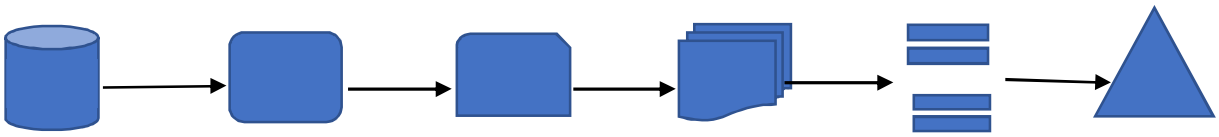
I. INTRODUCTION

Data mining refers as the process of discovering meaningful into new collections, patterns and trends by digging into large amount of data stored in warehouses using Artificial Intelligence (AI), statistical and Mathematical technique. The process of extracting information to identify patterns trends and useful data that would allow the business to take the data driven decision from huge amount of data is known as data mining.

It can be defined as the process of identifying meaningful new collection, patterns and trends by digging into large amount of data stored in ware houses using artificial intelligence (AI), statistical and mathematical technique. Information Technology has generated very huge number of databases and very large amount of data in various areas. The research in these technologies has given an approach to manipulate precious data for another use. And another name is this technique is knowledge discovery process.

And it is the logical process that is used to find through very big amount of data in order to find useful data.

The term data mining is not new. It is a synonymous with data dredging or fishing. This article refers to the history of KDD and data mining and their related tools. KDD is a real-world application; it provides general multistep process.



[2] fig:1 knowledge discovery process (author,year)

Knowledge discovery from data (KDD) refers to the overall process of discovering useful knowledge from data and it involves multiple steps.

II. DISCUSSION

A. [9]Need's of Data-Mining/Kdd

As today we know the growth of data is moving from Terabytes to petabytes. Means moving towards the big data, hence storing of data and getting some useful information from that process of data mining.

The tradational form of all data convert into knowledge relies on manual analysis for example, in health care industry, the specialists provide a report detailing the analysis to provide health care organization.

With the help of this technique, we can know the customer, and helps to our business growth.

And it also helps the develop smart market.

Evolution of Data Mining:

1) 1960's- Data Collection, Data Base Creation.
2) 1970's- Relational DBMS, Implementation.
3) 1980's – RDBMS, Advance Data Model, (Extended relational, Deductive, etc.) Application – Oriented DBMS (spatial, scientific engineering).
4) 1990's – Data Mining, Data Warehousing, Multimedia Database and Web Database.
5) 2000's – Stream, Data Management and Mining and its applications, wen technology (XML, Data and Integration) and Global Information System.

B. [1]KDD Process

1) KDD stands for Knowledge Discovery from Data. KDD refers to the overall process of discovering knowledge from Data.
2) It involves multiple steps.
a) Data Cleaning: It is Defined as removal of noisy and irrelevant data from collection.

- Cleaning missing values
- Cleaning Noisy data

b) Data Integration: It is defined as the heterogenous data from multiple sources combined into a common source (Data Warehouse)

- Multiple tools are used here to do the process.

c) Data Selection: It is defined as the process where Data relevant to the analysis is decided and retrieved from the warehouse.

- Neural Network or Decision trees or Noisy Bugs etc. are used.

d) Data Transformation: It is defined as the process of transforming Data into the appropriate form required by the mining procedure.

e) Data Mining: It is the process applied to extract patterns, potentially useful.

- It transforms task relevant data into patterns.

f) Pattern Evolution: Pattern Evolution is defined as identifying stats, increasing patterns, representing knowledge based on given measures.

- Both Summarization and Visualization is Done.

g) Knowledge Representation: Using Visualization tools to represent results.

- Reports, Tables and to generate Results, etc.
- KDD is an iterative process.

C. [4]Patters That Can be Mined Using Data Mining

1) Classification

- Here we separate the Data based on their properties.
- Classification is done into multiple modules or classes which makes their access easier and faster.
- Classification is done on the basis of data label we classify these data or if in values.
- Example – Role Based Classification, Pattern Based Classification, Decision Trees are Used.

2) Clustering

- It is the Division of the information into groups of the connected group.
- Previous in classification, we have labels for our data, but here it's an unsupervised learning. (Labels are Unknown)

- Based on their common properties the data are sorted into multiple groups and each group is a cluster.
- Applications of clustering are in – Text Mining, CRM, Web Analysis, Computational Biology.

3) Regression
- Regression is the Data Mining process is used to identify and analyze the relationship between various variables
- Example – We Might use it to project certain cost depending on other factors which are available.

4) Association
- his is the process in which we discover the link between two or more items.

5) Prediction
- It is nothing but analyzing the past event or instance data in the right sequence to predict future event.

6) Outer Analysis:
- A Database may contain Data Objects that do not compile with general behaviour or mode of the data.
- Then the Data objects are outlying.

D. [8]Technologies Used in Data Mining

1) Statistics: It is used for the Mathematical analysis for expressing representation and summarization of empirical data or real-world observation.
- Statistical Analysis involves the collection of methods, applicable to large amount of the data to conclude and report the trend.

2) Machine Learning: By using this technology the computer system adapts the ability to learn without being programmed.
- For Using this technique an Algorithm could be constructed, designed and developed.

3) Algorithm: An Algorithm can be Written for any application in Data Mining.
4) Applications: Applications can be Built and used as a technology using Data Mining.
5) Pattern Recognition: By Observing various patterns Data Mining can be performed.
6) Visualization: For Depth Analysis of Data, By Plotting in 2D Graph we do Visualization.
7) Database Technology: Using Various Database technology or language like SQL we retrieve Data efficiency.

Application Built Using Data Mining Technology:
a) Web Analysis- For Web Page classification, Page recognition.
b) Basket Data Analysis to target marketing.
c) Collaborative analysis.
d) Biological and medical Data Analysis, classification, clustering, analysis and Biological Sequence Analysis.

E. [3]Applications of Data Mining

Data mining is commonly word now a days, it is used for growing to our business. And it mainly focuses of retailing, and marketing the organization. Here's the some of the Important applications of data-mining-

A. Higher Education

Data Mining is also applied in the Field of the Education Sectors as There are a number of students in the country and their data to be processed and to derive the important details and knowledge of the student's interest and the course outcomes along with the study environment and the course curriculum. It is also known as Educational Data Mining (EDM), It basically deals with development of the methods for Discovering knowledge.

B. CRM

CRM stand for customer relationship management. It focused on the customer's strategies and loyalty and maintain a relationship.

C. Fraud Detection

A fraud detection system should protect the data of all the users. a method includes collection of virtual records. And these records are distributed fraudulent or non-fraudulent.

D. Manufacturing Engineering

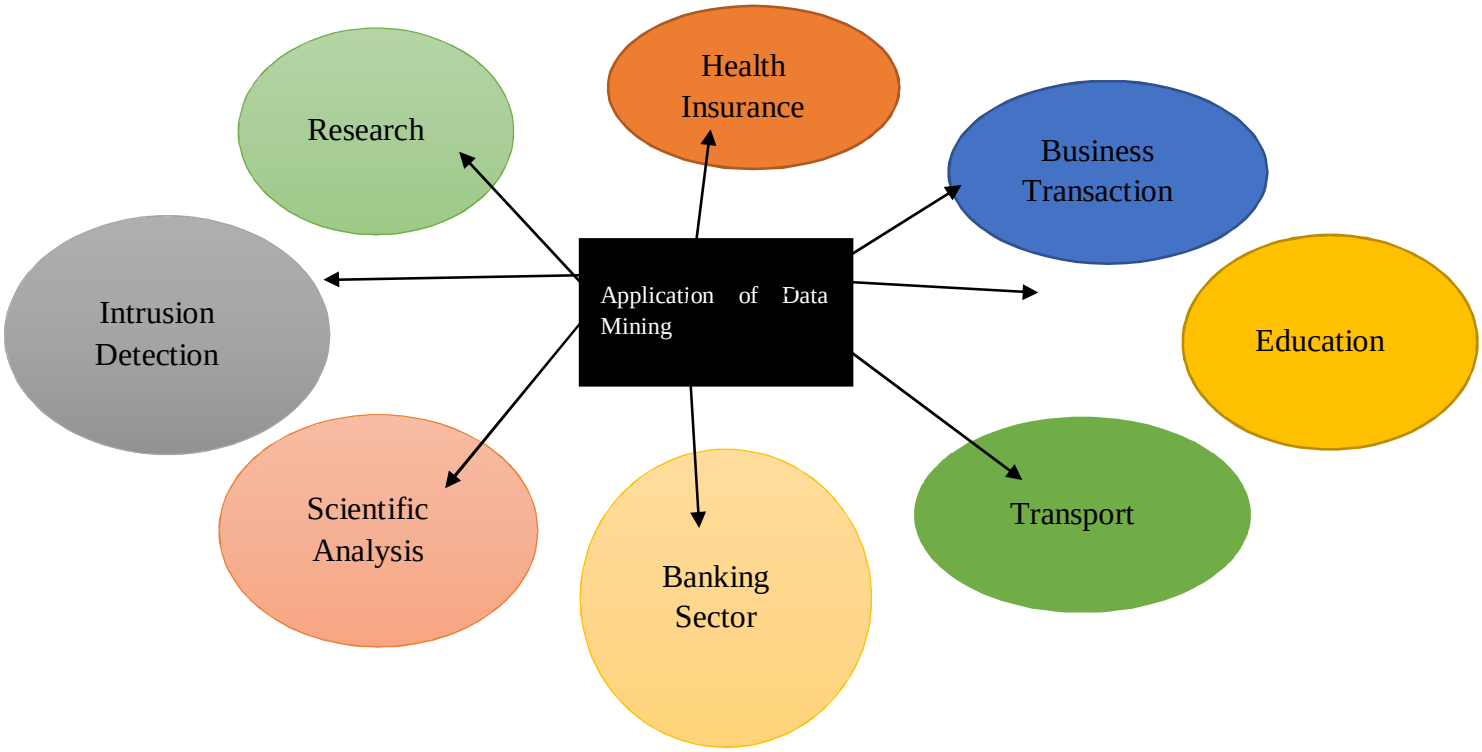
Data mining tools can be very useful to search new designs in multiple manufacturing process.

E. Market Based Analysis

This technique helps retailer to understand the customer needs and satisfaction

F. Banking Sector

To Analyze the Data of the Customer of the Bank and the Different Employees on a Huge Scale



[5]Fig:2 Application of data mining

III. LITERATURE REVIEW

A. [6]Rainfall

Now a days to predict rainfall is one of the challenging tasks in weather forecasting. Exact rainfall prediction will be beneficial to take effective measures earlier for e.g. In Aviation sector, multiple tasks in primary and secondary sector ,Diseaster management . Using data mining we can effectly predict.

Multiple natural phenomena by extracting data of last multiple natural phenomena by extracting data of last multiple years and analyzing that.It is often said that specially for India economy is gamble with Indian monsoon.

Depending on this everything is gamble with Indian Monsoon i.e. primary,secondary and tertiary sector.

But in India rain depends on major factor affecting reasons e.g. Topography,jet streams , shifting of ITCZ ,Elnino,Lanino,Enso effect etc.

So using data mining technique we can extract data from past trends and apply these factor affecting reason to predict future trends on rain.

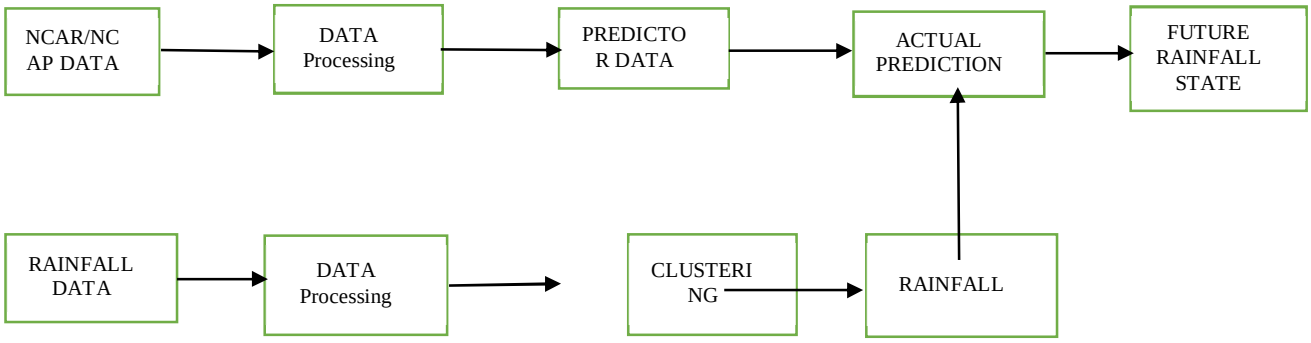
This rain can vary depending upon demographic condition.

B. [10]Conclusion On Rainfall

For existence of human being rainfall touches all aspects of life e.g. economy,polity, culture, Health, Education, Defence.
To be more precise all these sectors requires modern technological support.
Data mining is one of such technique to enhance predictions using data study and extract figures from that data. Rain and snow both are forms of precipitation and both need correct predictions. Siachin glacier which is highest peak for International border where Indian troops deployed everyyear and most of the soldiers suffer there and that oftenly causes causality also because temperature is there (-70) centigrade almost.
Using data mining technique it will be feasible to supply goods and resources which is essential for soldiers.

IV. CONCLUSION

1) Data Mining Technology Gets Consolidate with Internet of Things (IOT) for Decision Making support and Pattern Recognition.
2) Data Mining Presumes for the innovation of interesting and potentially useful pattern, form data and applying algorithms.
3) Data mining is a Knowledge Discovery from Data (KDD) process.



[7] System Architecture

V. RECOMMENDATIONS

This Thesis can be further used for the further researches using Data Mining like in field of the Healthcare and Education for determining the patterns and the trends of the world at a given time period.

VI. ACKNOWELDGEMENTS

We would Like to Thanks our Mentor and Respected Teacher Mrs. Manjot Kaur Bhatiya and his guidance in the preparation of this Paper and the JIMS College , Rohini Sec-5 ,which provide us this golden opportunity.

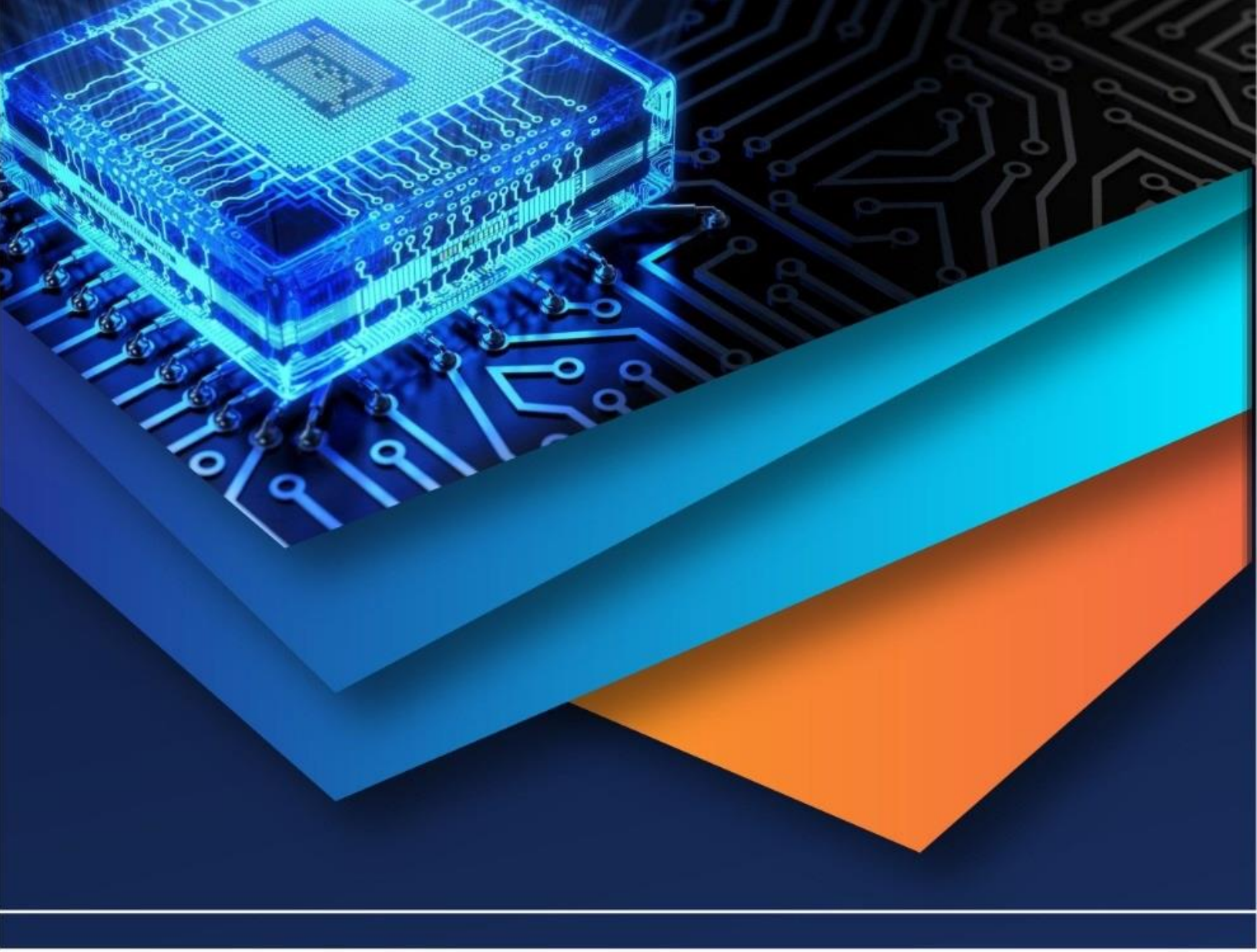
VII. AUTHOR'S CONTRIBUTION

The Study was designed, directed and coordinated by Sandeep Kumar, Siddhartha Shukla and Navdeep Sharma. The idea was thought of the Data Mining, its methodology and the Applications About Rainfall prediction, firstly we Understood the subject by getting depth in the subject further, idea and applications were researched related to the mining of the Data and how the functionalities of the Data is done and finally thought of the further researches and the technology that could be also combined with It for further Innovations.

REFRENCES

[1] Jiawei Han and Micheline Kamber (2006), Data Mining Concepts and Techniques, published by Morgan Kauffman, 2nd ed.
[2] http://www.kdnuggets.com/
[3] Crisp-DM 1.0 Step by step Data Mining guide from http://www.crisp-dm.org/CRISPWP0800.pdf.
[4] S.Kannan,Subimal Ghosh, Prediction of daily rainfall state in a river basin using statistical downscaling from GCM output. Stochastic environment research and risk assessment .May 2011,Volume 25,Issue 4,pp 457-474.
[5] Najmeh Khalili, Saeed Reza khodashenas et al. "Daily rainfall Forecasting for Mashhad Synoptic station using Artificial Neural Networks" 2011 International conference on Environmental and Computer Science IPCBEE vol.19(2011) (2011) IACSIT press,Singapore.
[6] M.Kannan, S.Prabhakaran,P.Ramachandran "Rainfall Forecasting Using Data Mining Technique" International Journal of Engineering and Technology Vol.2(6),2010,pp.397-401.

[7] The Science and Information organization " Rainfall prediction,data mining techniques,systematic literature review.
[8] Department of informatics engineering Faculty of Computer Science. Informatics and Business Institute Darmajaya "Rainfall prediction using data mining techniques"
[9] Journal of Physics "The Weather Forecast using Data mining Research Based on Computing".
[10] Ijraset journal for research in applied science and engineering technology.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY