

# AN INNOVATION DEVELOPMENT OF DATA TRACKING AND SECONDARY DATA ANALYSIS BY USING AUTOMATED MACHINE LEARNING MODEL

**V. Indu Nair and J. Gowri**

*Department of Bio-medical Engineering, Park College of Engineering and Technology, India*

## Abstract

*The data analysis refers to the science of analyzing or tracking source data to draw conclusions about a given piece of information. Most of the relevant processes and techniques of data analysis are currently automated in the respective machine mechanisms and processes that are known to work on a given source data series used by the consumer. In this paper, a automated machine learning model was proposed to track and analyze the data. The Data Analytics techniques related to data analysis help to reveal measurements and trends that may be lost in the given information. The information given is used to improve processes in the guise of increasing the overall efficiency of the organization or business. Data analysis is a broad term that encompasses many types of data analysis and their respective techniques. Any specific type of information can be used with the help of data analysis techniques to gain insights into processes and business development.*

## Keywords:

*Data Analysis, Automated, Machine Mechanisms, Machine Learning Model, Encompasses*

## 1. INTRODUCTION

The Companies there are known to record idle time, work order and uptime for multiple machines when analyzing data to better plan a given workload. This helps business owners to ensure that the machines work efficiently to achieve their respective peak capabilities. Data Analytics is capable of doing more than just pointing out obstacles in a given production process. For example, gambling companies are known to use data analytics to set up reward tables for their respective players. This technique helps to keep most players very active in a given game [4].

At the same time, content companies are known to use advanced data analytics to see users click, or rearrange existing content to get another click or another view [5]. The use of machine learning technology is playing an important role not only in self-propelled vehicles but also in various fields including medical and cyber-crime prevention [6]. Machine learning methods can be useful not only for learning that technology but also for solving people's problems based on it. Machine learning can be basically divided into different categories and the algorithms can be categorized accordingly [9].

The first type is the creation of algorithms for machine learning with analyzed data (Supervised Machine Learning). In this way, the data to be entered into the machine will be categorized under specific headings or sections [11]. By examining such data, the machine can develop its own decision-making mechanisms. For example, if the machine wants to detect the difference between Rs.1, Rs.5 and Rs.10 coins, it can create algorithms based on the weight of the coins. Through it, the machine will automatically detect the amount of coin input by the

weight of the coin. This includes data divided by the weight of the currency [12].

We have already provided such instructions to the machine based on that the machine makes its own decision. If the data thus entered is properly sorted and injected into the machine, the machine will easily deliver the result [13]. But there are some issues in this category. The machine will give the result only if the parsed data is input. An environment in which that machine cannot function can be created by keeping undivided data. Also, systematically splitting all the data and inputting it into the machine can be time consuming and costly.

Another type of application comes from it to get rid of such flaws. That is, creating algorithms for machine learning with unclassified data (supervised machine learning). In this way the data to be entered into the machine will not be parsed properly. The algorithms are built into the machine to function based on the data. In this way the key component is to systematically analyze the input data and race them rather than the machine having to confirm a particular result.

For example, suppose a cricket match scoreboard is fed into a machine. The data in it will be completely unclassified. The machine has to sort them automatically based on the nature of the data. Certain players on the scoreboard will score more runs and certain players will take more wickets. Based on this data, you can find out which player is the batsman, who is the bowler and which player is the all-rounder. The machine can automatically detect them and develop algorithms for data sorting. This type can also be used by the machine to automatically sort unwanted and unrelated emails into the "spam" section. Apart from these, we can also classify the mechanisms for machine self-learning (reinforcement machine learning) through mistake.

The machine learns it, so it can create algorithms for the machine to learn itself through mistakes, in which the machine makes mistakes. The inputs will be re-issued and errors will be corrected. Algorithms for machine learning can be developed in the above categories [14]. Nowadays artificial intelligence and associated machine learning are playing a vital role in taking technology to the next level.

## 2. RELATED WORKS

In many researchers collect new data for analytical purposes, but many rely on secondary data collected by someone else to conduct a new study. When a study uses secondary data, they are called secondary analysis for performance analysis [1]. The vast majority of secondary data resources and datasets are available for sociological research, many of which are public and easily accessible [2]. There are both positive and compassionate ways to use secondary data and to conduct secondary data analysis, but most can be minimized by learning about the methods used to collect and clean up in the first place. That and honest statement

[3]. There are some disadvantages to the second grade. In some cases, it may be out of date, unrelated or irregular. But a trained sociologist can diagnose, work, or correct such problems. To conduct a meaningful secondary analysis, researchers must spend significant time reading and learning about the origin of the data set [4].

Data analysis plays an important role in the scientific process. Data analysis is required to understand the nature of the problem taken up for study and to find appropriate solutions accordingly. It is important not only in academic studies but in all kinds of studies. Analysts can not only gather data relevant to their research problem research objectives, but also obtain meaningful information and conclusions related to the research problem while analyzing them appropriately [5]. In studies, not only quantitative and qualitative studies, but also mixed methods of mixing methods have been on the rise in recent times, especially in the field of education. It is important for those involved in research in the field of education to be familiar with data analysis methods. Therefore, it is very important to choose the appropriate quantitative data analysis techniques, especially about data analysis [6].

Quantitative analysis can summarize the past and present trends of numerical data and their general characteristics. For this, an explanation of descriptive statistics will suffice for the researcher. Here data is calculated, calibrated, described, indexed, sorted, etc [7]. In this, it is sufficient for the researcher to know how to use the ratio, radioactive forcing, centrifugal measurements, deviation measurements coefficient, etc. for analysis [8].

Therefore, it is important for the researcher to know the basics of statistics / statistics. Statistics can be a challenge for some / many of you. However, knowing the statistics is a step in the data analysis process. In this way, the researcher will be able to better practice the scientific methodological analysis of trends, behavioral patterns, characteristics, and predictions related to numerical data and to collect, submit, analyze, and interpret those numerical data [9]. However, although the statistical techniques are not fully known, it is important to be aware of the nuances involved in selecting the appropriate analytical techniques for the data collected in the study. With that in mind, it would be possible to carry out quantitative analysis with the help of emerging technologies [10].

### 3. PROPOSED MODEL

Unlike primary data collected by a researcher to accomplish a specific research goal, secondary data is data collected by other research researchers. Sometimes researchers or researchers share their data with other researchers. In addition, many government agencies in the United States and around the world collect data available for their secondary analysis. In many cases, this data is available to the general public, but in some cases, it is only available to authorized users.

The Secondary data will be in size and quality in form. Secondary measurement information is often available from official government sources and trusted research organizations. In addition, there are many analysts, including collection and judicial statistic. This information is collected for a variety of purposes, including budgeting, policy planning, and city

planning, and can be used as a tool for sociological research, among others. By analyzing and analyzing numerical data, social psychologists can often detect unnoticed types of human behavior and large-scale trends in society. Secondary standard information is commonly found in social art forms, including newspapers, blogs, diaries, letters, and emails. Such information provides contextual and detailed information for an excellent information and sociological analysis of individuals in the community. Some important steps in the process of utilizing data analysis are shown in Fig.1.

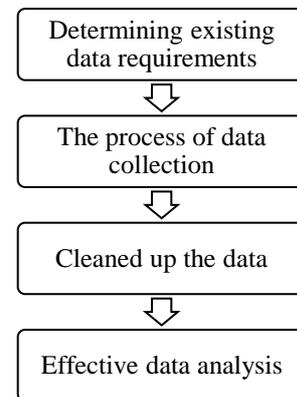


Fig.1. Proposed Data Analysis

- Determining existing data requirements or how grouping of data is done. The data can be divided into basic population, age, gender, income, and more.
- The next stage involves the process of data collection. It can be implemented in many ways or from various sources such as online sources, computers, cameras and others.
- In data collection, data is expected to be organized for effective data analysis. Organizing data with the help of a proper spreadsheet or some other software solution is known as providing statistical data and access to its structure.
- Before analyzing, existing data is cleaned up. This ensures there is no error or duplication. Before the data is sent to the data analysts for in-depth analysis it helps to eliminate any error as given.

Data analysis is important because it helps businesses improve their respective performance. In addition, businesses can expect to reduce overall costs by finding efficient ways to implement business operations. Secondary analysis is the procedure for applying secondary data. Research systematically saves both time and money and avoids unnecessary copies of research effort. The secondary analysis usually differs from the primary analysis, which is the analysis of the primary data collected by the researcher.

Secondary data represent a vast resource for sociologists. It is easy and free to use often. It can add information about very large people that can be expensive and very difficult. Also, secondary information is available from periods other than today. It is not easy to conduct primary research on events, attitudes, practices or ethics that do not exist in today's world. After careful reading and observation, researchers can determine shown in Fig.2:

- The purpose for which the material was collected or created
- Specific methods used to collect it

- Validation of population survey and sample captured
- Collector or creator reliability and trustworthiness
- Data package limitations (no information requested, collected or provided)
- Historical and/or political circumstances surrounding material creation or collection

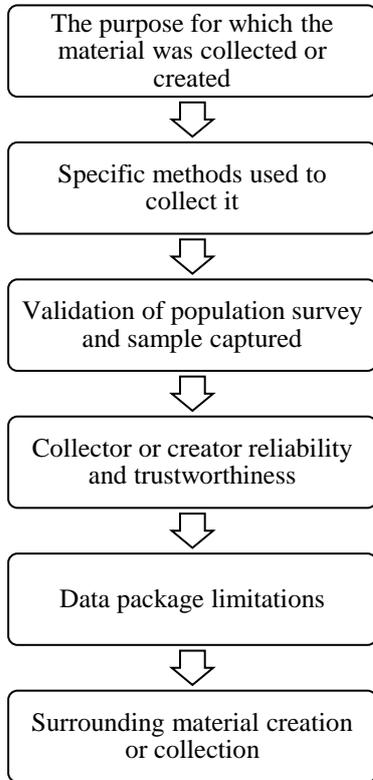


Fig.2. Proposed Model

In addition, before using secondary data, a researcher should consider how the data is encoded or categorized, and how this may affect the results of secondary data analysis. She should consider whether the data has been adapted or corrected before conducting her own analysis. Quality data is generated in known situations by individuals named for a specific purpose. It is very easy to analyze data by understanding data, spaces, social context and other issues. There are policies that help protect everyone.

Often, we think of malicious scenes and hope they never happen to us. Unfortunately, it worked looking at the companies where they took place; A bonus is a simple local filter on a dashboard for active accounts against all accounts at risk. A team that accesses data governed by administrative policy, but elevates the cloud to the database for self-service use outside the control of the ID. The risks associated with analytical management policy include which is shown in Fig.3.

- **Incorrect Results:** Incorrect analysis results or unreliable results
- **No Results:** Stuck in analysis
- **Wasted Cost:** Teams lose time making their own tools with their own tools
- **Brand Equity Loss:** Slow Market Responses, Bad Choices or Data Leaks Go General

Having open conversations around these topics can be difficult but bridging the gap between IT and business lines is essential for a successful and positive culture.

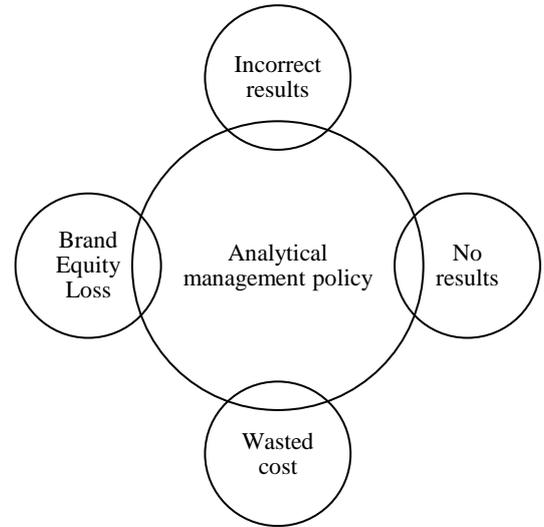


Fig.3. Analytical management policy

However, quantitative data may still require critical analysis. It is not clear when data will be collected, if some types of data are collected and others are not, or of no use in the development of tools used to collect data. Polls, questionnaires and interviews are all designed to have pre-determined outcomes. While pro data is very useful, it is very important for the researcher to be aware of pro, its purpose and its magnitude. The data analysis programs allow analysts to organize, cleanse, and issue pre-planned commands for their data, allowing analysts to use the most advanced forms of statistical analysis from the most basic. It will be useful for you to understand your data and what you would like to use when sharing it with others. There are many plans on the market, but unfortunately, they are too expensive to buy.

#### 4. RESULTS AND DISCUSSION

The proposed automated machine learning model (AMLM) was compared with the existing Scalable clustering algorithms (SCA), black-box medical algorithms (BMA), distributed learning algorithms (DLA) and Scalable machine-learning algorithms (SMLA)

##### 4.1 SYSTEMATIC DATA ANALYSIS

It is the process of systematically applying statistical and or logical techniques to describe, interpret, summarize, review and evaluate collected data. This allows the researcher to draw meaningful conclusions from the data collected. Data analysis is the process of preparing the collected data for analysis, performing the analysis, reporting the results, and interpreting them. There must be management models to ensure that the data and the resulting analyzes are accurate, concise, reliable and secure! Most companies believe this includes implementing a data management policy.

Table.1. Systematic Data analysis

Inputs	SCA	BMA	DLA	SMLA	AMLM
250	65.11	68.37	75.09	87.63	94.24
500	64.40	67.44	73.98	86.30	93.04
750	63.10	66.44	73.28	85.22	92.88
1000	62.19	65.49	72.31	83.97	92.03
1250	61.19	64.52	71.40	82.77	91.35
1500	60.18	63.56	70.50	81.56	90.67
1750	59.18	62.59	69.59	80.36	89.99

## 4.2 QUANTITATIVE DATA ANALYSIS

Such data is called quantitative data when it is found in terms of number or number of data blocks collected for analysis. Such data should be analyzed using mathematical or statistical techniques. Conversely, the researcher will need knowledge of inference (hypothetical) statistical techniques when the researcher attempts to analyze the data obtained from the sample and make a decision about the drink of the sample based on its results. Deciding on such hypothetical statistical techniques can be difficult for many researchers.

Table.2. Quantitative Data Analysis

Inputs	SCA	BMA	DLA	SMLA	AMLM
250	68.97	72.37	79.01	92.16	95.94
500	68.64	70.87	78.42	90.29	94.93
750	67.30	69.76	77.44	89.46	94.77
1000	66.16	69.38	76.23	88.55	93.81
1250	65.33	68.08	75.45	87.20	93.23
1500	64.35	67.07	74.51	86.03	92.57
1750	63.37	66.06	73.58	84.87	91.92

## 4.3 DATA MANAGEMENT

A data management policy systematically outlines how data processing and management should be done to ensure that data is accurate, accessible, stable and secure. This policy establishes who is responsible for the information in the various situations and specifies what procedures should be used to manage it. How data is managed and how it goes to the tool is managed, but once in a while the tool is in the dark and open as if it were self-service or getting the job done.

Table.3. Data Management

Inputs	SCA	BMA	DLA	SMLA	AMLM
250	76.24	75.15	66.68	77.24	87.44
500	76.08	73.95	65.06	77.37	87.47
750	75.34	72.30	63.26	76.10	87.47
1000	74.99	70.95	61.58	75.76	87.49
1250	74.54	69.53	59.87	75.19	87.51
1500	74.09	68.10	58.16	74.62	87.52

1750	73.64	66.68	56.45	74.05	87.54
------	-------	-------	-------	-------	-------

## 4.4 ANALYSIS MANAGEMENT

Analytical management policy systematically outlines what processing, modification, and editing is allowed beyond the data layer to ensure accurate, accessible, consistent, reproducible, secure and reliable results. Having an analytical management policy helps to avoid conflicting results when using multiple tools or unique authors. In the right world we are all aligned to the dashboard with input and confidence. An Analytics Governance Policy Alignments in the Dashboard can only be made by specific individuals.

Table.4. Analysis Management

Inputs	SCA	BMA	DLA	SMLA	AMLM
250	73.27	81.89	57.98	86.66	87.18
500	71.61	76.03	64.82	81.25	87.28
750	71.16	77.17	66.11	79.76	87.35
1000	75.74	76.03	68.25	76.52	87.40
1250	74.69	73.67	72.32	73.07	87.49
1500	75.38	72.03	75.53	69.88	87.56
1750	76.08	70.38	78.74	66.69	87.63

## 5. CONCLUSION

In general, data collection is a collection process whose purpose is to obtain information to establish results on a specific topic. This function can be used in any field; In social sciences, business, natural sciences, etc. Data collection is a function of collecting or storing important information on a specific topic; generally, this activity aims to provide the knowledge necessary for the development of a job or research. Data collection is carried out in different ways: it can be through surveys and interviews, by observing events, or by bibliographic advice (i.e., by reviewing books and materials in which data are recorded).

Moreover, this activity is not limited to researchers and scientists; It is also widely used in schools and educational institutions. This happens-for example- when teachers collect data on their students with the aim of getting to know a group of students' academic performance. This function is very important in any investigation because it brings truth to the work. That is, data collection is essential for people to take research seriously. This happens because the data allow the analyst to establish objective questions and answers for adequate guidance during their work. It is worth highlighting that data can be defined as quantity or quality; In the first case, it is information expressed in numeric characters. In contrast, standard data are properties expressed in alphabetical order.

## REFERENCES

- [1] H. Harris, "Content Analysis of Secondary Data: A Study of Courage in Managerial Decision Making", *Journal of Business Ethics*, Vol. 34, No. 3, pp. 191-208, 2001.

- [2] J.F. Chizmar and M.S. Walber, "Web-Based Learning Environments Guided by Principles of Good Teaching Practice", *Journal of Economic Education*, Vol. 3, No. 2, pp. 248-264, 1999.
- [3] Institute for Higher Education Policy, "Quality on the line: benchmarks for success in Internet-base distance education. Washington, DC: Institute for Higher Education Accreditation", Available at <http://www.ihep.org/Pubs/PDF/Quality.pdf>, Accessed at 1999.
- [4] L.A. Petrides, "Web-Based Technologies for Distributed (or Distance) Learning: Creating Learner-Centered Educational Experiences in the Higher Education Classroom", *International Journal of Instructional Media*, Vol. 29, No. 1, pp. 69-77, 2002.
- [5] V.A. Ayma, G. Costa and P. Gamba, "Classification Algorithms for Big Data Analysis, A Map Reduce Approach", *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 40, No. 3, pp. 1-17, 2015.
- [6] S. Athmaja, M. Hanumanthappa and V. Kavitha, "A Survey of Machine Learning Algorithms for Big Data Analytics", *Proceedings of International Conference on Innovations in Information, Embedded and Communication Systems*, pp. 1-4, 2017.
- [7] V. Cevher, S. Becker and M. Schmidt, "Convex Optimization for Big Data: Scalable, Randomized, and Parallel Algorithms for Big Data Analytics", *IEEE Signal Processing Magazine*, Vol. 31, No. 5, pp. 32-43, 2014.
- [8] G. Carra and G. Meyfroidt, "Data-Driven ICU Management: using Big Data and Algorithms to Improve Outcomes", *Journal of Critical Care*, Vol. 60, pp. 300-304, 2020.
- [9] S.K. Punia, M. Kumar and R. Patan, "Performance Analysis of Machine Learning Algorithms for Big Data Classification: ML and AI-based Algorithms for Big Data Analysis", *International Journal of E-Health and Medical Communications*, Vol. 12, No. 4, pp. 60-75, 2021.
- [10] Sarfraz Fayaz Khan, "Health Care Monitoring System in Internet of Things (IoT) by Using RFID", *Proceedings of International Conference on Industrial Technology and Management*, pp. 198-204, 2017.
- [11] Prosanta Gope and Tzonelih Hwang, "BSN-Care: A Secure IoT-Based Modern Healthcare System using Body Sensor Network", *IEEE Sensors Journal*, Vol. 16, No. 5, pp. 1368-1376, 2016.
- [12] Byung Mun Lee, "Registration Protocol for Health IoT Platform to Share the Use of Medical Devices", *International Journal of Bio-Science and Bio-Technology*, Vol. 7, No. 4, pp. 1-10, 2015.
- [13] A. Nayyar and V. Puri, "Comprehensive Analysis and Performance Comparison of Clustering Algorithms for Big Data", *Review of Computer Engineering Research*, Vol. 4, No. 2, pp. 54-80, 2017.
- [14] J.Y. Lee and L.M. Ryan, "Sufficiency Revisited: Rethinking Statistical Algorithms in the Big Data Era", *The American Statistician*, Vol. 71, No. 3, pp. 202-208, 2017.