Data Science Techniques, Tools and Predictions

Mujthaba G.M , Abdalla Al Ameen, Manjur Kolhar, Mohammed Rahmath

Abstract: Almighty created human being with numerous wants and needs which makes them associated with their own data, choices and preferences. To grow and develop any business or organizations it is very obligatory to know their clients requests or customer needs based on their data. The evolving role of data makes it very vital element in any organization and carried with convinced operations. In this paper we are going to present a study of Data Science and its relevance with Artificial Intelligence, machine learning and deep learning. The incorporation of these intellectual sciences in data science is useful for perming numerous operations in our research we tried to demonstrate the data science operations like data cleaning, data processing, data modeling, data visualization and data presentations techniques. To grow any business it is mandatory to know their customer needs and satisfy their future expectations by smart decision makings. The intellectual algorithms or data operations in the data science make the data to be more effective in decision making and decision polices. We also focus on how data science incorporates mathematical & statistical methods, logical reasoning with applications of Artificial Intelligence techniques. We also focus on various data operations tools which exists in the market like python, SAS, R and many others. At last we focusses on how data science field going to meet the future expectations of many businesses. This research paper may become as successful reference for the people to carry out their research and meet the expectations of data science field with business growing decisions.

Keywords : Artificial Intelligence (A.I.), Machine Learning (M.L.),Internet of Things (I.OT's) Data Science, Data Analysis, Data Processing, Data Presentations and Data Science Careers

I. INTRODUCTION

A. Artificial Intelligence and its relevance with Data Science:

Artificial Intelligence articulates about how to make the system as intelligent like a human being. Designing intelligent system is conceivable by incorporate the computers with learning, processing and decision making ability [1]. All these abilities deal with vast knowledge which helps the system to train with intelligent behavior. A.I speaks about numerous approaches of learning, understanding and processing techniques which can be applied on various problems or domains. The most popular A.I techniques are Heuristics, Support Vector Machines, Artificial Neural Networks, and Markov Decision Process [1]. Artificial Intelligence is well known for its applications like natural language processing, data retrieval by using intelligent systems, expert systems for various domains, theorem proving & game playing, Scheduling and combinatorial problems , robotics and so on[2]. Know question rises how the A.I is related to data science, as almost all humans' beings uses the data for their wide variety of applications in day to day life. These data will be gathered by the various businesses or sectors to figure out how can develop. In response these data science will plays as noticeable role from gathering to visualize data.

B. Data and its operations

Data is the basic component in transformation of any individual, organizations and businesses towards development in the future era [9]. Technology plays an emerging role in transmuting data into usefulness in all disciplines of the society [9]. The primary objective is to make the data usefulness by applying with statistical and logical techniques. These techniques define the scope, describe, process, modularize, exemplify and evaluate the data. Before learning into the depth like tools, operations, process, methodologies, algorithms and techniques to operate the data, it is very much required to do complete and through analysis of data. The types of data we have available with any individual or organization like text, numerical, pictorial, images, audio, video and sensitize data[8]. These data need to carry out with certain operations by which it can be transformed to usefulness or profitable to the society. Before operate on data be ensure that all these operations must not violate any social, professional and ethical values of the society or any law. As stated in below

* Correspondence Author

Mr.Mujthaba Gulam Muqeeth* Computer Science Department, College of Arts and Science, Prince Sattam Bin Abdul-Aziz University,Wadi Al Dawassir, Riyadh, K.S.A. Email: mujthaba.gmm@gmail.com

Dr.Manjur Kolhar Computer Science Department, College of Arts and Science, Prince Sattam Bin Abdul-Aziz University, Wadi Al Dawassir, Riyadh, K.S.A. Email: manjur.kolhar@gmail.com

Dr.Aballa AlAmeen, Computer Science Department, College of Arts and Science, Prince Sattam Bin Abdul-Aziz University, Wadi Al Dawassir, Riyadh, K.S.A. Email: abda711su@gmail.com

Dr.Mohammed Rahmath Computer Science Department, College of Arts and Science, Prince Sattam Bin Abdul-Aziz University, Wadi Al Dawassir, Riyadh, K.S.A. Email: mohammed480226@gmail.com

Retrieval Number: F9887038620/2020©BEIESP
DOI:10.35940/ijrte.F9887.038620
Journal Website: www.ijrte.org

5661

Published By:
Blue Eyes Intelligence Engineering
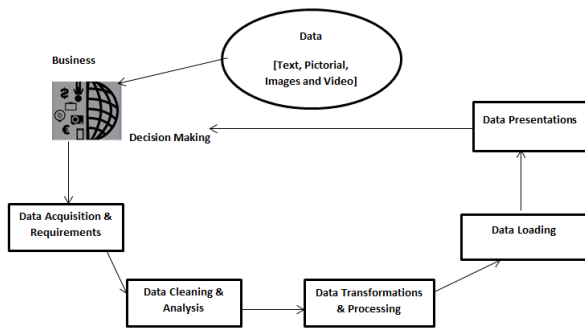*& Sciences Publication*

fig. 1 data and its operations. We need to learn the past developments of over six decades in 1950 are where Alan Turing initiates with an idea of machine computing and intelligence [3]. M.L. is considered as subset, practical approach and application of A.I based algorithms.

As the name implies machine deals with wide variety of data of various domains and design the system. This system will be able to identify the train the new set of data with the existing data samples or derive the new set of rules. Unlike algorithms to make the machine as efficient such as supervised, unsupervised and semi-supervised and reinforced algorithms [3]. There are numerous techniques proposed by M.L like game analytics, software, voice recognition, stock trading, and internet of things (I.O.T's)[3]. The data science plays an important role by providing the data in good means to have effective M.L algorithms. Machine learning techniques are used to routinely find the appreciated primary patterns inside complex data that we would otherwise brawl to determine.

C. Machine Learning relevance with Data Science

To develop the definition of Machine learning (M.L.) we need to learn the past developments of over six decades in 1950 are where Alan Turing initiates with an idea of machine computing and intelligence [3]. M.L. is considered as subset, practical approach and application of A.I based algorithms. As the name implies machine deals with wide variety of data of various domains and design the system. This system will be able to identify the train the new set of data with the existing data samples or derive the new set of rules. Unlike algorithms to make the machine as efficient such as supervised, unsupervised and semi-supervised and reinforced algorithms [3]. There are numerous techniques proposed by M.L like game analytics, software, voice recognition, stock trading, and internet of things (I.O.T's)[3]. The data science plays an important role by providing the data in good means to have effective M.L algorithms. Machine learning techniques are used to routinely find the appreciated primary patterns inside complex data that we would otherwise brawl to determine

D. Role of Data Science with Artificial Intelligence and Machine Learning

To meet the growing business needs of individuals life it is very much mandatory to make use of data in effective means is the primary concern. Another major concern is to correct the drawbacks depicted in the previous projects or mishandling of data [8]. These data can be analyzed according to its type like text, statistical, predictive and perspective Data Science consists of countless statistical practices whereas A.I relates how use of computer algorithms

in an intelligent way. AI shows how reporting sovereignty to the data model. It can be regarded as a union of traditional studies like statistics, data mining, distributed systems and databases [4][7]. Continuing research studies need to be incorporate with data science to benefit the individuals, organizational sectors, business, society & community and educations for various purposes [4]. Data Science deals with notion of to tackle big data which includes certain operations like data cleaning, training, analysis, process, modeling and so on . A data scientist collects the data from various ways and incorporated with machine learning algorithms. Data science is a subset of machine learning which develops the critical thinking, predictive analytics, domain knowledge and sentiment analysis. Data science is expected to do lot of innovations in the areas like applied computing, medical sciences, professionals & social life activities, computing paradigms, Data management systems and many more to have a better decision making[5]. Influence the new methods of improving intellectual thinking of how to use, organize, process, load, and model or visualize the data [6]. The emerging existing professions in the field of data science is the data scientist who draws an medium salary of $124,00 and stated this profession may be on the peak of in the coming years[7]. The tool selection to implement the data science activities like we have SAS, Orange, R, Python, Tableau, Tanagra, Rapid Miner, and Weka [7]. The primary operations which can be performed in the data science like cleaning raw data, loading the data at the server side, process data, visualize data and acquire data by various stake holders. Detailed explanation on techniques of data requirements, data analysis, data processing, visualize or model data are given below. We also have the look on various tools to support these operations of data science to a data scientist.

E. Significance of Data science with Artificial Intelligence and Machine Learning
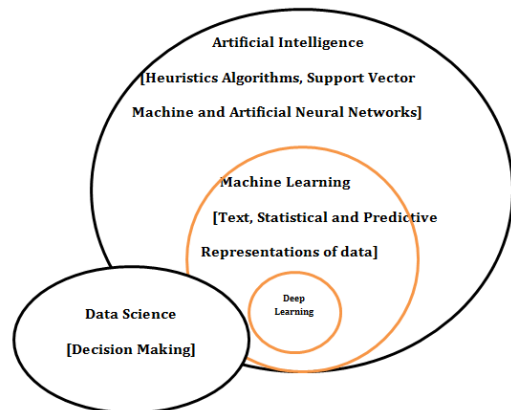

Figure 2. Significance of Data Science with A.I and M.L

As stated in the above figure 2. The Data science field will make use of A.I algorithms and machine learning in order to make the effective and useful decisions. These decisions will be based on the user choices that how they need their data presentations like statistical, pictorial, textual and any other form. These representation of data is directly proportional with data processing by using Machine learning and A.I algorithms. These algorithms applied by using statistical, analytical and mathematical approaches.

II.   LITERATURE REVIEW

The study of Artificial Intelligence is not only thinking and analyzing but also intelligent systems which can perform intelligent functions as said by Peter Norvig in et al 2016. Intelligent functions mainly thinking & acting rational also consider the performance factors like reduce cost, no replicating jobs and many more. All these intelligent rules which can draw valid conclusions to and from computer uncertain information's as said by Stuarts Russel et al (2016). A Traditional approach to artificial intelligence will connect the gap between theory and practice as said by Nilsson et al (2014). These A.I ideas underlines various applications in the areas like natural language processing, automatic processing, robotics, machine vision, automatic theorem proving and data retrieval. Alan Turing is a British mathematician and logical philosopher raises the questions why machines cannot think by its own? And Samuel discussed machine learning is the study of the ability to learn without much programming skills. The problems which can solve by machine learning are manual data entry, medical diagnosis, financial analysis and many logical operations on data sets over clouds. Many organizations could not able to make effective use of the data collected by their customers and that data is termed as big data. Various operations of processing capabilities can perform on the data sets or big d**ata** is like saving, processing, transformations, visualizations, loading at server and presentation was said by van der Aalst wil et al (2016). These big data deals with data growth, data storage, authenticate data, securing data and organizational resistance. These operations or data processing capabilities give rise to a position of data science who perform these operations. The data scientist does in collecting the data, cleaning or analysis data, process or evaluate the data, load the data at server side and model or present the data.   These operations involves the various studies like mathematics, deep learning or machine learning, artificial intelligence techniques, statistical operations, analytical reasoning, data bases and optimization techniques as said by Dhar, Vasant et al (2013).

The issues which rises in data science like separations of unrelated data, lack of experience and knowledge of particular domain, structuring the data according to user preference, selection of appropriate algorithm & its implementations and presentations of the results or output. Know a days a group of software professionals are involved in data acquisition, inspiring new ways of thinking how data can be analyzed, data organizations, evaluating data and presentation of data as highlighted by Hazen, Benjamin et al (2014). Achieving the performance in the operations of data across over internet as discussed as another important issue. To implement these operations technology rises and develops much various tools for acquire, analysis, process, load and present data. These tools issues which can found like is it well suited for big data, memory related issues in performing SQL statements, in capabilities of interactive environment, inappropriate selections of algorithms and unstructured data as said by Sumathi, S. Subhitsha1 S. Selvakumar2 et al (2017). Islam, Mohaiminul said et al (2020) to meet the business requirements or demands it is very important to ensure the effective use data of customers initiate from data acquisition to data presentation. The key approach is to ensure the data capabilities and inefficiencies with proper mechanisms. To

evaluate these operations several tools exits in the market which possess their merits and demerits. Nicolae, Bogdan & . Park, Yoonho et al (2020) explained and focuses on key issues related with data operations which give rise to data science study. These studies examine various modern technical factors like connectivity, mobile communications and social media interactions like youtube, whatsup and others. Abas, Zuraida Abal,et al (2020) focuses on 12 rising technologies which implements A.I techniques, machine learning on various Internet of Things. Logical and analytical reasoning is the way of transforming the knowledge into valuable decision makings. There are various types of analytics like knowledge or descriptive analytics, interpret or predict analysis and perceptive analysis. Abas also focus on features of business intelligence a principle which clarifies how business organizations or individual business can grow. Nowadays, there are several advanced institutions in the world offering undergraduate and postgraduate degree in analytics which can perform and useful for data processing operations. Numerous specialized certifications are obtainable for those who want to be familiar as certified data science and analytics professional. Rani bindu and Shri Kant at al (2020) setups how different sources possess various qualities and ascertain decision making process. To gain this decision making process how information can be used in correct means involves in mapping or analyzing the internal data with external data. Due to exponential growth of massive data much effective and appropriate algorithms need to be devised. Bejjam, Suvarnamukhi & Seshashayee at el (2018) illustrates on how to understand the big data, Arrange the big data, structuring the data, stages of data extraction and transformations of data. It also focusses on how Hadoop (Hadoop Distributed File System) , Map reducing programming frameworks and the mapper step which explains the data operations and its effective implementations. Choudhury, Amitava, and Kalpana Rangra et al (2020) elucidate the technologies that track on big data changes over its transformation. The technologies that exhibit on big data are with read access only from the perspective of analytical system. The data which changes over time knowingly and unknowingly which demands appropriate mechanisms to build which can protect such data. With the invent of various Internet of Things (IoT's) and internet its become so important to be more active towards data processing operations. Though tools of data processing exists in the market beside their advantages it also possess few disadvantages which are highlighted by S.Subhitsha, S.Selvakumar, V.P.Sumathi et al (2017). These author emphases on software tools weka , Orange and Rapid Miner along with their features, advantages and disadvantages.

III.   METHODOLOGY FOR DATA ANALYSIS

As discussed the data is the primary artifact in any organization so it's mandatory to look inside the data like clear & precise definition of data, visibility of data scope, arranging the data using proper data structure, model the data via tables, images, pictorial representations, statistical tables and evaluation of data . Complete and through analysis of data can be happened by appropriate selection of analytical and statistical skills.

Proper prevention of errors and recovery mechanism should be properly ensured. Be ensuring about the reliability and validity of data sources from where it is obtained

A. Data Analysis Methods

Exercise and follow good process in collecting the data by using various qualitative and quantitative approaches. Data Analysis [8] can be divided into

a. Textual analysis which can also referred as data mining it is to arrange the data into large data sets using mining tools. The main aim of textual analysis is to map the data into business data using business intelligence tools.

b. **Descriptive Analysis:** It is to interpret, model and process the previous collected data which can be done in statistical analysis.

c. Inferential Analysis: In which we can investigate various inferences from the same data various samples.

d. Diagnostic analysis: These methods are to investigate the statistical analysis and find the cause for why it happens.

e. Predictive analysis: In this analysis we try to predict what can happen by using statistical data. For example in day to day life how the person does save on his predictable earning income.

f. Prescriptive Analysis: This form analysis is used to collaborate all the previous analysis reports to decide what decision could be taken based on current situation.

g. Factor Analysis: This analysis speaks about how the variables form the relationships within the data set.

h. Discriminant Analysis: This analysis is used to find the relationships between different variable of different groups.

i. Time Series Analysis: Measurement is done based on time series for the variables of data sets**.**

B. Data Analysis Tools

As stated in [16][8] the growing need in the market for Information technology professionals demands from data analytics. It becomes considerably essential to deploy the various data analytics tools in accordance with rising need of society. Below is the list of top 10 of data analytics tools which are open source and as well as paid versions to improve the performance and learning of the system. The below fig: 3 are the following few tools which exists in the market to perform data an**a**lysis tools

Figure 3. Data Analysis Tools

a. Excel: This is product of Microsoft suite and developed under Microsoft Office family for performing mathematical, statistical and analytical operations. Excel is the essential and important entity as analytical tools used in various organizations. It plays an important role by analyzing the complete user requirements and précis in way which is useful to users. It also used for business analytics which helps in presenting of automatic relationship detraction. It can be used in creating budget sheets for personal and business purposes as come up in [8][16].

b. R Programming Language: It is free software programming language and reinforced R foundations for statistical computing. The R Language is widely used data analysists by mining the data and statistical information. R is used as analytical tool which can be used in various ways to extract and present the data of the many organizations as stated in [8].

c. Tableau Public As discussed in [8][16] It is free interactive environment which allows various users to visualize their data over web. This software is used to visualize the presentations known as vizzes can be entrenched into web pages, blogs and can be shared using social media. No much programming is required to run the desktop applications of tableau public software. This software also links with various databases to produce and displays the information.

c. Python: Python is developed by Guido van Rossum created it in the early 1980s, dynamic all-purpose purpose high programming language supports both structured and object oriented programming. It stated in [8][16] Python also rich in library & open source and considered for functional & structured techniques which is used to implement various tasks. Python can assemble in & from any platform such as Mango DB, JSON, SQL, server and many more.

d. SAS: With reference to [8] it is abbreviated as Statistical Analysis System developed in between the year 1980's & 1990's by SAS institute. SAS is a programming environment for managing the data and analytical operations. This programming language is used to manage the data from various sources can be analyzed which can be serve to client profiling and future opportunities. This SAS modules used for Web, Social and market analytics.

e. Apache Spark: As come up in [8][16] Apache spark was created in university of California in the year 2009 AMP lab of barkely., Spark rummage-sale for micro- batching for real time streaming by analyzing large amount of data from various resources. Like Hadoop it also works with the system by distributing the data over various clusters and processes them in parallel**.**

IV. METHODOLOGY FOR DATA PROCESSING

Data gathered from various resources possess numerous potentials which help in decision making to grow any organizations. To define the success for any organization solely depends upon his data and its correct usage. After collecting and analyzing data then next step goes to process that data in a productive means. Following below are the steps to be ensure while processing the data. As stated in paper [12] there are numerous big data technologies have been advanced and classified into data processing concepts. There is bulk of data collected from and extracted to attend knowledge requirements of various business organizations. Hadoop is one of the common and best examples for storing the big data for many organizations.

A. Data Processing Operations

a. Data Grouping and storage: Data need to be collected from various resources and store at appropriate places. Organizing data according to its usage of applications is important.

b. Cross substantiation: After collecting the data from various resources its time to verify the resources from where the data is collected or produced.

c. Data conversion: Conversion of the data is according to its specific format which depends upon its application.

d. Data cleaning and removal: Data cleaning is very mandatory as unwanted data may leads to improper output.

e. Data separations and data sorting: Data should be grouped under different subsets and proper mapping need to be done between them. Example drawing patterns and forming the relationships between the groups.

f. Selection of techniques: choose the right technique as per the requirement which leads to gain output. Also ensure there is proper mechanism to avoid mistakes and recovery from the loo holes. Always apply E.T.L. functions to revalidate your data sets groups.

g. Data summarization and reporting: Obtained result from different groups needs to combine.

h. Data Presentation: After all the operations of data processing data need to be present or model in proper way.

i. Maintenance: Test your OUTPUT again with the initial requirement for a better delivery.

B. Data Processing operations

Data collection is depends upon the type of the applications for which it is being used. Data is processed according to its specific applications as stated below.

a. Data processing for scientific purpose: Data will be collected for the use of scientific study and research activities. It is mainly important to collect and process data without any errors especially for scientific and research activities.

b. Commercial Data Processing: This is a influential system which groups and process bulk volumes of data at high speed. There are multiple users connected to this system and data found with in this system are with minimal errors For example: Airport, University, Super market, online shopping and many more.

c. Automatic versus Manual Data Processing: Manual Data processing allows the user to perform the data processing without aid of any tool or machine. In this processing all the logical, arithmetical, statistical and analytical operations manually or with using any machine. Similarly data is transferred manually from one place to another so this is the reason why it is time consuming and contains. This type of manual processing is suitable for small scale organizations and not suitable for large businesses

d. Automatic Data Processing: This type of contemporary data processing technique which process the data automatically as per the instructions of computer. This is the fastest and best technique which process as per the instructions of computer. So the chances of errors are less and best suited for big and small organizations also.

e. Batch Processing: It is extensively used data processing method which is also termed as sequential or queued offline processing. This is best suitable for the jobs of various users for processing in the order as they received. All the jobs are collected from the users into the stack and given for processing in the same order as they received. It is helpful in decreasing the processing cost and proved to be economical.

f. Real Time Data Processing: As the name implies it **is** suitable for real time applications by processing real time data. It is best suited for displaying the results or output in s short duration. The data stored at the servers or data centers will be used immediately by the software for processing. As it is real time its mandatory to have connection with internet to process and store the data online. Due to these processing deals with spontaneous display of output it needs latest software and hardware capabilities which make it be expensive compared with batch **data processing. Examples** suited for this type of processing is banking sectors, financial institutions, airline reservations and so on.

g. Distributed processing: This method is extensively used by remote workstations connected to central workstations or server. ATM machines are best suited example for this type of processing where all the end user machines are connected to one central machine. All the connected end users machines run according to the central machine. As these works in distributed environment needs continuous internet connection

h. Online processing: Online data processing is a computerized way to enter data in to the system, process and generate reports. These reports are continuously available online and identified by a bar code. A best example of online processing system is automated sales system where the products we select and buy on basis of barcodes. This processing is best over batch processing as it provides continuous data for management in order to provide accuracy and sales statistics. Even reports can be generated on hourly, weekly, monthly and year basis which gives complete details of sales and maintain records.

i. **Multi-Processing:** In this processing systems works by a group of CPU's and connected to each other. The work or job is distributed parallel on each CPU to increase the efficiency and throughput. As CPU is working in parallel it does not effect if any of the CPU failures or fail to work.

j. Time sharing: Time sharing data processing system is based in single core CPU is connected with multiples users allocated with time slice. Each user is allocated with certain amount of time intended to execute their process or jobs. Since multiple users connected it is also referred as computer resources allocated on time basis to have interaction with multiple users.

C. Data Processing tools

Data processing is the gathering and operation of data into the practical and wanted form. The operation is nothing but processing, which is approved either manually or automatically in a predefined order of processes. In the past data is collected and processed manually which is time consuming so it is mandatory to use data processing tools. Following are the below data processing tools listed as listed in [17]:

Retrieval Number: F9887038620/2020©BEIESP
DOI:10.35940/ijrte.F9887.038620
Journal Website: www.ijrte.org

5665

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Figure 4: Fig: Data processing tools

a. **Google Big Query**: This product is from google and it is complete –manageable enterprise data warehouse for analytics. Google offers a fully-managed enterprise data warehouse for analytics via its Big Query product. The solution is server less, and enables organizations to analyze any data by creating a logical data warehouse over managed, columnar storage, and data from object storage and spreadsheets. Big Query captures data in real-time using a streaming ingestion feature, and it's built atop the Google Cloud Platform. The product also provides users the ability to share insights via datasets, queries, spreadsheets, and reports.

b. **Amazon Web Services:** It come up in [17] Amazon Web Services offers Amazon Redshift, a fully managed, petabyte-scale data warehouse that analyzes data using an organization's existing analytic software. Redshift's data warehouse architecture allows users to automate common administrative tasks associated with provisioning, configuring, and monitoring cloud data warehousing. Backups to Amazon S3 are continuous, incremental, and automatic. Redshift also includes Redshift Spectrum, allowing users to directly run SQL queries against large volumes of unstructured data without loading or transforming.

c. **Hortonworks**: As focus on [17] the development and support of Apache Hadoop. Hortonworks Dataflow (HDF) manages streaming data by securely acquiring and transporting it to the Hortonworks Data Platform. The solution organizes and oversees all data types. Hortonworks has a partnership with Microsoft for hybrid deployments, but offers a version of HDP on Amazon Web Services as well.

d. **Cloudera:** Cloudera discusses in [17] offers a data storage and processing platform based on the Apache Hadoop ecosystem, as well as a proprietary system and data management tools for design, deployment, operations and production management. Cloudera differentiates itself from other Hadoop distribution vendors by continuing to invest in specific capabilities, such as improvements to Cloudera Navigator (which provides metadata management, lineage and auditing), while at the same time keeping up with the Hadoop open-source project.

V. METHEDOLOGY FOR DATA PRESENTATIONS

Data presentations play as a front end after data analysis and data processing operations. It is exactly the end product of the customer who wills to have his data in his preferred manner. Yet several tools exist in the market to visualize data still we efforts on few of the tools listed. Google Public Data Explorer provides an free interface for sharing various data sets. It also as an interactive tool for visualization

and presentation which permit changes to be followed over data sets. This tool permits the user to create pie charts, bars, represent colors, graphs, shapes and so on as shown in paper [18].

A. Tableau Public

This tool as discussed in [18] is used to create an interface to access over the web, download and represent visual pictures. It provides a free version represent the data across various data sets. Tableau can provide access the represent the data by using various multiple formats. All the data will not only save at the local computer but also over the internet to access any time. Tableaus have 57000 Plus accounts from various organizations to save, access and represent the data.

b. Wordle

Cloud clouds shows in [19] are also term as wordle or word collage or tag cloud which is used for visual representations of words. A user can design wordle gallery to interact and save the data across over various clouds.

VI. ROLE OF DATA SCIENTIST

Data Scientist is the person mainly involves who make use of logical progression of data into cherished or valuable form [10]. With enormous advancement of numerous forms of data involves the data scientist to operate the data into multiple or various levels such as data cleaning, data loading, data modeling, data processing and evaluation of data. As the data is a gathered from various fields so it's mandatory to make use of advancement of skills in various fields like A.I, Machine learning, robotics, biotechnology, statistical approaches, analytical methods, medical sciences, mathematical procedures and IoT's [10].

A. Data Scientist perspectives towards organizations:

- Effective use of data for growing business
- Proper mechanisms to be develop for acquiring the data from various sources
- Cleaning data
- Process and evaluate data
- Proper A.I algorithms need to be device
- Involve Deep machine learning algorithms
- Develop Analyze, statistical and logical reasoning methods

A. Data science uses and applications

As discussed in [15] Data Science has conquered maximum all the organizations of the globe today. There is no such business across the globe which does not use data to improve their organizations. As such, data science has become important aid for organizations to make effective use of data. There are various organizations like banking, financial institutions, automations **and** engineering, conveyance, e-commerce, edification sectors, etc. that use data science.

B. Role of Data science in Banking or Financial institutions

Banking is one of the leading sectors which can make use of beneficiaries or customers data in effective means. These institutions can make better decisions and predict future preventions of frauds in an intelligent way. Management of customer's data involves much analytical, statistical, mathematical reasoning incorporated with A.I techniques, or algorithms, deep learning and machine learning.

This also supports in maintaining customer, **predicting the plans** according to their usage & savings, investment plans and so on.

C.   Role of Data science in Edifications sectors

Data science plays a decisive role in the development of all activities involved in education sectors. The learning ability of the students will be improved by knowing their data and skills they possess. Depending on the skills of the student's data new learning mechanisms can be devised to cope up and attain learning objectives. Data acquired from the students helps data scientist in analyzing student requirements, building their emotional & social skills, developing or in cultivate their learning parameters & cognitive skills, monitoring regular student performances, measuring parameters of instructions and maintaining the community relationships.

D.   Role of Data Science in Health care or medical provisions

After successful registration of the customer data science involves in extracting the meaningful information's to maintain the records of patient. These meaningful insights help us to create the patient domains and perform predictive modeling like classifications, reversions and visualizations or presentations of data. It also helps in developing the scalable algorithms or procedures like indexing, streaming and sampling the data. It incorporates and makes uses of computing platforms like cloud to store & access the data. Input and Output data to and from the data bases involve much data science operations as discussed above. Sensors also benefitted by making use of data science operations like imaging, spectroscopy and microscopic operations.

E.   Data Science in Digital Marketing

With the arrival of data science study much advancement have been brought in the field of marketing to promote their respective business. To grow any organizations we can't deny that marketing plays an vital role which is possible by means of many social media applications. It allows customer connections in web based environment by means of Facebook, amazon and various e-commerce sites.

F.   Role of Data Science in Automated Language Analysis

Various organizations invite automated language analysis operations and hiring relevant professionals. To promote and achieve good organizational setup data science will proves to be helpful for the advertisements depends upon the moods of the customers.

G.   Role of Data Science

In whether prediction weather models are needed to be predicted and whether forecasting from time to time .Data science incorporates various deep machine learning techniques to achieve this objective of forecasting and prediction. Data acquisition should be properly acquired from various sources in order to take accurate decisions. These proper predications help to organize the events like sports, meeting, public addressing, examinations and cultural events properly. Satellite images of various shapes & sizes operate in white & black spectrum which can identified by data science operations.

H.   Data Science Contributions for the future

Data Science comprehends many advances technologies like Artificial Intelligence, Internet of Things (IoT's), Deep learning, machine learning and so on. With the advancements in technology demands to incorporate & implement the statistical, mathematical and logical reasoning concepts. Proper mechanisms need to stratagem to make the organizations to handle the data with operative use. There are numerous reasons to give for which we need data science operations to be performed in business like

- Organizations how they mishandle the data
- Data protection to formulate the regulations & policies, a surprising incline in data growth
- Much demand for the data scientists
- Natural Language Processing (NLP) will be used for information retrieval
- Data purgative should be computerized
- Much need to improve business intelligence
- Used to predict sports, whether, banking sector, stocks and shares
- Need much improvement in social media applications

I.   Data Science Careers

Exponential growth of any organizations is complete depends and demands to have right decisions which is possible by hiring good or sound data scientists. Following below are few Professions in the field of data science:

- Business Intelligence Developer
- Data Architect
- Applications Architect
- Infrastructure Architect
- Enterprise Architect
- Data Analyst
- Data Scientist
- Data Engineer
- Machine Learning Scientist
- Machine Learning Engineer
- Statistician

VII.   RESULTS

The below table is based upon the methodologies discussed for Data analysis, processing & presentation various tools or software's can be used. This table also focuses on data science perspective, applications of data science over various fields to grow the organizations. It also focusses on data scientist carrier option for the future

Table 1: Data Science tools and operations

Sl.No	Data Science operations	Tools and operations
1	Data Analysis	Rapidminer, Qlikview, Excel, SAS, Python, Tableau public R and Splunk.
2	Data processing	Hadoop, Cassandra, Cloudera, Flink, Qubole, Statwing, Storm and couchDB.
3	Data presentation	Tableu public

4	Data Scientist role	To breed organizations like medical diagnostics, financial institutions, edification sectors, health care, digital marketing, automated language processing and many more.
5	Future Expectations	Develop natural language processing, business intelligence, social media, whether furcating, stock market predications and others
6	Data scientist professions	Business intelligence developer, data architect, data analysis, data scientist, machine learning scientist and others.

VIII. CONCLUSION

Know a day's data science becomes as a mandatory field which coordinates between multi disciplines like mathematics, statistical approaches, mathematical methods, logical reasoning, intelligence algorithms and machine learning practical's. All these fields correlate to access the data from various business or organizations and make use of them in effective means. These effective use of data leads to perform proper decision making to grow business further on the basis of customer chooses and satisfaction. Hence we can conclude that rise of data science field can demand more positions of data scientists to grow in each organization. At last we focus on how successful carriers can be built in the field of data science. The main beauty of this field it used to grow all businesses.

REFERENCES

1. Russell, Stuart J., and Peter Norvig. Artificial intelligence: a modern approach. Malaysia; Pearson Education Limited,, 2016.
2. Nilsson, Nils J. Principles of artificial intelligence. Morgan Kaufmann, 2014.
3. 3. Bell, Jason. Machine learning: hands-on for developers and technical professionals. John Wiley & Sons, 2020.
4. Van Der Aalst, Wil. "Data science in action." Process mining. Springer, Berlin, Heidelberg, 2016. 3-23.
5. Dhar, Vasant. "Data science and prediction." Communications of the ACM 56.12 (2013): 64-73.
6. Hazen, Benjamin T., et al. "Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications." International Journal of Production Economics 154 (2014): 72-80.
7. Wimmer, Hayden, and Loreen Marie Powell. "A comparison of open source tools for data science." Journal of Information Systems Applied Research 9.2 (2016): 4.
8. Islam, Mohaiminul. "Data Analysis: Types, Process, Methods, Techniques and Tools." International Journal on Data Science and Technology 6.1 (2020): 10.
9. Nicolae, Bogdan, et al. "Park, Yoonho. Leveraging Adaptive I/O to Optimize Collective Data Shuffling Patterns for Big Data Analytics. IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS. PP (99) pp: 1-13." (2020).
10. Abas, Zuraida Abal, et al. "Analytics: A Review Of Current Trends, Future Application And Challenges." Journal of Advanced Computer Technology. PP 3560 (2020): 3565.
11. Rani, Bindu, and Shri Kant. "An Approach Toward Integration of Big Data into Decision Making Process." New Paradigm in Decision Science and Management. Springer, Singapore, 2020. 207-215.
12. Bejjam, Suvarnamukhi & Seshashayee, M.. (2018). Big Data Concepts and Techniques in Data Processing. International Journal of Computer Sciences and Engineering. 6. 712-714.
13. Choudhury, Amitava, and Kalpana Rangra. "Trends and Technologies in Big Data Processing: An Overview." Innovations, Algorithms, and Applicatftions in Cognitive Informatics and Natural Intelligence. IGI Global, 2020. 17-42.
14. Sumathi, S. Subhitsha1 S. Selvakumar2 VP. "Comparative Analysis of various Data Mining Tools."
15. Zuccolotto, Paola, and Marica Manisera. Basketball Data Science: With Applications in R. CRC Press, 2020.K. Elissa, "An Overview of Decision Theory," unpublished. (Unplublished manuscript)
16. Data Flair" Data Science Tools" (2019) available at https://data-flair.training/blogs/data-science-tools/
17. Timothy King," Data Management Solutions Review" (2018), Available at https://solutionsreview.com/data-management/the-4-best-big-data-processing-software-tools-to-consider/
18. In, Junyong, and Sangseok Lee. "Statistical data presentation." Korean journal of anesthesiology 70.3 (2017): 267.
19. Creative Blog," Data Visualization Tools", (2019), available at https://www.creativebloq.com/design-tools/data-visualization-712402

AUTHORS PROFILE

Mr. Mujthaba Gulam Muqeeth holding B.Tech (Kakatiya University), M.Tech (J.N.T.U.H) with specialization in Computer Science and Pursuing is PhD in Artificial Intelligence. Having a vast experience of 17years in the teaching field of computer science engineering courses in various reputed engineering colleges at both national and international level. Presently working as a Lecturer of Computer Science Department at Prince Sattam Bin Abdul-Aziz University, kingdom of Saudi Arabia since nine years. Besides teaching working with Quality Control Department accredited with NCAAA. His area of interest includes Cloud Computing, Software Engineering and Artificial Intelligence. He published papers and books on Cloud Computing and Computer Networks.

Dr. Manjur Kolhar did his Master of Application master degree and Ph.D from Karnataka university, Dharwad, Karnataka, and Malaysia respectively. Presently he is working as an Associate Professor of Computer Science Department at Prince Sattam Bin Abdul-Aziz University, Kingdom of Saudi Arabia. He has more than 20 years of experience in software development and academics. His area of interest includes Computer Networks Cloud Computing, Software Engineering, Artificial Intelligence and Neural Networks. He published various papers and books on his research areas. Presently on working image processing, focused on smart parking, using convolution network. He has received various grants from deanship of research, Prince Sattam Bin Abdulaziz University.

Dr. Abdalla Al Ameen holding Ph.D in Computer Science is working as an Associate Professor and Head of Computer Science Department at Prince Sattam Bin Abdul-Aziz University, K.S.A. His area of interest includes Databases Computer Networks Cloud Computing, Software Engineering and computer security. He published various papers and books on his research areas.

Dr.Rahmath Mohammed holding B.Tech in Electronics and Communication from J.N.T.U.H. M.S from New York Institute of technology (U.S.A),Ph.D from V.B.S.Purvanchal university. He is currently working as Lecturer in Department of Computer Science at College of Arts & Sciences, Prince Sattam Bin Abdul Aziz University, and K.S.A. His area of interest is Computer Networks, Communication Systems, Cloud computing, Middle ware technologies. He published various papers in international journals.