International Journal of Engineering, Science and Mathematics
Vol.9Issue 05, May2020
ISSN: 2320-0294 Impact Factor: 6.765
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

RESEARCH ON DATA SCIENCE, DATA ANALYTICS AND BIG DATA

Rahul Reddy Nadikattu

University of the Cumberlands United States

BIG DATA:

Big Data refers to a huge volume of data of various types, i.e., structured, semi structured, and unstructured. This data is generated through various digital channels such as mobile, Internet, social media, e-commerce websites, etc. Big Data has proven to be of great use since its inception, as companies started realizing its importance for various business purposes. Now that the companies have started deciphering this data, they have witnessed exponential growth over the years.Impact on various sectors like Retail, Banking and investment, Fraud detection and analyzing, Customer-centric applications and Operational analysis

DATA SCIENCE

Data Science deals with the slicing and dicing of the big chunks of data, as well as finding insightful patterns and trends from them using technology, mathematics, and statistical techniques. Data Scientists are responsible for uncovering the facts hidden in the complex web of unstructured data so as to be used in making business decisions. Data Scientists perform the aforementioned job by developing heuristic algorithms and models that can be used in the future for significant purposes. This amalgamation of technology and concepts makes Data Science a potential field for lucrative career opportunities. McKinsey once predicted that there will be an acute shortage of Data Science Professionals in the next decade. Impact on various sectors like Web development, Digital advertisements, E-commerce, Internet search, Finance, Telecom, Utilities.

DATA ANALYTICS:

Data Analytics seeks to provide operational insights into complex business situations. The concept of big data has been around for years; most organizations now understand that if they capture all the data that streams into their businesses, they can apply analytics and get significant value from it. But even in the 1950s, decades before anyone uttered the term "big data," businesses were using basic analytics (essentially numbers in a spreadsheet that were manually examined) to uncover insights and trends. he new benefits that big data analytics brings to the table, however, are speed and efficiency. Whereas a few years ago a business would have gathered information, run analytics and unearthed information that could be used for future decisions, today that business can identify insights for immediate decisions. The ability to work faster – and stay agile – gives organizations a competitive edge they didn't have before. Looking into the historical data from a modern perspective, finding new and challenging business scenarios and applying methodologies to find a better solution are the prime concerns of a Data Analyst. Not only this, but a Data Analyst also predicts the upcoming opportunities which the company can exploit. Data Analytics has shown such a tremendous growth across the globe that soon the Big Data market revenue is expected grow by 50 percent.Impact on various sectors like Traveling and transportation, Financial analysis, Retail, Research, Energy management, Healthcare.

Data is the baseline for almost all activities performed today, whether it is in the field of education, research, healthcare, technology, retail, or any other industry. The orientation of businesses has changed from being product-focused to data-focused. Even a small piece of information is valuable for companies nowadays, making it essential for them to derive more and more information possible. This necessity gave rise to the need for experts who could bring meaningful insights.It is evident from this table how these areas impact our economy. Actually, technologies are helping diverse sectors in a great way, allowing them to put each and every piece of insight into use. While Big Data is helping the retail, banking, and other industries by providing some of the important technologies such as fraud-detection systems, operational analysis systems, etc., Data Analytics allows the industries of healthcare, banking, traveling and transport, energy management, etc. to come up with new advancements using the historical trends. On the other hand, Data Science is letting the companies get into web development, digital advertisements, e-commerce, etc., and dive deep into the granular information for different purposes. Data Scientists perform the most challenging jobs among the three. Data Scientist is one of the most trending profiles in the 21st century. There is considerable overlapping of roles between Data Analysts and Big Data Professionals. Data Science is booming like anything and hence has been tagged as the sexiest job of the 21st century by Forbes. Skill sets required to become Data Scientists, Data Analysts, and Big Data Professionals are different. Though there are some skills that are common in all the three profiles, the level of proficiency varies as per the job roles. Therefore, you should clearly know what you want to become and what skills you need to have for that. In order to become a Data Scientist, you need to be proficient in mathematics, statistics, programming, and business strategies. You should have good communication skills, as a Data Scientist needs to distribute the information to various departments of an organization. Similarly, a Big Data Professional would require having a good grasp of technology (such as Hadoop and Java), mathematics, and statistics, as well as analytics. However, a Data Analyst needs to be good in programming, Artificial Intelligence, and data wrangling. While many people use the terms interchangeably, data science and big data analytics are unique fields, with the major difference being the scope. Data science is an umbrella term for a group of fields that are used to mine large datasets. Data analytics is a more focused version of this and can even be considered part of the larger process. Analytics is devoted to realizing actionable insights that can be applied immediately based on existing queries. Data science isn’t concerned with answering specific queries, instead parsing through massive datasets in sometimes unstructured ways to expose insights. Data analysis works better when it is focused, having questions in mind that need answers based on existing data. Data science produces broader insights that concentrate on which questions should be asked, while big data analytics emphasizes discovering answers to questions being asked.More importantly, data science is more concerned about asking questions than finding specific answers. The field is focused on establishing potential trends based on existing data, as well as realizing better ways to analyze and model dataThe two fields can be considered different sides of the same coin, and their functions are highly interconnected. Data science lays important foundations and parses big datasets to create initial observations, future trends, and potential insights that can be important. This information by itself is useful for some fields, especially modeling, improving machine learning, and enhancing AI algorithms as it can improve how information is sorted and understood. However, data science asks important questions that we were unaware of before while providing little in the way of hard answers. By adding data analytics into the mix, we can turn those things we know we don’t know into actionable insights with practical applications Big Data for financial services: Credit card companies, retail banks, private wealth management advisories, insurance firms, venture funds, and institutional investment banks use big data for their financial services. The common problem among them all is the massive amounts of multi-structured data living in multiple disparate systems which can be solved by big data. Thus big data is used in several ways like: Customer analytics, Compliance analytics, Fraud analytics, Operational analytics. Big Data in Communications: Gaining new subscribers, retaining customers, and expanding within current subscriber bases are top priorities for telecommunication service providers. The solutions to these challenges lie in the ability to combine and analyze the masses of customer-generated data and machine-generated data that is being created every day.Big Data for Retail: Brick and Mortar or an online e-tailer, the answer to staying the game and being competitive is understanding the customer better to serve them. This requires the ability to analyze all the disparate data sources that companies deal with every day, including the weblogs, customer transaction data, social media, store-branded credit card data, and loyalty program data.Healthcare: The main challenge for hospitals with cost pressures tightens is to treat as many patients as they can efficiently, keeping in mind the improvement of the quality of care. Instrument and machine data is being used increasingly to track as well as optimize patient flow, treatment, and equipment used in the hospitals. It is estimated that there will be a 1% efficiency gain that could yield more than $63 billion in global healthcare savings. Travel: Data analytics is able to optimize the buying experience through mobile/ weblog and social media data analysis. Travel sights can gain insights into the customer's desires and preferences. Products can be up-sold by correlating the current sales to the subsequent browsing increase browse-to-buy conversions via customized packages and offers. Personalized travel recommendations can also be delivered by data analytics based on social media data. Gaming: Data Analytics helps in collecting data to optimize and spend within as well as across games. Game companies gain insight into the dislikes, the relationships, and the likes of the users.Energy Management: Most firms are using data analytics for energy management, including smart-grid management, energy optimization, energy distribution, and building automation in utility companies. The application here is centered on the controlling and monitoring of network devices, dispatch crews, and manage service outages. Utilities are given the ability to integrate millions of data points in the network performance and lets the engineers use the analytics to monitor the network.We have made tremendous progress in the field of Information & Technology in recent times. Some of the revolutionary feats achieved in the tech-ecosystem are really commendable. Data and Analytics have been the most commonly used words in the last decade or two. As such, it's important to know why they are inter-related, what roles in the market are currently evolving and how they are reshaping businesses.Technology often regarded as a boon to those already aware of its potential, can also be a curse to audiences who can't keep up with it's rapid growth. Each era has had it's moments of breakthrough and an equal share of victims (or as I'd like to call them collateral damage). As of today, every monetary-driven industry completely relies on Data and Analytics for it's survival.This was the uprising of Data warehouse where customer (Business) and production processes (Transactions) were centralized into one huge repository like eCDW (Enterprise Consolidated Data Warehouse) . A real progress was established in gaining an objective, deep understanding of important business phenomena — thereby giving managers the fact-based comprehension to go beyond intuition when making decisions.The data surrounding eCDW was captured , transformed and queried using ETL & BI tools. The type of analytics exploited during this phase was mainly classified as Descriptive (what happened) and Diagnostic (why something happened).However , The main limitations observed during this era were that the potential capabilities of data were only utilized within organizations , i.e. , the business intelligence activities addressed only what had happened in the past and offered no predictions about it's trends in the future.The certain drawbacks of the previous era became more prominent by the day as companies stepped out of their comfort-zone and began their pursuit of a wider (if not better) approach towards attaining a sophisticated form of analytics. Customers surprisingly reacted well to this new strategy and demanded information from external sources (clickstreams , social media , internet , public initiatives etc) . The need for powerful new tools and the opportunity to profit by providing them — quickly became apparent. Inevitably , the term 'Big data' was coined to distinguish from small data, which is generated purely by a firm's internal transaction systems.What companies expected from their employees was to help engineer platforms to handle large volumes of data with a fast-processing engine . What they didn't expect was a huge response from an emerging group of individuals or what is today better known as the "Open Source Community". This was the hallmark of Analytics 2.0.With the unprecedented backing of the community , Roles like Big-Data Engineers & Hadoop Administrators grew in the job-sector and were now critical to every IT organization. Tech-firms rushed to build new frameworks that were not only capable of ingesting , transforming and processing big-data around eCDW/Data Lakes but also integrating Predictive (what is likely to happen) analytics above it. This uses the findings of descriptive and diagnostic analytics to detect tendencies, clusters and exceptions, and to predict future trends, which makes it a valuable tool for forecasting. In today's tech-ecosystem , I personally think the term big-data has been used, misused & abused on many occasions. So technically, 'big data' now really means 'all data'.The pioneering big data firms began investing in analytics to support customer-facing products, services, and features. They attracted viewers to their websites through better search algorithms, recommendations , suggestions for products to buy, and highly targeted ads, all driven by analytics rooted in enormous amounts of data. The outbreak of the Big-Data phenomena spread like a virus. So, now it's not just tech-firms and online companies that can create products and services from analysis of data, it's practically every firm in every industry. On the other hand, the wide-acceptance for big-data technologies had a mixed impact . While the tech-savvy giants forged ahead by making more money, a majority of other enterprises & non-tech firms suffered miserably at the expense of not-knowing about the data. As a result, a field of study Data Science was introduced which used scientific methods, exploratory processes, algorithms and systems to extract knowledge and insights from data in various forms.Indeed, an interdisciplinary field defined as a "concept to unify statistics, data analysis, machine learning and their related methods" in order to "understand and analyse actual phenomena" with data. In other words , a well-refined data combined with good training models would yield better prediction results. The next-generation of quantitative analysts were called data scientists, who possessed both computational and analytical skills.The tech-industry exploded with the benefits of implementing Data Science techniques and leveraged the full power of predictive & prescriptive (what action to take) analytics ,i.e, eliminate a future problem or take full advantage of a promising trend. Companies began competing on analytics not only in the traditional sense — by improving internal business decisions — but also by creating more valuable products and services. This is the essence of Analytics 3.0. There has been a paradigm shift in how analytics are used today. Companies are scaling at a speed beyond imagination, identifying disruptive services, encouraging more R&D divisions — many of which are strategic in nature. This requires new organizational structureThere have always been four types of analytics: descriptive, which reports on the past; diagnostic, which uses the data of the past to study the present; predictive, which uses insights based on past data to predict the future; and prescriptive, which uses models to specify optimal behaviors and actions. Although , Analytics 3.0 includes all of the above types in a broad sense, it emphasizes the last . And it introduces — typically on a small scale — the idea of automated analytics.Analytics 3.0 provides an opportunity to scale decision-making processes to industrial strength. Creating many more models through machine learning can let an organisation become much more granular and precise in its predictions. Having said that ,the cost & time for deploying such customised models wasn't entirely affordable and necessitated a cheaper or faster approach. The need for automation through intelligent systems finally arrived , and this idea (once deemed as beyond-reach) that loomed on the horizon is where Analytics 4.0 came into existence .There is no doubt that the use of artificial intelligence, machine learning and deep learning is going to profoundly change knowledge work. We have already seen their innovative capabilities in the form of Neural Machine Translation, Smart Reply, Chat-bots, Meeting Assistants etc ,which will be extensively used for the next couple of years. The data involved here originated from vast heterogenous sources consisting of indigenous types — one that requires complex training methods — and especially those that can sustain (make recommendations, improve decision-making, take appropriate actions) itself. Employing data-mining techniques and machine learning algorithms along with the existing descriptive-predictive-prescriptive analytics comes to full fruition in this era. This is one reason why Automated Analytics is seen as the next stage in analytic maturity.filled with the promise of a utopian society run by machines and managed by peace-loving managers and technologists. We could reframe the threat of automation as an opportunity for augmentation : combining smart humans and smart machines to achieve an overall better result.

Now, instead of pondering “What tasks currently employed by humans will soon be replaced by machines?” I’d rather optimistically question “What new feats can companies achieve if they have better-thinking machines to assist them? or How can we prevent death tolls in a calamity-prone area with improved evacuation AI routines ? or Why can’t AI-driven e-schools be implemented in poverty-ridden zones ?” Most organisations that are exploring “cognitive” technologies — smart machines that automate aspects of decision-making processes — are just putting a toe in the water. Some are doing pilots to explore the technology. While others are working on the concept of building a Consumer-AI-Controlled platform. These platforms use the idea of Personal AI agents that communicate with other AI services or so called bots to get the job done. There will be no more manual interventions necessary with just an AI-powered system to steer your personal day-to-day activities. I wouldn’t be surprised to see either of these technologies making giant leaps in the future. Surely, there’s an element of uncertainty tied to them but unlike many, I’m rather optimistic about the future.