Data Analytics For Beginners

Introduction To Data Analytics

By Anthony S. Williams

Table Of Contents

Conclusion

Copyright © 2021 By Anthony Williams - All Rights Reserved.

This document is geared towards providing exact and reliable information in regards to the topic and issue covered. The publication is sold with the idea that the publisher is not required to render accounting, officially permitted, or otherwise, qualified services. If advice is necessary, legal or professional, a practiced individual in the profession should be ordered.

From a Declaration of Principles which was accepted and approved equally by a Committee of the American Bar Association and a Committee of Publishers and Associations.

In no way is it legal to reproduce, duplicate, or transmit any part of this document by either electronic means or in printed format. Recording of this publication is strictly prohibited, and any storage of this document is not allowed unless with written permission from the publisher. All rights reserved.

The information provided herein is stated to be truthful and consistent, in that any liability, in terms of inattention or otherwise, by any usage or abuse of any policies, processes, or directions contained within is the solitary and utter responsibility of the recipient reader. Under no circumstances will any legal responsibility or blame be held against the publisher for any reparation, damages, or monetary loss due to the information herein, either directly or indirectly.

Respective authors own all copyrights not held by the publisher.

The information herein is offered for informational purposes solely and is universal as so. The presentation of the information is without contract or any type of guarantee assurance.

The trademarks that are used are without any consent, and the publication of the trademark is without permission or backing by the trademark owner. All trademarks and brands within this book are for clarifying purposes only and are owned by the owners themselves, not affiliated with this document.

Introduction

Even if you know nothing about data analytics, you have probably heard of these two words, especially if you are interested in any computer science field. So, what is data analytics? And is there any difference between data analysis and data analytics?

Data analytics, especially big data analytics, has been and will continue to change and modify the world we live in.

More specifically, data analytics is transforming the ways companies and businesses use the raw data they gather to make valuable conclusions about their products and services.

In this digital world, making the right decisions makes a huge difference between succeeding and failing, and this is why data analytics is necessary.

At a basic level, data analytics is a computer science that involves gathering various kinds of raw data in order to detect and analyze different trends and draw valuable conclusions based on massive data batches collected.

Data analytics involves numerous techniques, and many of these techniques revolve around transforming raw data collected into other forms, which make it possible for businesses and companies to detect and analyze those important business metrics.

If no data analytics techniques are used, all of those important metrics and information that raw data has will be lost or buried under piles and piles of other not-so-important information. Essentially, data analytics increases businesses and companies' efficiency and makes it much easier to detect and analyze trends.

Needless to say, pursuing a career in this field is an excellent choice if you are ready to constantly refine your skills, learn new things, and of course, challenge yourself. Learning new things when you are a data analyst is a must if you want to stay on top of your game and benefit from more job prospects.

For a high-income data analyst career, learning new skills and refining those skills you already possess is also needed, and this is why starting with the basics is the best route to take.

Chapter 1:

Becoming a Data Analyst

All renowned, bigger and smaller, more and less successful companies and businesses are hiring experts in data analytics, so pursuing a career as a data analyst is definitely something worth considering.

Pursuing a career as a data analyst most certainly makes sense if you are looking for high-income job opportunities.

In fact, there will never be a shortage of such job opportunities as companies and businesses will always focus on data analytics to boost their overall productivity and efficiency.

Moreover, skilled, professional data analysts fall into the category of individuals with the most job opportunities. The demand for expert data analysts has been growing for a couple of years now and will continue to grow in the years after.

The demand for data analysts allows every single expert in the field to make a living. While the demand is extraordinary, the supply of individuals who are truly capable of doing data analytics jobs is quite limited.

Add to this high demand amazing perks and huge salaries, becoming a data analyst is a wonderful career to pursue even though it is not for everyone.

Numerous data analytics jobs are posted on a daily basis throughout all sorts of industries and companies. Every single business or company out there that uses any kind of data needs an expert data analyst to analyze this raw data.

Before big companies make any sort of investment decision, they need a data analyst by their side to do the magic and assess any potential risks associated with making big investment choices.

What Do Data Analysts Do?

Now that you know about data analytics job opportunities and prospects, let’s examine what data analysts actually do. As mentioned in the previous section, data analysts are individuals who rely on various kinds of data analytics techniques and tools in order to scrutinize the raw data they collect.

Using various data analytics tools and techniques, they are able to bring meaningful, valuable information from the raw data collected, and this way, they help their companies and businesses with making those important decisions.

Data analysts are also responsible for collecting raw data and using this data to identify and analyze current trends in their respective industries.

In particular, data analysts collect and then organize raw data and finally analyze statistical information in order to make it valuable for their companies and businesses.

Some of their main duties include using computerized models to collect, organize, and interpret the raw data needed, removal of any kind of corrupted or less meaningful data, performing data analysis to examine its quality, undertaking analysis to determine the meaning of data, and performing analysis for data screening.

Data analysts are also responsible for preparing reports that they will present to companies and business management teams.

Data analysts are employed in all sorts of industries and fields, including software development businesses, banking institutions, consultancy agencies, telecommunication businesses, various organizations from the public sector, pharmaceutical businesses, and of course, various manufacturing businesses.

Data analysts collect large bulks of raw data, process that data, and use it to perform various kinds of statistical analyses.

Essentially, they use those large bulks of raw data collected and figure out how that data can be used to solve crucial problems and answer important questions.

By taking mountains of raw data, data analysts further probe that data to analyze different trends, make predictions or forecasts, as well as extract crucial information to help their clients or companies make much better and much more informed decisions.

In addition to working in those previously mentioned industries and fields, data analysts also work with hedge funds, private equity companies, retail, marketing, and the healthcare industry. In fact, data analysts are needed everywhere, and this is why pursuing a data analytics career is always a good idea.

All of those big tech firms such as Google and Facebook need data analysts. All investment banks rely on data analysts. Governments also rely on data analysts who gather data, do data mining, and of course, data analysis, as previously mentioned.

Data Analytics Sector and Qualifications

Generally, data analysts possess an extremely dynamic set of skills, and they are very good when it comes to working with details and numbers.

Data analysts are very organized, very confident, and generally very successful when it comes to juggling many different tasks at once.

Jobs offered within the data analytics sector provide great salaries, and there are numerous different career paths to explore. The data analytics sector has a wealth of working opportunities throughout many different corporate levels.

With data analytics skills, you can become a financial analyst, you can do market research, focus on machine learning and big data, and even more.

One of the most diverse data analytics fields is definitely financial data analytics. Jobs within this field usually include management analytics, business analytics, and various kinds of investment analytics.

Financial analysts usually have hourly salaries ranging between $25 and up to $80. On a yearly basis, financial analysts can earn at least $100,000. This sector of data analytics is forecasted to grow at a staggering rate of 5% in the next five years.

Then, there is the extremely profitable and dynamic market research field. Data analysts working in this field earn around $35 per hour. This data analytics field is expected to grow at a rate of 18% in the next five years.

Following the growing business world, the demand for big data analysts and technologies revolving around big data is also growing. In addition to big data, machine learning technologies are evolving as well, and data analysts working within this field have amazing opportunities.

This leads us to the qualifications required to get a job as a data analyst. One of the major qualifications is finishing a data analytics program. With this degree, you will have no issues getting an entry-level position.

Having a degree in statistics, mathematics, and even economics can be enough to get started as well. However, you can get a job as a data analyst even if you do not have this type of degree. It is not a secret that people get hired as data analysts without academic backgrounds, but your chances are much higher if you have an academic background in data analytics, statistics, or some other math-related program.

Choose Your Career Path

If you possess strong data analytics skills, you have limitless career opportunities, and this is the biggest benefit of learning about data analytics and working on refining your skills and knowledge.

There are numerous career paths at your disposal. You can work as a business analyst, and business analysts are responsible for analyzing massive bulks of business-associated data. You can do management reporting as well.

Data analysts working within this field analyze and report data needed by business and management teams. Data analysts also devise different corporate and business strategies, and individuals working at these positions usually work with big data. They focus on acquisitions and mergers and generally advise management teams on business strategies.

Various human resource departments also employ data analysts to analyze benefits and compensation data. In those cases, there are many budget data analysts, and they focus on analyzing data related to budgeting.

Insurance companies also employ data analysts who work on supporting, enhancing, and of course, optimizing the overall sales process. Data analysts are also web analysts, and individuals working within this field are responsible for analyzing big data associated with websites.

Data analysts can also work on analyzing and monitoring fraud data while professionals working for credit companies do credit monitoring and examine data for lending approvals and credit reposting.

Professionals working as business product data analysts explore the characteristics and attributes of products. They also advise management teams on product pricing depending on different market factors.

Social media data analysts also have fantastic working opportunities. All tech and social media companies rely on data analysts to advance, monitor, and even build the technology that drives their businesses.

Those working within the machine learning field usually focus on crafting technology, programming, data feeds, data preparation, and more that assist companies and businesses in making proper decisions.

Chapter 2:

What is Data Analytics?

As mentioned at the beginning, data analytics is the science of extracting meaningful, valuable information from raw data.

To do so, data analysts rely on various computer systems that help with collecting, transforming, and of course, modeling the raw data that is then used to identify certain patterns and draw conclusions.

Essentially, data analytics may be very simple, but in today's world, data analytics usually revolves around collecting and analyzing massive data volumes, including high-velocity information, and in these cases, data analytics is more challenging and requires more skills and expertise.

Generally speaking, data analysts who deal with high-velocity, high-volume data need to possess strong expertise in statistics, and this is why they are also called data scientists.

While many believe that data analytics only revolves around big data, the truth is that any kind of processing data is considered data analytics, whether it be smaller or larger chunks of data collected, transformed, and processed.

However, data analytics has been and will continue to evolve to be more capable of dealing with big data and associated computer systems.

Essentially, data analytics is a science based on strategy and statistics. Collected raw data is used to answer important questions, detect trends, and draw crucial conclusions from large chunks of data.

Numerous techniques are involved in data analytics. Some of these convert collected raw data into another form that organizations and companies can use to identify important metrics. This process, in particular, lets them increase their businesses and companies' efficiency on a bigger scale.

Data analytics and all associated strategies and techniques are essential when it comes to identifying different patterns, finding anomalies and relationships in large chunks of data, and making the data or information collected more meaningful and more understandable.

Data analytics is understood as the collection, organization, and transformation of raw data facts to make predictions and draw conclusions to make better-informed business decisions.

When it comes to data analytics processes and techniques, both quantitative and qualitative processes and techniques are used to reach business goals faster and enhance overall productivity and efficiency.

You can also think of data analytics as an amazing science-based practice that uses big or smaller chunks of data to drive performance and strategy.

As it is, data analytics and all of the processes involved can answer four very important questions, including what occurred, why it occurred, what is most likely to happen in the future, and what can be done for that something to occur again.

When used the right way, data analytics and everything it involves can significantly boost the capability of all fields or departments within a company or organization, including customer assurance and human resources departments.

With many different techniques involved, data analytics is used for many different goals that vary from one company to another. No matter the business goals and strategies used, data analytics processes always involve a set of common components, and these components are used to draw important answers and get assistance with many different initiatives.

More specifically, by combining these different components, data analysts are able to get a clear picture of their businesses, including where their businesses are right now and in which direction they should be moving.

Data of all sorts can help companies and businesses to better understand their business goals and their customers. Data analytics can also help companies and businesses with their advertising and marketing campaigns.

Data analytics can also help businesses and companies with improving their services and products.

Raw data, as it is, offers amazing potential, but to unlock this, data analytics and associated tools, strategies, and practices are required.

Chapter 3:

Types of Data Analytics

Descriptive Analytics

This leads us to several different data analytics types. As mentioned in the previous section, the process of collecting raw data is the very first step towards using it to draw conclusions, analyze different trends, and get help with making business decisions.

In most cases, this process starts with what we call descriptive analytics. Descriptive analytics is responsible for collecting, analyzing, and describing crucial historical trends in raw data.

The main goal of descriptive analytics is to find answers to the question “What occurred?” Essentially, descriptive analytics measures various forms of traditional indicators such as ROI (return on investment), and this and other historical indicators can be used across many industries and fields.

Instead of making direct predictions as you would think, descriptive analytics is more focused on collecting and summarizing raw data in a more descriptive, understandable, and meaningful way.

Descriptive analytics is essentially used to collect, analyze, and finally describe the important features of raw data collected. To draw meaningful, important conclusions, past data is used.

Drawing conclusions from part data usually revolved around two different data analytics techniques, including data mining and data aggregation.

The raw data is collected before it can be organized with the help of other data analytics techniques. The next step is to identify raw data meanings and patterns, and this is done with the help of data mining.

Over ninety percent of the companies and organizations across the globe turn to descriptive analytics not only because it is the simplest form of data analytics but also because it allows experts in the field generate more insightful data from massive big data chunks.

Descriptive analytics is one of the most commonly used techniques or tools for driving crucial information, and leveraging this type of data analytics allows businesses and companies to decode that inner context and find the reasoning behind their past success or failure.

In this sense, descriptive analytics is extremely helpful when it comes to collecting and extracting the most useful pieces of information through various data mining techniques.

Business intelligence systems commonly rely on descriptive analytics as this kind of data analytics is perfect for analyzing both the historical and real-time data required to extract insights.

Naturally, extracting crucial insights is what helps businesses when devising their future approach.

As previously mentioned, descriptive analytics answers that important question, “What occurred?” For a business or company, analyzing their monthly income or monthly revenues helps them with boosting their efficiency.

Analyzing past data is always involved with descriptive analytics, and this type of data analytics is perfect for juggling data from many different sources to get valuable insights into companies’ past performance.

Descriptive analytics findings generally signal when something is right or wrong without giving reasoning or explanation.

This is why big, successful companies and businesses never settle for just descriptive analytics. Instead, they combine descriptive analytics with other types of data analytics.

Even though descriptive analytics does not provide answers to the question, “Why did something occur?”, it is still the foundation of reporting.

In fact, it is almost impossible to make use of BI (business intelligence) tools without turning to descriptive analytics and addressing those most basic questions, including “What happened and when did it happen?”

Descriptive analytics can be divided into two sub-categories, including canned reports and ad hoc reporting. Canned reports are reports which have already been designed and include important information about specific subjects.

A canned report can be a monthly report which lists performance metrics on businesses or companies’ latest ad efforts sent by ad teams or ad agencies in detail.

Unlike canned reports, ad hoc reports are not scheduled and are usually designed by a team working within a company.

Essentially, ad hoc reports are designed when certain questions must be answered. Ad hoc reports are very important and extremely useful when it comes to collecting in-depth information regarding a certain query.

For instance, ad hoc reports can answer important questions regarding your company’s social media presence and the kind of individuals who visit your company’s social media pages.

Ad hoc reports are also very valuable when it comes to gathering demographic and customer engagement information.

In today’s world, ad hoc reports are generally used to giving a more insightful picture of companies’ social media presence and audience.

If you are looking for descriptive analytics in action, check out Google Analytics, as this is an excellent example of how this data analytics type works.

If you have a website, check out Google Analytics to see everything that may interest you about your website’s traffic and visitors. Google Analytics is also a great tool to see how many individuals visited your site in a certain period or from where they visited your site.

Predictive Analytics

As mentioned in the previous section, companies and businesses combine different types of data analytics to get the best results.

You will rarely come across businesses and companies that only rely on descriptive analytics, and in most cases, descriptive analytics is combined with predictive analytics.

In fact, alongside descriptive analytics, predictive analytics is one of the most commonly used data analytics types. As suggested, predictive analytics use various data analytics tools and strategies to identify different trends, causations, and correlations.

Essentially, predictive analytics is mainly used to draw predictions on possible future outcomes, and to do so, companies and businesses rely on historical or past data. In other words, historical or past data is used to determine what the future may hold.

This means that descriptive analytics are closely related to predictive analytics as both these types of data analytics rely on historical data to draw conclusions.

For predictive analytics, big bulks of data are required. Big bulks of data are collected and carefully mined with the use of different predictive models, and the process ensures making accurate predictions.

Predictive analytics can be broken down into two sub-categories, including statistical modeling and predictive modeling, and these two are usually utilized together.

A company that promotes its products on social media can turn to predictive analytics to determine how its conversion rate associates with its target audiences depending on their income, preferences, and geographic areas.

With this information available, the company can easily turn to predictive analytics to determine the statistics corresponding to one or multiple target audiences and, consequently, predict revenues for each demographic.

For companies and businesses, forecasting what may happen or what is most likely to happen in the future is always extremely important and equally fascinating.

Predictive analytics is there to help companies and businesses with predicting market trends. Predictive analytics is also crucial when it comes to examining and analyzing changes in the behaviors of customers.

When companies know more about future trends and reasons behind their customers changing their habits or behaviors, they can turn to other data analytics tools and strategies and start working on maximizing their revenues and business outcomes.

As expected, predictive analytics is all about forecasting. In most cases, predictive analytics tools and strategies are used hand in hand with crucial insights gathered from diagnostic and descriptive analytics, and these are used together to devise a more efficient business model.

More efficient business models based on the data collected with predictive, diagnostic, and descriptive analytics are used with machine learning tools and advanced statistical models.

While descriptive analytics tells you “What happened,” predictive analytics tells you “what is more likely to happen in the future.”

To answer this question, predictive analytics relies on the crucial findings of diagnostic and descriptive analytics to detect and analyze different exceptions and clusters.

While forecasting or predicting what may happen in the future is just making assumptions or estimates, predictive analytics, when used the right way, has a high accuracy rate.

In fact, the accuracy of predictive analytics models always depends on the data stability and quality of the particular situation, and every single model requires special treatment to get the best results.

As previously mentioned, predictive analytics also relies on a collection of variables or variable data, and these variables are needed to make forecasts of any kind.

The variability of each data component is also associated with what is more likely to happen. For instance, older individuals are more likely to suffer from a heart attack than younger people.

In this case, age is associated or correlates with a higher risk of suffering from a heart attack. In predictive analytics, each component of raw data is compiled together and used to make forecasts.

Needless to say, we live in a world where almost everything is uncertain, and for companies and businesses, the ability to make predictions can make a huge difference between succeeding and failing.

Predictive analytics is very useful since it allows businesses and companies to plan ahead and center their businesses goals around what is more likely to happen in the future.

No matter which business you run, predictive analytics always relies on the relationship between different variables to drive insightful conclusions. If you are running a company that sells stuff online, you may turn to predictive analytics to predict when your sales are most likely to drop.

To draw any kind of prediction, you need to rely on the correlation between your sales figures over a certain time period and seasonality. If you have a predictive model that suggests your sales will go down during the summer months, the next logical step would be coming up with a special summer promotion or campaign to boost those sales.

Using the same information your predictive model uncovered, you want to work on decreasing expenditure to make up for sales going down in summer.

A restaurant owner can also use predictive analytics to forecast how many takeaways he or she is most likely to get on a Sunday night.

Depending on what predictive models conclude, hiring another employee or a delivery driver may be a good idea.

Besides being used to make predictions, predictive analytics is also very handy when it comes to classification. When it comes to classification and predictive analytics, logistic regression is the most used model.

Essentially, logistic regression is valuable when it comes to predicting any kind of binary outcome that is correlated with multiple independent variables.

For example, a company that issues credit cards may turn to predictive models on top of logistic regression in order to classify its customers into two categories, those who are more likely to default and those who are not likely to default.

Using these predictive models and information, the credit card company can easily access which individuals out of all applicants are proper candidates for getting a credit card.

Machine learning and predictive analytics also go hand in hand.

Humans turn to predictive analytics to forecast what may happen in the future and devise predictive models, while machine learning algorithms are capable of identifying different data patterns and automatically making quite accurate predictions.

Even though there are many differences between machine learning and human-led predictive analytics, both can make correct, detailed, and on-target predictions.

However, predictive models can’t be one hundred percent correct at all times, and this is expected. At the same time, predictive models help you eliminate guesswork, and you do not want to guess anything when it comes to your business.

Instead, you want to make your business decisions the right way, and the only way to do so is to examine what has happened and predict what is most likely to happen in the future.

When your business decisions correlate to predictive models, your chances of succeeding and reaching your business goals are much higher.

Diagnostic Analytics

Analyzing historical data may not be as important as predicting what may happen in the future, but it doubtlessly serves a critical purpose for companies and businesses. This is where diagnostic analytics comes in very handy.

Essentially, diagnostic analytics revolves around the process of collecting and examining historical or data that correlates to the past. Diagnostic analytics revolves around examining why something occurred or understanding the reasoning behind something occurring.

Many different diagnostic analytics tools are used, including data discovery, drill down, data mining, and of course, correlations that go hand in hand with other valuable techniques and tools.

While predictive analytics answers the question of “What may happen in the future,” diagnostic analytics is used to give answers to the question “Why something happened?”

Just like other types of data analytics, diagnostic analytics can be broken down into several sub-categories, including drill downs and query, discover, and alerts.

More specifically, drill downs and queries are generally used to gather more details from data analytics reports.

For instance, a company that makes sales can turn to diagnostic analytics to examine why it closed significantly fewer business deals the previous month. Drill downs performed by the same company could potentially show less than usual workdays due to a national holiday.

Diagnostic analytics also relies on discoveries and alerts, and as suggested, these tools show potential issues that may happen in the near future. Diagnostic analytics is also crucial when it comes to gathering information about potential employees and finding the one who is the most qualified.

Moreover, this type of data analytics is the second most important data analytics form as it assists companies and businesses with solving the most crucial challenges by providing answers to the questions "If something is occurring," Why it is happening," and "What could be the main reason behind that happening?"

Needless to say, diagnostic analytics is a major part of business intelligence or any business that relies on dashboards with business intelligence involved in some way.

Turning to diagnostic analytics helps companies and businesses drill down within the raw data to uncover the major factors and reasons for something crucial happening within the organization.

In most cases, diagnostic analytics tools and strategies are used alongside tools and strategies related to descriptive analytics.

With these tools and strategies combined, companies and businesses can easily find the architecture and relations of the raw data and do super quick comparisons to generate the most reliable and most informative decision models.

Historical data can be easily measured or compared against other collected data to answer the most important question of “Why something occurred.”

For instance, a company may turn to diagnostic analytics and historical data to examine how employees can drill the gross profit and sales down to sub-categories to find out why employees have failed to reach their business goals in the first place.

Diagnostic data analytics is widely used in the healthcare industry as gathering historical data allows the examination of the influence of certain medications. In these cases, several specific filters are used, like prescribed medication and diagnoses.

Essentially, diagnostic analytics gives extremely valuable and detailed insight into a specific issue. A company of any kind and size should have crucial information always at its disposal, and this is where diagnostic analytics does the job.

A wide range of diagnostic analytics tools and strategies are used, and their main goal is to empower data analysts and data science professionals to drill down deeper into collected raw data and identify the main issue.

Such analyses are possible when business intelligence dashboards incorporate time-series data with appropriate filters.

Diagnostic analytics is also crucial when it comes to identifying and responding to various kinds of anomalies that may happen within the data you have collected.

When there is any kind of issue, such as dropping sales, companies and businesses perform diagnostic analytics to discover why their sales have dropped.

As expected, to get to the bottom of the issue, data analysts need to identify different chunks of data or extra data sources that can potentially provide further details and insight into why something has happened.

In the case of sales dropping, data analysts may use drill down techniques to find out that there has been a very healthy volume of site visitors and even a solid amount of add to cart behaviors, but just several visitors proceeded to make a purchase.

In this case, further inspections are required, and data analysts may find out that numerous visitors to the site decided to leave after coming to the point when they need to enter their delivery address.

This instantly suggests that there was some kind of issue revolving around the website's address form. This may be anything from the address form failing to load correctly on certain devices or the address form taking too long.

With diagnostic analytics, it is all about digging and looking for answers in different places before eventually coming across the most logical reasoning behind issues emerging or data anomalies.

Therefore, diagnostic analytics is used to fix problems, but there is so much more to diagnostic analytics than you may think.

For instance, diagnostic analytics is very handy when it comes to examining positive business trends.

A company may enjoy a major increase in sales, and turning to diagnostic analytics can explain why this happened.

When data analytics drill down to examine why positive trends have happened, they may find out that the website's traffic increased massively following a big celebrity post on his or her social media account about the company's product or service.

When discussing diagnostic analytics, there is a wide range of tools to use, including filtering, regression analysis, probability theories, and of course, time-series analysis.

Lastly, there is no diagnostic analytics without descriptive analytics, as you need to know what has happened in part to find out why that happened in the first place, whether it be something that positively or negatively affects your business.

Prescriptive Analytics

Big data and artificial intelligence combine to form prescriptive data analytics models. Prescriptive analytics provides answers to the question, “What is the very best course of action?”

To get these answers, data analytics uses prescriptive models, big data, and artificial intelligence to predict certain outcomes and finally identify the best actions.

Prescriptive analytics can be divided into two sub-categories, including random testing and optimization.

Using different prescriptive models, you can easily test the appropriate variables or add some new variables that offer higher chances of enjoying a positive outcome in the future.

In addition to giving answers to the “What is the best route to take?” question, prescriptive analytics can also assist with exploring new business models or options.

Prescriptive analytics also comes after predictive analytics. It helps businesses and companies to generate prescriptions based on what predictive models suggested.

In other words, prescriptive analytics can solve issues depending on the gathered factors and variables from the raw data collected.

While it is not always possible to predict the most accurate inputs, prescriptive analytics most certainly helps with finding cases of issues emerging in the first place.

Companies and businesses use prescriptive data analytics models to examine the most possible outcomes and devise the best business actions that will most likely lead to maximizing business outputs.

Essentially, you can think of prescriptive analytics as a data analytics process used for business optimization since prescriptive data analytics models provide crucial insights regarding what a business should do to solve an issue.

As suggested, the main purpose of this type of data analytics is to tell companies or businesses what business actions to take to eliminate certain issues happening in the future or to take full advantage of positive trends.

For instance, a company may turn to prescriptive analytics to examine future business opportunities based on its sales history and customer analytics from the last six months.

Just like other types of data analytics, prescriptive analytics relies on many different technologies and tools, including machine learning algorithms.

To make accurate prescriptions, prescriptive models rely on historical data alongside external data, and data analytics who use prescriptive analytics tools usually compare the required business action efforts with the expected values of those actions.

There is no prescriptive analytics without other forms of data analytics to give insights into what actions should be taken.

All prescriptive models need a complete understanding of what has occurred in the past and why it occurred. Prescriptive models also rely on “what may happen” variables and data analysis to help companies and businesses determine the very best action to take.

This means that prescriptive analytics usually requires a series of actions rather than a single one.

A great example of prescriptive analytics in action is a mobile app that suggests the best route to take.

To give advice, the app takes in different variables, including current traffic conditions, the speed at which you can travel on different roads, and of course, the distance of every route from your destination.

With prescriptive analytics, you can also gather important information about capitalizing on emerging trends, steps to take to avoid issues in the future, or simply the best business actions to take to avoid coming across issues that may have happened in the past.

When compared to those three previously mentioned data analytics types, prescriptive analytics is much more complex as it involved machine learning, various kinds of algorithms, computational modeling techniques, and different statistical models.

All of these techniques and tools are needed as prescriptive data analytics must consider every single possible decision-making pattern a business or company may take and all of the outcomes that may happen.

As complex as it is, prescriptive analytics is extremely important and valuable as prescriptive models enable you to examine how every single combination of decisions and conditions impacts or affects the future of your business.

Prescriptive analytics and associated models also allow you to examine the impact of any decision you make on your business and its future.

With all possible outcomes and scenarios provided by prescriptive models, companies and businesses can easily decide the best pathway to take for the best outcome in the future.

Just like mentioned previously, a great example of prescriptive models and analytics in action is using a traffic app or Google Maps.

When suggesting which route you should take to get from A to B, Google Maps takes into consideration various forms of transportation, including driving, walking, or taking a bus.

Google Maps also takes into consideration the traffic conditions at the moment, and all possible road works to calculate the best route to take to get from point A to point B.

Prescriptive models are similar. They identify all of the possible outcomes or routes a business or company may take to reach its business goals and finally determine the very best route or action to take.

Prescriptive models work in any industry. In the healthcare industry, prescriptive models are used to manage the population of patients by measuring the number of individuals struggling with obesity.

These kinds of prescriptive analytics models support all sorts of filters, and in the healthcare industry, these filters can be anything from LDL cholesterol levels to diabetes. The main goal is to determine the best and most accurate treatment based on all those different factors.

As it is, prescriptive analytics relies on the findings of predictive analytics models to prescribe or determine further actions based on predicted, desired outcomes, and such models are crucial to companies striving to reach their business goals as fast as possible.

Prescriptive models constantly gather new data or learn through different feedback mechanisms, and these feedback mechanisms are responsible for analyzing different actions that could be taken and the potential outcomes of these actions.

Different outcomes and different routes must be considered so that prescriptive models can recommend or prescribe the best and optimal solution.

Prescriptive analytics models can easily examine all of the most crucial factors and criteria to ensure the most likely outcome achieves the expected business goals.

Machine learning, artificial intelligence, and even neural network algorithms are commonly employed to enhance prescriptive data analytics models.

More specifically, machine learning, artificial intelligence, and neural network algorithms assist prescriptive models with making accurate suggestions that correlate with business goals, nuanced patterns, limitations in place, and other influencing factors.

Needless to say, knowing which business actions bring the highest chances of succeeding is one of the biggest advantages, and it truly makes the difference between reaching business goals and not reaching them.

All things considered, it’s no wonder why prescriptive analytics has one of the biggest roles in companies of all kinds in any industry.

Cognitive Analytics

The most advanced and complex form of data analytics is cognitive analytics. This type of data analytics combines many different highly intelligent techniques and technologies, including deep learning, machine learning, and artificial intelligence,

The main purpose of cognitive analytics is to process huge chunks of data and deliver highly technologically advanced results that mimic human thinking.

Essentially, cognitive analytics is inspired by human thinking, and it mirrors humans' cognitive abilities to make advanced applications more effective and smarter over time.

Chapter 4:

The Evolution of Data Analytics

The most commonly accepted definition of data analytics is collecting, gathering, and analyzing data related to a business or company's part performance, possible future performance, and everything else in between.

Thanks to amazing data analytics tools and strategies, companies and businesses can devise valuable recommendations and insights to boost their companies' future performance. The accuracy and quality of data analytics tools determine the future results.

Essentially, three major functions are needed in order to be able to devise crucial business solutions, including statistics, math skills, and technological skills. These are the three main pillars, and they have been the fundamental approach in data analytics since the very beginning.

Over the past twenty years, data analytics has evolved massively, and most certainly, it will continue to evolve with new technologies emerging.

What has changed in the world of data analytics mainly corresponds to the changes made in the fields of statistical and technology techniques, and these have been changing rapidly.

Data analytics requires that companies, businesses, and all data scientists involved refine their already acquired skills and constantly work on adopting new ones. Now, let’s see how this scientific field has evolved.

Data Analytics Then

Back in the day when companies and businesses had just started using data science and related techniques and skills, the primary goal was to locate more reliable and more accurate business solutions than the standard business heuristics techniques can provide.

Back in the day, companies and businesses that just started turning to data science and data analytics have tried their best to keep the obtained solutions as simple as possible to avoid overwhelming their customers.

Naturally, the choice of techniques, skills, and technology was also kept very straightforward and simple. When keeping it as simple as possible, the companies and businesses that started using data science were able to easily implement businesses solutions.

The choice of data science techniques and technologies that was kept simple also allowed easier consumption and implementation, and the same was true for math and statistical models that, when kept simple, allowed easier explanation and development.

The earliest usage of data analytics tools and techniques was explanatory rather than predictive, which is the case today.

This naturally impacted the selection of techs and tools used back in the day. In those early days, the availability of data analytics tools and techniques was quite limited, and not many people were working as data analysts.

Back in the early 2000s, data processing techniques emerged on the scene, with the SAS company being one of the first to embrace them.

More specifically, SAS embraced data processing to build powerful backend data for modeling and reporting. Several other companies such as Oracle, Teradata, and IBM also used data processing techniques for similar purposes. In its early days, data analytics also incorporated predictive modeling techniques.

These were quite commonly used by companies and businesses as they allow the building of quite simple and straightforward statistical models such as those revolving around linear regression.

Besides linear regression, early predictive modeling techniques used logistic regression, Naïve Bayes, CHAID, alongside different time series models including ARIMA, smoothing, and ARIMAX. At the time, powerful toolkits, including IBM CPLEX, emerged on the scene as well.

Data analytics reports were in most cases developed and delivered on VBA and Excel, while large companies and businesses used Cognos and other visualization tools.

With this in mind, the skillset needed back in the day was quite limited and narrow, and companies and businesses usually hired individuals with degrees in statistics and proceeded to train them in VBA, SAS, and SQL programming.

Data Analytics Now

When it comes to data analytics and data processing today, Python and R remain the most crucial technologies that data scientists utilize to analyze raw data and devise accurate, reliable business recommendations and solutions.

Data Processing

Both Python and R are powerful open-source tools, and they have always-evolving libraries. Both of them are very easy to integrate with raw data platforms, and both are equally versatile and competent when it comes to processing massive chunks of data for many different purposes.

However, many data analysts prefer using R, especially when the main goal is to derive valuable insights for a company or business using predictive modeling or exploratory analysis.

When it comes to Python, this programming language is mostly used when crafting applications that come packed with analytics engines.

Then, there is the field of disrupted processing, and the most commonly used disrupted processing framework was developed by Hadoop and Spark or Apache Open Source Projects.

Hadoop emerged on the scene back in the early 2010s, and it is very popular even today. Back in the day when it just emerged on the scene, its capabilities were quite limited, especially when compared with popular relational database systems.

Nonetheless, it became very popular since it is very affordable, extremely flexible, and can easily scale. Even more importantly, Hadoop brings numerous other benefits thanks to PIG, Hive, and other enablers.

Today, Hadoop is one of the main options for many companies whose business operations revolve around data-driven analytics.

Hadoop is one of the pioneers when it comes to data analytics and data processing, but its performance is lacking when it comes to iterative data analytics, machine learning, real-time processing, and predictive modeling.

This is why Apace Spark was introduced to perform better. This is a powerful in-memory data analytics framework that, as suggested, can hold the crucial data in its memory and still perform full data analytics operators.

Naturally, this makes Apache Spark much faster than Hadoop. Apache Spark also integrated with many different programming languages, including Japan, Python, and Scala.

Predictive Modeling

When it comes to data analytics and predictive modeling today, most data analytics tools and technologies that revolve around predictive modeling rely on machine learning.

In the past several years, the amazing field of machine learning has gone through advancements, with companies, businesses, and data analysts relying on more advanced and more complex machine learning techniques.

Today, machine learning technologies and techniques are much more powerful, reliable, and accurate than the standard logistics and linear regressions.

More specifically, today's machine learning models are capable of uncovering extremely complex data parameters, all sorts of variable interactions, as well as non-linearity, and much more.

Machine learning models that are used today also provide more accurate results, especially when using Random Forests, Support Vector Machines, Parametric GAMs, XGBoost, GBM, Restricted Boltzmann, Latent Dirichlet Allocation, Maximum Entropy, and other advanced techniques.

When discussing advanced machine learning techniques, these are used for improving forecasting accuracy not only by bigger but various kinds of smaller companies and businesses.

For sentiment and classification, companies mainly rely on TensorFlow, Maximum Entropy, and Latent Dirichlet Allocation.

Supervised machine learning techniques include Multilayer Perception, Support Vector Machines, Parametric GAMs, Random Forests, XGBoost, and GBM.

Unsupervised machine learning techniques include Restricted Boltzmann Machines, Autoencoders, Matric Factorization, and K-nearest Neighbors.

For optimization, businesses use Tabu Search, Simulated Annealing, and Genetic Algorithms, and various other techniques are used for blending.

All of the machine learning techniques mentioned above are quite complex when compared to what was used before.

However, they can’t compare to the techniques related to deep learning, which today, power loads of different applications related to computer vision and artificial intelligence.

At the same time, many different deep learning techniques such as convolutional networks can be easily trained on raw but structured data to solve the most common business operational cases, but they are still most commonly used in fields of image feature learning, recognition, and image classification.

One of the biggest reasons why deep learning technologies will not make it into standard companies and businesses is the fact that deep learning models require more resources to develop and implement.

Deep learning models also usually require GPUs and for many businesses, developing such complex models is not worth the trouble unless justified by their return on investment or significantly increased accuracy.

At the same time, several non-tech companies and businesses have embraced complex deep learning models for developing non-artificial intelligence predictive models, which, even though they require more resources, bring a bigger return on investments.

Visualization Technologies

Today, many companies and businesses still use visualization products from the past, such as ElasticSearch Kibana, QlikView, and Tableau. Some companies have already adopted more technologically advanced visualization techniques such as Angular and D3, which make great low-cost options.

With almost all low-cost options, companies can easily develop very visually appealing, easily customizable, and highly interactive networks and mobile dashboards, and the development of these is quite fast with numerous modules and reusable components.

Thanks to the amazing advancements in algorithm and technology industries, companies and businesses began looking for data analysts with basic programming skills, open-minded thinking, and basic statistics and mathematic skill sets.

Individuals that have these skills are extremely flexible when it comes to learning new skills, embracing new technologies, and very agile at identifying and solving business issues. These individuals can also easily master Python or R, and this is what companies and businesses need and want.

Data Analytics in the Future

Considering current trends and everything that has been done in the future, the future of data analytics looks very promising.

Given the current trends and the latest use cases that companies and businesses have already embraced, the data analytics industry will be focused on real-time development, automation, big data, and embedded data science.

In the near future, there will be a greater need for new data analytics paradigms in the fields of programming ecosystems, database handling, and newer data analytics algorithms.

It will also become extremely important for data to consistently learn about the merging tools, techniques, and trends.

To stay in the game, data analysts will need to embrace new technologies and refine their skills.

Besides those technologies that are already in use, Spark Ecosystem, Pyspark, Python, and Scala will probably remain the most common choice of data analytics technologies in the next five or more years.

Then, there is Google's programming language Go, and in the future, this is expected to see more and more data analysts using these programming languages besides those that are currently used.

The Go programming language has many benefits, including enabling data analysts to craft data science applications, services, and codes that are production-ready.

With single-threaded languages, the codes are not production-ready, and they usually require a huge effort to transition data analytics models from machines into production platforms.

More specifically, testing must be done alongside error handling and much more. This is where Google's Go programming language comes in very handy.

It has already performed amazingly, and it allows data analysts to easily craft very efficient, scalable applications right from the start, and this is not possible with Python or other similar programming languages.

Moreover, the ability of Google's Go programming languages to handle and report all sorts of errors as they happen boosts the integrity of applications and allows easier maintenance over time.

When it comes to the thriving algorithm field, in the future, we can expect to see more adoption and development, especially in the field of deep learning.

Companies and businesses will probably further embrace deep learning. While today, deep learning is mostly used in the fields of image feature analysis, image recognition, and artificial intelligence, in the future, we will probably see companies and businesses using deep learning models for common business solutions.

We will probably see more development and adoption among companies that use boosting-based algorithms.

Traditionally, these companies and businesses have relied on machine learning models, but thanks to the latest technological advancements including LightGBM, and XGB, we will see companies and businesses using improved boosting-based algorithms.

The big data field would also see more development and adoption, and this will require more versatile skillsets.

All in all, the data analytics industry has been and will continue to evolve at a rapid pace, and the key to staying relevant will be hiring experts who are not afraid to explore and embrace new technologies.

With new technologies, the data analytics industry changes, evolves, and becomes even more powerful.

Chapter 5:

What is Big Data?

Big data is usually defined as a massive combination of unstructured, structured, and semi-structured data that companies and businesses mined as well as used in predictive modeling, machine learning, and other c0mplex data analytics projects.

Powerful systems which store and process such massive big data collections have become one of the most crucial features of data management departments in many companies.

Big data always refers to the large data volume of both unstructured and structured data; however, the amount of data is not as important as the raw data that matters to their business performance.

As suggested, big data can and should be analyzed to get valuable insights that will lead to making better and more informed business moves and decisions, and this is why big data analytics matters.

Data that is extremely huge, very complex, and constantly changing is usually referred to as big data. As huge as it is, big data can't be analyzed using standard methods.

While the act of storing and analyzing huge chunks of data has been around for quite some time, the big data concept gained its momentum back in the 2000s.

At the time, Doug Laney formed the commonly accepted big data definition. According to his definition, big data is always the very large data volume present in multiple environments.

Big data also includes a wide variety of different data types stored and analyzed across different systems.

The same definition of big data also explains that the velocity at which raw data is generated and processed is much faster than with other types of data. Therefore, the volume, variety, and velocity form the three V’s of big data.

Big Data Volume, Velocity, and Variety

Organizations and companies collect raw data from many different sources such as Internet of Things devices, their past transactions, social media platforms, industrial equipment, and much more.

Back in the day, storing such large bulks of raw data was a huge problem. However, Hadoop and similar, cost-effective storage platforms have made it easy to collect and store huge volumes of data.

With the advent of the IoT, data streams occur at high speed, and companies must handle huge data collections as quickly as possible.

For this reason, smart meters, sensors, and RFID tags are making it possible to analyze big data chunks in a timely manner.

Raw data comes in all formats. It can be numeric data, structured data, or unstructured data. Data also comes in the form of financial transactions, stock ticker data, audios, videos, emails, and text documents.

Big Data Variability, Value, and Veracity

Besides those three major V’s of big data, including the volume, variety, and velocity, other V’s regarding big data have been included recently, such as viability, value, and veracity.

Besides the increasing varieties and velocities of data, data streams are quite unpredictable. Data streams tend to change quite often, and data collections vary greatly.

This is why it can be challenging for businesses and companies to collect and analyze big data chunks, but this must be done to examine the latest trends.

This also needs to be done to manage seasonal, daily, and certain event-activated data peaks.

When it comes to big data veracity, this concerns data quality. As raw data is collected from many different sources, it is crucial that companies and businesses match, link, clean, and transform big data chunks across different data systems.

More specifically, companies have to connect as well as correlate data relationships, data linkages, and hierarchies. If this is not done, raw data collected can easily get out of control, and it would be extremely hard to make any reasonable, meaningful connections between different data types.

Using big data alongside predictive modeling and other data analytics strategies and technologies is one of the major keys to truly understanding how different products and services are made, what they bring, and how they work.

Every industry is very competitive, and staying on top of the game means everything to companies and businesses.

The importance of delivering only high-quality, premium products and services has never been so great as it is today, with so many different companies and businesses doing the same.

Why Big Data Matters?

As mentioned in the previous section, the amount of raw data that companies and businesses collect does not matter as much as what they do with said raw data.

Analyzing big chunks of data collected from many different sources can give very valuable information about time and cost reductions, optimized products and services, and new products and services, as well as smart decisions.

When big data is effectively combined with data analytics tools and strategies, many different tasks leading to business optimization can be accomplished.

More specifically, collecting and analyzing big data can help companies and businesses determine major causes of issues, defects, and failures almost in real-time.

Big data, alongside powerful data analytics tools, can also help companies and businesses analyze their customers' purchasing habits, recalculate risk portfolios, and even detect fraudulent activities and behaviors before such activities negatively affect their businesses.

The way businesses generate and analyze big data to derive valuable insights from it has completely changed the way we use information.

In today’s world, big data is everything when it comes to building powerful, efficient businesses optimization strategies. Big data is also one of the most crucial factors when building powering data analytics ecosystems.

Big data is a huge deal for all industries simply because it is needed to identify hidden patterns and, finally, to find important answers to crucial business performance-related questions.

In today’s world, the more high-quality raw data you have, the more positive results you get. Before companies can actually put big data into action and expect it to start working for them, they need to examine how big data chunks flow among different locations, users, owners, systems, and sources.

How Does Big Data Work?

So, the main question is how does big data actually work? To put big data into action so it can work its magic, there are five crucial steps to take, and these five steps are always the same regardless of you dealing with structured, unstructured, or semi-structured data.

The very first step is to set a business strategy or a plan, and this plan must be carefully designed to help companies improve the data they collect, store, share, and finally use.

Having an effective big data strategy can make a huge difference when it comes to companies' and businesses' future performances. Having an effective big data strategy is also required if companies want to set the stage for success.

When devising a big data strategy, it is extremely important that you consider your future and current business goals as well as future and already existing technology initiatives.

With this in mind, big data should be treated like any other extremely valuable asset businesses acquire rather than a meaningful byproduct.

Once companies have devised their big data strategies, they need to start working on uncovering the major sources of big chunks of data.

As already mentioned in this section, big data comes from many different sources such as IoT and all sorts of other devices connected to the internet that stream into complex IT systems for smart cars, industrial equipment, wearables, and much more.

When dealing with big data, the best option is to analyze it in real-time as it arrives. This approach, in particular, requires instantly deciding which pieces of big data are worthy of keeping and which are not.

Those that will be kept will also be analyzed further. Big data also comes from social media platforms such as Instagram, YouTube, and Facebook.

Needless to say, social media platforms feature vast amounts of big data, and this comes in the form of voice messages, videos, images, text messages, and more.

This kind of big data is extremely useful when it comes to sales, marketing, and support functions.

It is generally semi-structured or unstructured, so analyzing it takes more time than when dealing with structured big data streams.

Big data collections that are accessible to everyone come from a variety of different sources, such as the EU Open Data Portal. Other big data collections may come from different cloud data sources, data lakes, customers, and suppliers.

Once companies have identified the source of big data, they work on accessing those massive chunks of raw data.

Fortunately, very modern data analytics systems boast the flexibility, power, and speed needed to quickly, in real-time, access massive big data collections packed with various kinds of data.

In addition to working on being able to easily access big chunks of data, businesses also require effective methods that will allow them to ensure the high quality of data collected and easy integration of the data collected.

Companies also need to work on storing big data collections the right way so that the raw data they have collected can be analyzed correctly.

Some companies and businesses store data in standard data warehouse places, while others prefer storing big data in more complex data warehouses places.

Either way, when it comes to storing massive data collections, there are great options that are cost-effective such as Hadoop.

Once companies and businesses have stored big data collections in traditional warehouses or cost-effective cloud solutions, the next step is to analyze them.

Thanks to amazing technologies such as in-memory analytics and grid computing, companies and businesses can easily determine which pieces of big data are worthy of further analysis and which are not.

In addition to in-memory analytics and grid computing, companies and businesses can use other big data analytics options, and many of these let me determine the importance of each piece of information before investing time and energy into analyzing it.

No matter which approach companies and businesses take, analyzing big data is the best option for gaining extremely important insights and value that are needed to boost future performance.

Big data analytics, alongside the complex data analytics models and algorithms associated with artificial intelligence and machine learning, most certainly feed all sorts of technologically advanced data analytics projects.

Once big data collections have been analyzed, companies and businesses explore the value and quality of big data and draw insights and conclusions to make appropriate decisions.

The decisions they make are naturally data-driven and correspond to the evidence that emerges as big data is analyzed. Instead of relying on their gut instinct, companies rely on the value of big data.

There are numerous benefits to making data-driven business decisions. Companies and businesses that rely on big data generally perform better, generate bigger profits, and operate more successfully.

Chapter 6:

Data Mining

Data mining is closely related to data analytics and big data. In the simplest terms, data mining is a process companies employ to collect or extract meaningful data from big data chunks or raw data collections.

You can also think of data mining as a process of analyzing massive data collections to uncover those pieces of data that are truly valuable to businesses to mitigate risks, solve issues, and seize future businesses opportunities.

In computer science, data mining is a process that involved uncovering useful, interesting, and meaningful relationships and patterns in massive data collections.

Using different data mining techniques, tools, and strategies, companies and businesses can analyze various pieces of information from many different sources and effectively summarize them into only meaningful, useful data.

With meaningful, useful data carefully collected, companies can use it to decrease their future costs, increase their future revenues, and most importantly, find relationships or patterns among dozens and dozens of large data collections.

Data mining is just one of many extremely important data analytics tools and strategies used for analyzing and reading data.

With the help of data mining tools and techniques, data analysts can easily access big data from different angles, work on categorizing it, and finally summarizing it once all of the crucial patterns and relationships have been identified.

There are many different goals of data mining, but the primary one is always discovery and predictions.

Data mining processes revolve around looking for systematic correlations between different variables, looking for distinct patterns, and finally validating those important findings with the help of newly created patterns.

The Process of Data Mining

The process of data mining is usually divided into five steps, and the very first step revolves around extracting, modifying, and finally loading data into some kind of big data warehouse.

The next step required storing and managing the extracted data in a database system, while the third step includes providing big data access to IT professionals and data analysts.

The fourth step includes analyzing the extracted data using some kind of application software, and finally, the fifth step requires presenting the extracted data in a table, graph, or some other useful format.

Therefore, the process of data mining is quite simple. The first stage always starts with data preparation, including cleaning out extracted data, modifying data, and uncovering subsets of data sets among many different variables.

The next step starts with identifying important variables using exploratory data analytics tools alongside statistical and graphical methods.

Data Mining Requirements and Techniques

If you will be mining big chunks of data soon, you need to make sure that your big data warehouse is fully ready, as you need a place to store all those massive pieces of data.

You also need to learn more about data mining tools to find those that you are comfortable with using. You also want to make sure that the big data collections you will be analyzing can be integrated with different data analytics systems.

Equally important is investing time and energy into examining different types of data mining techniques.

Every single company operational today can collect huge bulks of big data even if it does not have a big budget.

However, the main goal is not collecting as much data as possible but being able to compete by extracting everything that is meaningful and valuable from raw data, and this is where different data mining techniques come in very handy.

Some of the most commonly used data mining techniques include exploratory, predictive, diagnostic, descriptive, prescriptive, mechanistic, inferential, and causal data mining.

The main goal of the descriptive data mining technique is to uncover and summarize what has occurred in the near past within a company or organization.

More specifically, this data mining technique is used to analyze the content within the raw data extracted to examine what has occurred.

When compared to other commonly used data mining techniques, the descriptive data mining technique is the most important when it comes to solving issues that occurred within a company or business in the past.

Then, there is the diagnostic data mining technique, and as suggested, this technique is used to examine why something occurred in the past.

To do so, companies and businesses combine different tools and strategies to extract the main causes behind both positive and negative trends.

The prescriptive data mining technique assists companies and businesses with gathering meaningful insights.

Uncovering any kind of meaningful insight using this data mining technique helps companies and businesses devise business plans that will be data-driven.

The exploratory data mining technique is used to identify and classify raw data collected to uncover crucial, important features that other data mining techniques can’t uncover so easily.

More specifically, this data mining technique is used to understand where and why something in the past occurred. The exploratory data mining technique also examines all sorts of variables from different environments, and knowing these makes it easier to make future decisions.

The predictive data mining technique uses the collected raw data with machine learning and statistical models and helps companies and businesses with forecasting that may happen in the future.

In other words, the predictive data mining technique is used by companies and businesses when they want to determine the likelihood of certain future outcomes happening according to the collected historical data and past trends.

When this data mining technique is used the right way, it helps companies and businesses take the right steps towards making their business goals come true.

Then, there is the mechanistic data mining technique, and this one allows data analysts to easily examine and truly understand variables, protocols, and procedures that may happen in the near future as an outcome of changing crucial factors.

The causal data mining technique allows data analysts to examine what event is most likely to occur in the near future with only one crucial variable changing.

The inferential data mining technique is more complex since it usually takes into account different theories with smaller data samples. With the use of smaller data samples alongside general theories, data analysts can analyze bigger chunks of data.

Data Cleaning

As suggested, data cleaning, data scrubbing, or data cleansing is one of the most crucial steps taken by companies when they want to make appropriate data-driven business decisions.

Data cleaning refers to the process of removing or fixing incorrect, corrupted, incomplete or duplicate data within a big data collection.

When extracting big chunks of data from many different sources, data can easily be mislabeled or duplicated, and this is something data analysts deal with quite commonly.

If the data you have extracted is mislabeled, incomplete, or incorrect, algorithms and outcomes will be unreliable even if everything looks fine.

The data cleaning process varies from one dataset to another, so data cleaning steps also vary. However, one step is crucial no matter which dataset you are dealing with, and that is establishing a data cleaning template.

Once you have your data cleaning template, you use it every single time to remove chunks of data that are incorrect, corrupt, or mislabeled.

Data cleaning should not be confused with data transformation. Data transformation refers to the process of modifying data or converting data collection from one structure or format into another.

To clean the data you have extracted, the best option is to start with removing irrelevant or duplicate observations, as duplicated most commonly occur when collecting data from different sources.

Data cleaning also includes fixing any kind of structural errors that may occur and filtering unwanted outliers. Handling incomplete or missing data is also one of many valuable data cleaning steps.

You can’t afford to ignore incomplete or missing data as many data analytics algorithms can’t work effectively with missing data.

Once you have fixed structural errors, filtered unwanted data outliners, and removed irrelevant observations and duplicates, you finish the data cleaning process by making sure that the data you have extracted and analyzed makes sense for your business goals.

You also need to make sure that the data you have extracted follows the regulations and rules within its field and that it helps you with reaching your business goals.

The bottom line, to have high-quality data that can provide valuable insights, removing what does not belong is crucial, and to do so, you clean your data.

Data Visualization Tools

Data visualization refers to the process of graphically representing the raw data you have extracted. There are many different data visualization tools you can turn to include maps, graphs, and charts.

Data visualization tools offer an easily accessible way to truly understand patterns, outliers, and trends in raw data, and knowing these doubtlessly helps with forming appropriate, data-driven decisions that will boost your business' performance in the future.

In addition to graphs, maps, charts, and tables, companies and businesses also use a wide range of other methods to visualize extracted data, including dashboards, infographics, bar charts, area charts, bullet graphs, bubble clouds, matrix, radial trees, histograms, heat maps, timelines, and much more.

Cluster Analysis

Clustering is one of the most used and most popular data analytics classification techniques, as it saves a lot of time and energy when dealing with big chunks of data.

In the most basic form, clustering refers to a machine learning method that is used to group and identify similar patterns and relationships within bigger data collections.

Clustering is an unsupervised data analytics technique, and it is usually used when big data chunks need to be classified into easily manipulated and understood structures.

Clustering analytics categorizes or groups different instances or structures within the same dataset that are similar or related to each other in some way.

Structures that are not similar at all are also grouped together using the same data analytics technique of clustering.

Clustering data instances that are related to each other is one of the most popular machine learning unsupervised techniques.

It is performed without supervision since there is no right or wrong outcome. When it comes to representing similar or associated data strictures using several different methods.

Some of the most popular clustering analytics methods include density-based, density-based spatial, ordering points, hierarchical density-based spatial, fuzzy, and hierarchical clustering analysis.

Needless to say, there are numerous benefits to using one or more of these techniques. When dealing with massive chunks of raw data, you want to be able to uncover different patterns and relationships between different data structured.

Different clustering techniques and methods also allow you to outline and distinguish different data structures, especially if they have not been visible before.

When dealing with large chunks of business data, we recommend turning to K-means algorithms, and these are the most widely used.

The K-means algorithm is a statistical clustering analysis method that involves taking a look into the center cluster point and minimizing the distance from the center to other parts of the same cluster.

There are also other clustering techniques such as neural networks, but K-means are more used simply because comparing structured clusters is never an easy task.

There is no right or wrong number of structured clusters, and K-means algorithms are the simplest and easiest to implement. K-means are also very efficient.

There are also some downsides to using K-means, and one of them is being unable to identify clusters that are not hyper-spheres. When using this clustering technique, you also need to specify a beginning value of K.

If you are looking for a more advanced clustering analysis technique, then turn to neural networks.

Conclusion

Becoming a data analyst is a wonderful career if you are ready to constantly learn new things and refine your skills. Pursuing a career in this field, in particular, is a great option as there is a high demand for data scientists and data analysts out there.

If you have the right set of skills and the needed knowledge, you will have no issues with finding a job. Data analysts also earn more than average per year, and this is something to consider.

Data analytics is evolving and will continue to evolve in the future with new technologies emerging on the scene. Data analytics will always be there as businesses need to make sense of raw data.

Data analytics tools and technologies allow companies to make better, more informed, data-driven decisions, and only data-driven decisions can help them succeed in the long run.

The applications of tools, strategies, and techniques related to data analytics are truly endless, with every single company relying on those tools and techniques every single day. With more and more data being extracted daily, data analytics is here to stay.