Leibniz FH

Research Paper Series
Nr. 10 | Dezember 2023

Data Science Essentials for Business

Exploring Analytics and Data Scientists' Contributions

Autor:
Mohsen Afsharian

Leibniz-Fachhochschule

Autor:

Mohsen Afsharian
Leibniz-Fachhochschule
mohsen.afsharian@leibniz.fh.de

ISSN 2511-7491

Redaktion: Robin Christmann

Leibniz FH
University of Applied Sciences

Research Paper No. 10

Data Science Essentials for Business: Exploring Analytics and Data Scientists' Contributions

MOHSEN AFSHARIAN

Zusammenfassung / Abstract:

This paper presents an overview of data science, exploring the role of data scientists and the methods they employ to investigate analytical problems within the field of business administration. It discusses the responsibilities of data scientists, emphasizing their multidisciplinary nature and the importance of technical skills, domain knowledge, and effective communication. The paper outlines commonly utilized methods by data scientists, drawing from the fields of computer science, statistics, and operations research. It illustrates the practical application of these methods in addressing analytical problems, including descriptive analytics, predictive analytics, and prescriptive analytics. The paper also highlights potential pitfalls and challenges encountered in the analytic process, such as errors in problem formulation and interpretation of results.

Schlagworte / Keywords: Data Science; Data Scientists; Analytics; Business Administration

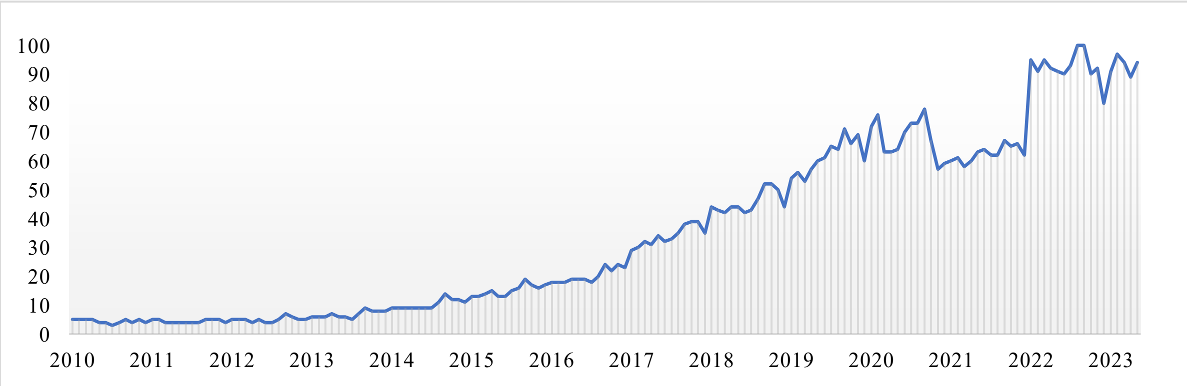
1. Introduction

In our modern times, terms like "data-driven decision-making", "big data", and "data science" have become a regular part of our everyday conversations. The increasing availability and easy access to vast amounts of data, coupled with advancements in technology and computing power, have created a wealth of opportunities for leveraging data in various fields and industries. Data has permeated every sector, from education to business, prompting us to explore essential questions:

- What exactly is data science?
- Who are the key individuals driving its development?
- why is being a data scientist often referred to as "*The Sexiest Job of the 21st Century*" as aptly coined by Davenport and Patil in their article published in Harvard Business Review in 2012 (see Davenport and Patil 2012)?

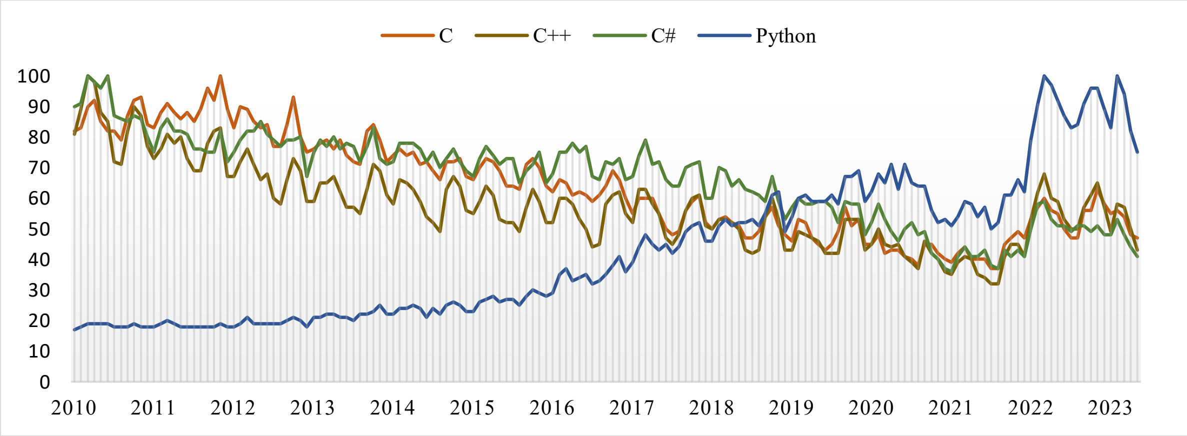
These questions highlight our shared enthusiasm in understanding our data-centric world and the growing fascination with data science. This interest is supported, e.g., by the evidence we observe in Google search trends as depicted in Fig. 1. As can be seen, the trend clearly demonstrates a noteworthy and consistent upward trajectory in searches related to data science since 2010. This growing interest highlights the strong desire of individuals and organizations to comprehend and harness the transformative potential of data science in addressing complex challenges.

Fig. 1. Global Google search interest over time for the term "data science"



The impact of this trend extends far beyond mere curiosity. It also illustrates the extensive utilization of data in various fields, including but not limited to education, energy, entertainment, finance, healthcare, marketing, telecommunications, and transportation. The programming landscape vividly portrays this phenomenon. For example, Python – a versatile language widely used in data science (www.python.org) – has experienced a surge in popularity, surpassing previously dominant languages such as C, C++, and C# (see Fig. 2). Python's rise as the preferred language for data-related tasks reflects the growing adoption of data-driven approaches and the importance of data analysis, manipulation, and interpretation. This shift highlights the pervasive influence of data in driving innovation and decision-making across industries, emphasizing its significant role in these advancements.

Fig. 2. Global Google search interest over time for the term "data science"



As data science continues to gain traction, a myriad of businesses, startups, and organizations are keen on integrating its methodologies, techniques, and tools into their operations. This growing demand and usage highlight the critical need for a clear and comprehensive understanding of the field. However, defining data science has proven difficult due to its interdisciplinary nature and the constant evolution of its boundaries. To navigate this complexity, we put forth the definitions provided by two esteemed institutions. The Data Science Association (www.datascienceassn.org) posits that:

> *"Data science means the scientific study of the creation, validation, and transformation of data to create meaning."*

On the other hand, IBM (www.ibm.com) offers an alternate viewpoint on Data Science:

> *"Data science combines the scientific method, math and statistics, specialized programming, advanced analytics, AI, and even storytelling to uncover and explain the business insights buried in data."*

In comparing the two perspectives, the Data Science Association emphasizes the scientific nature of data analysis, while IBM takes a more applied approach, emphasizing the multidisciplinary nature of data science. The Data Science Association focuses on the study of data and its transformation to derive meaningful insights, underscoring the importance of rigorous analysis and validation. On the other hand, IBM showcases the combination of diverse disciplines, methodologies, and tools in data science to uncover and explain business insights. This perspective highlights the practical application of data science techniques and the integration of various domains. Both perspectives contribute valuable insights into the field of data science, with the Data Science Association highlighting the scientific study and transformation of data, and IBM emphasizing the practical and multidisciplinary nature of the field.

To gain deeper insights into the realm of data science, in this paper we set out to explore fundamental aspects such as the roles of data scientists, the underlying approaches driving their projects, the methods they employ, their origins, and the analytical problems they investigate. In particular, we will address three pivotal questions:

- What do data scientists do, and what approach underlies their projects?
- Which methods do data scientists apply, and where do they originate from?
- What types of problems do data scientists investigate from an analytical perspective?

By answering these questions, we aim to foster a richer understanding of data science, its diverse applications, and the integral role of data scientists, particularly focusing on applications in business administration practice.

In Section 2, we will explore the roles, responsibilities, and tasks commonly performed by data scientists. Section 3 will delve into the methods utilized by data scientists, examining their origins and sources such as computer science, statistics, and operations research. In Section 4, we will focus on the various types of problems that data scientists investigate from an analytical perspective, specifically exploring descriptive analytics, predictive analytics, and prescriptive analytics. Section 5 concludes the paper.

2. Decoding the tasks of data scientists

Data science teams typically follow a specific procedural framework consisting of distinct stages during project execution.

- *Stage 1 (Recognize the situation & grasp objectives)*

At this stage, data scientists understand project goals, defining business problems while communicating regularly with domain experts or sponsors. This phase is vital for a successful solution. Continuous involvement of experts, reviewing interim results, and maintaining project trajectory are key. By the end, data scientists can define problems analytically.

- *Stage 2 (Data requirements, acquisition, & comprehension)*

Data scientists determine required data, identify relevant sources, and address gaps. They assess data quality, occasionally investing in hard-to-access data. Afterward, they typically use some initial descriptive statistics and visualization techniques to understand the data content and gain initial insights into the data.

- *Stage 3 (Data preparation & cleansing)*

In this phase, data scientists create the dataset that will be used for modeling in the following stage. This involves data cleaning activities, such as eliminating missing or invalid values, removing duplicates, properly formatting data, and combining data from various sources. It also involves transforming data into more useful variables.

- *Stage 4 (Data preprocessing & modeling)*

Modeling Data scientists apply alternative models and algorithms to find out which model is better suited in this context. For this, the prepared data set is often split into training and test data.

- *Stage 5 (Model validation/ Proof-of-concept)*

Once the models are developed and the initial results are convincing, it is time for a pilot trial to check functionality in practice. The functionality is evaluated by new data (validation data) or by user tests in a limited framework. If necessary, the models may need to be further developed.

- *Stage 6 (Communication & Implementation)*

In the event of a successful preceding stage, the developed system or concept moves forward to large-scale implementation. Now the organization can receive feedback on the model's performance and its impacts on the model's operational environment.

Understanding these stages is vital because it helps us see that data science is not just about having the right skills or tools. It is also about understanding the business and its problems and applying scientific methods to help solve those problems. As we move into the next sections, we will further explore the tools and methods data scientists use and discuss the types of problems they address from an analytical perspective. This will provide a comprehensive overview of data science as a field, with a specific focus on its applications in business administration practice.

Building on this, let us examine deeper into the potential pitfalls that data scientists may encounter in their work, specifically errors in the analytic process. We are all familiar with the first and second types of errors from statistics; false positives and false negatives (see, e.g., Anderson et al. 2016):

- Type I error: Incorrectly rejecting the null hypothesis.
- Type II error: Incorrectly failing to reject the null hypothesis.

But there is another the third kind of error, wherein an inaccurately formulated problem is solved correctly (see, e.g., Schwartz and Carpenter 1999):

- Type III error: Correctly solving a misformulated problem.

This might sound strange, but it happens more often than one might think. A data scientist might have the correct methodology, and the solution might be technically accurate, but if the initial problem was not defined correctly, the whole process is rendered irrelevant. For instance, you could have a flawless analysis regarding a customer group that is not a strategic focus for the company. Despite the technical accuracy of the analysis, it will not contribute to the company's goals and objectives.

This emphasizes the crucial point that before delving into the data and beginning the analysis, data scientists must first understand the problem from a business perspective. It is not enough to be a technical expert; a competent data scientist should also have a strong understanding of business principles, while maintaining regular communication with domain experts or sponsors.

A striking statistic revealed by VentureBeat in 2019 (www.venturebeat.com) indicated that a staggering 87% of data science projects never make it into production. This statistic further underscores the disconnect that can occur between a data science project's technical side and its practical, business-oriented application. While a project might be statistically sound and technically impeccable, if it does not align with the business objectives or cannot be practically implemented due to organizational constraints or a lack of understanding of the business context, it may end up being shelved.

Data scientists must therefore also be skilled at understanding and interpreting business problems. They must have the ability to communicate effectively with different stakeholders, translating complex statistical findings into actionable business insights. It is a challenging role that requires a balanced skillset, but when done correctly, the impact of data science can be transformative for businesses.

3. Methods applied by data scientists

The methods and tools commonly applied by data scientists are primarily rooted in three overarching disciplines: *Computer Science*, *Statistics*, and *Operations Research* (see Fig. 3). Each of these disciplines, along with the corresponding intersections such as *Decision Support & Analytics*, *Data Mining & Machine Learning*, and *Simulation & Risk Analysis*, imparts fundamental skills and techniques essential to the field of data science, as outlined in the following.

- *Computer Science*

Data scientists should possess a strong proficiency in programming languages such Python. Computer Science also provides tools and methodologies for data management, knowledge discovery, artificial intelligence, and big data technology, among others. They should be equipped to handle different types of data (including big data), processing, combining, storing, and so on.

- *Statistics*

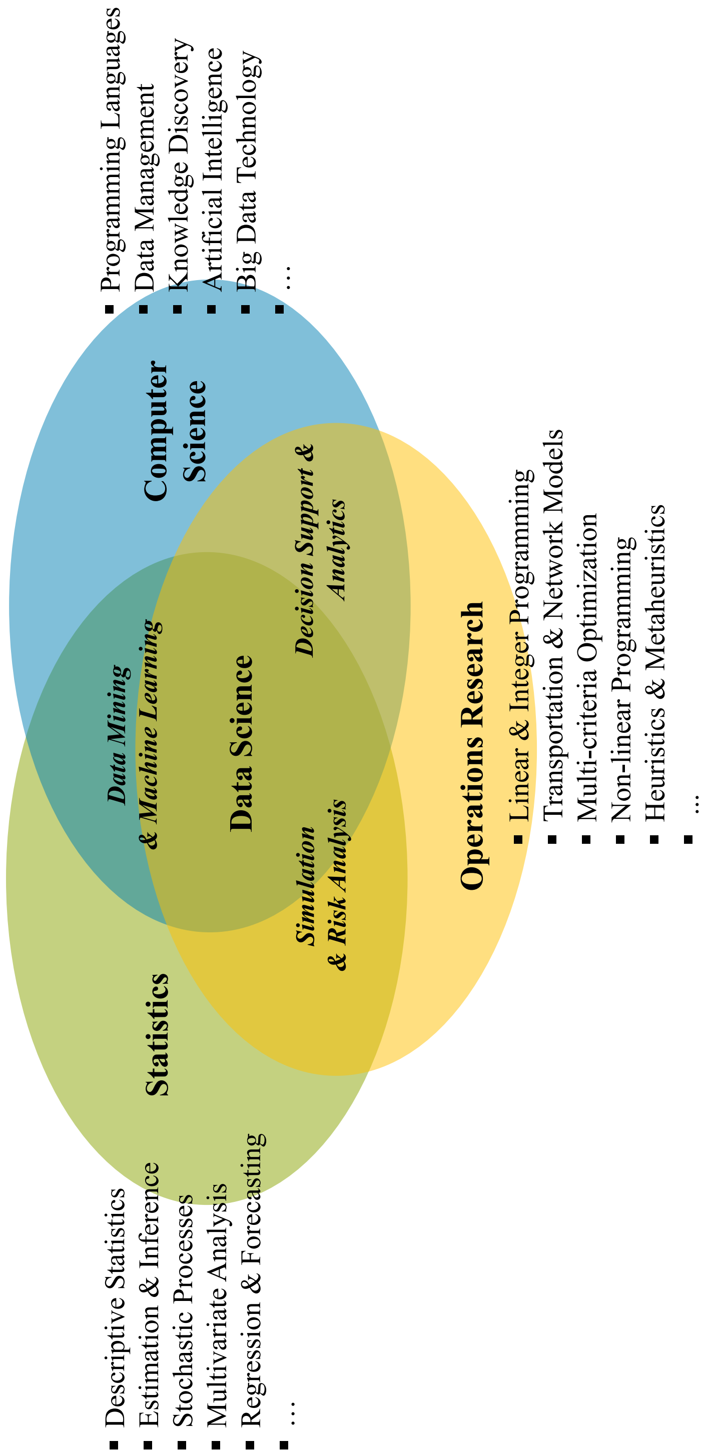
Data scientists require a strong grasp of statistical methodologies. Statistical methods allow them to gain a deeper understanding of the data and the relationships within the data. They should be able to utilize descriptive statistics effectively and understand concepts such as regression, estimation, inference, and so on. Methods such as descriptive statistics, estimation & inference, stochastic processes, multivariate analysis, regression & forecasting are essential in this context.

- *Operations Research*

Many complex business problems are decision-making problems and directly or indirectly related to optimization. Therefore, data scientists need to be familiar with important modeling and optimization methods. This includes linear & integer programming, transportation and network models, multi-criteria optimization, non-linear programming, heuristics & metaheuristics, and so forth.

Even though data scientists should possess a good understanding of the above-exemplified tools in the corresponding disciplines, the core of data science methodologies resides at the intersection of these fields as shown in Fig. 4.

Fig. 3. Fundamental skills and techniques essential to the field of data science (adapted from Evans 2017)



- *Decision Support & Analytics*

This is the intersection of Computer Science and Operations Research. The techniques involve developing analytically based information systems to support projects in decision-making.

- *Data Mining & Machine Learning*

This area represents the intersection of Computer Science and Statistics. These methods focus on recognizing characteristics and patterns between variables in data.

- *Simulation & Risk Analysis*

This is the intersection of Statistics and Operations Research. Methods here examine the effects of the uncertainty of estimates and their potential interactions on the desired output variable.

Returning to the potential pitfalls faced by data scientists in their work, it is important to address another type of error: Type IV error (see, e.g., Glen 2023). It occurs when a data scientist successfully formulates a problem and applies the appropriate analytical techniques to obtain accurate results. However, despite the correctness of the solution, there is a misinterpretation of the findings.

- Type IV error: Correctly solving a properly formulated problem, but with a misinterpretation of the results.

This error highlights the significance of not only solving problems correctly but also deriving meaningful insights from the results. Moreover, it is crucial to emphasize the practical implications such a pitfall. According to a study conducted by Gartner in 2019 (www.gartner.com), a staggering 80% of analytical insights fail to translate into tangible business outcomes. This finding underscores the challenge of bridging the gap between analytical findings and their implementation in real-world business contexts. It underscores the importance of not only generating accurate insights but also ensuring their relevance and applicability to the organization's goals and objectives.

This discussion highlights the need for data scientists to possess not only technical expertise but also a deep understanding of the business domain. They must actively engage with stakeholders, ask the right questions, and align their analyses with the strategic objectives of the organization. By doing so, they can enhance the likelihood of converting analytical insights into valuable business outcomes.

4. Problem types investigated by data scientists?

4.1. Descriptive analytics

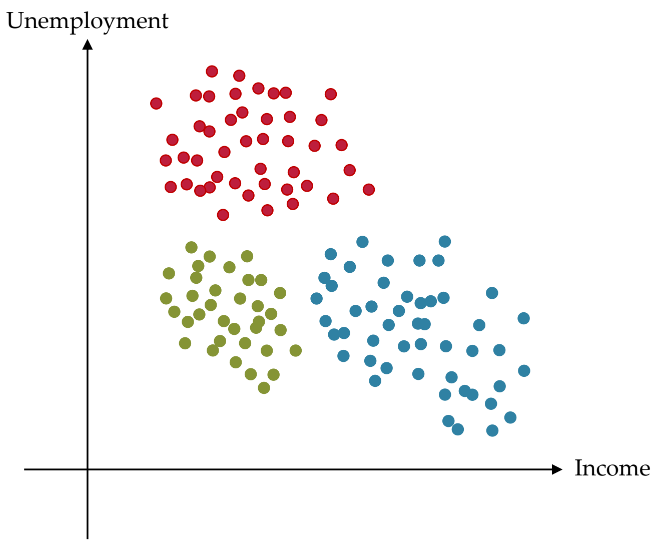
We delve into the domain of descriptive analytics, addressing key questions such as: *What is happening, and why is it happening?*

Known as the most frequently utilized and best-understood type of analysis, descriptive analytics is often the starting point of numerous projects. The aim is to comprehend and scrutinize both historical and current business performances, using various techniques from statistics and data mining to recognize patterns, associations, and more. We present in the following three examples to illustrate this type of analytics.

- *Example 1: Socio-economic analysis*

When assessing the social features of cities, we may consider a multitude of variables such as education, security, income, poverty, housing, transport, and health. This complex, multi-dimensional analysis can benefit from statistical techniques like clustering (for an overview of clustering techniques, see, e.g, Kubat). Simplifying the scenario, e.g., the *k*-means clustering categorizes cities with similar traits, focusing on variables such as unemployment and income, as shown in Fig. 4. It is then crucial to decipher the shared characteristics within each group, and why they exist. Furthermore, we might explore if additional features, such as geographical or cultural factors, can further illuminate these patterns.

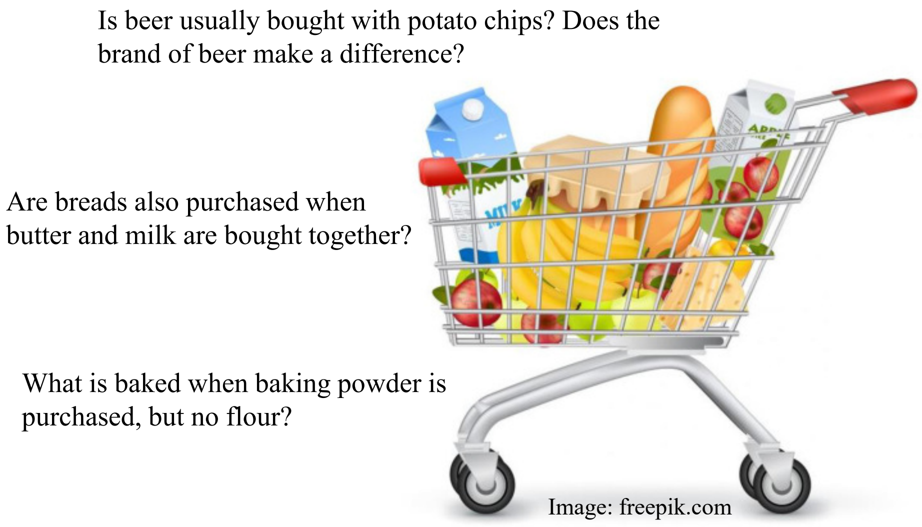
Fig. 4. Socio-economic analysis



- *Example 2: Shopping cart analysis*

Consider shopping at a grocery store; each visit introduces more data to their database. By applying techniques such as market basket analysis and association rules (see, e.g., Ünvan 2021), this cumulative data provides valuable insights into customer preferences and product associations. These insights guide improvements in marketing strategies, enabling businesses to make data-driven decisions and enhance customer experiences. As depicted in Fig. 5, exemplary questions in this context could be: Are potato chips typically purchased with beer? Does the brand of beer influence the purchase? When butter and milk are bought together, is bread also likely to be part of the basket? What is baked when baking powder is purchased, but no flour? Answering such questions helps optimize store layouts and promotional offers, ensuring that the right products are placed and priced strategically to maximize sales.

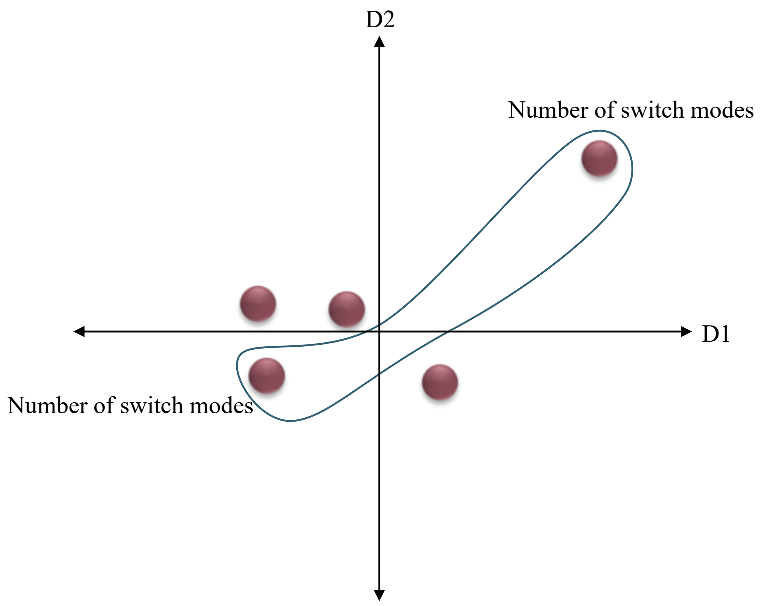
Fig. 5. Shopping cart analysis



- *Example 3: Visualizing associations*

The final example of this kind of analytics could be the use of multi-dimensional scaling for visualizing the association of factors within, e.g., a two-dimensional space (for an overview of the method of multi-dimensional scaling, refer to Tan et al. 2019). Consider a dataset in the context of benchmarking electricity transmission operators. Through multi-dimensional scaling, variables like "Number of switch modes" and "Number of substation equipment" are visibly separated, as illustrated in Fig. 7. This clear distinction signifies divergent behaviors for these variables when compared to all others. This technique also serves in reducing dimensionality when the number of variables needs to be minimized for technical reasons (see, e.g., Tan et al. 2019).

Fig. 6. Visualizing associations



- *A word of caution: The unintended consequences of descriptive analytics*

While descriptive analytics helps us understand what has happened and why it has happened, it is important to remember that this understanding relies on human interpretation. In other words, automated tools can describe data, but they cannot yet fully understand context or appropriateness without human guidance. Consider this example: an automated algorithm was created to generate variations of the famous "KEEP CALM AND CARRY ON" slogan for products to be sold on Amazon. This algorithm used descriptive analytics to identify popular trends and phrases, then created products accordingly. However, the algorithm lacked the ability to differentiate between appropriate and inappropriate phrases, leading to the creation of offensive slogans. This incident serves as a stark reminder of the limitations of descriptive analytics when used without careful human oversight. These tools can describe and generate based on data, but they lack the human ability to understand context, nuance, and appropriateness. It emphasizes the necessity of pairing analytics tools with careful human judgment. Understanding the limitations and thoughtfully implementing these tools are essential aspects of responsible data analytics.

4.2. Predictive analytics

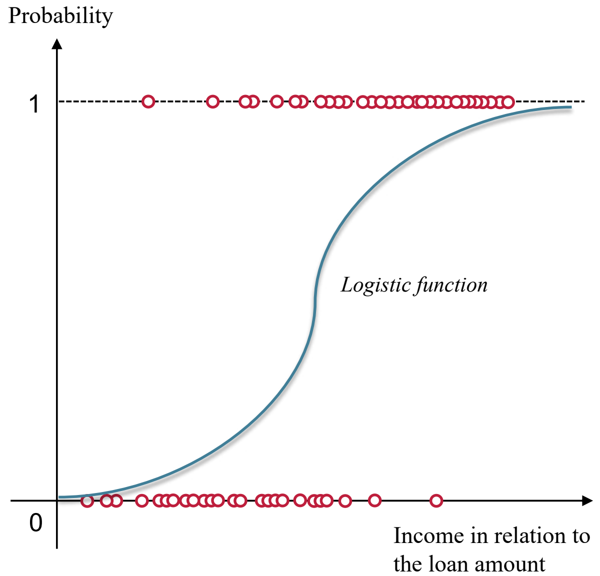
Predictive analytics forms the next level in the hierarchy of analytics. Following the descriptive analytics phase, where questions about what has happened and why are answered, the key question becomes: *What could happen in the future?*

Predictive analytics employs statistical models and forecasting techniques to anticipate future events. It analyzes historical data to identify patterns and trends, using this knowledge to predict future outcomes. While these predictions may not be completely accurate, they provide valuable guidance for decision-making and planning. The following examples illustrate the practical applications of predictive analytics in various business contexts, showcasing its utility and versatility.

- *Example 1: Credit scoring*

Consider a financial institution keen on predicting the credit repayment ability of their applicants using data science methods. They can utilize historical data about clients to determine who was able to repay the loan and who was not. One potential method to predict this is through logistic regression (for an overview of logistic regression, see, e.g., Burger 2018), which can provide a probability estimate of an individual being able to repay a loan (see Fig. 7). For instance, with a single variable scenario, the model might reveal an income-to-loan ratio equates to a lower chance of repayment.

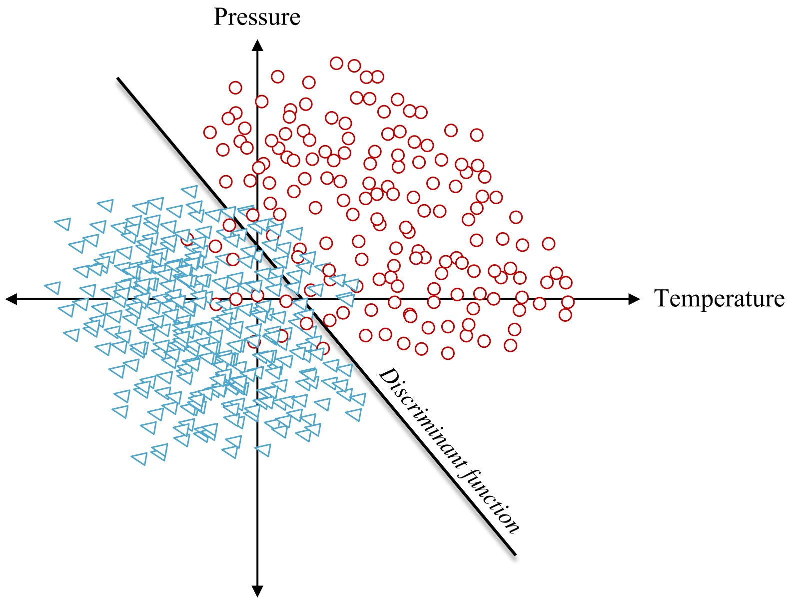
Fig. 7. Credit scoring



- *Example 2: Defect detection*

Consider the graphical example in Fig. 8. In the realm of semiconductor manufacturing, the timely detection of defective products on the production line is critical. By leveraging historical data from defective and non-defective productions and accounting for various relevant variables such as “temperature” and “pressure”, it becomes possible to predict if a newly produced item is defective. One viable approach for this task is discriminant analysis (see, e.g, Huberty and Olejnik 2016). Discriminant analysis aids in comprehending the distinctions between defective and normal products by identifying the key variables contributing to these differences.

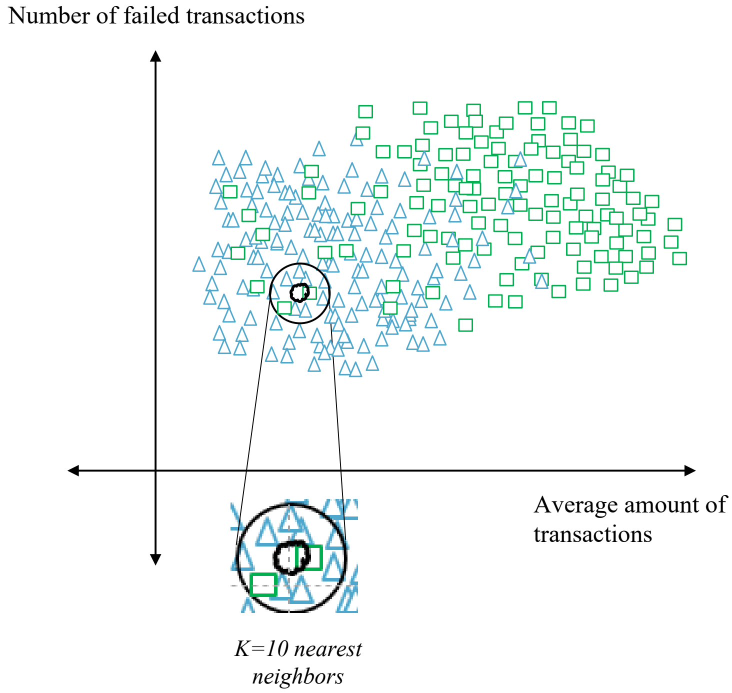
Fig. 9. Defect detection



- *Example 3: Fraud detection*

One of the most intriguing applications of predictive analytics lies in fraud detection and prevention, specifically credit card fraud detection from transactional records (see, e.g, Patil et al. 2018). Given the task's urgency, overlapping characteristics between fraudulent and secure observations often pose challenges. A pertinent method employed in this context is the *k*-nearest neighbors (for an overview see, e.g., Provost and Fawcett 2013). As demonstrated in Fig. 9, with the help of two variables “number of failed transactions” and “average amount of transactions”, it becomes feasible to assess the probability of fraud, through the majority vote of neighboring data points, indicating the potential for fraud in a new observation. The choice of *k* is context-dependent and integral to the success of this method.

Fig. 9. Fraud detection



These examples illustrate how predictive analytics, even with its inherent uncertainty, can provide crucial insights that inform and improve business decisions.

- *A word of caution: The pitfalls of predictive analytics*

Predictive analytics holds great potential, allowing us to forecast future outcomes based on historical data. However, it is crucial to apply robust analytical thinking to avoid any missteps. One notable example of predictive analytics gone awry is the Google Flu Trends project. Launched in 2008, Google Flu Trends sought to predict the spread of the flu using data from Google searches. This was not about the typical seasonal flu, but other, less visible epidemics. By analyzing internet searches related to the flu, the aim was to predict the onset of flu epidemics, potentially weeks before the virus became widespread in a specific region. In a publication in Nature (Ginsberg et al., 2009), successful predictions of all flu epidemics since the inception of the search engine were even showcased. However, the accuracy of the model soon came into question. Their predictive method, as it turned out, was well suited for past epidemics but fell short in anticipating future ones – a phenomenon known as underfitting. To address this, Google made numerous adjustments to their algorithm. Yet, this led to another issue: overfitting. The revised method started predicting more flu epidemics than there actually were, resulting in numerous false positives. In light of these shortcomings, the Google Flu Trends project was eventually discontinued, highlighting the importance of refining predictive models based on a nuanced understanding of data nuances and context.

4.3. Prescriptive analytics

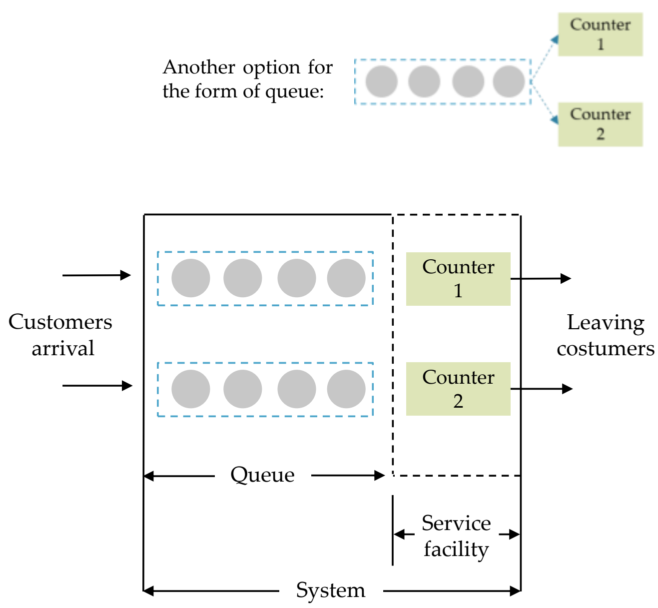
Prescriptive analytics is the final stage in the analytics journey, where the key question is: *What action should be taken to achieve a particular objective?*

As we transition from predicting potential outcomes to prescribing specific actions, we are confronted with a multitude of alternatives and choices. Among these extensive possibilities, the challenge is to identify the optimal alternatives that either minimize or maximize a specific objective. In the following, we discuss three examples highlighting the application of prescriptive analytics in diverse business scenarios.

- *Example 1: Queue management*

Imagine a fast-food restaurant chain where managers are eager to reduce waiting times and queue lengths while keeping service costs acceptable. As demonstrated in Fig. 10, through an exploration of diverse queuing disciplines or queue management strategies, such as the comparison between a single queue with two servers versus two separate queues each with its own server, the most efficient approach can be identified. Several methods are available for analyzing the flow of people, objects, or information through queues. Choosing the appropriate method can significantly enhance both customer experience and operational efficiency (for an overview of queueing theory, modeling, and applications, please refer to Bhat 2015).

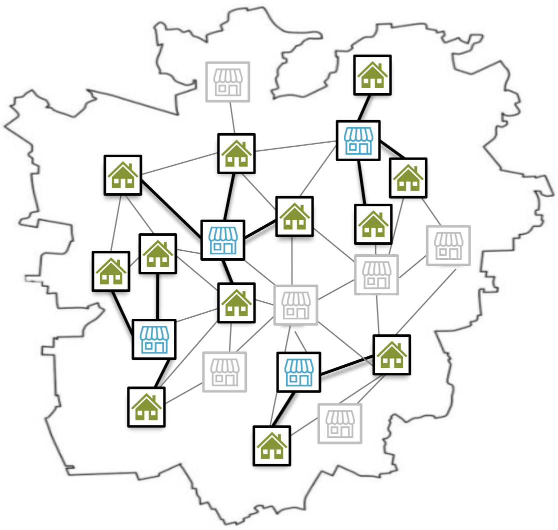
Fig. 10. Queue management



- *Example 2: Spatial separation analysis*

Consider a franchise business aiming to strategically position its stores to optimize spatial separation and minimize internal competition. A crucial factor is the distance between potential store locations and customers, as well as the distances to other outlets. A graphical illustration is provided in Fig. 11, depicting potential store locations (shown by square symbols), demands (represented by home symbols), and the corresponding network of distances (indicated by arrows) along with a hypothetical solution. We should note that the decision-making process can also be influenced by additional economic and ecological factors. In such scenarios, employing optimization techniques, such as facility location-allocation analysis methods, becomes essential to identify the most ideal solution (for an overview of the models and application areas of location analysis, see, e.g., Afsharian 2022).

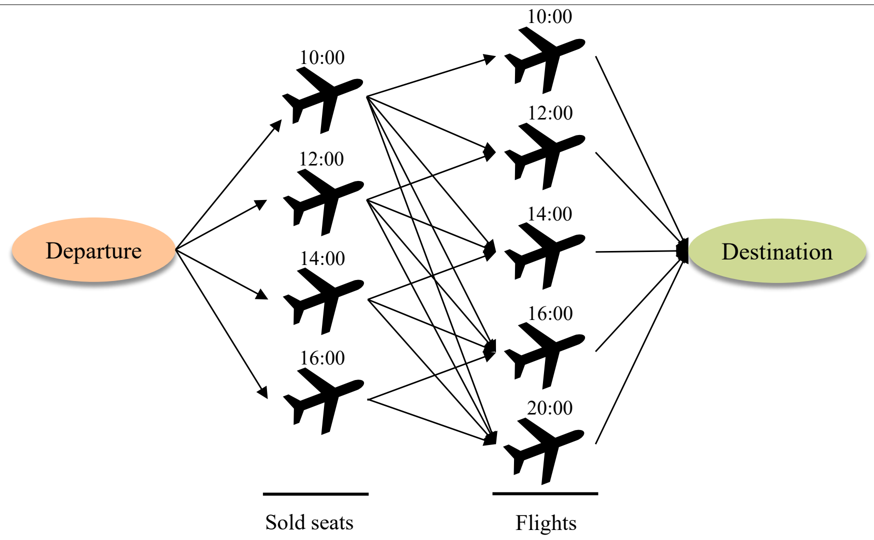
Fig. 12. Spatial separation analysis



- *Example 3: Overbooking management*

Overbooking is a common phenomenon in sectors like aviation and hospitality. For example, airlines often have to deal with no-shows or cancellations. To fully utilize their capacities, they resort to overbooking, based on historical data. However, the challenge lies in controlling the number of overbookings to prevent issues like passengers having to be rebooked onto later flights, as depicted in Fig. 12. At the same time, airlines have to adequately incentivize the impacted passengers in cases of over-sale. Network optimization methods are frequently employed to solve such problems and strike a balance between capacity utilization and customer satisfaction (see, e.g., Fard et al. 2019).

Fig. 12. Overbooking management



These examples underscore how prescriptive analytics can play a critical role in decision-making by suggesting the best course of action based on the available data and the defined objective.

- *A word of caution: The limitations of prescriptive analytics*

Prescriptive analytics is a powerful tool that can help decision-makers optimize their operations. However, it should be used with caution and the results should be interpreted with a full understanding of the method's limitations. An infamous example of this can be found in the overbooking scenario experienced by United Airlines (www.theguardian.com). United Airlines overbooked a flight to maximize capacity utilization, based on the prediction that some passengers would not show up. However, in this instance, more passengers showed up than the flight could accommodate, and none agreed to be rebooked onto a later flight. This led to a situation where the airline forcibly removed a passenger. While overbooking is a strategy derived from prescriptive analytics, its execution in this case failed to consider crucial aspects, notably how to handle a situation where all passengers arrive. Typically, airlines gradually increase incentives for passengers willing to be rebooked onto a later flight. In this case, the provided incentives were not enticing enough, leading to the unfortunate incident.

We conclude this section by returning to the potential pitfalls faced by data scientists in their work and addressing another type of error: Type V error. We define it as follows:

- Type V error: A properly formulated problem is solved accurately and interpreted correctly. However, delayed implementation renders the insights ineffective or irrelevant by the time they are applied.

In other words, even though the data scientists have formulated the problem correctly, executed the analysis accurately, and derived meaningful insights, the delay in implementing those insights renders them ineffective or irrelevant by the time they are put into action. This type of error can arise due to various reasons. It could be a result of organizational inefficiencies, delays in decision-making processes, lack of alignment between data science initiatives and business strategies, or even external factors that prevent the timely implementation of the recommended actions. For example, imagine a data scientist working on a project to optimize supply chain operations for a retail company. The data scientist successfully identifies bottlenecks in the distribution network, proposes efficient routing strategies, and provides actionable insights to improve overall efficiency. However, due to organizational inertia, bureaucratic hurdles, or other operational constraints, the recommendations are not implemented until much later, by which time the business landscape may have changed, rendering the insights less relevant or effective.

To mitigate this error, it is crucial for data scientists to not only deliver accurate analyses and meaningful interpretations but also actively engage with stakeholders and decision-makers throughout the process. Effective communication, collaboration, and a proactive approach to driving change are essential to ensure that the insights derived from data science efforts are acted upon promptly, leading to tangible business outcomes. By being aware of this potential error and taking proactive steps to address it, data scientists can help organizations avoid missed opportunities, seize the right moment for implementing data-driven decisions, and maximize the value derived from their analytical endeavors.

5. Conclusions

In this paper, we have provided an overview of data science, concentrating on the role of data scientists and their methods in tackling analytical challenges in business administration. The paper delved into the multidisciplinary nature of data scientists, emphasizing their technical skills, domain knowledge, and communication abilities. In particular, we explored the three main categories of data analytics: descriptive, predictive, and prescriptive. Descriptive analytics provides insights into what has happened, predictive analytics estimates what might happen in the future, and prescriptive analytics suggests actions to take for optimal outcomes.

It is essential to acknowledge that each type of analytics has its constraints and should be wielded with care and comprehension.

- Descriptive analytics can present an incomplete or even misleading picture if data are not adequately considered or if the context is not fully understood.
- Predictive analytics, while powerful, is limited by the quality of the historical data and the assumption that future patterns will follow those from the past.
- Prescriptive analytics, which strives to advise on potential actions, can lead to unforeseen consequences if not carefully considered and implemented.

The discussions in this paper provide an overview of the potential of analytics, while also highlighting the importance of analytical thinking, comprehensive understanding, and prudent application. It is crucial to remember that analytics tools and methods are not infallible and should be combined with expert judgment, particularly in intricate real-world situations. In essence, data analytics is a robust tool, yet its effectiveness hinges on responsible and thoughtful application.

References

Anderson, D. R., Sweeney, D. J., Williams, T. A., Camm, J. D., Cochran, J. J. (2016). Statistics for business & economics. Cengage Learning.

Bhat, U. N. (2015). An Introduction to Queueing Theory: Modeling and Analysis in Applications. 2nd edition. Springer Sciences+Business, New York.

Burger, S. V. (2018). Introduction to machine learning with R: Rigorous mathematical analysis. O'Reilly Media, Inc.

Data Science Association/ Code of Conduct. Available at: https://www.datascienceassn.org/code-of-conduct.html. Accessed on 30.09.2023.

Davenport, T. H., Patil, D. J. (2012). Data scientist. Harvard business review, 90.5, 70-76.

Evans, J.R. (2017): Business Analytics. 2nd edition. Pearson, London.

Fard, F. A., Sy, M., Ivanov, D. (2019). Optimal overbooking strategies in the airlines using dynamic programming approach in continuous time. Transportation Research Part E: Logistics and Transportation Review, 128, 384-399.

Gartner (2019). Business Insights and Trends. Available at: https://www.gartner.com/en/insights. Accessed on 01.08.2023.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. Nature, 457(7232), 1012-1014.

Glen, S. (2023). Type III Error and Type IV Error in Statistical Tests. Available at: https://www.statisticshowto.com/type-iii-error-in-statistical-tests/. Accessed on 30.09.2023.

Google Trend. Available at: https://trends.google.com/trends/?geo=DE. Accessed on 01.08.2023.

Huberty, C. J., Olejnik, S. (2006). Applied MANOVA and discriminant analysis. John Wiley & Sons.

IBM/ Data Science. Available at: https://www.ibm.com/cloud/learn/data-science-introduction. Accessed on 30.09.2023.

Kubat, M. (2021): An Introduction to Machine Learning. Springer, New York.

Patil, S., Nemade, V., Soni, P. K. (2018). Predictive modelling for credit card fraud detection using data analytics. Procedia computer science, 132, 385-395.

Provost, F., Fawcett, T. (2013): Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking. O'Reilly and Associates.

Schwartz, S., Carpenter, K. M. (1999). The right answer for the wrong question: consequences of type III error for public health research. American journal of public health, 89(8), 1175-1180.

Tan, P. N., Steinbach, M., Karpatne, A., Kumar, V. (2019). Introduction to data mining. Pearson, UK.

Ünvan, Y. A. (2021). Market basket analysis with association rules. Communications in Statistics-Theory and Methods, 50(7), 1615-1628.

VentureBeat (2019). VentureBeat's Data and AI Insider's Event. Available at: https://venturebeat.com/ai/why-do-87-of-data-science-projects-never-make-it-into-production/. Accessed on 01.08.2023.

Bisher erschienene Research Papers:

Nr. 1 | Februar 2017
Berlemann, Michael; Christmann, Robin.
The Role of Precedents on Court Delay. Evidence from a Civil Law Country.

Nr. 2 | September 2017
Matthes, Roland.
A note on the Saito-Kurokawa lift for Hermitian forms.

Nr. 3 | Februar 2018
Christmann, Robin.
Prosecution and Conviction under Hindsight Bias in Adversary Legal Systems

Nr. 4 | März 2019
Broere, Mark; Christmann, Robin.
Takeovers, Shareholder Litigation, and the Free-riding Problem

Nr. 5 | März 2019
Matthes, Roland.
Vector-valued Cusp Forms and Orthogonal Modular Forms

Nr. 6 | März 2020
Matthes, Roland.
A Geometric View on Linear Regression and Correlation Tests

Nr. 7 | März 2020
Matthes, Roland.
A Note on the Geometry of Partial Correlation and the Grassmannian

Nr. 8 | April 2020
Christmann, Robin; Kirstein, Roland.
You Go First! Coordination Problems and the Standard of Proof in Inquisitorial Prosecution

Nr. 9 | Juli 2021
Christmann, Robin.
Plea Bargaining and Investigation Effort: Inquisitorial Criminal Procedure as a Three Player Game

Nr. 10 | Dezember 2023
Afsharian, Mohsen.
Data Science Essentials for Business: Exploring Analytics and Data Scientists' Contributions

Nr. 11 | Dezember 2023
Broere, Mark; Matthes, Roland.
The Error of Present Value Discounting Conventions in the Case of Uniform Intra-Period Cash Flow

Nr. 12 | Dezember 2023
Rödder, Wilhelm; Dellnitz, Andreas.
Planning a Joint Venture: Decision Support by Linear Vector Duality Theory

Nr. 13 | Dezember 2023
Christmann, Robin; Klein, Dennis.
Compliance, Liability and Corporate Criminal Law: A Setup for a Three-Player Inspection Game

Leibniz FH

Leibniz-Fachhochschule
Expo Plaza 11
30539 Hannover

leibniz-fh.de