Contents

Acknowledgments

This book is prepared for teaching Analytics Project course at the School of Interdisciplinary Management and Technology, Institute of Technology Sepuluh Nopember (ITS), Surabaya, Indonesia.

I would like to thank the students who have taken the course and provided feedback on the materials. Their enthusiasm and curiosity have been a source of inspiration for me to create this book.

My deepest gratitude goes to Pak Nyoman Pujawan, the head of the school, for his vision and leadership in promoting interdisciplinary education and research. His guidance has been instrumental in shaping my teaching philosophy and approach to analytics education and its applications in engineering and technology.

To the staff and administration of the school, I am grateful for their hard work and dedication in supporting the students and faculty all these years.

My special thanks to Pak Muhammad Atok, my colleague, mentor, and partner in teaching the course. His insights and expertise have been invaluable in shaping the content of the course, and therefore, this book.

The book is inspired by the excellent work of many authors and educators in the field of business analytics, engineering optimization, and machine learning. Much of the content in this book is based on Mykel Kochenderfer and Tim Wheeler's book, *Algorithms for Optimization* (MIT Press, 2019). Mykel's teaching and research have been a source of inspiration for me, and I am grateful for his mentorship and guidance over the years. The template for this book is based on the Tufte-LaTeX class, which is designed to produce elegant books. I am grateful to the Stanford Intelligent Systems Lab for making it available to the public. This book would not have been possible without the support of the open-source community, and I am grateful for their contributions to the field of education and research.

PART I:

PRESCRIPTIVE ANALYTICS

1 Prescriptive Analytics

We will review the basic components and concepts of prescriptive analytics. We will start by reviewing how prescriptive analytics differs from descriptive and predictive analytics. We will then discuss the prescriptive analytics framework, including problem scoping and definition, data collection, optimization process, and implementation and monitoring. By the end of this chapter, you should have a good understanding of the key concepts and processes involved in prescriptive analytics and the expected outcomes of the prescriptive analytics projects.

1.1 Prescriptive Analytics vs. Descriptive and Predictive Analytics

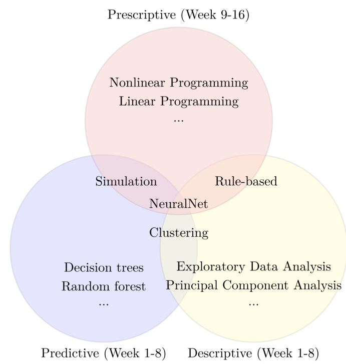

Figure 1.1. Algorithms and models for descriptive, predictive, and prescriptive analytics.

Prescriptive analytics is the process of using data and analytics to *recommend actions* that will *optimize outcomes*. It goes beyond predicting future outcomes by also suggesting actions to benefit from the predictions. In contrast, descriptive analytics focuses on what has happened in the past, while predictive analytics focuses on what is likely to happen in the future.

Although the scope of prescriptive analytics is broader than that of descriptive and predictive analytics, it is important to note that all three types of analytics are interconnected and build on each other. Descriptive analytics provides the foundation for understanding past performance, predictive analytics helps forecast future outcomes, and prescriptive analytics suggests actions to improve future outcomes.

It is worth mentioning that the algorithms and models used in prescriptive analytics are often more complex than those used in descriptive and predictive analytics. This is because prescriptive analytics involves optimization, which requires finding the best solution among a set of possible solutions. Though, given the pervasiveness of neural networks and deep learning, the line between predictive and prescriptive analytics is becoming increasingly blurred.

Figure 1.1 illustrates the relationship between descriptive, predictive, and prescriptive analytics. Descriptive analytics focuses on summarizing historical data, while predictive analytics uses statistical models and machine learning algorithms to forecast future outcomes. Prescriptive analytics suggests actions to optimize future outcomes.

Exercise 1.1. Let's consider a simple example to illustrate the differences between descriptive, predictive, and prescriptive analytics. Suppose you are a sales manager and tasked to increase sales for the upcoming quarter. Your colleague sends over a report with the following information:

1. List of products sold in the past year and their sales volume aggregated by month and by store location
2. Statistical summary of sales data, including average sales, standard deviation, and trends for each product
3. Clustering of products based on sales volume and their co-occurrence in transactions
4. Forecast of sales for the product clusters for the next quarter
5. Simulation and what-if analysis of different pricing and promotion strategies
6. Recommendations for pricing and promotion strategies to increase sales
7. Optimal allocation of marketing budget across different product categories

Based on the information provided,

a. **Identify which of the above items correspond to descriptive, predictive, and prescriptive analytics!**
b. **Given your task, which of the above items would be most useful for you and why?**
c. Let's say you succeed in increasing sales by 10% for the quarter. Your boss now wants you to scale this success across all regions. **How would you go about doing this?**

Solution: We will discuss this question in the first class.

Prescriptive analytics involves a set of processes and techniques to recommend actions that will optimize outcomes. The prescriptive analytics framework consists of the following major steps: problem scoping and definition, data collection, optimization process, and implementation and monitoring.

1.2 *Problem Scoping and Definition*

Crucial to the success of prescriptive analytics projects is the clear definition of the problem and the expected outcomes. This involves understanding the business objectives, identifying the key performance indicators (KPIs) that will measure the success of the project, and defining the available resources and stakeholders involved in the project. The problem scoping and definition phase is essential because it sets the foundation for the rest of the prescriptive analytics process. It helps ensure that the project is aligned with the business goals and that the results will be useful and actionable for the decision makers or stakeholders.

To define the problem, we need to answer the following questions:

- What are the business objectives of the project?
- What are the key performance indicators (KPIs) that will measure the success of the project?
- Who are the stakeholders involved in the project, and what are their roles and responsibilities?
- What data sources are available? How much can we trust the data? Are there any limitations or biases in the data that we need to be aware of?
- What are the expected outcomes of the project, and how will the results be used to make decisions?

While these questions may seem straightforward, they are often challenging to answer in practice, especially in complex and dynamic business environments that involve multiple stakeholders and data sources. For data-driven business analytics, it is also crucial to understand the limitations and coverage of the available data to ensure the problems are well-defined and the results are reliable and actionable. It is not uncommon for the problem scoping and definition phase to take up a significant amount of time and effort in prescriptive analytics projects, as it lays the groundwork for the rest of the process.

Let's consider an example to illustrate various aspects of problem scoping and definition in prescriptive analytics projects.

Exercise 1.2. Let's revisit the sales manager example, in which you are tasked with increasing sales for the upcoming quarter. You are responsible for optimizing the allocation of the marketing budget across different product categories to achieve this goal. You are assigned to work with the marketing team, the sales team, the inventory management team, and the data and IT for the project. During one of the initial meetings with these teams, a few key points are discussed:

1. The marketing team has historical data on the marketing budget allocation for different product categories and the corresponding sales volume for the past year, but they only use customer engagement metrics to evaluate the effectiveness of the marketing campaigns. It is unclear how these metrics are related to actual sales.

2. The sales team has data on the sales volume for each product item (not aggregated by product category) and can only record sales if the product is in stock. The current website and point of sales system are integrated in real-time with the inventory management system to ensure that the products are in stock. But, this also means that the sales team do not track backlogs or lost sales due to stockouts.

3. The inventory management team focuses on delivering products to the stores and customers on time. They have data on the inventory levels for each product item and the order received from the sales team, but they also do not have data on the lost sales.

4. The data and IT team maintains the data infrastructure and systems for the company. They can provide historical data and summary of correlations between the marketing budget allocation, customer engagement metrics, sales volume, and customer reviews post purchase. They can also provide predictive models for sales volume based on historical data for the upcoming quarter with 100% accuracy.

Based on the information provided,

a. **Identify a proper scope for your 1-month project to increase the sales for the upcoming quarter! What are the business objectives that can coordinate these teams? What are the KPIs? Are there any trade-offs?**
b. **What are the potential limitations and biases in the data? How would you address these limitations and biases?**

Solution: Given the information provided, we can consider the following answer (you may have a different answer, that's absolutely fine).

- Given a 1-month (pretty short time period to coordinate multiple teams while also working on their main responsibilities),
  - the scope of the project could focus to optimize the allocation of the marketing budget across different product categories to increase sales for the upcoming quarter.
  - The business objectives could be to increase sales volume for the quarter to increase market share to maintain a strong presence in the market and induce potential business growth. This will need coordinations from sales, marketing, and inventory management teams to better capture the demand (less stockouts, etc.). Supports from the data and IT team are also crucial to provide accurate and reliable data for the project.
  - The KPIs could be the montly or total sales revenue or volume for the next quarter.
  - The potential trade-offs could include the trade-off between customer engagement metrics and actual sales, the trade-off between maintaining stocks and inventory costs, and the trade-off between the marketing budget with renewing order fulfilment equipments, IT infrastructure, and other operational costs.

- The potential limitations and biases in the data could include the lack of data on lost sales due to stockouts, the use of customer engagement metrics to evaluate the effectiveness of the marketing campaigns, and the lack of correlation between these metrics and actual sales. To address these limitations and biases, the project could collect additional data on lost sales, customer reviews post purchase, and other relevant metrics, and use this data to build more accurate predictive models for sales volume. The project could also work with the marketing team, the sales team, and the inventory management team to improve the data collection process and ensure that the data is accurate, complete, and representative of the system or process.

1.3 Data Collection

Once we have defined the problem and the expected outcomes, the next step is to collect the data needed to build the prescriptive analytics model. Data collection is a critical step in the prescriptive analytics process because the quality and quantity of the data will determine the accuracy and reliability of the results.

The data collection process involves identifying the relevant data sources, extracting the data from these sources, cleaning and preprocessing the data, and storing the data in a format that is suitable for analysis. It is important to ensure that the data is accurate, complete, and representative of the system or process. If the data collection process involves integrating data from multiple sources, such as customer databases, sales records, marketing campaigns, and inventory management systems, as well as external sources, such as market research reports, social media, and industry benchmarks, it is essential to ensure that the data is consistent and compatible with each other.

Common data issues that may arise during the data collection process include missing data, outliers, errors, inconsistencies, and biases. These issues can affect the quality and reliability of the results and may require additional data cleaning and preprocessing steps to address. It is also important to consider the privacy and security of the data, especially if the data contains sensitive or confidential information. Data collection is an iterative process that may require multiple rounds of data collection, cleaning, and validation to ensure that the data is accurate and reliable.

1.4 Optimization Process

Optimization is the process of finding the best solution for a particular problem. The best solution is defined as the one that maximizes or minimizes a certain objective while satisfying a set of constraints. In engineering systems, data-driven business analytics, and even AI development and training, optimization is a crucial tool to improve performance and efficiency.

The process can be broken down into the following steps:

1. *Modeling*: Define the problem and create a mathematical model that represents the system or process.
2. *Verification and Validation*: Ensure that the model represents the system accurately and that the results are useful for the model's intended purpose.
3. *Solving the Model*: Use optimization algorithms to find the best solution to the model.
4. *Evaluation*: Evaluate the solution and compare it to other possible solutions or benchmarks.
5. *Interpretation*: Analyze the results and interpret the implications of the solution, taking into account the simplified nature of the model and the assumptions made.
6. *Iteration and Refinement*: Refine the model and the solution based on the interpretation and feedback from the users or stakeholders.

1.4.1 *Modeling*

We often start by a set of *design specifications*[1] and *performance measures*[2] that define the requirements and goals of the optimization problem. The goal at this stage is to find the right mathematical representation of the problem that captures the essential features of the system or process. Oftentimes, these specifications are defined as the *objective function* and the *constraints* of the model. It is also common to identify the *decision variables* that represent the choices we can make to improve the system or process.

[1] Design specifications are the requirements that the solutions must satisfy.

[2] Performance measures are the metrics that quantifies the performance of the system or process.

- The *objective function* is a mathematical function that we want to maximize or minimize. It represents the performance measure that we want to optimize.
- The *constraints* are a set of conditions that the solution must satisfy. These conditions can be related to the design specifications, physical limitations, or other requirements.

- The *decision variables* are the parameters that we can adjust to optimize the objective function. These variables represent the choices that we can make to improve the system or process.

Exercise 1.3. Let's revisit the sales manager example from the previous section. Suppose you are tasked with optimizing the allocation of the marketing budget across different product categories to increase sales. **Identify the objective function, constraints, and decision variables for an optimization model that can help you achieve this goal!**

Solution: There can be many ways to model this problem, and the choice of the objective function, constraints, and decision variables will depend on the specific requirements and goals of the sales manager. An example of an optimization model for this problem could be as follows:

- The objective function is the performance measure that we want to optimize. In this case, the objective function could be the total sales revenue or the sales volume for the quarter.
- The constraints are the conditions that the solution must satisfy. For example, the constraints could be the total marketing budget available, the minimum and maximum budget allocation for each product category, or the expected increase in sales for each product category.
- The decision variables are the parameters that we can adjust to optimize the objective function. In this case, the decision variables could be the marketing budget allocation for each product category.

1.4.2 *Verification and Validation*

Once we have defined the problem and created a mathematical model, we need to check whether the model captures the logic of the system that we want it to represent. This process is known as *verification*. We also need to ensure that the results obtained from the model are sufficiently accurate and useful for the intended purpose. This process is known as *validation*. Verification and validation are crucial steps in the optimization process because they help ensure that the model is reliable and that the results are trustworthy. If the model is not verified and validated properly, the results may be inaccurate or misleading, leading to suboptimal or even harmful decisions. We will discuss verification and validation in more detail in the next chapter.

1.4.3 *Solving the Model*

Once we have verified and validated the model, we can use optimization algorithms and solvers to find the best solution to the model. The goal of the solvers is to find the values of the decision variables that optimize the objective function while satisfying the constraints as defined in the model. There are many different optimization algorithms that we can use to solve the model, depending on the complexity of the problem and the requirements of the optimization process. The solutions obtained from the solvers are known as the *optimal solutions* of the model, which represent the best possible outcomes given the constraints and the objective function.

1.4.4 *Evaluation*

After we have obtained the optimal solutions from the solvers, we need to evaluate the solutions to determine their quality and usefulness. This evaluation process involves comparing the optimal solutions to other possible solutions or benchmarks, analyzing the results, and interpreting the implications of the solutions. The goal of the evaluation process is to ensure that the solutions are reliable, accurate, and useful for the decision makers or stakeholders.

Important to note that the evaluation process is not just about comparing the results to the initial design specifications or performance measures. It is also about understanding the limitations of the model and the assumptions made, and interpreting the results in the context of the real-world system or process that the model represents. In data-driven business analytics, it is also crucial to know the limitations and coverage of the data used to build the model. If the data is biased, incomplete, or not representative of the system under study for the time period of interest, the results may be inaccurate, suboptimal, or even harmful.

1.4.5 *Interpretation*

Once we have evaluated the solutions, we need to interpret the results and analyze the implications of the solutions. This interpretation process involves understanding the trade-offs between different solutions, analyzing the sensitivity of the solutions to changes in the model parameters, and identifying the key factors that influence the outcomes. The goal of the interpretation process is to provide insights and recommendations to the decision makers or stakeholders, based on the results of the optimization model.

In most cases, the raw solution of the optimization model may not be directly actionable or understandable to the decision makers or stakeholders. The interpretation process involves simplifying the results and presenting them in a way that is easy to understand and act upon. This may involve visualizing the results using charts, graphs, or other visual aids, or summarizing the results in a report or presentation. The interpretation process is crucial because it helps ensure that the results are communicated effectively and that the decision makers or stakeholders can make informed decisions based on the results.

1.4.6 *Iteration and Refinement*

Finally, after we have interpreted the results and provided recommendations to the decision makers or stakeholders, we need to refine the model and the solution based on the feedback received. It is important to accommodate the feedback and suggestions from the users of the model, as well as to update the model with new data or information that becomes available. This iterative process of refinement helps ensure that the model remains accurate, reliable, and useful. It also helps improve the performance of the system or process over time, as new information and insights are incorporated into the model.

Exercise 1.4. Imagine you are an inventory manager for a retail store and you are tasked with building an optimization model the inventory levels for different products to minimize costs while ensuring that the products are always in stock.

a. **Identify the modeling components (objective function, constraints, and decision variables) for your model!**

b. **What data sources would you need to build and validate the model?**

c. **Who are your stakeholders? How would you evaluate and interpret the results of the model to communicate with them?**

d. **What are some foreseeable refinements that you could make to the model?**

1.5 *Implementation and Monitoring*

The final step in the prescriptive analytics process is the implementation and monitoring of the results. During the implementation process, we will need to work closely with the decision makers, stakeholders, and users of the model to ensure that the solutions are implemented effectively and that the expected outcomes are achieved. This may involve making changes to the existing processes, systems, or workflows, or introducing new tools, technologies, or strategies to improve the performance of the system. Because it may introduce changes to the existing processes, it is important to communicate the changes effectively and provide training and support to the users as needed.

Also, it is important to note that hhile the solutions of the prescribed by our prescriptive analytics project may be optimal at a given point in time, the real-world system or process is dynamic and may change over time. It could also be the case that the business shifts their priorities or goals, which may require adjustments to the solutions. Data and information may also become outdated or inaccurate, which may affect the reliability and accuracy of the results.

To address these challenges, it is important to establish a monitoring and evaluation process that tracks the performance of the system or process, gathers feedback from the users and stakeholders, and makes any necessary refinements or adjustments to the solutions. Recent advances in AI and machine learning have made it easier to automate the monitoring and evaluation process, using real-time data and analytics to track the performance of the system and make adjustments as needed. There are also techniques to monitor the consistency and distribution of the data (often referred to as *runtime monitoring*) to ensure that the results remain reliable and accurate. In the case of outliers[3], rare-events[4], or data-drift[5], the system can be designed to trigger alerts or notifications to the users or stakeholders, so that they can take appropriate actions to address the issues. Such a system is out of the scope of this book, but it is worth mentioning that it is an active area of research and development.

[3] In this context, outliers are data points that are significantly different from the rest of the data.

[4] Rare-events are events that occur infrequently but have a significant impact on the system or process.

[5] Data-drift is the change in the distribution of the data over time, which may affect the reliability and accuracy of the results.

1.6 Summary

In this chapter, we reviewed the basic components and concepts of prescriptive analytics. We discussed how prescriptive analytics differs from descriptive and predictive analytics, and how it involves using data and analytics to recommend actions that will optimize outcomes. We also discussed the prescriptive analytics framework, including problem scoping and definition, data collection, optimization process, and implementation and monitoring. The main objective of this chapter is to provide you with a good understanding of the key concepts and processes involved in prescriptive analytics and the expected outcomes of the prescriptive analytics projects. In the next chapter, we will discuss details about optimization modeling and discuss a few examples that can be useful for your prescriptive analytics projects in the future.

1.7 End of Chapter Exercises

Exercise 1.5. What are the outputs of the prescriptive analytics that are different from the descriptive and predictive analytics?

Exercise 1.6. What additional challenges do data-driven presciptive analytics (i.e., where the problem parameters may be inferred from the available data) face compared to traditional optimization problems in which problem parameters are assumed known?

Exercise 1.7. Give an example of data collection challenges that may arise in a prescriptive analytics project in your organization! What are the potential solutions to address these challenges? If these challenges are not addressed, what are the potential consequences given the solutions are implemented based on the data collected?

Exercise 1.8. An optimization model outputs a solution that describes the optimal routes for a delivery service given a real-time traffic data. The output is a huge matrix that contains binary values (0 or 1) that represent the nodes that the delivery service should visit at a given time. Clearly, this output is not directly actionable by the delivery service drivers nor is it interpretable by the managers. What are the potential solutions to address this issue?

Exercise 1.9. A prescriptive analytics project is often constrained by time and resources. How should a project manager manage the risks of not meeting the project deadlines or budget constraints? What are the potential trade-offs that the project manager should consider?