4. MACHINE LEARNING

Author: Dejan Mirčetić

There are a lot of questions about what is Machine learning (ML). Is it a really process in which machines learn from the external environment by themselves, or it is a formalised process via mathematical algorithms which allow computers to „figure out" rules in the outside world? What tools is ML using? How does the typical data flow look in the ML pipeline? Is it applicable to traditional industries, not just in IT and internet-related industries? Where is the place of ML in the context of business? How to systematically use it for solving business issues? Is there any architecture on how to apply it to SCs?

On these and similar questions, we will try to provide answers in the following chapter, closing with a real case study example of the application of ML algorithms in the food supply chain.

4.1. What is machine learning?

Machine learning is a discipline focused on two interrelated questions: How can one construct computer systems that automatically improve through experience? and What are the fundamental statistical computational-information-theoretic laws that govern all learning systems, including computers, humans, and organizations? The study of machine learning is important both for addressing these fundamental scientific and engineering questions and for the highly practical computer software it has produced and fielded across many applications (Jordan & Mitchell, 2015).

ML arises from this question: could a computer go beyond "what we know how to order it to perform" and learn on its own how to perform a specified task? Could a computer surprise us? Rather than programmers crafting data-processing rules by hand, could a computer automatically learn these rules by looking at data? This question opens the door to a new programming paradigm (Chollet, 2021).

The ML enables a fundamental shift in the programming paradigm (Figure 4.1). In classical programming human programmer inputs rules (programme) and the data that is analysed and processed in agreement to those rules. As a result, the answers are provided at the end.

On the other hand, with ML human programmer inputs the data with answers expected from the data, and outcome the rules.

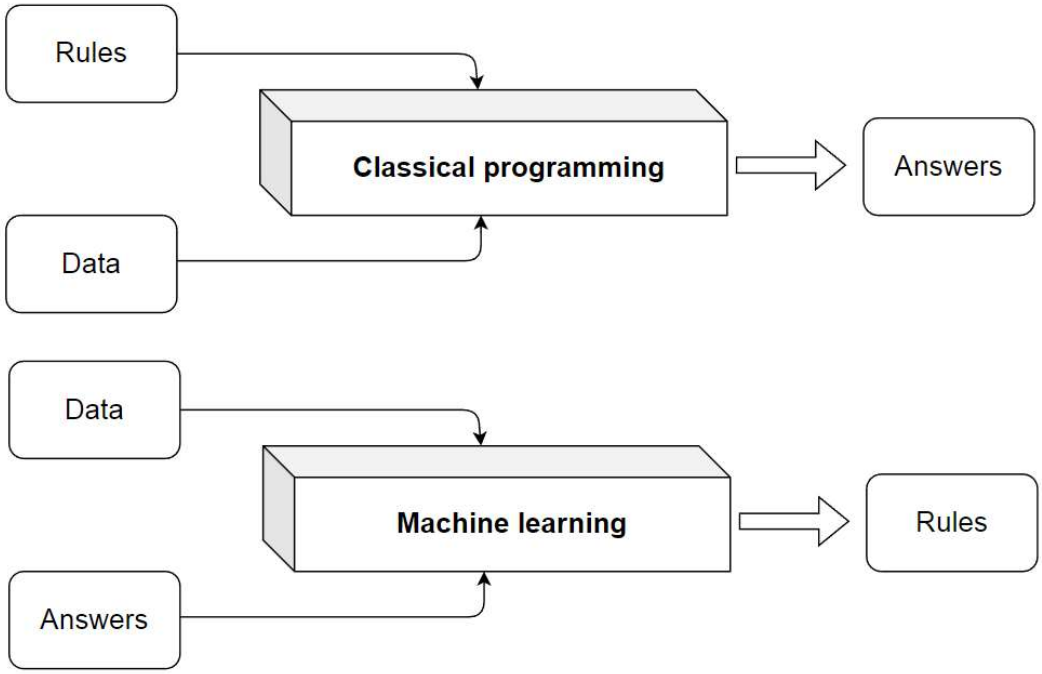

Figure 4.1 The classical programming vs. machine learning system training

Source: Chollet (2021).

Classical programming could be understood as imperative programming since the programmer predefines all the rules and execution of code is performed accordingly, while ML could be understood as declarative programming where we express higher-level goals or describe important constraints, and rely on mathematical algorithms to decide how and/or when to translate that into action.

Today, ML is the foundation of countless important applications, including web search, email anti-spam, speech recognition, product recommendations, and more (Ng, 2017). Many developers of AI systems now recognize that, for many applications, it can be far easier to train a system by showing it examples of desired input-output behaviour than to programme it manually by anticipating the desired response for all possible inputs. The effect of ML has also been felt broadly across computer science and across a range of industries concerned with data-intensive issues, such as consumer services, the diagnosis of faults in complex systems, and the control of logistics chains (Jordan & Mitchell, 2015).

4.2. Foundations and theoretical assumptions of ML

The background of ML lies in mathematics, more specifically in statistics. Therefore ML uses a theory background and algorithms developed in **statistical learning** and there is also debate is the ML a real area of its own or it is just part of statistics. In practice ML algorithms usually lack a certain level of mathematical rigidity and sometimes easily go above some mathematical constraints present in statistics. For example, ML algorithms do not pay a lot of attention to confidence intervals when optimizing the coefficients in parametric algorithms, although this is one of the most important topics in statistics. Generally, there is a **big overlap of ML and statistics** and some of the most notable ML algorithm creators and professors claim that it is just part of statistics (Hastie et al., 2009). Nevertheless, being the area of its own or part of statistics, ML consists of several steps in acquiring knowledge from the data. There is no general consensus about these steps but generally, they can be represented as transformation of different data sources to business intelligence insights.

In a business context, the ML models are useless, without proper support regarding the data pre-processing, data mining and application of the insights to the actual processes. Therefore, creating the ML algorithms without the possibility of updating the model and using its output for an actual decision-making process, does not bring any value to modern companies. Accordingly, in modern-day business analytics, the quantitative ML process is usually part of the business intelligence workflow. More specifically, it is part of business intelligence's important subprocesses (data science and data analytics part of business intelligence). The details about the role of ML in these processes and the actual processes itself of generating the values for the business via ML will be provided in the upcoming subchapter.

4.3. Business intelligence and ML in SCs

Business intelligence, in the context of SCs, is the process of making conclusions about the observed SC processes, based on the modelling of the data from those processes. It is mostly based on statistics, but other mathematical areas come into play: operation research, linear algebra, fuzzy logic (in a case when data is scarce or missing), numerical optimization, metaheuristics, etc. Additionally, new disruptive technologies are also becoming an important

aspect for analysing the data and delivering conclusions: **machine learning, artificial intelligence, digital twins, smartisation, living labs**, etc.

There are no strictly organized procedures on how the business intelligence procedure and ML workflows should be organized, but there are some useful guidelines in the practice and literature which have been proven successful when conducting the analysis. The procedure for conducting the business intelligence is also diverse to the origin of the software that is used for the analysis. For example, Microsoft offers several tools via its channel Microsoft Business Intelligence package, that conduct different tasks: **data ingestion, data storage, data integration, data management, data processing, reporting, data sharing and data science** (Figure 4.2).

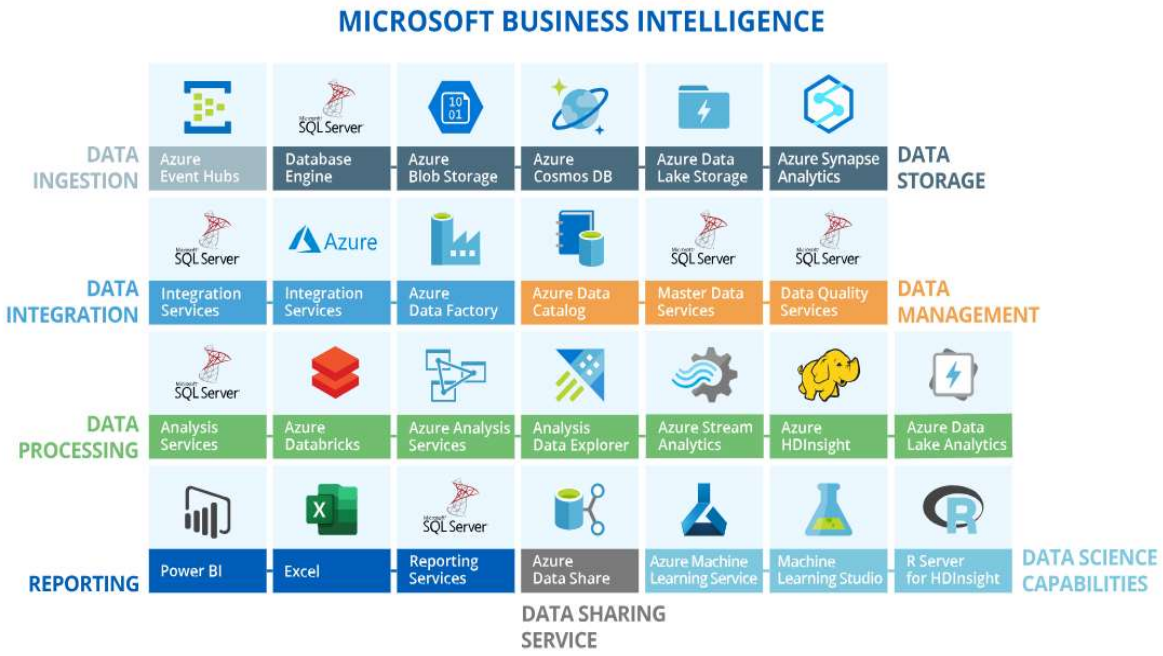

Figure 4.2 Microsoft business intelligence architecture

Source: ScienceSoft (n.d.).

In a given architecture the ML procedures are applied only at the data science level via several tools: Azure ML services, ML studio and R Server for HDInsight. The general procedure of how the data analysis in the context of ML is performed in R Server is presented in Figure 4.3.

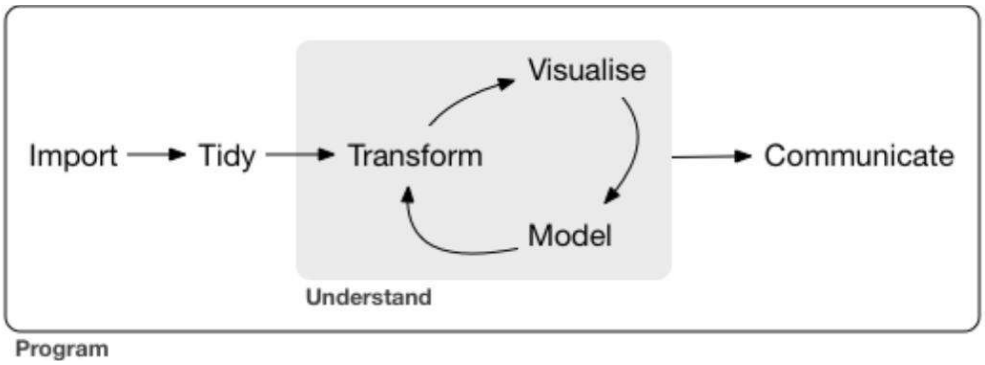

Figure 4.3 ML Data analysis steps in R

Source: Wickham et al. (2023).

When dealing with ML, there is usually a **misconception that the majority of time and effort is spent on actually building the ML algorithms**. The reality is totally the opposite, the majority of time is usually spent on wrangling with the data and pre-processing tasks rather than to the modelling process. Sometimes, all the processes before the modelling process are much more challenging and demanding. That is why there is no conses on how these steps need to be performed. Figure 2.9 presents an example of a good approach to transforming the data into business insights and general knowledge. The procedure starts with the import step, which is one of the most important steps in building ML models, since without importing data to the software it is not possible to conduct any kind of analysis. This typically means that you take data stored in a file, database, or web application programming interface (API), and load it into a data frame in R (Wickham et al., 2023). The second step is related to tidying the data which is a procedure unique to R and relates to transforming the data into a specific form for further analysis (each column is variable, and each row is an observation– tibble data frame). The next step is related to the transformation of the data which usually includes narrowing the set of observations to the subsample of interest. Additionally, it may also include creating new variables as combinations of several existing ones or generating summary statistics. Visualization and modelling serve distinct but complementary roles in the realm of data analysis. Visualization is a profoundly human-centric activity, offering insights that may elude more formalized approaches. A well-crafted visualization can reveal unexpected patterns, prompt new inquiries, and even suggest that the original questions may need refinement or different data. In contrast, models provide a mathematical or computational framework for answering precisely formulated questions. They offer scalability and efficiency, making them suitable for handling large datasets. However, models (in which ML are also

included) come with inherent assumptions, and they cannot question or challenge these assumptions. Consequently, models may not have the capacity to surprise or unveil unforeseen insights. The synergy between visualization and modelling is evident in their collaborative role in data analysis. Visualization aids in the initial exploration, encouraging the formulation of precise questions, while models systematically provide answers within the defined parameters. Recognizing the strengths and limitations of each approach is crucial, leading to a more comprehensive and informed data analysis process. The last step represents communication which is vital for the success of data analysis, since if the information is not provided to the decision maker in a right and consistent way, then the whole analytics could be in vain. The key element in data analytics are the ML models, without which, conclusions about the business processes could not be inferred. In order to tackle the specific SC problems, the architecture for general-purpose business intelligence applications (presented in Figure 4.2) needs to be better tuned, as well as the ML models. Accordingly, to transform how supply chains operate via enhancing operational efficiency, improving decision-making, and driving towards the achievement of corporate objectives, the company **Equilibrium AI** developed the AI & ML platform presented in Figure 4.4.

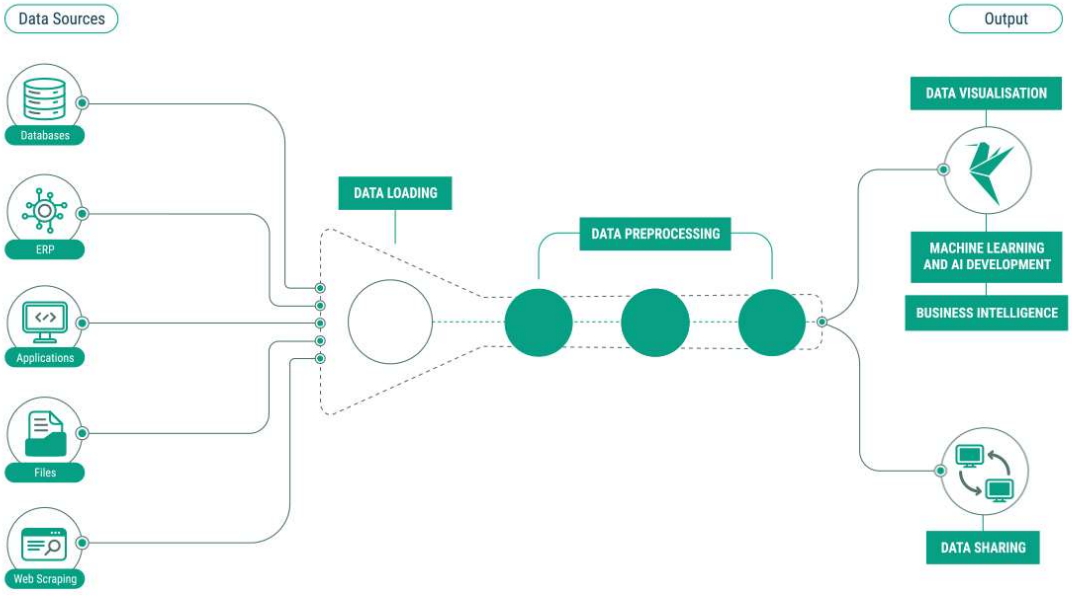

Figure 4.4 The ML data & knowledge pipeline for company Equilibrium AI

Source: Equilibrium AI (n.d.).

The figure represents a good example of everyday practice of how the extraction of knowledge and insights are generated in SC applications. Generally, the process consists of backend and frontend operations in order to create value (business insights) for users. The backend process starts with extracting the data from different data sources usually found in SCs:

- Databases;
- Enterprise resource planning systems (SAP, Navigator, Microsoft Dynamics, etc);
- Applications (web APIs);
- Flat files (csv, xlsx, JSON, etc);
- Web, internet and other online sources.

Each of the data sources has a different structure, protocols and accordingly procedures for how the data is extracted and loaded for cleaning and pre-processing before applying the ML algorithms. Accordingly, this process is performed via data loaders which have pre-programmed code for data mining of different data sources and transitioning the row data to the new database, which is structured and arranged for the application of ML models. Before applying the ML models, there is one additional step called data pre-processing. In this step row data gathered from the companies is checked for the wrong inputs, non-logical values, correct structure of inputs, outliers, double entries, NA, NaN, etc. The procedure continues with merging the outside data with company data. This data is usually related to external factors which can potentially influence the observed SC business process, for example, weather data, consumer price index, average income in a given region, specific demographic characteristics in a given area, gas prices, pandemic outbreaks, social network comments about company products, etc. This is very important because it holistically collects all the possible factors (internal and external) which are possibly influencing on a given business process, which increases the chance that the ML models will find the right signal in the data and be able to make the right conclusion and rules what are the root cause reasons why is business process behaving as observed.

After merging internal and external data, pre-processing consists of signal detection, removing the noise from data, feature engineering and randomly dividing the data on train and test (sometimes on validation data if the neural network model are developed). Data which comes out of the pre-processing step is cleaned and structured for the application of ML models.

The output part consists of data visualization, ML & AI development and data sharing. Sometimes, this data is shared without applying the ML models to other platforms which conduct different kinds of analysis (just reporting to stakeholders or government agencies). The visualization process is performed via the frontend part of the platform which is user-centric and allows users to make requests on what data, how and in what settings they want to see the observed SC data (for example Figure 4.5).

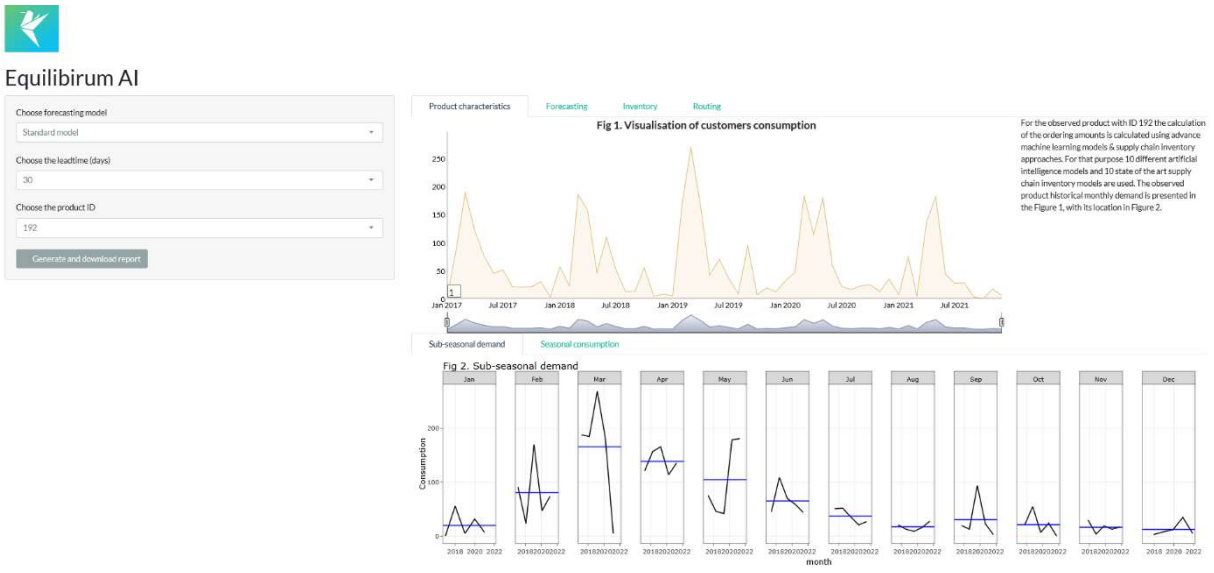

Figure 4.5 The typical visualization part of ML platform in SCs

Source: Author.

Conversely, the ML part of processing the data is hidden from the eyes of the users and it is not easy to understand. That is why the ML models are sometimes regarded **as black box** models in which there is no clear understanding of how exactly the machine connected the observed input with the observed output. This is one of the obstacles which is preventing the broader usage of ML models in practice, especially ones that are complex to interpret (Rostami-Tabar & Mircetic, 2023). Accordingly, we could divide the ML models into those with high interpretability-low flexibility and low interpretability-higher flexibility (Figure 4.6). In general, as the flexibility of an ML method increases, usually accuracy of the ML model increases and interpretability decreases (Mirčetić et al., 2016).

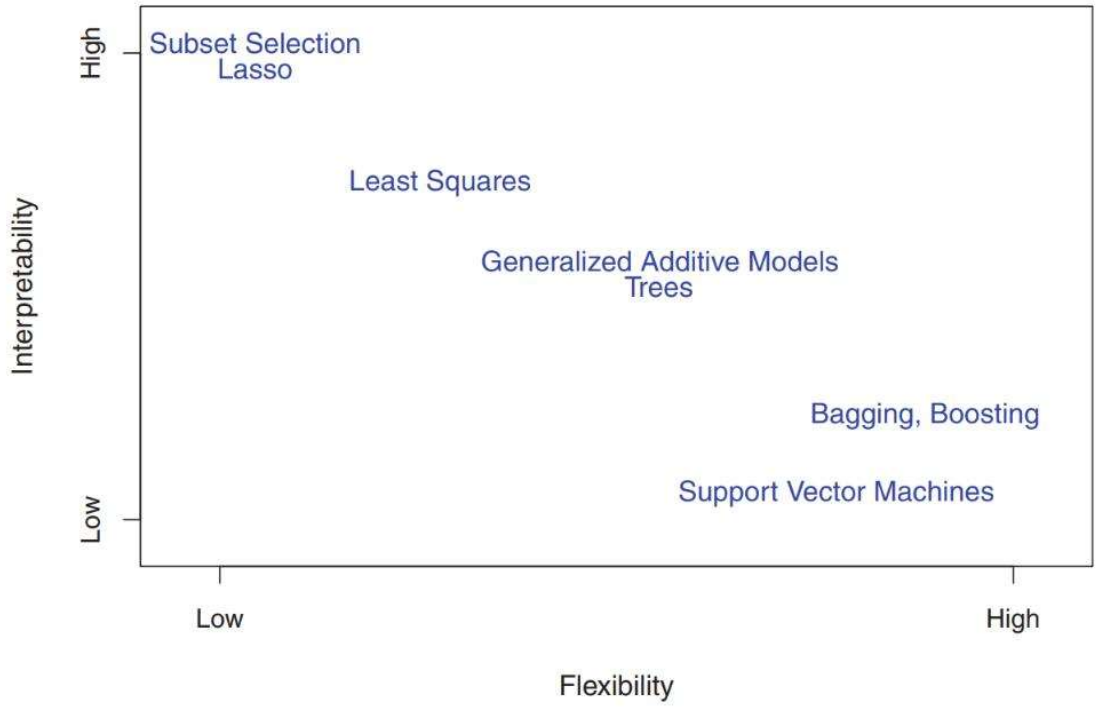

Figure 4.6 A representation of the trade-off between flexibility and interpretability, using different ML methods

Source: Hastie et al. (2009).

ML and SC business data

If the users better understand the visualization and graphics like Figure 4.5, why do we need ML at all, and could we skip modelling the data with ML and just make informative graphics? Unfortunately no. Maybe the main reason why we need ML models is because it is not possible in all situations to have easily readable and detectable patterns in the data seen via graphics (like in Figure 4.5). The more common situation is that graphics cannot usually reveal the mystery of what is happening in the observed SC business data and we need stronger tools in the form of ML algorithms to dig deeper into data and search for **data-generating rules** (Figure 4.7).

Figure 4.7 Statistical characteristics of products in the food supply chain (summarized for all products)

Source: Author.

It is very hard to make easy conclusions from Figure 4.7 and derive business rules about the data-generating process. To find patterns in the data from the figure, we have developed an ML model which could be used to summarize the characteristics and detect important signals in data. Accordingly, the developed ML model for a food supply chain is presented in Equation 1. The basic driver and the backbone of this ML model is the autoregressive integrated moving average model, with the general following form:

$$y_t' = c + (\phi_1 y_{t-1}' + \ldots + \phi_p y_{t-p}') + (\theta_1 e_{t-1} + \ldots + \theta_q e_{t-q}) + e_t; \quad (1)$$

$$y_t - y_{t-1} = c + \phi_1 (y_{t-1} - y_{t-2}) + \ldots + \phi_p (y_{t-p} - y_{t-p-1}) + (\theta_1 B e_t + \ldots + \underbrace{\theta_q e_{t-q}}_{\theta_q B^q e_t} + e_t);$$

$$y_t - B y_t = c + \phi_1 (y_{t-1} - B y_{t-1}) + \ldots + \phi_p (y_{t-p} - B y_{t-p}) + (e_t (1 + \theta_1 B + \ldots + \theta_q B^q));$$

$$(1 - B) y_t = c + \phi_1 (1 - B)(y_{t-1}) + \ldots + \phi_p (1 - B) y_{t-p} + e_t (1 + \theta_1 B + \ldots + \theta_q B^q);$$

$$(1 - B) y_t = c + \phi_1 (1 - B) B y_t + \ldots + \phi_p (1 - B) B^p y_t + e_t (1 + \theta_1 B + \ldots + \theta_q B^q);$$

$$\underbrace{(1 - B)^d y_t}_{\substack{differencing \\ d\_\deg ree}} \cdot \underbrace{(1 - \phi_1 B - \ldots - \phi_p B^p)}_{AR(p)} = c + \underbrace{e_t (1 + \theta_1 B + \ldots + \theta_q B^q)}_{MA(q)}.$$

The ML model clearly demonstrates its low interpretability and black box characteristics. It is hard for an average business user to understand the connections between input and output data. Moreover, for the average business user when confronted with the presented model, the question emerges! What is Equation (1)? We could argue that Equation (1) represents the rules from Figure 4.1, generated by ML data & knowledge pipeline, which reveal the mystery about data-generating processes in a given SC business setting.

At first look, the developed ML model in Equation (1) doesn't seem to improve our understanding of the data. We are still confused as with Figure 4.7, but the ML model has a crucial advantage over the figure. In essence, the ML model is a **mathematical formula** which may not be easily understandable to a human user but is completely understandable to a computer, which **can be programmed to use a given formula** and make **business decisions** based on discovered rules.

REFERENCES

1. Chollet, F. (2021). Deep learning with Python. Simon and Schuster.
2. Equilibrium AI (n.d.). Equilibrium AI Data Pipeline [available: https://eqains.com/, access: January 23, 2024]
3. Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). The elements of statistical learning: Data mining, inference, and prediction (Vol. 2). Springer.
4. Jordan, M. I. & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), pp. 255-260.
5. Mirčetić, D., Ralević, N., Nikoličić, S., Maslarić, M. & Stojanović, Đ. (2016). Expert system models for forecasting forklifts engagement in a warehouse loading operation: A case study. Promet-Traffic & Transportation, 28(4), pp. 393-401.
6. Ng, A. (2017). Machine learning yearning. [available: http://www.mlyearning.org/, access: January 23, 2024]
7. Rostami-Tabar, B. & Mircetic, D. (2023). Exploring the association between time series features and forecasting by temporal aggregation using machine learning. Neurocomputing, 548, 126376.

8. ScienceSoft (n.d.). Microsoft Business Intelligence to Drive Robust Analytics and Insightful Reporting [available: https://www.scnsoft.com/services/business-intelligence/microsoft, access: January 23, 2024]
9. Wickham, H., Çetinkaya-Rundel, M. & Grolemund, G. (2023). R for data science. O'Reilly Media, Inc.