3 SPECIAL EDITION

BUSINESS

COACHING

# THINKING THROUGH DATA

"

Without big data analytics, companies are blind and deaf, wandering out onto the web like deer on a freeway

— Geoffrey Moore

# EDITORIAL

The editorial team of the "Business coaching" magazine is constantly trying to improve the content, design, but also the format of the magazine as well. Thus, in addition to the paper version, we also created a website where we first published the content of the paper versions. Later, as we grew along with our readers, there was a need for articles to be published more often and with topics that we did not cover in the paper edition.

We experimented with formats in June, when we worked hard to "revive" the magazine in an electronic, interactive version.

The latest format that we are publishing this time is the "special edition" format of the magazine. This magazine is smaller in format, but we tried to keep it large in content and design.

Why this format? We wanted to collect and share with you all the articles published by the authors only in the online edition as an added value to your knowledge. The thematic edition of the magazine always focuses on a single topic, and this time it is Data Science.

To learn more, please visit us:

www.bcoaching.online

# CONTENT

# ABOUT **AUTHOR**

Valentina Djordjevic is a huge Data Science enthusiast. Graduating from the Faculty of Organizational Sciences, she faced the challenge of packing her (then irreconcilable) interests into a unique career path-cutting machine. Still, she is a real "fox" when it comes to her areas of interest, so we have no doubt that this has given her good signposts. Valentina finds motivation in challenging projects and in working with equally enthusiastic team members. She believes that working on the Data Science challenges helps her to continually grow, examine (her) boundaries, and develop her imagination. Still, she has not yet determined whether the fact that she has never solved the same problem in exactly the same way is a good or bad thing. She hopes her colleagues, or maybe someone in the community, will help her finally come up with the correct answer.

# DATA SCIENCE & CO.

Unless you have spent the last ten years trapped into the deep cave surrounded by nothing else but darkness - the chances are you have heard of Data Science. You may have even tried to reveal the mystery of what Data Science really is - and found everything and - nothing at all. Why is it so hard to find a single point of truth? Well, we'll try to demystify this for you in a series of texts dedicated to this field.

**Data Science** is a broad field of study of phenomena within some specified domain, observing their occurrence and behavior. The occurrence of these phenomena is best described by data. Being a "science of data", it has the basic elements of scientific and methodological principles, such as defining hypotheses, testing them by conducting experiments, rejecting and formulating new hypotheses, all with the aim of expanding existing, and creating new knowledge. In addition, it is very important to understand that this field involves different stages, from data collection, data consolidation and storage, to processing itself, which involves exploration and modeling, that should describe some past patterns in behavior, and generate future patterns of behavior. This approach allows not only the analysis of the past, but also getting insights into the future, which is crucial for any business, as well as the decision making process and strategic planning. Precisely because it gives us the possibility to answer the questions of what, where, when, why, and how - data science has become a silver bullet for every manager aiming to leverage the business.

## THE EVOLUTION OF DATA SCIENCE

The evolution of Data Science is mostly directed by the evolution of computer science. Today we have enough processing power and memory capacities to perform complex computational tasks in a reasonable amount of time. What this means for the business is that analytics is now not limited to manual work and human ability to dig into a bunch of papers presenting some numbers. We have machines that can give us answers within minutes, or even seconds.

**Computational efficiency** enables the usage of more sophisticated methods of analysis, like machine learning algorithms. **Machine learning algorithms** are able to detect patterns within data, and provide predictions of some future behavior. In a competitive world as it is today, the right timing is crucial for companies that aim to keep on track and prosper, while the possibility to get projections of the future means having the power to always be one step ahead of the competition, but one step ahead of the customer as well.

In the world of analytics, there are three main types of analysis - descriptive, predictive and prescriptive analysis. In traditional systems, the analytics is mainly focused on reporting and descriptive analysis. Descriptive analysis is good for analysing the past and evaluating the outcomes of some decisions. But it does not provide the answers to what is going to happen. That's where the predictive analysis steps in. In predictive analysis, we deal with the projections of the future, in order to define next actions and make decisions. Prescriptive analysis is one more step ahead - provide projections of the future, taking into account different what-if scenarios, for a given context defined by its conditions and constraints. And that is what Data Science is all about - predictive and prescriptive analysis, often being referred to as advanced analytics.

Another great progress of advanced analytics is moving from silo-oriented to data lake architectures. Companies are now using data from different departments and channels, in order to get a 360 view of business. However, having "big data" sets does not guarantee the power and dominance, but it can be very useful in order to analyze business from different contexts and to conjoin that information, in order to determine how they affect each other, and how can the malfunction of one part produce low performances on other parts, etc.

In the series of posts that follows, we will try to dig deeper into the world of Data Science, machine learning and artificial intelligence, so stay tuned.

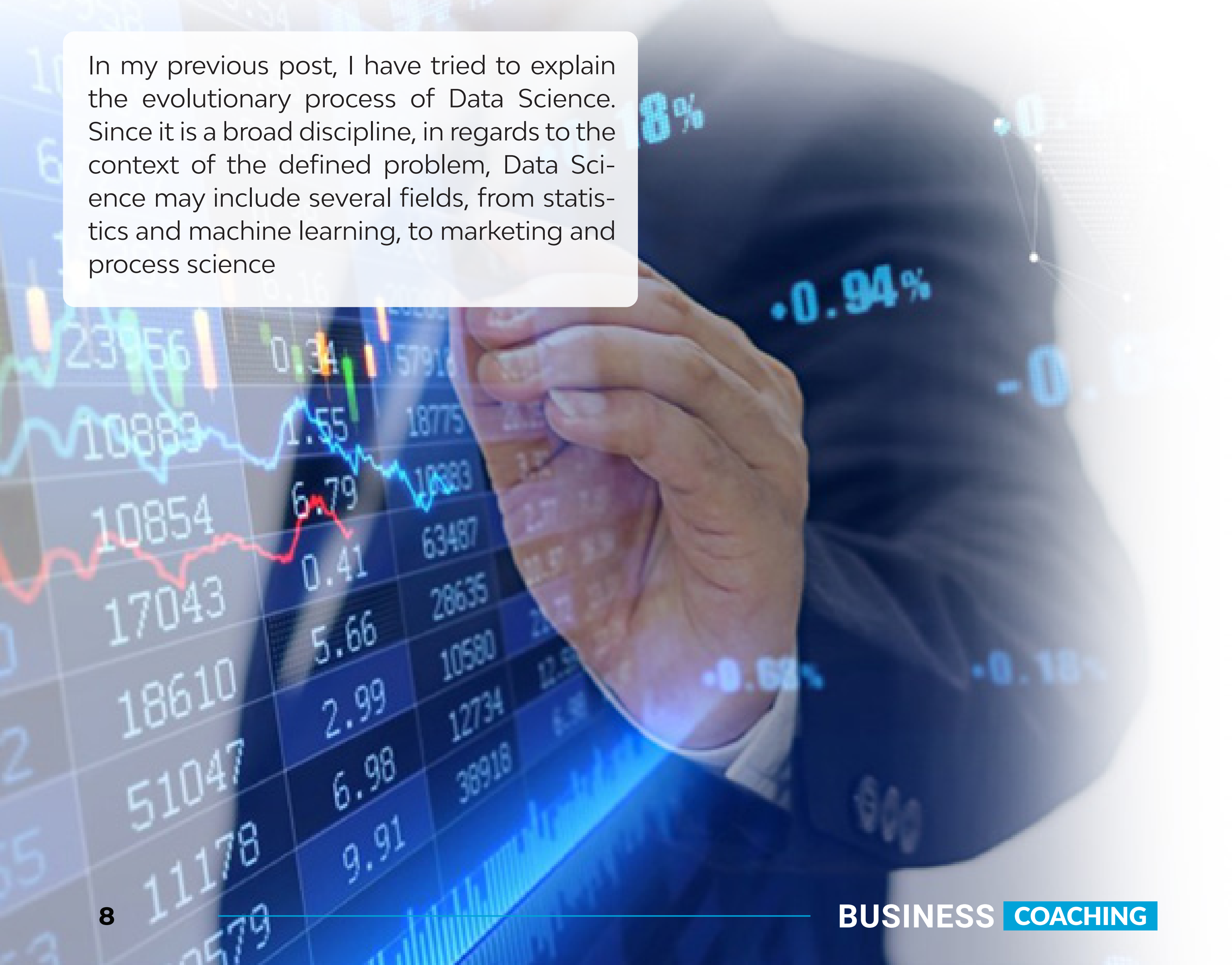
❝

You can have data without information, but you cannot have information without data.

- Daniel Keys Moran

# DATA SCIENCE IN BUSINESS

In my previous post, I have tried to explain the evolutionary process of Data Science. Since it is a broad discipline, in regards to the context of the defined problem, Data Science may include several fields, from statistics and machine learning, to marketing and process science



In general, Data Science is a multidisciplinary field which includes combining business with statistics and software development. What does that mean? In the business world, Data Science is all about data and turning it into a real value. The possibilities are endless. Using predictive analytics for what-if scenarios and decision-making purposes has grown to be the essence of strategic planning.  On the other side, advanced analytics is crucial for the optimization of operations pillars within the organization. But how is all that being realized?

Well, the first step in Data Science is to define a business problem that needs to be solved. For example, a business problem may be defined through some KPIs measured. Like low conversion rate, increasing churn rate, or high operational costs. There are methodologies explaining what does it mean to define a business problem, but in short - besides the identification of the problem itself, it includes broader definition - in terms of all the data sources and KPIs that describe the problem and should be used in the analysis. It is also important to answer the main questions - why that particular problem should be solved, and what are the main benefits? One of the most frequent mistakes managers do is that they catch a buzzword (that in most cases they don't fully understand) and believe that it is a magical wand that could solve every problem. What is important is to prevent the "l'art pour l'art" kind of objectification.

The solution of the business problem lays within data. It is often said - data tells a story. One should just dig deep enough, and the answers will be revealed. However, in reality - that is not always the case. Data only reflects history. Which in most cases is driven by some previous actions and reactions, defining a given set of circumstances, by which a defined problem is observed. Statistics is there to help us ensure we have a representative dataset and understand the main dependencies among the KPIs. By using statistical analysis, we can determine the predictive power of the dataset and understand its constraints. Furthermore, statistics is the heart of machine learning and a crucial part of advanced analytics, through which we can analyze patterns within the data, model the historical behaviour, and get projections on the future.

**Software development** plays an important role in Data Science. Managers often want solutions that are reliable, scalable and automated - which requires incorporating software development into the analytics process.

That means that all the algorithms should be placed within scripts, wrapped in fancy containers and callable though some API services. This is where one of the biggest problem of applying Data Science in business shows up - managers think that Data Science can be handled by one person only, while it should be the a task for a whole team, including individuals with different set of complementary skills, from business and soft skills, through programming and IT, to math and statistics. Another issue is that a lot of organizations suffer from a lack of data literacy, no defined data strategy and assessment, and poor (or no) experience with the business digitalization and transformation. These are the main prerequisites of applying Data Science in business - successfully.

When done right, the applications of advanced analytics in business context are numerous. From predictive analytics, to what-if scenarios and process mining and optimization. From retail, banking, insurance, to oil and gas, aviation, telecommunications,... Everything you can think of - can be a fertile ground for advanced analytics. The idea is simple - the data tells a story. A story that may give the answers to the questions like:

- **how to increase customer satisfaction/ improve customer experience?**
- **how to maximize ROI (return of investment)?**
- **how to optimize CAPEX (capital expenditures?)**
- **what can be done in order to increase performances?**
- **why is the churn rate so high?**

This list is only a short preview of questions managers often ask. However, the advanced analytics and the existing methods are not almighty, and not everything can be answered by looking into some data samples, but one thing is for sure - "all models are wrong, some are useful", and that is what it is all about, in this infinite game of enduring in the business - to search for the useful models and apply them in order to leverage the business.

In the next blog post, we will talk about Machine Learning and its applications, as well as its connection to Data Science and Artificial Intelligence. Stay tuned!

“

Numbers have an important story to tell. They rely on you to give them a voice.

- Stephen Few

# MACHINE LEARNING IN BUSINESS

***"A field of study that gives computers the ability to learn without being explicitly programmed".***

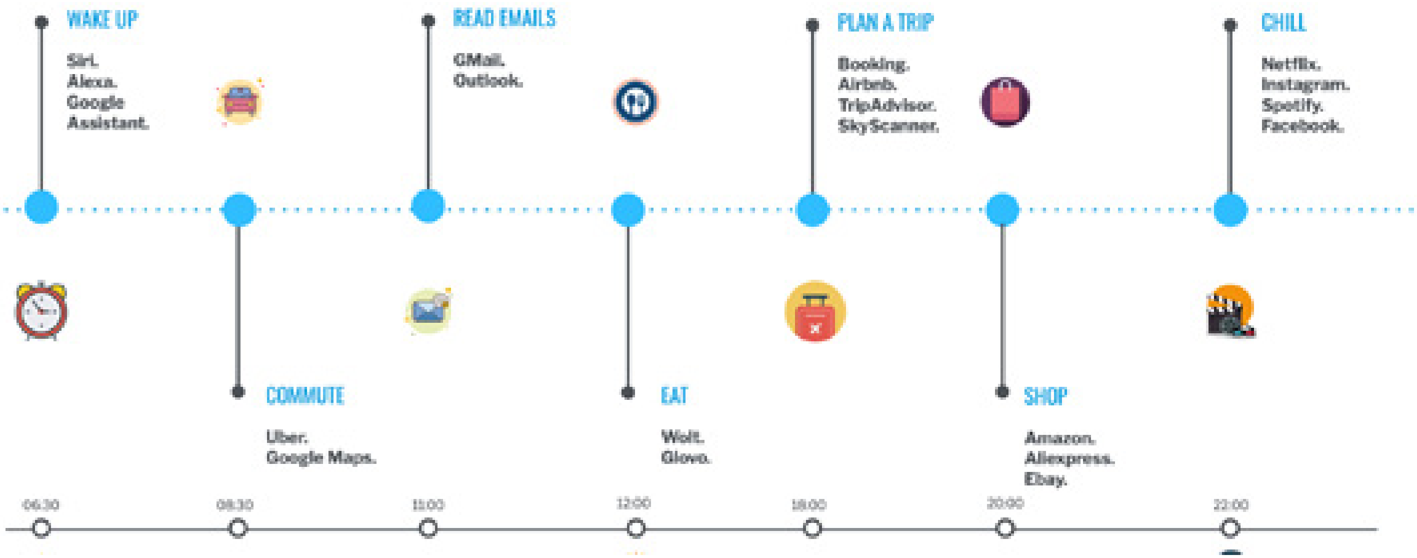
That is how Arthur Samuel defined machine learning by the end of the 1950s. As one of the pioneers in the field of Artificial Intelligence, he was the first to popularize this term as one of the subsets of AI. On the other hand, **Machine Learning** is a set of techniques used in the field of Data Science, for extracting insights.

Technically speaking, machine learning includes defining a mathematical model which represents the problem we want to analyze, in order to find an approximate solution.
In most cases, machine learning models are prediction-based: trained on some (hopefully representative) sample of data, they are learned to provide an output (used for decision making purposes, for example - the answer to the question "Is a particular customer going to churn?), based on some input features (for example, a set of attributes describing the customer, his preferences and activity level). This is the so-called supervised learning.
There is also unsupervised learning - where the output is not predefined, but extracted as insight from the algorithm (for example, in clustering we don't know what the clusters are, prior to running the model, but the model itself identifies groups of similar instances, called clusters and returns the cluster labels as an output).
Besides these two groups, there are other variations on the topic - like semi-supervised and reinforcement learning, as hybrid and modified versions developed to tackle other types of problems.

The choice of the appropriate algorithm depends on several things: a problem being defined, data and its nature, model pre-assumptions, computational power and resources needed. Enough with the technicalities - what does machine learning mean to the business today?

Let me depict the ML application in one picture (a slight digression: all the credits go to my dear colleague, mr. Milos Milovanovic, for drawing this picture, making this easier for me).

# DOES MACHINE LEARNING DRIVE OUR LIVES?




This picture is a showcase of our everyday life. It gives a sneak peek on how the biggest leaders in the AI world are using machine learning. You wake up, commute to the work, eat, plan a trip, etc - without even noticing how every application that you use is driven by one or more machine learning algorithms developed to customize and personalize the content you are watching/using, in order to leverage your experience and satisfaction. For example, **Google** integrates ML in all its applications, for natural language processing purposes (Google search engine, speech recognition and language translation), image processing (Google photos) and traffic prediction (Google maps).

**Uber** sets dynamic drive prices based on the traffic jams and demand rates.

**Porhub** uses computer vision and recommender systems to place the right content for its visitors.

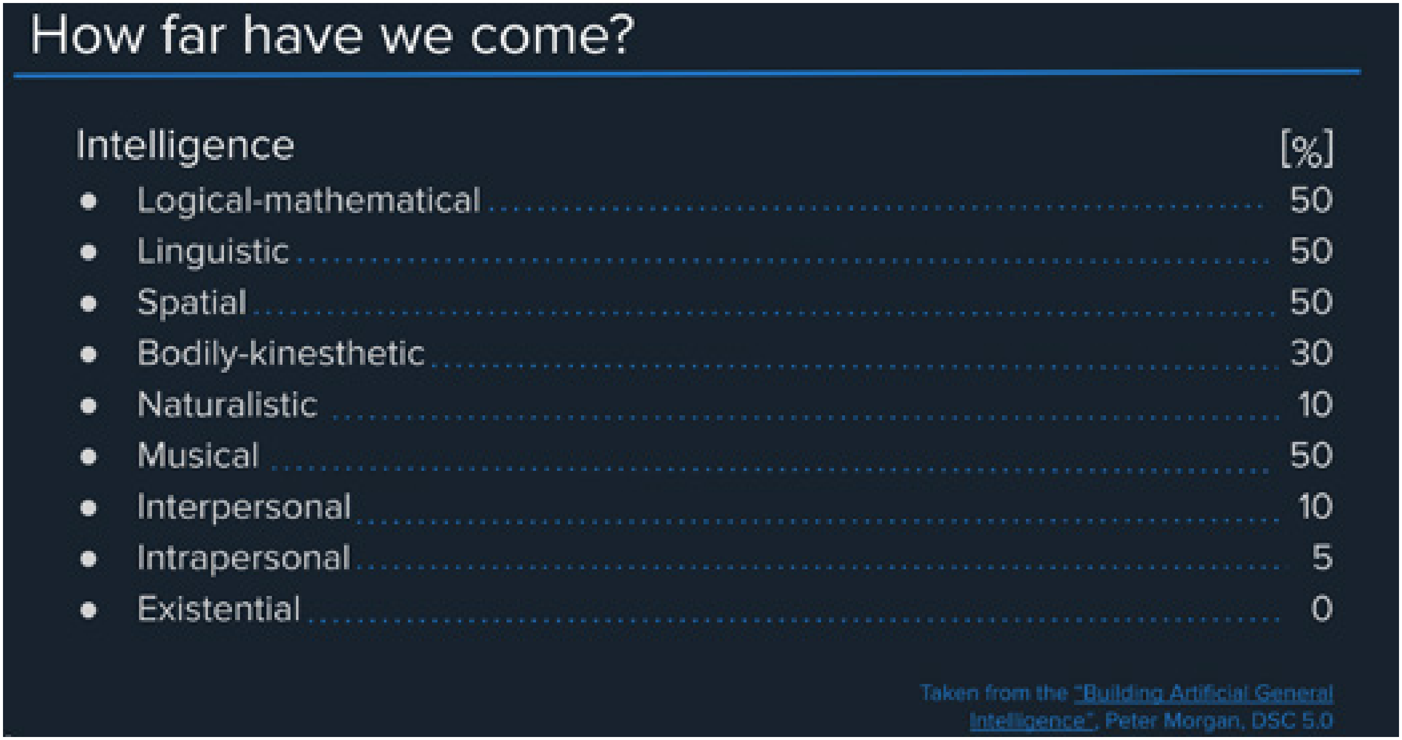
**Netflix** is also using collaborative filtering recommender systems for tailor-made recommendations to the users.

**Facebook** is one of the biggest leaders in AI research, integrating ML in all its services, while simultaneously open-sourcing some of the algorithms (like Facebook Prophet, used for forecasting).

More and more companies are becoming aware of the importance and benefits of ML applications within the business. Applications often include:  extracting predictions of the future - classification, regression and forecasting; and extracting insight explaining underlying patterns, similarities and drivers of some events that happened. Having this as a tool in the decision-making process can be a really powerful asset that could affect the perception of the business and its future.

However, machine learning strongly relies on the experience. It requires lots of historical data and lots of observations - in order to catch all the possible patterns. Machine learning is only as strong as the prediction power of the data is (if the dataset is not representative, or there are not enough observations, the model will most possibly fail).

We already mentioned that ML is a subset of one much broader discipline - Artificial Intelligence. **Artificial Intelligence** is a branch of computer science that deals with (ML) algorithms inspired by various facets of natural intelligence. It is a system able to completely autonomously perform tasks that normally require human intelligence, like visual perception, speech recognition, problem solving, language translation,... Besides being autonomous, it is able to make decisions and define actions in dynamic environments, where the conditions and constraints are changing (which is not the case with traditional Machine Learning, strongly relying on the experience and already seen patterns). The next picture represents the level of intelligence we've artificially reached so far.




However, as much as we talk about it nowadays, we are extremely far from the AI at its finest. Thus, in most cases, when companies talk about AI, they usually refer to the application on machine learning and switching to data-driven decision making (a.k.a. Data Science).

I would like to conclude this post with one quote authored by one of the biggest businessmen in the world - Ray Dalio.

***"If the future can be different from the past and you don't have a deep understanding, you should not rely on AI."***

Machine learning can be powerful, but only when wisely taken and deeply thought through. Being aware of this could make a significant difference between success and failure of integrating machine learning in business.

“

Data Scientist is a person who is better at statistics than any software engineer and better at software engineering than any statistician.

- Josh Wills

# DATA SCIENCE IN MARKETING

In the previous posts, we got introduced with Data Science and some of the most famous use cases from the industry leaders. In our future posts, we will talk about the applications in specific industries. Today on the menu - marketing.

**Marketing** is one of the most fertile fields for Data Science. There are many different techniques and directions which marketing teams are striving to realize, but the two most frequently discussed are - up sell and cross sell. For those of you who are less familiar with the terms, up sell stands for the action of convincing a customer to buy (or upgrade to) a more expensive product/service, and thus - spend more money, while cross sell stands for making a customer buy one additional diverse product/service besides the one already chosen and thus - spend more money. I will leave the further explanation here, for the reference.

But is there a missing link that could enhance this process of "convincing" a customer to make a purchase, and increase upselling and cross selling? If you ask me - that is Data Science. To be honest - we are light years away from the period when there was no big deal if we lose a customer. Nowadays - things are way different. The situation on the market is as follows - numerous and strong competitors, mass production, high customer acquisition costs, high churn rates and - poor loyal relationships. And the evergreen claim - it is less expensive to keep the current customer, than to acquire a new one.

The attractiveness of Data Science to the marketing teams comes from several factors, all sinking into one funnel: tailor-made marketing. Life would be much easier if you could know what the customer would like to buy, right? But then again, it's not only about the preferences, but the ability to afford it (from the financial aspect). Besides, maybe he cannot afford it currently, which does not mean he could not afford it by the end of the month, when he gets his salary. On the other hand, maybe he's waiting for the special discounts and actions... How would you know? The list goes on and on. What to do with all the pressure? Well, look into the data, there lies the answer.

The main benefits of applying advanced analytics and machine learning in marketing are:

- **personalization - understand customers preferences and needs, in order to determine the content, frequency and channel to target them and nurture their customer journey**
- **prioritization - identify high-value customers and improve decision making process including time, resource and budget allocation**
- **performance - monitor and evaluate transitions and progress over the time, in order to quantify how promoters and detractors are growing or shrinking through the time**
- **predictiveness - the possibility to project and influence the future customer behaviour, based on their historical behavioural patterns, preferences and habits**

It is clear that there is no all-mighty algorithm, or tool that can give you the answer to all these questions, in such a way that you could use that information to solve all the problems. Here are few ideas on how marketing could leverage the underlying campaigning processes.

## RECOMMENDER SYSTEMS

Recommender systems are maybe the most powerful, yet certainly the most attractive, asset that can be embedded into campaigning tools. The main purpose of using recommenders is to analyze customers preferences and similarities, in order to evaluate which products are highly likeable to be bought by the customer in the future. This makes recommenders a good call for cross-selling campaigns. You can extract the rating from recommender engines describing the affinity level of the customer towards each product, and use that information to create the "next best offer".

## SEGMENTATION

Segmentation (or in machine learning terms - clustering) is used for grouping customers with similar characteristics and purchasing habits into small segments. There are plenty of segmentation types that can be developed - activity segmentation, behavioural segmentation, brand awareness segmentation, etc. The type of the segmentation depends on the information we would like to obtain and the way in which we would like to use it. Dividing customers into small segments makes the process of prioritization and focused targeting a lot easier (for example, one would define a different campaign for the segment of "overachievers" and the one with "dormant" customers).

## PROPENSITY TO PURCHASE

Analyzing the probability that the customer will make a purchase in the future can further direct the actions taken towards. In a nutshell, based on the historical data of activity level, spending habits, periodicity, preferences and final outcome (made purchase, didn't make purchase) - an algorithm can be trained in order to retrieved the probability or a flag indicating the propensity to purchase in the future, given current state describing customers features.

## SURVIVAL ANALYSIS

The dark horse of tailor-made marketing certainly is - the survival analysis. The main goal of this analysis is to estimate the time to a given event (e.g. purchase), and to quantitatively explain how this time depends on various properties of the treatment (campaign), customers and other variables. Sometimes, having estimated the propensity to make a purchase isn't enough. If we want to create and send a promotion - we need to know when is the right time to do that. Why send the promotion, if a customer will come and make a purchase either way? On the other hand - if you wait for too long, the customer will give up. This time estimation can help target customers in a timely manner.

## CUSTOMER LIFETIME VALUE

Estimating customer's lifetime (monetary) value is performed by taking into account all previously mentioned features - spending habits, periodicity, activity level, spending trend, demographic features, etc. The information of the expected revenue from the given customer in some future period can help in identifying high-value customers and the potential that can be (out) reached if the customer is well nurtured and targeted. CLV can be a valuable insight for maintaining the cash flow and strategic planning regarding marketing and sales campaigns.

# Many, many more...

Integrating all these use cases and joining their outputs could help in wrapping up the answers to these questions:

- **what to target the customer with?**
- **how important is this customer for our business?**
- **is this customer going to make a purchase in the future time period?**
- **when is it expected for this customer to make a purchase?**
- **how much money is the customer likely to spend in some future period of time?**
- **how sensitive is this customer to the specific campaign/discount value?**

Having answers to all these questions means being able to define and direct future marketing activities in order to differentiate from others and nurture loyal relationships with customers by placing them products, content and promotions they could really be interested in. The main prerequisite is to find the answers though clearly defined use cases, and well engineered and developed machine learning algorithms.

There are plenty of other use cases that can be defined and developed in order to leverage marketing processes focused on upselling and cross selling. Hopefully this post will wake up some new ideas in which data science could be applied in marketing.

"Every company has big data in its future and every company will eventually be in the data business.

- Thomas H. Davenport

# AGILE & DATA SCIENCE AS A PERFECT MATCH?

As I got familiar with the agile manifesto methodology a few months ago - agile values and principles, the most interesting thing for me was to think about its application within Data Science projects. The idea is very simple, however, I am deeply convinced that each of the values has its own disclaimer - applicable in one context, not in another, because of this and that, etc. **The Agile Manifesto** as a step forward in relation to traditional methodologies such as the so-called "waterfall" approach, should make it possible to reduce the difference in communication that exists between the client and the vendor, while on the other hand - respond to the increasingly rapid development of new technologies, and introducing changes in initial requirements, resulting from work dynamics, digitalization and competitive market.

Let us recall agile values through the prism of Data Science.

**Individuals and interactions over processes and tools.** In Data Science, interactions are key to understanding and defining business problems, that is - extracting maximum value based on analysis. It is very important to keep a critical mindset, and point out irregularities. Some insights often lead to a change of direction and resolving things that were not defined by the scope, which sometimes means going out of the defined process. On the other hand, limiting analytics to certain tools and technologies can result in truncated analysis and unusable insights.

**Working software over comprehensive documentation.** This is perhaps the value that can be most discussed through the prism of Data Science. Having applicable software and delivery is very important, but writing detailed documentation and explanations of how the data was sampled and prepared, which models were used and why, what is behind those models and how to interpret their output, what are the expected performances - all this it is very important to elaborate in detail in order to ensure the value in addition to the delivered solution.

**Customer collaboration over contract negotiation.** As in any software development project, having good collaboration with the client is a prerequisite for everything. In Data Science, this is very important, both for understanding the domain through interaction with the client, and for the interpretation and testing of the solution that is delivered by the client. Since it is a specific area, it is very important to establish cooperation with the client, which involves making joint efforts to create a solution that will have a value - most often the value of the solution directly depends on the knowledge of domain experts.

**Responding to change over following a plan.** This is where Agile and Data Science coincide the most. It is very often the case that as a result of the analysis a new idea of future steps, improvements or adjustments of the existing plan is awakened, and therefore it is necessary to be agile and not adhere to a blindly defined plan, in order to successfully respond to such requirements. The plan is very important, but it becomes obsolete and unusable as the goals are redefined and changed.

To make it clear - the fact that Agile and Data Science, as I characterized them in the title, are a “perfect combination” does not mean that they always lead to a perfect outcome in realization. What I wanted to emphasize is precisely that Agile allows Data Scientists to be - Data Scientists. This means that they can dedicate themselves to research, to change the direction of movement and redefine goals, depending on the course of analysis and the insights gained, to work closely with clients in attempts to find a solution, and so on. Further, if we are talking about agile principles (the famous twelve), there is a good chance that every Data Scientist / developer will agree with each of them at first. That is the beauty of agile principles - they are defined so that they can be successfully applied to any project. When you think about it more thoroughly, there are some principles that are debatable - e.g. a principle that says that the best architecture, requirements and design come from self-organizing teams. I believe in this. But one very important precondition for this is - the way these teams are made. If there are no people in that team who are adorned with innovation, “growth mindset”, autonomy and responsibility - it is very likely that this idea will fall apart. Simply, it usually happens that the teams are made - as they had to be made, and sometimes it is evident that the team lacks a leader who will supervise and lead it - which is not how self-organizing teams work. I could do this about each principle individually, but I will only dwell on this, and allow you to think about the pros and cons of each (or situations where a principle could be challeged)

However, there are several (more serious) problems that can occur as a result of this compound, and they are:

- **poor and pruned (or even no) documentation of the research process, because the focus is on insights and results - which can be a problem if someone else needs to be involved in the process**
- **very frequent changes in requirements can take the analysis in a completely different direction, which makes it difficult to define "acceptance" criteria and deadlines - sometimes the process of developing a module takes several months (unnecessary)**
- **clients do not always have an understanding of the lean results of predictive models, which then affects communication and the quality of collaboration**
- **also, clients often believe that Data Science is a magical weapon that will solve all their business problems - which in turn affects communication, the quality of collaboration and the practical use of the solution**
- **Data Scientists often have a problem with feeling a lot of pressure - their solution is difficult to materialize, and when it is materialized, it is critically dependent on the input data, which they cannot influence**
- **communications on a daily basis can be depressing, as it often happens that in some Data Science tasks there is no significant progress for several days in a row, where the idea of frequent and incremental shifts is lost**

Since Data Science is so diverse, depending on what the Data Science project you're working on includes - the ability to apply agile methods will vary from project to project. If you are working on product development, Data Science in that sense becomes a niche of software engineering, where the application of agile methodologies and scrum prove to be very useful. On the other hand, if you are working on one-time projects or solutions - the application can be much milder and meaningful only in certain phases. The most important thing is to recognize what are the good sides that you could use, in order to improve your way of working and achieve the best possible results.

# Business and People Development

www.cetbalkan.com