Wealth Management and Financial Data Science: a short guide

How Financial Data Science improves productivity in Wealth Management

VIRTUAL B

WHITE PAPER

CONTENT

The Financial Data Science Frenzy 3

What is Financial Data Science 6

Data is everywhere: where, what and how 8

Machine Learning, automation and artificial intelligence 10

Supervised Learning 12

Unsupervised Learning 15

Reinforcement Learning 17

A Holistic approach to Financial Data Science 21

Forget investing and let’s focus on ROI 24

How Financial Data Science Helps Compliance & Product Governance 26

How Financial Data Science Helps Marketing 28

How Financial Data Science Helps Communication 32

How Financial Data Science Helps Financial Advisor 33

How Financial Data Science Helps Executives 35

Putting Financial Data Science to work 36

The Financial Data Science Process 37

The team: not only pure Data Scientist 39

In-house or outsourcing? Make or buy? 40

Our Solution: SideKYC® 41

The Financial Data Science Frenzy

Data Science is everywhere. Data is the fuel of the digital economy. There has been an explosion in the velocity, variety and volume of data of all kinds: social media activity, mobile interactions, real-time market feeds, customer service records, transaction details, information from existing databases – there's no end to this tidal wave.

Compared to the past, nowadays companies can deal with this huge amount of data sets by using Data Science techniques thanks to four key trends:

- The cloud-based data storage;
- The declining cost of computing (Figure 1);
- The improving computing power;
- The sophistication of analytic tools (often open-source, and relatively easy-to-use).

Technology has never been so cheap

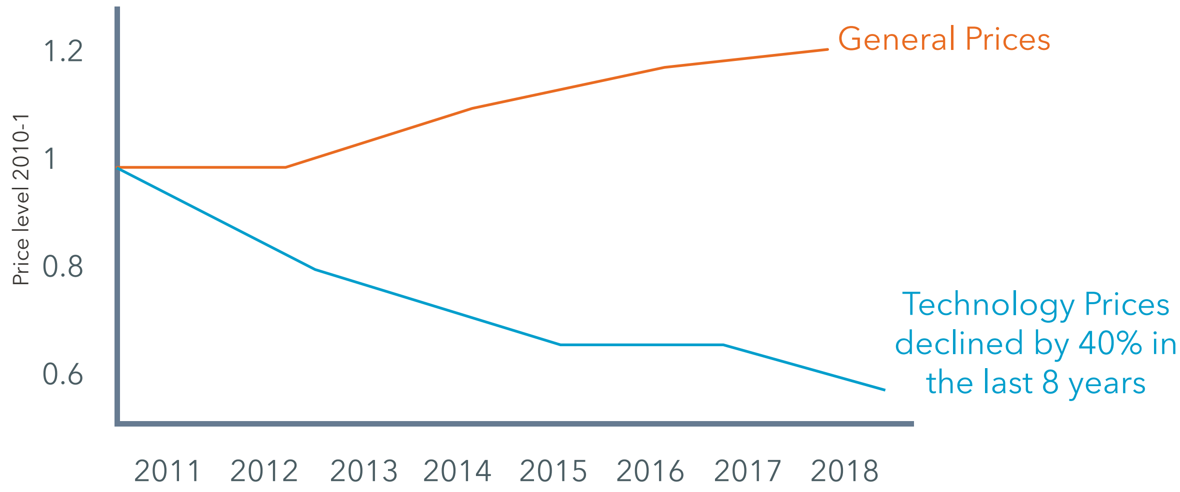


FIGURE 1: The US Bureau of Labor Statistics tracks prices for broad categories of goods over time and, as this chart shows, prices have dropped dramatically in the technology sector.

Virtually every sector of the economy now has access to more data than would have been imaginable even just a decade ago. The financial industry - and thus wealth management, which is the main focus of this white paper - is not an exception.

Since the great financial crisis of 2008, the financial industry has been subjected to more and more intense regulatory scrutiny, enforcing and introducing numerous strict rules. Now, banks, insurance companies and asset managers face huge challenges to revise their compliance and governance infrastructure to meet regulatory criteria in a proper and timely manner.

Thus, the applications of Artificial Intelligence and Machine Learning in the asset and wealth management industry are endless, and can impact the entire value chain. Basically, a confluence of circumstances opens the way to face different wealth management issues, with the support of data and algorithms, such as:

- Capturing and analysing new sources of data;
- Coping with regulation requirements in real time;
- Understanding client needs and life-goals;
- Enhancing user experience at lower cost;
- Improving user segmentation and offering real personalised investment solutions and services;
- Building predictive models of client behaviour;
- Promoting targeting distributed models and doing a “people-based marketing”;
- Supporting advisory services;
- Exploiting lower net worth market segments;
- Running live simulation of market events, analysing their impact on costs and revenues, allowing for true risk management at an overall business level;
- And so on.

In short, this breed of technology named Financial Data Science boosts one thing: productivity. Hence, Data Science is the new fever in the financial industry. Wealth management is one of the fields where this new kind of know-how and technology has the most interesting and useful applications.

The aim of this white paper is to provide an overview of Financial Data Science, outlining how it can spread value in the wealth management business across the investor’s whole lifecycle.

What Is Financial Data Science

Financial Data Science is an incredibly broad subject, and means different things to different people. As many elements of the financial industry become automated and new techniques arise, the domain includes investments, trading, signal extraction, risk management, wealth management. But, essentially, and unsurprisingly, Financial Data Science is just Data Science applied to finance.

The field of Data Science corresponds to the intersection of the fields of Statistics, Information Technology, Computer Science, Social Sciences (such as Economics), and Data Visualization (DataViz). Data crunching alone is not enough to solve business problems, the domain's knowledge and the business acumen are essential. In general, Data Science needs a deep domain expertise and horizontal business knowledge, thus Business Analytics or Business Intelligence play a central role in the applications of Financial Data Science.

In our opinion, understanding the economics behind the data and the signals is more important than just developing complex technical solutions.

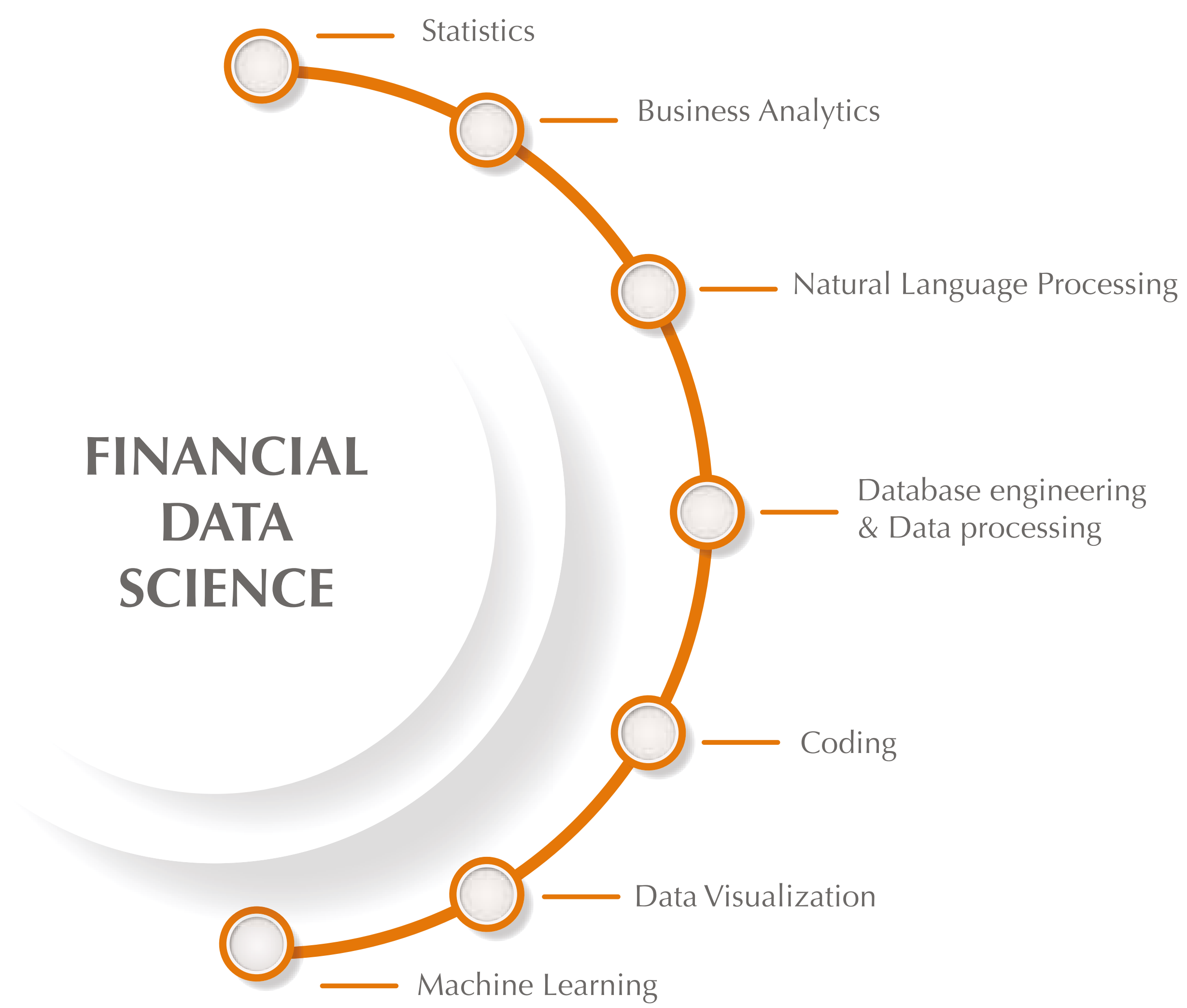


FIGURE 2: Financial Data Science is an interdisciplinary field that requires different skills.

Data is everywhere: where, what and how

Financial firms can extract useful information from various internal and external sources, including “alternative data sets”. Data can be structured (numerical/categorical) or unstructured (images, video, text).
Data sources may include:

- Client financial positions and transactions;
- Socio-demographic data, such as age, place of birth, sex, family situation;
- Mifid questionnaire answers;
- Legal documents, emails, instant messages, news archives, analyst reports, etc;
- Data coming from smart engagement activities, e.g. quizzes, gaming, simulation tools;
- Social media data, from Linkedin, Facebook, Twitter, etc;
- Data from the web and mobile devices;
- Payment/credit card streams – transaction data able to capture trends in consumer purchasing habits;
- Records of individual online experiences - e.g. web searches, product reviews, web searches (often obtained through scraping);
- Crowdsourced data (opinions from large groups of people, especially from online communities/specialised social networks offering insights from the “wisdom of the crowd”);
- Financial market data;
- Company-level data;
- Macro-economic data;
- Geolocation data;
- Sentiment data (often from external providers).

In order to provide sound results, the data should be gathered from all the available relevant sources. Business relevancy of the data sources is strategic. Thus, probably, collecting the data on tequila sales will not help to understand how to increase the sales of mutual funds.

Furthermore, following the “law of parsimony" and keeping the dataset size close to the minimally appropriate is crucial. Having more data is not necessarily a good thing if data is not properly stored, cleaned, managed, crunched and, last but not least, understood. It is mainly a matter of process, and business know-how.
Anyway, even leaving aside the alternative data sources, most financial institutions already own valuable datasets able to generate key business information.

A note on Big Data

There is a common misconception around the idea of Big Data: many people think that every time you apply predictive or otherwise advanced data analytic methods that extract value from data, it is Big Data Analysis. However, this is not the case.

Big Data can be described by the following characteristics, the so called “Three Vs”:

- **Volume** - The quantity of data determines whether we can talk of Big Data or not: the size of a genuine Big Data dataset is usually measured in zettabytes (10^21bytes);
- **Variety** - The type and nature of the data: Big Data includes structured data (numbers, labels), and unstructured data (e.g. draws from text, images, audio and video files);
- **Velocity** - The speed at which the data is generated and processed: Big Data is often real-time.

In accordance to the “Three Vs rule”, Big Data requires a set of dedicated technologies to store, handle and process data. Having said this, what counts as "Big Data" varies depending on the capabilities of the users, and expanding computational capabilities make Big Data a moving target over time.

Many Financial Data Science applications are not Big Data. They are just data. Often a lot of data. The value of data is not in amassing huge amounts of information, but in extracting actionable insights through sensible analysis of that data. This is the most important thing: extracting useful information and making worthy business decisions.

Machine Learning, automation and artificial intelligence

When facing real world problems all these disciplines tend to amalgamate: basically you collect, model, and interpret data, make projections and make business decisions. In Virtual B we love to think of Financial Data Science as a superset of all the disciplines working on financial data (latu sensu) using analytical tools.

At the very core of Financial Data Science is Machine Learning (ML). ML means using data (i.e. experience) to improve a system's performance by means of computations. So ML teaches computers to do what comes naturally to humans: learning from experience. ML algorithms use computational methods to "learn" directly from data, given a task, and the algorithms adaptively improve their performance as the number of samples available for learning increases

Machine Learning defintion

D = DATA
T = TASK
P = P(T,D) = PERFORMANCE

$$\frac{\Delta P(T,D)}{\Delta D} > 0 \Longleftrightarrow \text{Machine Learning}$$

FIGURE 3: Machine Learning implies that performance must be improved thanks to more data. We have Machine Learning when a computer program learns from data D, ie its performance at some task T, measured by a performance measure P, improves using some data D.

ML is about inference, automatic reasoning, and representing knowledge. ML needs automation, that is automated computing through a mathematical sequence of steps, i.e. an algorithm.

ML systems crunch raw data and generate an output, such as insights that help you make better decisions as well as predictions. ML systems are made up of four major steps, which are:

- **Model** - the mathematical/statistical model chosen to make forecasts;
- **Parameters** - numbers that define the structure of the model precisely;
- **Forecast** - the ML model takes an input and tries to predict an output;
- **Learner** - the system adjusts the parameters — and in turn the model — by looking at differences in predictions versus actual outcome.

So an ML system loops, in the following systematic way:

Essential Components of a Machine Learning System

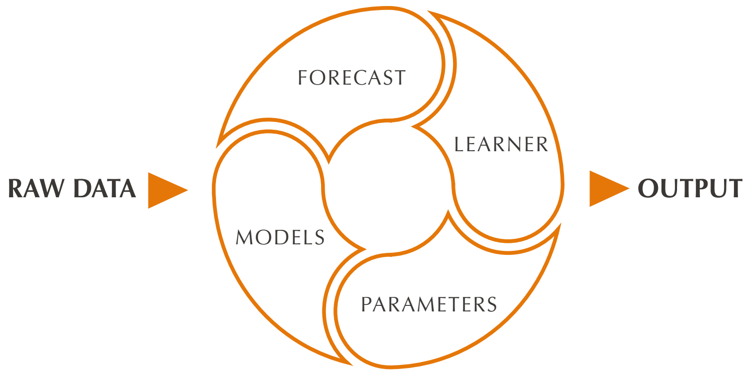


FIGURE 4

There are plenty of ML applications: in wealth management, for example, ML can be used to find a new cluster of clients or new investment needs and products; to create the right map between investment solutions and clients; or to analyse a client conversion funnel.

Today too many people tend to confuse ML with automation or Artificial Intelligence (AI), which is a common, yet big misunderstanding since they are not the same thing. Let's make some clarity.

Automation means automated computing through a mathematical sequence of steps, i.e. an algorithm. For this reason, automation is a process with minimal human intervention but with a huge difference with ML, because there is no learning at all. Therefore, on a practical basis, if, for example, you put a MIFID questionnaire online, it is not ML, it is just a useful automation of a repetitive task that goes digital.

Let's look at the difference between ML and AI. In order to really learn something from data (that is describe, classify, analyse and predict) ML uses algorithms and statistical models that, in some sense, look "intelligent". But, strictly speaking, it is not Artificial Intelligence (AI), because AI needs that the computer act like a human. This means being able to interact with humans, using Natural Language Processing (NLP), a specific Data Science field. To be more precise, NLP is a branch of Data Science that helps computers understand, interpret and manipulate human language. NPL is strongly based on Computational Linguistics in its search to fill the gap between human written and oral communication and computer understanding.

Examples of AI are visual perception, speech recognition, true decision-making. A well-known, classic application is ChatBot. Or personal financial assistants, i.e. something that helps clients, or financial advisors.

Machine Learning & Artificial Intelligence Map

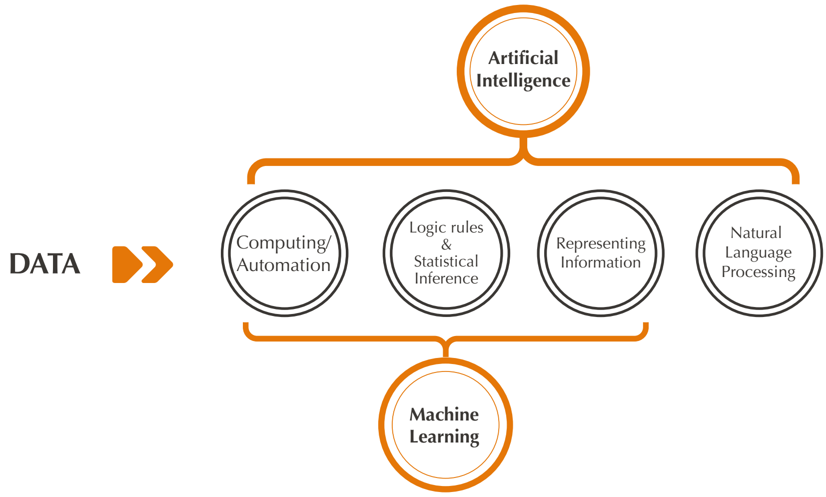


FIGURE 5: Strictly speaking, Artificial Intelligence is different from Machine Learning because it uses Natural Language Processing (NPL) – ie speech recognition, natural language understanding, and generation -allowing computers to act like a human.

Coming back to ML algorithms, they are typically classified into broad categories, depending on whether there is a learning "signal" or "feedback" available to a learning system.

Supervised Learning

What it is all about

In short: it is learning with a teacher.

Supervised Learning trains a model on given input and known response data (output), so that it can produce recommendations, prescriptions, or predictions. The situation is, you know how to classify the input data X and the output Y you want to predict, but you need the algorithm Y=f(X), a mapping function.

The algorithm, fed with input and known output, is trained on the data to find the connection f(.) between the input variables X and the output Y. Once training is complete–normally when the algorithm is adequately accurate, or when early stopping time is reached–the algorithm is applied to new data.

Let us assume you want to predict assets under management (AUM) changes based on clients' behaviour, financial prices, products' performances, etc. These are the main steps of Supervised Learning:

1. A person or an algorithm "labels" the known input data X (e.g. clients' behaviour, financial prices, products' performance, etc) and defines the known output variable Y (e.g. net AUM);
2. The algorithm is trained on the data to find the connection f(.) between the input variables X and the output Y;
3. Once training is complete and the mapping function y=f(X) is estimated with the desired accuracy, the algorithm is applied to new data in order to generate the expected output, e.g. to predict AUM.

Supervised Learning uses classification and regression techniques, that identify the map Y=f(X):

1. **Classification techniques** predict discrete responses (categorical variables)—for example, whether a client has a certain investment need, or whether that person is promising or not, or whether an investment solution is suitable for a given client;

2. **Regression techniques** predict continuous responses— for example, changes in assets under management due to market fluctuations and behavioural inputs.

How Supervised Learning Works

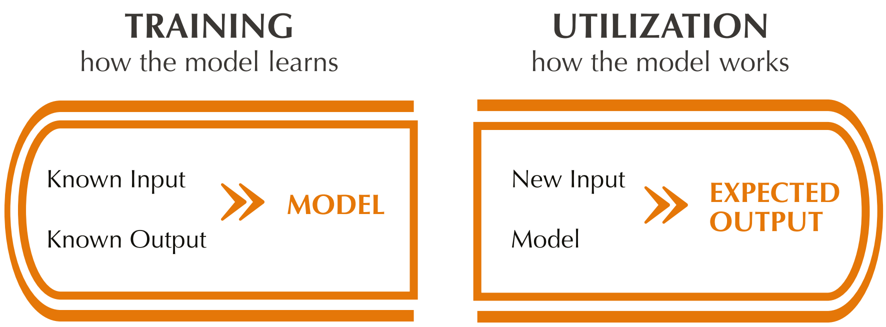


FIGURE 6: A Supervised Learning algorithm learns how to map an input onto an output based on example input-output pairs. It is learning with a teacher.

When to use it

Here are a few examples:

- Given former prospects labelled as good/not good (based on their past evolution as clients), learning how to identify the most promising prospects;
- Understanding assets under management (AUM) drivers, given financial prices, such as marketing campaigns, change in distribution policies, advertisement, macro-economic environment;
- Forecasting asset returns;
- Detecting fraudulent activities;
- Evaluating the probability of success of a communication campaign based on the knowledge of factors that might affect it;
- Given a dataset of clients having a given financial need/goal or not, learning to classify new, different clients as having that need/goal or not;
- Classifying clients based on how likely they are going to invest in a given financial product;
- Analysing market/client sentiment;
- Understanding financial product attributes that make it most probable to be purchased;
- Authenticating and onboarding through face/voice recognition.

Unsupervised Learning

What it is all about

Short answer: it is learning without a teacher.

Using Unsupervised Learning, an algorithm explores input data without being given an explicit output variable. Doing so, Unsupervised Learning finds hidden patterns or intrinsic structures in input data (such as similarities/dissimilarities). Analytically, you have X but you do not have Y, and the most general output of this kind of model is a mapping function g(X) which identifies which particular latent factor/structure underlies X.

The main aim of Unsupervised Learning is to interpret input data based only on input data, without a human to provide guidance along the way, thanks to the following steps:

1. The algorithm receives unlabelled input data X (e.g. data set describing investor behaviour);
2. The algorithm infers a structure g(X) from the data using mathematical modeling and statistical techniques, from scratch;
3. Using g(x), the algorithm identifies groups of data that exhibit similar behaviours (e.g. clusters of investors that exhibit similar behaviours).

Thus, if Supervised Learning is when the data you feed your algorithm is "tagged well", then Unsupervised Learning are types of algorithms that try to find similarities, correlations, associations, and other "structures" below the data surface, without any external inputs other than the raw data.

How Unsupervised Learning Works

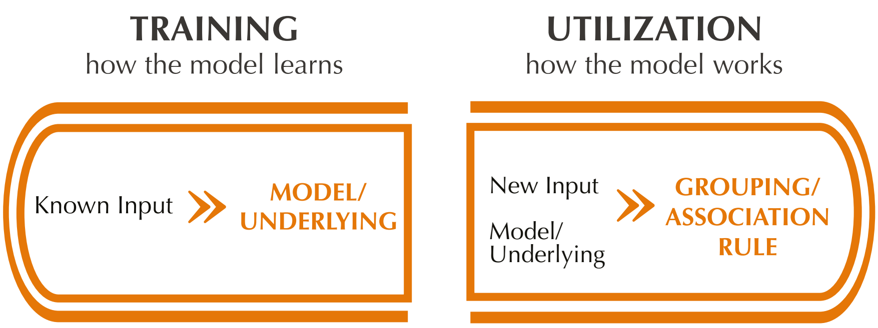


FIGURE 7: With Unsupervised Learning the algorithm infers a function that describes the structure of "unlabelled" data (i.e. data that has not been categorised or classified). It is learning without a teacher.

When to use it

Here are a few examples:

- Pattern recognition problems, e.g. finding latent variables that explain certain behaviours, such as subscribing/unsubscribing investments, or automatically discovering new market segments;
- Discovering groups of similar items within the data, e.g. clustering of financial instruments, or clients -the algorithm decides how to group items into classes that share common properties;
- Recommendation systems, e.g. recommending which documents financial consultants should read based on attributes of clients they are going to meet with, or recommending articles/news a client might be interested in, based on his actual investments and his financial know-how;
- Describing and simulating complex, probabilistic systems, such as networks of clients and financial consultants, or specific market segments (e.g. the Italian banking system);
- Looking for securities that are statistically similar to an illiquid security which is difficult to price;
- Data mining and dimension reduction, e.g. packing several raw market variables in a compact sentiment index.

Reinforcement Learning

What it is all about

Short answer: the algorithm learns to react to an environment.

In Reinforcement Learning (RL), the algorithm ("agent") is fed with an unlabelled set of data, then chooses an action for each data point, and receives feedback (from the environment, sometimes from humans) that helps the algorithm learn how to maximise some defined reward function – fundamentally, a kind of utility function, for those with a background in financial economics.

RL is a general class of algorithms in the field of ML that allows an agent to learn how to behave in a stochastic and possibly unknown environment, where the only feedback consists of a scalar reward signal coming back from the environment.

The goal of the agent is to learn by trial-and-error (during a loop) whose actions maximise his long-run reward. Thus, RL algorithms can be seen as computational methods aimed to solve complex sequential decision problems by directly interacting with the environment.

Reinforcement Learning

Basically, RL works in the following way:

1. The algorithm takes an action on the environment (e.g. makes a change in an investment portfolio);
2. It receives a reward if the action gets closer to maximising the total rewards available (e.g. the highest total return on the portfolio, i.e. the main goal);
3. The algorithm corrects itself, over time and over several iterations, searching for the best policy, that is, the best series of actions.

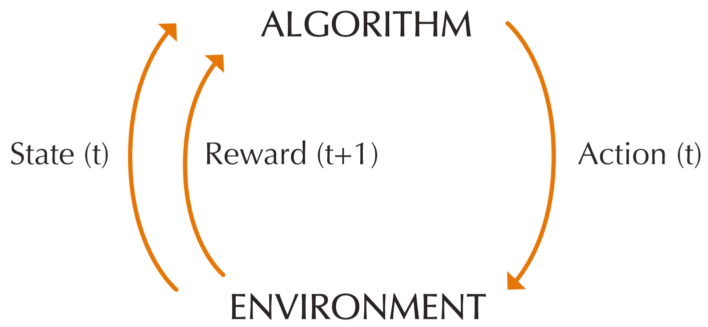


FIGURE 8: How Reinforcement Learning works.

When to use it

Widely used with success in online video games, where agents actively interact with their environment (where the environment is game-specific), RL is now entering finance, especially investing. It is quite a new frontier. Some examples are:

- Finding optimal investment strategies, e.g. the algorithm maximises points it receives for increasing returns of a portfolio;
- Optimal execution policies, e.g. the algorithm maximises points it receives for reducing trading costs and market impact.

A Map Of Machine Learning For Wealth Managers

Below you find a “cheat-sheet” of the ML techniques, with the main algorithms, and some business use cases.

Supervised Learning

EXAMPLE OF PRACTICAL USE	INPUT/DATA	ALGORITHMS	USE
- Customer segmentation - Clients product mapping - Product segmentation	Client positions, financial trades, MiFid questionnaire, basic personal data, legal documents, e-mails,instant messages, transactions/payments, social media, smart engagement,financial market, products, features, crowdsourced & institutional data	**Classification** Naïve Bayes, SVM, Decision Trees, Random Forest, Gradient Boosting Trees, Bernoully Mixtures, Neural Networks, etc...	- Classify - Recognize - Rank
- Behaviour prediction - Sentiment analysis - Market prediction - Client scoring		**Regression** Linear, GLM, SVM, Logistic, Neural Network, Ensemble of Trees, etc...	- Find casual relationships - Identify trends - Analyse trends - Predict - Describe

Unsupervised Learning

EXAMPLE OF PRACTICAL USE	INPUT/DATA	ALGORITHMS	USE
- Customer segmentation - Stock/bond picking - Product segmentation	Client positions, financial trades, MiFid questionnaire, basic personal data, legal documents, e-mails,instant messages, transactions/payments, social media, smart engagement,financial market, products, features, crowdsourced & institutional data	**Clustering** K-means, k-mades, Hierarchical clustering, fuzzy c- means, etc..	- Group/ungroup - Find similarities/ dissimilarities
- Identify behavioural drivers - Identify market drivers - Sentiment analysis		**Dimension reduction** Principal component analysis, factor analysis, singular value decomposition,etc..	- Simplify - Find latent variables - Analyze the hidden structure of a system
- Risk analysis & management - Budget simulation		**Density estimation** Parametric densities, kernel densities, copulas, etc..	- Estimate probabilities of events - Analyze risks
- Financial consultants network analysis - Client funnel analysis		**Complex network analysis** Graph algorithms	- Analyze complex systems

Reinforcement Learning

EXAMPLE OF PRACTICAL USE	INPUT/DATA	ALGORITHMS	USE
- Asset allocation - Algorithmic trading	Client positions, financial trades, MiFid questionnaire, basic personal data, legal documents, e-mails,instant messages, transactions/payments, social media, smart engagement,financial market, products, features, crowdsourced & institutional data	Dynamic programming genetic algorithms, grid search, etc..	- Analyze & manage complex systems

In addition, ML can roughly be divided into two categories: Classical Machine Learning - a natural extension of Statistics - and **Deep Learning**, that is, cutting-edge Neural Network applications.

Deep Learning is key in image and speech recognition, language translation, and processing other unstructured data. For example, Deep Learning is used by Google in its voice and image recognition algorithms. Note that not all Neural Network techniques are Deep Learning: "normal" Neural Networks usually have one to two hidden layers and are mainly used for supervised prediction or classification, while Deep Learning Neural Networks differ from "normal" ones because they have more hidden layers. Deep Learning could be useful in discovering and modelling complex relationships and patterns. For example, Deep Learning could be used to identify financial crisis triggers, starting from rates and prices, macroeconomic data, news, social media data, etc.

Anyway, in the near future most of the Financial Data Science applications in wealth management fall (or will fall) into the category of Classical Machine Learning (even if investment firms could buy signals obtained with methods of Deep Learning), and in the rest of the paper we will focus on this.

A Holistic Approach To Financial Data Science

According to Gartner Data & Analytics Summit 2018 "Machine Learning is giving organizations the power to classify and predict in ways workers cannot achieve alone.” However, ML is not an "out of the box solution" that just magically appears in an organization.

Financial Data Science can be very powerful only if properly used. Unfortunately, in Virtual B we have the feeling that too many clients are still approaching Financial Data Science in the wrong way: instead of taking a holistic approach, they aim to solve limited, individual operational problems.

What does “taking a holistic approach to Financial Data Science” mean? To us, it means creating a data driven organisation where the whole wealth management process is empowered by the data life-cycle. Hence, it is a firm-wide process. The main benefit of the holistic approach is that business decisions driven by financial data analysis are easily spread around the organisation in a transparent and multidirectional way. In addition, the data driven process is not static, but dynamic. It means that the organisation can learn from feedback, adapting faster to the changing business environment.

The “learning as you go” approach provides a blueprint for responding to the major business challenges of the following years, from regulation to lead generation.

Financial Data Science May Improve Business Productivity

Now let us briefly examine in a very practical way how Financial Data Science can impact the different parts of a typical wealth management process.

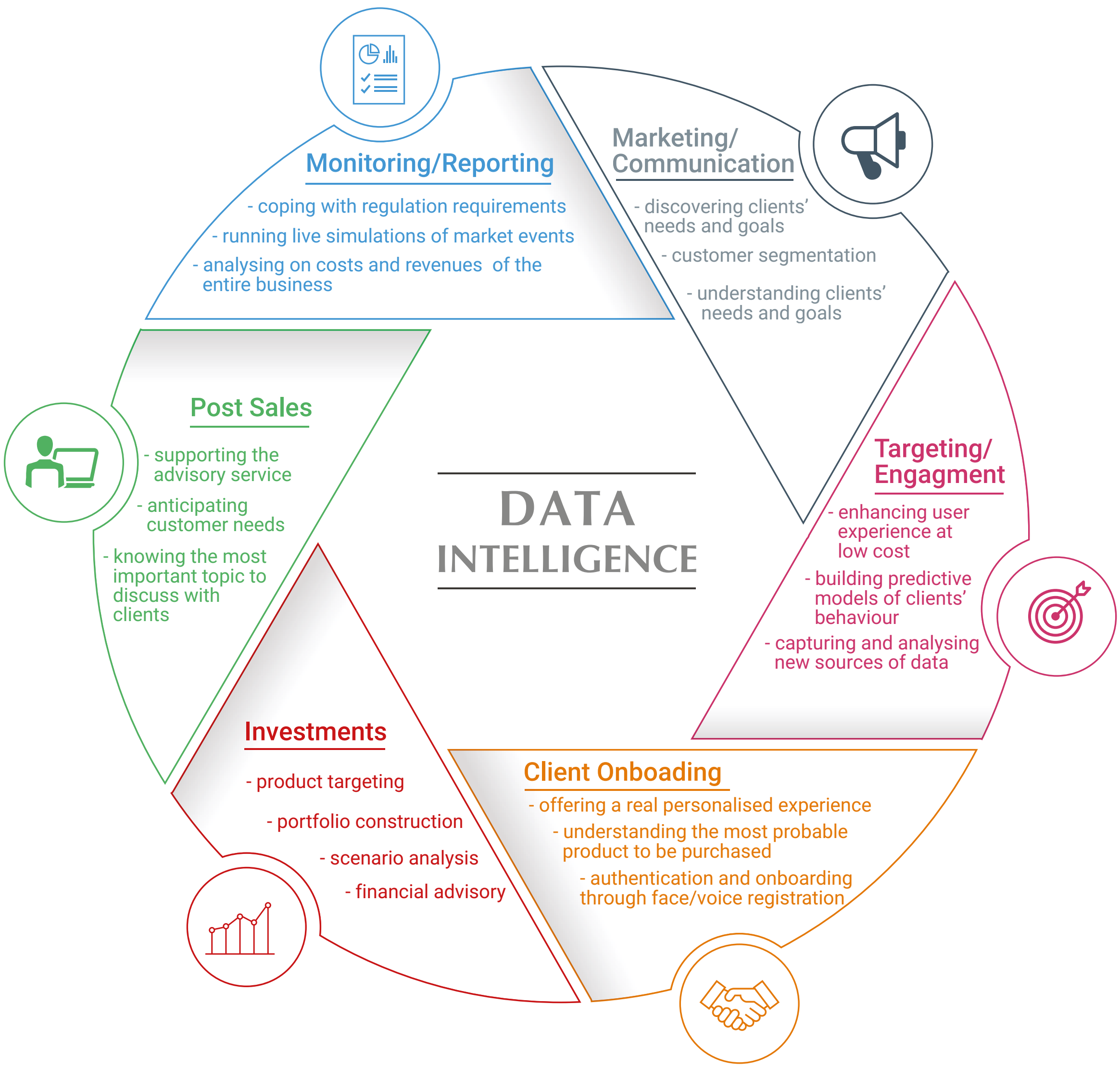


FIGURE 9: Data should be at the heart of the wealth management process, supporting and driving all the main actions.

Forget investing and let's focus on ROI

In principle, in the investment area of wealth management, Financial Data Science has several applications, including:

- Capturing and analysing relatively new sources of data (e.g. social media);
- Data quality assessment and control (e.g. in back-office operations, mitigating operational risks);
- Finding signals for higher (and possibly uncorrelated) returns, i.e. forecasting asset returns and validating models;
- Analysing market sentiment and detecting early warning signs of crisis/ systemic risk;
- ML-based credit scoring models for bond picking;
- Optimising trading execution, e.g. analysing the market impact of unwinding large positions;
- Looking for securities that have characteristics similar to an illiquid security which is difficult to price.

However, according to our opinion, this is the area where the ROI of Financial Data Science is at its lowest level. There are a number of reasons why we think that Data Science is of relatively limited help with investments, at least in terms of signal search and portfolio construction.

1) The intrinsic nature of the financial market:

- It is a complex adaptive system;
- The underlying Data Generation Process (DGP) is widely unknown;
- The DGP changes continuously;
- The financial market is largely unpredictable.

That is (Figure 10), the signal-to-noise ratio is unfavourable, at least for short-to-medium investment horizons: making this one of the worst milieu for the application of ML and AI techniques

2) As a corollary, due to the very low signal-to-noise ratio, by identifying and relying on asset patterns that were predictive of outcomes in the past, Data Science techniques (especially "black boxes") are highly susceptible to overfitting and data snooping. That is, false information.

3) The financial community is always trying to find new signals and to increase

portfolio efficiency, so the market is particularly quick to adapt. After all, the quant community has used Financial Data Science techniques since the dawn of time, when the term "Data Science" was still unknown. Thus, in the investment area of the wealth management process any relevant new edge -i.e. any new source of alpha - is likely to have a fast decay factor, as data is more widely used and hence becomes less valuable for gaining an advantage over other investors.

In addition, large extreme events happen, and can abruptly destroy the gains of investment strategies based on Data Science. The situation is different for large trading firms such as investment banks, that can, for example, profitably use trading book data and Big Data analysis to manage risks, estimate signals and optimise their use of capital.

In the wealth management field, in the investment area, the application with the highest probability of success (and the best ROI) is likely to be where there are simple, repetitive tasks, for example, in back-office operations and credit scoring. Situations where ML techniques can streamline (relatively) low-noise processes.

Financial market is very hard to predict: signal-to-noise ratio (US equities, 1899-2017)

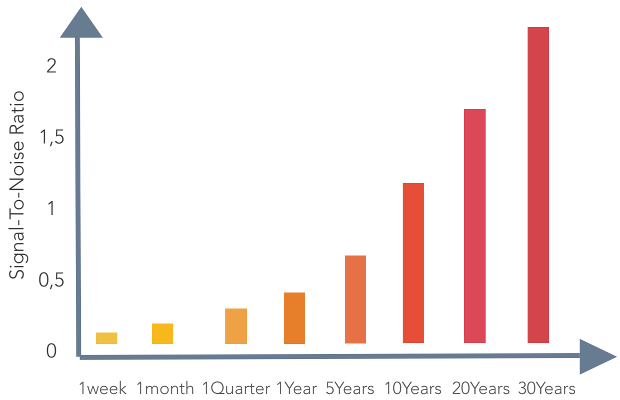


FIGURE 10: The financial market is a complex system with an unfavourable signal-to-noise ratio (having measured the reciprocal of the coefficient of variation, i.e. the ratio of mean to standard deviation). Thus, market predictions are very hard (that is the reason why average active managers across market cap segments and styles have underperformed their respective benchmarks – see https://us.spindices.com/spiva).

How Financial Data Science Helps Compliance & Product Governance

As a consequence of the Eurozone crisis, the regulatory environment has changed significantly. And even if lawmakers and regulators have definitely made certain adjustments, financial companies need to cope constantly with the risk of meeting applicable laws, regulations, and supervisory expectations.

From the financial product perspective, Mifid II and PSD2 directives are re-designing the landscape of relationships between customers and financial companies from scratch.. Basically, product design and distribution must be managed with a specific focus on client needs. According to ESMA, financial manufacturer distributors should use the following five main “categories” (i.e. variables) as a basis for defining the target market for their products:

1. The type of clients to whom the product is targeted;
2. Knowledge and experience, based on product type, product features and/or knowledge in thematically related areas that help to understand the financial product;
3. General risk tolerance and compatibility of the risk/reward profile of the product with the target market;
4. In particular, the clients' ability to bear losses;
5. The clients’ objectives and needs, including, for example, a specific age demographic, tax efficiency, or specific investment objectives such as capital protection, long term care risk hedging, green or ethical investment.

Under this structured regulatory-environment, Financial Data Science is extremely helpful in dynamically monitoring the full compliance metrics using a single dashboard, as well as in encouraging a transparent and profitable dialogue with regulators.

For example, managers and executives may want to check and monitor to what extent the organisation, as a whole, is compliant with the given rules (Figure 11).

So, using specified compliance metrics for risk, liquidity, complexity, investment duration, client financial knowledge and experience, they may want to control if a client and actual portfolios are aligned. And, whenever it is appropriate, they may want to drill down to any single clients, or group of clients, and a single financial product, and quickly step in to correct any kind of conceivable anomaly.

Risk Compliance Analysis

FROM OVERALL COMPANY

TO SINGLE CLIENT

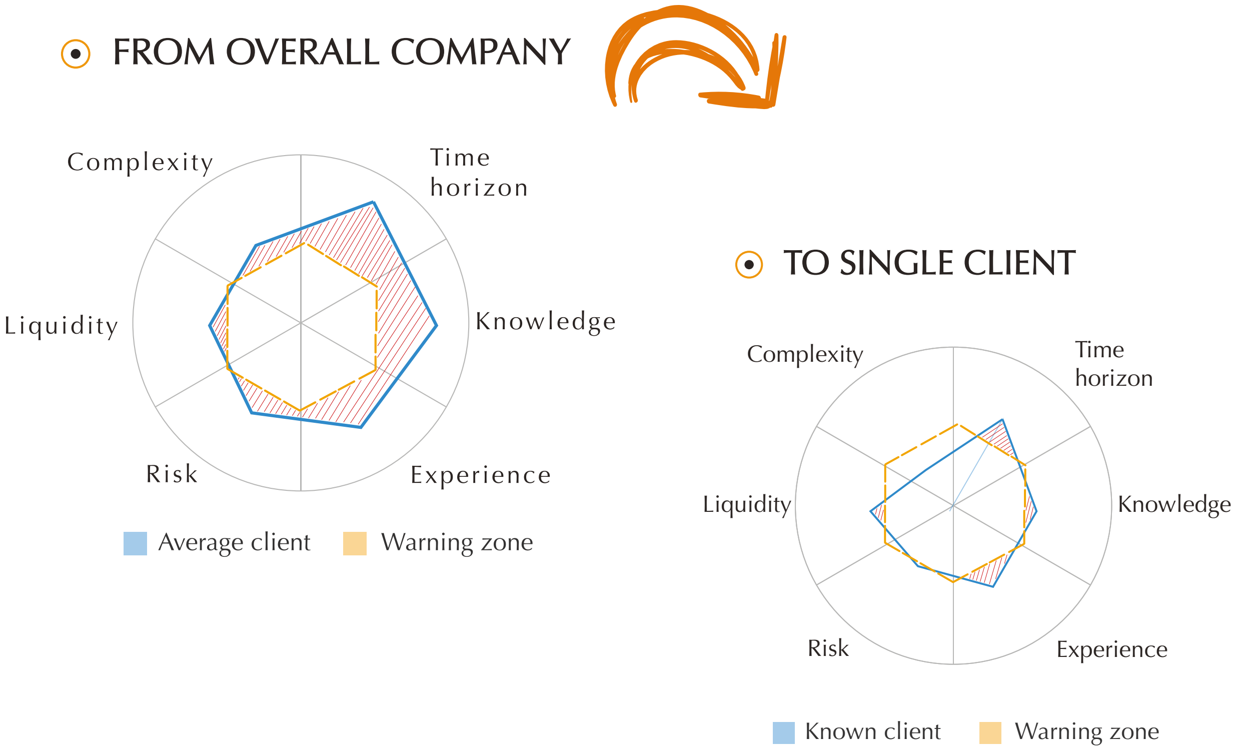


FIGURE 11: Financial Data Science allows real time control of key compliance metrics (i.e. related to knowledge-experience, risk propensity, investment horizon, etc), starting at firm-level and drilling-down to any single client, identifying warning signs and anomalies, and quickly addressing problems

Product governance means that financial institutions must put sound processes in place for the design and the distribution of financial products and services, properly identifying target investors ("target market").

Hence they need detailed profiles of each client, and a financial product, pursuing precision-target offers and perfects matching over time, reviewing their product governance process regularly: it is a typical application of Financial Data Science. The algorithms allow for coping with regulation requirements in real time.

Organisations must keep records of all activities undertaken for a minimum of five years, including all transactions, client phone calls, text messages, instant messages and social media interactions. This is not a challenge for Compliance: it is a big opportunity, and it is a gold mine from a Financial Data Science perspective.

Due to the customer-centric evolution of the market and of the rules, Compliance can no longer be seen as a bureaucratic unit: it plays an active role. The holistic view facilitates this kind of approach, sharing information extracted from data on several levels of the organisation, for different purposes.

How Financial Data Science Helps Marketing

One of the major impacts of the digitalisation process is that financial products have been turned into "commodities", with massive consequences on customer loyalty. For this reason, financial companies are shifting away from a product-centric model, moving towards a customer-centric model, where customer satisfaction is the key metric in creating value.

In order to generate value, financial companies need to start from client investment needs and life-goals. We are in the era of "people-based marketing". Financial Data Science goes in this direction because it allows for several key actions, like the following ones.

- Building predictive models of client behaviour, and actions. ML could use the gigantic mass of data on each person to profile customers and to analyse behaviour patterns, dynamically.
- Understanding products and AUM drivers, given explanatory variables such as product features, marketing campaigns, financial prices, market events, macro-economic and political environment, and so on.
- Analytically mapping the client funnel, simulating and optimising it.
- Identifying specific customer needs through cluster analysis. Clusters mean user segmentation. And user segmentation means offering real personalised products/services, as algorithms allow matching data to clients, and clients to products, to target consumers and measure results at the level of a single client, across all channels and devices. Financial institutions can thus understand if needs are either explicit or, "hidden", already satisfied, or not, so that they are able to plan proper strategies in order to address clients' needs.

Exploring client needs

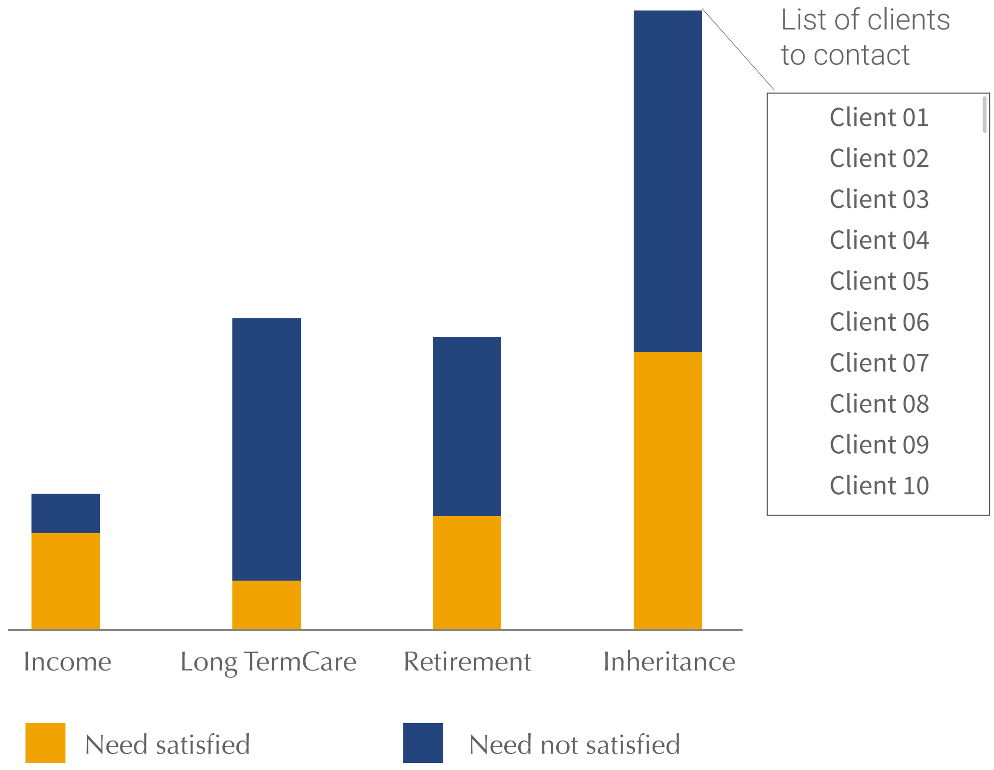


FIGURE 12: Starting from raw data, customer needs can be clearly identified using Business Intelligence and Machine Learning tools, checking if they are already satisfied by the current investment solutions. If not, needs are quickly addressed with proper marketing and/or sales actions.

Business Case: Exploiting "The Long Tail"

Exploiting the large portion of retail investors with small AUMs is a common problem among banks and wealth management firms.

So far, large, incumbent financial firms have mostly paid little attention to small clients. However, some of them have a great future potential: some might have money in other banks, others are young with huge human capital and good probability to become High Net Worth Individuals (HNWI) in the future. Thus, the surface of "The Long Tail" hides some very promising clients.

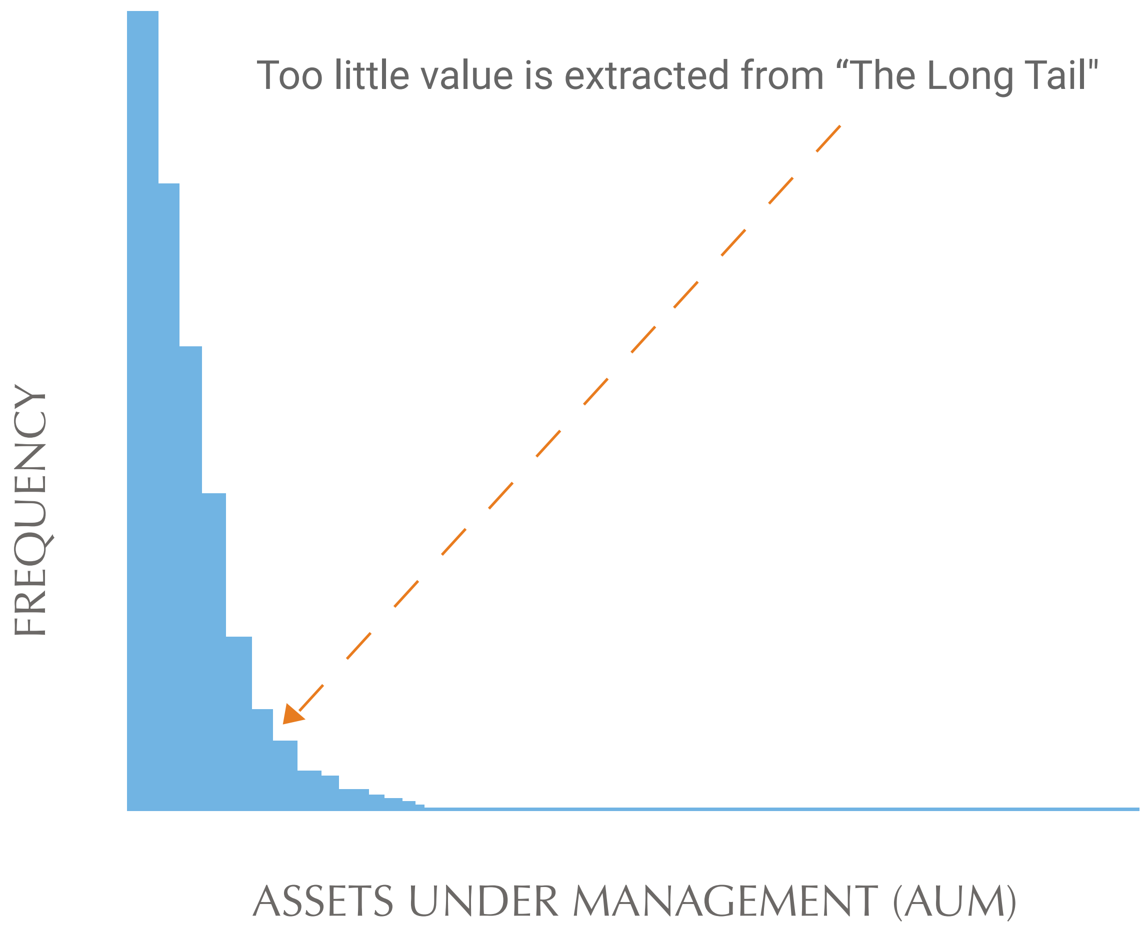


FIGURE 13: Wealth follows a so called "power law": there is a huge number of clients with small AUM. The "Long Tail" has been snubbed so far because small clients are not considered profitable: Financial Data Science can make them profitable.

Financial Data Science techniques can find these clients: following this idea, we developed a proprietary algorithm, the Future Potential Index. It is based on a combination of Supervised Machine Learning and Business Intelligence techniques and aimed to search future HNWI among small retail clients. Trained on past, existing data, the algorithm learns to identify the most promising clients or prospects, based on a number of features. Depending on the dataset, the typical probability of success ranges from 80% to 90%.

Once the high potential clients are recognised, the next step is engaging them in the proper way with a well targeted strategy. As the algorithm estimates the risk preference and investor profile of clients, each individual can be targeted according to her specific goals, e.g. paying for children's education, retiring with ease, sustaining a certain lifestyle, buying a house, coping with rising healthcare costs of aging parents, transfer of wealth from one generation to the next, and so on.

Communicate	Engage	Close the deal
Different channels/styles/time	Different tools (eg behavioural tests, investment simulations, chats, events)	The right product for the right client
Measure (eg A/B test), collect and learn	Store every feedback, any data point	Smooth onboarding

Results and feedbacks from these communication and marketing actions are new data for the Machine Learning engine, so the procedure gets more and more precise over time, understanding client profiles and measuring client's propensity to purchase various investment solutions.

The upside potential is high, as the probability of finding profitable clients is high. On the contrary, when the model is wrong, there are no major drawbacks.

How Financial Data Science Helps

Communication

In the digitalised world, the one size-fits-all model does not work anymore. Today, customers have direct access to any sources of information they wish. As content has grown progressively ample and instantly available, attention becomes the limiting factor in the consumption of information. Therefore, communication strategies need to convey the right message, at the right time, to the right audience, according to the principles of Attention Economics.
Through powerful predictive analytics and measurement services, Financial Data Science helps to better understand a customer's journey and to meet his needs anywhere, anytime, delivering true omnichannel seamless experiences. Communication is a key step of the process.

Examples of Financial Data Science to communication range from optimal design of communication campaigns, evaluating their probability of success based on knowledge of factors that might affect it, to clustering base communications (Figure 14). This means that, according to the cluster of membership (for example based on client financial DNA), each client receives a specific communication, or report, through the channel she favours (e.g. whatsapp), at the time she is most likely to pay attention to it (e.g. Saturday morning).

Essentially, a full data driven communication strategy allows financial companies to customise any campaign, improving the lead generation rate thanks to a deeper understanding of the customer profile.

Customer segmentation makes a marketing campaign more effective

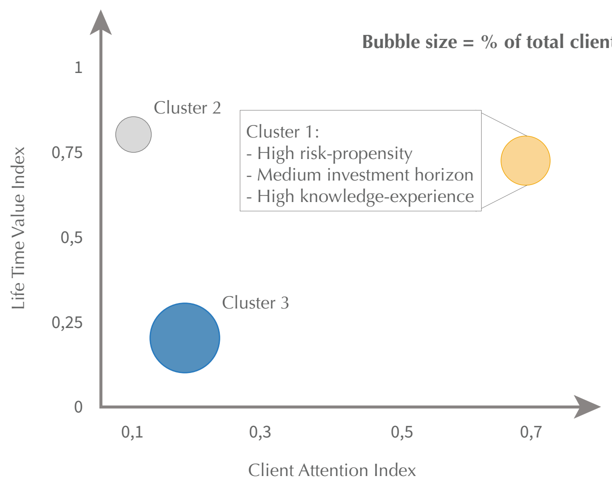


FIGURE 14: Using different Financial Data Science tools (classification, pattern recognition, clustering, etc) it is possible to segment the customer base, getting a clear picture, based on a small number of key variables extracted from raw data. For example, a customer can be a segment considering the long-term value for the company (life time value) and how much they pay attention to promotional messages. This allows for optimal marketing strategies and product governance.

How Financial Data Science Helps Financial Advisor

Since financial products are becoming less important and cost transparency is a key factor, financial advisors will play a central role in terms of customer satisfaction. The question is: how might financial institutions enhance the role of the financial advisor in order to improve customer satisfaction? How can they help Financial advisors?

Financial companies should leverage on their dataset and data analytics to help financial advisors to be as close as possible to their customers. Financial advisors have to work on the long-term relationship with their clients, built on a constant dialogue. Financial Data Science is a powerful instrument that increases the role of financial advisors, giving them all the relevant insights to improve customer satisfaction, personalise the relationship and increase productivity. For example:

- Helping to detect customer financial DNA, discovering the latent investment needs, having all the relevant client data and facts at one's fingertips;
- Identifying which investment solutions are most likely to fit clients' needs, based on client profiling and an estimate of the probability of investing;
- Recommending what documents (e.g. "conversation starters") a financial consultant should read based on attributes of clients they are going to meet with, or recommending articles/news a client might be interested in, based on his actual investments and his financial know-how;
- Using Chatbots and virtual personal assistants.This kind of supporting advisory services enhances user experience at a lower cost, and also allows for exploiting lower net worth market segments.

Client Dashboard for Financial Advisors

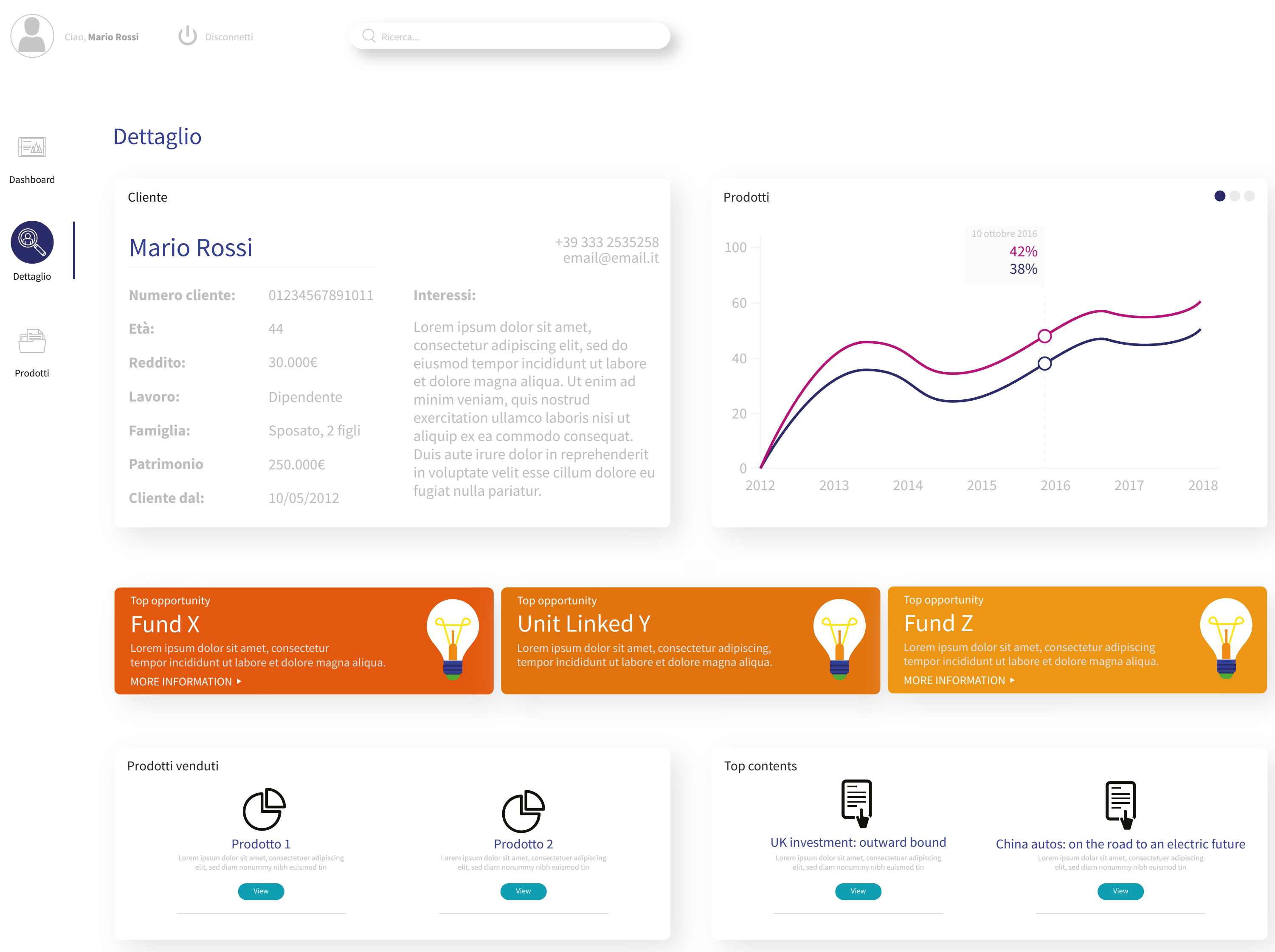


FIGURE 15: Financial Data Science empowers financial advisors with a suite of tools that enables them to serve their customer in the best way, from portfolio management and risk analysis, to product recommendations and conversation starters.

How Financial Data Science Helps Executives

In a global and circular economy, executives have to deal with complexity. The speed of technology is changing and the increasing pressure of competition makes the job even harder. Under this market condition executives need to better identify, measure, and mitigate risk- the concept of risk management has to move beyond financial and compliance definition -, and consider the importance of customer satisfaction as well.

Financial Data Science radically improves executives' decision models and helps to commit leaders to shaping the strategic agenda. Executives can now conveniently access all the relevant data (KPI), while proper Data Visualization techniques permit an easy interpretation.
Thus, employing Financial Data Science is a strategic hedge, because it encourages executives to keep full control of the entire organisation and process, measuring efficiency.

Putting Financial Data Science to Work

Nowadays, Financial Data Science should be an essential toolkit for improving the decision process and leading the whole organisation.

The Fintech revolution is putting traditional players under pressure in many aspects, but incumbents should remember that they have an incredible hedge: the client financial history, and a lot of data.

However, this huge advantage usually has to cope with old and binding IT legacy systems. Financial institutions have a long history of data trapped in old and inaccessible databases, with stiff IT departments and legacy systems, normally with the inability to handle dynamically big volumes of information. Thus, looking beyond the hype of Data Science, the gap between desire and execution is wide at most wealth management companies.

The new contest should make executives rethink the way information is processed between core and non-core activity, coming up with an appropriate IT framework. Every business of the future would rely on rapid transmission and assimilation of data, but the ROI of Financial Data Science applications is different along the value chain: executives must choose wisely how and where to invest.

In the medium-to-long term, finding the most strategic application of Financial Data Science will determine the split between winners and losers among wealth managers. Executives should avoid using ML and AI just for the sake of its use, they should instead set up a data driven organisation.

The Financial Data Science Process

The quote "power is nothing without control" is extremely relevant with Financial Data Science projects and applications.

Even if computing power is cheap and open source libraries allow everyone to easily access the best ML techniques, it is simply not enough. You need a sound process. The process has to be put in context within the organisation; each company has different sources of data and business objectives.

Anyway, from a methodological standpoint, a sound Financial Data Science application generally involves these phases:

1. Main problem definition;
2. Data acquisition, collection, and storage;

3. Discovery and specific goal identification (i.e. ask the right questions);
4. Accessing, pre-processing and cleaning data;
5. Exploratory data analysis;
6. Choosing potential models and algorithms;
7. Applying Data Science methods and techniques (e.g. ML, statistical modelling, Business Intelligence, AI, …);
8. Validation (i.e. measuring results) and fine tuning of the process;
9. Data visualisation, i.e. delivering, communicating, and/or presenting results with proper SW tools;
10. Making business decisions;
11. Loop.

On practical ground, in a realistic situation this process must be industrialised. That is, implemented in a well-engineered, easy-to-use, software solution, which can be customised, web-based, or on premise, or using off-the-shelf data science platforms. In order to do that, you need people (the team) and technology and unless you are Google or Amazon, you should consider the outsourcing option.

The Team: not only pure Data Scientist

Within a complex organisation, a decision-making process requires a cross understanding of the business. Since data should be at the core of business activity, the Financial Data Science team should be ideally composed by people with different competencies and backgrounds. After all, Data Science is multi-disciplinary, and in the real world nobody possesses all the hypothetical Financial Data Science core strengths (otherwise she would be a unicorn).

Therefore a good team is key. When building a data science team, the set-up is a function of the environment in which the team operates. But, generally speaking, a lean, minimal team should include the following roles/professional skills (please note that many skills may intersect between roles):

- Team leader, with strong domain expertise, mathematical modelling skills, good technical/coding grasp, leadership and vision;
- Technical Data Scientist, with robust coding/computer science skills (e.g.: Python, R, Matlab, SAS, SQL, noSQL, Hive, Pig, Hadoop, Spark, etc), super-strong knowledge of ML techniques and some exposure to the business sphere;
- Business Data Scientist, with a solid business domain, excellent communication skills, some quantitative background, DataViz and data interpretation expertise, and at least some basic tech and coding skills.
- Data Engineer, a concrete software developer (that means some programming languages expertise, databases, cloud computing, distributed frameworks like Hadoop) with enough exposure to algorithms, quantitative topics, and business domain.

A mix of backgrounds is good: people with a variety of backgrounds and experience tend to bring greater insight.

In-house or outsourcing?
Make or buy?

These are old questions: hiring in-house or outsourcing a Financial Data Science team, and making or buying a solution? The perfect solution, of course, does not exist, so let us share some simple considerations.

- Custom development is aimed at building a solution that exactly fits firm-specific-requirements and gives flexibility for future evolution. But it is usually more expensive, more time-consuming and riskier than off-the-shelf solutions: the risk of reinventing the wheel, deploying standard models is relatively high.
- Purchasing a solution is typically much faster and less expensive. But also potentially less flexible, even if good existing solutions allow for flexibility in terms of algorithms and data visualisation.
- Large scale financial firms sitting on massive data sets may prefer to have a strong data science presence in-house, with an in-house team that constantly tries to create value within the organisation. Hiring a team is obviously the more expensive option, at least in the short term. But an internal team can build a knowledge base, consistently pushing the firm to became data driven. However, the risk is to underuse the team.
- Outsourcing, at least in part, to an external data team is cheaper, but a lot of knowledge might not be transferred. Thus, the impact of Data Science on the organisation could be negligible.

In reality, using a mixed strategy is a good option: a firm starts with some limited internal resources and outsourced partners/solutions. Then, as the projects grow, the firm progressively increases their internal capabilities. This should ensure fast deployment, knowledge transfer, and creating value from data in the long term.

Our Solution:

SIDEKYC®

a software developed by Virtual B and GFT

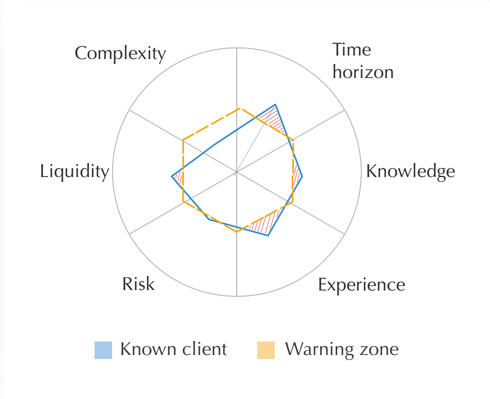
SideKYC® is an advanced **data analytics** dashboard that empowers **different business functions** with the use of Data Science and Business Intelligence.

A cutting edge **omni-profiling tool** based on Financial Data Science that can **close the existing gap** of client knowledge inside banks, asset managers and insurers.

SideKYC® Features

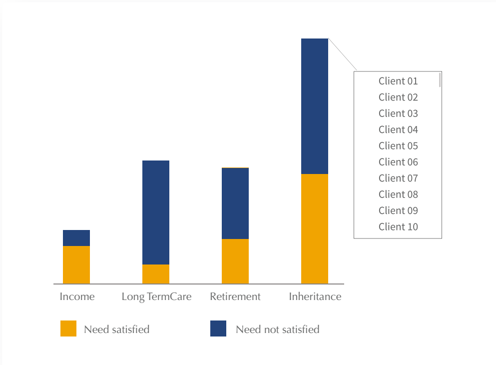
1. Financial Fitness Tracker®

The FFT is a tool for evaluating the adherence of an investment, or an investment portfolio, to the financial profile of a current or prospective client.
The FFT is therefore particularly useful in responding to business needs and customers.



2. Future Potential Index®

The Future Potential Index is an algorithm that uses Supervised Machine Learning and Business Intelligence techniques to improve customer acquisition. The algorithm learns to identify the most promising clients, based on a number of features.
The typical probability of success ranges from 80% to 90%.

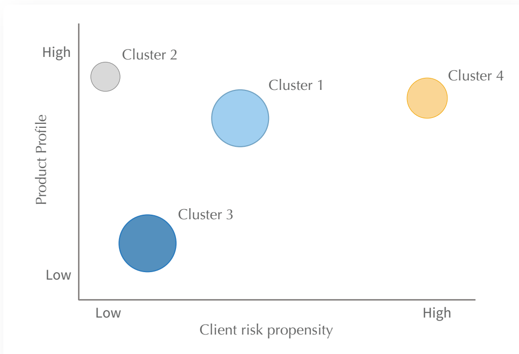


3. Profiling

Clients' needs, objectives, constraints, psychological characteristics, time horizon, and any other factors that may be relevant in the investment choices.
All client information in a single comprehensive dashboard.

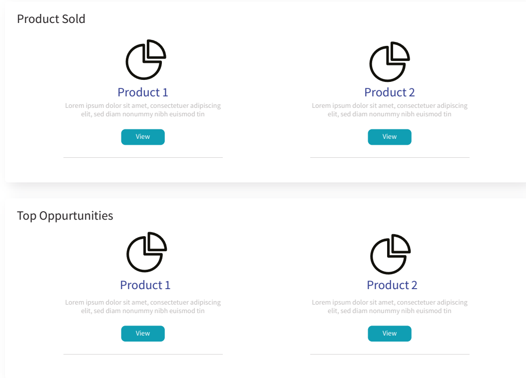
4. Client Clusters

Clustering enables proper client segmentation and a precise analysis of the relationship between clients and products. The Sales and Compliance Departments can understand if customers are served with the right products in relation to the risk profile.



5. Product Recommendation

All the products sold to the client and all the products that he might need to buy in order to satisfy his needs.
Product recommendations can be based on SideKYC proprietary algorithms or tailor made according to the algorithms of the client.



SideKYC: solving real business issues

1. Who are my clients?
CLUSTERS (KYC)

Marketing
Sales
Compliance
Business Management

2. Are we compliant to MIFID?
MAPPING
CLIENTS/PRODUCTS
CLIENTS/GOALS

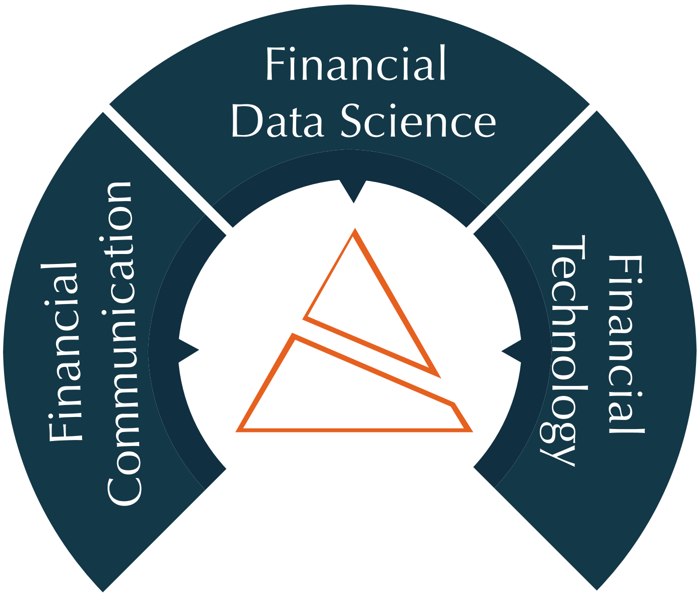
Compliance
Product factory
Sales
Business Management

3. Who are the best prospects?
PREDICTIONS

Marketing
Product factory
Sales
Business Management

About Virtual B

Virtual B is a tech firm offering digital solutions for the wealth management industry. In 2011 the company eveloped the first robo advisory platform in Europe.



If you're thinking about putting into practice a digital solution in wealth management, find out how VirtualB can help you.

Via San Gregorio 40
20124, Milano - Italy

info@virtualb.it

References

Barber, D., (2012), Bayesian Reasoning and Machine Learning, Cambridge University Press.

Bengio, Y., Courville, A., Goodfellow, I. (2017), Deep Learning, MIT Press.

Bishop, C.M., (2006), Pattern Recognition and Machine Learning, Springer.

Bughin, J., (2017), Ten big lessons learned from Big Data analytics, Applied Marketing Analytics, Volume 2, Number 4.

Fawcett, T., Provost, F., (2013), Data Science for Business: What you need to know about Data Mining and Data-Analytic thinking, O'Reilly Media.

Financial Stability Board, (2017), Artificial Intelligence and Machine Learning in financial services.

Friedman, J., Hastie, T., Tibshirani, R., (2017), The elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, Springer.

McKinsey Global Institute, (2016), The age of analytics: competing in a data-driven world.

Norvig, P. (2015) Artificial Intelligence: a modern approach, Pearson India.

World Economic Forum, (2015), The Future of Financial Services - How disruptive innovations are reshaping the way financial services are structured, provisioned and consumed.