Data Analytics in Investment Banks

Basma Iraqi[1]*, Lamia Benhiba[2], Mohammed Abdou Janati Idrissi[3]

Mohammed V University in Rabat
ENSIAS, Rabat
Morocco

Abstract—**Capital Markets are one of the most important pillars of worldwide economy. They gather skilled finance and IT professionals as well as economists in order to take the best investment decisions and choose the most suitable funding solutions every time. Data analytics projects in Capital Markets can definitely be very beneficial as all optimizations and innovations would have a financial impact, but can also be very challenging as the field itself has always incorporated a research component, thus finding out what could really be of an added value might be a tricky task. Based on a comprehensive literature review, this paper aims to structure the thoughts around data analytics in investment banks, and puts forward a classification of relevant data analytics use cases. Lastly, it also discusses how transforming to a data-driven enterprise is the real change investment banks should aim to achieve, and discusses some of the challenges that they might encounter when engaging in this transformation process.**

Keywords—*Capital markets; data analytics; data analytics use cases; data-driven transformation; investment banks*

I. INTRODUCTION

Capital markets are one of the most important and complex fields in worldwide economy that involves multiple stakeholders, among which investment banks are the main actor. These later play the role of the mediator between enterprises looking for funding and others willing to invest. They advise their clients on the best approaches to collect funds and manage their investments, thanks to a very thorough understanding of the client profile, but also the ups and downs of the markets and the economy.

Faced with troubled economic times, investment banks, like all other financial actors, have been pushed, more than ever, to rethink their products and services and improve their quality by analyzing further all the available data on clients and markets [1]. The interest towards data analytics comes therefore naturally as this is where magic happens with the data. Through the different types of analytics possible, the algorithms and techniques developed, there is opportunity for investment banks to improve their relationship with their clients and enhance their overall revenues.

Nevertheless, starting data analytics projects can be a challenging step as these are usually initiatives where the output is not known from the beginning and the return on investment not determined. Thus, it is important to do thorough research beforehand to identify the types of analytics that could be of interest to the organization, as well as find out similar success stories in order to define the scope and get inspired before starting the experimentation.

Research on data analytics in investment banks have continuously focused on specific use cases' implementations, but the topic has never been addressed from a broader perspective. For the financial services sector, no comprehensive guides are available to help investment banks in approaching data transformation from the beginning, by identifying the problems they can resolve and the types of analytics they can implement in order to choose use cases from.

This paper aims therefore to provide a classification of relevant data analytics use cases for investment banks, and discusses the importance of engaging in a data-driven transformation and some of the challenges that could be faced in this process. The rest of this paper will be organized as follows: Section II presents capital markets and investment banks, Section III outlines the motivations behind adopting data analytics in investment banks and Section IV proposes a classification of the most relevant use cases that can be deployed in this sector based on a review of the literature. In the last section, a discussion is initiated about some of the challenges that investment banks may encounter when initiating the data-driven transformation, in order to provide better guidelines for investment banks in this process.

II. INTRODUCTION TO CAPITAL MARKETS AND INVESTMENT BANKS

Capital markets are markets where demand and supply for funding based on different instruments are organized. These instruments are grouped in different asset classes and quoted in different markets [2].

The main objectives of Capital Markets are to secure funding for national economy, structure liquidity and savings, enable firm restructuring, mergers and acquisitions, cover multiple kinds of risks and provide and represent the main references for all asset values [3].

Out of all the activities performed in capital markets, portfolio management is one of the most important ones. It focuses on decision making with regards to investments for individuals or institutions, taking into account the investment policy, objectives, asset allocation, risk and performance [4]. The science behind it enables measuring performance, and consequently making consistent improvements based on continuous learning, controlling risk and ensuring consistency and discipline throughout the process and the portfolio lifecycle.

---

*Corresponding Author

In order for the different stakeholders in the Capital Markets to decide on the strategies they will work with and manage their portfolios appropriately, they also rely greatly on different economic indicators which represent some of the most important signals and raw data influencing the capital markets' activities. That is because it is these same indicators that explain the motivation behind the need for funding, investing or restructuring among others. These indicators could be of the following three types: (1) Supply indicators, (2) Demand indicators and (3) Inflation, Money and financial indicators [5].

Modeling and predicting how changes in the values of economic indicators impact the investment strategies adopted and how the change in the prices of the different instruments due to economic variations affect the value of the investment portfolio and its risk vulnerability is how Capital Markets function on a daily basis.

In capital markets, investment banks play a very important role. In fact, they are the main intermediary between large corporations or governments that need funding, and the market investors. Based on the financial challenges of each, investment banks advise their clients on the best approaches to raise capital taking into account the current investment climate as well as the client's needs [6]. This also applies to mergers and acquisitions' scenarios where businesses are often looking to restructure by purchasing their competitors.

As such, the main departments of an investment bank are mainly [6]:

- Trading and Sales: the department in charge of executing deals on behalf of the clients, but also the one responsible of trading the bank's own money on the different financial instruments.

- Asset Management: referring to portfolio management for large institutions, insurance companies or pension funds.

- Wealth Management: the function of advising wealthy families and individuals in ways to fulfill their financial long-term objectives and needs by investing in capital markets.

- Securities Services: the department responsible for all services related to assets held or issued by the clients. These services include custody, clearing, settlement and fund administration among others.

- Research: the function of collecting all kinds of information about the market and the different corporations in order to generate recommendation reports on some stock buy or sell.

- In addition to the risk management, internal controls, accounting and IT functions among others.

Given their very diverse missions, investment banks' departments often have opposite views on market trends, look for different opportunities, offer various services to the same clients, and thus collect distinct data inputs as well as analyze and process them differently.

III. DATA ANALYTICS IN INVESTMENT BANKS

Analytics is a buzz word. It is used almost everywhere and in very different contexts. There are even many research papers which only aimed at finding the right definition to analytics. One of the best descriptions one could find in the literature regarding analytics is the one provided by CETIS (the Center for Educational Technology and Interoperability Standards) which defines "Analytics [as] the process of developing actionable insights through problem definition and the application of statistical models and analysis against existing and/or simulated future data" [7].

Actionable insights shed the light on the importance of action in analytics. Analytics' results should go beyond reporting or descriptions to opportunities and conclusions that prompts users for action. That's actually the main difference between Analytics and Analysis, which are the most confused concepts in this regard. Indeed, while "Analytics aim to project the future based on past performance, analysis rather presents a historical overview of past data" which is equivalent to what descriptive or diagnostic analytics could render [8]. An example of an actionable insight is the outcome that goes with A/B testing initiatives of E-commerce websites (such as Amazon or eBay) in which they test variations of their websites in samples of users in order to analyze their sentiments and reactions to the changes made.

The literature distinguishes four types of analytics [9]:

- Descriptive analytics: which answers questions such as: what is happening? What is it about? What does it mean? What are its attributes? About the data and content being analyzed.

- Diagnostic analytics: are answers to the following questions: why did it happen? What are the triggering events? What are the preliminary conditions for each context?

- Predictive analytics: are more focused about the future, thus answering questions like: what is likely to happen? What happened in similar situations throughout the past? What could be the results of such actions or decisions?

- Prescriptive analytics: come after predictive analytics to discuss and suggest potential alternatives fix problems or capitalize on opportunities. Questions answered in prescriptive analytics include: what should I do about it? What can I do prevent it? How can I best exploit it?

The capital markets, and investment banks more particularly, have been one of the first and most important users of data analytics. That is not a surprising finding as the field does not manufacture any physical product and data is really its main asset [10]. In fact, statistics in 2016 have even showed that the field is the number 1 consumer of data services with a share of 13% of the global data analytics use cases implemented [11], and that is because data is used and generated tremendously every second of the day in this field, and that analytics could definitely help making sense of data that could never be processed exhaustively.

In addition, the interest in Data Analytics within investment banks is also the result of a tight conjuncture which led to weak returns compared to earlier times. Indeed, according to a study by Accenture, the financial sector and capital markets more specifically, has witnessed since a few years a stagnation of revenues given the fall of margins and the rising complexity of regulations. Also, the Fixed Income, Currencies and Commodities business, which has historically filled the largest share of revenues, witnessed an important share shrinking for the same reasons. Last, new bank competitors have arisen, making the pressure on old banks even tighter [1].

Data has therefore made it at the center of the banks' analytics projects because it promises [12]:

- A better risk management: investment banking is the sector where resources are heavily invested in risk because the consequences of a bad risk assessment could be devastating. The 2008 financial crisis and its impact on the global economy is the best example to illustrate the high importance of this business line. This applies to all types of risks, whether it be related to [1]: (1) Fraud detection: where big data could be leveraged in order to identify patterns of fraudulent transactions or atypical operations and alert the appropriate personnel in order to investigate further. (2) Liquidity: where analytics could be used to keep track of the short- and long-term liquidity every time, assess the impact of transactions on liquidity in real time and run simulations and stress tests regularly in order to make sure that the necessary funds for banks to function correctly are always available. (3) Credit risk: where internal data about clients and counterparties is enriched with external data from the web, social media and the news in order to get an exhaustive feel of their financial situation and ensure that the hazards are well managed and known if any.

- A pleased and Loyal Customer: this is where sentiment analysis comes to play in order to better understand the needs of the customers and address them correctly. Thanks to all the data available in the web, including the news, social media, research reports and corporate websites, it is now possible to anticipate what the client might or might not appreciate, and direct them to the most suitable products (cross and up selling) at the right time. This not only enables an improved customer loyalty, but also makes attracting prospects a more successful process.

- A secure ecosystem: Analytics enables a very large and thorough monitoring where patterns of incidents and problems are identified through machine learning algorithms, thus making their handling and resolution a much straightforward process.

Since it is usually very expensive to start collecting all relevant data for the bank before making any use of it, all organizations, whether they are in financial services or in another industry, usually proceed in a different way. Companies start by defining a business objective linked to a strategic goal they want to reach thanks to analytics, and then reverse the process to identify the data to capture in order to achieve the goal. This specific application of data analytics is what a use case refers to, and it is the most common approach to address analytics as it offers a scope and a project perspective to the data application and enables the association of KPIs and governance that are very important to the monitoring of data analytics initiatives.

Despite these promising benefits, there are unfortunately fewer Analytics initiatives deployed in the capital markets sector, and especially in investment banks. That is because the challenge of finding interesting use cases is real, and the technical complexity that comes along the execution of these use cases is also more pronounced and different from the challenges that could be faced in other environments. So, what business problems could Data Analytics fix or optimize in investment banks? And what problems could be faced when implementing them?

In fact, based on the literature review of the use of analytics in investment banking, it came out that a brainstorming around business use cases of data analytics in investment banking is actually missing. As much as there is a lot of material when it comes to fields like retail banking, insurance or even industry, there is significantly less work in the investment banking sector given its business complexity and specific characteristics.

The focus of the remainder of this section is to describe data analytics use cases that can be of interest to investment banks, based on a review of research papers, of data analytics use cases worldwide as well as the authors' professional experience. These papers will be discussed following a proposed classification, which puts each use case in one of the 7 categories described below (from A to G) and indicates whether it is a deployed solution or is still the object of ongoing research. The research papers included in this study are the authors' selection of the most relevant papers in each category, based on the assessment of their added value according to the authors' financial expertise. Last, throughout this section, both internal bank data and external economic data will be proven to be equally important in implementing data analytics in investment banks, and the methods and data processing techniques that can be used to generate insight from this data will also be discussed.

A. Prediction of Financial Market Patterns

The investment banking sector is very tied to market data and economic indicators. It is based on this data that traders and portfolio managers decide each day whether to buy or sell. Therefore, predicting the market shifts and movements is a very valuable information for the actors to anticipate their investment strategies and act accordingly.

Predicting market patterns can be done based on the history of transactions, current and historical market data as well as indicators of the economic sector. The idea behind this use case is to identify if the current scheme is somehow similar to an old one in order to learn from the mistakes made by then and make the most out of it. This use case can be implemented by labeling the dataset based on sets of conditions and feature engineering in order to output the

knowledge needed. Then techniques like clustering to group alike data together or game theory to evaluate how each variable in the dataset influences the model positively or negatively can be used [13].

Another method to predict the patterns of financial markets is to build models which, based on input financial indicators, predict values of the output ones. An example of this approach is the article "Big Data Analytics for Financial Market Volatility Forecast based on Support Vector Machine" [14] which zooms on high frequency data or real time data, which is data that is generated on a very high velocity. This type of data has been proved to be very rich in terms of information, but its treatment has always been difficult given the high number of problems that must be fixed beforehand. In this article, the focus was put on volatility as an important indicator of high frequency data on financial markets, and it suggests predicting the short-term volatility via a model of the jump volatility. The technique used for this purpose is Support Vector Machine. It takes eight input indicators in order to model the price and predict the next day's price, and thus its short-term volatility. It also applies several kernel functions, which happen to have different prediction results on different markets.

On this category of use cases, JPMorgan Chase has been able to deploy a NLP project for equity investing. Indeed, in order to improve the research capabilities of their portfolio managers on equity, JPMorgan launched a Natural Language Processing project specialized on equity investing. The tool is fed with news articles on equity markets and generates insights on the data such as identifying enterprises which have new products to launch or the ones currently in difficulty. According to JPMorgan, the investments based on this tool have outperformed those based on the NASDAQ50 index [15].

B. Detection of Crisis Situations

Detecting crisis situations is technically a very similar use case to the prediction of market patterns. Indeed, it just focuses on a specific label of the data which is the crisis or the most alarming ones. The how-to is therefore the same, and the only difference consists in shedding light on the weak signals only in order to anticipate any future challenges or problems that may occur and set a backup plan in advance. This use case can be particularly of interest to the risk teams who need to anticipate and adjust their measures in case of problems or disasters.

An example of use cases in crisis detection is the paper "Monitoring Banking System Connectedness with Big Data" [16], where the authors came up with a solution to model connectedness between individual firms based on firm-level information in order to ensure better supervision and monitoring of financial stability since connectedness between economic agents has become one of the most addressed topics in the sector. The methodology applied for this purpose is a decomposition of the outcome measure into multiple components, and then a network construction based on indicators of connectedness in the US banking system. The first step is simply based on linear regression of CAPM framework using the French-Fama model, and the second step is about network construction and analysis. This latter is done through many measures in order to provide the best insight possible to the banking system stability state.

C. Client Risk Profiling

In portfolio management, there are two main inputs which define how the investment strategy should be: The first is the performance level wanted and the second is the risk appetite. Determining how much risk the client is willing to take is yet another challenging task that portfolio managers have to realize. Indeed, a portfolio manager could find cases where a client is unable to evaluate their risk appetite or one who underestimates his willingness to take risks or vice versa.

For such problems, there are multiple solutions. Scoring algorithms can be used which, based on questionnaires, determine the weights and risk scores to assign to every question and answer. An example of algorithms that does so is the self-organizing map, which belongs to the artificial neural network family and is based on unsupervised learning for model training. Another solution to the risk profiling use case is to apply classification models to the client data in order to put them on a risk group based on a set of attributes. Here, algorithms like KNN, Decision Trees or SVM can be used. And last, clustering can also be used on the historical transactions authorized by the client in order to actually find out what other clients they are similar to and adjust their risk profile accordingly [17].

One paper that addressed this topic in a different way is "The value of big data for credit scoring: Enhancing financial inclusion using mobile phone data and social network analytics" [18]. This paper comes up with a new approach to estimate credit scores, which is based on mobile call records and social data. In fact, the paper suggests that the use of mobile data just like social media data can have an interesting impact on the clients' credit scores. The data is modeled as a network where nodes represent people in the log. Labels as well as other information are extracted from by using link-based features as well as a personalized page rank. The methodology is then evaluated statistically and economically. In fact, the tests proved that the combination of call network data with conventional data has given better results in credit scoring. The data by itself (without conventional data) has even rendered satisfying outputs. The paper claims that this methodology is especially interesting in developing countries where historical financial data is rare or inexistent. The only limitation to this approach is the availability of data, especially given the tight regulations in place in many developed countries.

On this family of use cases, CitiBank has been able to deploy a Credit-Sentiment Monitor from news media. This tool, which name is CitiVelocity, is based on the news articles published in Thomson Reuters. The objective behind such a tool is to provide Citi staff with insights on the companies that are mentioned positively and negatively in the press. This information would then be used to evaluate indices of Credit Default Swaps in which these companies are included [19].

D. Product Recommendation for Clients and Prospects

Investments banks, at the opposite of retail banks, do not have huge numbers of clients as the business is mainly B2B (Business To Business), thus a thorough understanding of the client needs and profile is required to maintain a good and continuous relationship with the client, especially for the senior bankers and the client relationship management team, who represent the front interface of the bank with its clients. For that, a 360° vision of the client is necessary. This latter should be composed of the client history, which consists of all client transactions and exchange logs with the bank, its risk profile and its strategic orientations. These could be deduced from the sentiment analysis of the client activity on their website or social media accounts or its interaction with the bank's website.

Recommendation systems in such case are based on collaborative filtering algorithms, and the model used could be matrix factorization, neural matrix factorization or singular value decomposition among others. The rationale behind is to give an implicit score to all products for every client, such that the higher the score, the higher are the chances that the client will be satisfied from the recommendation [20].

The paper "an investment portfolio recommendation system for individual e-commerce users" has come with a new recommendation system implementation, where aside of the investor's risk preference, parameters such as gender, education or location are taken into consideration. The system is built based on the collaborative filtering algorithm and the VaR (Value at Risk) method for risk measurement. The particularity of the recommendation system used is based on the rapprochement with the behavior of other e-commerce investors with similar preferences [21].

Among investment banks, UBS has succeeded in providing a free investment consulting service based on recommendation engines. They implemented a tool called SmartWealth based on machine-learning algorithms which is meant to provide prospects with a free advisory service on online investment in order to increase their chances of becoming clients. In order to realize such a service, UBS asks the client to answer a set of questions, and then the algorithm assigns the client a risk segment and invests its money in an adequate fund [22].

Goldman Sachs had also transformed its whole approach of identifying investment opportunities for the clients based on three aspects: (1) Momentum, where techniques of industry sentiment are applied in order to identify correlations between businesses in the same sector; (2) value, which takes into consideration sector information in addition to financial data in order to define the intrinsic value of the firm; and last (3) profit, which concerns e-commerce companies and enables to evaluate them based on the patterns of web traffic [23].

E. Client Satisfaction follow up

Client satisfaction is also an important brick of the bank client relationship, but also a difficult one as it is not straightforward to evaluate the client satisfaction from the bank services. There are three main sources from which client satisfaction could be inferred: the bank portal, where client complaints are processed, emails and phone calls exchanged with the client and their online posts and articles or interaction with the bank's social media accounts.

Determining whether the client feedback is positive or negative is processed using machine learning and sentiment analysis, where after cleaning the data extracted from the identified sources, embedding is performed, then model training is realized through algorithms like decision trees, neural networks or naïve bayes among others. Such insights could be of interest to all the bank departments, but to the senior bankers more especially who can use the lessons learnt from the client satisfaction measures to better approach them and handle their inquiries.

The fanciest way to follow up on client satisfaction today is to have chatbots. The role of the chatbot is to answer client questions automatically or to log and route their complaints to the relevant staff. The paper "AI-Based chatbot service for financial industry" has suggested an approach based on machine learning in order to automate the communication between the bank and the client regarding product sales or customer support, such that the robot agent keeps accumulating knowledge throughout the process [23].

On the same application of chatbots, J.P.Morgan and Morgan Stanley have all implemented chatbots for their production environments. The beginning is not easy, but the result is worth it. Indeed, thanks to conversational agents, Morgan Stanley for instance, has been able to answer many of the Research-related questions of their clients (analyst profile, research coverage, etc) thanks to well-trained agents who answer efficiently all the questions [25].

F. Simulation and Recommendation of Market Trades and Investment Strategies

For their daily activities, traders and portfolio managers are always handling the trade-off between their stock and the changes in market data and economic indicators. Market data being on movement in real time, it is just a matter of seconds before a good buy is not anymore or the opposite. That is why, the main struggle in such a use case is to constantly have at hand data that is enriched by market data, transformed taking into consideration all the market and economic changes and customized by filtering the economic indicators that have an impact on the positions on hand.

For that purpose, applications for real-time computations of risk metrics, limits or price predictions are all interesting. Machine learning models like K-means for clustering are a good fit to select assets while taking portfolio diversification into consideration. Reinforced learning or deep learning can also be used to control the dynamics of the system in an optimized way [26]. Last, recommendation systems can be applied, such that based on historical transactions, returns and economic factors, recommendations of new operations can be made based on techniques like matrix factorization [27]. A paper that has also worked on the topic is "An equity fund recommendation system by combining transfer learning and the utility function of the prospect theory". Actual transaction data is used to test the technique suggested, and the result is to

answer the main question "what to buy?" given inputs of the operations and the market [28].

On the side of investment banks, MAN Group has also used artificial intelligence algorithms in order to define the best investment strategies on her AI-Managed funds. Based on millions of data points including transactional data, the tool developed by MAN identifies patterns to improve and adapt to the market changes. According to Bloomberg, this algorithm is behind more than 50% of MAN profit on its biggest funds [29].

G. Optimization of Internal and Support Processes

Finally, besides all the benefits that the main business lines can realize, the support functions can also make huge gains out of business analytics. Audit and Internal Control departments can get their controls automated, IT can have their system and technical incidents and problems automatically detected and resolved, and compliance and legal teams can make use of better fraud detection models among others. And for this family of use cases, all kinds of techniques could be used depending on the problem at hand.

One paper that handled a similar use case is "Automatic Detection of Relationships between Banking Operations using Machine Learning" [30], which solved the problem of matching operations in different systems. In fact, because there are different contexts in which the banks make use of the same operation (such as accounting, risk, internal control), banks end up having the same operation in different formats in multiple systems, and it is impossible for humans to figure out which transactions actually point out to the same operation. The contribution of this article is therefore a framework that enables operation matching thanks to machine learning mechanisms. The framework proposed is divided into 4 stages: Stage 1 is pre-processing, which consists of preparing the data from the data lake by selecting candidate operations, identifying possible records, merging and filtering. Stage 2 is machine learning based on ANN algorithms, and stage 3 is post-processing where a summary and combination of all results of the classifiers is performed in order to generate an exhaustive report of all matching operations. In order to test this framework, the authors applied it to 3 different scenarios, and they all ended up with at least 99,58% accuracy.

BNP Paribas Securities Services automatic trade matcher is definitely one of the best illustrations of this use cases' category. In fact, BNP Paribas built a tool named Smart Chaser in order to match trades between the Clients and Brokers. The tool, based on predictive analysis, has allowed BNP Paribas to identify patterns of mismatches in order to avoid them beforehand, thus reducing manual intervention on the trades [31].

Table I summarizes all the research and enterprise use cases discussed within this paper:

TABLE I.   DATA ANALYTICS USE CASES IN INVESTMENT BANKS

Use Case Family	*Use Case*	*Problem*	*Technique Used*	*In production or ongoing research?*
Prediction of financial market patterns	Big data analytics for financial Market volatility forecast based on support vector machine [14]	Analytics on high-frequency data	Support Vector Machine	Ongoing research
	NLP project for equity investing – JP Morgan Chase [15]	Research in equity investing	Natural Language Processing	In production
Detection of crisis situations	Monitoring Banking System Connectedness with Big Data [16]	Monitoring of firm connectedness	Linear Regression and Network Construction	Ongoing research
Client risk profiling	The value of big data for credit scoring: Enhancing financial inclusion using mobile phone data and social network analytics [18]	Credit scores' estimation	Network modeling	Ongoing research
	CitiVelocity – CitiBank [19]	Credit-sentiment monitor from news media	Sentiment Analysis	In production
Product recommendation for clients and prospects	An investment portfolio recommendation system for individual e-commerce users [20]	Investment portfolio recommendation	Collaborative filtering algorithm	Ongoing research
	SmartWealth – UBS [22]	Product recommendation for prospects	Machine learning algorithm	In production
	Investment opportunities approach – Goldman Sachs [23]	Investment opportunities definition for clients	Sentiment Analysis	In production
Client satisfaction follow-up	AI-Based chatbot service for financial industry [24]	Customer support	Machine Learning	Ongoing research
	Conversational agents – Morgan Stanley [25]	Client research questions	Machine Learning	In production
Simulation and Recommendation of market trades and investment strategies	An equity fund recommendation system by combining transfer learning and the utility function of the prospect theory [28]	Asset recommendation	Transfer Learning and Utility Function	Ongoing research
	AI-Managed funds – MAN Group [29]	Fund management	Artificial Intelligence	In production
Optimization of internal and support processes	Automatic Detection of Relationships between Banking Operations using Machine Learning [30]	Operation matching in different systems	Machine learning (ANN algorithms)	Ongoing research
	Smart Chaser – BNP Paribas Securities Services [31]	Automatic trade matching	Predictive analysis	In production

Based on this summary, the most important dimensions for an investment bank transformation emerge and structure the business and strategic objectives behind the data analytics. In fact, from a high-level perspective, one can isolate the external environment from the internal one within the bank; and within each environment, another detailed partitioning can be made. On one hand, the external environment includes market and economic data in addition to client information, whether it be their financial results or their perception of their relationship with the bank. On the other hand, in the internal environment side, one can distinguish between three main components: the product, the transaction and the risk metrics. All these dimensions combined give a 360° view of the investment bank's processes and enable a transformation impacting all the key factors. It is then up to each bank and depending on its context and pain points that the focus can be put on one or some of the use case families instead of all of them.

This being said, data today has become the backbone of all innovative and successful banking projects and the key element in order to take advantage of any of the use cases previously mentioned is to switch the whole decision-making process to one that is based primarily on figures and facts generated through analytics. As such, information technology projects within the bank should be strategically aligned with the business objectives, and all means and technological capabilities to achieve this cross-disciplinary cooperation should be deployed. In other words, the turning point for investment banks would rather be to transform to data-driven enterprises where data lies at the core of any process and is the most crucial factor in all business choices. This transformation is the only genuine change that can guarantee that all data sources, whether they are structured or unstructured, are being used and combined in order to create rich and insightful input data, and that analytics are achieving the best possible outputs in a timely manner through very fast and efficient data processing. However, this transformation does not come easily as there are various challenges to overcome, some of which will be discussed briefly in the next section.

IV. CHALLENGES OF BECOMING A DATA-DRIVEN INVESTMENT BANK

A. Use Case Prioritization

One of the first challenges investment banks face when starting analytics use cases is how to prioritize them. In fact, as illustrated by the use cases listed in the above section, there are so many inter-dependencies between the use cases because they mostly rely on the same data: a mix of internal deals and operations with market and economic data. Thus, deciding on what use cases should go first is not only a matter of business priorities but also a matter of technical constraints, mainly related to data availability.

B. Data Availability

This first point leads to the second concern of data analytics projects in the field, which is ensuring that all data is accessible. In fact, all the magic of analytics in capital markets lies in the correct combination of internal and external data, which is not always available in internal databases, and is rather found in data providers' platforms, social networks and websites of regulatory entities, ministries, national agencies and clients.

C. Cloud Integration

Besides, given the large amounts of data investment banks may end up processing due to the broad scope of external data needed for analytics, the nature of the data repository / data lake used can also be a difficult decision to make, and one that significantly impacts the long-run and cost of the analytics initiatives. There is an important trade-off to make between the strict regulations on investment banking and data confidentiality they require, and the big data volumes to process, in which a big chunk is mainly public and already available to everyone on the net.

D. Data Architecture

Another problem to consider is the technical and technological architecture of the environment hosting data analytics use cases. In fact, the peculiarity of the external data needed for investment banks is that it is mostly presented in numerous files (pdf, word or excel) of small or medium sizes, since every publication comes in a separate document. The block storage, that is mostly used in data analytics in all other domains, is not necessarily the best or most straightforward option as it is not initially designed for the storage of small files.

E. Event Sourcing

Last, when it comes to internal data, reproducing and replaying historical events is a must for analytics projects. However, the investment banks mostly rely in their IT departments on vendor software which are developed based on architectures that do not enable such options. Therefore, in order to get the most out of internal data too, a lot of work has to be done on the data within the IT systems in order to transform it to fit into an event-driven architecture, which is the only architecture that can answer all the analytics needs.

III. CONCLUSION AND FUTURE WORK

This paper describes why, now more than ever, launching analytics initiatives is important in investment banks. It presents also use cases that could be of interest to them and the processing techniques used by these use cases. It is meant to be a first-hand guideline for investment banks that are willing to engage in the process of data-driven transformation, and raises the challenges and problems they might face in this process, especially in terms of use case prioritization, data availability, cloud integration, data architecture and event sourcing.

As a future work, a deeper investigation of investment banks' data driven transformation will be performed by shedding light on its principles and key aspects and proposing solutions to the different challenges raised in order to complete the guideline. On the challenges' part, a special attention will be given to use case prioritization as it should serve all types of investment banks (those who are willing to transform to data-driven enterprises or those who are just looking for data analytics use cases to implement). This will be done by defining the most significant criteria to consider in this process depending on the bank's profile and strategic

orientations and building a model to sort the different alternatives accordingly.

REFERENCES

[1] Jelf, Owen. "Capital Markets Technology 2022." *Accenture*, 2018, www.accenture.com/t20180124t060525z__w__/us-en/_acnmedia/pdf-69/accenture-capital-markets-technology-2022.pdf.

[2] Morvan, Jérémy. "(PDF) Chapitre 1 - Les Marchés Financiers - ResearchGate." *Researchgate*, www.researchgate.net/publication/318641258_Chapitre_1_-_Les_marches_financiers.

[3] "Exposé Sur Le Marché Des Capitaux." *Economie-Monetaire-Et-Marche-Des-Capitaux*, economie-monetaire-et-marche-des-capitaux-23.webself.net/file/si267430/download/Exposé sur le marché des capitaux-fi4489230.pdf.

[4] Bilaus, Bogdan. "Portfolio Management for Institutional Investors." *Cfasociety*, June 2010, www.cfasociety.org/romania/Files/Analiza Financiara si de Investitii, Etica si Standarde Profesionale, Asociatia Brokerilor, June 2010/Bogdan Bilaus - PM for institutional investors.pdf.

[5] Laharach, Youssef. "Marchés Financiers Et Gestion De Portefeuille." *Laharach-Youssef*, 2014, www.laharach-youssef.com/medias/files/support-en-pdf-partie-i.pdf.

[6] Epstein, Adam, et al. Schroders, Vault Career Intelligence, 2015, www.schroders.com/en/sysglobalassets/digital/careers/2015-european-guide.pdf.

[7] Taylor-Sakyi, K. (n.d.). Big Data: Understanding Big Data (Tech.).

[8] Shende, Vikram. "Project Analytics to Improve Project and Portfolio Decision Making." *PMI*, 2017, www.pmi.org.in/conference2017/pdfs/papers-pdfs/theme-3-rapidly-changing-world/21-Project-Analytics-to-Improve-Project.pdf.

[9] Zikopoulos, P. (2012). Understanding big data: Analytics for enterprise class hadoop and streaming data.

[10] IBM® Institute for Business Value, and Saïd Business School at the University of Oxford. "Analytics: The Real-World Use of Big Data in Financial Services ." *IBM*, 2013, www.ibm.com/downloads/cas/E4BWZ1PY.

[11] Trelewicz, Jennifer Q. "Big Data and Big Money: The Role of Data in the Financial Sector." *InfoQ*, InfoQ, 17 Oct. 2017, www.infoq.com/articles/big-data-in-finance/.

[12] "(PDF) Big Data Analytics Enabled Smart Financial Services ..." Researchgate, 2017, www.researchgate.net/publication/321282806_Big_Data_Analytics_Enabled_Smart_Financial_Services_Opportunities_and_Challenges.

[13] Fan, Weiguo, and Michael D Gordon. "The Power of Social Media Analytics." *ACM*, 1 June 2014, cacm.acm.org/magazines/2014/6/175163-the-power-of-social-media-analytics/fulltext.

[14] Yang, Rongjun, et al. "Big Data Analytics for Financial Market Volatility Forecast Based on Support Vector Machine." *International Journal of Information Management*, Pergamon, 13 June 2019, www.sciencedirect.com/science/article/pii/S0268401218313604.

[15] "Innovation with Machine Learning." Jpmorgan.com, www.jpmorgan.com/insights/research/machine-learning.

[16]

[17] Ertek, Gurdal, and Murat Kaya. "Scoring and Predicting Risk Preferences." *Researchgate*, 2012, www.researchgate.net/publication/233755791_Scoring_and_Predicting_Risk_Preferences.

[18] Óskarsdóttir, María, et al. "The Value of Big Data for Credit Scoring: Enhancing Financial Inclusion Using Mobile Phone Data and Social Network Analytics." *Applied Soft Computing*, Elsevier, 9 Oct. 2018, www.sciencedirect.com/science/article/pii/S156849461830560X.

[19] "2020 Annual Report." Citigroup, Citigroup, 2020, www.citigroup.com/citi/investor/quarterly/2021/ar20_en.pdf.

[20] O. Isinkaye, Folasade, et al. Recommendation Systems: Principles, Methods and Evaluation, Researchgate, 2015, www.researchgate.net/publication/283180981_Recommendation_systems_Principles_methods_and_evaluation.

[21] Li, Xiang, and Chunxia Yu. "An Investment Portfolio Recommendation System for Individual E-Commerce Users ." *Semantics Scholar*, International Conference on Production Research , 2017, pdfs.semanticscholar.org/d5b7/53157004b28d892f18 d508e8e3aa715b2da3.pdf.

[22] Wealth and Asset Management 2022: The Path to Digital Leadership, Oracle, 2017, www.oracle.com/assets/wealth-report-summary-full-report-4010572.pdf.

[23] *The Role of Big Data in Investing*, www.gsam.com/content/gsam/us/en/institutions/market-insights/gsam-insights/gsam-perspectives/2016/big-data/gsam-roundtable.html.

[24] Okuda, T. & Shoda, S.. (2018). AI-based chatbot service for financial industry. Fujitsu Scientific and Technical Journal. 54. 4-8.

[25] Global Digital Wealth Management Report, BCG, 2019, media-publications.bcg.com/BCG-Global-Digital-Wealth-Management-Report-2019-2020-ENG.pdf.

[26] Y. Ng, Andrew, and Stuart Russel. "Algorithms for Inverse Reinforcement Learning." *Ai.stanford*, ai.stanford.edu/~ang/papers/icml00-irl.pdf.

[27] Moylan, and Anderson. Nonlinear Regulator Theory and an Inverse Optimal Control Problem - IEEE Journals & Magazine, ieeexplore.ieee.org/document/1100365.

[28] Zhang, Li, et al. "An Equity Fund Recommendation System by Combing Transfer Learning and the Utility Function of the Prospect Theory." *The Journal of Finance and Data Science*, Elsevier, 14 Feb. 2018, www.sciencedirect.com/science/article/pii/S240591881730020X.

[29] "AI Pioneers in Investment Management." CFA Institute, CFA Institute, www.cfainstitute.org/-/media/documents/survey/AI-Pioneers-in-Investment -Management.ashx.

[30] González-Carrasco, Israel, et al. "Automatic Detection of Relationships between Banking Operations Using Machine Learning." *Information Sciences*, Elsevier, 12 Feb. 2019, www.sciencedirect.com/science/article/pii/S0020025519301409.

[31] "BNP Paribas Securities Services Brochure." BNP Paribas Securities Services, 2021, securities.cib.bnpparibas/app/uploads/sites/3/2021/05/ss-doc-broch-glance.pdf.

Hale, Galina, and Jose A. Lopez. "Monitoring Banking System Connectedness with Big Data." *Journal of Econometrics*, North-Holland, 16 Apr. 2019, www.sciencedirect.com/science/article/pii/S030440761930082X.