Deep Science Publishing
https://doi.org/10.70593/978-93-49910-98-0_6

Chapter 6: Applications of big data in identifying market trends, managing risks, and detecting financial fraud

6.1. Introduction

Big Data is emerging as a transformational tool in finance. Throughout the world, the financial sector has turned to Big Data as a source of new revenue generation, operational cost saving, and risk management for existing products. In this chapter, we present advances in the use of Big Data across different subsectors in finance, namely, asset management, banking, and insurance (Kumar, 2023; Jiang, 2024; Rao, 2024).

Big Data is typically defined by the so-called three Vs – vast amount of data, ranging from terabytes to zettabytes; diverse data sources, including structured transactional data and unstructured data generated from sensors and web activities; and velocity of data flow, which makes real-time processing of the data more desirable yet extremely challenging. Coupled with cheap data storage and processing tools enabled by cloud computing, organizations are increasingly leveraging Big Data to gain valuable business insights and further their objectives. Examples include online e-commerce companies recommending products based on users' activities; hospitals improving healthcare quality by analyzing drug prescription data; and manufacturers forecasting product demand, based on supply chain and sales data (Udeh et al., 2024; Bansal et al., 2025).

However, Big Data applications in finance are still nascent. In the case of risk management, for instance, regulators are only beginning to employ data obtained through data-sharing broker services to cross-validate banks' stress testing results. The use of social media data in finance remains sporadic and unregulated, despite its demonstrated porosity to bubble and crash information. At the same time, existing revenue streams like asset and portfolio management, trading, and detection of money laundering and fraud are also being increasingly challenged.

6.2. Understanding Market Trends

Understanding the financial trends in play in capital markets is often one of the key concerns for both issuer and investor. An investment strategy is almost always built on some assumption or an expectation on how the market will behave; whether it is bullish all the way or is due for correction. This becomes even more important for issuers who use markets both for raising capital as well as for liquidity in the stock and company. Apart from the research usually carried by issuers and investors on the current trends and their movement, big data today has significantly helped everyone concerned with capital markets to develop a sound understanding of the trends which help better investment decisions.

Fig 1: Applications of big Data

Before we delve into the role of big data in understanding market trends, let us first discuss what a market trend is and its important components that need to be understood correctly before identifying it. In any business, the market trend can be described as the general direction in which most of the prices once in an upward direction and on the other is constantly moving downward. Each market in finance, i.e., equity, bond and forex, usually has two key components: the cyclical component and the seasonal component. There exists some pattern with a propulsive increase or decrease in prices leading to the trend and the cycle as shown in market prices that move up or down. Each trend theory has its own concept of whether a trend is in price, volume, or both; but the general trend of price should go in each active market.

Over this period, prices increase more than they decrease, often leading to new highs; falling prices also would tend to make new lows and the net movement almost generally upward, during a major down trend. Such trends in financial assets can be identified and validated with big data, a technology capable of sifting billions of data points in real market prices and rising downward movements that happen over a period of time. Capital market players, analysts and technocrats alike have today developed skills, models and tools to assess what the market trends are, the forecast for the patterns in prices and cyclical movements, whether long-term or short-term or seasonal. Building trend forecasts has successfully employed the new technologies made available by big data that enables the market players to sift through a multitude of data points.

6.2.1. Defining Market Trends

Market price time-series are generally very confusing and difficult to analyze, as they are often mixed with noises and, because of their highly volatile nature, appear very irregular. In many cases, their underlying structures are not easy to discover. Some of the dominant approaches to analyze market prices can be divided into two groups: fundamental analysis and technical analysis. Fundamental analysis attempts to measure the intrinsic value of a security by examining related economic and financial factors, in order to find investment opportunities that already are undervalued or overvalued according to those measurements. Technical analysis attempts to analyze price movements of a security based on historical patterns and price trends, with a focus primarily on technical indicators such as moving averages or rates of relative strength. A common assumption employed in both analyses is the efficient market hypothesis, which states that financial markets are informationally efficient and that asset prices always reflect all available information. Because of this, it is increasingly important to leverage additional information and alternative methods, such as those from the field of behavioral finance, to capture asset mispricings.

Both of these analyses aim to define and analyze market trends, those unfamiliar and invisible paths that market prices trace over time. A market trend usually refers to the general price direction during a given time horizon. A trend could be defined upward, downward or sideways. These definitions could be made more precise in case market prices have to pass above or below specific thresholds, and the trend could last over some time. Furthermore, market trends can also embrace different markets. Moreover, actual market price movements are generally very noisy and irregular and could be described hierarchically by combining several trend components at different frequency levels.

6.2.2. Role of Big Data in Market Analysis

Before large volumes of digital data, market analysts relied on traditional sources for consumer sentiment analysis. Today, consumers increase the amount of data available by sharing their views on products and services through reviews and social media. The volume, velocity, variance, and veracity of Big Data originating from the digital universe is significantly larger than traditional data sources. Market analysts have tools available that allow them to process this vast trove of data and translate it into real-time business intelligence, which is critical for making informed decisions. Software applications that dissociate and detangle consumer-driven digital information allow the data analyst to look for trends contained within that information and create a picture of consumer sentiment. For any company, the voice of the consumer is the best measure of brand loyalty. Social media, which is continually updated, creates a clear picture of trends relating to a product, a competitor, or an entire industry. By following hashtags and commenters, market analysts gather excellent insight into consumer views and obtain a measure of brand loyalty during specific business cycles. Similarly, monitoring this information allows the identification of trends that relate to competitors and potentially impact company business.

6.2.3. Case Studies of Successful Market Trend Identification

Using deep machine learning, a group of researchers were able to predict market trends very effectively. To predict a market trend, they employed three different approaches. The first approach is technical analysis, the second approach is the fundamental growth rate, and the third is momentum using predictor variables. These approaches offer various perspectives for predicting stock trends, and the study found that combining them yields significantly better performance than a single approach. Their model, a fully convolution neural network, which utilizes density predictor variables, predicts the upward or downward trend of S&P500 index one day ahead with a 95% accuracy and 73% return over holding the index for a year. The model's prediction is highly correlatable with itself, and assists in deep downstream applications including portfolio optimization problems, appointment of algorithmic trading when the index trend is up, or sitting on the sideline and using big stock cash.

A novel method using a convolutional neural network is proposed to solve finance prediction problems using 2D CNN images with a 2D matrix scheme. The model predicts the movement of the price of the S&P500 index one day ahead based on two key predictor variables: "The Day-ahead Price Movement (%)" and the Correlation Coefficient. The proposed model's prediction accuracy, F1 score, and Matthews correlation coefficient are 73.55%, 0.4696, and 0.5717, respectively, and excel in the current finance prediction situation. The model output can be usefully applied to derive a portfolio optimization solution by diving the "The Day-ahead Price Movement (%)" into groups.

6.3. Risk Management in Financial Institutions

Applications of Big Data in Finance consists of ten main chapters focusing on the different areas of finance and a wide variety of sub-sections exploring these fields in depth, covering over 40 applications. Additionally, in each of these applications, we focus on relevant case studies where businesses are actively using these applications of Big Data in Finance to assess their decision-making process and financial models. In the world today, almost all sectors of financial markets are affected by technological innovation and are undergoing a fundamental change through the development of technology processes, enhancing the availability of data as well as computer resources. Advanced technology allows financial institutions to analyze and assess portfolios, client characteristics, and market risks with better precision reducing the risk of volatility and potential losses in financial sectors, including banks and insurances. Consequently, financial organizations, especially risk and portfolio managers are reviewing their strategies and processes to implement new technologies and apply Big Data and predictive analytics for better analysis of historical information to pre-emptively identify possible future incidents. In addition, financial firms are searching for experts with knowledge in Big Data technologies and predictive analytics to lead their projects as well as hiring third-party consulting firms to assist the internal team on assessing their Big Data use. One of the major applications of Big Data in financial institutions is risk management. Financial organizations have a large amount of historical and time delayed data through which analytics can be used to access predicted long-term future results based on current financial market events. Particularly for insurance firms, their existing large amount of data can be assisted with third-party data and machine learning tools to develop predictive models which can significantly reduce prediction dead time.

6.3.1. Types of Financial Risks

Financial Psychology provides an opportunity for risk managers to understand the mental limitations that prevent stakeholders from fully grasping the realities of risk. One of the central tasks of risk management is to reduce large, client-invisible and risky events to manageable risk entities that can be hedged or even passed on to third-party providers of risk capacity. These risk entities are not without limitations and must therefore be modeled accurately using up-to-date understanding and innovative resources. They must be as accurate as possible in order to minimize capital drain on the financial institution while ensuring optimized capital allocation. The risk entities must be transparent such that both clients and end-users of the risk calculating units understand enough about them to ensure no negative hits on reputation and at worst survive-managed regulatory and compliance reporting. They must be highly reliable to avoid the extreme systematic events of market explosions. And lastly, they have to be flexible enough to allow for configuration for special exposures, as modeled econometric risks cannot be implemented uniformly over the balance sheet of a financial institution.

Risk landscape encompasses a collection of risk types which take to account the operating environment of a bank. The following list presents a collection of various risk types, all applicable to financial institutions to varying degrees insomuch as they reflect the portfolios and specialization of the bank. These are the usual suspects but not necessarily a comprehensive list as new risk types constantly appear. Market risk – the risk for trading desks, causing capital erosion. Credit risk – the classic pitfall of banks related to the collateral nature of loans and insufficient due diligence. Strategic risk – a new entry in the risk landscape that reflects the potential threat for a bank stemming from wrong positioning and increased loss potential as a result of reduced activities or liquidation due to a wrong business model. Business risk – a risk very much akin to strategic risk, which through too volatile business results impedes reliable risk modeling. Reputation risk – the Bank's Achilles heel, where erosion of reputation results in a heightened sensitivity of stakeholders, with a potentially real impact on capital. Economic capital risk – the risk that economic capital will not be sufficient to prevent multiple insolvencies due to portfolio limitations. Compliance risk – a potentially overwhelming loss even for the most functional organization that derives from breaching laws and regulations. Systems risk – a risk that if triggered would not only affect one bank but potentially bring the whole financial system to its knees.

6.3.2. Big Data Techniques for Risk Assessment

Financial risk assessment is a broad area that includes credit, liquidity, market, operational, solvency, and systemic risk as components of overall operations risk. In this section, we focus on data processing approaches. Innovations in this area are rather methodological. Specific applications are shown in the section, which deals with specific risk areas. A variety of Financial service activities make Financial institutions combine multiple risk types. Consolidating risk assessments meant initially for specific risk types brings substantive methodological challenges.

The revolution in the data processing potential in business moves and disturbs Financial economists out of their comfort zone. There has always been an urge to rely more on models. The body of mathematical and statistical modeling seems extremely rich to refuse its role. It is unlikely that the task of Financial modeling will ever converge to a central limit, and the models converge to a small number of expressions that investors overly and unsuccessfully apply in their strategies. However, many would agree that modeling was limited by the restrictions in data processing potential in other domains of business, resulting in the frameworks used in finance being rather basic compared to innovations that occurred in applied probability, econometrics, statistics, mathematics, and physics.

Research must deliver sound methods that either use the existing domain knowledge for Financial applications and form the basis for empirical validation, or pull Financial institutions enough out of their past experience to explore results of new methods. There is a cost and risk attached to nailing Financial institutions to models that work in other domains but are not validated in finance. Depending on data usage, risk management could use qualitative methods, build simple semi-quantitative models, or rely on data-intensive quantitative methods.

6.3.3. Predictive Analytics in Risk Management

Big data brings with it the potential for upheaval in the mechanisms available to financial institutions for risk management oversight and governance. The business models of the institutions that help manage credit and market risk for their corporate customers may not, however, necessarily change as dramatically. The financial situation is this.

Advances in computational power and the availability of open source packages that allow for easy use of machine learning tools means that large amounts of unstructured data can allow for significant enhancements to work in risk management. The analytical methods used at most financial institutions for these tasks, however, will require much more than a bit of tuning, even with the huge availability of data capital. Ultimately, beyond a small number of esoteric institutions in specific geographical locations, those organizations helping clients assess and hedge corporate credit and market risk should not have to fear for what role they will play going forward as their apparatus for helping stabilize the economy is enhanced. Global macroeconomic models that have been employed would greatly benefit from the use of big data. Our models that track the effect of credit and market risk on the global economy could use more timely and accurate forecasts of economic data.

While many financial institutions have begun to experiment with big data to find ways to enhance the work they are currently doing, their models for predicting the likelihood of events triggering extreme moves in economic risk need refinement. Findings have shown big advantages in using a mixture of models approach to short-term high frequency prediction errors in stock price indexes. With the increasing prevalence of algorithmic trading in financial markets, accurate predictions in high frequency volatility have important ramifications for such areas as monetary policy and the implementation of monetary policy through open market operations.

6.4. Detecting Financial Fraud

Big Data analytics have helped increase efficiency in many activities, including in the field of detecting financial fraud. Given the sheer number of transactions executed in the financial services industry, conducting a diligent investigation to search for evidences of financial dirty dealings, such as financial statement fraud or money laundering, is not an easy task. At the same time, if nothing is done to efficiently mitigate these problems, the costs of any omissions will increase sharply, either for financial institutions or for society.

Financial fraud is a major source of loss, humiliation and sustainability problems for financial institutions, governments, and customers. Typical types of financial fraud include credit card fraud, asset misappropriation, money laundering, identity fraud, and accounting fraud. Recent estimates suggest that the value of losses due to different types of fraud is around US$700 billion a year. Fraud detection, especially in the case of money laundering, is an extremely time-consuming and costly endeavor. Current supervisory procedures to minimize financial crime are instead based on case studies and heuristic criteria to detect outliers.

The evolution of Big Data technologies that provide fast computational and predictive capabilities in a context of working with complex and dynamic data of various types is making possible to use a new generation of data mining techniques. These enabling techniques have the potential of elevating substantially the capabilities of financial institutions to prevent fraud cases, either through intelligent decision-making or through alerting to possible fraud cases. These techniques will not substitute for the judgment and scrutiny of money laundering compliance officers in detecting fraud. The idea is instead to assist anti-fraud efforts with advanced predictive capabilities that make it easier and faster to investigate possible dirty cases.

6.4.1. Overview of Financial Fraud Types

Fraud is described as “any intentional act or omission designed to deceive others, therefore gaining a benefit for oneself or causing a loss to another party.” Fraud is an unlawful and methodical act, especially carried out for monetary gain by a particular group of people, perpetrated for their own benefit, regardless of its consequences on people or societies. The impact of fraudulent activities is legion globally, affecting people financially and psychologically. Areas of incidence include business; finance industry; and financial markets. Fraud case studies report about the impact on economic sustainability, growth, cost burden of companies, and loss of employment.

Fraud could take place in various industries; however, some of the sectors highly prone to fraud are: financial services; government, including public administration; health; manufacturing; insurance; accommodation, food, and beverage; professional services; real estate; transportation; wholesale and retail; and energy, mining, and utilities. Financial fraud refers to schemes that induce a financial institution or its customers to deliver money or property fraudulently for failure to receive money or property which justifies the transaction. Financial fraud adheres to specific consequences that could have occurred, such as bankruptcy of an organization and increased cost to the consumer. Financial fraud is broadly classified into five types: credit card fraud; investor fraud; insurance fraud; money laundering; and mortgage fraud.

6.4.2. Big Data Tools for Fraud Detection

Fraud detection focuses on identifying fraudulent activities typically using pseudonymous data. It has received significantly more attention from the big data community than fraud prevention. Aside from the volume of fraud detection data, the dimensions of fraud detection problems make traditional detection methods expensive and unscalable. Fraud detection needs to automatically deal with massive volume, kolossal velocity, and extremely high-dimensional systems without a priori knowledge on the qualitative distributions of the data. Big detection and inferring on massive amounts of fraud would still be data costly. New hardware architectures such as graphic card processing units speed up many core machine learning, majority, and nearest neighbors at low latency and high volume.

Fraud detection tools differ in many ways. Data scientists have roots in statistics, which provides a solid foundation for designing and validating discovery tools on one or two dimensional data challenges. Besides the tool design aspects of flexibility and ease of use, what distinguishes statistical fraud detection based on estimation, evaluation, and hypothesis tests from graph-based tools based on probabilistic graphical models, what distinguishes tree-based tools based on prediction and classification designed for single-dimensional, hierarchical decision tasks from collective analysis tools. What distinguishes similarity and nearest neighbor tools designed for small tasks from structured learning tools designed for sequential, noisy data is the power of the user questions as analyzed here.

6.4.3. Machine Learning Algorithms in Fraud Prevention

In recent years, practical uses of big data in the financial industry have largely shifted to data analysis and semantic analytics. Today, banks and other financial institutions are increasingly looking to new resources and innovative techniques. As a result, the business landscape is changing, forcing these organizations to adapt, innovate, and be more agile than ever before. Balancing growth and profitability is at the core of the financial decision-making process, and organizations are beginning to recognize that new advances in big data technologies can positively impact executive-level decisions by providing data-driven insights. Sophisticated algorithms associated with structured and unstructured big data processing can help with a multitude of quantitative and qualitative financial decisions tied to risk management.

An unusual spike of transactions, abnormal patterns in the cash flow of a company, the increase in repetitive requests for regulation infringements, and other similar red flags relying both on quantitative metrics and qualitative assessments are the basis of several supervised learning kernels from the broader set of financial fraud machine learning models. Besides these, anomaly detection is one of the techniques to automatically detect fraud through big data. First, by utilizing a dimensionality reduction mechanism using unsupervised learning or by the calculation of pattern aggregate features using a semi-supervised learning kernel, exploratory techniques are applied so that the more abnormal observation and activity coordinates can be devised. Then, based on the residuals of the dimensionality reduction method or the activity deviation from the aggregate featurizing mechanism, anomalous observations can be detected using one of several supervised or unsupervised outlier detection techniques.

6.5. Data Sources for Financial Analysis

The production and availability of data have reached unprecedented levels. Financial analysts have had to adapt to new approaches, methods, and techniques to extract value from the substantial amount of data that is available. Finance has traditionally relied on well-established sources of structured historical data from index providers, market data vendors, central banks, and statistical agencies to develop tools that support decision-making.

Most of the available big data in the world is unstructured data in sources such as news, earnings calls, discussion boards, social media, satellite images, and blogs. The emergence of large unstructured data sources is adding challenges but also opportunities for financial analysis. Firms and investors need to explore new data sources to capitalize on opportunities, detect new risks, and improve investment decisions. They also have to manage the complexity when dealing with various structured and unstructured data across public and private sources in different frequency formats and with various levels of availability and quality issues and try to develop innovative methods to address such challenges.

The variety of available data means that firms cannot ignore alternative data sources. Out of the market and non-market data, unquantifiable information in alternative data sources and its underlying drivers are often neglected in financial analysis, leading to inferior decisions. This may also imply that firms are slow to react to opportunities and threats. The use of alternative data did reduce the information disadvantage of small firms with respect to large ones. The increasing use of scraped and crawled news articles, feeds, and specialized services suggests that the financial industry as a whole is becoming more quantitative and analytical in nature.

6.5.1. Structured vs Unstructured Data

Given the advances in technology over the last several decades, the volume and variety of data available are unprecedented and are growing exponentially. High volume is not new to financial data users. Consider the thousands of transactions that occur daily in the NYSE. Transaction price and volume data are reported every second, and electrically adjusted for corporate actions like stock splits, dividends, or security mergers. These data represent structured information, as each is characterized by a well-defined number of simple attributes taken from a precisely defined universe. The data are easily manipulated, since they present themselves as standard tabular arrays in relational databases. Service providers compile this information into financial data vendor services, which present real-time financial information for many millions of prices on equities, indices, options, commodities, exchange-traded funds, mutual funds, and foreign exchange.

By contrast, consider the terabytes of text, audio, and video files that are generated daily by various media providers that record spoken words, visual images, or written text. Very much as the huge amount of structured trading data represent daily activities on the lifeblood of the global economy, these unstructured sources reflect human feelings, thoughts, and ideas. What do we believe are the most significant factors that influence the markets? What moves them? Are we concerned about macroeconomic conditions like interest rates, monetary and fiscal policy, and economic growth? Are we focused on microeconomic challenges that firms face, especially corporate governance issues, such as financing, capitalization, and regulation? Are we optimistic or pessimistic about equities, bonds, real estate, commodities, or currency? Are we worried about crime, terrorism, war, bankruptcy, inflation, trade, or government debt? These sources of qualitative data do not simply represent the quantitative measures of our economy and society; they express our opinions and emotional states.

6.5.2. Public vs Private Data Sources

In finance, we can rely on the fact that many data sources have been widely used because of the trustworthiness of the data. Financial statements, as required by law, are published by companies and their data reflect their true financial status.

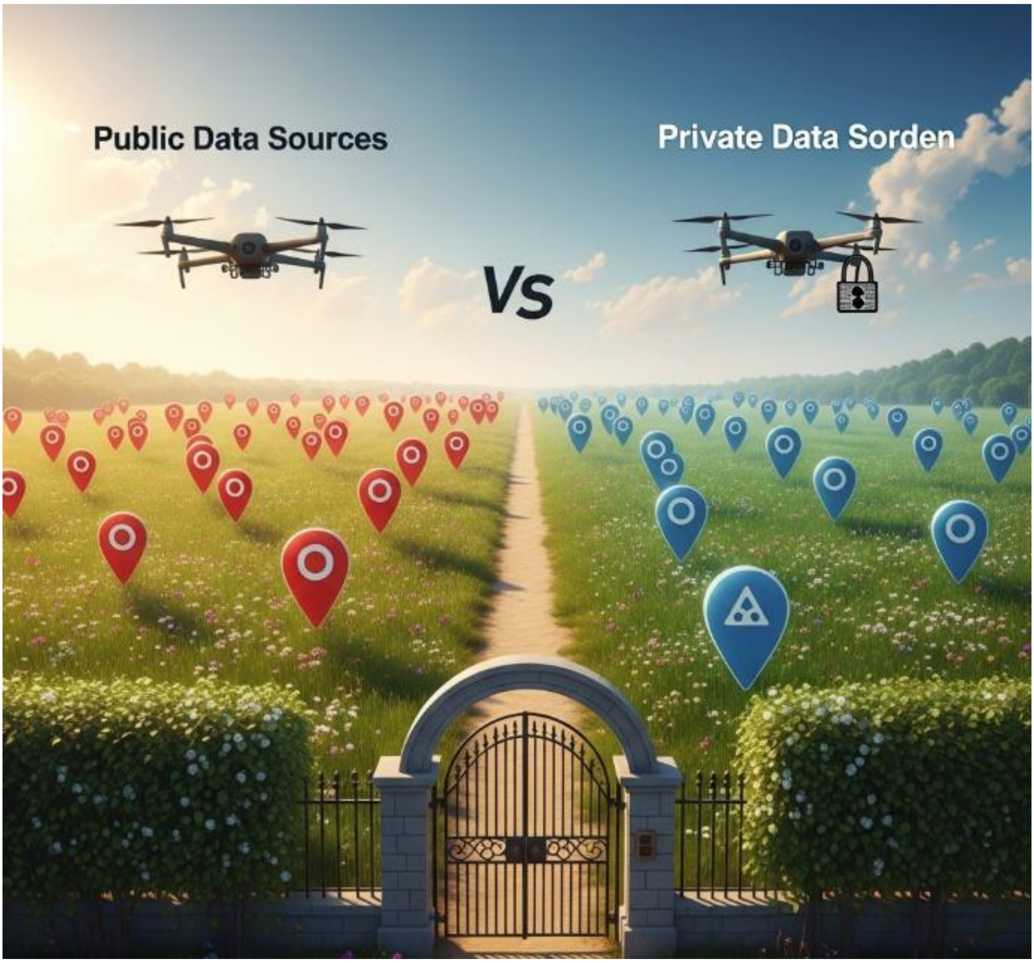

Fig 2 : Public vs Private Data Sources

Data integrity is, of course, an important issue and we will discuss that. However, finance has a wealth of data that is in the public domain. Examples include key economic data, exchange rates, interest rates, inflation data, statistics related to the issuing of debt and money, economic reports, and balance of payment statistics, all of which are published by central banks and government agencies. Financial Time Series from various exchanges that provide ticker and historical data are also publicly available. These data predict financial distress and bankruptcy. They also have an effect on returns from stocks, bonds, currencies, and commodities, as well as OTC derivatives such as interest rate and foreign exchange swap contracts. Macro data are also used to predict investment policy, changes in GDP, and tax policy. Macro data at the multicountry level are used to forecast currency return levels.

That said, access to private data can help increase the accuracy of our analysis. Information is what gives an investor an edge in the market. Financial data are considered a commodity; they are widely available. Alternative data that are not commonly available may be the difference between superior knowledge and herd mentality. Proprietary data that are obtained privately may result in abnormal returns, or may be highly correlated with metrics that are predictive of performance yet are unavailable in the public domain. Data on hedge fund contracts, litigation contracts and tax filings, class-action lawsuits, earnings calls, pitch books, and private equity transactions are some examples of private data. Investing in private data may be expensive; however, knowing how to create an edge by investing in a niche may generate huge excess returns.

6.5.3. Real-Time Data Streaming

The importance of being the first in the market to have an opinion or relevant information on financial movements has led to the emergence of companies that offer market news and updates in real-time. Financial investors, hedge funds, and institutions can subscribe to their services, where they obtain news at the same time as it is produced in the media, in order to analyze and react in seconds. After facing a system failure a few years ago, regulations were introduced that require exchanges to communicate their information in real-time, pertaining to volume, prices, and quotes of stocks. This exposes traders to the risk of receiving outdated information or no data at all.

From a theoretical perspective, the existence of institutional participants in the market that analyze an event, process it, and take investment decisions, should lead to an instant price adjustment, mitigating the profits to be made from potential upcoming anomalies or inefficiencies. This was supported by prior literature foresighting that high-frequency trading strategy following common earlier proponed models would get diminished effectiveness and be unfeasible. However, these studies did not take into account the actual existence of time stamps hyper-concentrated and limited over the financial event happening. A fact that allows particular categories of agents in the market to transform such an occurrence into financial fortune, when "just-in-time information" is exploited. This discrepancy of treatment between the ways in which the asset pricing model predicts price adjustment speeds and the observation of excessive rewards associated with particular traders seems paradoxical.

The resolution of this incongruence is the hypothesis of a dual speed of adjustment, to the point where agents subject to the predictive model capital market assumptions react almost instantaneously, while those agents staying outside are subject to inertia and result in price adjustments in a sluggish manner. This concept has been supported by various studies. The high-frequency trading activity includes hedge funds, proprietary shops and brokerage either acting in customers' favor or market making. Nevertheless, there are also institutions active at longer time spans who exploit slow market making for their benefit.

6.6. Data Analytics Techniques

Advances in computational storage and processing infrastructure have enhanced the scale and availability of data available for analysis, which in turn has increased its usage in a variety of industries. This increased availability of data has led to the birth and maturation of data analytics, which is defined as the use of a variety of tools and techniques to extract information from data to support decision-making. Data analytics supports better decision-making because of its ability to provide timely and relevant information and insights that would not be available without conducting such analytical processes on the data. Simply stated, data analytics transforms data into insights. Within data analytics, there are three broad categories. Descriptive analytics answers the question "What happened?" by finding patterns in historical data. Predictive analytics answers the question "What is likely to happen?" by using the patterns from historical data to model the behavior of real-world systems. Prescriptive analytics answers the question: "What should we do?" by combining descriptive and predictive analytics with optimization, simulation, and other techniques to provide specific advice based on the predicted patterns of likely future outcomes. These three analytics types are increasingly integrated with each other, because the value in predictive and prescriptive analytics is often enhanced by the use of descriptive analytics. In business applications, descriptive analytics is often presented in a visualization format, usually through dashboards, scorecards, or other reporting tools, which allow end-users to "see" the important insights present in the data.

6.6.1. Descriptive Analytics

The term data analytics refers to a set of techniques that allow you to analyze data in order to understand it and create a better and richer information about the real world than the original data. Data analytics encompasses a series of different types of analysis, ranging from the very simple to the very complex. Data analytics can be classified following several criteria, such as what's the input data, what's the aim of the analysis and what techniques have been used. The objectives of porting out these analyses can vary widely. The most common ones are: supporting decision-making processes, improving process efficiency and effectiveness, discovering and predicting trends and patterns in data. In business, decision-making is often supported by dashboards and ad hoc reports. However, dashboards require a considerable upfront investment.

Descriptive analytics can be defined as the set of techniques that allows us to better understand the past and answer questions such as: ”what happened?”, ”where did it happen?” and ”how many times did it happen?”, describing the characteristics of the recorded data. They can use how much and how many of the traditional statistics functions, in addition to exploring and visualizing the behaviors and relations within the data. Exploratory Data Analysis includes offeratives, number of violations, visualizations, and statistical measures for missing data; outliers; extremes; news; relations between dependent and independent variables; stratifications and subgroups; distributions; shape; count data; dates; and temporal sequences. Because its exploratory nature, descriptive analytics can give a very large variety of results, which may include unprocessed and incomplete data; and may take very different forms, such as scatter plots; correlation matrices; bar charts; and so forth.

6.6.2. Predictive Analytics

The knowledge obtained through descriptive analytics can be employed to infer the possible future. Predictive models rely on past data and present conditions to establish expectations for a dependent variable of interest. They are built on supervised learning techniques, where the algorithm is provided with a historical sample containing the discrepancy between the model objective and its predicted value. The trained algorithm then uses new input data to generate estimates that can be timely and useful in several financial applications.

Predictive models have been widely implemented in many areas, from credit scoring to portfolio optimization, fraud detection, corporate bankruptcy prediction, crisis prediction, insurance premium pricing, mortgage default and prepayment prediction, stock price prediction, risk assessment of non-listed companies, stock market forecasting, etc. However, due to the non-stationarity of the financial market, prediction is a challenging task. Predictive models based on traditional statistical methods may fail in capturing the complexity of the dynamics of financial series. In contrast, ML algorithms are capable of handling a wealth of historical information in a data-driven way by automatically detecting different relationships, features, and patterns in data, which may lead to more accurate decision rules for prediction. Accordingly, numerous studies have explored the prediction power of ML on various financial and alternative financial assets. Banks are using Neural Networks to change their internal structures for profit prediction.

6.6.3. Prescriptive Analytics

The last knowledge of analytics is working with "what is the best to do?" Prescriptive analytics is a decision-making technique that can be informed by descriptive analytics and predictive analytics. In line with decisions, where specific demands or parameters apply and which can be modeled with a clearly defined mathematical formulation, prescriptive analytics can assist by determining the optimal decision. Intensive research in the area of operations research has brought some success in providing optimal decision recommendations for a clearly defined decision area. A more general application of prescriptive analytics is possible if the particular modeling of the decision also incorporates predictive analytics.

Prescriptive analytics relies on predictive analytics to provide information on which is the most likely possible or probable scenario for the future, providing additional information for the analysis. PWM is focusing also on providing additional analyses and enabling investigations of special questions in both demand and lead time processing. Typical applications are seasonal and cyclic variations, features and requirements for seasonally changed products, and providing additional analysis of product lead time by testing short reduced lead time increases. Prescriptive analytics is nearer related to the underlying algorithms used in predictive analytics and are originating from the area of operations research. In its basic understanding, prescriptive analytics is producing a recommendation for the decision maker by informing him about the results of mathematical optimization algorithms.

Prescriptive analytics relies on the fact that the decision problem can be defined by a mathematical model. The basic structure of the mathematical optimization model for prescriptive analytics can be summarized as follows. An objective function describes the goal of the decision process and describes the performance of the alternative decisions. Constraints define the decision area and possible decisions that can be taken.

6.7. Challenges in Implementing Big Data Solutions

Despite the wide availability of tools to support Big Data implementations, companies in general, and financial services in particular, report that such implementations are a greater challenge than anticipated. In financial services, institutions have been using large volumes of data for decades; however, these were structured datasets generated by the internal systems of organizations. The constraints of processing these data limited data use to simple statistical data. In writing rules to approve loan applications, for example, institutions did not consider non-structured information such as social media, as this was not available in the past. The current explosion of non-structured data, generated in real-time by customers of financial services are serving for a new set of applications in finance. The low cost of web-based technology has made them the preferred platforms used by financial services companies.

The implementation of Big Data solutions in finance faces unique constraints that could lead to less-than-optimal solutions. Some of the major challenges in implementing Big Data solutions consist of ensuring data privacy, integrating these solutions with existing legacy systems, and the shortage of workforce with the necessary skills to implement the desired analytics.

Customers of financial services are concerned about how the information companies collect about their habits is used. For example, when a bank considers the loan application of a customer, and examines the customer's social media posts to understand the risk, it is invading the virtual space of the customer. Although this is legal today, banks will have to be transparent and disclose their actions in order to maintain the privacy of customers. Today, however, no one really knows how customers will react to discovering that organizations are tracking what they do. The historical experience of banks can only be assessed through an analysis of actual implementations.

6.7.1. Data Privacy Concerns

Data is at the heart of Big Data, and it is the use of unprecedented volumes of personal data recorded by sensors, digital devices or fingerprints in complex combinations to define the financial lifeblood of practically every economic agent from citizens, businesses to states that can result in both impressive gains and breaches of data privacy. The use of Big Data to track individual behavior at a highly detailed level and represent socio-demographic, mobility, consumption, creditworthiness, health, purchasing ability, etc. profiles to market goods or provide services but also to predict how they would react to a certain situation is everywhere. In addition, the interpretation of patterns in the data can allow insights and consequences that go beyond their intended purposes, such as the potential return of a previously unknown medical condition, or prove difficult to unearth or discard once exposed. Therefore, data privacy is crucial in teaching ethical practices in the collection and usage of such data.

Almost anything nowadays leaves a digital trace, but privacy is in the act of making choices, and citizens are usually very concerned about data usage for utilization without their consent or knowledge of financial and consumer behavior. Although private companies have information security policies in place, security is not a guarantee that a data threat will not take place, be it a terrorist attack or a cyber attack that expose confidential data. Given past hacks and breaches, many financial and insurance companies are not willing to embrace Big Data technology for fear of undesirable consequences. Such reluctance raises many questions for the future of the financial industry and business generally, which is largely predicated on trust, as to how such concerns will affect the introduction of such technologies.

6.7.2. Integration with Legacy Systems

Introducing a new technology in an old infrastructure often requires a long, painful integration process and it takes time and energy to make the two coexist peacefully, possibly forever. Unfortunately, the distributional differences between big data systems and legacy systems are so deep as to make seamless integration virtually impossible. In spite of this, it is an undeniable fact that, with very few exceptions, financial institutions still keep their legacy systems as the single most important source of data processing and storage. It is, therefore, imperative to make big data systems capable of smooth integration with these legacy systems. Integrating big data systems and legacy systems can take place on two levels: the data level and the engineering level, the former being obviously a child of the latter.

At the data level, we want the big data platform to support the various data formats and modes of operations of old systems without needing either heavy alterations to the legacy systems or expensive data migrations towards the big data pipeline. These alterations and migrations are very expensive and mostly impossible and defeat the purpose of using big data technology, which is supposed to help the company reduce costs and be more efficient. Meeting data integration requirements is important not only for the company implementing big data technologies but also for solution providers, which, to remain competitive, want their products to be compatible with legacy systems. Integrating big data systems and legacy systems can take place at the engineering level via connectors, which are software tools bridging the two environments.

6.7.3. Skilled Workforce Shortage

While there has been no shortage of interest in the applications of Big Data technologies, there has been a shortage of sufficiently skilled workers to actually implement analytics initiatives. Organizations across industry sectors recognize the need to adopt Big Data technology strategies to keep pace with competitors who are already capitalizing on the potential value of their data. A rising number of business and technical job offerings are explicitly calling for experience with the Hadoop ecosystem or specific Big Data skills, such as real-time data analysis.

The market for Big Data technology and services will grow from its current $16 billion to $232 billion by 2016. The runaway growth of business and technical job postings for analysts, scientists and Big Data generalists indicates how difficult it is for organizations in any industry to find talent. It is not just the number of postings that has skyrocketed, but also the staggering pace at which the postings are growing. Online job postings requiring skills associated with Hadoop have more than doubled in the past year. Moreover, the need for both the individual industry and technical skills is projected to increase dramatically. Estimates suggest that, by 2018, the U.S. could face a shortage of up to 190,000 workers with "deep analytical" skills and a shortage of more than 1.5 million managers and analysts with the know-how to use the analysis of Big Data to make effective decisions.

Given that it's unlikely that many universities will rapidly change their curriculums to train large numbers of Big Data workers quickly, employers are likely to have a difficult time finding highly trained employees in the near future. Instead, organizations will have to rely on creative employee training and development programs that foster internal talent. Leaders should also look for IT generalists who specialize in a range of different technologies, and who have the aptitude and desire to grow their Big Data skills.

6.8. Regulatory Considerations

The evolution of technology has enabled research in many projects related to innovative financial services, which are now available to consumers around the world. From the free-flowing information of social media to the sophisticated sentiment analysis involving facial expressions at international airports, finite and uncontrolled data are now embedded in financial decisions. However, these recent developments related to big data usage are raising greater interest from regulators and supervisors. Combined with the growth of open finance, the analysis of a multitude of datasets that leverage artificial intelligence algorithms has resulted in the emergence of innumerable successful innovative solutions, offering subscription-less or very low-cost financial and non-financial services. From bank accounts to investments strategies and even creditworthiness detection, fintechs have been naively and successfully paved the way for digitally-poised consumers to enjoy their uninterruptible use of every day-to-day service they are exposed to. It is thus imperative to keep in check the protection of customers. Anyone's activity generates online traces, which could allow any institution, private or governmental, to anticipate that individual's interest in an event-related product or information – and potentially exploit that importance with no restraint.

The first governmental propensity to interrogate the influence of online data in the financial ecosystem was expressed with the release of regulations, which was drafted by the necessity to keep an egoistic control related to personal data usage and that ultimately obliges any financial institution to comply to a strict set of regulations before employing datasets that involve such a magnitude of consumers or partners. While big datasets drawn from social media and tech firms would in turn allow banks to develop services specifically tailored upon the users' wants, desires, expectations or future actions, the dissonance with regulations would limit their applications to a handful of activities behind their scope of action. Such paths are now widely used by companies to minimize potential consequence of breaches.

6.8.1. Compliance with Financial Regulations

Big data technologies afford financial services firms the capacity and insight necessary to stem offenders years not possible, but the gilded edge is how such technologies threaten financial stability objectives and consumer trust if misused. The need for solid risk governance has been an urgent call to action for the industry ever since advanced approaches for credit and operational risk, the data security and consumer financial protection reforms followed by heavy surveillance formed a second line of defense that should have been brought to bear for risk IT prior to the crisis. Surveillance systems for detecting insider money laundering and other offenses are increasingly AI-led, with new big data technologies focusing attention on developing suspicious transaction analysis and prediction algorithms, but these methodologies are only as strong as the data quality that drives them. Critical perspectives enumerate the serious issues that operationalize such recommendations and work have created a taxonomical framework of supervisory recommendations to ensure the economic potential of AI in financial services and security for consumers, investors and the risk management and governance infrastructure of the financial industry are not in opposition.

While the strong association for using big data in finance comes from machine learning's high-technology deployment in credit scoring, more conventional uses of designed analytical systems can also be derived from big data. These recourses can be used to help meet objectives and levels for stable funding ratios, excessive growth, structural liquidity risk, maturity mismatches, and sudden withdrawal of liquidity exposure through new risk signal systems that can create models that respond rapidly to early signs of risk level movement across the supply chain as increased risk prompters for policy action to the line controllers.

6.8.2. Impact of GDPR on Financial Data Usage

Data privacy legislation has been nearly nonexistent over the last two decades, which is surprising given the rapidly growing number of data breaches. Governments have implemented tailored laws for verticals, but a general framework has been lacking. In the private sector, organizations have set up their own regulations via compliance with standards or through use of models at many universities and facilitating organizations for research data access.

But now the European Court of Justice has overturned previous frameworks allowing data sharing convenient for businesses. More generally, the EU General Data Protection Regulation is the first broad-based regulation governing how organizations treat personal data privacy. It prioritizes individual privacy over free flow of data across national boundaries. Although only applicable to organizations doing business in the EU or with EU citizens, it will redefine business practices worldwide as organizations seek a single set of policies instead of a patchwork of laws. Enterprises in particular will need to rethink how they optimize value from the vast stores of personal data they have instead of just complying with the letter of the law.

6.9. Future Trends in Big Data Applications

The trends of Big Data applications in the financial sector, more or less combined, such as the automation process by Artificial Intelligence, Blockchain, and Big Data coherent adoption by the user. In academia, the applications of Machine Learning in Finance or Data Analytics and Artificial Intelligence in financial service and operations are topics evidently growing in importance. In practice, the appraisal of Artificial Intelligence is also growing. For example, a consultancy company published that a significant percentage of companies operating in financial services were implementing Artificial Intelligence.

On the other hand, the attention around the Blockchain relation with the Big Data phenomenon has also been increasing. Admittedly, there is room to better explore the relationships on a concrete level, including the Big Chain technology (that is the relation between Big Data with Blockchain). The conclusion is that research directions still open exploring the theoretical relation; the applications linked to blockchain mining and optimizing blockchain-based data sharing; the tokenizing Big Data for social good using blockchain technology; the Big Data play in blockchain governance; Tokenomics and Big Data's role on shaping blockchain governance policies; the Ethical design collateral; or the Technical solutions considering the Big Data dimension. Moreover, the theoretical, empirical, and technological analyses to reduce the cost of Big Data in blockchains; to increase the activity to deliver Big Data provided by blockchain; and the synergies to integrate Bitcoin, Data, and Energy. Lastly, for the economy experts, including the direction Big Data Analysis and the economic limit for blockchains into the Taxonomy of blockchain research issues is a future direction of research. In practice, one significant trend is the establishment and the growing of a new Fintech standard: Tokenomics.

6.9.1. Artificial Intelligence in Finance

The dramatic evolution of Artificial Intelligence (AI) technologies is dramatically affecting the finance industry with significant transformations in consumer-facing and enterprise-level processes. AI-powered technology can develop various acoustic, sensation, cognitive, and emotion capabilities that are very near to human capabilities and thus hijack traditional finance processes involving human workers. Additionally, these AI capabilities are continually improving by leveraging one or more tools which make them more affordable for most enterprises. AI technologies are widely used in different areas of finance processes including credit card cybersecurity, loan allocation, automated trading robots, the digital anonymization of sensitive financial data, the customer-centric chatbots, diversified asset portfolio management, reduced human intervention automation, insurance risk detection, fraud analytics, accounting processes automation, robo-advisors, insurance scoring services, loan risk assessment, cross-border payment services, human resources and finance departments automation, and many others. AI-powered automated systems are employed for the services that could be standardized and thus needs lower cost under pressure. However, many traditional finance processes still require human attention and these processes could be optimized but not made entirely automated in the short term.

Sophisticated customer service chatbots can significantly improve customer experiences creating human-like conversations powered by natural language processing and augmented data generation capabilities. Chatbot services incorporate the critical area of AI innovations where voice, text, and image-based access to sensitive financial data and services is increasingly implemented. Innovations in automation are still speedy in accounting, insurance, risk management, regulatory compliance, and loan assessment areas with a major optimization potential. AI job displacement will accelerate in the finance departments around the world in the near-term because of predictions showing that its cost is dropping and the industry is still returning to the pre-COVID levels of expense structure.

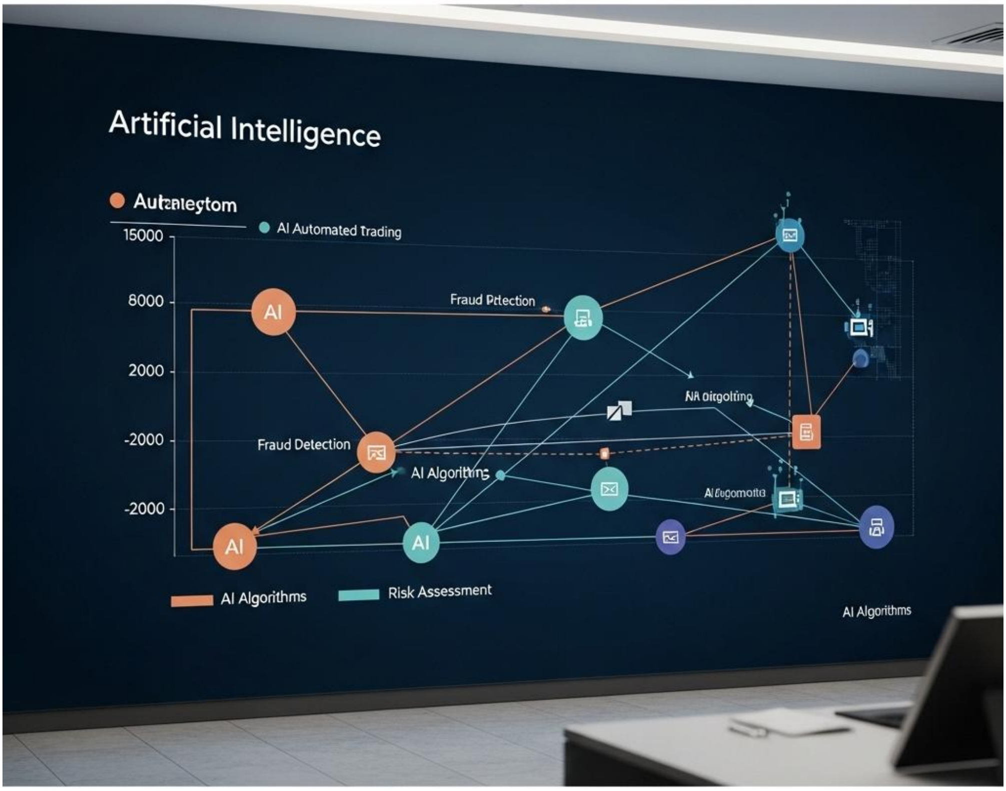

Fig : Artificial Intelligence in Finance

6.9.2. Blockchain and Big Data

The intersection of blockchain and Big Data possesses the potential to bring revolutionary improvement in current data management systems. Blockchain's open-access yet verifiable systems can address the growing privacy concerns over the use of centralized personal data, countering the threat towards the business model of technology giants who basically create Big Data monopolies by accumulating masses of user data. In organizing and storing massive amounts of data across networks securely and efficiently, blockchain serves as a new infrastructure for utilizing Big Data and enriches the financial services whether to process Big Data for businesses or to provide a Big Data-powered smoothly connected digital ecosystem. What blockchain effectively does is construct a decentralized business ecosystem in which any businesses who are trusted partners of each other can work together for mutual benefit by sharing parts of their respective Big Data to improve the overall business efficiency while keeping the rest of their Big Data proprietary due to the native structure of blockchain. Provided that its data privacy has room for improvement, blockchain has a great prospect to be adopted in multiple scenarios such as the finance service, cross-organization, and administrative services such as taxation whereby efficient collaboration across organizations is key to improving overall efficiency. Banks can enhance their credit risk decision-making process by sharing sensitive data of clients via a borrowing credit chain protocol executed on private blockchain. Real-time monitoring of companies' liquidity would become true if banks join effort by embracing a corporate loan blockchain. On the other hand, companies also benefit from the move.

6.9.3. The Evolution of Financial Technologies

The new development of financial technologies has surpassed recent years. The possible confluence of AI systems with big data applications can lead the way to a real evolution of existing service offers. Indeed, a deeper integration of both artificial intelligence and big data analytics technologies is already restructuring many of the most traditional finance functions worldwide. There is a need to understand possible evolutions of the finance industry, both in a local and in a global perspective, and the role that fintech can play. To better deepen the analysis, we would comment on microsegmentation and digital credit scoring, blockchain services encompassment, the turnover of initial global coin offering offerings, and shared digital finance platforms. In terms of emerging technological trends, the obvious sophistication of information and communication technologies is one of the most considered factors. There are other factors that promote disruption in common segments of consumer credit, but that reflect key aspects of the overall fintech ecosystem as well. These two factors cannot be separated. More sophisticated, yet more affordable, technologies foster new players with deep resources that can innovate on common service flows in retail and corporate markets. Shared digital finance platforms capable of integrating multiple financial services will slowly turn into full serviced global players.

There is consensus on the idea that emerging financial technologies will likely dictate the financial ecosystem revolution. Recent years' performance by incumbent players has shown that their business model is sufficiently stable to withstand relatively small amounts of innovative threats. In this view, the deepest riding wave of fintech will result in shared, vertically specialized platforms, integrating different types of services, rather than monopoly models based on super platforms. These systems will probably allow niche service specialists to share common points of customer annoyance or weakness. The balance between progressively shared and monopolistic solutions also seems to indicate that issues stemming from institutional setup will also be at stake in determining digital finance organizational structures.

6.10. Conclusion

Despite the increasing popularity of applications from both academia and the industry, there are still several hurdles that need to be overcome before widespread adoption of big data in the finance industry can become a reality. For instance, electronic health records can provide hospitals, clinics, payers, and even patients a complete view of the patient's health history in near real-time. Data ownership of medical financial data such as health insurance claims may still reside with insurers limiting ability of providers and even patients to make use of big data applications that rely on insurance claims to provide analysis of benefit design, coding accuracy, or quality improvement work. Such issues are not just limited to the healthcare industry. New rules that protect the consumer of a financial product from exploitation could limit the applicability of big data in finance. Anti-violation of privacy laws will also make banks and lenders hesitant to act aggressively on these opportunities. Indeed, financial institutions have been slow to embrace big data applications in vulnerable markets. Banks and lenders have found ways to rely on traditional methods to reach borrowers targeted by third-party big data vendors. With the potential risk and rewards, the financial services industry is forced to adapt. Risk departments and compliance functions are sitting with boards eager to evaluate big data applications and mitigate both financial and reputational consequences. Start-ups are beginning to emerge all over the United States and abroad to know activity in this space. Incubators are beginning to see success. Banks are embracing partnerships in hopes to test the capabilities of using big data applications and wait for the regulatory environment to catch up.

References

J. Udeh et al., "Big Data in Detecting and Preventing Financial Fraud in Digital Transactions," IJFDR, 2024.

S. Kumar, "Big Data Analytics in Financial Fraud Detection," ResearchGate, 2023.

H. Jiang, "Big Data Analytics in Anti-Money Laundering and Financial Fraud Detection," OJApplSci, 2024.

P. Rao, "Leveraging Big Data for Predictive Financial Analysis," SSRN, 2024.

A. Bansal et al., "Big Data-Driven Fraud Detection Using Real-Time Stream Processing," arXiv:2505.07841, 2025.