A Big Data Analytics Framework to Improve Financial Inclusion's Usability

Hoda Ismaeel Ibarhim[1] Ahmed Ibrahim Bahgat ElSeddawy[2]
Ahmed Gamal Elish[3]

Abstract

The significant portion of the population that lacks access to formal banking services represents a serious challenge to most developing countries, especially in achieving both social and economic development goals. For that, enhancing financial inclusion has become a critical objective for fostering equitable economic growth. By facilitating access to financial systems for all individuals, this will ensure financial inclusion help to reduce income disparity, particularly between affluent and economically disadvantaged groups, as well as boosting capital flows into and between the different economy sectors. This paper explores the essential role of financial inclusion in driving economic growth and examines the various variables that should be considered more by financial institutions in prediction of targeted different segments attraction factors successfully. And, we examined the accuracy of chosen algorithms of prediction to optimize the decisions relating to achieving better benefits for all related parts in the economy, ranging from government agencies to financial institutions and individuals.

Keywords: Big Data Analytics (BDA), Digital Banking, Financial Inclusion, Financial Services, Insurance, Decision Tree algorithm, Random Forest algorithm.

[1] Business Information Systems Department, Faculty of commerce and business, Helwan University. Cairo, Egypt.

[2] Information Systems Department, Arab Academy for Science, Technology and Maritime Transport (AAST). Cairo, Egypt.

[3] Faculty Of Commerce & Business Administration Department of Economics. Helwan University, Helwan University. Cairo, Egypt

إطار عمل لتحليلات البيانات الضخمة لتحسين قابلية استخدام الشمول المالي

الملخص

تمثل النسبة الكبيرة من السكان الذين يفتقرون إلى القدرة على الوصول إلى الخدمات المصرفية الرسمية تحديًا خطيرًا لمعظم البلدان النامية، وخاصة فيما يتعلق بتحقيق أهداف التنمية الاجتماعية والاقتصادية. ولهذا السبب، أصبح تعزيز الشمول المالي هدفًا بالغ الأهمية لتعزيز النمو الاقتصادي العادل. ومن خلال تسهيل الوصول إلى الأنظمة المالية لجميع الأفراد، يمكن التأكيد على أن يساعد الشمول المالي في الحد من التفاوت في الدخل، وخاصة بين الفئات الثرية والمحرومة اقتصاديًا، فضلاً عن تعزيز تدفقات رأس المال إلى القطاعات الاقتصادية المختلفة وفيما بينها. ويستهدف هذا البحث استكشاف واستعراض الدور الأساسي للشمول المالي في دفع النمو الاقتصادي، وأيضاً استكشاف المتغيرات المختلفة بالتنبؤ الناجح بعوامل الجذب المختلفة للفئات المستهدفة؛ والتي يجب على المؤسسات المالية مراعاتها بشكل أكبر. كما تم في هذا البحث فحص دقة خوارزميات التنبؤ المختارة، وهذا بهدف تحسين القرارات المتعلقة بتحقيق فوائد أفضل لجميع الأطراف الاقتصادية ذات الصلة، بدءًا من الوكالات الحكومية ومروراً بالمؤسسات المالية وحتى مختلف الأفراد.

الكلمات المفتاحية: تحليلات البيانات الضخمة، الخدمات المصرفية الرقمية، الشمول المالي، الخدمات المالية، التأمين، خوارزميات شجرة القرارات (DT) والغابة العشوائية (RF).

1. Introduction

Financial services all over the world have been transformed and evaluated by big data analytics (BDA). Businesses are being able to leverage data on a larger scale more easily, quickly, and reliably as a result of advancements in computing capacity and data analytics. By utilizing big data analytics efficiently, financial organizations and new entrants from other industries can provide a greater number of and superior financial services. Governments are also a major part of the economy that get benefits of (BDA) looking into methods to more systematically utilize big data produced by the banking industry to gain a better understanding of the overall picture of how financial system operates in the economy. Despite all the advantages mentioned, the increased usage of big data has created a concerned issues about customer privacy, security, discrimination, data veracity, and competitiveness. As a result, policymakers have begun to regulate and supervise financial organizations' use of big data, as well as consider how to utilize big data. [1]

There is no doubt that there are advantages and challenges linked to the use of big data (BD) and artificial intelligence (AI) in promoting financial inclusion. The benefits of artificial intelligence and big data for financial inclusion, including (1) the production of credit scores for unbanked individuals based on alternative information, (2) the availability of intelligent financial products and services for individuals with bank accounts, (3) the streamlining of the account opening procedure for those without bank accounts, and (4) the enhanced operational effectiveness and risk mitigation for financial service providers. [2]

On the other hand, a number of challenges related to artificial intelligence and big data in financial sector generally and for financial inclusion specifically appeared to require more

attention, including (1) the scarcity of proficient AI professionals who owns efficient knowledge in both AI and business, (2) the increasing rates of unemployment within the financial sector (caused mainly by replacement of AI over individuals), (3) the inherent bias in the architecture of artificial intelligence systems, and other obstacles arising from stringent data privacy regulations. [2]

An analysis of the benefits and consequences of using (AI) in relation to financial inclusion requires a clear understanding of three fundamental concepts: big data, AI, and financial inclusion. Big data is the term used to describe extremely large and complex datasets that provide challenges for conventional data processing systems to effectively manage the necessary calculations for interpretation. Big data is characterized by its large volume, fast generation rate, and extensive range of origins, whereas artificial intelligence (AI) is the replication of intelligent human behaviour by machines. In layman's terms, artificial intelligence is the intelligence that is demonstrated by computers. The replication of human intellect in computers that are programmed to think and act in the same manner as humans is sometimes referred to as artificial intelligence. [3]

2. Big Data Analytics and Its Significance in the Financial Sector

Due to the revolution in technology and globalizing business operations, a lot has changed, and the huge flow of information is no longer something to ignore. In the digital age, Big Data has revolutionized the financial sector and banking industry significantly by driving data-driven decision-making. Both corporate and individual investors now is relying on wide and various data trends analysis and forecasting methodologies while making financial decisions. Efficient working capital management, risk assessment, and fraud detection are just a few of the areas enhanced by Big Data [4]

For example, banks nowadays, as a leading doers in the financial sector, are heavily investing in Big Data analytics to survive in the competitive environment. For instance, banks use text mining to extract valuable insights from emails, reports, and official documents, allowing them to monitor patterns and classify data for qualitative or quantitative analysis. The data-driven approach has empowered the banking industry to optimize operations, manage risks, and better understand customer behavior [5].

At present, every corporation or individual, when engaging in financial sector investments, they make decisions driven by data analysis. Similarly, the business sector employs different strategies to make efficient working capital management and to improve financial performance by optimizing the use of existing cash. For such importance of financial decisions, Banks as the predominant force in the financial sector are allocating substantial resources to Big Data analytics. Therefore, analytics of the vast amount of data generated by the banking industry, including customer data, financial products data, and transaction data, have become essential for survival in the current competitive environment. Thus, for this survival goal, the financial industry now employs apps that encompass social media, internet, risk management, security intelligence, and even counterfeit detection or fraud risk assessment. [6]

Based on the above, text mining considered to be one of the most often utilized methods for retrieving information and data inside these sectors. The primary objective of data mining through emails, reports, official documents and different websites and applications is to extract diverse information for usage in various decision-making processes, in presence of strong monitoring and analysis patterns. Mainly, the data is predominantly structured to facilitate qualitative or quantitative analysis. Use of financial data obtained and extracted from

these resources enables different institutions to classify them according to the specific requirements of the business challenge for thorough analysis.

The primary objective can be represented in seeking different economic units to comprehend the patterns in a way that helps optimize decision making. A comprehensive analysis of the application of Big Data in banking sector can be glanced from the research undertaken by (Enas Al-lozi, and others, 2022), illustrated in figure (1). [7]

As figure (1) shows, the highest use of big data in the banking sector goes for introducing new products for traditionally unprofitable customers (which comes in line with the goals of financial inclusion). The usage that comes in the 2nd place was digital experience with human support comes, and the 3rd place is shared by (brokerage service, investment decision support) and (crowdsourcing, investment decision). The lowest percentage according to the study was 3% for custom investment solutions and 5% for asset allocation automation.

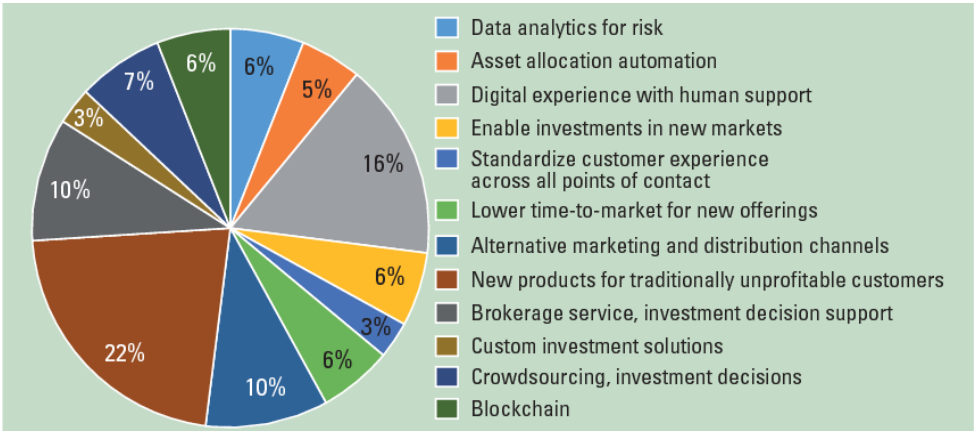

Fig. 1. Big Data Usage in Banking Sector [7]

Last but not least, we can conclude that applications of Big Data analysis yielded financial advantages in terms of reducing costs or enhancing productivity. Simultaneously, it also

expands the possibilities for new commercial prospects. Future company prospects are contingent upon judgments that rely on the analysis of Big Data and are undoubtedly more efficient. The use of Big Data Analytics involves the processing of existing data using sophisticated techniques to evaluate the vast amount of structured, unstructured, and semi-structured data collected from sources with comparable characteristics. [8]

3. Overview of Big Data and Big Data Analytics

For a comprehensive understanding of Big Data and Big Data analytics, it is advisable to first grasp the concept of Big Data. In 1997, NASA scientists coined the phrase "Big Data" and provided the initial scientific explanation of this concept. The concept of data volume gave rise to the need for computer systems to possess substantial memory capacity for storing Big Data. This task necessitates more resources such as external storage devices and internal storage devices like discs and memory cards. Consequently, this issue was regarded significant [8].

Following a four-year period of research, analyst Goug Laney of the META group broadened the concept by including issues associated with data growth and the corresponding opportunities. This was consolidated into three categories: volume, velocity, and diversity, which were then increased to five [9] . Without any doubt, the abundance and speed at which Big Data is generated offers efficient and novel forms of easily accessible information that ensures enhanced comprehension, efficient, optimized and accurate decision-making and processing operations [8]. The application of Big Data analytics facilitates the organization of unstructured data. Big Data encompasses a broad spectrum of ideas, including gathering, aggregation, conversion, and analysis of data to get insights on human behavioral patterns. [10]

The primary obstacle that organizations are now encountering in achieving real-time decision making is the development of patterns derived from previous data and the ability to obtain real-time corresponding data. The objective of this activity is to facilitate the acquisition of information and enhance decision-making processes by utilizing Big Data analytics. Over the past twenty years, Big Data analytics has become a dominant technology paradigm. Conceptual framework of the large data volume chain.

And to enhance comprehension of the concept of big data analytics and its benefits, Table (1) provides a clarifying overview of the terminology employed by various writers, together with its associated advantages.

Table (1) definition of big data analysis [7]

Authors	Definitions
(Abkenar, Kashanib, Mahdipour,& Jameii, 2021)	Big data analytics are techniques and frameworks categorized in contextual and network approaches that help firms to grow.
(Akoka, Comyn-Wattiau, & Laoufi, 2017)	Big data is massive data with large structures that identify the data flow while relying on data flow using IT functions to improve the value of firm.
(Al-Sai, Abdullah, & Husin, 2020)	Big data is an asset that has attracted the attention of the top management to gain foster insight and high revenue.
(Asad, Altaf, Israr, & Khan, 2020)	Big data analytics is a term that can help businesses regardless of size, if appropriately utilized for decision making and drawing inferences by appropriate techniques.
(Bag, Wood, Xu, Dhamija, & Kayikci, 2020)	Five dimensions of big data namely, volume, veracity, variety, velocity, and value that help in gaining competitive advantage.

Authors	Definitions
(Dong & Yang, 2020)	Big data analytics having marketing capabilities drives the organizations towards market-oriented innovations using social media trends.
(Favaretto, Clercq, Schneble, & Elger, 2020)	Big data analytics describe a wide range of concepts from collecting and aggregating huge data sets to an understandable usable manner for identifying human behavior.
(Mauro, Greco, & Grimaldi, 2016)	Big data analytics are characterized by high volume, velocity, and variety used with specific technology to retrieve specific information.
(Pappas, Mikalef, Giannakos, Krogstie, & Lekakos, 2018)	Big data analytics has the capability to portrait business analytics eco-system towards digital transformation and sustainability in improving firm performance.

Definitions of Big Data analysis in the table focuses on BD as the process of analyzing vast and intricate datasets in order to find patterns, trends, and insights. While table (2) most likely summarizes the essential components of big data, including the amount, velocity, and diversity of data as well as the applications of machine learning, data mining, and predictive analytics to the analysis of financial data. This definition is important in understanding how Big Data is utilized in sectors like finance and banking to drive decision-making. [11]

Table 2: Overview of Big Data Analytics and its Significance in Financial Sector [12] [13] [14]

Term	Definition	Applications in Finance	Benefits
Big Data Analytics	The process of analyzing vast and intricate datasets to discover patterns, trends, and insights.	Used in financial data analysis, customer analytics, risk management, fraud detection, and predictive analytics.	Helps with decision-making, optimizing operations, managing risks, and enhancing customer experience in banking and finance.
Volume	Refers to the sheer amount of data generated, often measured in terabytes or petabytes.	Analysis of large datasets from transactions, social media, customer behavior, etc.	Enables comprehensive data analysis for insights that were previously inaccessible due to data volume constraints.
Velocity	The speed at which data is generated, processed, and analyzed in real time.	Real-time monitoring of transactions, fraud detection, and personalized marketing based on customer activity.	Improves response times to market changes, fraud alerts, and customer needs.
Variety	The different types of data (structured, semi-structured, and unstructured) generated from	Combines data from emails, reports, social media, and customer transactions to offer a full view of financial	Provides a holistic view of customer behavior, operational

Term	Definition	Applications in Finance	Benefits
	various sources.	operations.	efficiency, and market trends.
Machine Learning	A subset of AI that uses algorithms to identify patterns and make predictions based on large datasets.	Used in predictive analytics for investment strategies, customer segmentation, and fraud detection.	Enhance predictive accuracy, improves customer targeting, and reduces fraud by identifying hidden patterns.
Data Mining	The process of extracting useful information from large datasets by identifying patterns and correlations.	Commonly used in financial decision-making by analyzing historical data, trends, and financial transactions.	Help in optimizing investments, reducing risks, and improving capital management.

Adding to the above, the three modifications documented in figure (2) have been employed by several writers in their respective research endeavors. Over the previous five years, the writers have examined and analyzed five distinct categories of Big Data. The two other fundamental aspects of Big Data are veracity and value. Veracity refers to the level of trust or uncertainty associated with the Big Data, which is inherent in the judgements made throughout its analysis.

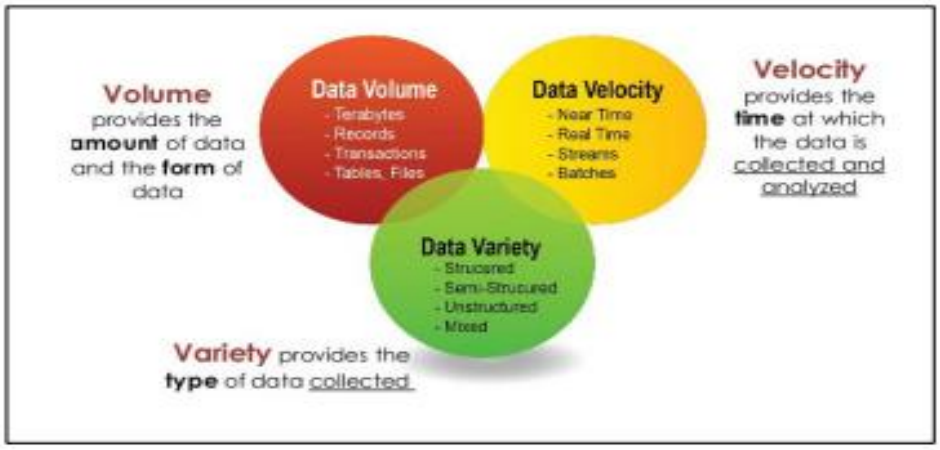

Fig. 2. The Common Characteristic of Big Data Analytics [11]

Conversely, value refers to the insights obtained by the company using Big Data. Acquiring value requires the implementation of efficient operational processes, improved tactics, and scalability. Therefore, in the present day, the widely acknowledged attributes of Big Data are volume, velocity, variety, veracity, and value, often referred to as the 5 V's. The five Vs of Big Data are the fundamental characteristics that distinguish traditional data management from modern data analytics principles. The current standard practice is to utilize a concise image to symbolize the 5 V's, as seen in Figure (3). The Figure illustrates the ultimate 5 key V's that corporations are now using to evaluate Big Data Analytics. [15]

Fig. 3. Five V's of Big Data [15]

4. Related work

In his work, S. H. Oluwasola investigated methods to enhance the effectiveness of the banking system using behavior-driven design (BD). The analysis encompassed consumer happiness by identifying the primary client of the bank through the analysis link, enhancing customer feedback, detecting customer attrition potential, and improving security measures. Study revealed that the banking sector may enhance the efficiency of the banking system by successfully implementing all analytical techniques and algorithms, including feeling, aggregation, correlation, survival, decision tree, analysis, and data file. [16]

Ravi and Kamarudin conducted a research that discussed analytically the financial services sector, specifically examining the banking industry, emphasizing certain business challenges that have been successfully addressed using quantitative data analysis. Particularly, the banking sector has transitioned from the traditional paradigm of journal and ledger entry to a data and analytics-driven banking operations that encompass both online and offline client behavior. They revealed that the

services of financial sector is rapidly transitioning from the conventional paradigm to the advanced digital methods of consumer interaction. This phenomena of digital transformation is affecting both components of the financial industry services, namely the **financial service provider** and the **client**. [17]

Furthermore, it addressed cutting-edge technology that, when coupled with extensive data analytics, have greatly transformed the quality and efficiency of the banking sector. Prognostic and directional analyses have been extensively investigated in the fields of the stock market and insurance. The study examines many scenarios within the domains of baking, financial services, and insurance (BFSI), wherein the significance of big data analytics is increasingly realised. Thus, the study highlights the potential benefits of contemporary technologies such Internet of Things (IoT), Blockchain, Chatbots, and robots. The study also identified some obstacles that impede the complete adoption of Data Driven Analytics in the financial services sector to improve business value. [17]

A study by A. A. Alzaidi, clarified that BD analysis is being applied in several areas of the banking industry, enabling enterprises to deliver enhanced services to their clients, both internal and external, while also enhancing their performance. This paper also shed a light on the influence and benefits of big data analytics on the banking industry by examining transactional and sentiment analysis specifically for the banking sector. Further methods exist for banks and other financial institutions to acquire information pertaining to their clients for the purpose of analyzing emotions. These methods include utilizing social media platforms as channels for marketing research. [6]

The study conducted by Anshari, Et. al., demonstrated the correlation between Customer Relationship Management

(CRM) and Business Development (BD). While CRM focuses on comprehending human behavior and interests, BD is expected to enhance customer relationships by enabling multi-directional communication and the ability to personalize customers and clints experiences. Business discovery using (BD) improves CRM strategies by gaining a deeper understanding of customer habits and behaviors, enabling businesses to provide more personalized and tailored CRM solutions for each customer. Ultimately, CRM will enhance the personalization of large data through advanced tools and improved strategies, as it comprehends the well-defined target audiences and the desired messages to be delivered. The research also demonstrated the significance of implementing business development in banks, particularly in the formulation of marketing strategies. [18]

On the study made by Brown, Wilson and Johnson, they investigate the diverse functions of big data in enhancing customer experiences, with an emphasis on its effects on personalization, customer service, operational efficiency, and strategic decision-making. By evaluating recent developments and implementations, the study underscores how big data analytics facilitates highly tailored recommendations, focused marketing initiatives, and adaptive content that greatly boost customer engagement. Furthermore, it explores advancements in customer service through immediate support, sentiment analysis, and predictive assistance, all of which contribute to a more agile and customer-oriented strategy. [19]

The study further explores how big data analytics enhances operational efficiency by optimizing workflows, improving resource distribution, and lowering expenses, thereby boosting overall organizational performance. Strategic decision-making is enriched by data-driven insights into market dynamics, pricing approaches, and competitive positioning,

enabling organizations to adapt and succeed in a competitive landscape. Nonetheless, the research also recognizes the challenges related to data quality, privacy, and investment, stressing the necessity for strong measures to tackle these concerns. [19]

Additionally, the convergence of big data with emerging technologies such as artificial intelligence, the Internet of Things, and blockchain introduces new opportunities and complexities in improving customer experience. In summary, this research highlights the transformative capabilities of big data analytics while emphasizing the critical need to address its associated challenges to fully leverage its advantages. [19]

Finally, according to the scholars Indriasari, Elisa & Soeparno, Haryono & Gaol Ford, the prevalent concerns in modern times are the problems associated with automated systems and intelligent tracking of transactions and banking customers. Conversely, the majority of transactions that have already transitioned to digital and online platforms require more advanced capabilities in the Predictive Analyst. The rise of predictive analytics (PA) is one of the biggest disruptions in financial services. [20]

According to the authores, PA methods in the last years grew sharply to give the baseline of data driven decision-making process for forecasting future business situation. And Because PA uses various algorithms to discover different patterns in the big data environment that might create more value for businesses including in financial institutions, the prospective application of data science in the financial sector remains to be investigated. In their paper, they present a concise overview of the research conducted on Predictive Analytics within the financial sector. Furthermore, they delve into the prospective implementation of Predictive Analytics and

examine the future utilization of data science and big data in financial institutions. [20]

5. Research Objectives

The main objective of the research is to assess the government's efforts in financial inclusion using its policy instruments, considering the substantial development and contribution achieved in the last three years, as well as the evident results accomplished. The significance of the study arises from the extensive impact that financial inclusion might have on Egypt's economic development, as well as the diverse range of intended recipients. For that, the study aims to conduct an exploratory assessment of the present condition of financial inclusion in Egypt, identify the barriers and challenges faced, and evaluate the recent advancements made towards establishing an inclusive financial system.

6. The Analysis and The Model

This study proposes a strategy to attract investors who own small businesses to involve their operations and transactions into commercial banks, in order maximize the benefits of investment decisions in the Egyptian economy within the framework of financial inclusion targeted by the government. So, government efforts using both fiscal and monetary policies, represented clearly for example in (interest rate policies, offering tax exemptions and taxation system reforms, optimizing private financial institution’s role, etc.) will meet its goals.

The proposed model will be built on financial data, and it aims to analysis all data and extract the relations between all companies and banks which working in this field. After finishing the analysis, we will predict the future financial performance according to organizations behavior and recent initiatives to enhance financial inclusion that have been implemented through government policies. The most important

benefit we seek from the model is to detect and monitor the future behavior of the financial market and companies, to find the effective part of them using present data, so we can enhance making decisions regarding financial inclusion.

It will be useful to shed the light on the financial inclusion in Egypt from different prospectives, before starting to use the model and to analyze the data.

6.2 Exploratory Analysis of Financial Inclusion in Egypt: Assessing the Role of Key Institutions in Supporting Households, SMEs, Women, and Entrepreneurs

There is no doubt that financial inclusion requires a combined efforts from different economic units, especially (CBE, FRA, and the Ministry of Finance)[4], in advancing financial inclusion for certain groups of recipients including households, SMEs, women, and entrepreneurs. The main reason for financial inclusion is basically targeting the excluded groups from banking and unified financial services. The differences occur in accessing financial services between targeted units of financial inclusion worries different governments for many reasons, on top of them to monitor effectively and efficiently the monetary

[4] **CBE**: **Central Bank of Egypt** – Is the primary regulatory authority overseeing Egypt's monetary policy, banking system, and financial stability. It manages currency, controls inflation, and regulates the country's banking sector.

FRA: **Financial Regulatory Authority** – The FRA is responsible for supervising and regulating non-banking financial markets and instruments in Egypt, such as capital markets, insurance, mortgage finance, and leasing.

Ministry of Finance: This government institution is responsible for the country's fiscal policy, including the management of public finances, budgeting, taxation, and overseeing economic reforms and financial legislation.

flow within the economy[5], which will achieve a lot of strategic, social and economic objectives for the government.

To know which categories are of targeted financial inclusion, we can illustrate the difference between access to financial services and the usages, as it shown in figure No (4). [21]

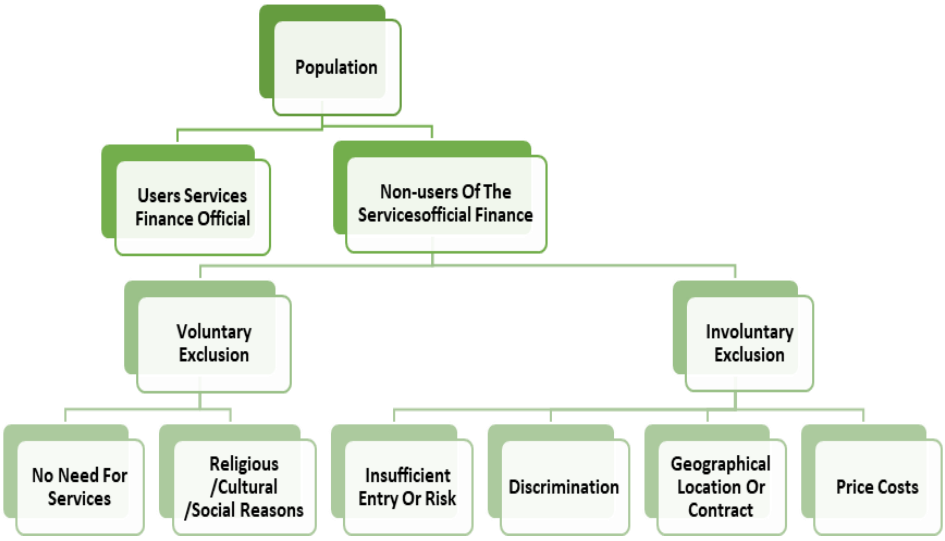

Fig. (4) Users of financial services

As Figure (4) shows that the population is divided into two parts: 1) Users of official financial services, and 2) non-users of formal financial services, which also divided into two groups:

I. **Voluntarily excluded**: Individuals who do not use financial services because:

[5] Economists depend on explaining how the economy works through what they called **Circular Flow of Income**, which divide the economic transactions into 2 types, **1). Physical flows** "represented in the move of factors of production and produced goods and services between different economic sectors", and **2). monetary flows** the shows how income is created by these transactions to different sectors.

- They do not need the services provided.
- The religious, cultural, or social beliefs prevent them from using formal financial services.

II. **Involuntarily excluded** Individuals who are unable to access financial services due to:

- Insufficient money to save or invest.
- High perceived risks.
- Discrimination based on various factors.

- Legal requirements, such as difficulties in providing necessary documentation.
- Geographical barriers, limiting access to financial institutions.
- High costs associated with opening and maintaining accounts.

6.3 The evolution of the Egyptian banking sector, and its efforts in financial inclusion.

The Egyptian banking sector showed outstanding resilience in the face of the economic and financial challenges that the world and Egypt witnessed in light of the spread of the Corona pandemic, and it was one of the most important pillars on which the Egyptian economy relied to face the repercussions of the pandemic and get out of it in record time.

The Egyptian banking sector has achieved significant growth during the past three years, as it has become the third among the Arab banking sectors in terms of the volume of assets, and the first among the banking sectors of Arab non-oil countries. It should also be noted that while the assets of the Egyptian banking sector constituted 8.1% of the total assets of the Arab banking sector in 2017, this ratio gradually increased to 8.8% in 2018, 9.8% in 2019, and 11.2% in the year 2020. , to reach 13.1% at the end of 2021. [22]

As for the assets of the Egyptian banking sector, it amounted to about 8,627 billion pounds (about 542.3 billion dollars) at the end of the year 2021, achieving an increase of 22.9% during the mentioned year, compared to 19.9% in the year 2020 and 7.8% in the year 2019, the combined assets of banking banks would have increased by more than 50% in just three years. [22]

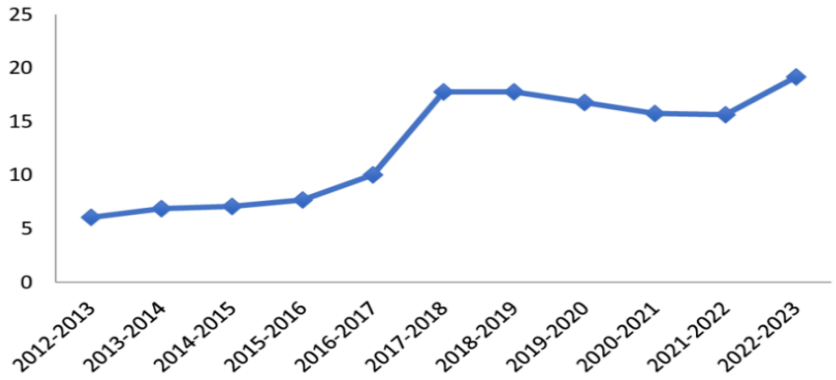

Fig. (5): Growth rate of volume of assets on Egyptian banking sector. Made by the researchers using CBE data [22]

Also deposits were amounted to about **6,433 billion EGP** at the end of 2021, achieving an increase of 24.2% compared to an increase of 22.6% in 2020, and by 11% in 2019. Thus, the total deposits of Egyptian banks have increased by 57.9 % during the period 2019-2021. As for lending and discount balances to customers, it reached about **3,099 billion EGP** at the end of 2021, achieving an increase of 24.3% during the mentioned year, compared to 31.9% in 2020 and 4.2% in 2019. Accordingly , Egyptian banks have increased the volume of credit provided to customers by 60.4% within three years, which indicates their quest for a significant expansion in financing the national economy. [22]

Finally, according to CBE annual report (Fiscal Year 2021-2022), the growth in domestic credit can be shown in figure (6), and was primarily attributed to [22] :

- Net credit extended to the government increased by LE 905.4 billion or 25.7 percent. Such an increase reflected the surge in banks' holdings of government securities by LE 900.9 billion and credit facilities by LE 491.0 billion, which was mitigated by the rise in government deposits by LE 486.5 billion.
- Credit to the private business sector went up by LE 287.2 billion or 24.0 percent.
- Credit to the household sector moved up by LE 138.7 billion or 24.9 percent.
- Credit to the public business sector scaled up by LE 5.9 billion or 4.0 percent.

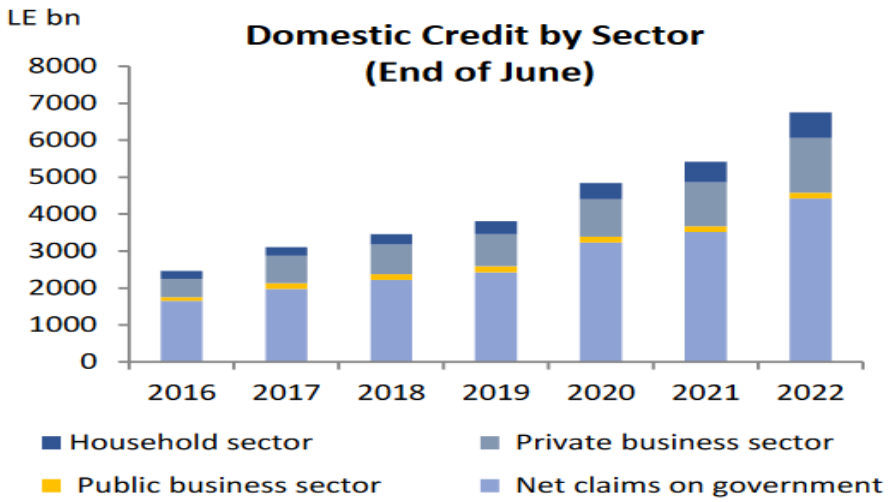

Fig. (6). Distribution of Domestic Credit By Sector [22]

Addionally, table (3) shows some of the tasks carried out during the period from the end of June 2022 to end of June 2023.

Item	Accomplisment by the end of June 2022	Accomplisment by the end of June 2023	Change %
No. of branches added to the registry of banks	Adding (19) new branches for (8) banks	Adding (21) new branches for (9) banks	-
No. of debit cards	22,959,816 cards	24,427,638 cards	6.4%
No. of prepaid cards	28,274,864 cards	30,311,634 cards	7.2%
No. of credit cards	4,810,227 cards	5,249,225 cards	9.1%
No. of ATMs	21,459 machines	22,708 machines	5.8%
No. of P.O.S2 machines	188,429 machines	215,923 machines	14.6%
No. of internet banking subscribers	13,247,557 subscribers	14,656,241 subscribers	10.6%
No. of mobile banking subscribers	12,120,034 subscribers	13,795,889 subscribers	13.8%
Requests related to new electronic banking services	(11) banks obtained an approval for providing (21) new electronic banking services.	(15) banks obtained an approval for providing (81) new electronic banking services.	-

Table. (3). Tasks carried out by Egyptian Banks 2022-2023. Made by the researchers using CBE data [22]

And according to the following figure (7), it presents financial metrics and the economic indicators related to the economic activity of banking sector in Egypt, as reported by the Union of Arab Banks. It involve measures such as credit distribution, financial inclusion, or growth in banking activities based on data from the Central Bank of Egypt. Data of the distribution of credit facilities provided by Egyptian banks to clients indicates that the government sector acquired 38% of these facilities at the end of 2021, followed by the family sector by 20%, the industry sector by 18%, and the services sector by 16% The trade sector increased by 7%, and finally the agriculture sector by almost 1%.

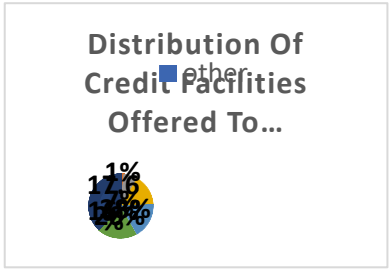

Fig. (7). Distribution of Credit Facilities Offered to Customers from Egyptian banks. Made by the researchers using CBE data [22]

Figure (7) displays the global distribution of financial capacities, potentially showing how financial resources, investments, or banking services are spread across different regions of the world. It illustrates disparities between countries in terms of available capital or financial infrastructure. Such a figure helps to understand global financial inequality or the concentration of financial power in specific regions.

6.4 Systematic Review Methodology

The purpose of the current investigation was to conduct a comprehensive examination of the implementation of big data analytics within the banking sector. Consequently, in order to gain a more comprehensive understanding of the concept of big data analytics and its primary characteristics, literature has been investigated in its entirety. The concept of big data analytics was initially developed from the viewpoint of numerous academicians. Additionally, the domain of finance has been the subject of an examination of the application of big data analytics for trend analysis and forecasting. Ultimately, the exhaustive application of big data analytics in finance across the five aforementioned categories has been demonstrated through the implementation of big data analytics in a specific domain.

6.5 Regression

Regression analysis is a statistical technique used to examine the relationship between one or more independent variables, sometimes referred to as predictor variables, and a dependent variable, also called the response variable. Using the values of the independent variables as a guide, the goal of this approach is to forecast the value of the dependent variable. Regression analysis, to put it simply, explains how changes in the values of one variable affect the values of other variables. The ability to model and quantify the link between variables makes it an

effective tool in a variety of sectors, including social sciences, finance, economics, and medicine. [23]

6.5.1 CLASSIFICATION

Classification is the process of learning a function that can divide data items into subsets of a given class set. In data mining, classification is regarded as one of the most well-known challenges that researchers face. Finding a decent generic that can reliably predict the class of unknown data objects for each class is the first step in a number of classification tasks. The methods that were utilized in this study to get the desired results: Random Forest (RF), Decision Tree (DT) and Pairplot for RF tree diagram. [14]

6.5.2 RANDOM FOREST

In general, **Random Forest (RF)** algorithm, which is a machine learning model is used for classification and regression. In this context, the tree diagram shows how the model splits data into different branches or "trees" based on various features or attributes, eventually leading to a decision or prediction. The RF algorithm is useful in financial analysis for making predictions, such as credit scoring or investment decisions, by analyzing multiple factors. Random forest is a tool for help that produces an outcome that is thought to have been combined from a set of decision trees. It performs superbly in a range of real-world prediction problems & scenarios, including as the medical field. The reasons are as follows: it is not impacted by noise in the data set, it is not overfitted in any kind, it is built by combining the predictions of numerous trees, and it operates rapidly. Numerous additional tree-based algorithms, such decision trees, typically exhibit significant increases in performance. [24]

Using data of Egyptian banking sector, we used RF algorithm to summarize the data from high level features to low level features which will help achieving the suitable classification of the data to be analyzed accurately. Figure (8) show the tree diagram that shows the probability distribution of RF which shows the analysis of the tree based on statistical descriptive (correlation, std, mean), these leaves from master node represented the data classified from analysis.

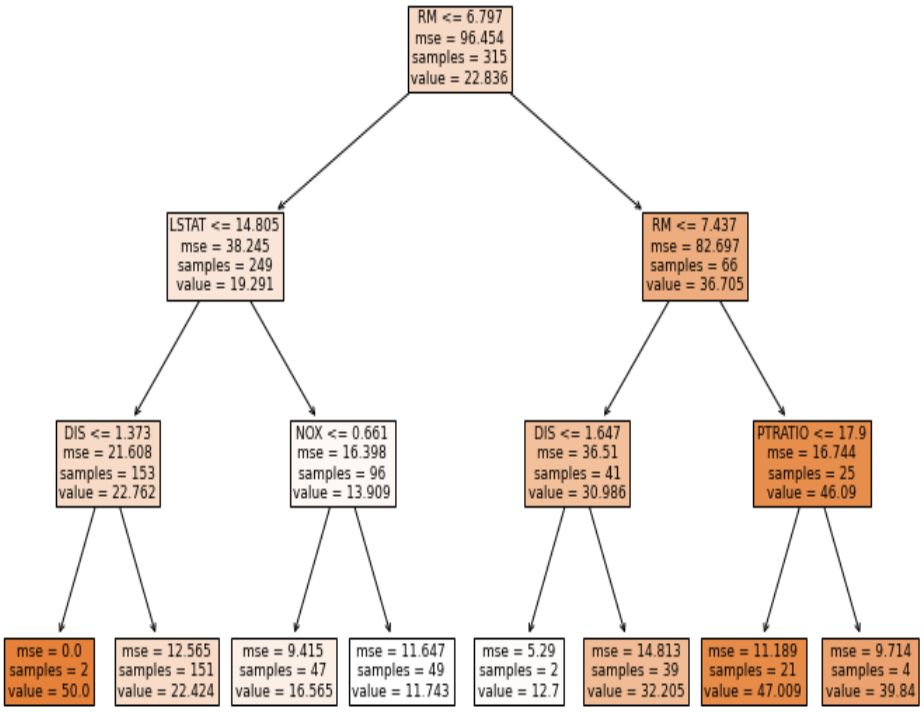

Fig. (8) RF tree diagram. Made by the researchers using CBE data [22]

Figure (9) shows a **Decision Tree** model, which is a simpler form of the Random Forest algorithm. It visualizes the process of decision-making by splitting data into branches based on features, leading to a final decision or classification. In the financial context, a Decision Tree can be used to decide whether to approve a loan application, assess credit risk, or predict customer behavior. The diagram shows the step-by-step process of decision-making in this model. The figure also shows the nodes represented in the tested data from analysis which divided into nodes and leaves, which one based on the higher classification of data which shows the classifiers of financial ways.

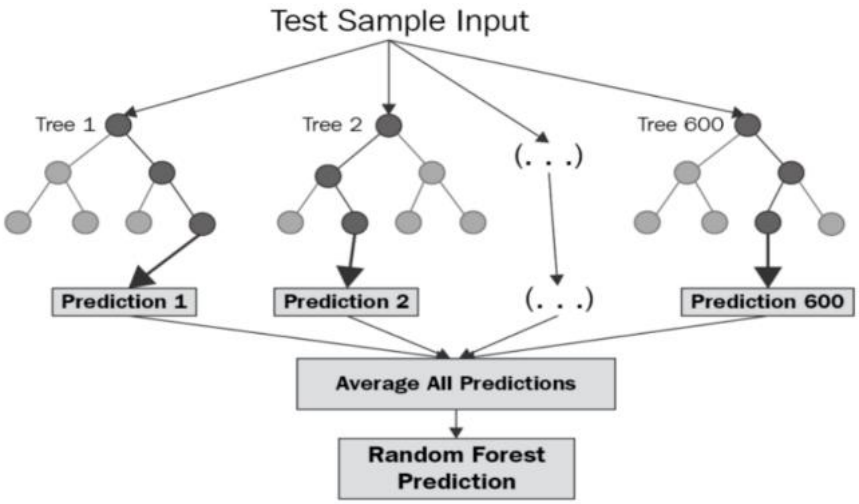

Fig. (9) decision tree test diagram DT. Made by the researchers using CBE data [22]

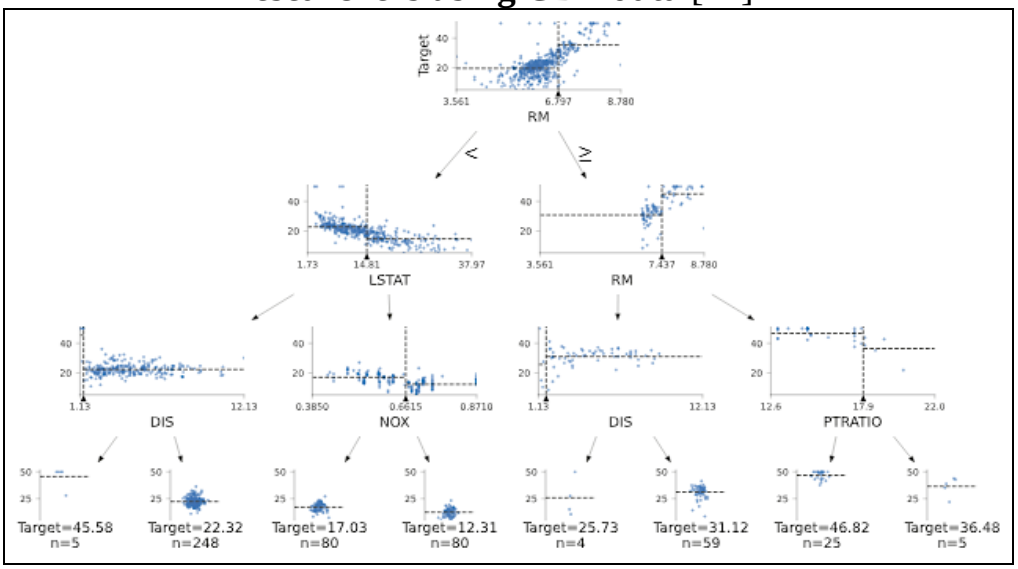

Fig. (10) Pairplot for RF tree diagram. Made by the researchers using CBE data [22]

Figure (10) shows the relations between all data and how this data distributes. A **pairplot** is a way of visualizing relationships between multiple variables in a dataset. This figure likely shows the relationship between different financial features used in the **Random Forest** model. By plotting the variables against each other, the pairplot can help identify correlations, trends, or interactions between different features, which can then be used to improve the predictive accuracy of the model.

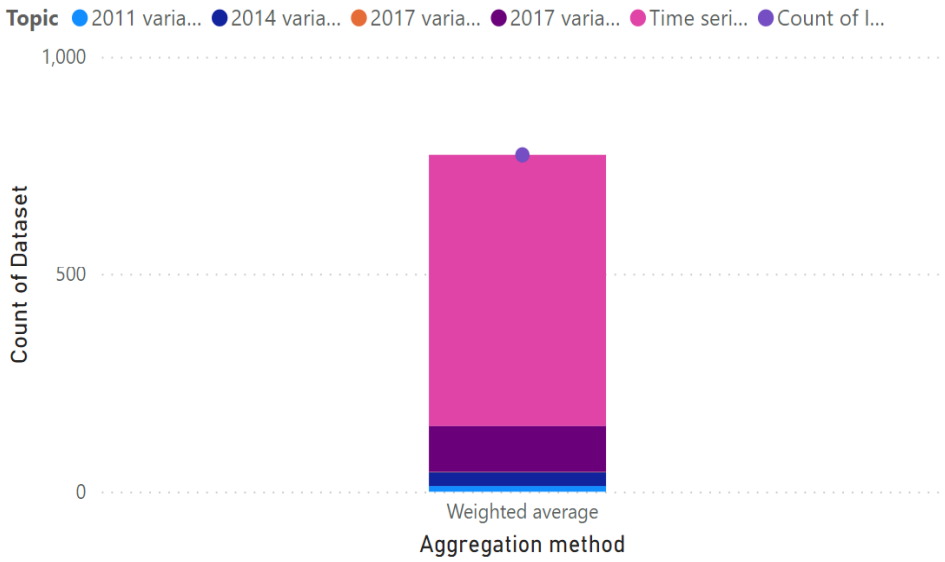

Fig. (11) relation between database and Aggregation. Made by the researchers using CBE data [22]

Figure (11) shows the avg. of financial distributed over years. It figure shows the relationship between a **database** (a structured collection of data) and **aggregation** (the process of combining or summarizing data). In financial analysis, aggregation is used to summarize large amounts of data for easier interpretation, such as by averaging transaction values or summing total investments. This figure likely visualizes how raw data is transformed into aggregated insights for decision-making.

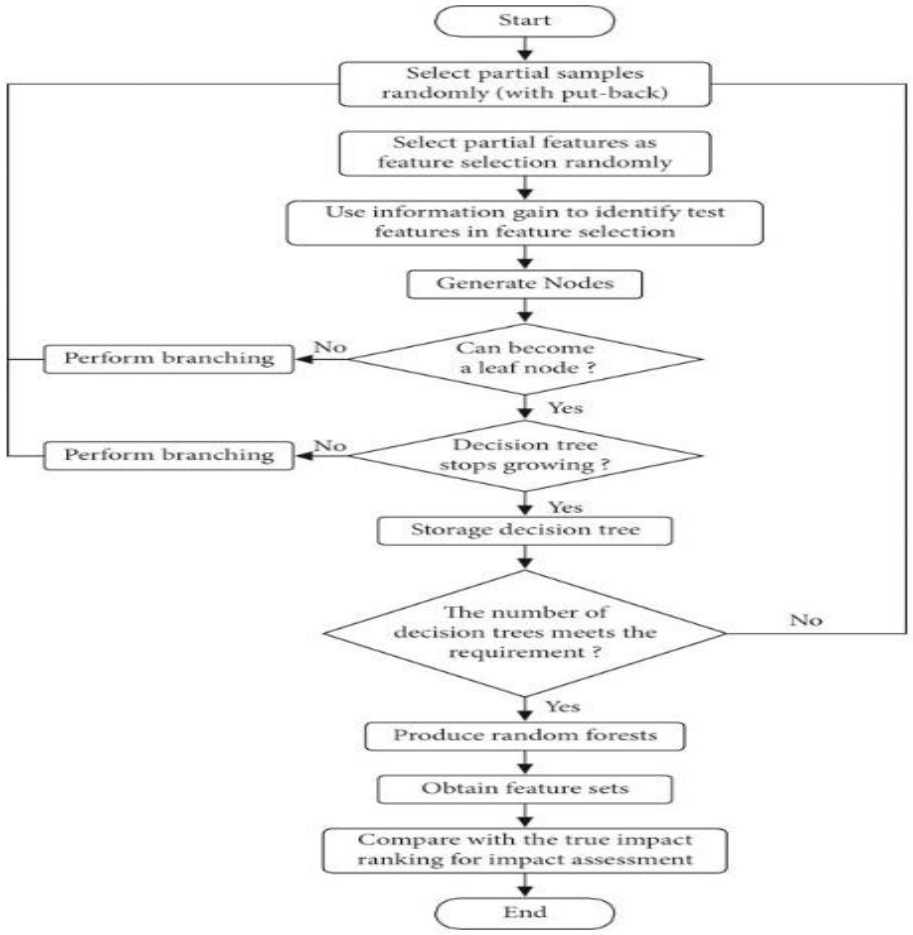

Fig. (13) flow chart of DF algorithm. Made by the researchers using CBE data [22]

This figure shows the tree structure of data analysis, at which the chart shows the structure of a **Decision Forest (DF)** algorithm, which is similar to a Random Forest model. The Decision Forest consists of multiple Decision Trees, where each tree analyzes different parts of the data and the final decision is made based on the consensus of the trees. The flow

chart likely outlines the steps taken to input data, split it into trees, and output a final prediction or decision.

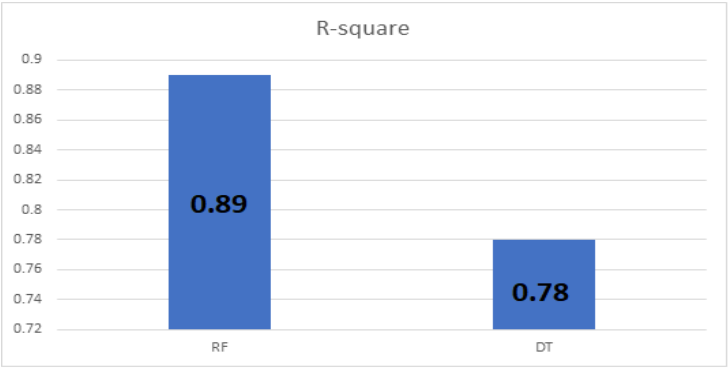

Fig. (14) Difference between RF , DT
Made by the researchers using CBE data [22]

This figure shows the difference between two algorithms RF , DT of accuracy, it compares the accuracy of two algorithms: **Random Forest (RF)** and **Decision Tree (DT)**. It shows that the **R-square value** for RF is **0.89**, while for DT it is **0.78**. The **R-square value** is a measure of how well the model explains the variability of the data. The higher value for RF indicates that the Random Forest model has better predictive accuracy compared to the Decision Tree model, making it more reliable for financial predictions or classifications.

7. Conclusion

In this study, we proposed a Big Data Analytics Framework aimed at enhancing the usability of financial inclusion initiatives through predictive analytics. Our framework leverages the power of advanced machine learning algorithms, specifically Random Forest and Decision Trees, to forecast various aspects relevant to financial inclusion. Through the utilization of these algorithms, we have demonstrated the

capability to predict key indicators such as customer behavior, financial product adoption rates, and potential barriers to inclusion. By analyzing large volumes of data encompassing diverse socioeconomic factors, demographic information, and transactional histories, our framework provides valuable insights for policymakers, financial institutions, and other stakeholders.

The findings derived from our predictive models provide practical suggestions for enhancing the development and execution of financial inclusion strategies. By identifying patterns and trends within the data, decision-makers can tailor their approaches to better meet the needs of underserved populations, enhance accessibility to financial services, and ultimately promote inclusive economic growth.

Furthermore, the integration of Random Forest and Decision Trees into our framework allows for robust and interpretable predictions, enabling stakeholders to understand the underlying factors driving inclusion or exclusion. This transparency fosters trust and confidence in the decision-making process, facilitating the adoption of evidence-based policies and initiatives.

However, it is essential to acknowledge the limitations of our study. While Random Forest and Decision Trees are powerful tools for predictive analytics, their effectiveness relies heavily on the quality and quantity of input data. Therefore, continuous efforts to collect, clean, and update datasets are crucial for maintaining the accuracy and relevance of our predictive models over time.

8. REFERENCES

[1] H. Chen, R. H. L. Chiang and V. C. Storey, "Business intelligence and analytics: From big data to big impact," *MIS Quarterly,* vol. 36, pp. 1165 - 1188, 2012.

[2] P. K. Ozili, "Big Data and Artificial Intelligence for Financial Inclusion: Benefits and Issues," *SSRN Electronic Journal,* 2021.

[3] S. J. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed., Pearson, 2020.

[4] M. M. Hasan, J. Popp and J. Oláh, "Current landscape and influence of big data," *Journal of Big Data,* vol. 7, no. 1, 12 3 2020.

[5] S. Sharma, "The Role of Big Data Analytics in Financial Decision Making: An Indian Perspective," *International Journal of Innovative Research and Creative Technology,* vol. 4, no. 1, pp. 1-7, 1 2018.

[6] A. A. Alzaidi, "Impact of use of big data in decision making in banking sector of Saudi Arabia," *International Journal of Computer Science and Network Security,* vol. 18, no. 12, 12 2018.

[7] E. Al-lozi, A. Alfityani, A. A. Alsmadi, A. M. Al_Hazimehd and J. A. Al-Gasawneh, "Retracted: The role of big data in financial sector: A review paper," *International Journal of Data and Network Science,* vol. 6, pp. 1319 - 1330, 2022.

[8] Z. Z. Noor, H. Hussin and M. N. M. Nazri, "Big Data Initiatives by Governments - Issues and Challenges: A Review," in *6th International Conference on Information and Communication Technology for The Muslim World (ICT4M),* 2016.

[9] J. E. a. J. J. Curto, "A Risk and Benefits Behavioral Model to Assess Intentions to Adopt Big Data," 2013.

[10] M. Favaretto , . E. De Clercq , C. O. Schneble and B. S. Elger , "What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. PLoS ONE 15(2): e0228987," *PLoS One,* 25 2 2020.

[11] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," *International Journal of Information Management,* vol. 35, no. 2, pp. 137-144, 2015.

[12] T. I. Journal, "Big Data Phenomenon in Bankin," *Texila International Journal of academic Research,* 2019.

[13] M. A. Amakobe, "The Impact of Big Data Analytics on the Banking

Industry," *Colorado Technical University,* 7 2015.

[14] R. Sethy and M. Panda, "Big Data Analysis Using Hadoop: A Survey," *International Journal of Advance Research in Computer Science and Software Engineering,* vol. 5, no. 7, 8 2015.

[15] G. Yao, "How is Big Data related to information security - A literature review," 2020.

[16] S. H. Oluwasola, "How to Improve Efficiency of Banking System with Big Data (A Case Study of Nigeria Banks)," *International Journal of Science and Research,* vol. 6, no. 6, pp. 944 - 947, 6 2017.

[17] R. Vadlamani and K. Sk, Big Data Analytics Enabled Smart Financial Services: Opportunities and Challenges, In: Reddy, P., Sureka, A., Chakravarthy, S., Bhalla, S. (eds) Big Data Analytics. BDA 2017. Lecture Notes in Computer Science(), Springer, Cham, 2017, pp. 15 - 39.

[18] M. Anshari , M. N. Almunawar, S. A. Lim and A. Al-Mudimigh , "Customer Relationship Management and Big Data Enabled: Personalization & Customization of Services," *Applied Computing and Informatics,* vol. 15, no. 2, pp. 94 - 101, 7 2019.

[19] W. Brown, G. Wilson and O. Johnson, "Understanding the Role of Big Data Analytics in Enhancing Customer Experience," 24 8 2024.

[20] E. Indriasari, H. Soeparno, F. L. Gaol and T. Matsuo, "Application of Predictive Analytics at Financial Institutions: A Systematic Literature Review," in *8th International Congress on Advanced Applied Informatics (IIAI-AAI)*, Toyama, Japan, 2019.

[21] A. Atkinson and F.-A. Messy, "Promoting Financial Inclusion through Financial Education: OECD/INFE Evidence, Policies and Practice," *OECD Working Papers on Finance, Insurance and Private Pensions,* Vols. 34, OECD Publishing, Paris, 11 10 2013.

[22] C. B. o. Egypt, "Annual Report. miscellaneous releases.".

[23] A. Kumar, "Comprehensive Guide to Descriptive vs Inferential Statistics!," simplilearn, 11 2024. [Online]. Available: https://www.simplilearn.com/difference-between-descriptive-inferential-statistics-article. [Accessed 11 2024].

[24] H. A. Salman, A. Kalakech and A. Steiti, "Random Forest Algorithm Overview," *Babylonian Journal of Machine Learning,* pp. 69 - 79, 6 2024.