

https://dspace.org/

Big Data Analytics as a Customer Retention and Acquisition Strategy within South African Retail Banking

Kharidzha, Muano

2024
University of the Witwatersrand, Johannesburg
https://hdl.handle.net/10539/43916

Kharidzha, Muano. (2024). Big Data Analytics as a Customer Retention and Acquisition Strategy within South African Retail Banking [Master's dissertation, University of the Witwatersrand, Johannesburg].WireDSpace.https://hdl.handle.net/10539/43916

Downloaded from a DSpace repository. https://dspace.org/

Big Data Analytics as a Customer Retention and Acquisition Strategy within South African Retail Banking

Muano Kharidzha

Student number: 386432

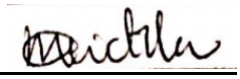
Email: 386432@students.wits.ac.za and mobile: 0823609163

A research proposal submitted to the Faculty of Commerce, Law and Management, University of the Witwatersrand, in partial fulfilment of the requirements for the degree of Master of Management in the field of Digital Business

Johannesburg, 2023

DECLARATION

I, Muano Kharidzha, declare that this research paper is my own work except as indicated in the references and acknowledgements. It is submitted in partial fulfilment of the requirements for the degree of Master of Management in Digital Business in the Wits Business School, University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in this or any other university.



Muano Kharidzha

Signed at Johannesburg, South Africa

On the 25th Day of February 2024

ACKNOWLEDGEMENTS

I would like to dedicate this paper to my son, Bumabumelani Ndzeru, 5 years old, who has been patient with me as I focused on my studies. I want to thank my Lord Jesus Christ for granting me the opportunity to further my studies and for seeing me through the completion of my degree. I would also like to thank my supervisor, Dr. Nakuze Chalomba, for always availing herself for consultations and for her timely responses to my queries.

ABSTRACT

Big Data analytics in banking can lead to superior performance, but the industry is still working to understand its relationship with organisational impact. To stay relevant, banks should invest in big data analytics, offering customer-centric experiences for retention and acquisition strategies, as threats from non-banks threaten the banking space. This study offers a novel approach to investigating how the application of BDA in the form of personalised customer engagements, product offerings, and fraud detection in the South African banking environment can be used as a customer acquisition and retention strategy. A quantitative method of study was used, where digitally active banking customers were requested to complete an online survey. The study's overall conclusion is that, thanks to the considerable impacts of the above-mentioned BDA applications on customer experience, there is a positive correlation between BDA and customer acquisition and retention.

KEYWORDS

Banking, Big Data Analytics, Customer Experience, Customer Retention, and Customer Acquisition

TABLE OF CONTENTS

LIST OF TABLES ..... *ix*

LIST OF FIGURES ..... *x*

LIST OF ACRONYMS ..... *xi*

CHAPTER 1. INTRODUCTION *1*

1.1 Statement of purpose **1**

1.2 Background of the study **1**

1.3 Research problem. **4**

1.4 Research questions. **8**

1.5 Rationale **8**

1.6 Delimitations of the study **10**

1.7 Definition of terms **11**

1.8 Assumptions **12**

1.9 Chapter Outline **12**

CHAPTER 2. LITERATURE REVIEW *13*

2.1 Introduction **13**

2.2 Definition of topic or background discussion **14**

2.2.1 What are big data Analytics? 14

2.2.2 Characteristics of Big Data 14

2.2.3 Types of big data analytics Methods 15

2.2.4 Challenges associated with Big Data Analytics 17

2.2.5 BDA in banking 18

2.2.1 Credit Risk Analysis through BDA 21

2.2.2 Sentiment Analysis And BDA 26

2.3 Primary research question. **29**

2.4 THEORETICAL FRAMEWORK **29**

2.4.1 Theoretical Framework 29

2.5 Hypothesis development And First research question 31

2.5.1 Customer experience and BDA applications 31

2.6 Hypothesis development and Sub research questions 34

2.6.1 Personalisation of products, services, and customer engagements through BDA and Customer Experience 34

Personalised Product and services 37

2.6.2 Fraud Detection through BDA AND Customer experience 40

2.7 Hypothesis development And Second research question 43

2.7.1 Customer retention and new customer acquisitions because of Improved customer experience 43

2.7.2 Conceptual Framework 49

CHAPTER 3. RESEARCH METHODOLOGY 50

3.1 Research approach 50

3.2 Research design 50

3.3 Data collection methods 51

3.4 Population and sample 51

3.4.1 Population 51

3.4.2 Sample and sampling method 52

3.5 The research instrument. 52

3.6 Procedure for data collection 53

3.7 Data analysis strategies and interpretation 53

3.8 Quality Assurance 53

3.8.1 External validity 54

3.8.2 Internal validity 54

3.8.3 Reliability 55

3.9 Ethical considerations 55

CHAPTER 4. PRESENTATION OF RESULTS/ FINDINGS 56

4.1 Introduction 56

4.2 Descriptive statistics 56

4.3 Measurement Instrument Assessment 58
4.3.1 Reliability Analysis/ Testing 60
4.3.1.1 Cronbach Alpha Coefficient 61
4.3.1.2 Composite reliability 61
4.3.2 Validity Testing 61
4.3.2.1 Convergent Validity 62
4.3.2.3 Average Value Extracted (AVE) and Highest Shared Variance (HSV) 63
4.4 Structural Equation Modelling (SEM) 64
4.4.1 Model Fit Indices 64
4.5 The structural Model (Path modelling and Hypotheses Testing) 65
4.6 Results pertaining to Hypothesis 1 67
4.7 Results pertaining to Hypothesis 2 67
4.8 Summary of the results/findings 68
CHAPTER 5. DISCUSSION OF THE RESULTS OR FINDINGS 69
5.1 Introduction 69
5.2 Discussion pertaining to Hypothesis 1: BDA applications improve customer experience. 69
5.2.1 Customer experience and BDA applications 69
5.3 Discussion pertaining to Hypothesis 2: Improved customer experience resulting from BDA applications has a positive impact on customer retention and new customer acquisitions. 74
5.4 Conclusion 77
CHAPTER 6. CONCLUSIONS & RECOMMENDATIONS 78
6.1 Introduction 78
6.2 Conclusions regarding research question 1 78
6.2.1 Customer experience and BDA applications 78
6.3 Conclusions regarding research question 2 79
6.3.1 BDA applications AND Customer Retention 79
6.4 Overall Conclusions of this study 80

6.5 Managerial implications and Recommendations 80

6.6 Limitations of the study 81

6.7 Suggestions for further research 82

REFERENCES 87

APPENDIX A 103

LIST OF TABLES

Table 1. List of Acronyms xi

Table 2. Definitions of terms 11

Table 3. Proposed schedule and timelines 107

Table 4: Overall descriptive statistics 57

Table 5: *Descriptive statistics for each construct* 57

Table 6: Overview of the final research constructs with a focus on the Factor Loadings 60

Table 7: Composite Reliability (CR) and Cronbach Alpha (α) 61

Table 8: Convergent Validity (Factor Loadings) 62

Table 9: Inter-Item correlation matrix 63

Table 10: Highest Shared Variance (HSV) 63

Table 11: Average Value Extracted (AVE) and Highest Shared Variance (HSV) 64

Table 12: Model Fit results 65

Table 13: Model Fit results cut-off criteria 65

Table 14: Summary of the SEM results (Standardized Regression Weights) ... 66

Table 15: Squared Multiple Correlations: R squared values 68

Table 16. Consistency table: research questions, propositions, data collection and data analysis 84

Table 17. Research Questionnaire 103

LIST OF FIGURES

Figure 2: Antecedents for BDA sustained value 19

Figure 3: The strategy for the utilization of social media to improve client experience. 28

Figure 4: Customer Focus Theory 29

Figure 5: A strategic framework for CXM based on CX insights resulting from BDA. 31

Figure 6: Primary perspective about big data applications in banking. 21

Figure 7: knowledge fusion strategy/ taxonomy. 38

Figure 8: Customer experience and customer loyalty relationship 46

Figure 9: Customer experience and customer retention with psychological comfort as a mediator. 47

Figure 10: High Level conceptual framework. 49

Figure 11: Measurement Model 59

Figure 12: The Structural Model 66

LIST OF ACRONYMS

Table 1. List of Acronyms

Acronyms	Meaning
BD	Big Data
BDA	Big Data Analytics
CFA	Confirmatory Factor Analysis
DQ	Data Quality
CX	Customer Experience
CXM	Customer Experience Management
FinTech	Financial Technology
SEM	Structural Equation Modelling

CHAPTER 1. INTRODUCTION

1.1 Statement of purpose

This quantitative study examines the impact of big data analytics applications within the South African retail banking industry and their impact on business performance in an increasingly competitive environment. Specifically, the study examines the impact of personalisation of product offerings, services, and customer engagements, as well as fraud detection, as enabled by BDA, on the customer experience and how that can influence customer retention as well as the acquisition of new customers.

1.2 Background of the study

First National Bank, Standard Bank, Absa, Nedbank, and lastly Capitec are the South African banking industry’s five key players. "In 2023, Standard Bank was the leading South African banking service provider from a tier 1 capital perspective," reaching a capital of approximately 169.9 billion U.S. dollars. First National Bank and Absa Bank came in 2nd and 3rd with capital of around 93.4 billion and 87.1 billion U.S. dollars, respectively. Statista, 2024 (Cowling, 2023). Camarate and Brinckmann (2017) highlight a threat to South African incumbent banks, which they call the "marketplace without boundaries," where non-traditional banks are seen to be looking for opportunities within the financial services spaces. with retail companies, large employers, taxi associations, etc. as potential entrants. Camarate and Brinckmann (2017) suggest that one of the key strategies for incumbent banks to remain relevant is through digital transformation and delivering customer-centric products and services using big data and analytics capabilities. Insights obtained from transactional data can be used to understand customer spending habits, which can be used to tailor product offerings. Banks can use the insights to create customer retention strategies for their high-value customers.

In their South African bank results analysis annual report (2023), PricewaterhouseCoopers (PWC) has noticed an intentional change in the configuration of costs within the major South African banks. Intentional IT expense growth and talent spending continue to grow. PWC (2023) attributes this to South African banks' digitally led focus, customer experience enhancement, boosting productivity, and efforts to streamline their IT infrastructure so as to improve on their legacy systems. The efforts to be digitally led are no doubt leading to the generation of big data. According to PWC (2023), the South African retail and business banking industries continue to face intense competition, as the top management of the various banks has acknowledged. This can be attributed to niche lenders with agile technology stacks (i.e., fintech). In response to this threat, South African retail banks intend to focus on opportunities presented by technology and competitive disruption. From a competition perspective within the South African context, we've also seen the launch of digital banks such as Timebank, Discovery, and Bank Zero in the past few years.

The operating environment for South African retail banks continues to present challenges. The South African economy continues to be under pressure, with StatsSA reporting a 1.3% decline in the GDP in the 4th quarter of 2022. This has raised concern about a recession. In an effort to curb inflation and the cost of living, the South African Reserve Bank has also been increasing its bench repo rate; a recent hike of 50 basis points to 7.75% was made in March 2023, which is noted to be the 9th consecutive hike since November 2021, resulting in increases in the cost of borrowing (Trading Economics, 2023).

PWC (2023) also notes that complex global geopolitics (i.e., Ukraine's inversion by Russia and China's COVID-19 lockdown policies) have exerted inflationary pressures on African countries and the rest of the world. The most unique problem that South Africa faces is loadshedding, which is having an impact on the economy, predominantly businesses, and may soon result in job losses. The above-mentioned challenges will most likely result in reduced deposits and the inability to pay back loans. Carletti et al. (2020), who did a study on the COVID

shock and equity shortfall in an Italian firm or bank, identified that the number of defaults resulted in a significant decrease in profitability as a result of drained market liquidity.

According to Murinde et al. (2022), both digital banks and Neobanks provide traditional banking products such as current and savings accounts, credit products, and payment services. These banks thrive by leveraging technology with no physical presence, such as branches; however, what differentiates a Neobank from a digital bank is that it does not hold a banking license.

They further suggest ways in which banks can remain competitive through digital transformation. Murinde et al. (2022) paint a picture of what the future of banks should look like and present opportunities for development; these include a platform business model, mastering of the data economy, selling experiences instead of products, personalisation, financial inclusion, and social responsibility. Based on the presented opportunities, it's clear that BDA is key to enabling most of the suggested opportunities, i.e., mastering the data economy, personalisation, and financial inclusion (i.e., from a credit risk assessment perspective).

Navaretti et al. (2018) suggest that to compete with new entrants, incumbent banks need to adopt new strategies in order to face rising competition. Both Navaretti et al. (2018) and Murinde et al. (2022) concluded that fintech will not replace banks but rather that both fintech and incumbent banks will most likely co-exist in the future. This is as a result of the fact that Fintech has limitations in diversifying their product offerings without tapping into the regulated space, thus also becoming banks. In the same vein, large banks have leveraged technology to develop fintech-like offerings by digitising their processes and having fintech-like platforms that are cheaper and more efficient. As a result, one can thus conclude that banks need to continue to innovate as competition from fintech gets steeper.

According to McKinsey & Company (2018), advanced analytics have become an enabler of superior performance for organisations that have made proper

commitments and investments. However, the banking industry has not fully realised the full potential of BDA by embedding the capability into its culture, decision-making processes, and business operations. McKinsey & Company (2018) notes that data analytics foundations have been created within banks; however, they believe that there is still much room for improvement in data-driven banks' performance. The study revealed that banks need to improve in the following areas: "Align analytics to the strategy of the organisation," "Need to embed BDA in decision making." "Need to develop BDA assets that can be scaled instead of standalone initiatives," "Need to invest in critical BDA roles," and "Need to enable the user revolution."

The challenges mentioned above mean that incumbent banks ought to act or risk losing their competitive advantage. We've selected BDA as a source of big data value creation for this research paper. This paper sheds some light on the extent to which customers perceive the value created by the use of BDA within South African retail banking. By presenting a conceptual model that identifies the key big data analytics applications, their impact on customer experience within the South African retail banking industry, and the changes necessary to make the transformation, this paper also adds to the body of scholarly knowledge.

1.3 Research problem.

The problem that this research solves is to determine whether big data analytics applications that can be used by South African retail banks have a positive relationship with customer experience, resulting in customer retention and new customer acquisitions. This was achieved by investigating whether some of the BDA applications in banking identified in the literature has a positive relationship with customer experience.

South African retail banks' strategic and financial performance is likely to be negatively impacted by competition, unfavourable economic and geopolitical conditions, and rising customer expectations; as a result, banks must act quickly

or risk losing market share to rivals and new market entrants, according to PWC (2017). Even though BDA has been recommended by many as a key enabler to superior business performance, the banking industry has not fully realised the full potential of BDA through embedding the capability into their culture, their decision-making processes, and their business operations (McKinsey & Company, 2018). McColl-Kennedy et al. (2018) also indicate most organisations still struggle to generate valuable insights even though big data is captured. Holmlund et al.'s (2020) study based on customer experience management (CXM) in the age of BDA was done as a result of identifying a knowledge gap on the importance of BDA in the improvement of customer experience (CX) in its infancy stages.

Murinde et al. (2022), who conducted a study on the threat from fintech and digital banks against incumbent banks, recommend digital transformation as a response strategy in which the BDA is a key driver. Camarate and Brinckmann (2017) highlighted a rise in non-traditional banks entering the South African banking space and, as a result, recommended the use of BDA to provide customer-centric products and services. McKinsey & Company (2018) argues that advanced analytics have become an enabler of superior performance for organisations that have made proper commitments and investments.

South African researchers: Vallentyn (2021), Buldoo (2018), Matsimbi (2020), and Ogochukwu (2022) all did their research related to BDA within the South African retail banking industry; however, none of them directly focused on the BDA applications and their impact on customer experience improvement and business performance. As a result, this study intends to contribute to the empirical knowledge pertaining to BDA within the retail banking space.

Vallentyn (2021) explored how South African incumbent banks can improve their competitiveness through investing in BDA. The research findings revealed that there’s room for SA incumbent banks to improve on pricing, operational efficiency, and customer value management to improve their competitiveness through BDA. The researcher further recommends ways in which SA's incumbent

banks can leverage BDA to become more competitive and avoid price competition.

The use of BDA for a competitive edge within the South African banking sector was also investigated by Buldoo (2018). The scope is wide, with the aim of answering four research questions, including the use and application of BDA in banking, the resources required for BDA, the challenges faced in BDA implementations, and the inter-relationships between BDA assets. The research findings highlight the importance of top management, organisational designs, interdepartmental dynamics, and organisational culture as key enablers for successfully leveraging BDA capabilities. The author also notes that organisations still struggle to create BDA capabilities for competitiveness, despite the promise that BDA can provide superior organisational performance.

Regulation, legacy systems, and the traditional silo structure have been noted as key inhibitors of leveraging BDA. Key resource requirements for BDA were identified as human resources, technology, partnerships, and training. To understand how BDA is being used by South African banks for competitive advantage, Buldoo (2018) found that BDA is predominately being used for personalisation and customisation, target marketing, and operational efficiencies. The author came to the conclusion that BDA adoption in the South African banking sector is still in its early stages.

Matsimbi (2020) also conducted a qualitative case study on big data adoption challenges within a South African bank. With the objective of understanding how big data is collected and processed, the study revealed that the bank is utilising BDA tools that meet industry standards. However, what he found is that skills scarcity is the number one challenge, followed by organisational structure and governance issues, leading to a lack of alignment on how big data should be packaged as well as how BDA projects should be run.

The difficult market conditions under which South African retail banks operate may drive them into a cost-cutting strategy, which is most likely to have an impact

on the bottom line and has been deemed unsustainable. This study recommends a differentiation strategy through building a defendable data asset (Buldoo, 2018). The study aims to solve the problems faced by South African banks as well as the under-realisation of BDA benefits for improving business performance. Big data use and applications as one of the big antecedents of business will be the focus, since none of the previous researchers fully focused on and unpacked the use of big data.

A qualitative study done by Ogochukwu (2022) examines BDA's competitive advantage within the South African financial services sector, specifically banking, with a key focus on understanding BDA's strategies, resources, technologies, skills, privacy, and security. The study is broad as it explores BDA challenges, applications, adoption, and technologies. From a BDA application perspective, customer retention and customer service were explored. Therefore, this study seeks to fill some of these gaps by investigating how BDA applications can influence customer experience and business performance in the South African retail banking industry.

1.4 Research questions.

What is the impact of BDA applications on customer experience for customer retention and new customer acquisition within the South African Retail Banking Industry?

I. Does the use of BDA improve customer experience?
    a. Will the customer experience improve because of utilising BDA for the personalisation of products, services, and customer engagements?
    b. What is the relationship between Fraud detection and customer experience?
II. Will the use of BDA to improve customer experience result in customer retention and new customer acquisitions?

1.5 Rationale

According to Porter's three generic strategies, an organisation may choose one of three approaches to achieving and sustaining competitive advantage: cost leadership, differentiation, or a focus strategy. According to Amstrong and Lee (2021), a cost leadership strategy targets a broad market by being the largest in the industry at the lowest cost, with supplier bargaining power, cost control, or technological advantage, and a focus on operational effectiveness. However, this strategy is argued to provide the least sustainable competitive advantage and result in price wars (Tanwar, 2013). A firm can also pursue a differentiation strategy targeting a broad market with a unique product offering; this strategy is said to provide a more durable competitive advantage. Lastly, a firm can pursue a focus strategy that focuses on a niche group or subset of customers, either through cost leadership or differentiation from a product perspective (Amstrong and Lee, 2021).

In an effort to provide insights on whether big data can improve customer retention and the acquisition of new customers through improved customer

experience (i.e., as a source of differentiation), this study uses the Customer Focus Theory (Lohan et al., 2011) to develop multiple hypotheses.

Big data analytics have gained enormous popularity over the past ten years, but organisations still struggle to understand how to use big data to add value to their operations (Vidgen et al., 2017). The author concentrated on the difficulties organisations face in gaining performance benefits from big data and business analytics in order to fill the research gap. Key issues like a lack of a data-driven culture, poor data quality, and the requirement for a BDA strategy that is aligned with or included in the organisational strategy have become prominent, according to Vidgen et al. (2017). According to Sharma et al. (2014), while there is some evidence that BDA may be able to generate business value, further investigation is needed to show that it actually does.

Both Popoviaz et al. (2016) and Wamba et al. (2017) noted that many organisations continue to struggle to realise value from their BDA investments. Popoviaz et al.'s 2016 study found that the utilisation of BDA in manufacturing operations enhances performance. They argue that big data capability (i.e., data sourcing, data processing technologies and methodologies, and people's expertise) and organisational factors (i.e., BDA strategies, organisational culture, management support, et al.) can make or break the effective utilisation of BDA and the performance benefits. According to a 2017 study by Wamba et al., personnel capabilities, infrastructure flexibility, management capabilities, and data analytics capabilities all directly contribute to an increase in business performance.

Big data analytics (BDA) has the potential to revolutionise established business practices. The performance of a company is still unknown to be impacted by BDA capabilities, according to Rialti et al. (2019). The author argues that the complex relationship between BDA and organisational impact is still being understood by scholars and that part of the research is still theoretical or qualitative at most, with quantitative studies still in their infancy. BDA capabilities being discussed by Wamba et al. (2017) include the flexibility of the BDA infrastructure, management abilities, and organisational personnel capabilities. Even though more companies

are launching big data initiatives, little is understood about how they actually turn the potential of these technologies into business value (Mikalef et al., 2019).

1.6 Delimitations of the study

The study is limited to the South African retail banking industry and how they're using BDA to improve customer experience.

Delimitations of the study include:

- Theories
    - The customer Focus Theory
- Conceptual framework
    - Big Data Analytics applications (use cases)
        - Fraud Detection
        - Credit Risk Analysis
        - Personalised Customer engagements
        - Customer sentiment analysis
        - Personalised product and service offerings
    - Business performance because of BDA
    - Customer experience as a mediator

Delimitations of the study exclude:

- Big data maturity
- Big Data Capabilities (i.e., How big data is being collected, processed and the relevant tools, people, culture, leadership).
- Big data quality and implementation challenges
- Change management pertaining to the implementation of BDA projects.

1.7 Definition of terms

Table 2. Definitions of terms

Term	Definition
Big Data	Data that cannot be processed using conventional methods due to its high volume, velocity, variety, variability, veracity, and value (Sivarajah et al., 2017).
BDA (Big Data Analytics)	Techniques used to extract knowledge from sizable datasets are referred to as BDA, which is a sub-process of the insight extraction process from BD (Labrinidis and Jagadish, 2012) as cited in as cited by Sivarajah et al. (2017).
BDA (Big Data Analytics) methods	These, which include descriptive analytics, inquisitive analytics, predictive analytics, prescriptive analytics, and pre-emptive analytics, enable the generation of insights from big data. (Sivarajah et al., 2017).
CX (Customer Experience)	Responses from customers to interactions with a company through various channels and over time, these may occur before, during, or after a purchase or consumption. (Holmlund et al., 2020).
CXM (Customer Experience Management)	A higher-order resource of cultural mindsets toward customer experience, strategic directions for CX design, and organizational capacities for continuous CX improvement, with the objectives of achieving and maintaining long-term customer loyalty. (Homburg et al., 2017)
FinTech	FinTech, or financial innovation made possible by technology, has an impact on financial markets, organizations, and the provision of financial services and has the potential to produce new business models, items, methods, and applications. (The Financial Stability Board).

1.8 Assumptions

I. Big data analytics capabilities already exist within the South African retail banking space, as highlighted by Ogochukwu (2022) and Buldoo (2018). According to Wamba et al., 2017 big data capabilities can be broken down into three components namely, Infrastructure, personnel expertise, and management.

II. South African retail banks generates and source big data (i.e., large volumes, with high veracity and at high velocities) as seen in Matsimbi (2020).

III. According to Buldoo (2018), there is a strategic alignment between the organisational strategy and that of BDA within the South African retail banking sector.

1.9 Chapter Outline

The first chapter outlines the statement of purpose, followed by the background of the study, which aims to give context around where the study is being conducted and reasons as to why there is a need to conduct the study. The conceptual model then follows with the hypotheses to be tested. The research questions that the study aims to answer are then outlined. The rationale of the study touches on the theory upon which it is based as well as gaps identified by the literature within the topic. The delimitations of the study indicate the scope of the study within the context of big data analytics within banking. Lastly, the definitions of terms that are being used regularly and the assumptions made in the study are also highlighted.

CHAPTER 2. LITERATURE REVIEW

2.1 Introduction

The topic of big data analytics is well researched, especially at an international level. There is a combination of systematic literature review studies, qualitative and quantitative studies, as well as experimental studies across BDA topics that have been cited in this study. Even though there are a lot of studies on this topic, e.g., Côrte-Real et al. (2019), Delgosha et al. (2019), Ghasemaghaei & Calic (2019), many authors have asked the question of whether the value of big data analytics is being created and realised at an organisational level.

The literature review within this paper covers the definition of big data analytics, the challenges associated with BDA, the methods and capabilities of BDA, as well as the applications of BDA in banking. Sivarajah et al. (2017) and Kshetri (2016) covered the characteristics of big data. Sivarajah et al. (2017) also covered BDA methods, citing several authors, together with Akter et al. (2019) and Gupta et al. (2018). A number of BDA challenges were identified by the likes of Vidgen et al. (2017) and Delgosha et al. (2019), including BDA skills, management, strategy, data availability, and data quality, to name a few.

A study by Côrte-Real et al. (2019) introduces antecedents for sustained BDA value in the form of business performance within retail banking. One of the key antecedents of BDA is BDA use; as a result, this study aims to further unpack BDA use and applications. With the understanding that there is a lack of empirical studies in big data applications but rather reports on big data analytics use cases by practitioners, with the goal of examining key drivers and strategic applications of big data analytics within banking, Delgosha et al. (2019) conducted a study on BDA in banking.

The literature review then delves into each BDA application, including the use of BDA for fraud detection, the personalisation of products and services, and customer engagements. Credit risk assessment is another key BDA application

that was reviewed. As one unpacked the applications identified by Delgosha et al. (2019), one came to the realisation that some of the listed BDA applications are interdependent and therefore need to be combined. One also came to the realisation that the customer is at the centre of BDA applications, and as a result, one proposes that customer experience needs to be a mediator within the research model.

To understand customer experience and BDA, Holmlund et al. (2020) provided a framework, and Anshari et al. (2019), whose study focuses on the use of CRM and BDA to provide personalised customer experiences, were reviewed. Lastly, with the objective of this study being to understand customer retention and acquisition as a result of an improved customer experience from BDA, the relevant studies were also reviewed.

2.2 Definition of topic or background discussion

2.2.1 *What are big data Analytics?*

Sivarajah et al. (2017) believe that we're living in an era of data and bring up a phenomenon called "data divulge," resulting from the large amounts of complex and heterogeneous data being generated from different sources in real time from any device. "BDA refers to techniques used to extract knowledge and insights from large datasets." (Labrinidis and Jagadish, 2012), as cited by Sivarajah et al. (2017).

2.2.2 *Characteristics of Big Data*

Characteristics of big data include volume, velocity, variety, variability, veracity, visualisation, and value. According to Sivarajah et al. (2017), "volume" refers to the enormous amounts of data generated daily from different sources and devices. The large volumes of data are said to be dynamic in nature, heterogeneous, and ambiguous in most instances. The different data sources

can include web clicks, social media, business transactions, emails, blogs, IOT devices, the news, websites, and reports.

Data can come in many formats, which may be structured or unstructured (i.e., image, video, multimedia, audio, text, transactional, etc.). This is called variety. Veracity refers to the inconsistencies, complexities, and trustworthiness (i.e., stemming from potential biases, fabrications, and people's opinions) of large datasets. According to Vasarhelyi et al. (2015), veracity is used to determine the usefulness of data in analysis by measuring its accuracy.

The potential inaccuracies of data present another challenge that needs to be dealt with using BDA tools (Gandomi & Haider, 2015). Velocity is the speed at which heterogeneous human and machine data is generated. Variability is another characteristic of data relating to the constant and rapid changes in its meaning. Data visualisation is about presenting the processed data in a readable manner for the understanding of insights. Last but not least, Sivarajah et al. (2017) define value as the information and understanding gained from big data sets.

2.2.3 *Types of big data analytics Methods*

Sivarajah et al. (2017) identified that there are four methods of BDA that can be conducted by analytics software such as Hadoop and Dyrad. The simplest form of BDA is descriptive analytics, which is responsible for giving hindsight knowledge of what happened through the use of simple statistical methods. This method is useful for making business decisions as one can do correlations and root cause analysis through the discovery of useful patterns; the authors also call it inquisitive analytics.

Predictive analytics is said to provide insights into what the future holds for the organisation through forecasting enabled by statistical modelling. According to Gandomi and Haider (2015), predictive analysis can be categorised into

regression techniques (e.g., for making extrapolations using current and historical patterns) and machine learning techniques based on various algorithms.

BDA was the subject of a systematic literature review by Sivarajah et al. (2017), which produced a comprehensive analysis of BDA techniques and difficulties. The authors created a conceptual framework for categorising big data challenges, which includes issues with data characteristics (such as volume, velocity, etc.), data processing issues (such as data acquisition and warehousing, mining and cleaning, aggregation and integration, analysis and modelling, and finally data interpretation), and management issues (such as issues with understanding and analysing the data—privacy, security, data governance, etc.).

Vidgen et al. (2017) explore the challenges that management faces in creating value from big data analytics and changing the organisation's decision-making culture. The study was motivated by the fact that even though big data analytics has gained popularity over the last decade, organisations still face challenges on how to create business value by leveraging big data.

Vidgen et al.'s Delphi study identified key challenges in becoming a data-driven organisation, including a lack of a data-driven culture, data ethics, and alignment of BDA strategy with organisational strategy. Challenges included managing data quality, using analytics for decision-making, creating BDA strategies, ensuring data availability, and developing data skills. Data quality is seen as the most prevalent issue associated with BDA, which may be associated with legacy systems and the way data is managed.

Ghasemaghaei & Calic (2019) examined the impact of big data quality (DQ) on an organisation's decision-making. They used organisational learning theory and Wang and Strong's DQ framework to identify four elements of DQ: intrinsic, contextual, representational, and accessibility. The study found that intrinsic DQ, representational DQ, and contextual DQ significantly impact data diagnosticity, while accessibility DQ did not. The authors suggest that accessing valuable data is crucial for generating valuable insights. Ghasemaghaei & Calic's (2019)

findings indicate that organisations should have high utilisation of big data with high DQ and dignosticity as mediators. Firms with the highest diagnosticity and decision quality result from their use of big data.

Delgosha et al. (2019) also identified big data challenges during their Delphi study; these included data, organisational, and technological categories. "Data silos and information" was ranked as the most important challenge, followed by "lack of technical skills." Privacy and security concerns are also at the top of the list, followed by implementation costs.

2.2.4 *Challenges associated with Big Data Analytics*

BDA was the subject of a systematic literature review by Sivarajah et al. (2017), which produced a comprehensive analysis of BDA techniques and difficulties. The authors created a conceptual framework for categorising big data challenges, which includes issues with data characteristics (such as volume, velocity, etc.), data processing issues (such as data acquisition and warehousing, mining and cleaning, aggregation and integration, analysis and modelling, and finally data interpretation), and management issues (such as issues with understanding and analysing the data—privacy, security, data governance, etc.).

Vidgen et al. (2017) explore the challenges that management faces in creating value from big data analytics and changing the organisation’s decision-making culture. The study was motivated by the fact that even though big data analytics has gained popularity over the last decade, organisations still face challenges on how to create business value by leveraging big data.

Vidgen et al.'s Delphi study identified key challenges in becoming a data-driven organisation, including a lack of a data-driven culture, data ethics, and alignment of BDA strategy with organisational strategy. Challenges included managing data quality, using analytics for decision-making, creating BDA strategies, ensuring data availability, and developing data skills. Data quality is seen as the most

prevalent issue associated with BDA, which may be associated with legacy systems and the way data is managed.

Ghasemaghaei & Calic (2019) examined the impact of big data quality (DQ) on an organisation's decision-making. They used organisational learning theory and Wang and Strong's DQ framework to identify four elements of DQ: intrinsic, contextual, representational, and accessibility. The study found that intrinsic DQ, representational DQ, and contextual DQ significantly impact data diagnosticity, while accessibility DQ did not. The authors suggest that accessing valuable data is crucial for generating valuable insights. Ghasemaghaei & Calic's (2019) findings indicate that organisations should have high utilisation of big data with high DQ and dignosticity as mediators. Firms with the highest diagnosticity and decision quality result from their use of big data.

Delgosha et al. (2019) also identified big data challenges during their Delphi study; these included data, organisational, and technological categories. "Data silos and information" was ranked as the most important challenge, followed by "lack of technical skills." Privacy and security concerns are also at the top of the list, followed by implementation costs.

2.2.5 *BDA in banking*

"Although big data analytics (BDA) is considered the next "frontier" in data science by creating potential business opportunities, the way to extract those opportunities is unclear." (Côrte-Real et al., 2019, p. 160). As a result, the authors conducted a Delphi study to understand the antecedents of BDA value creation in an organisation. To explain the antecedents of the BDA value, an empirically supported model was developed. Dynamic capabilities, strategic IT and business alignment, the strategic role of BDA, environmental volatility (i.e., market turbulence and technological turbulence), and business use are the most critical contributors to business value in the form of strategic and financial performance.

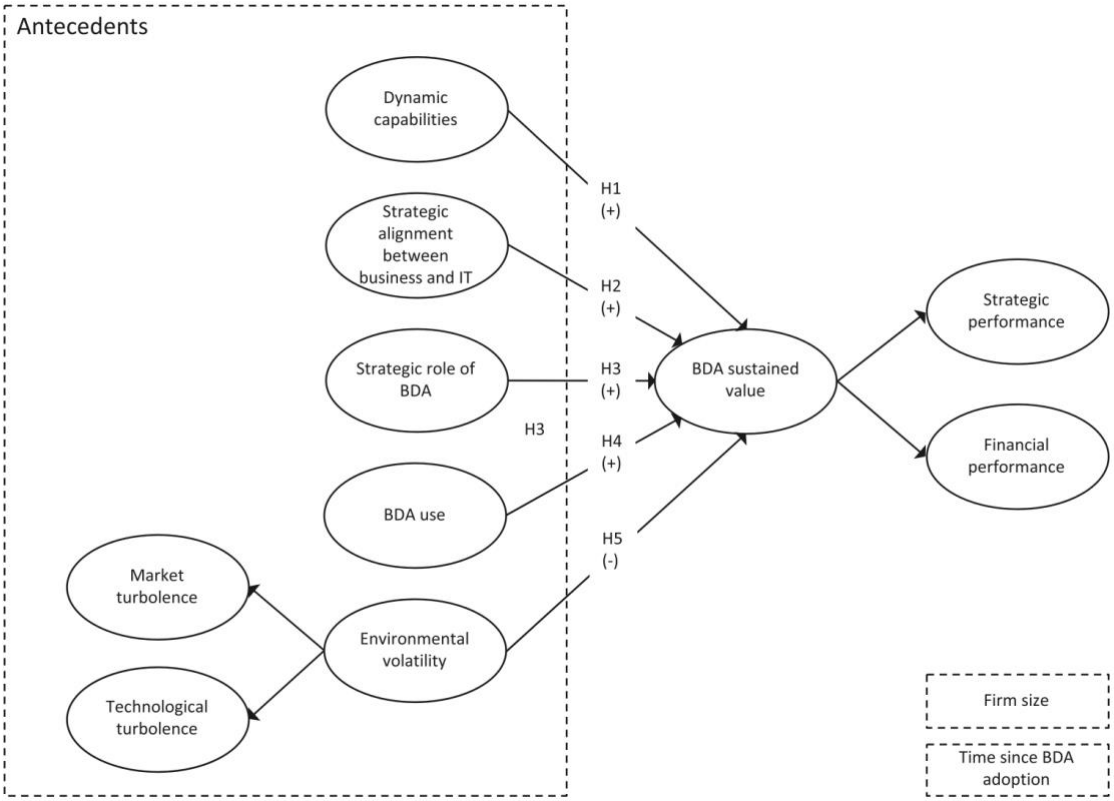


Figure 1: Antecedents for BDA sustained value

Source: Adopted from Côrte-Real et al. (2019)

These are considered crucial to creating and maintaining business value and achieving a competitive advantage as a result. The most important antecedent as a result of the Delphi study is dynamic capabilities, which refer to the firm's ability to respond rabidly to changes through the use of BDA capabilities. Agility comes in second, and it is the result of a strategic and holistic look at business processes to improve them (i.e., automation of business processes through the use of knowledge gathered from big data).

Côrte-Real et al. (2019) argue that a BDA strategy motivated by top management is critical, as their results indicate that a top-down approach is preferred as the top ranking antecedents are strategic, followed by managerial, and lastly operations; otherwise, BDA implementations will most likely fail. The study also

discusses the value of a data-driven culture and proposes further studies on the topic.

BDA usage is the actual application of BDA, which refers to activities such as sales and customer relationships, marketing, product and service enhancements, production and operations, supplier relations, etc. Côrte-Real et al. (2019) highlighted TG4 as a platform company that uses Big Data Analytics (BDA) tools to analyse customer attributes and business transactions, developing predictive models for real-time customer targeting and personalised messages. The study focuses on BDA use as it is considered the most effective antecedent for long-term business value, as stated by a European FMCG big data director.

Hock et al. (2022) conducted a study pertaining to big data analytics in the German retail banking industry with the aim of examining how big data is being implemented and future development opportunities. To determine whether the German banks have been using a cost-focused strategy or a differentiation strategy (through BDA), they examined any evidence of cost-cutting vs. differentiation. An empirical study was conducted by testing six hypotheses regarding the developments in the German banking sector from 2003 to 2019. Their findings show that the number of employees within the sector and the number of credit institutions have declined significantly. The operating conditions have been stagnant throughout the years, with a positive trend on the balance sheet. The cost-to-income ratio has been showing a long-term deteriorating trend.

Lastly, whether the use of big data has an influence on the developments within the German retail banking sector could not be concluded, meaning that a positive effect of big data analytics could not be measured by 2019, so as a result, Hock et al. propose a similar study at a later stage. As a result of this study, one can conclude that the German banking sector is still utilising the cost strategy and has not fully utilised big data capabilities as a differentiation strategy.

With the aim of understanding and prioritising the strategic applications of BDA, challenges, and main drivers of BDA implementations in banks, Delgosha et al. (2019) conducted a Delphi study; see results pertaining to BDA applications in figure 6 below.

Rank	Application	Round 3 Rank mean	Round 4 Rank mean
1	Fraud detection	1.94	1.72
2	Credit risk analysis	2.33	2.17
3	Customer 360	2.50	2.50
4	Customers segmentation	4.06	3.89
5	Designing new business models and new products and service development	4.83	4.94
6	Increasing loyalty and decreasing customer attrition	6.50	6.56
7	Customer recommender systems	7.33	7.33
8	Increasing efficiency and reducing costs	7.61	7.78
9	Customers sentiment analysis	7.78	8.00
	Kendall's W	0.63	0.78

Figure 2: Primary perspective about big data applications in banking.

Source: Adopted from Delgosha et al. (2019)

"Fraud detection" and "credit risk" were found to be the most important applications of BDA in banks, followed by "Customer 360 (customer behaviour)," "Customers segmentation," "Designing new business models and new products and service development," "Increasing loyalty and decreasing customer attrition," "Customer recommender systems," "Increasing efficiency and reducing costs," and lastly, "Customers sentiment analysis" in order of ranking. "Decision-making enhancement" and "new product or service development" were seen as the key drivers, followed by "survival" and "creation of competitive advantage."

2.2.1 *Credit Risk Analysis through BDA*

Having recognised that there is a gap in the provision of financial services by financial institutions due to a lack of readily available credit information from credit bureaus, some companies have resorted to the use of big data analytics as a way

to determine creditworthiness. Tech companies such as QQ, developed by Tencent within the Chinese market, have resorted to big data to determine credit worthiness in order to provide credit products to such customers. For SMEs, the lack of audited bank statements and credit histories, which would typically be in the bureau databases, makes it hard for financial institutions to determine the creditworthiness of such companies, thus not providing them with the required credit assistance; the same goes for customers (Kshetri, 2016).

As a result of the identified gap in the Chinese financial market, the Kshetri (2016) study examines the role of big data analytics in accessing financial products for SMEs and low-income individuals. The author started by exploring a few cases within the Chinese market, including Alibaba, Tencent, JD.com, China CITIC Bank, and China's state-owned banks, and how they've managed to use their big data assets to assess customers' creditworthiness.

Alibaba is said to have used transaction and payment data from their huge online ecosystem (i.e., their e-commerce platform) to develop its own credit ratings and risk control models (Li, 2014). Tencent, on the other hand, also has a huge online ecosystem called "WeChat," which provides a variety of services that generate big data. The data asset from "WeChat" was used to birth "WeBank," an online-only bank that gives credit ratings based on data from online activities like social media, gaming websites, and online shopping (Zhou, 2015), as cited in Kshetri (2016).

JD.com, a Chinese online retailer with over 100 million active customers in 2015, had a joint venture with an organisation called ZestFinance. The pair used algorithms that use customers' recent and historical online shopping patterns to predict default, taking into account factors such as times of purchase, types and prices of items being purchased, etc. (Rao, 2015), as cited in Kshetri (2016). In order to improve its marketing precision to target credit card systems, China CITIC Bank has developed a customer database that takes into account customers' profile information, demographic information, transaction data, online

shopping preferences, and social media activities. This improved their online application approval rate by 60% (Chng & Zhang, 2014).

China's state-owned banks have decided to respond to the big data trend by diversifying their business models to move into the e-commerce space either through partnerships or building in-house big data capabilities with the aim of collecting customer transaction data. CCB and BOCOM built e-commerce platforms; BOC in 2013, followed by ICCB in 2014, followed by ABC, which developed an e-commerce platform targeting rural China. The move into the e-commerce space then resulted in a rise in demand for financial products, creating an online financing market (Hongyuran et al., 2015).

Kshetri (2016) also touched on the Chinese government's credit code, which aims to provide each adult with a credit code based on their financial standing, social media behaviour, criminal record, etc. This is regarded as the most spectacular instance of the application of technology and BDA to governance, and it serves as an example of a method of population control and manipulation that is effective. (Kshetri, 2016). They also expanded on the use of financial and non-financial information to assess creditworthiness, which includes identity information, ability to repay information such as income and current debt, and willingness to pay as per customers historical credit behaviour.

Business models related to the use of big data to enhance financial access for those whose information is not in the credit bureaus. Having recognised the value of information about potential borrowers, financial institutions such as banks are diversifying their product offerings through e-commerce platforms. Big Data companies like Alibaba and Tencent in the Chinese market have tapped into financial products and services such as loans (Kshetri, 2016). A huge base of clients, large volumes of transactions, and big data analysis have been seen as Alibaba's major competitive advantages.

Kshetri (2016) also looked at the role of big data in the lack of credit information and transaction costs. Digitising activities between the lender and the borrower

are said to reduce transaction costs, and various data sources are said to reduce the lack of credit-related information. This has the potential to improve SMEs and poor people’s access to credit.

According to the discussion above, banks and financial institutions in emerging economies like China lack the data, information, and capabilities to provide financial services to low-income families and microbusinesses. As a result, this group's creditworthiness cannot be assessed, not because they are uncreditworthy (Kshetri, 2016).

Yuan et al. (2022) conducted an experimental study on a default prediction model using k-means clustering and support vector domain description (SVDD) techniques for creditworthiness calculations in an effort for banks and other financial institutions to make better lending decisions. The study focused on Chinese listed companies rather than individuals; their hybrid model is able to predict default for a period of 5 years. Yuan et al. (2022) research framework starts with a double feature selection model for the removal of redundant features to improve the performance of the default prediction model, with 80% training data and 20% test data. Key features or variables that resulted from the feature selection processes include “total assets/earnings," “audit opinion type,” and "revenue/expenses”.

The authors managed to test the superiority of their hybrid model (with the optimal number of clusters for k-means and kennel value for SVDD) over a single model. They tested the models using K-means clustering only and another test using SVDD only and found the performance to be lower for both scenarios. AUC, G-mean precision, and lower Type-II error techniques were used to measure the performance of the model, resulting in an AUC greater than 0.85, which proves high performance.

Pérez-Martín et al. (2018) also conducted an experimental study using big data techniques, where credit risk was assessed for home equity loans. According to the authors, credit scoring refers to any credit rating system that enables the

automatic evaluation of the risk connected to credit-based banking products such as credit cards, vehicle finance, home loans, personal loans, etc. Numerous customer and credit characteristics, including solvency, credit type, maturity, loan amount, and additional factors that are part of financial operations, may affect this risk.

Eight BDA techniques were compared to find the best model for credit risk evaluation. By simulating random data sets and training and testing the various models accordingly, Pérez-Martín et al. (2018) found the linear mixed model to have the best performance from a credit risk evaluation and average time of prediction perspective.

A reliable method for estimating risk would result in fewer anticipated losses, increased company profit, and perhaps more credit loans. The potential profits of a company increase with the amount of credit it extends, which has a positive impact on company performance. Pérez-Martín et al. (2018) propose the use of BDA techniques to segment risk groups as the scientific and traditional credit risk methods are insufficient in predicting customer payment behaviour, excluding other potential customers from accessing credit. A method of approving credit that is impartial and independent of the analyst's judgement.

Hasan et al. (2020) conducted a study on the influence of big data in finance, focusing on internet banking, financial service management, and financial markets. They found that big data is linked to lending companies, banks, online marketplaces, and financial markets, which generate millions of pieces of data from daily transactions, demographics, and customer accounts. These large volumes can be processed to predict customer preferences based on previous purchases and credit risk.

Hasan et al. (2020) then examined how big data analytics is impacting the creation of value at an online credit service provider, which essentially includes banks and FinTech companies, and how they provide their services as a result of tech advancements. According to Hasan et al. (2020), the efficiency of risk-

based pricing and risk management can be improved through big data while alleviating information asymmetry problems. They further stated that Internet credit companies are also able to assess customers without a good financial status through the use of BDA. Fraud detection is also one of the applications of big data that was mentioned.

In Hasan et al. (2020), Kshetri (2016), and Pérez-Martín et al. (2018), customers stand to benefit from credit risk analysis through BDA. As a result, one can argue that there is a positive relationship between customer experience and credit risk analysis through BDA. It leans more towards new customer acquisition, as this method of credit analysis mostly caters to customers that do not have access to credit. Given that the subject of credit risk assessment and its relationship with customer acquisition from a customer experience perspective is fairly new, as such, reliable research instruments will need to be developed to test this construct. As a result, this topic will not be tested in this research paper.

2.2.2 *Sentiment Analysis And BDA*

Rodriguez-Ibañez et al. (2023) found that sentiment analysis is a rapidly growing research area, with most publications focusing on Twitter. The high interest in this field is attributed to the easy availability of data on social networks. The authors found that in the last five years, more than 2400 patents pertaining to social network sentiment analysis have been issued, primarily by large tech companies like Google, Twitter, and Facebook.

Rodriguez-Ibañez et al. (2023) analysed sentiment analysis methods and found that the feature extraction technology that stood out was the "Embeddings" technology. Natural Language Processing (NLP) is the most popular, followed by transformer-based and embedding technologies. Recent methods include auto-regressive and large-language models. Deep learning and machine learning are the most common processing technologies. They also emphasise causality in

sentiment analysis on social networks, e.g., an increase in prices may cause an uproar within social media.

Birjali et al. (2021) used sentiment analysis to extract opinions and sentiments from texts using natural language processing (NLP), with the aim of classifying sentiments as positive, negative, or neutral. Applications include business intelligence and recommendation systems. Business intelligence is mostly used for product and service improvements based on customer feedback resulting from the perception of a product or service. Emotions like anger, happiness, fear, sadness, and surprise are analysed using the NCR emotion lexicon.

Birjali et al. (2021) outline a sentiment analysis process that includes data extraction, text pre-processing, feature extraction, feature selection, sentiment classification, and output presentation. This method can be applied to various sources like social media, review websites, blogs, and interview scripts, removing grammatical and spelling errors and irrelevant features. Similar to Rodriguez-Ibañez et al. (2023), the authors categorise sentiment analysis into machine learning, dictionary-based, and hybrid approaches, with machine learning being the most common. Challenges include spam detection, negation handling, and sarcasm detection. Evaluation metrics like accuracy, precision, sensitivity, F1 score, and specificity are used to measure model effectiveness.

An experimental study on sentiment analysis was carried out by Botchway et al. (2020) on a sizable dataset of tweets involving a well-known bank in sub-Saharan Africa. As the definition of competitive advantage has shifted from purely competing on price and product to developing loyalty and trust, their strategy showed that in today's empowered customer environment, businesses need to focus on customer engagement to improve the customer experience via social media channels (such as Twitter). Therefore, one can infer that there is a positive link between customer experience and sentiment analysis.

Askar et al. (2022) conducted a study that examined how banks can use social media to increase customer engagement and highlighted the opportunities that

social media can present to financial institutions. Given the intense competition within banking, their main aim is to maintain and grow their customer base. According to the author, they need to achieve their goal by providing personalised service and products that consider each client's preferences and investment habits. They proposed the use of social CRM (i.e., social media analytics and CRM) to acquire the necessary information on target customers.

Askar et al. (2022) developed a strategy for the utilization of social media to improve client experience. Outlined in Figure 3 below.

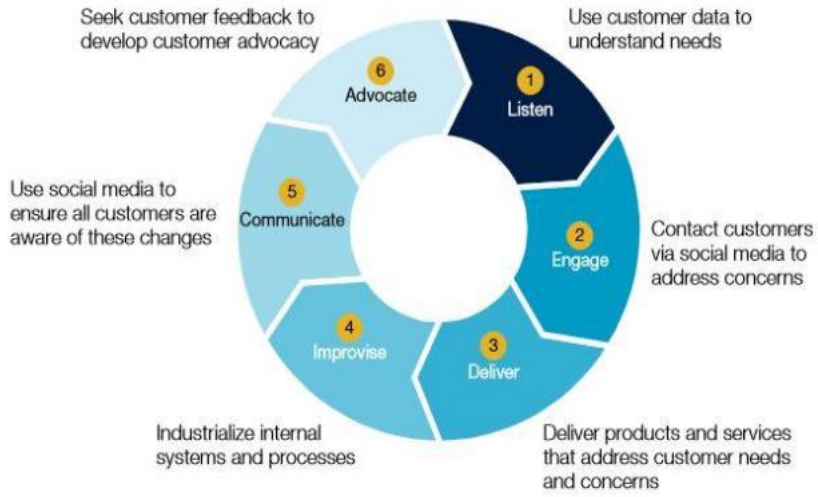


Figure 3: The strategy for the utilization of social media to improve client experience.

Source: Askar et al. (2022)

Additionally, Askar et al. (2022) propose that banks use a combination of other customer data (i.e., spending habits, investment preferences, dependents, etc.) that they collect with social data to create a thorough behavioural picture of the customer. This identifies the needs and preferences of the customer and predicts the products and services that will be needed in the future to meet those needs.

2.3 Primary research question.

What is the impact of BDA applications on customer experience for customer retention and new customer acquisition within the South African retail banking industry?

2.4 THEORETICAL FRAMEWORK

2.4.1 *Theoretical Framework*

In an effort to provide insights on whether big data can improve customer retention and the acquisition of new customers through an improved customer experience, this study uses the Customer Focus Theory (Lohan et al., 2011) to develop multiple hypotheses.

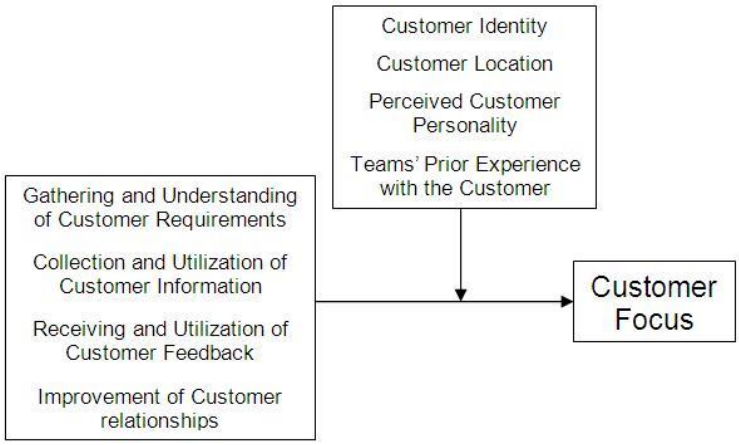


Figure 4: Customer Focus Theory

Source: Lohan et al., 2011

According to the theory, businesses should constantly monitor customer needs and preferences to spot fresh opportunities for development and innovation. The ability of an organisation to collect and analyse customer data is crucial to the customer focus theory, which is the very thing that BDA enables. For this study,

BDA is considered a resource that an organisation acquires over time through its BDA capabilities to acquire, analyse, and use customer information. An organisation can use its BDA capabilities to learn about its customers' needs and respond accordingly. Big data analysis can be seen as a continuous and disruptive synthesis used to find patterns and link them to potential corrective actions (Ghasemaghaei and Calic, 2020).

In a variety of contexts, including marketing, product development, and customer service, customer focus theory has been used. The responses can be in the form of new products and services that meet specific customer needs, credit worthiness assessments through the use of big data, alternative methods of fraud detection to protect customers online and in POS transactions, customer sentiment analysis, etc., which are ultimately used as a source of differentiation.

Customer Retention theory

Reichheld (1996) argues that customer loyalty is driven by a combination of loyal employees and investors supported by a good value proposition, which in turn leads to profitability.

Arnold et al. (2010) look at both customer acquisition and retention orientations and their impact on an organisation's innovation performance. According to the authors, "customer acquisition orientation" is used to describe a company's emphasis on gathering data on potential clients, estimating their potential value, and allocating resources to hire those with higher long-term value. Customer retention orientation involves putting an emphasis on gathering data, identifying key players, and investing in relationships with current customers in light of their long-term value. Arnold et al. (2010) argue that the depth (i.e., amount of knowledge an organisation has about its customers) and diversity (i.e., variety and range in customer information) of customer knowledge increase with an organisation's increased focus on customer acquisition and retention due to the culture and leadership within the organisation, which positively increases innovation performance.

2.5 Hypothesis development And First research question

2.5.1 *Customer experience and BDA applications*

- To what extent should South African retail banks use BDA for fraud detection, and the determination of customer needs through customer segmentation to develop personalised products, services, and engagements to improve the customer experience?

Holmlund et al.'s (2020) study based on customer experience management (CXM) in the age of BDA stemmed from the realisation that research on the importance of BDA for the improvement of customer experience (CX) is in its infancy stages.

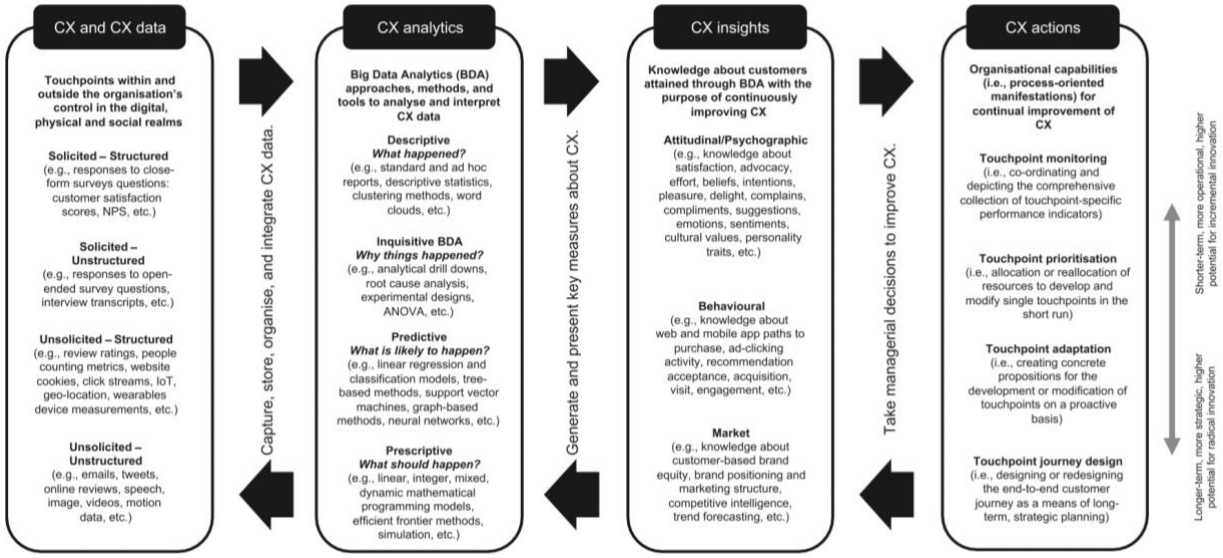


Figure 5: A strategic framework for CXM based on CX insights resulting from BDA.

Source: Adopted from Holmlund et al. (2020)

Holmlund et al. (2020) developed a framework grounded on data obtained from different touch points that a customer has with the organisation (e.g., data generated when a customer interacts with the organisation's physical channels,

web/digital channels, and/or social media) to generate actionable insights. CX data is also collected from customer satisfaction scores, Net Promoter Scores, surveys or interviews, review ratings, clicks, website cookies, tweets, speeches, emails, and complaints. The data is then analysed using Business Data Analysis (BDA), followed by implementing CX actions based on the insights. BDA methods yield descriptive, inquisitive, predictive, and prescriptive analytics, providing insights for improving CX. According to the authors, CX actions involve touchpoint monitoring, prioritisation, adaptation, and journey design based on BDA insights. Piloting decisions include success measures, deployment, scaling, and monitoring for issues.

Anshari et al. (2019) highlight the importance of big data in customer relationship management (CRM) strategies, particularly for service personalisation and customization. They propose social CRM, a technique for identifying customer behaviour patterns across multiple channels. This method allows businesses to target their services to customers' interests and preferences, leveraging user-generated content on social media. As a result, they developed a framework for CRM and BDA, analysing unstructured and structured data to provide customer profiling based on demographics, facial recognition, online activities, geolocation, social network activities, and e-commerce purchase history. This enables CRM solutions to provide personalised, customised, quick responses, and multichannel interactions, providing competitive value.

Anshari et al. (2019) highlight the benefits of using big data in CRM, including accurate customer profiling, trend analysis, personalised marketing, and improved customer service. They provide real-life examples of how companies like Starbucks and Amazon.com use social media to gather customer feedback and identify patterns in shopping behaviour for better deals and discounts.

Based on the findings of their study, Delgosha et al. (2019) recommend that banks consider investing in BDA applications in risk and fraud and operational or customer analysis categories in order to benefit from large amounts of internal and external data.

Tseng (2023) conducted a quantitative empirical study on an organisation’s sensing and responding capabilities to customer needs through big data analytics. The study was done by interviewing 116 managers from different organisations across the world in order to examine the proposed model. They tested a few hypotheses, the first being to test the effectiveness of big data analytics to create better sensing capabilities. Sensing capabilities were defined as the ability of an organisation to understand its customers' needs and wants. Their second hypothesis was to test if the effective use of big data analytics positively impacts customer response capabilities, which are defined as an organisation's ability to respond to market demands and offer products or services as required by customers. They further tested whether both the sensing and response capabilities have a positive impact on new product success.

Their findings suggest a beneficial relationship between the two studied capabilities and the efficient application of big data analytics. As the success of new products is positively impacted by the customer's sensing capability, this positively affects the performance of the company.

The above literature on BDA applications focuses on customer experience improvement, understanding customer behaviour and segmentation, and enabling new product and service development. Sentiment analysis is crucial for organisations to listen to customer opinions on social media, as customers can significantly impact a brand (Birjali et al., 2021; Botchway et al., 2020; Askar et al., 2022). By utilising insights from social media platforms, organisations can enhance the customer experience across all touchpoints, ensuring relevance and meeting customer needs.

Anshari et al. (2019) and Holmlund et al. (2020) indicate that there is a positive relationship between big data analytics and customer experience. Delgosha et al. (2019) highlight various BDA applications within the banking environment that directly relate to customers. Because of this, one can hypothesise that there is a positive relationship between various BDA applications and customer experience and satisfaction.

H1: BDA applications improve customer experience.

2.6 Hypothesis development and Sub research questions

2.6.1 *Personalisation of products, services, and customer engagements through BDA and Customer Experience*

I. Sub question 1

- Will the customer experience improve because of utilising BDA for the personalisation of products, services, and customer engagements?

Personalised Engagements

An experimental study was conducted by Hosseini et al. in 2022 to examine customer behaviour in traditional and online banking. The approach was used to segment customers, better understand each customer segment, and enhance customer engagement through the bank's CRM. It was found that customers that use both traditional banking services and electrical channels have the highest Customer Lifetime Value (CLV) and, as a result, are considered the most valuable to the bank. Hosseini et al. (2022) argue that strategies to target and maintain these customers need to be developed.

It can also be argued that the bank should also target low-CLV customers in an effort to improve their CLV through enhanced CRM activities in order to provide the right product to the right customer. Customer segmentation is most likely the most effective way to retain and attract new customers, as a result of a bank's better understanding of customer behaviour and effective communication.

Kovács et al. (2021) conducted an experiment to segment customers in the investment space using machine learning clustering. Results showed that affluent customers prefer a balanced portfolio with low risk and high yield; those with little savings prefer a government bond portfolio; and those with significant savings do not want an investment portfolio. Kovács et al. (2021) concluded that the machine

learning two-stage clustering method can effectively cluster customers according to behavioural patterns and factors.

According to Kovács et al. (2021), this method can effectively be used by banks to cluster their customers for any of their product offerings. Patterns are generated from customer information and, as a result, can be used to make recommendations on products and services that customers might be interested in. This would result in a bank’s ability to customise product and service offerings based on customers’ personal needs.

Cuomo et al. (2023) utilised predictive and exploratory analysis to analyse Italian households' savings behaviours using machine learning models. They identified demographic and environmental factors such as age, gender, location, household income, consumer confidence, employment status, education, financial knowledge, habits, personality traits, and physiological habits as influencing factors. Digital knowledge also played a role in these variables. Cuomo et al. (2023) utilised K-means to analyse data from 20.1 million Italian savers, creating 12 clusters based on customer profiling variables like age, financial product type, assets, and active accounts.

Cuomo et al.'s (2023) study found that customers tend to repeat purchases and exhibit similar behaviour within clusters. The study also revealed that customers who have historically closed their accounts tend to reduce their assets, redeem funds early, and close their accounts. The authors suggest targeting each cluster differently to recommend relevant products and actions for portfolio growth.

Abedin et al. (2023) conducted an experimental study using feature engineering and classification techniques to model bank customer behaviour from 24,000 active and non-active datasets. They proposed various feature selection techniques to select the best features and remove redundant ones. The Random Forest ML classification method was found to be the best performing, with 73.79% accuracy, demonstrating the importance of feature engineering in knowledge mining.

Egorova et al. (2022) conducted an experimental analysis study focusing on bank customer transaction behaviour analysis using a knowledge graph embedding method. Historical transactional data was used to predict the customer's next purchase at a store. K-means was used for clustering customers based on their behaviour based on the number of purchases per purchase category (i.e., customer spend on beauty, clothing, food, fuel, housing and maintenance, restaurants, pharmacies, supermarkets, and transport). It was also found that most customers bought from the same store.

Egorova et al.'s (2022) study identified customer clusters, including those who prefer dining out and buying beauty products without spending on petrol, those who travel and spend on beauty, and those who shop at home stores. Based on the above examples of clusters, one can thus argue that it's possible to target each cluster with product offerings related to their needs instead of following an umbrella approach.

Abedin et al. (2023) argue that banks can use their model to understand their customers' demands, identify risk, and predict factors associated with customer behaviour to minimise customer churn. The author further argues that the same techniques can be used for financial fraud detection, loan default prediction, and marketing. Businesses are utilising cutting-edge technologies to gather customer-centric data as they develop, which will help them create offerings that are most in line with the requirements and preferences of their customers (Gupta et al., 2020).

Marketers can take advantage of new technologies to tailor advertisements to consumers based on their personal information, location, and past online navigation, thanks to the growing popularity of GPS-enabled online communication tools and user experience monitoring practices (Lamprinakos et al., 2021).

Personalised Product and services

Tseng et al. (2022) studied customer agility and big data analytics within the new product context. The findings show that new products will succeed if data interpretation tools are used correctly. These big data analytics tools are capable of revealing business insights and trends. Tseng et al. (2022) further explain that feasible product improvement opportunities can be identified through systematic and comprehensive near-real-time or real-time reporting from data analysis tools. This, as a result, will improve an organisation's customer sensing and response capabilities, which enable customer agility. Determining additional customer needs can also be improved through customer sensing and response capabilities.

Xu et al. (2016) conducted a systematic review of the effects of traditional marketing analytics and big data analytics on the success of new products. The authors proposed that with the generation of large amounts of data that give organisations insights into customer behaviour and preferences, traditional marketing techniques ought to be fused with BDA techniques to promote new product success (NPS). NPS is said to be measured through revenue projections, cost estimates, profit projections, risk evaluations, and finally, the potential for sale cannibalization.

According to Xu et al. (2016), BDA offers businesses strategies for creating goods that customers are more likely to purchase. BDA allows businesses to monitor competitors, observe customers, conduct surveys, test prototypes, and gather feedback. It helps in understanding competitors' product features, pricing strategies, and customer feedback. Product managers can use real-time data for product development. BDA also aids in automated knowledge management, web analytics, search analytics, Search Engine Optimisation (SEO), and customer analytics.

The authors developed a knowledge fusion strategy/ taxonomy as per Figure 7 below.

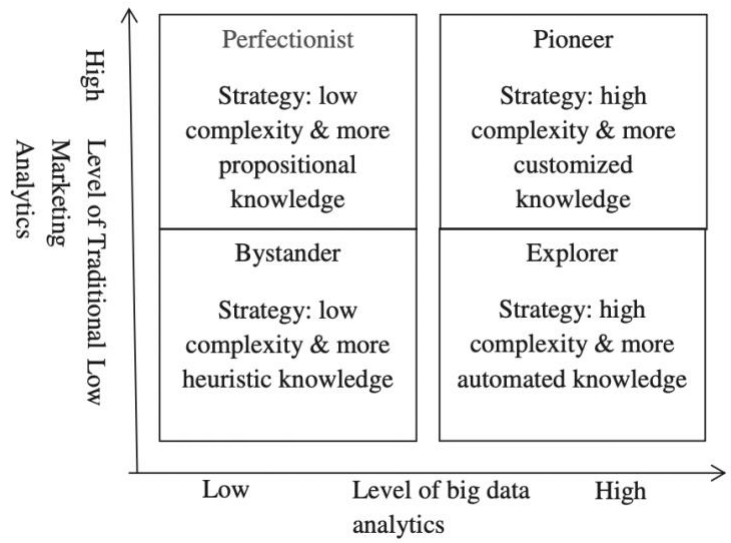


Figure 6: knowledge fusion strategy/ taxonomy.

Source: Adopted from Xu et al. (2016)

Xu et al. (2016) define a "bystander" as an organisation that is both low on traditional marketing and BDA and, as a result, will have low New Product Success (NPS) due to having low knowledge fusion. "Explorers" and "perfectionists" are considered to have medium knowledge fusion, resulting in medium NPS, as they have adopted either high levels of BDA or traditional marketing analytics. An organisation that adopts high levels of both BDA and traditional marketing analysis is considered a "pioneer," which is most likely to have the highest knowledge fusion and consequently the highest NPS.

Ghasemaghaei and Calic (2020) explored the impact of big data's main characteristics and their impact on the firm’s innovation performance, with a focus on product innovation. The performance includes financial returns, customer perspective, and operational excellence. "Using big data to identify patterns and connect them to potential corrective actions can be thought of as a continuous, disruptive blend of abduction, deduction, and induction." Ghasemaghaei and Calic (2020, p. 149).

Ghasemaghaei and Calic's (2020) study reveals that the most important data for organisations to learn from is a large variety of real-time data in order to generate new insights. Customer needs can be better understood from both structured and unstructured data and different sources. including previous purchase data, web data, social media data, etc. Firms can develop products that fit customer preferences with reduced time and effort as a result of collecting a variety of data from various sources and, as a result, improve their financial returns and customer perspective.

Lamprinakos et al. (2021) found that when people have mainly positive brand-related thoughts in response to an advertisement (also known as cognitive responses), this leads to more positive brand evaluations and increases purchase intentions for products linked to the relevant brand.

One can argue that product success, whether new or old, is determined by high levels of customer satisfaction. Xu et al. (2016) have indicated the need for BDA to understand customer needs. Tseng et al. (2022) emphasise the need to have better sensing capabilities to meet customer needs in the development of products. One can also conclude that customer segmentation can be used to understand customer behaviour and engage with customers in a personalised way, which in turn can be used to improve the customer experience in order to reduce customer attrition and improve loyalty. Thus, my hypothesis is that there is a positive relationship between customer experience and the personalisation of products, services, and customer engagements.

H1a: Personalisation of products, services, and customer engagements through BDA improves customer experience.

2.6.2 *Fraud Detection through BDA AND Customer experience*

II. Sub question 2

- What is the relationship between fraud detection and customer experience?

The digital transformation in the financial services industry has accelerated the development of more intelligent fraud schemes that also require intelligent means to detect fraud (Karpoff, 2020). Mckinsey & Co. (2022) highlight the rise of sophisticated fraud as more people use digital devices. Businesses must address this while maintaining a seamless digital experience. Techniques like phishing, malicious software, social engineering, deep fakes, and exploit kits for fraud-as-a-service are commonly used. Some of the frauds that are closely related to banks include mortgage fraud, loan defaults, credit card fraud, and money laundering. The three broad categories of data used for financial fraud detection are basic quantitative structured data, diverse semi-structured data, and complex unstructured data. These data types are constantly changing (Zhu et al., 2021).

Zhu et al. (2021) conducted a systematic review on the topic of machine learning techniques, highlighting the development of deep learning and neural network classifiers. These techniques extract statistical features from user profiles, credit histories, and historical transaction data to train classifiers. However, they face challenges in understanding complex and covert fraud activities, as well as model flexibility, robustness, and interpretability. Despite these challenges, these data-driven ML models have proven useful in fraud detection.

McKinsey & Co. (2022) found that businesses lose sales due to fraudulent transactions, with false-positives accounting for up to two-thirds of declined sales. Fraud also harms customer trust and willingness to use services, making transactions difficult and increasing anxiety. A McKinsey survey found that 70% of banking customers experienced stress, anxiety, anger, or frustration when warned about potential fraud. Mckinsey & Co. (2018) recommends data analytics

as a fraud detection method, e.g., using similar analytics and typical use cases, bank transactions should be risk-scored based on timing, destination, source, value, frequency, device, and geolocation intelligence.

Leading companies employ machine-learning algorithms and make use of all available data to improve the precision of fraud detection. They aim to lower the noise (false positives) and the possibility of missing fraudulent transactions (false negatives). If not reduced, false negatives and positives lead to a bad customer experience and hinder their ability to transact (Mckinsey & Co., 2022).

Shahana et al. (2023) conducted a systematic review of fraud detection on financial statements. Data mining, statistical analysis (regression), and machine learning techniques have been found to be the most popular in fraud detection, with less emphasis on natural language processing. with supervised learning methods being more popular than unsupervised learning. A good fraud detection model has a huge dependency on the data types and sources used; these include financial and non-financial variables as well as text data.

Shahana et al. (2023) also touched on the importance of evaluating the performance of the fraud detection model through the use of various evaluation metrics. A fraud detection model may result in true positives (where a statement is correctly classified as positive, i.e., fraudulent), true negatives (correctly classified as not fraudulent), false positives (falsely classified as fraudulent), and false negatives (falsely classified as not fraudulent). False predictions represent a prediction error.

Gupta et al. (2023) conducted an experimental study on credit card fraud detection on unbalanced data using machine learning techniques. The experimental procedure started with the pre-processing process, followed by feature selection, model training, and lastly, the evaluation of the model's performance. The pre-processing and feature selection stages remove data attributes that may not add value to the ML model; they are necessary for the removal of inconsistent, partial, duplicate, and incorrect data.

The classification techniques (used to cluster fraudulent transactions against non-fraudulent transactions) used in the study include logistic regression, decision trees, and the XGBoost classifier, which is the best-performing fraud detection model with an accuracy greater than 99%. According to Gupta et al. (2023), a very high accuracy implies that the data used is unbalanced (the training data consisted of only 0.17% fraudulent data). data balance methods such as oversampling and under-sampling, after which the accuracy score was still high at 0.998. Based on the findings of the study, one can argue that a study with more balanced input data is required.

Jurgovsky et al. (2018) also conducted an experimental study on credit card fraud detection using the "sequence classification technique," covering both online (e-commerce) and face-to-face transactions (point of sale). According to the authors, financial institutions and service providers are facing a significant financial challenge because of the rising volume of electronic payments, which forces them to continuously enhance their fraud detection systems.

Nevertheless, despite being widely used in other fields, contemporary machine learning-based methods are still making slow progress in business applications. For accurate classification and overall model performance when building a credit card fraud detection system, features must be carefully chosen, known as "feature engineering." Features identified by Jurgovsky et al. (2018) include time, transaction type, entry mode, amount, merchant group, country (merchant), country (card holder), and card type. Classifier models (i.e., Long Short-Term Memory Network (LSTM) and Logistic Regression) were trained and tested for each feature set, data set, and sequence length.

Zoldi (2015) suggests that businesses can differentiate between significant demographic changes and temporary ones, like travel, by combining real-time insights. This will ensure customers are not prevented from using their cards until fraud is evident. To understand customer demands, financial service providers should monitor Net Promoter Scores and enhance their fraud strategies.

Additionally, customers can be informed about fraud using their preferred channel for real-time transaction approvals, enhancing the customer experience.

As a result of the above literature, one can argue that fraud detection may be used to protect both the customer (e.g., the use of stolen identity information to apply for credit or the use of stolen credit card information to make unauthorised transactions) and the organization. In the studies above, it was found that there is a positive relationship between fraud detection and customer experience.

H1b: Fraud detection through BDA improves customer experience.

2.7 Hypothesis development And Second research question

Will the use of BDA to improve the customer experience result in customer retention and new customer acquisitions?

2.7.1 *Customer retention and new customer acquisitions because of Improved customer experience*

Mikalef et al.'s (2019) study looks at big data analytics and its contribution to a firm's performance. The study contends that while the literature suggests that big data analytics has a positive impact on business, little is known about the organisational factors to take into account when attempting to get a return on big data analytics investments because big data capability implementation presents many difficulties for organisations. Rialti et al. (2019) also argue that the complex relationship between BDA and organisational impact is still being understood by scholars and that part of the research is still theoretical or qualitative at most, with quantitative studies still in their infancy.

Mikalef et al.'s (2019) research framework consists of seven elements, including data, process, technology, organisational, people, and context, whose level of maturity determines the performance gains derived from big data analytics. The study uses a multimethod approach, including a survey-based quantitative

approach and semi-structured interviews, to identify key factors driving performance from big data analytics. Three case studies reveal that a well-defined big data analytics strategy is crucial for business performance. The strategy should include a clear roadmap and deadlines for leveraging big data for value. The study also highlights the importance of resistance to change and the implications of ethics and legislation for successful big data analytics implementations.

Mikalef et al. (2019) conclude that big data analytics needs to be in line with the company's business strategy and should be seen as an organisational challenge rather than a technical one. The context within which an organisation operates, and, as a result, different factors must be emphasised in accordance with the firm’s context and environment. The authors conclude by highlighting organisational structure, size class, the industry, and top managers' support as factors that determine how big data analytics strategies are created and carried out.

Wamba et al. (2017) examines factors contributing to firm performance as a result of big data analytics implementation by looking at how big data capabilities are measured and how they’re linked to the performance of a firm. They also examine the relationship between big data analytics capabilities and firm performance and whether process-oriented capabilities play a role in the relationship. The study draws from previous literature on big data analytics capabilities and the resource-based view theory, where big data analytics is argued to be a resource that a firm acquires in order to create a competitive advantage.

The study used a survey questionnaire to examine big data analytics infrastructure flexibility, management capabilities, personnel capabilities, process-oriented dynamic capabilities, and firm performance-related questions. Results showed a causal relationship between these capabilities and firm performance, with dynamic process capability (PODC) being a critical mediator. Personnel expertise capability was identified as the most important, requiring organised efforts to develop technical, business, management, and relational

skills related to big data analytics. Improving PODC and BDAC (Big Data Analytics Capability) is crucial for enhancing firm performance and competitive advantage (Wamba et al., 2017).

To address the gap on the impact of BDA capabilities on an organisation's performance, Rialti et al. (2019) conducted a study to understand the relationship between BDA and organisational performance with organisational ambidexterity (i.e., the firm's ability to use its existing resources to adapt to changes while seeking new opportunities) and agility as mediators. The study focuses on large organisations, as they're said to have large data sets and can invest in BDA despite the costs. Organisational resistance to the implementation of BDA technologies and the fit of the BDA technologies with the organisation played a moderator's role. The framework developed allowed several hypotheses to be tested.

The most important finding from the study is that organisational performance is positively impacted by ambidexterity and agility derived from BDA capabilities; this allows firms to become more dynamic and responsive to change. As a result of their findings, Rialti et al. (2019) advise that firms should invest in BDA infrastructure as well as the people to promote the organisation's ability to exploit big data. The authors argue that with the ability to exploit insights from customers, rivals, and manufacturers, organisations can make informed decisions, giving them a competitive advantage.

Fernandes and Pinto (2019) conducted a quantitative study with the aim of understanding the impact of customer experience on relationship quality and how that affects loyalty-related outcomes such as customer retention, word of mouth, and tolerance (i.e., the willingness of customers to overlook minor errors, even price increases) in retail banking. The authors tested the validity of four antecedents of customer experience, including the environment (i.e., bank branches), frontline personnel, moments of truth (i.e., impactful customer interactions with a brand where flexibility and speed, in being receptive to customers' concerns, are particularly important (Klaus and Maklan, 2012), as

quoted in Fernandes and Pinto (2019), and product offerings. One can infer that BDA-enabled engagements play a role in Moments of Truth and product offerings.

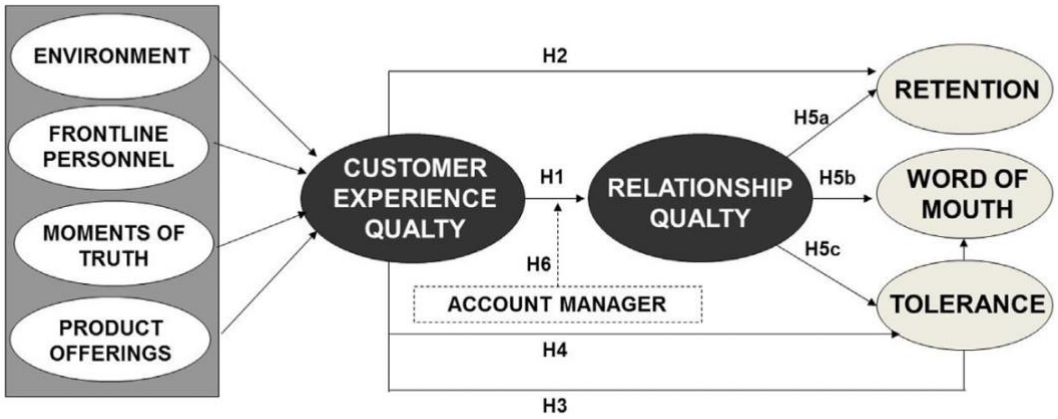


Figure 7: Customer experience and customer loyalty relationship

Source: Fernandes and Pinto, (2019)

Fernandes and Pinto's (2019) research findings indicate that there is a strong relationship between customer experience and customer retention, word of mouth, as well as tolerance for relationship quality as a mediator. They found the relationship to be particularly strong for customers with dedicated account managers.

Radia et al. (2022) conducted a quantitative study expanding on Fernandes and Pinto's (2019) study with a focus on physical comfort (PC), which is said to reflect a customer’s mental clarity and sense of ease during a service interaction, which supports the maintenance of long-lasting relationships.

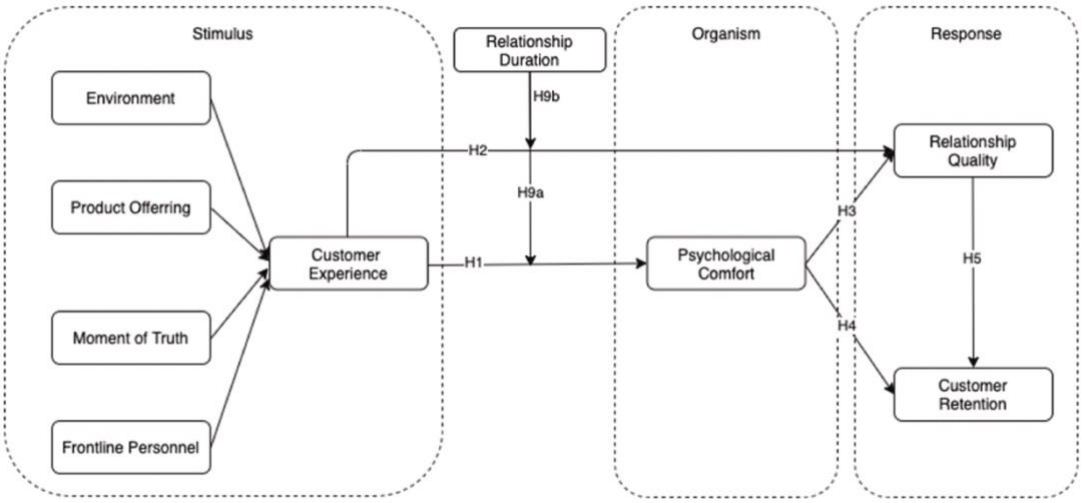


Figure 8: Customer experience and customer retention with psychological comfort as a mediator.

Source: Radia et al. (2022)

The authors found that customer experience has the greatest and most beneficial effect on PCs because it is successful in creating a favourable emotional state in customers. Out of the 4 antecedents and dimensions of CX, they found "Moment of Truth" to have the most impact on customer experience as it is associated with the bank's ability to address customer complaints and improve customer experience quickly and flexibly.

In line with Fernandes and Pinto's (2019) findings, Radia et al. (2022) also found customer experience to have a significant influence on relationship quality. The results show that between customer experience and relationship quality and between customer experience and retention, physical comfort has a mediating effect. These findings imply that for retail banks to increase customer retention, they need to focus on ensuring PC. They also concluded that customers with long-standing relationships with a retail bank are more likely to be retained.

Ascarza et al. (2017) define customer retention as “the customer continuing to transact with the firm." The authors suggest that customer satisfaction is one of the predictors of customer churn. Lamrhari et al.'s (2022) study focuses on the

improvement of customer acquisition, retention, and conversion. As a result, they developed a social CRM framework that uses machine learning and sentiment analysis, which they argue will increase customer experience.

According to the authors, social CRM consists of a series of procedures that enable decision-makers to evaluate consumer data in order to initiate a marketing plan that is effective, affordable, and focused on the needs of the consumer. They argue that a social CRM seeks to not only retain existing clients but also acquire new ones through social media customer engagements and collaborations through electronic word-of-mouth (eWOM).

H2: Improved customer experience resulting from BDA applications has a positive impact on customer retention and new customer acquisitions.

2.7.2 *Conceptual Framework*

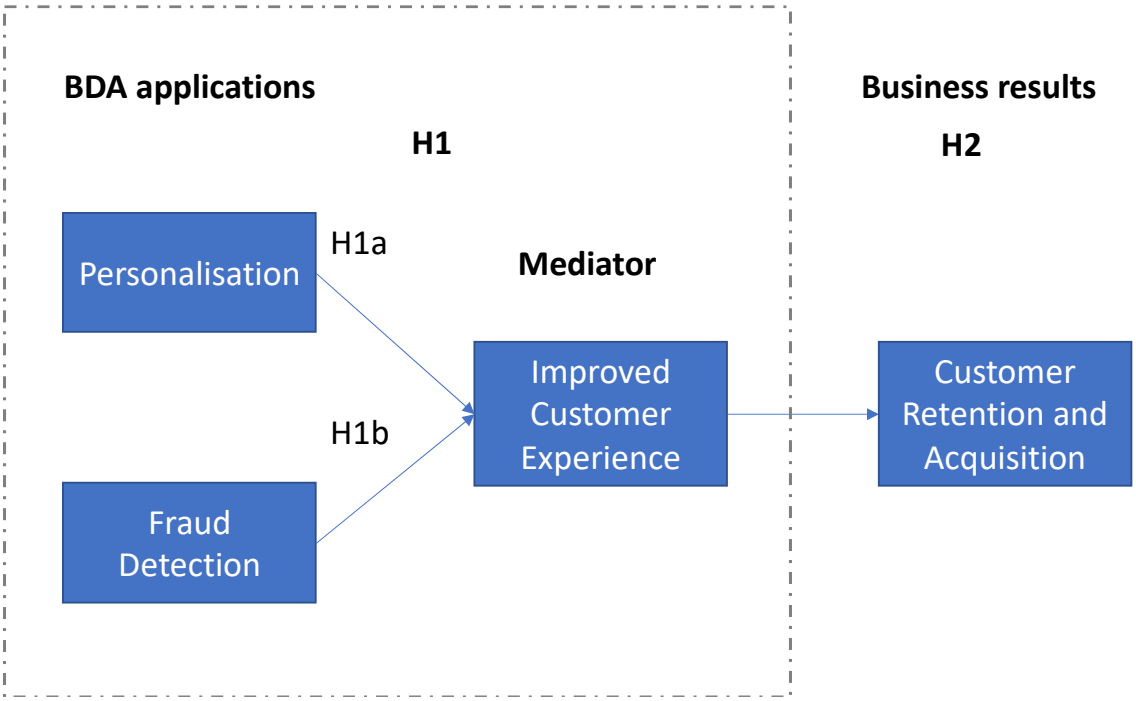


Figure 9: High Level conceptual framework.

The above conceptual framework aims to test a relationship of causality between BDA applications and business results in the form of customer retention and the acquisition of new customers, with customer experience as a mediator within the South African retail banking sector.

I. **H1: BDA applications improve customer experience.**
   a. H1a: Personalisation of products, services, and engagements through BDA improves customer experience.
   b. H1b: Fraud detection through BDA improves customer experience.

II. **H2: Improved customer experience resulting from BDA applications has a positive impact on customer retention and new customer acquisitions.**

CHAPTER 3. RESEARCH METHODOLOGY

This section outlines the approach that was taken to address the hypotheses that emerged from the literature review, which are supported by the conceptual framework that was proposed as a potential answer to the research questions. We discuss the research approach, research design, data collection technique, target population, and sampling technique in this chapter. The research instrument used is also presented in this section, as are the data analysis strategies, quality assurance methods, and ethical considerations that were taken during the study.

3.1 Research approach

In this research, the positivist paradigm was used, which assumes generalizability. In an effort to test if Dig Data analytics factors can be used to predict customer retention and acquisition in banking, a quantitative method was used for this study in order to test the hypotheses and relationships between constructs within the South African context. The reason why this method was deemed appropriate is because the BDA topic seemed to be well researched, clarified, and understood, so a more structured approach was appropriate (Kumar, 2014). As a result, the aim was to investigate relationships between the analysed constructs.

3.2 Research design

Given that the study is meant to investigate whether a change in one variable will cause a change in the other variable (Kabir, 2016), a research questionnaire in the form of a Likert scale survey was put together for use as a research instrument. Qualtrics survey software was used to create and distribute the survey online, with the link being shared via email or WhatsApp. The reason for

using an online survey is that the study would cover a large sample in a quick and cheap manner and achieve a greater level of anonymity (Sue and Ritter, 2012).

Kumar (2014) has outlined some of the disadvantages of an online survey that were taken into consideration during data collection. These include a low response rate, no opportunities for clarity, the influence of responses by others if the respondent consults with them, etc.

3.3 Data collection methods

An online survey was distributed among digitally active bank customers within South Africa. For each hypothesis that needed to be tested, questions that needed to be answered through the study were formulated. According to Kumar (2014), the scale is said to demonstrate the relative strength of one respondent's opinion in light of another's, rather than their overall attitude. The attitudinal scores pertaining to each question will be used to assess relationships between the variables being tested (refer to the framework, Figure 1).

3.4 Population and sample

3.4.1 *Population*

Wagner et al. (2012) define a population as the full cluster of individuals from which data can be collected. The target population for this study was digitally active bank customers. The reason for choosing digitally active customers is that they are the ones that will be impacted the most by BDA applications. A report by PWC for 2023 indicates a gradual increase in digitally active customers, up 18.4% from the 2021 financial year, currently sitting at 17 million customers. They further define a "digitally active customer" as one who uses one or more of a bank’s digital channels.

3.4.2 *Sample and sampling method*

A non-probability sampling method was used. Given that the focus was specifically on digitally active bank customers, a purposive sampling method was used in combination with the snowball method to cater for any difficulties in obtaining responses. To test our hypotheses, data from a sample of customers using digital banking channels was collected through an online survey.

Kass and Tinsley (1979) recommend between 5 and 10 respondents per variable or scale item up to 300 maxima. Given that there are 22 items on the scale, the sample size should be 110 participants at minimum and 220 at maximum. Thus, at least 579 questionnaires should be distributed, based on the usual survey response rate of 38% (Baruch, 1999), as cited by Rialti et al. (2019). For this study, thousands of online surveys were distributed.

3.5 The research instrument.

Data was gathered through a cross-sectional survey, after which the research model was evaluated. For new research questions, Wagner et al.'s (2012) principles of research design were followed, which include the title of the survey, explaining the aim of the research, question formatting, and input from experts. Some questions were adopted from McKinsey and Co.'s survey. A portion of the survey's multi-item scales that were previously published were used in the study (Klaus and Maklan, 2012) with favourable psychometric properties, reporting high convergent validity with t-values > 0.5 and a root mean square error of approximation (RMSEA) of 0.05.

A 7-point Likert scale (1 = strongly disagree, 7 = strongly agree) was used to help in answering each of the research questions covered in this study, with the assumption that each statement is of equal importance (Kumar, 2014). The research instrument consists of four research constructs. The personalisation construct has seven items, sourced from McKinsey & Co. (2021) and Fernandes and Pinto (2019). Fraud detection has four items that were all sourced from

McKinsey & Co. (2022). Both the customer experience construct and retention and acquisition have three items each, all sourced from Fernandes and Pinto (2019). The questionnaires for the study are captured in the appendix.

3.6 Procedure for data collection

Data was collected through an online questionnaire accessible through a link that was sent to a sample of bank customers who are digitally active. The link was also sent to thousands of Wits university students through their student emails, as most students are digitally active bank customers, with approval from the registrar's office, and through WhatsApp to non-Wits students.

3.7 Data analysis strategies and interpretation

Once data was collected through the online survey, SPSS 16.0 (Statistical Programme for Social Science) and IBM SPSS Amos 26 (a structural equation modelling tool) software were used to conduct statistical analysis. A descriptive analysis of all the variables was conducted to calculate the mean and standard deviation and understand the responses. Confirmatory factor analysis was also conducted in order to investigate the relationship between constructs and ultimately test the hypotheses.

3.8 Quality Assurance

The concepts of validity and reliability, external validity (i.e., generalisability), and objectivity are mainly used to measure the quality of a quantitative study. According to Kumar (2014), the aim is to determine whether the methods used for answering the research questions were appropriate, good, and accurate. From a research instrument perspective, it's a question of whether it measures what it's intended to measure.

3.8.1 *External validity*

External validity, or generatability, refers to the extent to which research findings can be generalised to different populations, environments, or situations. Kumar (2014) suggests the use of logic (i.e., links between questions and the backing of experts) as well as statistical evidence (i.e., to calculate correlations between variables and research questions) to test the validity of an instrument. In order to determine whether findings from the sample will be generalised, the minimum sample size, as recommended by Kass and Tinsley (1979), of respondents who meet the requirements was met.

3.8.2 *Internal validity*

How well a study can establish causality between variables while minimising confounding factors is known as "internal validity." The research instrument will undergo relevant convergent validity calculations. The aim is to reach high convergent validity with standard factor loadings > 0.5 and a root mean square error of approximation (RMSEA) of 0.05, as recommended. The data received from the questionnaire from a representative sample was used in confirmatory factor analysis to validate the scale, confirming its validity (Hair et al., 1998, as quoted by Klaus and Maklan, 2012; Creswell and Creswell, 2018). The instrument's validity was further assessed using the correlation matrix, average value extracted (AVE), and highest shared value (HSV). To demonstrate the distinction between constructs, for this study, the inter-item correlation matrix was used to assess the discriminant validity of the measurement instrument (O'Rourke & Hatcher, 2013). The AVE was used to measure the degree to which multiple items used in the research to measure the same concept agree; this is known as converging validity (Hair et al., 2010). Discriminant validity was verified by comparing the AVE for each construct with the HSV for each construct (Nusair & Hua, 2010).

3.8.3 *Reliability*

Whether the instrument would reproduce consistent findings if replicated by another researcher or repeated on another occasion is a measure of reliability. The reliability of the scale was calculated through the Cronbach coefficient Alpha (α), as it is said to measure the commonalities and uniqueness of the items or variables being tested by the instrument. The composite reliability (CR) was also calculated for each construct. The aim was to achieve a Cronbach coefficient (α) and composite reliability (CR) greater than or equal to 0.7, indicating 30% error or 70% (Creswell and Creswell, 2018). Errors may include participant errors, participant bias (false responses), researcher error, or bias (in recording responses), etc.

3.9 Ethical considerations

It is critical that one adhere to research ethics and ensure that all participants are protected. Ethical issues, as outlined by Creswell and Creswell (2018), were considered by ensuring that approval is received prior to collecting data. Participants were not pressured into providing consent, the purpose of the study was disclosed, and harmful and sensitive information was not collected. In analysing the data, all the results are being reported. The anonymity and privacy of participants were respected, as the instrument itself did not require identity information. Once the report is compiled, it will be made public.

CHAPTER 4. PRESENTATION OF RESULTS/ FINDINGS

4.1 Introduction

In this chapter, a brief overview of how missing data was solved as well as the reverse coding of negative statements is given. This is followed by the presentation of descriptive statistics. Reliability and validity assessments are then conducted on the instrument, along with a model fit analysis. The final results of the hypotheses that were being tested are then presented.

Missing values

In order to ensure that there aren't any missing values in the data, the expectancy maximisation algorithm on SPPS was used to conduct Little's test to determine if the data is missing at random or not. The results of Little's MCAR test show that the P value or significance value is greater than 0.05 (95% confidence in our estimations), thus the data is missing at random. In this case, P = 0.057, thus > 0.05. Thus, given that the values are missing at random, the most likely value in the dataset was computed through the use of an algorithm (the EM method of computing, given that the number of missing values is less than 20%).

Negative worded items/ questions of likert scale questionnaire

To reverse code negative items, new variables were computed. Given that a 7-likert scale was being used, the responses were reversed such that 7 = 1, 6 = 2, 5 = 3, 4 = 4, 3 = 5, 2 = 7, and 1 = 7.

4.2 Descriptive statistics

The sample of digitally active bank customers consisted of 133 (40.5%) male and 190 (57.9) female respondents, for a total of 328 (N = 328) respondents. 23.8% of those respondents were Gen Z (18–26 years old), 53.7% were millennials (27–

42 years old), and 20.7% were Gen X (43–58 years old), with only 1.2% being baby boomers and 0.6% being Gen Alphas. From a qualification perspective, 17.4% of the respondents only have a high school certificate, with 12.5% being diploma holders, 22.9% having undergraduate degrees, and 47.1% being postgraduate degree holders.

Table 3: Overall descriptive statistics

Descriptives	Frequencies	Percentages
Gender		
Male	133	40.5
Female	190	57.9
Non-binary	2	.6
Prefer not to say	3	.9
Total	**328**	**100.0**
Age		
Below 18	2	.6
18-26	78	23.8
27-42	176	53.7
43-58	68	20.7
Over 58	4	1.2
Total	**328**	**100.0**
Education		
High School certificate	57	17.4
Diploma	41	12.5
Undergraduate Degree	75	22.9
Post Graduate Degree	154	47.0
Total	**327**	**99.7**

Table 4: *Descriptive statistics for each construct*

Research Construct	N Statistic	Minimum Statistic	Maximum Statistic	Mean Statistic	Std. Deviation Statistic
Personalisation	328	1.00	7.00	5.6045	1.29083
Fraud_Detection	328	1.00	7.00	6.1978	1.16743
Customer_Experience	328	1.00	7.00	6.1668	1.18868
Customer_Acquisition_Retension	328	1.00	6.33	5.8720	1.03750

The results also reveal an overall mean score of 5.60 (SD=1.29) for the personalisation construct, which shows a positive perception towards ‘personalised products, services, and engagements’ from banking customers, with the ‘I need my bank to celebrate my milestones and reward me accordingly’ scale item having the highest mean value. Banking customers' overall perception

of fraud detection was found to be positive, with a mean score of 6.20 (SD = 1.17) for this construct.

Additionally, the results show that the Customer Experience Construct has an overall mean score of 6.17 (SD = 1.19), indicating that bank customers have a positive perception of the improvement of the "Customer Experience," with three of the four scale items having very high mean values. There is a positive correlation between customer experience and customer retention or acquisition, as evidenced by the mean score of 5.87 (DD = 1.04) for customer retention or acquisition.

4.3 Measurement Instrument Assessment

The analysis and discussion of the validity and reliability of the measurement instrument are conducted in this section. A statistical technique for analysing a measurement model's validity and structure is called confirmatory factor analysis (CFA). It's a type of structural equation modelling (SEM) that deals with measurement models specifically, and it’s employed to assess the degree to which the survey questions and observed variables match the contracts they are meant to measure (Brown and Moore, 2012). A conceptual model with pre-set variables and factors and their relationships was created for this research paper. A CFA is used to test the model and validate the measurement items within the South African banking context, as validated measurement scales were adopted for this study.

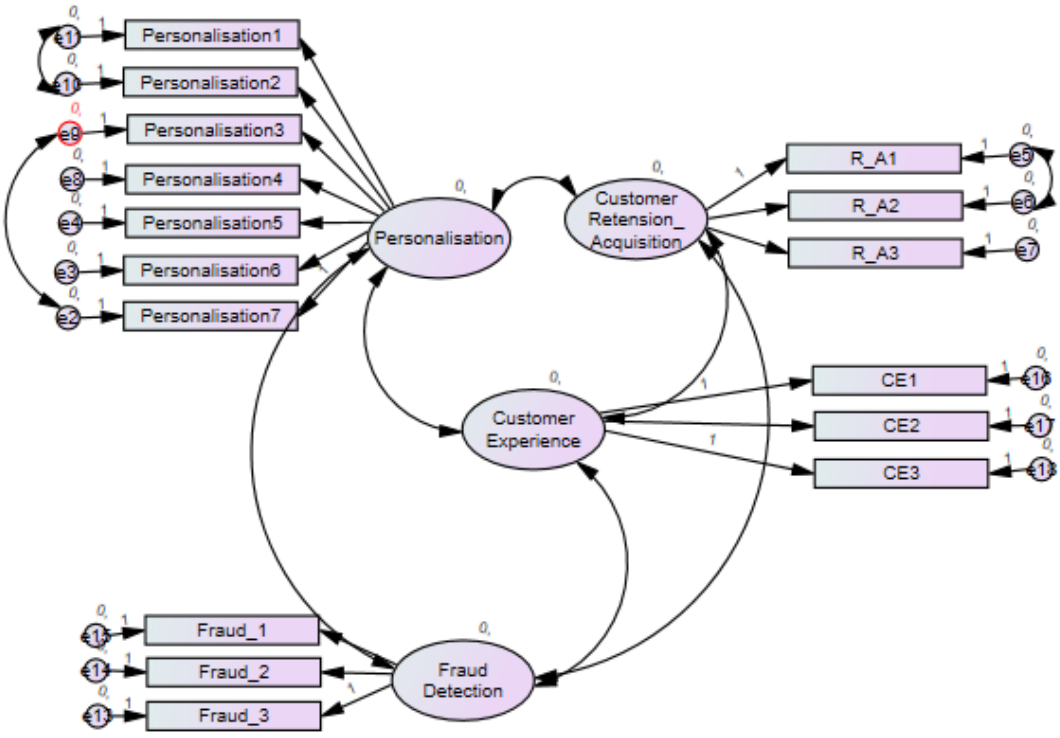


Figure 10: Measurement Model

The proposed conceptual research model consists of two independent constructs, a mediator and a dependent construct. The first construct, 'Personalisation' was measured by seven items (Personalisation1–Personalisation7). The second construct, 'Fraud Detection' was measured by four items (Fraud1–Fraud4). 'Customer Experience' was measured by 4 items (CE1–CE4). Lastly, the 'Customer Retention/Acquisition' construct was measured by 3 items (CRA1–CRA3).

The factor loadings, which are estimates of the underlying construct, of each scale item of the measurement instrument meant to test the proposed conceptual model were assessed. Hair et al. (2010) and Vinzi et al. (2009) recommend factor loadings >0.7; all the items meet the minimum factor loading. After running structural equation modelling, all the items also met the minimum AVE of 0.5 for their constructs (Hair et al., 2016), thus proving the validity of the model.

Table 5: Overview of the final research constructs with a focus on the Factor Loadings

Research Construct		Descriptive Statistics				Cronbach's Test		Composite Reliability (C.R) Value	AVE Value	Highest Shared Variance	Factor Loading
		Mean Value		Standard Deviation		Item-total	a value				
Personalisation	Personalisation1	5,50	5,641	1,768	1,675	0,866	0,884	0,855	0,526	0,520	0,699
	Personalisation2	5,67		1,726		0,863					0,733
	Personalisation3	5,68		1,682		0,863					0,788
	Personalisation4	5,71		1,539		0,868					0,711
	Personalisation5	5,21		1,872		0,873					0,665
	Personalisation6	6,02		1,597		0,875					0,702
	Personalisation7	5,70		1,544		0,864					0,769
Fraud Detection	Fraud_1	6,62	6,580	1,306	1,254	0,938	0,941	0,945	0,851	0,723	0,883
	Fraud_2	6,56		1,192		0,883					0,967
	Fraud_3	6,56		1,264		0,923					0,915
Customer Experirnce	CE1	6,12	6,047	1,331	1,398	0,776	0,875	0,894	0,739	0,673	0,921
	CE2	6,04		1,403		0,779					0,890
	CE3	5,98		1,459		0,911					0,760
Customer Acquisitions/ Retension	CRA1	4,72	5,870	0,826	1,103	0,887	0,915	0,924	0,803	0,723	0,884
	CRA2	6,29		1,322		0,889					0,845
	CRA3	6,60		1,161		0,855					0,956

4.3.1 *Reliability Analysis/ Testing*

By looking at the Cronbach Alpha Coefficient, Composite Reliability, and Average Value Extracted, the reliability of the measurement scale was evaluated. The composite reliability (CR) and Cronbach Alpha (α) reliability coefficients must both be ≥0.7 for every measurement scale. Additionally, the DV, or discriminant validity, or square root of the average variance extracted (AVE), must be higher than the correlations of the latent variables in the CFA (Hair et al., 2010; Henseler et al., 2015; Hu & Bentler, 1999).

Table 6: Composite Reliability (CR) and Cronbach Alpha (α)

				Composite reliability (CR)			
					summation of error terms		CRη=(Σλyi)2/[(Σλyi)2+(Σεi)]
			Estimate	(∑λYi)²	éi	∑éi	CR
Personalisation	<---	Personalisation1	0,699	25,674	0,511	3,321	0,885
	<---	Personalisation2	0,733		0,463		
	<---	Personalisation3	0,788		0,379		
	<---	Personalisation4	0,711		0,494		
	<---	Personalisation5	0,665		0,558		
	<---	Personalisation6	0,702		0,507		
	<---	Personalisation7	0,769		0,409		
Fraud Detection	<---	Fraud_1	0,883	7,645	0,220	0,448	0,945
	<---	Fraud_2	0,967		0,065		
	<---	Fraud_3	0,915		0,163		
Customer Experirnce	<---	CE1	0,921	6,610	0,152	0,782	0,894
	<---	CE2	0,890		0,208		
	<---	CE3	0,760		0,422		
Customer Acquisitions/ Retension	<---	CRA1	0,884	7,209	0,219	0,591	0,924
	<---	CRA2	0,845		0,286		
	<---	CRA3	0,956		0,086		

4.3.1.1 Cronbach Alpha Coefficient

As shown in Table 7, the Cronbach Alpha coefficients exceed the recommended value of 0.7. These range between 0.875 and 0.941 for all the research contexts, thus proving the reliability of the instrument.

4.3.1.2 Composite reliability

The composite reliabilities of each construct within measured by the measurement scale are shown in Table 7. These range between 0.885 and 0.945 for all the research contexts, all exceeding the recommended value of 0.7, thus further indicating the reliability of the instrument.

4.3.2 *Validity Testing*

For this study, the instrument's validity was assessed using the Factor Loadings, Correlation Matrix, Average Value Extracted (AVE), and Highest Shared Value (HSV).

4.3.2.1 Convergent Validity

Standard factor loadings were used to assess the convergence validity of the measurement model (Figure 11). These loadings show the extent to which indicators of a construct converge on that construct. According to Comrey and Lee (2013), factor loadings are: 0.32 (poor), 0,45 (fair), 0.55 (good), 0.63 (very good), and 0.71 (excellent) based on the correlation between the item and the construct. Factor loadings below 0.5 were eliminated per the recommendations of Bryne (2000). The model shows adequate convergent validity, with factor loadings ranging from 0.665 to 0,956.

Table 7: Convergent Validity (Factor Loadings)

Research Construct		Factor Loading
Personalisation	Personalisation1	0,699
	Personalisation2	0,733
	Personalisation3	0,788
	Personalisation4	0,711
	Personalisation5	0,665
	Personalisation6	0,702
	Personalisation7	0,769
Fraud Detection	Fraud_1	0,883
	Fraud_2	0,967
	Fraud_3	0,915
Customer Experirnce	CE1	0,921
	CE2	0,890
	CE3	0,760
Customer Acquisitions/ Retension	CRA1	0,884
	CRA2	0,845
	CRA3	0,956

4.3.2.2 Discriminant Validity: Correlation Matrix

To demonstrate the distinction between constructs, for this study, the inter-item correlation matrix has been used to assess the discriminant validity of the measurement instrument. According to O'Rourke & Hatcher (2013), the values within the inter-item correlation matrix should deviate from 1, thus proving the

discriminant validity of a scale. From the below table, it can be seen that discriminant validity has been met between all the constructs, with the highest being met between customer retention/acquisition and personalisation at 0,583 and the lowest between customer retention/acquisition and fraud detection at 0,784.

Table 8: Inter-Item correlation matrix

Research Construct	Personalisation	Fraud Detection	Customer Experience	Customer Retention OR Acquisition
Personalisation	1,000			
Fraud Detection	0,622	1,000		
Customer Experience	0,633	0,663	1,000	
Customer Retention OR Acquisition	0,583	0,784	0,775	1,000

4.3.2.3 Average Value Extracted (AVE) and Highest Shared Variance (HSV)

Table 9: Highest Shared Variance (HSV)

Research Construct	**Personalisation**	**Fraud Detection**	**Customer Experience**	**Customer Retention OR Acquisition**
Personalisation	1			
Fraud Detection	0,387	1		
Customer Experience	0,401	0,440	1	
Customer Retention OR Acquisition	0,340	0,615	0,601	1

The average variance extracted (AVE) is used to measure the degree to which multiple items used in the research to measure the same concept are in agreement; this is known as converging validity (Hair et al., 2010). As recommended by Hair et al. (2010), the minimum cut-off for AVE is 0.5 to prove that the construct explains more than half of its indicators' variance. In this study, the AVE values ranged between 0.526 and 0.851, suggesting that there are no discriminant validity issues. As shown in Table 3, the AVE for each construct is higher than the HSV for each construct, thus further verifying discriminant validity (Nusair & Hua, 2010).

Table 10: Average Value Extracted (AVE) and Highest Shared Variance (HSV)

Research Construct		AVE Value	Highest Shared Value	Factor Loading
Personalisation	Personalisation1	0,526	0,400	0,699
	Personalisation2			0,733
	Personalisation3			0,788
	Personalisation4			0,711
	Personalisation5			0,665
	Personalisation6			0,702
	Personalisation7			0,769
Fraud Detection	Fraud_1	0,851	0,614	0,883
	Fraud_2			0,967
	Fraud_3			0,915
Customer Experirnce	CE1	0,739	0,600	0,921
	CE2			0,890
	CE3			0,760
Customer Acquisitions/ Retension	CRA1	0,803	0,614	0,884
	CRA2			0,845
	CRA3			0,956

4.4 Structural Equation Modelling (SEM)

4.4.1 *Model Fit Indices*

The following thresholds were used to subject the measurement scale to CFA: Comparative Fit Index (CFI)≥0.95, Tucker-Lewis Index (TLI)≥0.95, Incremental Fit Index (IFI)≥0.95, $\mathrm{CMIN/DF}\ x^2/df \leq 3.0$, Standard Root Mean Squared Residual (SRMR)≤0.08, Root Mean Error of Approximation (RMSEA)≤0.06 (Hu & Bentler, 1999; Hair et al., 2010).

As can be seen from table 12 below, the model was found to have adequate fit according to all the measures as recommended by Hu & Bentler (1999).

Table 11: Model Fit results

Measure	Estimate	Threshold	Interpretation
CMIN	356.944	--	--
DF	96.000	--	--
CMIN/DF	3.718	Between 1 and 3	Acceptable
CFI	0.942	>0.95	Acceptable
IFI	0.942	>0.95	Acceptable
TLI	0.927	>0.95	Acceptable
SRMR	0.057	<0.08	Excellent
RMSEA	0.059	<0.06	Acceptable

Table 12: Model Fit results cut-off criteria

Measure	Terrible	Acceptable	Excellent
CMIN/DF	> 5	> 3	> 1
CFI	<0.90	<0.95	>0.95
SRMR	>0.10	>0.08	<0.08
RMSEA	>0.08	>0.06	<0.06

Note: the above cutoff criteria are defined by Gaskin et al., 2022 "Model Fit Measures", AMOS Plugin. Gaskination's StatWiki

4.5 The structural Model (Path modelling and Hypotheses Testing)

This section aims to indicate the results of the hypotheses made for this research. According to Keith (2015), path modelling can analyse direct, indirect, and total effects between constructs. The final structural model (Figure 12), which resulted

from the CFA, has two independent variables, namely personalisation through products, services, and engagements and fraud detection, with customer experience as a mediator and customer retention and acquisition as a dependent variable. The final hypotheses that have been tested are H1a, H1b, and H2.

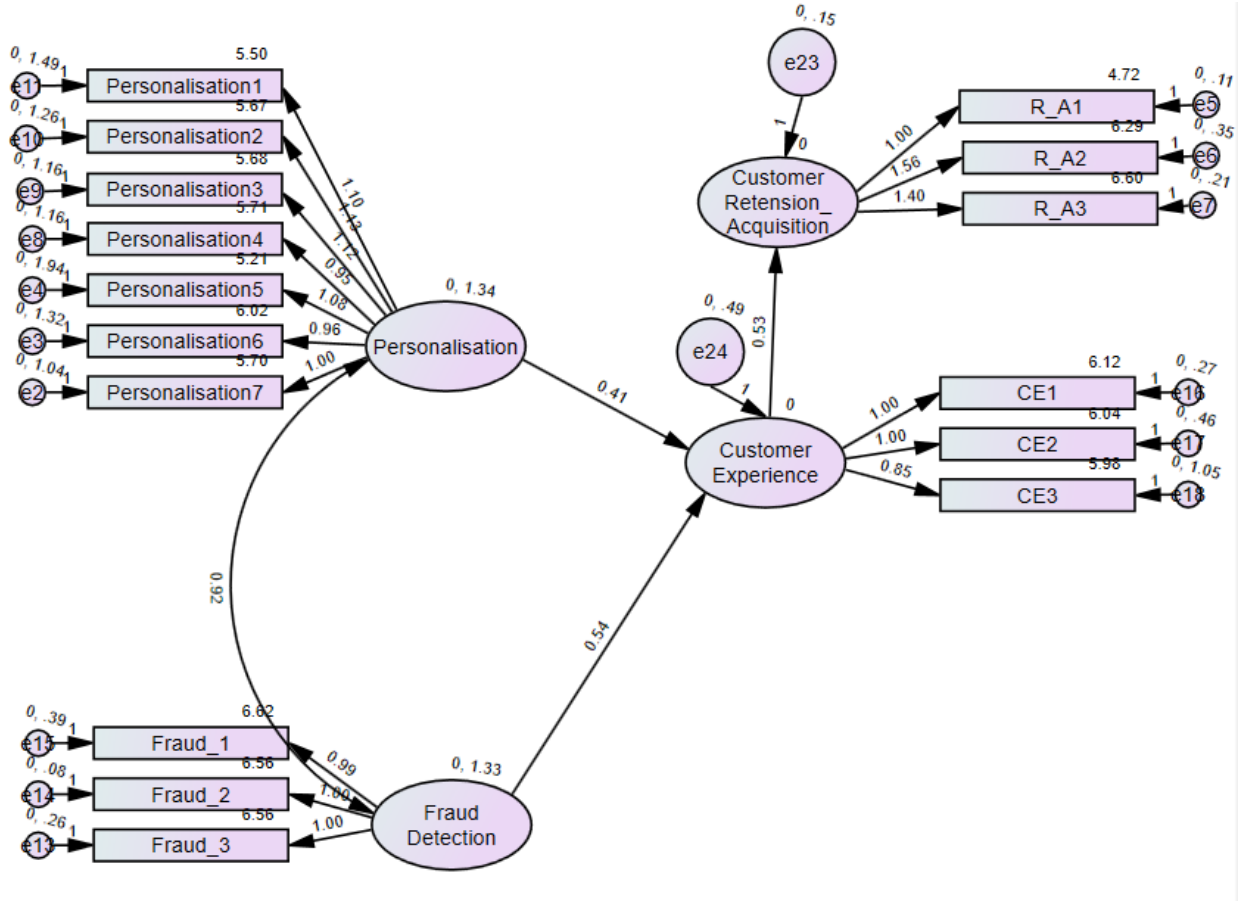


Figure 11: The Structural Model

Table 13: Summary of the SEM results (Standardized Regression Weights)

Parameter			Estimate	Lower	Upper	P	Interpretation
Customer Experience	<---	Personalization	.385	.203	.588	.001	H1e- Accepted
Customer Experience	<---	Fraud Detection	.507	.234	.786	.001	H1a- Rejected
Customer Retention/ Acquisition	<---	Customer Experience	.863	.702	.968	.002	H2- Accepted

Note: P value less than 0.05 (statistically significant).

4.6 Results pertaining to Hypothesis 1

H1: BDA applications improve customer experience.

a. H1a: Personalisation of products, services, and engagements through BDA improves customer experience.
b. H1b: Fraud detection through BDA improves customer experience.

As can be seen from Table 14 above, SEM results show that fraud detection within banks has a greater contribution to the customer experience (standardised beta = 0.507) than personalisation (standardised beta = 0.385). Given that the P-value is statistically significant (P<0.05) for both hypotheses, one can conclude that personalisation in the form of product, services and engagements through BDA improves the customer experience. One can also conclude that the ability of a bank to detect fraud has a positive relationship with customer experience.

4.7 Results pertaining to Hypothesis 2

H2: Improved customer experience resulting from BDA applications has a positive impact on customer retention and new customer acquisitions.

The SEM results also indicate that customer experience contributes (standardised beta = 0.863) to customer retention and/or acquisition in banking within the South African context. The P-value between customer experience and customer retention and/or acquisition shows that the positive relationship is statistically significant.

Testing Indirect effects (Mediation effect) of customer experience

The total effect of one variable on another is equal to the sum of all the direct and indirect effects between the two variables.

Table 14: Squared Multiple Correlations: R squared values

Parameter	Estimate	Lower	Upper	P
Customer Experience	.676	.386	.919	.002
Customer Retention/Acquisition	.745	.493	.938	.002

Furthermore, the indirect effect (mediation effect) of customer experience was tested, resulting in the squared multiple correlations indicated in Table 15. “A multiple correlation coefficient (R) yields the maximum degree of linear relationship that can be obtained between two or more independent variables and a single dependent variable” (Arthur R. Jensen, 2006). The predictive ability of the constructs of the model is also indicated by the R-square values. Table 15 indicates the squared multiple correlations for the endogenous variables. Customer experience scored an R-square of 0.676, and customer retention/acquisition scored 0.745, both statistically significant, thereby confirming the predictive ability of the model and both hypotheses related to the BDA use in this study.

4.8 Summary of the results/findings

The reliability and validity of the measurement instrument were indicated through the use of confirmatory factor analysis. It was also proven that the data fits the model through the model fit analysis. The testing of the hypotheses also indicated that the relationship between BDA applications; fraud detection, and personalisation will improve the banks’ ability to retain and acquire customers as a result of customer experience improvement.

CHAPTER 5. DISCUSSION OF THE RESULTS OR FINDINGS

5.1 Introduction

In this chapter, the findings of the results presented in Chapter 4 are discussed within the context of the literature related to the study. The gap that the study addresses is the impact of BDA capabilities on an organisation's performance. Having investigated the antecedents of BDA within organisations, Côrte-Real et al. (2019) suggest that the use of BDA creates the most value for an organisation. Hock et al. (2022) believe that BDA can be used as a differentiation strategy, though the authors were unable to measure the positive effect of big data analytics within the German banking industry by 2019 and suggested a future study. One can argue that this study addresses the issue to a certain degree, given the customer perspective that was obtained during the study.

5.2 Discussion pertaining to Hypothesis 1: BDA applications improve customer experience.

5.2.1 *Customer experience and BDA applications*

H1: BDA applications improve customer experience.

a. H1a: Personalisation of products, services, and engagements through BDA improves customer experience.
b. H1b: Fraud detection through BDA improves customer experience.

The findings of this study showed not only that personalisation through customer engagements and product offerings is a factor that has a significant relationship with customer satisfaction among banking customers in the South African context, but also that fraud detection by banks in relation to customer bank

accounts is the strongest dimension of big data use affecting customer experience.

H1a: Personalisation of products, services, and engagements through BDA improves customer experience.

The findings of this study regarding personalisation through product offerings and customer engagements as enabled by the application of big data analytics having a positive relationship with the improvement of customer experience are supported in various ways by previous research.

Côrte-Real et al. (2019) suggest the use of BDA predictive models to target customers and send personalised messages in real time. Rodriguez-Ibañez et al. (2023) encourages the use of sentiment analysis in order to understand causalities between actions taken by an organisation with regards to their products and what customers are saying (positive or negative). Birjali et al. (2021) suggest that BDA should be used for product and service improvements based on customer feedback resulting from the perception of a product or service. The authors proposed that banks should provide personalised services and products that consider each client's preferences and habits through the use of social CRM, which will enable banks to grow and maintain their customer base. Anshari et al. (2019) also suggest the use of big data in CRM to provide personalisation, customisation, and quick responses to customers.

To address the question of whether there are causalities between customer experience and BDA use, authors such as Holmlund et al. (2020) developed a comprehensive framework that indicates how BDA can be used to improve customer experience. From strategizing the type of customer experience actions needed, assessing the kind of customer experience insights needed, determining the type of data required, and implementing a customer experience solution. Even though Holmlund et al. (2020) simply gave a framework, this study has provided guidance on how the value of using BDA to improve customer experience can be obtained.

Tseng (2023) found that the use of big data analytics positively impacts customer response capabilities, given that an organisation is able to obtain and analyse information in real time. Both Holmlund et al. (2020) and Anshari et al. (2019) indicate that there is a positive relationship between big data analytics and customer experience. According to Mukherjee et al. (2018), an additional area in which technology can significantly influence a positive customer experience is the calibre of client engagements during a buying event or obtaining a new banking product.

Hosseini et al. (2022) argue that customer segmentation is most likely the most effective way to retain and attract new customers as a result of a bank's better understanding of customer behaviour and effective communication. Kovács et al. (2021) suggest that patterns are generated from customer information and, as a result, can be used to make recommendations on products and services that customers might be interested in. This would result in a bank’s ability to customise product and service offerings based on customers’ personal needs.

Some authors focused on experimental research. Cuomo et al. (2023) used predictive and exploratory analysis to provide extensive customer profiling in an effort to assist organisations in personalising customer engagements. Egorova et al. (2022) also developed a similar model. Harvard Business Review Analytic Services (2018) found that personalisation can be used in order for a firm to obtain a competitive advantage. Personalisation has emerged as a key element in enhancing business performance, according to a 2018 Harvard Business Review survey of more than 600 business executives.

According to Ameen et al. (2021), the degree to which information is customised to meet the needs of a single user is known as personalisation, and it plays a significant role in determining positive experiences. They suggest that by analysing a customer's past purchases and preferences, machine learning technology can personalise services and product recommendations, and that the focus should be on the improvement of predictive models and accurate algorithms. Ameen et al. (2021) found that personalisation has a positive

relationship with the AI-enabled customer experience, with trust, perceived sacrifice, and relationship commitment mediating the path between the two.

Given that personalisation is an element associated with AI-enabled services (Zanker et al., 2019), this finding is similar to Hoyer et al.'s (2020), who also found a positive relationship between AI-powered technology (the Internet of Things, virtual reality, chatbots, and virtual assistants) and customer experience.

The findings of this study are also consistent with the recommendations made by Gupta et al. (2020), who contend that real-time, continuous, personalised engagement should be designed and delivered using digital analytics. According to Gupta et al. (2020), because personalisation and customer engagements are so important to customer management strategies, businesses must learn how to smoothly incorporate new technologies into their current operations in order to produce insights that can be put to use.

Therefore, managers should invest in and speed up the adoption of big data analytics technology to improve the customer experience through personalisation from a product offering and customer engagement perspective.

H1b: Fraud detection through BDA improves customer experience.

The findings of this study regarding a bank’s ability to detect fraud as enabled by the application of big data analytics having a positive relationship with the improvement of customer experience are supported in various ways by previous research.

Delgosha et al. (2019) recommend that banks consider investing in BDA applications in risk and fraud and operational or customer analysis categories in order to benefit from large amounts of internal and external data. Abedin et al. (2023) argue that banks can use the BDA model that they developed to understand their customers' demands, identify risk or fraud, and predict factors associated with customer behaviour to minimise customer churn.

Zhu et al. (2021) and Mckinsey & Co. (2022) argue that an increase in the use of digital devices for banking and electronic payments has accelerated fraud and made it more sophisticated. The findings of this study are supported by other authors, such as Mckinsey & Co. (2022), who found that willingness and trust in using certain banking services are harmed by fraud incidents. They also found that customer satisfaction increases where fraud happens after a customer is warned about potential fraud with regards to their account.

Shahana et al. (2023) and Jurgovsky et al. (2018) also conducted experimental studies by developing models for BDA that organisations need to consider for fraud detection purposes in order to protect themselves and their customers.

According to Zoldi (2015), reducing false positive rates can lead to fewer customer cards and transactions being blocked, which is one of the major factors that affects customer experience in fraud detection. He continues by saying that in the fight against fraud, providing an excellent customer experience should be the top priority. False positives are defined by Zoldi (2015) as the quantity of legitimate transactions that banks suspect to be fraudulent.

Vorobyev and Krivitskaya (2022) developed a machine learning model or algorithm with the goal of lowering false positive rates in order to address the problem of fraud detection in bank payments transactions suffering from a high number of false positives. The authors believe that developing ways to protect client funds against fraudsters is part of banking corporations' strategy for improving the customer experience. Their research verified that using AI techniques to recombine the conditions that make up the expert rules used in fraud monitoring systems is both feasible and effective. They also believe that their approach could enhance the client experience in e-commerce and banking.

Wedge et al. (2019) observed that false positives plague the fraud prediction industry, and as a result, they used feature engineering to develop an improved machine learning solution to drastically reduce the "false positives" through the use of large datasets from a bank containing 900 million transactions. According

to research, only one in five cases that are reported as fraud are found to be fraudulent, and one in six consumers report having had a legitimate transaction declined in the previous year (Wedge et al., 2019). They highlighted that organisations currently use 24/7 fraud monitoring systems that are costly and only reduce 5% of false positives.

Zohar (2018) noted that a survey conducted by the 2017 Global Fraud Survey found that 2.6% of all incoming online store orders are declined due to fear of fraud, which results in the potential loss of good customers. He argues that the value of orders that were mistakenly declined, the lost customer lifetime value, and the wasted customer acquisition costs should all be considered when calculating the financial impact of falsely declined transactions. According to Javelin Strategy (2015), a false-positive decline has the potential to severely damage a customer's loyalty by encouraging many of them to stop doing business with the issuer or merchant that led to their transaction being blocked.

The implication of this finding on business is that with an increase in fraudulent activity with electronic payments, as highlighted by Zhu et al. (2021) and Mckinsey & Co. (2022), businesses need to firm up their efforts to prevent fraud through the use of BDA in order to improve customer experience. However, businesses should not hinder customers from transacting as a result of false positives, as this will have a negative impact on the customer experience (Zoldi, 2015); instead, they must utilise robust BDA-enabled machine learning models to ensure that they increase fraud detection rates while reducing false positive rates (Wedge et al., 2019; Vorobyev and Krivitskaya, 2022).

5.3 Discussion pertaining to Hypothesis 2: Improved customer experience resulting from BDA applications has a

positive impact on customer retention and new customer acquisitions.

The findings of this study also showed that there is a significant relationship between the willingness of customers to stay or join a bank (customer retention or acquisition of new customers) and the priority of customer experience as a result of BDA applications such as fraud detection and personalisation. The findings address the gap highlighted by Rialti et al. (2019), who found that the complex relationship between BDA and organisational impact is still being understood by scholars and that part of the research is still theoretical or qualitative at most, with quantitative studies still in their infancy.

In trying to understand how big data adds value to organisations, Mikalef et al.'s (2019) study focused on the contribution of big data analytics to a firm's performance. They found that BDA needs to form part of the company's business strategy. Wamba et al. (2017) also found that there is causality between the three big data analytics capabilities and firm performance, with dynamic process capability as a mediator. Rialti et al. (2019) also found that organisational performance is positively impacted by ambidexterity and agility derived from BDA capabilities, giving them a competitive advantage.

Fernandes and Pinto's (2019), whose study focused on customer experience, found that there is a strong relationship between customer experience and customer retention, word of mouth, as well as tolerance with relationship quality as a mediator. In line with Fernandes and Pinto's (2019) findings, Radia et al. (2022) also found customer experience to have a significant influence on relationship quality. The results show that there's a relationship between customer experience and relationship quality and a relationship between customer experience and retention, with physical comfort as a mediating effect.

Becker and Jaakkola (2020) found that customer experience is central to a firm's competitiveness. As a result of their systematic review process of going through 136 articles, they define customer experience as "non-deliberate, spontaneous

responses and reactions to particular stimuli (touchpoints within the customer journey, e.g., product or service offerings)," which can be measured as ordinary (weak) or extraordinary (strong). Becker and Jaakkola (2020) suggest that businesses must understand a wide variety of touchpoints, both inside and outside of their control, in offline and online contexts that have a dynamic impact on customers.

Although businesses cannot create the customer experience, they can control, direct, and observe a variety of factors that influence it. In the context of this research, one can argue that fraud detection, personalised product offerings, and engagements are both stimuli along the customer journey that the customers respond to and react to, thus impacting whether their experience was ordinary or extra ordinary.

To personalise retention efforts, businesses can tailor and maximise the incentives they provide to customers by gathering real-time data on them from a variety of sources, including unstructured data from sources like online chat sessions and call centre conversations and structured data like product usage and browsing behaviour (Ascarza et al., 2017).

Cambra-Fierro et al. (2021) argue that customer experience is critical to customer retention, ultimately leading to positive financial returns for firms that invest in customer experience. They suggest that a customer's perceptions of an experience they had with a brand determine customer retention. According to a global survey conducted by Gartner in 2019 across a variety of industries, 74% of organisations anticipated an increase in customer experience budgets in 2020 and predicted that this would remain one of the major areas of investment in the years to come.

The results of this study reveal that a positive customer experience positively influences customer retention and the acquisition of new customers. Even though previous authors looked at different contexts in terms of the antecedents of customer experience, this idea is not new and has already been suggested by

several previous authors (e.g., Becker and Jaakkola (2020); Cambra-Fierro et al. (2021); Fernandes and Pinto (2019)).

The findings of this study also align with Lamrhari et al. (2022), who argue that their framework shows how social media data can be used to better understand customer satisfaction, identify the true needs of customers, enhance customer acquisition and retention, and improve specific product aspects. Lamprinakos et al. (2021) found that online advertisements that are tailored to the individual receive more positive feedback than those that are not. Personalised advertisements were found to have a greater impact on consumers' intentions to purchase and evaluate products, thus leading to the acquisition and retention of customers.

From a management perspective, businesses need to invest in BDA applications to improve their customer experience, as this will result in customer retention and the acquisition of new customers through targeted and personalised advertisements and electronic word of mouth from existing customers.

5.4 Conclusion

No previous study empirically tested or validated the impact of BDA applications on customer retention or acquisition with positive experiences on customers as a mediator. Other authors have taken various angles to assess the business value of investing in BDA. By proving that there is a positive relationship between BDA applications (personalisation and fraud detection) and customer retention and acquisition, this study has added to the body of knowledge by showing that BDA will improve business performance by taking a customer-focused approach.

CHAPTER 6. CONCLUSIONS & RECOMMENDATIONS

6.1 Introduction

This chapter concludes the paper by providing possible answers to the research questions. Recommendations are then made to banking executives and senior managers on how they can use BDA as a customer retention and acquisition strategy in order to compete within the industry based on the discoveries made. Lastly, suggestions for future research are made based on the gaps that were identified on the topic of BDA applications.

6.2 Conclusions regarding research question 1

6.2.1 *Customer experience and BDA applications*

- To what extent should South African retail banks use BDA for fraud detection, credit risk analysis, and the determination of customer needs through customer segmentation to develop relevant products and services to improve the customer experience?

This study aimed to examine the possibility of using big data analytics as a customer retention and acquisition strategy within the South African retail banking context. As a result, the research was customer-focused and explored whether BDA applications would improve the customer experience. The study followed a quantitative approach, where a survey was distributed among South African bank clients. The data was analysed using SEM with AMOS. Various studies have been conducted in the field of BDA; however, most of them either focused on the implementation challenges or the development of ML models from BDA, and some aimed to understand the value brought about by BDA in an organisation

and how it can be used to get a competitive advantage, as was shown in the literature review; however, none of the authors' focus was on a customer-focused approach, as is done in this study.

Findings suggest that the three hypotheses that were successfully tested in this study were supported by the data, and the main contribution of this study was that the ability for a bank to provide personalisation (from an engagement and product offering perspective) and its ability to detect fraud are both predictors of customer satisfaction within South African digitally active bank customers, thus improving the ability for the bank to retain their existing customers and potentially acquire new customers.

To answer the research question, one can conclude that South African retail banks should use BDA for fraud detection, provide personalisation through the determination of customer behaviour through customer segmentation, and offer products in order to improve the customer experience.

The study aligns with previous research highlighting the importance of personalisation in enhancing customer experience. It suggests that technology, big data, and social CRM can improve customer engagement, growth, and retention, while reducing false positive rates can enhance fraud detection.

6.3 Conclusions regarding research question 2

6.3.1 *BDA applications AND Customer Retention*

- Will the use of BDA to improve the customer experience result in customer retention and new customer acquisitions?

With regards to the second research question, the aim was to analyse the role played by the application of BDA on customers' retention and acquisition in retail banking, with customer experience as a mediator variable. When testing the indirect effects of BDA applications that were successfully tested in this study

(personalisation and fraud detection) on customer retention and acquisition through customer experience as a mediator, the results of this research indicate that there is a positive relationship.

From a customer retention and acquisition perspective, the findings of this study also align with those of previous authors, who have identified a strong correlation between customer experience, retention, and word of mouth. Previous research suggest that social media data can enhance customer satisfaction, identify needs, and improve product aspects. Personalised online advertisements receive positive feedback, thereby increasing customer acquisition and retention.

6.4 Overall Conclusions of this study

The research presents a ground-breaking attempt to examine how a cutting-edge technology, Big Data Analytics, can be used as a customer retention and acquisition strategy in banking through understanding customer needs from a personalisation of product offerings and engagements perspective and fraud detection within the South African banking context. The overall conclusion of the study is that there is a positive relationship between BDA applications and customer retention and acquisition through the significant effects of customer experience.

6.5 Managerial implications and Recommendations

With the rife competition within the South African retail banking industry, senior management within banking can use the findings of this empirical study as reassurance that BDA investments should form part of the organisational strategy and should also take into consideration the challenges that come with BDA, as highlighted by other authors. This study allows managers to understand BDA applications within the banking context and how customers feel about those that were successfully tested.

The findings of this study have important implications for senior executives and managers in banking and financial services who are interested in acquiring new customers and retaining their current customer base. This study suggests that big data applications such as personalisation of product offerings and engagements and fraud detection have a positive impact on the customer experience. As a result, the recommendation is that managers should invest in BDA technology with the aim of using it to provide a personalised customer experience and to detect fraud from both structured and unstructured data sources.

Fraud detection was found to have a stronger relationship with customer experience; as a result, senior management should prioritise the use of BDA to detect fraud. The lack of stringent fraud detection methods may result in a bank losing customers to competitors as customers are afraid of losing their money (Mckinsey & Co., 2022; Zoldi, 2015). The provision of personalisation through engagements and product offerings also improves the customer experience; as a result, banks should also invest to ensure that they meet customer expectations.

Organisations should also invest in the use of tried and tested BDA and machine learning models, both for personalisation and for detecting fraud. Shahana et al. (2023), Jurgovsky et al. (2018), and Vorobyev and Krivitskaya (2022) developed fraud detection models that minimise false positives. Authors such as Cuomo et al. (2023) and Egorova et al. (2022) developed models to assist organisations in personalising customer engagements.

6.6 Limitations of the study

Potential weaknesses of the study were taken into consideration during data collection, given the use of a questionnaire as outlined in Bryman et al. (2019). Additionally, anything that reduces the ability to establish a relationship between dependent and independent variables reduces the study's internal validity. These

include selection bias, sample size, history, participants' maturity, instrumentation, practicing effects, attrition (dropping out of the study), and confounding (when a variable not controlled for has an impact on the results).

This study is among the first to focus on customers' experiences in the context of big data analytics-enabled applications. As a result, there is a need to improve the quality and reliability of the research instrument, as suggested by Patel (2009), who recommends the Total Survey Error Framework and outlines the steps to be followed (i.e., expert reviews, cognitive interviews, another round of expert interviews, a pilot survey, and factor analysis).

6.7 Suggestions for further research

To further extend this research, it is recommended that a similar study be conducted in other countries and in other industries, such as insurance, that stand to benefit from the use of BDA as a way to cross-validate the research. The data were collected primarily using the snowball sampling method, which warrants caution in generalising results, as this method may result in sampling bias and a margin of error since people are most likely to refer to those whom they know and have similar traits as them. As a result, it is recommended that future studies be done using probability sampling methods, thus allowing for strong statistical inferences about the whole group to be made.

Given that this study could only validate two constructs, other dimensions of BDA applications can also be explored. One of the applications of BDA that was explored during the literature review is the use of BDA to conduct credit risk analysis for financially excluded customers and those without credit records through the use of customer transaction data. Its relationship with customer retention or acquisition could not be tested given the poor reliability and validity of the measurement scale.

This study focuses on the customer's perspective; future research could potentially study the perceptions of senior management and the differences between management and customers' perceptions.

Table 15. Consistency table: research questions, propositions, data collection and data analysis

RQ #		State Research Question	Prop/hyp #	State Proposition or Hypothesis	Data collection detail	Data analysis method
1		Does the use of BDA improve customer experience?	1	**H1: BDA applications improve customer experience.**	Questionnaire 7 Likert scale Strongly agree, Agree, somewhat agree, neither agree nor disagree, somewhat disagree, disagree, strongly disagree	Confirmatory Factor analysis
1.1		Will customer experience improve because of utilizing BDA to personalise, products, services and engagements?	1.a	H1a: Personalised products, services and engagements through BDA improve customer experience.	Questionnaire 7 Likert scale Strongly agree, Agree, somewhat agree, neither agree nor disagree, somewhat disagree,	Confirmatory Factor analysis

RQ #		State Research Question	Prop/hyp #	State Proposition or Hypothesis	Data collection detail	Data analysis method
					disagree, strongly disagree	
1.2		What is the relationship between Fraud detection and customer experience?	1.b	H1b: Fraud detection through BDA improves customer experience	Questionnaire; 7 Likert scale Strongly agree, Agree, somewhat agree, neither agree nor disagree, somewhat disagree, disagree, strongly disagree	Confirmatory Factor analysis
2		Will the use of BDA to improve customer experience result in customer retention and new customer acquisitions?	2.	**Improved customer experience resulting from BDA applications has a positive impact on customer retention and acquisition.**	Questionnaire; 7 Likert scale Strongly agree, Agree, somewhat agree, neither agree nor disagree, somewhat disagree,	Confirmatory Factor analysis

RQ #		State Research Question	Prop/hyp #	State Proposition or Hypothesis	Data collection detail	Data analysis method
					disagree, strongly disagree	

REFERENCES

A new approach to fighting fraud while enhancing customer experience. (2022, November). McKinsey & Company. Retrieved February 23, 2024, from https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/a-new-approach-to-fighting-fraud-while-enhancing-customer-experience.

Abedin, M. Z., Hájek, P., Sharif, T., Satu, M. S., & Khan, M. I. (2023). Modelling bank customer behaviour using feature engineering and classification techniques. Research in International Business and Finance, 65, 101913. https://doi.org/10.1016/j.ribaf.2023.101913

Akter, S., Bandara, R., Hani, U., Wamba, S. F., Foropon, C., & Παπαδόπουλος, Θ. (2019). Analytics-based decision-making for service systems: A qualitative study and agenda for future research. International Journal of Information Management, 48, 85–95. https://doi.org/10.1016/j.ijinfomgt.2019.01.020

Ameen, N., Tarhini, A., Reppel, A., & Anand, A. (2021). Customer experiences in the age of artificial intelligence. Computers in Human Behavior, 114, 106548. https://doi.org/10.1016/j.chb.2020.106548

Armstrong, B., & Lee, G. J. (2021). A digital business primer (2nd ed.). Silk Route Enterprises.

Arnold, T. J., Fang, E., & Palmatier, R. W. (2010). The effects of customer acquisition and retention orientations on a firm's radical and incremental innovation performance. Journal of the Academy of Marketing Science, 39(2), 234–251. https://doi.org/10.1007/s11747-010-0203-8

Anshari, M., Almunawar, M. N., Lim, S. A., & Al-Mudimigh, A. S. (2019). Customer relationship management and big data enabled: Personalization & customization of services. Applied Computing and Informatics, 15(2), 94–101. https://doi.org/10.1016/j.aci.2018.05.004

Askar, M. a. A., Aboutabl, A. E., & Galal, A. (2022). USING SOCIAL MEDIA ANALYTICS IN THE BANKING INDUSTRY TO DRIVE BUSINESS ADVANTAGE. Xinan Jiaotong Daxue Xuebao, 57(4), 405–415. https://doi.org/10.35741/issn.0258-2724.57.4.36

Ascarza, E., Neslin, S. A., Netzer, O., Anderson, Z., Fader, P. S., Гупта, С., Hardie, B. G. S., Lemmens, A., Libai, B., Neal, D. T., Provost, F., & Schrift, R. Y. (2017). In pursuit of enhanced customer retention Management: review, key issues, and future directions. Customer Needs and Solutions, 5(1–2), 65–81. https://doi.org/10.1007/s40547-017-0080-0

Becker, L., & Jaakkola, E. (2020). Customer experience: fundamental premises and implications for research. Journal of the Academy of Marketing Science, 48(4), 630–648. https://doi.org/10.1007/s11747-019-00718-x

Bihani, P., & Patil, S. T. (2014). A comparative study of data analysis techniques. International Journal of Emerging Trends & Technology in Computer Science, 3(2), 95–101.

Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges, and trends. Knowledge-Based Systems, 226, 107134. https://doi.org/10.1016/j.knosys.2021.107134

Botchway, R. K., Jibril, A. B., Oplatková, Z. K., & Chovancová, M. (2020). Deductions from a Sub-Saharan African Bank's Tweets: A sentiment analysis approach. Cogent Economics & Finance, 8(1), 1776006. https://doi.org/10.1080/23322039.2020.1776006

Brown, T. A., & Moore, M. T. (2012) Confirmatory factor analysis. In R. H. Hoyle (Ed.), Handbook of structural equation modelling. New York: Guilford Press. Pp. 361-379.

Bryan, A., Bell, E., Hirschsohn, P., Dos Santos, A., Du Toit, J., Masenge, A., Van Aardt, I., & Wagner, C. (2019). Research Methodology (2nd ed.).

Buldoo, D. R. (2018). Leveraging big data analytics for competitive advantage in South African banking. University of Pretoria. http://hdl.handle.net/2263/68831

Byrne, B. M. (2000). Structural equation modeling with AMOS: basic concepts, applications, and programming. http://cds.cern.ch/record/1320709

Camarate, J., & Brinckmann, S. (2017). A marketplace without boundaries. The future of banking: A South African perspective. PWC Strategyand, 1–19. https://www.pwc.co.za/en/assets/pdf/strategyand-future-of-banking.pdf

Cambra-Fierro, J., Gao, L., Melero-Polo, I., & Trifu, A. (2021). How do firms handle variability in customer experience? A dynamic approach to better understanding customer retention. Journal of Retailing and Consumer Services, 61, 102578. https://doi.org/10.1016/j.jretconser.2021.102578

Carletti, E., Oliviero, T., Pagano, M., Pelizzon, L., & Subrahmanyam, M. G. (2020). The COVID-19 Shock and Equity Shortfall: Firm-Level Evidence

from Italy. The Review of Corporate Finance Studies, 9(3), 534–568. https://doi.org/10.1093/rcfs/cfaa014

Chng, J., Zhang, J. (2014). Big data: the next leading edge in the financial industry, <http://www.atkearney.com.tr/financial-institutions/ideas-insights/article/-/ asset publisher/LCcgOeS4t85g/content/big-data-the-next-leading-edge-in- the-financial-industry/10192? 101 INSTANCE LCcgOeS4t85g redirect=%2Ffinancial-institutions%2Fideas-insights%2F-%2Fasset publisher%2FSxO47XYyZBF5%2Fcontent%2Fthe-new-reality-of-gcc- banking%2F10192>.

Comrey, A. L., & Lee, H. B. (2013). A first course in factor analysis. In Psychology Press eBooks. https://doi.org/10.4324/9781315827506

Côrte-Real, N., Ruivo, P., Oliveira, T., & Popovič, A. (2019). Unlocking the drivers of big data analytics value in firms. Journal of Business Research, 97, 160–173. https://doi.org/10.1016/j.jbusres.2018.12.072

Cowling, N. (2023, October 19). Leading Banks in South Africa as of 2023, by assets. Statista. Retrieved February 23, 2024, from https://www.statista.com/statistics/1346961/leading-banks-in-south-africa-by-assets/

Creswell, J. W., & Crewell, J. D. (2018). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (5th ed.). SAGE.

Cuomo, M. T., Tortora, D., Colosimo, I., Celsi, L. R., Genovino, C., Festa, G., & La Rocca, M. (2023). Segmenting with big data analytics and Python: A quantitative exploratory analysis of household savings. Technological

Forecasting and Social Change, 191, 122431. https://doi.org/10.1016/j.techfore.2023.122431

Delgosha, M. S., Hajiheydari, N., & Fahimi, S. M. (2020). Elucidation of big data analytics in banking: a four-stage Delphi study. Journal of Enterprise Information Management, 34(6), 1577–1596. https://doi.org/10.1108/jeim-03-2019-0097

Egorova, E., Glukhov, G., & Shikov, E. (2022). Customer Transactional Behaviour Analysis through embedding interpretation. Procedia Computer Science, 212, 284–294. https://doi.org/10.1016/j.procs.2022.11.012

False-positive card declines push consumers to abandon issuers and merchants. (2017, August 17). Javelin. Retrieved February 24, 2024, from https://www.javelinstrategy.com/press-release/false-positive-card-declines-push-consumers-abandon-issuers-and-merchants

Fernandes, T., & Pinto, T. (2019). Relationship quality determinants and outcomes in retail banking services: The role of customer experience. Journal of Retailing and Consumer Services, 50, 30–41. https://doi.org/10.1016/j.jretconser.2019.01.018

Gandomi, A. H., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Gartner Says 74% of Customer Experience Leaders Expect Budgets to Rise in 2020. (2020, January 15). Gartner. Retrieved February 24, 2024,

from https://www.gartner.com/en/newsroom/press-releases/2020-01-15-gartner-says-74--of--customer-experience-leaders-expe.

Gaskin et al., 2022 "Model Fit Measures", AMOS Plugin. Gaskination's StatWiki

Ghasemaghaei, M., & Calic, G. (2020). Assessing the impact of big data on firm innovation performance: Big data is not always better data. Journal of Business Research, 108, 147–162. https://doi.org/10.1016/j.jbusres.2019.09.062

Glicksohn, J. (2007). Review of Arthur R. Jensen (2006)--Clocking the mind: Mental chronometry and individual differences. Canadian Journal of Experimental Psychology, 61(2), 154–158. https://doi.org/10.1037/cep2007_2_154

Gupta, P., Varshney, A., Khan, M. R., Ahmed, R., Shuaib, M., & Alam, S. (2023). Unbalanced Credit Card Fraud Detection Data: A Machine Learning-Oriented Comparative Study of Balancing Techniques. Procedia Computer Science, 218, 2575–2584. https://doi.org/10.1016/j.procs.2023.01.231

Gupta, S., Leszkiewicz, A., Kumar, V., Bijmolt, T. H. A., & Dmitriy, P. (2020). Digital Analytics: modeling for insights and new methods. Journal of Interactive Marketing, 51, 26–43. https://doi.org/10.1016/j.intmar.2020.04.003

Hair, J. F. (2010). Multivariate data analysis : a global perspective. https://ci.nii.ac.jp/ncid/BB03463866

Hair, J. F., Black, W. C., & Anderson, R. E. (2010). Multivariate data analysis (7th ed.). Pearson, New York.

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2014). A primer on partial least squares structural equation modeling (PLS-SEM). http://ci.nii.ac.jp/ncid/BB15179462

Hasan, M., Popp, J., & Oláh, J. (2020). Current landscape and influence of big data on finance. Journal of Big Data, 7(1). https://doi.org/10.1186/s40537-020-00291-z

Havard. (2018). THE AGE OF PERSONALIZATION crafting a finer edge. Havard Business Review. Retrieved February 10, 2024, from https://hbr.org/sponsored/2018/09/the-age-of-personalization

Henseler, J., Ringle, C. M., & Sarstedt, M. (2014). A new criterion for assessing discriminant validity in variance-based structural equation modeling. Journal of the Academy of Marketing Science, 43(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Hock, K., & Author_Id, N. (2022). Big data analytics in the German banking sector using the example of retail banking. Account and Financial Management Journal, 07(02). https://doi.org/10.47191/afmj/v7i2.01

Holmlund, M., Van Vaerenbergh, Y., Ciuchita, R., Ravald, A., Sarantopoulos, P., Ordenes, F. V., & Zaki, M. (2020). Customer experience management in the age of big data analytics: A strategic framework. Journal of Business Research, 116, 356–365. https://doi.org/10.1016/j.jbusres.2020.01.022

Hongyuran, W., Yuzhe, Z., & Ling, W. (2015, September 30). Online financiers banking on China's future. Caixin Global. Retrieved February 24, 2024,

from https://www.caixinglobal.com/2015-09-30/online-financiers-banking-on-chinas-future-101012157.html

Hosseini, M., Abdolvand, N., & Harandi, S. R. (2022). Two-dimensional analysis of customer behavior in traditional and electronic banking. Digital Business, 2(2), 100030. https://doi.org/10.1016/j.digbus.2022.100030

Hoyer, W. D., Kroschke, M., Schmitt, B. H., Kraume, K., & Shankar, V. (2020). Transforming the customer experience through new technologies. Journal of Interactive Marketing, 51, 57–71. https://doi.org/10.1016/j.intmar.2020.04.001

Hoyle, R. H. (2012). Handbook of structural equation modeling. http://cds.cern.ch/record/1615690

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal, 6(1), 1–55. https://doi.org/10.1080/10705519909540118

Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P., He-Guelton, L., & Caelen, O. (2018). Sequence classification for credit-card fraud detection. Expert Systems With Applications, 100, 234–245. https://doi.org/10.1016/j.eswa.2018.01.037

Karpoff, J. M. (2021). The future of financial fraud. Journal of Corporate Finance, 66, 101694. https://doi.org/10.1016/j.jcorpfin.2020.101694

Kass, R. A., & Tinsley, H. E. A. (1979). Factor Analysis. Journal of Leisure Research, 11, 120-138. https://doi.org/10.1080/00222216.1979.11969385

Kovács, T., Kő, A., & Asemi, A. (2021). Exploration of the investment patterns of potential retail banking customers using two-stage cluster analysis. Journal of Big Data, 8(1). https://doi.org/10.1186/s40537-021-00529-4

Kshetri, N. (2016). Big data's role in expanding access to financial services in China. International Journal of Information Management, 36(3), 297–308. https://doi.org/10.1016/j.ijinfomgt.2015.11.014

Klaus, P., & Maklan, S. (2012). EXQ: a multiple-item scale for assessing service experience. Journal of Service Management, 23(1), 5–33. https://doi.org/10.1108/09564231211208952

Kumar, R. (2014). Research Methodology: A Step-by-Step Guide for Beginners (4th ed.). SAGE Publications Ltd, London.

Lamprinakos, G., Μαγρίζος, Σ., Kostopoulos, I., Drossos, D., & Santos, D. (2022). Overt and covert customer data collection in online personalized advertising: The role of user emotions. Journal of Business Research, 141, 308–320. https://doi.org/10.1016/j.jbusres.2021.12.025

Lamrhari, S., Ghazi, H. E., Oubrich, M., & Faker, A. E. (2022). A social CRM analytic framework for improving customer retention, acquisition, and conversion. Technological Forecasting and Social Change, 174, 121275. https://doi.org/10.1016/j.techfore.2021.121275

Li, J. P., Zhang, Y. J., Wu, D. S., & Zhang, W. (2014). Impacts of big data in the Chinese financial industry. The Bridge, 44(4) https://www.nae.edu/Publications/Bridge/ 128772/129193.aspx

Lohan, G., Conboy, K., & Lang, M. (2011). Examining Customer Focus in IT Project Management: Findings from Irish and Norwegian case

studies. Scandinavian Journal of Information Systems, 23(2), 2. http://iris.cs.aau.dk/tl_files/volumes/volume23/no2/Lohanetal-23-2.pdf

Matsimbi, J. C. (2020). Big data adoption challenges and benefits: The case of a South African bank [University of the Witwatersrand]. https://wiredspace.wits.ac.za/items/e672f366-78de-4514-80a0-69a474b78ee8

McColl-Kennedy, J. R., Zaki, M., Lemon, K. N., Urmetzer, F., & Neely, A. (2018). Gaining customer experience insights that matter. Journal of Service Research, 22(1), 8–26. https://doi.org/10.1177/1094670518812182

Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019). Big data analytics and firm performance: Findings from a mixed-method approach. Journal of Business Research, 98, 261–276. https://doi.org/10.1016/j.jbusres.2019.01.044

Milford, T. M. (2016). Keith, T. Z. (2015). Multiple Regression and Beyond: An Introduction to Multiple Regression and Structural Equation Modeling (2nd ed.). New York, NY: Taylor & Francis. Journal of Educational Measurement, 53(2), 248–250. https://doi.org/10.1111/jedm.12108

Mukherjee, A., Smith, R. J., & Turri, A. M. (2018). The smartness paradox: the moderating effect of brand quality reputation on consumers' reactions to RFID-based smart fitting rooms. Journal of Business Research, 92, 290–299. https://doi.org/10.1016/j.jbusres.2018.07.057

Murinde, V., Rizopoulos, E., & Zachariadis, M. (2022). The impact of the FinTech revolution on the future of banking: Opportunities and risks.

International Review of Financial Analysis, 81, 102103. https://doi.org/10.1016/j.irfa.2022.102103

Navaretti, G. B., Calzolari, G., Mansilla-Fernández, J. M., & Pozzolo, A. F. (2018). Fintech and banking. friends or foes? Social Science Research Network. https://doi.org/10.2139/ssrn.3099337

Nobanee, H., Dilshad, M. N., Dhanhani, M. A., Neyadi, M. A., Qubaisi, S. A., & Shamsi, S. A. (2021). Big Data Applications The Banking Sector: A Bibliometric Analysis approach. SAGE Open, 11(4), 215824402110672. https://doi.org/10.1177/21582440211067234

Nusair, K., Hua, N., & Li, X. (2010). A conceptual framework of relationship commitment: e-travel agencies. Journal of Hospitality and Tourism Technology, 1(2), 106–120. https://doi.org/10.1108/17579881011065029

Ogochukwu, P. L. (2022). Big data analytics for competitive advantage in the financial services sector [University of Johannesburg]. https://ujcontent.uj.ac.za/esploro/outputs/9921904707691

O'Rourke, N., & Hatcher, L. (2013). A Step-by-Step Approach to Using SAS for factor analysis and Structural equation Modeling, Second edition. https://support.sas.com/publishing/authors/extras/61314.pdf

Pérez-Martín, A., Pérez-Torregrosa, A., & Vaca, M. (2018). Big Data techniques to measure credit banking risk in home equity loans. Journal of Business Research, 89, 448–454. https://doi.org/10.1016/j.jbusres.2018.02.008

Popoviăź, A., Hackney, R., Tassabehji, R., & Castelli, M. (2016). The impact of big data analytics on firms' high value business performance. Information

Systems Frontiers, 20(2), 209–222. https://doi.org/10.1007/s10796-016-9720-4

PricewaterhouseCoopers. (2017, September 1). The future of banking: A South African perspective. PwC. https://www.pwc.co.za/en/publications/the-future-of-banking.html

Provost, F., D'Alessandro, B., Hook, R., Zhang, X., & Murray, A. T. (2011). Audience selection for On-Line brand advertising: Privacy-Friendly social network targeting. Social Science Research Network. https://doi.org/10.2139/ssrn.1852644

Radia, K. N., Purohit, S., Desai, S., & Nenavani, J. (2022). Psychological comfort in service relationships: A mixed-method approach. Journal of Retailing and Consumer Services, 64, 102824. https://doi.org/10.1016/j.jretconser.2021.102824

Reichheld, F.F. (1996) The Loyalty Effect. Harvard Business School Press, Boston.

Rialti, R., Zollo, L., Ferraris, A., & Alon, I. (2019). Big data analytics capabilities and performance: Evidence from a moderated multi-mediation model. Technological Forecasting and Social Change, 149, 119781. https://doi.org/10.1016/j.techfore.2019.119781

Rodríguez-Ibáñez, M., Casánez-Ventura, A., Castejón-Mateos, F., & Cuenca-Jiménez, P. (2023). A review on sentiment analysis from social media platforms. Expert Systems with Applications, 223, 119862. https://doi.org/10.1016/j.eswa.2023.119862

Shahana, T., Vilvanathan, L., & Bhat, A. R. (2023). State of the art in financial statement fraud detection: A systematic review. Technological Forecasting and Social Change, 192, 122527. https://doi.org/10.1016/j.techfore.2023.122527

Sharma, S., Tim, U. S., Wong, J., Gadia, S. K., & Sharma, S. C. (2014). A brief review on leading big data models. Data Science Journal, 13(0), 138–157. https://doi.org/10.2481/dsj.14-041

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017a). Critical analysis of Big Data challenges and analytical methods. Journal of Business Research, 70, 263–286. https://doi.org/10.1016/j.jbusres.2016.08.001

Smarter analytics for banks. (2018, September 19). McKinsey & Company. https://www.mckinsey.com/industries/financial-services/our-insights/smarter-analytics-for-banks

South Africa interest rate. (2023). Trading Economics. Retrieved May 8, 2023, from https://tradingeconomics.com/south-africa/interest-rate

South Africa - Major Banks analysis: Thriving through turbulence PWC's analysis of South Africa's major banks' results for the reporting period ended 30 June 2023. (2023, September). PWC South Africa. Retrieved February 23, 2024, from https://www.pwc.co.za/en/publications/major-banks-analysis.html.

Sue, V. M., & Ritter, L. A. (2012). Conducting Online Surveys. SAGE.

Tanwar, R. (2013). Porter's Generic Competitive Strategies. IOSR Journal of Business and Management, 15(1), 11–17. https://doi.org/10.9790/487x-1511117

Tena, M. Á. M., Monferrer, D., & Guillén, M. E. (2019). Customer engagement, non-transactional behaviours and experience in services. International Journal of Bank Marketing, 37(3), 730–754. https://doi.org/10.1108/ijbm-04-2018-0107

The value of getting personalization right—or wrong—is multiplying. (2021, November 12). Mckinsey & Company. Retrieved February 23, 2024, from https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/the-value-of-getting-personalization-right-or-wrong-is-multiplying.

Tseng, H. (2023). Customer-centered data power: Sensing and responding capability in big data analytics. Journal of Business Research, 158, 113689. https://doi.org/10.1016/j.jbusres.2023.113689

Tseng, H., Aghaali, N., & Hajli, N. (2022). Customer agility and big data analytics in new product context. Technological Forecasting and Social Change, 180, 121690. https://doi.org/10.1016/j.techfore.2022.121690

Vallentyn, G. (2021). Exploring How Traditional Banks In Sa Can Improve Their Competitiveness Through Investment In Big Data Technology. Research Gate. https://www.researchgate.net/publication/356791034

Vasarhelyi, M. A., Kogan, A., & Tuttle, B. (2015). Big Data in Accounting: An Overview. Accounting Horizons, 29(2), 381–396. https://doi.org/10.2308/acch-51071

Vidgen, R., Shaw, S., & Grant, D. (2017). Management challenges in creating value from business analytics. European Journal of Operational Research, 261(2), 626–639. https://doi.org/10.1016/j.ejor.2017.02.023

Vinzi, V. E., Trinchera, L., & Amato, S. (2009). PLS Path modeling: from foundations to recent developments and open issues for model assessment and improvement. In Springer eBooks (pp. 47–82). https://doi.org/10.1007/978-3-540-32827-8_3

Vorobyev, I. A., & Anna, K. (2022). Reducing false positives in bank anti-fraud systems based on rule induction in distributed tree-based models. Computers & Security, 120, 102786. https://doi.org/10.1016/j.cose.2022.102786

Wagner, C., Kawulich, B., & Garner, M. (2012). Doing Social Research: A Global Context. UK Higher Education Humanities & Social Sciences Sociology.

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J. F., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. Journal of Business Research, 70, 356–365. https://doi.org/10.1016/j.jbusres.2016.08.009

Wedge, R., Kanter, J. M., Veeramachaneni, K., Rubio, S. M., & Pérez, S. I. (2019). Solving the false positives problem in fraud prediction using automated feature engineering. In Lecture Notes in Computer Science (pp. 372–388). https://doi.org/10.1007/978-3-030-10997-4_23

Xu, Z., Frankwick, G. L., & Ramirez, E. (2016). Effects of big data analytics and traditional marketing analytics on new product success: A knowledge

fusion perspective. Journal of Business Research, 69(5), 1562–1566. https://doi.org/10.1016/j.jbusres.2015.10.017

Yuan, K., Chi, G., Zhou, Y., & Yin, H. (2022). A novel two-stage hybrid default prediction model with k-means clustering and support vector domain description. Research in International Business and Finance, 59, 101536. https://doi.org/10.1016/j.ribaf.2021.101536

Zanker, M., Rook, L., & Jannach, D. (2019). Measuring the impact of online personalisation: Past, present and future. International Journal of Human-Computer Studies, 131, 160–168. https://doi.org/10.1016/j.ijhcs.2019.06.006

Zhu, X., Ao, X., Qin, Z., Chang, Y., Liu, Y., He, Q., & Li, J. P. (2021). Intelligent financial fraud detection practices in post-pandemic era. The Innovation, 2(4), 100176. https://doi.org/10.1016/j.xinn.2021.100176

Zohar, S. (2018, June 17). The true cost of declined orders. Riskified. Retrieved February 24, 2024, from https://www.riskified.com/blog/true-cost-declined-orders.

Zoldi, S. M. (2015). Using anti-fraud technology to improve the customer experience. Computer Fraud & Security, 2015(7), 18–20. https://doi.org/10.1016/s1361-3723(15)30067-1

APPENDIX A

a. Research Questionnaire

Table 16. Research Questionnaire

7-point Likert scale (1 = strongly disagree, 7 = strongly agree)

Consent	
Demographic Information	
Gender	Male/ Female/N/A
Age Range	18-26, 27-42, 43-58, above 58
Highest qualification	Secondary School, Diploma, Degree, Post graduate degree
Please indicate the level of expectations you have on your bank on their product/ service offerings, personalised engagements, credit risk assessments, and fraud detection when it comes to improving your experience as a customer.	
Product/ service offerings offered by the bank include cheques and savings accounts, credit card, vehicle finance, home loans etc.	
Personalised engagements relate to communications from the bank to customers	
fraud detection refers to the bank's ability to prevent theft of customer's money or identity	

Questionnaire		
Original	**Developed/ new**	
Measures	**Measures**	**Source**
Personalisation	**Product and services**	
The bank keeps me up to date new about offers	I need my bank to keep me up to date with new offers	Fernandes and Pinto (2019)
The bank provides all the products I need	I need my bank to provide me all the products I specifically need/ tailored offers	Fernandes and Pinto (2019)
Tailor messaging to my needs	I need my bank to tailor messaging to my needs	McKinsey & Co. 2021
Offer me targeted promotions	I need my bank to offer me targeted promotions	McKinsey & co. 2021
celebrate my milestones	I need my bank to celebrate my milestones and reward me accordingly	McKinsey & co. 2021
Send me timely communications tied to key moments	I need my bank to send me timely communications tied to key moments	McKinsey & Co. 2021
Send triggers based on my behaviour	I need my bank to send triggers based on my behaviour	McKinsey & Co. 2022

Fraud detection	**Fraud detection**	
Had a good Fraud experience	I would like my bank to notify me when there is potential fraud on my account	McKinsey & Co. 2022
	I would like my bank to have stringent measures in verifying my identity to prevent fraudulent transactions	
	I would like my bank to proactively prevent fraudulent transactions on my behalf	
Customer Experience (Net Promoter's Score: Word Of Mouth)	**Customer Experience (based on NPS: the likelihood that you'll have a good word of mouth)**	
I would speak positively about the [bank] to others	I would recommend a bank that tailors their products and services to my needs	Fernandes and Pinto (2019)
I would recommend the [bank] to someone who seeks my advice	I would recommend a bank that has high levels of personalisation	Fernandes and Pinto (2019)
I would recommend the [bank] to family members and close personal friends	I would recommend a bank that has a flexible credit assessment approach	Fernandes and Pinto (2019)
	I would recommend a bank that has the ability to track/ stop fraud	
Customer Retention and acquisition	**Customer Retention and acquisition**	

The bank will look after me for a long time	I will join or stay with a bank that looks after me	Fernandes and Pinto (2019)
The bank knows me and takes good care of me, I won't leave it	I will join or stay with a bank that knows me and takes care of my needs	Fernandes and Pinto (2019)
I rather stay with my bank because I'm not confident about alternative providers	I will join or stay with a bank that protects me from fraudsters	Fernandes and Pinto (2019)

b. Table 17. Proposed schedule and timelines

Activity	Description	Timelines
Proposal Submission	• Research Proposal and data collection instruments submission • Ethical Clearance Forms and accompanying documents	30 June 2023
Panel presentation	Present proposal before panel committee	10- 17 July 2023
Title Approval	Title tabled before PGC	10- 17 July 2023
Ethics Approval	Ethics applications submitted to the WBS ethics committee	10- 17 July 2023
Instrument Pilot	To test the validity of the instrument	August 2023
Research Progress	Data collection, analysis, and write up	August 2023- January 2024
First Final Draft	Completion of the first final draft project report	January 2024

Final Research Report Submission	Final editing and electronic submission to the faculty office	28 ebruary 2024

c. Electronic Respondent consent form

Research Topic: Big Data Analytics as a Customer Retention and Acquisition Strategy within South African Retail Banking

Name of Researcher: Muano Kharidzha

This consent form asks you to allow the researcher to collect the level of expectations you have on your bank when it comes to improving your experience in the various interactions you have with them. Your answers will enhance the understanding of the topic being researched.

I agree to participate in the study. YES NO

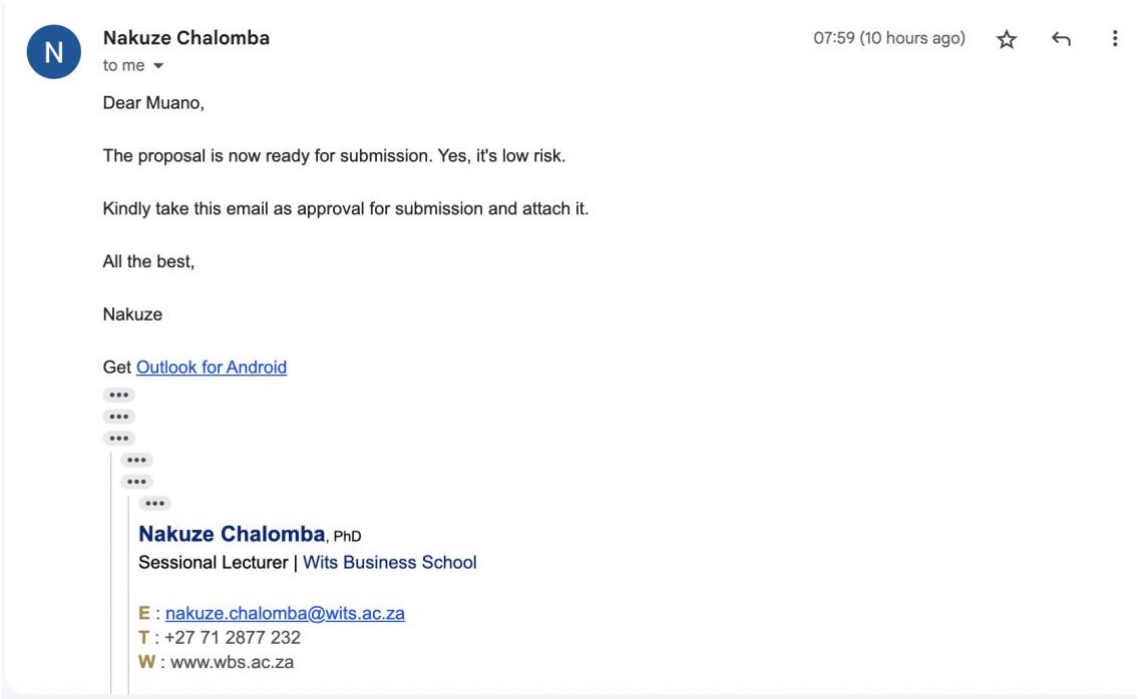
I understand what the study is about. YES NO

I understand that I am volunteering to take part in the study and that there will be no negative consequences if I choose not to participate or decide not to answer any specific question. YES NO

I understand that my participation in this survey is anonymous (with no identity information required). YES NO

I understand that the researcher will use my responses to the survey questions in their research report. YES NO

d. Supervisor proposal support letter



Nakuze Chalomba 07:59 (10 hours ago)

to me

Dear Muano,

The proposal is now ready for submission. Yes, it's low risk.

Kindly take this email as approval for submission and attach it.

All the best,

Nakuze

Get Outlook for Android

Nakuze Chalomba, PhD
Sessional Lecturer | Wits Business School

E : nakuze.chalomba@wits.ac.za
T : +27 71 2877 232
W : www.wbs.ac.za