KENYA BANKERS
ASSOCIATION
One Industry. Transforming Kenya.

WPS/07/22

# The impact of Artficial Intelligence and how it is shaping Banking

By Joseph Theuri and John Olukuru

February 2022

KBA Centre for Research on Financial Markets and Policy®
Working Paper Series

61

KENYA BANKERS
ASSOCIATION
One Industry. Transforming Kenya.

# Working Paper Series

## Centre for Research on Financial Markets and Policy

The Centre for Research on Financial Markets and Policy® was established by the Kenya Bankers Association in 2012 to offer an array of research, commentary, and dialogue regarding critical policy matters that impact on financial markets in Kenya. The Centre sponsors original research, provides thoughtful commentary, and hosts dialogues and conferences involving scholars and practitioners on key financial market issues. Through these activities, the Centre acts as a platform for intellectual engagement and dialogue between financial market experts, the banking sector and the policy makers in Kenya. It therefore contributes to an informed discussion that influences critical financial market debates and policies.

The Kenya Bankers Association (KBA) *Working Papers Series* disseminates research findings of studies conducted by the KBA Centre for Research on Financial Markets and Policy. *The Working Papers* constitute "work in progress" and are published to stimulate discussion and contribute to the advancement of the banking industry's knowledge of matters of markets, economic outcomes and policy. Constructive feedback on the *Working Papers* is welcome. *The Working Papers* are published in the names of the author(s). Therefore their views do not necessarily represent those of the KBA.

# The Impact of Artificial Intelligence and how it is Shaping Banking

By Joseph Theuri and John Olukuru


## Abstract

*The importance and adoption of Artificial Intelligence schemes in supporting business operations in risk management and spurring revenue growth continues to gain traction globally. While this has been exacerbated by the disruptions caused by COVID-19 pandemic on the traditional sources of information, its utilization remains low particularly across many countries. In advanced economies however, as AI gains popularity in banking, financial institutions (FIs) are building on their existing solutions to transform customer experiences to solve increasingly complex challenges and expectations.*

*This paper illustrates the potential of employing AI in banking to reduce costs, including through opportunities it offers to banks to leverage algorithms on the front end to smooth customer identification and authentication, mimic live employees through chatbots and voice assistants, deepen customer relationships, and provide personalized insights and recommendations. Further, AI can also be used by banks within middle-office functions to assess risks, detect and prevent payments fraud, improve processes for anti-money laundering (AML) and perform know-your-customer (KYC) regulatory checks. The main output involves an interactive dashboard illustrating application of the descriptive and predictive analytics at a click for a given business unit of a bank.*

# 1.0 Introduction

**Artificial Intelligence is typically defined as the ability of a machine to perform cognitive functions we associate with human minds**, such as perceiving, reasoning, learning, interacting with the environment, problem solving, and even exercising creativity.

However, Artificial Intelligence (AI) is actually a combination of advanced computational technologies in varying degrees of maturity. Some of these technologies have been around for decades while others are relatively new. It is however an incremental technological evolution, sometimes based on old technologies, which has now been made possible by access to large volumes of data and new capacities in processing these volumes of data. The wording "Artificial Intelligence" is based on terms that can sometimes seem overused as it is used generically to cover multiple technologies. Most of the time, the use of algorithms is limited to mimicking scenarios; reproducing and automating the processing of repetitive tasks that a human being could perform. By AI it is generally "cognitive technologies that rely on large volumes of structured or unstructured data (big data)" that is meant. In this sense, "cognitive intelligence" is defined as any unstructured data processing, modelling that emulates and/or allows to augment and enhance the cognitive abilities of humans. For example, here is a non-exhaustive list of some areas of technological research:

a. Natural language processing: understands language by attributing meaning and purpose. It is often associated with Automatic Speech Recognition and Text to Speech.

b. Cognitive computing: support in the realization of cognitive tasks and decision making. These are interactive, iterative and evidence-based systems.

c. Smart analytics/ processing: predictive analysis and simulations. These provide support for rule-based automatic actions (e.g. recommendation engines).

d. Deep learning and reinforcement learning: structure that moves the focus of machine learning from pattern recognition to a process of sequential and experience-driven decision making.

01
ONE

Therefore, AI encompasses a set of principles, problem definitions, algorithms, and processes for extracting non-obvious, useful patterns and actionable insight from large data sets. **AI** have become the simulation of human intelligence processes by machines, especially computer systems. Banks can use AI to transform the customer experience by enabling frictionless, 24/7 customer service interactions. Hence, the unique purpose of AI technologies is to mimic the human brain, mainly within the following categories: visual perception, speech recognition, decision making and translation.

AI technologies can be classified into two types: First, Artificial machine intelligence which Refers to a top down, human engineered approach in which, the computer executes tasks and makes decisions based on predefined rules and requirements. Secondly, natural machine intelligence that Implies a bottom-up approach in which the computer uses self-improvement algorithms to increase its performance and abilities for each task handled.

Artificial Intelligence in finance is more than about chat bots. Artificial Intelligence has taken over numerous sectors including banking industry. The principal thought behind this investigation was to comprehend the impact of AI on present day banking. This paper mainly focuses on the concept of AI in the field of banking, how it has brought revolutionary changes in banking and its impact on human manpower. As we are aware that humans tend to commit errors, but the world is evolving so does the innovations, there is lack of skilled talents required to handle the automation.

Several digital disruption is redefining industries and changing the way businesses function. Every industry is assessing options and adopting ways to create value in the technology-driven world. The banking sector is witnessing groundbreaking changes: foremost being the rise in customer-centricity. Artificial Intelligence is the future of banking as it brings the power of advanced data analytics to combat fraudulent transactions and improve compliance. Artificial intelligence is now becoming more widespread in the current market. It is used in various sectors; banking industry is one among them. Banking industry is using AI in a very innovative way which save plenty of time and money. The banks use algorithms to generate accurate results which in turn help in enhancing customer service and generate better sales performance to deliver profits. AI includes machine learning and profound learning which helps to reduce errors caused by emotional and psychological factors. One of the most important task AI performs is to channelize key information from wide variety of data and draw conclusions.

Over the past decades, banks have been improving their methods of interacting with customers. They have tailored modern technology to the specific character of their work. As an example, in the 1960s, the first ATMs were installed, and ten years later, there were already cards for doing transactions and payment. At the beginning of this century, users learned about round-the-clock online banking, and in year 2010, they heard about mobile banking. But the development of the financial system didn't stop here, as the digital age is opening up new opportunities — the use of Artificial Intelligence in banking and financial institutions. By 2023, banks are projected to save approximately $447 billion by developing and implementing AI applications.

The Impact of AI in banking and how AI is being used has been on the rise and especially in 2021 as post covid 19 era has moved most customers to be digital. Adoption of AI solutions in banking has become more mainstream: A majority of financial services companies say they've implemented the technology in business domains like risk management (56%) and revenue generation through new products and processes (52%), per the Cambridge Centre for Alternative Finance and the World Economic Forum. As AI gains popularity in banking, financial institutions (FIs) are building on their existing solutions to solve increasingly complex challenges. Banks in developed world are using AI to transform the customer experience by enabling frictionless, 24/7 customer service interactions — but AI in banking applications isn›t just limited to retail banking services. The back and middle offices of investment banking and all other financial services for that matter could also benefit from AI. A majority of financial services firms have implemented AI in risk management or revenue generation.

The aggregate potential cost savings for banks from AI applications is estimated at $447 billion by 2023, with the front and middle office accounting for $416 billion of that total, per Autonomous Next research seen by Business Insider Intelligence. Analysts also estimated that AI would save the banking industry over $1 trillion by 2030 (Maskey, 2018). Moreover, with the increasing customer demand of convenience and quickness not only at work but in all aspects of everyday life, AI technologies have become indispensible tools for banks to step their game up in this highly competitive market. Accenture (2018) suggested that AI would profoundly change the way banks work, the products they sell and the way they interact with customers and employees. With the emergence of AI, the banking work - force will be able to move away from repetitive, process-driven tasks towards the more strategic and innovative kinds of work that will ultimately drive the industry forward. It will create seamless interplays among customers, employees and AI-led services. And it will break through the silos and practices of process-driven banking, allowing banks to become analytics-driven entities, using data to dynamically inform and shape what they do in real time.

The trend of increasing use of AI in banking is clear. Banks in Hong Kong have been integrating the technology into various key functional areas including front-line businesses, risk management, back office operations and customer services. According to a survey conducted by the HKMA in August 2019, over 80% of the participating banks view AI adoption as a way of reducing operating costs, improving efficiency and strengthening risk management. Reflecting optimism about the prospects of broader AI adoption, some 80% of the banks plan to increase investment in the technology over the next five years. The broader use of AI will create new opportunities, but also pose new risks and challenges to banks, including the lack of quality data and data protection, and difficulty in explaining and validating AI models. Banks participating in the survey are aware of these risks and they have regular reviews to identify AI risks (68% of AI-utilising banks) and clear procedures to address model defects (70%). Banks in Hong Kong also highlight additional challenges including issues related to development such as shortage of talent, technical aspects such as increased complexity of AI models, and issues associated with the evolving regulatory environment. The growing use of AI applications in online and mobile banking may expose banks to new cyber threats. Banks will need to identify potential weaknesses in their cyber defence systems by conducting regular tests, and assess the resilience of their AI applications to more sophisticated cyberattacks. Strengthening the cybersecurity of the most important and vulnerable operations of banks and enhancing the security features of cloud computing will become increasingly important.

# 2.0 Literature Review

**The application of artificial intelligence (AI) in banking has advanced to a new level, thanks to maturing techniques in big data analytics and machine learning, as well as enhancements in computational power.** In Hong Kong, the use of AI in the banking industry is expanding to cover key functional areas including frontline businesses, risk management and back office operations. These new technologies are increasingly used to perform more sophisticated tasks such as credit assessments and fraud detection. They also enable banks to better serve their customers, providing the convenience of remote onboarding for account opening and lending.

Hand & Henley, 1997 argued that "credit scoring is the term used to describe formal statistical methods which are used for classifying applicants for credit into "good" and "bad" risk classes". Credit scoring models are multivariate statistical models applied to economic and financial indicators to predict the default risk of individuals or companies. These indicators are assigned a weight relative of importance in predictions, and are fed as input to arrive at an index of creditworthiness. This numerical score serves as a measure of the borrower's probability of default. The support vector machine technique was concluded as being the most widely applied in credit risk evaluations. Hybrid SVM models have been proposed to improve the performance by adding methods for the reduction in the feature subset. These, however, only classify, and don't provide an estimation of the probability of default (Keramati & Yousefi, 2011).

The dramatic growth in consumer credit has increased the importance of credit scoring models. The bulk of the research appears to be focussed on credit scoring techniques, as seen in the number of papers focussing on this area (Ala'raj & Abbod, 2016a; Ala'raj & Abbod, 2016b; Bellotti & Crook, 2009; Zhou et al., 2013; Van Gestel et al., 2003; Addo et al., 2018; C.-L. Huang et al., 2007; Keramati & Yousefi, 2011; Lai et al., 2006; Lessmann et al., 2015; Van Sang et al., 2016; Chen & Huang, 2003; Wang et al., 2015; Wójcicka, 2017). Predominantly, the focus is on classification and the application of algorithms that enable this. A number of papers evaluate the various algorithms in an attempt to identify the most efficient and accurate prediction algorithm. The papers make a case that machine learning delivers comparable accuracy and is better equipped to capture non-linear relationships common to credit risk (Bacham & Zhao, 2017; Zhang et al., 2017).

G. Zhou & Wang, 2012 propose allocating weights to decision trees for better prediction. They put forward an improved random forest algorithm for predictions. The algorithm, during aggregation, allocates weights which are calculated based on out-of-bag errors in training to the decision trees in the forest. They attempt to address the binary classification problem, and their experiment shows that the proposed algorithm beats the original random forest and other popular classification algorithms (SVM, KMM, C4.5) in terms of balanced and overall accuracy metrics.

Mhlanga and Denhere (2021) investigated the application of credit risk assessment using machine learning in Southern Africa. Using the logistic regression, the study discovered that financial inclusion is driven by many factors including age, education level, income, race gender and marital status.

Bussmann et al., 2021 proposed an explainable AI model that can be used in the analysis of credit risk and management particularly when credit is borrowed in peer-to-peer lending platforms. The model that they proposed is to apply correlation networks to Shapley values to ensure that AI predictions are grouped concerning similarity in the underlying explanations. Using this model they analysed 15,000 small and medium firms by asking for credit that reveals both risky and not risky borrowers to be grouped with their financial characteristics. Bussmann et al., 2021 believe that it is possible to explain the credit score of the borrowers and to subsequently predict their future behaviour. Punniyamoorthy & Sridevi, 2016 also stated that credit risk assessments have gained a lot of attention in recent years motivated by the global financial crisis and credit crunch. As a result, various financial institutions seek to support credit rating agencies to come up with predictions on the ability of the creditor to meet financial obligations. Therefore, they prepared a neural network (NN) and fuzzy support vector machine (FSVM) to discriminate between good creditors and bad creditors and came up with the best classifier for credit risk assessments. They discovered that the FSVM model performs better than the backpropagation neural network.

Moscatelli et al., 2020 came forth with an analysis performance for machine learning models in predicting default risk. The study used standard statistical models like logistic regression as a benchmark. The study discovered that machine learning models give meaningful gains in discriminatory power and precision compared to statistical models. The benefits of these models diminish when confidential information like credit behavioural indicators is also available, and they become negligible when the data set is small. Moscatelli et al., 2020 also assessed the consequences of the use of credit allocation rule based on machine learning ratings on the overall supply of credit and the number of borrowers gaining access to credit. Machine learning models proved to assist lenders to offer credit towards safer and larger borrowers which result in lower credit losses for lenders.

# 3.0 Methodology

## 3.1 Introduction

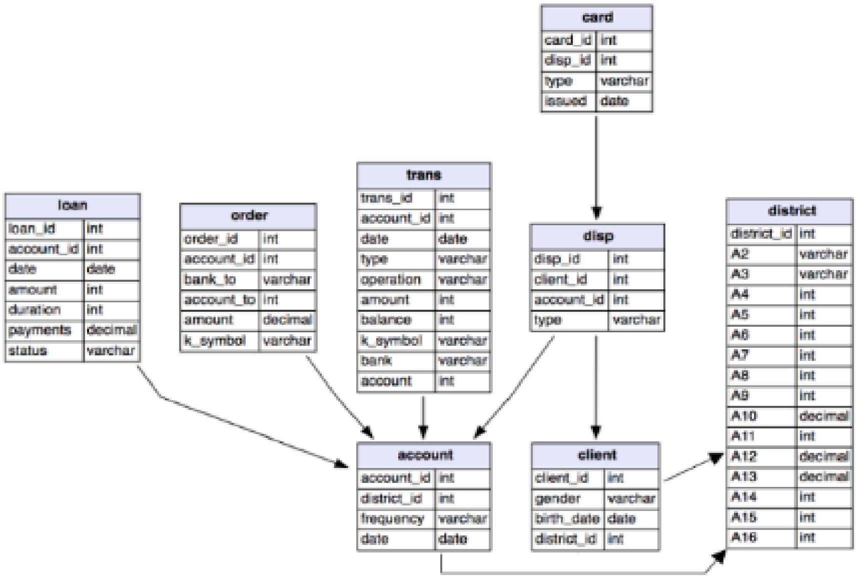
**Given the emphasis on technology aspects of AI and ML applications to the banking industry we develop a dashboard containing key information as well as results from a ML model running in the backend.**

## 3.2 Data

In this study we utilize the Berka Dataset, or the PKDD'99 Financial Dataset, which is a collection of real anonymized financial information from a Czech bank, used for PKDD'99 Discovery Challenge.

In the dataset, 8 raw files include 8 tables:

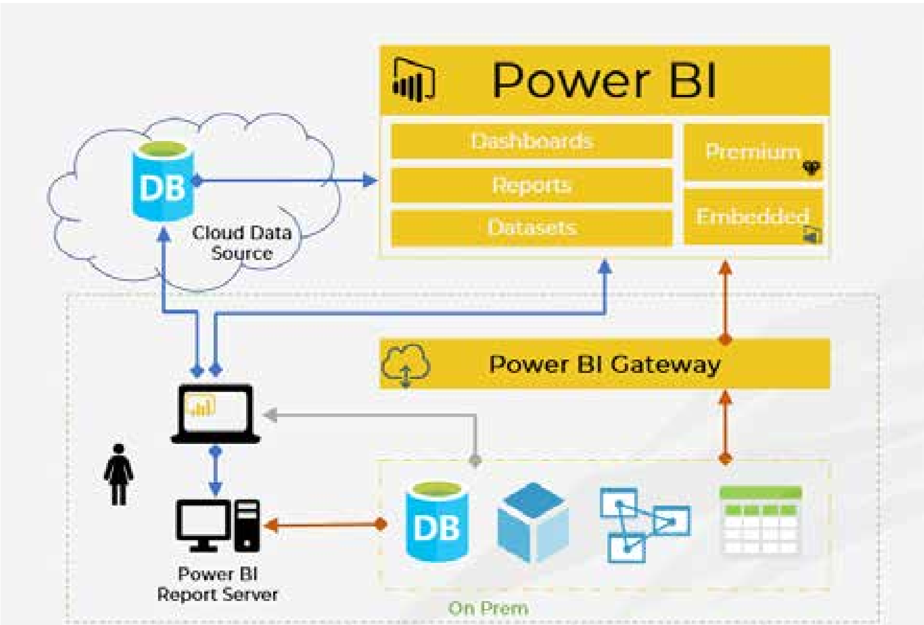
- account (4500 objects in the file ACCOUNT.ASC) – each record describes static characteristics of an account.
- client (5369 objects in the file CLIENT.ASC) – each record describes characteristics of a client.
- disposition (5369 objects in the file DISP.ASC) – each record relates together a client with an account i.e. this relation describes the rights of clients to operate accounts.
- permanent order (6471 objects in the file ORDER.ASC) – each record describes characteristics of a payment order.
- transaction (1056320 objects in the file TRANS.ASC) – each record describes one transaction on an account.
- loan (682 objects in the file LOAN.ASC) – each record describes a loan granted for a given account.
- credit card (892 objects in the file CARD.ASC) – each record describes a credit card issued to an account.
- demographic data (77 objects in the file DISTRICT.ASC) – each record describes demographic characteristics of a district.




- Each account has both static characteristics (e.g. date of creation, address of the branch) given in relation "account" and dynamic characteristics (e.g. payments debited or credited, balances) given in the relations "permanent order" and "transaction".
- Relation "client" describes the characteristics of persons who can manipulate the accounts.
- One client can have more accounts, more clients can manipulate with a single account; clients and accounts are related together in relation "disposition".
- Relations "loan" and "credit card" describe some services which the bank offers to its clients.
- More than one credit card can be issued to an account.
- At most one loan can be granted for an account.
- Relation "demographic data" gives some publicly available information about the districts (e.g. the unemployment rate); additional information about the clients can be deduced from this.

We first import the data into a MYSQL database, cleaned the variables and connect the database to Power BI for visualization

## 3.3 Business Intelligence

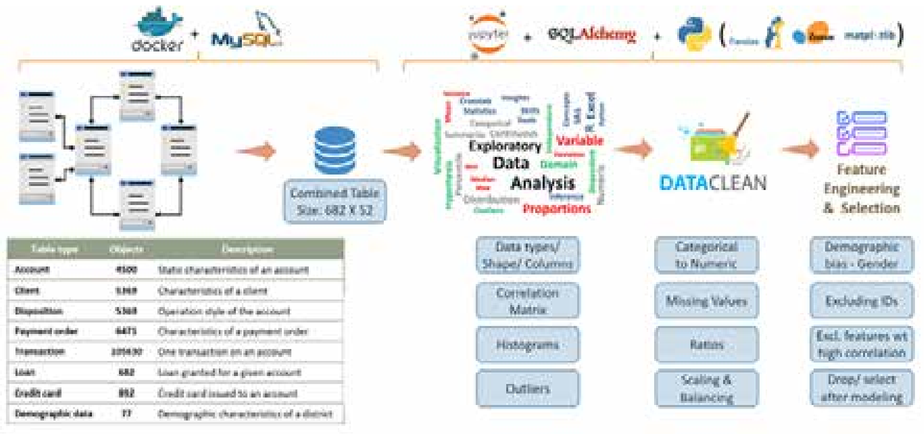



Microsoft Power BI is a business intelligence platform that provides nontechnical business users with tools for aggregating, analysing, visualizing and sharing data. It can be used to find insights within an organization's data. Power BI can help connect disparate data sets, transform and clean the data into a data model and create charts or graphs to provide visuals of the data. In the end, all of this can be shared with other users within the organization.

The data models created from Power BI can be used to create data stories through charts and data visualizations and examining "what if" scenarios within the data. The dashboard reports can also answer questions in real time and help with forecasting to make sure departments meet business metrics.

The software also provides executive dashboards for administrators or managers, giving management more insight into how departments are doing.

### 3.4 Machine Learning




There are several performance measures to compare the performance of the models including AUC, Gini, RMSE and Akaike information criterion (AIC); in addition to different metrics like the F-score, the recall and the precision. In this paper, we will mainly present results on the AUC and RMSE criteria, although the results of the other metrics can be made available.

The Gini index was introduced by Dorfman, 1979 and extended by Lerman and Yitzhaki (1984). This index, in essence, permits one to compare several algorithms. It is based on the decision tree methodology and entropy measure. The work in Raileanu & Stoffel, 2004 discussed the possibility to compare algorithms using classification systems. From the empirical point of view, this problem has been discussed greatly. It seems that no feature selection rule is consistently superior to another and that Gini can be used to compare the algorithms; nevertheless, we will not use it in the paper, focusing on the ROC curve for which the interpretation in terms of risks is more efficient.

For each loan, we build the ROC curve. An ROC curve is then used to evaluate the quality of the model. The ROC approach can be often associated with the computation of an error from a statistical point of view. If we want to associate the AUC value (coming from the ROC building) to Type I and Type II errors, we need to specify the test we consider and, thus, to determine the hypotheses, the statistics, the level and the power of the test. In our case, the objective is to know if a bank can provide a loan to the individual, in the sense that the individual will not default. To answer this question, we use several models or algorithms, and the idea is to find the algorithm that permits answering this question with accuracy. The AUC criteria are used to do it; here, we explain why. To analyze, in a robust way, the results obtained with this indicator, we specify the risks associated with it. When a bank provides a loan to an individual, it faces two types of errors: (1) to refuse a loan to a loan whose probability of default is much lower than the one obtained with the model (Type I error); (2) to agree to provide a loan to a loan whose probability of default is much higher than the value obtained with the selected model (Type II error). We want these two errors to be as small as possible. We compute the probability of the corresponding events under the null and the alternative hypotheses. We assume that a bank provides a loan, and the null hypothesis is $H_0$ : The loan can reimburse the loan, $\alpha$ = P[the bank does not provide a loan |the loan can reimburse it] = $P(H_0)$ [the bank does not provide a loan]; this is the Type I error; thus, the alternative is H1 : The loan does not reimburse the loan, and $\beta$ = P[the bank provides a loan |the loan cannot pay it back] = $P(H_1)$ [the bank provides a loan]; this is the Type II error.

Considering the dataset, the bank could provide a loan as the probability of default of the target loan is sufficiently low (the model outcome has the value of one) or the bank could decide not to provide a loan as the probability of default of the target loan is not low enough (the outcome is now zero). These outcomes, one or zero, depend on many variables that we use to compute the risks $\alpha$ or $\beta$. Note that to build the ROC curve, we make $\alpha$ varying (it is not fixed as it is in the general context of statistical tests). When we build the ROC, on the x-axis, we represent $1 - \alpha$, also called specificity (in some literature). We want this number close to one4. On the y-axis, we represent $1 - \beta$, which corresponds to the power of the test, and we want it to be close to one. It is usually referred to as sensitivity. In practice, when the ROC curve is built, all the codes are done under two kinds of assumptions on the data: the data are independent, and the distributions under the null and the alternative are Gaussian; these assumptions can be far from reality in most cases. From the ROC curve, an AUC is built. The AUC represents the area under the curve. How can we interpret its value? If the curve corresponds to the diagonal, then the AUC is equal to 0.5; we have one chance out of two to make a mistake. If the curve is above the diagonal, the value will be superior to 0.5, and if it attains the horizontal at one, for all $(1 - \alpha)$, the optimal value of one is obtained. Thus, as soon as the AUC value increases from 0.5–1, it means that we have less and less chance to make a mistake, whatever the value of $(1 - \alpha)$ between zero and one (which means that the Type I error diminishes). It is assumed that the test becomes more and more powerful as the probability for the bank to provide a loan to an individual that does not default is very high. Each algorithm provides a value of AUC. To be able to compare the results between the algorithms, we need to verify that we use the same variables to get the outputs one or zero. If that is not the case, the comparison will be difficult and could be biased Manilow et al., 2017.

Another question affects the quality of the results: it concerns imbalanced data. The presence of a strong imbalance in the distribution of the response (which is the case for our exercise) creates a bias in the results and weakens the estimation procedure and accuracy of the evaluation of the results. A dataset is imbalanced if the classification categories are not approximately equally represented. Examples of imbalanced datasets have been encountered in many fields, and some references are Grobelnik, 1999, among others. Several approaches are used to create balanced datasets, either by over-sampling the minority class and under-sampling the majority class; diverse forms of over-sampling can be used such as the Synthetic Minority Over-sampling Technique (SMOTE) algorithm developed by Chawla, 2003. In this paper, the latter methodology has been implemented, blending under-sampling of the majority class with a special form of over-sampling of the minority class associated with a naive Bayes classifier, improving the re-sampling, modifying the loss ratio and class prior approaches.

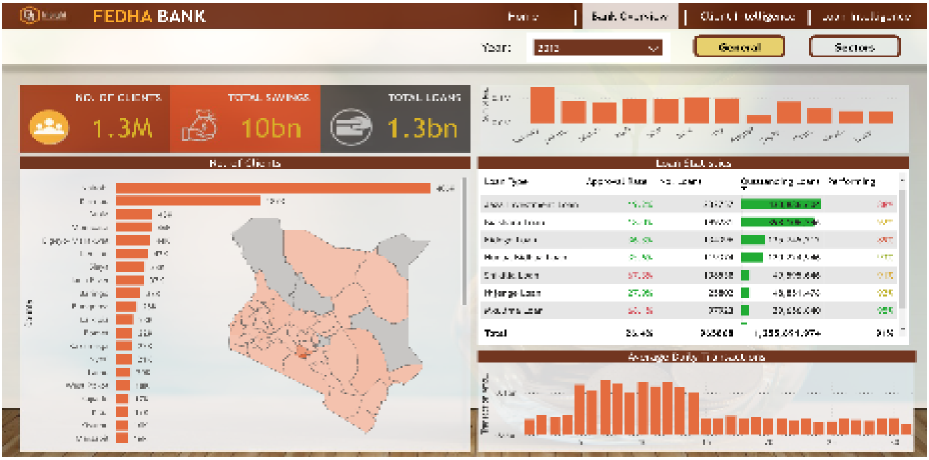
The approach that we use, the SMOTE algorithm, proposes an over-sampling approach in which the minority class is over-sampled by creating "synthetic" examples rather than by over-sampling with replacement. The synthetic examples are generated by operating in "feature space" rather than in "data space". The minority class is over-sampled by taking each minority class sample and introducing synthetic examples along the line segments joining any/all the k minority class neighbors randomly chosen. To generate the synthetic examples, we proceed on the following path: we take the difference between the feature vector under consideration and its nearest neighbor. We multiply this difference by a random number between zero and on and add it to the feature vector under consideration. This causes the selection of a random point along the line segment between two specific features. This approach effectively forces the decision region of the minority class to become more general. The majority class is under-sampled by randomly removing samples from the majority class population until the minority class becomes some specified percentage of the majority class. This forces the learner to experience varying degrees of under-sampling; and at higher degrees of under-sampling, the minority class has a large presence in the training set.

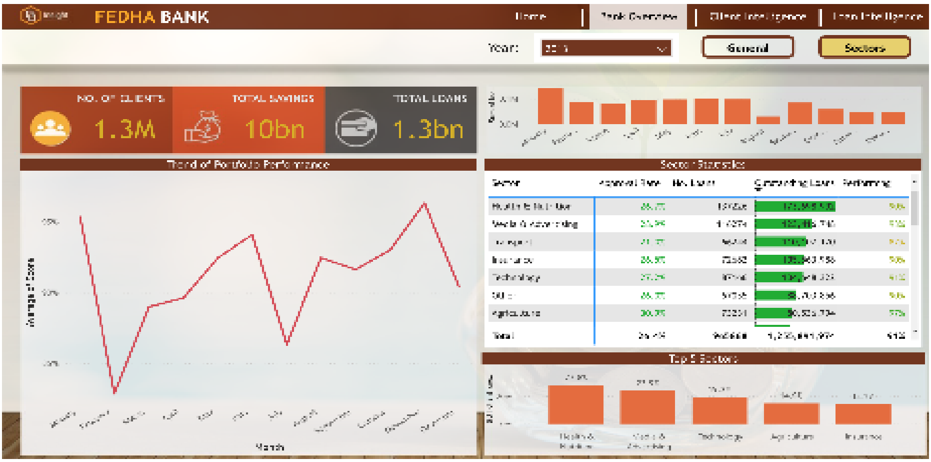
# 4.0 Model Results

**The application of AI can easily be illustrated by the following screenshots of a dashboard that will enable a decision maker in the bank to make decisions at a click of a button.** The dashboard is very dynamic and will dependent on the timely data collection. If the bank has a platform that collects real time data, then the dash board is also update real time.

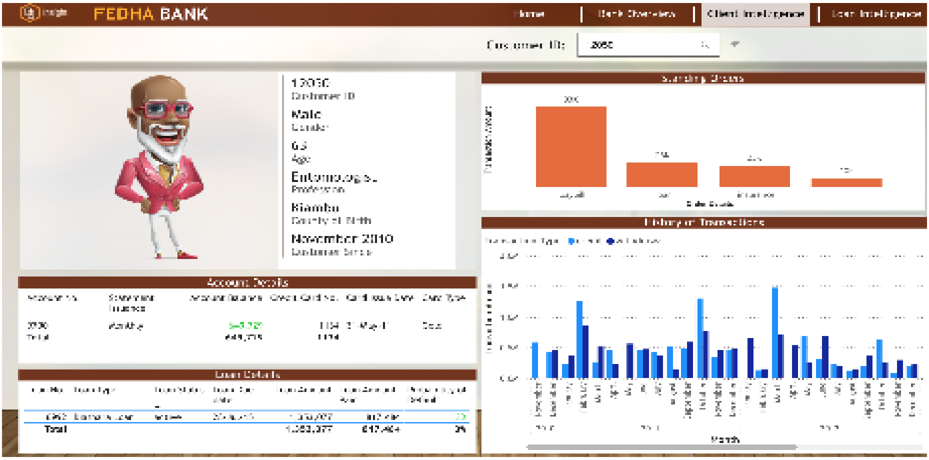
The image above shows the landing page for the Fedha Bank dashboard. It allows users to navigate to any of the three modules including: Bank overview, client intelligence and loan intelligence.

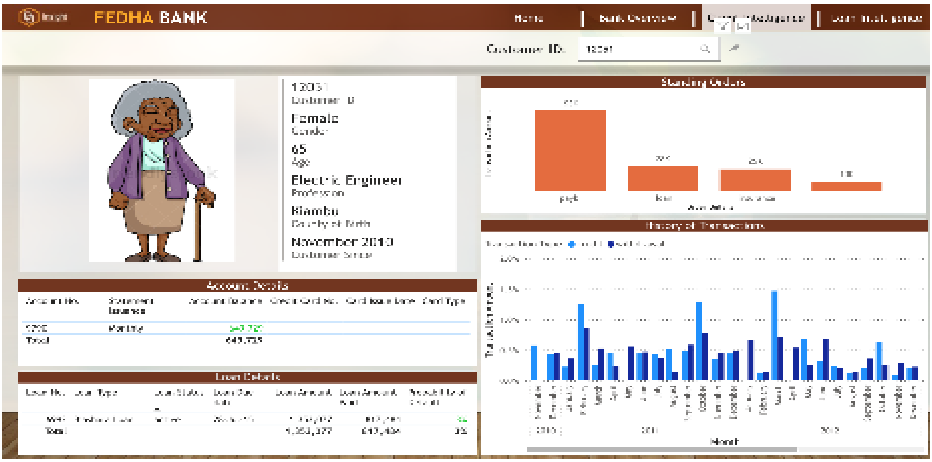
Within the bank overview module two sub-modules exist: General and Sectors. Under the General sub-module summary statistics such as the client base, lending and savings statistics are displayed along with demographic and product specific information. This tab can be extended to display capital adequacy ratios, staffing information or even social media news regarding the bank. On the Sector sub-module information is summarized by sector and allows each sector to be examined across time.

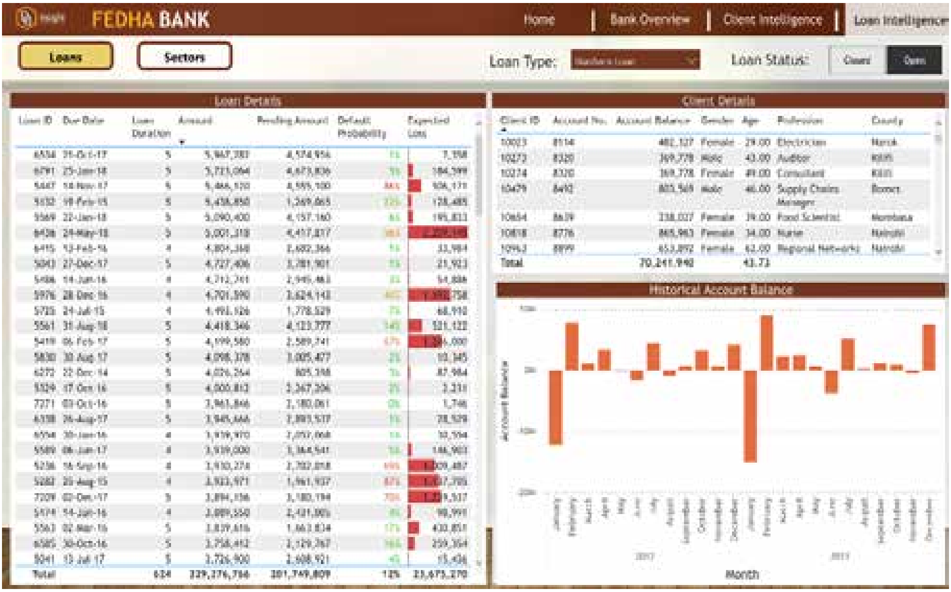



The client intelligence module aims to give a single snapshot of the information pertaining to any bank customer. The example above shows personal information on customer 12050, their account details, loans taken, banking instructions issued as well as historical transactions with the bank. What is interesting to note here is that under the loans section a probability of default is calculated using machine Learning and updates periodically after each customer transaction.


FEDHA BANK
Customer ID:
Male
Entomologist
Biambu
November 2010
Account Details
Loan Details
Standing Orders
History of Transactions
Month


FEDHA BANK
Customer ID:
Female
65
Age
Electric Engineer
Biambu
November 2010
Account Details
Loan Details
Standing Orders
History of Transactions
Month




The Loan Intelligence module allows credit officers to have a view of individual loan performance and develop strategies that could minimize expected loss. Here the default probability is calculated by a Machine Learning algorithm that uses historical loan records. Each Loan is given a default probability based on the loan characteristics as well as the customer attributes and interactions. Credit officers can then sort by highest default probability as well as expected loss then focus on those loans and later perform random sampling on the rest to ensure everything is going okay.

The sector sub-module examines expected loss across all product lines and sectors dealt with by the bank. In addition, it incorporates geographic information to map out regions that may have higher expected-loss compared to other.


Along with the sector performance forecasts for future performance are also computed by a Machine Learning algorithm to show what the likely path in the future would look like based on historical performance.

Below is the list of some of the predictive models used to calculate the probability of default and their accuracy levels

| | Model | Accuracy | AUC | Recall | Prec. | F1 | Kappa |
|---|---|---|---|---|---|---|---|
| 0 | Logistic Regression | 90.2% | 84.9% | 98.1% | 91.5% | 94.7% | 0.32 |
| 1 | CatBoost Classifier | 89.9% | 82.4% | 98.1% | 91.3% | 94.6% | 0.29 |
| 2 | Extreme Gradient Boosting | 89.7% | 82.6% | 96.9% | 92.0% | 94.4% | 0.35 |
| 3 | Linear Discriminant Analysis | 89.5% | 83.0% | 95.8% | 92.7% | 94.2% | 0.41 |
| 4 | Light Gradient Boosting Machine | 89.5% | 80.2% | 96.9% | 91.8% | 94.3% | 0.31 |
| 5 | Ridge Classifier | 89.1% | 0.0% | 99.5% | 89.4% | 94.2% | 0.07 |
| 6 | Quadratic Discriminant Analysis | 88.9% | 50.0% | 100% | 88.9% | 94.1% | - |
| 7 | Extra Trees Classifier | 88.7% | 67.6% | 99.5% | 89.0% | 94.0% | 0.02 |

| | Model | Accuracy | AUC | Recall | Prec. | F1 | Kappa |
|---|---|---|---|---|---|---|---|
| 8 | Gradient Boosting Classifier | 88.7% | 80.2% | 96.5% | 91.3% | 93.8% | 0.27 |
| 9 | Random Forest Classifier | 88.5% | 69.0% | 98.6% | 89.5% | 93.8% | 0.09 |
| 10 | SVM - Linear Kernel | 88.4% | 0.0% | 93.6% | 93.4% | 93.5% | 0.42 |
| 11 | Ada Boost Classifier | 88.3% | 73.9% | 95.3% | 91.8% | 93.5% | 0.31 |
| 12 | K Neighbors Classifier | 86.8% | 64.1% | 97.2% | 89.0% | 92.9% | 0.01 |
| 13 | Decision Tree Classifier | 84.1% | 62.3% | 90.3% | 91.7% | 91.0% | 0.22 |

After model Tuning the Logistic Regression Model achieved the following final out of sample results:

| Accuracy | AUC | Recall | Prec. | F1 | Kappa |
|---|---|---|---|---|---|
| 88% | 83% | 97% | 91% | 94% | 0.24 |

The top three features used to determine Probability of default included:

- Average account balance over last 4 periods
- Total withdrawal amount over last 4 periods
- Type of Loan

# 5.0 Opportunities and benefits that banks in Kenya can tap

**The use cases for AI in banks are organised in three categories, highlighting the potential areas of opportunities for the banking sector.** First, it is about enhancing customer interaction and experience: e.g., chatbots, voice banking, robo-advice, customer service improvement, biometric authentication and authorisation, customer segmentation (e.g., by customized website to ensure that most relevant offer is presented), targeted customer offers; Secondly, enhancing the efficiency of banking processes: e.g., process automation/optimisation, reporting, predictive maintenance in IT, complaints management, document classification, automated data extraction, KYC (Know-Your Customer) document processing, credit scoring, etc; Then, it used to enhancing security and risk control: e.g., enhanced risk control, compliance monitoring, any kind of anomaly detection, AML (Anti-Money Laundering) detection and monitoring, system capacity limit prediction, support of data quality assurance, fraud prevention, payment transaction monitoring, cyber risk prevention.

## 5.1 Banking on Artificial Intelligence

Harnessing cognitive technology with Artificial Intelligence (AI) brings the advantage of digitization to banks and helps them meet the competition posed by FinTech players. In fact, about 32% of financial service providers are already using AI technologies like Predictive Analytics, Voice Recognition, among others, according to a joint research conducted by the National Business Research Institute and Narrative Science[1]. Artificial Intelligence is the future of banking as it brings the power of advanced data analytics to combat fraudulent transactions and improve compliance. AI algorithm accomplishes anti-money laundering activities in few seconds, which otherwise take hours and days. AI also enables banks to manage huge volumes of data at record speed to derive valuable insights from it. Features such as AI bots, digital payment advisers and biometric fraud detection mechanisms lead to higher quality of services to a wider customer base. All this translates to increased revenue, reduced costs and boost in profits.

1 https://narrativescience.com/Offers/The-Rise-of-AI-in-Financial-Services

**AI is strengthening competitiveness of banks through:** Enhanced customer experience: Based on past interactions, AI develops a better understanding of customers and their behavior. This enables banks to customize financial products and services by adding personalized features and intuitive interactions to deliver meaningful customer engagement and build strong relationships with its customers.

**Prediction of future outcomes and trends:** With its power to predict future scenarios by analyzing past behaviors, AI helps banks predict future outcomes and trends. This helps banks to identify fraud, detect anti-money laundering pattern and make customer recommendations. Money launderers, through a series of actions, portray that the source of their illegal money is legal. With its power of Machine Learning and Cognition, AI identifies these hidden actions and helps save millions for banks. Similarly, AI is able to detect suspicious data patterns among humungous volumes of data to carry out fraud management. Further, with its key recommendation engines, AI studies past to predict future behavior of data points, which helps banks to successfully up-sell and cross-sell.

**Cognitive process automation:** This feature enables automation of a variety of information-intensive, costly and error-prone banking services like claims management. This secures ROI, reduces costs and ensures accurate and quick processing of services at each step. Cognitive process automation fundamentally automates a set of tasks that improvises upon their previous iterations through constant machine learning.

**Realistic interactive interfaces:** Chatbots identify the context and emotions in the text chat and respond to it in the most appropriate way. These cognitive machines enable banks to save not only time and improve efficiency, but also help banks to save millions of dollars as a result of cumulative cost savings.

Effective decision-making: Cognitive systems that think and respond like human experts, provide optimal solutions based on available data in real-time. These systems keep a repository of expert information in its database called knowledge database. Bankers use these cognitive systems to make strategic decisions.

Robotic automation of processes: AI reviews and transforms processes by applying Robotic Process Automation (RPA). This enables automation of about 80% of repetitive work processes, allowing knowledge workers to dedicate their time in value-add operations that require high level of human intervention.

### *Challenges /limitations*

There are quite a number of aspects that we can consider as challenges. These include;

- **Ethical concerns:** the EBF welcomes the ongoing discussions on ethics and recommends the Ethics by design approach and a set of abstract and high-level principles to allow for flexibility in practice.

- **Lack of transparency:** the EBF recommends a risk-based approach to transparency and explainability, based on the impact of the outcomes of the systems.

- **Asymmetric regulation between industries:** the EBF recommends enforcing the principle of "same services, same risks, same rules and same supervision";

- **Asymetric regulation in a global context:** the EBF encourage the creation of research centres in an opensource environment and notes the importance of ensuring a strong AI eco-system in Europe;

- **A dated regulatory framework:** the EBF recommends a full-fledged AI fitness check of the current regulatory framework in order to adapt rules where relevant and remove potential obstacles

- **Ambivalence and low perception of AI relevance and context in the banking sector:** the EBF belives an important step is to ensure regulators, policy makers and supervisors have the necessary knowledge and undertanding of the technology and its impact in the banking industry. The EBF recommends collaboration and dialogue between the European Commission, banking supervisors and regulators, data protection authorities, and lawmakerss as well as with the industry.

- **Education:** the EBF recommends encouraging the development of programs to foster the sills and knowledge needed by data scientists, engineers, mathematicians, etc. and to ensure a proper system for re-skilling. •Competition with other sectors for recruiting: the EBF would welcome further analsysis on the way the banking sector's remuneration rules impact its ability to compete when recruiting data scientists. Unfortunately, while it is easier than ever to run state-of-the-art ML models on pre-packaged datasets, designing and implementing the systems that support ML in real-world applications is increasingly a major bottleneck. In large part this is because ML-based applications require distinctly new types of software, hardware, and engineering systems to support them. Indeed, modern ML applications have been referred to by some as a new "Software 2.0" to emphasize the radical shift they represent as compared to traditional computing applications. They are increasingly developed in different ways than traditional software—for example, by collecting, preprocessing, labeling, and reshaping training datasets rather than writing code—and also deployed in different ways, for example utilizing specialized hardware, new types of quality assurance methods, and new end-to-end workflows. This shift opens up exciting research challenges and opportunities around high-level interfaces for ML development, low-level systems for executing ML models, and interfaces for embedding learned components in the middle of traditional computer systems code.""Modern ML approaches also require new solutions for the set of concerns that naturally arise as these techniques gain broader usage in diverse real-world settings. These include cost and other efficiency metrics for small and large organizations alike, including e.g. computational cost at training and prediction time, engineering cost, and cost of errors in real-world settings; accessibility and automation, for the expanding set of ML users that do not have PhDs in machine learning, or PhD time scales to invest; latency and other runtime constraints, for a widening range of computational deployment environments; and concerns like fairness, bias, robustness, security, privacy, interpretability, and causality, which arise as ML starts to be applied to critical settings where impactful human interactions are involved, like driving, medicine, finance, and law enforcement."

Data protection and privacy: A lot has been said about the new data protection regulation and its impacts on innovation. Notably, automated decision-making rules (and rights of individuals) under article 22 of the GDPR may hinder banks from embracing AI to provide better services and safer solutions, since significant manual processes may still be necessary. AI-based decision making should be subject to oversight and control, but efficiencies may not be realised if human intervention in individual cases becomes significant. In this regard, the exemptions provided by Article 22(2) of the GDPR are welcome.

# 6.0 Conclusion

**AI-driven future is very clear. AI will not only empower banks by automating its knowledge workforce, it will also make the whole process of automation intelligent enough to do away with cyber risks and competition from FinTech players.** AI, integral to the bank's processes and operations, and keeps evolving and innovating with time without considerable manual intervention. AI will enable banks to leverage human and machine capabilities optimally to drive operational and cost efficiencies, and deliver personalized services. All of these benefits are no longer a futuristic vision to accomplish for banks. By adapting AI, leaders in the banking sector have already taken actions with due diligence to reap these benefits. The more extensive use of AI in banking will also present new opportunities and challenges to regulators seeking to safeguard financial stability while enhancing consumer protection and nurturing innovation. The landscape of compliance and supervision for the banking sector is likely to evolve with more widespread adoption of AI. The technology can be harnessed to streamline the compliance process, introduce machine-readable regulations and automate data collection for supervisory purposes. Additional insights may also be generated from various types of data collected from banks. The prospects for a broader and more advanced use of AI in banking, compliance and supervision appear promising, encouraged by gains in efficiency and enhancement in risk management. Policymakers are exploring further use of AI in improving compliance (Regtech) and supervisory capacity (Suptech), which is mutually beneficial to banks and regulators.

Artificial Intelligence (AI) presents opportunities to increase prosperity and growth. For the banking sector, it provides great opportunities to enhance customer experience, democratize financial services, improve cybersecurity and consumer protection and strengthen risk management. Transparency and explainability, are important in maintaining trust in AI, given the statistical nature of this technology. However, we would highlight that a risk-based approach should be preferred to maintain a high-level of customer protection and trust. Ensuring a level playing field for all industries and geographies is of capital importance to ensure the uptake of AI in the European banking sector and to maintain a strong level of customer protection, ensuring customers are empowered. AI is an evolving technology and it is paramount to ensure that the regulatory environment is fit for the use of AI by promoting innovation and legal certainty. We highlight the need for a "future-proof"/ "technology-ready" legal and regulatory framework. We thus stress the need to maintain a high level of consumer protection while ensuring a level playing field and an activity-based/technology-neutral approach to regulation.

This paper has illustrated a number of key takeaways:

a. AI applications offer the greatest cost savings opportunity across digital banking. These include; banks leveraging algorithms on the front end to smooth customer identification and authentication, mimic live employees through chatbots and voice assistants, deepen customer relationships, and provide personalized insights and recommendations.

b. AI is also being implemented by banks within middle-office functions to assess risks, detect and prevent payments fraud, improve processes for anti-money laundering (AML) and perform know-your-customer (KYC) regulatory checks.

c. The winning strategies employed by banks that are undergoing an AI-enabled transformation reveal how to best capture the opportunity. These strategies highlight the need for a holistic AI strategy that extends across banks' business lines, usable data, partnerships with external partners, and qualified employees.

The paper emphasizes the possibility the banks can utilize the huge amount of data across the various levels of analytics – descriptive, diagnostic, predictive and prescriptive analytics. This is captured in one whole unit operational dashboard developed from simulated data that imitates a local Kenyan Bank.

# Reference

1. Addo, P. M., Guegan, D., & Hassani, B. (2018). Credit risk analysis using machine and deep learning models. *Risks, 6*(2), 38.

2. Ala'raj, M., & Abbod, M. F. (2016a). A new hybrid ensemble credit scoring model based on classifiers consensus system approach. *Expert Systems with Applications, 64*, 36–55.

3. Ala'raj, M., & Abbod, M. F. (2016b). Classifiers consensus system approach for credit scoring. *Knowledge-Based Systems, 104*, 89–105.

4. Bacham, D., & Zhao, J. (2017). Machine learning: Challenges, lessons, and opportunities in credit risk modeling. *Moody's Analytics Risk Perspectives, 9*, 30–35.

5. Bellotti, T., & Crook, J. (2009). Support vector machines for credit scoring and discovery of significant features. *Expert Systems with Applications, 36*(2), 3302–3308.

6. Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2021). Explainable machine learning in credit risk management. *Computational Economics, 57*(1), 203–216.

7. Chawla, N. V. (2003). C4. 5 and imbalanced data sets: Investigating the effect of sampling method, probabilistic estimate, and decision tree structure. *Proceedings of the ICML, 3*, 66.

8. Chen, M.-C., & Huang, S.-H. (2003). Credit scoring and rejected instances reassigning through evolutionary computation techniques. *Expert Systems with Applications, 24(*4), 433–441.

9. Dorfman, R. (1979). A formula for the Gini coefficient. *The Review of Economics and Statistics*, 146–149.

10. Grobelnik, M. (1999). Feature selection for unbalanced class distribution and naive bayes. *ICML '99: Proceedings of the Sixteenth International Conference on Machine Learning,* 258–267.

11. Hand, D. J., & Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: A review. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 160(3)*, 523–541.

12. Huang, C.-L., Chen, M.-C., & Wang, C.-J. (2007). Credit scoring with a data mining approach based on support vector machines. *Expert Systems with Applications, 33*(4), 847–856.

13. Keramati, A., & Yousefi, N. (2011). A proposed classification of data mining techniques in credit scoring. *The Proceeding of 2011 International Conference of Industrial Engineering and Operations Management, Kuala Lumpur, Malaysia, Jurnal,* 22–24.

14. Lai, K. K., Yu, L., Wang, S., & Zhou, L. (2006). Neural network metalearning for credit scoring. *International Conference on Intelligent Computing,* 403–408.

15. Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research, 247*(1), 124–136.

16. Manilow, E., Seetharaman, P., Pishdadian, F., & Pardo, B. (2017). Predicting algorithm efficacy for adaptive multi-cue source separation. *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA),* 274–278.

17. Moscatelli, M., Parlapiano, F., Narizzano, S., & Viggiano, G. (2020). Corporate default forecasting with machine learning. *Expert Systems with Applications, 161*, 113567.

18. Punniyamoorthy, M., & Sridevi, P. (2016). Identification of a standard AI based technique for credit risk analysis. *Benchmarking: An International Journal.*

19. Raileanu, L. E., & Stoffel, K. (2004). Theoretical comparison between the gini index and information gain criteria. *Annals of Mathematics and Artificial Intelligence, 41*(1), 77–93.

20. Van Gestel, I. T., Baesens, B., Garcia, I. J., & Van Dijcke, P. (2003). A support vector machine approach to credit

scoring. *FORUM FINANCIER-REVUE BANCAIRE ET FINANCIAIRE BANK EN FINANCIEWEZEN-*, 73–82.

21. Van Sang, H., Nam, N. H., & Nhan, N. D. (2016). A novel credit scoring prediction model based on Feature Selection approach and parallel random forest. *Indian Journal of Science and Technology, 9*(20), 1–6.

22. Wang, H., Xu, Q., & Zhou, L. (2015). Large unbalanced credit scoring using lasso-logistic regression ensemble. *PloS One, 10*(2), e0117844.

23. Wójcicka, A. (2017). Neural Networks in Credit Risk Classification of Companies in the Construction Sector. *Econometric Research in Finance, 2*(2), 63–77.

24. Zhang, Y., Wang, D., Chen, Y., Shang, H., & Tian, Q. (2017). Credit risk assessment based on long short-term memory model. *International Conference on Intelligent Computing*, 700–712.

25. Zhou, G., & Wang, L. (2012). Co-location decision tree for enhancing decision-making of pavement maintenance and rehabilitation. *Transportation Research Part C: Emerging Technologies, 21*(1), 287–305.

26. Zhou, H., Wang, J., Wu, J., Zhang, L., Lei, P., & Chen, X. (2013). Application of the hybrid SVM-KNN model for credit scoring. *2013 Ninth International Conference on Computational Intelligence and Security*, 174–177.

**Kenya Bankers Association**
13th Floor, International House, Mama Ngina Street
P.O. Box 73100– 00200 NAIROBI
Telephone: 254 20 2221704/2217757/2224014/5
Cell: 0733 812770/0711 562910
Fax: 254 20 2221792
Email: research@kba.co.ke
Website: www.kba.co.ke

KENYA BANKERS
ASSOCIATION

One Industry. Transforming Kenya.