International Journal of Economics and Finance; Vol. 17, No. 2; 2025
ISSN 1916-971X E-ISSN 1916-9728
Published by Canadian Center of Science and Education

Assessing the Creditworthiness of Lebanese Banks Using Bayesian Networks

Georges Chlela[1] & Hasan Mousawi[1]

[1] Doctoral School of Law, Political, Administrative, and Economic Sciences (EDDSPAE), Lebanese University, Sin-El-Fil, Beirut, Lebanon

Correspondence: Georges Chlela, Doctoral School of Law, Political, Administrative, and Economic Sciences (EDDSPAE), Lebanese University, Sin-El-Fil, Beirut, Lebanon. E-mail: georges.chlela@st.ul.edu.lb

Received: November 12, 2024 Accepted: December 12, 2024 Online Published: December 20, 2024

doi:10.5539/ijef.v17n2p1 URL: https://doi.org/10.5539/ijef.v17n2p1

Abstract

This research evaluates the creditworthiness of Lebanese banks using the Bayesian Naïve Classifier (BNC) in the CAMELS framework. Using the CAMELS indicators—capital adequacy, asset quality, management efficiency, earnings, liquidity, and sensitivity to market risk—the study examines data from 2012 to 2022, a period also marked by financial instability. The complex interdependencies between these variables are modeled using the BNC, a machine learning technique that provides a probabilistic approach that improves prediction accuracy. In order to assess how well the BNC predicts banks' ratings, training and testing datasets are created. The findings indicate that the most important elements influencing bank ratings are capital adequacy, management efficiency, and asset quality. Liquidity and sensitivity to market risk become more significant during economic downturns, especially following the 2019 financial crisis in Lebanon. With a predicted accuracy of more than 98%, the BNC proved its resilience and dependability in identifying patterns that traditional models would miss. By incorporating machine learning into the CAMELS framework, this study presents an innovative approach to credit risk assessment and offers insightful information to investors, regulators, and decision-makers who are keeping an eye on the stability of financial institutions. To further confirm this model's resilience in many economic contexts, future studies should extend its use to more industries and geographical areas.

Keywords: Bayesian Naïve Classifier, CAMELS framework, Lebanese banks, creditworthiness, machine learning, risk management

1. Introduction

The Lebanese banking sector has been at the core of a number of economic and financial difficulties, especially over the past couple of years. Economic crises, regulatory changes, and political instability have all had a substantial effect on the stability and trust in the nation's banking institutions. Following the 2019 financial crisis, these challenges have intensified, resulting in hyperinflation, currency devaluation, and rising concerns about the solvency and operational stability of Lebanese banks. For regulators, investors, and financial industry decision-makers, determining a bank's trustworthiness has become an increasingly important responsibility in this setting.

The Bayesian Naïve Classifier (BNC) is applied to a dataset from 2012 to 2022, which includes important economic events such as the 2019 Lebanese financial crisis, regulatory changes by Basel III framework, and the Syrian refugee crisis. By including this time period, the model is able to account for a variety of economic circumstances, providing a strong and flexible evaluation instrument that can forecast bank health more precisely than conventional techniques.

Bayesian Belief Networks (BBNs), which are often referred to as Belief Networks or Bayesian Networks, are probabilistic graphical models that use a directed acyclic graph (DAG) to depict a set of variables and their conditional relationships. They are effective tools for reasoning under uncertainty because they give a clear, concise representation of the joint probability distribution over a set of variables.

A Bayesian Belief Network (BBN), according to Ghribi and Masmoudi (2013), is a graphical depiction of a probabilistic model that encodes a collection of conditional independence relationships. Abid, Zaghdene, Masmoudi, and Ghorbel (2017), Hassen, Masmoudi, and Rebai (2008), Bouchaala, Masmoudi, Gargouri, and

Rebai (2010), and Dbouk and Zaarour (2017) are just a few of the sectors where it has gained popularity as a tool for decision-making systems. These fields have found use for Bayesian Belief Networks: medicine, where they are used to model the probabilistic relationships between symptoms and diseases; genomics, which uses them to understand gene interactions and predict gene expression levels; risk assessment, which assesses and manages risks in the fields of engineering and finance; natural language processing, which uses them for tasks like language modeling and machine translation; and decision support systems, which help with decision-making under uncertainty.

In fact, according to Gevaert et al. (2006), the BBN is one of the most thorough and reliable formalisms for the collection and modeling of complex systems, exceeding logistic regression in terms of diagnostic prediction. According to Saad et al. (2013), BBNs are effective instruments for knowledge inference and representation. Others noted that a well-known feature of statistical learning algorithms with a probabilistic inference and good classification accuracy is the majority of BBN (Manolopoulos, Kirkos, & Spathis, 2007; Zaarour et al., 2015).

Nodes and edges are the two main parts of a Bayesian belief network. A discrete or continuous random variable is represented by each node in the network. Conditional dependencies between the variables are represented by directed edges connecting nodes. An edge connecting node A to node B indicates that A directly affects node B. A Conditional Probability Table (CPT) that measures the influence of the parents on each node in the network is linked to each node. Based on the states of its parent nodes, the CPT calculates the likelihood of every potential state for the node. For instance, the CPT for node B defines $P(B \mid A1, A2, \ldots, A\text{n})$ where B has parents $A1$, $A2$, …, An.

Given any observed data, inference in Bayesian Belief Networks entails calculating the posterior distribution of a set of query variables. Many algorithms can be used to accomplish this, including approximate inference techniques like Markov Chain Monte Carlo (MCMC) and Loopy Belief Propagation and exact inference techniques like Variable Elimination, Belief Propagation, and Junction Tree algorithms. The answer to this research question will help inform West African authorities and investors about the measures to be taken in the event of exogenous shocks and co-movements on the various stock exchanges in the zone.

BBNs have a number of benefits. They make it simpler to comprehend and explain connections between variables by giving a clear graphical representation of them. The models' modular nature makes it simple to update them with new data as it becomes available. In addition, compared to deterministic models, BBNs allow for robust probabilistic inference and improved handling of uncertainty. Bayesian Belief Networks have a number of drawbacks in spite of their benefits. Both the network's complexity and the computing expense of inference rise exponentially with the number of variables. Large volumes of data are necessary for accurate CPT estimate, and in certain fields, obtaining that data might be challenging. Selecting the right network structure and settings also takes a lot of work and domain knowledge. By creating more effective inference algorithms, automated structure learning strategies, and methodologies for integrating disparate data sources, future research in BBNs seeks to address these issues.

2. Naïve Bayes

The Bayes theorem was first used to study the Naïve Bayes approach. Using labeled data—that is, data where each occurrence is allocated to a class—supervised statistical classifiers are estimated under the premise of conditional independence (Bradley, 1997). The Bayesian classifier employs mutual conditional probability distributions to enable class-based conditional independencies to operate across variables. The underlying relationships are represented in a graphical model (Davutyan & Ozar, 2006, Witzany, 2017). While the attribute in the data may have an actual Boolean variable to establish the relationship, the random variables make either a continuous or discrete relationship (Demerjian, 2007). All variables are independent of the nondescendant, and each arc in the acyclic graph indicates the dependence probability.

The following expressions will be defined to calculate the posteriori probability using the Bayes theorem (Beltran et al., 2014):

- P(y) is prior probability that, in the absence of additional observed data (x), the hypothesis (y) is fulfilled.
- P(y/x) is the posterior probability that, when the data x is known, the hypothesis y will be fulfilled. It illustrates how the data has affected hypothesis y.
- P(x/y): the likelihood function, or probability that the data x will be observed once the hypothesis y is known.

If we consider the probability distribution shown below:

$$P(y \cap x)=P(y)*P(x/y)=P(y/x)*P(x) \tag{1}$$

What is posterior probability?

$$P(y/x)=P(y \cap x)/P(x)=P(y)*P(x/y)/P(x) \tag{2}$$

If there are two nodes (variables), x and y, which are the response and prediction variables, respectively, then this function would be utilized. The dependency relationships between the variables, in which the parent variables transmit information to the offspring variables, are represented by the arcs. According to Tuya et al. (2007), Bayesian networks operate under the assumption that nodes are directly dependent on their parent nodes. Each node is connected to a conditional probability table that indicates the likelihood of each variable's status given the status of potential parent nodes.

To obtain the most optimal network, we should gather every possible case; nonetheless, we should distinguish between two kinds of learning (Castillo et al., 1997):

- Learning the graphic structure, or dependency, also known as structure learning.
- Learning the parametric structure (probabilities) also known as parameter learning.

Two components make up these two learning styles (Castillo et al., 1997):

Quality: A group of Bayesian networks can be categorized based on their quality. Its network's quality (both graphic and parametric structure) will be evaluated.

Search measure: The optimal network will be chosen from a subset of high-quality Bayesian networks using the search method.

After the Bayesian network is chosen, the so-called classifiers are used to determine the most likely prediction for a given data set, which helps estimate the probability of an occurrence based on data outside of the network design.

3. Methodology

This study evaluates the creditworthiness of Lebanese banks using a hybrid methodology that combines the Bayesian Naïve Classifier (BNC) with the CAMELS framework. Eight structured steps make up the methodology: problem definition, period selection, data collection and preparation, data selection, dataset partitioning for training and testing, machine learning technique selection, and finally model learning, testing, and evaluation. Each process to guarantee the validity and robustness of the research findings is thoroughly explained in this section.

CAMELS composite scores serve as the classification targets for the BNC model. The model was trained using a variety of performance metrics and financial data from ten Lebanese banks from 2012 to 2022. The dependent variable was each bank's CAMELS overall rating, whereas the inputs were variables like capital adequacy, asset quality, management efficiency, earnings, liquidity, and sensitivity.

Step 1: Problem Definition

Numerous difficulties have been faced by the banking industry in Lebanon, such as political instability, regulatory reforms, and economic downturns. These difficulties highlight how crucial it is to forecast bank creditworthiness with accuracy in order to maintain financial stability and investor confidence. The goal of this study is creating a Bayesian Belief Network (BBN) model that will be used to assess the credit ratings of Lebanese banks using the CAMELS rating framework. Recent severe issues and uncertainty in the Lebanese banking system have made a thorough creditworthiness evaluation necessary. The objective is to create a probabilistic model that, by capturing the interdependencies between these elements, can forecast bank credit ratings.

Traditional linear models have difficulty predicting bank credit ratings because of the intricate and dynamic interdependencies among these components. A Bayesian Belief Network, a probabilistic graphical model that illustrates the conditional interdependence among a set of variables is used in this study to overcome this difficulty. Because they can include both expert knowledge and empirical data, BBNs are especially well-suited for this kind of study. This enables a more thorough and nuanced understanding of the interactions between the CAMELS components. While data-driven techniques, such structure learning algorithms, were used to develop the network structure based on the gathered data, expert knowledge was utilized to identify the important relationships and dependencies among the CAMELS components. The integration of both practical and theoretical components of bank performance is ensured by the mix of empirical data and expert knowledge in the model.

This research ensures a structured and thorough analysis of the credit ratings of Lebanese banks using the CAMELS framework and Bayesian Belief Network by precisely defining the problem, and objectives of the study. It does this by methodically moving on to the subsequent stages of the 10-stage approach. A more sophisticated and precise estimate of bank creditworthiness which effectively captures the intricacies and interdependencies between the CAMELS components. This methodology not only augments our comprehension of the ratings of Lebanese banks but also provides significant perspectives and suggestions for diverse stakeholders, hence augmenting the general steadiness and adaptability of the banking industry.

Step 2: Period Selection

The Lebanese banks' CAMELS rating analysis period of choice is 2012-2022. In this research, a crucial stage is determining the right time frame for evaluating Lebanese banks' using the CAMELS rating system. The years 2012– 2022, which have been selected, represent a decade that has been characterized by noteworthy financial, political, and economic developments that have had a substantial effect on the banking industry in Lebanon. By capturing both the long-term trends and the impacts of different external shocks, this timeframe provides a thorough understanding of the resilience and performance of the banks.

Year 2012 was selected as the launch year because of its importance in the aftermath of the global financial crisis. The 2008 financial crisis was progressively being recovered from by the world economy by 2012, and the banking industry had been stabilized after surviving the early shocks and new regulations were set out globally. The start of this year's comparatively tranquil period offers a starting point for observing how banks modified their operations and strategy in response to the crisis' aftereffects and subsequent events.

To guarantee that the research accurately represents the current status of the banking industry, the ending point of 2022 is chosen to incorporate the most recent data available. The years 2019 to 2022 are especially important because that's when Lebanon's financial crisis peaked, starting in late 2019. The banking industry and the economy have suffered greatly as a result of this crisis, which has been marked by severe currency depreciation, hyperinflation, and widespread bank closures.

Several significant events and trends that are necessary for a comprehensive creditworthiness evaluation occurred in the Lebanese banking sector between 2012 and 2022. Among these are the Basel III regulatory reforms, which are intended to bolster risk management procedures and bank capital requirements. Beginning in 2013, banks' capital adequacy and risk profiles were impacted by the implementation of Basel III requirements. Bank compliance and adaptation to these worldwide regulatory standards can be evaluated by analyzing data collected during that time.

The influence of geopolitical tensions in the region and how they affect Lebanon indirectly is another important aspect. The 2011 start of the Syrian civil war resulted in a significant migration of refugees into Lebanon, placing a burden on the nation's resources and stability of its economy. The choice of time frame considers the long-term financial effects of the humanitarian crisis on the banking industry in Lebanon, including elevated risks and alterations in asset quality.

A number of domestic political events that have an immediate impact on the banking sector are also included in the chosen period. Political instability has resulted in uncertainty about policy and economic disruptions. It is typified by frequent changes in government and extended periods without a functioning government. For example, there were major economic consequences following the 2014-2016 political vacuum and the 2016 election of President Michel Aoun. Political unrest in 2019 was exacerbated by Prime Minister Saad Hariri's resignation and protests, which led to a catastrophic financial crisis. The analysis can assess the impact of policy changes and political instability on bank performance by examining these occurrences.

The chosen time period also considers technology developments and how they affect the banking industry. To improve operational effectiveness and customer service, Lebanese banks have embraced digital banking services and technology advancements more and more during the last ten years. This technical evolution is important for evaluating the quality of management and the ability of banks to adjust to shifting market conditions.

Significant changes in monetary policy and currency also occurred throughout this decade. The devaluation that followed the Lebanese pound's peg to the US dollar in 2019 and 2020 had a significant impact on bank liquidity and market risk sensitivity. The monetary policies implemented by the Central Bank throughout that time, such as changes to interest rates and capital restrictions, are essential to comprehending how the banks react to problems with liquidity and foreign exchange.

The selected time frame offers insights into the institutions' future prospects in addition to capturing the historical background and history of the banking industry. It is essential to comprehend how banks have handled difficulties in the past and adjusted to shifting circumstances in order to forecast their performance and creditworthiness in the future. Thus, the analysis will contribute to the general stability and resilience of the Lebanese banking system by providing insightful analysis and recommendations for stakeholders, including bank managers, investors, regulators, and policymakers.

Step 3: Data Collection

Careful preparation is needed for the data collecting procedure that will be used to create a Bayesian Belief Network (BBN) model and evaluate the creditworthiness of Lebanese banks using the CAMELS rating system. Collecting thorough, precise, and pertinent information about the six essential elements of the CAMELS framework—capital adequacy, asset quality, management quality, earnings, liquidity, and sensitivity to market risk—is the main goal. Data was gathered for this purpose from several sources, including regulatory filings, publications from the Central Bank of Lebanon, annual reports, and financial databases like Bloomberg and Reuters. The main source for the comprehensive and consistent financial data used for this analysis was annual reports.

The first step in the data collection procedure was to determine which particular metrics were pertinent to each CAMELS component. For example, the capital adequacy ratio (CAR) was used to measure capital adequacy; the ratio of non-performing loans (NPL) to total loans was used to assess asset quality; efficiency ratio was used to evaluate management; return on equity (ROE) and return on assets (ROA) were used to analyze earnings; the loan-to-deposit ratio (LDR) was used to measure liquidity; and the total securities to total assets ratio was used to gauge the level of sensitivity risk.

Reports and statistical bulletins from the Central Bank of Lebanon provided more information on monetary policies, economic indicators, regulatory changes, and the general performance of the banking industry. Furthermore, information on capital adequacy, risk exposures, and compliance was obtained from regulatory filings. And finally, annual reports provided information on basically the rest where external audited reports were used.

Carefully gathering these indicated metrics from the cited sources was required for data extraction. To ensure accuracy and completeness, a detailed examination of the financial statements, management discussions, and notes to accounts in the annual reports was necessary for this phase.

This study guarantees the robustness and reliability of the dataset used in the development of the Bayesian Belief Network model by implementing a strict data gathering procedure. The basis for precisely forecasting the creditworthiness of Lebanese banks is this extensive dataset, which provides important information about the banks' resilience and financial stability in the face of difficult political and economic conditions in Lebanon between 2012 and 2022.

Step 4: Data Preparation

To guarantee that the dataset used to create the Bayesian Belief Network (BBN) model is reliable, consistent, and appropriate for analysis, data preparation is an essential step. Data cleansing, addressing missing values, standardization, and data integration are some of the crucial procedures involved in this step. By working together, these procedures guarantee the accuracy and utility of the information gathered, providing a strong basis for evaluating Lebanese banks' creditworthiness using the CAMELS rating system.

To identify and correct any errors or discrepancies in the dataset, data cleaning is crucial. Anomalies including typographical errors, duplicate entries, and inconsistent formatting are frequently found in financial data that has been retrieved from regulatory filings, annual reports, financial databases, and other sources. To resolve these problems, the data was carefully examined and cleansed. Typographic problems were fixed by cross-referencing the rectified data with the original data sources, such as inaccurate mislabeled data. To preserve consistency throughout the dataset, inconsistent formatting—such as different date formats and unit measures—was standardized. Managing missing values is essential to preserving the dataset's trustworthiness and completeness. Financial datasets, especially ones covering ten years, can contain gaps for a variety of reasons, such as missing certain measures in particular years. Depending on the kind and volume of missing data in audit reports, data was gathered for this purpose from other several sources, including regulatory filings, and publications from the Central Bank of Lebanon.

Furthermore, by maintaining uniformity in units and formats amongst banks and historical periods, standardization improves data comparability even further. To reflect real values, this process entailed converting all financial figures to a single currency. The base currency was the Lebanese pound (LBP). To guarantee uniform computation and reporting across all institutions and years, financial ratios and percentages were also standardized. For instance, uniform definitions of the numerator and denominator components were used in the calculation of the ratios such as ROACE, ROAA, etc.

Integrating data from multiple sources into a single database requires cleaning, and standardizing. This procedure makes sure that all pertinent data points—like bank names, reporting periods, and certain financial metrics—are combined based on shared identifiers. A thorough dataset that offers a comprehensive picture of each bank's performance during the chosen period across all CAMELS components is guaranteed by the integration process. The integrity of the data was given great consideration during data integration. To make sure that no data was lost or duplicated and that all integrated data points matched appropriately, cross-referencing was done. To guarantee reproducibility and openness, thorough documentation was kept throughout the data preparation phase.

The stages used during data integration, standardization, and thorough records of data cleaning procedures are all included in this documentation. The thorough documenting of the research is essential to its legitimacy and facilitates the replication of the study by other researchers, should the necessity arise. This study makes sure that the dataset used to create the Bayesian Belief Network model is robust and dependable by carefully preparing the data. This meticulously created dataset offers a strong basis for precisely projecting the creditworthiness of Lebanese banks, offering insightful information about their resilience and financial stability in the face of the difficult political and economic environment in Lebanon from 2012 to 2022. The meticulous and rigorous processing of the data highlights the dedication to generating reputable, high-caliber research that can educate stakeholders and support the general stability and resilience of the banking industry in Lebanon.

Step 5: Data Selection

In order to guarantee the validity and applicability of the study utilizing the Bayesian Belief Network within the CAMELS framework, a thorough selection of Lebanese banking institutions was conducted at this phase. The ten chosen banks are representative of a wide range of Lebanese banking institutions, including those with different market capitalizations, operating sizes, and ownership arrangements. A thorough understanding of each institution's strategic approaches and financial health is made possible by this varied collection.

Publicly traded banks were included in order to benefit from their accessibility and transparency, as they are obligated to periodically release comprehensive financial reports. A trustworthy source of financial information that may be used to deduce risk variables and operational patterns is also made available by these disclosures, which improve transparency. One important finding of the research is that publicly traded banks typically implement governance models and initiatives that coincide with shareholder expectations.

Conversely, banks that are privately traded provide a different but no less useful viewpoint. Their financial strategies and governance frameworks can indicate differing operational dynamics, despite the fact that they are not subject to the same strict reporting standards. These banks might use alternate financial methods and more adaptable management techniques, which help to provide a more thorough awareness of the diversity of the industry. Their inclusion offers a comprehensive picture of the banking environment by enabling the investigation of unconventional methods to governance and risk management.

Through the introduction of various financial strategies, governance frameworks, and risk management procedures, this combination of publicly and privately listed banks enhances the dataset. The use of the Bayesian Belief Network in evaluating the CAMELS composite scores requires this kind of diversity. In addition to reflecting the complexity of the banking industry, the dataset provides a solid basis for predictive modeling, allowing for the examination of diverse risk indicators and financial habits across different bank types.

The dataset's diversity is further increased by the inclusion of institutions with varying market capitalizations and sizes. Generally speaking, banks with higher market capitalizations have greater access to capital and resources, which affects their financial resilience and strategic choices. Smaller banks, on the other hand, could have fewer resources at their disposal, which forces them to use more creative or cautious financial techniques. A thorough examination under the CAMELS framework requires that the research encompass a broad range of operational models, from conservative approaches to more aggressive growth strategies. This is ensured by the balance between large and small banks.

By choosing these ten banks, the analysis's data will be typical of the Lebanese banking industry, encompassing its entire spectrum of operational, financial, and strategic diversity. Building a trustworthy Bayesian Belief Network model that can accurately evaluate and forecast these banks' risk profiles and performance based on their CAMELS composite scores requires this step.

Step 6: Data Partition

Data partition is an essential next step in this study after the careful data selection process. To make it easier to design, validate, and test the Bayesian Belief Network (BBN) model, data partitioning entails splitting the gathered dataset into discrete groups. In order to guarantee that the model is accurate and generalizable and that it can offer trustworthy insights into the stability and performance of the ten chosen Lebanese banks between 2012 and 2022, proper data partitioning is crucial.

The main goal of data partitioning is to establish a framework that enables rigorous validation, objective model training, and reliable testing. The study tries to prevent overfitting, a situation in which the model works well on training data but is unable to generalize to new data, by breaking the data up into discrete subsets. Efficient partitioning facilitates the evaluation of the model's prediction precision and practicality in the Lebanese banking industry.

The dataset is divided into two main subsets for this study:

Training Set (80% of the data): The training set is assigned the majority of the data, usually 80%. The BBN model is trained on this subset in order to uncover underlying correlations and patterns in the data. In order to best fit the data during training, the model modifies its parameters, revealing important variables that affect bank ratings. The model continuously improves its predictions based on the training data during the iterative training phase.

Test Set (20% of the data): The test set, which is the last 20% of the data set, is used to assess the predicted accuracy and resilience of the model. This subset is used for testing the model to evaluate its performance in an actual scenario once it has been trained and validated. By simulating future data, the test set offers an accurate assessment of how well the model predicts bank performance based on past trends. The test set results are essential for assessing how well-suited the model is for real-world use and how well it can produce insights that can be put into practice.

Random sampling was used during the data partitioning process to guarantee that each subset is representative of the full dataset. Because random sampling makes sure that the subsets accurately represent the distribution of the data overall including different market conditions, economic conditions, and bank-specific features, it reduces the possibility of adding bias into the model. This strategy is especially crucial in the context of the Lebanese banking industry, since outside variables like political unrest and economic downturns can have a big influence on bank performance.

During the validation stage, cross-validation techniques were used to further improve the model's robustness. To be more precise, k-fold cross-validation was applied, in which the training and validation procedures were carried out several times using various data divisions. The training and validation sets are split into 'k' (5) equal-sized folds using this strategy. 'k-1' folds are used to train the model, while the remaining fold is used to validate it. Every fold serves as the validation set once during the 'k' iterations of this operation. Because the effects of any one partition on the final result are minimized, the cross-validation results yield a more accurate evaluation of the model's performance.

Logistic regression was used to perform the K-Fold cross-validation, and the accuracies for each of the five folds are as follows:

Fold 1: 95.45%

Fold 2: 100%

Fold 3: 95.45%

Fold 4: 100%

Fold 5: 100% Over all folds, the accuracy average is about 98.18%. The following procedure was used to determine each fold's accuracy:

Data division:

The K-Fold cross-validation method was used to split the dataset into 5 subsets, or folds. Four subsets were utilized as the training set and one subset as the testing set for each iteration (fold).

Training Models:

Using the training data from the four subsets, a logistic regression model was developed.

Forecast:

The target variable (CAMELSRANKING) for the testing set (the remaining subset) was then predicted using the trained model.

Calculating Accuracy:

The ratio of accurately predicted instances to all instances in the testing set was used to determine the accuracy.

Repetition:

Five times through this process, a different subset was used as the testing set while the remaining subsets were used as the training set.

The mean precision:

In order to determine the overall ranking of the model, the accuracies from each of the five folds were averaged.

Because the data was longitudinal and covered a ten-year period, temporal relationships had to be taken into consideration when splitting the data. The relationships between data points over various time periods are referred to as temporal dependencies. For example, a bank's financial success in a certain year can affect how well it does in the following years. In order to remedy this, great care was taken to guarantee that the data's temporal order was maintained throughout dividing. When necessary, the data were divided chronologically to preserve the integrity of the time-dependent relationships.

In summary, the data partitioning procedure plays a crucial role in this research and has a direct impact on how well the Bayesian Belief Network model assesses the rating of Lebanese banks. The study establishes a solid framework for the following modeling stages by meticulously segmenting the dataset into training, and test sets and by using methods to manage dependencies and guarantee representativeness. The strict partitioning methodology guarantees the accuracy and generalizability of the BBN model, which can offer significant insights into the variables influencing bank rating in the Lebanese banking industry.

Step 7: Machine Learning Technique Selection

Choosing the right approach for a particular research is essential to the model's application success in the quickly developing field of machine learning. The ability of a system to find patterns and connections within a dataset is the foundation of machine learning. This is achieved through the use of algorithms that can adjust and get better over time as new data is added. In complex frameworks where prediction accuracy and dependability are critical, the choice of machine learning technique has a direct impact on the success of the model. The main objective of the CAMELS approach is to categorize financial institutions according to their overall stability and risk profile. The Bayesian Naïve Classifier (BNC) was found to be the best machine learning method for this problem because of the intricate and varied nature of the data involved. The BNC was chosen because of its track record of success in solving real-world classification issues, especially in financial settings where it's important to identify and take advantage of minute patterns in data in order to make precise predictions.

Based on the assumption of feature independence, the Bayesian Naïve Classifier is a probabilistic classifier that utilizes Bayes' Theorem (Davutyan & Ozar, 2006, Witzany, 2017). The BNC, for all its simplicity, is a strong tool when the independence assumption is roughly met. This makes it especially useful in situations such as the CAMELS framework, where an institution's health is evaluated using a variety of financial indicators. An important benefit in this area is the BNC's capacity to offer concise probabilistic interpretations of classification results, which facilitates the assessment of uncertainty, an essential component of financial decision-making.

Several factors led to the decision to choose the BNC over other machine learning techniques:

Robustness in Noisy Data: Because of the influence of multiple external factors, financial datasets frequently contain noise. The BNC's technique for weighing evidence from different indicators enables it to mitigate the impact of such noise, leading to more reliable classification results.

Computational Efficiency: The BNC is computationally efficient, which makes it suitable for applications involving large datasets, like the CAMELS approach. Its ability to learn incrementally is especially advantageous, as it permits ongoing improvement and updates the model with new data without totally retraining from scratch.

Interpretability: Because the BNC produces probabilistic results, each classification's confidence level can be obtained. This capability is especially helpful in financial settings where risk management tactics can be influenced by knowing the level of certainty underlying a classification.

As the network absorbs data, learning takes place in the BNC, modifying the probabilistic correlations between variables to more accurately represent real-world situations. Although access to enormous amounts of data should theory enable more precise learning, problems like noisy or incomplete data frequently limit the process (Heckerman, 1998). The Expectation Maximization (EM) technique is used to efficiently handle missing data and enhance the learning process. The BNC's capacity to learn from incomplete datasets which are typical in financial settings where not all variables may be regularly recorded is dependent on this approach.

Step 8: Learning, Testing, and Evaluation

The Bayesian Naïve Classifier (BNC) relies heavily on learning, which is the process of improving the network's predicted accuracy by fine-tuning it using past data. The study's learning phase was carried out with financial data from 2012–2022, with a focus on the CAMELS indicators, which evaluate many facets of banks' ratings. The Expectation Maximization (EM) technique as shown in figure 1 was used to efficiently handle missing data and enhance the learning process. The BNC's capacity to learn from incomplete datasets which are typical in financial settings where not all variables may be regularly recorded is dependent on this approach.

The CAMELS indicators—Capital Adequacy, Asset Quality, Management Quality, Earnings, Liquidity, and Sensitivity to Market Risks—were subjected to the learning process in this study. When evaluating the entire risk profile of banks, these factors are essential. The BNC was taught to identify trends in these measures.

The BNC was tasked with determining the correlations between the CAMELS indicators and the overall CAMELS ranking of each bank using the training dataset, which included instances from the 2012-2022 period. This learning process was aided by the use of the software program Netica, which made use of its sophisticated Bayesian network modeling and EM algorithm implementation capabilities.

The BNC's efficacy in differentiating between stable and unstable banks was demonstrated by its ability to correctly categorize institutions according to their CAMELS scores. This was especially clear in the way the BNC was able to forecast results. The learning stage's success highlights the BNC's value. The BNC is a useful tool for analysts and regulators alike since it can retain its accuracy and relevance in changing financial conditions by consistently adding new data to its model.

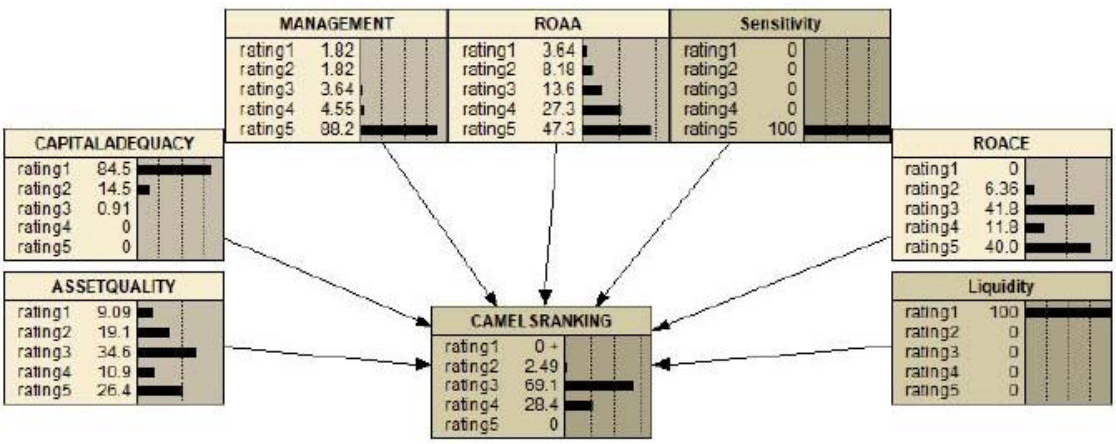

Figure1. Learned-training data set using EM Algorithm

The individual CAMELS components are represented by the nodes in this BNC shown in figure 1, and conditional dependencies are indicated by the directed edges (arrows), which show how changes in one indication may potentially affect other indicators and, ultimately, the overall CAMELS ranking. The arrows indicate directional influences, however as these are probabilistic relationships rather than precise causal lines, care must be taken when interpreting the causal link. Furthermore, in figure 1, we can note that the bulk probabilities stand between Rating 3 (69.1%) and Rating 4 (28.4%).

Figures 2-7 reveal the probability of achieving a better rating when adjusting various parameters thus investigating potential future developments of these linkages under different possible parameters.

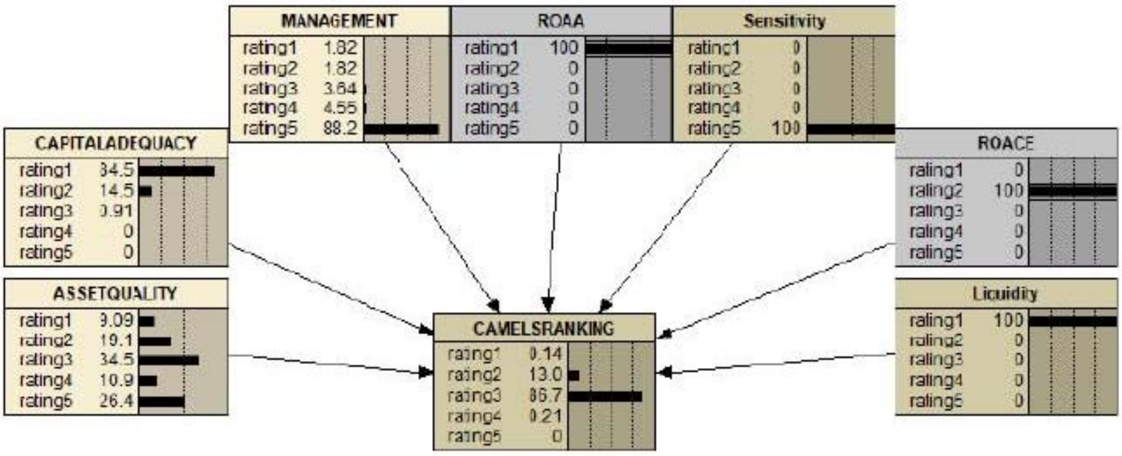

Figure 2. Findings of P (Camelsranking | ROAA=1, ROACE=2)

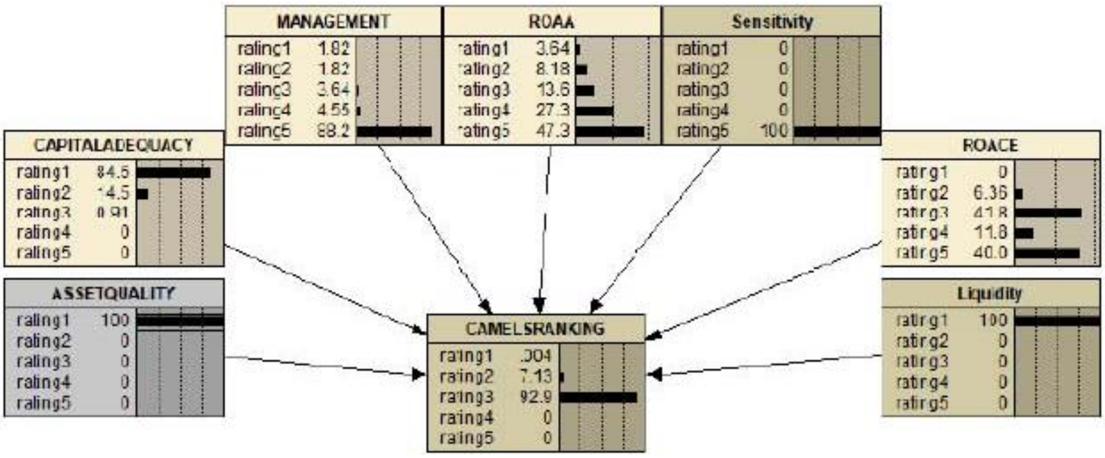

Figure 3. Findings of P (Camelsranking | Assetquality=1)

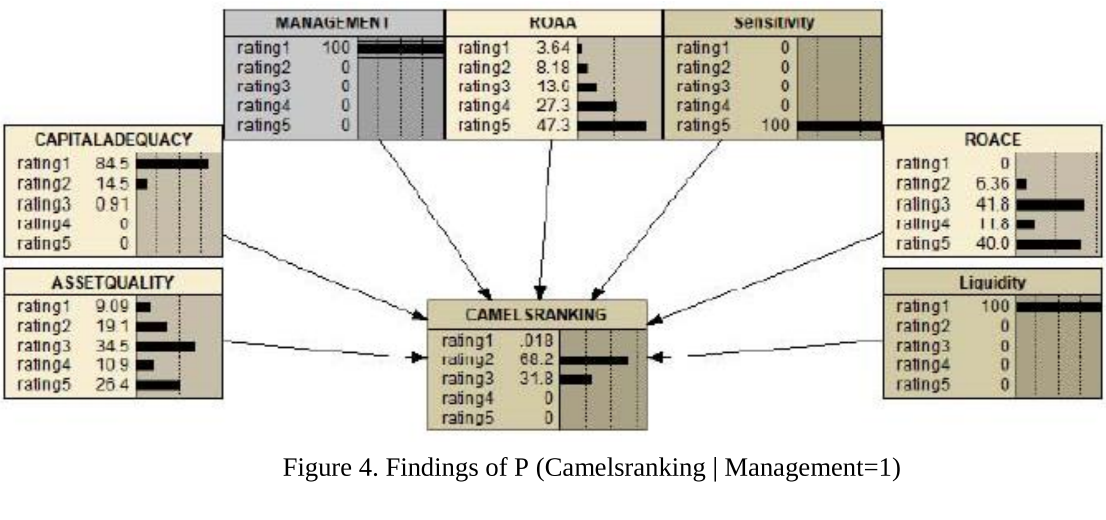

Figure 4. Findings of P (Camelsranking | Management=1)

As per Figure 2, given the ratings of ROAA and ROACE, we can note that the network shows and predicts that the bank will be highly classed as rating 3 with a rating 2 percentage of 13% instead and a rating 3 percentage of 86.7%. Moreover, figure 3 indicates that, with a 92.9% overwhelming probability, the most likely CAMELS ranking in this case is 3. A rating of two also has a lower probability of 7.13%, and ratings of one, four, and five have little to no probability. This shows that the bank is most likely to obtain a CAMELS rating of 3, which represents an average overall performance, while having a very high Asset Quality rating of 1. However as shown in Figure 3, A CAMELS rating of 2 in this case has a higher probability of 68.2% than the case before, where the rating of 3 constitutes 31.8%. According to this, the bank is most likely to obtain a CAMELS rating of 2, which indicates strong overall performance, when the Management rating is at its highest level (1). A rating of 3, on the other hand, may be obtained, indicating average performance.

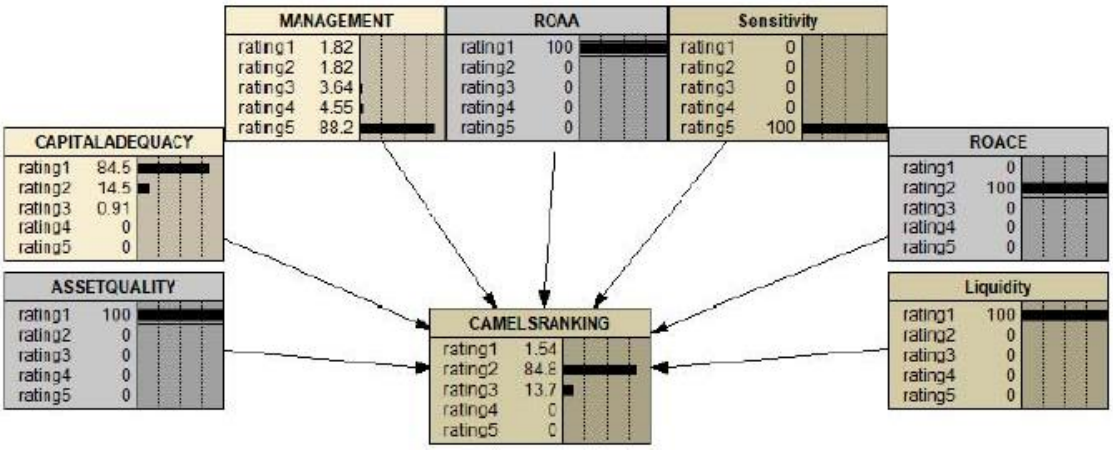

Figure 5. Findings of P (Camelsranking | Assetquality=1, ROAA=1, ROACE=2)

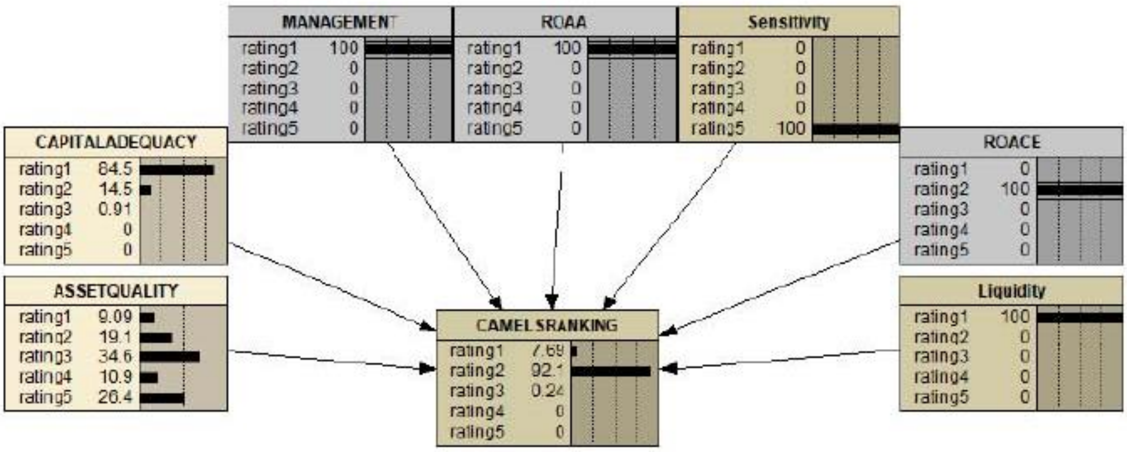

Figure 6. Findings of P (Camelsranking | Management=1, ROAA=1, ROACE=2)

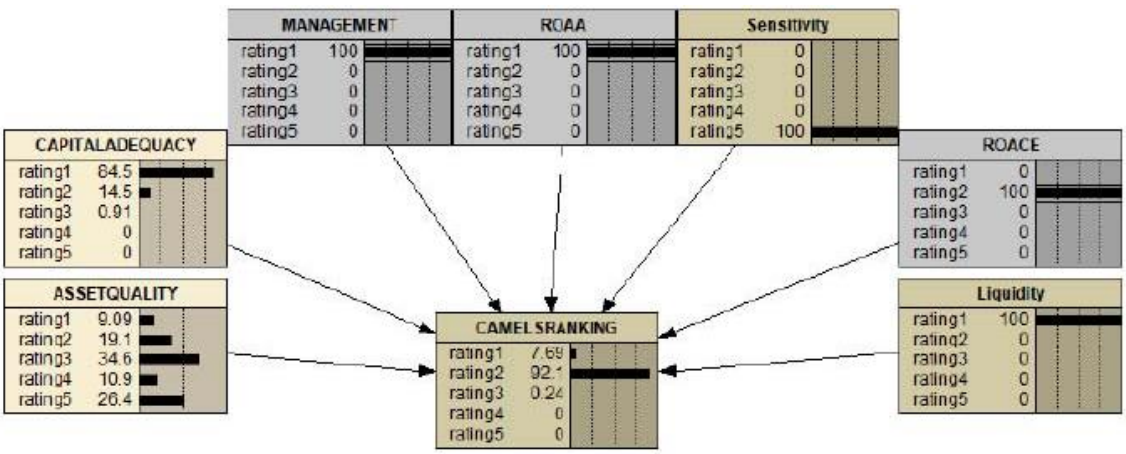

Figure 7. Findings of P (Camelsranking | Management=1, ROAA=1, ROACE=2, Assetquality=1)

According to Figure 5, there is an 84.8% probability that the most likely CAMELS rating in this scenario will be 2. A CAMELS rating of 3 has a lower likelihood of occurring at 13.7%, while the probability of a rating of 1 is quite low at 1.54%. Ratings of 4 and 5 have no probability of occurring at all. Moreover, as indicate in figure 6, there is a high probability of 92.1% that the most likely CAMELS rating in this situation will be 2. A CAMELS rating of 1 also has a lower chance of occurring (7.69%), however the likelihood of a rating of 3 is extremely low (0.24%). Four-and five-star ratings have no chance at all. This implies that the bank is very likely to receive a CAMELS rating of 2, which denotes good performance overall, if Management, ROAA, and ROACE are evaluated favorably (1 and 2). However, Figure 7 algorithm projects an 84.5% chance of a CAMELS rating of 1 in this case. A CAMELS rating of 2 also has a lower probability of 15.5%, but ratings of 3, 4, and 5 have nil probabilities. This indicates that the bank has a strong chance of receiving the best possible CAMELS rating of 1 when ROACE is rated at 2, asset quality, management, and ROAA are all rated at the highest level.

Expanding upon our model discussion, it is crucial to emphasize that the model was developed using data gathered from 2012 to 2022. But there are certain restrictions on the criteria that are offered, especially when it comes to the sensitivity ratings (which range from 1 to 4) and the liquidity ratings (which range from 2 to 5). For the original model, we added new parameter ratings to solve these problems. To get around these limitations, we used the Expectation-Maximization (EM) technique, which estimates missing values and fills them in with statistically inferred data. This made it possible for us to include more factors and build a more complex and thorough model.

By improving prediction reliability and taking into consideration new data gaps, the revised model in figure 8 offers a richer understanding of the CAMELS rating system. For instance, there were some limitations in the first model because the CAMELS ranking predictions were primarily based on the available data set. Nonetheless, the research is more reliable and more accurately captures the dynamics of bank rankings under more different conditions thanks to the extended model that considers additional ranking parameters variables. With this enhancement, forecasts are protected from being unduly distorted by missing or insufficient data, providing decision-makers with more precise and insightful information.

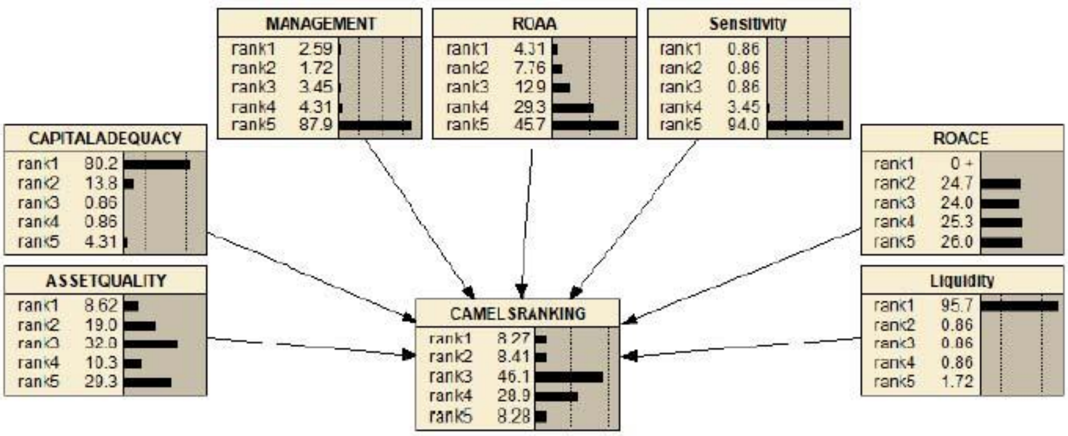

Figure 8. Revised trained data model using expectation maximization

4. Findings

The combined results of our investigation point to a number of important understandings regarding the CAMELS grading system, which is based on both the original and improved models. Although the original model offered insightful information, it also pointed up areas where more information could improve predictions, especially in the sensitivity and liquidity parameters. In order to solve this, we estimated missing data points using the Expectation-Maximization (EM) algorithm and added extra parameters, creating a more complete model.

The results highlight how crucial it is to have constant data availability across all criteria in order to guarantee the most precise and significant conclusions.

The present research contributes to a deeper comprehension of the CAMELS framework and offers a more resilient instrument for forecasting bank stability and performance under diverse circumstances. As such, it is an invaluable resource for risk mitigation and strategic decision-making within the banking industry.

5. Conclusion and Discussions

This study has successfully utilized Bayesian Belief Networks (BBN) in conjunction with the CAMELS framework to assess the financial stability of Lebanese banks between 2012 and 2022. In order to gain a better predictive understanding of bank stability, the BBN technique made it possible to conduct a thorough investigation of the interdependencies among the different CAMELS factors: capital adequacy, asset quality, management, earnings, liquidity, and sensitivity to market risk.

The results highlight that although capital adequacy ratios at Lebanese banks have been reasonably constant, other aspects, have demonstrated variable degrees of risk, which may affect the banks' long-term stability. A more dynamic and predictive model has been made possible by the inclusion of machine learning techniques like as BBN, providing important insights into the ways in which these components interact and impact total bank ratings.

Future research can expand on this by integrating more real-time data and exploring additional external factors to further enhance the predictive power of the models. The findings underscore the importance of strong governance, risk management practices, and the use of advanced data-driven decision-making tools to strengthen the resilience of the banking sector in challenging environments.

Furthermore, the study's recommendations for actions are:

Improved practices of Risk Management: Particularly in areas like management efficiency and susceptibility to market risks that have been identified as weak, banks should concentrate on fortifying their risk management frameworks. These hazards could be reduced by putting in place stronger internal controls and stress testing procedures.

Data-Driven Decision Making: To improve their decision-making processes, banks should make advantage of cutting-edge analytical tools and machine learning models, such as the BBN technique employed in this work. Banks are able to make well-informed strategic decisions and proactively handle emerging risks by consistently monitoring important financial indicators.

Boosting Corporate Governance: Banks' ability to withstand financial crises depends on their ability to exercise effective governance. Prioritizing governance improvements that promote accountability, openness, and long-term strategic planning is a good idea for boards and senior management.

Creating Crisis Contingency Plans: Banks ought to create and update thorough crisis plans that cover a range of crisis situations, such as natural disasters, political unrest, and economic downturns. Comprehensive procedures for managing liquidity, communicating with customers, and maintaining operational continuity should be included of these plans. These strategies should be tested and updated frequently to guarantee that they are ready for unanticipated circumstances.

Cooperation with International Financial Institutions: Lebanese banks could gain from better cooperation with international financial institutions given the global nature of financial markets. By providing them with access to best practices, and technical know-how, these collaborations can increase their adaptability and competitiveness on a worldwide scale.

References

Abid, L., Zaghdene, S., Masmoudi, A., & Ghorbel, S. (2017). Bayesian network modeling: A case study of credit scoring analysis of consumer loan's default payment. *Asian Economic and Financial Review, 7*(9), 846-857. https://doi.org/10.18488/journal.aefr.2017.79.846.857

Beltran, M., & Muoz, A. (2014). Redes bayesianas aplicadas a problemas de credit scoring. *Una aplicacion practica, Cuadernos de Economia, 37*(104), 73-86. https://doi.org/10.1016/j.cesjef.2013.07.001

Bouchaala, L., Masmoudi, A., Gargouri, F., & Rebai, A. (2010). Improving algorithms for structure learning in Bayesian Networks using a new implicit score. *Expert Systems with Applications, 37*(7), 5470-5475. https://doi.org/10.1016/j.eswa.2010.02.065

Bradley, A. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition, 30*(7), 1145-1159. https://doi.org/10.1016/S0031-3203(96)00142-2

Castillo, E., Gutierrez, J. M., & Ali. H. (1997). Sistemas expertos y modelos de redes probabilsticas. *Monografas de la Academia de Ingeniera, Madrid.*

Davutyan, N., & Ozar, S (2006). A credit scoring model for Turkey's micro & small enterprises (MSE's). *In 13th Annual ERF Conference*, 16–18, Kuwait, 2006.

Dbouk, B., & Zaarour, I. (2017). Towards a machine learning approach for earnings manipulation detection. *Asian Journal of Business and Accounting, 10*(2), 215-251.

Demerjian, P. (2007). Financial Ratios and Credit Risk: The Selection of Financial Ratio Covenants in Debt Contracts. *AAA 2007 Financial Accounting & Reporting Section (FARS), 2007.* https://doi.org/10.2139/ssrn.929907

Gevaert, O., De Smet, F., Kirk, E., Van Calster, B., Bourne, T., Van Huffel, S., & Condous. G. (2006). Predicting the outcome of pregnancies of unknown location: Bayesian networks with experts prior (398) information compared to logistic regression. *Human Reproduction, 21*(7), 1824-1831. https://doi.org/10.1093/humrep/del083

Ghribi, A., & Masmoudi, A. (2013). A compound poisson model for learning discrete Bayesian networks. *Acta Mathematica Scientia, 13*(6), 1767-1784. https://doi.org/10.1016/S0252-9602(13)60122-8

Hassen, H. B., Masmoudi, A., & Rebai, A. (2008). Causal inference in biomolecular pathways using a Bayesian network approach and an Implicit method. *Journal of Theoretical Biology, 253*(4), 717-724. https://doi.org/10.1016/j.jtbi.2008.04.030

Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications, 32*(4), 995-1003. https://doi.org/10.1016/j.eswa.2006.02.016

Saad, A., Zaarour, I., Edine, A. Z., Ayache, M., Bejjani, P., Guerin, F., & Lefebvre, D. (2013). A preliminary study of the causality of freezing of gait for Parkinson's disease patients: Bayesian belief network approach. *International Journal of Computer Science Issues, 10*(3), 88-95.

Tuya, J., Ramos, I., & Dolado, J. (2007). Dolado, Tecnicas cuantitativas para la gestion en la ingenieria del software. *NetBiblio, La Corua, 2007.*

Witzany, J. (2017). Credit Risk Management Pricing, Measurement, and Modeling. *Credit Risk Management*. https://doi.org/10.1007/978-3-319-49800-3_7

Zaarour, I., Saad, A., Eddine, A. Z., Ayach, M., Guerin, F., Bejjan, P., & Lefebvre, D. (2015). Methodologies for the diagnosis of the main behavioral syndromes for Parkinson's disease with Bayesian belief networks. *Emerging Trends in Computational Biology, Bioinformatics, and Systems Biology*, 467-485. https://doi.org/10.1016/B978-0-12-802508-6.00026-0