International Journal of Scientific & Engineering Research, Volume 14, Issue 8, August 2023
ISSN 2229-5518

A Smart Data-Driven Approach to Improving Public Fiscal Performance

Haitham M. Khalifa, Samir I. Al-Said, and Samir A. El-Fotouh

Abstract— Modern technologies, such as (Big Data analytics, AI, Blockchain, etc.) have emerged to keep pace with the needs of society and companies that provide products and services of various kinds. The interest in data quality and employment information to extract practical knowledge has increased to help transform people's lives in businesses, industry, government, and services. In developing countries, the financial sector has already applied (FMIS) and Performance-based Budgets to ensure proper scrutiny of budget estimates and to ensure (the accuracy, effectiveness, and efficiency) of government revenues and expenditures. However, it cannot face the financial obstacles resulting from external crises. Therefore, this research aims to investigate emerging artificial intelligence techniques in a financial analysis prediction to improve enterprise and local institutes' performance. we explored the studies that provide a novel early warning system to strengthen financial management by exploiting Data Science and AI techniques to face these difficulties. To accomplish the research objective EKB databases including Science Direct, Springer-Link Journal, IEEE Xplore, and Emerald are employed. Our study will shed light on the various research in this field and provide a pathway for us to analyze recent financial distress prediction models to enhance financial management decisions. Finally, we suggested a smart FMIS by emerging a data-driven AI-Based approach Based on a comparative analysis of recent research related to financial prediction. The contribution of artificial intelligence to finance has several forms, including ANN and its branches, integrated forecasting and prediction models with ANN, decision trees, sentiment analysis, and an AI-explainable approach. We discovered that the best techniques for predicting financial distress are Integrated Z-score, MLP, ANN, and Hybrid CNN, LSTM, and AM model. This article is the first study that explores opportunities for applying AI techniques to local government financial information systems.

Index Terms— Public Budget, Economy Crisis, Bankruptcy**.** Financial Information System. Artificial Intelligence, F-Distress Prediction

---

1 Introduction

Now, state-of-the-art technologies (IoT, AI, Blockchain, etc.) Have emerged to keep pace with society and companies' exigencies that provide products and services of various kinds. The interest in data quality and employment information to extract proper knowledge has increased to help transform people's lives in businesses, industry, government, and services. Moreover, in recent days, the world has been exposed to multiple threats such as (COVID-19) and economic crises that are the largest after the 2008 crisis. These challenges have posed high instability to governments and their citizens [1]. In summary, Economists and financial analysts expected that the countries of Africa and the Middle East, which have the largest import platforms for basic commodities, would achieve huge profits from oil and food supplies, but price fluctuations and the instability of the exchange rate of currencies were not taken into account, which may lead to a sharp decline in investment and the trade exchange [2]. Recent statistics have demonstrated the impact of the (COVID-19) pandemic, which was the number one reason for digital transformation. Therefore, "it has become imperative for governments worldwide to strengthen their capacity to strategically use new intelligent technologies and develop innovative, smart public services to confront and overcome the pandemic" [3].

At the other time, the modernization of financial management operations, the implementation of programs, and the performance of budgetary systems are the most important recommendations of the international organizations for developing countries such as (the Egyptian, Jordanian, and Turkish Republics). Developing countries have responded to international organizations' desire to undertake economic reforms. According to [4], IFMIS provides good fiscal performance (i.e., clearness, dependability, velocity, timeliness, validity, and accountability) by supporting the automation and integration of PFM processes to enable the publication of past, present, and future public finance positions [5]. These elements are critical to the effectiveness of fiscal policy. In the context of the programs and performance budget system, the budget formulation is the step that involves allocating resources before submitting them to the Legislative Assembly for consideration and final approval. It is the local organization's directors' responsibility to ensure proper scrutiny of budget estimates and ensure (the accuracy, effectiveness, and efficiency) of government revenues and expenditures. In Egypt, in a speech to the head of the program budget and performance unit, "The PPB Unit at the Ministry of Finance is effectively the government focuses on it to crystallize planning programs and development projects and ward off the risks of economic and social fluctuations. It also helps to allocate public resources to implement programs with specific goals, while subjecting them to criteria for measuring effectiveness and efficiency" [6].

Public financial performance refers to the measure of government-level performance and how to measure and assess whether macroeconomic objectives have been achieved and

- *Haitham M. Khalifa. Researcher in Business Information Systems, Helwan University, Cairo, Egypt. email: hytham00kh@gmail.com*
- *Samir I. Al-Said. Faculty of Commerce and Business Administration, Helwan University, Cairo, Egypt. Email: drsamironline99@gmail.com*
- *Samir A. El-Fotouh Faculty of Commerce, Mansoura University, Mansoura, Egypt. Email: prof_samir@hotmail.com*

whether public finances can be considered healthy on a macroeconomic level [7]. Many studies shed light on financial performance considerations based on business and technology. Financial ratios of financial/non-financial corporates assist in diagnosing the corporates' current/periodic financial situation along with putting insight into the corporates' potentiality to reach conducting business in the future. Based on financial distress articles, the most popular fiscal ratios are categorized as (financial position, liquidity, solvency, and fiscal capacity). Moreover, public sector debt management differs from private entities because of the potential for many local governments to collect revenue through taxation, access to additional financing from central governments in distress, and their redistributive role in economic management [8]. According to recent statements by many international organizations, the rapid increase in government debt is one of the main problems in the global economic crisis, which the novel coronavirus has exacerbated, and another pandemic [9]. International organizations also referred, by the way, to improve the performance of public finances through the Public Expenditure and Financial Accountability model, which focuses on forecasting and control practices that predict payment obligations to public sector entities arising from loan contracts.

Despite The evolution of the financial services reform being closely associated with progress in several frontier technologies, including some critical software-oriented technologies such as blockchain, data analytics, and AI, there are limitations in financial performance measurement. According to [10], The articles on financial performance measurement highlight several limitations and constraints with financial indicators that relate to the incompleteness of purely financial information, its reliability, its comparability in a different context, and the impact of external factors.

This paper is organized as follows: Section 2 examines the relationship works. The comparative analysis of Almost techniques in the literature will be given in Section 3. The procedure for proposing a smart data-driven for financial system will be presented in Section 4. The paper's conclusion and potential future research directions will be provided in Section 5.

2 Related Work

Since the global financial crisis, many researchers have sought to determine the effects of financial pressures on economic activity. It was mentioned [11] that among these pressures is corporate bankruptcy, which is one of the main drivers of credit risk and receives immediate attention from creditors and investors. However, the financial damage caused by corporate bankruptcy cannot be accurately determined. For these reasons, researchers are looking for more effective forecasting models to predict bankruptcy and financial distress. Predicting bankruptcy papers routinely adopt measures, including algorithmic trading, credit scoring, information, and accounting data from company financial statements to predict bankruptcy. According to [12], prediction is an essential yet challenging part of time series data analysis. Financial series analysis has long been a leading financial engineering and enterprise risk management area. Other industry-dependent factors such as seasonality, economic shocks, and unexpected events internal and produce the data that affect the forecast, and statistical tools or techniques that reveal rules and predict future circumstances through financial risk analysis that have a guiding signature of both governments and corporations to forecast revenues, and costs to avoid known financial risks.

2.1 Fiscal Distress Prediction

Economic globalization and capital market expansion are among the main reasons companies fall into competition struggles and bankruptcy threats. In this case, it is necessary to exercise caution and caution on the part of managers and decision-makers by analyzing the financial position of institutions periodically and early detection of financial risks and the reasons for their occurrence [13]. Otherwise, they will cause huge losses to financial institutions and will cause harm to investors, disrupting the global economic and political system. Financial hardship refers to several cases, including the case in which the company fails to fulfill its debt obligations towards its creditors when due, and from it the inability of institutions to provide the necessary resources to achieve its goals, which appears mainly when reviewing the estimated budgets.

FDP is a multi-perspective controversy in benchmarking corporate fiscal stability. Its essential goal is to discriminate between balanced companies compared to those at risk of financial distress [14]. From an enterprise perspective, FDP may assist in identifying hazards early, making strategies based on the current circumstances, and adjusting corporate strategy. From an investor's perspective, FDP can assist organizations in determining their financial risks and selecting investment projects based on their risk preferences. From the regulator's perspective, FDP may assist in understanding each company's financial position promptly, performing well in monitoring and management and ensuring the financial market's stability.

Traditionally, the assessment of the financial distress situation of companies was mainly based on the subjective judgment of experts. Therefore, how to effectively predict enterprises' financial distress has become a hot topic in academic and business circles [15]. Typically, financial distress prediction uses mathematical, statistical, or intelligent models to predict whether an enterprise will be in financial distress based on current financial data. Based on the literature the most famous FDP models shown in Table 1 are Z-Score [16], O-score [17], Probit Model [18], Hazard Model [19], and D-Score Model [20].

However, these statistical models have strong data standards, and the data must match tight parameters such as normal distribution, high sample size, and elimination of overlap. Unlike traditional statistical techniques, machine learning does not presuppose a certain data distribution and may automatically extract information from training samples. By and large, the studies that are interested in financial risk prediction are split into 2 categories: statistics and AI-Based Models. In the next sections, much recent research has demonstrated that ensemble learning strategies result in improved accuracy

Table 1
Statistical Financial Distress Prediction Models

Models	Analysis Technique	F-Ratio
Z-Score [16]	MLPDA	*WCTA, RETA, EBITTA, MCTL,STA*
O-Score [17]	Logit	OSIZE, TLTA, WCTA, CLCA, OENEG, NITA, FUTL.
Probit [18]	Probit	NITA, TLTA, CACL
Hazard [19]	Hazard	NITL, TLTA, RSIZE, LEXRETURN, LSIGMA
D-Score [20]	Logit	NITA, TDME, META, CLTA

2.2 F-Distress Prediction by ML/DL

Machine learning (ML) is the primary technology underlying artificial intelligence. Machine learning approaches enable machines to do complicated tasks such as recognizing faces, comprehending voices, and responding to communications. Given the potential of machine learning technology, it is logical to wonder whether ML approaches may be used elsewhere. This section discusses the application of ML/DL approaches to problem-solving in financial research [21]. Deep learning is a subset of machine learning that is essentially a three- or more-layered neural network. These neural networks seek to imitate the behavior of the human brain, albeit with limited success, allowing it to "learn" from enormous volumes of data. While a single-layer neural network may still produce approximate predictions, additional hidden layers are required. Machine learning/Deep learning methods make realizing many complex financial models or analysis and forecasting methods simple and possible [22]. [23] Presented a combination prediction model based on LSTM, CNN, and AM. LSTM is used to solve the gradient explosion problem in long-term sequence prediction. CNN employed for feature selection, and select the most significant indicators related to financial risk prediction. The attention mechanism (AM) can learn autonomously, choose suitable parameters, optimize the model, and increase the CNN-LSTM model's prediction accuracy. The independent variables are classified into five categories: liquidity, leverage, profitability, activity, and non-financial. The results show that the most significant financial distress indicator is liquidity with (0.71%), and the best accuracy score was for LSTM-CNN-AM by (0.9843%).

Furthermore, merchandise costs are reaching record highs over all regions. The Food and Agriculture Organization of the United Nations (FAO) issued its third consecutive peak inflation index. Food prices were 34% higher than in the same period last year and have never been higher since the FAO began keeping track of them. Similarly, crude oil costs have risen by around (60%), while petrol and fertilizer prices have more than doubled. [24] Offer dual EW approaches for oil-related fiscal hazards, particularly the Bayesian network prediction model and the BPNN model. A (65%) connection existed between the NN prediction and the BPNN prediction results. The BN outperformed the BPNN in terms of prediction, with a fitting effect of (80%). As a result, the BN is robust to perform risk EW in oil finance. They split the sample data set from (1987 to 2015) into two halves. The first sets of samples were inserted into the NN model for training, as with the autoregressive model, The remaining four groups of data were simulated and evaluated to assess the effect of the NN fitting tool in the matlab2015b program, and the forward NN toolbox was used to finish the data fitting. The results of the test samples during the previous four years show that (R = 0.65) was the correlation between the actual and predicted production. Certain capabilities are provided by network simulation. Also, the simulated output results were (0.860, 0.893, 0.915, and 0.880), whereas the risk levels after de-normalization were (1.251, 1.353, 1.424, and 1.311), which were quite close to the predicted output results.

[25] Provided an exploratory study to discover Malaysian publicly listed corporations over time-series data. This study compares the performance of (LR, NN, SVM, DT, RNN, LSTM, and GRU) using annual data before the state's year. The data used for this study are the financial profile of corporations listed on Bursa Malaysia that were classified under PN17 status as the worst case and non-PN17 as the best case from (2011 to 2017). The authors used the machine learning library SCI-kit-learn and Keras framework in Python as a programming language and Jupiter notebook as the algorithmic platform. For RNN, LSTM, and GRU Keras platform is used. The other model LR, NN, and SVM used by the SCI-kit-learn and Keras framework. The study uses accuracy, recall, and precision of performance measurements. The results show that all of the deep learning models used in this study had (90%) accuracy or higher, but that LSTM and GRU had the best accuracy performance with (93%) accuracy, which is (3%) better than the RNN model. They also had better recall and precision than RNN. Finally, Long-term memorization works better with LSTM or the updating gate in GRU.

In the context of the government financial sector, three strong elements may significantly affect financial management performance: the budget's reliability, the matching of the expenditure balance with the revenue, and finally the government's ability to repay the debt. In China, [26] constructed a framework to present a risk assessment system for LG debt. This study uses the analytic hierarchy process and the entropy approach of comprehensive weighing to make index weight calculations more objective and acceptable. To examine the data more objectively, a thorough risk assessment value model was built in conjunction with the TOPSIS approach. Because debt risk is the outcome of numerous variables acting in concert, this article selects the index based on three parameters: economic condition, fiscal revenue and expenditure, and LG debt status. The growth rates of GDP and regional fixed asset investment have been chosen as indices of regional economic development, and the alert value of this indicator is set to 6%. To assess the financial position of regions, four indicators are chosen: the ratio of fiscal revenue to GDP, the ratio of fiscal spending to GDP, the ratio of fiscal deficit to GDP, and the rate of political self-sufficiency, and the warning lines were set at (10% and 15 %), respectively. LG's debt condition is assessed using three indicators: debt ratio, debt-to-burden ratio, and new debt ratio, and the alert value of this indicator is set to (100%). This paper adopts BP and CART machine learning algorithms to carry out early warnings on local government debt risk. The NN toolbox in MATLAB software was utilized for training and testing. The accuracy of the systemically important government EW forecast based on the monthly value of the recent year is as high as (85.72%), and the failure probability is (14.28%). The simulation value and the real value of the test samples are close, with a fitting degree of (97.9%) and a mean square error of just (0.00749), showing that the simulation effect is good and the warning effect is optimal.

2.3 F-Distress Prediction by Hybrid Models

Most of the studies are concerned with analyzing and measuring the efficiency and accuracy of basic time series analysis/financial distress analysis models such as (ARIMA, GARCH, Altman, etc.) And ML/DL models such as (ANN, RNN, CNN, LSTM, SVM) concluded the following: Working with traditional models separately is not accurate enough to predict financial deficits. Therefore, many studies have tried to use the hybrid model methodology founded by [27] to diagnose the accuracy of the model output data, increase the chance of catching linear patterns with nonlinear patterns in the data, and improve the prediction result performance.

[28] criticized the traditional algorithms in financial forecasting by the time series method and mentioned to ANN has a perfect self-learning ability, adaptive ability, generalization ability, and non-linear mapping ability. Hence, they constructed internal forecasting by adding a fuzzy neural network algorithm. The results of this study proved that the algorithm can achieve high classification accuracy, and to compare it with the other, the researcher used the MATLAB program, and the result showed that the accuracy rate of FNN was (96.31%), which is better than the expected accuracy of the CNN algorithm by (9.84%).

In this context, [29] stated that a high-accuracy fiscal forecasting framework is critical for corporations that have fiscal and quality control issues, as stated due to the report on the health score of listed companies in China (2021), to improve financial performance and maximize investment decisions. They integrate the Z-Score and MLP-ANN algorithms to anticipate the health of Chinese A-share companies. They independently examine the predicted results of the Z-Score model, MLP-ANN model, and combination model to analyze the difference in prediction capabilities of each approach and indicate which is most suited for the Chinese stock market. They used a CSMAR database system that was built using financial statement data for all businesses listed on the Shenzhen and Shanghai stock exchanges between (2016 – 2020). A total of observations was obtained by utilizing average values to supplement missing data. Solvency, operational capacity, profitability, and development capacity are the enterprise performance ratios. The five predictors (or independent variables) are as follows: WCTA (working capital to total assets), RETA (retained earnings to total assets), EBITDA (EBITDA to total assets), MVETA (market value of equity to total liabilities), and STA (sales to total assets). The popular SPSS program will be used to implement the Z-score model, MLP-ANNs model, and hybrid model. The empirical results reveal that the novel hybrid model outperformed the Z-score model (86.54%) and the pure neural network technique (98.26%) in terms of average accurate classification rate. While the MLP-ANN model had a better overall classification success rate, it was skewed since it was applied to a highly unbalanced data set. While that model was not degenerate (calling all situations safe from bankruptcy), it did forecast bankruptcies only seldom. The integrated model's Z-score component handled the dataset imbalance issue.

In light of the difficult economic conditions, most of the financial decisions may be complex and may result in conditions of uncertainty and errors in estimates of financial indicators. The exploitation of innovative management accounting techniques such as a sustainable balanced scorecard, target costing, and life cycle costing can improve the decision-making process and contribute to effective management control by providing comprehensive data enabling the adoption of control procedures, ultimately contributing to the firm's competitiveness. [30] And [31] provided an early warning system for management costs. [30] Employed a balanced scorecard approach and LSTM algorithm in their study. The BSC model provides accurate data about management costs. LSTM's function is to predict the distress of financial indicators for long-term conditions. The authors used monthly data on economic activities in Turkey from 1996 January – November (2009) period Among the selected indicators, Loans, Reserves, Portfolio Investment, Net Errors and omissions, and Domestic Debt Stock. The results using the MATLAB program found LSTM's success in predicting the financial crisis of Turkey from (2002 to 2008) with (95.45%) accuracy.

[31] Established an FCEW model using PSO-SVM. He stated that an SVM has the issue that the unpredictability of weight and threshold parameters impacts its early warning effect and that Particle Swarm Optimization (PSO) may discover the ideal particle position to compensate for the flaws of random solutions to enhance accuracy. The initial stage in building the model was to identify the indicators that indicate a financial crisis. As a result, the following (24) financial indicators are chosen from six categories: innovation and development capacity, cash flow ability, profitability, operational ability, solvency, and equity structure. From (2015 to 2017), research data was acquired. After removing the samples that had missing data, The prediction accuracy is assessed using the evaluation index for unbalanced samples, and the evaluation indexes for unbalanced sample classification are geometric average accuracy rate G and minority class measure F. The experimental findings reveal that the PSO-SVM model outperforms other models for the two types of samples in terms of fitting and prediction performance under varied data partitioning ratios. In detail, as the penalty factor is increased, the accuracy of the PSOSVM model steadily climbs to (100%). When the number of nodes in the hidden layer of test samples reaches roughly (16), the accuracy hits its peak and begins to stabilize. Whereas the number of hidden layer nodes in the SVM model is (24), the training accuracy is (80%), and there is a significant difference between training and test sample accuracy. The comparison demonstrates that the SVM model is less resilient than the PSO-SVM model.

2.4 F-Distress Prediction by Decision Trees

On the other hand, data mining and decision trees get a chance to improve financial distress by focusing on management costs. [32] Optimize the decision tree algorithm ID3 by the PCA algorithm to be suitable for enterprise cost control. Many factors related to enterprise cost were chosen such as: (main business cost, management cost, and sales cost). Finally, the authors tested KNN and Naïve Bayes in risk detection accuracy. The study results show that the optimized algorithm's accuracy is (2.2%) higher than the traditional ID3 and (1%) higher than the traditional PCA algorithm. The accuracy of the KNN algorithm was (87.5%) and the accuracy of the Naïve Bayes algorithm is (93.7%). In terms of detection ability, the overall index for the logistic early warning model (the mean of average detection rate, average accuracy) gets a (72.2%). In business and market conditions of the economic context, Businesses' ability to avoid bankruptcy depends on a variety of financial and non-financial factors, yet failure or non-bankruptcy can be assessed using financial ratios.

On the other hand, [33] suggested the random forest method to rank the importance of some financial indicators based on screening financial data from Chinese listed companies. Furthermore, he employed the CART decision tree algorithm for certain financial warning indicators can greatly improve the model's functionality. Mean decrease accuracy (MDA), out-of-bag rate (OOB), and Area of the under-ROC curve (AUC) are the assessment methods for financial risk indicators ranking and evaluating the model's results accuracy. Research results show that the early warning model has a low out-of-bag error rate (OOB ERROR=8.41%) as well as a high AUC Value (AUC=0.909), also specifically indicates significance ranking as 1) net assets per share, 2) total asset, 3) net profit, 4) total assets turnover ratio, 5) earnings per share. Finally, he indicates that the constructed model has a good predictive ability.

2.5 F-Distress Prediction by Explainable AI

The most current generation of AI algorithms with a rationale as a black box that has limited visibility into financial risk identification findings. Complex machine learning models are prominent in FDP, although most scientists have not thoroughly investigated their interpretability. Previous AI research has been conducted by [34] and [35]. The purpose is to assist machine learning developers debug models or offer the observer a clear explanation of the activities, neglecting other explanatory demands of external users. The first paper by [34] provides an Explainable AI (XAI) model, and they have deemed it necessary to understand the reasons for making decisions and to trust them. CBR is a significant type of knowledge-based system, its function is to solve new problems based on past experiences. The explainable CBR model includes four queries related to transparency, justification, relevance, and learning. The experiment research was adopted to evaluate the model's performance, they first determined the almost classifier methods namely (logistic regression, K-nearest neighbor, decision tree, Gaussian naïve Bayes, multilayer perception, and equally-weighted CBR), until achieving evaluation purposes they employed several types of financial risk indicators such as (Credit Card default, Bank Churn, and financial Distress). The results show that the equally-weighted CBR and multilayer perception have the best accuracy index of (0.83%, and 0.75%).

Another XAI technique was developed by [35]. They combine a whole process aggregation methodology and an explainable structure to fulfill the interpreter needs of third-party stakeholders while maintaining strong prediction performance. They investigate external stakeholders and their interpretive desires before developing an explainable framework. To address the interpreter's needs, Shapley Additive explanations (SHAP), Partial Dependence Plots (PDP), and counterfactual explanations are used. The authors studied (LR, SVM, DT, RF, and GB) in the FDP instance. Their method is divided into two parts: 1) a full ensemble model and 2) an explainable model for external users. In the first step, the significance test is used to examine if there are significant variances between the characteristics of financial distress and healthy firms. On the other hand, classifier construction LR, SVM, and DT are the candidate base classifiers. Profitability, solvency, operational capacity, growth capacity, cash liquidity, structural ratios, and per-share ratios are among the seven financial factors considered. They gathered information from Chinese A-share businesses listed on the Shanghai and Shenzhen stock exchanges. They are gathered from the years (2007 through 2020). The overall evaluation metrics in this study are (ACC, AUC, and KS). The performance outcomes of the three models are selected in the Ensemble strategy. The LightGBM model performs best (ACC=84.7%), followed by the XGBboost model (ACC=83.5%), while the RF model performs poorest (ACC=82.9%). Second, the LightGBM has the highest KS value (KS=0.702), indicating that it has the best capacity to differentiate between positive and negative samples. Finally, the AUC value shows the classifier's ability to rate persons properly, with the LightGBM having the greatest AUC. In terms of overall performance, the LightGBM performs best overall. Regardless of the evaluation index, the DT model outperforms the SVM and LR models in predicting performance. The Operating profit growth rate, the Growth rate of net flow from operating activities, the Cash ratio, and the Tangible asset-liability ratio has a bigger influence than the other features.

2.6 F-Distress Prediction by Sentiment Analysis

In recent years, both academic and industrial disciplines have paid close attention to text mining in the financial field. This is because text mining is an excellent method for analyzing financial markets or economic events [36]. Annual reports, financial news, analyst reports, and associated social media content may contain information that does not show clearly in numeric-financial statistics. As a result, various researchers have extracted important characteristics from these financial texts and utilized them to address classic financial issues such as stock prediction, financial distress prediction (FDP), and fraud detection. FDP success is primarily determined by predictive criteria, which include both financial and non-financial characteristics. Furthermore, in recent years, many academics have focused on extracting and optimizing non-financial aspects.

[37] Constructed a domain sentiment lexicon and conducted a sentiment analysis for the FDP. They employed Python to undertake the SL construction, SA, and FDP. This study collected data from (214) Chinese-listed companies that had received the ST label from (2012 to 2018). The method integrates dictionary-based and corpus-based approaches and includes three steps: data preprocessing, word vector model construction, and classifier construction. Data preprocessing focuses on generating a collection of sentiment words from a specific domain, based on domain corpus and general sentiment lexicons. Word vector model construction calculated the similarity of words in the seed words collection in the financial domain corpus respectively obtained word vectors through Word2Vec and the BERT

pre-training model. In Classifier Construction they developed three deep learning-based classifiers, namely: deep neural networks (DNN), multi-head attention-based DNN (MA-DNN), and bidirectional long short-term memory (Bi-LSMT). They applied three evaluation metrics: accuracy, precision, and recall. To compare and optimize the predictive models of the FDP. This study applied SVM, DT, Xgboost, and DNN which are common and widely used algorithms in the FDP. After that, this study adopted DT, SVM, Xgboost, and DNN as the predictive models to achieve FDP and evaluate the experiment results during the different periods. This finding indicates that the sentiment factors generated by this study affect financial distress. According to the experimental results, predictive models achieved the best performance. DNN reached the optimal value for predictive models under each evaluation metric, followed by Xgboost. Also, SVM and DT. DNN obtained the highest accuracy (85.71%). Also, Xgboost demonstrated the highest recall (85.03%). It is possible to conclude from these findings that sentiment features can independently achieve relatively satisfactory predictive performance in the FDP.

[38] Developed another framework comprising data collecting, preprocessing, feature development, and financial distress prediction models. In terms of data gathering, the dataset comprises accounting data as well as patent data. A patent is made up of organized and unstructured material (abstracts, claims, citations, and specifications). In terms of data preparation, they analyze missing values in accounting data and eliminate characteristics with a high number of missing values. In terms of feature building, statistical features are built from structured data, whereas semantic features are built from unstructured data using multi-natural language processing approaches. They employed four exemplary machine learning methods for FDP models, including extreme gradient boosting (XGB), logistic regression (LR), random forest (RF), and gradient-boosting decision trees (GBDT). In the case of document embedding, they use BERT to create semantic characteristics from the patent text. These characteristics indicate the firm's stock, growth capability, profitability, and operational capacity. Patient characteristics in this study are classified into two types: statistical features and semantic features. From patent-structured data, statistical characteristics are created. The linear model LR is one of them, while the nonlinear ensemble learning models XGB, RF, and GBDT are others. The models' performance evaluation is calculated by the (ROC) curve, the (KS) of the models, and the H-measure, which uses different misclassification cost distributions for different classifiers to compensate for the AUC measure's deficiency. Shapley Additive explanations (SHAP) are used to calculate feature significance. The results show that the combinations A + PT and A + PS outperform A, implying that patent characteristics (statistical features and semantic features) provide more information for predicting financial hardship than accounting factors. The discriminating performance of the feature set A + PS + PT exceeds three others (A, A + PS, and A + PT). The findings show that semantic aspects of patents may accurately detect financial hardship and can be integrated with the statistical features to improve the discrimination performance of FDP models. Furthermore, the A + PT beats A + PS, demonstrating the significance of mining technical semantic traits.

3 A Comparative Analysis

In this section, we will review and compare various techniques employed in the financial domain. The main indicators for this comparison are author, field, financial indicators ranking, prediction algorithm, and accuracy index.

Table 2
A Comparative Review of Recent FDP Techniques

Author	Domain	Indicators Rank	Prediction Algorithm	Accuracy Index
[23]	Stock Market	Current Ratio Quick Ratio Cash Ratio	CNN+ LSTM+ AM	AUC: 0.9843
[24]	Energy Industry	Oil Financial Risk	Bayesian Net BPNN	0.860 0.880
[25]	Public Co.	Gross Margin Leverage Quick Ratio Debt/Equity	LSTM GRU RNN	93% 93% 90%
[26]	Government Debt Management	Debt Rate Debt Burden Ratio New Debt Ratio	BPNN CART	85.72%
[28]	Financial Market	Purchase Quantity Purchase Time	FNN	96.31%
[29]	Stock Exchanges	The ratio of working capital /total assets The ratio of equity/total liabilities The ratio of sales / total assets	Integrated Z-score MLP ANN	99.4%
[30]	Management Accounting	Banking Loans Gross Reserves Debt Stock	Integrated SBSC and LSTM	95.41%
[31]	GEM Industry	Cash Flow Ability Profitability Solvency	Integrated PSO-SVM	For Sample 16: 100% For Sample 24: 80%
[32]	Enterprise Cost Control	Business Cost Sales Cost Management Cost	Optimized ID3 PCA	95.1%
[33]	A-Share Companies	Net Assets Per Share Total Assets Growth Rate Net Profit Total Asset Turnover	CART	OOB_ERROR: 8.41% AUC: 0.909

[34]	Financial Industry	Credit Card Fraud Credit Bank Churn Financial Distress	CBR-E MLP DT	0.8379 0.7561 0.6475
[35]	A-Share Companies	Operating profit Growth rate Cash ratio Liability ratio	LightGBM	0.92
[37]	OTC market	Debt assets ratio Beneficial effects Accounts receivable turnover days	LR RF GBDT XGB	0.917% 0.908% 0.985% 0.890%
[38]	Capital Markets Stock Markets	Net Assets Per Share Financial Behavior	DNN XGBOOST DT SVM	IN T4: 85.71% IN T3: 84.41% IN T3: 75.25% IN T4: 73.97%

Table 2 shows that there is a wide range of applications of artificial intelligence in the field of financial management performance, particularly financial forecasting, which is one of the most essential parts that reflect financial performance. This is especially crucial during this trying time. To begin, we can see that the majority of the application areas were in the private sector, except for two studies that concentrated on the public sector, such as [25], and [26]. As a result, we can see that it is vital to pay greater attention to the financial environment of the government because it is the source of the country's financial foundations. Although most studies focused heavily on financial indicators, which some saw as an important part of analyzing an organization's financial position, there were a few studies that focused on arranging financial indicators based on the strength of their impact on financial failures, such as ([37], [27], [30], [33], [36], [37], and [38]). They used the following methods (AHP, Delphi Method, ID3, MDA, AUC, and SHAP value) as well as according to these studies, the most relevant factors for financial failure are as follows (Debt Rate, The ratio of working capital /total assets, Business Cost, Net Assets Per Share, Operating profit, Debt assets ratio Beneficial effects, and Net Assets Per Share). In light of the researchers' criticism of traditional statistical approaches for analyzing financial failure, they devoted special attention to smart methods based on artificial intelligence and data science. Finally, it is intriguing to discover a smart system that analyses the financial situation of the institution while also taking care of its external elements, as demonstrated by studies [35], and [36] which are concerned with inputs and outputs from others, which considered these systems a black box whose task is only to analyze financial data and submit reports on it. In terms of inputs, they may be found in studies on elicitation from daily news and newspapers, as well as past organizational reports. The outputs, for the most part, are concerned with providing judgments on the causes of financial failure. As a result, decision-makers may benefit from this feature in the presence of quick and precise financial crisis solutions.

4 Smart Data-Driven Approach

4.1 FMIS Architecture

According to [39], Automation of Government Financial Management (GFM) Systems is a key component of the reform program because timely and accurate information is critical to the management of government finances and public funds; And it may simply be impossible to obtain the information required for economic management promptly without some degree of automation, given the large transaction volumes involved and their dispersal across multiple systems.

Some governments believe that the systems' scope and coverage should be expanded to include elements of the GFM chain such as budget preparation, payroll and position management, debt management, and auditing. There is also a need to include performance criteria in budgeting and to transition from pure cash accounting to increasingly sophisticated practices that include parts of accrual accounting. This might necessitate considerable system upgrades. Fig 1 depicts a complete IFMIS and its primary users. Innovative IFMIS solutions also allow more detailed analysis by providing dynamic query options to many users, both internal (MOF- Line Ministries, and MOF-

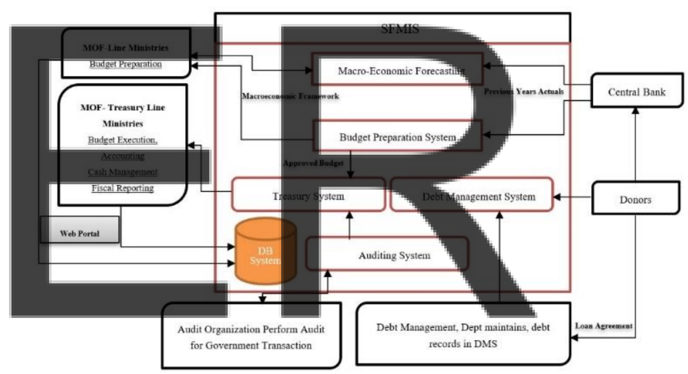

Fig 1. IFMIS Architecture

Treasury Line Ministries) and external (Central Bank, and Donors), and they support the publication of open budget data.

Through MOF- Line Ministries users can access to forecasting system, and a budget preparation system to predict the yearly expenditure and revenues. MOF- Treasury Line Ministries have access to the treasury system to execute the approved budget, monitor accounting records, and submit reports. The Central bank feeds into the system's previous actual data.

Donors and debt management offer and control loan procedures. Nevertheless, according to [34] long-term trends such as aging populations, rising inequality, climate change, and unpredictable challenges such as migration and security impact citizens' well-being and resilience. It necessitates governments developing a long-term vision, making evidence-based investments, and coordinating their efforts. Governments are becoming more aware of how better data use can support a response to these challenges while also improving policy productivity, performance, and inclusiveness.

International Journal of Scientific & Engineering Research, Volume 14, Issue 8, August 2023
ISSN 2229-5518

4.2 AI-Based Data Driven

A Data-Driven Public Sector (DDPS) transforms the design, delivery, and monitoring of public policies and services through data management, sharing, and use. Using data as a strategic asset is critical for governments to improve public sector intelligence and, as a result, increase the capability of developing long-term policies and services that are as inclusive and trusted as possible [40]. We suggest developing the analytical process in FMIS by employing intelligent Data-driven systems see Fig 2. Smart Data-Driven software solutions for information and data management. The two primary functions of these systems are the acquisition and presentation of information. Information acquisition is typically performed using data entry forms or interfacing with external data sources.

Information presentation involves retrieving and displaying Stored information to the user with appropriate navigation and querying facilities. Data-driven systems also require intensive user interaction for acquiring and retrieving information. Data beyond doubt is among the most common types of customized software systems in use today.

4.2.1 External Data Source

According to the studies presented above, we have added data sources

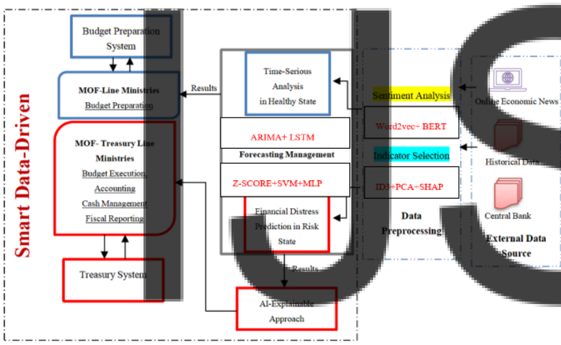

Fig 2. A Proposed Smart FMIS by Authors

represented in some sources such as newspapers and economic news that may be useful in warning financial institutions to beware of falling into severe financial crises such as changing the currency exchange rate or changing prices Consumer Goods in the section related to external sources of financial data. And the Central Bank's periodic reports may consider earlier difficulties whose problems have been remedied. Finally, past financial data from examining annual reports, the balance sheet, and the institution's financial indicators of the income statement, the balance sheet, and the institution's financial indicators, may become inputs for the modified forecast models.

4.2.2 Data processing

In this section, after collecting data, the first step of data processing is how a financial analyst can determine and normalize collected data. As we know there are multi types of data (structured, unstructured), also most collected data has noise issues that should be resolved and fixed. As shown in Fig 2 the most data processing actions in FDP study the indicator selection and sentiment analysis. Both results should alert financial management to take the necessary actions to avoid the crisis before it occurs. Classification of the most important financial ratio by ID3, PCA, SHAP value, and CART algorithms, able managers to clarify whether the financial ratio is related to financial distress.

On the other hand, sentiment analysis is combining machine learning, statistics, and linguistics to identify textual patterns and trends in unstructured data. Sentiment analysis may yield more quantitative insights by translating data into a more organized manner via text mining and text analysis. Then, data visualization tools may be used to disseminate findings to a broader audience. Word2Vec and the BERT are useful to convert text attributes into measurable values that can be used as predictive model input.

4.2.3 Smart Financial Forecasting Management

After the data processing is achieved the financial analyst and developers build or construct prediction models based on time series models, machine learning, or deep learning. Each technique has a specific situation as viewed in Fig 2. We suggest a time-series analysis of a normal state for corporates, Garph for linear prediction, and ARIMA for non-linear prediction. In the risk state, we suggest based on the literature LSTM, and SVM. The authors suggest the LSTM technique for long-term prediction, and they refer to SVM the is the highest accuracy results tool.

4.2.4 AI-Explainable System

It's very important to understand the critical to meet the interpretative expectations of external stakeholders for machine models, an explainable framework is presented. External stakeholders require interpretability, however, most XAI research overlooked their explanatory requirements. Using FDP as an example, this research examines external stakeholders and their interpretive desires before establishing an explainable framework. Shapley Additive explanations (SHAP), Partial Dependence Plots (PDP), and counterfactual explanations are used to address the interpreter's needs. Local explanations assist particular businesses in identifying the primary factors that contribute to their financial difficulty, and counterfactual explanations are developed to give improvement options. Global explanations can increase the transparency and believability of 'black box' models for external use by analyzing the relevance of features and the influence of feature interaction on the findings.

5 Conclusion

Even after the automation of planning and implementation of discretionary budgets in developing countries, particularly the Middle East and South Africa, the situation continued to follow fiscal policies based on the traditional budget, which international organizations such as the World Bank and the United Nations have described as opaque and lacking in efficiency and effectiveness. Also, management cannot still analyze and forecast financial performance during a global crisis. This study provided insight into recent studies that employ Machine Learning/Deep Learning to enhance fiscal performance. The capacity of corporations to properly estimate their fiscal state is reflected in their fiscal performance. The contribution of artificial intelligence to finance has several forms, including ANN and its branches, integrated forecasting and prediction models with ANN, decision trees, sentiment analysis, and an AI-explainable approach. We discovered that the best techniques for predicting financial distress are Integrated Z-score MLP ANN and Hybrid CNN+LSTM+AM. This work is a part of Ph.D. Study in future work we will empirically use this model in Egypt's electrical and energy industry.

References

[1] Whitelaw, S., Mamas, M. A., Topol, E., & Van Spall, H. G. C. (2020). Applications of digital technology in COVID-19 pandemic planning and response. *The Lancet Digital Health*, *2*(8), e435–e440.

[2] Fletcher, G., & Griffiths, M. (2020). Digital transformation during a lockdown. International Journal of Information Management, July, p. 55.

[3] Agostino, D., Arnaboldi, M, & Lema, M, D (2021). New development: COVID-19 as an accelerator of digital transformation in public service delivery. The UK. The Journal of Public Money & Management, 41(1), 69–72.

[4] Rozner F. & Gallagher M. (2008). Integrated Financial Management Information Systems. Practical Guide. No. 06. United States Agency for International Development.

[5] Denver, C., & Young Min, S. (2013). Financial management information systems and open budget data: Do governments report on where the money goes? Washington, DC: The International Bank for Reconstruction and Development/The World Bank.

[6] Ministry of Finance in Egypt (2020) State General Budget Retrieved from https://mof.gov.eg/ar/archive/stateGeneralBudget/ Accessed on 01/2023.

[7] Pulay, G., & Simon, J. (2020). Measuring the Macroeconomic Performance of Public Finance Management. Budapest. Journal of Public Finance Quarterly, 65 (Special Issue), 23–43.

[8] Caruana, J., Brusca, I., Caperchione, E., Cohen, S., Manes Rossi, F. (2019). Exploring the Relevance of Accounting Frameworks in the Pursuit of Financial Sustainability of Public Sector Entities: A Holistic Approach. In: Caruana, J., Brusca, I., Caperchione, E., Cohen, S., Manes Rossi, F. (eds) Financial Sustainability of Public Sector Entities. Public Sector Financial Management (1st Edition, pp. 1-18). Palgrave Macmillan

[9] International Monetary Fund. 2021. Debt management responses to the pandemic. Special series on COVID-19.

[10] Navarro, A., Lara, J., Buendía, D., & Rayo, ´ S. (2020). Analyzing political and systemic determinants of financial risk in local governments. Transylvanian Review of Administrative Sciences, 16(59), 104–123.

[11] Iacuzzi, S. (2022). An appraisal of financial indicators for local government: a structured literature review journal of Public Budgeting, Accounting & Financial Management. 34(6), 69-94.

[12] Zahra F. And Achchab S. (2019). A hybrid neural network model based on improved PSO and SA for bankruptcy prediction. Journal of Computer Science Issues, 16 (1).

[13] Yu, L., & Li, M. (2023). A case-based reasoning-driven ensemble learning paradigm for financial distress prediction with missing data. *Applied Soft Computing*, *137*, 110163.

[14] Ashraf, S., G. S. Félix, E., & Serrasqueiro, Z. (2019). Do Traditional Financial Distress Prediction Models Predict the Early Warning Signs of Financial Distress? *Journal of Risk and Financial Management*, *12*(2), 55.

[15] Qu, Y., Quan, P., Lei, M., & Shi, Y. (2019). Review of bankruptcy prediction using machine learning and deep learning techniques. *Procedia Computer Science*, *162*, 895–899

[16] Altman, E. I. (1968). Financial Ratios, Discriminant Analysis, and The Prediction of Corporate Bankruptcy. *The Journal of Finance*, *23*(4), 589–609.

[17] Ohlson, J. A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, *18*(1), 109.

[18] Zmijewski, M. E. (1984). Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research*, *22*, 59.

[19] Shumway, T. (2001). Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *The Journal of Business*, *74*(1), 101–124.

[20] Blums, M., & Macalester (2004). D-Score: Bankruptcy Prediction Model for Middle Market Public Firms.

[21] Hoang, D., & Wiegratz, K. (2022). Machine Learning Methods in Finance: Recent Applications and Prospects. *SSRN Electronic Journal*.

[22] Schulz, H. & Behnke, S. (2012). Deep Learning. Künstliche Intelligenz: 26(4), 357-363.

[23] Li, J., Xu, C., Feng, B., & Zhao, H. (2023). Credit Risk Prediction Model for Listed Companies Based on CNN-LSTM and Attention Mechanism. *Electronics*, *12*(7), 1643.

[24] Wei, L., Yu, H., & Li, B. (2023). Energy financial risk early warning model based on Bayesian network. *Energy Reports*, *9*, 2300–2309.

[25] Halim, Z., Shuhidan, S. M., & Sanusi, Z. M. (2021). Corporation financial distress prediction with deep learning: analysis of public listed companies in Malaysia. *Business Process Management Journal*, *27*(4), 1163–1178.

[26] Chen, D. (2021). Risk Assessment of Government Debt Based on Machine Learning Algorithm. *Complexity*, *2021(7)*, 1–12.

[27] Zhang, G. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, *50*, 159–175.

[28] Gong, Y., Ai, H., Gao, Z., & Wang, M. (2022, August 17). Optimization of Local Prediction Algorithm of Financial Time Series Based on Fuzzy Neural Network. *Mobile Information Systems*, *2022*, 1–9.

[29] Wu, D., Ma, X., & Olson, D. L. (2022). Financial distress prediction using integrated Z-score and multilayer perceptron neural networks. *Decision Support Systems*, *159*, 113814.

[30] Ayvaz, E., Kaplan, K., & Kuncan, M. (2020). An Integrated LSTM Neural Networks Approach to Sustainable Balanced Scorecard-Based Early Warning System. *IEEE Access*, *8*, 37958–37966.

.[31] Li, W. (2022). Design of Financial Crisis Early Warning Model Based on PSO-SVM Algorithm. *Mathematical Problems in Engineering*, *2022*, 1–8.

[32] Hong, S., Wu, H., Xu, X., & Xiong, W. (2022). Early Warning of Enterprise Financial Risk Based on Decision Tree Algorithm. *Computational Intelligence and Neuroscience*, *14(7)*, 1–9.

[33] Wang, Z. (2022). A Study on Early Warning of Financial Indicators of Listed Companies Based on Random Forest. *Discrete Dynamics in Nature and Society*, *2022*, 1–12.

[34] Li, W., Paraschiv, F., & Sermpinis, G. (2022). A data-driven explainable case-based reasoning approach for financial risk detection. *Quantitative Finance*, *22*(12), 2257–2274.

[35] Zhang, Z., Wu, C., Qu, S., & Chen, X. (2022,). An explainable artificial intelligence approach for financial distress prediction. *Information Processing & Management*, *59*(4), 102988.
[36] Wei, L., Li, G., Zhu, X., & Li, J. (2019). Discovering bank risk factors from financial statements based on a new semi-supervised text mining algorithm. *Accounting & Finance*, *59*(3), 1519–1552.
[37] Li, S., Shi, W., Wang, J., & Zhou, H. (2021). A Deep Learning-Based Approach to Constructing a Domain Sentiment Lexicon: a Case Study in Financial Distress Prediction. *Information Processing & Management*, *58*(5), 102673.
[38] Jiang, C., Lyu, X., Yuan, Y., Wang, Z., & Ding, Y. (2022). Mining semantic features in current reports for financial distress prediction: Empirical evidence from unlisted public firms in China. *International Journal of Forecasting*, *38*(3), 1086–1099.
[39] Hashim A. (2014). A handbook on financial management I formation systems for the government. Working Paper No.100509. World Bank. Washington, D.C.
[40] OCED, (2019) A data-driven public sector Enabling the strategic use of data for productive, inclusive, and trustworthy governance. Working Papers No. 33. OECD.

.