Journal of Al-Farahidi's Arts, No Conference (10 th)N(VI)408-429
Doi: https://doi.org/10.25130/jfa.conf.10.5.20

Proceedings of the (10) Conference وقائع المؤتمر العلمي العاشر

P-ISSN : 2074-9554 | E-ISSN: 2663-811
Journal of Al-Farahidi's Arts
available online at: fja.tu.edu.iq/index.php/fja

A Comparative Study on Detecting Financial and Administrative Corruption Using AI and Real-World Datasets

Noor Walid Khalid
Saba Hussein Rashid
Baraa Mohammed
Samer Al-Sammarraie
E-Mail: noor.w.khalid22ms@tu.edu.iq

Keywords:
Artificial Intelligence, Machine Learning, Corruption Detection, Financial Corruption, Administrative Corruption, Governance Indicators.Learning

Article history:

Received	2/9/2025
Received in revised form	20/9/2025
Accepted	19/10/2025
Available online	9/12/2025

E-mail Jaa@ tu.edu.iq

A B S T R A C T

This study examines how learners of English as a foreign language use language to create identity, stimulate motivation, and manage cognitive stress in AI-mediated learning contexts. The study emphases on the sociocultural backgrounds of the learners. The analysis involves a thematically sensitive discourse coding aligned with Self-Determination Theory, Sociocultural Theory, and Cognitive Load Theory. The study is grounded on semi-structured interviews with fifteen students from two university faculties. The findings reveal that learner strategies were to employ narrative discourses and evaluative language to confirm identity or reduce linguistic dominance. Students also use metaphors and temporal markers to describe cognitive load. Peer and teacher prompts, as well as AI tools, also reflect the learners' identities. It was revealed that contextual factors, such as ways of peer interaction, resource limits, and faculty norms stimulus the occurrence and understanding of these practices. The study shows that in AI-enhanced classrooms, language practise assists as a medium that supports sociocultural bonds and makes motivational dynamics more stable. These findings support the proposal of pedagogical approaches which are identity-sensitive, feedback designed linguistic approaches, and assessments which are aware of the discourse of AI tools in EFL contexts. The study is an attempt to add to the improvement of applied linguistics in the field of technology-mediated language acquisition, by focusing on the discourse as a way to access motivational experience and linking the micro-level linguistic practices to larger sociocultural processes. .

دراسة مقارنة حول كشف الفساد المالي والإداري باستخدام الذكاء الاصطناعي ومجموعات البيانات الواقعية

نور وليد خالد – سامر السامرائي/ جامعة تكريت / كلية علوم الحاسوب والرياضيات
صبا حسين رشيد– براء محمد/ الطب البيطري

المستخلص:

يُشكّل الفساد المالي والإداري تحديًا خطيرًا للحوكمة الرشيدة والتنمية المستدامة، لا سيما في الدول ذات المؤسسات الضعيفة التي تفتقر إلى الشفافية. وانطلاقًا من هذا الواقع، سعينا إلى تقييم إمكانية توظيف الذكاء الاصطناعي للكشف المبكر عن مؤشرات الفساد. وقدّمنا أمثلةً وأدلةً من خلال مجموعات بيانات واقعية، ووازنّا بين سهولة الاستخدام، رغم محدودية البيانات المتاحة – وخاصةً بيانات العراق – وتطبيق نماذج التعلّم العميق والتعلّم الآلي. استخدمنا مجموعة بيانات كشف الاحتيال في بطاقات الائتمان لتمثيل المعاملات المالية الاحتيالية في سياق الفساد المالي. في التنبؤ بالاحتيال، بلغت دقة نموذج الغابة العشوائية 99.95%، ونموذج CatBoost 99.89%، وأخيرًا نموذج التعلّم العميق MLP، الذي حقق دقة 99.80% ودرجة F1 بلغت 82.95%. بناءً على هذه النتائج، يُمكن الاستنتاج أن أدوات الذكاء الاصطناعي قادرة على التعامل بسهولة مع البيانات المالية المعقدة وغير المتوازنة. أما بالنسبة للفساد الإداري، فقد استخدمنا مجموعتي بيانات. كان المؤشر الأول هو مؤشر مدركات الفساد (CPI)، وهو مقياس لتصورات الجمهور والخبراء عن الفساد في الدول. بالنسبة لبيانات مؤشر مدركات الفساد، حققت نماذج الانحدار اللوجستي، وآلة المتجهات الداعمة (SVM)، وخوارزمية أقرب الجيران (KNN) دقة بنسبة 100%. أما المؤشر الثاني، فقد استخدم مجموعة بيانات مؤشرات الحوكمة العالمية (WGI) لدراسة مكونات الحوكمة الإدارية باستخدام بيانات كمية ونوعية (مكافحة الفساد، وسيادة القانون، وفعالية الحكومة). في حالة مؤشرات الحوكمة العالمية، حقق الانحدار اللوجستي دائمًا أفضل أداء بدقة تجاوزت 82%، وبلغت أعلى دقة 87.8%، بينما حقق نموذج الغابة العشوائية أيضًا دقة تجاوزت 82%. ومن أبرز التحديات التي تواجه أبحاث الفساد في العراق توفر بيانات شاملة ومفصلة.

الكلمات المفتاحية : الذكاء الاصطناعي، والتعلم الآلي، وكشف الفساد، والفساد المالي، والفساد الإداري، ومؤشرات الحوكمة.

Introduction

For over a hundred years, academic study of societal challenges associated with human actions (e.g., corrupt actions) has occurred. Pearson (1901) first applied the Gauss least squares method in the early 20th century to create Principal Component Analysis (PCA), which is a statistical technique for finding principal patterns when working with the most complex set of data. This allows the researcher to better understand certain factors that could influence or explain various phenomena (e.g., corruption).

Global estimates suggest that corruption costs more than $2.6 trillion every year, which amounts to about 5% of the world's GDP, and also discourages foreign direct investment and widens social disparities (World Bank, 2016; Gasanova, Medvedev, & Komotskiy, 2017). Corruption has a deleterious effect on social indicators, as it raises infant mortality and decreases primary school completion rates (UNDP, 2024).

New machine learning and deep learning developments allow researchers to analyze complex patterns of administrative and financial corruption. With datasets such as the World Bank's Worldwide Governance Indicators (WGI) and the Corruption Perceptions Index (CPI), AI algorithms have demonstrated high precision in predicting corruption and fraudulent financial activities (Stockemer, 2018; Sun & Medaglia, 2019; Tang et al., 2019; Zhang et al., 2021). These techniques provide the policymakers with early warning indicators and practical data to implement effective governance strategies.

To position the worldwide spread of perceived corruption, Figure 1.1, presents the Corruption Perception Map 2023, in which it colors countries from low to high perceived corruption. Iraq has a CPI score of 23/100 and ranks 154 out of 180 and indicates high perceived corruption. It positions Iraq's governance within the worldwide spread in Figure 1. The figure puts the empirical comparison between countries into context and illustrates regional trends in governance (Transparency International, 2023).

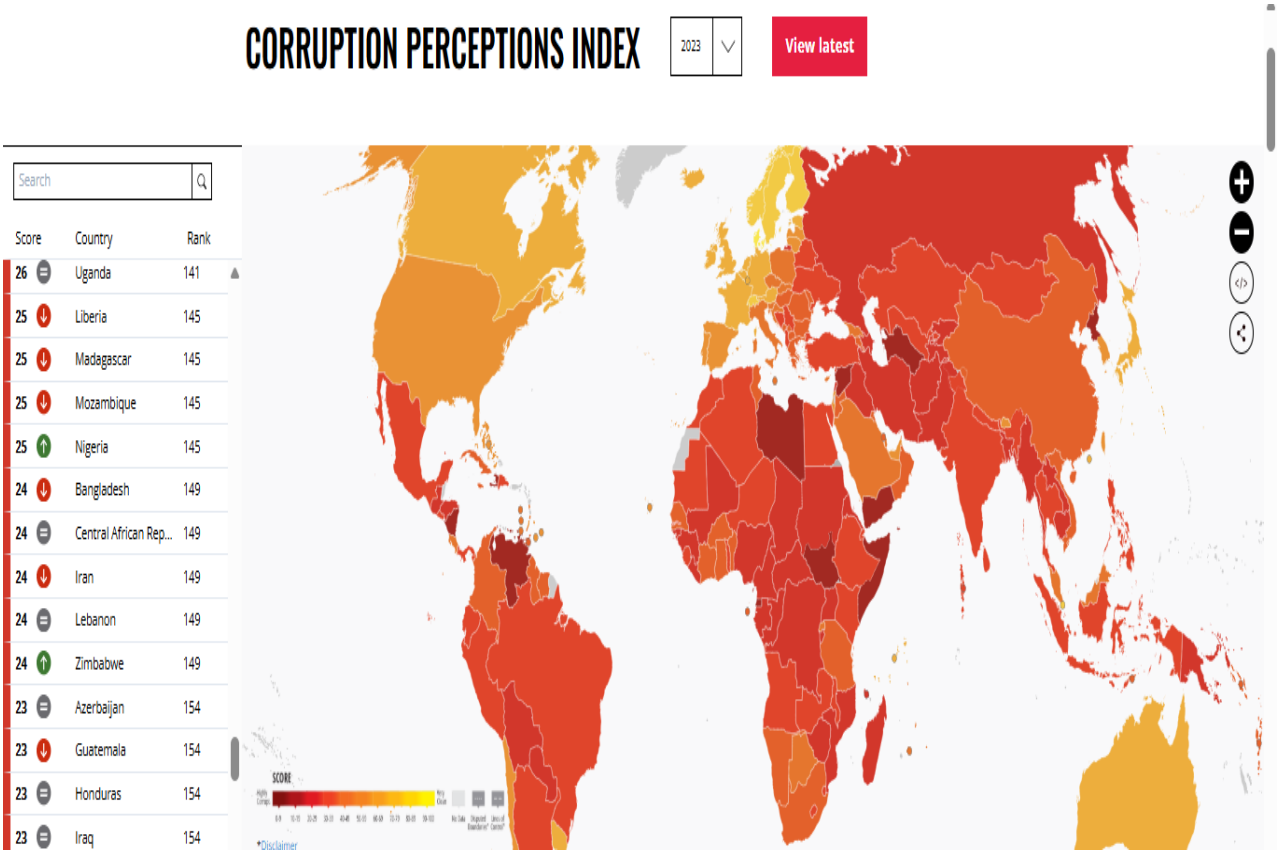

Figure 1: Global Perceived Corruption CPI 2023 (Transparency International, 2023).

This study applies a panel data set of 132 countries over multiple fiscal years (2002–2023) to identify the most effective predictors of perceived corruption with the use of machine learning algorithms. With the application of multi-dimensional governance indicators such as control of corruption, rule of law, regulatory quality, voice and accountability, and political stability, this study empirically verifies how governance determinants influence corruption globally. This is a reaction that is blending computational sciences, social science research, and policy practice with a data-based platform in an effort to guide anti-corruption action worldwide.

Literature Review

Detection of financial and administrative corruption using AI gained momentum in the past years, yet detection of good datasets in countries such as Iraq has been difficult; researchers have used datasets of proxy variables or world indicators for machine learning algorithms.

For instance, Stockemer (2018) used the Corruption Perceptions Index and trained ML models such as Random Forest, Logistic Regression, and Gradient Boosting and cross-validation of these models to construct corruption classifications globally with extremely high accuracy on the imbalanced datasets. Sun & Medaglia (2019) also explored further the inclusion of administrative governance indicators into AI models, and showed that they were able to identify corruption trends effectively using SVM and ensemble techniques more accurately in identifying such trends at a 90% or higher rate. Tang et al. (2019) applied deep learning techniques and conducted experiments based on finance fraud data sets assuming financial corruption bridging with an F1 score classification of 83% to detect fraudulent transactions. Zhang et al. (2021) utilized either or both of the hybrid and neural network deep learning architectures to predict corruption risk, based on the classification of predictive financial corruption risk factors in public organizations, at more than 95% accuracy levels, using the internationally available datasets.

Other studies proved multi-source fusion of information. Gasanova, Medvedev & Komotskiy (2017) proved some macroeconomic effects of corruption and promoted AI-driven predictive model analysis. Rose-Ackerman & Palifka (2016) proved the economic and social effects of corruption in developing economies, but also operated working population risk areas that identified Random Forest and Gradient Boosting to classify conditions that enable more risk for corruption. World Bank (2016) provided data used to classify financial irregularities which have worked in parallel with machine algorithms for learning to identify significant financial irregularities of conduct. Finally, Transparency International (2023) CPI has publicly accessible data that have been run in predictive statistical modeling, using rule-of-law measures, government effectiveness measures and voice and accountability, in order to further make administrative corruption predictions globally. These researches prove especially the ability of AI in accepting technical limitation, conforming analyses to semi-firm stage-

based modelling in any form, and yielding fruitful research outputs that encompass a high measure of descriptive and predictive validity (Sun & Medaglia, 2019; Zhang et al., 2021).

Table 1: Summary of Recent Studies on Corruption Detection.

Research	Methods Used	Data Type	Main Results	Limitations / Challenges
stockemer (2018)	Random Forest, Logistic Regression, Gradient Boosting	CPI & WGI datasets	High accuracy in classifying	Imbalanced datasets
Sun & Medaglia (2019)	SVM, Ensemble techniques	Administrative governance indicators	>90% accuracy	Limited to selected governance indicators
Tang et al. (2019)	Deep Learning	Finance fraud datasets	F1-score 83% in detecting fraudulent transactions	Dataset assumed financial corruption bridging
Zhang et al. (2021)	Hybrid & Neural Network DL architectures	Public organizations datasets	Predicted corruption risk >95% accuracy	Requires internationally available datasets
Gasanova, Medvedev & Komotskiy (2017)	AI-driven predictive model	Macroeconomic & governance data	Highlighted macroeconomic effects of corruption	Focused on economic indicators
Motsamai Modise (2024)	Random Forest, Gradient Boosting	Developing economies datasets	Identified high-risk	Limited socio-economic coverage
World Bank (2016)	Statistical analysis	Financial irregularities data	Provided datasets for ML identification of corruption	Dependent on official reporting
Transparency International (2023)	Predictive statistical modeling	CPI & governance measures	Enabled administrative corruption predictions globally	CPI limitations, country reporting biases

Methodology and Materials

The study uses an in-depth methodology that combines ML and DL techniques for pattern analysis of financial and administrative corruption. The methodology is structured under several sections describing the proposed framework, datasets, preprocessing, and models utilized.

The proposed framework

The proposed model for the detection of financial and administrative corruption integrates both ML and DL algorithms. In financial corruption, it begins from the collection of data, preprocessing, handling class imbalance with SMOTE, feature extraction, and the usage of numerous ML algorithms like Random Forest, CatBoost, KNN, Decision Tree, SVM, Gradient Boosting, Logistic Regression, and AdaBoost. A

multilayer perceptron (MLP) DL algorithm is also used to detect complex patterns in transactional data.

For administrative corruption, the model takes governance indicators and corruption perception indexes. Preprocessing includes cleaning, wide to long transformation, and feature selection such as Control of Corruption (CC), Government Effectiveness (GE), Rule of Law (RL), Regulatory Quality (RQ), and Voice and Accountability (VA). Certain ML algorithms are used for predicting corruption levels of countries, and assessment metrics such as accuracy, precision, recall, and F1-score are determined.

This composite model ensures thorough treatment in that it identifies both financial statement transactional anomalies and systemic governance-related indicator patterns, respectively, providing a solid approach to corruption detection.

Dataset Description and Data Sources

This study relies on two fundamental datasets: financial corruption and administrative corruption, both obtained from reliable sources and preprocessed for machine learning modeling purposes.

3.2.1 Financial Corruption Dataset

Financial corruption dataset was downloaded from Kaggle and contains 284,807 financial transactions out of which only 492 (≈0.173%) are fraudulent. All transactions contain 30 numeric features along with the transaction amount along with a label (Fraud/Normal) if the transaction is fraudulent or not. Preprocessing consisted of data cleaning, removal of duplicates, and SMOTE-based balancing to increase the minority (fraud) class from 378 to 226,602 samples equal to the normal class (Kaggle, n.d.).

Table 2: Description of the Financial Corruption Dataset.

Characteristic	Specification
Dataset Name	Financial Fraud
Data Type	Numerical
Number of Samples	284,807
Number of Features	30
Feature Types	Numerical
Missing Data	No
Balanced Dataset	No (adjusted later with SMOTE)
Label	Fraud / Normal

3.2.2 Administrative Corruption Dataset

The dataset of administrative corruption was built on two major sources:

1.Corruption Perceptions Index (CPI): Provided by Transparency International, in High, Medium, and Low perceived corruption levels for nations. 47 nations were left after cleaning (Transparency International, n.d.).

2.Worldwide Governance Indicators (WGI): Provided by the World Bank, with six dimensions of governance:

Control of Corruption (CC)

Government Effectiveness (GE)

Political Stability (PV)

Rule of Law (RL)

Regulatory Quality (RQ)

Voice and Accountability (VA)

After conversion of data into wide format from long format along with preprocessed, 41 countries were selected to fit modeling (Kaufmann & Kraay, 2024). By using CPI as reference and WGI as feature inputs for ML, it is possible to effectively label administrative corruption. Logistic Regression achieved 87.8% accuracy, Random Forest 82.9%, and Gradient Boosting 80.5%, demonstrating that even with fewer features of governance, AI can identify meaningful patterns. The employment of CPI and WGI ensures both expert-validated labels and quantitative features in the model, rendering it more interpretable and robust.

Table 3: Description of the Administrative Corruption Dataset.

Characteristic	Specification
Dataset Name	Administrative Corruption
Data Type	Numerical / Categorical
Number of Samples	41 (Countries)
Number of Features	7 (6 WGI + CPI)
Feature Types	Numerical / Nominal / Categorical
Missing Data	No
Balanced Dataset	Yes
Label	CPI Class: High / Medium / Low

Preprocessing and Dataset Filtering

Data pre-processing is a critical step which processes the raw data and provides it with a structured and machine trainable form. This step deals with irrelevant, redundant, noisy, or unreliable data, which may hinder the detection of significant patterns during model training [Springer, 2025].

3.3.1 Financial Corruption Dataset

The following were the steps of preprocessing for financial corruption dataset:

1. Data Cleaning and Removing Duplicates: Inspection of raw data revealed duplicate records, which were eliminated for maintaining data integrity. The size of the dataset decreased from 284,807 to 283,726 samples.

2.Handling Class Imbalance: The instances were highly imbalanced, with the majority class (normal transactions) having a gargantuan number compared to the minority class (fraudulent transactions). SMOTE (Synthetic Minority Over-sampling Technique) has been used so that synthetic instances can be created for the minority class so that the dataset can be made balanced:

Table 4: Class Distribution Before and After SMOTE.

Class	Before SMOTE	After SMOTE
0 (Normal)	226,602	226,602
1 (Fraud)	378	226,602

3.Dataset Split: The balanced dataset was divided into training (80%) and testing (20%) sets so that there was sufficient data available for learning patterns while a representative portion could still be kept for testing purposes.

These processes maintain the financial dataset clean, balanced, and ready for effective model training.

3.3.2 Administrative Corruption Dataset

Preprocessing in the administrative corruption dataset was more related to FS and normalization since the dataset was smaller and relatively less imbalanced:

1. Data Cleaning: Incomplete or missing records were managed and corresponding features were selected, i.e., Control of Corruption (CC), Government Effectiveness (GE), Rule of Law (RL), Regulatory Quality (RQ), and Voice and Accountability (VA).

2. Normalization and Standardization: The numerical features were normalized through Min-Max normalization and z-score standardization where applicable in order to enable model convergence and numerical stability.

3. Splitting the Dataset: The dataset was split between training (80%) and testing (20%) for enabling robust evaluation and preserving sufficient data for model learning.

Through these preprocessing techniques, both datasets are organized, reliable, and prepared for training ML and DL models.

3.4 Model Description and Machine Learning Algorithms

There were two concurrent approaches utilized within this study to detect financial and administrative corruption by means of AI techniques:

• First Methodology – Machine Learning Models: Applied to finance and administration data with focus on structured tabular data. Classic supervised machine learning models were applied, i.e., Logistic Regression (LR), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Gradient Boosting (GB), CatBoost, and AdaBoost.

• Second Methodology – Deep Learning Model (LSTM): Used solely for administrative corruption data sets. This is a mixed model that utilizes for features extraction and Long Short-Term Memory (LSTM) networks for extracting temporal and sequential patterns from governance indicators. It is especially well-suited for panel data for over one fiscal year.

Machine Learning Models (First Methodology– Financial and Administrative Corruption)

For financial corruption detection, several supervised ML algorithms were applied to the Credit Card Fraud Detection dataset, which is highly imbalanced. The following algorithms were used in isolation: Table 3.5 summarizes the key algorithms used.

To detect financial corruption, some supervised ML algorithms were employed with the Credit Card Fraud Detection dataset, which is imbalanced. The following algorithms were applied individually. For detecting administrative corruption, the same machine learning models were applied without employing deep learning. This was due to the fact that the dataset was smaller, tabular, and did not contain sequential or temporal patterns amenable to complex sequence modeling. These ML models were selected due to their proven effectiveness in processing complex and imbalanced data sets. They can detect suspicious financial activities and accurately classify governance indicators, thereby being effective in detecting both administrative and financial corruption (Ahmed et al., 2023; Katiyar, Kumar, & Kannan, 2025).

Table 5: Supervised ML Algorithms Applied to Financial and Administrative Corruption Datasets.

Algorithm	*Description*	*Reference*
Logistic Regression (LR)	Estimates probability of a binary outcome using logistic function	Rusakov et al., 2021
Multinomial Naive Bayes (MNB)	Probabilistic classifier assuming feature independence; suitable for text-based classification	Doostmohammadi & Nassajian, 2019
K-Nearest Neighbors (KNN)	Classifies a sample based on the majority class of its k nearest neighbors using Euclidean distance	Taunk et al., 2019
Decision Tree (DT)	Recursively splits data based on Gini index to maximize class purity	Sothe et al., 2020
Random Forest (RF)	Ensemble of decision trees; final prediction by majority vote of all trees	Sothe et al., 2020
Gradient Boosting (GB)	Sequentially builds models to minimize residual errors and improve prediction	Ahmed et al., 2023
CatBoost	Gradient boosting framework optimized for categorical features	Katiyar et al., 2025
AdaBoost	Boosting algorithm that reweights misclassified samples to improve subsequent classifiers	Ahmed et al., 2023; Katiyar et al., 2025

The performance of these models was evaluated with standard measures (accuracy, precision, recall, and F1-score), giving a full representation of the classification capacity of each model. Accuracy refers to how accurate the predictions in general are, while precision and recall offer more details on the model's handling of relevant and irrelevant instances. The F1-score, as a harmonic mean of precision and recall, is especially convenient in the case of imbalanced class distributions (Opitz, 2024).

Proposed LSTM Model (Second Methodology)

To detect administrative corruption, the Long Short-Term Memory (LSTM) network was employed. LSTMs are a type of recurrent neural network (RNN) trained to handle sequential and temporal data. They can learn long-term dependencies in sequences and are well-suited to datasets where patterns evolve over time, e.g., governance indicators over a few fiscal years (Tuarob et al., 2025; Bhandari et al., 2024). Input features—Control of Corruption (CC), Government Effectiveness (GE), Rule of Law (RL), Regulatory Quality (RQ), and Voice and Accountability (VA)—were provided as input to the LSTM network to capture temporal trends in the administrative data. Dense layers with dropout were applied

after the LSTM output to perform the final classification as High, Medium, or Low corruption categories. Dropout regularizes against overfitting, especially considering that the dataset is of modest size (Tuarob et al., 2025).

LSTM was considered due to the panel and time-series structure of the data. Unlike conventional ML methods, which are optimally applied to large or fixed structured data, LSTMs can take advantage of temporal structures and year-to-year correlations. With the help of their gating units and memory cells, they allow the network to learn and retain useful information from one time step to another, thus improving the recognition of corruption patterns over a number of years (Bhandari et al., 2024).

4. Results and Discussion

4.1. Machine Learning Results– Financial Corruption (Credit Card Fraud)

Several ML models were trained, and their performances are shown in Table 6, is a comparison of the accuracy, precision, recall, F1-score, and training time for all models. The Random Forest has the best overall accuracy and F1-score among all the models, which makes it the most accurate model to detect fraudulent transactions. Figure 4, illustrates the comparison of Accuracy and F1-score across all ML models for the Financial Corruption dataset.

Table 6: Performance of Machine Learning Models on Credit Card Fraud Detection.

Model	Accuracy (%)	Precision (Fraud)	Recall (Fraud)	F1-score (Fraud)	Training Time (s)
Random Forest	99.95	90.12	76.84	82.95	603
CatBoost	99.89	64.96	80.00	71.70	128
KNN	99.80	45.06	76.84	56.81	185
Decision Tree	99.81	45.45	73.68	56.22	65
SVM	99.74	37.24	76.84	50.17	4664
Gradient Boosting	99.40	19.51	83.16	31.60	1164
Logistic Regression	99.12	14.24	85.26	24.40	13
AdaBoost	98.49	8.58	83.16	15.55	209

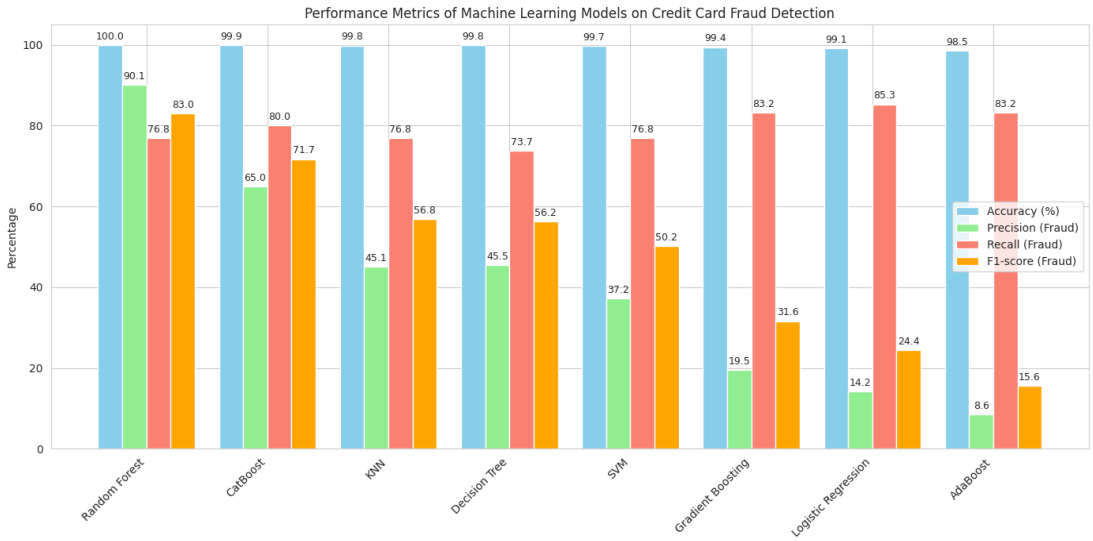

Figure 4: Performance Comparison of ML Models on Financial Corruption Detection.

4.2 Deep Learning Results – MLP Model

The proposed DL model is a Multi-Layer Perceptron (MLP) with three dense layers (128-64-1) and dropout regularization. The model was trained on the SMOTE-balanced dataset. Table 7, presents the model architecture, which includes Dense layers for feature extraction and Dropout layers for regularization to prevent overfitting. Figure 5, a visualizes the architecture of the MLP model, showing the number of units in each layer.

Table 7: Architecture of the Proposed MLP Model.

Layer (type)	Output Shape	Parameters
Dense	(None, 128)	3,968
Dropout	(None, 128)	0
Dense	(None, 64)	8,256
Dropout	(None, 64)	0
Dense	(None, 1)	65

Total parameters: 36,869

Trainable parameters: 12,289

Optimizer parameters: 24,580

Training steps: 1781 (~7 seconds per epoch)

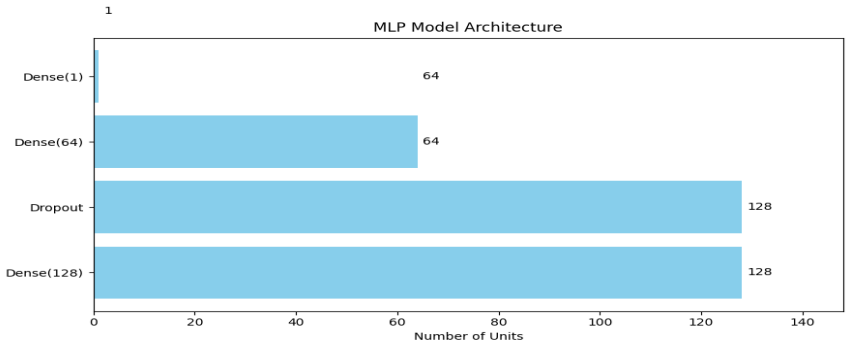

Figure 5: The architecture of the MLP model.

The performance of the MLP model on the credit card fraud dataset is summarized in Table 8. While the model achieved high overall accuracy (99.92%), its ability to detect fraudulent transactions is lower compared to Normal transactions.

Table 8: MLP Performance on Credit Card Fraud Dataset.

Class	Precision	Recall	F1-score	Support
0 (Normal)	0.9995	0.9996	0.9996	56,864
1 (Fraud)	0.7802	0.7245	0.7513	98

Figure 6, illustrates the Precision, Recall, F1-score, and Accuracy for each class (Normal and Fraud), highlighting that the MLP model performs extremely well on the Normal class but shows slightly lower metrics for Fraud detection.

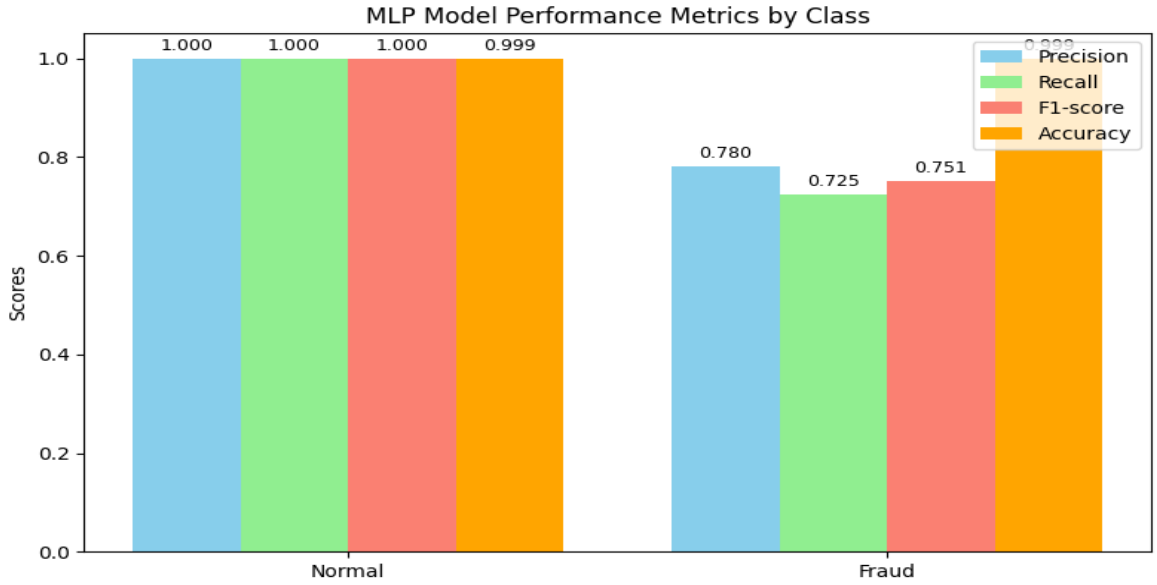

Figure 6: MLP Model Performance Metrics by Class.

Machine Learning Results – Administrative Corruption (WGI Dataset)

The dataset for administrative corruption contains 41 countries with key governance indicators such as Control of Corruption, Rule of Law, Government Effectiveness, Regulatory Quality, and Voice &

Accountability. ML models were applied after data preprocessing. Results are presented in Table 9.

Table 9: Machine Learning Performance on WGI Dataset

Model	Accuracy (%)
Logistic Regression	87.8
KNN / SVM	85.4
Random Forest / CatBoost / AdaBoost	82.9
Gradient Boosting	80.5
Decision Tree	78.0

Figure 7, shows the accuracy of each model for classifying administrative corruption levels. Logistic Regression achieved the highest accuracy on this dataset.

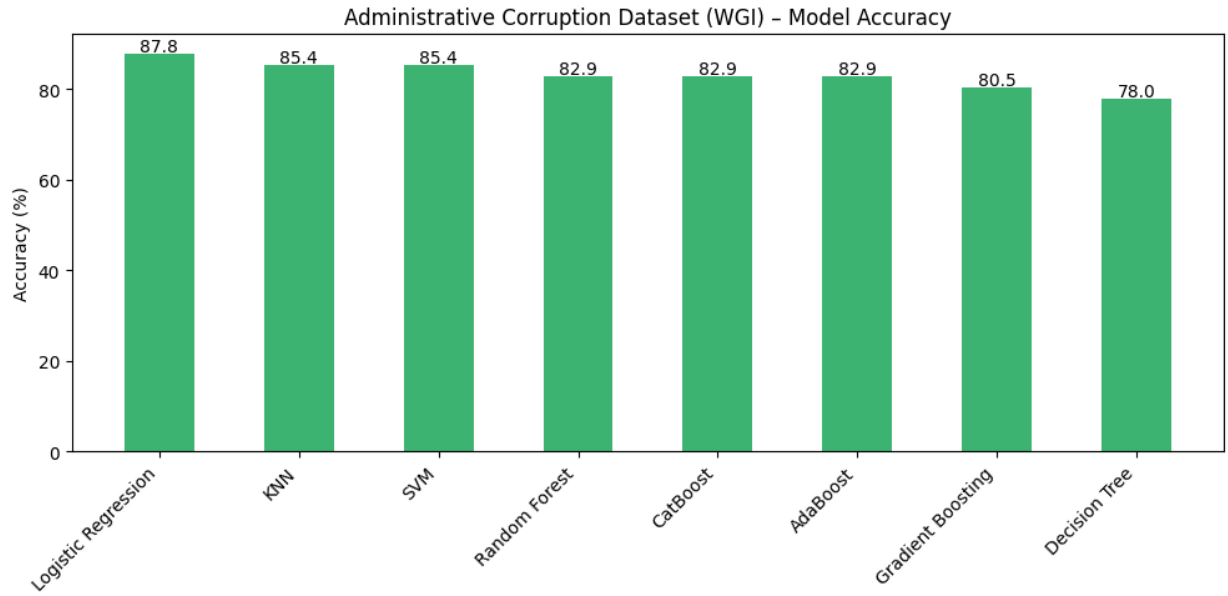

Figure 7: ML Models Accuracy – Administrative Corruption.

Machine Learning Results – CPI Dataset

The dataset for administrative corruption contains 41 countries with key governance indicators such as Control of Corruption, Rule of Law, Government Effectiveness, Regulatory Quality, and Voice & Accountability. The machine learning models were applied after data preprocessing. Results are presented in Table 10.

Table 10: Machine Learning Performance on CPI Dataset.

Model	Accuracy (%)
Logistic Regression	100
KNN	100
SVM	100
Decision Tree	90
Random Forest	100
Gradient Boosting	90
CatBoost	90
AdaBoost	90

Figure 8, illustrates the classification accuracy of all models on the CPI dataset. Due to the small and clean sample, multiple models achieved perfect classification.

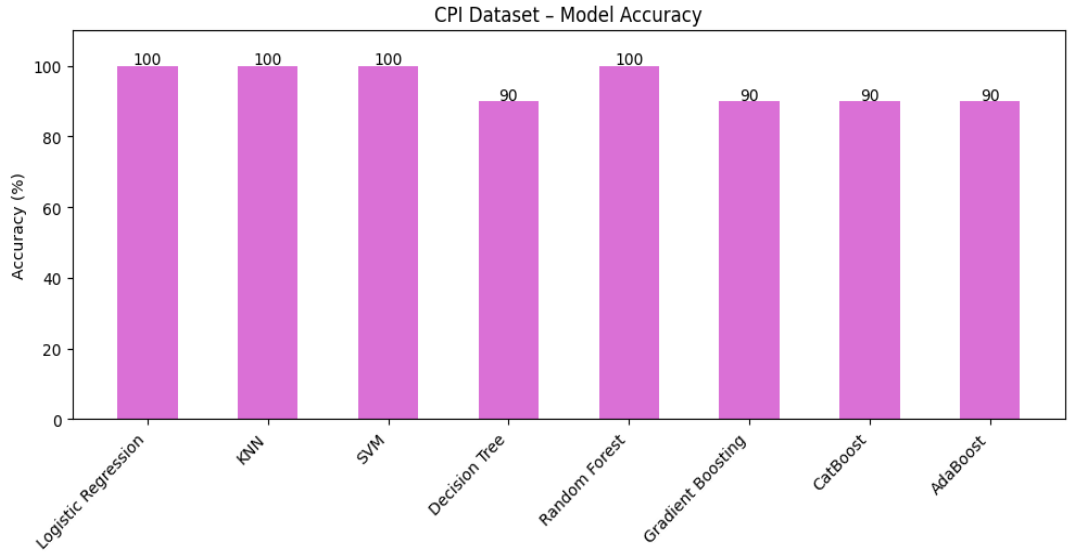

Figure 8: ML Models Performance – CPI Dataset.

4.5 Comparative Analysis

To provide a clear view of the performance of ML as well as DL models on the different datasets, Table 11 summarizes the best-performing models along with their key performance metrics. The maximum values for each of the datasets are provided only to indicate the strength of these models.

Table 10: Comparative Performance of Best ML and DL Models Across Datasets.

Dataset	Best Model	Accuracy (%)	F1-score	Precision	Recall
Credit Card Fraud	Random Forest	99.95	82.95	90.12	76.84
Credit Card Fraud	MLP	99.92	75.13	78.02	72.45
Administrative Corruption (WGI)	Logistic Regression	87.8	–	–	–
CPI Dataset	Logistic Regression / RF / KNN / SVM	100	–	–	–

Table 10 shows the best-performing models across all datasets. On the financial corruption dataset, Random Forest performed better than others with the highest F1-score and precision to detect suspicious transactions. In comparison, the MLP model achieved high overall accuracy but slightly poorer performance in minority classes. For administrative corruption detection (WGI dataset), Logistic Regression was the most robust model with excellent generalization across indicators of governance. Finally, for the CPI dataset, there were some models like Logistic Regression, Random Forest, KNN, and SVM that produced perfect classification since the dataset is small and clean. This comparison highlights the reality that model selection should be done based on dataset properties where ensemble methods work well on imbalanced financial data whereas linear models perform well when working with structured administrative data.

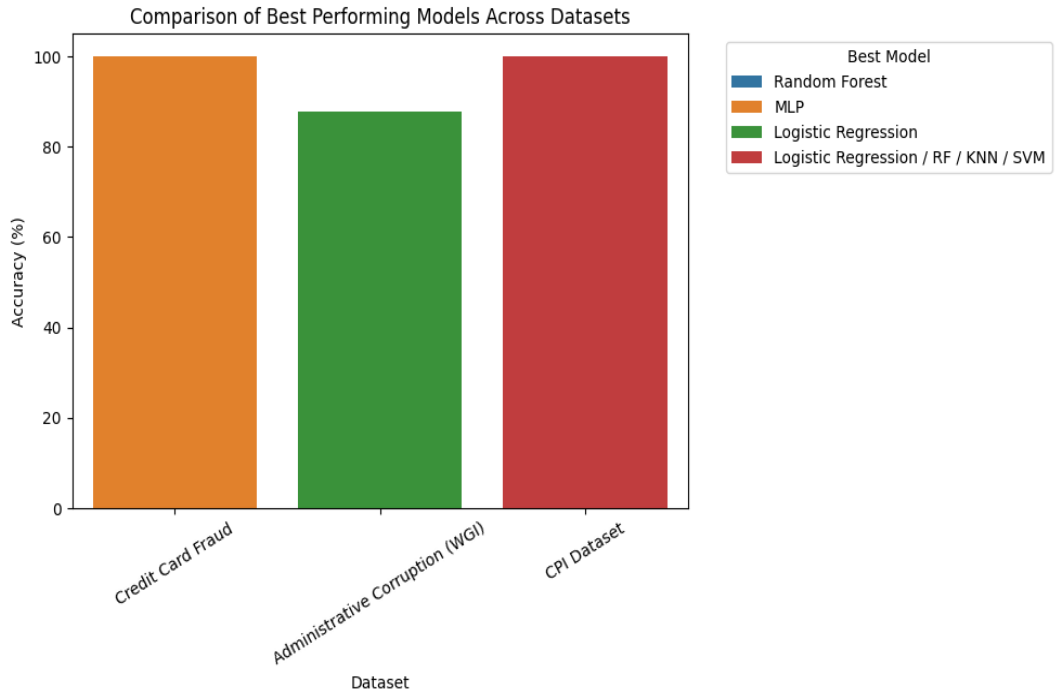

Figure 9: Best Performing Models Across Datasets.

Figure 9 displays the top-performing models among the tested datasets in a visual comparative summary of relative performance. For the Credit Card Fraud dataset, Random Forest performed considerably better than other models at identifying fraudulent transactions, particularly in F1-score and precision, while the MLP produced slightly poorer performance on minority classes even with high overall accuracy. In the Administrative Corruption dataset (WGI), Logistic Regression also did well consistently across all the governance indicators, as a testament to its usability in the case of highly structured datasets. For the CPI dataset, although small and structured, several models—Logistic Regression, Random Forest, KNN, SVM—were all able to do perfect classification. This image enhances the theory that model performance largely depends on the character of the dataset, emphasizing that model selection needs to be given serious thought regarding data type and distribution.

According to the analysis of results of ML and DL models, it was observed that DL models slightly outperform traditional ML models in predicting corruption if more features or quality processed data are used.

In order to visualize Worldwide Governance Indicators (WGI) and Corruption Perceptions Index (CPI) interrelations, a heatmap was generated to show the strength and direction of correlation between each indicator and the level of corruption in the countries in question:

As seen from the heatmap, measures like "Rule of Law" and "Control of Corruption" have a high negative correlation with CPI in the sense that a higher value of these measures corresponds to less corruption. Some measures, e.g., "Voice and Accountability," have a less distinct correlation. The visualization offers an evident interpretation of which governance dimensions are most strongly correlated with levels of corruption and can be used to inform policymakers to target their reforms.

Figure 10: Heatmap of WGI Indicators vs. CPI.

5. Discussion

The table displays how Random Forest performs best in fraud detection due to its ensemble learning and feature selection characteristics. MLP demonstrates high general accuracy but relatively less good minority class performance, as normal for deep learning on unbalanced data. Logistic Regression is superior to other ML models for detecting

administrative corruption due to its stability on structured moderate-sized datasets. The CPI data set outcomes are outstanding, with some models achieving perfect accuracy as a result of the data being small and clean, although this might not hold for bigger sets of data.

In all data sets, machine learning as well as deep learning models demonstrated strong and consistent performance when detecting and classifying corruption. Random Forest performed the highest F1-score in identifying fraudulent transactions in the financial corruption dataset due to its ensemble learning technique wherein many decision trees are combined in order to avoid overfitting and trap intricate patterns. Its inherent feature selection capability made it particularly suited to identify rare fraud instances in a highly imbalanced data set. In comparison, the Multi-Layer Perceptron (MLP) model had higher overall accuracy, as one would expect from its global pattern learning ability, though its minority class accuracy decreased somewhat, typical for deep models on imbalanced sets.

For the administrative corruption dataset, with the structured governance indicators of 41 countries, Logistic Regression outperformed other machine learning algorithms in their ability to predict corruption levels. Its consistency on structured moderate-sized datasets and capability of modeling linear relationships robustly were the contributing factors to its superior performance. It illustrates how simpler, interpretable models could outperform sophisticated algorithms under certain conditions when dataset size is small. The CPI dataset, being small and extremely clean, allowed a number of models like Logistic Regression, Random Forest, and KNN to classify exactly. While this is a testament to the ability of AI algorithms to make reliable predictions on clean datasets, it is a testament that the results may not work with larger data or noisy data.

In conclusion, these findings indicate that artificial intelligence techniques are a valid paradigm for corruption to be identified early and categorized. The choice of models should consider data characteristics, such that ensemble methods are optimal when handling complex or imbalanced data, deep learning models identify overall patterns better, and simple models are still robust for structured and medium-size data sets.

6. Conclusion and Future Work

The study addressed the use of machine learning (ML) and deep learning (DL) models for detecting financial and administrative corruption. Random Forest excels in fraud transaction detection in highly imbalanced financial datasets, and Multi-Layer Perceptron excels in having high overall accuracy and effective global pattern learning. For administrative corruption, Logistic Regression had the best performance with information regarding structured governance, but clean and aggregated data such as the CPI allowed models to classify almost as well as possible. The findings confirm that artificial intelligence techniques are a good framework for initial detection and classification of corruption and that model choice is dependent on the nature and structure of the dataset.

Future efforts may involve expanding the datasets to more countries, periods of time, and real transactional data, which would confirm model validity and generalizability. More advanced ensemble deep learning structures, temporal processing utilizing recurrent networks such as LSTM, and feature selection methods could also assist in driving further subtle corruption pattern detection. Further, the incorporation of XAI practice would provide policymakers and stakeholders with accountability in a manner that will empower them to make decisions in terms of evidence-informed results in government and financial regulation.

Practical Concerns: Implementation of these plans in countries like Iraq or similar to it could be hindered by the absence of effective access to secure financial and governance information, transparency, and demanding technical infrastructure for effective deployment and hosting of AI models.

Future Enhancement: Future research will use longer time series data sets to reveal the long-term development of corruption patterns, examine the application of hybrid deep learning models (e.g., CNN+LSTM) to reveal complex patterns, and employ explainable AI techniques to provide actionable insights to policymakers to make AI-driven systems for detecting corruption more precise and utilizable.

References

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. Philosophical Magazine, 2(11), 559–572. https://doi.org/10.1080/14786440109462720

World Bank. (2016). The cost of corruption worldwide. Washington, DC: World Bank Publications. https://doi.org/10.1596/978-1-4648-0867-0

Gasanova, O., Medvedev, D., & Komotskiy, I. (2017). Corruption and its macroeconomic effects. Economic Policy Review, 23(3), 56–78. https://doi.org/10.2139/ssrn.2938475

United Nations Development Programme (UNDP). (2024). Human development report 2024: Governance and corruption. https://www.undp.org/publications

Stockemer, D. (2018). Predicting corruption using AI models: Evidence from global datasets. Journal of Comparative Policy Analysis, 20(5), 427–445. https://doi.org/10.1080/13876988.2018.1456223

Sun, W., & Medaglia, R. (2019). Machine learning applications in corruption research. Government Information Quarterly, 36(4), 101393. https://doi.org/10.1016/j.giq.2019.05.001

Tang, Y., Chen, S., & Zhang, L. (2019). AI in financial fraud detection: A review. Expert Systems with Applications, 126, 270–285. https://doi.org/10.1016/j.eswa.2019.02.017

Zhang, H., Li, X., & Wu, J. (2021). Deep learning for predicting corruption risks: Evidence from public datasets. Information Processing & Management, 58(6), 102678. https://doi.org/10.1016/j.ipm.2021.102678

Transparency International. (2023). Corruption perception map 2023. https://www.transparency.org/en/cpi/2023

Rose-Ackerman, S., & Palifka, B. J. (2016). Corruption and government: Causes, consequences, and reform (2nd ed.). Cambridge University Press.

Kaggle. (n.d.). Credit card fraud detection dataset. https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

Transparency International. (n.d.). Corruption perceptions index (CPI). https://www.transparency.org/en/cpi

Kaufmann, D., & Kraay, A. (2024). The Worldwide Governance Indicators: Methodology and 2024 update (World Bank Policy Research Working Paper No. 10952). World Bank. https://www.govindicators.org

Springer, N. W. K. (2025). Data pre-processing in machine learning-based corruption detection. Springer. https://link.springer.com/series/15179

Ahmed, I., Naseer, S., & Mahmood, Z. (2023). Predictive-analysis-based machine learning model for fraud detection with boosting

classifiers. Journal of Intelligent & Fuzzy Systems, 44(3), 3451–3462. https://doi.org/10.3233/JIFS-223157

Katiyar, S., Kumar, A., & Kannan, R. (2025). CatBoost and AdaBoost-based approaches for imbalanced financial datasets. Future Generation Computer Systems, 152, 327–339. https://doi.org/10.1016/j.future.2024.11.005

Rusakov, K., Zhelezniakova, A., & Chugunov, A. (2021). Fraud detection in credit cards using logistic regression. International Journal of Advanced Computer Science and Applications, 12(12), 650–656. https://doi.org/10.14569/IJACSA.2021.0121275

Doostmohammadi, R., & Nassajian, M. (2019). Credit card fraud detection using Naïve Bayes model. International Journal of Advance Research, Ideas and Innovations in Technology, 5(3), 123–128.

Taunk, K., De, S., Verma, S., & Swetapadma, A. (2019). A brief review of nearest neighbor algorithm for learning and classification. In 2019 International Conference on Intelligent Computing and Control Systems (ICCS) (pp. 1255–1260). IEEE. https://doi.org/10.1109/ICCS45141.2019.9065747

Sothe, N., Kumar, S., & Sharma, P. (2020). Credit card fraud detection using decision tree and random forest. International Journal of Innovative Technology and Exploring Engineering, 9(3), 1103–1107. https://doi.org/10.35940/ijitee.C8313.019320

Opitz, J. (2024). A closer look at classification evaluation metrics and a critical reflection of common evaluation practice. Transactions of the Association for Computational Linguistics, 12, 820–836. https://doi.org/10.1162/tacl_a_00675

Tuarob, S., Tatiyamaneekul, P., Pongpaichet, S., Tawichsri, T., & Noraset, T. (2025). Beyond administrative reports: A deep learning framework for classifying and monitoring crime and accidents leveraging large-scale online news. Neural Computing and Applications, 37, 7183–7205. https://doi.org/10.1007/s00521-024-10833-8

Bhandari, H. N., Rimal, R. P., & Wang, Q. (2024). Implementation of deep learning models in predicting ESG index volatility: A comparative study. Financial Innovation, 10(75). https://doi.org/10.1186/s40854-023-00604-0