Litrature survey

Date	28 April 2023
Team ID	NM2023TMID01937
Project Name	Project – Audit AI: A Machine Learning for Detecting Fraud in Audit Data

Customer Problem Statement Template

1	**Paper title**	MATIN N. ASHTIANI AND BIJAN RAAHEMI "Intelligent Fraud Detection in Financial Statements Using Machine Learning and Data Mining: A Systematic Literature Review"Date of publication 13 July 2021, date of current version 15 July 2022.Digital Object Identifier 10.1109/ACCESS.2021.3096799
	Problem definition	Financial statement fraud is a serious issue that can cause significant financial losses for individuals and organizations. Detecting fraudulent financial statements is a challenging task due to the complexity and diversity of the data involved. To address this problem, researchers have been investigating machine learning and data mining approaches to develop intelligent systems that can detect financial statement fraud efficiently and accurately.
	Methodology/ Algorithm	the Kitchenham methodology to analyze the problem regression methods, ensemble methods, and clustering
	Advantages	Using machine learning and data mining approaches to detect financial statement fraud has several advantages. For example, these techniques can process large amounts of data quickly and accurately, which is critical in detecting fraudulent activities. Additionally, these techniques can identify complex patterns and relationships in the data that might be missed by human analysts. Using machine learning and data mining techniques can also lead to more efficient fraud detection and reduce the time and resources required to investigate potential fraud cases.
	Disadvantages	There are also some disadvantages to using machine learning and data mining techniques for financial statement fraud detection. For example, these techniques require significant amounts of high-quality data to train the models effectively. Additionally, the models may not always be transparent, making it challenging to understand how they make decisions. Finally, these techniques are not foolproof and may still miss some fraudulent activities, leading to potential losses.

2	**Paper title**	WU XIUGUO AND DU SHENGYONG School "An Analysis on Financial Statement Fraud Detection for Chinese Listed Companies Using Deep Learning "February 22, 2022, date of current version March 4, 2022. Digital Object Identifier 10.1109/ACCESS.2022.3153478
	Problem definition	The research aims to develop models using deep learning algorithms to detect financial statement fraud in Chinese listed companies' annual reports. The traditional approaches used for detecting fraud are ineffective due to companies' tactics, resulting in significant losses to different stakeholders. Therefore, the research aims to build models that have high classification performance and develop a classification framework to detect fraud using both textual and numerical data in annual reports.
	Methodology/ Algorithm	CNN (Convolution Neural Network) RNN (Recurrent Neural Network)
	Advantages	The research presents a framework for using deep learning algorithms to detect financial statement fraud in Chinese listed companies' annual reports. The use of deep learning algorithms helps to achieve high classification performance, which traditional approaches cannot achieve due to companies' tactics. The study also highlights the importance of textual analytics in detecting fraud in financial documents.
	Disadvantages	The research has some limitations, such as the sampling period of the study being limited to five years, and some companies may have been delisted for various reasons. The study also eliminates some companies' annual reports because of their incompleteness, which may affect the prediction results. Additionally, the data source only involves Chinese listed companies, excluding those in other markets, so the applicability of the models may need further study.

3	Paper title	SHI QIU AND HONG-QU HE"Machine Learning- and Evidence Theory-BasedFraud Risk Assessment of China's Box Office"Date of publication November 27, 2018,date of current version December 27, 2018. Digital Object Identifier 10.1109/ACCESS.2018.2883487
	Problem definition	Box-office fraud in China has become a significant problem for the movie market, misleading consumers and investors and potentially harming the developing motion picture industry and shadow movie market in the country. More accurate supervision and auditing are needed to regulate the market and detect financial fraud.
	Methodology/ Algorithm	Ordered logistic regression The framework includes iterative aggregation of different evidence and the use of ordered logistic regression to calculate the basic probability assignment
	Advantages	The proposed NFM method is an effective complementary method to improve the efficiency of auditing and supervising box-office fraud in China. It uses publicly available data from different websites, including Web 2.0 websites, to assess fraud risk. The evidence theory-based framework provides a convenient iterative assessment procedure, and the ordered logistic regression algorithm improves the accuracy of fraud risk assessment.
	Disadvantages	One potential disadvantage of the proposed NFM method is that it may not capture all aspects of box-office fraud. It relies heavily on publicly available data and may not be able to capture fraudulent activities that are not easily detected through these sources. Additionally, the algorithm used for fraud risk assessment may have limitations in certain situations, and the method may require continuous updates to remain effective in detecting new fraud schemes.

4	Paper title	BADR OMAIR AND AHMAD ALTURKI "A Systematic Literature Review ofFraud Detection Metrics in Business Processes" date of publication February 4, 2020, date of current version February 12, 2020.Digital Object Identifier 10.1109/ACCESS.2020.2971604
	Problem definition	The literature review highlights the need for a comprehensive approach to detecting possible business process fraud (PBF). While metrics for possible fraud detection have been studied, there is a lack of attention given to PBF specifically, which can be detected when business processes deviate from standard operating procedures. Only a few detection metrics have been identified, and they do not fully address the different conceptual perspectives on business processes. Therefore, future research is needed to extend the current detection metrics for PBF and to evaluate their effectiveness in exposing common fraud risks.
	Methodology/ Algorithm	CNN (Convolution Neural Network) RNN (Recurrent Neural Network) Ordered logistic regression
	Advantages	The content analysis methodology used in this study allows for a thorough review of relevant literature, providing a solid foundation for future research. The identified conceptual perspectives for detecting PBF provide a comprehensive framework for organizations to assess their business processes and identify potential fraud risks. The study's focus on extending current detection metrics for PBF highlights the need for a more targeted approach to fraud detection, which can improve the effectiveness of fraud prevention measures.
	Disadvantages	The content analysis methodology has limitations, as it relies on the availability and quality of the literature reviewed. The study's focus on PBF may not address other types of fraud, and the identified detection metrics may not be applicable to all business contexts. Additionally, the study does not provide specific guidance on how organizations can apply the identified detection metrics in practice.

5	Paper title	AKIB MASHRUR , WEI LUO , NAYYAR A. ZAIDI ,AND ANTONIO ROBLES-KELLY" Machine Learning forFinancialRiskManagement: A Survey"date of publication November 5, 2020, date of current version November 19, 2020.Digital Object Identifier 10.1109/ACCESS.2020.3036322
	Problem definition	Traditional statistical models for financial risk management have limitations, and there is a need for advanced machine learning models to improve accuracy and robustness. However, challenges such as long training times, the need for large amounts of data, and the lack of explainability and fairness of machine learning models need to be addressed.
	Methodology/ Algorithm	GARCH or stochastic volatility models Long Short Term Memory (LSTM) Natural Language Processing (NLP) techniques
	Advantages	Advanced machine learning models, including deep learning, can significantly improve the accuracy and robustness of financial risk management compared to traditional statistical models. Machine learning models can also incorporate unstructured textual data to improve volatility forecasting accuracy. Federated learning systems can ensure private and secure learning using sensitive financial data.
	Disadvantages	The challenges of long training times and the need for large amounts of data for some machine learning models, such as LSTMs, may limit their applicability in some situations. Machine learning models also lack explainability and fairness, which is a concern for financial risk management. Additionally, the use of unstructured textual data for volatility forecasting may require sophisticated NLP techniques to extract relevant information accurately.