Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6, 835-845
2025
Publisher: Learning Gate
DOI: 10.55214/25768484.v9i6.7955

Integration of NLP, AI-driven data analysis, risk assessment, and electronic whistle-blowing systems in fraud detection

Chelsea Tan[1*], Calrsen Cyntia[1], Bambang Leo Handoko[1]
[1]Accounting Department, School of Accounting, Bina Nusantara University, Jakarta, Indonesia, 11480
chelsea.tan@binus.ac.id (C.T.)

Abstract: The rapid development of technology in Industry 4.0 today has encouraged the integration of Artificial Intelligence (AI), the Internet of Things (IoT), and big data in helping the operations of various industrial sectors, especially in the start-up sector. This study aims to determine whether there are factors such as Natural Language Processing, AI-driven data analysis, risk assessment, and electronic whistleblowing systems that will affect the way the system detects fraud, and to determine whether these factors cause several start-up companies to use the integration of AI, NLP, and E-WBS to accelerate the fraud disclosure process. This study involved 113 employee respondents who worked in start-up companies. The results of the respondent data were processed using SMART-PLS 4.0, which involved the reliability and validity methods, discriminant analysis, r-squared adjusted, and outer loading. The results of the study showed that Natural Language Processing, AI-driven data analysis, risk assessment, and electronic whistleblowing systems did have a positive impact or increase the accuracy of fraud disclosure in real-time, effectively, and efficiently. Early identification of fraud patterns can prevent greater losses, and parties who are aware of fraudulent actions will report them and have reporting channels that create a sense of security for the reporter.

Keywords: *AI-driven data analysis, Natural language processing, Whistle blowing systems.*

1. Introduction

1.1. Research Background

In recent periods, technology has developed very rapidly in everyday life and made industrial areas change drastically. In industry 4.0, many business operations have combined the main functions of Artificial Intelligence (AI), Internet of Things (IoT) and big data automation to create a more efficient business operational environment [1]. This combination of functions opens many opportunities and innovations, which are marked by the emergence of many startups (e.g. Gojek, Tokopedia, Ovo, Dana, and Kredivo), especially in the financial sector that use financial technology (FinTech). FinTech enhances business efficiency, reduces costs, and improves customer satisfaction by enabling faster digital transactions (i.e., QRIS), benefiting SMEs in particular [2]. However, this also causes negative impacts, such as the potential for fraud in the form of data falsification, account hijacking, phishing, and lack of transparency in financial reports based on peer-to-peer lending which does not yet have clear accounting standards [3]. This negative impact will certainly lead to asset misappropriation or fraudulent financial reporting, which will harm many parties such as companies, investors, creditors, customers, and regulators. Therefore, it is important for companies and regulators to implement effective fraud detection systems to mitigate potential fraud from the start.

In the digital era, leveraging technology for fraud detection has become essential. Traditional methods often lack the efficiency and accuracy that artificial intelligence (AI) can offer. AI techniques can adapt to evolving fraud tactics and identify hidden patterns in real-time [4]. AI-driven data analysis and AI-driven risk assessment enable organizations to detect anomalies and potential fraud more

History: Received: 14 March 2025; Revised: 2 June 2025; Accepted: 6 June 2025; Published: 10 June 2025
* Correspondence: chelsea.tan@binus.ac.id

effectively. Natural Language Processing (NLP) enhances fraud detection by analyzing textual data, identifying suspicious patterns in reports, and improving investigative processes [5]. Additionally, an electronic whistle-blowing system (e-WBS) ensures a more secure and confidential reporting mechanism, allowing whistleblowers to submit evidence—such as photos, videos, and audio recordings—without compromising their anonymity. Despite these advancements, AI-based fraud detection remains underutilized. Research has shown that integrating these technologies significantly improves fraud detection accuracy and efficiency. Therefore, adopting AI-driven solutions and secure digital reporting mechanisms is crucial for strengthening fraud detection systems.

While recent advancements in AI-driven fraud detection systems demonstrate significant potential, several key limitations persist across current research. In Qatawneh [5] examination of NLP's moderating role in AI-driven auditing systems fails to integrate whistle-blowing mechanisms with AI detection capabilities and inadequately explores practical NLP implementation challenges in real-world fraud scenarios. Yang, et al. [6] discusses the FinChain-BERT model that despite its novel architecture and high accuracy, it faces scalability constraints due to substantial computational demands that could hinder real-time detection applications, while its effectiveness remains unverified for non-financial fraud contexts. Boulieris, et al. [7] addressed dataset scarcity through their public release, encountered detection challenges from excessive data anonymization that restricted access to crucial financial indicators, compounded by their reduced reliance on traditional features despite compensating with NLP techniques. The limited number of studies specifically addressing the application of NLP in fraud detection highlights the need for further research, motivating the development of a study focused on the integration of NLP, AI-driven data analysis, AI-driven risk assessment, and whistle-blowing systems in fraud detection.

Based on the limitations identified in previous research, the researcher plans to further explore what factors will influence the system's way of detecting fraud from the integration of Artificial Intelligence (AI) and Whistle-Blowing System. Researchers also want to examine further the reasons why several company operations choose to use Artificial Intelligence and Natural Language Processing which are then integrated using a Whistle-Blowing System to speed up the fraud disclosure process. In detecting fraud, this research offers a new, more promising solution, namely by synergizing AI, NLP, and a whistle-blowing system. This step is expected to produce a system that is sophisticated, and relevant to the needs of the times. This research will provide contributions to several parties, such as practitioners who can use it as a reference for developing a more sophisticated and effective fraud detection system with the integration of AI, NLP, and Whistle Blowing System. This research can provide knowledge about how AI can be used effectively and efficiently to detect fraud for start-ups and companies that want to improve their internal monitoring systems. Apart from that, for the academic community it is to enrich the literature in the field of forensic accounting and fraud examination regarding new perspectives in the integration of artificial intelligence, natural language processing, and electronic whistle blowing systems in disclosing fraud.

1.2. Research Question

Based on the background of the research, the research questions that the researcher will test in this study are as follows:

1) Does Natural Language Processing (NLP) affect fraud detection?
2) Does AI-driven data analysis affect fraud detection?
3) Does AI-driven risk assessment affect fraud detection?
4) Does the electronic whistle-blowing system affect fraud detection?

2. Literature Review and Hypothesis Development

2.1. Artificial Intelligence and Natural Language Processing

Artificial Intelligence (AI) is the simulation of human intelligence in machines, enabling them to perform tasks that typically require human cognitive abilities [8]. AI systems use machine learning

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 835-845, 2025
DOI: 10.55214/25768484.v9i6.7955

algorithms to develop cognitive abilities and perform tasks autonomously or semi-autonomously. As AI advances, it is improving efficiency and accuracy in complex tasks. This technology continues to evolve, promising significant impacts across various domains. AI encompasses various technologies, including machine learning, natural language processing, computer vision, and robotics [9]. NLP aims to bridge the gap between human communication and computer comprehension by developing models and algorithms capable of analyzing and interpreting natural language data. Key techniques in NLP include tokenization, part-of-speech tagging, named entity recognition, sentiment analysis, machine translation, and text classification [10]. NLP has applications in various fields, such as computational linguistics, email spam detection, information extraction, summarization, and question answering [11]. Recent advancements in deep learning, particularly neural networks like CNNs and RNNs, have significantly improved NLP system performance. The field has evolved rapidly due to the availability of large datasets, increased computing power, and sophisticated machine learning models.

AI-driven risk assessment and data analytics have revolutionized fraud detection in various industries, particularly in finance. These technologies leverage machine learning algorithms and advanced analytics to process vast amounts of data rapidly, identifying patterns and anomalies that may indicate fraudulent activities. Unlike traditional methods relying on historical data and static models, AI-driven approaches offer real-time insights and continuously learn from latest information, providing dynamic and precise risk evaluations [12]. This capability enables financial institutions to enhance their fraud detection accuracy, optimize decision-making processes, and adopt a more proactive approach to risk management.

AI's ability to analyze large datasets comprehensively surpasses human limitations, allowing for more efficient resource allocation by focusing on high-risk areas. However, challenges related to data quality, availability, privacy, and ethical considerations persist, necessitating robust governance frameworks for AI implementation in fraud detection. Despite these challenges, AI-driven risk assessment and data analytics significantly improve operational efficiency and decision-making in fraud detection, with potential applications across various sectors including finance, healthcare, and manufacturing [13].

2.2. Electronic Whistle Blowing System

Whistle Blowing System (WBS) is a system used by company operations to mitigate fraud, by receiving anonymous complaints of suspicion by employees, vendors, customers, investors, and other stakeholders. Scherbarth and Behringer [14] indicates that the whistleblowing system is one of the company's internal instruments that is highly recommended by academics and practitioners to prevent and detect compliance violations committed by company personnel. However, in today's sophisticated digital world, company operations need to implement a digital-based whistleblowing system, which can reduce anonymity and the risk of data leakage. Therefore, many companies have implemented a digital-based whistleblowing system, such as using websites, chatbots, blockchain, and artificial intelligence, or what we now know as the Electronic Whistle Blowing System (EWBS).

EWBS or commonly known as Electronic Whistle Blowing System, is a digital version of the whistleblowing system, which uses the help of sophisticated technology today to report suspected potential fraud through a special platform. Berendt and Schiffner [15] indicates that in today's digital era, reporting using a whistleblowing system experiences many risks and challenges, such as the limitations of anonymity that can be provided by the system, and the possibility of consequences from outsourcing used to reveal fraud to the public. Many company operations also move from the conventional whistleblowing system to the electronic whistleblowing system, because EWBS can respond faster or in real-time than WBS. In addition, EWBS uses encryption, or data is converted into a code that cannot be read by the public without a description key that explains it, so that the security and confidentiality of the reporter are more guaranteed. Fraud reporters only need to upload evidence in the form of photos, voice recordings, or videos which will later be able to immediately detect the fraud

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 835-845, 2025
DOI: 10.55214/25768484.v9i6.7955

pattern using big data or AI. Cloud-based data storage in EWBS makes the storage costs incurred by several companies cheaper, making it attractive in the eyes of company stakeholders.

Many company operations use artificial intelligence alone to detect fraud by monitoring user patterns and profiles but can only reduce fraud by 40% [16]. This shows that by only using AI, company operations will not be able to understand someone's motives for carrying out a transaction or action, because AI only works based on existing patterns and historical data. Therefore, AI must be combined with EWBS to improve the accuracy of fraud detection and effectively reduce the occurrence of fraud cases by a larger percentage. The use of AI in collaboration with EWBS can increase fraud accuracy, because it will capture fraud reports in real-time which will later be aligned with predetermined acceptance criteria, and this has succeeded in overcoming 90% of reports within 48 hours [17]. Artificial Intelligence will identify reports sent anonymously via EWBS by matching them to similar incidents that have occurred before, so that the response given will be faster, increase accuracy, increase protection from the reporter, and make operational performance more efficient.

2.3. Natural Language Processing and Fraud Detection

Natural Language Processing (NLP) has emerged as a significant tool in fraud detection, particularly within the financial and insurance sectors. Qatawneh [5] indicates that NLP, when integrated with AI and machine learning, can enhance fraud detection accuracy and efficiency. This is reinforced by research by Hermawan and Novita [18] stating that NLP's ability to analyze textual data enables the identification of linguistic patterns associated with fraudulent behavior, improving the detection of deceptive practices in various contexts, including news media. In insurance, NLP facilitates the automated analysis of written claims, identifying keywords and phrases indicative of potential fraud [19]. The integration of NLP with other advanced technologies has shown promising results, including a 30% increase in fraud detection accuracy and a 40% improvement in compliance documentation processing efficiency [19]. These advancements suggest that NLP plays a crucial role in modernizing fraud detection strategies across multiple industries. Based on this explanation, we formulate the first hypothesis as follows:

H_1: Natural Language Processing Significantly Affects Fraud Detection

2.4. AI-Driven Data Analysis and Fraud Detection

Artificial Intelligence (AI) is revolutionizing financial fraud detection by offering more accurate, scalable, and adaptive systems compared to traditional methods [20]. Some parties assume that the use of Artificial Intelligence (AI) will reduce the detection of fraud that does not actually exist, because there are adjustments to the algorithm that can reduce errors. Artificial intelligence can process large amounts of data on a fast time scale, allowing for real-time handling, which creates more accurate and timely information [21]. Several other studies also say that [22]. Artificial Intelligence can increase the accuracy of fraud detection and reduce false positives or false negatives in the financial industry.
And so, we formulate the second hypothesis as follows:

H_2: AI-Driven Data Analysis Significantly Affects Fraud Detection

2.5. AI-Driven Risk Assessment and Fraud Detection

AI-driven risk assessment models significantly improve real-time fraud detection and adaptability to evolving fraud patterns, with one study reporting a 98.7% detection accuracy and the ability to process over 5 million transactions per second [23]. These technologies enable financial institutions to process vast amounts of data rapidly, identifying patterns and anomalies that might be missed by human analysts. AI-driven approaches offer dynamic and precise risk evaluations, continuously updating and learning from added information, which is particularly beneficial in credit scoring and default prediction as stated by Ikudabo and Kumar [24]. In auditing, AI enhances strategic planning by providing real-time insights and allowing auditors to focus on high-risk areas [12]. Based on this explanation, we formulate the third hypothesis as follows:

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 835-845, 2025
DOI: 10.55214/25768484.v9i6.7955

H_5. AI-Driven Risk Assessment Significantly Affects Fraud Detection.

2.6. Electronic Whistle Blowing System and Fraud Detection

Criminal acts involving fraud cases have not yet decreased in various countries, and in fact are still increasing with various new modes used by the perpetrators [25]. In this era of digital technology, company operations must implement a sophisticated, effective, and efficient electronic violation reporting system to spread fraud cases. Nwakeze, et al. [26] shows that 30% of companies throughout the world experience fraud because of their own employees or company management. Researchers also indicated that electronic violation reporting systems are important in detecting and spreading fraud and preventing violations of the law. Apart from that, the use of an electronic whistleblowing system as a system that will detect fraud is believed to be a system that can help auditors to carry out investigative audit activities, because the incoming report will be the first step to prevent violations [25]. So, to find out more about whether the use of EWBS an effective way is to detect fraud, we formulated the fourth hypothesis:

H_6. Electronic Whistle Blowing System Significantly Affect Fraud Detection.

3. Method, Data, and Analysis

3.1. Research Method

Quantitative method is the main method that will be used in the research with the utilization of primary data. The population that will be used in this research is an unknown population that will be focused on employees working in start-up companies. The research uses unknown population because of the high level of workforce fluctuation, where every day there must be new employees who join, resign, have completed their contracts, and are transferred to branch offices. This is the main factor that made researchers decide to use the method from [27]. Hair suggests that in an unknown population, the sample size used should be at least 5-10 times the variables in each indicator.

The researcher will use the convenience sampling method as the sampling method in the study because this method makes it easier to achieve a higher level of participation, speeds up the data collection process, and the data obtained using the convenience sampling method can be analyzed more deeply easily. The researcher will use Google Form to conduct an online survey. The survey will use a Likert scale of 1-5, where 1 is strongly disagrees, and 5 is strongly agrees. Data analysis in the study will be carried out using Structural Equation Modeling Partial Least Square (SEM-PLS). The researcher decided to use SEM PLS, because it is able to test models with many variables and indicators in the study, is very relevant for many types of research, the output of SEM-PLS is simpler than other methods, so it will be easier to interpret, and can be used on independent variables that have a high level of correlation [27]. Dataset of this paper is available at https://zenodo.org/records/15468997

3.2. Operation of Variable

To ensure the variables in this study are clearly defined and quantitatively measurable, we operationalize them based on predetermined standards. Table 1 presents the operationalization of variables to facilitate a structured quantitative assessment.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 835-845, 2025
DOI: 10.55214/25768484.v9i6.7955

Table 1.
Operation of Variable.

Variable	Indicator	Reference
Natural Language Processing (NLP)	1. NLP algorithms effectively extract relevant features from unstructured text data in fraud detection cases. 2. NLP models improve the system's ability to classify fraudulent patterns in textual data. 3. NLP systems demonstrate high accuracy in detecting inconsistencies in reported cases based on textual analysis.	Shah [21]
AI-Driven Data Analysis (ADA)	1. The system integrates large datasets from diverse sources to identify fraud patterns. 2. AI-driven algorithms continuously analyze data in real time to detect anomalies indicative of fraudulent activities. 3. The AI system can predict potential fraud incidents based on historical and real-time data patterns.	Mrida, et al. [13]
AI-Driven Risk Assessment (ARA)	1. The AI model assesses potential risks by evaluating financial transactions and behaviors indicating fraud. 2. Employees believe that the AI system can accurately assess the level of risk in each fraud case based on historical data and trends. 3. The AI risk assessment system provides real-time risk scores to prioritize fraud investigations.	Kalluri [23]
Electronic Whistleblowing System (EWS)	1. Effectiveness of the whistleblowing system 2. How to report fraud 3. Whistleblowing system benefits	Handoko, et al. [28]
Fraud Detection (FD)	1. The AI system demonstrates a high accuracy rate in identifying fraudulent transactions in financial systems. 2. The system's fraud detection capabilities are assessed based on its ability to minimize false positives and negatives. 3. The system's fraud detection performance improves with continuous learning and feedback from human investigators.	Adhikari, et al. [20]

4. Result and Discussion

4.1. Identity of Respondents

From the questionnaire distributed on social media, we have obtained 113 answers from respondents (see Table 2). Based on table 2, it can be concluded that most of the respondents from the researcher's study were men who worked as IT Security or Cybersecurity with an age range of 21-30 years old.

Table 2.
Identity of Respondents.

Gender	Amount	Age	Amount	Position	Amount
Male	68	21-30 years	52	Founder/Co-Founder	13
Female	45	31-40 years	18	CEO/C-Level Executive	14
		41-50 years	19	Head of Department	16
		51-60 years	20	Data Analyst	23
		>61 years	4	IT Security/Cybersecurity	25
				Fraud Analyst	14
				Operation Intern	1
				Software Engineer	1
				Marketing	1
				Finance Officer	1
				Developer	1
				Mandarin Teacher	1
				Staff	1
				Business Development	1

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 835-845, 2025
DOI: 10.55214/25768484.v9i6.7955

4.2. Outer Model Test

The outer model test or commonly known as outer loading using SMART-PLS is carried out with the aim of finding out how well the research indicators represent latent variables. Latent variable is a variable that cannot be measured directly, but can be measured with the help of other variables/indicators [29]

According to Ghozali [30] in using SMART-PLS, the value that reflects that the indicators in the study are highly correlated is ideally 0.70 or higher indicating that the indicators represent the latent variables strongly, 0.50-0.70 indicating that the representation of the latent variables is still acceptable, and below 0.50 indicates a weak representation [31]. Therefore, the researcher has removed several indicators (FD3 and FD8) that do not meet the requirements and ensured that all indicators in this study are considered to represent the existing latent variables (refer to table 3).

Table 3.
Outer Model Test.

Indicator	Outer Loading	Indicator	Outer Loading	Indicator	Outer Loading	Indicator	Outer Loading	Indicator	Outer Loading
ADA.1	0.751	ARA.1	0.667	EWS.1	0.746	FD.1	0.545	NLP.1	0.771
ADA.2	0.691	ARA.2	0.774	EWS.2	0.783	FD.2	0.796	NLP.2	0.764
ADA.3	0.734	ARA.3	0.838	EWS.3	0.656	FD.4	0.635	NLP.3	0.734
ADA.4	0.830	ARA.4	0.790	EWS.4	0.773	FD.5	0.714	NLP.4	0.744
ADA.5	0.778	ARA.5	0.828	EWS.5	0.753	FD.6	0.730	NLP.5	0.770
ADA.6	0.747	ARA.6	0.787	EWS.6	0.654	FD.7	0.808	NLP.6	0.757
ADA.7	0.762	ARA.7	0.808	EWS.7	0.779	FD.9	0.834	NLP.7	0.755
ADA.8	0.797	ARA.8	0.772	EWS.8	0.789			NLP.8	0.782
ADA.9	0.722	ARA.9	0.781	EWS.9	0.707			NLP.9	0.738
								NLP.1	0.742

4.3. Convergent Validity

Convergent validity is used to ensure that the variables measured by several existing indicators are correlated with each other and can increase accuracy in research [32]. It is said that indicators are highly correlated in convergent validity if the AVE and Rho_C values meet the minimum values ≥0.50-≥0.70. In this study, the indicators used by researchers have reached values >0.80->0.90 which indicates that the existing indicators have represented the research model very well (refer to table 4).

Table 4.
Convergent Validity.

Variable	Composite Reliability (Rho-C)	Average Variance Extracted
ARA	0.935	0.615
ADA	0.931	0.574
NLP	0.930	0.571
EWS	0.915	0.547
FD	0.887	0.532

4.4. Discriminant Validity

Discriminant validity assesses the ability of measurement instruments and theoretical concepts to differentiate between attributes that should not be measured together. It involves comparing correlations between measures after accounting for measurement error [33]. This study presents the discriminant validity through the Heterotrait-Monotrait Ratio (HTMT) in Table 5.

HTMT can detect potential indiscrimination among latent variables that may be missed by other methods [33]. If the HTMT value is below the recommended HTMT threshold of 0.85, then the discriminant validity is achieved [32].

Table 5 indicates that the variables' HTMT are below the 0.85 threshold. This demonstrates that the model has successfully passed the discriminant validity test.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 835-845, 2025
DOI: 10.55214/25768484.v9i6.7955

Table 5.
Heterotrait-Monotrait Ratio.

Variable	AI-Driven Data Analysis	AI-Driven Risk Assessment	Electronic Whistleblowing System	Fraud Detection	Natural Language Processing
AI-Driven Data Analysis					
AI-Driven Risk Assessment	0.279				
Electronic Whistleblowing System	0.121	0.157			
Fraud Detection	0.577	0.492	0.417		
Natural Language Processing	0.143	0.133	0.161	0.574	

4.5. Hypothesis Testing

It is a fundamental statistical method used to draw conclusions about population parameters based on sample data. This process involves comparing the computed t-value with the critical t-value from a table. The alternative hypothesis is accepted if the computed t-value exceeds 1.96 and the p-value is less than 0.05 [16]. The original samples refer to the initial data set collected from the population. The hypothesis testing results for this study are presented in Table 6.

To arrive at the hypothesis testing, the researchers have used SMART PLS to calculate the data, as shown in Figure 1. The data in Table 7 reveal that all hypotheses significantly impact fraud detection. The P-values—7.593, 8.883, 7.025, and 10.045—are all greater than 1.96, and the p-values for each variable are below 0.05. This indicates that natural language processing, AI-driven data analysis, AI-driven risk assessment, and electronic whistleblowing system all have a significant effect on fraud detection. As a result, hypotheses H1, H2, H3, and H4 can be accepted. Additionally, the coefficient of determination shown in Table 7 reveals that the adjusted R-squared value is 75.1%, indicating that the model has a high explanatory power, with an R^2 above 50% [34].

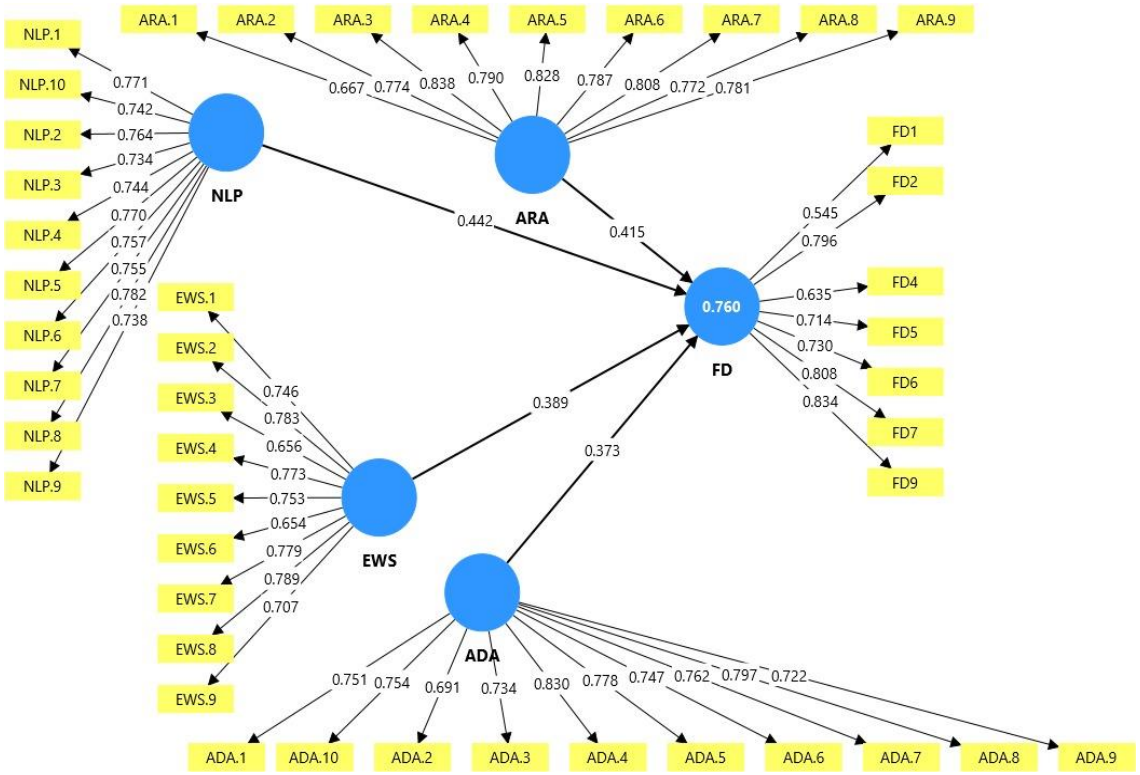

Figure 1.
Path Coefficient.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 835-845, 2025
DOI: 10.55214/25768484.v9i6.7955

Table 6.
Hypothesis Testing.

Variable	Original Sample	T Statistics	P Values
H1: Natural Language Processing Significantly Affects Fraud Detection	0.373	7.593	0.000
H2: AI-Driven Data Analysis Significantly Affects Fraud Detection	0.415	8.883	0.000
H3: AI-Driven Risk Assessment Significantly Affects Fraud Detection	0.389	7.025	0.000
H4: Electronic Whistle Blowing System Significantly Affect Fraud Detection	0.442	10.045	0.000
R-Square/R-Square Adjusted	0.760/0.751		

5. Conclusion and Suggestion

Based on the results of data processing obtained in chapter 4, researchers can conclude that the integration of NLP, AI-Driven Data Analysis, Risk Assessment, and Electronic Whistle-Blowing System will affect the results of fraud detection. All existing indicators, namely Natural Language Processing, AI-Risk Assessment, AI-Driven Data Analysis, Electronic Whistle Blowing Systems and Fraud Detection have a fairly strong correlation with each other, so that integration will produce a more sophisticated, effective, and efficient system, and in accordance with the needs of the current era of technology. NLP has been proven to increase accuracy in identifying fraud patterns in information, AI-Driven Data Analysis can help process data that contains indications of fraud in large volumes in real-time, and risk assessments have been proven to increase more accurate predictions for historical and current data, and E-WBS which can be a more efficient and secure reporting channel for reporters.

This research primarily focuses on fraud detection within startup companies, offering insights and strategies that can be directly applied to this sector. However, the findings and methodologies discussed are not limited to startups alone. The frameworks and conclusions presented in this study are also relevant to other industries, as fraud detection remains a critical concern across various business environments. Hence, it can serve as a valuable reference for organizations in diverse sectors seeking to strengthen their fraud detection mechanisms.

Although this research provides empirical evidence that natural language processing, AI-driven data analysis, AI-driven risk assessment, and electronic whistleblowing system significantly impact fraud detection, there remains potential to explore additional factors which make up the 24.9% that could further detect fraud, which could be explored. Furthermore, this study could serve as a foundation for future research with a larger sample size to offer a more accurate representation of fraud detection.

Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Acknowledgements

Contributor of this paper Conceptualization, C.T., C.C., B.L.H.; methodology: B.L.H.; software: M.A.; validation: B.L.H., K.D.; formal analysis: B.LH.; investigation: C.T., C.C.; resources: C.T., C.C.; data curation: B.L.H.; writing-original draft: C.T., C.C.; writing-review and editing: B.L.H; visualization: C.C.; supervision: B.L.H.; project administration: B.L.H., C.C., C.T.; funding acquisition, N/A. All authors have read and agreed to the published version of the manuscript.

References

[1] M. Javaid, A. Haleem, R. P. Singh, R. Suman, and E. S. Gonzalez, "Understanding the adoption of Industry 4.0 technologies in improving environmental sustainability," *Sustainable Operations and Computers*, vol. 3, pp. 203-217, 2022. https://doi.org/10.1016/j.susoc.2022.01.008

[2] N. O.-O. Anwuli, E. A. Edith, and R. C. Njideka, "Fintech integration in small and medium enterprises: Enhancing economic resilience and operational efficiency," *Finance & Accounting Research Journal*, vol. 6, no. 8, pp. 1485–1500, 2024. https://doi.org/10.51594/farj.v6i8.1435

[3] Y. Oktaviani and M. K. Dewi, "Is information transparency important for funders? A case study of sharia P2P lending companies in Indonesia," *Journal of Accounting and Investment*, vol. 24, no. 2, pp. 462-486, 2023. https://doi.org/10.18196/jai.v24i2.17220

[4] N. Jain and S. Patil, "Artificial intelligence models for fraud detection : Advancements , challenges , and future prospects," *International Journal of Global Innovations and Solutions*, pp. 1–6, 2024. http://dx.doi.org/10.21428/e90189c8.6d8ab5f6

[5] A. M. Qatawneh, "The role of artificial intelligence in auditing and fraud detection in accounting information systems: moderating role of natural language processing," *International Journal of Organizational Analysis*, 2024. https://doi.org/10.1108/IJOA-03-2024-4389

[6] X. Yang *et al.*, "FinChain-BERT: A high-accuracy automatic fraud detection model based on NLP methods for financial scenarios," *Information*, vol. 14, no. 9, pp. 1–26, 2023. https://doi.org/10.3390/info14090499

[7] P. Boulieris, J. Pavlopoulos, A. Xenos, and V. Vassalos, "Fraud detection with natural language processing," *Machine Learning*, vol. 113, no. 8, pp. 5087-5108, 2024. https://doi.org/10.1007/s10994-023-06354-5

[8] F. Morandín-Ahuerma, "International journal of research publication and reviews what is artificial intelligence?," *International Journal of Research Publication and Reviews*, vol. 3, no. 12, p. 5500, 2022.

[9] D. Mochihashi, "Natural language processing in robotics," *Journal of the Robotics Society of Japan*, vol. 39, no. 5, pp. 399–404, 2021.

[10] Vanita, "An extant of natural language processing," *International Journal For Multidisciplinary Research*, vol. 6, no. 2, pp. 62–69, 2024.

[11] A. Anam, S. Ali, and V. K. Shandilya, "AI-natural language processing (NLP)," *International Journal For Research in Applied Science and Engineering Technology*, vol. 9, pp. 1–8, 2021. http://dx.doi.org/10.22214/ijraset.2021.37293

[12] E. R. Onwubuariri, B. O. Adelakun, O. P. Olaiya, and J. E. K. Ziorklui, "AI-Driven risk assessment: Revolutionizing audit planning and execution," *Finance & Accounting Research Journal*, vol. 6, no. 6, pp. 1069-1090, 2024. https://doi.org/10.51594/farj.v6i6.1236

[13] H. Mrida, A. Rahman, and S. Alam, "Ai-driven data analytics and automation: A systematic literature," *Journal Oof Strategic Data Management and Innovation*, vol. 02, no. 01, pp. 21–40, 2022. https://doi.org/10.71292/sdmi.v2i01.9

[14] S. Scherbarth and S. Behringer, "Whistleblowing systems: A systematic literature review on the design specifications and the consideration of the risk for organizational insiders to blow the whistle," *Corporate Ownership and Control*, vol. 18, no. 2, pp. 60-73, 2021.

[15] B. Berendt and S. Schiffner, "Whistleblower protection in the digital age--why'anonymous' is not enough," *The International Review of Information Ethics*, vol. 31, no. 1, pp. 1–16, 2021. https://doi.org/10.29173/irie479

[16] A. L. Mhatre, "Using AI to combat medicaid fraud, waste, and abuse," *International Journal of Advancements in Computational Technology*, vol. 2, pp. 2–5, 2024. https://doi.org/10.56472/25838628/IJACT-V2I4P112

[17] R. Reznikov, "Enhancing project management success through artificial intelligence," *International Journal of Multidisciplinary Research and Growth Evaluation*, vol. 6, no. 6, pp. 1036–1046, 2025.

[18] A. Hermawan and N. Novita, "The effect of governance, risk management, and compliance on efforts to minimize potential fraud based on the fraud pentagon concept," *Asia Pacific Fraud Journal*, vol. 6, no. 1, pp. 82-95, 2021.

[19] I. D. Ssetimba, J. Kato, E. O. Pinyi, E. Twineamatsiko, H. N. Nakayenga, and E. Muhangi, "Advancing electronic communication compliance and fraud detection through machine learning, NLP and generative AI: A Pathway to enhanced cybersecurity and regulatory adherence," *World Journal of Advanced Research and Reviews*, vol. 23, no. 2, pp. 697-707, 2024. https://doi.org/10.30574/wjarr.2024.23.2.2364

[20] P. Adhikari, P. Hamal, and F. B. Jnr, "Artificial Intelligence in fraud detection: Revolutionizing financial security," *International Journal of Science and Research Archive*, vol. 13, no. 01, pp. 1457-1472, 2024. http://dx.doi.org/10.30574/ijsra.2024.13.1.1860

[21] Q. Shah, "Predictive analytics and AI integration in fraud detection and risk assessment for financial services," *International Journal of Financial Technology and Risk Management*, vol. 12, no. 3, pp. 45–62, 2024.

[22] N. A. Davis and S. A. Harris, "Leveraging machine learning models for real-time fraud detection in financial transactions," *International Journal of Computer Technology and Science*, vol. 1, no. 1, pp. 1–6, 2024. https://doi.org/10.62951/ijcts.v1i1.56

[23] K. Kalluri, "AI-driven risk assessment model for financial fraud detection: A data science perspective," *International Journal of Scientific Research and Management*, vol. 12, no. 12, pp. 1764-1774, 2024. https://doi.org/10.18535/ijsrm/v12i12.ec01

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 835-845, 2025
DOI: 10.55214/25768484.v9i6.7955

[24] A. O. Ikudabo and P. Kumar, "AI-driven risk assessment and management in banking: Balancing innovation and security," *International Journal of Research Publication and Reviews*, vol. 5, no. 10, pp. 3573-88, 2024. https://doi.org/10.55248/gengpi.5.1024.2926

[25] H. Hamilah, A. Adji Suratman, and S. Saeful Alam, "The effectiveness of the internal whistleblowing system in fraud detection and prevention (Empirical Study at PT Asuransi Jiwasraya (Persero)," *Akuntansi*, vol. 3, no. 1, pp. 1-20, 2022. https://doi.org/10.38035/dijefa.v2i6.1142

[26] E. O. Nwakeze, S. Oshiole, and J. C. Orajekwe, "Whistle blowing system (Wb-S) and fraud prevention and detection in companies: Proposing a conceptual model for the protection of whistle blowers," *International Journal of Research Publication and Reviews*, vol. 4, no. 4, pp. 1017-1027, 2023. https://doi.org/10.55248/gengpi.2023.4.4.35142

[27] J. Hair and A. Alamer, "Partial least squares structural equation modeling (PLS-SEM) in second language and education research: Guidelines using an applied example," *Research Methods in Applied Linguistics*, vol. 1, no. 3, pp. 1–16, 2022. https://doi.org/10.1016/j.rmal.2022.100027

[28] B. L. Handoko, E. C. Febriyanne, and N. Ayuanda, "Enhancing fraud prevention: Exploring the interplay of internal control system, organizational culture, internal audit roles and online whistleblowing mechanisms," presented at the 2024 6th Asia Conference on Machine Learning and Computing (ACMLC 2024), July 2628, 2024, Bangkok, Thailand, 2024.

[29] R. Van Bork, M. Rhemtulla, L. J. Waldorp, J. Kruis, S. Rezvanifar, and D. Borsboom, "Latent variable models and networks: Statistical equivalence and testability," *Multivariate Behavioral Research*, vol. 56, no. 2, pp. 175-198, 2021. https://doi.org/10.1080/00273171.2019.1672515

[30] I. Ghozali, *Structural equation modeling: Alternative method with partial ;east squares (PLS)*, 4th ed. Semarang: Diponegoro University Publishing Agency, 2015.

[31] P. G. Subhaktiyasa, "PLS-SEM for multivariate analysis: A practical guide to educational research using SmartPLS," *EduLine: Journal of Education and Learning Innovation*, vol. 4, no. 3, pp. 353-365, 2024.

[32] G. W. Cheung, H. D. Cooper-Thomas, R. S. Lau, and L. C. Wang, "Reporting reliability, convergent and discriminant validity with structural equation modeling: A review and best-practice recommendations," *Asia Pacific Journal of Management*, vol. 41, no. 2, pp. 745–783, 2024.

[33] M. Rönkkö and E. Cho, "An updated guideline for assessing discriminant validity," *Organizational Research Methods*, vol. 25, no. 1, pp. 6–47, 2022.

[34] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *Peerj Computer Science*, vol. 7, p. e623, 2021. https://doi.org/10.7717/PEERJ-CS.623