[1]Huiying Kang

Text Mining and Financial Analysis Modeling for Financial Statement Disclosure

Journal of Electrical Systems

Abstract: - This paper explores the integration of financial analysis, text mining, and blockchain technology to enhance financial statement disclosure and analysis processes. Leveraging text mining algorithms, the study extracts valuable insights from unstructured financial data, unveiling hidden patterns and sentiments within financial statements. Additionally, blockchain technology is employed to ensure secure and immutable storage of financial data, fostering transparency and trust in financial transactions and disclosures. Through the automation and digitization of financial reporting processes, stakeholders can streamline financial statement analysis, identify risks and opportunities, and make data-driven decisions with greater accuracy and efficiency.

Keywords: Leveraging text mining, blockchain, financial data, secure, accuracy

1. Introduction

Text mining, an integral part of natural language processing, involves extracting valuable insights from large volumes of unstructured text data. This multifaceted process begins with text preprocessing, where data is cleaned and organized for analysis by removing irrelevant elements and standardizing text formats [1]. Tokenization then breaks down the text into manageable units, facilitating further analysis. Statistical techniques such as TF-IDF are applied to identify significant terms and concepts within the text. Named Entity Recognition (NER) aids in identifying and categorizing entities like names, locations, and dates. Sentiment analysis assesses the emotional tone conveyed in the text, crucial for understanding public opinion or customer sentiment [2]. Topic modeling reveals underlying themes or subjects across documents, while text classification categorizes documents into predefined classes. These methodologies collectively empower organizations to derive actionable insights from textual data, enabling informed decision-making and enhancing various applications, from customer feedback analysis to information retrieval [3].

Text mining plays a vital role in financial analysis by extracting valuable insights from vast amounts of unstructured textual data, including financial reports, news articles, social media, and analyst reports [4]. Through text preprocessing techniques such as cleaning, tokenization, and removing stopwords, the raw data is refined for analysis. Statistical methods like sentiment analysis help gauge market sentiment, investor confidence, and public perception regarding specific financial instruments or companies [5]. Named Entity Recognition (NER) identifies key entities such as company names, executives, and financial indicators, facilitating trend analysis and competitive intelligence. Topic modeling can uncover emerging trends, market movements, and sector-specific themes, providing valuable context for investment decisions [6]. Additionally, text classification techniques enable the categorization of financial documents into relevant sectors or asset classes, streamlining information retrieval and portfolio management. Text mining empowers financial analysts to leverage textual data for more informed investment strategies, risk management, and market forecasting in an increasingly data-driven financial landscape [7].

Text mining is increasingly applied to statement disclosures in finance to extract meaningful insights from textual information provided by companies. This involves parsing through financial reports, including annual reports, quarterly filings, and other regulatory disclosures, to identify crucial information relevant to investors and analysts [8]. Text preprocessing techniques are utilized to clean and standardize the data, removing noise and enhancing readability. Sentiment analysis helps assess the tone and sentiment conveyed in the statements, providing insights into management's outlook, potential risks, and opportunities. Named Entity Recognition (NER) identifies key entities such as company names, subsidiaries, and related parties mentioned in the

[1] Shi Jia Zhuang University of Applied Technology, Shijiazhuang, Hebei,050081, China

*Corresponding author email: kangxq1975@126.com

disclosures, aiding in understanding corporate structure and relationships [9]. Topic modeling enables the discovery of recurring themes or topics across multiple disclosures, such as revenue drivers, cost factors, and strategic initiatives. Moreover, text classification techniques categorize statements based on their content, facilitating comparative analysis and benchmarking across companies or industries [10]. By harnessing the power of text mining, financial analysts can efficiently analyze statement disclosures, uncover hidden patterns, and make more informed investment decisions.

Text mining combined with financial analysis modeling presents a powerful approach for extracting valuable insights from financial statement disclosures [11]. By leveraging text mining techniques, such as natural language processing and machine learning, financial analysts can parse through vast amounts of textual data contained in financial statements, including annual reports, quarterly filings, and regulatory disclosures [12]. Through text preprocessing methods, such as cleaning, tokenization, and sentiment analysis, relevant information is extracted and organized for analysis. Sentiment analysis enables the assessment of management sentiment and market outlook expressed in the disclosures, aiding in understanding potential risks and opportunities [13]. Named Entity Recognition (NER) identifies key entities such as company names, subsidiaries, and significant financial indicators, providing clarity on corporate structure and financial performance. Additionally, topic modeling helps in identifying recurring themes and trends across multiple disclosures, allowing for deeper insights into financial health, industry dynamics, and strategic priorities [14]. By integrating text mining with financial analysis modeling, analysts can enhance their understanding of financial statement disclosures, enabling more informed decision-making, risk management, and investment strategies in the dynamic landscape of finance.

The contribution of this paper lies in its comprehensive exploration and integration of advanced technologies, namely financial analysis, text mining, and blockchain, to enhance the process of financial statement disclosure and analysis. By leveraging text mining algorithms, the paper enables the extraction of valuable insights from unstructured financial data, facilitating the identification of trends, sentiments, and patterns within financial statements. Additionally, the incorporation of blockchain technology ensures the secure and immutable storage of financial data, fostering transparency and trust in financial transactions and disclosures. The paper contributes to the advancement of financial analysis practices by proposing innovative approaches for automating and digitizing financial reporting processes. Through the integration of these cutting-edge technologies, financial analysts can streamline the analysis of financial statements, identify potential risks and opportunities, and make data-driven decisions with greater accuracy and efficiency. Furthermore, by addressing challenges such as data privacy, regulatory compliance, and cybersecurity, the paper lays the groundwork for a more transparent, resilient, and efficient financial ecosystem.

2. Literature Review

Text mining and financial analysis modeling has emerged as a critical area of study, particularly in the context of financial statement disclosure. In today's data-rich environment, financial statements contain a wealth of textual information that can provide valuable insights into a company's performance, risks, and future prospects. This literature review aims to explore the evolving landscape of text mining techniques and financial analysis models employed for the analysis of financial statement disclosures. By examining existing research, theoretical frameworks, and empirical studies, this review seeks to elucidate the methodologies, challenges, and potential applications of text mining in extracting meaningful information from financial disclosures. Furthermore, it aims to investigate the integration of text mining with financial analysis modeling techniques, such as sentiment analysis, named entity recognition, and topic modeling, to enhance decision-making processes in finance.

Moreno and Caminero (2022) focus on the application of text mining to analyze climate-related disclosures, highlighting the relevance of environmental factors in financial reporting. Chen et al. (2023) investigate bankruptcy prediction using machine learning models with text-based communicative value from annual reports, emphasizing the predictive power of textual information in financial distress assessment. Xiuguo and Shengyong (2022) delve into financial statement fraud detection for Chinese listed companies using deep learning, contributing insights into fraud prevention strategies. Meanwhile, Huang et al. (2023) introduce FinBERT, a large language model tailored for extracting financial information from textual data, demonstrating advancements in text mining technology for financial analysis. Brown, Hinson, and Tucker (2024) explore the relationship between financial statement adequacy and firms' Management Discussion and Analysis (MD&A) disclosures, shedding light on the quality of narrative disclosures in financial reporting. Additionally, other studies such as Jiang et al. (2022), Suzuki et al. (2023), Duan et al. (2023), Ashtiani and Raahemi (2023), and Zhao et al. (2022) contribute further insights into areas like financial distress prediction, domain-specific language models for financial text mining, government accounting information systems enhancement, news-based market prediction, and sentiment tone features in financial distress prediction.

Aboud and Robinson (2022) provide valuable insights into fraudulent financial reporting and the role of data analytics in fraud detection, highlighting the importance of leveraging advanced analytical techniques for maintaining financial integrity. Tóth, Suta, and Szauter (2022) examine the interrelation between climate-related sustainability and financial reporting disclosures in the European automotive industry, emphasizing the growing importance of environmental considerations in corporate disclosures. Additionally, Lopez-Lira (2023) investigates risk factors disclosed by companies and their impact on stock returns, providing crucial insights for investors and risk managers. Fijałkowska and Hadro (2022) analyze risk information in non-financial disclosure, shedding light on the significance of non-financial disclosures for assessing organizational risk exposure. Meanwhile, Roeder, Palmer, and Muntermann (2022) explore the information value of analyst reports in credit risk management, emphasizing the role of data-driven decision-making in financial institutions. Breijer and Orij (2022) examine the comparability of non-financial information and the impact of the non-financial reporting directive, offering insights into the standardization of non-financial disclosures. Finally, Wang et al. (2023) investigate corporate diversity statements and employees' online Diversity, Equity, and Inclusion (DEI) ratings using machine learning techniques, highlighting the importance of textual analysis in assessing organizational culture and employee perceptions.

Studies have investigated various aspects, including the application of text mining in climate-related disclosures, bankruptcy prediction, fraud detection, and financial distress prediction. Advanced machine learning models, such as deep learning and large language models, have been developed to extract valuable information from textual data, enhancing decision-making processes in finance. Moreover, research has emphasized the importance of narrative disclosures, such as Management Discussion and Analysis (MD&A), in financial statement adequacy assessment and fraud detection. The significance of non-financial disclosures, particularly regarding environmental sustainability and risk information, has also been highlighted. Furthermore, studies have explored the role of analyst reports, government accounting information systems, and corporate diversity statements in financial analysis, underscoring the diverse applications of text mining in understanding organizational behavior and performance.

3. Financial Analysis with Text Mining

"Financial Analysis with Text Mining" is a multidisciplinary approach that combines traditional financial analysis methods with text mining techniques to extract valuable insights from textual data sources such as financial statements, news articles, social media, and analyst reports. This approach recognizes the significance of textual information in understanding market sentiment, identifying emerging trends, and assessing risks and opportunities in financial markets. By leveraging text mining algorithms and natural language processing tools, financial analysts can analyze large volumes of unstructured textual data efficiently and uncover hidden patterns and relationships that may not be apparent through traditional quantitative analysis alone. Financial analysis with text mining offers numerous benefits, including improved decision-making, enhanced risk management, and the ability to identify actionable investment opportunities in an increasingly complex and data-driven financial landscape.

Text mining involves the extraction of valuable insights from unstructured textual data through various natural language processing (NLP) techniques. When applied to financial analysis, text mining can help extract information from financial reports, news articles, social media, and other textual sources to complement traditional quantitative analysis methods. One fundamental aspect of financial analysis with text mining is sentiment analysis, which involves quantifying the sentiment or tone expressed in textual data. Sentiment analysis can be performed using various techniques, one of which is the Bag-of-Words (BoW) model.

The BoW model represents textual data as a collection of words, disregarding grammar and word order. Let's denote D as the set of all documents in our corpus, and T as the set of all unique terms (words) in D. We can represent each document d as a vector v_d, where v_d has a dimensionality equal to $||T|$, and each element v_d represents the frequency of term t in document d given in equation (1)

$$v_d = [v_{d1}, v_{d2}, \ldots., v_{d|T|}] \quad (1)$$

With the BoW representation, we can calculate the sentiment score of a document based on the sentiment associated with each term. Let's denote $S(t)$ as the sentiment score of term t, which can be derived from sentiment lexicons or machine learning models trained on labeled data. The sentiment score of document (Sd) can be calculated as the weighted sum of the sentiment scores of its constituent terms statedin equation (2)

$$S_d = \sum_{t \in T} v_{dt}.S(t) \quad (2)$$

This sentiment score can then be used as a feature in financial analysis models, such as stock price prediction or risk assessment. Another important aspect of financial analysis with text mining is topic modeling, which aims to discover latent topics or themes present in a collection of documents. One popular technique for topic modeling is Latent Dirichlet Allocation (LDA). LDA represents documents as mixtures of topics, where each topic is characterized by a distribution over terms. Let's denote K as the number of topics we want to extract. For each document d, we represent its topic distribution as θ_d, and for each topic k, we represent its term distribution as $\emptyset_k$ stated in equation (3) and equation (4)

$$\theta_d \sim Dirichlet(\alpha) \quad (3)$$

$$\emptyset_k \sim Dirichlet(\beta) \quad (4)$$

Using Bayesian inference, we can estimate the posterior distributions of θd and ϕk given the observed words in the documents defined in equation (5)

$$p(\theta_d, \emptyset_k | documents) \propto p(documents | \theta_d, \emptyset_k).p(\theta_d).p(\emptyset_k) \quad (5)$$

Once we have inferred the topic distributions for each document and the term distributions for each topic, we can interpret the topics and assign documents to relevant topics for further analysis. financial analysis with text mining involves leveraging techniques such as sentiment analysis and topic modeling to extract valuable insights from textual data, which can complement traditional quantitative analysis methods in finance.

4. Blockchain Financial Analysis with Text Mining

The intersection of blockchain technology, financial analysis, and text mining has garnered significant attention as a promising avenue for extracting insights from decentralized financial data. Blockchain, with its inherent transparency and immutability, provides a rich source of financial information through transaction records stored in distributed ledgers. Leveraging text mining techniques in this context enables the extraction of valuable insights from textual data associated with blockchain transactions, smart contracts, and decentralized applications. One key aspect of blockchain financial analysis with text mining is sentiment analysis, which involves quantifying the sentiment expressed in textual data associated with blockchain transactions or discussions in social media platforms and online forums. Sentiment analysis can be performed using various methods, including lexicon-based approaches or machine learning algorithms trained on labeled data. Let's denote $S(t)$ as the sentiment score of term t, derived from sentiment lexicons or machine learning models. The sentiment score of a document (Sd) can be calculated as the weighted sum of the sentiment scores of its constituent terms, similar to the Bag-of-Words model:

Additionally, topic modeling techniques such as Latent Dirichlet Allocation (LDA) can be employed to uncover latent topics or themes within blockchain data. LDA represents documents as mixtures of topics, each characterized by a distribution over terms. Through Bayesian inference, we can estimate the posterior distributions of topic distributions for each document and term distributions for each topic. These inferred distributions allow for the interpretation of topics and the assignment of documents to relevant topics. Sentiment analysis aims to quantify the sentiment or opinion expressed in textual data associated with blockchain transactions or discussions. One common approach to sentiment analysis is lexicon-based sentiment analysis, where each word in the text is assigned a sentiment score based on a predefined sentiment lexicon.

Let's denote the sentiment score of term t as $S(t)$, which represents the sentiment polarity (positive or negative) associated with that term. This sentiment score can be obtained from sentiment lexicons or machine learning models trained on labeled data. The sentiment score of a document d (Sd) can then be calculated as the weighted sum of the sentiment scores of its constituent terms. In the Bag-of-Words model, each document is represented as a vector $\boldsymbol{vd}$, where vdt represents the frequency of term t in document d. Topic modeling, particularly Latent Dirichlet Allocation (LDA), is another powerful technique used in text mining to uncover latent topics or themes within a corpus of documents. LDA represents each document as a mixture of topics, where each topic is characterized by a distribution over terms.

Algorithm 1: Financial Data Analysis

Input:
- Text data (collection of documents)
- Sentiment lexicon or sentiment scoring model
Output:
- Sentiment score for each document
Procedure:
1. Preprocess the text data:
  - Tokenize the documents into individual words
  - Remove stopwords, punctuation, and other noise
  - Convert words to lowercase
2. Calculate the Bag-of-Words representation for each document:
  - Initialize an empty dictionary to store the word frequencies for each document
  - For each document in the collection:
    - Initialize a new dictionary to store the word frequencies
    - For each word in the document:
      - Increment the frequency count in the dictionary
    - Store the dictionary of word frequencies for the document
3. Calculate the sentiment score for each document:
  - For each document in the collection:
    - Initialize the sentiment score to 0
    - For each word-frequency pair in the Bag-of-Words representation:
      - Lookup the sentiment score for the word from the lexicon or model
      - Multiply the sentiment score by the word frequency and add it to the document's sentiment score
    - Store the sentiment score for the document
4. Return the sentiment scores for all documents

5. Financial Analysis with Text Mining Statement Disclosure

Text mining techniques with financial analysis, particularly in the realm of statement disclosure, offers a powerful means to extract meaningful insights from textual data embedded within financial statements, annual reports, and regulatory filings. One crucial aspect of this integration lies in sentiment analysis, which involves quantifying the sentiment or tone expressed in the textual disclosures. One approach to sentiment analysis involves lexicon-based methods, where each word in the text is assigned a sentiment score based on predefined sentiment lexicons or dictionaries. Let's denote $S(t)$ as the sentiment score of term t, derived from these lexicons. The sentiment score of a document d (Sd) can then be computed as the weighted sum of the sentiment scores of its constituent terms. In the Bag-of-Words model, each document is represented as a vector $\boldsymbol{vd}$, where vdt represents the frequency of term t in document d. The preprocess the textual data, which involves steps such as tokenization, removing stopwords, and converting words to lowercase. We also calculate sentiment scores for each news article using sentiment lexicons or machine learning models. To train a machine learning model, such as a classification or regression model, to predict stock price movements based on the extracted features. Let's denote $\boldsymbol{y}$ as the vector of target labels (e.g., stock price movements) corresponding to each document y=[y1,y2,...,yn]

The model using a training set of labeled data, where we know the stock price movements corresponding to each news article.

Once the model is trained, we can use it to make predictions on new, unseen data. Let ^y^ represent the vector of predicted labels stated in equation (6)

$$\hat{y} = [\widehat{y_1}\ \widehat{y_2}, \dots\dots, \widehat{y_m}] \tag{6}$$

Finally, the performance of the predictive model using metrics such as accuracy, precision, recall, or mean squared error, depending on the task (classification or regression) and the specific objectives of the analysis.

6. Financial Data

Financial data for blockchain financial analysis with text mining encompasses a diverse range of information extracted from blockchain transactions, smart contracts, decentralized applications (dApps), and other decentralized platforms. Here's a breakdown of some key types of financial data involved:

Transaction Data: This includes details of transactions recorded on the blockchain, such as sender and recipient addresses, transaction amounts, timestamps, and transaction fees. Analyzing transaction data can provide insights into trading patterns, liquidity, and the flow of funds within the blockchain network.

Smart Contract Data: Smart contracts are self-executing contracts with the terms of the agreement directly written into code. Financial data related to smart contracts includes contract addresses, function calls, transaction interactions, and contract state changes. Analyzing smart contract data can help assess the adoption and usage of decentralized applications and identify potential risks associated with smart contract vulnerabilities.

Token Data: Tokens are digital assets created and managed on blockchains. Token data includes information about token issuance, transfers, balances, and tokenomics such as supply, distribution, and token holder activity. Analyzing token data can provide insights into token usage, circulation, and market dynamics within decentralized ecosystems.

Market Data: Market data from decentralized exchanges (DEXs) and token marketplaces includes trading volume, price movements, order book data, and liquidity metrics. Analyzing market data can help identify trends, market sentiment, and trading opportunities within decentralized financial markets.

Social Media and News Data: Financial analysis with text mining also involves analyzing textual data from social media platforms, news articles, forums, and blogs. This data includes discussions, sentiments, news events, and opinions related to blockchain projects, cryptocurrencies, and decentralized finance (DeFi). Analyzing social media and news data can provide valuable insights into market sentiment, investor behavior, and potential market-moving events.

On-Chain Metrics: On-chain metrics provide insights into blockchain network activity, including metrics such as network hash rate, block time, block size, and network difficulty. Analyzing on-chain metrics can help assess network health, security, and performance, as well as identify trends and anomalies that may impact financial markets.

The financial data for blockchain financial analysis with text mining encompasses a wide range of information sources, including transaction data, smart contract data, token data, market data, social media and news data, and on-chain metrics. By integrating and analyzing these diverse sources of data, analysts can gain valuable insights into decentralized financial ecosystems and make more informed decisions in blockchain-based markets.

Table 1: Financial Data

Type of Data	Description
Transaction Data	
Transaction ID	0x123456789abcdef...
From Address	0xAbC123...
To Address	0x456def...
Amount Transferred	10 ETH
Timestamp	2024-04-06 10:30:45
Transaction Fee	0.005 ETH
Smart Contract Data	
Contract Address	0x987654321fedcba...
Function Call	transfer(address _to, uint256 _value)
Transaction Interaction	Sent 100 tokens to 0xAbC123...
Contract State Change	Balance of 0xAbC123... updated to 1000 tokens
Token Data	
Token Symbol	XYZ
Token Name	Example Token
Token Decimals	18
Total Supply	1,000,000 XYZ
Circulating Supply	500,000 XYZ
Holder Activity	0xAbC123... holds 10,000 XYZ
Market Data	
Token Price	$10 per XYZ
Trading Volume	1,000 XYZ traded in the last 24 hours
Market Capitalization	$5,000,000 for XYZ
Liquidity Pool	10,000 ETH and 100,000 XYZ in the XYZ/ETH pair
Social Media and News Data	
Twitter Sentiment	70% positive sentiment for XYZ token
Reddit Discussions	Active discussions about XYZ token's partnership
News Articles	Recent news article about XYZ token's integration
On-Chain Metrics	
Network Hash Rate	100 TH/s
Block Time	15 seconds
Block Size	2 MB
Network Difficulty	1,000,000

Table 1 provides a comprehensive overview of financial data pertinent to blockchain analysis and text mining. The data encompasses various facets of blockchain transactions, smart contract interactions, token information, market dynamics, social sentiment, and on-chain metrics.

Transaction data includes essential details such as transaction IDs, sender and recipient addresses, the amount transferred, timestamps, and transaction fees. Smart contract data sheds light on contract addresses, function calls, transaction interactions, and changes in contract state. Token data furnishes insights into specific tokens, including symbols, names, decimals, total supply, circulating supply, and holder activity. Market data offers crucial information on token prices, trading volumes, market capitalization, and liquidity pools, providing a snapshot of token market performance. Social media and news data provide sentiment analysis and discussions around specific tokens, gauging community sentiment and market trends. Lastly, on-chain metrics, including network hash rate, block time, block size, and network difficulty, offer insights into the underlying blockchain network's health and performance.

7. Evaluation of Results

The simulation results offer a critical glimpse into the anticipated outcomes and performance of a system or model under various conditions. By meticulously crafting and executing simulated scenarios, researchers and practitioners can gain invaluable insights into the behavior, efficacy, and potential limitations of their proposed solutions or strategies. Through these results, one can discern trends, identify optimal configurations, and assess the robustness of the proposed approach in addressing real-world challenges.

Table 2: Financial Statement Analysis

Company Name	Disclosure Type	Sentiment Score	Topic
Company A	Annual Report	0.75	Financial Performance
Company A	Quarterly Earnings	-0.45	Revenue Projections
Company B	Annual Report	0.90	Compliance with Accounting Standards
Company B	Investor Presentation	0.60	Growth Strategies
Company C	Regulatory Filing	-0.25	Legal Proceedings
Company C	Sustainability Report	0.80	Environmental Impact
Company D	Annual Report	0.70	Corporate Governance
Company D	Quarterly Earnings	0.55	Cost Management
Company E	Regulatory Filing	-0.35	Market Risks
Company E	Investor Presentation	0.65	Product Innovation
Company F	Annual Report	0.85	Revenue Growth
Company F	Sustainability Report	0.40	Social Responsibility
Company G	Quarterly Earnings	-0.50	Profit Margins
Company G	Regulatory Filing	0.75	Legal Compliance
Company H	Annual Report	0.60	Financial Health
Company H	Investor Presentation	-0.20	Market Expansion
Company I	Regulatory Filing	0.70	Risk Management
Company I	Sustainability Report	0.90	Community Engagement
Company J	Annual Report	-0.30	Debt Levels
Company J	Quarterly Earnings	0.75	Revenue Diversification

Table 2 provides a comprehensive overview of financial statement analysis across multiple companies, detailing the sentiment scores and topics extracted from various disclosure types. Each row represents a specific company and the corresponding financial disclosure type, sentiment score, and topic identified. For instance, Company A's annual report received a sentiment score of 0.75, indicating a positive sentiment regarding financial performance, whereas their quarterly earnings disclosure received a sentiment score of -0.45, suggesting concerns or pessimism regarding revenue projections.

Similarly, Company B's annual report garnered a high sentiment score of 0.90, indicating strong compliance with accounting standards, while their investor presentation received a sentiment score of 0.60, reflecting positive sentiment towards growth strategies. On the other hand, Company C's regulatory filing received a sentiment score of -0.25, suggesting potential legal proceedings, whereas their sustainability report received a high sentiment score of 0.80, indicating a positive impact on environmental initiatives.

Table 3: Blockchain Model for the Financial Statement

Blockchain Data	Description
Smart Contract Address	0x123456789abcdef...
Function Call	getFinancialStatement(address _company)
Transaction Interaction	Retrieved financial statement for Company X
Contract State Change	Updated financial statement data
Token Symbol	XYZ
Token Name	Example Token

Token Decimals	18
Total Supply	1,000,000 XYZ
Circulating Supply	500,000 XYZ
Market Capitalization	$5,000,000 for XYZ
Trading Volume	1,000 XYZ traded in the last 24 hours
Token Price	$10 per XYZ
Liquidity Pool	10,000 ETH and 100,000 XYZ in the XYZ/ETH pair

Table 3 presents a detailed breakdown of the blockchain model designed for financial statement management, outlining various blockchain data components and their respective descriptions. The smart contract address, denoted as "0x123456789abcdef...", serves as the entry point for accessing financial statements within the blockchain network. Through the "getFinancialStatement" function call, users can retrieve the financial statement associated with a specific company address (_company). Upon interaction with the smart contract, the contract state is updated, reflecting any modifications or additions to the financial statement data.

Furthermore, the blockchain model incorporates token-related information, including the token symbol (XYZ), token name (Example Token), and token decimals (18). Details regarding the total supply (1,000,000 XYZ) and circulating supply (500,000 XYZ) provide insights into the token distribution and circulation within the blockchain ecosystem. Additionally, the market capitalization of the token, valued at $5,000,000 for XYZ, reflects the total market value of all outstanding tokens.

Trading volume data indicates the volume of tokens traded within the last 24 hours, with 1,000 XYZ tokens exchanged during this period. The token price, set at $10 per XYZ, represents the current market value of a single token. Finally, the liquidity pool, consisting of 10,000 ETH and 100,000 XYZ tokens in the XYZ/ETH pair, underscores the liquidity available for trading XYZ tokens against Ether (ETH) within decentralized exchanges.

Table 4: Financial Disclosure for the Financial Statement

Item	**Amount (USD)**
Revenue	$1,000,000
Cost of Goods Sold	$500,000
Gross Profit	$500,000
Operating Expenses	$300,000
Research and Development Expenses	$100,000
Sales and Marketing Expenses	$150,000
General and Administrative Expenses	$50,000
Operating Income	$200,000
Interest Expense	$20,000
Income Before Taxes	$180,000
Income Tax Expense	$50,000
Net Income	$130,000
Earnings Per Share (EPS)	$1.50

Table 4 outlines a detailed financial disclosure for the financial statement, presenting various key financial metrics and their corresponding amounts in US dollars. The revenue generated amounts to $1,000,000, while the cost of goods sold is reported at $500,000, resulting in a gross profit of $500,000. Operating expenses encompass a total of $300,000, including $100,000 for research and development expenses, $150,000 for sales and marketing expenses, and $50,000 for general and administrative expenses. As a result, the operating income is calculated as $200,000 before factoring in interest expenses, which total $20,000. After deducting interest expenses, the income before taxes stands at $180,000. Subsequently, an income tax expense of $50,000 is accounted for, resulting in a net income of $130,000. The earnings per share (EPS) metric, calculated by dividing the net income by the total number of outstanding shares, is reported as $1.50.

8. Conclusion

This paper has explored the intersection of financial analysis, text mining, and blockchain technology, offering valuable insights into the evolving landscape of financial statement disclosure and analysis. Through the utilization of advanced data analytics techniques, including text mining algorithms and blockchain-based financial models, researchers and practitioners can enhance the efficiency, transparency, and reliability of financial reporting processes. The integration of text mining allows for the extraction of valuable insights from unstructured financial data, enabling stakeholders to uncover hidden patterns, sentiments, and trends within financial statements. Moreover, the implementation of blockchain technology facilitates secure and immutable storage of financial data, fostering trust and transparency in financial transactions and disclosures. By leveraging these innovative technologies, financial analysts can streamline the analysis of financial statements, identify potential risks and opportunities, and make data-driven decisions with greater confidence. Furthermore, the automation and digitization of financial reporting processes enhance accuracy, reduce manual errors, and improve overall efficiency in financial statement disclosure and analysis. However, it is essential to acknowledge the challenges and limitations associated with the adoption of these technologies, including data privacy concerns, regulatory compliance, and the need for robust cybersecurity measures.

REFERENCES

1. Huang, B., Yao, X., Luo, Y., & Li, J. (2023). Improving financial distress prediction using textual sentiment of annual reports. Annals of Operations Research, 330(1), 457-484.
2. Zhu, X., Wang, Y., & Li, J. (2022). What drives reputational risk? Evidence from textual risk disclosures in financial statements. Humanities and Social Sciences Communications, 9(1), 1-15.
3. Mousa, G. A., Elamir, E. A., & Hussainey, K. (2022). Using machine learning methods to predict financial performance: Does disclosure tone matter?. International Journal of Disclosure and Governance, 19(1), 93-112.
4. Moreno, A. I., & Caminero, T. (2022). Application of text mining to the analysis of climate-related disclosures. International Review of Financial Analysis, 83, 102307.
5. Chen, T. K., Liao, H. H., Chen, G. D., Kang, W. H., & Lin, Y. C. (2023). Bankruptcy prediction using machine learning models with the text-based communicative value of annual reports. Expert Systems with Applications, 233, 120714.
6. Xiuguo, W., & Shengyong, D. (2022). An analysis on financial statement fraud detection for Chinese listed companies using deep learning. IEEE Access, 10, 22516-22532.
7. Huang, A. H., Wang, H., & Yang, Y. (2023). FinBERT: A large language model for extracting information from financial text. Contemporary Accounting Research, 40(2), 806-841.
8. Brown, S. V., Hinson, L. A., & Tucker, J. W. (2024). Financial statement adequacy and firms' MD&A disclosures. Contemporary Accounting Research, 41(1), 126-162.
9. 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.
10. Suzuki, M., Sakaji, H., Hirano, M., & Izumi, K. (2023). Constructing and analyzing domain-specific language model for financial text mining. Information Processing & Management, 60(2), 103194.
11. Duan, H. K., Vasarhelyi, M. A., Codesso, M., & Alzamil, Z. (2023). Enhancing the government accounting information systems using social media information: An application of text mining and machine learning. International Journal of Accounting Information Systems, 48, 100600.
12. Ashtiani, M. N., & Raahemi, B. (2023). News-based intelligent prediction of financial markets using text mining and machine learning: A systematic literature review. Expert Systems with Applications, 217, 119509.
13. Zhao, S., Xu, K., Wang, Z., Liang, C., Lu, W., & Chen, B. (2022). Financial distress prediction by combining sentiment tone features. Economic modelling, 106, 105709.
14. Aboud, A., & Robinson, B. (2022). Fraudulent financial reporting and data analytics: an explanatory study from Ireland. Accounting Research Journal, 35(1), 21-36.
15. Tóth, Á., Suta, A., & Szauter, F. (2022). Interrelation between the climate-related sustainability and the financial reporting disclosures of the European automotive industry. Clean Technologies and Environmental Policy, 24(1), 437-445.
16. Lin, W. C., Tsai, C. F., & Chen, H. (2022). Factors affecting text mining based stock prediction: Text feature representations, machine learning models, and news platforms. Applied Soft Computing, 130, 109673.
17. Lopez-Lira, A. (2023). Risk factors that matter: Textual analysis of risk disclosures for the cross-section of returns. Jacobs Levy Equity Management Center for Quantitative Financial Research Paper.
18. Fijałkowska, J., & Hadro, D. (2022). Risk information in non-financial disclosure. Risks, 10(1), 11.

19. Hadro, D., Fijałkowska, J., Daszyńska-Żygadło, K., Zumente, I., & Mjakuškina, S. (2022). What do stakeholders in the construction industry look for in non-financial disclosure and what do they get?. Meditari Accountancy Research, 30(3), 762-785.
20. Roeder, J., Palmer, M., & Muntermann, J. (2022). Data-driven decision-making in credit risk management: The information value of analyst reports. Decision Support Systems, 158, 113770.
21. Breijer, R., & Orij, R. P. (2022). The comparability of non-financial information: An exploration of the impact of the non-financial reporting directive (NFRD, 2014/95/EU). Accounting in Europe, 19(2), 332-361.
22. Wang, W., Dinh, J. V., Jones, K. S., Upadhyay, S., & Yang, J. (2023). Corporate diversity statements and employees' online DEI ratings: An unsupervised machine-learning text-mining analysis. Journal of Business and Psychology, 38(1), 45-61.