

Assessing the Impact of ESG-Related News on Stock Trading in the Indonesian Market: A Text Similarity Framework Approach

Okiriza Wibisono¹, Ali Akbar Septiandri², Reinhard Denis Najogie¹

¹Independent Researcher, ²University College London
okiriza.w@gmail.com, ali.septiandri.21@ucl.ac.uk, najomath@gmail.com

Abstract

Environmental, Social, and Governance (ESG) perspectives have become integral to corporate decision-making and investment, with global regulatory mandates for ESG disclosure. The reliability of ESG ratings, crucial for assessing corporate sustainability practices, is compromised by inconsistencies and discrepancies across and within rating agencies, casting doubt on their effectiveness in reflecting true ESG performance and impact on firm valuations. While there have been studies using ESG-related news articles to measure their effect on stock trading, none have studied the Indonesian stock market. To address this gap, we developed a text similarity framework to identify ESG-related news articles based on Sustainability Accounting Standards Board (SASB) Standards without the need for manual annotations. Using news articles from one of the prominent business media outlets in Indonesia and an event study method, we found that 17.9% out of 18,431 environment-related news are followed by increased stock trading on the firms mentioned in the news, compared to 16.0% on random-dates datasets of the same size and firm composition. This approach is intended as a simpler alternative to building an ESG-specific news labeling model or using third-party data providers, although further analyses may be required to evaluate its robustness.

Keywords: ESG, sentence embeddings, stock trading

1. Introduction

Environmental, social, and governance (ESG) perspectives have become one of the most prominent factors in corporate decisions and investment (Edmans and Kacperczyk, 2022). Many regulators worldwide have mandated public firms to disclose their ESG practices regularly. Firms are likewise concerned about the perception of their ESG practices as research has shown that it can drive firm value, e.g. through its impact on sales and market investment on firms. The number and size of investment funds which participate in the UN's Principles for Responsible Investment (PRI) also have expanded greatly, reaching more than \$112 trillion in assets under management by 2023 (Principles for Responsible Investment, 2023).

Despite the increased focus on ESG, the reliability of ESG ratings, the primary source of ESG information, remains under examination. Serious inconsistencies are documented relating to differences in data sources and methodologies across rating agencies, and also internally (i.e. across time within a rating agency) (Berg and Rigobon, 2019; Johnson, 2023; Temple-West, 2023). Such issues have raised concerns about the reliability of ESG ratings in capturing firms' sustainable practices and their subsequent impact on valuations.

As an alternative, recent research suggests that monitoring ESG-related news in the media can offer valuable insights into the impact of ESG practices on firms' market prices. However, identifying ESG-related news that is relevant and also potentially market-moving for a firm requires expertise in ESG

topics as well as time-consuming manual effort. Previous studies addressing this challenge have relied on third-party data providers, which often use automatic classifications based on machine learning to label news articles as positive or negative with respect to a firm's ESG performance.

In this study, we aim to quantify the effect of ESG-related news articles on the Indonesian market by using a text similarity framework and causal analysis, thus eliminating the need for manually annotating the articles. In doing so, we made two main contributions:

1. We collected a large dataset of 119k news articles pertinent to financial and economic issues from January 2016 to July 2023 with the corresponding stock ticker symbols to measure the effect of these articles on the Indonesian stock market (§3). Using Sentence-BERT (SBERT) fine-tuned on Indonesian text and the SASB Standards, we identified 11,920 environment-related articles that served as the basis for our analysis.
2. We found that 17.9% out of 18,431 news articles (adjusted by the number of firms mentioned per publication date, referred to as "firm-dates") are followed by increased stock trading on the firms mentioned in the news for five trading days after the observation period. This is higher than the 16.0% increase observed from a set of dummy datasets (§4). Furthermore, the percentage of the excess trading rose to 25.6% when we extended the number of trading days post-period to 20.

Building on our findings, we explore the implications of leveraging news articles to assess the impact of ESG factors on firms listed in the Indonesia Stock Exchange (IEX). Additionally, we propose four avenues for further research, emphasizing enhancements in model robustness, inclusion of social and governance aspects, analysis of news sentiment, and the exploration of varied datasets.

2. Related Work

2.1. Using ESG news to predict stock market

Several research have examined the significance of ESG publication and ESG news on firms' value. [Capelle-Blancard and Petit \(2017\)](#) used 126 thousand ESG news, covering 100 listed firms from Dow Jones Sector Titans indexes, to study the impact on stock price cumulative average abnormal returns (CAAR). The dataset was obtained from a third-party provider, including the positive/negative labels on the ESG news. They found, among other hypotheses, that negative ESG news is associated with a price decline of about 0.1%, while positive ESG news does not lead to a gain in value.

Similar to [Capelle-Blancard and Petit, 2017](#)), [Serafeim and Yoon \(2022\)](#) further studied the kinds of ESG news that result in market reaction using a third-party news dataset with positive/negative labels, comprising more than 100 thousand firm-date observations for 3,109 firms. However, they found that excess market-adjusted returns are more substantial for positive ESG news, for news related to social factors (as opposed to environmental or governance factors), and if there is more news on the same day. Pertinent to our research, the reaction is significant only when the news discusses financial material issues for a given industry – according to the standards published by the Sustainability Accounting Standards Board (SASB) ([IFRS Foundation, 2023](#)).

Our study is the first to utilize ESG news for predicting trends in the Indonesian stock market. While we identified studies confirming the impact of ESG factors on the Indonesian stock market, such as those by [Nareswari et al. \(2023\)](#) and [Lubis and Rokhim \(2021\)](#), none of these previous studies have employed Indonesian news articles as a measure of these factors.

2.2. Identifying ESG factors from news articles

One way to identify ESG factors from news articles involves training a classifier based on the articles' content. Previous studies have successfully applied this approach using articles not only in En-

glish but also in French and Chinese ([Chen et al., 2023](#); [Tseng et al., 2023](#); [Pontes et al., 2023](#); [Lee et al., 2023](#); [Wang et al., 2023](#); [Billert and Conrad, 2023](#); [Mashkin and Chersoni, 2023](#); [Glenn et al., 2023](#)). Another way involves extracting ESG-related terms, as demonstrated by [Sandwidi and Pallitharammal Mukkolakal \(2022\)](#). However, both approaches typically require costly human expert annotations.

To bypass the need for manual annotation, one could use Sentence-BERT (SBERT) ([Reimers and Gurevych, 2019](#)) to generate embeddings for the news articles and ESG standards such as SASB, comparing the similarity values between the two sets of embeddings. This method was employed by [Sen et al. \(2023\)](#) to analyze the internal sustainability efforts of major US companies by comparing text from online reviews with the United Nations (UN) Sustainable Development Goals (SDGs). A more directly relevant example to our study is the work by [Pontes et al. \(2023\)](#), which also used SBERT, but in their case, to train a support vector machine classifier using the embeddings as inputs. Our approach differs by directly comparing the embeddings from the news articles with the SASB standards and filtering out irrelevant matches, as described in the subsequent section.

3. Methodology

3.1. Data

News We collected Indonesian news articles from [Kontan.co.id](#)¹, one of Indonesia's largest news portals specializing in financial news. In total, we collected 119k articles, spanning January 2016 to July 2023. These articles often contain stock ticker symbols to identify firms listed in the Indonesia Stock Exchange (IDX).

Stock trading We used IDX trading data in the same timespan as the news articles, i.e. January 2016 to July 2023. The data we used is the daily trading volume for each firm in IDX. In addition, we also used trading data for each sector in IDX.

SASB Standards To anchor the text embeddings to (globally-recognized) ESG concepts, we utilized sustainability standards published by the Sustainability Accounting Standards Board (SASB) ([IFRS Foundation, 2023](#)), as previously done in the literature ([Taleb et al., 2020](#); [Consolandi et al., 2022](#)). In total, there are 77 industry standards grouped into 11 sectors; e.g. Coal Operations industry standard in the Extractives & Minerals Processing sector.

¹<https://www.kontan.co.id>

Each industry standard is associated with several ESG topics². We selected only environmental topics as the focus of our study as suggested by Sandwidi and Pallitharammal Mukkolakal (2022). For each environmental topic in each industry standard, we took the Topic Summary section, which describes relevant points about that topic for this industry. There are in total 244 industry-topic combinations. Since we were using an Indonesian news dataset, we translated the topic summaries into Indonesian with Google Translate API.

3.2. Identifying relevant news articles

We employed IndoSBERT (Diana and Khodra, 2023), an SBERT-based (Reimers and Gurevych, 2019) model fine-tuned on Indonesian texts, to generate embeddings of the news articles and the SASB Standards. SBERT is a framework derived from BERT (Devlin et al., 2019) designed to generate fixed-size sentence embeddings that can be compared using cosine similarity (Reimers and Gurevych, 2019). SBERT uses the siamese network architecture (Chicco, 2021) and a pooling operation on BERT’s output, facilitating the fixed-size embeddings generation. This approach enables the efficient comparison of sentence pairs of varying lengths (Reimers and Gurevych, 2019).

For each article, we identified the SASB environmental topic standard with the highest article-standard similarity score (Equation 1) to represent the relevance of article a to standard s :

$$\alpha_a = \max_s \text{sim}(v_a, v_s) \quad (1)$$

where $\text{sim}(v_a, v_s)$ is the cosine similarity between the embeddings of article v_a and the embeddings of standard v_s .

We then selected the 90th percentile as the threshold for identifying relevant articles to ensure robustness against false positives, as suggested in a previous study (Chaturvedi et al., 2023). Given that each article was assigned a similarity value in the preceding step, this filtering is needed as the standard deemed most similar to a specific article might still be unrelated to it. Consequently, we ended up with 11,920 news articles focused on environmental topics, with an average of 1,490 articles annually (SD = 353).

3.3. Extracting firm mentions

We collected a comprehensive list of firms in the Indonesian stock market from IDX website³. This list,

²Each topic can be associated with multiple industry standards, although the topic’s summary within each industry standard can be different.

³<https://www.idx.co.id/id/data-pasar/data-saham/daftar-saham>

which contains 888 firms as of September 2023, serves as our reference for identifying relevant firms within news articles. We used both ticker symbols and full official firm names to search for firm mentions. We then extracted the firm mentions using a regular expression.

Additionally, we excluded irrelevant mentions by filtering lines containing phrases like “*Baca juga:*” (“Also read:”), “*Menarik dibaca:*” (“Interesting to read:”), and “*Selanjutnya:*” (“Further reading:”) within the articles. These phrases are typically links to other news articles, thus extracted firms mentioned in these lines are irrelevant.

3.4. Event study

To validate the usefulness of the text embeddings in identifying environment-related news articles, we employed a causal analysis of the impact of the filtered articles on stock trading using an event study. Specifically, we tested whether a firm’s stock trading volume increases for several days after environment-related news about that firm is published.

In our event study framework, we considered the publication date of a news article as the pivotal event from which we began observing changes in stock trading volumes. Given that an article may mention multiple firms, we associated each article with pairs of firms and publication dates, referred to as “firm-dates”. Each firm-date then represented a time series comprising stock trading volume and several control variables, allowing us to analyze periods before (pre-periods) and after (post-periods) the event. This methodology yielded a dataset of 18,431 firm-dates (18,431 time series), using days as the unit of measurement.

For each firm-date, we used the standardized stock trading volume (setting the mean to 1) from 66 trading days before the article publication date (pre-periods), which reflects approximately 3 months or 22 trading days per month. We then evaluated the effect of the environment-related news on the stock trading volume after 5, 10, and 20 trading days (post-periods).

To validate the observed effects, we incorporated the sectoral index price from the IDX as a control variable for each firm, alongside the daily return and volatility (= return²) of the sectoral index. This comprehensive approach ensured a thorough analysis of the news articles’ impact on stock trading dynamics.

Finally, we ran R CausalImpact library (Brodersen et al., 2015) with the previous specifications on each generated time series. From each output of CausalImpact, we extracted the p-value and the effect, which is the difference between the actual and predicted values of the main time series during

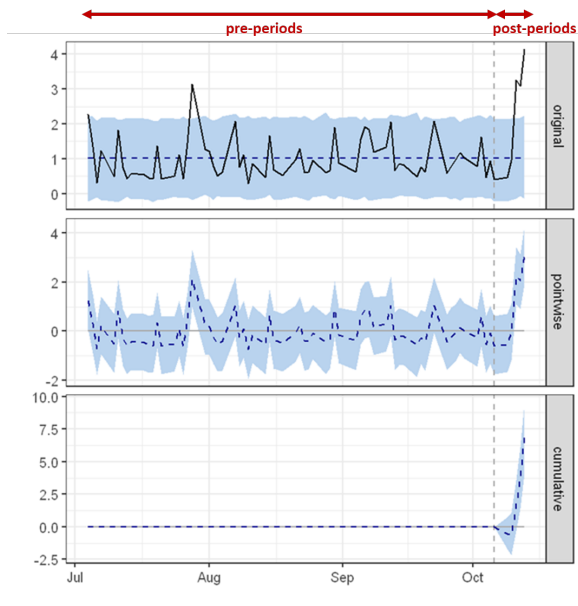


Figure 1: Example `CausalImpact` output for the trading volume of stock `IDX:JSMR` on 9 Oct 2017 (news similarity score to environment topic = 96th percentile). The dashed vertical line denotes the event/news date.

post-periods. If environment-related news significantly affects stock trading, then we would expect the resulting p -values to be low and the effects to be large and positive, i.e. since we tested on stock trading volume, a large and positive effect means that the news generates excess stock trades. An example output of the `CausalImpact` library for a given firm-date can be seen in Figure 1.

As an additional test, we ran the same analysis on 5 dummy datasets constructed by replacing dates in the original dataset with random dates from the same year, in which none of the top 10% environment-related news about the firm was published.

4. Results

The number of environment-related news (firm-dates) which result in a positive and significant effect (i.e. larger than predicted trading volume, $p < 0.05$) is tabulated in Table 1 and can be distilled into three main observations. First, the top 10% news articles most similar to environmental topics led to higher-than-predicted trading volumes for 17.9% of firm-dates over the next 5 trading days. Conversely, 9.7% of firm-dates experienced reduced trading (not detailed in Table 1), while for 72.4% of firm-dates, trading volumes did not significantly deviate from the pre-period levels. For context, control datasets (outlined in §3.4) revealed that $16.0\% \pm 0.4\%$ of firm-dates were followed by excess trading (t -statistic of difference in means

Table 1: Percentage of events followed by excess trading, by the number of trading days post-periods and the proportion of news articles sorted by the similarity to environment topic by SASB standards

Number of trading days in post-periods	Top 10% news	Top 5% news	Top 1% news
5	17.9%	18.2%	18.1%
10	21.3%	21.4%	21.9%
20	25.0%	25.2%	25.6%

Table 2: Percentage of events followed by excess trading per sector

IDX sector	N. firms	N. news	% events
Technology	32	328	24.4%
Consumer Cyclical	132	1375	20.0%
Cons. Non-Cyclical	103	2814	18.4%
Energy	75	3069	18.1%
Finance	98	2254	18.0%
Basic Materials	93	2278	17.9%
Industrials	51	845	17.5%
Infrastructure	129	4279	16.9%
Transportation	29	392	16.1%
Healthcare	26	797	16.1%
Total	768	18431	17.9%

= 6.69, $p < 0.0001$, $df = 25,593$), and $10.6\% \pm 0.2\%$ by reduced trading, indicating that our embedding methodology successfully identifies market-relevant environmental news associated with increased trading activity.

Second, the proportion of news articles followed by excess trading escalates with the length of the post-period: from 17.9% with a 5-day window to 21.3% for 10 days and 25.0% for 20 days. However, this trend may be influenced by other unaccounted events, which are more likely in longer post-periods.

Finally, increasing the topic similarity threshold by only including the top 5% or top 1% of news most similar to environmental topics does not significantly increase stock trading volume in the following days compared to the top 10%, with percentages inching from 17.9% to only 18.2% and 18.1% for 5, 10, and 20 post-periods, respectively.

In an additional analysis, we examined whether the effects vary across sectors. The results, summarized in Table 2 using IDX's firm classification, show that the significance of effects differs between sectors, ranging from 16.1% of firm-dates in the IDX Healthcare sector to 24.4% in IDX Technology.

5. Conclusion

In this paper, we have described how using SBERT to generate embeddings of sustainability standards can be used to identify ESG environment-related news. We tested our framework on an Indonesian news dataset between 2016 and 2023. The results show that the identified sustainability-related news

tends to be followed by increased stock trading volumes on firms mentioned in the news. Thus, this approach can be expanded upon as a simpler alternative to building an ESG news labeling model or using third-party data providers. Firms and regulators could also use such an approach to monitor ESG news that is potentially market-moving.

We acknowledged the limitations of this study and identified four ways to extend it. First, to explore long-term effects, one could examine the impact of ESG news on firm values using regression analysis. While we opted for an event study methodology based on `CausalImpact` library, conducting regression analysis with varying numbers of pre- and post-periods could enhance the results' robustness. Second, the scope of our study is limited to environment-related news; however, the methodology should be readily applicable to social and governance pillars of ESG as well. Third, future studies could investigate whether significant price movements are associated with the sentiments of the preceding ESG-related news articles. Finally, our approach was tested solely on a dataset from one news media outlet. Thus, future studies could include datasets from various news media outlets or other stock markets.

6. Bibliographical References

- Kölbl Julian F Berg, Florian and Roberto Rigobon. 2019. Aggregate confusion: The divergence of esg ratings. *Review of Finance*, 26(6):1315–1344.
- Fabian Billert and Stefan Conrad. 2023. [Team HHU at the FinNLP-2023 ML-ESG task: A multi-model approach to ESG-key-issue classification](#). In *Proceedings of the Fifth Workshop on Financial Technology and Natural Language Processing and the Second Multimodal AI For Financial Forecasting*, pages 146–150, Macao. -.
- Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik. 1992. [A training algorithm for optimal margin classifiers](#). COLT '92, page 144–152, New York, NY, USA. Association for Computing Machinery.
- Kay H. Brodersen et al. 2015. Inferring causal impact using bayesian structural time-series models. *Annals of Applied Statistics*, 9(1):247–274.
- Gunther Capelle-Blancard and Aure'lien Petit. 2017. Every little helps? esg news and stock market reaction. *Journal of Business Ethics*, 157:543–565.
- Sugat Chaturvedi, Ekaterina Prytkova, Tommaso Ciarli, and Önder Nomaler. 2023. [What is the future of automation? using semantic analysis to identify emerging technologies](#). Technical report.
- Chung-Chi Chen, Yu-Min Tseng, Juyeon Kang, Anaïs Lhuissier, Min-Yuh Day, Teng-Tsai Tu, and Hsin-Hsi Chen. 2023. [Multi-lingual ESG issue identification](#). In *Proceedings of the Fifth Workshop on Financial Technology and Natural Language Processing and the Second Multimodal AI For Financial Forecasting*, pages 111–115, Macao. -.
- Davide Chicco. 2021. Siamese neural networks: An overview. *Artificial neural networks*, pages 73–94.
- Costanza Consolandi, Robert G Eccles, and Giampaolo Gabbi. 2022. How material is a material issue? stock returns and the financial relevance and financial intensity of esg materiality. *Journal of Sustainable Finance & Investment*, 12(4):1045–1068.
- François Derrien, Philipp Krueger, Augustin Landier, and Tianhao Yao. 2021. [ESG News, Future Cash Flows, and Firm Value](#). Technical report.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Kadek Denaya Rahadika Diana and Masayu Leylia Khodra. 2023. [IndoSBERT: Enhancing Indonesian Sentence Embeddings with Siamese Networks Fine-tuning](#). In *2023 10th International Conference on Advanced Informatics: Concept, Theory and Application (ICAICTA)*. IEEE.
- Alex Edmans and Marcin Kacperczyk. 2022. Sustainable finance. *Review of Finance*, 26(6):1309–1313.
- Parker Glenn, Alolika Gon, Nikhil Kohli, Sihan Zha, Parag Pravin Dakle, and Preethi Raghavan. 2023. [Jetsons at the FinNLP-2023: Using synthetic data and transfer learning for multilingual ESG issue classification](#). In *Proceedings of the Fifth Workshop on Financial Technology and Natural Language Processing and the Second Multimodal AI For Financial Forecasting*, pages 133–139, Macao. -.

- IFRS Foundation. 2023. [SASB® Standards](#).
- Steve Johnson. 2023. [Hundreds of funds to be stripped of ESG rating](#).
- Hanwool Lee, Jonghyun Choi, Sohyeon Kwon, and Sungbum Jung. 2023. [EaSyGuide : ESG issue identification framework leveraging abilities of generative large language models](#). In *Proceedings of the Fifth Workshop on Financial Technology and Natural Language Processing and the Second Multimodal AI For Financial Forecasting*, pages 127–132, Macao. -.
- MFF Lubis and R Rokhim. 2021. [The effect of environmental, social, and governance \(ESG\) disclosure and competitive advantage on companies performance as an implementation of sustainable economic growth in Indonesia for period of 2015-2019](#). In *IOP Conference Series: Earth and Environmental Science*, volume 940, page 012059. IOP Publishing.
- Ivan Mashkin and Emmanuele Chersoni. 2023. [HKESG at the ML-ESG task: Exploring transformer representations for multilingual ESG issue identification](#). In *Proceedings of the Fifth Workshop on Financial Technology and Natural Language Processing and the Second Multimodal AI For Financial Forecasting*, pages 140–145, Macao. -.
- Ninditya Nareswari, Małgorzata Tarczyńska-Łuniewska, and Rizqi Umar Al Hashfi. 2023. [Analysis of Environmental, Social, and Governance Performance in Indonesia: Role of ESG on Corporate Performance](#). *Procedia Computer Science*, 225:1748–1756.
- Hilal Pataci, Yunyao Li, Yannis Katsis, Yada Zhu, and Lucian Popa. 2022. [Stock price volatility prediction: A case study with AutoML](#). In *Proceedings of the Fourth Workshop on Financial Technology and Natural Language Processing (FinNLP)*, pages 48–57, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.
- Elvys Linhares Pontes, Mohamed Benjannet, and Lam Kim Ming. 2023. [Leveraging BERT language models for multi-lingual ESG issue identification](#). In *Proceedings of the Fifth Workshop on Financial Technology and Natural Language Processing and the Second Multimodal AI For Financial Forecasting*, pages 121–126, Macao. -.
- Principles for Responsible Investment. 2023. [Data Portal](#).
- Nils Reimers and Iryna Gurevych. 2019. [Sentencebert: Sentence embeddings using siamese bert-networks](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.
- Blaise Sandwidi and Suneer Pallitharammal Mukkolakal. 2022. [Transformers-based approach for a sustainability term-based sentiment analysis \(STBSA\)](#). In *Proceedings of the Second Workshop on NLP for Positive Impact (NLP4PI)*, pages 157–170, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.
- Indira Sen, Daniele Quercia, Licia Capra, Matteo Montecchi, and Sanja Šćepanović. 2023. [Insider stories: analyzing internal sustainability efforts of major US companies from online reviews](#). *Humanities and Social Sciences Communications*, 10(1).
- George Serafeim and Aaron Yoon. 2022. [Which Corporate ESG News Does the Market React To?](#) *Financial Analysts Journal*, 78(1):59–78.
- Walid Taleb, Théo Le Guenedal, Frédéric Lepetit, Vincent Mortier, Takaya Sekine, and Lauren Stagnol. 2020. [Corporate ESG news and the stock market](#). Available at SSRN 3723799.
- Patrick Temple-West. 2023. [Sp drops esg scores from debt ratings amid scrutiny](#).
- Yu-Min Tseng, Chung-Chi Chen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2023. [DynamicESG: A Dataset for Dynamically Unearthing ESG Ratings from News Articles](#). In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pages 5412–5416.
- Weiwei Wang, Wenyang Wei, Qingyuan Song, and Yansong Wang. 2023. [Leveraging contrastive learning with BERT for ESG issue identification](#). In *Proceedings of the Fifth Workshop on Financial Technology and Natural Language Processing and the Second Multimodal AI For Financial Forecasting*, pages 116–120, Macao. -.