Text-Based Corporate Default Prediction with BERT Models

Tae Seon Yoon[1] (SKKU) and Byung Hwa Lim[2] (SKKU)

2024년 한국금융공학회 추계학술대회
2024.11.29

1) Department of FinTech, Sungkyunkwan University, E-mail: tjs1244@g.skku.edu
2) Department of FinTech, Sungkyunkwan University, E-mail: limbh@skku.edu

Contents

1. Introduction

- **Adopting machine learning models in credit risk modeling (predicting corporate defaults) widely emerges!**

  - Accurate prediction in corporate credit risk is considered **especially important in fields such as risk management, corporate lending, and credit rating (scoring).**

  - **Machine learning-based models** have demonstrated **exceptional performance** in addressing prediction problems, including asset pricing, fraud detection, bankruptcy prediction, etc.

  - Many financial institutions in the United States already leverage **AI techniques** to measure credit risk. In Korea, institutional frameworks are also being established **to support this approach**.

- As a result, the importance of predicting credit risk has been highlighted in the field of credit risk management, raising significant questions **about *how to apply these techniques effectively***.

1. Introduction

- **Alternative data is necessary for the efficient prediction of a corporation's financial distress.**
    - Existing machine learning-based models are primarily financial data-driven, **relying on structured data** such as accounting, market, and macroeconomic indicators.
    - However, such numerical data alone **cannot fully represent a company's situation.**
        - There may also be environments or periods where such variables do not perform effectively
    - It is necessary to incorporate **unstructured or non-financial data into the models** (multi-modal).
        - **Diversity in form and source, sophisticated methods for utilizing new types** of data
- In this study, we have focused **on leveraging textual data** in the context of **corporate default prediction**.

1. Introduction – Motivation

Research Questions

1. Is it possible to predict corporate bankruptcy using text-based data?
2. Can we enhanced model performance by mitigating data imbalance?
3. Can the performance of corporate default prediction be improved through a multimodal model?

- How to apply machine learning effectively to a ***firm's default prediction***
  - Listed companies in U.S. stock markets
  - Alternative data – textual data (MD&A)
  - Synthetic data – mitigate the data imbalance using Gen AI (ChatGPT)
  - Robust tests for subsample analysis (mid-cap companies)

1. Introduction – Alternative data and Gen AI

1. **MD&A** (Management's Discussion & Analysis) textual data.
   - ✓ SEC Regulation S-K requires firms to discuss their *liquidity needs and sources* in the MD&A section.
   - ✓ Reliability of being subject to accounting standards
   - ✓ Compatibility with existing datasets

2. **Gen AI using ChatGPT** to mitigate the imbalanced data problem
   - ✓ The imbalance data is the typical challenge faced by data-driven credit model
   - ✓ Using ChatGPT (3.5), we generate synthetic MD&A data to reduce the imbalance problem.
   - ✓ Also, EDA and focal loss function

3. **Multimodal model**
   - ✓ Examining a multimodal model *combining traditional financial accounting data with textual data* can achieve better predictive performance than existing models.

Contents

2. Literature Review

- **<u>Financial data-driven credit modeling</u> (Machine learning based models)**
  - ✓ Barboza et al. (2017): machine learning models offer a 10% improvement in predictive accuracy.
  - ✓ Kim et al. (2021): Focusing on sequential data could improve performance over traditional cross-sectional data.
  - ✓ Korangi et al. (2023): Multimodal models(accounting, market, marco data) based on sequential data.
  - ✓ Veganzones et al. (2018): Significant performance drops can occur in datasets with an imbalance ratio below 20% (SMOTE technique is the most effective solution)
- However, a significant limitation in applying these machine-learning techniques is the **heavy reliance on structured data,** such as market and accounting data.
- Furthermore, the data augmentation methodologies proposed in existing studies are **challenging to apply to text data such as MD&A**.

2. Literature Review

- **Text-based credit modeling (Focus on corporate disclosure)**

  - ✓ Mai et al. (2019): Embed MDA data and utilize word2vec for the embedding.
  - ✓ Chen et al. (2023): Utilize the characteristics of 10-K text as variables (frequency of specific words, file size, and the number of sentences)
  - ✓ Kim et al. (2023): BERT to extract sentiment

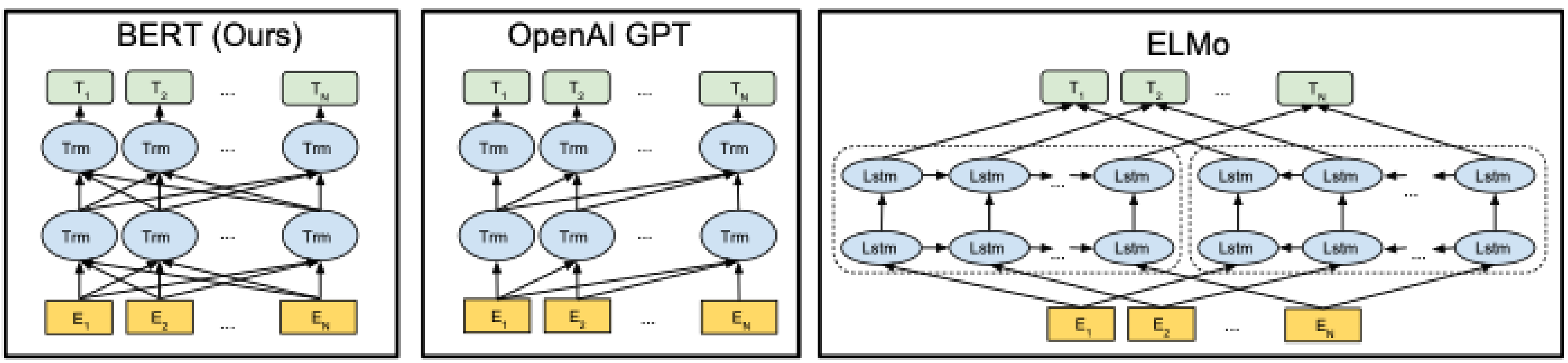
✓ These studies, however, are based **on only a portion of the information** contained in text data (e.g., sentiment analysis) and fail to fully incorporate the specific circumstances of individual companies into the predictions.

✓ Moreover, the studies neither address the **issue of data imbalance** nor clearly articulate the **value of multimodal models** that leverage text data in the prediction process.

Contents

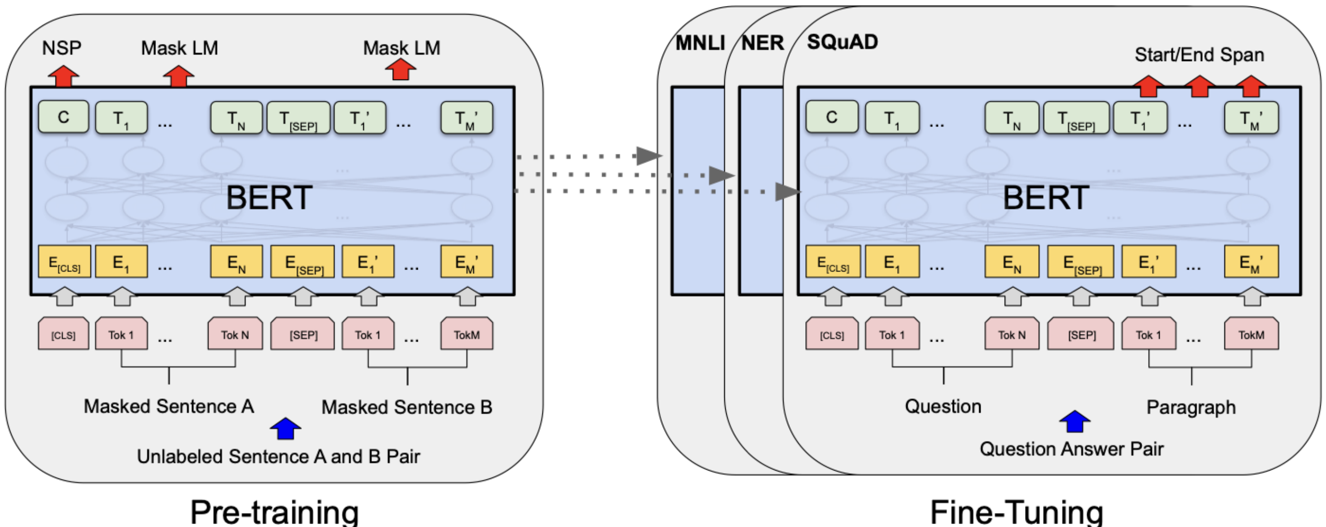
3. Models - Embedding

- **BERT (Bidirectional Encoder Representations from Transformers)**
  - ✓ BERT is a model developed by Google (Devlin et al., 2018) as a framework for **NLP tasks**.
  - ✓ Unlike Word2Vec and GloVe, BERT is designed **to read text bidirectionally**
  - ✓ **Transformer: self-attention mechanism**



Source: Devil et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," 2018

3. Models - Embedding



Source: Devil et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," 2018

- **FinBERT** (For finance domain)

  1. FinBERT by **ProsusAI** (Araci et al., 2019): Use a large financial corpus (**Financial PhraseBank** by Malo et al., 2014, containing 4,841 sentences) and fine-tuned for financial sentiment classification.

  2. FinBERT-tone by **Yiyanghkust** (Huang et al., 2023): Fine-tuned on 10,000 manually annotated sentences from **analyst reports**.

3. Models - Prediction models

- A deep learning architecture to process **sequential and high-dimensional** data effectively

1. **LSTM (Long Short-Term Memory)**
   - ✓ A type of recurrent neural network (RNN) structure designed to capture long-term dependencies in sequential data effectively.

2. **GRU (Gated Recurrent Unit)**
   - ✓ Simplified version of LSTM that enables more efficient computation.

3. **Transformer**
   - ✓ Considers the entire sequence at once, regardless of the sequence position, and operates based on a self-attention mechanism.

4. **TCN (Temporal Convolutional Network)**
   - ✓ Models sequential data using convolutional neural networks.

5. **DNN (Deep Neural Network)**
   - ✓ The sequence information is compressed into a **single sequence by averaging** the individual elements.
   - ✓ A model that **only uses the data from the last quarter** (annual text data) will also be trained to compare with other models.

- All models in our study have a **single hidden layer.**

Contents

4. Data

- We explore the prediction of default risk **for publicly traded U.S.** companies.
  - ✓ **Financial industry** firms are excluded.
  - ✓ **Quarterly** data is collected.

- In this study, default is defined as instances where a company undergoes **bankruptcy or liquidation**, as provided by the **Compustat** database.
  - ✓ To identify these cases, we use the **"DLRSN"** (Research Company Reason for Deletion), selecting cases where DLRSN is 02 (Bankruptcy) or 03 (Liquidation).
  - ✓ Also, we used the **asset footnote codes** of TL or GL to identify companies in bankruptcy or liquidation.
  - ✓ Specifically, we focused on companies that experienced bankruptcy **within a one-year period**.
  - ✓ Predicting defaults immediately after a quarter would offer **limited value** in terms of **implications**.

4. Data

- We download 10-K and 10-Q filings in text format from the **SRAF** (Software Repository for Accounting and Finance) for the years 1993–2023.
  - ✓ Using filing patterns, we employ Python to **extract the MD&A** section.
  - ✓ To maximize data on defaulted companies, **some cases are manually collected.**
  - ✓ We have removed special characters, line breaks, and numbers

	Character Count		Word Count	
	Non-bankrupt	Bankrupt	Non-bankrupt	Bankrupt
count	370824	3868	370824	3868
mean	51137.7	44656.82	7893.89	6945.14
std	55276.47	47133.98	8503.04	7303.18
min	1202	4063	185	208
25%	19223	15228.75	2991	2382.75
50%	34350	29559	5318	4620.5
75%	59677	54061.75	9207	8386.75
max	486670.61	250735.3	75520.16	38548.27

4. Data

No.	Variables	Description
1	(EBIT+DP)/AT	EBIT + Amortization & Depreciation / Total Asset
2	(LCT-CH)/AT	(Current Liabilities - Cash) / Total Asset
3	ACTLCT	Current Assets / Current Liabilities
4	APSALE	Accounts Payable / Sales
5	CASHAT	Cash and Short-term Investment / Total Assets
6	CASHMTA	Cash & E / (Market Equity + Total Liabilities)
7	CHAT	Cash / Total Assets
8	CHLCT	Cash / Current Liabilities
9	EBITAT	Earnings Before Interest and Tax / Total Assets
10	EBITSALE	Earnings Before Interest and Tax / Sales
11	FAT	Total Debts / Total Assets
12	INVTSALE	Inventories / Sales
13	LCTAT	Current Liabilities / Total Asset
14	LCTLT	Current Liabilities / Total Liabilities
15	LCTSALE	Current Liabilities / Sales
16	LOG(AT)	Log of Total Assets
17	LOG(SALE)	Log of Sales
18	LTAT	Total Liabilities / Total Assets
19	LTMTA	Total Liabilities / Market Equity + Total Liabilities
20	NIAT	Net Income / Total Asset
21	NIMTA	Net Income / Market Equity + Total Liabilities
22	NISALE	Net Income / Sales
23	OIADPAT	Operating Income / Total Asset
24	OIADPSALE	Operating Income / Sales
25	QALCT	Quick Assets / Current Liabilities
26	REAT	Retained Earnings / Total Asset
27	RELCT	Retained Earnings / Current Liabilities
28	RSIZE	Log of Market Capitalization
29	SALEAT	Sales / Total Assets
30	SEQAT	Equity / Total Asset
31	WCAPAT	Working Capital / Total Assets
32	MB	Market-to-Book Ratio
33	MC	Market Capiatalization

- The financial ratio for constructing the **multimodal model** is organized as follows.

✓ In their study on corporate default prediction, **Mai et al. (2019)** compile a set of financial ratios.

✓ We use **33** of these financial ratios in our research.

✓ We performed **standardization** on the financial ratios.

- Missing values are replaced with the **industry median**.

4. Data

cik	conm	fyearq	fqtr
320193	APPLE INC	2021	1
320193	APPLE INC	2021	2
320193	APPLE INC	2021	3
320193	APPLE INC	2021	4
320193	APPLE INC	2022	1
320193	APPLE INC	2022	2
320193	APPLE INC	2022	3
320193	APPLE INC	2022	4
320193	APPLE INC	2023	1
320193	APPLE INC	2023	2
320193	APPLE INC	2023	3
320193	APPLE INC	2023	4

⇒ Sample 1

⇒ Sample 2

⇒ Sample 3

YEAR	N	Non-bankrupt	Bankrupt	YEAR	N	Non-bankrupt	Bankrupt
1993	4	3	1	2009	3477	3451	26
1994	408	401	7	2010	3335	3316	19
1995	1270	1259	11	2011	3330	3308	22
1996	2614	2571	43	2012	3534	3516	18
1997	3471	3412	59	2013	3629	3611	18
1998	3620	3557	63	2014	3533	3514	19
1999	3608	3556	52	2015	3331	3304	27
2000	3790	3729	61	2016	3224	3201	23
2001	3856	3780	76	2017	3083	3066	17
2002	3854	3787	67	2018	3150	3124	26
2003	3649	3621	28	2019	3190	3161	29
2004	3335	3300	35	2020	3100	3092	8
2005	3209	3190	19	2021	3346	3320	26
2006	3117	3098	19	2022	3494	3462	32
2007	3187	3142	45	2023	360	360	0
2008	3565	3494	71	**SUM**	93673	92706	967

- We focus on **sequential data(panel)**.

✓ Each sample consists of **4 periods.**

✓ **window size=4, step=4**

Number of companies: 12,166 (CIK)

Years: 1993~2023

Default ratio: 1.032% (967 default/93,673 total samples)

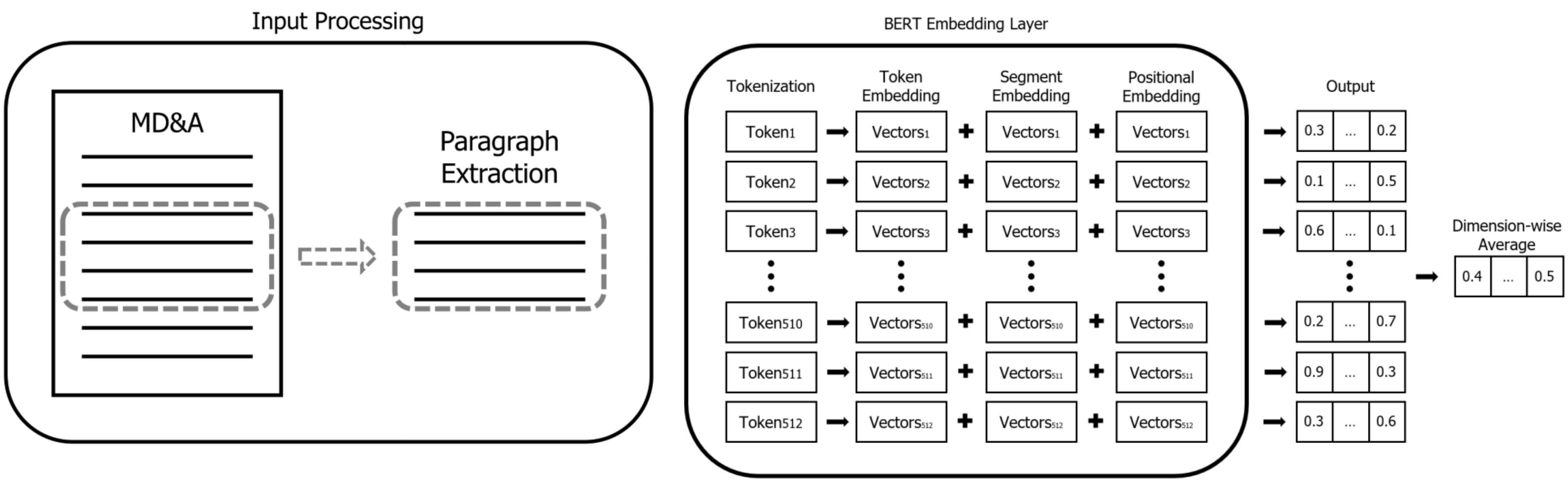
Total data points: 374,692 (93,673 samples*4 periods each)

Contents

5. Methodology - Preprocessing

- **First**, to accommodate the limited maximum token count of the BERT base model, we need to **extract a portion of the MDA** data.
  - ✓ Following Hoberg et al. (2015), we measure financial constraints by using the **frequency of synonyms** related to "financing" and "delay."
  - ✓ Therefore, we aim to extract the **3,000 characters** that contain the **highest frequency of relevant words**, with a step size of 500 characters. (approximately to 512 tokens)

Delay list: delay* OR abandon OR eliminate…
Equity-focused list: issuing equity securities OR expects equity securities…
Debt-focused list: increased borrowings OR use line of credit OR expanded borrowings…



5. Methodology – Synthetic Data

- The following **four approaches** are compared to address the issue of data imbalance:

1. ChatGPT(GPT-3.5-turbo):
   - ✓ ChatGPT is **a transformer-based conversational AI** model developed by OpenAI.
   - ✓ We apply method proposed by Dai et al. (2023), **rephrasing the text**.

2. Easy Data Augmentation (EDA):
   - ✓ We also augmented the data using the Easy Data Augmentation (EDA) method proposed by Wei et al. (2019).
   - ✓ EDA is a simple yet effective technique for data augmentation, which involves **synonym replacement(sr), random word deletion(rd), insertion(ri), and swapping(rs)**. (sr = 0.3, rd = 0.05, ri = 0.1, rs = 0.1)

3. GPT Combined with EDA:
   - ✓ To introduce **additional noise**, we combined GPT-generated data with EDA.

Please rephrase the following paragraph while strictly keeping the length of the text similar: {original MDA}

if our products are not perceived as being effective at reducing the risk of covid transmission or if covid is determined to spread in ways other than through airborne transmission if we do not successfully anticipate market needs...

→

Our business may suffer if our products are not perceived as effective in reducing the risk of covid transmission or if covid spreads in other ways than through airborne transmission Failure to anticipate market needs...

cik: 1872356 fyearq: 2022 fqtr: 4

Change in the proportion of bankrupt data:

GPT(1) and EDA(2): 1% > **21%**
(1 sample to 25 samples)

GPT+EDA(3): 1% > **44%**
(1 to 25 GPT × 3 EDA)

5. Methodology - Focal Loss

4. Focal loss function:

- ✓ Traditional cross-entropy loss, widely used in tasks like default prediction. This **approach may be less effective** in scenarios where **data imbalance** and sample simplicity are significant issues.
- ✓ The Focal Loss, introduced by Lin et al. (2017), addresses imbalanced classification by assigning **greater weight to hard-to-classify** cases while **reducing the impact of easily classified ones**.
- ✓ This **characteristic aligns closely** with the challenges present in our research data.
- ✓ Our goal is to investigate whether applying focal loss can effectively enhance the performance of corporate default prediction models by better handling the class imbalance.

Cross-Entropy Loss:

$$CE(p_t) = -\log(p_t)$$

Focal Loss:

$$FL(p_t) = -(1-p_t)^{\gamma}\log(p_t)$$

•p_t: The predicted probability for the true class label.

•γ: The focusing parameter, where γ >0 reduces the relative loss for well-classified examples, allowing the model to prioritize harder examples.

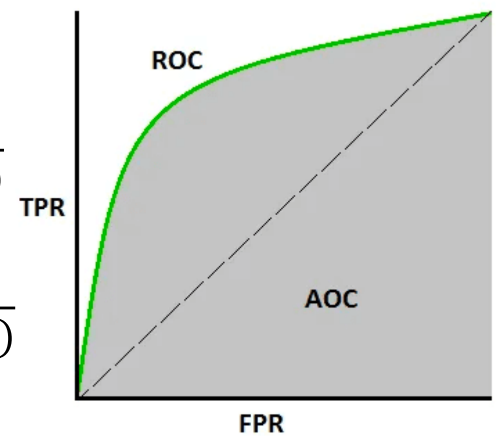
5. Methodology - Training

1. Training method:
   - ✓ We applied **early stopping**, **monitoring the AUC** of the validation set with a **patience of 10** epochs, and **restored the best weights** to evaluate the results.

2. Performance measure:
   - ✓ We use **AUC** (Area Under the Curve) to measure the performance of models.
   - ✓ The ROC curve represents the relationship between TPR and FPR, and AUC refers to the area under this curve.
   - ✓ AUC allows for comparing model performance on **imbalanced datasets**.
   - ✓ The AUC ranges from 0 to 1, where **1 represents perfect classification** performance, and 0.5 represents random guessing.

$$TPR(True\ Positive\ Rate) = \frac{True\ Positive\ (TP)}{True\ Positive(TP) + False\ Negatives(FN)}$$

$$FPR(False\ Positive\ Rate) = \frac{False\ Positive\ (FP)}{False\ Positive(FP) + True\ Negatives(TN)}$$



5. Methodology - Training

3. Multimodal model:

- Through our research, we aim to validate the utility of multimodal approaches, particularly in scenarios where predicting defaults using **accounting data alone proves challenging**.

- We hypothesize that multimodal models incorporating **MDA will provide greater value in such cases**.

  - ✓ First, during **financial crises**, corporate credit risk generally increases across the board.

  - ✓ Second, studies such as Bechworth et al. (2010), Lin et al. (2011), and Feldhutter et al. (2018) suggest that **credit risk for mid-cap** companies is influenced by **a relatively greater** number of factors, many of which exhibit highly nonlinear relationships.

  - ✓ In these cases, it is judged that predicting defaults is **relatively difficult**, and we aim to explore whether this issue can be addressed through multimodal approaches.

Before Financial Crisis: train(1993~2000), validation(2001~2003), test(2004~2006)
After Financial Crisis: train(1993~2003), validation(2004~2006), test(2007~2009)

Before Covid-19: train(1993~2014), validation(2015~2017), test(2018~2020)
After Covid-19: train(1993~2017), validation(2018~2020), test(2021~2023)

Contents

6. Result

Table 1. Results from the BERT Models with Random Split (6:2:2)

	LSTM	GRU	TF	TCN	DNN(avg)	DNN(4qtr)
BERT-base	0.761	0.763	0.745	0.756	0.757	0.723
FinBERT(2023)	**0.779**	**0.775**	**0.778**	**0.778**	**0.779**	**0.748**
FinBERT(2019)	0.764	0.761	0.742	0.766	0.752	0.727

- First, we confirmed that **defaults can be predicted** using **text data**.
- Among the three BERT models, **FinBERT-tone by Yiyanghkust (2023)** showed **slightly superior** performance compared to the others.
- Using the DNN (4qtr) model, we observed that applying **quarterly data** to the model yields **better performance** than using annual data.

6. Result

Table 2. Results from the FinBERT(2023) with Random Split (6:2:2)

	LSTM	GRU	TF	TCN	DNN(avg)
Basic	0.779	0.775	0.778	0.778	0.779
GPT	0.773	0.776	**0.786**	0.761	0.774
EDA	0.742	0.736	0.709	0.733	0.727
GPT+EDA	**0.791**	**0.791**	0.776	0.774	**0.782**
Focal	**0.788**	**0.779**	**0.784**	**0.785**	**0.780**

- Ultimately, **not all methods** resulted in significant performance improvements.
  - ✓ We speculate that this issue arises because most existing studies on text data augmentation **focus on short texts**.
- The **GPT+EDA** method, which introduced the most noise, showed some potential, but its impact was **limited**.
- However, the application of **focal loss** consistently demonstrated better results across all methods, highlighting its effectiveness.

6. Result

Table 3. Results from the Accounting & Mutimodal model with year split

		LSTM	GRU	TF	TCN
accounting	Before Crisis	0.847	0.844	0.853	0.834
	After Crisis	0.827	0.810	0.825	0.815
multimodal	Before Crisis	0.850	0.868	0.853	0.858
	After Crisis	0.862	0.855	0.856	0.845
accounting	Before Covid	0.883	0.892	0.862	0.890
	After Covid	0.819	0.783	0.812	0.803
multimodal	Before Covid	0.882	0.875	0.861	0.885
	After Covid	0.869	0.860	0.846	0.836

- We observe that, for both the 2008 crisis and COVID-19, the predictive performance of the models using accounting data **decreases after the crisis**.

- However, the improvement seen in the **multimodal model is more pronounced.**
  - ✓ This suggesting that, in situations like a crisis where the overall credit risk increases and accounting data **fails to perform effectively.**
  - ✓ Text data can be particularly **useful and more effectively** leveraged.

6. Result

Table 4. Results from the Accounting & Mutimodal model across cap-sizes

		LSTM	GRU	TF	TCN
accounting	All sizes	0.882	0.875	0.861	0.885
	Mid-Cap	0.862	0.867	0.843	0.866
multimodal	All sizes	0.883	0.892	0.862	0.890
	Mid-Cap	0.886	0.877	0.859	0.882

- We compare all sizes with under mid-cap. To **exclude the impact of the crisis**, we focus on the period **before COVID-19**.
  - ✓ As seen in previous studies on mid-cap credit risk, we find that predicting defaults becomes **slightly more difficult for mid-cap** companies.
  - ✓ The performance improvement through the multimodal approach appears to be **more significant for mid-cap** companies.

Contents

7. Conclusion

- This study used Fin-BERT to embed MD&A (Management Discussion and Analysis) text data to predict corporate defaults.
  - We confirm that a corporation's defaults can indeed be predicted with textual information.

- To address the **issue of data imbalance** within MD&A data, we applied **data augmentation techniques using ChatGPT** and a focal loss function, leading to noticeable performance improvements.
  - Synthetic data using ChatGPT

- We showed that a **multimodal model** incorporating text data can effectively supplement and enhance predictions.
  - This trend was especially noticeable after COVID-19 and mid-cap firms

- <u>Discussions</u>
  - ✓ (Longer texts) We can apply the advanced capabilities of **GPT-4 or higher version** to handle more complex instructions and generate better-augmented data for longer texts.
  - ✓ (Longer data sequence) In this study, the dataset is limited to only four sequences. We can manually collect **longer data**.
  - ✓ (Longer extraction) The basic BERT model is limited to 512 tokens. **BERT-large** or **Longformer** can be applied for a larger number of tokens for better extraction methods.

Reference

[1] Araci, Dogu. 2019. "FinBERT: Financial Sentiment Analysis with Pre-trained Language Models." arXiv Preprint, arXiv:1908.10063. https://arxiv.org/abs/1908.10063.

[2] Bai, Shaojie, J. Zico Kolter, and Vladlen Koltun. 2018. "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling." arXiv Preprint, arXiv:1803.01271. https://arxiv.org/abs/1803.01271.

[3] Barboza, Flavio, Herbert Kimura, and Edward Altman. 2017. "Machine Learning Models and Bankruptcy Prediction." Expert Systems with Applications 83: 405-417. https://doi.org/10.1016/j.eswa.2017.04.006.

[4] Beckworth, David, Kyoung Philip Moon, and Jeremy H. Toles. 2010. "Monetary Policy and Corporate Bond Yield Spreads." Applied Economics Letters 17 (12): 1139-1144. https://doi.org/10.1080/00036840902845368.

[5] Chen, Tsung-Kang, Hsien-Hsing Liao, Geng-Dao Chen, Wei-Han Kang, and Yu-Chun Lin. 2023. "Bankruptcy Prediction Using Machine Learning Models with the Text-Based Communicative Value of Annual Reports." Expert Systems with Applications 233: 120714. https://doi.org/10.1016/j.eswa.2023.120714.

[6] Cho, Kyunghyun, Bart van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. "On the Properties of Neural Machine Translation: Encoder-Decoder Approaches." arXiv Preprint, arXiv:1409.1259. https://arxiv.org/abs/1409.1259.

[7] Dai, Haixing, Zhengliang Liu, Wenxiong Liao, Xiaoke Huang, Yihan Cao, Zihao Wu, Lin Zhao, Shaochen Xu, Wei Liu, Ninghao Liu, Sheng Li, Dajiang Zhu, Hongmin Cai, Lichao Sun, Quanzheng Li, Dinggang Shen, Tianming Liu, and Xiang Li. 2023. "AugGPT: Leveraging ChatGPT for Text Data Augmentation." arXiv Preprint, arXiv:2302.13007. https://arxiv.org/abs/2302.13007.

[8] Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." arXiv Preprint, arXiv:1810.04805. https://arxiv.org/abs/1810.04805.

[9] Feldhütter, Peter, and Stephen M. Schaefer. 2018. "The Myth of the Credit Spread Puzzle." The Review of Financial Studies 31 (8): 2897-2942. https://doi.org/10.1093/rfs/hhy032.

[10] Hochreiter, Sepp, and Jürgen Schmidhuber. 1997. "Long Short-Term Memory." Neural Computation 9 (8): 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735.

[11] Hoberg, Gerard, and Vojislav Maksimovic. 2015. "Redefining Financial Constraints: A Text-Based Analysis." The Review of Financial Studies 28 (5): 1312-1352. https://doi.org/10.1093/rfs/hhu089.

[12] Huang, Allen H., Huan Wang, and Yang Yang. 2023. "FinBERT: A Large Language Model for Extracting Information from Financial Text." Contemporary Accounting Research 40: 806-841. https://doi.org/10.1111/1911-3846.12832.

[13] Kim, Alex, and Sangwon Yoon. 2023. "Corporate Bankruptcy Prediction with Domain-Adapted BERT." arXiv Preprint, arXiv:2312.03194. https://arxiv.org/abs/2312.03194.

[14] Kim, Hyeongjun, Hoon Cho, and Doojin Ryu. 2021. "Corporate Bankruptcy Prediction Using Machine Learning Methodologies with a Focus on Sequential Data." Computational Economics 59: 1213-1240. https://doi.org/10.1007/s10614-021-10126-5.

[15] Korangi, Kamesh, Christophe Mues, and Cristián Bravo. 2023. "A Transformer-Based Model for Default Prediction in Mid-Cap Corporate Markets." European Journal of Operational Research 308 (1): 306-320. https://doi.org/10.1016/j.ejor.2022.10.032.

[16] Lin, Hai, Junbo Wang, and Chunchi Wu. 2011. "Liquidity Risk and Expected Corporate Bond Returns." Journal of Financial Economics 99 (3): 628-650. https://doi.org/10.1016/j.jfineco.2010.10.004.

[17] Mai, Feng, Shaonan Tian, Chihoon Lee, and Ling Ma. 2019. "Deep Learning Models for Bankruptcy Prediction Using Textual Disclosures." European Journal of Operational Research 274 (2): 743-758. https://doi.org/10.1016/j.ejor.2018.10.024.

[18] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. "Attention Is All You Need." Advances in Neural Information Processing Systems (NeurIPS) 30: 5998-6008. https://arxiv.org/abs/1706.03762.

[19] Veganzones, David, and Eric Séverin. 2018. "An Investigation of Bankruptcy Prediction in Imbalanced Datasets." Decision Support Systems 112: 111-124. https://doi.org/10.1016/j.dss.2018.06.011.

[20] Wei, Jason, and Kai Zou. 2019. "EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks." arXiv Preprint, arXiv:1901.11196. https://arxiv.org/abs/1901.11196.

[21] Lin, Tsung-Yi, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. 2017. "Focal Loss for Dense Object Detection." Proceedings of the IEEE International Conference on Computer Vision (ICCV): 2980-2988. https://doi.org/10.1109/ICCV.2017.324.

Thank you.