Can publicly available information detect deteriorating credit health?

Chan Yin Tsung Lawrence 3035712169

Li Hao (Fiona) 3035666308

Chiu Tsz Chun (Toby) 3035712195

Lu Yuzhen Steven 3035713943

Wan Man Lau Lawrence 3035666396

Could we have predicted credit rating downgrades?

LEHMAN BROTHERS

中國恆大集團
CHINA EVERGRANDE GROUP

中国华融资产管理股份有限公司
CHINA HUARONG ASSET MANAGEMENT CO., LTD.

Business/external factors

adani

luckin coffee

Fraud

Could we have predicted credit rating downgrades?

LEHMAN BROTHERS

中國恆大集團
CHINA EVERGRANDE GROUP

中国华融资产管理股份有限公司
CHINA HUARONG ASSET MANAGEMENT CO., LTD.

Business/external factors

Key Questions

1. Were there signals present that predicted each credit rating downgrade?
2. Were these signals present in publicly available information?

Could we have predicted credit rating downgrades?

Why is this important (our interest)

Academic Interest (backward looking) **with** **Profit Implications (forward looking)**

Business/external factors

Key Questions

1 Were there signals present that predicted each credit rating downgrade?

2 Were these signals present in publicly available information?

Our Methodology: Bag-of-Words Model

Data Sources

Outlook/downgrade reports from Moody/Fitch/S&P

Company filings, such as annual reports

BoW Model

- Create a timeline of changes in credit health and label the reports respectively
- Consider the rationales/driving factors for downgrades and compile bag of words that indicate potential risks of downgrading

- Pre-process the filings using NLP techniques, such as removing stop words
- Count the frequency of the bag of words to predict the probability of downgrade
- Adjust the bag of words to improve the accuracy

Limitations

It might be difficult to generalize a bag of words that fit all sectors and all geographies, and we may consider narrowing down the scope of our data source

Technical Implementation

Methodology	Technical Implementation
1 Scrape relevant credit downgrade reports from Moody/Fitch/S&P	Utilise **Requests** and **BeautifulSoup** as web scraping package given little JS. **RoboBrowser** can be used to skip signup details on Moody's database
2 Pre-process filings using basic cleaning techniques	**Removing stop words:** excluding low-value terms (e.g. "the", "a", etc.) **Treating "Special Words":** manually tagging phrases (e.g. "joint ventures")
3 Create Bag-of-Words (BoW) model	**Bag-of-Words** model can count frequency of words in downgrade reports; over a large sample size, we can create a generalised word frequency model
4 Adapt existing model through NLP techniques to create better fit	**N-grams:** Analysing data based on combination of words **Sentiment**: Assigning sentimentality to prioritise deeply negative terms
5 Utilise existing model to analyse relationships and draw insights	Data can be used in **Pandas** to provide probabilistic predictions – establishing a connection between qualitative financial data and credit health

Expected Outcome and Challenges

Expected Outcome

- Reading select annual reports prior to bond downgrading
- Research the timeline of notable bond downgrades
- Highlighted notable bond downgrade for isolated analysis

We believe there is a relationship between wordings of annual reports and bond downgrades

Challenges

1

Data robustness

- Wording in reports can be drastically different due to independent writers have different types

2

Complexity in bond rating

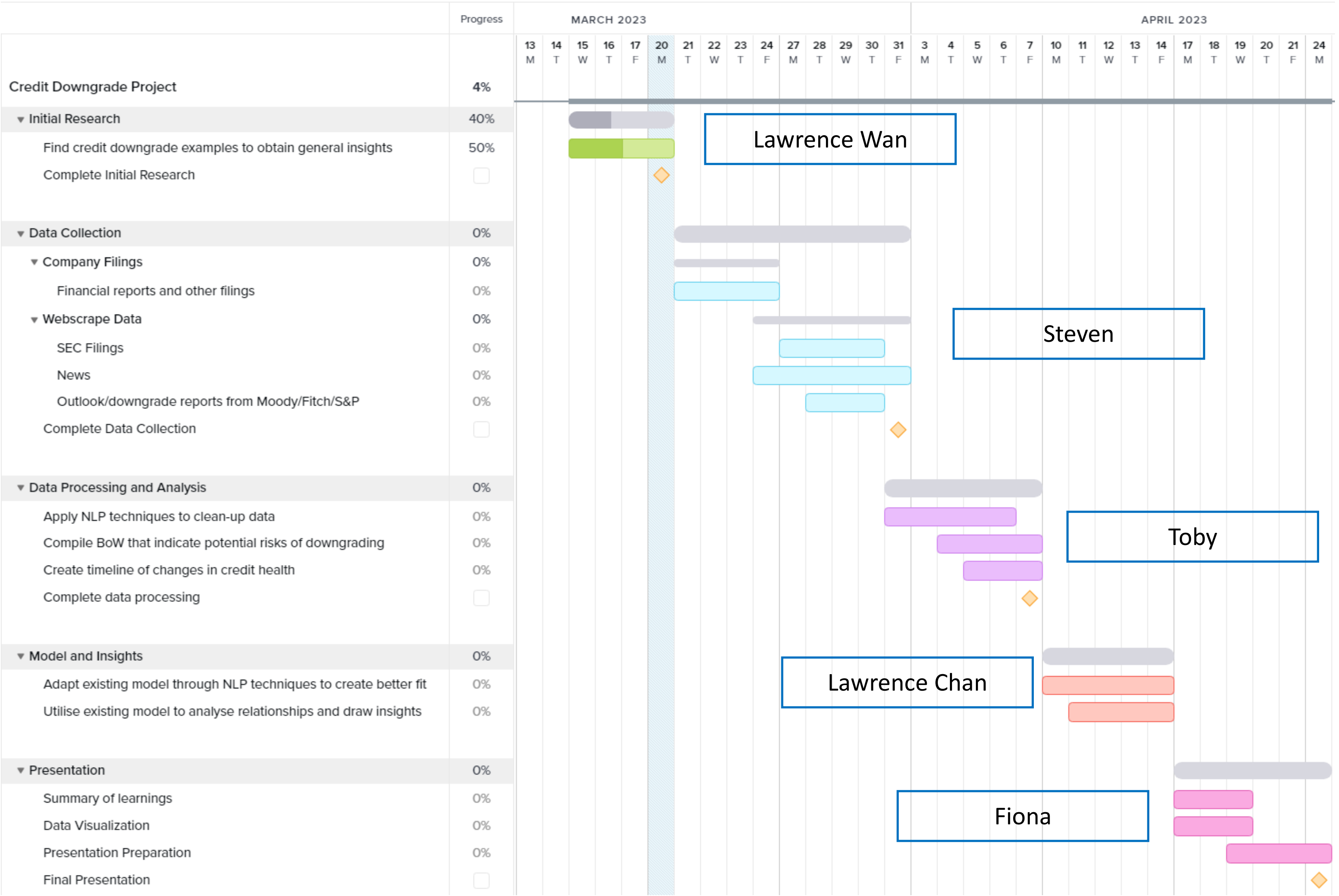
- Bond grading is done behind closed doors with multidimensional considerations which may not be reflected through wordings in reports

3

Correlation is not causation

- This project may be effective in discovering the correlation between reporting and downgrade, but does not offer explanations

Timeline and Distribution of Work



Credit Downgrade Examples

S&P downgrades China Evergrande again to 'CCC'

Reuters

DBRS Morningstar cuts Credit Suisse credit rating to 'BBB'

Reuters

Source(s): Reuters

Can publicly available information detect deteriorating credit health?

Li Hao (Fiona) 3035666308

Chan Yin Tsung Lawrence 3035712169

Chiu Tsz Chun (Toby) 3035712195

Lu Yuzhen Steven 3035713943

Wan Man Lau Lawrence 3035666396

Introduction

Key Questions

- Were there signals present that predicted each credit rating downgrade?
- Were these signals present in publicly available information?

Methodology

Collect links for downgrade reports for training data

Web scrape for text data, and preprocess such data

Use unsupervised machine learning to generate clusters and build model

Clustering Approach

Apply SCM algorithm to find similarities given contextual word embeddings

Similarity Approach

Appy the model/algorithm to test data

Data Collection

Training Data:
Downgrade Reports of 30 Key Chinese Developers

Testing Data:
MD&A in Annual Reports

Moody's revises Fantasia's rating outlook to negative; affirms B2/B3 ratings

Hong Kong, July 16, 2021 -- Moody's Investors Service has changed the ratings outlook on Fantasia Holdings Group Co., Limited (Fantasia) to negative from stable.

At the same time, Moody's has affirmed Fantasia's B2 corporate family rating (CFR) and B3 senior unsecured ratings.

RATINGS RATIONALE

Fantasia's B2 CFR reflects the company's long track record in property development in the Chengdu-Chongqing Metropolitan Area, diversified income streams from its property management business, and good liquidity.

On the other hand, the B2 rating is constrained by Fantasia's improving but still-high debt leverage and weak interest coverage, resulting from debt-funded growth; and its concentrated funding structure, which relies heavily on offshore USD bonds.

FACTORS THAT COULD LEAD TO AN UPGRADE OR DOWNGRADE OF THE RATINGS

An upgrade of the rating is unlikely in the near term, given the negative rating outlook.

However, the rating outlook could return to stable if the company (1) reduces its reliance on offshore debt and improves its onshore funding access, (2) improves its EBIT interest coverage to above 1.5x consistently, and (3) maintains stable property sales growth supported by a high level of cash collections and good liquidity.

MANAGEMENT DISCUSSION AND ANALYSIS

OVERALL PERFORMANCE

For the year ended 31 December 2020 (the Year), the revenue was RMB507.25 billion (2019: RMB477.56 billion), representing a year-on-year growth of 6.2%. Gross profit was RMB122.61 billion (2019: RMB132.94 billion).

Core business profit for the Year was RMB30.13 billion, which is based on the net profit excluding exchange gains or losses, fair value gains on investment properties, fair value gains or losses on financial instruments, donations and certain non-property development business losses.

REVENUE

Revenue of the Group was RMB507.25 billion for the Year, representing an increase of 6.2% as compared with 2019. Revenue generated from the property development segment increased by 6.5% to RMB494.55 billion. The increase was mainly due to the 11.6% increase in delivered area for the Year as compared to 2019. Revenue generated from property management amounted to RMB6.56 billion, an increase of 49.8% from 2019, which was mainly due to the increase in area under the Group's management service for the Year. Revenue generated from investment properties amounted to RMB1.28 billion, down by 6.5%, which was mainly due to the preferential measures for rent reduction and exemption provided by the Group during the first half of 2020 in order to help tenants tide over the pandemic. Rental income significantly decreased during the first half of the Year. However, the rental income for the second half of the Year increased by 51.2% as compared to the second half of 2019.

GROSS PROFIT

Gross profit of the Group was RMB122.61 billion for the Year. Gross profit for the Year decreased mainly due to a decrease in the average selling prices caused by the nationwide sales promotion activities and sales price concessions of the Group as a result of COVID-19. In addition, the average delivery costs per square meter during the Year increased by 0.8% as compared to 2019. Therefore, the revenue for the Year grew but the gross profit margin decreased. Gross profit margin during the Year was 24.2%.

FAIR VALUE GAIN ON INVESTMENT PROPERTIES, NET

Fair value gain on investment properties of the Group for the Year was RMB1.28 billion, which decreased by 15.8% as compared with RMB1.52 billion in 2019. Investment properties of the Group mainly include commercial podiums in living communities and office buildings with total gross floor area of about 9.22 million square meters, and approximately 356,000 car parking spaces.

OTHER INCOME

Other income of the Group for the Year was RMB10.25 billion, which was mainly attributable to the interest income, forfeited customer deposits and management and consulting service income.

OTHER (LOSSES)/GAINS, NET

Other net losses for the year were RMB5.1 billion mainly due to the losses in relation to the restructures of investors' equity interests in a subsidiary and foreign exchange losses. Other net gains for the last year were RMB1.73 billion which mainly represented net gains from disposal of subsidiaries, joint ventures and associates.

Insights from data scraping

Steps	Problems Encountered	Workaround
1 Identifying sources of financial data to analyze	S&P website did not provide **free access** to credit downgrade reports	Only used Moody's and Fitch reports
2 Identifying lines of HTML that were relevant	Our code could not deal with procedural loading, as locating expected XPATH was inconsistent	Added explicit waits that depended on triggers to progress the code
3 Webscraping- Coding	**Javascript:** Requests could not deal with JS **Website format:** lots of pop-ups such as login pages, cookie requests	Had to use Selenium instead Lots of trial and error to see if code could run consistently
4 Webscraping- Process	**Website format changed** during the project, meaning we had to recode the scraping	Recode the data scraping process; takes more time but buffer time allocated for this step

```
# allow JS/Ajax to load
WebDriverWait(driver, 10).until(EC.presence_of_element_located((By.XPATH, '//form/div/div/div/input')))

#remove login pop-up
element = driver.find_element(By.XPATH, '//form/div/div/div/input')
element.click()
element.send_keys("fina4350.nlp@gmail.com") # temp email
element = driver.find_element(By.XPATH, '//form/div/button')
element.click()
WebDriverWait(driver, 5).until(EC.presence_of_element_located((By.XPATH, '//form/div/div[2]/div/input')))
element = driver.find_element(By.XPATH, '//form/div/div[2]/div/input')
element.click()
element.send_keys("fina4350") # temp password
element = driver.find_element(By.XPATH, '//form/div/button')
element.click()
```

Insights from data pre-processing

Steps	Problems Encountered	Workaround
1 Keeping only alpha-numeric characters	n/a	n/a
2 Removing stop-words, short (1-character words) or irrelevant words	During project we kept updating preprocessing steps; meaning we had to constantly rerun the program to get the updated tokenized data	Allocated more buffer time as we expected this to happen
3 Lemmatization	There are words the program over-lemmatized, leading to possible data errors	n/a
4 NLTK Tokenisation	Scraping picked up words that missed spaces (previously carriage returns) – these words did not separate when processed	Had to replace all carriage returns with spaces during scraping process

```
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from nltk.stem import  WordNetLemmatizer

def processtext(text):

    # tokenization and pre-processing
    wnl = WordNetLemmatizer()
    arr = list([w for w in word_tokenize(text.lower()) if w.isalpha()])
    stemmed_arr = [wnl.lemmatize(word) for word in arr if not word in set(stopwords.words('english'))]
    print(stemmed_arr)

    return stemmed_arr
```

Insights from Clustering Approach

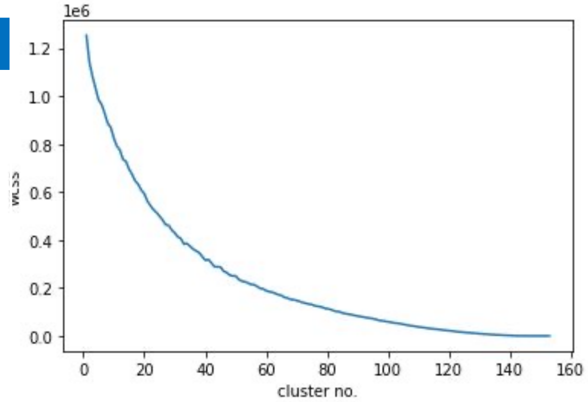
Tf-Idf Weighted Bag of Words

- Determining importance of terms based on **frequency** in each document and across documents
- Generating **vectors** based on weighting for ML model
- **Visualisation** of BoW through word cloud

K-Means Clustering Model

- Finding **clusters** of similar words using unsupervised ML model
- **Optimise** no. of clusters using elbow method
- Next Steps:
  - **Classification analysis** to determine which clusters can more effectively predict defaults
  - **Increasing accuracy** of model generation process per set of training data through further training

Cluster 1	Cluster 2	Cluster 3
Country	Advance	Disclosure
Garden	Debt	Rated
Moody	Sale	Please
Rating	Rate	Regulatory
Metric	Liquidity	Issued



Insights from Similarity Approach

Doc2Vec Neural Network

- **Dense neural network** to estimate relationship between words
- Word similarities are guessed through **context** with neighbouring terms
- Next Steps:
  - **Visualisation** of relationships based on current model
  - **Optimise model** based on parameters to minimise noise
  - **Increasing efficiency** of model generation process by increasing size of corpus and training time

```
model = Doc2Vec([TaggedDocument(doc, [i]) for i, doc in enumerate(textList)], workers = 2, min_count = 2)
similar_word = model.wv.most_similar('default')
print(similar_word)
```

```
[('uncured', 0.9126061797142029), ('definition', 0.8995665311813354), ('c', 0.8960571885108948), ('idrs', 0.8956425189971924),
('experienced', 0.8928045630455017), ('summaryevergrande', 0.882044792175293), ('scenery', 0.8810537457466125), ('b',
0.8798206448554993), ('triggered', 0.8730372190475464), ('idr', 0.8699644804000854)]
```

Soft Cosine Measure (SCM)

- Predicting similarity between documents based on **contextual relationships from words**
- Negate vector geometry difficulties with high neighbor distance through **word embeddings (Doc2Vec)**
- Next Steps:
  - **Bug-fix** for model – currently unable to deal with errors regarding terms **out of model vocabulary (OOV)**
  - **Refactoring** code - current input data structures are not aligned
  - Providing **larger set of training data** to optimize similarity values

Results

Clustering Approach

```python
# testing data
with open('testing-data.pickle', 'rb') as myfile:
    testData = pickle.load(myfile)
myfile.close()

for ind in range(len(testData)):
    joinedText = ' '.join(testData[ind])
    testData[ind] = joinedText

# making predictions
temp = v.fit_transform(testData)
prediction = model.fit_predict(temp)
print(prediction)
```

```
[0 0 0 2 2 2 2 1 1 1]
```

- Our model can effectively **classify testing documents into appropriate clusters**
- Accuracy improvements: larger set of training data, increasing no. of clusters, or increasing set of testing data (to identify nuanced differentiation of clusters)

Similarity Approach

```python
# create dictinoary
documents = []
for doc in testData:
    if doc in model:
        documents.append(doc)
dictionary = Dictionary(documents)

documents = [dictionary.doc2bow(doc) for doc in documents]

# create tf-idf model
tfidf = TfidfModel(documents)
documents = [tfidf[doc] for doc in documents]

# SCM
termsim_index = WordEmbeddingSimilarityIndex(model)
termsim_matrix = SparseTermSimilarityMatrix(termsim_index, dictionary, tfidf)

def SCM(docA, docB):
    similarity = termsim_matrix.inner_product(docA, docB, normalized=(True, True))

# testing

for ind in range(len(textList)):
    joinedText = ' '.join(textList[ind])
    textList[ind] = joinedText

for ind in range(len(testData)):
    joinedText = ' '.join(testData[ind])
    testData[ind] = joinedText

print([SCM(textList[0], test) for test in testData])
```

- Currently still working on **bug-fixing SCM approach**
- Expected ability to **determine similarities between downgrade reports and similar documents in China RE field**

Application

Potential Applications

1

Application of Model on Other Data (within company)

- Can be used in earnings transcripts, and other textual data within the same company

2

Application of Model on Outside Data (other companies)

- Can be used in other textual data for different companies in the China Real Estate industry

3

Application of Model on Outside Data (other industries/geographies)

- Using in other textual data for different companies outside of China Real Estate industry

Adjustments/Challenges

1

Webscraping

- Webscraping code needs to be adjusted as it is specific to each medium of text/source

2

Pre-processing

- Different companies tend to use different jargon to convey the same idea; preprocessing needs to be adjusted which will take more time

3

More training needed

- Machine learning model is only trained to analyze credit downgrades in the context of China Real Estate
- More training is needed to properly understand contexts from other industries

Technical Limitations - Approach

Models → **Limitations**

Clustering Approach

- Tf-Idf Bag of Words
  - Ignores contextual relationships between words
- K-Means clustering
  - Dependency on initial clustering value
  - Difficulty in clustering data of varying sizes and density

Similarity Approach

- Doc2vec Neural Network
  - Context limited to neighboring window
  - Vocabulary limitations - OOV
- Soft Cosine Measure
  - Sensitivity to term frequency

Technical Limitations – Efficiency/Scalability

Current Limitations

Technical limitation – Efficiency:
- Many parts of the program are not fully optimized for speed and memory usage
  - Scraping takes too long – **40 minutes per run**
  - Model-building are not optimized around streaming – **RAM usage is too high**

Technical limitation – Scalability:
- Code is unoptimized for Big-O **–** increasing size of training/testing data will cause **scraping/memory usage to increase exponentially** - $\theta(n^2)$

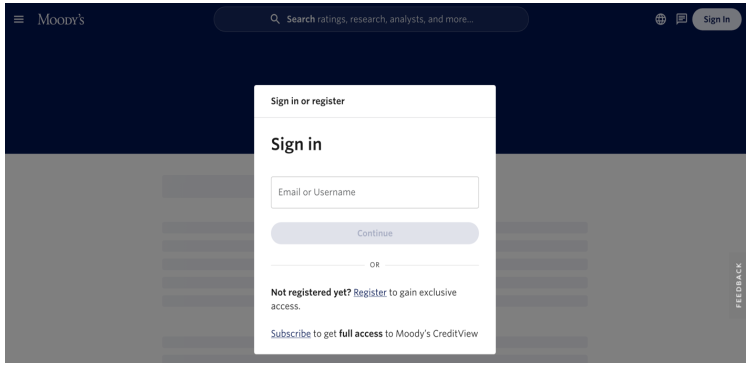
Improvement Areas

1. **Refactor current code to improve efficiency**
   - Cut down on inefficient/obsolete aspects
   - **Multi-threading** could allow for multiple scrapers at once
2. **Improve data pipeline to allow for streaming**
   - Integrate codes together to allow for smoother data pipeline – **increase compatibility** of different parts
3. **Utilise more efficient data structures**
   - Include only structures with faster access capabilities (e.g. dataframes)

Challenges

Difficulty in bypassing login prompts / CAPTCHA

Example



Solution

Automatically enter login prompts with click/send keys

Consistency of code outputs when pre-processing

Established in 1996, Logan Group Company Limited is a property developer based in Shenzhen. The company focuses mainly on residential projects in Shenzhen, Shantou, Nanning and Huizhou.

Logan listed on the Hong Kong Stock Exchange in December 2013. As of the end of June 2021, its land bank totaled 85.6 million square meters in gross floor area in several cities across China, including Shenzhen; Shantou; Nanning; Hong Kong SAR, China; and other Greater Bay Area cities, as well as Singapore.

REGULATORY DISCLOSURES

For further specification of Moody's key rating assumptions and sensitivity analysis, see the sections Methodology Assumptions and Sensitivity to Assumptions in the disclosure form. Moody's Rating Symbols and Definitions can be found at: https://www.moodys.com/researchdocumentcontentpage.aspx?docid=PBC_79004.

Optimize code during scraping to only take useful HTML elements

Removing irrelevant noise in both approaches


```
'affected', 'fluctuation', 'foreign', 'exchange', 'rate', 'group', 'managed', 'exposure', 'fluctuation', 'foreign', 'exchange',
'rate', 'implementation', 'certain', 'foreign', 'exchange', 'swap', 'arrangement', 'continue', 'closely', 'monitor', 'fluctuation',
'foreign', 'exchange', 'rate', 'actively', 'take', 'corresponding', 'measure', 'minimise', 'foreign', 'exchange', 'risk',
'contingent', 'liability', 'financial', 'guarantee', 'group', 'provides', 'guarantee', 'bank', 'mortgage', 'loan', 'certain',
'property', 'purchaser', 'ensure', 'purchaser', 'perform', 'obligation', 'mortgage', 'loan', 'repayment', 'amount', 'guarantee',
'approximately', 'billion', 'december', 'compared', 'approximately', 'billion', 'december', 'guarantee', 'terminate', 'upon',
'earlier', 'transfer', 'real', 'estate', 'ownership', 'certificate', 'purchaser', 'generally', 'occur', 'within', 'average',
'period', 'six', 'month', 'property', 'delivery', 'date', 'ii', 'satisfaction', 'mortgage', 'loan', 'purchaser', 'property',
'period', 'guarantee', 'provided', 'group', 'start', 'date', 'mortgage', 'granted', 'b', 'litigation', 'date', 'report', 'various',
```


Improving pre-processing and tokenization to minimize "garbage"

Challenges

Understanding of Credit / Real Estate Industry

Specific knowledge required for understanding context of credit

- Credit rating systems
- Industry-specific conventions
- Complex financial terminologies

For Example…

On 18 February 2022, Zhenro announced a consent solicitation to its securities holders for the USD senior perpetual capital securities senior notes due in March 2022 with a total principal amount of USD200 million. The proposal seeks holders' consent to (1) waive any default if the company does not redeem the senior perpetual capital securities and (2) delay a contractual coupon step-up until March 2023. The company also said that it may not be able to fully redeem the securities if the consent solicitation is not successful.

Application / Practicality Concerns

Obtain data from credit rating action reports

↓

Data analysis and operations to generate results

↓

Company financial statements and reports -> Is it really applicable to other types of documents since the training data was downgrade repots

↓

Application to other geographies and industries?

Future Implications

Implication on Credit Valuation

Development of Improved Credit Health Valuation

Credit health valuation from rating companies are an important consideration for market stakeholders

Improved accuracy of rating:

- Access to accurate and efficient predictors helps investors make more rational decisions
- Reduces information asymmetry

Expand access of capital to companies:

- Increased accuracy of rating can allow companies previously perceived as "too-risky" can gain easier access to capital

Profit Implications

Improved Investment Strategy for Asset Managers

Change in credit rating often cause change in the market price of securities

The ability to credit change in credit rating will allow asset managers to:

- Improve scope of investment strategies
- Increase returns/minimize losses
- Improve rationale for investment decisions
- Provide quantitative measurements as credit health becomes more quantifiable