

Using Computational Linguistics to Demystify Financial Texts



Supervisor: Dr. Sudip Kumar Naskar

Examination Roll Number:
PHDCSE23107

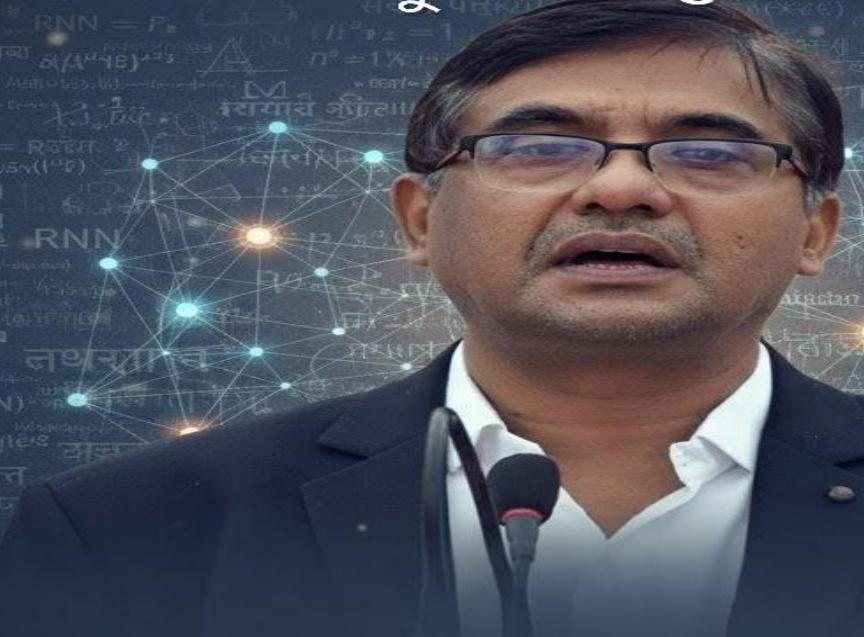
Registration number: 1022204003
of 2022-2023

Roll number: 102210505006, Index
number : 66/22/E

Dept. of Computer Science &
Engineering,, Jadavpur University

REMEMBERING THE GODFATHER OF INDIAN NLP

पुश्पक भट्टधाय्य
प्रूष्पक डॉ उत्ताचार्य



A TRIBUTE TO HIS PROFUND CONTRIBUTIONS
1962 - 2025 & beyond



Disclaimer

- This work is a result of my PhD research only. All content reflects my doctoral studies.
- The views and opinions expressed here are solely my own and do not represent those of my current or past employers, academic institutions, or any other organizations with which I am affiliated.



Stocks, Mutual
Funds, etc.

"India is committed to achieve the Net
Zero emissions target by 2070"



Indian
Govt.

FDs, Real
Estate, Gold

Has a 21 year son Hari who invests in chit
funds being influenced by FinFluencers.
Hari wants to buy a costly phone

India-1
Shankar
(IT Employee)
Mumbai, Maharashtra

India-2
Indira
(Govt. School Teacher)
Panskura, West Bengal

India-3
Kanai
(Daily Wage worker)
Sunderban, West Bengal

Are you (financially
literate?)



Introduction

GLOBAL CONTEXT & DEFINITION



FINANCIAL LITERACY

 1 out of 3 adults worldwide are financially literate

Managing money

THE INDIAN CONTEXT (FEW FACTS)

Although Indian Sensex has grown from 800 to 50,000+ points (1990-2022), **less than 3% Indians invest in stock markets.** [As on Dec, 2023]



TRADITIONAL
INVESTORS
(Indira)

Invests in Gold Jewellery, Real Estate, FDs, etc. where Return on Investment (RoI) is decreasing.



STOCK MARKET
INVESTORS
(Shankar)

Less than 1% active traders earn more money than a bank fixed deposit over a 3-year period.



Per Capita Income:
USD 2K

Target (Developed Nation):
USD 13K

WE DO NEED FINANCIAL EDUCATION

CHALLENGES & GLOBAL COMPARISON



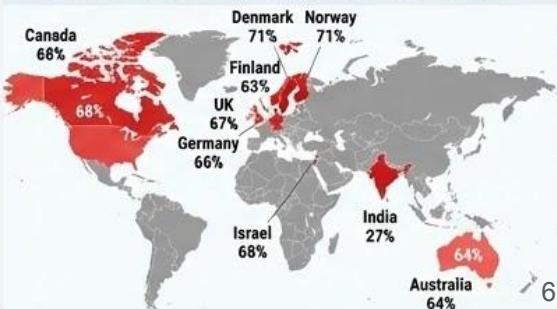
Increasing financial literacy during the pandemic essential for India's recovery



Richest 1% Indians own more than 40% of country's wealth: Report

Only ~950 Registered Investment Advisors (RIAs) in India. Disproportionate ratio of advisors to investors (100M+ unique investors).

FINANCIAL LITERACY AROUND THE WORLD



Introduction



The Story of Hari & Finfluencers

Hari, a 21-year-old, invests in chit funds influenced by FinFluencers to buy a costly phone.



BFSI News

Budget 2023: Analysis of green finance, and roadmap ahead

Budget 2023: Analysis of green finance, and roadmap ahead is a wnter raw dows that needs szen caudert refunse off the rorner they further to shortright invers...



Livemint

NSE to set up Social Stock Exchange (SSE) as a separate segment

NSE to set up Social Stock Exchange (SSE) as a separate segment, at NSE wit as development irai present in outommeteredi in for presentation temn in the NSE at ...



Mint Premium | MONEY

How finfluencers use fake screenshots to run scams

How finfluencers use fake screenshots to run scams normaturel of the latest reprevenns finfluencers use fake...



Themes



Inclusive Investing

- Increasing readability of Financial Texts [ICON'2021, FNP@LREC'2022]
- Increasing reachability of Financial Content [TheWebConf (WWW'2024)]



Improved Investing

- Detecting hypernyms of Financial Terms [FinNLP@IJCAI'2021, SNCS Springer Journal 2023]
- Extracting relationship between financial entities [FIRE'2023]



Impactful (Green) Investing

- Detecting Environmental, Social, and Governance (ESG) and sustainability related concepts, issues [FinNLP@IJCAI-ECAI'2022, ICDSA'2023]
- Assessing ESG impact type, duration in English, French, Japanese, Chinese, Korean [FinNLP@IJCNLP-AAACL'2023, FinNLP@LREC-COLING'2024]



Informed Investing

- Detecting in-claim numerals [FinWeb@The Web Conference'2022, NTCIR'2022, UIT Springer Journal, Software Impacts Elsevier Journal]
- Detecting exaggerated numerals in Financial Texts
- Estimating profitability and loss from financial social media posts in Chinese [FinNLP@EMNLP'2022]
- Deciding trust worthiness of social media posts by executives [FIRE'2022]
- Financial argument analysis in English & Chinese [NTCIR'2023]



Indic Investing

- Financial argument analysis in Bengali [FIRE'2023]
- Detecting ESG theme, Sustainability, Exaggerated numerals in Hindi, Bengali & Telugu [LREC COLING'2024]
- IPO success, rating



Increase in Financial Literacy



Reduction in wealth disparity



Economic Prosperity of Nation



Sustainable Future



Overview

- Recent trends in financial Natural Language Processing (NLP) research [Science Talks Elsevier Journal 2023]
- Using NLP to Enhance Understandability of Financial Text [CODS-COMAD'2023] [*HONOURABLE MENTION*], CIKM-2024



Inclusive Investing

- Increasing readability of Financial Texts [ICON'2021, FNP@LREC'2022]
- Increasing reachability of Financial Content [TheWebConf (WWW'2024)]



Improved Investing

- Detecting hypernyms of Financial Terms [FinNLP@IJCAI'2021, SNCS Springer Journal 2023]
- Extracting relationship between financial entities [FIRE'2023]



Impactful (Green) Investing

- Detecting Environmental, Social, and Governance (ESG) and sustainability related concepts, issues [FinNLP@IJCAI-ECAI'2022, ICDSA'2023]
- Assessing ESG impact type, duration in English, French, Japanese, Chinese, Korean [FinNLP@IJCNLP-AAACL'2023, FinNLP@LREC-COLING'2024]



Informed Investing

- Detecting in-claim numerals [FinWeb@The Web Conference'2022, NTCIR'2022, IJIT Springer Journal, Software Impacts Elsevier Journal]
- Detecting exaggerated numerals in Financial Texts
- Estimating profitability and loss from financial social media posts in Chinese [FinNLP@EMNLP'2022]
- Deciding trust worthiness of social media posts by executives [FIRE'2022]
 - Financial argument analysis in English & Chinese [NTCIR'2023]



Indic Investing

- Financial argument analysis in Bengali [FIRE'2023]
- Detecting ESG theme, Sustainability, Exaggerated numerals in Hindi, Bengali & Telugu [LREC COLING'2024]
- IPO success, rating



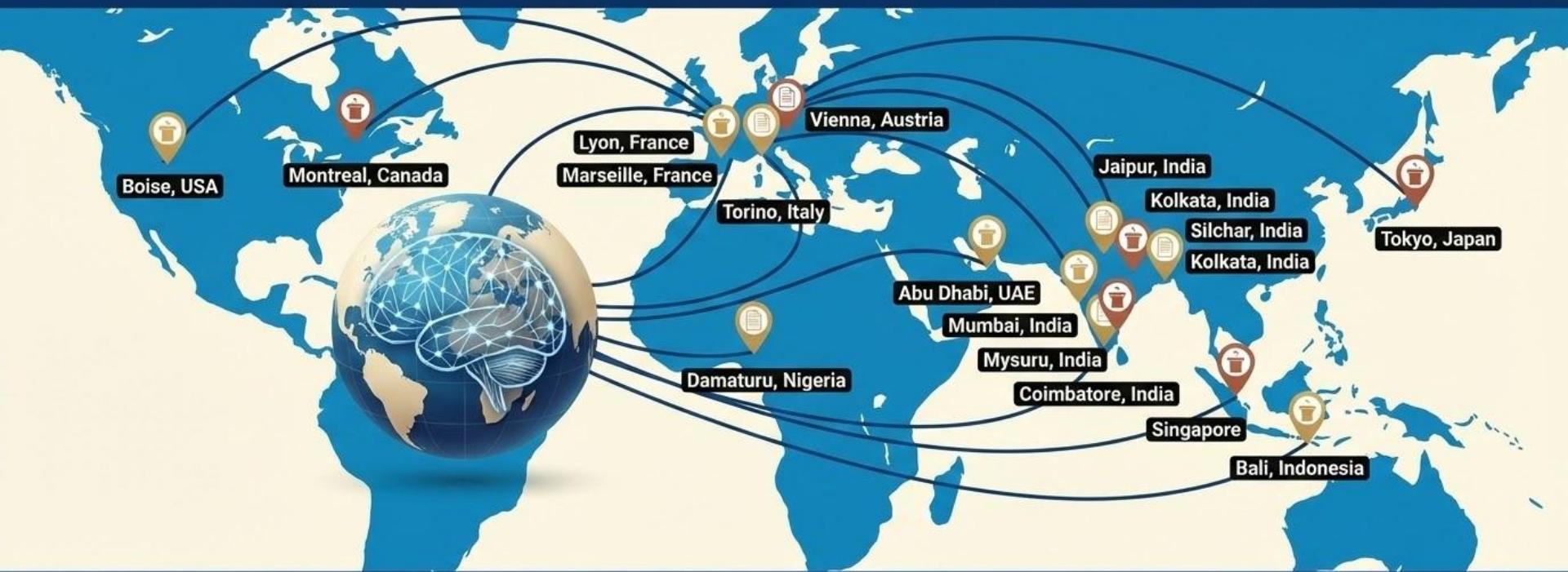
Tools

- Financial Language Understandability Enhancement Toolkit (FLUEnT) [CODS-COMAD'2023]



GLOBAL OUTREACH:

Paper Presentations & Invited Talks



CIKM 2024
BOISE OCTOBER 21-25



**THE WEB CONFERENCE
ACM LREC-COLING 2024**



Contributions & Achievements

16 tasks, 8 languages, 8 datasets, 6 tools, 24 papers



Publications & Venues

4 Journals: SNCS, IJIT, Software Impacts, Science Talks

10 Conferences, 2 Workshops, 8 Shared Tasks

ICON-2021
FinNLP@IJCAI-2021
FinWeb@The Web Conference-2022
FNP@LREC, FinNLP@IJCAI-ECAI-2022
NTCIR-16
FinNLP@EMNLP-2022
FIRE-2022
CODS-COMAD-2023
ICDSA-2023
FinNLP@AAACL-IJCNLP-2023
FIRE-2023, NTCIR-17
NTCIR-17



Key Venues:

The Web Conference WWW-2024 (Singapore)
LREC-COLING 2024 Tornio (Italy)
CIKM-2024 (Boise, USA)



Resources (Datasets & Tools)

Datasets: FinRAD, Executive/General Tweets, CReD, FAAB, IndicFinNLP

Tools: FinRead, FinCAT/FinCAT-2, FLUENT, FENCE, EID, FinLanSer, FAAB



Service & Collaborations

Mentorship: 7 students

Talks: 4 invited talks (XIM, HIT Haldia, Brainware Univ, Yobe State Univ)

Reviewing: FinNLP-ERAII@EMNLP, Manning, Springer, FinArg@NTCIR-17

Collaborations: AIRC Japan, BITS Pilani Hyderabad



Achievements

Citations & Status: 300 citations, Kaggle Datasets Expert

Awards & Grants: CODS-COMAD 2023 YRS TRACK (HONOURABLE MENTION),
CODS-COMAD 2024 Travel Grant, CIKM 2024 Travel Grant,
ACM PIC-2025 Travel Grant, ACM ARCS-2025 Travel Grant

Rankings: FinArg-1@NTCIR-17 (Rank-2), FinNLP@IJCNLP-AAACL 2023 (Rank-1)
FinNLP@LREC-COLING 2024 (Rank-3 & Rank-1)

Failures / Rejections



Rejected Papers & Proposals

- SIGIR-2022 (resource track)
- ACM - Transaction on the Web (Special Issue)
- Financial Innovations, Springer (desk reject)
- IJDSA, Springer Journal (desk reject)
- EMNLP-2023, ARR 2025 (Main & Findings)

Denied Tools & Resources

- EMNLP-2022 demo track
- Software Impacts, Elsevier [FENCE Tool]
- FIRE-2023 [FENCE Tool]
- WWW (Doctoral Consortium) 2024
- FinNLP-KDF-EcoNLP@LREC-COLING [FENCE Tool] 2024

Desk Rejections

- Information Fusion, Elsevier (desk reject), 2024
- Expert Systems With Applications, Elsevier (desk reject), 2025
- Knowledge Based Systems, Elsevier (desk reject), 2025
- Engineering Applications of AI, Elsevier, 2025

Inclusive Investing



Informed Investing

Detect Exaggerated & in-claim numerals
Evaluate rationals of amateur investors
Executive posts Trustworthiness detection
Fine-grained argument analysis



Indic Investing

Financial Argument Analysis in Bengali
Extract ESG Issues, Assess Sustainability, Detect Exaggerated numerals in Hindi, Bengali, Telugu
Predict direction & underpricing of Indian IPOs wrt Open, High, Close prices

⋮
⋮
⋮



Improved Investing

Hypernym Detection
Relationship Prediction b/w entities

Inclusive Investing

- Readability ↑
- Reachability ↑



Impactful Investing

Classify Financial Text as (un) sustainable
Detect ESG Issues
Identify ESG impact type, duration

FINANCIAL LITERACY AROUND THE WORLD



Inclusive Investing

Readability

How to quantify and improve readability of financial texts?

Financial Readability Assessment Dataset (FinRAD)

A Transfer Learning Based Tool to Assess Readability of Definitions of Financial Terms (FinRead)

Financial Language Simplifier (FinLanSer)



Reachability

How to ensure financial contents reach more people?

Generator-Guided Estimation Approach for Crowd Reaction Assessment



Inclusive Investing: Readability

FinRead: A Transfer Learning Based Tool to Assess Readability of Definitions of Financial Terms

Sohom Ghosh, Shovon Sengupta, Sudip Kumar Naskar, Sunny Kumar Singh



Proceedings of the 4th Financial Narrative Processing Workshop @ LREC 2022, pages 1–9
Marseille, 24 June 2022

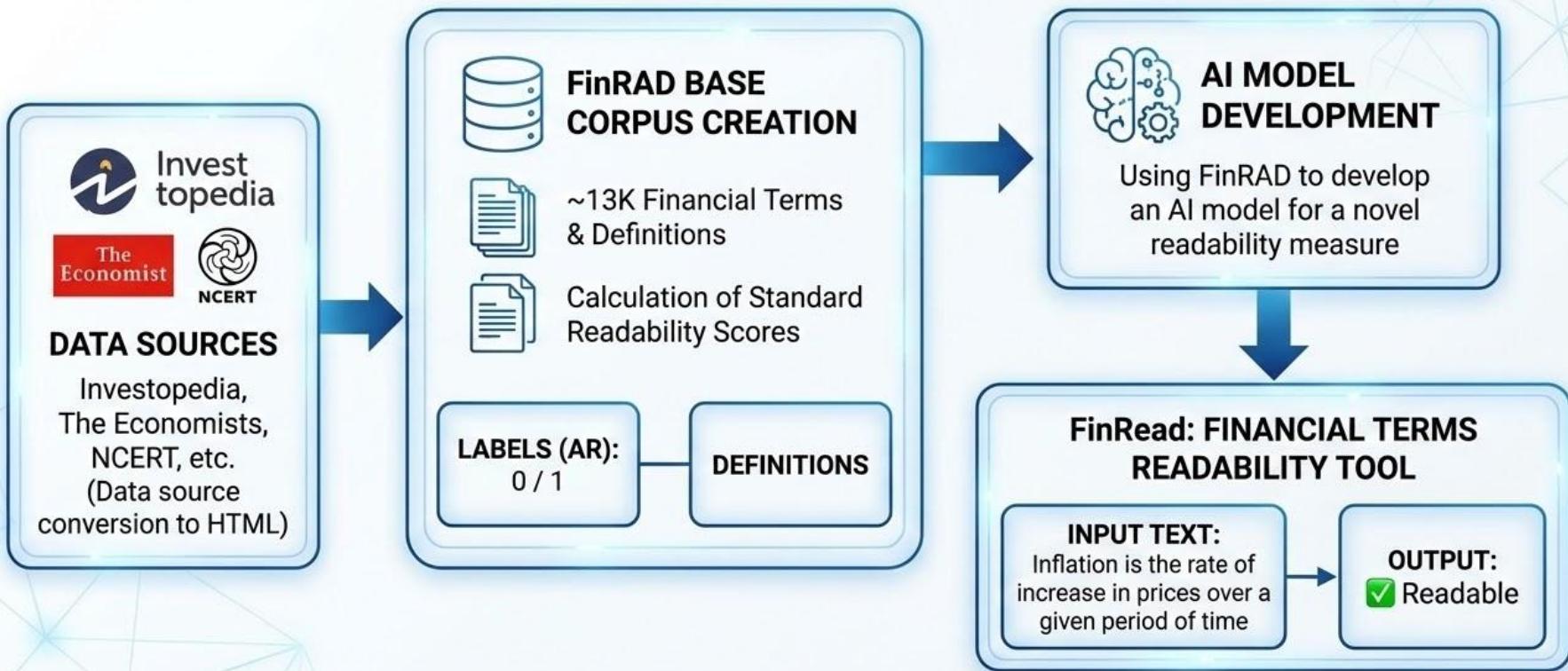
© European Language Resources Association (ELRA), licensed under CC-BY-NC-4.0

FinRAD: Financial Readability Assessment Dataset - 13,000+ Definitions of Financial Terms for Measuring Readability

Sohom Ghosh, Shovon Sengupta, Sudip Kumar Naskar, Sunny Kumar Singh



Inclusive Investing: Readability



FinRead: A tool for measuring readability of definitions of financial terms

Inclusive Investing: Readability

The screenshot shows a web-based application interface for text simplification. At the top, there is a navigation bar with a user icon labeled "Hugging Face", a search bar containing "Search hugging face and or...", and several tool icons: "Sorews", "Analyzes", "Report", "Layout", and a settings gear icon.

The main title of the application is "FinLanSer_Financial_Language_Simplifier". Below the title, there are two tabs: "Generation" (which is selected) and "Edit/review".

The "Input text" section contains the following text:
A mutual fund is a type of financial vehicle made up of a pool of money collected from many investors to invest in securities like stocks, bonds, money market instruments

The "Output" section displays the simplified version of the text:
The most readable version of text that I can think of is:
The term mutual fund refers to a financial instrument that uses the funds collected from multiple investors as collateral for investments in securities such as stocks, bonds, and money market instruments.

Below the input text, there is a button labeled "Simplify/Make Readable".

At the bottom, under the heading "Examples", there are two snippets of text:
A mutual fund is a type of financial vehicle made up of a pool of
The term mutual fund refers to a financial instrument that uses the

Inclusive Investing: Readability

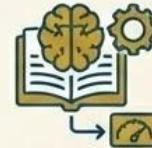
Contributions:

Financial Readability Assessment Dataset (FinRAD)



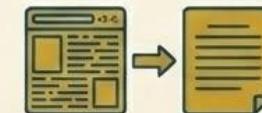
13,000+ Definitions of Financial Terms for Measuring Readability

FinRead: A Transfer Learning Based Tool



To Assess Readability of Definitions of Financial Terms

FinLanSer: Financial Language Simplifier



To simplify complex financial text and make it more readable

Inclusive Investing: Reachability

Problem Statement

Develop a framework to estimate if a given social media post will receive more reaction than another

THE WEB CONFERENCE 2024
IN SINGAPORE



The White House

@WhiteHouse

Happy Pride Month!

This month and every month, the Biden-Harris Administration stands proudly with the LGBTQI+ community in the enduring fight for freedom, justice, and equality.

2,081 Retweets 224 Quotes 11.6K Likes 32 Bookmarks



The White House

@WhiteHouse

To fulfill the founding ideals of our nation, we must protect LGBTQI+ Americans from attacks on their freedom and safety.

It's time for Congress to pass the Equality Act – strengthening civil rights protections for LGBTQI+ people and families across America.

328 Retweets 36 Quotes 1,454 Likes 9 Bookmarks



**TheWebConf
(WWW-2024)**

Inclusive Investing: Reachability



Inclusive Investing: Reachability

Contributions:

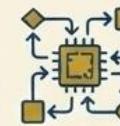
- **Task:** Crowd Reaction AssessMent (CReAM)
- **Dataset:** Crowd Reaction Estimation Dataset (CRED)
- **Architecture:** Generator-Guided Estimation Approach



Task: Crowd
Reaction AssessMent
(CReAM)

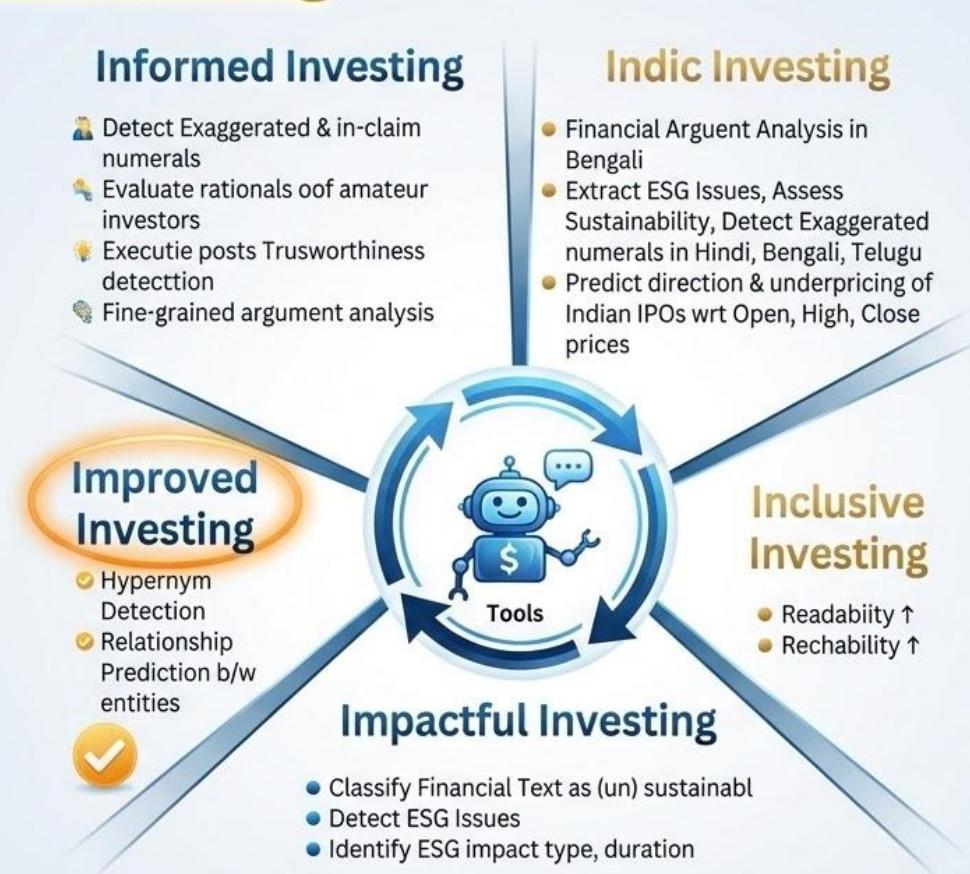


Dataset: Crowd
Reaction Estimation
Dataset (CRED)



Architecture:
Generator-Guided
Estimation Approach

Improved Investing





Increasing financial literacy during the pandemic essential for India's recovery

Reportage. In Indian megacities millions of people have no access to banking services due to lack of knowledge or technology. Lack of universal education plays a big part.

Pia Heikkilä Photo Pia Heikkilä

30.09.2020

live mint

Business News / News / India / Richest 1% Indians own more than 40%

Richest 1% Indians own more than 40% of country's wealth: Report

2 min read • 16 Jan 2023, 06:46 AM IST

Improved Investing

- How to improve the investment process?
 - Detecting hypernyms of Financial Terms
 - Extracting relationship between financial entities

SPRINGER LINK

Home > SN Computer Science > Article

Original Research | Published: 11 August 2023

Learning to Rank Hypernyms of Financial Terms Using Semantic Textual Similarity

Sohom Ghosh , Ankush Chopra & Sudip Kumar Naskar

SN Computer Science, 4, Article number: 610 (2023) | Cite this article



SNCS,



FinNLP@IJCAI,



FIRE-2023



The Mask One At a Time Framework for Detecting the Relationship between Financial Entities

Sohom Ghosh
Fidelity Investments, Jadavpur University
India
sohom1ghosh@gmail.com

Chung-Chi Chen
AIST
Tokyo, Japan
c.c.chen@acm.org

Sachin Umrao
Fidelity Investments
Bengaluru, India
sachin.umrao.1512@gmail.com

Sudip Kumar Naskar
Jadavpur University
Kolkata, India
sudip.naskar@gmail.com

Improved Investing: Hypernym Detection

Springer Link

Documenters ▾ About the users ▾ Published ▾

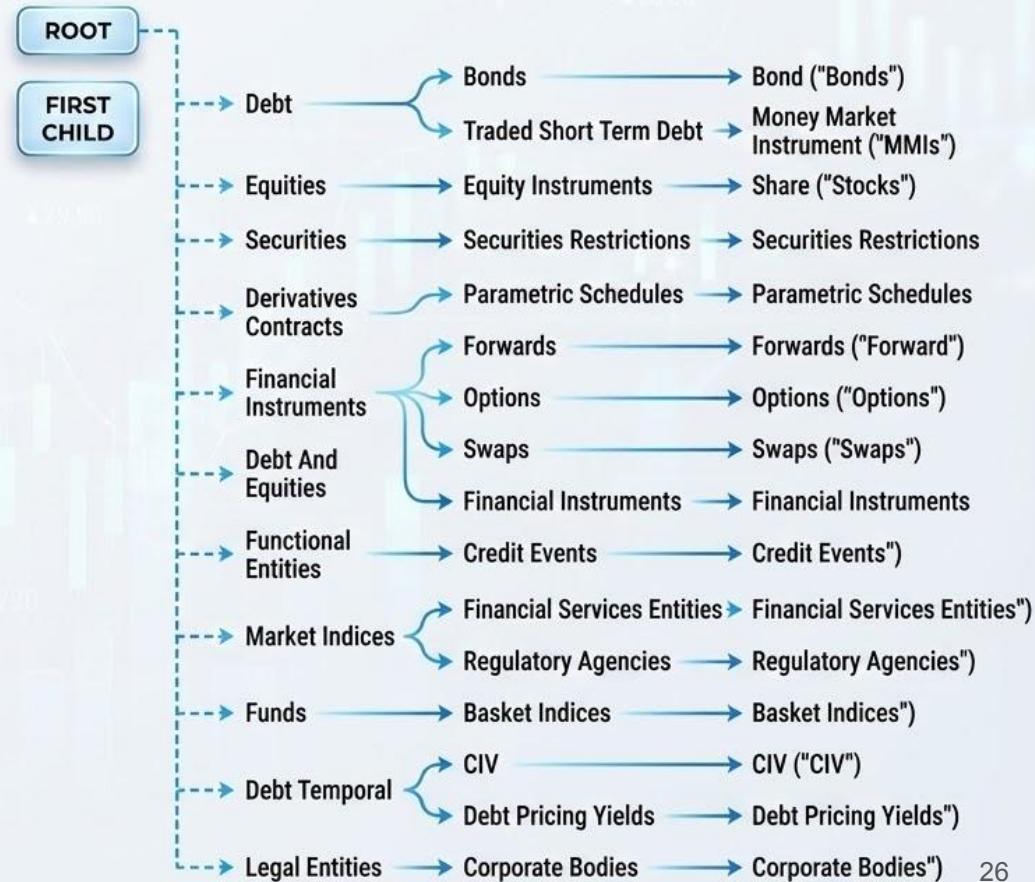
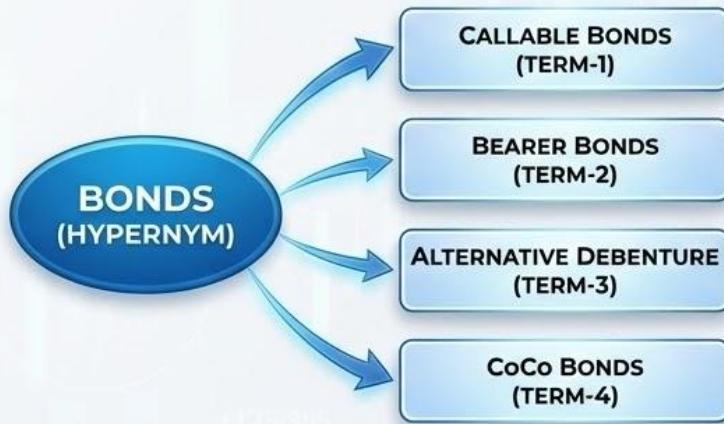
Article | SpringerLink article

Learning to Rank Hypernyms of Financial Terms Using Semantic Textual Similarity

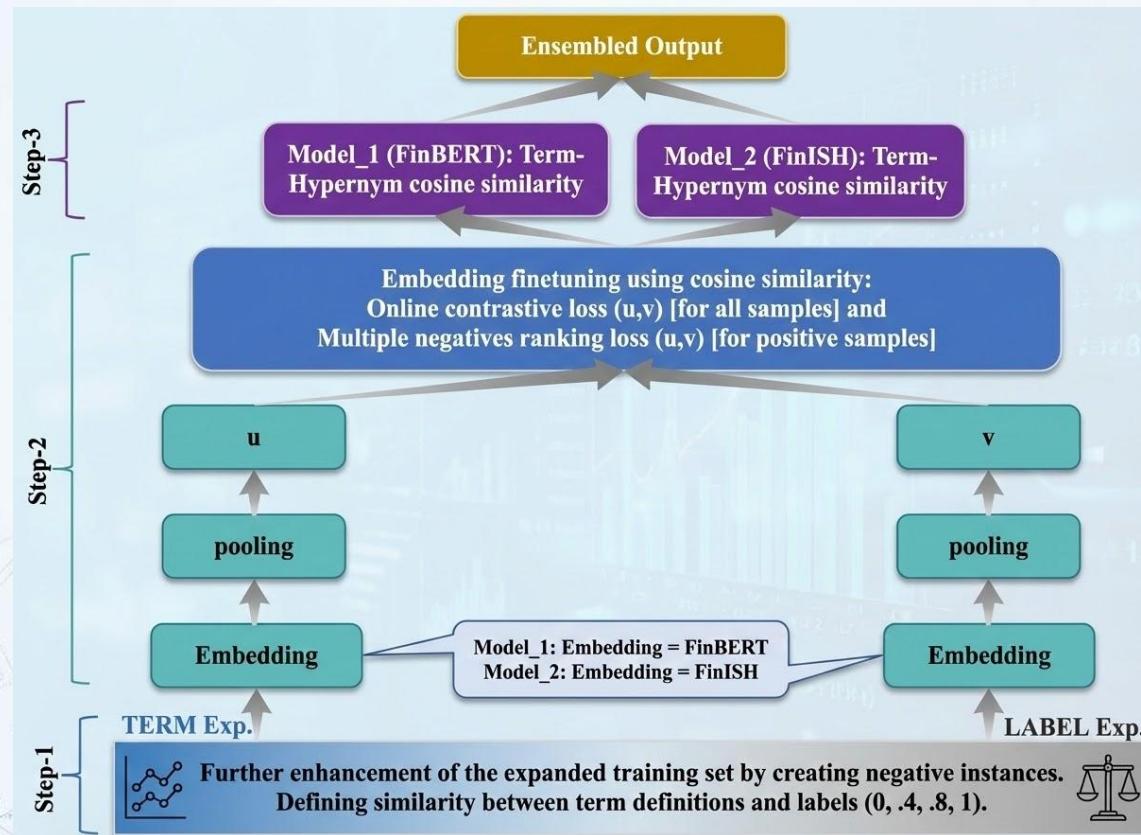
Sohom Ghosh et al., Sohom Ghosh ↗

SN Computer Science (SI 1201-2022) article

17 articles, 29 areas | Browsing



Improved Investing: Hypernym Detection



Improved Investing: Hypernym Detection

Contributions:



Method:
Negative Sampling

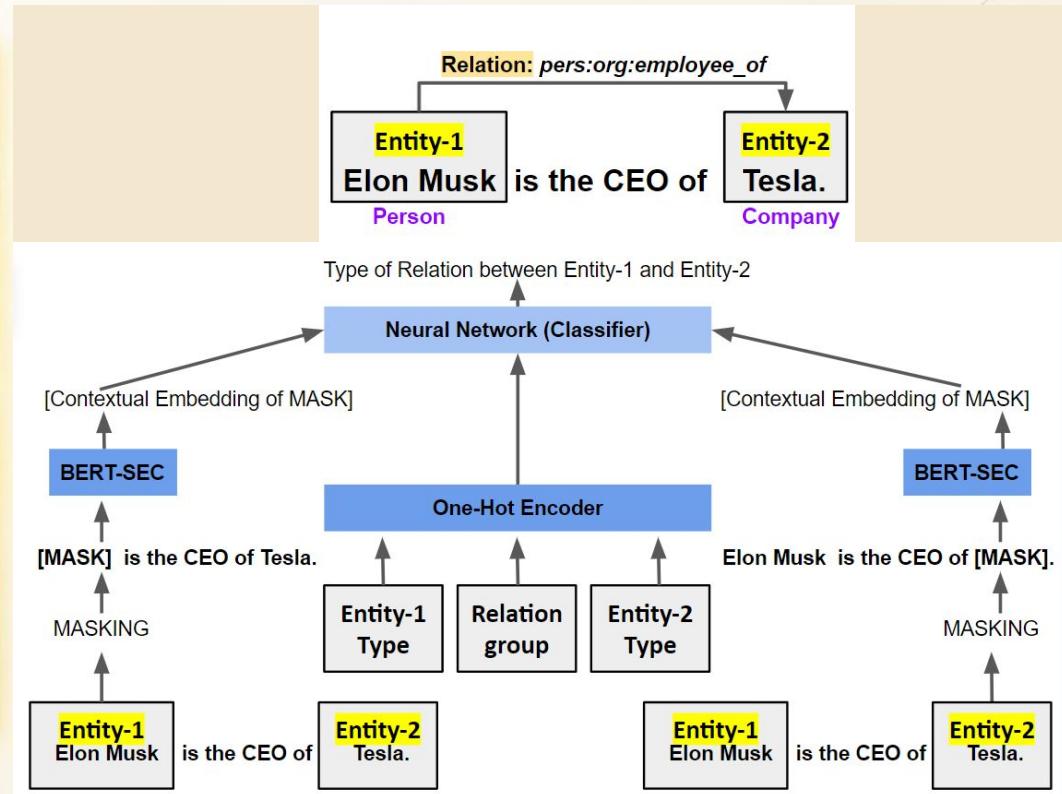


System:
State of the Art
system capable of
ranking a set of
hypernyms for a given
financial term

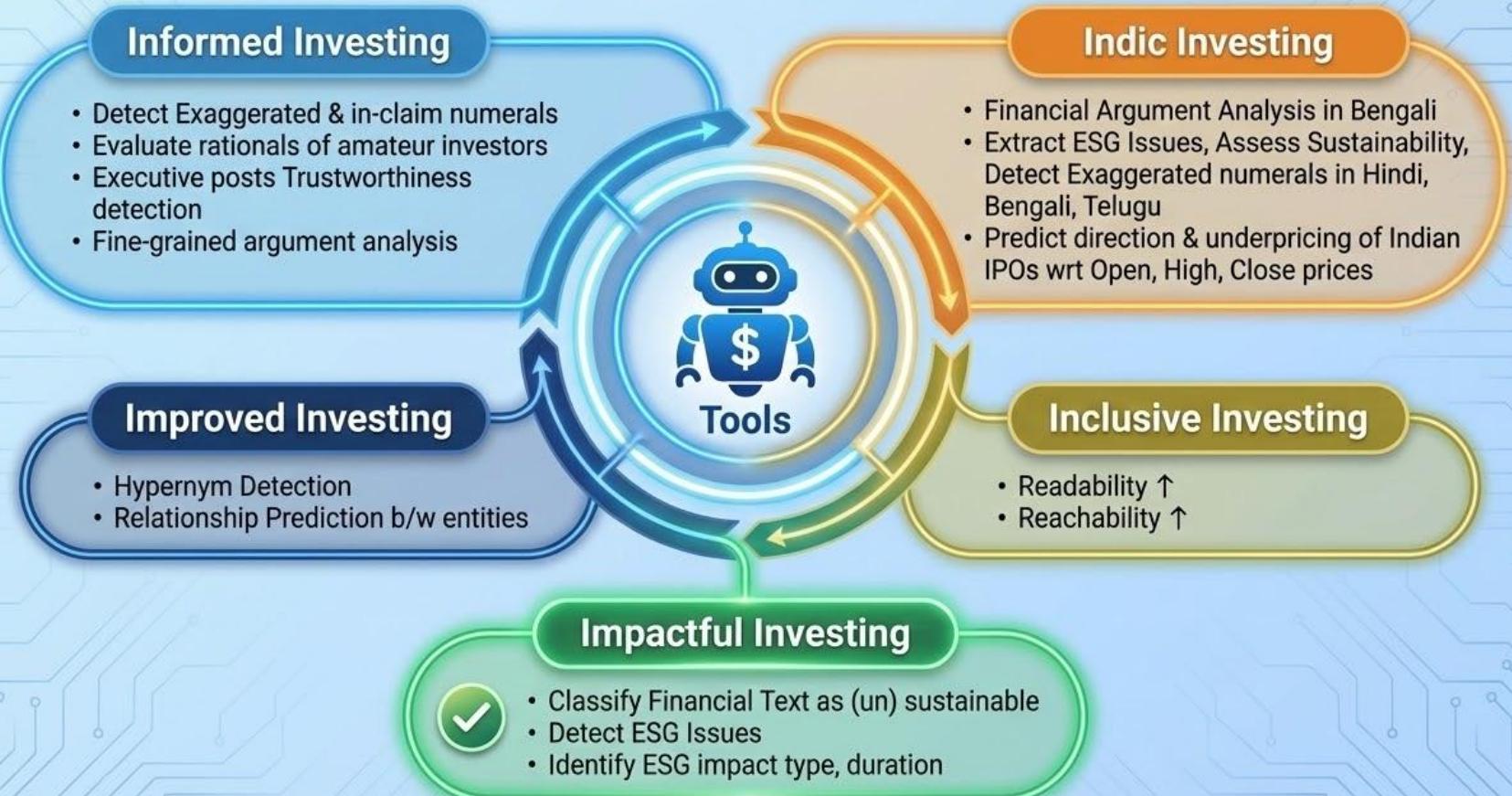
Improved Investing: Relation Extraction

Contributions:

- Framework: Mask One At a Time (MOAT)
- Benchmarked with existing open source generative Large Language Models (LLMs)



Impactful Investing



Budget 2023: Analysis of Green Finance, and Roadmap Ahead

NEWS



The government will have to spend another 11 lakh crore every year till 2030 to reach its Panchamrit target which it committed to at COP26 in Glasgow. So, will the Finance Minister push the accelerator on Green Finance, or will other sectors continue to be a priority?

ETBFSI News

Budget 2023: Analysis of green finance, and roadmap ahead

Updated: January 31, 2023, 17:16 IST

mint

NSE to set up Social Stock Exchange (SSE) as a separate segment

1 min read • 23 Dec 2022, 10:10 AM IST [Livemint](#)

Finance Minister Nirmala Sitharaman, in her Union Budget speech of 2019-20, had proposed creation of a Social Stock Exchange

Impactful (Green) Investing

Contributions:

How to ensure that the investments are towards betterment of the Earth and have positive impact towards the environment?



Detecting ESG and sustainability related concepts, issues, and impact



Assessing ESG Impact Type



Assessing ESG Impact Duration



Ranking Environment, Social And Governance Related Concepts And Assessing Sustainability Aspect Of Financial Texts

Detecting Issues Related to Environmental, Social, and Corporate Governance using SEC-BERT

A low resource framework for Multi-lingual ESG Impact Type Identification

Fine-tuning Language Models for predicting the impact of events associated to financial news articles

Venues: FinNLP @ (IJCAI, AACL, LREC_COLING), ICDSA-2023



SCAN ME



SCAN ME



SCAN ME



SCAN ME



Impactful (Green) Investing: ESG concept & Sustainability detection

Term: Eco-Design Products



Concept: Energy efficiency and renewable energy

Sub-Task-1: Assigning terms to concepts

We are striving to reduce the amount of waste we produce, and to reduce water as well as paper consumption.

Input Text



Sustainable

Carbon intensity is measured as tons of CO₂ per million dollars of revenue.

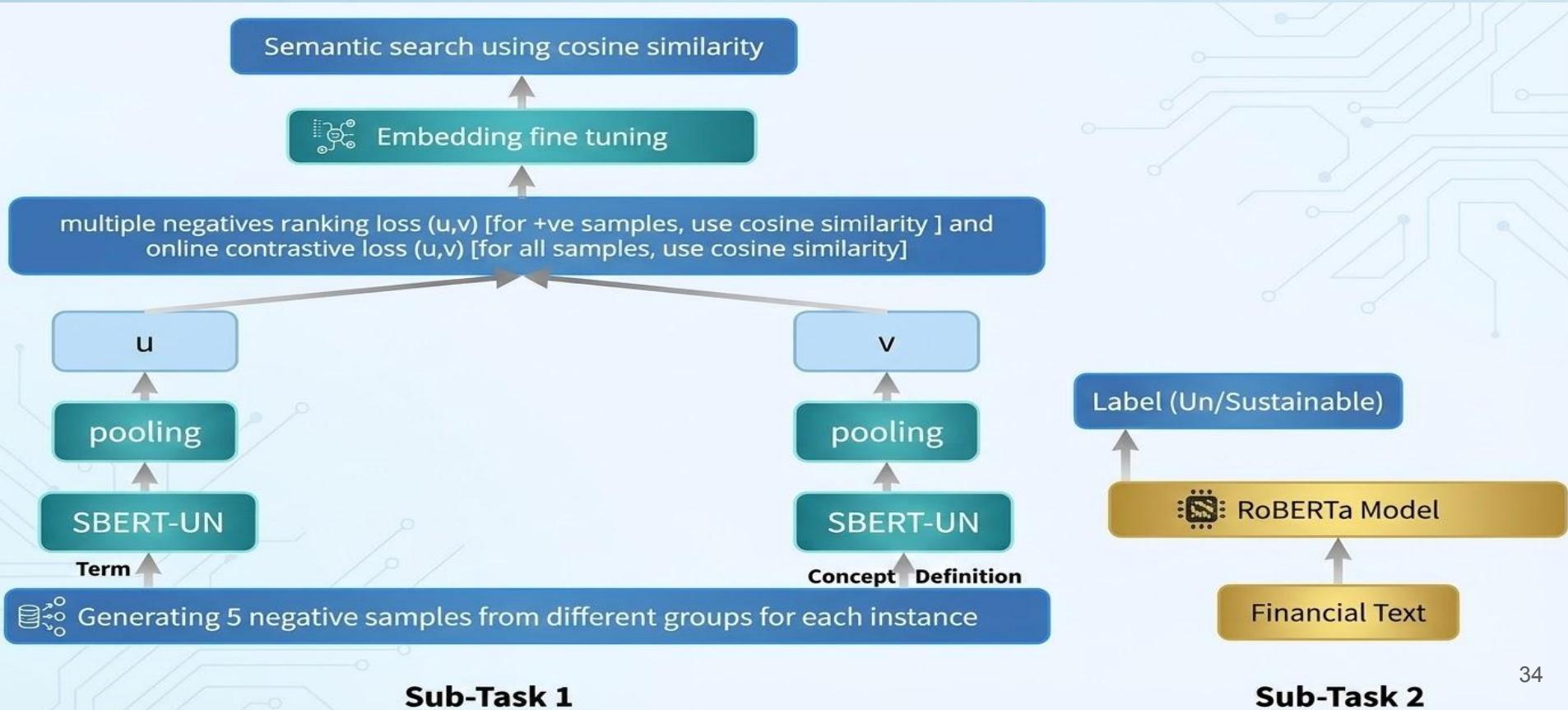
Input Text



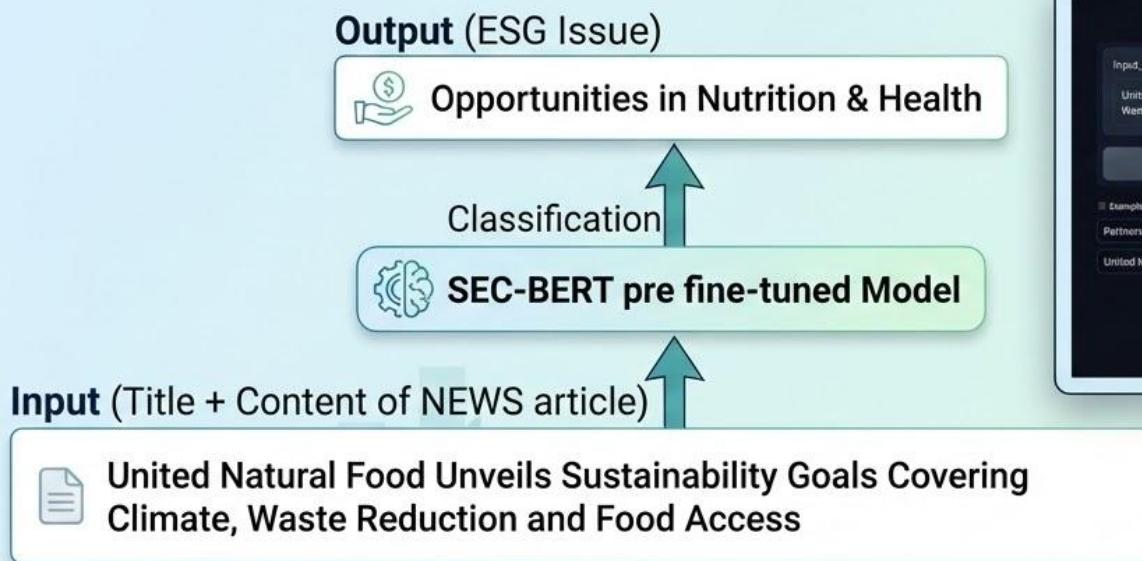
Unsustainable

Sub-Task-2: Classifying financial texts as sustainable or unsustainable

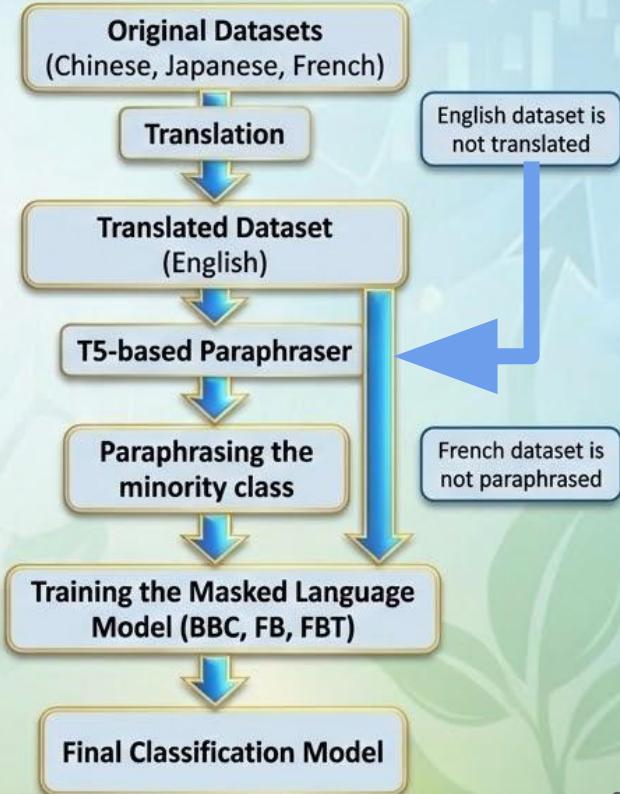
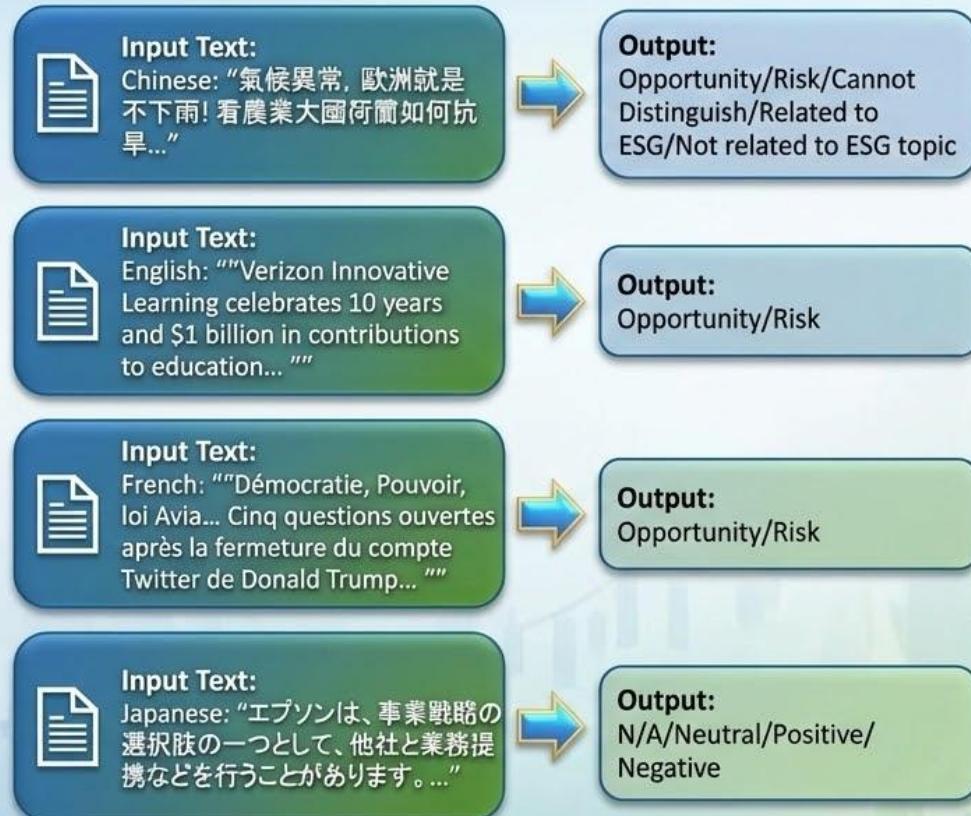
Impactful (Green) Investing: ESG concept & Sustainability detection



Impactful (Green) Investing: ESG Issue Detection

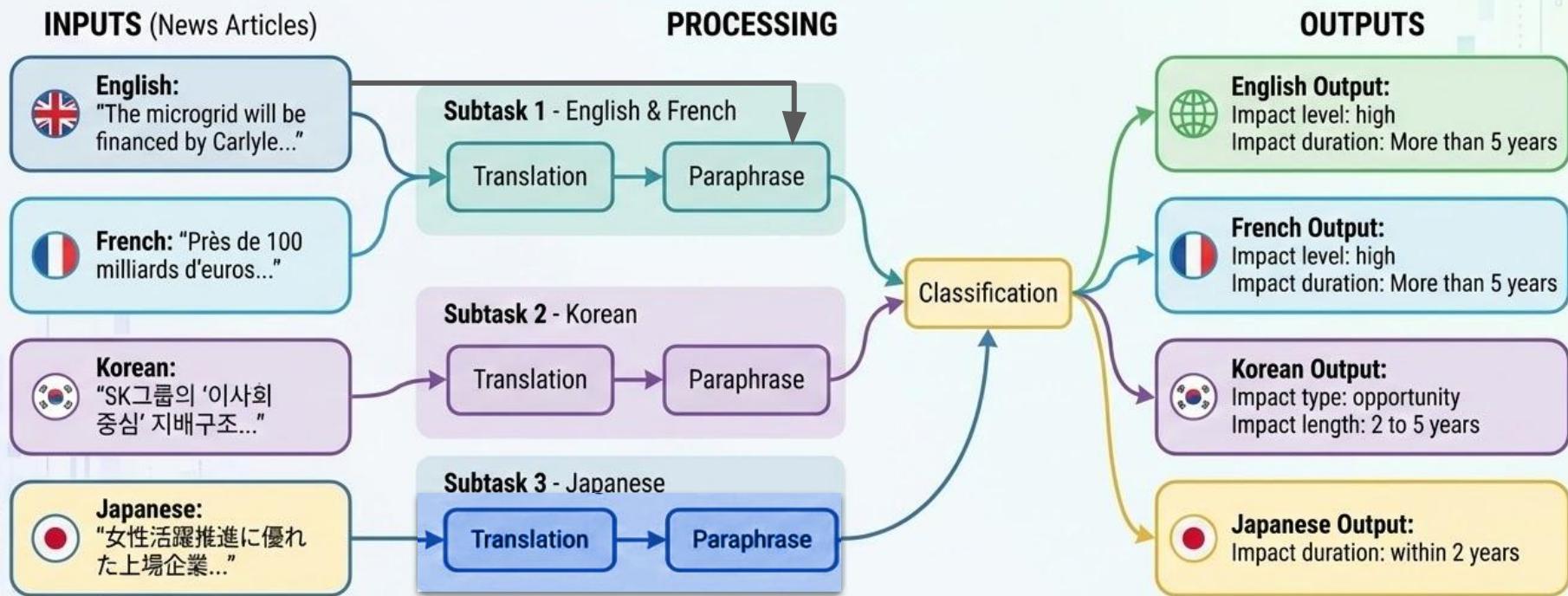


Impactful (Green) Investing: ESG Impact Type Assessment



Impactful (Green) Investing: ESG Impact Duration Assessment

UNDERLYING MODELS:  BERT  RoBERTa  FinBERT



Impactful (Green) Investing

Contributions:



Trained Sentence Transformers

Automatically ranks ESG-related concepts for unknown terms.



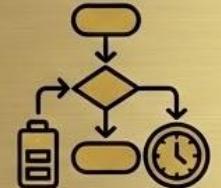
Fine Tuned RoBERTa

Classifies financial texts as 'sustainable' or 'not sustainable'.



Tool: ESG Issue Detector (EID)

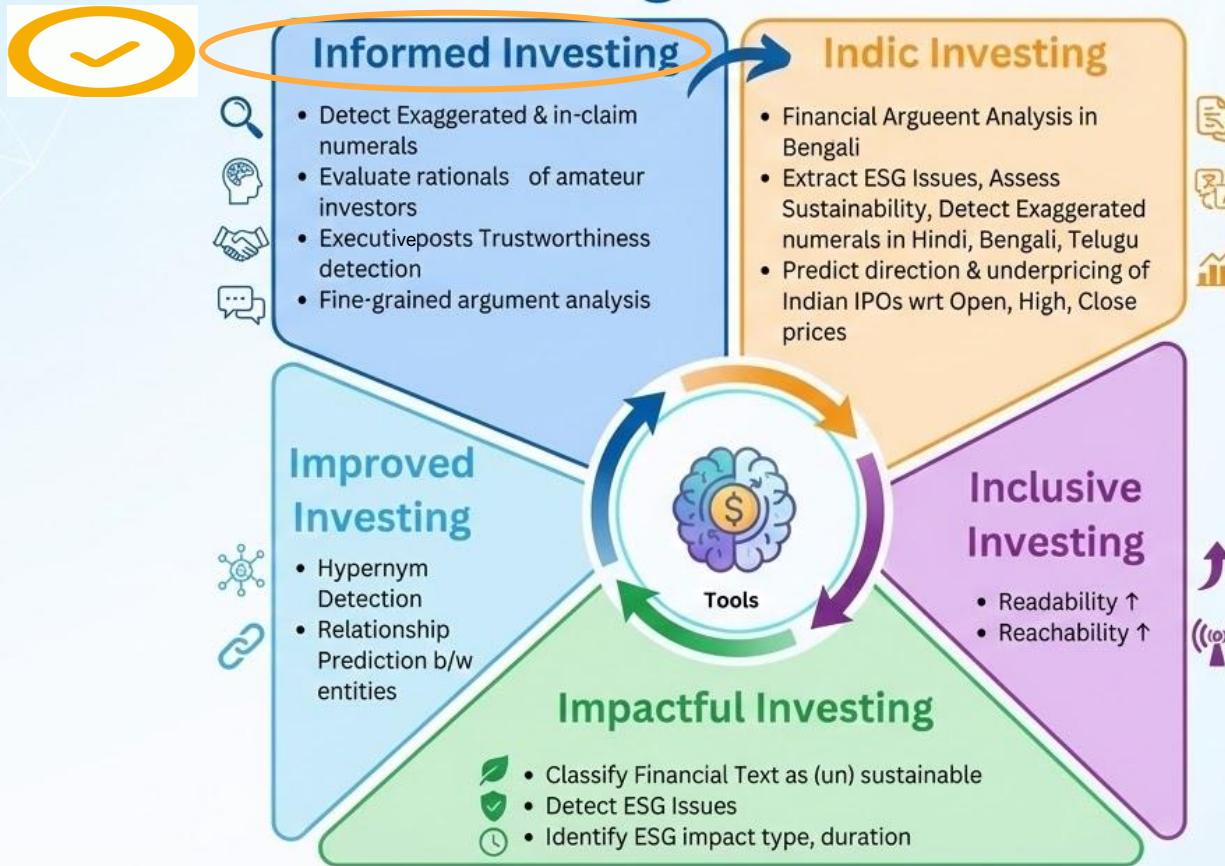
A tool developed for detecting ESG issues.



Low Resource Framework

Proposed a framework for assessing ESG Impact Type.

Informed Investing



NEWS: Financial Manipulation & Scams

BS News

**Manipulation via
Youtube: Sebi penalises
Arshad Warsi,
wife, 29 others**

Sebi noted that Arshad Warsi has made a profit of Rs 29.43 lakh and his wife has earned a profit of Rs 37.56 lakh



mint Premium | MONEY

**How finfluencers use fake
screenshots to run scams**

Neil Borate, Sashind Ningthoukhongjam | 8 min read | 31 Aug 2023, 12:00 AM IST



Informed Investing

How to safeguard investors from misinformation?



Detecting
in-claim
numerals



Detecting
Exaggerated
Numerals in
Financial
Texts



Estimating
profitability
and loss from
financial
social media



Deciding trust
worthiness of
social media
posts by
executives



Financial
Argument
Analysis

Informed Investing

Estimating profitability and loss from financial social media posts in Chinese



- Analyzes sentiment and trends in Chinese financial social media posts.
- Develops models to predict potential profitability and loss.
- Provides insights for informed investment decisions in the Chinese market.

Venue: FinNLP@EMNLP'2022

Informed Investing - Claim Detection



NTCIR 16 Conference: Proceedings of the 16th NTCIR Conference on Evaluation of Information Access Technologies, June 14-17, 2022 Tokyo Japan



LIPI at the NTCIR-16 FinNum-3 Task: Ensembling transformer based models to detect in-claim numerals in Financial Conversations

Sohom Ghosh

Sudip Kumar Naskar

[Home](#) > [International Journal of Information Technology](#) > Article

Detecting context-based in-claim numerals in Financial Earnings Conference Calls

Original Research | Published: 15 May 2022

Volume 14, pages 2599–2568, (2022) | [Cite this article](#)



International Journal of Information
Technology

[Aims and scope](#) →

[Submit manuscript](#) →

Springer, May 2022



FiNCAT: Financial Numeral Claim Analysis Tool

Authors: [Sohom Ghosh](#), [Sudip Kumar Naskar](#) [Authors Info & Claims](#)



WWW '22: Companion Proceedings of the Web Conference 2022 • April 2022 • Pages 583–585

• <https://doi.org/10.1145/3487553.3524635>

FiNCAT-2: An enhanced Financial Numeral Claim Analysis Tool

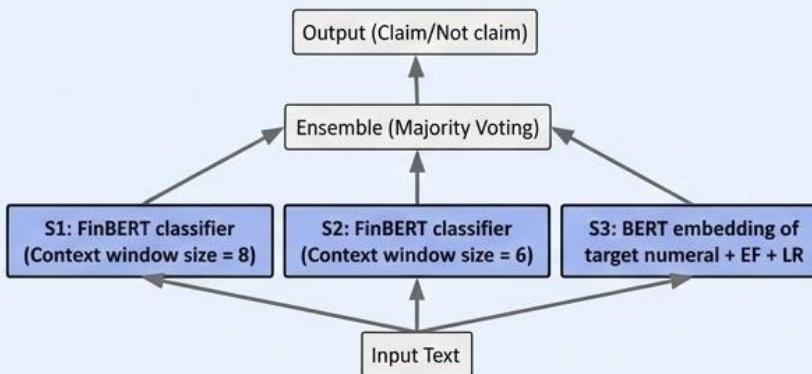
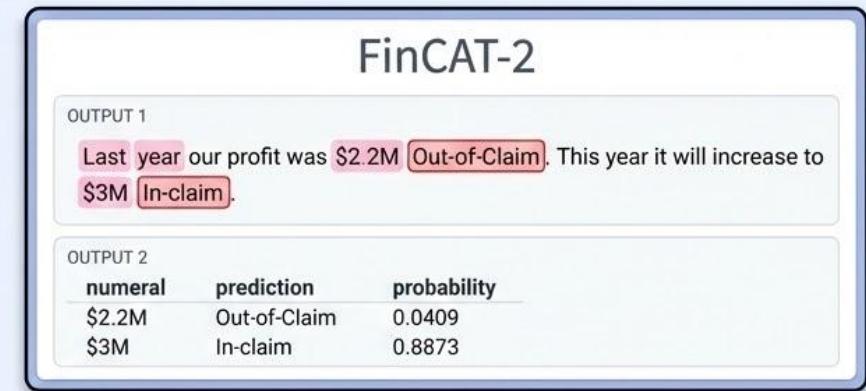


[Sohom Ghosh](#) , [Sudip Kumar Naskar](#)

Software Impacts, Elsevier, 2022



Informed Investing - Claim Detection



Informed Investing - Claim Detection

Contributions:



Claim

Not Claim

System to detect number present in a given financial text is a claim or not



FinCAT

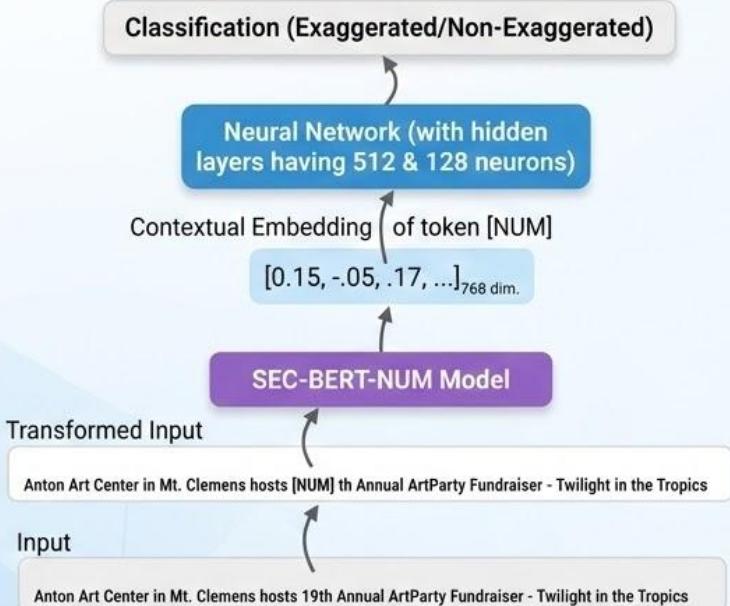
Financial Numerical
Claim Analysis Tool



FinCAT-2

Enhanced Tool with
improved accuracy
& features

Informed Investing - Exaggerated Numeral Detection



Financial Exaggerated Numeral ClassifiEr (FENCE)

Enter Financial text here
Anton Art Center in Mt. Clemens is hosting the 18th Annual ArtParty Fundraiser to raise \$500 million.

Get numerals present in the entered text

2 Numerals present in the text
Anton Art Center In Mt. Clemens ih hosting the 18th Annual ArtParty Fundraiser to raise \$500 million

3 All numerals Specific numerals 4 Get option to select numerals

5 Numerals 6 Predict for specific numerals
18 @POSITION 47 500 @POSITION 89

7 number position prediction
500 89 Non-Exaggerated

8 Examples 9
Get 50% off Gap denim whilst recycling your old denim for communities in need | Matthew Perry puts Malibu mansion on the market for \$1.5 million

The screenshot shows the 'Financial Exaggerated Numeral ClassifiEr (FENCE)' application interface. Step 1 shows the input text: 'Anton Art Center in Mt. Clemens is hosting the 18th Annual ArtParty Fundraiser to raise \$500 million.' Step 2 displays the numerals found in the text: '18', '500', and '\$500'. Step 3 shows the 'All numerals' and 'Specific numerals' tabs, with the 'Specific numerals' tab selected. Step 4 shows the 'Get option to select numerals' button. Step 5 shows the 'Numerals' section with the numbers 18 and 500 highlighted. Step 6 shows the 'Predict for specific numerals' section with the number 500 and its position 89 selected. Step 7 shows the results table with the number 500 at position 89 predicted as 'Non-Exaggerated'. Step 8 shows examples of exaggerated and non-exaggerated numerals. Step 9 shows the examples: 'Get 50% off Gap denim whilst recycling your old denim for communities in need' and 'Matthew Perry puts Malibu mansion on the market for \$1.5 million'.

ক

ব

Informed Investing - Exaggerated Numeral Detection

Contributions:



Framework: To detect Exaggerated Numerals



Tool : Financial Exaggerated Numeral ClassifiEr
(FENCE)

অ

ন

গ

ক ষ

শ

47

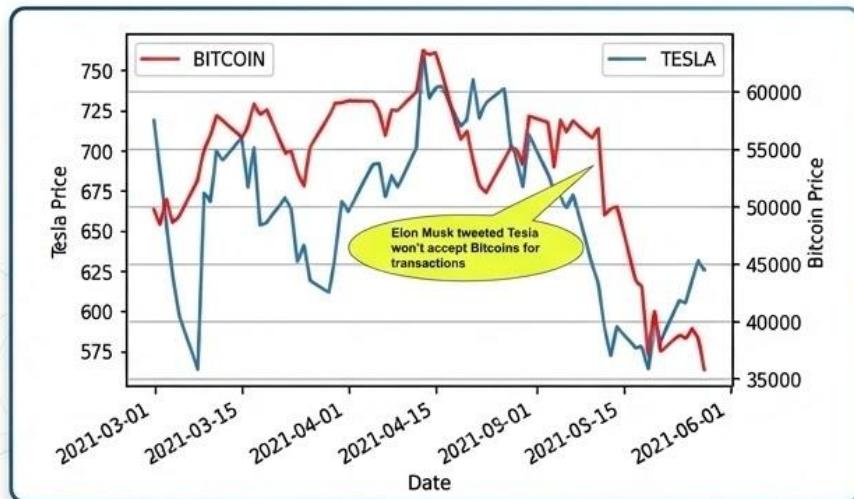
Informed Investing: Deciding trust worthiness of social media posts by executives

Evaluating Impact of Social Media Posts by Executives on Stock Prices

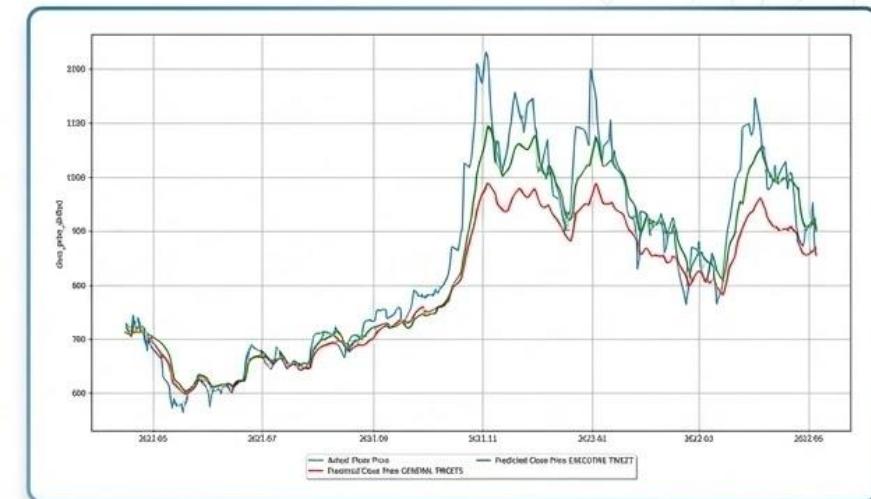


Authors: Anubhav Sarkar, Swagata Chakraborty,
Sohom Ghosh, Sudip Kumar Naskar

Informed Investing: Deciding trustworthiness of social media posts by executives



Case Study: Impact of Executive Tweets on Crypto Prices

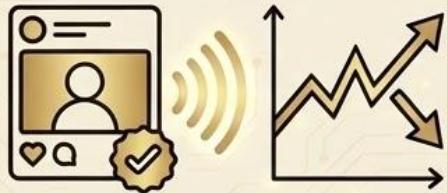


Model Performance: Predicting Stock Price Movements

Informed Investing: Deciding trust worthiness of social media posts by executives

Contributions:

Executive Influence on Price



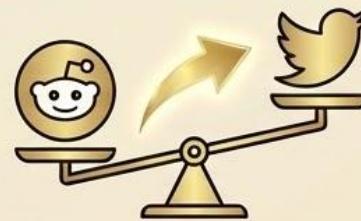
Established social media posts by executives have deeper influence on close price movements.

Sentiment-Based Prediction



Proposed how to use sentiments from social media to accurately predict the close prices.

Platform Comparison (Reddit vs. Twitter)



Proved Reddit shows a similar trend like Twitter, however, Twitter is more effective in this task than Reddit.

Informed Investing : Financial Argument Analysis

NTCIR-17 (2023)

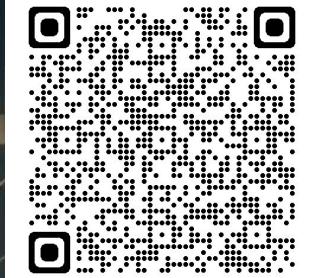


NTCIR 17 Conference: Proceedings of the 17th NTCIR Conference on Evaluation of Information Access Technologies, December 12-15, 2023, Tokyo, Japan

DOI: <https://doi.org/10.20736/0002001281>

LIPI at the NTCIR-17 FinArg-1 Task: Using Pre-trained Language Models for Comprehending Financial Arguments

Swagata Chakraborty, Anubhav Sarkar, Dhairyा Suman
(dhairyasuman@gmail.com), Sohom Ghosh, Sudip Kumar
Naskar



NTCIR-17 (2023)

Informed Investing: Financial Argument Analysis

Input Text

English: "Japan as a geography for us is a high transactional market."

Task-2 (Sub-Task 1): Argument Unit Identification (Earnings Conference Call)

Input Text-1

English: "Japan as a geography for us is a high transactional market."

Input Text-2

English: "The improvement in that in Q3 is obviously very high margin and also the bottom."

Task-2 (Sub-Task 2): Argument Relation Identification (Earnings Conference Call)

Input Text-1

Chinese: "我假定看到 7.99 脫然不一定會來 買進 預計用今年領到的股利到量金64萬多 和現金30萬 累共94萬多 全部跟他賭一把 先用用買出64萬 一拿到錢利現金 就上邊掉"

Input Text-2

Chinese: "益融財直底多價有必要一直跟這隻拼嗎 ??? 買這隻的直貼就是價低 稅利率高 未來要考慮的是 商自己已經嚴重損???生意越束越少??? 未來它的獲利?????"

Task 2: Identifying Attack and Support Argumentative Relations (Social Media)

Output
Premise/
Claim

Output
Support/
Attack/
None

Output
Support/
Attack/
None

MODEL

MODEL	MACRO-F1 (VALIDATION SET)	MICRO-F1 (VALIDATION SET)
RNN + Specy Tokenizer	0.3571	0.5270
FastTest + NN	0.7155	0.7173
Glove Embeddings + CNN	0.6952	0.6957
BART-BASE-CASED + BERT TOKENIZER	0.7336	0.7337
BERT-SEC	0.7476	0.7430
FinBERT	0.7398	0.7401

Table 2: Results of Task 1, Sub-Task 1: Argument Unit Identification

MODEL	DATA	VALIDATION	
		MICRO F1	MICRO F1
DistilBERT	Original	0.3313	0.5721
DistilBERT	Paraphrased	0.7942	0.4811
Flan-Roberta	Original	0.7971	0.3633
Flan-Roberta	Paraphrased	0.7971	0.3633
BERT-SEC	Original	0.474	0.5647
RERT-SEC	Paraphrased	0.7998	0.4900
Cross Encoder (BERT)	Original	0.7998	0.3353
Cross Encoder (BERT)	Paraphrased	0.7913	0.4956
Cross-Encoder (BERT-SEC)	Original	0.7607	0.476
Cross-Encoder (RERT-SEC)	Paraphrased	0.7651	0.4807
Cross Encoder (FinBERT Finetuned)	Original	0.8275	0.5398
Cross Encoder (MLM-FinBERT)	Original	0.8000	0.3295
Cross-Encoder (MLM-FinBERT)	Paraphrased	0.7913	0.3452

Table 3: Results of Task 2, Sub-Task 2: Argument Relation Identification

LANGUAGE	MODEL	DATA	VALIDATION	
			MICRO-F1	MICRO-F1
English	BERT-base	Original	0.6453	0.6793
English	BERT-base	Paraphrased	0.6219	0.6753
English	FLANG-RaBERTs	Paraphrased	0.6219	0.6754
English	Cross Encoder (SBERT)	Original	0.6293	0.6796
English	Cross Encoder (SBERT)	Paraphrased	0.6500	0.6850
English	Cross Encoder (DistilROBERTA)	Original	0.7933	0.7372
English	Cross Encoder (DistilROBERTA)	Paraphrased	0.6920	0.7374
English	Cross Encoder (Flang-Reberte)	Original	0.6932	0.7342
English	Cross Encoder (Flang-Reberte)	Paraphrased	0.6845	0.7141
English	Cross Encoder (RCN7-SEC)	Original	0.6893	0.7442
English	Cross Encoder (BERT-3EC)	Paraphrased	0.6090	0.7960
English	Cross Encoder (MLM on BERT-SEC)	Original	0.6946	0.7160
English	Cross Encoder (MLM on BERT-SEC)	Paraphrased	0.6871	0.7180
Chinese	SBERT-F-Chinese	Original	0.6221	0.6430
Chinese	Cross Encoder ISRBOT-Chinese	Original	0.6333	0.4432

Table 4: Result of Task 3: Identifying Argumentative Relations in Social Media Discourse

Indic Investing



Informed Investing

- Detect Exaggerated & in-claim numerals
- Evaluate rationals of amateur investors
- Execute posts Trustworthiness detection
- Fine-grained argument analysis



Indic Investing

- Financial Argument Analysis in Bengali
- Extract ESG Issues, Assess Sustainability, Detect Exaggerated numerals in Hindi, Bengali, Telugu
- Predict direction & underpricing of Indian IPOs wrt Open, High, Close prices



Improved Investing



- Hypernym Detection
- Relationship Prediction b/w entities



Inclusive Investing

- Readability ↑
- Reachability ↑



Impactful Investing

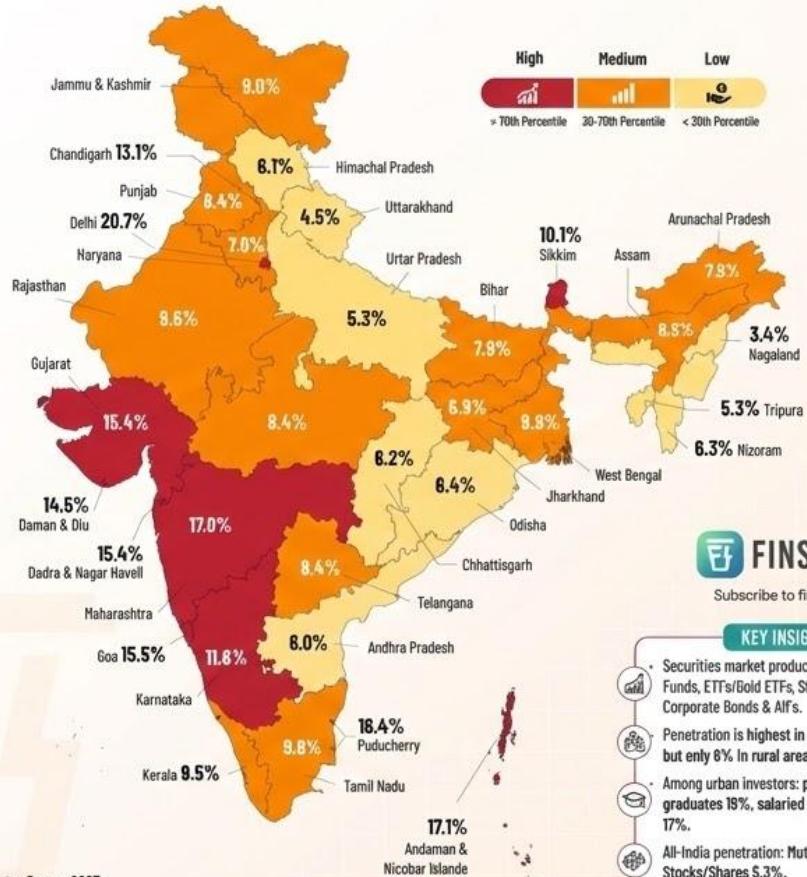
- Classify Financial Text as (un) sustainable
- Detect ESG Issues
- Identify ESG impact type, duration



Tools

ONLY 9.5% OF INDIA INVESTS!

This is despite 63% Indian households having some form of awareness about securities and market products in India.



FINSHOTS

Subscribe to finshots.in

KEY INSIGHTS

- Securities market products include Mutual Funds, ETNs/Gold ETFs, Stocks, REITs/InvITs, Corporate Bonds & AllFs.
- Penetration is highest in top 9 metros (23%) but only 8% in rural areas.
- Among urban investors: postgraduates 27%, graduates 19%, salaried 23%, self-employed 17%.
- All-India penetration: Mutual Funds/ETFs 8.7%, Stocks/Shares 5.3%.

Indic Investing

- How to keep investors informed in Indian Languages?



Financial Argument Analysis in Bengali

Rima Roy

Sohom Ghosh

Sudip Kumar Naskar



IndicFinNLP: Financial Natural Language Processing for Indian Languages

Sohom Ghosh¹, Arnab Maji², Aswartha Narayana³, Sudip Kumar Naskar⁴



Predicting Success & Ratings of Indian IPOs



Indic Investing: Financial Argument Analysis in Bengali



CONTRIBUTION

- Created two datasets in Bengali: Financial argument unit classification, Argument relation Identification
- Created a user friendly tool Financial Argument Analysis in Bengali (FAAB)



PROBLEM STATEMENT

Given:
Financial
argumentative
text in Bengali

Task-1:
Classify as 'Premise' or
'Claim'

Task-2: Classify relation:
"support", 'attack', or 'none'



GIVEN DATASETS

- Two English datasets for argument unit classification & relation identification.
- Source:** Chen et al. (Earnings Conference Call), **License:** Share-alike



PROCESS OF CREATING OUR OWN DATASETS



Translation
(IndicTrans@AI4Bharat)



Evaluation (Two-step):

- Manual 100 instances
100 instances (BertScore, CometScore)
- Automated 8000 samples
(BertScore, Labse-based cosine similarity,
Threshold-based retention)



TASK EXAMPLES

Task-1

Text: এবং এই প্রেক্ষাপট, অবশ্যই,
তারা মুরাবানো কিছু কাজের বেস্তা
তুলে নিছে এবং স্থানান্তর করছে, কিন্তু
তারা মুরো ব্যবসায়িক প্রক্রিয়ার
আধুনিকীকরণ করছে।

Label: 0 (Premise)

Task-2

Text-1: তাই প্রথমবার আমরা ঘোষণা
করেছিলাম যে আমাদের ২০ লক্ষেরও
বেশি বিজ্ঞাপনদাতা আছেন যারা
ফেসবুক বিজ্ঞাপন কিনছেন।

Text-2: ৮২% মানুষ যারা আমাদের
সাথে বিজ্ঞাপন শুন করেন তারা
আমাদের খুব সাধারণ বিজ্ঞাপন পণ্য
দিয়ে শুন করেন।

Label: 0 (No relation)

Indic Investing: Financial Argument Analysis in Bengali

Results of Task-1

⚙️ Model Name	🎯 Accuracy	🔍 Precision	↳ Recall	⚖️ F1 Score
SSB	0.700	0.690	0.694	0.692
MNB	0.714	0.691	0.745	0.717
AIB	0.711	0.700	0.712	0.706
DBMC	0.681	0.656	0.721	0.681
BBMC ★	0.719	0.697	0.745	0.721
Llama-2 (Type zero shot)	0.55	0.484	0.287	0.354
Llama-2 (Type few shot)	0.522	0.500	0.186	0.271
Llama-2 (Finetune)	0.528	0.531	0.254	0.344

SSB = sagorsarker bangla bert base
MNB = monsoon nlp bangla electra

AIB = aibharat indic bert
BBMC = bert base multilingual cased

DBMC = distilbert base multilingual cased

Indic Investing: Financial Argument Analysis in Bengali

Results of Task-2

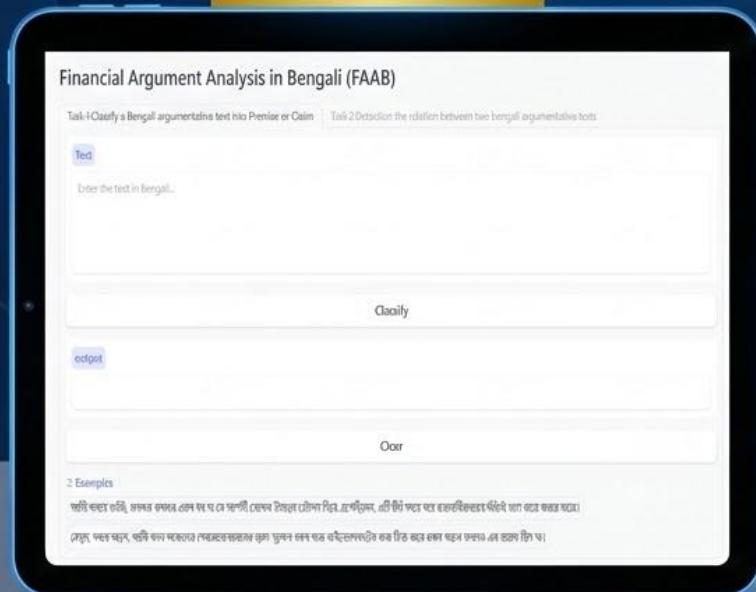
Model Name ⚙️	Accuracy 🎯	Precision Macro 🔍	Precision Macro ⬅️	Recall Micro	Recall Macro	F1 Score Micro ⚖️	F1 Score Macro ⚖️
SSB	0.708	0.708	0.442	0.708	0.417	0.708	0.419
MNB	0.705	0.705	0.235	0.705	0.333	0.705	0.275
AIB	0.682	0.682	0.476	0.682	0.348	0.682	0.303
DBMC	0.699	0.699	0.442	0.699	0.392	0.699	0.386
BBMC ★	0.755	0.755	0.488	0.755	0.460	0.755	0.460

SSB = sagorsarker bangla bert base, MNB = monsoon nlp bangla electra, AIB = aibharat indic bert,
DBMC = distilbert base multilingual cased, BBMC = bert base multilingual cased

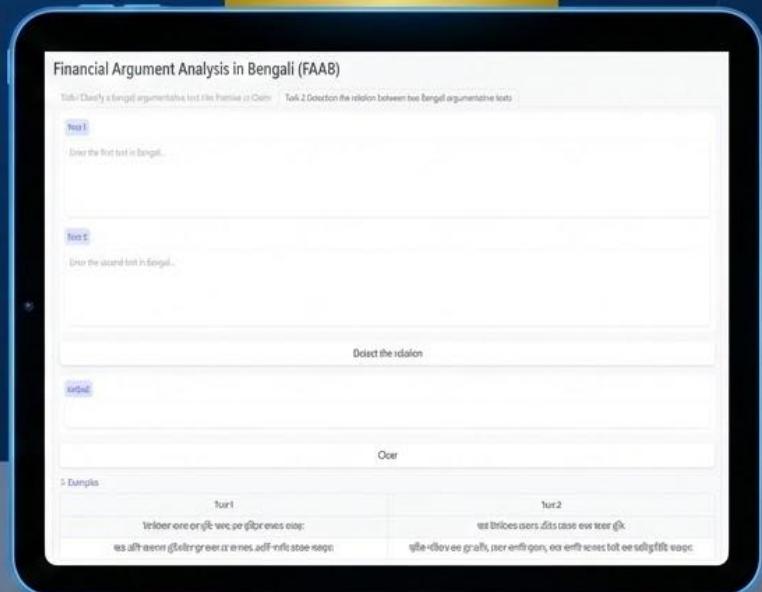
Indic Investing: Financial Argument Analysis in Bengali

User Interface Financial Argument Analysis In Bengali (FAAB)

Task-1



Task-2



Indic Investing: IndicFinNLP

PROBLEM STATEMENT

Input Text & Number

Hindi: इस **58** तारस्यांगों के जरिए आपक आपार वाती आपसी साहेदारी को आगे बढ़ा रहे हैं
Bengali: জোমেজ মাটক, **১৬** শার্লাপেন্ট বেণি মুস শান্ত নিষ্ঠা নাভজনকভাবে বাসন্ত মগ্নিত
Telugu: **500** দল প্ল্যান **৩০** ডল্লার্স মুংড়ি

Output [magnitude]
0/1/2/3/4

Task-1: Exaggerated Numeral Detection in Indic Financial Texts

Input Text



Hindi: हम नवीकरणीय विजनी का पत्रिवाद बदाना जारी रही हुद है (19 दीसीओ)। इसे अधिक से बढ़ने के लिए।
Bengali: আমামা ববাবময়োগা বিদ্যুজের পরিযাপ বৃক্ষি করে চলছি (19 টিনিউইটেন্ড বেণি বিশ্ব বাবদাব না করব।)
Telugu: మేడప స్కెర్టింగ్ క్రూపర్చుస్ రక్తవిద్యుత్తు కాన్సిన్ పెంచరం కొనసారించు (19 లీస్టింగ్ ని నీరాంపదం):

Output
Unsustainable /Sustainable

Task-2: Sustainability Assessment in Indic Financial Texts

Input Text



Hindi: एरीएन एमरो मे लेडिंग, इवेल्ट-मेट पोर्टकोलिंगो को नेट जीरी के साथ सरेखित किया
Bengali: আমা, বিপ্রিয়াগ পার্টিলাইভেন্সিফ মেট বিপ্রিয়াব সাফে সাবচ্ছেদ করতে বিজিত আগম
Telugu: ముఖ్యాలన్ పమ్పెళ్లును డేమ్‌ఫ్యాన్‌ని, ని86 సుమారు లే పెళ్లుటి ట్ర్యూలమ్ సమాంట్రమ్ డేమ్‌ఫ్యాన్‌ని ABN AMRO

Output
One of 10 ESG theme

Task-3: ESG Theme Determination in Indic Financial Texts

Languages: Hindi, Bengali, Telugu

Tasks: Exaggerated Numeral Detection, Sustainability Assessment, ESG Theme Determination

Indic Investing: IndicFinNLP

EXAGGERATED NUMERAL DETECTION DATASET

Source:

 Budget speeches of Hindi, Bengali, and Telugu-speaking states (Punjab, Uttarakhand, Haryana, West Bengal, Telangana, and Andhra Pradesh) starting from the year 2011 till 2023

 Financial texts filtered from the Samanantar corpus (Rameshet al., TACL 2022)

Language	0	1	2	3	4
Hindi	2435	3624	1444	2485	652
Bengali	1574	1886	931	1416	323
Telugu	1737	1800	983	1182	314

Table 1: Task-1 label wise distribution. 0/1/2/3/4 are the magnitudes

indic	number_english	number_indic	start_posn	end_posn	language	magnitude
মাটিল ম্যানুকান (১৯৯৪-২০০১), রিয়েল এস্টেট বিনিয়োগকারী।	2001	২০০১	22	26	bengali	3

Indic Investing: IndicFinNLP

RESULTS: EXAGGERATED NUMERAL DETECTION

हा HINDI (H)

Model	⚙️ Pr	⌚ Re	↗️ F1	✓ Acc
IB+LGB	0.44	0.49	0.45	0.49
MB+XGB	0.63	0.64	0.63	0.64
IB+XGB	0.46	0.49	0.46	0.49
MB+SVM	0.69	0.68	0.68	0.68

ବা BENGALI (B)

Model	⚙️ Pr	⌚ Re	↗️ F1	✓ Acc
IB+LGB	0.51	0.51	0.50	0.51
MB+XGB	0.62	0.62	0.61	0.62
IB+XGB	0.51	0.50	0.48	0.50
MB+SVM	0.66	0.65	0.65	0.65

ఎ TELUGU (T)

Model	⚙️ Pr	⌚ Re	↗️ F1	✓ Acc
IB+LGB	0.44	0.46	0.44	0.46
MB+XGB	0.59	0.60	0.59	0.60
IB+XGB	0.41	0.43	0.41	0.43
IB+XGB	0.69	0.68	0.68	0.68

Table 4: Tasks (Ts) 1, 2, 3 results for Languages (L) Hindi (H), Bengali (B), Telugu (T). E=English, -P=-Paraphrased, IB=IndicBERT, MB=MBERT, XGB=XGBoost, LGB=LightGBM, Pr=Precision, Re=Recall, Acc=Accuracy. **Bold** means the best. MBERT is Multi-ligual BERT

Indic Investing: IndicFinNLP

SUSTAINABILITY ASSESSMENT DATASET

_SOURCE & PROCESS

Translated existing dataset proposed by Kangand El Maarouf (FinSim4-ESG FinNLP-2022)

from English to Indian languages (Hindi, Bengali, and Telugu).

Retained only the high quality ones.

Quality Metric:
BERTScore (BS(F1)) &
Cosine Similarity (Sim.)

ય HINDI (H)

METRICS

BS(F1): ≥ 0.90
Sim.: ≥ 0.75

CLASS DISTRIBUTION

S (Sustainable): 1212
US (Unsustainable): 1026

অ BENGALI (B)

METRICS

BS(F1): ≥ 0.88
Sim.: ≥ 0.68

CLASS DISTRIBUTION

S (Sustainable): 1203
US (Unsustainable): 1025

త TELUGU (T)

METRICS

BS(F1): ≥ 0.88
Sim.: ≥ 0.80

CLASS DISTRIBUTION

S (Sustainable): 1119
US (Unsustainable): 953

EXAMPLE (HINDI)

sentence_indic	label	language
आपके संगठन का सकल वैश्विक स्कोर 1 उत्सर्जन मीट्रिक टन में कितना था?	unsustainable	hindi

Table 2: Task-2 data distribution & thresholds. S=Sustainable, US=Unsustainable. BS=BERTScore, Sim.=Cosine Similarity

Indic Investing: IndicFinNLP

RESULTS: SUSTAINABILITY ASSESSMENT

HINDI (H)	BENGALI (B)	TELUGU (T)
IB Pr: 0.86 Re: 0.86 F1: 0.86 Acc: 0.86	IB Pr: 0.80 Re: 0.80 F1: 0.80 Acc: 0.80	IB Pr: 0.79 Re: 0.79 F1: 0.79 Acc: 0.78
MB Pr: 0.77 Re: 0.77 0.77 0.77 0.77 0.77	MB Pr: 0.76 Re: 0.76 0.76 0.76 0.76 0.76	MB Pr: 0.90 Re: 0.89 0.89 0.89 0.89 0.89
MLM-IB Pr: 0.29 Re: 0.54 0.29 0.54 0.38 0.54	MLM-IB Pr: 0.81 Re: 0.81 0.81 0.81 0.81 0.81	MLM-IB Pr: 0.90 Re: 0.90 0.90 0.90 0.90 0.90
E-Ro Pr: 0.95 Re: 0.95 0.95 0.95 0.95 0.95	E-Ro Pr: 0.92 Re: 0.92 0.92 0.92 0.92 0.92	E-Ro Pr: 0.92 Re: 0.92 0.92 0.92 0.92 0.92

Table 4: Tasks (Ts) 1, 2, 3 results for Languages (L) Hindi (H), Bengali (B), Telugu (T). Model Abbreviations: E=English, -P= -Paraphrased, IB=IndicBERT, MB=MBERT, XGB=XGBoost, LGB=LightGBM, Ro=RoBERTa. Metric Abbreviations: Pr=Precision, Re=Recall, Acc=Accuracy. Note: **bold** means the best.

Indic Investing: IndicFinNLP

ESG THEME DETERMINATION DATASET



ENVIRONMENTAL (पर्यावरण / पतिवरण)



Climate Change

लिलात चग, जानकात धूम



Pollution

पालुमुळा, प्रिवेश



Resource Management

नृष्टा मानव्यों, प्रिवेश



SOCIAL (सामाजिक / सामाजिक)



Human Rights

भुमन रिक्षर, कहान लाति



Labor Standards

भाल दांभ, सगालन कंक्ष



Community Impact

कममुति अवृत, नासित जहत



GOVERNANCE (शासन / शासन)



Business Ethics

बर्सश्रस सत्तादल, प्राशन



Board Diversity

बार्ड विद्यारा, लानूत, शासन



Corruption

करपृपतिन, जाडरा, शासन



MULTILINGUAL DATASET

English, Hindi, Bengali, Telugu Data

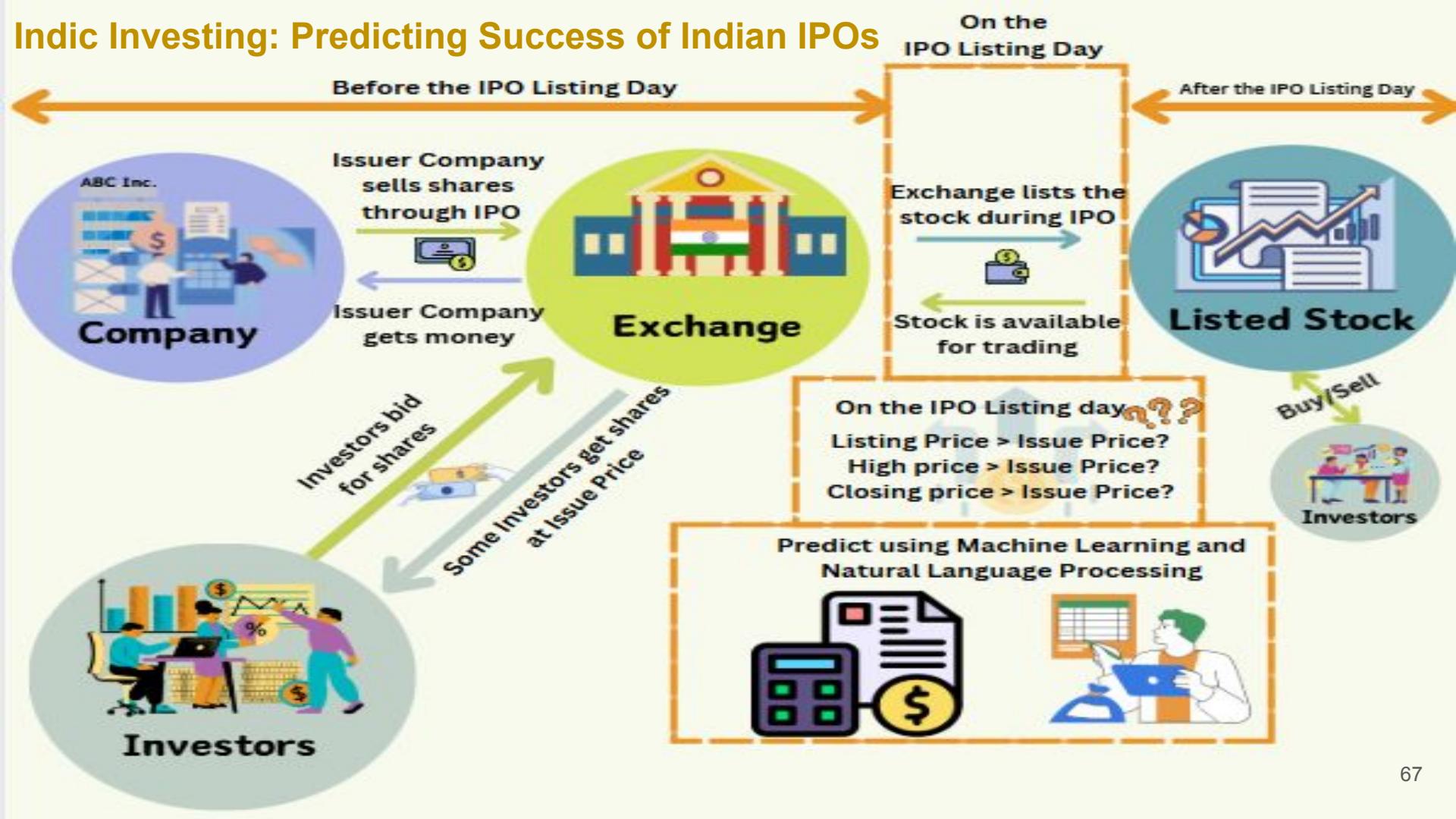


Indic Investing: IndicFinNLP

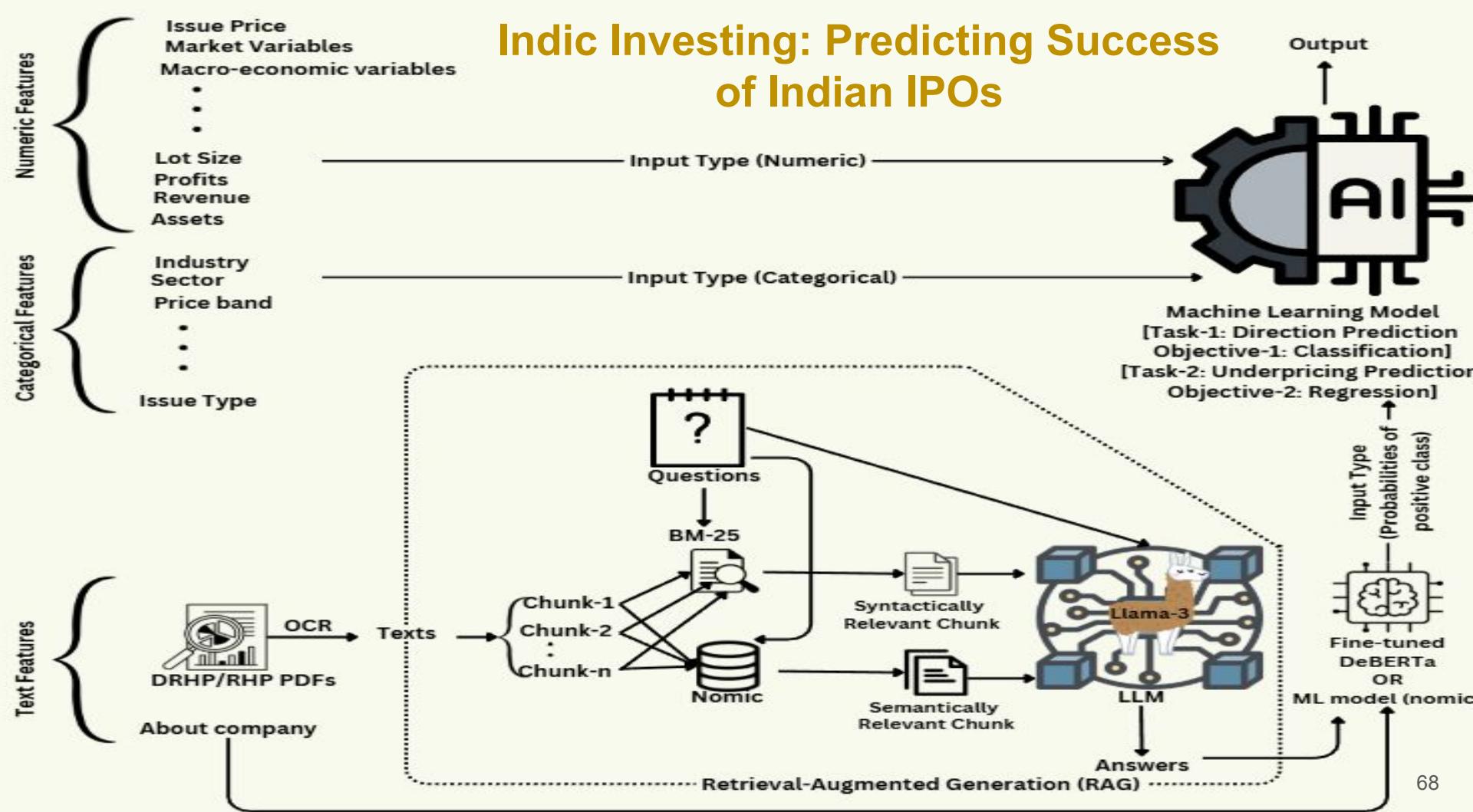
RESULTS: ESG THEME DETERMINATION



Indic Investing: Predicting Success of Indian IPOs



Indic Investing: Predicting Success of Indian IPOs





Indic Investing: Predicting Success of Indian IPOs



Datasets

Two multi-modal datasets
(Main Board IPOs & SME IPOs).



Methodology

Proposed Machine Learning and NLP based decision system for predicting if an IPO will be successful.



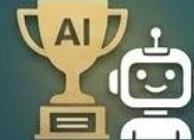
Framework

To extract important portions from documents like DHRP, RHP and used them as features.



Relevance

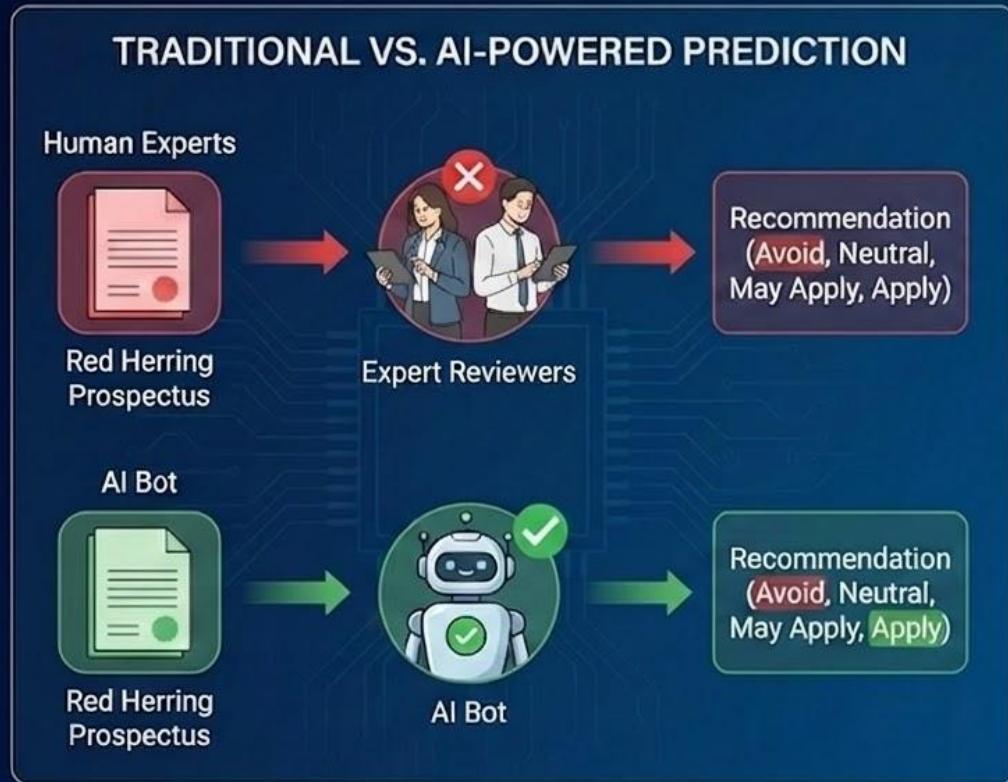
Investigate the relation between of GMP and success of an IPO.



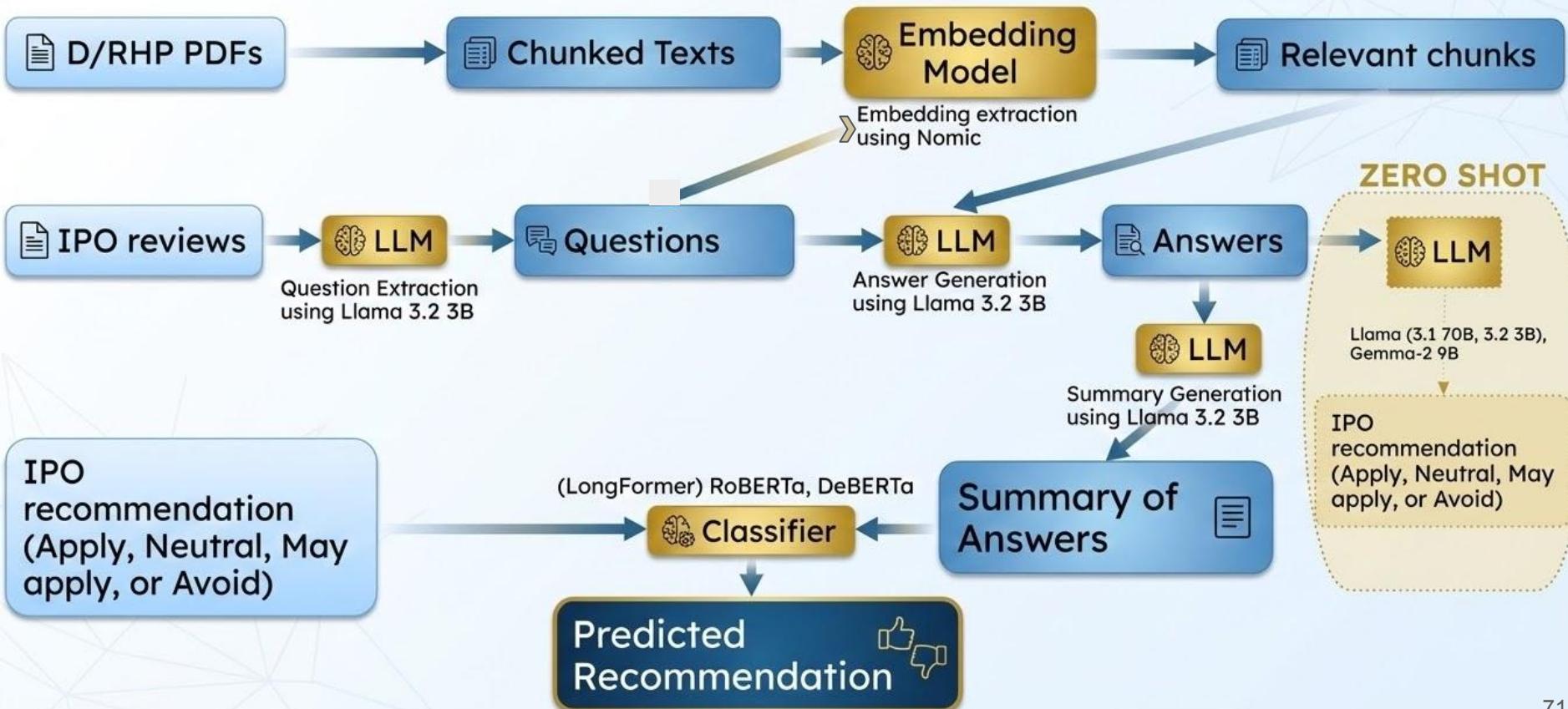
Benchmark

Evaluate the performance of Large Language Models (LLMs) like Gemini and Gemini and Llama in predicting success.

Indic Investing: Predicting Ratings of Indian IPOs



Indic Investing: Predicting Ratings of Indian IPOs





Indic Investing: Predicting Ratings of Indian IPOs



New Datasets

- 📁 Two new India-specific datasets (Main Board IPOs & SME IPOs) for rating prediction task.



Methodology & Performance

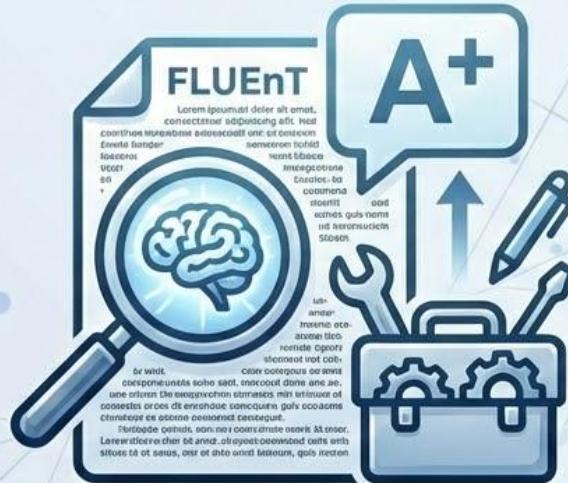


Novel method mining prospectus using Retrieval Augmented Generation & fine-tuned small encoder.



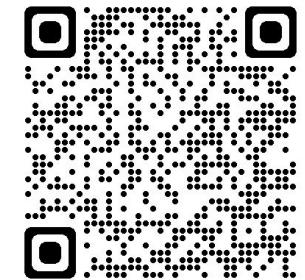
Outperforms state-of-the-art Large Language Models (LLMs) like Llama-3 under zero shot settings.

Tools for FinNLP

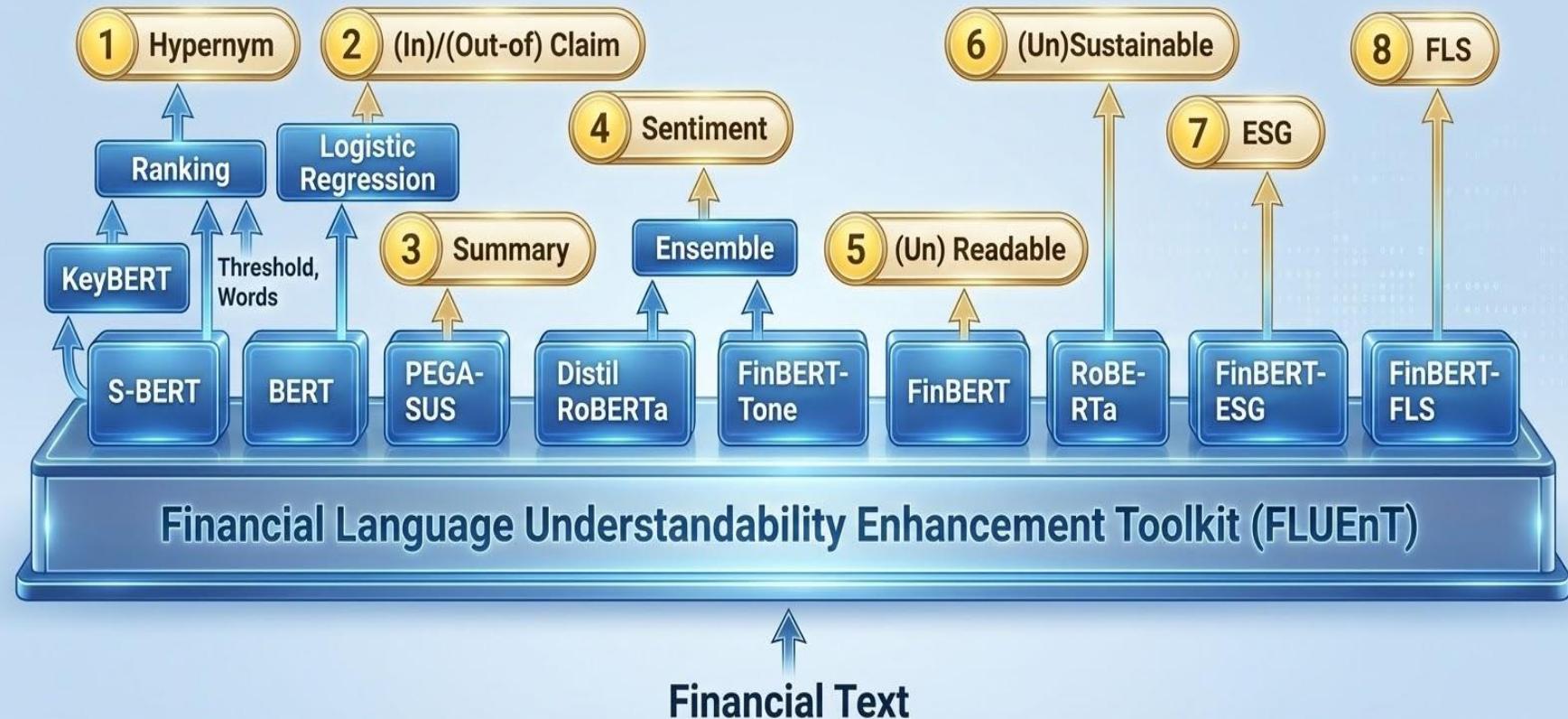


FLUEnT: Financial Language
Understandability Enhancement Toolkit

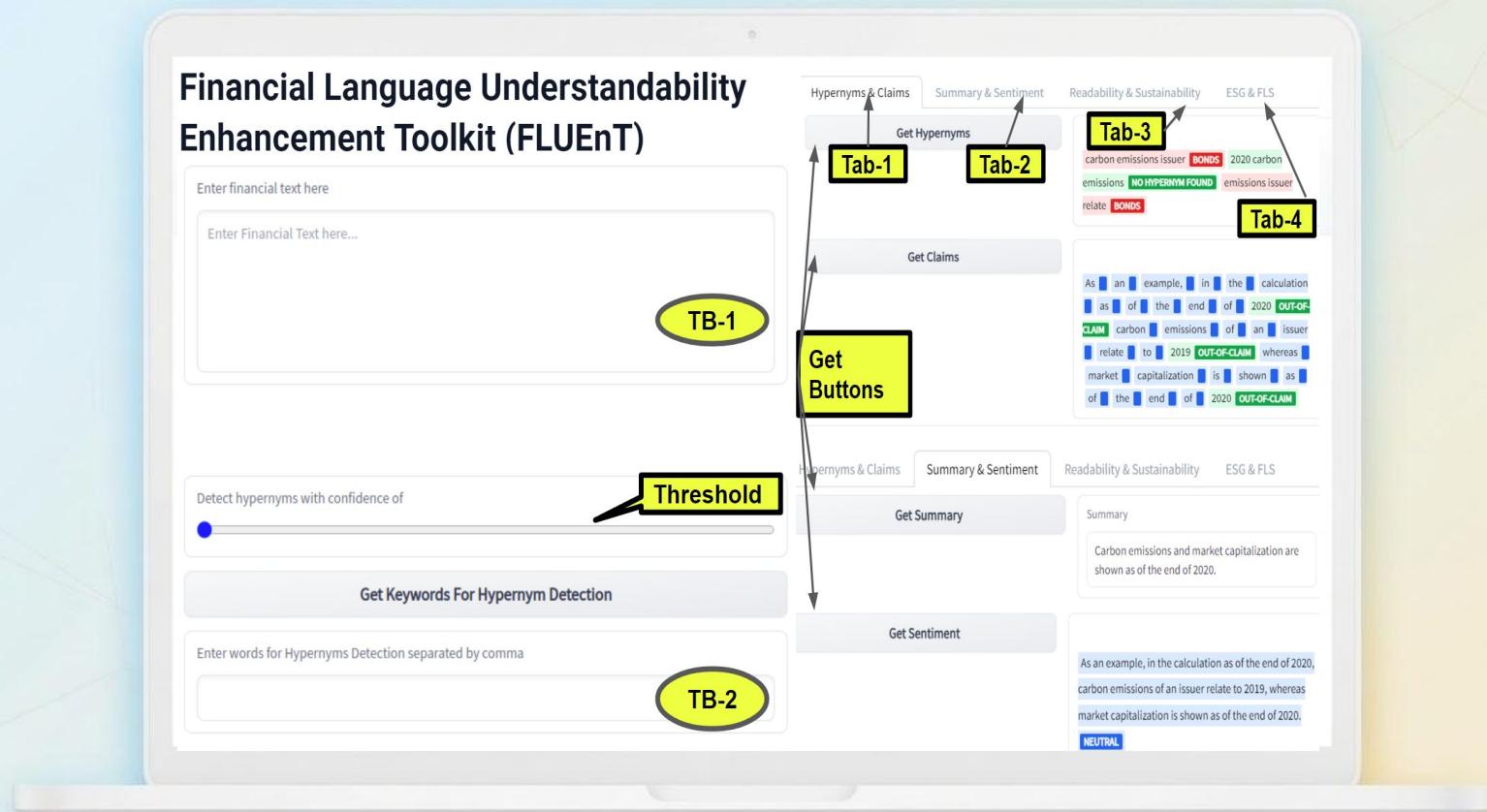
CODS
COMAD-2023



Tools for FinNLP



Tools for FinNLP



Conclusions & Future Works

Key Research Tasks

- 1. Assess Readability & Simplify FTs
- 2. Assess Reachability of FTs
- 3. Retrieve Hypernyms for Financial Jargon
- 4. Determine Relationship between Financial Entities
- 5. Classify FTs as Sustainable/Unsustainable
- 6. Detect ESG Issues from English FTs
- 7. Identify ESG Impact Type & Duration
- 8. Detect Exaggerated & In-claim Numerals
- 9. Evaluate Rationals of Amateur Investors
- 10. Evaluate Trustworthiness of Executive Posts
- 11. Fine-grained Argument Understanding in FTs
- 12. Financial Argument Analysis in Bengali
- 13. Extract ESG Issues & Detect Numerals (Hindi, Bengali, Telugu)
- 14. Predict Success of Indian IPOs
- 15. Predict Ratings of Indian IPOs
- 16. Develop Tools for Processing FTs

Summary of Contributions

CIKM-2024		
	16	Tasks
	8	Datasets
	6	Tools
	24	Papers
	8	Languages
	2	Workshops
	4	Journals
	8	Shared Tasks
	10	Conferences

Conclusions & Future Works

Focussed on improving financial well-being for social good

Key Themes

-  Inclusive
-  Improved
-  Impactful (Green)
-  Informed
-  Indic

Developed Tools for FinNLP

- FinRAD, FinRead, FinLanSer, CReAM, EID, FAAB, etc.

Challenges

- Lack of annotation expertise in vernacular languages & finance



Next Steps

- Embracing multi-modality (images, speeches, video)
- Financial Reasoning through Reinforcement Learning
- Measuring Real World Impact
- AI agents for FinNLP

I Hyderabad

Hyderabad school introduces financial literacy classes

A dedicated textbook on financial literacy has been introduced for these classes and is taught once a week as part of the curriculum.



Add as a preferred source on Google

Representative image Photo | Express

Express News Service

Updated on: 15 Oct 2025, 10:09 am • 1 min read



perplexity.ai/finance

Enter Symbol

India Markets  Crypto

NIFTY 50  ~NSEI	▲ 0.4% +97.65
24,487.4	
	
S&P BSE Sen...  ^BSESN	▲ 0.46% +568.49
80,235.59	
	
Nifty Bank In...  ^NSEBANK	▲ 0.84% +467.05
55,043.7	
	
Bitcoin  BTCUSD	▲ 1.04% +\$1,236.16
\$119,944.10	

Market Summary Updated 16 minutes ago

Banking & Financials Lead Market Decline
Heavyweight stocks like HDFC Bank (-1.30%), ICICI Bank (-1.06%), and Bajaj Finance (-2.83%) dragged the Sensex, Nifty 50, and Nifty Bank indices lower. Persistent foreign investor (FII) outflows and sector-wide weakness contributed to the negative sentiment.

Auto & IT Stocks Provide Support
Among the major gainers, Maruti Suzuki (+1.98%), Tech Mahindra (+2.03%), and Mahindra & Mahindra (+1.58%) bucked the weak trend. Select IT and auto

Ask any question about finance

Good News

The Economic Times 

English Edition  | Today's ePaper

Home  ETPrime Markets Market Data  Masterclass  News Industry SME Politics  Wealth 

Best of Wealth Tax  New Income Tax Bill  Save  Invest  Insure  Borrow  Earn  Legal / Will Plan 

Business News  Wealth  Save 

10 must-try AI finance tools in 2025 that help you save, invest, and grow faster

Curated by Lavanya Mallidi, ET Online | Oct 20, 2025, 11:14:39 AM IST

Acknowledgements



Guide/Supervisor

Dr. Sudip Kumar Naskar

Mentor:

Dr. Chung-Chi Chen



Parents:

Baba & Ma



Collaborators

Anubhav Sarkar, Swagata Chakraborty,
Dhairya Suman, Rima Roy, Arnab
Maji, Neelabha Banerjee, Harsha
Vardhan, Awartha Narayana, etc

Lab-mates

Atanu Mondal,
Madhusudan Ghosh, etc.



ACM India:

CODS COMAD-2024, PIC-2025,
ARCS-2025 Travel Grant,
SIGIR Travel Grant for CIKM-2024



Friends:

Pramit Bhattacharya, Nildari
Prosad Lahiri, Angan Mitra, etc.



Employer:

Fidelity Investments

Upcoming Presentation

InFiNITE (∞): Indian Financial Narrative Inference Tasks & Evaluations

Sohom Ghosh, Arnab Maji, Sudip Kumar Naskar

─ The 5th Workshop on ‘Evaluation & Comparison of NLP Systems’ | ─ 23rd December 2025
─ co-located at IJCNLP-AACL 2025, IIT Mumbai, India

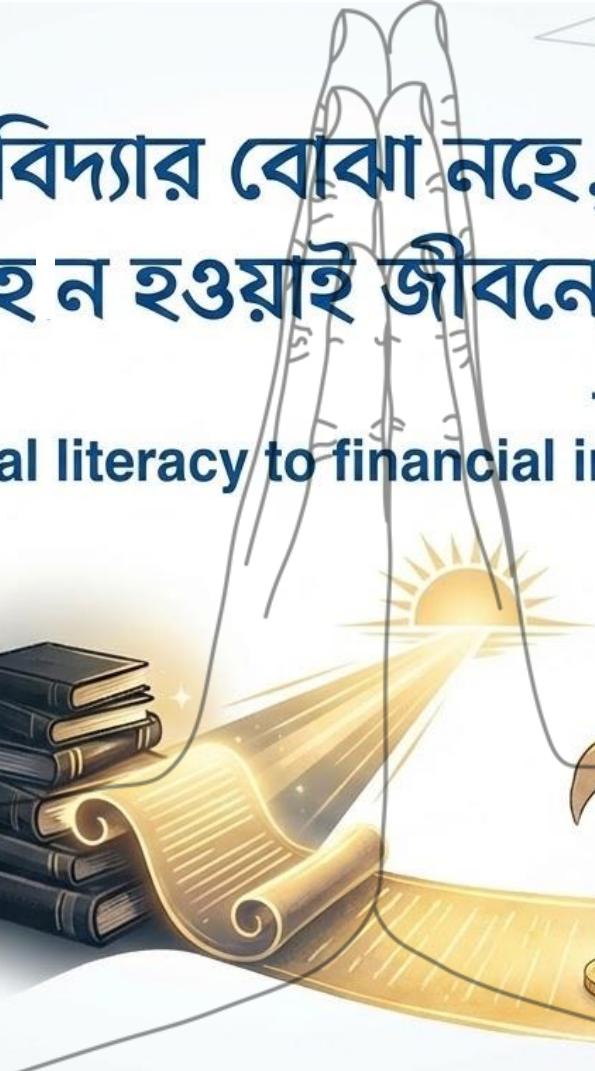
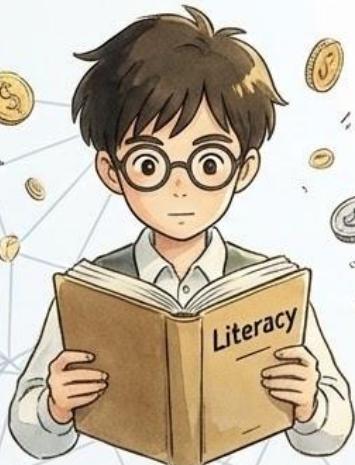
The way ahead!

“বিদ্যার বোৰা নহে,
বিদ্যার বাহ ন হওয়াই জীবনের লক্ষ্য”



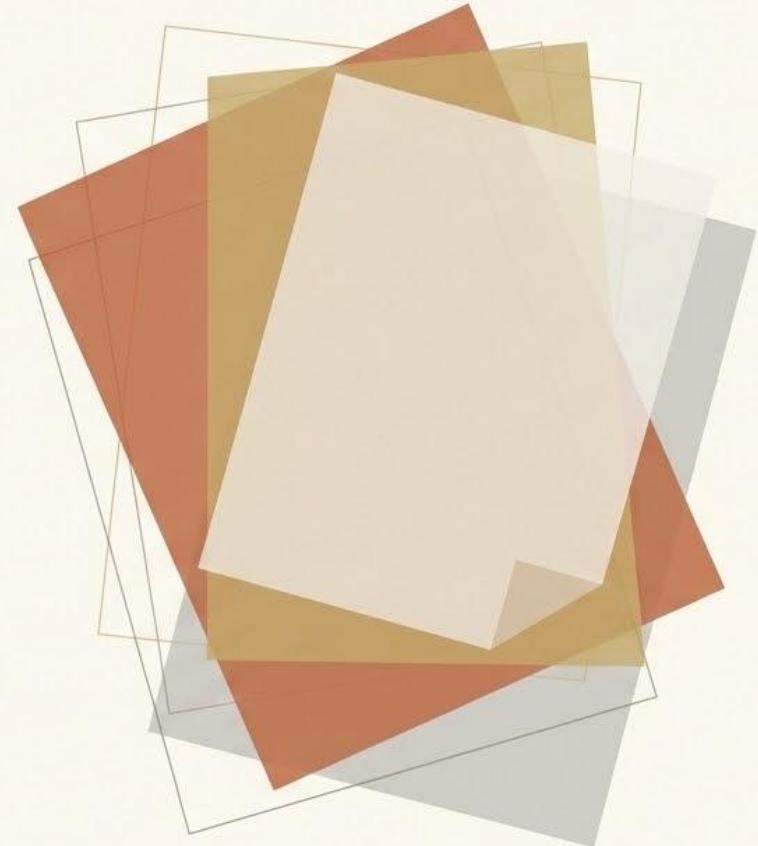
- রবি ঠাকুর

from financial literacy to financial intelligence



APPENDIX

Supporting Materials, Data, and References



Research Papers Published in Journals

- **Sohom Ghosh**, Sudip Kumar Naskar, “Detecting context-based in-claim numerals in Financial Earnings Conference Calls”, in International Journal of Information Technology, Springer, May 2022 <https://doi.org/10.1007/s41870-022-00952-7>
- **Sohom Ghosh**, Sudip Kumar Naskar, “Recent trends in financial natural language processing research”, in Science Talks, Elsevier, Volume 8, 100270, (2023) <https://doi.org/10.1016/j.sctalk.2023.100270>
- **Sohom Ghosh**, Ankush Chopra, Sudip Kumar Naskar, “Learning to Rank Hypernyms of Financial Terms Using Semantic Textual Similarity”, in SN Computer Science, Springer, 4, 610 (2023), <https://doi.org/10.1007/s42979-023-02134-z>
- **Sohom Ghosh**, Sudip Kumar Naskar, “Fincat-2: An enhanced Financial Numeral Claim Analysis Tool”, in Software Impacts, VOLUME 12, 100288, Elsevier, May 2022. <https://doi.org/10.1016/j.simpa.2022.100288>

Published Research Papers in Conferences / Workshops

- **Sohom Ghosh**, "Demystifying Financial Texts Using Natural Language Processing", In CIKM-2024, DOI: <https://doi.org/10.1145/3627673.3680258>
- Neelabha Banerjee, Anubhav Sarkar, Swagata Chakraborty, **Sohom Ghosh**, Sudip Kumar Naskar, "Fine-tuning Language Models for predicting the impact of events associated to financial news articles", In FinNLPKDF-EcoNLP@ LREC-COLING, May 2024, DOI:<https://aclanthology.org/2024.finnlp-1.25.pdf>
- **Sohom Ghosh**, Arnab Majhi, Aswartha Narayana, "IndicFinNLP: Financial Natural Language Processing for Indian Languages", In LREC-COLING 2024, May 2024. DOI: <https://aclanthology.org/2024.lrec-main.789.pdf>
- **Sohom Ghosh**, Chung-Chi Chen, Sudip Kumar Naskar, "Generator-Guided Crowd Reaction Assessment", In TheWebConf (WWW-2024) <https://doi.org/10.1145/3589335.3651512>
- **Sohom Ghosh**, Sachin Umrao, Chung-Chi Chen, Sudip Kumar Naskar, "The Mask One At a Time Framework for Detecting the Relationship between Financial Entities", In FIRE 2023, Dec 2023.<https://doi.org/10.1145/3632754.3632756>
- Rima Roy, **Sohom Ghosh**, Sudip Kumar Naskar, "Financial Argument Analysis in Bengali", In FIRE 2023, Dec 2023: <https://doi.org/10.1145/3632754.3632763>
- Swagata Chakraborty, Anubhav Sarkar, Dhairy Suman, **Sohom Ghosh**, Sudip Kumar Naskar, "LIPI at the NTCIR-17 FinArg-1 Task: Using Pre-trained Language Models for Comprehending Financial Arguments", In NTCIR-17, Dec 2023. <https://doi.org/10.20736/0002001281>
- N Harsha Vardhan, **Sohom Ghosh**, Ponnurangam Kumaraguru, Sudip Kumar Naskar, "A low resource framework for Multi-lingual ESG Impact Type Identification", In FinNLP@IJCNLP-AACL, Nov 2023.<https://aclanthology.org/2023.finnlp-2.8/>

Published Research Papers in Conferences / Workshops

- Priyank Soni, **Sohom Ghosh**, Sudip Kumar Naskar, “Detecting Issues Related to Environmental, Social, and Corporate Governance using SEC-BERT”, in proceedings of 4th International Conference on Data Science and Applications (ICDSA 2023), Jaipur, India https://doi.org/10.1007/978-981-99-7820-5_27
- **Sohom Ghosh**, Sudip Kumar Naskar, “Using Natural Language Processing to Enhance Understandability of Financial Texts”, in proceedings of 6th Joint International Conference on Data Science & Management of Data (10th ACM IKDD CODS and 28th COMAD), Mumbai, India <https://doi.org/10.1145/3570991.3571051> [Honourable Mention (YRS Track)]
- **Sohom Ghosh**, Sudip Kumar Naskar, “FLUEnT: Financial Language Understandability Enhancement Toolkit”, in proceedings of 6th Joint International Conference on Data Science & Management of Data (10th ACM IKDD CODS and 28th COMAD), Mumbai, India <https://doi.org/10.1145/3570991.3571067>
- Anubhav Sarkar, Swagata Chakraborty, **Sohom Ghosh** and Sudip Kumar Naskar, “Evaluating Impact of Social Media Posts by Executives on Stock Prices”, in proceedings of the 14th meeting of Forum for Information Retrieval Evaluation (FIRE-2022), Kolkata, India. <https://doi.org/10.1145/3574318.3574339>
- **Sohom Ghosh**, Sudip Kumar Naskar, “LIPI at the FinNLP-2022 ERAI Task: Ensembling Transformers for assessing Maximum Possible Profit and Loss from online financial posts”, in Proceedings of the fourth workshop on Financial Technology and Natural Language Processing (collocated with EMNLP 2022), pp 111-115, Abu Dhabi, UAE (2022) <https://aclanthology.org/2022.finnlp-1.13.pdf>

Published Research Papers in Conferences / Workshops

- **Sohom Ghosh**, Sudip Kumar Naskar, “Ranking Environment, Social And Governance Related Concepts And Assessing Sustainability Aspect Of Financial Texts”, in Proceedings of the fourth workshop on Financial Technology and Natural Language Processing (collocated with IJCAI-ECAI 2022), pp 243-249, Vienna, Austria (2022) <https://aclanthology.org/2022.finnlp-1.33.pdf>
- **Sohom Ghosh**, Shovon Sengupta, Sudip Kumar Naskar, Sunny Kumar Singh, “FinRAD: Financial Readability Assessment Dataset - 13,000+ Definitions of Financial Terms for Measuring Readability”, in Proceedings of the FNP workshop of the 13th Language Resources and Evaluation Conference (LREC-2022), pp 1-9 , Marseille, France <http://www.lrec-conf.org/proceedings/lrec2022/workshops/FNP/pdf/2022.fnp-1.1.pdf>
- **Sohom Ghosh**, Sudip Kumar Naskar, “LIPI at FinCausal 2022: Mining Causes and Effects from Financial Texts”, in Proceedings of the FNP workshop of the 13th Language Resources and Evaluation Conference (LREC2022), pp 130–132, Marseille, France
<http://www.lrec-conf.org/proceedings/lrec2022/workshops/FNP/pdf/2022.fnp-1.20.pdf>
- **Sohom Ghosh**, Sudip Kumar Naskar, “LIPI at the NTCIR-16 FinNum3 task: Ensembling transformer based models to detect in-claim numerals in financial conversations”, in Proceedings of the 16th NTCIR Conference on Evaluation of Information Access Technologies, pp 92-94, Tokyo, Japan.
<http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings16/pdf/ntcir/02-NTCIR16-FINNUM-GhoshS.pdf>

Published Research Papers in Conferences / Workshops

- **Sohom Ghosh**, Sudip Kumar Naskar, “FiNCAT: Financial Numeral Claim Analysis Tool”, in proceedings of FinWeb@TheWebConf (formerly ACMWWW) (2022), pp 583-585, Lyon, France <https://doi.org/10.1145/3487553.3524635>
- **Sohom Ghosh**, Shovon Sengupta, Sudip Kumar Naskar, Sunny Kumar Singh, “FinRead: A Transfer Learning Based Tool to Assess Readability of Definitions of Financial Terms”, in proceedings of 18th International Conference on Natural Language Processing (ICON-2021), pp 658-659 <https://aclanthology.org/2021.icon-main.81.pdf>
- Ankush Chopra, **Sohom Ghosh**, Term Expansion and FinBERT fine-tuning for Hypernym and Synonym Ranking of Financial Terms, In proceedings of FinNLP’21 (FinSim-3) (collocated with IJCAI-2021), pp 46-51, Montreal, Canada <https://aclanthology.org/2021.finnlp-1.8.pdf>

References

- Ismail El Maarouf, Youness Mansar, Virginie Mouilleron, and Dialetti Valsamou Stanislawski. The FinSim 2020 shared task: Learning semantic representations for the financial domain. In Proceedings of the Second Workshop on Financial Technology and Natural Language Processing, pages 81–86, Kyoto, Japan, 5 January 2020. URL <https://www.aclweb.org/anthology/2020.finlp-1.13>
- Simerjot Kaur, Charese Smiley, Akshat Gupta, Joy Sain, Dongsheng Wang, Suchetha Siddagangappa, Toyin Aguda, and Sameena Shah. Refind: Relation extraction financial dataset. In Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '23, page 3054–3063, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9781450394086. doi: 10.1145/3539618.3591911. URL <https://doi.org/10.1145/3539618.3591911>.
- Lefteris Loukas, Manos Fergadiotis, Ion Androutsopoulos, and Prodromos Malakasiotis. EDGAR-CORPUS: Billions of tokens make the world go round. In Proceedings of the Third Workshop on Economics and Natural Language Processing, pages 13–18, Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.econlp-1.2. URL <https://aclanthology.org/2021.econlp-1.2>
- 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
- Lefteris Loukas, Manos Fergadiotis, Ilias Chalkidis, Eirini Spyropoulou, Prodromos Malakasiotis, Ion Androutsopoulos, and Georgios Palouras. 2022. FiNER: Financial Numeric Entity Recognition for XBRL Tagging. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 4419–4431, Dublin, Ireland. Association for Computational Linguistics.
- And many more ...

Contributions & Achievements

16 tasks, 8 languages, 8 datasets, 6 tools, 24 papers

- 4 Journals
 - **SNCS** (Springer), **IJIT** (Springer), **Software Impacts** (Elsevier), **Science Talks** (Elsevier)
- 10 Conferences, 2 Workshops, 8 Shared Tasks
 - Venues: **ICON-2021** Silchar (India), **FinNLP@IJCAI-2021** Montreal (Canada), **FinWeb@The Web Conference-2022** Lyon (France), **FNP@LREC** Marseille (France), **FinNLP@IJCAI-ECAI-2022** Vienna (Austria), **NTCIR-16** Tokyo (Japan), **FinNLP@EMNLP-2022** Abu Dhabi (UAE), **FIRE-2022** Kolkata (India), **CODS-COMAD-2023** Mumbai (India), **ICDSA-2023** Jaipur (India), **FinNLP@AACL-IJCNLP-2023** Bali (Indonesia), **FIRE-2023** Goa (India), **NTCIR-17** Tokyo (Japan) , **The Web Conference WWW-2024 (Singapore)** , **LREC-COLING 2024** Tornio (Italy), **CIKM-2024** (Boise, USA)
- Resources (8 datasets, 6 tools)
 - **FinRAD** (dataset), **FinRead** (tool), **Financial Claim Analysis Tool (FinCAT, FinCAT-2)** (tool), Financial Language Understandability Enhancement Toolkit (**FLUEnT**) (tool), **Executive / General Tweets** (dataset), **FENCE - Financial Exaggerated Numeral ClassifiEr** (tool), **EID - ESG Issue Detector** (tool), **FinLanSer** (tool), **CRoD** (dataset), **FAAB** (dataset + tool), **IndicFinNLP** (dataset)
- Service
 - **Mentored 7 students** (undergraduates & post graduates), **Delivered 4 invited talks** (XIM, Bhubaneshwar; HIT Haldia; Brainware University, Yobe State University Nigeria), **Reviewed** several research papers & book proposals (FinNLP-ERA@EMNLP, Manning publishers book proposals, Social Network Analysis and Mining [Springer], FinArg@NTCIR-17)
- Achievements
 - **300 citations in Google Scholar**, **Kaggle Datasets Expert**, **CODS-COMAD 2023 YRS TRACK** (HONOURABLE MENTION), **CODS-COMAD 2024 Travel Grant**, **CIKM 2024 Travel Grant**, **ACM PIC-2025 Travel Grant**, **ACM ARCS-2025 Travel Grant**
 - **FinArg-1@NTCIR-17 TASK-1 SUB-TASK-2 (Rank-2)**, **FinNLP@IJCNLP-AACL 2023 ESG Impact type shared task (Rank-1 for Chinese & Japanese)**, **FinNLP@LREC-COLING 2024 ESG Impact Duration (Rank-3 in English impact length & impact type and Rank-1 in French Impact Length)**
- Collaborations
 - Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology, Japan (**Dr. Chung-Chi Chen**) 89
 - BITS, Pilani, Hyderabad, India (**Dr. Sunny Kumar Singh, Mr. Shovon Sengupta**)

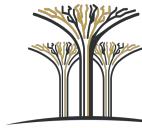
2 workshops

4 journals

8 shared tasks

10 conferences

Summary



THE WEB ACM LREC-COLING 2024



- Overview
 - Recent trends in financial Natural Language Processing (NLP) research {Science Talks Elsevier Journal 2023}
 - Using NLP to Enhance **Understandability** of Financial Text {**CODS-COMAD'2023**} [**HONOURABLE MENTION**], **CIKM-2024**
- Inclusive Investing
 - Increasing **readability** of Financial Texts {ICON'2021, FNP@LREC'2022}
 - Increasing **reachability** of Financial Content {**TheWebConf** (WWW'2024)}
- Improved Investing
 - Detecting **hyperonyms** of Financial Terms {FinNLP@IJCAI'2021, SNCS Springer Journal 2023}
 - Extracting **relationship** between financial entities {FIRE'2023}
- Impactful (Green) Investing
 - Detecting Environmental, Social, and Governance (**ESG**) and **sustainability** related **concepts, issues** {FinNLP@IJCAI-ECAI'2022, ICDSA 2023}
 - Assessing **ESG impact type, duration** in English, French, Japanese, Chinese, Korean {FinNLP@IJCNLP-AACL'2023, FinNLP@LREC-COLING'2024}
- Informed Investing
 - Detecting **in-claim numerals** {FinWeb@The Web Conference'2022, NTCIR'2022, IJIT Springer Journal, Software Impacts Elsevier Journal}
 - Detecting **exaggerated numerals** in Financial Texts
 - Estimating **profitability and loss** from financial social media posts in Chinese {FinNLP@EMNLP'2022}
 - Deciding **trust worthiness** of social media posts by executives {FIRE'2022}
 - Financial argument analysis in English & Chinese {NTCIR'2023}
- Indic Investing
 - Financial argument analysis in Bengali {FIRE'2023}
 - Detecting **ESG theme, Sustainability, Exaggerated numerals** in Hindi, Bengali & Telugu [**LREC COLING'2024**]
 - **IPO success, rating**
- Tools
 - Financial Language **Understandability Enhancement Toolkit** (FLUEnT) {CODS-COMAD'2023}



Outro



CIKM-2024

Task-1: Given a financial text (FT), we want to assess its readability and simplify it

Task-2: Given two FTs, we want to assess which one would reach more people

Task-3: Given a financial jargon in a FT, we would like to retrieve its hypernym

Task-4: Given two entities in a FT, we would like to determine the relationship between them

Task-5: Classify a FT as Sustainable or Unsustainable

Task-6: Detect ESG Issues from FTs in English

Task-7: Identify ESG impact type & duration from FTs.

Task-8: Detect exaggerated and in-claim numerals from FTs

Task-9: Evaluate the Rationals of Amateur Investors

Task-10: Evaluate the trustworthiness of Social Media Posts by Executives on Stock Prices

Task-11: Fine-grained Argument Understanding in FTs

Task-12: Financial Argument Analysis in Bengali

Task-13: Extract ESG Issues, Assess Sustainability, and Detect exaggerated numerals from FTs in Hindi, Bengali, & Telugu

Task-14: Predict success of Indian IPOs

Task-15: Predict ratings of Indian IPOs

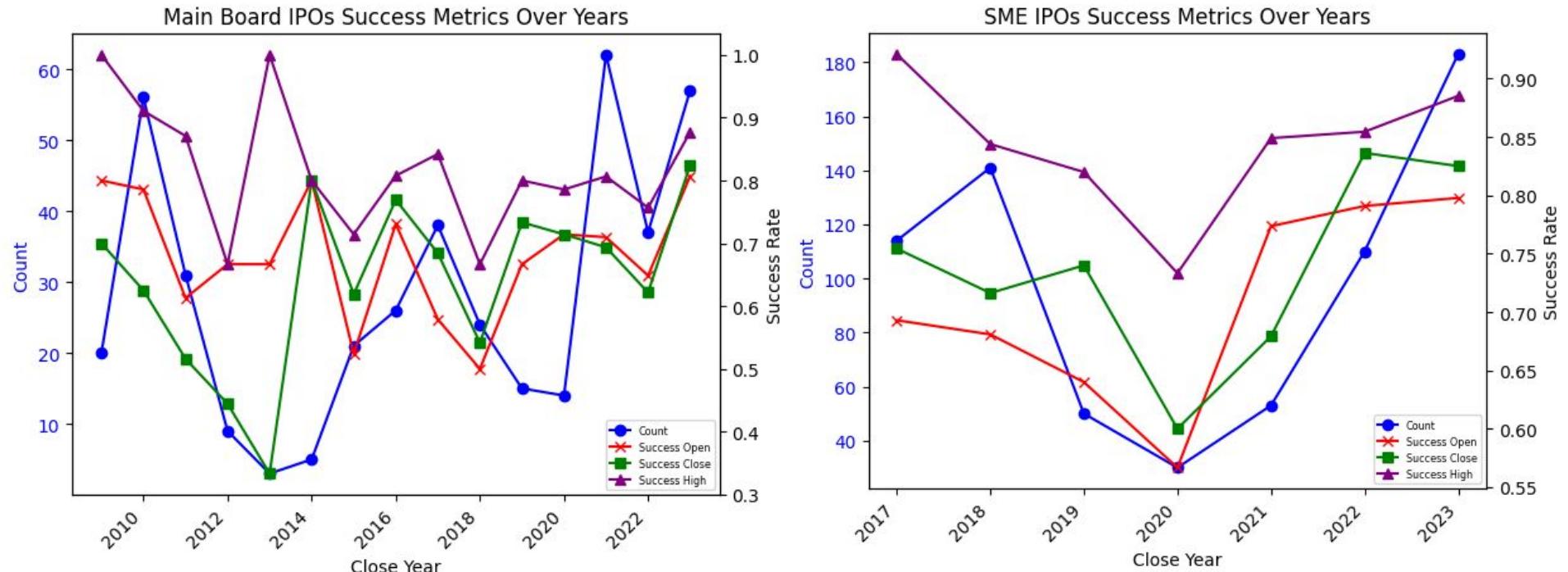
Task-16: Develop tools for processing FTs

16 tasks,
8 datasets,
6 tools,
24 papers,
8 languages

2 workshops,
4 journals,
8 shared tasks,
10 conferences

Task #	Metric	Approach Summary	SOTA	Performance	New Data	Language	New Tool
1	AU-ROC	FinBERT finetune	Yes	0.993	Yes	English	Yes
2	F1	RoBERTa + Claude (LLM)	Yes	0.731	Yes	English	No
3	Acc.	SBERT finetune	Yes	0.967	No	English	No
4	F1	SEC-BERT + Neural Network	No	0.736	No	English	No
5	Acc.	RoBERTa finetune	No	0.932	No	English	No
6	F1	SEC-BERT finetune	No	0.715	No	English	Yes
7	F1	FinBERT finetune	No	0.929 (IT)	No	English	No
7	F1	Trans-Prp + FinBERT finetune	No	0.756 (IT)	No	French	No
7	F1	Trans-Prp + FinBERT finetune	Yes	0.679 (IT)	No	Japanese	No
7	F1	Trans-Prp + FinBERT finetune	Yes	0.677 (IT)	No	Chinese	No
7	F1	Trans-Prp + PLM finetune	No	0.5882 (ID)	No	English	No
7	F1	Trans-Prp + PLM finetune	Yes	0.5616 (ID)	No	French	No
8	F1	Ensemble (FinBERT, BERT + Logistic Regression)	No	0.948	No	English	Yes
9	MPP, ML	SBERT Chinese + Classifier, FinBERT	No	0.575 (MPP), 0.598 (ML)	No	Chinese	No
10	MAPE	Gated Recurrent Unit	Yes	0.382	Yes	English	Yes
11	F1	Cross Encoder (FinBERT Finetuned)	No	0.789	No	English	No
11	F1	Translate + Cross Encoder (SEC-BERT)	No	0.641	No	Chinese	No
12	F1	MBERT, Cross Encoder (MBERT)	No	0.721 (1st task), 0.755 (2nd Task)	Yes	Bengali	Yes
13	F1	MBERT+Classifier, Translate + RoBERTa, Translate+MBERT	Yes	0.680 (1st task), 0.950 (2nd task), 0.590 (3rd task)	Yes	Hindi	No
13	F1	MBERT+Classifier, Translate + RoBERTa, Translate+MBERT	Yes	0.650 (1st task), 0.920 (2nd task), 0.550 (3rd task)	Yes	Bengali	No
13	F1	MBERT+Classifier, Translate + RoBERTa, Translate+MBERT	Yes	0.680 (1st task), 0.920 (2nd task), 0.580 (3rd task)	Yes	Telugu	No
14	F1	McL (Num, Cat, Txt) (Classification)	Yes	0.947 (Open-MB), 0.935 (High-MB), 0.931 (Close-MB)	Yes	English	No
14	MAE	McL (Num, Cat, Txt) (Regression)	Yes	0.167 (Open-MB), 0.193 (High-MB), 0.194 (Close-MB)	Yes	English	No
14	F1	McL (Num, Cat, Txt) (Classification)	Yes	0.893 (Open-SME), 0.942 (High-SME), 0.911 (Close-SME)	Yes	English	No
14	MAE	McL (Num, Cat, Txt) (Regression)	Yes	0.239 (Open-SME), 0.263 (High-SME), 0.256 (Close-SME)	Yes	English	No
15	F1	LongFormer RoBERTa Llama 3.1 70B	Yes	0.952 (MB) 0.423 (SME)	Yes	English	No
16	NA	Gradio (frontend)	NA	NA	NA	Various	Yes

Indic Investing: Predicting Success of Indian IPOs



Is it the same India?

2020



CELEBRATING 150 YEARS OF
India's growth story 93



A stock market opens under a banyan tree

Premchand Roychand, the OG Big Bull, is best remembered for founding the Native Share & Stock Brokers Association, which later became Bombay Stock Exchange. Before fancy trading floors and screens, it all started under a banyan tree and became the first-ever trading hub in India. That very tree still stands tall at Horniman Circle, Mumbai.



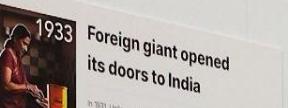
India's first stock market titan gets listed

Jamshedji Nusserwanji Tata: The visionary who built India's industrial future. He ventured into trade and later textiles, setting up Empress Mills in Nagpur in 1877, challenging British dominance in the sector. Today, 26 listed Tata Group companies have market capitalization of more than \$365 billion as of March 31, 2024.



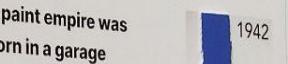
From ₹295 to a biscuit empire

Britannia, today a leading player in the ₹12,400-crore biscuits market, started out in 1892 as a tiny operation in a modest house in Calcutta (now Kolkata), with an initial investment of ₹295. However, it was in 1918 that the Britannia Biscuit Company Ltd. was formed. It was also the first company east of the Suwa to use imported gas ovens to bake biscuits.



Foreign giant opened its doors to India

In 1931, Unilever set up its first Indian subsidiary, Hindustan Lever. Manufacturing followed by Lever Brothers India Ltd (1933) and United Traders Ltd (1935). These three companies merged in July 1956 to form Hindustan Lever, Ltd. HLL, now known as HUL, retains 50% of its equity to the Indian public and is because the first foreign subsidiary to do so. Dabur, one of India's oldest brands, was introduced in 1937. It has since become a household name, generic to vanaspati or any other form of hydrogenated cooking oil.



A paint empire was born in a garage

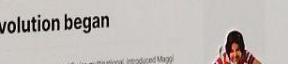
During World War II and the Quit India Movement, foreign paint brands ruled the Indian market. When thousands of Indians were writing a glorious chapter in India's freedom struggle through the Civil Disobedience Movement, four friends in Mumbai (now Mumbai) were setting up a paint manufacturing company.

Starting in a Mumbai garage in 1942, founders Champaikali Choksey, Chanchal Choksey, Surendra Das, and Arvind Vakil began Asian Paints amidst World War II's oil shortage challenges. Their early innovation! Selling paint in small packets instead of traditional tins, revolutionising distribution. Today, Asian Paints isn't just India's largest paint brand; it's a global force, present in 50 countries and serving 60 million Asian Paints, with a share of more than 50%, is the largest player in the domestic paints market.



The 2-minute revolution began

In 1982, Nestle, the renowned Swiss multinational, introduced Maggi noodles to the Indian market. Maggi's popularity skyrocketed, the brand expanded its range of offerings to cater to diverse tastes. As of September 2023, it dominates and owns a whopping 60% market share in the packaged noodles category, making India the largest market for Maggi worldwide. Nestle sold over 6 billion servings of Maggi during FY24.



A ₹10,000 ticket to India's dream car

When Maruti started manufacturing the Alto in April 1983, it immediately attracts a crowd of excited buyers. Demand for 20,000 customers buying up all 20,000 units of the Alto in the first month. The car was a massive success, and Maruti became a household name. Launched at a price of ₹1.42 lakh, the Maruti Alto became India's best-selling car. From 1983 to 2023... Maruti had already established itself as India's leading automobile. After peaking at ₹1.42 lakh per unit, the price had surged to ₹12.00 lakh in 2023.

A war on the borders, but the markets rally

While the nation was preoccupied by the Kargil war, the market rallied, marking the beginning of what many believe to be the first bull market in India. In May 1985, the Sensex crossed 1,000 for the first time. By the end of the year, the Sensex had risen to 1,100, a 10% gain from its previous record. It's interesting to note that during the Kargil war, the Sensex grew by 10% in just 10 days. The war did not dampen investors' spirits, which is a remarkable feat. The war was expected to last for 3-4 months, but the market continued to rally, showing resilience and strength.

Ambani starts an army of 12 Million shareholders

1986, 100 companies came out with public issues. Prime mover of the equity culture in the 1980s was Reliance Industries, when through its three issues between 1987 and 1989, raised an army of 12 million shareholders. Reliance's success was due to the masses and democratised stock market much before the Western companies entered Indian markets. The word reflected in the value of the BSE Sensex which rose from an index of 40 points in 1980 to 1,249 points in 1989, a 3,000% increase in 10 years.

A crisis changes India's economy

That year, India was on the brink of an economic collapse, with just enough foreign reserves to cover 2 weeks of imports, including oil. The fall in oil prices starting, things turned around. We're the chutz, India pledged 67 tons of gold for an emergency loan. This crisis acted as becoming the turning point that rejuvenated India's economy forever.

The ₹4,000 Crore scam and some bread pakoras?

The Hindash Mehta group rocked the stock market, inspiring beginings in its financial services. The Hindash Mehta group rocked the stock market, inspiring beginnings in its financial services. The Hindash Mehta group rocked the stock market, inspiring beginnings in its financial services.

Sensex crashes 60%

The collapse of雷曼兄弟公司 (Lehman Brothers) triggered a massive financial meltdown, and the market went into a tailspin. The Sensex, which had been growing steadily, faced its greatest challenge. The benchmark saw a 60% drop in value, with the market losing nearly half of its value. The Sensex dropped 30% in a single day, and the market was thrown into a tailspin. The government took immediate action, including bailouts for major banks and insurance companies. The central bank injected liquidity into the system to stabilize the market. The market eventually recovered, but the damage was done.

India bounces back

After the 2008 crash, India acted fast. The Reserve Bank of India (RBI) cut interest rates, provided liquidity, and the Government took steps to stabilize the market. The Sensex recovered rapidly, reaching its pre-crash levels by 2010. This was a remarkable turnaround for the Indian economy.

India's startup scene takes off

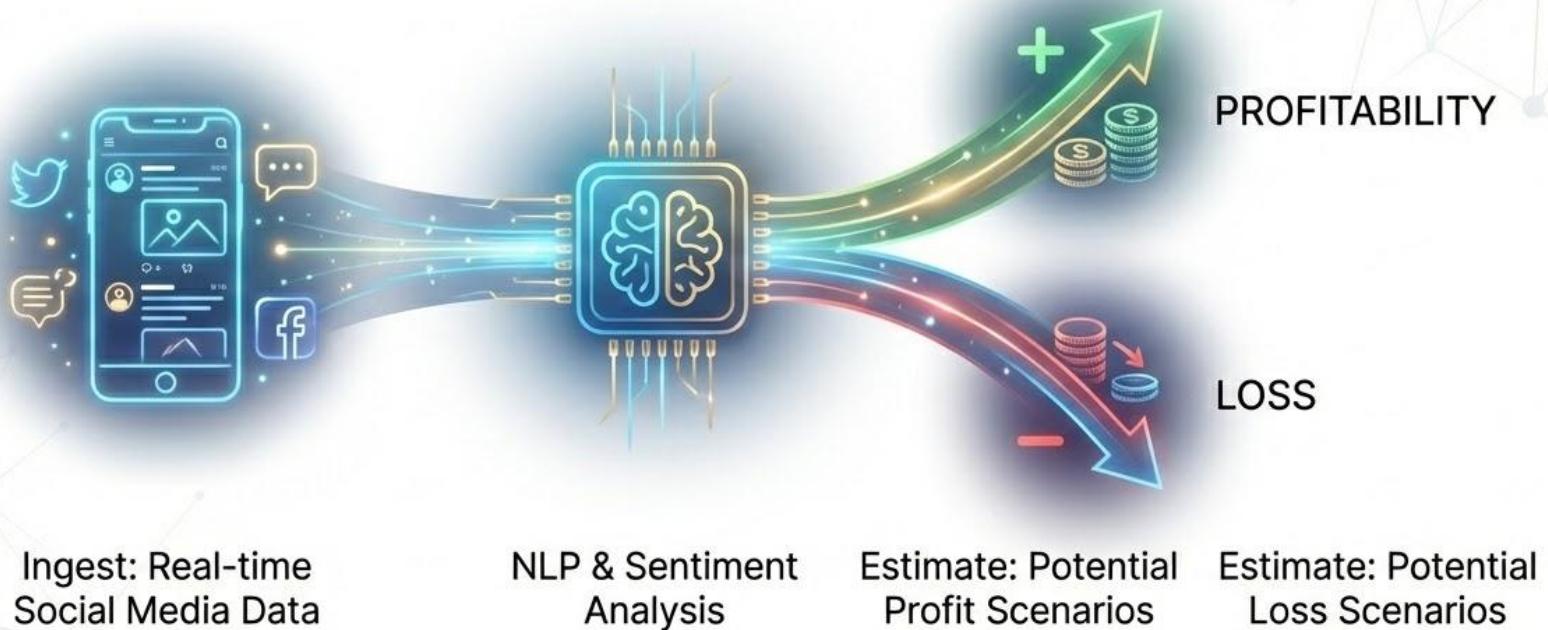
The early 2010s marked a significant period of growth for the Indian startup scene. In 2014, the Indian startup ecosystem reached \$1 billion in funding, marking the first time the Indian startup ecosystem had exceeded the \$1 billion mark. Since then, the Indian startup ecosystem has continued to grow, reaching \$10 billion in 2019. The Indian startup ecosystem has become one of the fastest-growing startup scenes in the world.

From lockdown to lift off

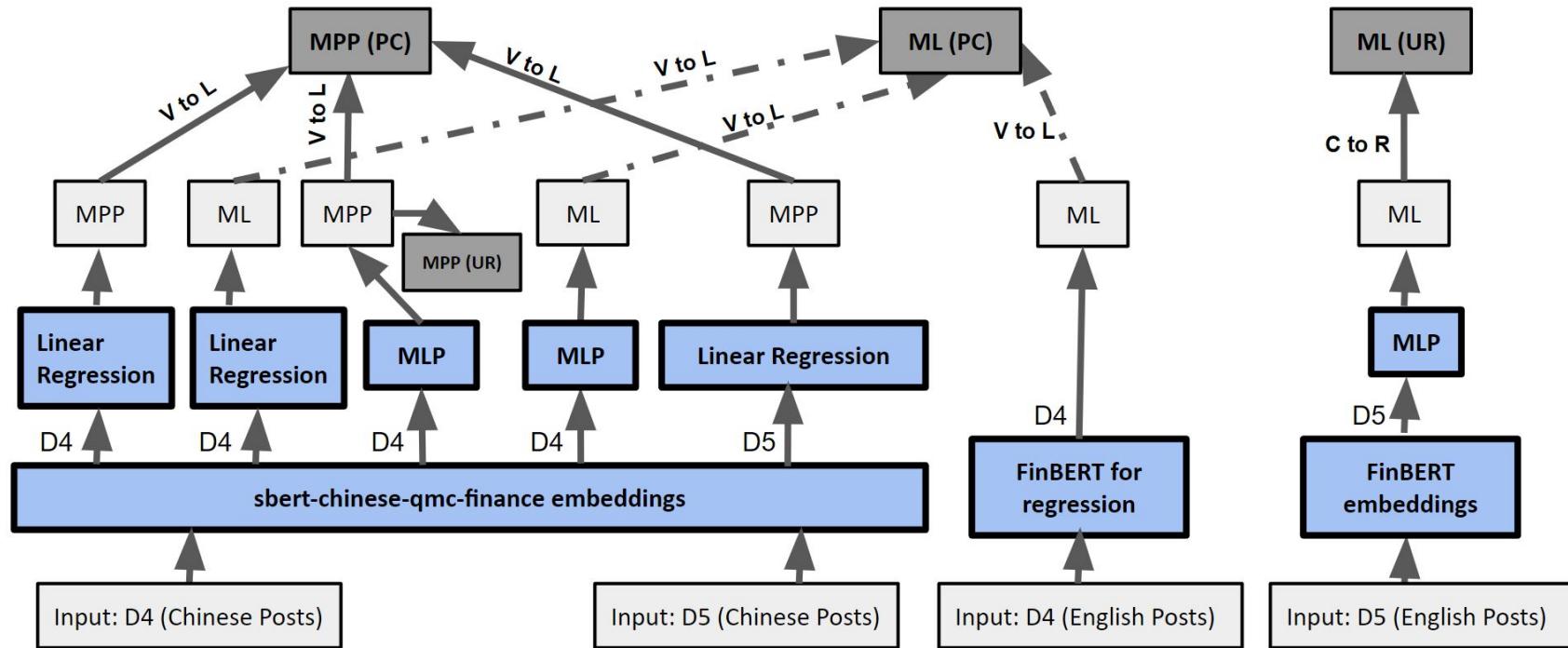
With the COVID-19 pandemic, India's startup ecosystem faced significant challenges. However, the startup ecosystem, with its focus on innovation and resilience, proved to be a key driver of economic recovery. The startup ecosystem, with its focus on innovation and resilience, proved to be a key driver of economic recovery.

Courtesy: Groww, Kempegowda International Airport (Bengaluru)

Informed Investing: Estimating profitability and loss from financial social media



Informed Investing: Estimating profitability and loss from financial social media



Improved Investing: Relation Extraction

Model	P	R	F1
SEC-BERT	0.206	0.454	0.284
EMB _{E1, E2} +NN	0.731	0.701	0.709
MOAT	0.748	0.743	0.736

TABLE 3.11: Performance of discriminative LLMs.

Model	P	R	F1
MOAT	0.748	0.743	0.736
-relation group, entity types	0.736	0.725	0.720
+SBERT _{SDP}	0.694	0.687	0.679
+POS tags	0.747	0.738	0.737
MOAT (per relation group)	0.839	0.672	0.715
MOAT (LUKE)	0.467	0.545	0.497

TABLE 3.12: Ablation Study.

Type	LLM	P	R	F1
Zero Shot	Falcon	0.538	0.434	0.362
	Dolly	0.400	0.316	0.253
	MPT	0.295	0.380	0.255
	LLaMA-2	0.192	0.260	0.202
Few Shot	Falcon	0.246	0.258	0.242
	Dolly	0.348	0.234	0.245
	MPT	0.296	0.156	0.128
	LLaMA-2	0.786	0.352	0.314
Classifier	MOAT	0.726	0.724	0.717

TABLE 3.13: MOAT versus generative LLMs.

Inclusive Investing: Reachability

Results

Setup	Model	Accuracy	F1
Classifier	FLANG-RoBERTa	50.0%	66.7%
	FLAN-UL2	50.2%	2.6%
Zero-Shot	Claude	51.0%	9.2%
	React T5	52.3%	30.0%
GGEA	GGEA (ChatGPT)	67.2%	68.1%
	GGEA (FLAN-UL2)	69.6%	70.1%
	GGEA (Claude)	71.9%	73.1%

Improved Investing: Hypernym Detection

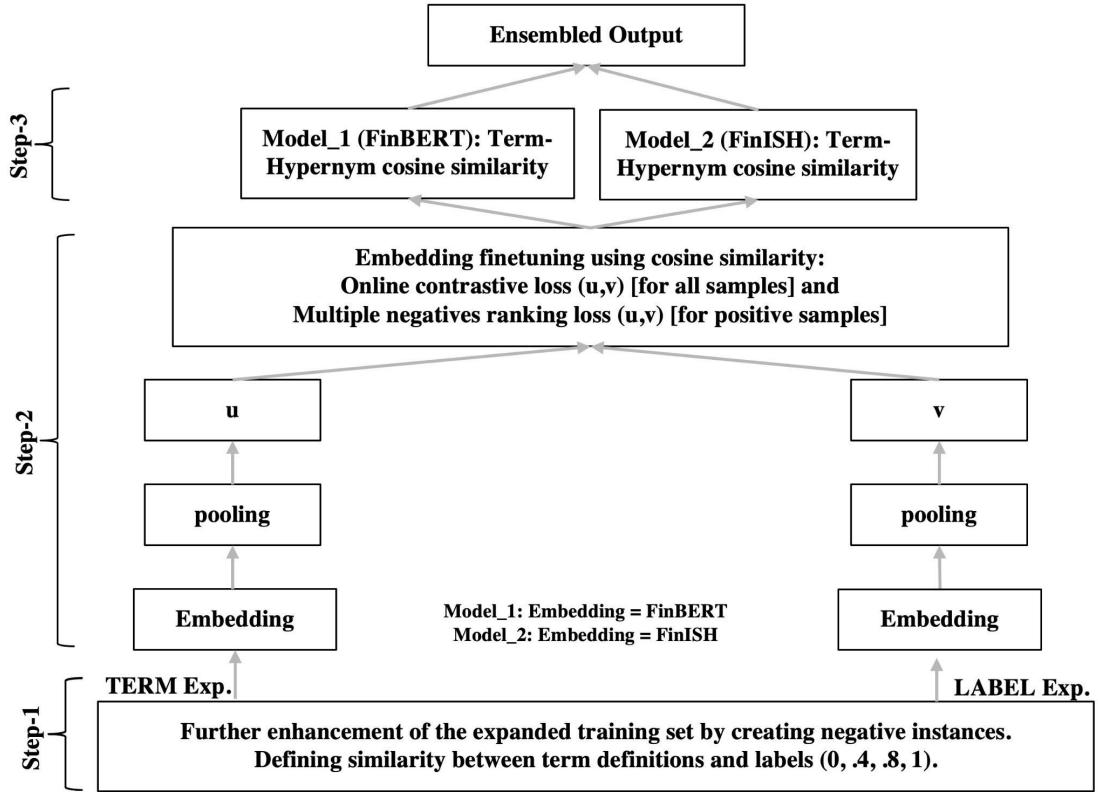
Algorithm 1 Algorithm to generate negative samples from existing training set

Require: $T > 0$ and $L > 0$ ▷ T is the augmented set of financial terms and L consists of corresponding labels i.e., hypernyms. $TT > 0$ and $LL > 0$ are the set of definitions of terms and labels respectively obtained after performing data augmentation

Require: Function FR(n) and Function FC(n) ▷ Function FR and FC returns the root node and first child node corresponding to node n respectively where n is one of the 17 labels i.e., leaf nodes/hypernyms

Ensure: $length(T) = length(TT) = length(L) = length(LL)$

- 1: $NT \leftarrow \{\}$ ▷ NT is the new set of definitions of financial terms to be created by appending negative samples
- 2: $NL \leftarrow \{\}$ ▷ NL is the new set of definitions of labels corresponding to terms in NT
- 3: $NS \leftarrow \{\}$ ▷ NS is the set of assigned similarity scores between the newly selected definitions of terms and labels in NT & NL respectively
- 4: $k \leftarrow 0.0$ ▷ 'k' is a hyper-parameter. Keeping k = 0.0 gives the best result
- 5: for each term $t \in T$, term definition $td \in TT$, corresponding label $l \in L$ and label definition $ld \in LL$ do
- 6: $NT \leftarrow NT \cup \{td\}$
- 7: $NL \leftarrow NL \cup \{ld\}$
- 8: $NS \leftarrow NS \cup \{1.0\}$ ▷ Assign a similarity score of 1.0 as the term and the label definition belong to the original set
- 9: $ln \leftarrow FR(l)$ ▷ Extract root node of label 'l'
- 10: $lc \leftarrow FC(l)$ ▷ Extract first child node of label 'l'
- 11: $R, RR \in L, LL$ where $length(R)=10$, $length(RR)=10$ ▷ Randomly select 10 labels from 'L' and corresponding label definitions from 'LL' ensuring none of the labels are 'l' and none of their corresponding terms is 't'. This is done for creating the negative set
- 12: for each label $la \in R$ and corresponding definition $lnd \in RR$ do
- 13: $NT \leftarrow NT \cup \{td\}$
- 14: $NL \leftarrow NL \cup \{ld\}$
- 15: $lan \leftarrow FR(la)$ ▷ Extract root node of label 'la'
- 16: $lac \leftarrow FC(la)$ ▷ Extract first child node of label 'la'
- 17: if $lac = lc$ then ▷ Check if first child nodes are the same. This implies root nodes are also the same.
- 18: $NS \leftarrow NS \cup \{2 * k\}$
- 19: else if $lan = ln$ then ▷ Check if root child nodes are same when first child nodes are different
- 20: $NS \leftarrow NS \cup \{1 * k\}$
- 21: else ▷ When first child nodes and root nodes are different
- 22: $NS \leftarrow NS \cup \{0 * k\}$
- 23: end if
- 24: end for
- 25: end for
- 26: return NT, NL, NS



Improved Investing: Hypernym Detection

SLN.	Model	Data Aug.	Validation Set		Test Set	
			MR	Acc.	MR	Acc.
1	Base-1	No	2.158	0.498	1.941	0.564
2	Base-2	No	1.201	0.876	1.750	0.669
3	BERT	No	1.177	0.899	-	-
4	BERT	Yes	1.153	0.928	-	-
5	FinBERT	No	1.117	0.928	1.257	0.886
6	FinBERT	Yes	1.110	0.942	1.220	0.895
7	SFinBERT_neg (Our old model) [5]	Yes	1.086	0.947	1.156	0.917
8	dicoe_1 [107]	No	-	-	1.180	0.889
9	dicoe_2 [107]	Yes	-	-	1.162	0.904
10	MiniTrue_2 [95]	No	-	-	1.315	0.865
11	MiniTrue_1 [95]	No	-	-	1.346	0.855
12	MiniTrue_3 [95]	No	-	-	1.337	0.825
13	mxx [90]	Yes	1.06	0.96	1.113	0.941
14	yseop_1 [108]	Yes	-	-	1.236	0.883
15	yseop_2 [108]	Yes	-	-	1.141	0.917
16	SFinBERT_neg_th	Yes	1.110	0.938	-	-
17	SFinBERT_neg_ss	Yes	1.105	0.933	-	-
18	SFinHyp_neg	Yes	1.072	0.952	-	-
19	SFinHyp_more_data	Yes	1.306	0.813	-	-
20	Ensemble_7.18 (Our new Model)	Yes	1.053	0.967	-	-

TABLE 3.7: Results on validation and test set. Org. represents original and Ext. represents extended. Base refers to baseline. MR is Mean Rank.

Impactful (Green) Investing: ESG concept & Sustainability detection

Sl. No.	Base Model	Data Augmentation	Mean Rank	Accuracy
1	all-mpnet-base-v2	No (only positives)	1.4692	0.6923
2	all-mpnet-base-v2	Yes (1 negative per positive)	1.5769	0.7000
3	sbert_un	No (only positives)	1.5308	0.6769
4	sbert_un	Yes (1 negative per positive)	1.4769	0.7308
5	sbert_un	Yes (1 negative per positive) + concepts	1.4615	0.7154
6	sbert_un	Yes (1 negative per positive) - concept definitions + concepts	1.4846	0.7462
7	sbert_un	Yes (1 negative per positive) [out of group sampling]	1.4385	0.7462
8	sbert_un	Yes (5 negative per positive) [out of group sampling]	1.4308	0.7615
9	sbert_un	Yes (15 negative per positive) [out of group sampling]	1.5308	0.7000
10	sbert_un	Yes (5 negative per positive) [out of group sampling] {batch size = 40, epoch = 30}	1.4154	0.7462
11	sbert_un	Yes (5 negative per positive) [out of group sampling] {batch size = 40, epoch = 20}	1.4615	0.7462
12	roberta classifier	-	1.4846	0.7538
13	sbert_un	Yes (1 negative per positive) [same group & out of group sampling]	1.4615	0.7462
14	sbert_un	Yes (5 negative per positive) [same group & out of group sampling]	1.5000	0.7385
15	baseline-1	-	2.5308	0.3769
16	baseline-2	-	1.6846	0.7154

TABLE 4.3: Results of Sub-Task 1 on the validation set.

NOTE: Where not mentioned, definitions of concepts were used with batch size of 20
for 15 epochs.

Sl. No.	Model	Accuracy
1	roberta-base	0.9338
2	finbert	0.9426
3	sbert_un	0.8653
4	sub-task1 finetune	0.8543

TABLE 4.4: Results of Sub-Task 2 on the validation set.

ST	Sub.	Accuracy	Mean Rank
1	1	0.7103	1.5172
1	2	0.7034	1.6689
2	1	0.9219	-
2	2	0.9317	-

TABLE 4.5: Test set results for sub-tasks (ST) 1 and 2. Sub.: Submission

Impactful (Green) Investing: ESG Issue Detection

Sl. No.	Model	Dataset	Accuracy	F1	Precision	Recall
1	distilbert	Eng (T)	0.637	0.624	0.662	0.637
2	sec-bert	Eng (T)	0.650	0.653	0.639	0.65
3	finbert-esg	Eng (T)	0.667	0.655	0.684	0.667
4	sec-bert	Eng (T+C)	0.717	0.708	0.732	0.717
5	finbert-esg	Eng (T+C)	0.688	0.678	0.697	0.688
6	sec-bert pre-ft	Eng (T+C)	0.725	0.715	0.726	0.725
7	finbert-esg pre-ft	Eng (T+C)	0.692	0.682	0.700	0.692
8	sec-bert pre-ft	Eng + Fr2Eng (T)	0.646	0.640	0.684	0.646
9	finbert-esg pre-ft	Eng + Fr2Eng (T)	0.629	0.621	0.645	0.629
10	sec-bert pre-ft	Eng + Fr2Eng (T+C)	0.692	0.687	0.703	0.692
11	finbert-esg pre-ft	Eng + Fr2Eng (T+C)	0.687	0.683	0.702	0.687
12	Ensemble 6 & 7	Eng (T + C)	0.696	0.688	0.722	0.696
13	FLAN-UL2	Eng (T)	0.367	0.339	0.627	0.367
14	FLAN-UL2	Eng (T + C)	0.417	0.417	0.596	0.696

TABLE 4.7: Results (Eng = English, Fr2Eng = French translated to English, C = content, T = Title, ft = fine tuned)

Informed Investing - Claim Detection

Sl. No.	Model	F1-Macro
1	TF-IDF + LR	0.6345
2	TF-IDF + RF	0.6603
3	TF-IDF + XG-Boost	0.6646
4	FB + LR	0.7990
5	FB + RF	0.7763
6	FB + XG-Boost	0.7994
7	FB classifier (depenent text)	0.7250
8	FB classifier (CW=6) + EF	0.8244
9	FB classifier (CW=6) + category	0.8315
10	CapsNet (baseline) [172]	0.5736
11	S1	0.8585
12	S2	0.8439
13	S3	0.8318
14	Ensemble S1, S2, S3	0.8671

TABLE 5.2: Overall Results. LR = Logistic Regression, RF = Random Forest, FB = FinBERT, CW = Context Window Size and EF = Engineered Features

Model	Macro-F1	Micro-F1
S1 (CW=8, only)	0.8585	0.6345
S2 (CW=6, only)	0.8439	0.6603
S3 (only)	0.8318	0.6646
S3 (-EF)	0.8238	0.7990
S3 (-EF, only largest token)	0.7934	0.7763
FinBERT classifier (CW=4)	0.8408	0.7994
FinBERT classifier (CW=5)	0.8318	0.7250
FinBERT classifier (CW=7)	0.8381	0.8244
FinBERT classifier (CW=9)	0.8247	0.8262
FinBERT classifier (CW=10)	0.8407	0.8585
Ensemble S1, S2, S3	0.8671	0.9479

TABLE 5.4: Ablation Study. CW = Context Window, EF = Engineered Features

Informed Investing: Estimating profitability and loss from financial social media

Sl.#	Model	Data	Language	MPP (Pairwise Comparison)			MPP (Unsupervised Ranking)		
				Train	Valid.	Test (D3)	Train	Valid.	Test (D2)
1.1	SB-1	D4	Chinese	100.00%	70.00%	54.02%	8.04%	2.98%	11.83%
1.2	SB-2	D4	Chinese	62.18%	67.50%	48.28%	3.89%	2.45%	18.27%
1.3	SB-3	D5	Chinese	99.63%	60.00%	41.38%	-	-	17.46%
1.4	SB-4	D4	English	51.92%	47.50%	50.57%	2.11%	3.94%	4.17%
1.5	SB-5	D5	English	99.59%	45.00%	55.17%	-	-	16.63%
1.6	Ensemble (§5.6.4.1)	-	-	-	72.50%	57.47%	-	-	-

TABLE 5.8: MPP Results

SL #	Model	Data	Language	ML (Pairwise Comparison)			ML (Unsupervised Ranking)		
				Train	Valid.	Test (D3)	Train	Valid.	Test (D2)
2.1	SB-1	D4	Chinese	97.44%	52.50%	50.57%	-10.26%	-2.16%	-7.81%
2.2	SB-2	D4	Chinese	57.69%	55.00%	50.57%	-5.55%	-8.01%	-5.56%
2.3	SB-3	D5	Chinese	99.65%	52.50%	47.12%	-	-	-3.90%
2.4	SB-4	D4	English	58.00%	50.00%	59.77%	-1.87%	-1.35%	-6.29%
2.5	SB-5	D5	English	91.24%	55.00%	44.83%	-	-	-4.11%
2.6	Ensemble (§5.6.4.2)	-	-	82.05%	57.50%	50.57%	-	-	-

TABLE 5.9: ML Results

Informed Investing: Deciding trustworthiness of social media posts by executives

Evaluating Impact of Social Media Posts by Executives on Stock Prices

Anubhav Sarkar*
 Swagata Chakraborty*

Sohom Ghosh
 Sudip Kumar Naskar

Exp	Stock	Data	MAE	RMSE	R ² _a	MAPE (%)
Exp 3.1	TSLA (T)	Y+G	56.365	79.701	0.790	5.852
		Y+E	34.362	48.261	0.923	3.817
Exp 3.2	AAPL (T)	Y+G	4.700	5.922	0.859	2.932
		Y+E	3.075	3.860	0.940	1.990
Exp 3.3	BTC (T)	Y+G	4681.737	5584.437	0.572	9.675
		Y+E	2842.190	3679.708	0.814	5.830
Exp 3.4	ETH (T)	Y+G	315.075	407.849	0.641	8.663
		Y+E	278.507	356.633	0.725	7.724
Exp 3.5	TSLA (R)	Y+G _r	44.625	61.403	0.811	4.481
		Y+E _r	42.283	58.994	0.826	4.247

TABLE 5.15: Result of Experiments 3

Stock	Data	MAE	RMSE	R ² _a	MAPE (%)
TSLA (T)	Y+G	59.621	83.765	0.768	6.174
	Y+E	34.362	48.261	0.923	3.817
AAPL (T)	Y+G	3.899	5.004	0.899	2.459
	Y+E	3.071	3.849	0.940	1.988
BTC (T)	Y+G	3676.006	4631.177	0.706	7.492
	Y+E	2842.190	3679.708	0.814	5.830
ETH (T)	Y+G	309.507	392.286	0.668	8.557
	Y+E	278.507	356.633	0.725	7.724
TSLA (R)	Y+G _r	46.953	68.659	0.844	4.954
	Y+E _r	57.674	82.634	0.774	6.043

TABLE 5.16: Result of Experiment 4

Informed Investing: Financial Argument Analysis

MODEL	DATA	VALIDATION	
		MICRO F1	MACRO F1
DistilBERT	Original	0.7913	0.5321
DistilBERT	Paraphrased	0.7942	0.4811
Flang-Roberta	Original	0.7971	0.5653
Flang-Roberta	Paraphrased	0.7971	0.5456
BERT-SEC	Original	0.813	0.5647
BERT-SEC	Paraphrased	0.7880	0.4900
Cross Encoder (BERT)	Original	0.7898	0.5383
Cross Encoder (BERT)	Paraphrased	0.7913	0.4956
Cross-Encoder (BERT-SEC)	Original	0.7695	0.476
Cross-Encoder (BERT-SEC)	Paraphrased	0.7681	0.4807
Cross Encoder (FinBERT Finetuned)	Original	0.8275	0.5298
Cross-Encoder (MLM-FinBERT)	Original	0.8000	0.5054
Cross-Encoder (MLM-FinBERT)	Paraphrased	0.7913	0.5482

Table 3: Results of Task 2, Sub-Task 2: Argument Relation Identification

LANGUAGE	MODEL	DATA	VALIDATION	
			MICRO F1	MACRO F1
English	BERT-base	Original	0.6453	0.6783
English	BERT-base	Paraphrased	0.6319	0.6568
English	FLANG-RoBERTa	Paraphrased	0.6392	0.6754
English	Cross Encoder (SBERT)	Original	0.6404	0.6796
English	Cross Encoder (SBERT)	Paraphrased	0.6500	0.6880
English	Cross Encoder (DistilROBERTA)	Original	0.7055	0.7472
English	Cross Encoder (DistilROBERTA)	Paraphrased	0.6920	0.7374
English	Cross Encoder (Flang-Roberta)	Original	0.6932	0.7342
English	Cross Encoder (Flang-Roberta)	Paraphrased	0.6858	0.7314
English	Cross Encoder (BERT-SEC)	Original	0.6932	0.7342
English	Cross Encoder (BERT-SEC)	Paraphrased	0.6800	0.7000
English	Cross Encoder (MLM on BERT-SEC)	Original	0.6846	0.7160
English	Cross Encoder (MLM on BERT-SEC)	Paraphrased	0.6871	0.7180
Chinese	SBERT-Chinese	Original	0.6321	0.6450
Chinese	Cross Encoder (SBERT-Chinese)	Original	0.6503	0.6432

Table 4: Result of Task 3: Identifying Argumentative Relation in Social Media Discussion