

Overview of the NTCIR-18 FinArg-2 Task: Temporal Inference of Financial Arguments

Chung-Chi Chen, Chin-Yi Lin, Cheng-Chih Chiu, Hen-Hsen Huang, Alaa Alhamzeh, Yu-Lieh Huang, Hiroya Takamura, and Hsin-Hsi Chen

AIST (Japan), NTU (Taiwan), Academia Sinica (Taiwan) NTHU (Taiwan), Universität Passau (Germany)

Impact Duration Matters in Financial Analysis

	Model	Feature	Accuracy	Precision	Recall	F1 Score
	PaLM 2	-	0.608	0.609	0.608	0.591
Zero-Shot	Gemini Pro	_	0.614	0.621	0.614	0.615
	ChatGPT	_	0.635	0.644	0.635	0.635
	GPT-4	_	0.640	0.662	0.640	0.638
	Longformer	-	0.705	0.723	0.705	0.699
	Llama-2-7B	_	0.696	0.708	0.696	0.696
	Mistral-7B	_	0.725	0.724	0.725	0.725
		-	0.731	0.732	0.731	0.728
C1		Conventional Argument Units	0.737	0.747	0.737	0.737
Supervised		Proposed Argument Units	0.749	0.761	0.749	0.741
	GNN	VADER Sentiment	0.738	0.739	0.738	0.738
		Argument-Based Sentiment	0.775	0.776	0.775	0.773
		Impact Duration	0.757	0.756	0.757	0.757
		All Labels in Equity-AMSA	0.798	0.801	0.798	0.796

Incorporating Impact Duration can enhance the assessment of forecasting skill

Argument-Based Sentiment Analysis on Forward-Looking Statements



Impact Duration Matters in Social Media Opinion

Type	Model	Accuracy	F1
	StockNet	46.72%	44.44%
SMP Model	HAN	57.35%	56.61%
	SRLP	61.76%	61.69%
	BERT-Chinese	58.36%	58.33%
	Multilingual-BERT	58.36%	58.16%
PLM	Chinese-BERT	60.32%	60.33%
	Mengzi-BERT	59.72%	59.70%
	Mengzi-BERT-Fin	57.61%	57.44%
	IDED-BERT	63.17%	62.92%
	IDED-mBERT	62.90%	62.84%
Proposed	IDED-CBERT	61.68%	61.36%
•	IDED-Mengzi	64.05%	64.09%
	IDED-Mengzi-Fin	64.18%	64.15 %

Pre-Finetuning with Impact Duration Awareness for Stock Movement Prediction



Participants





























Impact Duration Matters in Assessing Company's Operations

Expected Time frame for Risk/Opportunity to Materialize Short-Term Long-Term (5+ years) (<2 years) Industry is a major **Highest Weight** Level of contributor to impact Contribution to Environmental or Industry is a minor Social Impact **Lowest Weight** contributor to impact

The short-term events have a relatively greater impact

- **MSCI ESG Rating Methodology**
- DynamicESG: A Dataset for Dynamically Unearthing ESG Ratings from News Articles



FinArg-2 Dataset

Ten	nporal Tag	Earnings Call	Research Report	Social Media
	< a Quarter	150	_	_
_	1-2 Quarter	93	_	-
Past	3-4 Quarter	114	_	-
	> a Year	60	-	-
	< a Week	-	_	1,947
	> a Week	_	_	6,108
	< a Month	-	769	-
Future	1–3 Month	_	6,886	-
	4–6 Month	_	4,445	-
	7–12 Month	_	12,125	-
	> a Year	_	7,431	-
No Tem	poral Reference	1,767	_	705

Methods & Performances

Tr					Method			
Team		ECC				Socia	al Media	
FTRI [13]	BERT + TF-IDF + T	ime Sta	mp One-hot vector	-				
TMUNLPG1 [24]	FinBERT + Data Au		-	FinBERT + Ke	ywords (Log-Likelih	ood Ratio & I	Pointwise Mutual	Information)
		0		Fine-tuned GF			omitwise mataur	miormation,
IMNTPU [3]	RoBERTa, DistilBEI		-	rine-tunea Gr	71-40 mm			
Trustworthy	Logistic Regression	(TF-ID	F), BERT	-				
SCUNLP-2 [16]	Repeat-Error-Corre	ction		-				
SCUNLP-3 [14]	-			BERT + Data	Augmentation			
SCUNLP-1 [11]	BERT + Fine-tuned	GPT-4c	mini	-				
SCaLAR IT [21]	BERT + TF-IDF (Ter	mporal	Entity)	-				
AIDAVANCE [15]	mDeBERTa, GPT-40)		-				
	Team Name	Micro F1	Macro F1			14: T4	14 D4	
	FTRI_3	77.38%	75.07%		Team	Micro F1	Macro F1	
	FTRI_1	71.43%	68.58%		IMNTPU 1	76.83%	54.68%	
	SCaLAR IT Team_1 FTRI_2	70.24% 69.05%	67.85% 65.76%		_			
	IMNTPU_1	69.05%	67.06%		TMUNLPG1_3	74.20%	54.31%	
	TMUNLPG1_2	69.05%	66.13%		TMUNLPG1_2	73.92%	56.04%	
	AIDAVANCE_1 AIDAVANCE_3	69.05% 69.05%	67.11% 66.10%		IMNTPU 2	72.83%	53.40%	
	SCUNLP-1_3	67.86%	64.94%		_			
	SCUNLP-1_1	66.67%	63.06%		TMUNLPG1_1	72.77%	52.14%	
	AIDAVANCE_2 SCUNLP-2_2	66.67% 66.67%	61.05% 63.41%		IMNTPU 3	69.98%	53.50%	
	SCUNLP-2_3	66.67%	63.37%		-			
	IMNTPU_3	65.48%	62.44%		SCUNLP-3_2	66.04%	34.13%	
	TMUNLPG1_3 Trustworthy_1	65.48% 63.10%	64.45% 58.67%		SCUNLP-3 1	64.44%	33.39%	
	IMNTPU_2	63.10%	57.87%		SCUNLP-3_3			
	SCUNLP-2_1	63.10%	59.54%		SCUNLY-5_5	61.99%	33.97%	
	TMUNLPG1_1 Trustworthy 2	61.90% 60.71%	56.32% 52.75%					

Numerals Matter

	3-day	7-day	15-day	30-day
MDRM (Text Only)	1.431	0.439	0.309	0.219
MDRM (Text + Audio)	1.371	0.420	0.300	0.217
HTML (Text Only)	1.175	0.372	0.153	0.133
HTML (Text + Audio)	0.845	0.349	0.251	0.158
FinBERT	0.750	0.368	0.241	0.172
NAM (Text Only)	0.745	0.300	0.232	0.187
GNA-Vol + Discrete Strategy	0.700	0.322	0.252	0.207
GNA-Vol + Continuous Strategy	0.705	0.362	0.250	0.237

Crowd View: Converting Investors' Opinions into Indicators Enhancing Volatility Forecasting in Financial Markets: A

General Numeral Attachment Dataset for Understanding Earnings Calls. AACL-2023



Professionalism Matters

Dataset	Mobile01	PTT
Metrics	Accuracy	Top 10% MPP
GPT-3.5	43.68%	13.94%
Hand-Craft [26]	_	24.39%
Expert-Like [5]	_	17.61%
BERT-Chinese [11]	62.07%	9.92%
+ SL Pre-Finetune	52.87%	8.17%
+ WL Pre-Finetune	47.13%	14.12%
Mengzi-Fin [32]	55.02%	19.26%
+ SL Pre-Finetune	48.24%	11.84%
+ WL Pre-Finetune	52.07%	11.40%
SCQF [10]	52.94%	18.27%
+ SL Pre-Finetune	66.41%	23.39%
+ WL Pre-Finetune	61.75%	11.09%

Professionalism-Aware Pre-Finetuning for **Profitability Ranking. CIKM-2024**

Evaluating the Rationales of Amateur Investors



FinArg-3: Argument Quality Assessment of Financial Forward-Looking Statements

Iteration		Ta	sk	Year	Chung-Chi Chen - Hiroya Takamura
FinNum-1	Fine-g	rained Nume	eral Understanding	2018-2019	
FinNum-2	Nume	ral Attachme	ent	2019-2020	ninkullitu
FinNum-3	Fine-g	rained Claim	Detection	2021-2022	Agent Al for Finance
FinArg-1	Argun	nent-Based S	entiment Analysis	2022-2023	From Financial Argum Mining to Agent-Base
FinArg-2	Argun	nent-Based T	emporal Inference	2024-2025	Modeling
FinArg-3	Argun	nent Quality	Assessment	2025-2026	
Short Name	Language	Source Fornings Coll	Argument Unit/Deletion	Task	
	English	Earnings Call	Argument Unit/Relation	Identification	
Short Name FinArg-1			Argument-based Sentime	Identification ent Analysis	elations in Social Media Discuss
FinArg-1	English English	Earnings Call Analyst Report	Argument-based Sentime	Identification ent Analysis apport Argumentative R	elations in Social Media Discuss
	English English Chinese	Earnings Call Analyst Report Social Media	Argument-based Sentime Identifying Attack and Su	Identification ent Analysis apport Argumentative R erence Detection	elations in Social Media Discuss
FinArg-1	English English Chinese English	Earnings Call Analyst Report Social Media Earnings Calls	Argument-based Sentime Identifying Attack and Su Argument Temporal Refe	Identification ent Analysis apport Argumentative R erence Detection d Assessment	elations in Social Media Discuss
FinArg-1 FinArg-2	English English Chinese English English	Earnings Call Analyst Report Social Media Earnings Calls Analyst Report	Argument-based Sentime Identifying Attack and Su Argument Temporal Refe Premise's Influence Perio	Identification ent Analysis apport Argumentative Reference Detection d Assessment	elations in Social Media Discuss
FinArg-1	English English Chinese English English Chinese	Earnings Call Analyst Report Social Media Earnings Calls Analyst Report Social Media	Argument-based Sentime Identifying Attack and Su Argument Temporal Refe Premise's Influence Perio Claim's Validity Period A	Identification ent Analysis apport Argumentative Reference Detection d Assessment assessment	elations in Social Media Discuss

 Agent Al for Finance: From Financial Argument Mining to Agent-Based Modeling. 2025

 From Opinion Mining to Financial Argument **Mining. 2021**













