Commodity Market Forecasting using NLP

Dana Golden, Dikshya Mohanty, Khushboo Singh

CSE 538 - Natural Language Processing - **May 26, 2025**

Presentation Outline

Commodity Prices over Time

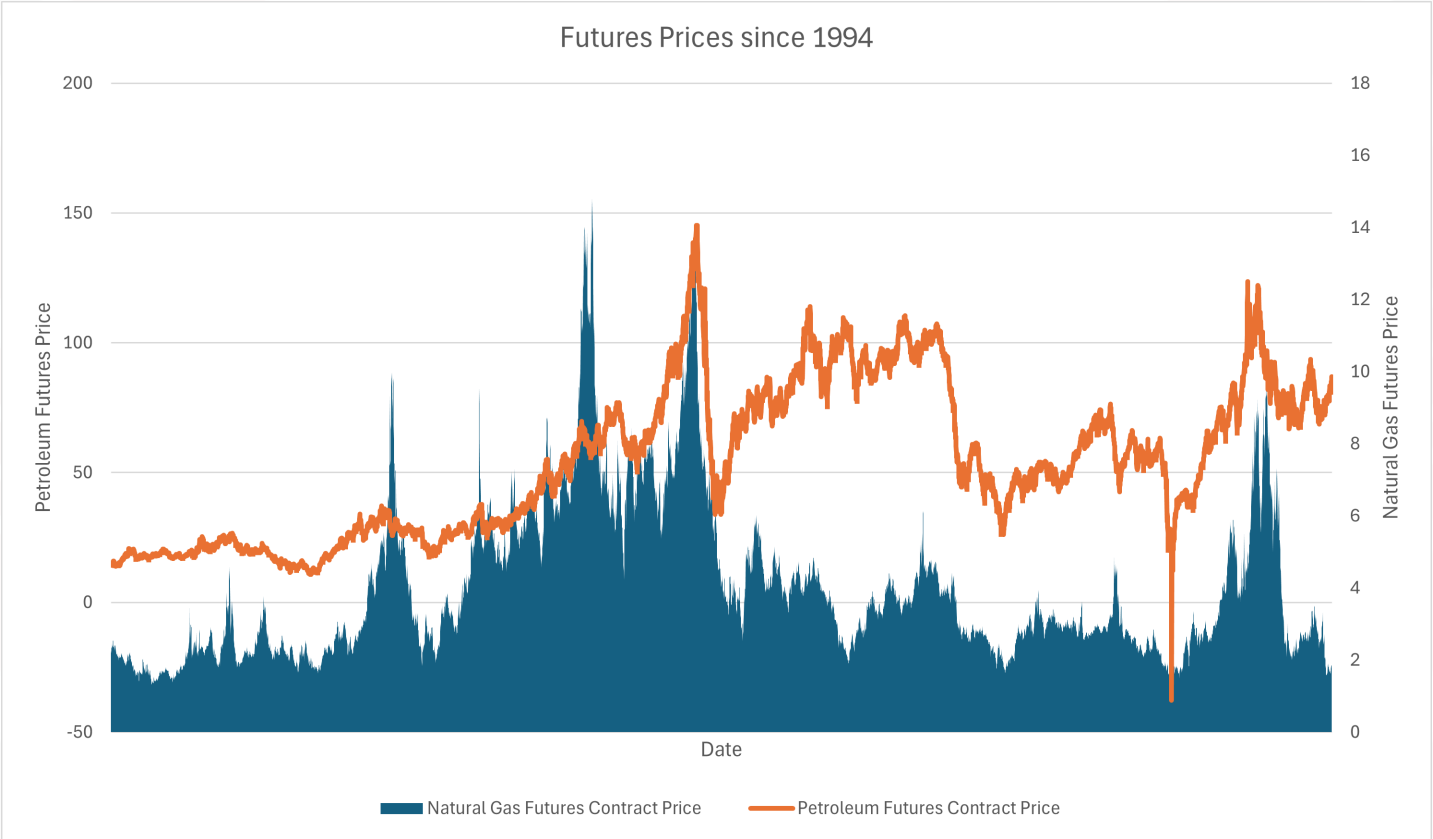


Figure 1: Future Prices over Time

Commodities Markets Reports

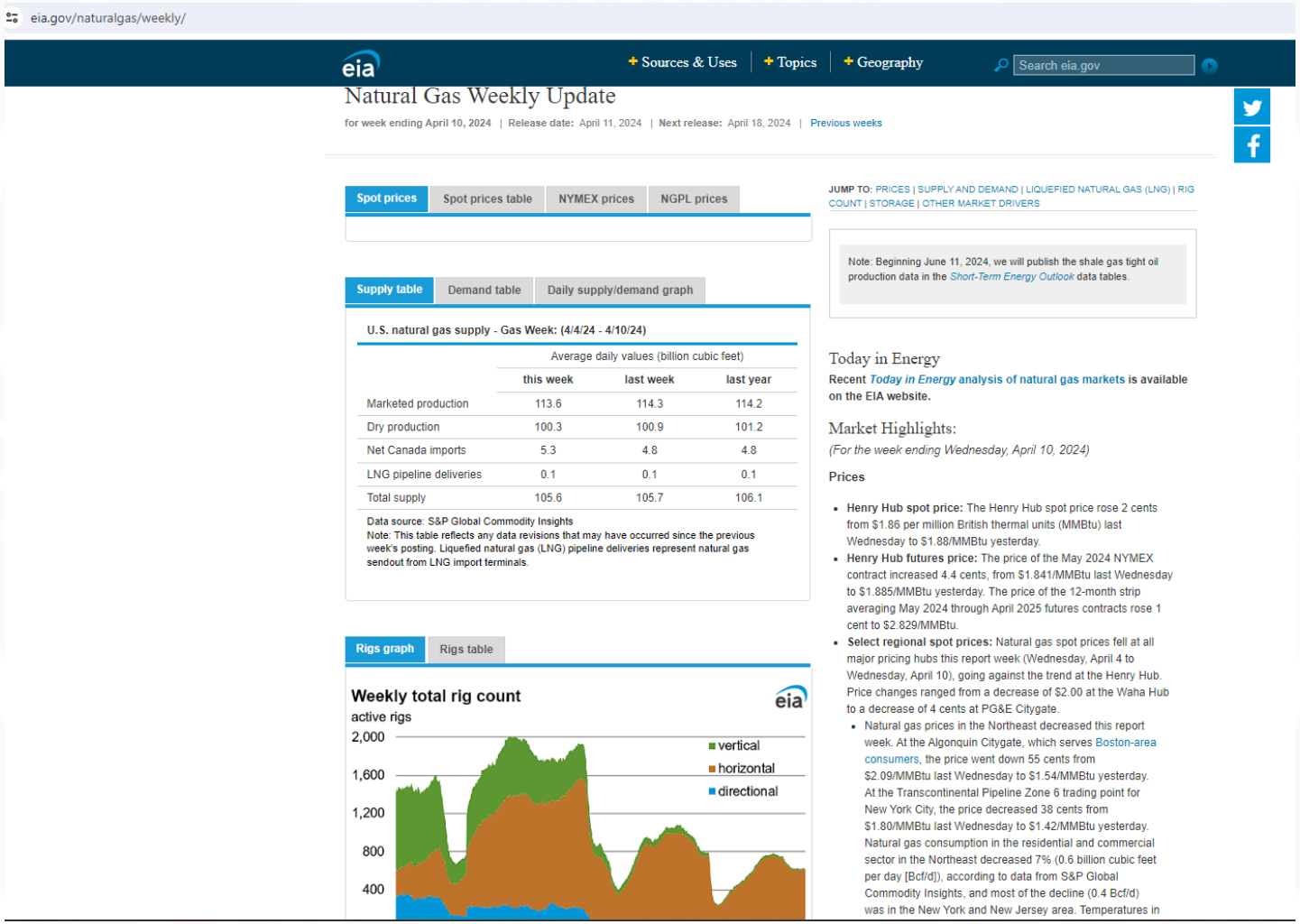
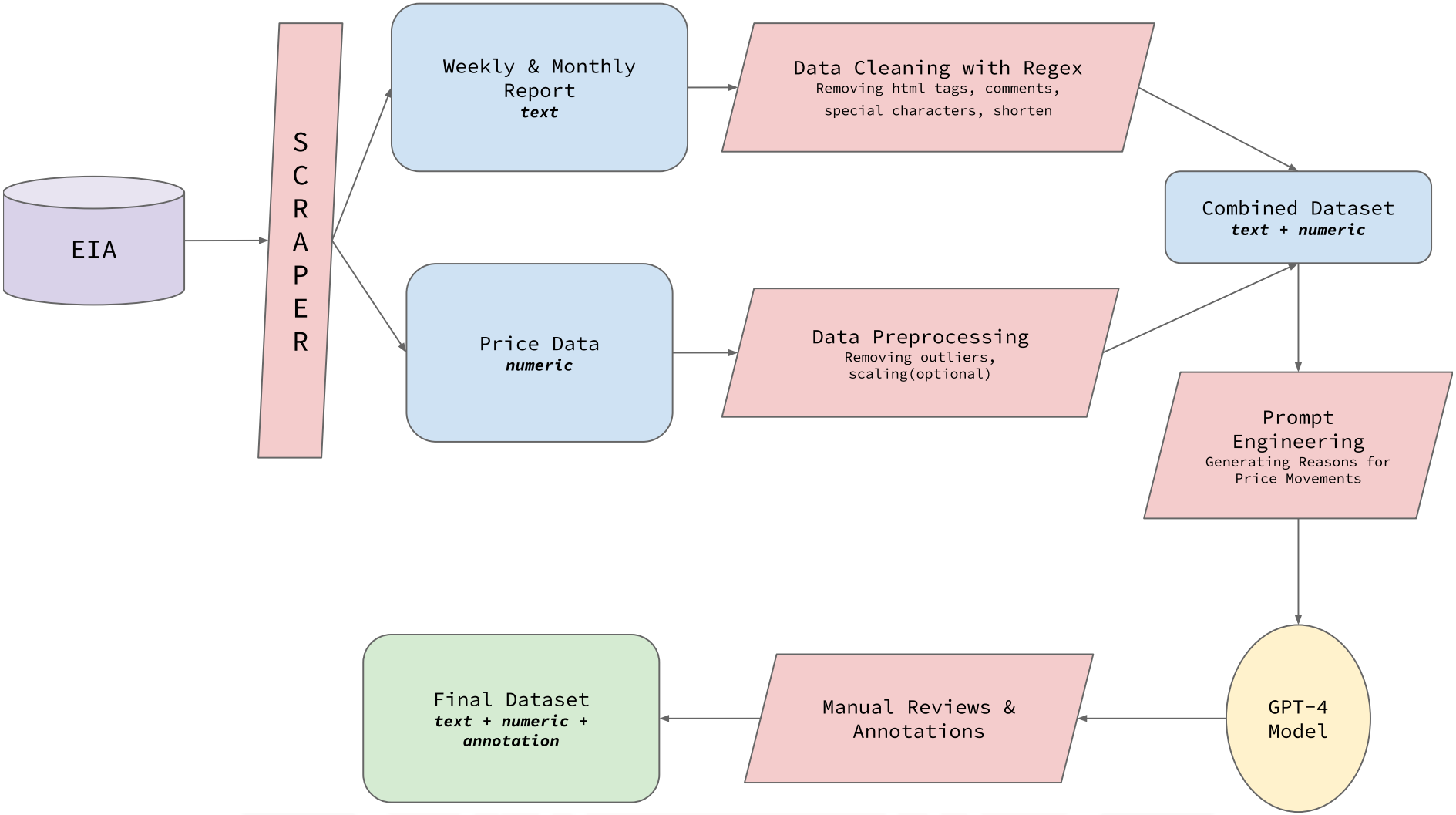


Figure 2: Example of Commodity Report in Natural Gas. Source: EIA

Plan of Work

- Download a) commodity markets reports, & b) futures price data and extract features for dataset
- Data annotation, feature engineering, data preprocessing & standardization
- Uni-Modal Baseline & Improved Models for time- and text- features
- Multi-Modal Model architectures for time- and text- features
- Evaluation
- Contribution:
  - A novel dataset for commodity price forecasting and reasoning
  - Combination of time series and text for a multi-objective model with mixed data inputs

Data Pipeline



Report Data Distribution

Table 1: Data Sources

Source of Data	Features			
	Type of Data	**Start**	**Documents**	**Size**
NG Reports	HTML Text	2001	1136	1.5M words
Oil Reports	Text	2011	533	580K words
STEO Reports	Text	1997	325	1M words
Futures Prices	Numerical	1983	NA	17,891 prices

Method Highlights

- **Time-series transformer:** transformer-architecture for time-series
- **LLM Fine-tuning LoRA:** reduced number of parameters to train
- **Cascaded Model Architecture:** using outputs of one model as inputs to the next model
- **Multi-objective joint model:** Combined prediction of text and forecasted price

Baseline Models

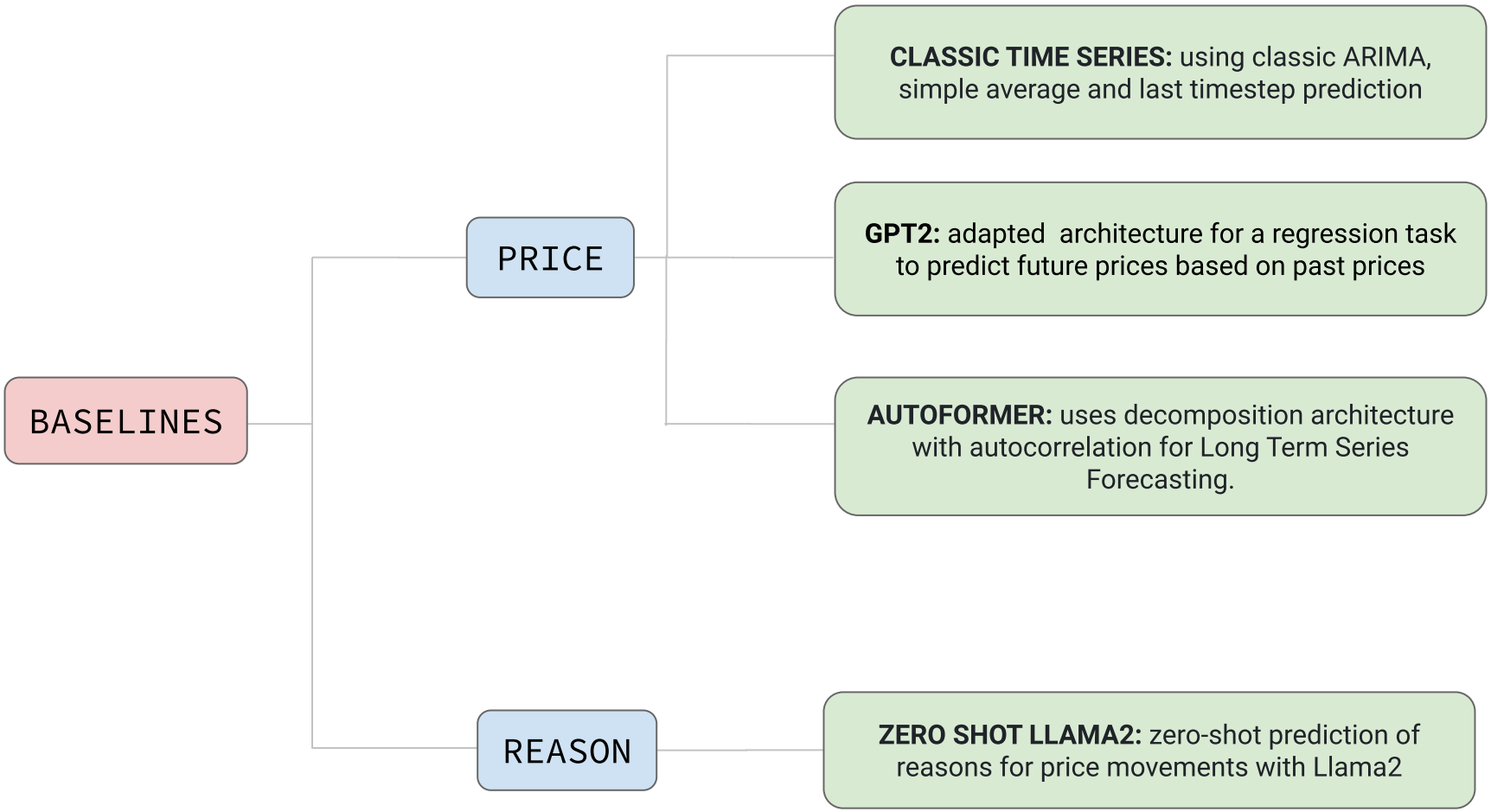
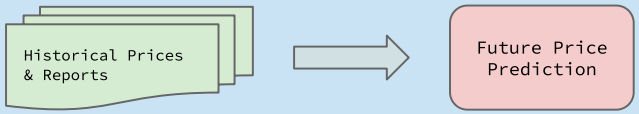


Figure 3: Baseline Models Tree Diagram.

Improved Models

Price Models



Eval Metrics: *MAPE, Pearson Coef, MSE, MAE*

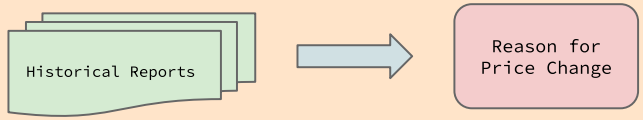
TIME SERIES TRANSFORMER
used for univariate probabilistic forecasting task

Parameters: learning_rate=6e-5, betas=(0.9, 0.95), weight_decay=1e-1, encoder_layers=8, decoder_layers=8, d_model=32, batch_size = 256

RoBERTa-base REGRESSION
adapted RoBERTa for predicting future prices from reports

Parameters: learning_rate = 1e-5, batch_size = 8

Reason Models



Eval Metrics: *Cross-entropy loss, Rouge Scores*

GPT-2-MEDIUM
generating reasons

Parameters: batch_size=4, weight_decay=0.01, learning_rate=5e-5

LLAMA-2-7b
fine tuned using LoRA to generate reasons.

Parameters: Lora: r=16, lora_alpha=32, lora_dropout=0.15, target_modules=["q_proj", "v_proj"]
trainable params= 8,388,608 / 6,746,804,224, trainable%= 12.4
Llama 2: max_steps=10000, learning_rate = 1e-4, batch_size=4

Figure 4: Improved Models Overview.

Final Models

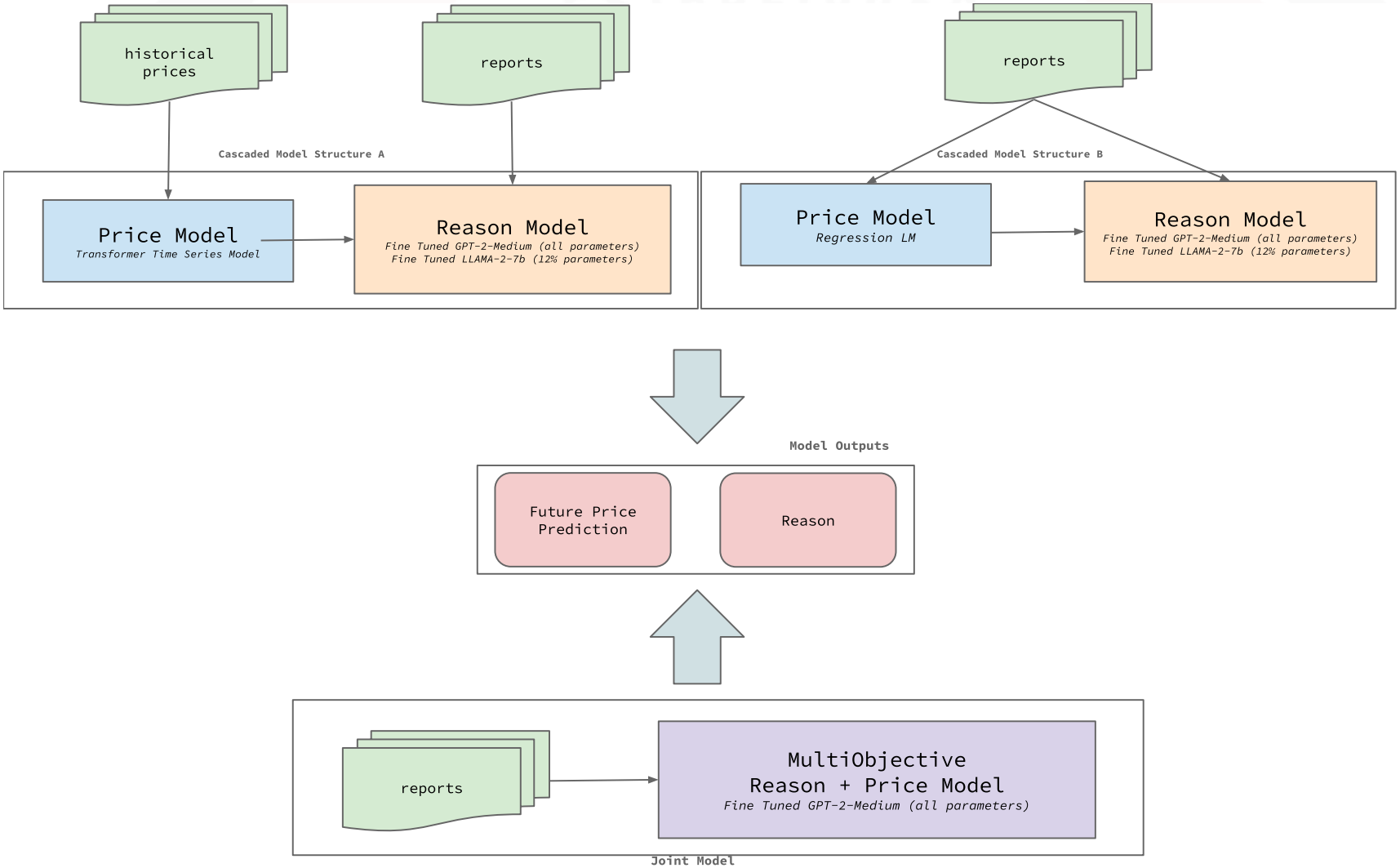


Figure 5: Extension Flow Chart.

Forecasting Results Performance

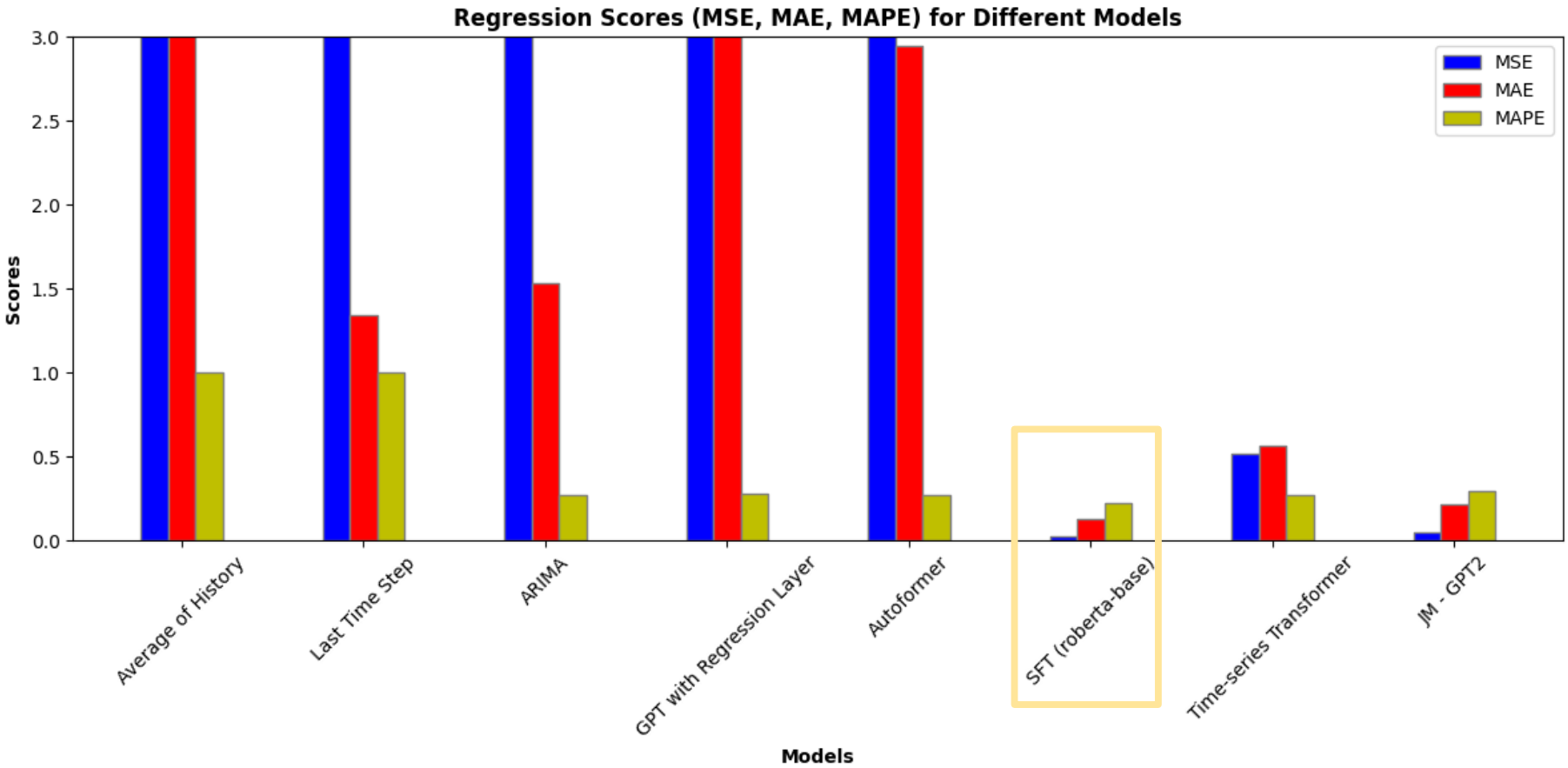


Figure 6: Performance by Model.

Forecasting Results: Predicted Vs. Actual

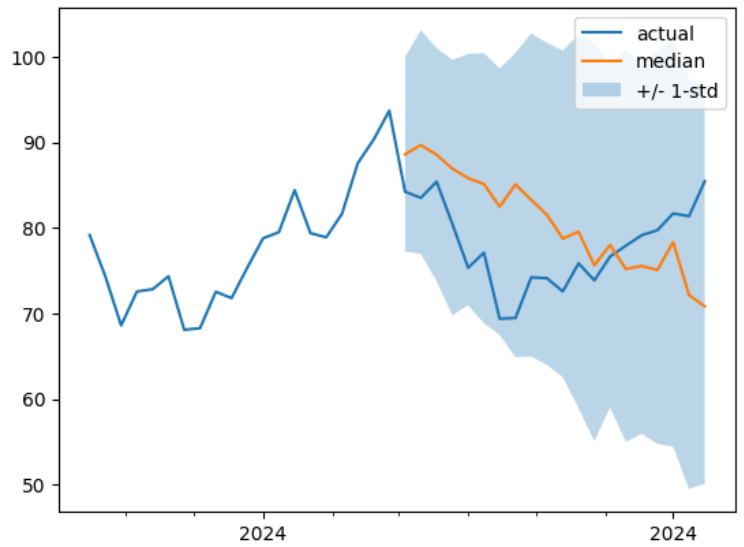


Figure 7: Predicted vs. Actual Transformer Time Series.

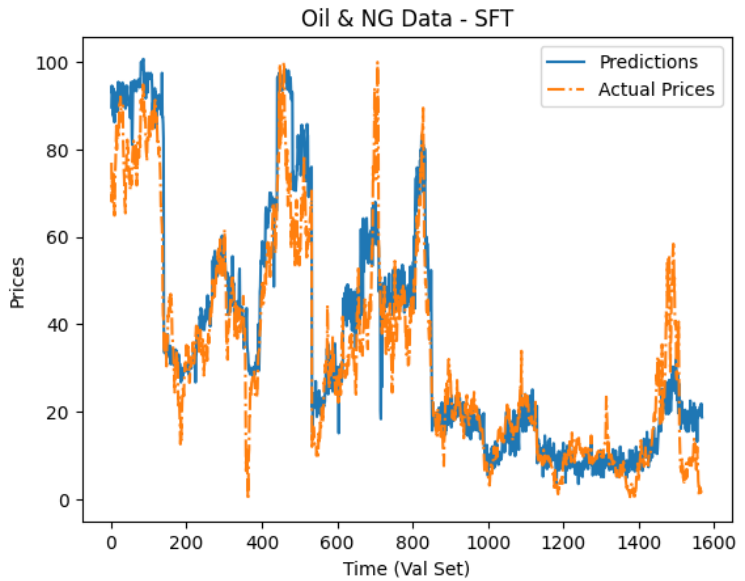


Figure 8: Predicted vs. Actual SFT.

Reason Model Performance

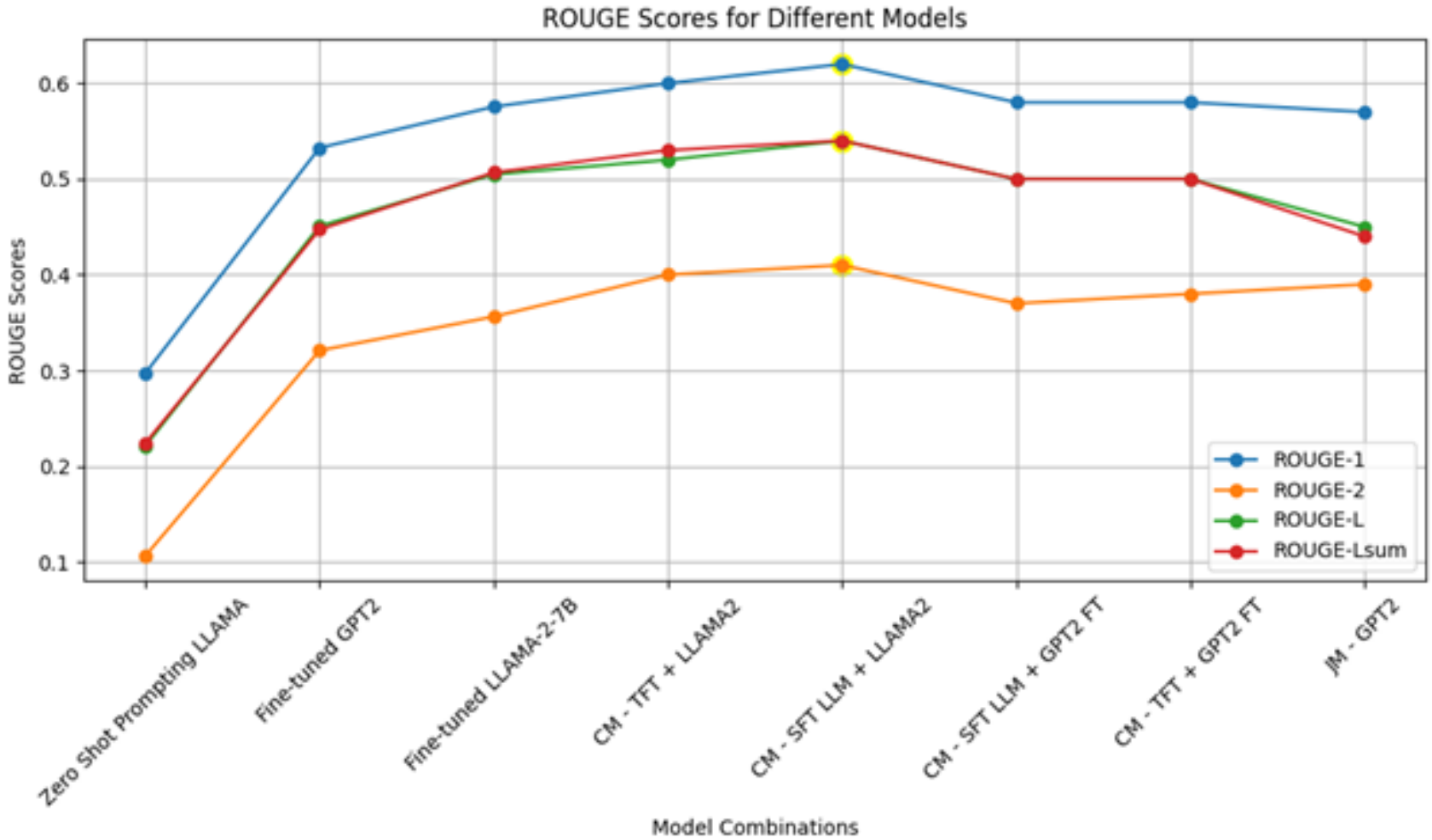
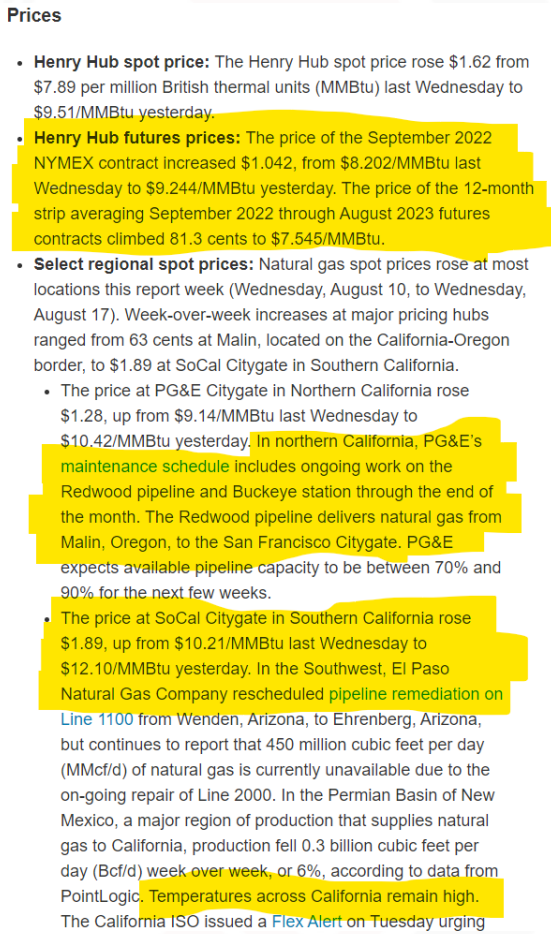


Figure 9: Rouge Scores by Model.

Reasons: Predicted Vs. Actual

- **Cascading TST+Llama2**
  Reason: "Today's natural gas price increased from last week due to volatile spot prices in the western United States driven by demand fluctuations in California and the Pacific Northwest, with SoCal Citygate and PG&E City"



Prices

- **Henry Hub spot price:** The Henry Hub spot price rose $1.62 from $7.89 per million British thermal units (MMBtu) last Wednesday to $9.51/MMBtu yesterday.
- **Henry Hub futures prices:** The price of the September 2022 NYMEX contract increased $1.042, from $8.202/MMBtu last Wednesday to $9.244/MMBtu yesterday. The price of the 12-month strip averaging September 2022 through August 2023 futures contracts climbed 81.3 cents to $7.545/MMBtu.
- **Select regional spot prices:** Natural gas spot prices rose at most locations this report week (Wednesday, August 10, to Wednesday, August 17). Week-over-week increases at major pricing hubs ranged from 63 cents at Malin, located on the California-Oregon border, to $1.89 at SoCal Citygate in Southern California.
  - The price at PG&E Citygate in Northern California rose $1.28, up from $9.14/MMBtu last Wednesday to $10.42/MMBtu yesterday. In northern California, PG&E's maintenance schedule includes ongoing work on the Redwood pipeline and Buckeye station through the end of the month. The Redwood pipeline delivers natural gas from Malin, Oregon, to the San Francisco Citygate. PG&E expects available pipeline capacity to be between 70% and 90% for the next few weeks.
  - The price at SoCal Citygate in Southern California rose $1.89, up from $10.21/MMBtu last Wednesday to $12.10/MMBtu yesterday. In the Southwest, El Paso Natural Gas Company rescheduled pipeline remediation on Line 1100 from Wenden, Arizona, to Ehrenberg, Arizona, but continues to report that 450 million cubic feet per day (MMcf/d) of natural gas is currently unavailable due to the on-going repair of Line 2000. In the Permian Basin of New Mexico, a major region of production that supplies natural gas to California, production fell 0.3 billion cubic feet per day (Bcf/d) week over week, or 6%, according to data from PointLogic. Temperatures across California remain high. The California ISO issued a Flex Alert on Tuesday urging

Figure 10: EIA Report. ▸ Report Link

Conclusion

- Combined economic and natural language processing insights to create explainable forecasting method for futures market
- Combined time-series transformer and llama2 to create cascading model
- Extensions and Next Steps:
  - Implement multi-task, multi-objective learning with time series and text features
  - Create fusion layer with time series and text features
  - Implementing cross-modality by adding cross-attention layers across both modals
  - Incorporating real-time news data

Thank You!

LoRa

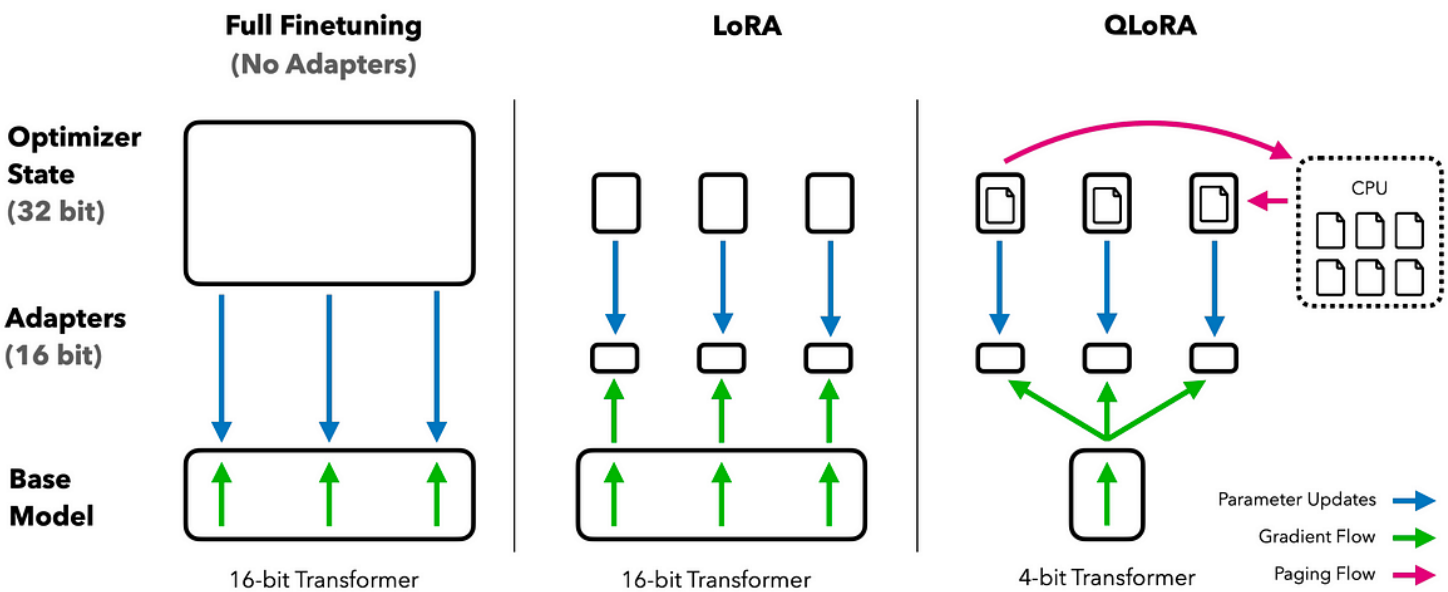


Figure 1: Different finetuning methods and their memory requirements. QLoRA improves over LoRA by quantizing the transformer model to 4-bit precision and using paged optimizers to handle memory spikes.

Figure 11: LoRa for fine-tuning model.

Time-series Transformer

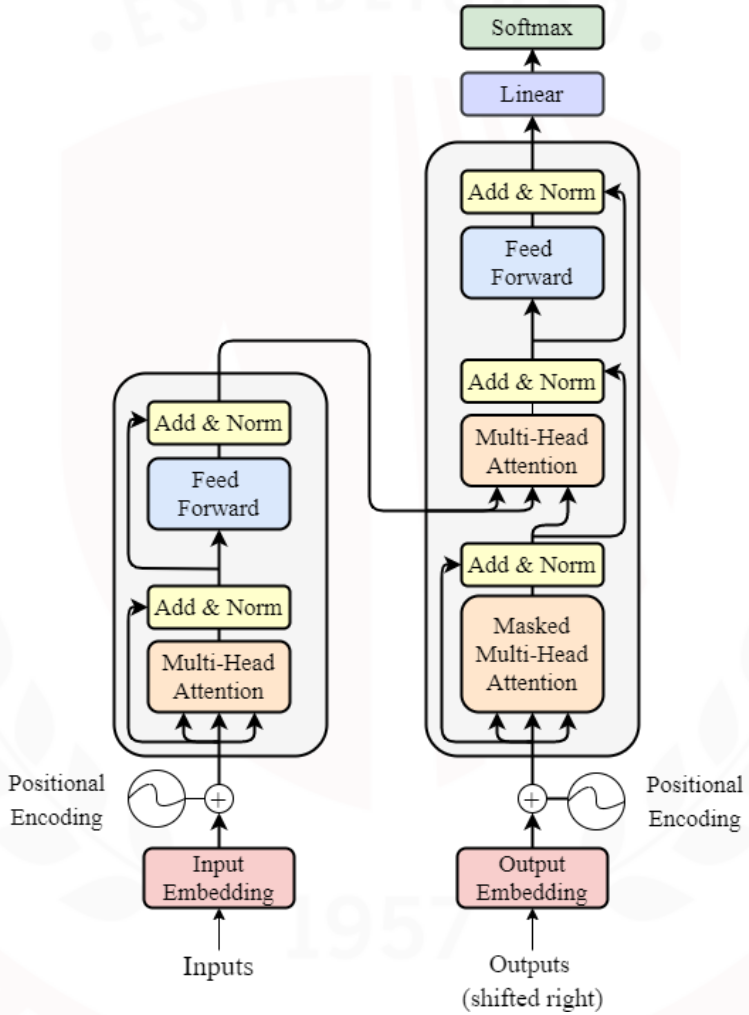


Figure 12: Transformer Architecture.

Pearson r Correlation

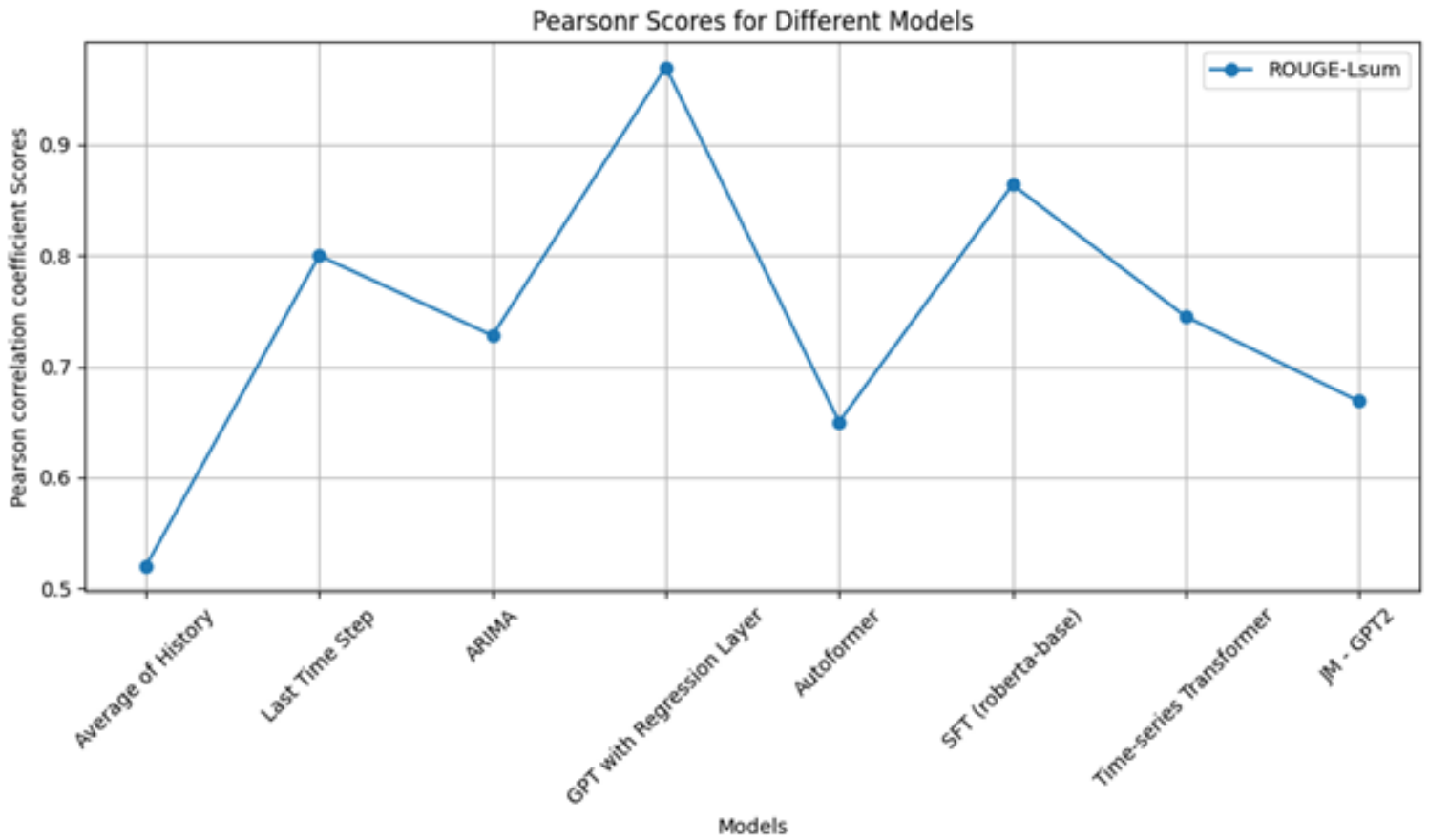


Figure 13: Pearson Correlation by Model.

List of References

[1] Michael K Adjemian. "Quantifying the WASDE announcement effect". In: *American Journal of Agricultural Economics* 94.1 (2012), pp. 238–256.

[2] Scott Baker et al. *What triggers stock market jumps?* Tech. rep. National Bureau of Economic Research Cambridge, 2021.

[3] Jean-Thomas Bernard et al. "Forecasting commodity prices: GARCH, jumps, and mean reversion". In: *Journal of Forecasting* 27.4 (2008), pp. 279–291.

[4] Gabriel David Bunek et al. "Characterizing the Effect of USDA Report Announcements in the Winter Wheat Futures Market Using Realized Volatility". PhD thesis. Montana State University-Bozeman, College of Agriculture, 2015.

List of References

[5] Qinkai Chen. "Stock movement prediction with financial news using contextualized embedding from bert". In: *arXiv preprint arXiv:2107.0872* (2021).

[6] CA Tapia Cortez et al. "Alternative techniques for forecasting mineral commodity prices". In: *International Journal of Mining Science and Technology* 28.2 (2018), pp. 309–322.

[7] Louis H Ederington et al. "EIA storage announcements, analyst storage forecasts, and energy prices". In: *The Energy Journal* 40.5 (2019), pp. 121–142.

[8] Barry Falk and Peter F Orazem. "A Theory of Future's Market Responses to Government Crop Forecasts". In: (1985).

[9] Luigi Gifuni. "NLP for analysis and forecasting of crude oil prices". PhD thesis. University of Glasgow, 2023.

List of References

[10] Yeong Hyeon Gu et al. "Forecasting agricultural commodity prices using dual input attention LSTM". In: *Agriculture* 12.2 (2022), p. 256.

[11] Girish Hegde, Vishwanath R Hulipalled, and JB Simha. "Price prediction of agriculture commodities using machine learning and NLP". In: *2021 Second International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*. IEEE. 2021, pp. 1–6.

[12] Joshua Huang, Teresa Serra, and Philip Garcia. "The Value of USDA Announcements in the Electronically Traded Corn Futures Market: A Modified Sufficient Test with Risk Adjustments". In: *Journal of Agricultural Economics* 72.3 (2021), pp. 712–734.

[13] Berna Karali et al. "Are USDA reports still news to changing crop markets?" In: *Food Policy* 84 (2019), pp. 66–76.

List of References

[14] Xuerong Li, Wei Shang, and Shouyang Wang. "Text-based crude oil price forecasting: A deep learning approach". In: *International Journal of Forecasting* 35.4 (2019), pp. 1548–1560.

[15] Scott C Linn and Zhen Zhu. "Natural gas prices and the gas storage report: Public news and volatility in energy futures markets". In: *Journal of Futures Markets: Futures, Options, and Other Derivative Products* 24.3 (2004), pp. 283–313.

[16] Kashif Rasul et al. "Lag-Llama: Towards Foundation Models for Probabilistic Time Series Forecasting". In: ().

[17] Adrien Rousset Planat Sarah Mouabb Evgenia Passari. "The Origins of Commodity Price Fluctuations". In: ().

List of References

[18] Xiaobin Tang and Nuo Lei. "Research on CPI Prediction Based on Natural Language Processing". In: *arXiv preprint arXiv:2303.05666* (2023).

[19] Haixu Wu et al. "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting". In: *Advances in neural information processing systems* 34 (2021), pp. 22419–22430.

[20] Shiyu Ye and Berna Karali. "The informational content of inventory announcements: Intraday evidence from crude oil futures market". In: *Energy Economics* 59 (2016), pp. 349–364.

[21] Jiahui Ying, Yu Chen, and Jeffrey H Dorfman. "Flexible tests for USDA report announcement effects in futures markets". In: *American Journal of Agricultural Economics* 101.4 (2019), pp. 1228–1246.

List of References

[22] W Yun. "Predictability of wti futures prices relative to eia forecasts and econometric models". In: *JOURNAL OF ECONOMIC RESEARCH-SEOUL-* 11.1 (2006), p. 49.

Sources

- Sources: [1, 17, 2, 4, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 19, 20, 21, 22]