Integrating Sentiment Analysis and Economic Indicators for GDP Forecasts and Prompt Engineering of GDP Prediction

Sai Sreenath, Kalva
UNM ID: 102027362
The University of New Mexico
Email: saisreenath@unm.edu

Gayathri, Nuthalapati
UNM ID: 102039336
The University of New Mexico
Email: ngnuthalapati1012@unm.edu

Rishitha, Kondrolla
UNM ID: 102029095
The University of New Mexico
Email: rkondrolla@unm.edu

***Abstract*—This project report explains about the integration of economic indicators and sentiment scores derived from news data to enhance GDP forecasting using time series models like Vector Autoregression (VAR), AutoRegressive Integrated Moving Average Exogenous model (ARIMAX), and Seasonal Autoregressive Integrated Moving Average Exogenous model (SARIMAX). In addition, the implementation of prompt engineering is based on the relationship between the trends of economic indicators, VAR forecasts, news sentiment scores across several categories like politics, business, etc, and overall sentiment scores of news to generate insights on GDP past trends and future forecasts.**

The proposed methodology will pre-process the structured and unstructured datasets and merge the economic indicators with the sentiment scores as exogenous variables for time series models in order to improve the forecasts of GDP. Major activities of this project are:

- **Sentiment analysis by the gpt-4o-mini model from OpenAI for deriving the sentiment scores of news data using LangChain.**
- **Integrating structured economic indicators with aggregated sentiment scores.**
- **VAR model along with other modified ARIMAX and SARIMAX for GDP forecasting.**
- **Generate insight based on the various visual trends and model outputs such as forecasting of GDP via appropriate prompt engineering.**

By effectively capturing sentiment scores in these important indicators in the economy, the forecast from the GDP was well predicted. The prompt engineering raises interpretability that offers insights for better economic decision-making.

***Index Terms*—Sentiment Analysis, GDP Forecasting, Vector Autoregression, ARIMAX, SARIMAX, Prompt Engineering, Big Data.**

I. Introduction

GDP forecasts allow policymakers, financial institutions, governments, corporations, and investors to plan and make appropriate decisions using the trends predicted. The increasing availability of unstructured data, such as news articles can be used for enhancing GDP forecasting. Sentiment analysis of news articles quantifies public opinion or sentiment from unstructured text can act as valuable information. Incorporating sentiment data together with certain economic indicators can gain economic insights potentially improving forecast accuracy.

This work explores the integration of structured economic indicators and sentiment scores derived from news data to enhance GDP forecasting. In this project, we used time series models like Vector Autoregression (VAR), ARIMAX, and SARIMAX models that incorporate both the structured data like economic indicators, S&P 500 index and finally sentiment scores from LLM by analyzing certain news articles that are used as exogenous variables. This work will include sentiment scores to capture the changes in economic performance through public sentiments, hence offering a better view of those factors that influence the rate of GDP.

Further, enhances interpretability by using prompt engineering techniques based on the relationships between economic indicators, GDP, and sentiment trends. Prompt engineering uses Large Language Models(LLMs), we used gpt-4o-mini model to generate textual insights that contextualize the outputs of the forecasting models. This approach makes the forecasts better and the corresponding actions that can be taken by decision-makers.

A. Objectives

The primary objectives of this project are:

- To pre-process and integrate structured economic indicators and unstructured sentiment data for forecasting using time series models.
- To evaluate the performance of VAR, ARIMAX, and SARIMAX models and also GDP forecasting with and without sentiment scores.
- To employ prompt engineering that can be used to get insights and interpretations based on visualizations of various important economic and sentiment trends.

B. Contribution

The contributions of this project are as follows:

- Demonstration of improved forecast accuracy by incorporating sentiment scores into time series models.
- Development of prompt engineering techniques to contextualize forecasting outputs and provide actionable insights.

- Exploration of ARIMAX and SARIMAX models alongside VAR for enhanced analysis of performance evaluation.

This project report below is organized as follows: Section II discusses related work in the fields of economic forecasting, sentiment analysis, and time series modeling. Section III describes the methodology of data preprocessing, model integration, and techniques of prompt engineering. Section IV presents the results and evaluation of the proposed approach. Section V discusses findings, challenges, and future directions. Finally, Section VI concludes the paper with contributions and implications of the findings.

II. Related Work

Traditional time series models used to forecast key indicators regarding the GDP, unemployment rate, and inflation rate have been one of the major research areas for a very long time. Big data technologies have enabled new methodologies to complement these traditional approaches. Therefore, with the increasing amounts of unstructured data there is an increasing area for the different applications using the data.

A. Economic Forecasting Using Time Series Models

Time series models such as Vector Autoregression (VAR), ARIMA, SARIMA and their variants (e.g., ARIMAX and SARIMAX) are mostly used in economic forecasting. VAR models are used to capture relationships between multiple endogenous variables, with applications in predicting GDP growth, trade volumes, and employment rates.

ARIMAX and SARIMAX models extend ARIMA by using exogenous variables, making them suitable for GDP prediction where external factors like economic indicators influence the dependent variable. These models are used to forecast economic indicators by combining behavioral factors that provide better performance [1]. These models often rely on structured data by limiting its ability to integrate other factors like the news sentiment.

B. Sentiment Analysis in Economic Forecasting

Sentiment analysis was used to extract public opinion and trends from unstructured data collected such as news articles, and social media posts. Previous research has shown the use of sentiment analysis in predicting stock market trends, consumer confidence, and business cycles [2]. For example, studies using Twitter sentiment data had shown positive correlations with short-term market movements [3].

Integration of sentiment data to the models was explored in different contexts like real-time stock price prediction [4]. However, its application in the prediction of annual economic indicators like GDP remains underexplored. This project addresses this by integrating sentiment scores as an exogenous factor in GDP forecasting models.

C. Integration of Structured and Unstructured Data

The integration of structured economic indicators with the unstructured sentiment data has different challenges like data alignment, feature engineering, and model integration. The aggregating techniques are used for the sentiment scores over time, aligning them with economic indicators such as annual GDP reports or unemployment rates [5]. In addition, advanced machine learning models and pre-processing methods have been developed to combine these different datasets [6].

Furthermore, this project incorporates prompt engineering by using GPT model to generate interpretive insights that provides both quantitative forecasts and qualitative understanding.

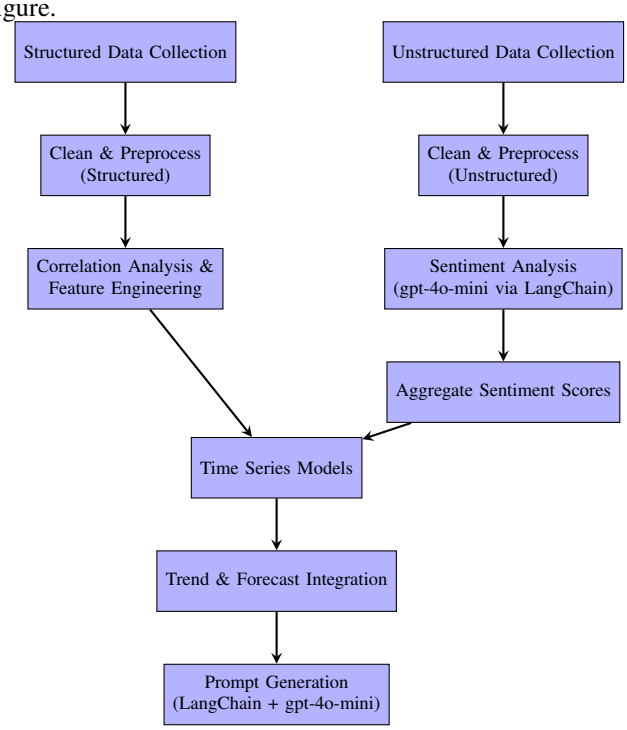
D. Gap and Contribution

The previous studies have shown the effective usage of sentiment analysis and advanced time series models independently but their combined application in forecasting GDP is relatively underexplored. This project contributes by:

- Introducing a combined framework that integrates structured economic and financial indicators and unstructured sentiment data for GDP forecasting.
- Evaluation of the performance of VAR, ARIMAX, and SARIMAX models and forecasting the GDP based on sentiment scores and other factors.
- Employing prompt engineering to predict and summarize the forecasting results.

III. Methodology

This section elaborate on the approach followed along the project, from collection/preprocessing to model implementation after analysis. The pipeline gives an overview of the next figure.



This project integrates structured economic data and unstructured news sentiment scores to forecast GDP using time

series models. This includes preprocessing structured and unstructured data, feature importance analysis, and implementing forecasting models. Additionally, trends and forecasts are leveraged with Large Language Models (LLMs) for prompt engineering.

A. Data Preprocessing

1) Structured Data: Economic indicators such as GDP, inflation, exports and imports, unemployment rates and many others were collected from World Bank datasets. Financial indicator namely S&P500 index information is collected from yfinance. Preprocessing steps include:

- Missing values in the structured data are handled using linear interpolation and values are ensured to be stationary by transformations using log and differencing methods.
- The features are normalized to be available for the models to be in a specific range as input for the models.

2) Correlation Analysis: A correlation analysis was performed to explore relationships between different economic indicators and S&P500 index rolling average values and volume to identify important features with respect to GDP. This analysis identified significant predictors of GDP. Figure 1 shows the correlation barplot with respect to GDP.

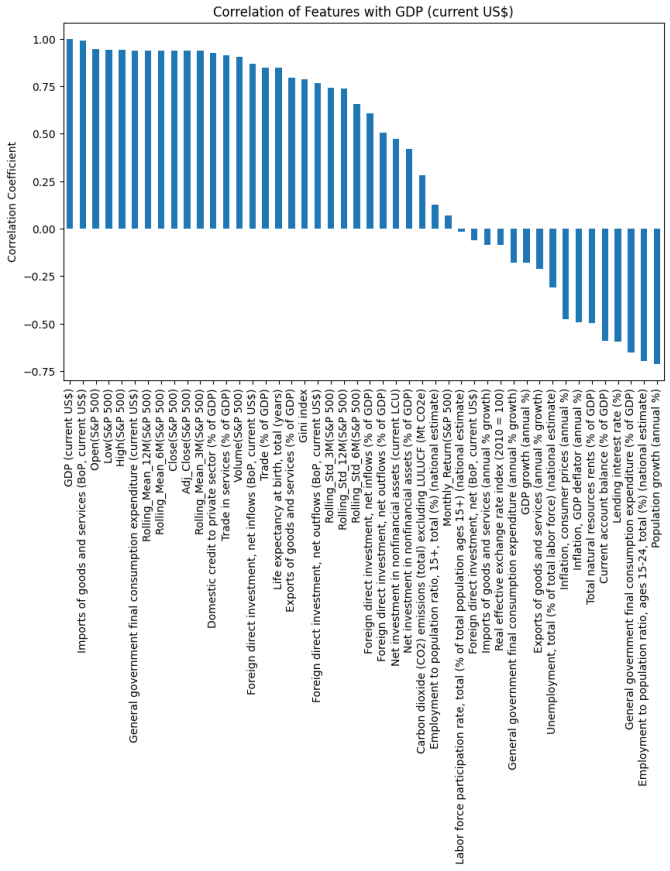


Fig. 1. Correlation Matrix of Economic and Financial Indicators

From the above plot, we can observe that with the increase of imports of goods and services and S&P 500 index values, the GDP rate increases. With the increase in inflation, employment to population, the GDP rate decreases indicating a strong negative correlation.

3) Seasonal Decomposition: Seasonal decomposition was performed on structured economic indicators to separate trends, seasonality, and residuals. This revealed underlying patterns in GDP rate. Figure 2 illustrates the decomposition of GDP.

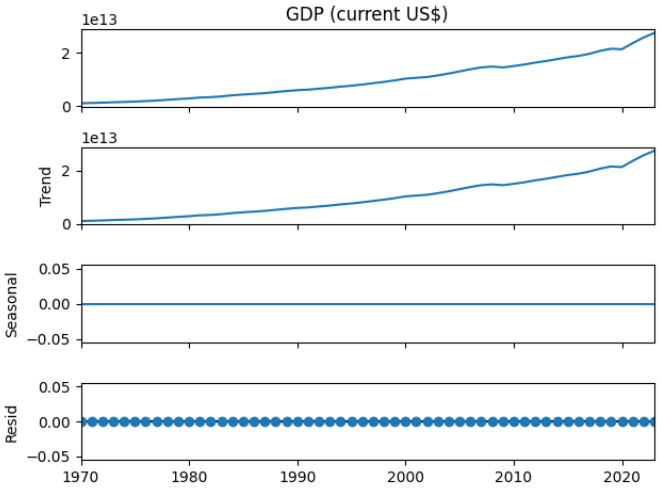


Fig. 2. Seasonal Decomposition of GDP

From the above plot, we can see the GDP rate only has the long term trend indicating that there is no seasonal trend and also there is no trend in residuals indicating zero residuals.

B. Feature Importance with XGBoost(Extreme Gradient Boosting)

The goal of this step is to rank the predictors based on their contribution to the accurate predictions of GDP values. XGBoost uses decision trees with gradient boosting and iteratively minimizing the loss function. To identify the most significant features that impact GDP, feature importance analysis was performed using the XGBoost machine learning algorithm. The analysis includes the structured economic indicators, such as unemployment rates together with the trade volumes and rank features with the importance based on weights. The top features that impact the change in GDP rate are the selected features for our project. Figure 3 highlights the feature importance scores.

	Feature	Importance
2	Domestic credit to private sector (% of GDP)	75.0
26	Unemployment, total (% of total labor force) (...	71.0
0	Carbon dioxide (CO2) emissions (total) excludi...	48.0
12	GDP growth (annual %)	46.0
17	Imports of goods and services (annual % growth)	44.0
6	Exports of goods and services (annual % growth)	43.0
5	Exports of goods and services (% of GDP)	20.0
24	Real effective exchange rate index (2010 = 100)	18.0
10	Foreign direct investment, net outflows (% of ...	16.0
30	Monthly_Return(S&P 500)	16.0
19	Inflation, consumer prices (annual %)	15.0
1	Current account balance (% of GDP)	14.0
23	Population growth (annual %)	13.0
21	Net investment in nonfinancial assets (% of GDP)	13.0
20	Lending interest rate (%)	11.0
4	Employment to population ratio, ages 15-24, to...	11.0
11	Foreign direct investment, net outflows (BoP, ...	9.0
29	Volume(S&P 500)	9.0
32	Rolling_Std_6M(S&P 500)	8.0
22	Net investment in nonfinancial assets (current...	8.0
13	General government final consumption expenditu...	8.0
14	General government final consumption expenditu...	6.0
7	Foreign direct investment, net (BoP, current US$)	6.0
31	Rolling_Std_3M(S&P 500)	5.0
33	Rolling_Std_12M(S&P 500)	5.0
18	Inflation, GDP deflator (annual %)	4.0
8	Foreign direct investment, net inflows (% of GDP)	4.0
3	Employment to population ratio, 15+, total (%)...	4.0

Fig. 3. Feature Importance Scores from XGBoost Model

C. Forecasting Models

The selected features are then fed to the time series models as the exogenous variables that are used for accurate GDP forecasting that can consider the interdependencies between different economic indicators as well as account for seasonal patterns and external influences. In this project we forecasted GDP using three time-series forecasting models: ARIMAX, SARIMAX, and VAR. Each model serves a specific purpose and provides complementary insights.

1) ARIMAX (Univariate TimeSeries model): The ARIMAX (AutoRegressive Integrated Moving Average with Exogenous Variables) model is the extension of the ARIMA model, which includes exogenous variables like economic indicators to improve the GDP prediction accuracy. The ARIMAX equation is:

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \sum_{k=1}^{n} \beta_k X_{t-k} + \epsilon_t \quad (1)$$

where:

- y_t: Target variable (GDP (Current US$).
- ϕ_i: AutoRegressive (AR) coefficients for lagged GDP values.
- θ_j: Moving Average (MA) coefficients for lagged error terms.
- X_{t-k}: Exogenous variables (e.g., Economic and financial indicators).
- β_k: Coefficients for exogenous variables.
- ϵ_t: Random error term at time t.

ARIMAX is effective in combining external influences which may not follow a pattern but impact the dependent variable. ARIMAX was applied to GDP data with values of economic and financial indicators as exogenous variables. Lag order was determined using AIC (Akaike Information Criterion) to minimize the complexity of the model while maintaining accuracy.

2) SARIMAX (Univariate TimeSeries model): The SARIMAX (Seasonal AutoRegressive Integrated Moving Average with Exogenous Variables) model extends the ARIMAX model by adding the seasonal components. The SARIMAX equation is:

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \sum_{k=1}^{n} \beta_k X_{t-k} + S_t + \epsilon_t \quad (2)$$

where S_t represents the seasonal component, modeled as:

$$S_t = \sum_{m=1}^{P} \Phi_m S_{t-m} \quad (3)$$

where P is the seasonal order, and Φ_m are seasonal coefficients.

SARIMAX is ideal for data with periodic fluctuations. Seasonal components were found using the ACF (AutoCorrelation Function) and PACF (Partial AutoCorrelation Function) plots. Performance metrics are shown in Table **??**.

3) Vector Autoregression (VAR) (Multivariate TimeSeries model): The VAR (Vector Autoregression) model is a multivariate time series model that considers the dependent variables (e.g., GDP, unemployment, trade volumes) as endogenous. This allows the model to capture the inter-dependencies among all the variables. The VAR model is expressed as:

$$Y_t = c + \sum_{i=1}^{p} A_i Y_{t-i} + \epsilon_t \quad (4)$$

where:

- Y_t: Vector of endogenous variables at time t.
- A_i: Coefficient matrices for lagged variables.
- c: Constant vector.
- ϵ_t: Error term vector.

VAR is particularly suited for understanding the dynamic relationships between variables, such as the lagged effect of sentiment scores on GDP. The optimal lag length for the VAR model was determined using the AIC and BIC criteria. Impulse response functions were used to analyze the effect of shocks in sentiment scores on GDP over time.

D. Unstructured Data Preprocessing

The unstructured dataset collected from Huffpost across several categories, consisting of news headlines and short descriptions was filtered and pre-processed to analyze sentiment scores using gpt-4o-mini model via langchain. The methodology for cleaning, analyzing, and aggregating sentiment data is explained in the following sections.

1) Data Cleaning: The preprocessing of text data includes preserving essential information that has text, whitespace, and some punctuation marks by removing the special characters and redundant whitespaces using regular expressions.

The preprocessing function used for cleaning text is shown in the code snippet below:

```python
import re

def preprocess_text(text):
    text = re.sub(r"[^\w\s.,!?]", '', text)  # To remove special characters
    text = " ".join(text.split())
    return text

df_news['headline_cleaned'] = df_news['headline'].apply(preprocess_text)
df_news['short_description_cleaned'] = df_news['short_description'].apply(preprocess_text)
```

Listing 1. Text Preprocessing Function

2) Sentiment Analysis: The headlines and descriptions that are cleaned in the previous step are fed to the gpt-4o-mini model via langchain hosted on openAI. Using LLM models sentiment analysis was done to categorize each news item as *Positive*, *Negative*, or *Neutral*. Additionally, the sentiment of each news article was quantified ranging from 0 (neutral) to 1 (highly intense). The sentiment analysis includes:

- The sentiment category (*Positive News*, *Negative News*, *Neutral*) and the polarities of the corresponding sentiment categories are 1, -1, and 0.
- The intensity of each sentiment category ranging from 0(neutral) to 1(highly intense)
- Computing a weighted sentiment score by multiplying polarity with the intensity.

The weighted sentiment scores are calculated using the formula below-

Weighted Sentiment Score = Polarity of Sentiment $\times$ *Intensity of Sentiment.*

3) Aggregation of Sentiment Scores: The calculated sentiment scores were aggregated by year and category to provide annual trends. The weighted sentiment scores were calculated as follows:

$$\text{Weighted Sentiment (Yearly)} = \frac{\sum(\text{Weighted Sentiment Score})}{\text{Number of Articles}} \quad (5)$$

These aggregated scores were pivoted by category for analysis, as summarized in Table I.

TABLE I
SAMPLE AGGREGATED SENTIMENT SCORES BY YEAR AND CATEGORY

Year	Politics	Business	U.S. News	World News	Overall
2020	-0.43	-0.42	-0.42	-0.42	-0.19
2021	-0.31	-0.16	-0.32	-0.29	-0.22
2022	-0.31	0.13	-0.28	-0.42	-0.05

4) Visualization of Sentiment Trends: The trends of average of weighted sentiment scores across years were visualized to identify patterns and their influence on economic indicators. Figure 4 shows the average of the overall news articles aggregated weighted sentiment scores annually.

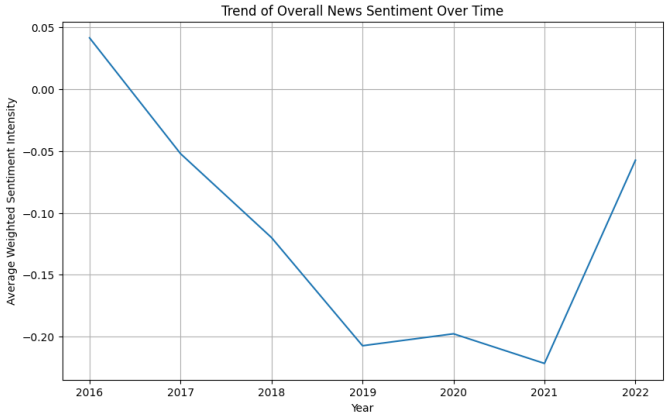


Fig. 4. Yearly Weighted Sentiment Intensity Across Categories

E. Integration of Structured and Unstructured Data

The sentiment scores were combined with structured economic indicators and used as dependent variables in the VAR model to forecast GDP. The combined dataset enhanced the model's ability to capture qualitative factors influencing GDP trends.

F. LLM-Based Prompt Engineering

Trends in economic indicators, sentiment scores, and GDP forecasts were fed into a Large Language Model (LLM) - gpt-4o-mini model to generate interpretive prompts.

The prompt used to feed images and analyze is -

```
message_prompt = [
        HumanMessage(
            content=[
                {"type": "text",
                    "text": f"""
                    Analyze the visualizations
                        of the following five
                        images:
                    1. The actual GDP trend
                    2. The GDP trend with log
                        transformation applied
                        following with the
                        first and second
                        differencing
                    3. The plots of top
                        economic indicators
                        that impact GDP (
                        current US$) in a
                        single image.
                    4. The VAR forecast past
                        and future predictions
                        with the economic
                        indicators as features.
                    5. The plots of various new
                         sentiment intensities
                        over the years and also
                         the average sentiment
                        intensity of various
                        categories


                    Answer the following
                        questions based on the
                        plots.
```

```
                    Question: {text}
                    Answer: (your response here
                        ...)
                    """
                },
                {"type": "image_url",
                    "image_url": {"url": f"data
                        :image/png;base64,{
                        GDP_TREND}"}
                },
                {"type": "image_url",
                    "image_url": {"url": f"data
                        :image/png;base64,{
                        GDP_LOG_DIFF2}"}
                },
                {"type": "image_url",
                    "image_url": {"url": f"data
                        :image/png;base64,{
                        ECONOMIC_INDICATOR_IMAGES
                        }"}
                },
                {"type": "image_url",
                    "image_url": {"url": f"data
                        :image/png;base64,{
                        VAR_FUTURE_FORECAST}"}
                },
                {"type": "image_url",
                    "image_url": {"url": f"data
                        :image/png;base64,{
                        NEWS_SENTIMENT}"}
                }
            ]
        )
    ]
```

Listing 2. Text Preprocessing Function

IV. Results

This section explains about the performance of both univariate time-series forecasting models namely ARIMAX, SARIMAX and multivariate time-series forecasting model namely VAR model. The integration of structured and unstructured data is evaluated for its impact on GDP forecasting and the prompt engineering result by the LLM model.

A. Forecasting Results

1) ARIMAX and SARIMAX Models: The performance of ARIMAX and SARIMAX models where economic and financial indicators were included as exogenous variables. The following are the rmse values for test data trained using both the models -

TABLE II
Performance Metrics of ARIMAX and SARIMAX Models

Model	RMSE
ARIMAX	0.038
SARIMAX	0.045

2) Vector Autoregression (VAR): The VAR model effectively captured interdependencies among GDP, unemployment rates, trade, and other important features. The following Figure 5 illustrates the past trends of GDP over the years and the future forecasts. The RMSE value for the test data is 0.04

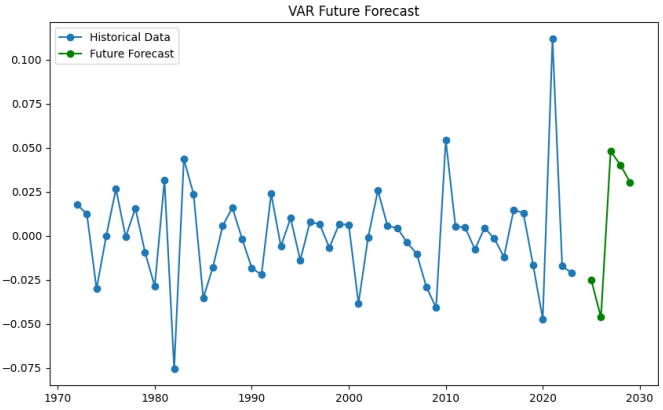


Fig. 5. VAR forecasts and historical trends of GDP

B. Integration of Sentiment and Economic Trends

The integration of sentiment scores with economic indicators enhanced the interpretability of forecasting models. Figure **??** shows the alignment between sentiment trends and GDP over time, indicating the potential of sentiment data to predict economic performance. The RMSE value of forecasts when sentiment scores are included is 0.05 and the trends of future GDP forecasts and past trends of GDP is shown in the figure below-

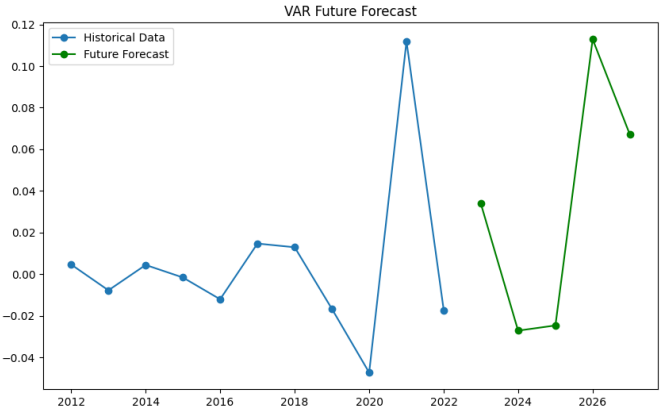


Fig. 6. Alignment of Sentiment Scores and GDP Trends

C. Prompt Engineering Results

Trends and forecasts were fed to a Large Language Model (LLM) - gpt-4o-min model to generate prompts. The generated responses provided actionable insights, linking sentiment and economic indicators to GDP forecasts. Sample prompts and responses are presented in the figure below- **??**.

V. Discussion

The integration of sentiment analysis with economic indicators in this project provides significant insights of GDP forecasting. This section discusses the key findings, strengths, limitations, and potential future directions based on the results.

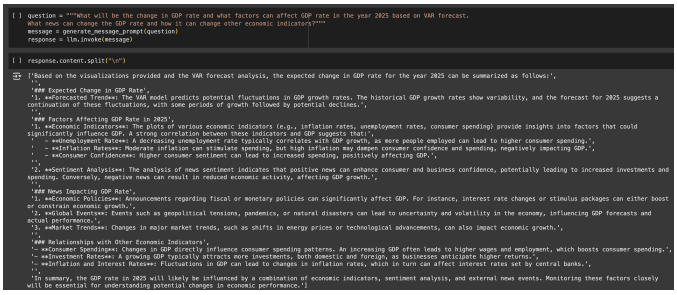


Fig. 7. The response of LLM based on the trends of sentiments and structured indicators

A. Key Findings

- **Enhanced Forecast Accuracy**: The inclusion of sentiment scores in VAR models increases the accuracy of GDP prediction. This shows the importance of using unstructured data for traditional forecasting models.
- **Feature Importance**: XGBoost analysis identified the top predictors of GDP for the traditional indicators like unemployment rates and trade volumes (Figure 3).
- **Interpretability through Prompts**: Prompt engineering effectively contextualized model outputs and provided the actionable insights.

B. Limitations

Despite its strengths, the project faced several limitations:

- **gpt-4o-mini Model Limitations**: The model which we used has certain limitations like 500 Requests/min and 10000 Requests/day. Therefore, we divided our data into a set of 10000 articles and added a delay of 60 sec for every 500 asynchronous calls made to the model.
- **Data Alignment**: Aligning the annual economic and financial indicators with daily sentiment scores required heavy preprocessing.
- **Sentiment Model Limitations**: The GPT-based model was trained on general datasets that might not completely capture nuances specific to economic or financial news.
- **Evaluation of Prompts**: The evaluation of prompt outputs was qualitative, and further work is needed to establish standardized metrics for interoperability.

C. Future Work

- **Improved Sentiment Models**: Training sentiment analysis models on domain-specific datasets for example, financial news trained model could improve the accuracy of sentiment scores that can improve GDP forecasting.
- **Advanced Time Series Models**: Exploring non-linear models namely Long Short-Term Memory (LSTM) networks or Transformer-based architectures, could capture more complex relationships in the data.
- **Real-Time Forecasting**: Implementing real-time data pipelines for sentiment scores and economic indicators can enable dynamic forecasting and better immediate insights.

VI. Conclusion

This project report focuses on the integration of sentiment analysis with traditional economic indicators in advanced time series models for GDP forecasting, and prompt engineering. By integrating structured data with unstructured sentiment scores, this project shows significant improvement in the accuracy and interpretability of forecasts.

A. Summary of Contributions

The key contributions of this work are as follows:

- **Enhanced Forecasting Models**: Incorporating sentiment scores into time series models improved forecasting accuracy.
- **Sentiment as a Predictor**: Sentiment scores were identified as a critical predictor of GDP, alongside traditional indicators such as unemployment rates and trade volumes.
- **Actionable Interpretability**: Prompt engineering provided contextual insights into model outputs, bridging the gap between quantitative forecasts and practical decision-making.
- **Unified Framework**: A methodology for integrating structured and unstructured data was presented, enabling a comprehensive approach to economic forecasting.

Combining structured and unstructured data for GDP forecasting offers better research opportunities to act immediately according to the news that we know. This project highlights the potential benefits of using different data sources and methods in data-driven economics.

References

[1] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.
[2] A. T. Stephen and J. Galak, "The complementary roles of traditional and social media in driving marketing performance," *Marketing Science*, vol. 31, no. 4, pp. 493–512, 2012.
[3] X. Zhang, H. Fuehres, and P. A. Gloor, "Predicting stock market indicators through Twitter 'I hope it is not as bad as I fear'," *Procedia Social and Behavioral Sciences*, vol. 26, pp. 55–62, 2011.
[4] R. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, no. 1, pp. 1–8, 2011.
[5] S. Mishra, "Combining structured and unstructured data for macroeconomic forecasting," *Big Data Research*, vol. 10, pp. 21–30, 2017.
[6] M. K. Narayan and K. Chaturvedi, "Machine learning for economic data integration," *AI and Economics*, vol. 8, no. 3, pp. 115–130, 2020.