# TSLA-Headlines Report

Hayden Kwok, Nathan Wang, Kliment Ho

Fall 2024

## 1 Dataset Identification and Exploratory Analysis

We are working with the following datasets:

- **Tesla Data (tsla.csv)**: Contains daily Tesla stock prices with columns such as `Date`, `Close`, `Open`, etc.
- **News Data**:
  - **Reuters**: Headlines and publication times.
  - **Guardian**: Headlines and publication times.
  - **CNBC**: Headlines and publication times.

### 1.1 Exploratory Analysis

**Descriptive Statistics**

- Total number of headlines: 53330.
- Average headline length: 8.08 words.
- Time range: 2017-12-21 to 2020-07-17.
- Average daily headlines: 15.68.
- Average Tesla closing price: \$26.57.
- Maximum Tesla closing price: \$103.07.
- Minimum Tesla closing price: \$11.93.

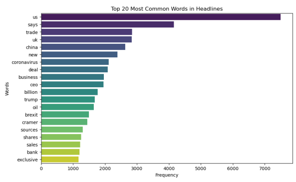



Figure 1: Top 20 Common Words in Headlines

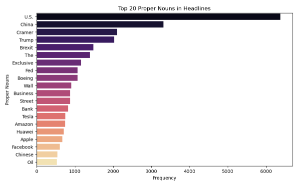



Figure 2: Top 20 Pronouns in Headlines

**Class Distribution**

The target variable (`Target`) which describes the stock movement is derived as:

- **Upward Trend (1)**: Next day's closing price (`Close(t+1)`) $>$ current day's closing price (`Close(t)`).
- **Downward Trend (0)**: Next day's closing price (`Close(t+1)`) $\leq$ current day's closing price (`Close(t)`).

The dataset is approximately balanced, with:

- Upward Trends: 51.17
- Downward Trends: 48.83

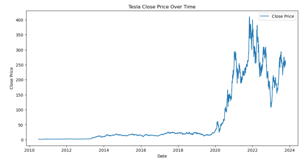



Figure 3: Tesla's Closing Price over Time

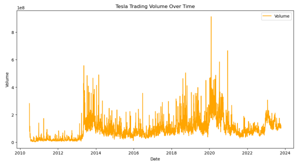



Figure 4: Tesla's Trading Volume

**Interesting Findings**

- Sentiment analysis showed no meaningful correlation ($r = -0.0036$) between positive/negative headlines and Tesla stock trends, indicating that sentiment alone may not drive price movements.
- Word frequency analysis revealed that the most common terms included "us," "says," "uk," "trade," and "china," reflecting a focus on international relations and economic topics. Other frequent words like "coronavirus," "deal," and "billion" suggest that headlines were heavily influenced by major global events and business agreements.
- Stock trends were more volatile on days with higher news volume.
- Tesla's closing price grew rapidly during 2020, the height of the global COVID-19 pandemic which 'coronavirus' appears as a common word.
- Tesla's trading volume spiked the highest during 2020.
- No meaningful correlation (Pearson $r = 0.00$) was observed between headline sentiment and upward trends. This suggests that sentiment alone does not significantly influence Tesla's stock movements.

# 2 Predictive Task and Evaluation

## 2.1 Predictive Task

Our task is to predict Tesla's stock price movement (upwards or downwards) based on the sentiment and content of the headlines. Given that the stock market is inherently volatile and its price movements are influenced by various factors, the headlines from the three previously established major media outlets were used as input features to understand the potential impact of news sentiment on stock prices.

We engineered several features to prepare the data for modeling. First, lag features were created by incorporating previous day's headlines to predict the stock price movement for the following day. This approach is common in time series forecasting, where past information helps predict future trends. Additionally, the data was scaled to ensure that there were no extreme gaps in the features, which could potentially distort model training. We employed a TF-IDF (Term Frequency-Inverse Document Frequency) Vectorizer to convert the headlines into numerical vectors. This vectorizer helps capture the frequency of relevant words in the headlines, such as terms closely associated with Tesla and the automotive industry, and weighs them according to their significance across the entire dataset.

The performance of the predictive model was evaluated using accuracy as the primary metric, which is standard for classification tasks. For comparison, a random guessing model, which assumes a 50% chance of either upward or downward movement, was used as a baseline.

### 2.2 Features

Several key features worth noting in relevance to predicting stock movement are opening price, closing price, highest/lowest prices, adjusted close and trading volume, key metrics determining the prediction model. Mentioned previously, engineering lag features aids to gauge stock growth based on the correct time occurrence of headline presented.

We also employed a TF-IDF vectorizer which aims to capture "importance" of words in relativeity to the model. We hope that words like "Tesla", which is one of the top 20 pronouns in news headlines, and "Elon Musk" would have large impact in making predictions.

Recognizing that news sentiment may impact investor behavior and market dynamics, we processed these headlines to extract meaningful information. It closely involved cleaning the text by converting it to lowercase, removing non-alphabetic characters, and eliminating stop words to focus on the most impactful terms.

## 3 Model Description

The model used in this study is the XGBClassifier, a gradient boosting classifier known for its effectiveness in handling tabular data and sparse features, which makes it an excellent fit for datasets processed using TF-IDF vectorization. The choice of this model was guided by its robustness in predictive tasks, particularly when relationships between features are nonlinear and complex.

To evaluate the model, accuracy was used as the primary metric, as the task involves binary classification (predicting stock price movement as upward or downward). This metric was chosen to directly compare model performance against a random baseline, which assumes a 50% chance of either outcome. Logistic regression and support vector machines (SVMs) were also implemented as baselines due to their simplicity and computational efficiency. However, these models struggled with sparse data and did not capture complex feature interactions, leading to lower accuracy compared to XGBClassifier.

TF-IDF proved to be an effective feature representation by capturing the importance of words like "Tesla" or "Musk" in the context of headlines. This representation outperformed simple bag-of-words and sentiment-only features, as it highlighted terms with high relevance to the stock market. Lag features, incorporating prior-day headlines, added temporal context, allowing the model to capture delayed impacts of news on stock price movements.

Hyperparameter tuning was performed using RandomizedSearchCV, exploring parameters such as learning rate, maximum tree depth, and regularization. This optimization helped mitigate overfitting and improved generalization, especially given the limited size of the dataset.

The dataset presented several challenges:

- Size: The limited timeframe (2017–2020) and small number of daily headlines constrained the model's ability to learn robust patterns, resulting in overfitting during training.
- Noise: Headlines often included general market or unrelated news, introducing noise into the dataset. Preprocessing steps reduced this but could not eliminate it entirely.
- Class Imbalance: Although the class distribution was approximately balanced, the variability in stock price movement influenced prediction consistency.
- Sparse Features: The TF-IDF vectorizer produced high-dimensional data, which, while informative, increased computational requirements and risked overfitting.

Other models considered include logistic regression and LSTMs. Logistic regression struggled with sparse data, achieving accuracy close to random guessing. LSTMs, while promising for temporal data, required more data points for effective training, making them unsuitable given the current dataset size. These comparisons highlight the balance between model complexity and dataset constraints.

Overfitting was partially mitigated by using regularization techniques and early stopping during training. However, the presence of noise in the data, such as irrelevant headlines or sentiment discrepancies, remained a limitation, likely contributing to the model's modest accuracy of 54%. Despite this, the XGBClassifier's results indicate some predictive power, albeit limited by dataset shortcomings.

The combination of TF-IDF features, lagged sentiment values, and gradient boosting provided a baseline for understanding the relationship between news headlines and Tesla's stock price movement. Future work could address dataset limitations by extending the time range, incorporating additional news sources, and exploring deep learning models capable of capturing richer text representations and temporal dependencies.

# 4 Literature Review

The task of predicting stock price movements based on news sentiment has been a subject of interest in the field of financial data science. Several studies have utilized sentiment analysis of financial news to forecast stock market trends. For example, research has shown that incorporating news articles into stock price prediction models can improve forecasting accuracy, particularly when combined with technical indicators and historical price data.

Datasets similar to the one we used, the historical stock price $TSLA, such as the Financial Times dataset or S&P 500 news datasets, have been continuously employed in previous and existing works to further optimize trading algorithm development. These datasets often contain news articles about publicly traded companies, and sentiment analysis has been applied to predict short-term price movements. In contrast to these datasets, our study utilized headlines from multiple international media outlets, which provided a broader view of how global sentiment might affect a specific company's stock price.

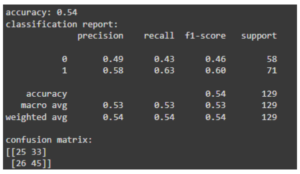
State-of-the-art methods in the NLP portion of this domain often involve deep learning models, such as transformers (BERT, GPT) and recurrent neural networks (LSTMs), which can capture both the semantic meaning of the text and temporal dependencies between news events and stock prices. While these models have shown success, their application requires large amounts of data to train effectively. The limitations of our dataset in terms of size and temporal scope may have constrained the performance of the model compared to more advanced models in existing literature.

In regards to the actual big picture regarding the concept of predicting stock price trends, the overall most efficient and optimized methodologies are still yet to be determined and are ever-changing, spearheaded by Quantitative Analysts and Traders. Therefore, much of the findings and results of other literature models share widely similar results in having a somewhat insignificantly unpredictable accuracy percentage, but often share different results based on targeted dataset.

# 5 Results and Conclusions

### Model Performance

- **Overall Performance:**
  - Accuracy: 54%


accuracy: 0.54
classification report:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.49 | 0.43 | 0.46 | 58 |
| 1 | 0.58 | 0.63 | 0.60 | 71 |
| accuracy | | | 0.54 | 129 |
| macro avg | 0.53 | 0.53 | 0.53 | 129 |
| weighted avg | 0.54 | 0.54 | 0.54 | 129 |

confusion matrix:
[[25 33]
[26 45]]

Figure 5: Model Performance Results

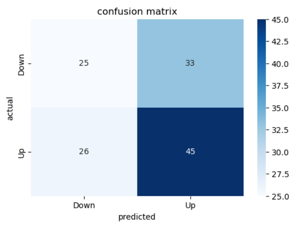



Figure 6: Confusion Matrix

- Weighted Precision: 54%
- Weighted Recall: 54%
- Weighted F1-Score: 54%

- **Class-Wise Performance:**
  - Class ‘0‘ (Downward Trend):
    * Precision: 49%
    * Recall: 43%
    * F1-Score: 46%
  - Class ‘1‘ (Upward Trend):
    * Precision: 58%
    * Recall: 63%
    * F1-Score: 60%

The XGBoost model outperformed the baseline accuracy of 51.17%, demonstrating its ability to capture meaningful patterns in the data.

### Comparison to Alternatives

- **Baseline Model**: Random and majority-class predictions achieved 51.17% accuracy.
- **Logistic Regression and Decision Trees**: These models performed worse than XGBoost due to their inability to capture nonlinear relationships and interactions.
- **XGBoost Significance**: Its 4% improvement over baseline is significant, as even small predictive advantages are valuable in financial modeling.

### Feature Representation and Analysis

**Effective Features:**

- **TF-IDF Features**: Several specific words extracted from the headlines using TF-IDF vectorization were among the most important predictors. For example:
  - `TFIDF_309`: This feature, representing a particular word, had the highest importance score, suggesting its strong association with stock trends.
  - `TFIDF_67`, `TFIDF_466`, and other TF-IDF features were also highly ranked, reflecting their predictive value in capturing nuanced patterns in the headlines.
- **Daily News Volume**: The number of headlines published each day proved to be a highly predictive feature, correlating strongly with stock volatility and trends.
- **Closing Price**: Historical closing prices provided essential context for understanding stock movements and significantly contributed to predictions.
- **Lagged Features**: Features such as `Close_Lag1`, `Volume_Lag1`, and `Sentiment` showed moderate importance, suggesting temporal patterns influence stock movements.

**Ineffective Features:**

- **Headline Sentiment**: Despite its intuitive appeal, sentiment analysis yielded a near-zero correlation with stock trends ($r \approx 0.00$), making it a weak predictor.
- **Most Common Words**: High-frequency terms like `"us"`, `"trade"`, and `"china"` lacked specificity and were not significantly associated with Tesla stock movements.
- **Volume Lag Features**: While recent volumes contributed to some extent, deeper lagged features (e.g., `Volume_Lag3`) had minimal impact on predictions.

**Insights:**

- TF-IDF features derived from specific words in the headlines were particularly effective, likely capturing subtle signals related to Tesla's stock trends.
- The ineffectiveness of sentiment analysis highlights the complex, non-linear nature of market dynamics, where simplistic correlations are insufficient.
- The findings suggest that combining robust numerical features (e.g., historical prices) with selective text-based insights leads to better performance.

**Reasons for Success and Failure**

**Why the Proposed Model Succeeded:**

- Effectively captured nonlinear interactions between features.
- Temporal and numerical features were well-engineered.
- Robust hyperparameter tuning reduced overfitting.

**Why Other Models Failed:**

- Simpler models (e.g., logistic regression, decision trees) could not capture complex patterns.
- Sentiment and text-based features lacked predictive power.

**Conclusions and Significance**

- **Market Efficiency**: Near-random correlation between sentiment and trends reflects the competitive nature of financial markets.
- **Significance of Results**: A 4% improvement over random predictions is significant and highlights the value of even small edges in trading.
- **Dataset Bottlenecks**: Expanding headline data is crucial to improve model performance.
- **Future Work**: Enhanced NLP features, incorporation of external factors, and improved datasets are essential for future improvements.

This work utilizes insights from ChatGPT [1] to discuss sentiment analysis in stock price prediction.

# References


[1] OpenAI. *ChatGPT (December 2024 version)*. OpenAI, 2024. Available at: `https://chat.openai.com/`.