Stock News Sentiment Analysis

By:

- Andreas I. Lukito - 2702211595
- Rayhan Permana - 2702215896

Project Objectives

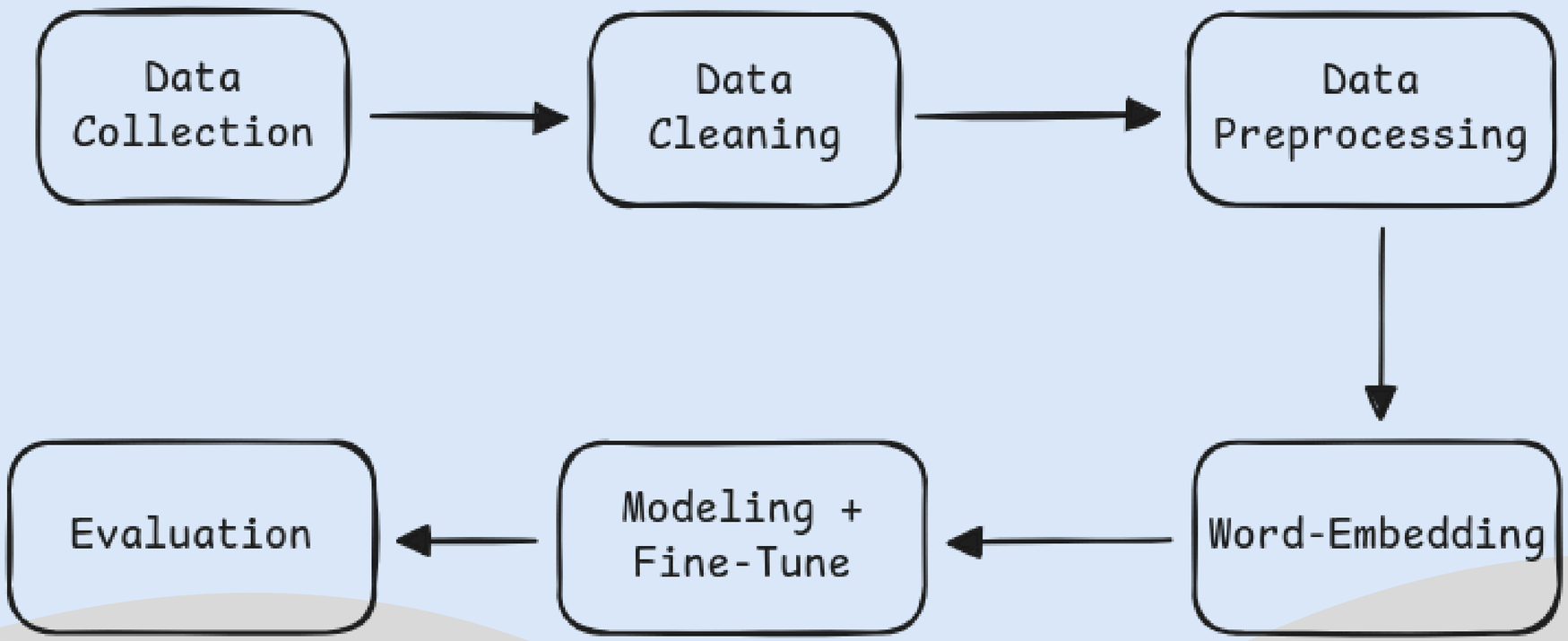
To develop a sentiment analysis system that automatically identifies and classifies the sentiment expressed in stock market-related news articles. The project aims to capture market sentiment that may influence investor behavior and stock price movements.

The results of this analysis support a better understanding of how news impacts financial markets and provide useful insights that can assist investors, analysts, and researchers in making more informed decisions based on current market sentiment.

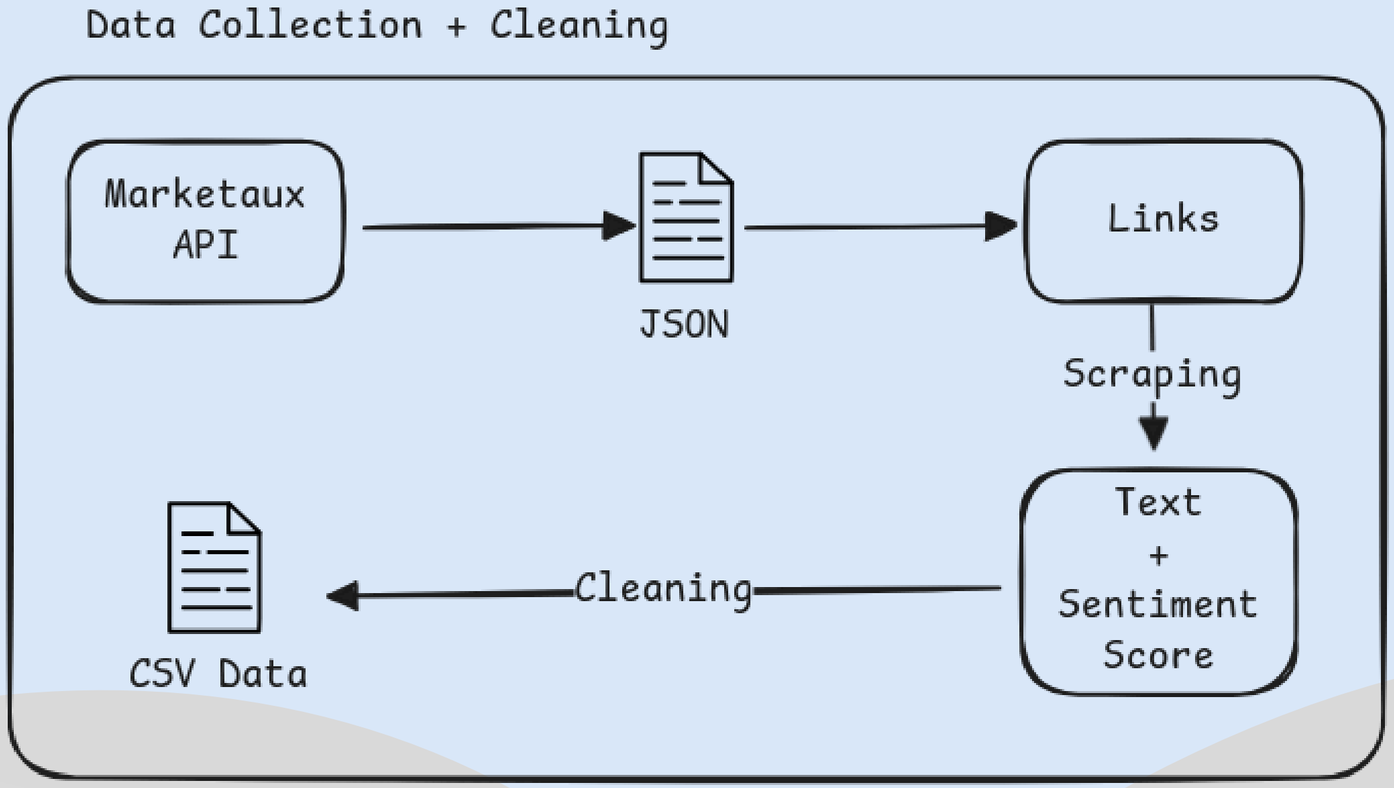
Project Objectives

- Custom Word Embedding
  - Word2Vec
  - FastText
- Model Comparison
  - LSTM Model
  - Pre-trained Bert Family (Albert-v2, FinBERT)

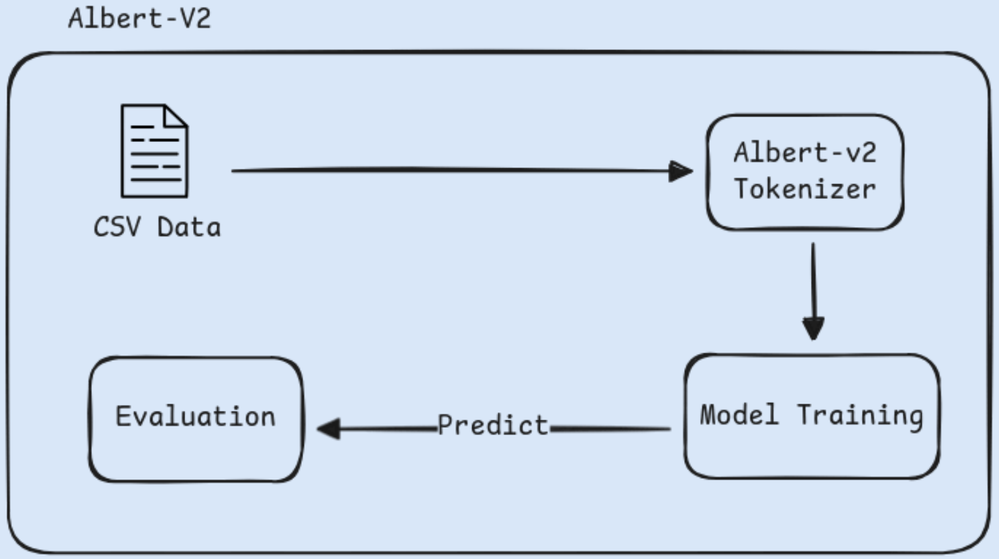
Our Process

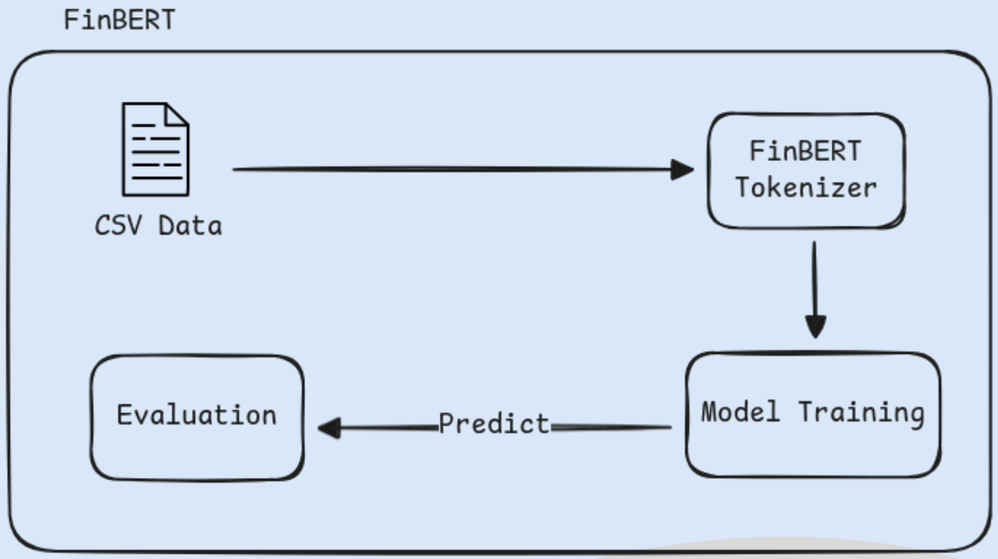

Data Collection
Data Cleaning
Data Preprocessing
Word-Embedding
Modeling + Fine-Tune
Evaluation

Our Process

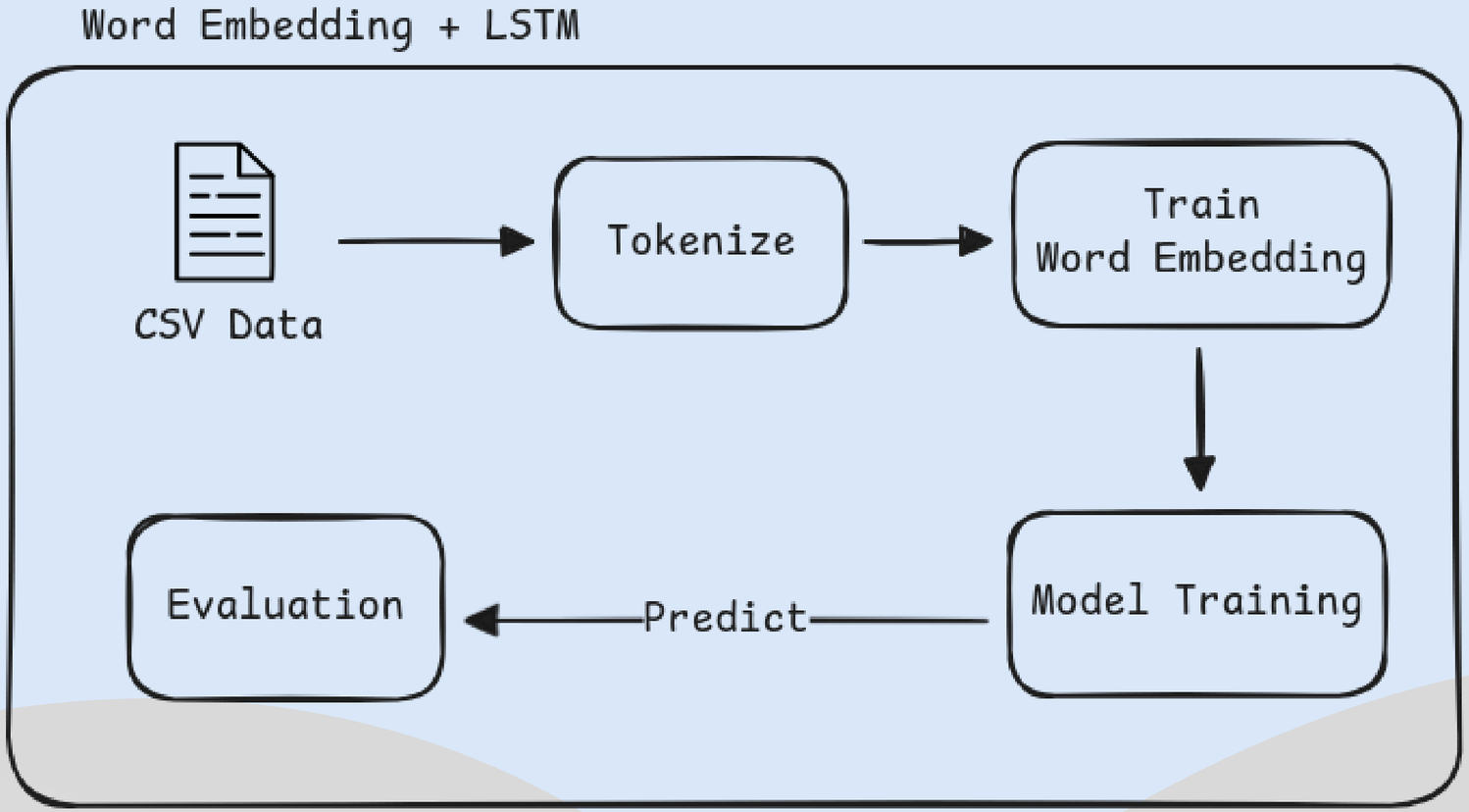

Data Collection + Cleaning
Marketaux API
JSON
Links
Scraping
Text + Sentiment Score
Cleaning
CSV Data

Our Process

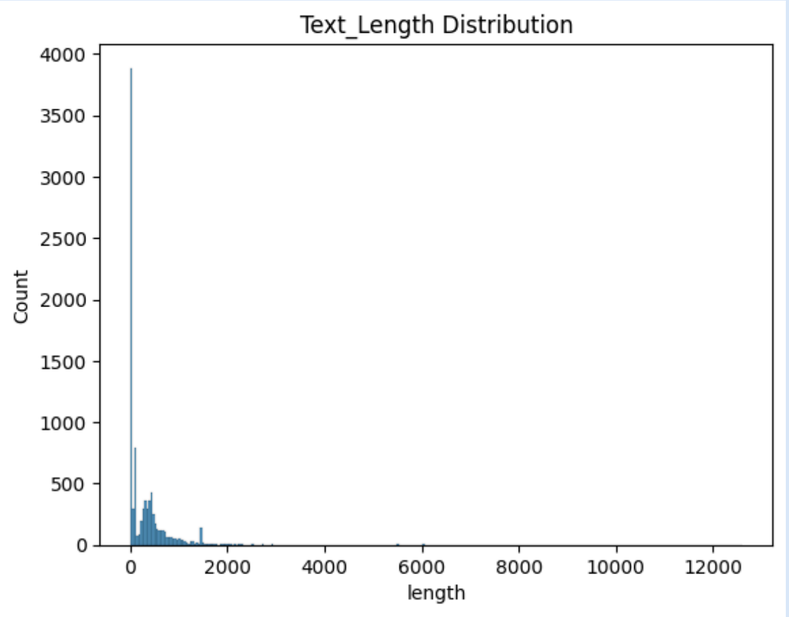

Albert-V2
CSV Data
Albert-v2
Tokenizer
Model Training
Predict
Evaluation

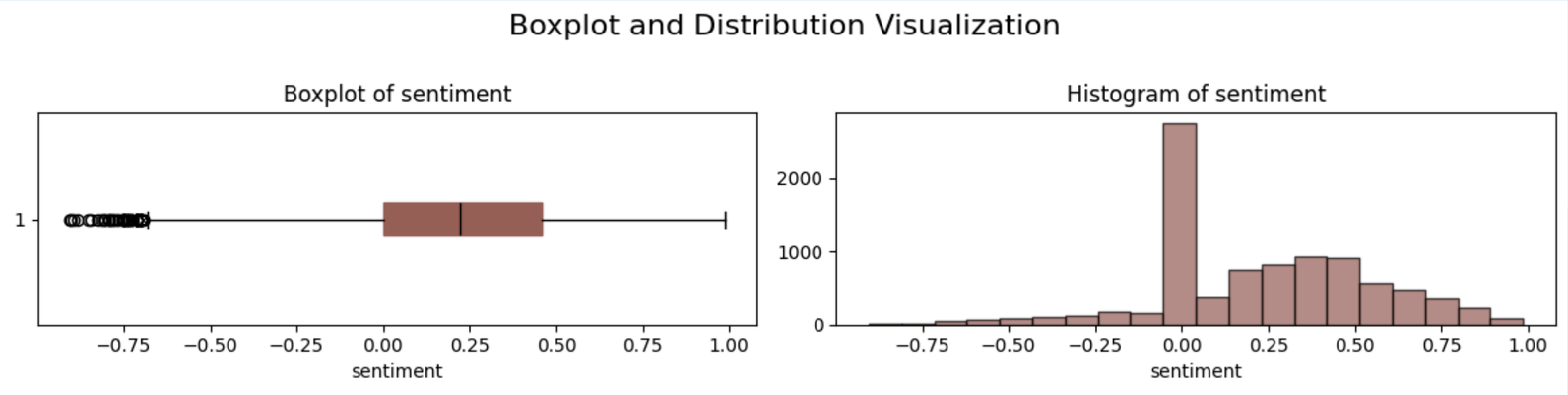

FinBERT
CSV Data
FinBERT
Tokenizer
Model Training
Predict
Evaluation

Our Process


Word Embedding + LSTM
CSV Data
Tokenize
Train
Word Embedding
Model Training
Predict
Evaluation

EDA


Text_Length Distribution
Count
length
0
500
1000
1500
2000
2500
3000
3500
4000
0
2000
4000
6000
8000
10000
12000


Boxplot and Distribution Visualization
Boxplot of sentiment
1
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
sentiment
Histogram of sentiment
0
1000
2000
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
sentiment

Results

Model	MSE	MAE	RMSE	r^2
FastText + LSTM	0.0895	0.245	0.2991	0.0042
Word2Vec + LSTM	0.0932	0.2539	0.3054	-0.0043
Albert-v2	0.0888	0.2182	0.298	0.0917
FinBert	0.0844	0.2066	0.2905	0.1093

- FinBERT demonstrates the **best overall** performance, achieving the lowest MSE, MAE, and RMSE, along with the highest (R^2) value.
- This suggests that FinBERT is **more effective at capturing sentiment** polarity and intensity in stock market news, likely due to its pretraining on financial-domain text, which allows it to better understand domain-specific terminology and contextual nuances.
- Despite this advantage, the model's **(R^2) value remains relatively low**, indicating that a large portion of sentiment variability is still unexplained.
- As a result, while FinBERT provides the most reliable sentiment estimates among the tested approaches, its predictions **should be interpreted with caution**.
- The model is best suited for extracting aggregate sentiment trends and supporting broader market analysis rather than delivering precise, article-level insights.

Next Steps

- Named Entity Recognition
  - Identify and rate the sentiment of each stock market ticker

- Combining with time series market data
  - Combines the sentiment result with the trend of closing prices

QnA

Thank You