Financial News Sentiment Analysis

Václav Stibor | doc. RNDr. Irena Holubová, Ph.D | bachelor thesis 2024

1. Introduction

In today's era of information explosion and constant flow, it becomes more time-consuming to keep track of associations and deeply understand the published content through online news, primarily when investing in a specific area. The thesis aims to design and implement a tool to provide users with sentiment analysis of news articles to aid investment decision-making.

2. Textual Data

First-party data providers are preferable to third-party providers because they offer higher **data quality** as direct sources.

Reliability | **Availability** | **Accessibility** | **Consistency** | **Terms of Service**

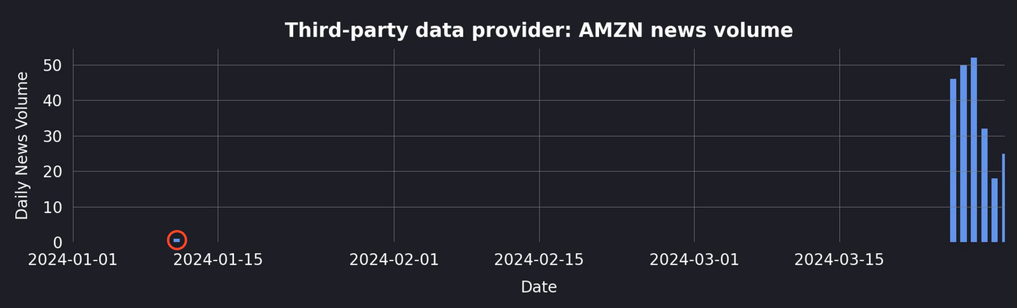

Figure 2.1. The Finnhub's daily volume of news articles mentioning Amazon.com, Inc. (AMZN) for the first quarter of 2024.

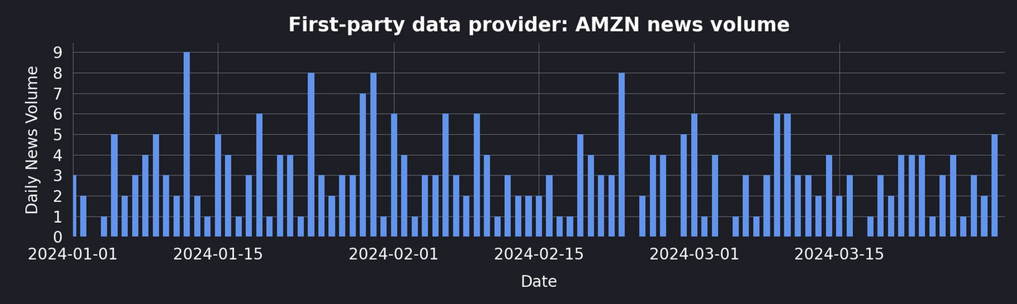

Figure 2.2. The Guardian's daily volume of news articles mentioning Amazon.com, Inc. (AMZN) for the first quarter of 2024.

5. Architecture

- **ETL service** is responsible for acquiring data from external sources, transforming it, and storing it in the database.

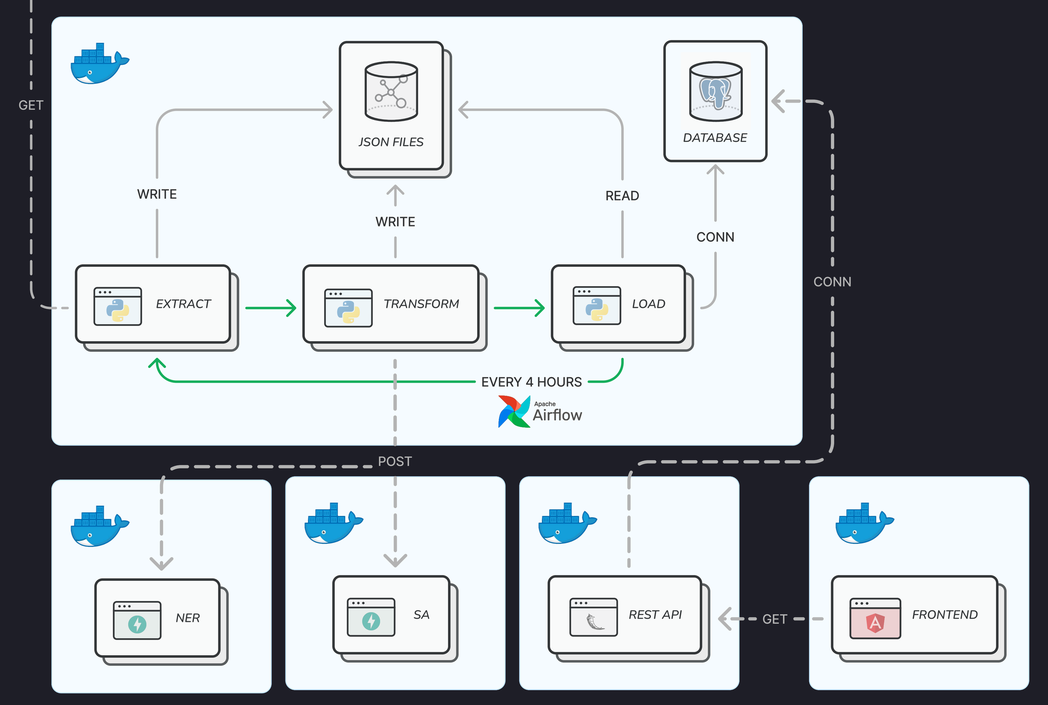

Figure 5.1. The application's high-level architecture containing individual services and their communication. The Named Entity Recognition (NER) and Sentiment Analysis (SA) services utilise FastAPI. The REST API service is implemented in Flask. The frontend is written in Angular, the database used is PostgreSQL, and the Extract Transform Load (ETL) process is implemented in Python and orchestrated using Airflow.

- **NER service** categorises company entities with their ticker in article content.
- **SA service** analyses text sentiment at the entity level.
- **REST API service** accesses data stored in the database.
- **FRONTEND service** provides a user interface for interacting with data from the database through the **REST API**.
  - The crucial frontend components include the **Home**, **Graphs**, **Companies**, **Company Graph**, and **Company Dashboard**.

3. Company to Symbol Linking

Amazon **ORG** → **Amazon.com, Inc.** | **AMZN** | **NASDAQ**

- **Discounted Levenshtein** | **Weighted Jaccard** | **Token Set Ratio**
  - ORG's name variations, exchanges datasets, and more **limitations**.
- **Trained knowledge base** on **Wikidata** offers a structured data repository.
  - **Spacy Entity Linker** library links entities to **Wikidata** entries.
  - Allowing to obtain **additional** entity **information** via **SPARQL queries**.
  - Achieving almost **100% success** depending on the current entity's relationship with the exchange.

4. Entity-level Sentiment Analysis

- We built the algorithm using the **FinABSA-Longer** trained model on the **SEntFiN 1.0** dataset.
- The model categorises sentiments into positive, neutral, and negative classes, providing numerical **entity-level sentiment analysis**.
- After significantly reducing the **FinEntity** dataset to ensure precise measurement, the algorithm's experiment achieved a **92% success** rate.

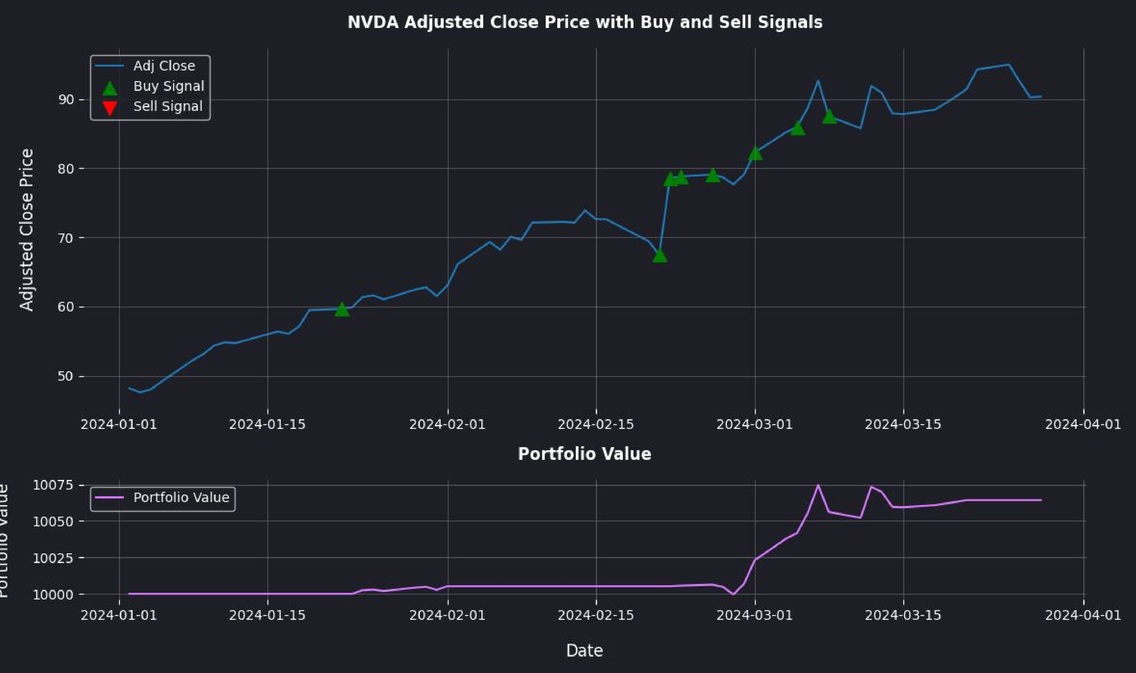

Figure 4.1. Nvidia Corp. (NVDA) adjusted close price with buy and sell signals for the first quarter of 2024 based on positive and negative sentiment. The portfolio evolution utilised a simple 7-day hold strategy based on signals.

6. Contributions

The tool enables the particular analysis and provides a **user-friendly** and **parameterisable visualisation** of the results and their correlation with the stock market using sentiment analysis on the **entity level**. Each crucial feature is verified using a selected set of historical data. The application architecture is **modular**, ensuring that the analytical approaches can be added or replaced and new input data types can be incorporated.

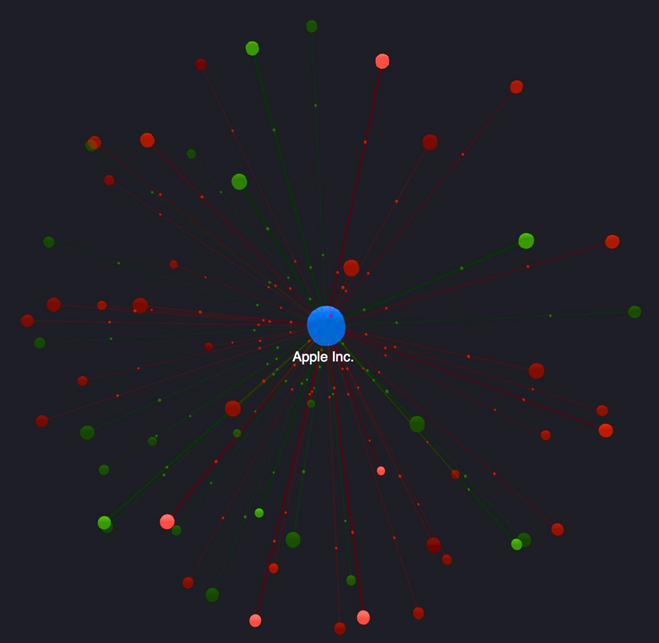

Figure 6.1. Apple Inc.'s company graph displaying all articles mentioning the company over the past three months.

According to Figure 6.1., an example of the Company Graph frontend component, each company is represented within an **individual graph**, with nodes corresponding to **articles** and **company**. The colour of each article node and its connecting edge **reflects** the company's **sentiment** in the article, detailed in the control panel. The company is coloured blue.

7. Acknowledgement

Special thanks to my supervisor **doc. RNDr. Irena Holubová, Ph.D.** for her expertise and guidance during the development of this work. We are also grateful to **the Guardian** for providing us with the data.

8. Contact

The Airflow webserver and the GlobeSense web application can be accessed by contacting **vasa20017@seznam.cz**.

Docker Hub