Author: Federico Durante

Sentiment Analysis of Financial Times Articles (2007-2009)

- **What does the code do?**

This code uses the *NLTK* sentiment analyzer to perform sentiment analysis on Financial Times news articles from 2007 to 2009, based on a partly preprocessed version of the Turenne et al. (2021) dataset, where files were merged, Chinese translations removed, and dates corrected and sorted. It processes the dataset by merging article titles and texts, calculates sentiment scores, and applies a moving average to smooth trends over time. The results are visualized in a time-series plot, highlighting key economic events from 2007 to 2009. The plot is then saved in the working directory. The code follows an object-oriented approach, organizing data loading, processing, analysis, and visualization within a reusable class.

- **How to use the code**

To run the code, ensure you have *Python* installed with the required libraries (*pandas*, *matplotlib*, *nltk*). Place the dataset file (*ft-articles.pkl.tar.gz*) in the same directory as the script. Execute the script, and it will automatically load and preprocess the data, compute sentiment scores, and apply a moving average. The output includes a time-series plot (Figure 1) displaying sentiment trends over time. The *window* parameter can be adjusted in the *main entry point* to modify the smoothing level of the sentiment trend. It defines the size of the moving average window used to smooth sentiment scores over time, controlling how many data points are averaged together to calculate each point in the trend line. In this code, the window is set to 500, meaning the sentiment score for each point on the graph is calculated as the average sentiment of the previous 500 articles. I chose the moving average window to be 500 because it smooths short-term noise while keeping major sentiment trends visible, ensuring that key financial crisis events remain clear.

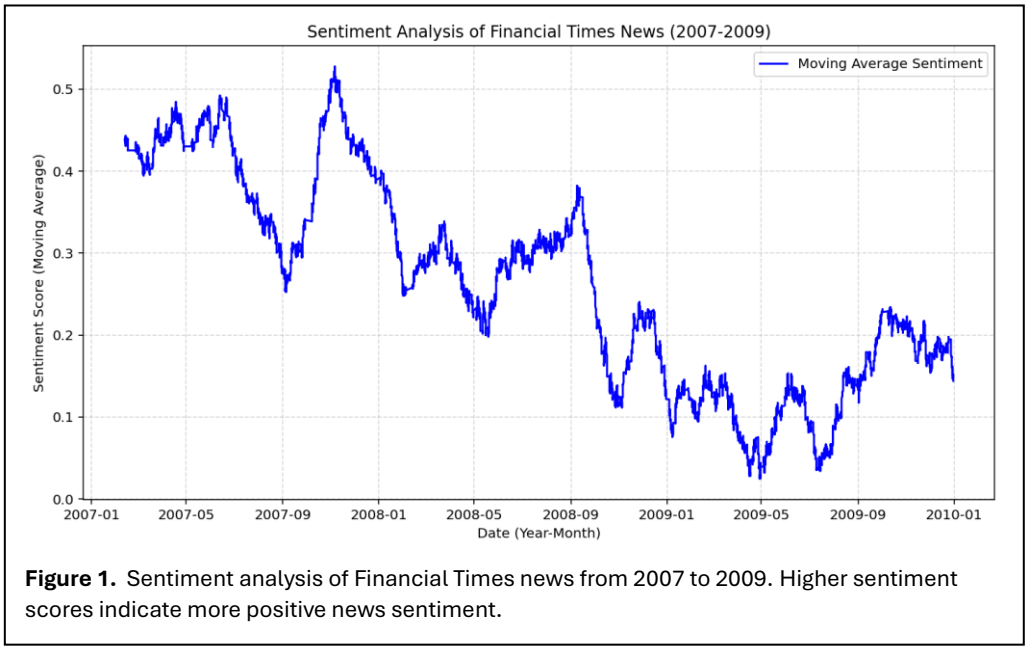

Figure 1. Sentiment analysis of Financial Times news from 2007 to 2009. Higher sentiment scores indicate more positive news sentiment.

- **Data processing**

The dataset consists of Financial Times news articles from 2007 to 2020, including their titles, text, and publication dates. However, the analysis focuses only on the period from 2007 to 2009. The *Title* and *Text* columns will be used to compute sentiment scores, while the *Date* column will track sentiment trends over time. The data processing will begin by loading the dataset and handling missing values by removing rows where essential fields are empty. To ensure meaningful sentiment analysis, the Title and Text columns will be merged into a single column. The *Date* column will be converted to a proper datetime format, and the dataset will be filtered to include only articles published between 2007 and 2009 before being sorted chronologically. Sentiment scores will be calculated using the NLTK sentiment analyzer, which assigns an aggregate sentiment score to each article. To smooth short-term fluctuations, a moving average will be applied to the sentiment scores.

- **Interpretation of the results**

The sentiment analysis shows big changes that match important events of the global financial crisis.

In mid-2007, sentiment starts to drop sharply around June, falling from about 0.45 to 0.25 by September. This matches the early crisis events, like BNP Paribas freezing some investment funds in August due to subprime mortgage problems and the bank run on Northern Rock in September.

In late 2007, sentiment briefly improves to about 0.50 by November but then starts falling again. This short recovery might be due to optimism about central bank actions.

In mid-2008, sentiment drops sharply to around 0.20 in March-April, aligning with Bear Stearns' collapse and rescue in March. The biggest drop happens in September-October 2008, when sentiment falls from 0.35 to below 0.15. This matches the worst part of the crisis, including Lehman Brothers' bankruptcy, AIG's bailout, and the global market crash.

In mid-2009, sentiment hits its lowest point (~0.04) around May, when the stock market was at its lowest and pessimism was at its peak.

From mid to late 2009, sentiment slowly recovers as stabilization efforts start working and fears of a total financial collapse decrease.

These results show that news sentiment changed in line with the financial crisis. This confirms that sentiment analysis can be useful for tracking economic uncertainty and how people feel about financial markets.