Financial Technology Final Project

Scraping

Obtain links — yahoo! finance — Stock Market News

Drop Duplicates — ❗ Webpage: limited news & every 5-hour update 100 news

Scrap Info — Contents: Abbreviations, Time, Title, News

NLP

Keywords — Set Parameters ⇨ max_ngram_size = 3
deduplication_threshold = 0.9
numOfKeywords = 20

Attitudes — Catch pos & neg

Classification

Finance Corpus { Transcoding
Tokenization ⇨ y: label
&
x: text
Sequencing

⇩

Word2Vec ➕ CNN ⇨ { Train model
Evaluate model ⇨ Labeling

Flask
AWS

HTML — Stock News Analysis — Query from Abbreviations

Connection — http://18.116.165.240/

Part	Knowledge	Main Packages
Scraping	▪ Webscraping ▪ Anti-scraping	• BeautifulSoup (bs4) • webdriver
NLP	▪ Sentiment Analysis ▪ Word2Vec ▪ Neural Network (CNN)	• Yake • Keras • SentimentIntensityAnalyzer • Tokenizer • sklearn
Flask	▪ HTML ▪ Flask	• Flask
AWS	▪ AWS EC2	