

# Information Retrieval Project

## Group 8

Group members:

Jared Nathaniel Chan Weng Wai (U2022954B)

Liau Zheng Wei (U2222032K)

Toh Yong Li (U2021290D)

Beh Ming Jun (U2022747E)

Ooi Yi Hang (U2022327K)

Wentai Nan (U2022249B)

# Group Member Roles

Name	Contribution
Jared Nathaniel Chan Weng Wai	Classification and Report write up
Liau Zheng Wei	Classification and Indexing
Toh Yong Li	Web Crawling and UI development
Beh Ming Jun	Classification and Report write up
Ooi Yi Hang	Classification and Video
Wentai Nan	Classification and Slides

# Tech Stocks Query System



Large amounts of data on stock news

Difficult for investors to obtain relevant and up to date information

Aims to provide a sentiment analysis on top tech stocks

# Crawling

Web scraping methods

Crawled data





# Crawling Methods



# Crawled Data

msn.com/en-us/money/topstocks/apple-stock-has-20-upside-according-to-1-wall-street-analyst/ar-BB1lw80D

 The Motley Fool [+ Follow](#) 258.7K Followers 

## Apple Stock Has 20% Upside, According to 1 Wall Street Analyst

20/04/23

Growth concerns keep dogging **Apple** (NASDAQ: AAPL). The Cupertino colossus [abandoned plans](#) to build its electric car, iPhone sales in China are drying up, and Apple even seems to have missed the boat on [artificial intelligence](#) -- to the extent it's had to [ask Google for help](#)). Share prices of Apple stock are down about \$20 from recent highs hit in January.

- Headline
- Source
- Posted Date
- Description
- Link

# Crawled Data



Number of Records	Number of Words
10,916	484,908

# Classification

Data preprocessing

Model training

Enhancements

Results



Positive



Negative



Neutral



# Data Cleaning

## Raw

In Finland 's Hobby Hall 's sales decreased by 10 % , and international sales fell by 19 % .

Rinkuskiai 's beer sales fell by 6.5 per cent to 4.16 million litres , while Kauno Alus ' beer sales jumped by 6.9 per cent to 2.48 million litres .

Sales in Finland decreased by 10.5 % in January , while sales outside Finland dropped by 17 % .

- tokenisation
- remove stopwords  
punctuations and  
special characters
- lower case
- stemming and  
lemmatisation

## Cleaned

finland hobby hall sale decrease 10  
international sale fall 19

rinkuskiai beer sale fall 6.5 per cent  
4.16 million litre kauno alus beer sale  
jump 6.9 per cent 2.48 million litre

sale finland decrease 10.5 january  
sale outside finland drop 17

# Data Labelling

Headline	Annotator 1	Annotator 2
Will ASML Be Worth More Than Apple by 2030?	NEUTRAL	NEUTRAL
Apple Inc. (NASDAQ:AAPL) Stake Lessened by Parkside Investments LLC	NEGATIVE	NEGATIVE
Apple (NASDAQ:AAPL) Trading Down 1.4%	NEGATIVE	NEGATIVE
Apple (NASDAQ:AAPL) Given Hold Rating at DZ Bank	POSITIVE	POSITIVE
Do the Magnificent Seven Still Rule the Market?	NEUTRAL	NEGATIVE

**Fleiss' Kappa score : 96.6%**

# Model Training

## Reference Model

SOTA Model

TF-IDF with SVM

## Feature Extraction

TF-IDF

N-Grams

BERT Word Embeddings

GloVe

## Classification Models

Naive Bayes

SVM

Logistic Regression

Random Forest

KNN

SGD

Cognitive Models

- Liquid Time Constant
- Spiking CNN

# Classification Results

## Subjectivity

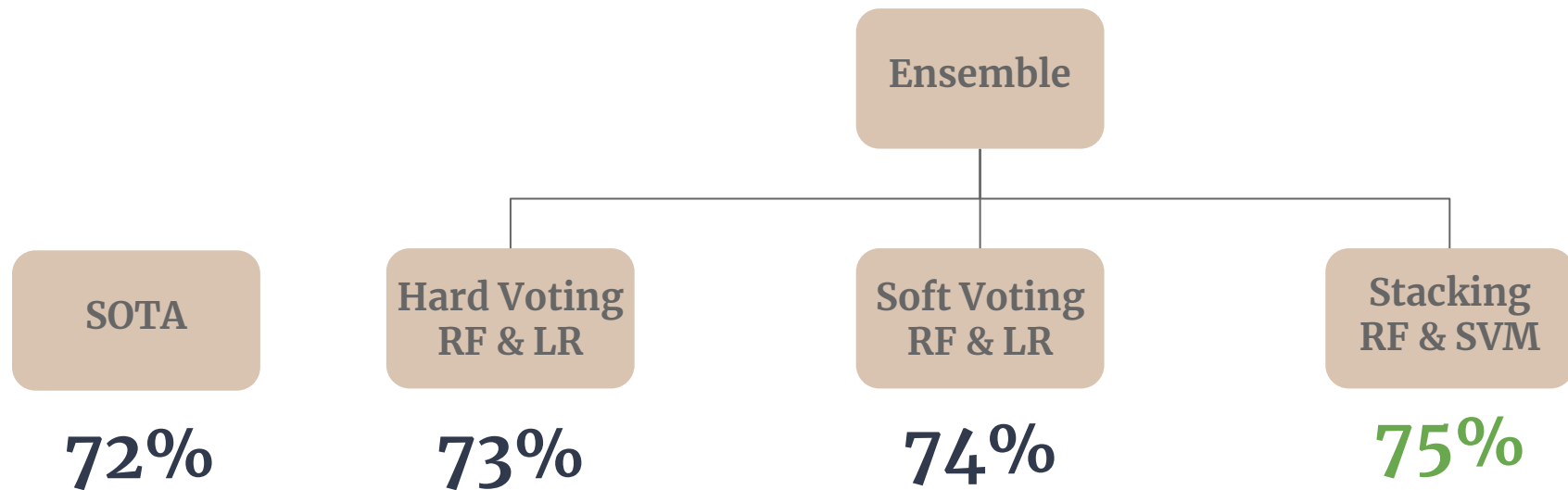
Model		Accuracy	F1	Precision	Recall
Count Vectoriser Unigrams	Naive Bayes	0.55	0.53	0.55	0.56
TF-IDF Unigrams	SVM	0.56	0.54	0.58	0.59
TF-IDF Unigrams + BERT word embeddings	Logistic Regression	0.58	0.57	0.60	0.62

# Classification Results

## Polarity

Model		Accuracy	F1	Precision	Recall
Count Vectoriser Unigrams	Naive Bayes	0.70	0.70	0.71	0.70
TF-IDF Unigrams	SVM	0.73	0.72	0.73	0.72
TF-IDF Unigrams + BERT word embeddings	Logistic Regression	0.79	0.78	0.79	0.79
TF-IDF Unigrams, Bigrams and Trigrams + BERT word embeddings	Logistic Regression	0.78	0.78	0.78	0.78

# Model Enhancements



# Final Model

Model		Accuracy	F1	Precision	Recall
TF-IDF Unigrams + BERT word embeddings	Logistic Regression	0.79	0.78	0.79	0.79
with Ensemble Voting	Random Forest	0.58	0.50	0.72	0.58
with Ensemble Stacking	Random Forest and Logistic Regression	0.78	0.78	0.78	0.78

## Random Accuracy

- random chance to assign sentiment
- model performs better

## Scalability

- matrix operations for linear computations
- parallelized and optimised for larger data

# Indexing

Data indexing

Query





# Data Indexing



- CSV row read as dictionary
- map dict keys to Solr schema field
- Solr documents added to index

# Query

## Query Field

- **Data Field**: Headline, Source, Posted, Description, Link



## User Preference

- 1) Choose a **Data Field**
- 2) (Optional) Filter **Start** and **End** date
- 3) (Optional) **Sort date** by asc or desc
- 4) Enter **query**
  - Eg. Headline, "apple", 2024-02-16 to 2024-03-31, asc



## Query Input

- a) **Query from pre-specified start and end (or max) index**

- **String:**
  - 1
  - 2
  - 3
  - 4

```
'Headline:"apple" AND Posted:["2024-02-16T00:00Z" TO "2024-03-31T23:59:59Z"]'
```
- **Parameter:** `**{'start': 0, 'rows': 20000, 'sort':'Posted' + asc}`
  - 3
  - 4

# Demo



# Overview



Deploy



## Query Finder

Display

Search



Search bar

Search

Start date

2024/02/16

End date

2024/03/31

Sort date

asc




## Stocks News Query

# ‘About Database’ Display

×

Query Finder

Display

About database 

Search bar

Search


Start date


2024/02/16

End date

2024/03/31

Sort date

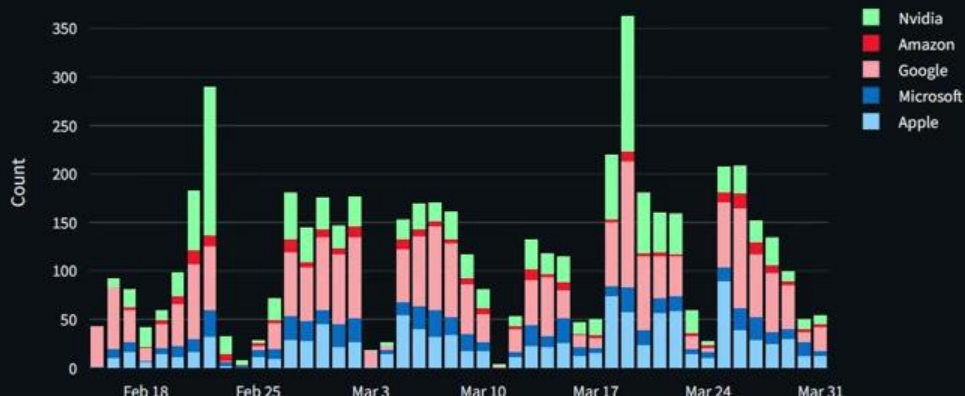
asc 

Deploy 

## Stocks News Query

### About database

Daily Document Count by Term



# Multilingual Search

×

Query Finder

Display

Search

Search bar

苹果

Search

Start date

2024/02/20

End date

2024/02/21

Sort date

asc

Deploy

## Stocks News Query

# Query Suggestion

×

### Query Finder

Display

Search

Search bar

inter

Press Enter to apply

Search

Start date

2024/02/16

End date

2024/03/31

Sort date

asc

## Stocks News Query

Deploy

# Sentimental Count Display

×

Query Finder

Display

Search

Search bar

apple

Search

Start date

2024/02/16

End date

2024/03/31

Sort date

asc

Positive Count: 210

## reports of new AI tool

Dated: 2024-02-16

Sentiment: POSITIVE

On Friday, Citi reaffirmed its Buy rating on Apple Inc. (NASDAQ ) with a steady price target of... [\(Read more...\)](#)

Source: Investing.com

Link: <https://in.investing.com/news/citi-maintains-buy-rating-for-apple-stock-at-225-amid-reports-of-new-ai-tool-93CH-4023059>

----- Result 4 -----

## ...Associates' Q4 Strategic Moves: Sells Netflix, Increases Nvidia Stake by 458% - Apple (NASDAQ:...

Dated: 2024-02-16

Sentiment: POSITIVE

Over the past few years, 74-year-old Dalio gradually scaled back his formal role at the firm.... [\(Read more...\)](#)

Source: Benzinga

Deploy





# Further Query Suggestions (Synonym and Antonym)

×

Query Finder

Display

Search

Search bar

happy

Press Enter to apply

Search

Start date

2024/02/16

End date

2024/03/31

Sort date

asc

Deploy



## Stocks News Query

~ The End ~

