WALLSTREETBETS VS WALL STREET - STOCK MARKET PREDICTION

#Team 11 #Chiaoya Chang #Tony Chu

#1 OBJECTIVES

#2 IMPACTS

#3 PREVIOUS WORKS

#4 DATA & APPROACHES

#5 OUR SCHEDULE

OBJECTIVES

#01

#OBJECTIVES

RECENT EVENT:
GAMESTOP (GME)

GameStop shares have soared 1,700% as millions of small investors last month.

EVALUATE THE SOCIAL
MEDIA IMPACTS

Social media can affect our daily lives. We want to observe how it can change the real-world events.

Use posts on reddit to forecast the stock market trends

IMPACTS

#02

#IMPACTS

FIND KEYWORDS ON REDDIT REGARDING THE STOCK MARKET

FIND KEY INFLUENCERS ON REDDIT WHO CAN INFORM USERS WITH THE RIGHT CONTENTS

HELP INVESTORS MAKE DECISIONS

NEW IMPACTS ON THE STOCK MARKET

PREVIOUS WORKS

#03

#PREVIOUS WORKS

- *Twitter mood predicts the stock market (2011)*

  time series analysis, causality analysis of mood vs. Dow Jones Industrial Average prices

- *Topic Modeling based Sentiment Analysis on Social Media for Stock Market Prediction (2015)*

  Topic-sentiment: TSLDA (Topic Sentiment Latent Dirichlet Allocation)

- *Artificial intelligence prediction of stock prices using social media (2021)*

  Twitter, VADER vs. word embedding (GloVe), LSTM

⇒ Previous Work: focus on **Twitter** and **sentiment analysis**
⇒ Our Work: focus on **Reddit** and **other features** (e.g., n-grams/vector embeddings)

DATA & APPROACH
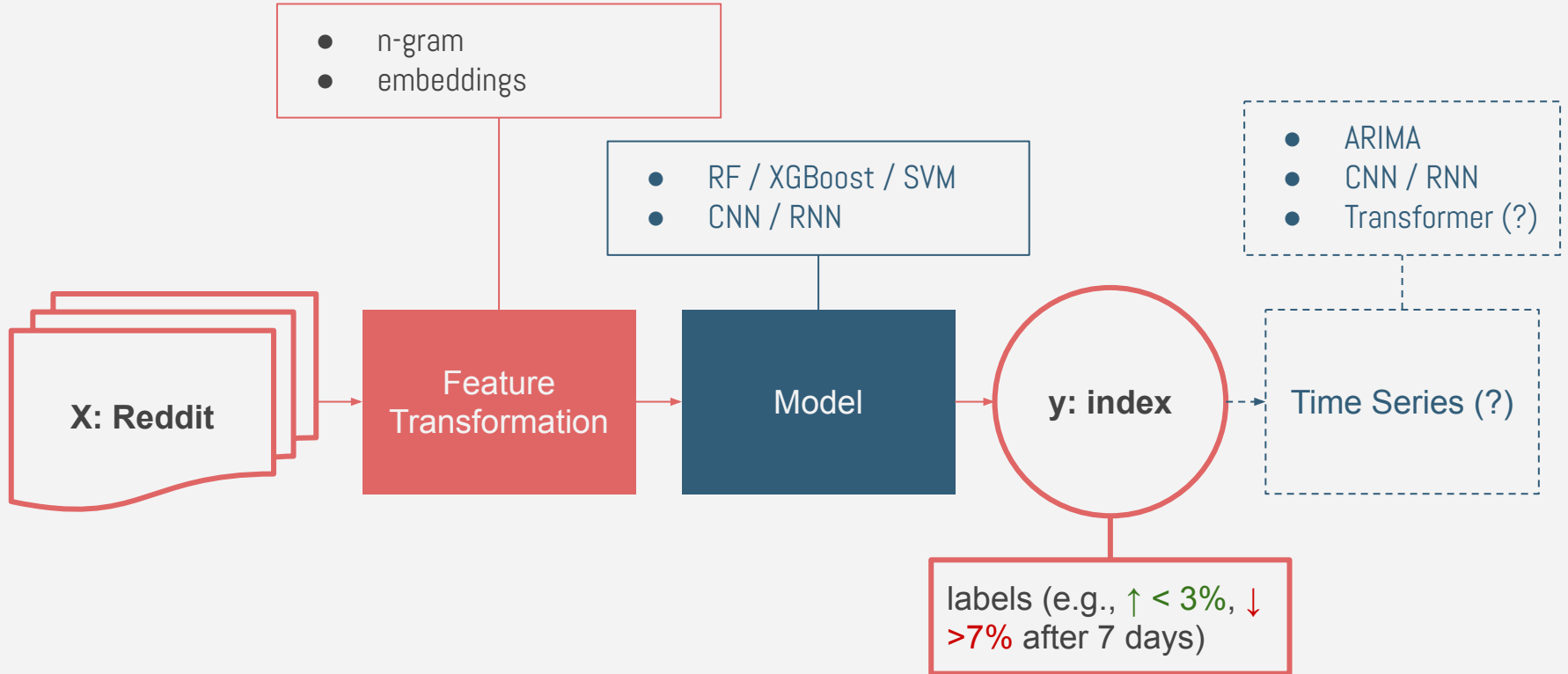
#04

#DATA SOURCES

#REDDIT

r/WallStreetBets on Reddit

#NASDAQ

NASDAQ-100 Technology Sector Index (NDXT)

#APPROACH

- n-gram
- embeddings

- RF / XGBoost / SVM
- CNN / RNN

- ARIMA
- CNN / RNN
- Transformer (?)

X: Reddit

Feature Transformation

Model

y: index
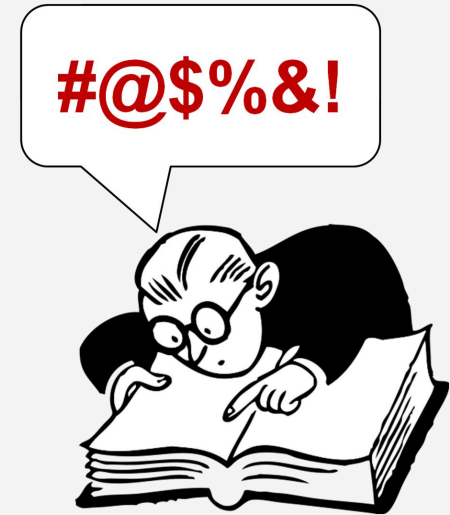
Time Series (?)

labels (e.g., ↑ < 3%, ↓ >7% after 7 days)

11

#POSSIBLE NLP ISSUES

- Abbreviations:

  - Finance: Return On Investment → ROI, Earnings Report → ER

  - Companies: Amazon → AMZN, Microsoft → MSFT

  - Special terms: bull, bear

- Trolling vs Serious posts

- Non-textual information (e.g, charts, tables, urls)

#@$%&!

OUR SCHEDULE

#05

#PLAN

#WEEK 6

DATA COLLECTION

- Collect reddit data using API
- Retrieve S&P 500 Daily Returns

#WEEK 7

DATA PROFILING

- Know data more
- Explore the data
- Assess data quality

#WEEK 8

FEATURE EXTRACTION

- Extract features we need
- Use pretrained word embeddings

#WEEK 9

DATA MODELING

- Implement the algorithms
- Tune our model
- Evaluate the model

#WEEK 10

VISUAL REPORT

- Provide insights
- Create the final report