Public Opinion Analysis of Airlines

Team 1:

Fan Wu

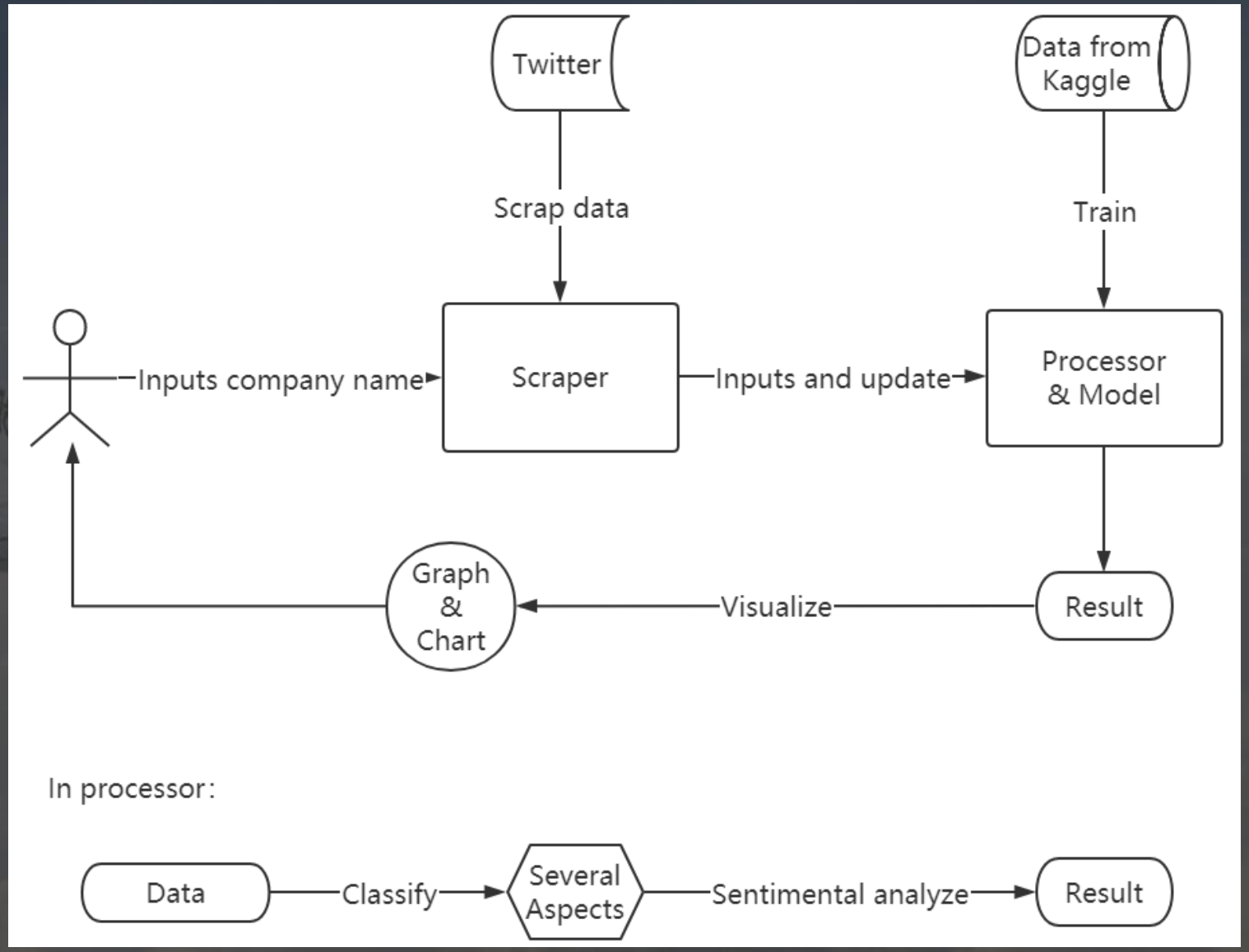
Dayu Jia

Bowen Jiang

Use Cases

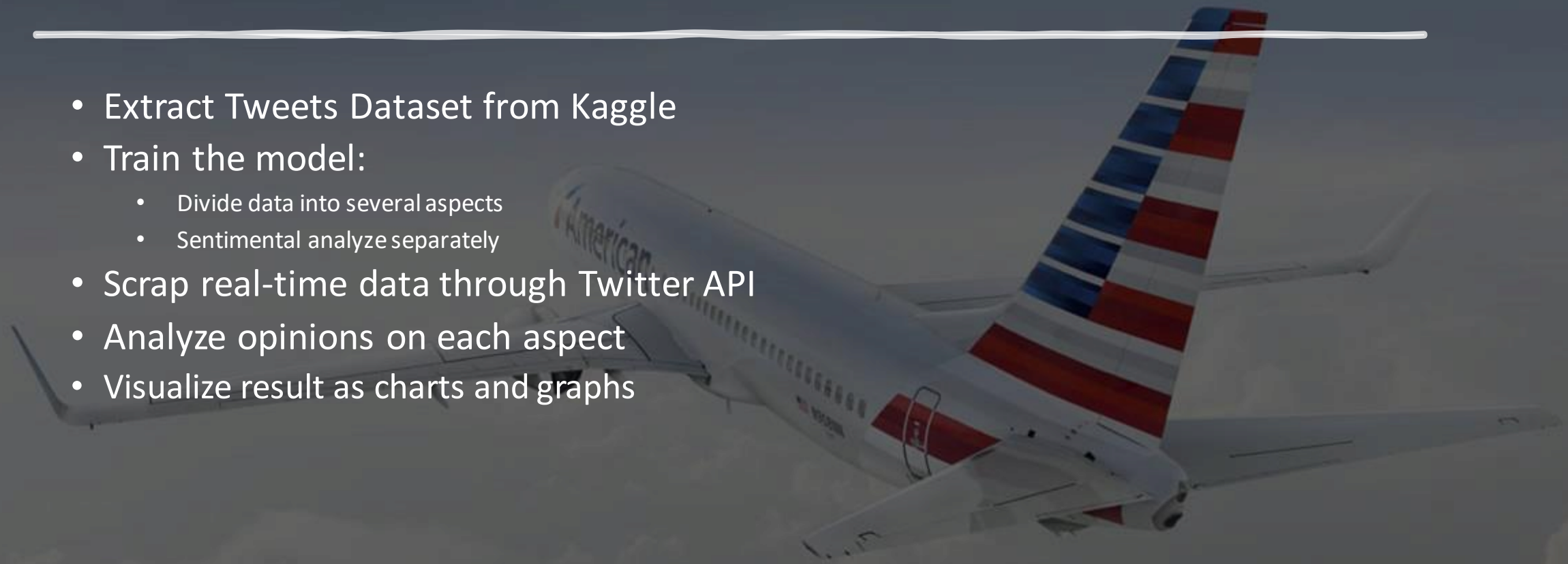
User inputs the name of an airplane company

and receives a comprehensive analysis on public opinion



Methodology

- Extract Tweets Dataset from Kaggle
- Train the model:
  - Divide data into several aspects
  - Sentimental analyze separately
- Scrap real-time data through Twitter API
- Analyze opinions on each aspect
- Visualize result as charts and graphs



Data Sources

Twitter US Airline Sentiment

This dataset has 14485 rows and 30 columns

Source from:

https://www.kaggle.com/crowdflower/twitter-airline-sentiment

Real-time data through Twitter API

Milestones

1st week:

Learn spark, get real-time data from Twitter

2nd week:

Clean data from kaggle and train the model

3rd week:

Analyze Twitter data with model, update model

4th week:

Visualize the result and improve the project

Programming in Scala

- Data extraction
- Data preprocessing
- Model implement
- Unit tests
- Visualization and UI design
- Repo: https://github.com/FanWu6/Spark-stream-twitter-analysis.git

Acceptance criteria

- 70% of data can be classified correctly
- The accuracy of sentimental analysis should reach 80%

Goals

- For any airline company, we can offer a real–time analysis on public opinion based on our model.

- For us, we want to learn the using of Scala and machine learning, and how to co-work on Github.