Drug Reviews: What do they mean?
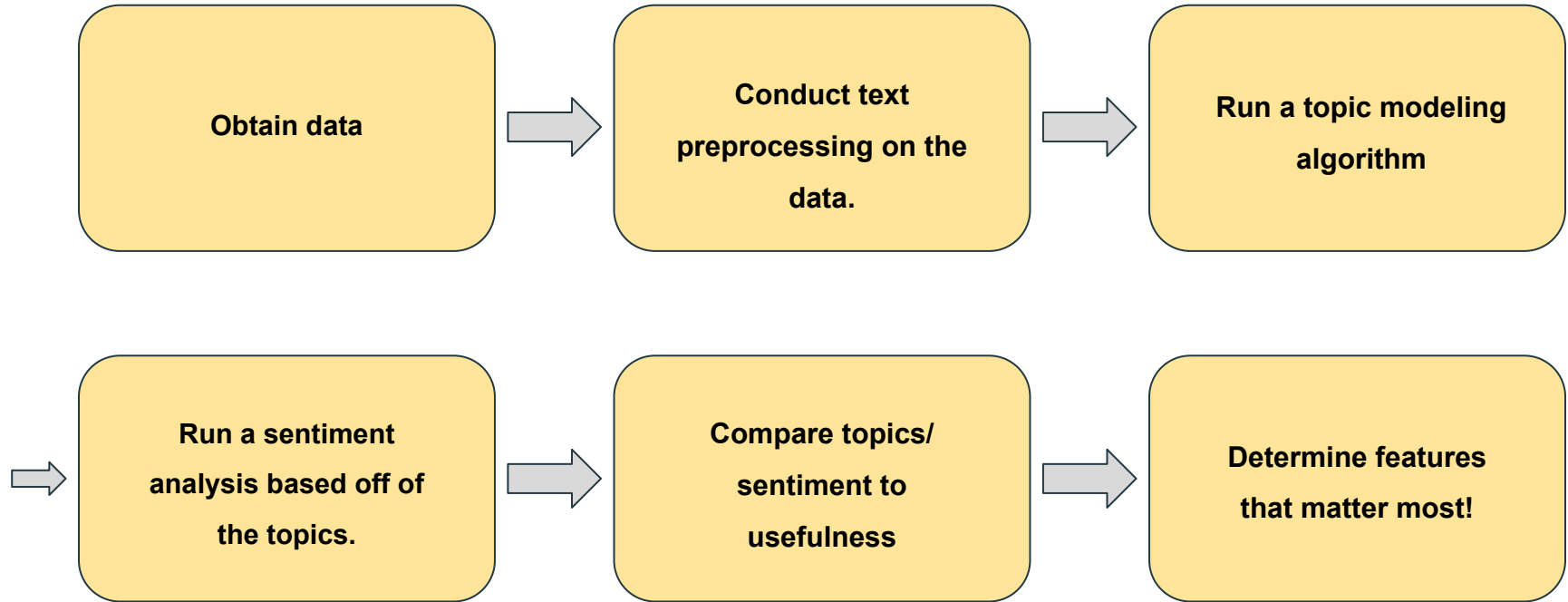
Che-Yu Liu

Introduction

- In a Harvard News Article from 2019, one of the largest concerns from new medicines were:
  - Unknown side effects
- Problem:
  - Not everyone reports when they do have a side effect.
- Goal:
  - Create a system in which people can easily see what side effects may be present.

Methodology

Obtain data → Conduct text preprocessing on the data. → Run a topic modeling algorithm → Run a sentiment analysis based off of the topics. → Compare topics/ sentiment to usefulness → Determine features that matter most!

Data

- Data was obtained from Kaggle dataset that uses drug reviews obtained from UC Irvine researchers.
    - Review
    - Usefulness
    - 10 star rating system
    - Drug name
    - Condition
- ~ 160000 responses

Topic Modeling

- Topics were obtained using various topic modeling algorithms:
    - CorEx
    - NMF
    - LSA

Most popular topics

- Reviews were taken between 2008 to 2017.
  - Reviews from 2017 had little to no people who found it useful.
- Sentiment was attained by ratings:
  - Ratings that were above a 5 were labeled as positive.
  - Below 5 were labeled as negative

Insights

- Reviews were taken between 2008 to 2017.
  - Reviews from 2017 had little to no people who found it useful.
- Sentiment was attained by ratings:
  - Ratings that were above a 5 were labeled as positive.
  - Below 5 were labeled as negative

Conclusion

- Reviews were taken between 2008 to 2017.
  - Reviews from 2017 had little to no people who found it useful.
- Sentiment was attained by ratings:
  - Ratings that were above a 5 were labeled as positive.
  - Below 5 were labeled as negative