# Drug Effectiveness

Cole Thorpen & Payton Burks

# Introduction

## Motivations

- Prescription pills for various ailments are on the rise
  - Adderall, Prozac, Opioids, etc.
- How satisfied are people with their medication?
- Can we predict effectiveness based on satisfaction? Other factors?

# Introduction

## Dataset

- 8 attributes
  - *drug* – name of drug
  - *condition* – condition drug is intended to treat
  - *date* – date of review and rating entry
  - *age* – age range of patient
  - *ease of use* – 5-star review of how easy to use
  - *satisfaction* – 5-star review of how well patient liked
  - *sex* – gender
  - *effectiveness* – 5-star patient review of drug (class attribute)

*Both 'useful count' and 'review' attributes were omitted; they will serve no use in the predictions

# Introduction

Dataset (cont.)

- Initially 7,015 instances
- Cleaning
  - Remove repeated instances
  - Drop non-useful attributes
  - *date* adjusted to *season*
- Aggregation
  - group by each individual drug (*name*)
  - 5-star reviews averaged out
  - *gender*, *age*, *condition*, *season* assigned to majority case
- Class attribute (*effectiveness*)
  - Discretized by Very Effective (VE), Effective (E), Moderately Effective (ME), Slightly Effective (SE), Not Effective (NE) - separated by star-rating by 1
- Clean dataset
  - 1017 instances

# Introduction

## Challenges

- Repeated instances (removed if *exactly* same)
- Disparity in ages
- Massive domain for *condition* attribute
- Aggregating instances and normalizing attributes

# EDA & Classification

Jupyter Notebook

# Flask Demo

- Random Forest Classifier
- https://drug-effectiveness.herokuapp.com/
- Example Predictions:
  - https://drug-effectiveness.herokuapp.com/predict/Drug=benzonatate&Age=0-2&Condition=Cough&Season=summer&EaseofUse=E&Satisfaction=E&Sex=Female
  - https://drug-effectiveness.herokuapp.com/predict?Drug=coumadin&Age=25-34&Condition=Blood%20Clot&Season=fall&EaseofUse=NE&Satisfaction=NE&Sex=Female

# Conclusion

- Some ideas for improvement:
  - our predictions rely heavily on EaseofUse and Satisfaction ratings, so we could remove these attributes and try to predict again.
- Partner Breakdown
  - Payton:
    - Forest tests, Random Forest, EDA (stats and charts), data preprocessing
  - Cole:
    - Random Forest, EDA (fit and predict), Heroku Deployment
- Dataset Source
  - https://www.kaggle.com/datasets/jessicali9530/kuc-hackathon-winter-2018