



THE UNIVERSITY  
*of* NORTH CAROLINA  
at CHAPEL HILL

# Term Project

June 11, 2018

HiDAV Boot Camp

# Term Project

- What you need to do
  - Form a group with **2** or 3 people
  - Assign roles among members
  - Select a health/biomedical-related topic
  - Select a particular task that interests your group members (e.g., classifying sub-topics of BRCA)
  - Use a dataset from PubMed or ClinicalTrials.gov from the provided software
  - Work on one or more hypotheses (or research questions)
  - Use the provided software on the server/computer
  - Do error analysis (that will be covered soon) and report your findings

# Term Project

- Project Deliverable
  - Poster
    - Research poster
    - Samples will be shared
    - We will demo the Adobe Illustrator
  - Presentation (June 29<sup>th</sup>)
    - About 30 minutes per group
    - Introduction, overview, description of the approaches tested, methods, results, discussion, and conclusion

# Details of Presentation

- Introduction
  - What is the problem?
- Overview
  - Current state of literature in the field
    - How is text mining used in this particular topic?
- Approaches tested
  - Ex: queries generated, information filtering, feature selection & evaluation
- Methods
  - With description of classification method and reasoning for choosing selected method
  - Ex: LightSIDE classification using Naïve Bayes with explanation on how it works
- Results
  - Metrics: accuracy, precision, and recall (will be discussed)
- Discussion/Conclusion
  - Explain your findings, limitations, and how you can improve

# Project Timeline

- **Note: Email Heejun all submissions**
- Fourth week (the week of June 11<sup>th</sup>)
  - Find a dataset and develop research question (Submit a proposal by June 13 11:00 pm)
  - Experiment with software on the server/computer
- Fifth week (the week of June 18<sup>th</sup>)
  - Investigate, test, and improve your approaches
  - Work with your instructors (PLEASE CONTACT US – piazza or email)
  - Start poster design (Submit preliminary design by June 21 11:00 pm)
- Sixth week (the week of June 25<sup>th</sup>)
  - Wrap-up
  - Finalize a poster and presentation slides (Submit final PDF by June 27 11:00 pm)
  - The presentation will be on June 29th

# Sample Project (Text Mining)

- Topic Classification
  - Multi-class classification (e.g., BRCA, Alzheimer's disease, heart disease, allergy)
  - Sub-topic classification (e.g., breast neoplasms vs. skin neoplasms)
- Classifying effects (positive/negative) of a certain treatment (or medicine) to a disease
  - Need to get PICO (Population, Intervention, Control, and Outcome) elements from ClinicalTrials.gov (by using a script)
  - Manually label the direction of effects (positive/negative)
- Exploratory text mining
  - Finding associations between diseases using information filtering

# What to Include for the Proposal

- A description of the problem you wish to investigate.
- Description of each member's role in the group.
- A list of 3-4 peer-reviewed papers you plan to survey to develop your approaches.
- The purpose of your experiments. What you are testing/exploring/associating?
- A description of the dataset you will use.
- A description of any risks or limitations.

Any Questions?

ENABLE



THE UNIVERSITY  
of NORTH CAROLINA  
at CHAPEL HILL

