

# Learning from memoirs: Classifying dementia using linguistic features extracted from non-clinical writing samples

Vaden Masrani  
Jacob Chen

# Background + Research Question

- What is Dementia?
  - Broad category of brain diseases which cause decrease in mental ability
  - **Causes speech and language difficulty** (among other symptoms)
- Previous Work
  - Supervised classification of dementia from linguistic features
  - **State of the art: 81% test accuracy**
    - Logistic regression
  - **Big weakness of previous work is small datasets**
- Research Question
  - **Can we improve test accuracy using writing samples from dementia patients?**
  - *“Non-clinical data”*: Writing or speech samples obtained outside a clinical setting, such as memoirs, books, blogs, emails, tweets, status updates, etc.
  - Siri could be a diagnostician!
  - Would allow for early detection and treatment of dementia

# Our proposed work

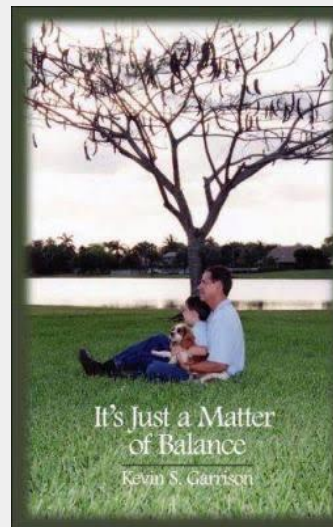
## 1. Extract text from books

- Welcome to Our World: A collection of life writing by people living with dementia
- It's Just a Matter of Balance: You Can't Put a Straight Leg on a Crooked Man

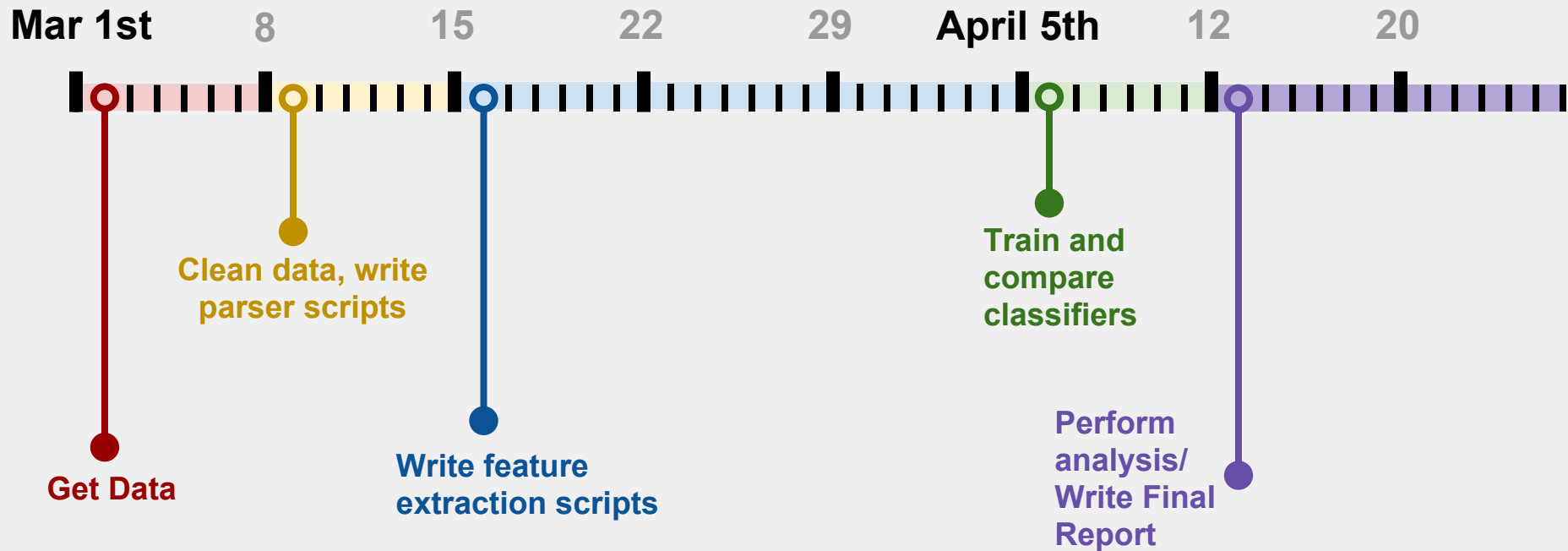
## 2. Use features proposed by Fraser (2015)

## 3. Train classifiers with and without added data

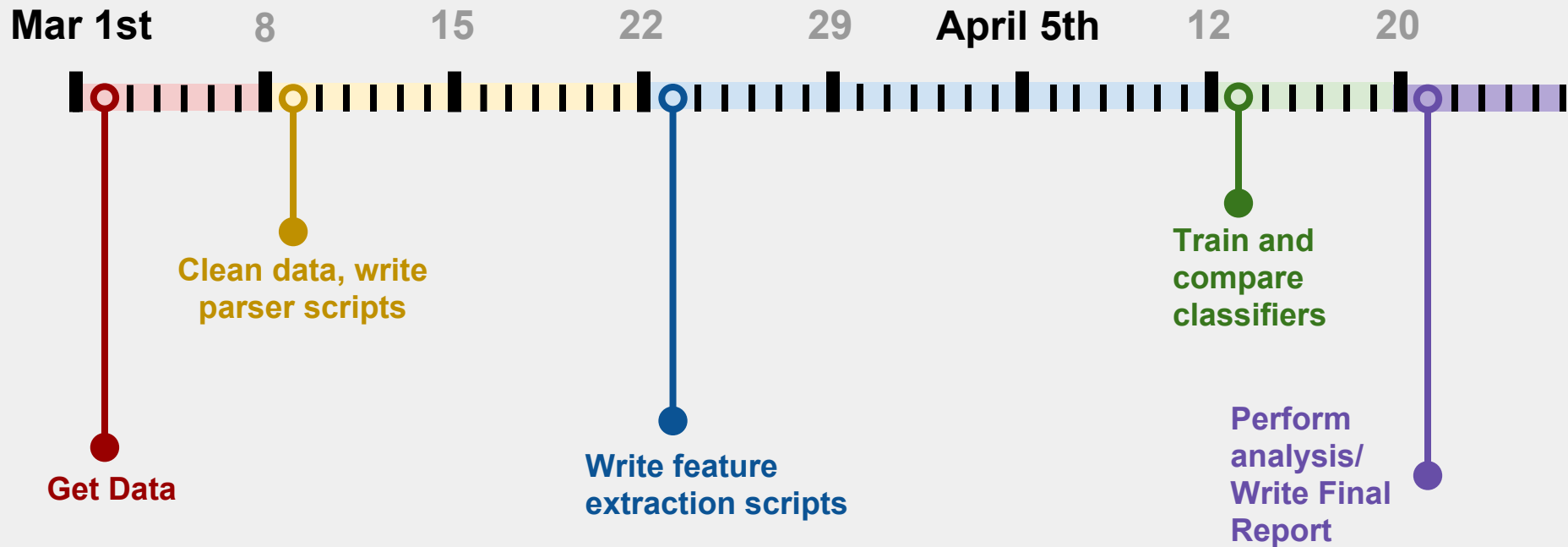
- Can we improve state of the art with extra “non-clinical” data?
- How do classifiers trained on clinical data do on non-clinical data?
- Can we reproduce Fraser (2015) accuracy of 81%?



# Proposed Research Plan



# ~~Proposed~~ Actual Research Plan



# Update

- Changes
  - Extracting and cleaning data took more time than planned
  - A lot (> 100) of features to extract from text!
  - Just started training
- What's left
  - Train and compare five classifiers on Weka
    - SVM, Naive Bayes, Decision Trees, Neural Networks, Bayes Nets
    - Train with and without added data
  - Compute F-Measure, Precision Accuracy
  - Write report

# References

1. Kathleen C. Fraser. and Jed A. Meltzer and Frank Rudzicz. Linguistic Features Identify Alzheimer's Disease in Narrative Speech, (2014).
2. Rentoumia, V. Raoufiana, L. Ahmed, S. A. de Jager C. and Garrard, P. Features and Machine Learning Classification of Connected Speech Samples from Patients with Autopsy Proven Alzheimer's Disease with and without Additional Vascular Pathology, Journal of Alzheimer's Disease, (2014).
3. Orimaye, Sylvester O. and Wong, Jojo S. and Golden, Karen J. Learning Predictive Linguistic Features for Alzheimer's Disease and related Dementias using Verbal Utterances. Association for Computational Linguistics, (2014).