CSE 4/587
Data Intensive Computing

Dr. Eric Mikida
epmikida@buffalo.edu
208 Capen Hall

Dr. Shamshad Parvin
shamsadp@buffalo.edu
313 Davis Hall

Project Overview

Basic Information

- Three Phases spanning the course of the semester
  - One continuous project – each phase builds on the previous
  - Each phase has its own due date and submission – no late submissions
- Can be done individually or as a team of two
  - Project requirements do not change based on team size
  - Project requirements will be slightly higher for 587 compared to 487
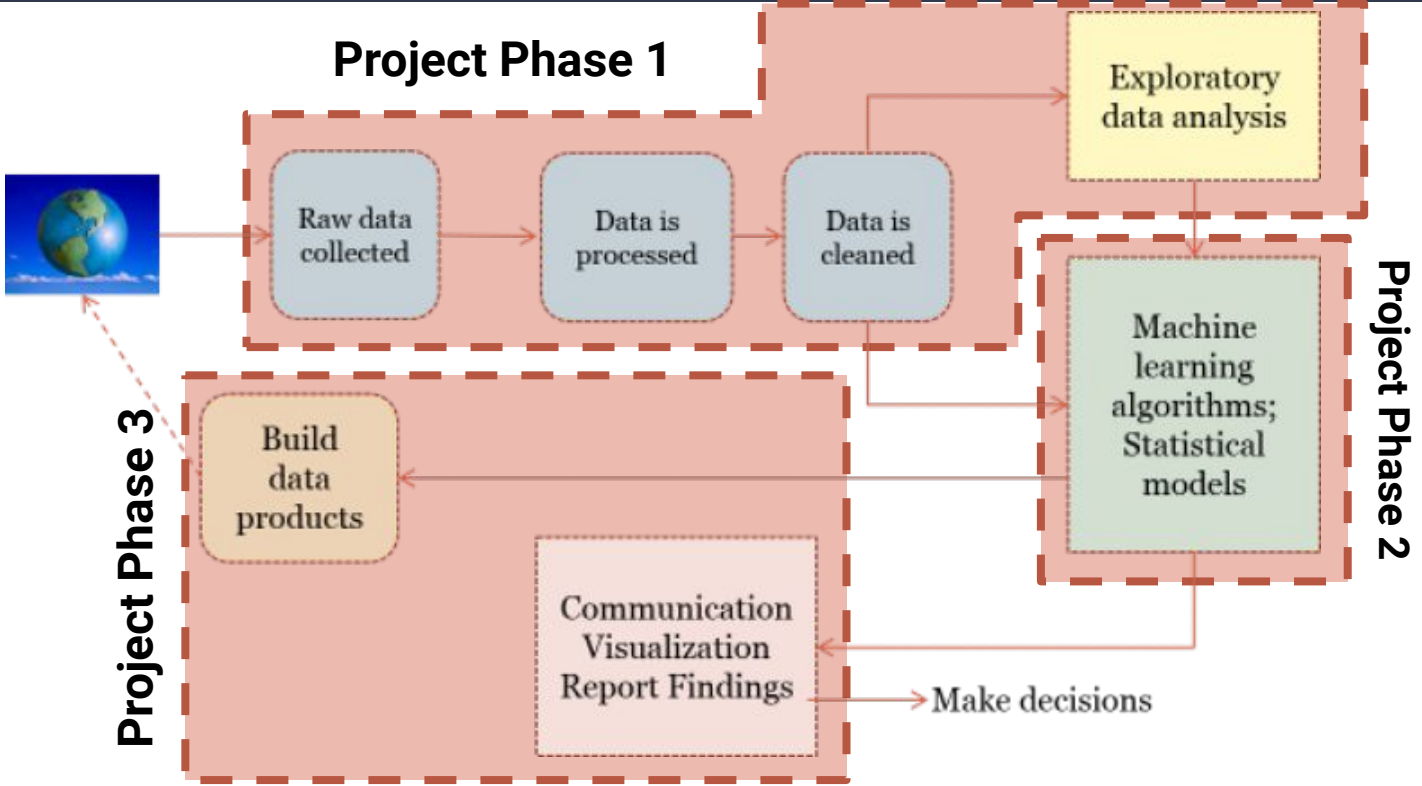
Project Phase Summary
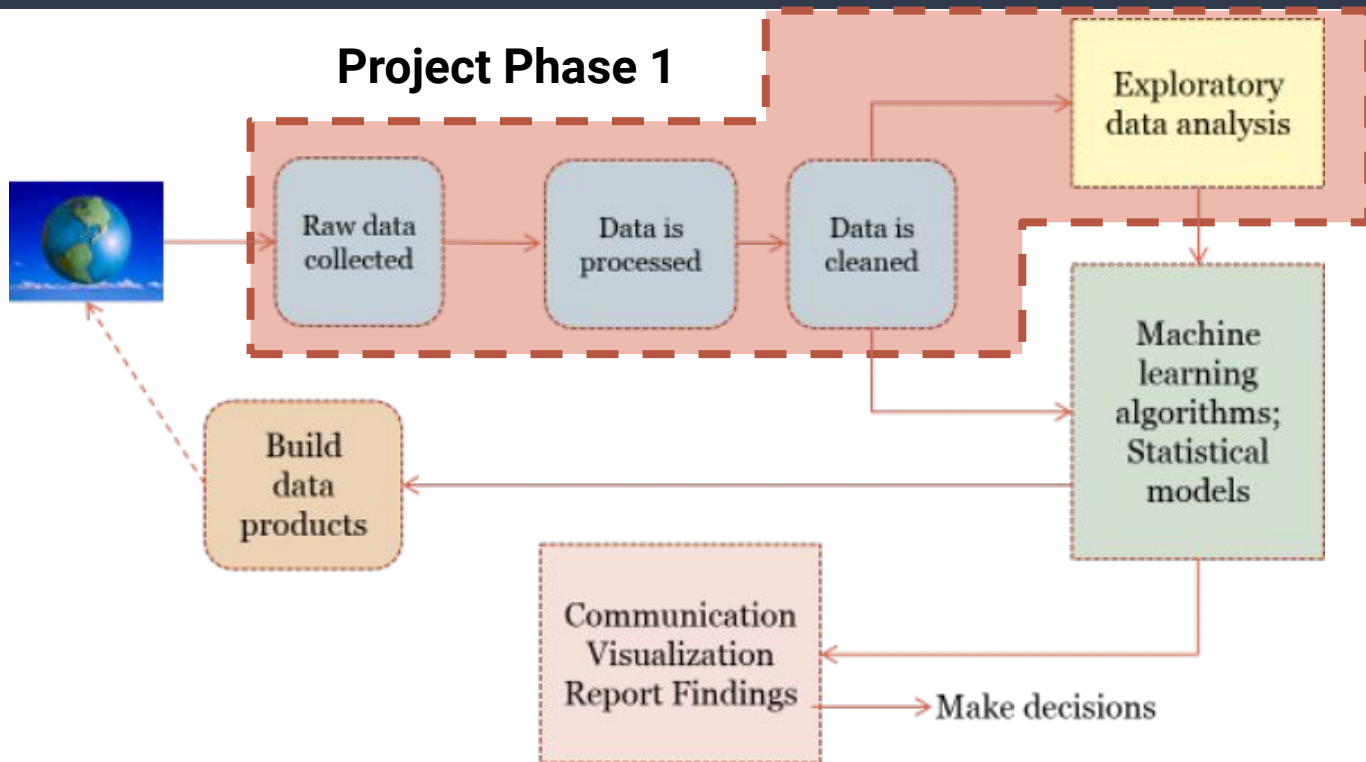
Project Phase 1

Project Phase Summary
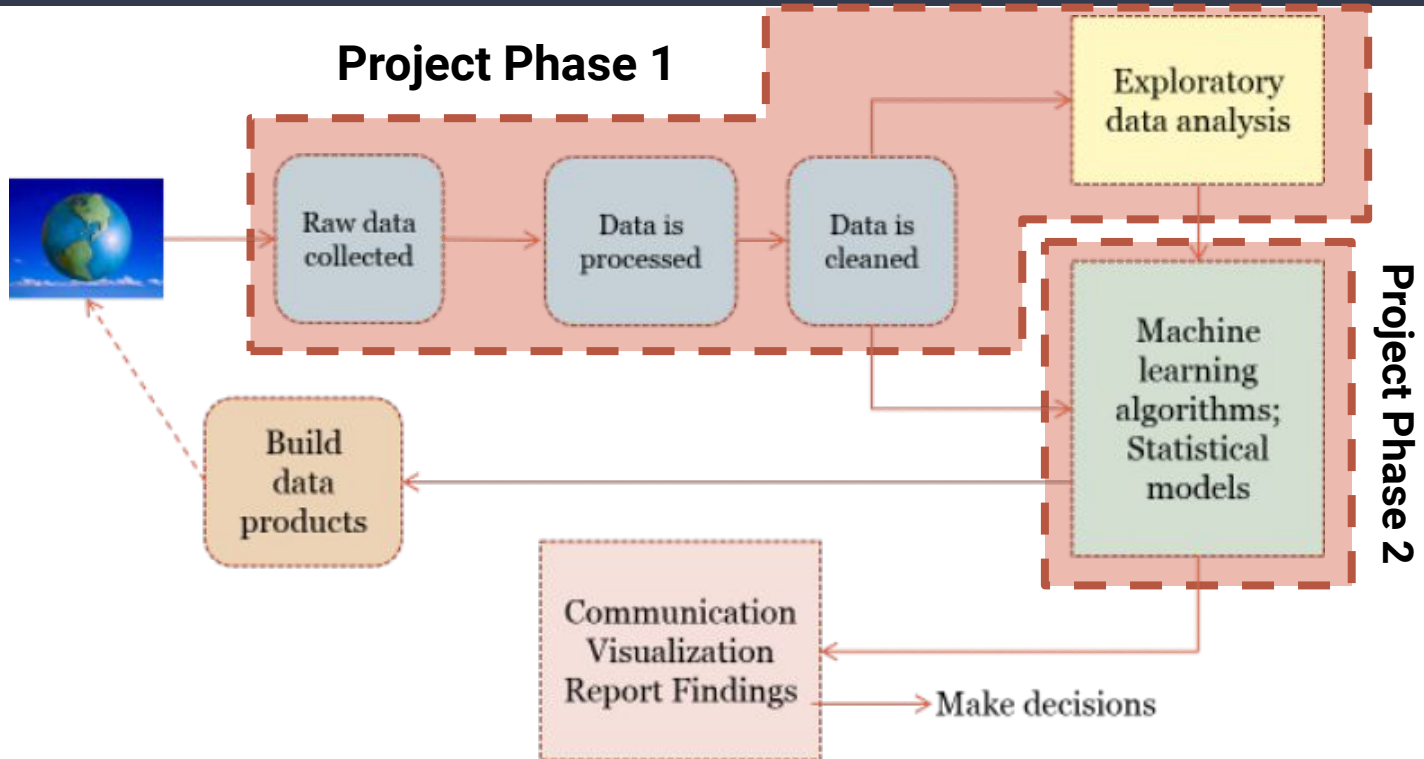
Project Phase Summary

First Steps

1. Find and register your team (will be a post on Piazza to do so)
2. Discuss possible problem statements/domains
   a. It should be a topic you find interesting/important
3. Phase 1 will be released at the end of the week

Example Problem Statement #1

Motivation

- Businesses are aware of how difficult it is to acquire new customers. This necessitates retention of customers as it can easily boost their profits.
- For banking and financial services, it is more crucial as Customers have a multitude of options
- Goal is to predict the possibility of customer attrition and target specific sections of the customer base to retain them instead of spending time on all customers

Problem Statement

- Analyze the bank data and predict the customers that are at the risk of attrition and could potentially stop being the bank's paying clients.
- Design data-driven solutions to address this problem. Leverage authentic and reliable bank data that includes all the necessary information of a customer.

Example Problem Statement #2

Motivation

- The way a doctor interprets your symptoms can impact whether your diagnosis is correct. Physicians have a lot of experience writing patient notes that detail the history of the patient's complaint, physical exam findings, probable diagnosis, and follow-up care by the time they're licensed. Learning and assessing the skill of writing patient notes necessitates feedback from other doctors, which is a time consuming process that could be streamlined with the use of machine learning

- We aim to identify clinical themes in patient notes and create an automated approach to link clinical ideas from an exam rubric (for example, "diminished appetite") to various ways these concepts are articulated in clinical patient notes provided by medical professionals (for example, "eating less," "clothing fit looser").

- If we're successful, we'll be able to help address some of the most pressing practical issues in patient note scoring, such as making the process clearer and more interpretable, as well as making the formulation and administration of such evaluations easier

Example Problem Statement #2

Problem Statement
- Develop a model to identify potential clinical features from the patient history notes taken by medical students.
- This model will be used to score patient note exams and thereby reducing physician intervention and financial resources.

Associated Research Questions
- How can we leverage NLP concepts to interpret free text and map it to a defined set of medical features?
- How do we associate a given medical symptom to multiple phrases in patient notes?
- How do we tackle the ambiguity problem in text? Ex: A statement can be related to different symptoms in different scenarios.