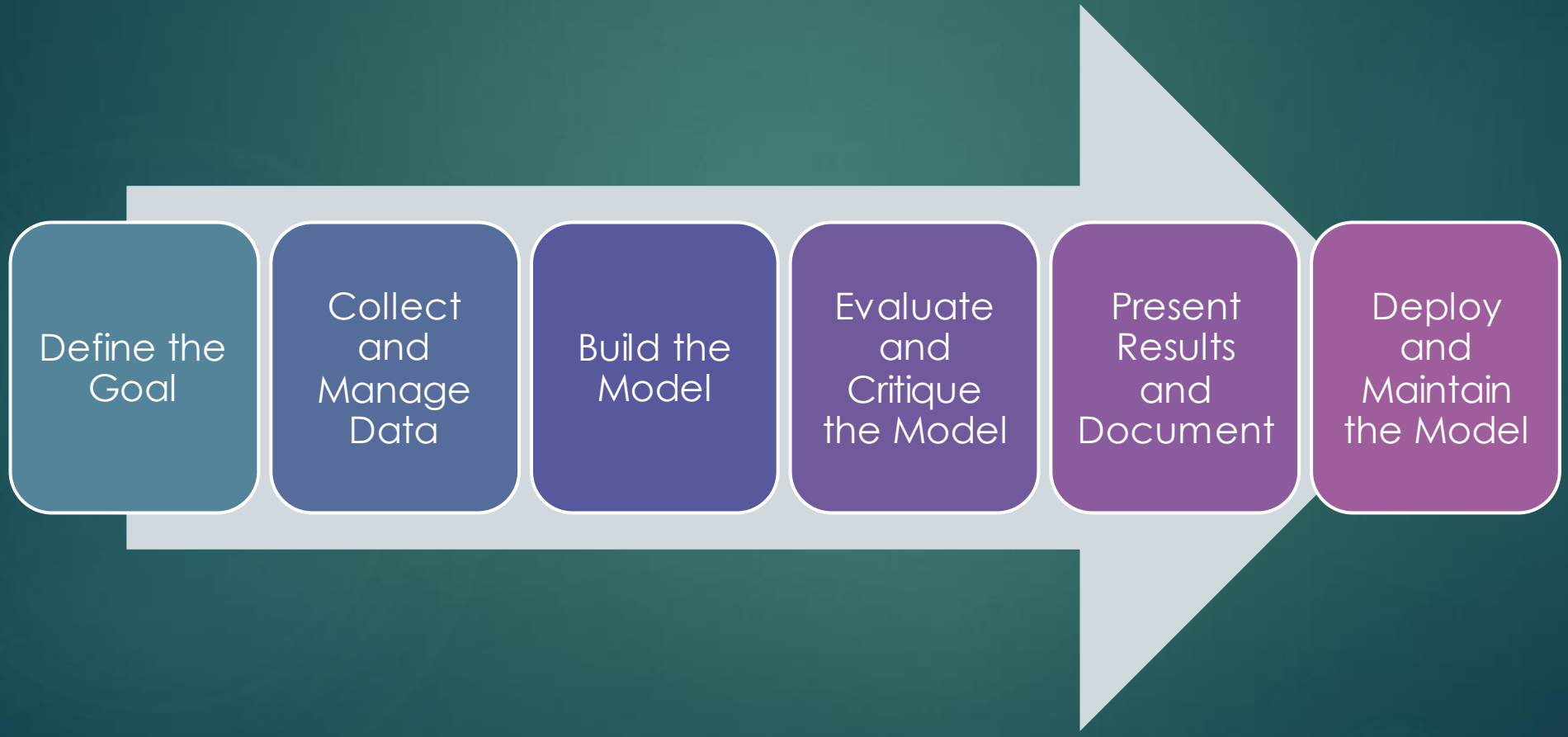
Credit One Project Introduction

BAILEY KEEL

Project Goals

- Our Goal is to give Credit One a model that gives them a better understanding of:
  - How much Credit to allow someone to use
  - If someone should be approved for credit at all
- We are looking to design and implement a creative, empirically sound solution for Credit One.

Framework Visualization



Framework-Define the Goal

- Why does Credit One want to do the project?
  - Credit One wants the project because they have seen an increase in the number of customers defaulting on their loans. Because of this they could lose business.
- What do they need from it?
  - Credit One needs a better analytical solution to understand how much credit to loan to an individual, or whether they should be loaned to at all.
- Why is their current solution inadequate?
  - Their current solution is inadequate because it is not an accurate model. They are getting wrong answers on how much credit to give someone and who to loan to.
- What resources do we need?
  - The resources needed are python, our local environment and the included modules.
- How will the results of our project be deployed?
  - The results of the data will be deployed on a company server, this will allow for new data to be added easily and also only allow people that have a "need to know" access it via credentials.

Framework-Collect and Manage Data

- What data is available?
  - We have data on 3670 creditors (2397 after duplicates are removed) , which includes education, age, sex, and payment/defaulting history.
- Will it help to solve the problem? Is it enough?
  - This data will be very beneficial in helping up solve this problem and should be enough for us to build an analytical model.
- Is the data quality good enough?
  - This data quality looks good in that it is full, but it looks like all the data types are not correct and there are also duplicate data points that needed to be removed (34% of the Data).

Framework-Build the Model

- Which technique might we apply to build the model?
  - The techniques used to build this model will be EDA to get a better understanding of the data. Then once we have that, Machine Learning to help build a model to accurately predict credit amounts.
- How many techniques should I apply?

Framework-Evaluate and Critique the Model

- Is the model accurate enough to meet the stakeholders' needs?
- Does it perform better than "the obvious guess" and any techniques being used correctly?
- Do the results of the model make sense in the context of the real-world problem domain?

Framework-Present Results and Documents

- How should stakeholders interpret the model?
- How confident should they be in its predictions?
- When should they potentially overrule the model's predictions?

Framework-Deploy and Maintain

- How is the model to be handed off to “Production”?
- How often, and under what circumstances, should the model be revised?

Description/Location of Data Resources

- This Data was from the Dept. of Civil Engineering at Tamkang University.
- Data was stored in a MySQL database
- The Data link was used to query this data into our own Pandas Framework
- Data had 23 Columns with info on:
  - Personal (Age, Gender, Education, Marital Status)
  - Credit history (Credit Given, Past Payments, Amount Billed, Previous Payment)
  - If the Client Defaulted or not

Data Management

- To Manage the Data, it will be accessed on a “need to know” basis.
- Stored on a secure server that is not open to everybody.

Data Issues

- Known issues are that the data has duplicates, all the data types are not appropriate, also the first row should be the column name.
- To fix this we will remove all the duplicate data, correctly assign data types, and remove the first row from the data.

Insights from the data

- From a quick cursory look at the data there are no major insights besides that the data will have to be cleaned before any learning can be done on it.

QUESTIONS?