DATA SCIENCE
FRAMEWORK REPORT

MAURICIO CARVAJAL

AGENDA

- Problem statement

- Proposal data science process framework

- Data management

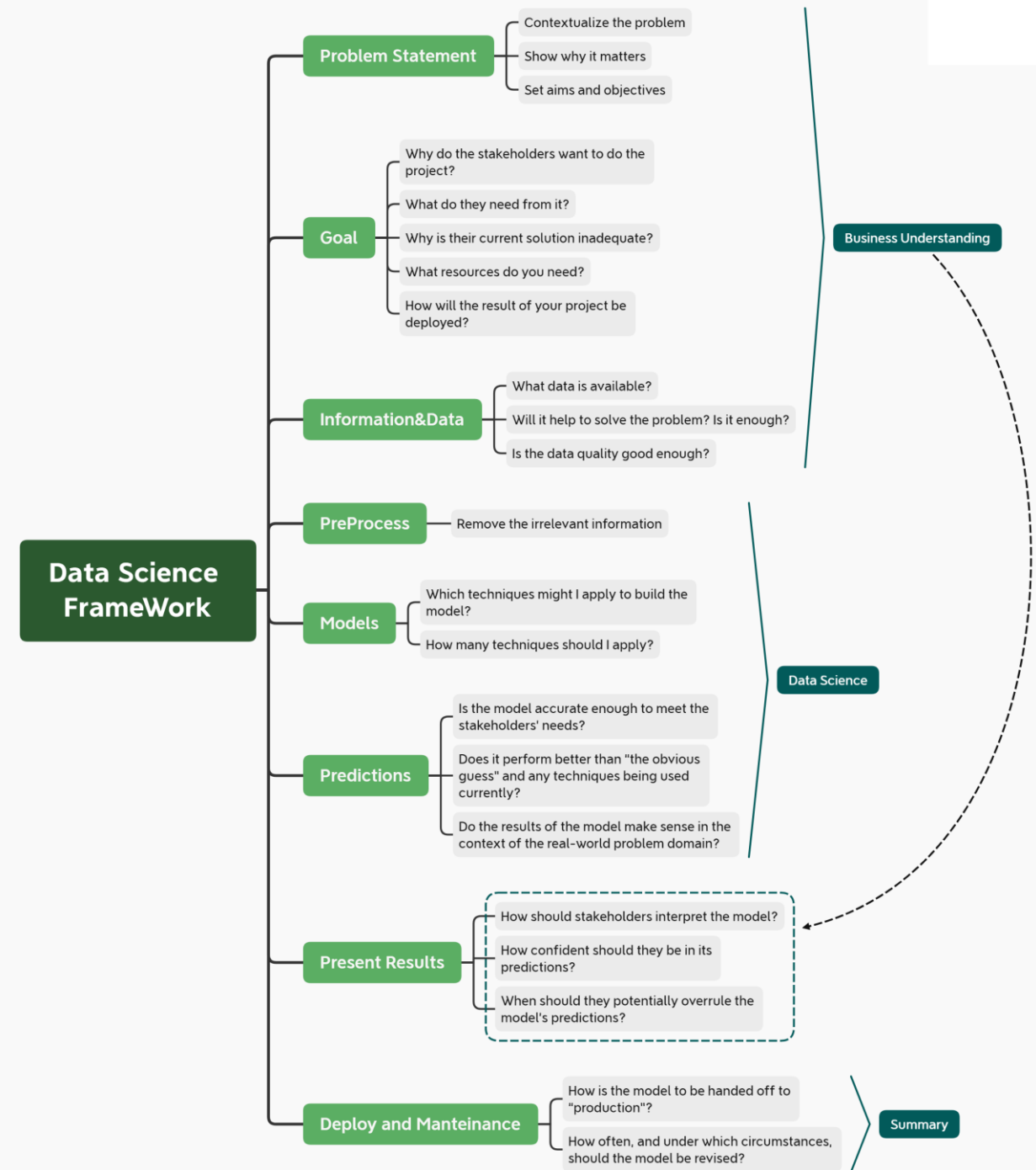- Risk management

- Insights

PROBLEM STATEMENT

- OVER THE PAST YEAR OR SO CREDIT ONE HAS SEEN AN INCREASE IN THE NUMBER OF CUSTOMERS WHO HAVE DEFAULTED ON LOANS THEY HAVE SECURED FROM VARIOUS PARTNERS, AND CREDIT ONE, AS THEIR CREDIT SCORING SERVICE, COULD RISK LOSING BUSINESS IF THE PROBLEM IS NOT SOLVED RIGHT AWAY.

GOALS

- GIVE A PREDICTION IN ARE GOING TO BE THE POSSIBLE NEW CUSTOMERS WITH PROBLEMS

- GIVE INSIDES ON WHAT THE ACTUAL SITUATION OF THE BUSINESS

- BASE ON DATA, RECOMMENDATIONS ON HOW IMPROVE THE SITUATION

DATA SCIENCE FRAMEWORK

- BASE ON THE MODEL :
  - ZUMEL AND MOUNT

- DIVIDED IN 3 GROUPS:
  - BUSINESS UNDERSTANDING
  - DATA SCIENCE (DATA MANIPULATION)
  - DEPLOY AND MAINTENANCE

INFORMATION OF THE DATA

- Amount of the given credit
- Gender
- Education
- Marital status
- Age (year)
- History of past payment
- Amount of bill statement
- Amount of previous payment
- client's behavior

DATA MANAGEMENT

- 1- Understanding Data
  - Need to convert the data, numerical to categorical
- 2-Proproces the information
  - Find a method to discard irrelevant information
    - Correlation, PCA, Feature variance, recursive feature elimination
- 3- Find the best model
  - Use different machine learnings models: RF,SVM,KNN C5.0
  - Evaluate the best performance
    - Kappa, accuracy
- 4-Use the best model to predict the results define in our problem statement

RISK DATA MANAGEMENT

- Need to have clear understanding of the type of the data.
  - Conversion from numerical to categorial as needed
- Understand if the problem is classification or regression problem
- The time that is needed to build the models

THANK
YOU

LMAURICIOCARVAJAL@GMAIL.COM