DIRECT MARKETING IN THE AGE OF BIG DATA

By: Gaby Lio, Lauren Stacy, Merideth Garcia

PROBLEM

DO NOT estimate likelihood of response or money spent

Mail catalogs to **ALL 5 MILLION** customers

INCREASE in mailing and printing costs

TARGET A SUBSET of customer

OBJECTIVES

Target as many buyers and as few non-buyers as possible

Target the subset of customers that yield the highest possible profit for Antixo

DATA

Total order amounts in $ in last 3 years

# of orders over past 3 years

Amount of most recent order

Date of most recent order

Date of first order

Average amount spent by customer

Frequency of order over last 24 months

Order amount category for the last 24 months

Target B: customers buying decision

Target D: Order margin in the last campaign

INFOGAIN TO RANK ATTRIBUTES

PRE-PROCESS

REMOVE TARGET D

ANALYSIS: OBJECTIVE 1

Calculated the base rate

~60%

Induced models with **ALL** variables and Class=Target B

Modified models by **deleting** variables

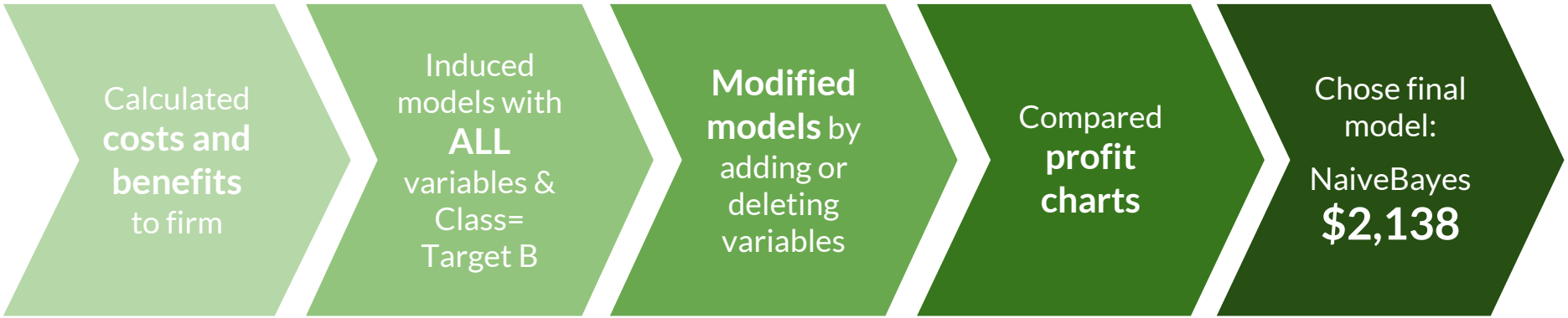
Compared **classification accuracy rates**

Chose final model:

Bagging with

62.011%

ANALYSIS: OBJECTIVE 2



CHALLENGES

Similar classification accuracies

Classification accuracy vs. cost/benefit

Could not use Target D as a class for cost/benefit

Finding significant variables

RECOMMENDATIONS

- Use final bagging model
- Calculate probability
- Rank customers on probability
- Target top 30% of customer
- Option: use minimum threshold

- Use final NaiveBayes
- Adjust cost/benefit for accuracy
- Calculate probability & rank customers
- Calculate minimum threshold
- Target customers at/above threshold
- Maximize profit

IMPLICATIONS

SAVE MONEY by only targeting a subset of customers

SAVE TIME by making faster decisions

INCREASE PROFIT by targeting those with highest probability of purchasing

Better than targeting customers at random (current practice)

QUESTIONS?