ANKIT SOBTI

SAAD KHAN

VARGHESE CHERIAN

HARNEET CHAWLA

KANIKA MIGLANI

PC MAILING CAMPAIGN – GROUP A8

BUSINESS ANALYTICS USING DATA MINING

BUSINESS GOAL
AN OVERVIEW

- **Client Profile**: The client is a diversified FMCG company with products in the Personal Care category
- **Business Objective:**
  - To find out the top 3 selling products last year in the Personal Care category amongst existing loyalty program customers
  - To run a mail-in-discount offer campaign of the top products to future loyalty program customers based on their likelihood of purchase
  - Reduce the costs of sending out mail-in-discount coupons to customers who are not likely to avail the offer
  - Increase loyalty of those customers who already buy PC products – Increase their switching costs & generate incremental volumes

DATA MINING GOAL
AN OVERVIEW

Supervised Learning

- **Data Mining Objective:**
  - Predict the likelihood of purchase of the top selling items in PC by a loyalty program customer

Customer Data	Basket Level Transactions
• Member No. • Age • Sex (Binary variable) • Marital Status (Binary variable) • City (Binary variable) • State (Binary variable) • Enrollment Duration	• Average Price • Average Quantity • Average SKU Count • Average Sub Department Count • No. of baskets

  - Input data was collected/modified for total (8475) customers based on the assurance that such data will be readily available in future

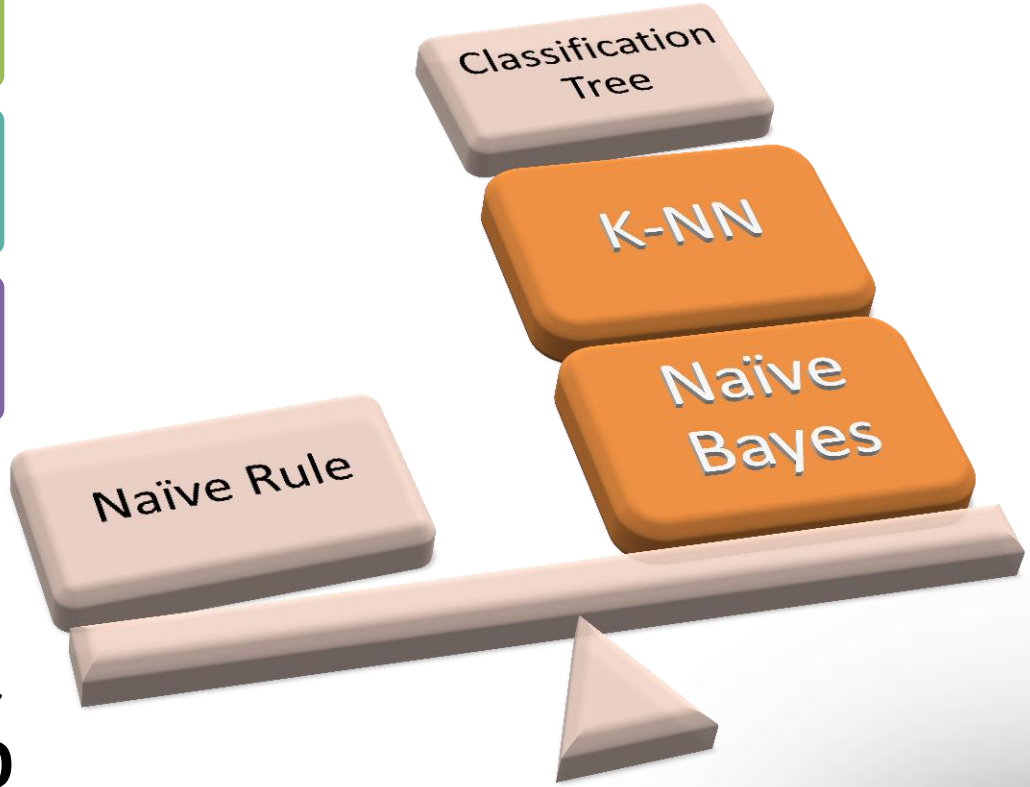
METHODS

AN OVERVIEW

Dove Soap REG 100 gm

Close-Up Red 150 gm

Pears Germ Shield 125 gm



Benchmark

6.63%

K-NN
AN OVERVIEW

Training Data scoring - Summary Report (for k=9)

Cut off Prob.Val. for Success (Updatable)	0.4

Classification Confusion Matrix		
	Predicted Class	
Actual Class	Buy	No Buy
Buy	25	243
No Buy	19	3950

Error Report			
Class	# Cases	# Errors	% Error
Buy	268	243	90.67
No Buy	3969	19	0.48
Overall	4237	262	6.18

Validation Data scoring - Summary Report (for k=9)

Cut off Prob.Val. for Success (Updatable)	0.4

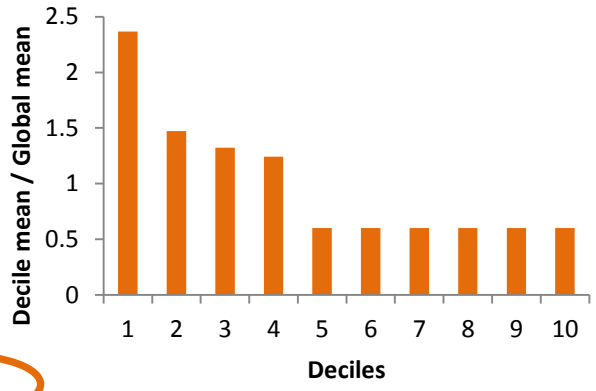
Classification Confusion Matrix		
	Predicted Class	
Actual Class	Buy	No Buy
Buy	12	247
No Buy	31	3947

Error Report			
Class	# Cases	# Errors	% Error
Buy	259	247	95.37
No Buy	3978	31	0.78
Overall	4237	278	6.56

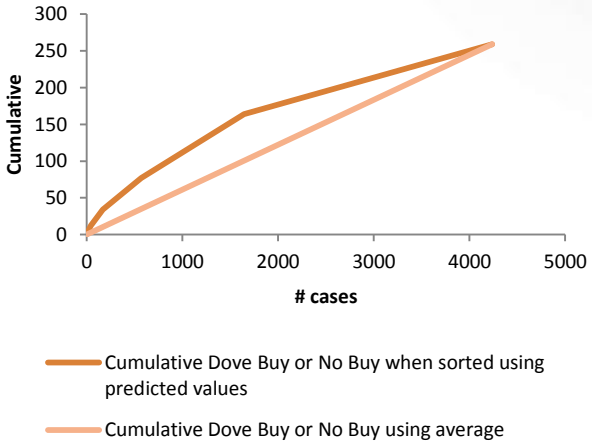
Validation error log for different k

Value of k	% Error Training	% Error Validation	
1	0.00	11.31	
2	5.83	15.06	
3	5.88	7.41	
4	6.07	8.59	
5	5.90	6.49	
6	5.99	6.82	
7	6.18	6.40	
8	6.14	6.51	
9	6.37	6.18	<--- Best k
10	6.30	6.25	

Decile-wise lift chart (validation dataset)



Lift chart (validation dataset)



- **Identify K observations in the training set which are similar to the new record that is to be classified**
- **K-NN does not make assumptions about the relationship between Dove (buy or not) & predictors**
- **Cut off value = 0.4 (Buy)**
- **K = 9 for highest classification performance**
- **Observing the lift chart for validation set, the K-NN method gives a lift of almost 2.4 for the first 10 %**
- **K-NN method is performing better than random assignment**
- **simple and lack of parametric assumptions and when the data set is large**

PC MAILING CAMPAIGN

NAIVE BAYES

AN OVERVIEW

Training Data scoring - Summary Report

Cut off Prob.Val. for Success (Updatable)	0.4

Classification Confusion Matrix		
	Predicted Class	
Actual Class	Buy	No Buy
Buy	92	176
No Buy	481	3488

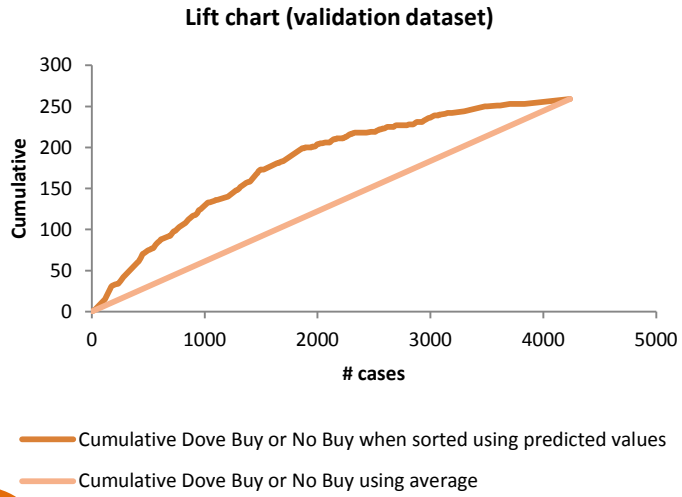
Error Report			
Class	# Cases	# Errors	% Error
Buy	268	176	65.67
No Buy	3969	481	12.12
Overall	4237	657	15.51

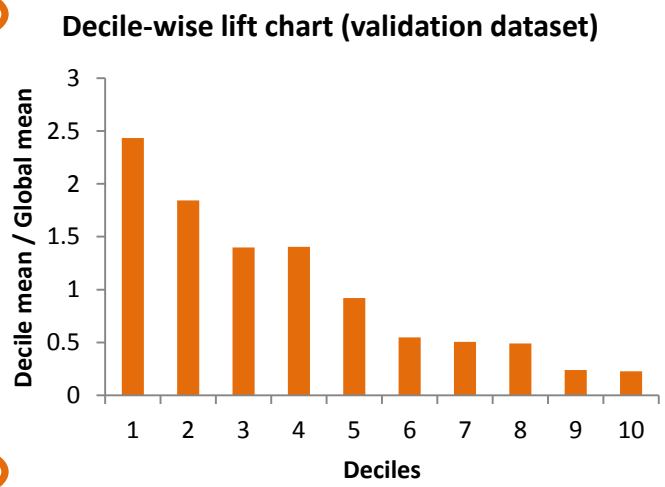
Validation Data scoring - Summary Report

Cut off Prob.Val. for Success (Updatable)	0.4

Classification Confusion Matrix		
	Predicted Class	
Actual Class	Buy	No Buy
Buy	81	178
No Buy	482	3496

Error Report			
Class	# Cases	# Errors	% Error
Buy	259	178	68.73
No Buy	3978	482	12.12
Overall	4237	660	15.58





- It is a more sophisticated method than Naïve rule as it integrates the information in the predictors
- To convert numerical variables into categorical, 4 bins were created
- Classification task is to estimate the probability of membership to buy/not buy given a certain set of predictor variables i.e. conditional probability
- Naïve Bayes assumes that the predictors are all mutually independent within each class
- Based on the results (with 0.4 as the cutoff probability of buy), it is observed that the overall error was around 15.55% for validation set
- Simplicity, computational efficiency and good classification performance

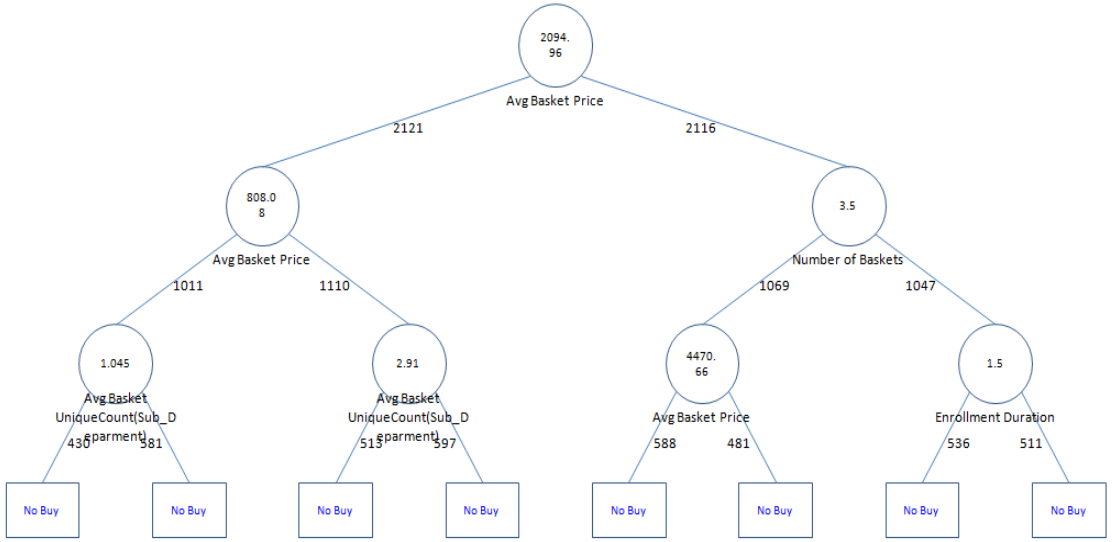
CLASSIFICATION
AN OVERVIEW

Training Data scoring - Summary Report (Using Full Tree)

Cut off Prob.Val. for Success (Updatable)	0.4

Classification Confusion Matrix		
	Predicted Class	
Actual Class	Buy	No Buy
Buy	0	268
No Buy	0	3969

Error Report			
Class	# Cases	# Errors	% Error
Buy	268	268	100.00
No Buy	3969	0	0.00
Overall	4237	268	6.33



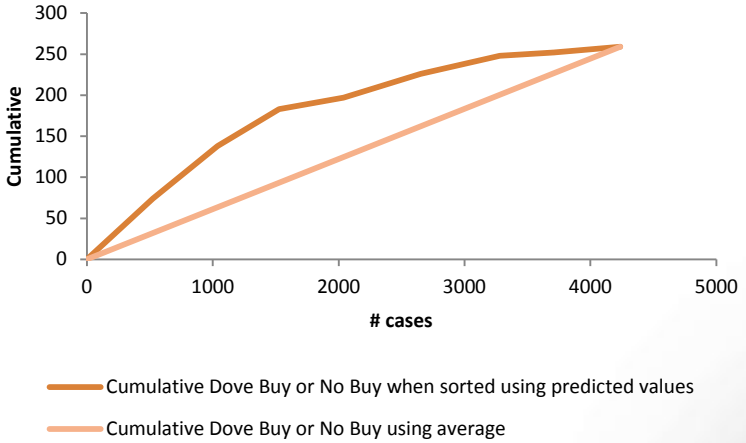
Validation Data scoring - Summary Report (Using Full Tree)

Cut off Prob.Val. for Success (Updatable)	0.4

Classification Confusion Matrix		
	Predicted Class	
Actual Class	Buy	No Buy
Buy	0	259
No Buy	0	3978

Error Report			
Class	# Cases	# Errors	% Error
Buy	259	259	100.00
No Buy	3978	0	0.00
Overall	4237	259	6.11

Lift chart (validation dataset)

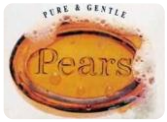


PC MAILING CAMPAIGN

RECOMMENDATIONS

AN OVERVIEW

- Deploy K-NN Model for mail-in-discount offers in PC category
- Expand model to Homecare Category – Top 3 products observed (Surf Excel detergent bar, Rin adv detergent bar & Sanifresh)
  - Buying pattern/behavior same for both categories thus extension of model viable
- Buy data from Hypermart for predicting future purchases
- Not viable to **all** categories
  - Buying cycle not as high
  - Cost-Benefit Analysis (incremental sales vs. cost of mailing)



THANK YOU!

QUESTIONS?