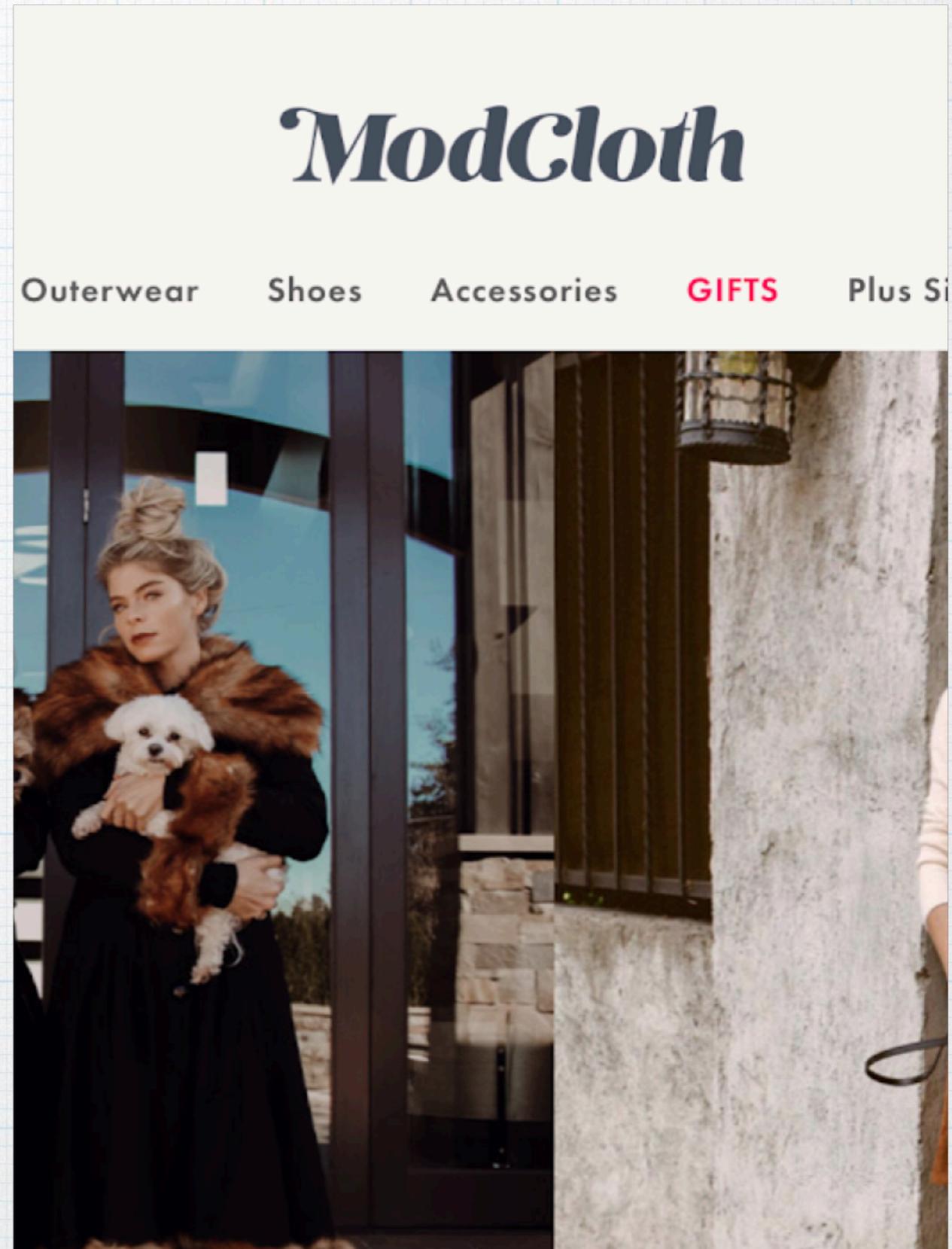


Clothing Fit

Presented By:
Analytics Triad

Ashna Gupta
Priyanka Raju
Kesha Shah



Introduction

Returned a cloth because it was not fit?



Price: **\$24.99** FREE Shipping on orders over \$25.00 shipped by Amazon or get **Fast, Free Shipping with Amazon Prime** FREE Returns

Fit: As expected (80%) ▾

Special Size: Standard

Standard

Plus

Size:

10 L

10 M

10 S

10 Large

12 Short

12

12 Long

12 L



Predict Fit of Customer & Infer/Recommend Size

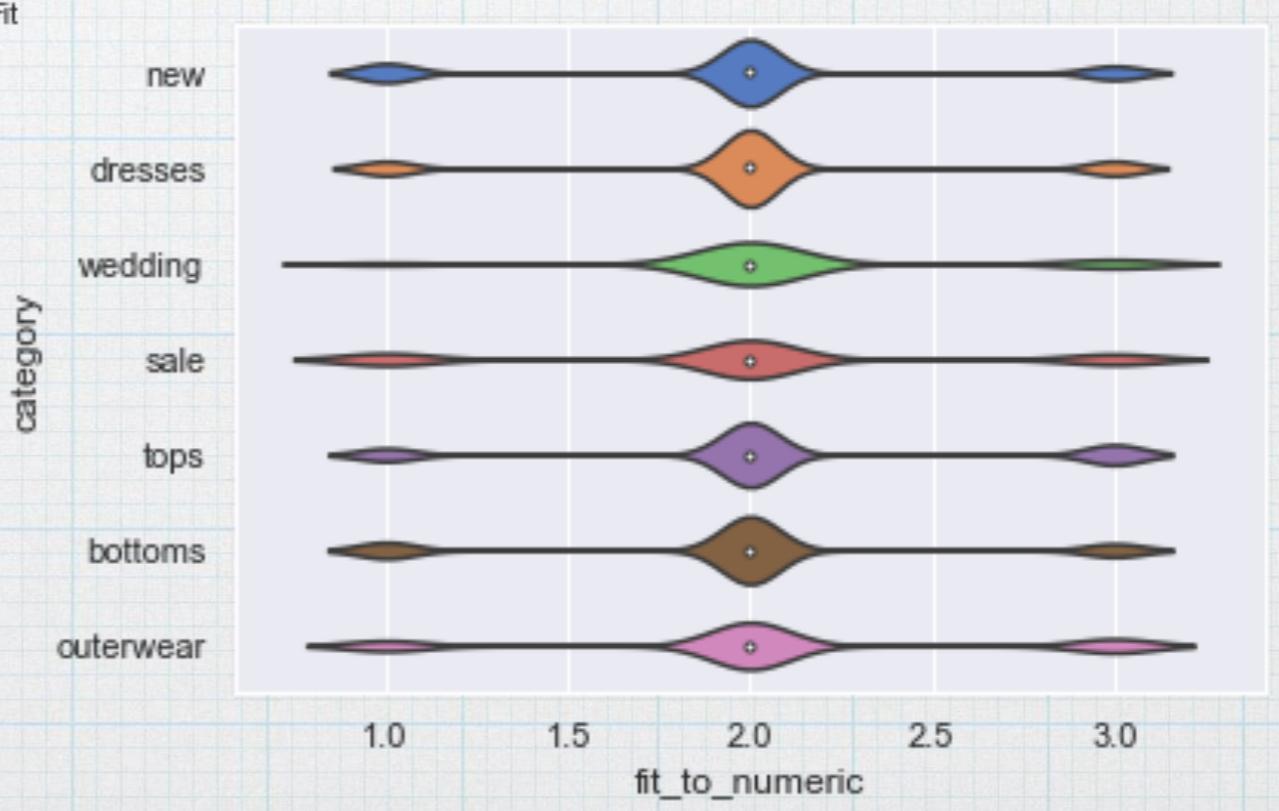
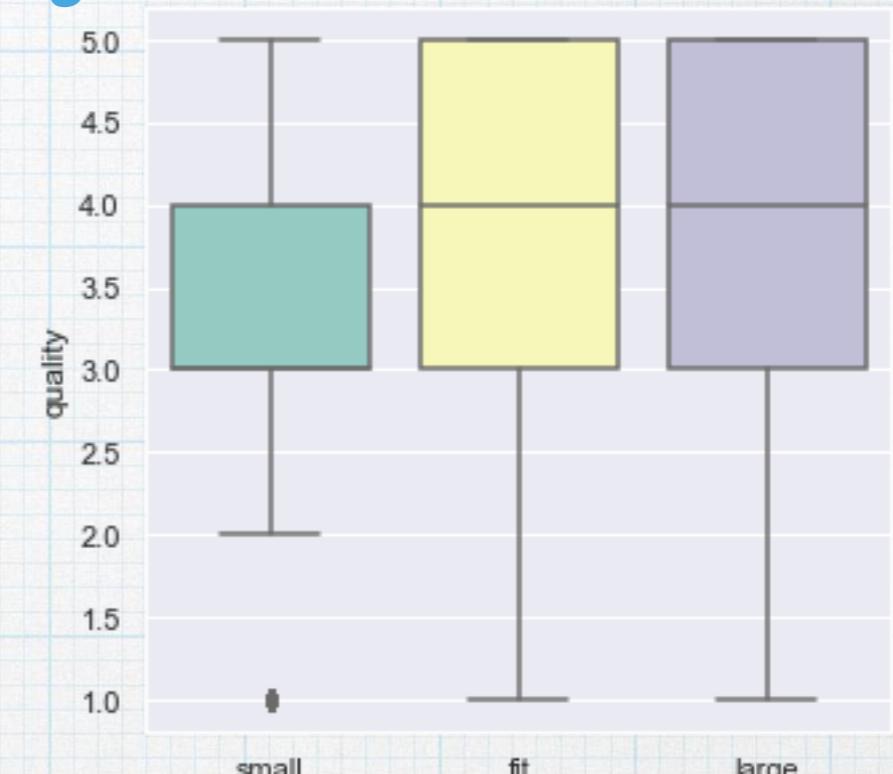
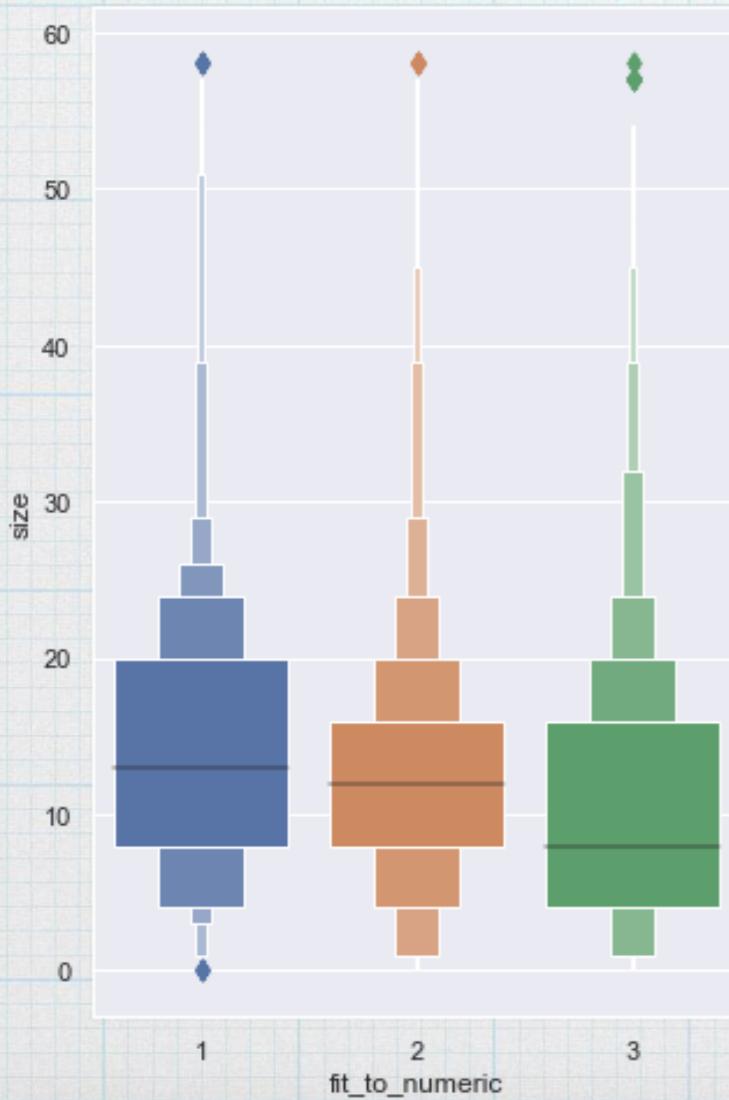
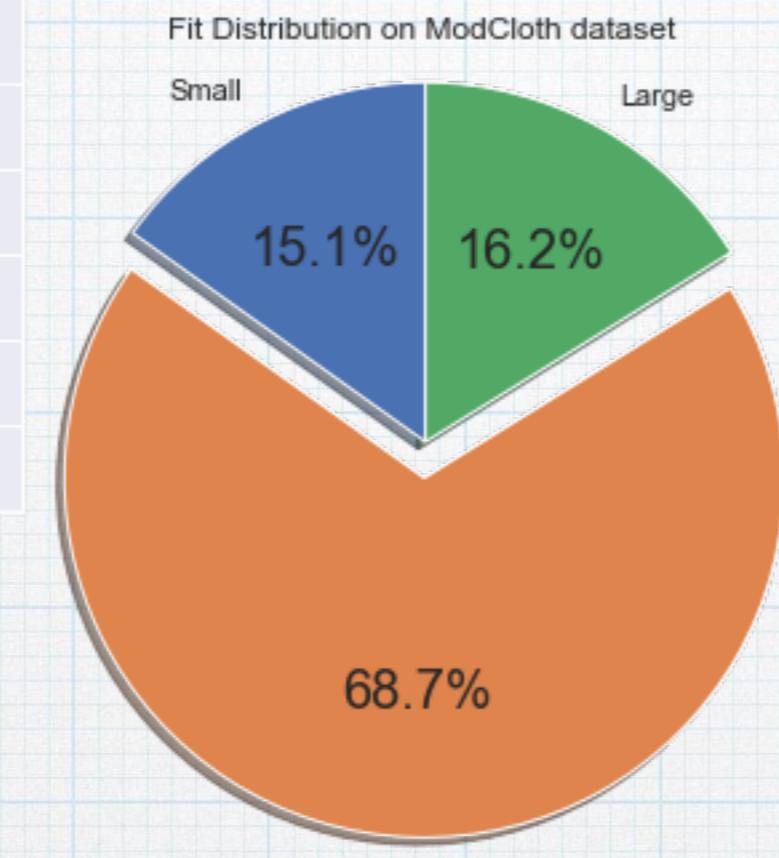
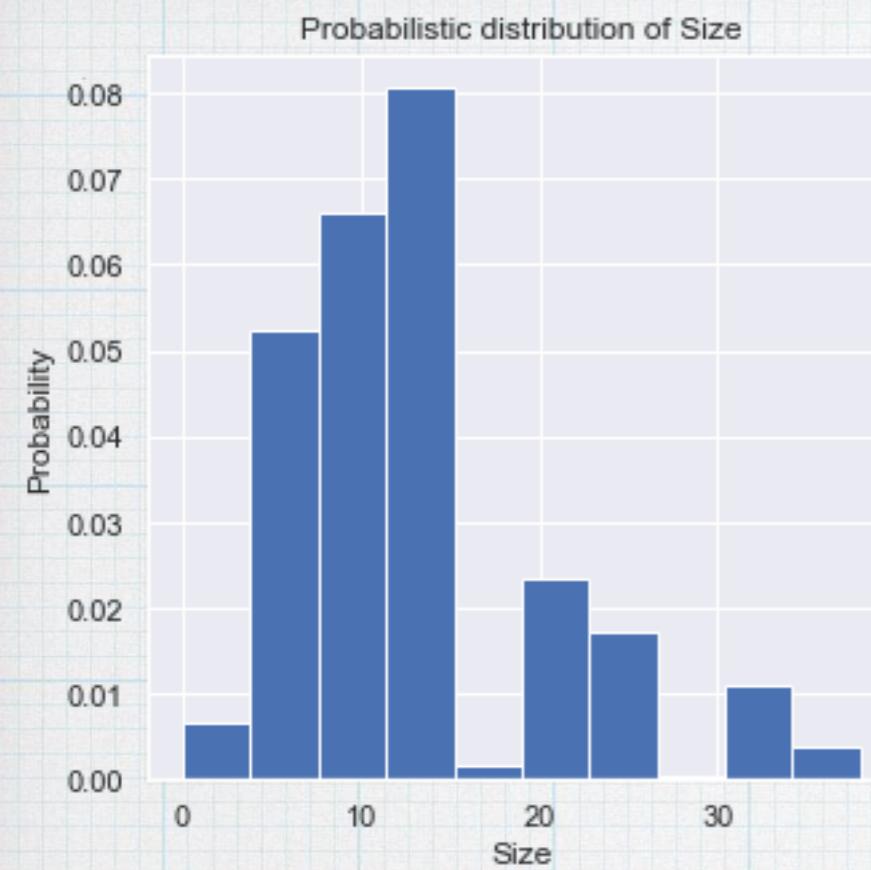
Dataset

- **ModCloth**
Size: 40MB
Features: 18
No. of Customers: 47,958
No. of Products: 1,378
- **RenttheRunway**
Size: 123MB
Features: 15
No. of Customers: 105,508
No. of Products: 5,850

Algorithms

- Logistic Regression
- Naive Bayes
- Large Margin Nearest Neighbor (Metric Learning)
- K-Nearest Neighbor

Data Exploration



Merging of Dataset

	Count of missing values	Percentage of missing values	Data Type
bra size	6018	7.27	float64
bust	70936	85.68	object
category	0	0.00	object
cup size	6255	7.56	object
fit	0	0.00	object
height	1107	1.34	object
hips	26726	32.28	float64
item_id	0	0.00	int64
length	35	0.04	object
quality	68	0.08	float64
review_summary	6725	8.12	object
review_text	6725	8.12	object
shoe size	54875	66.28	float64
shoe width	64183	77.53	object
size	0	0.00	int64
user_id	0	0.00	int64
user_name	0	0.00	object
waist	79908	96.52	float64
fit_to_numeric	0	0.00	int64
height_cms	1107	1.34	float64

	Count of missing values	Percentage of missing values	Data Type
age	960	0.50	float64
body type	14637	7.60	object
bust size	18411	9.56	object
category	0	0.00	object
fit	0	0.00	object
height	677	0.35	object
item_id	0	0.00	int64
rating	82	0.04	float64
rented for	10	0.01	object
review_date	0	0.00	object
review_summary	0	0.00	object
review_text	0	0.00	object
size	0	0.00	int64
user_id	0	0.00	int64
weight	29982	15.57	object

	Count	Percentage	Data Type
bra size	198562	72.12	float64
bust	263480	95.69	object
category	0	0.00	object
cup size	198799	72.20	object
fit	0	0.00	object
height	1784	0.65	float64
hips	219270	79.64	float64
item_id	0	0.00	int64
length	192579	69.94	object
quality	150	0.05	float64
review_summary	6725	2.44	object
review_text	6725	2.44	object
shoe size	247419	89.86	float64
shoe width	256727	93.24	object
size	0	0.00	int64
user_id	0	0.00	int64
user_name	192544	69.93	object
waist	272452	98.95	float64
age	83750	30.42	float64
body type	97427	35.39	object
bust size	101201	36.76	object
rented for	82800	30.07	object
review_date	82790	30.07	object
weight	112772	40.96	object

Preprocessing

* Missing Values

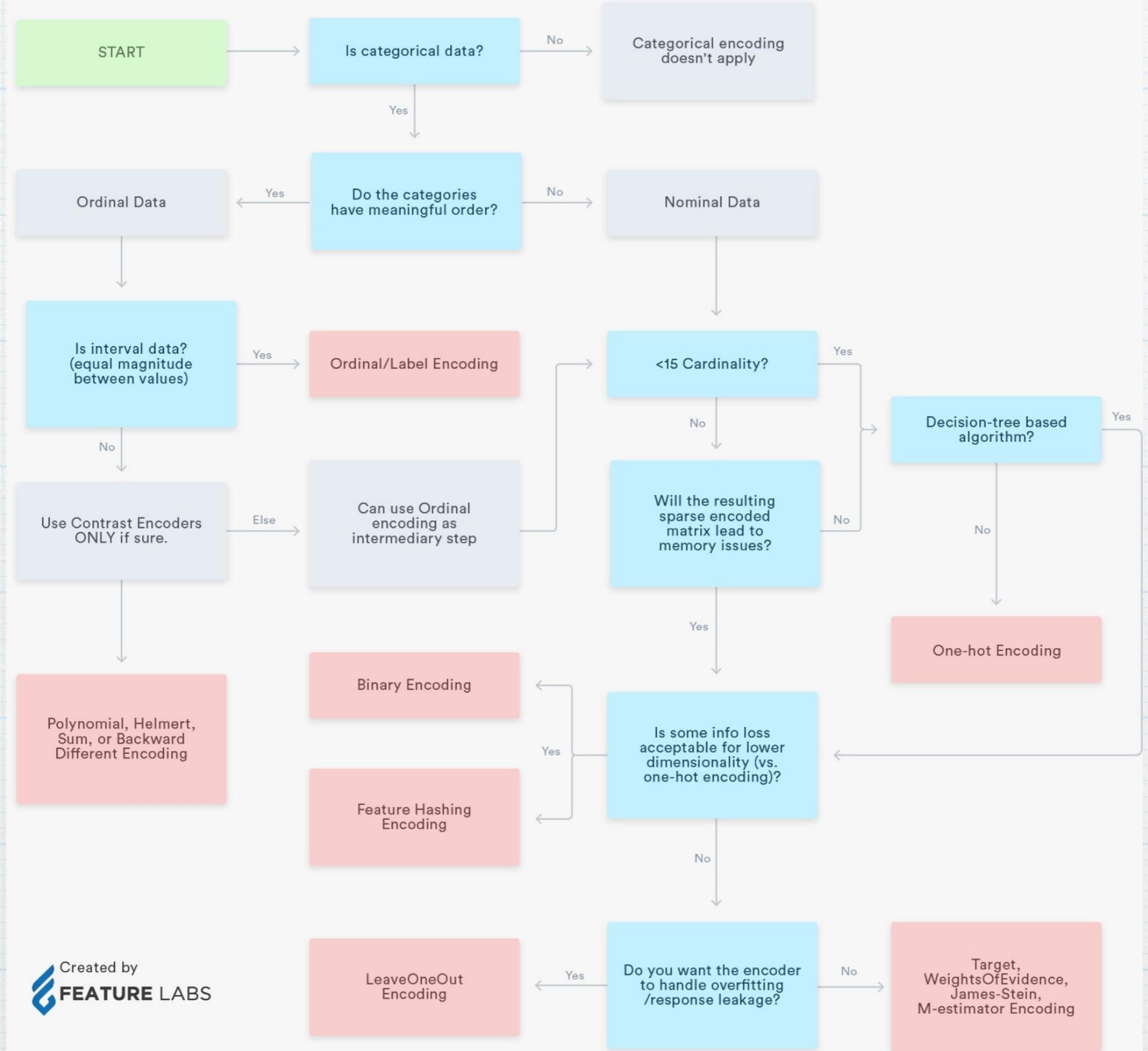
* Categorical Data

* Handling Duplicates

* Removal of features

	Count	Percentage	Data Type
bra size	0	0.0	float64
category	0	0.0	category
cup size	0	0.0	category
fit	0	0.0	category
height	0	0.0	float64
hips	0	0.0	float64
item_id	0	0.0	int64
length	0	0.0	category
quality	0	0.0	float64
shoe size	0	0.0	float64
shoe width	0	0.0	category
size	0	0.0	int64
user_id	0	0.0	int64
age	0	0.0	int64
body type	0	0.0	category
bust size	0	0.0	category
rented for	0	0.0	category
weight	0	0.0	float64
fit_to_numeric	0	0.0	int64

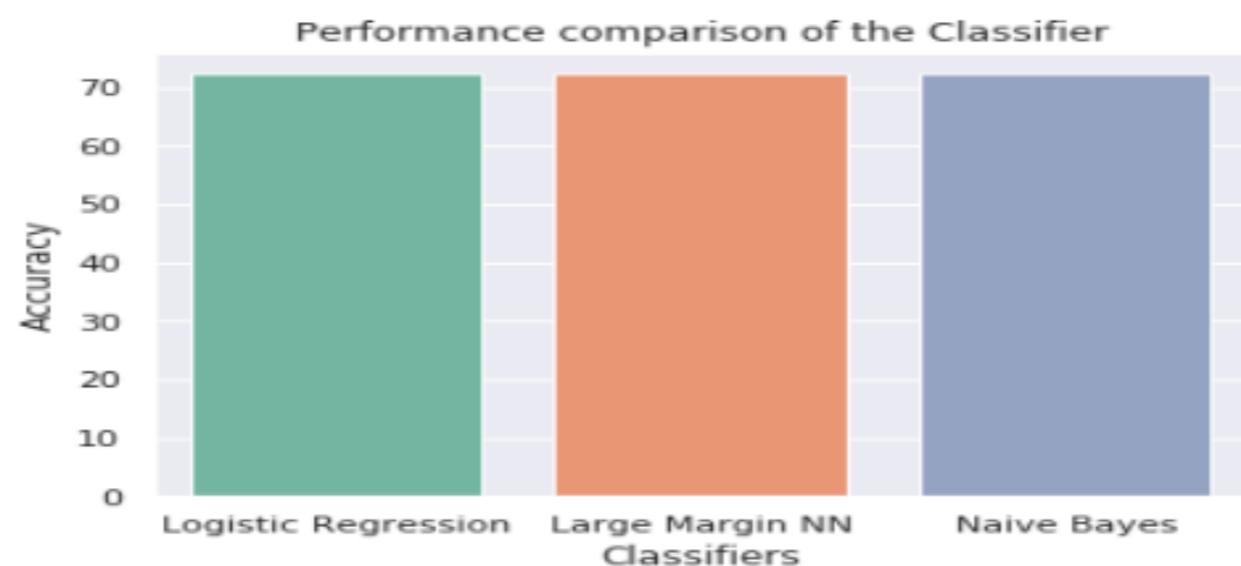
Categorical Encoding Methods Cheat-Sheet



Created by
FEATURE LABS

Evaluation

Algorithm	Accuracy	Parameters(tuned)
Logistic Regression Model	72.31	Max_iter = 2000 Solver = saga
Naive Bayes Model	72.07	Priors = None Var_smoothing 1e-09
Large Margin Nearest Neighbor	72.31	K = 50



Classification Report

Classification Report for Naive Bayes:

	precision	recall	f1-score	support
1	0.00	0.00	0.00	19366
2	0.72	1.00	0.84	99473
3	0.00	0.00	0.00	18828
accuracy			0.72	137667
macro avg	0.24	0.33	0.28	137667
weighted avg	0.52	0.72	0.61	137667

Classification Report for Logistic Model:

	precision	recall	f1-score	support
1	0.00	0.00	0.00	7661
2	0.72	1.00	0.84	39820
3	0.00	0.00	0.00	7586
accuracy			0.72	55067
macro avg	0.24	0.33	0.28	55067
weighted avg	0.52	0.72	0.61	55067

Classification Report for the best model:

Classification Report				
	precision	recall	f1-score	support
small	0.00	0.00	0.00	7661
fit	0.72	1.00	0.84	39820
large	0.33	0.00	0.00	7586
accuracy			0.72	55067

Things that didn't work

- LMNN being partly implemented
- SVM implementation
- Dataset constraint

Things that worked

- Achieve an above average accuracy for the problem defined
- Able to work on merged dataset even with limited availability of content

Conclusion

- Exploring large scale data
- Future Scope
 - 1. Better performing model
 - 2. Size inference feature for online portals

'S NEW

CLOTHING

DESIGNERS

HOLIDAY



SIGN IN



\$120 off Unlimited or 20% off your first order. Use Code: THEMOMENT [Details](#)



This is Your
Moment

Q&A
Thank You