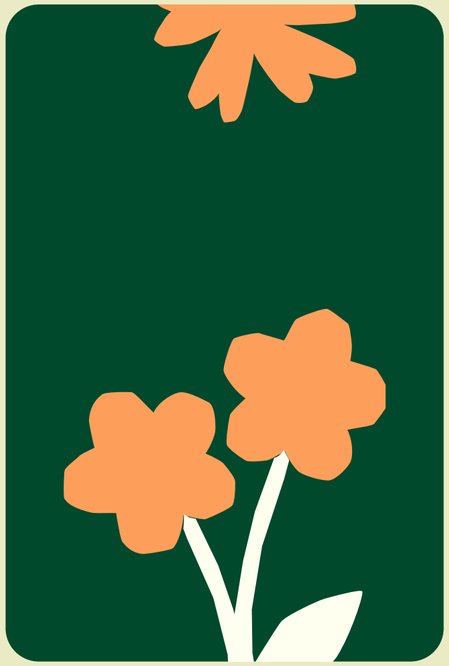
Model Comparison

ERSCOI LELIA

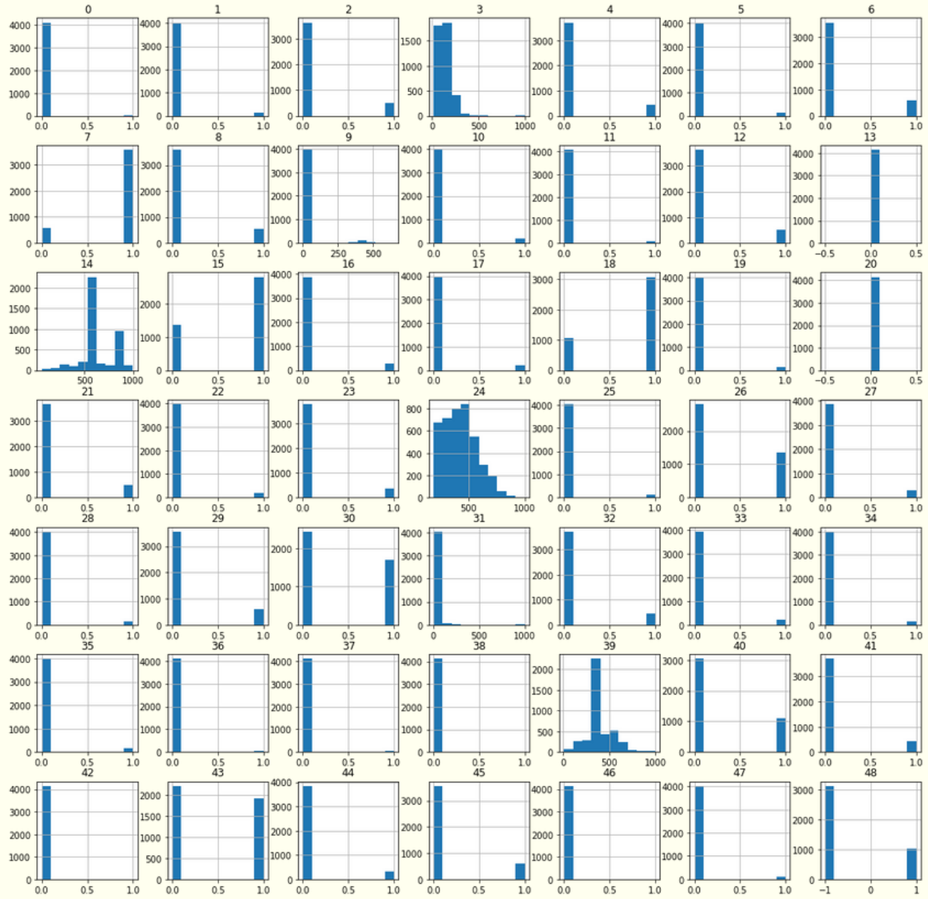



Train and Validation Data- ADA

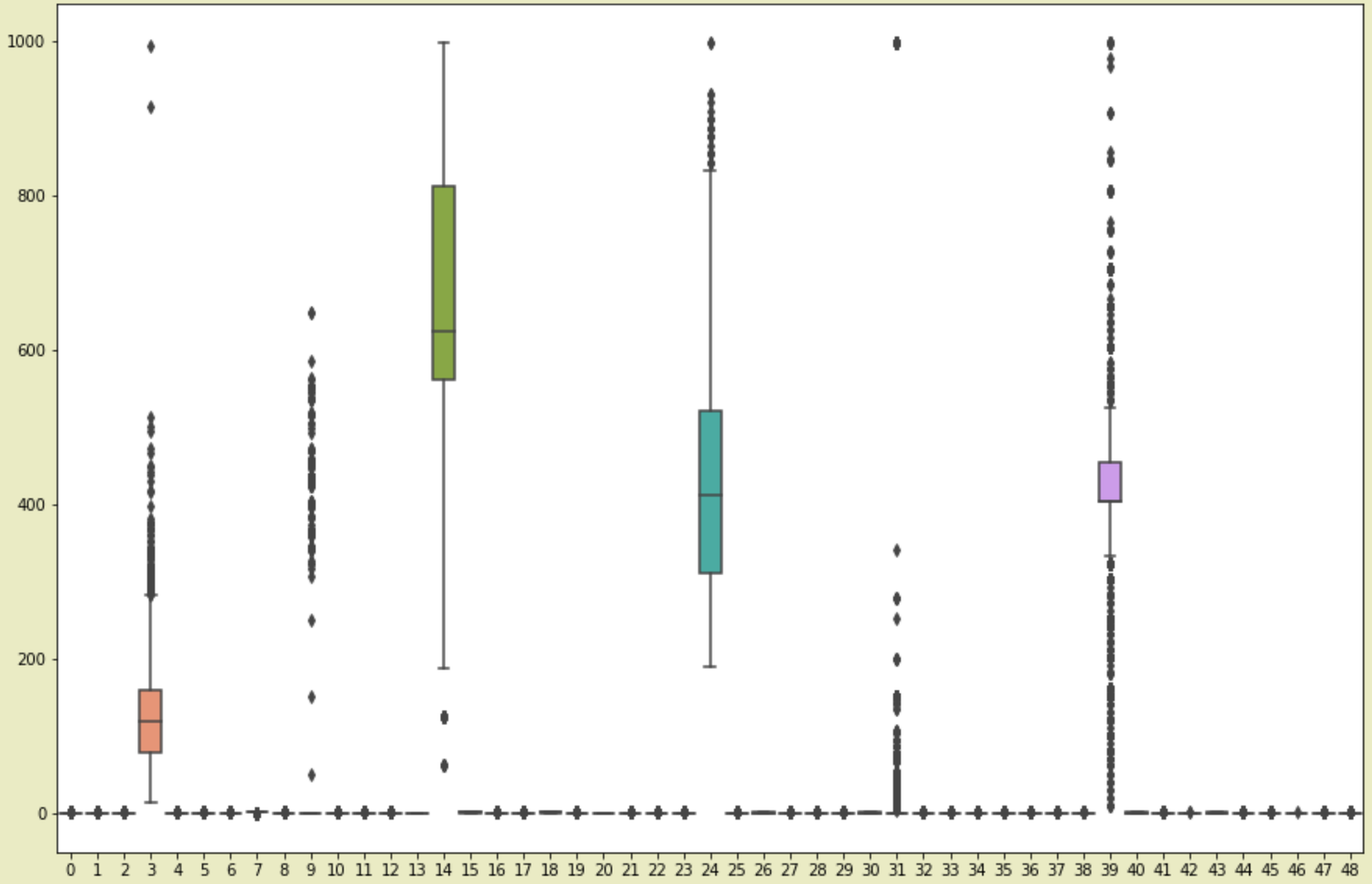
	0	1	2	3	4	5	6	7	8	9	...	40	41	42	43	44	45	46	47	48	49
0	0	1	1	32	0	0	0	1	0	0	...	1	0	0	0	0	0	0	0	NaN	-1
1	0	0	1	133	0	0	1	0	0	0	...	0	0	0	0	0	0	0	0	NaN	-1
2	0	0	0	109	0	0	0	1	0	0	...	1	0	0	0	0	0	0	0	NaN	-1
3	0	0	0	113	0	0	0	1	0	0	...	0	0	0	1	0	1	0	0	NaN	1
4	0	0	0	120	0	0	0	1	0	0	...	0	1	0	0	0	0	0	0	NaN	-1

Summary

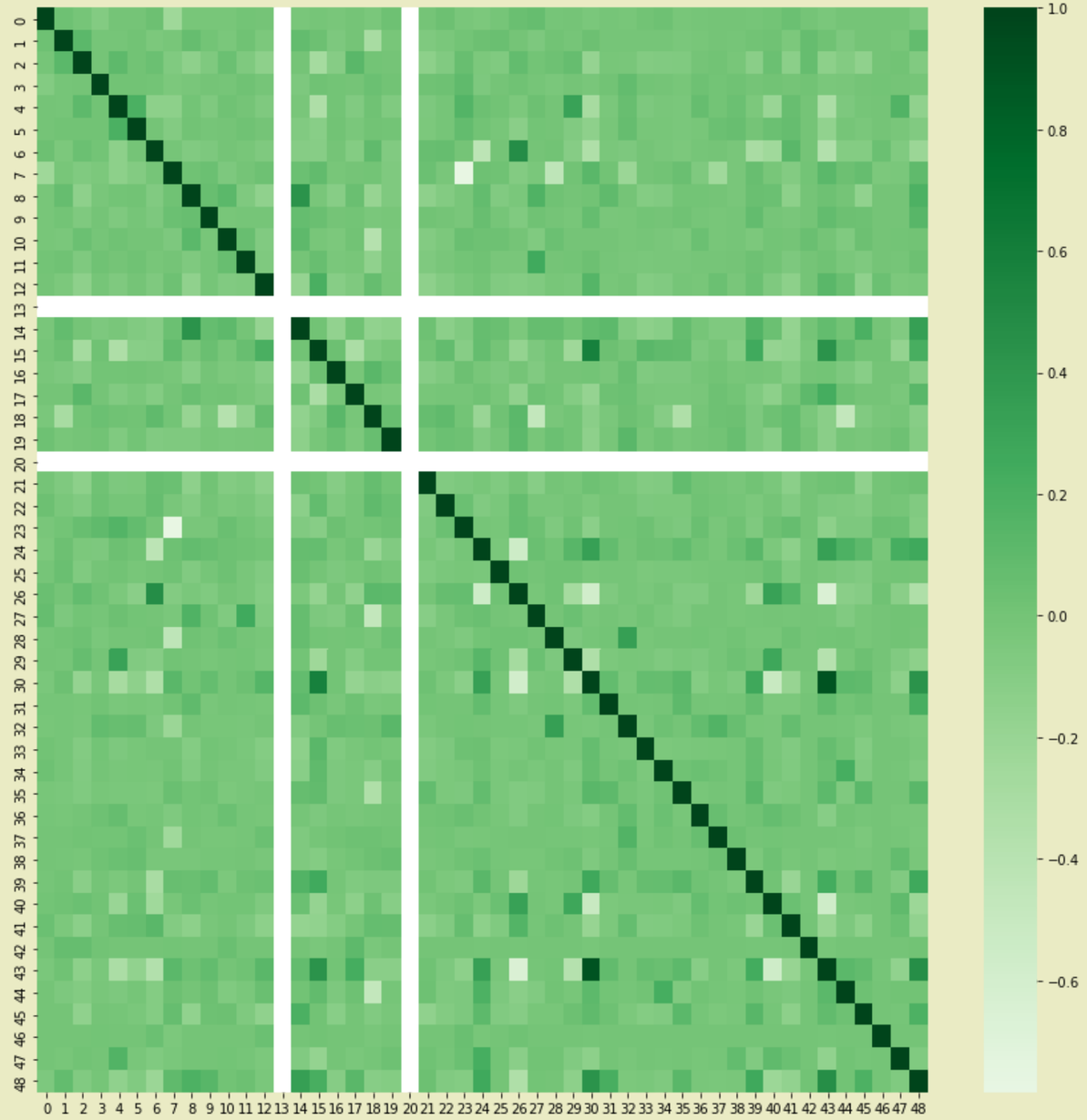


- Training and validation datasets contain 4147 rows and 50 columns.
- Columns with no obvious information
- Numerical data
- One empty column, few duplicate rows

Outlier Removal

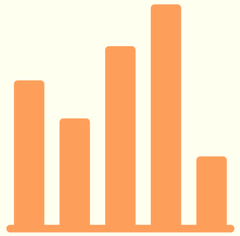


Correlation Matrix



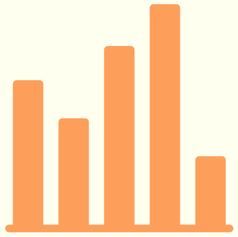
Models

KNN



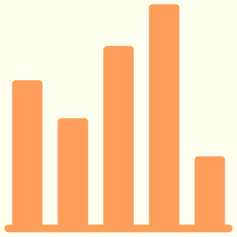
Accuracy: 79%
Precision: 83%
Recall: 93%

LDA



Accuracy: 84%
Precision: 88%
Recall: 90%

REG

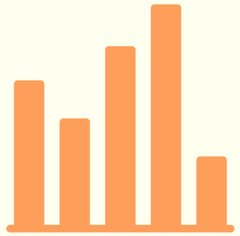


Accuracy: 83%
Precision: 87%
Recall: 91%

For class 1 with value -1

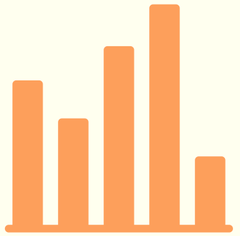
Models

Gini



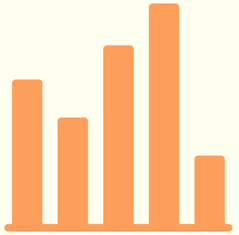
Accuracy: 82%
Precision: 86%
Recall: 89%

Entropy



Accuracy: 83%
Precision: 84%
Recall: 97%

Naive Bayes



Accuracy: 77%
Precision: 95%
Recall: 75%

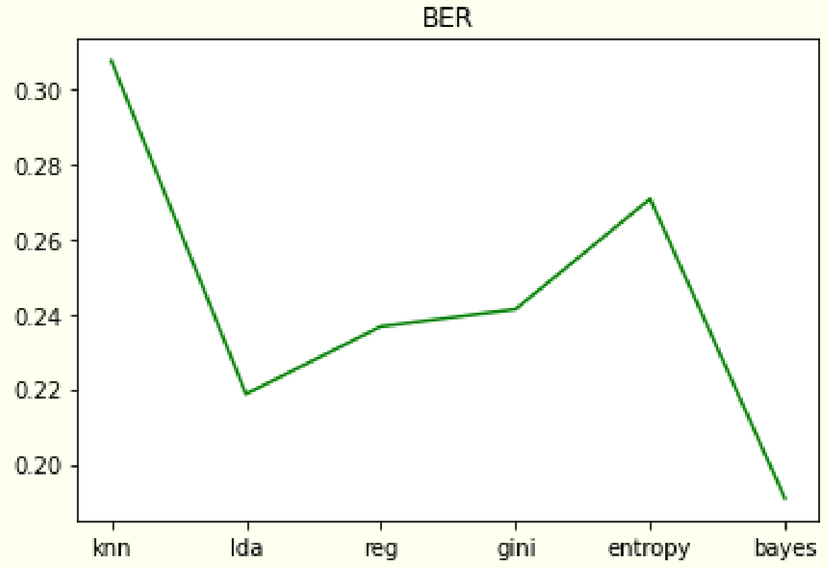
For class 1 with value -1

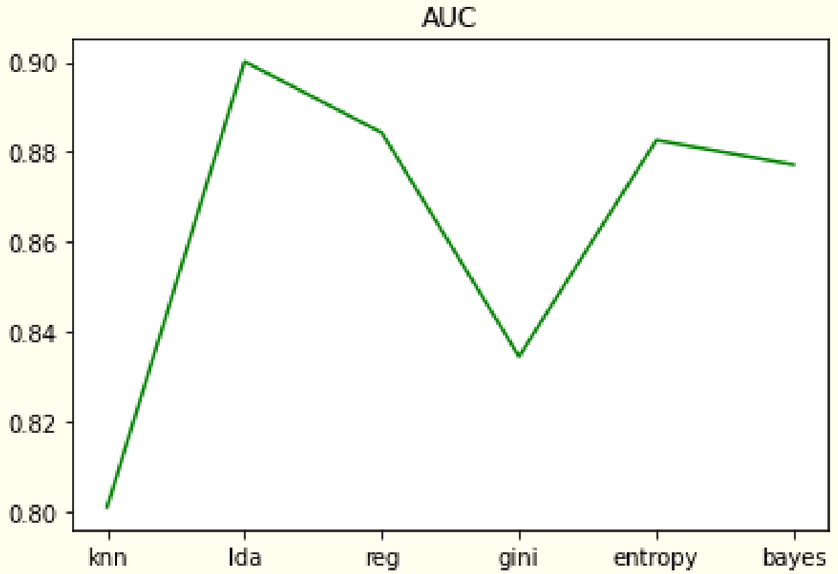
Evaluation

BER	BER Guess Error	AUC	Test Score
• KNN 0.3076 • LDA:0.2187 • REG: 0.2367 • gini: 0.2413 • entropy: 0.2708 • bayes: 0.1909 ★	• KNN 0.0048 ★ • LDA: 0.0127 • REG: 0.0105 • gini: 0.016 • entropy: 0.0152 • bayes: 0.0139	• KNN: 0.801 • LDA: 0.900 ★ • REG: 0.884 • gini: 0.835 • entropy: 0.883 • bayes: 0.877	• KNN: 0.31 • LDA: 0.23 • REG: 0.25 • gini: 0.26 • entropy: 0.29 • bayes: 0.2 ★

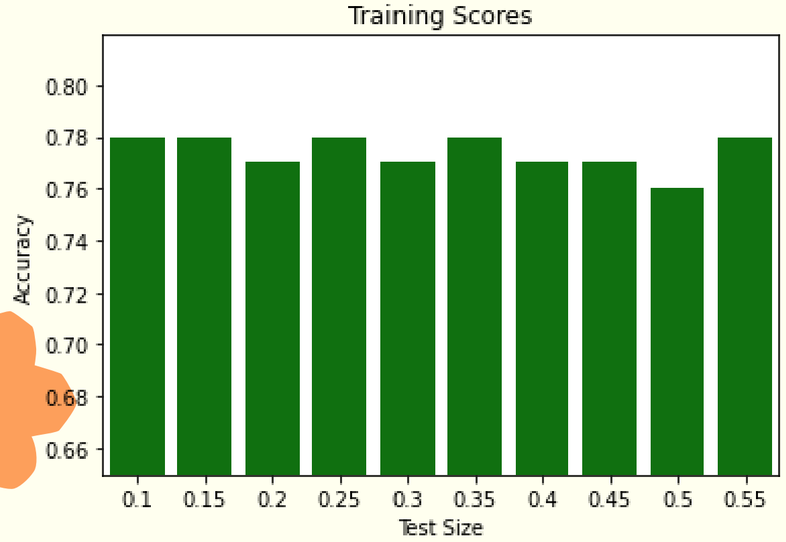
Best Model

(so far)

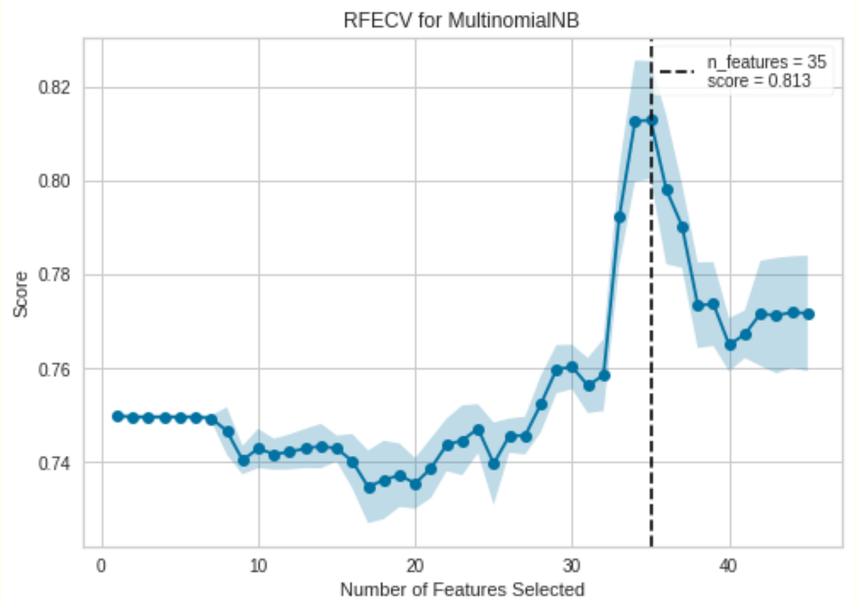


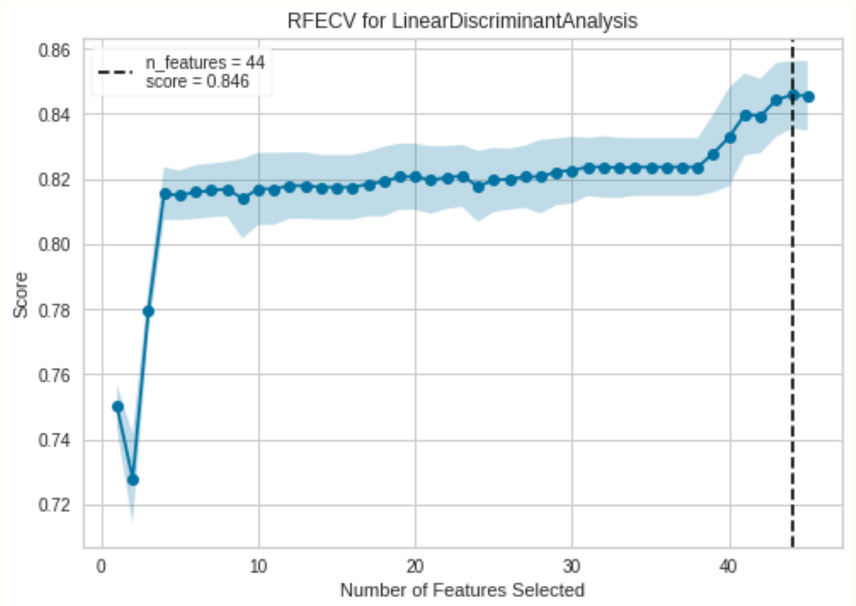


Why?



- easy linerarly separable
- easy to calculate probabilities
- hard to tell which variables are useful





Thank you!