CreditOne Loan Default Analysis

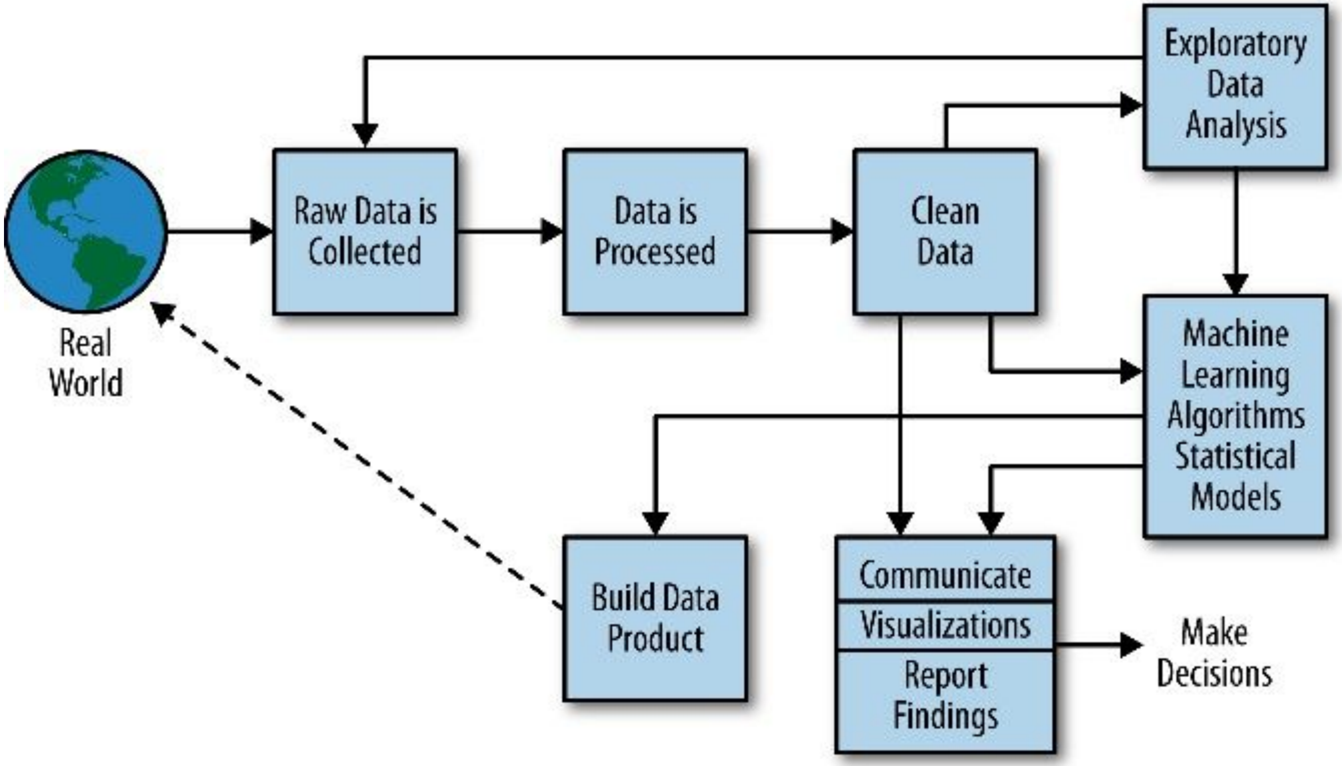
Presented by Don Bice

Credit ONE

Agenda

- Business Problem
- Data
- Insights
- Conclusions

Data Science Process



Business Problem

An increase in customer default rates is bad for Credit One since our business customers rely on us to score consumers for loans and an increase may result in the loss of Credit One's business customers.

Questions to Investigate:

- How do we ensure that customers can/will pay their loans?
- Can we approve customers with a high degree of certainty?

Facts Bearing on the Problem

- We cannot control customer spending habits

- We cannot always go from what we find in our analysis to the underlying "why"

- We must focus on the problems we can solve:
    - Which attributes in the data can we deem to be statistically significant to the problem at hand?
    - What concrete information can we derive from the data we have?
    - What proven methods can we use to uncover more information and why?

Data

30k records of consumer loans w/ demographic data and 6 month payment/billing history and whether defaulted

Attributes:

X1: Amount of the given credit

X2: Gender

X3: Education

X4: Marital status

X5: Age

X6-X11: History of past payment

X12-X17: Amount of bill statement

X18-X23: Amount of previous payment

Y: Default or no default

Source:
https://s3.amazonaws.com/gbstool/courses/910/docs/DataSourceUpdated5.18.pdf?AWSAccessKeyId=AKIAJBIZLMJQ2O6DKIAA&Expires=1539680400&Signature=%2FQ4%2FwuWGKyIAJC8V9TPOxh8rjSQ%3D

Most Important Features

Which attributes are most important in predicting whether a customer defaults?

X6-X11: History of past payment

Gender: 60% female

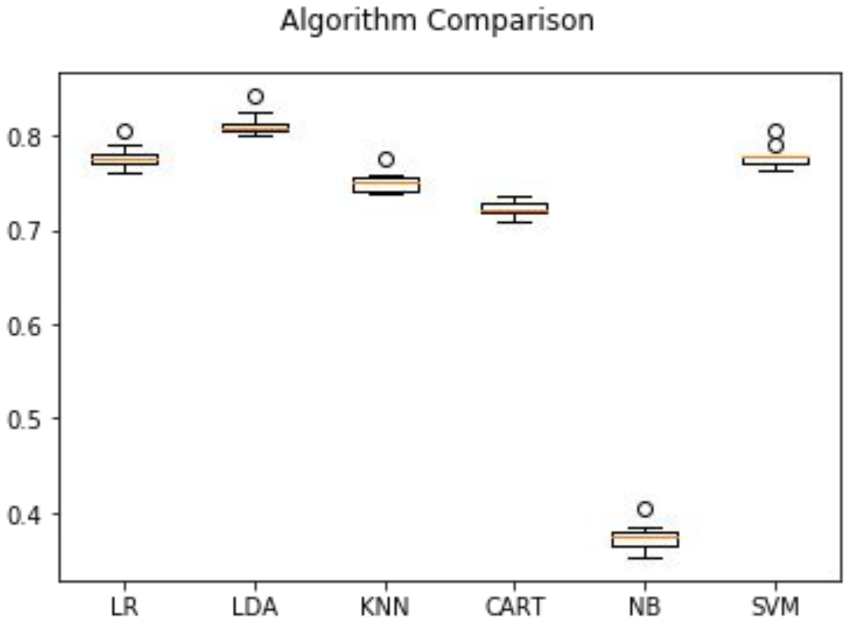
Education: Over 80% have college degree or higher

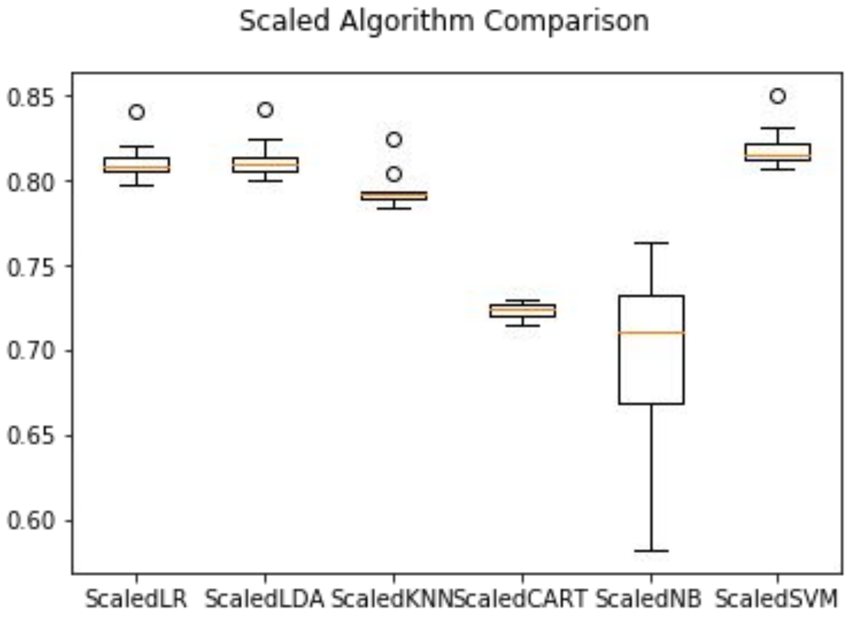
Marital status: 46% married

Age: Mean age is 49 years old

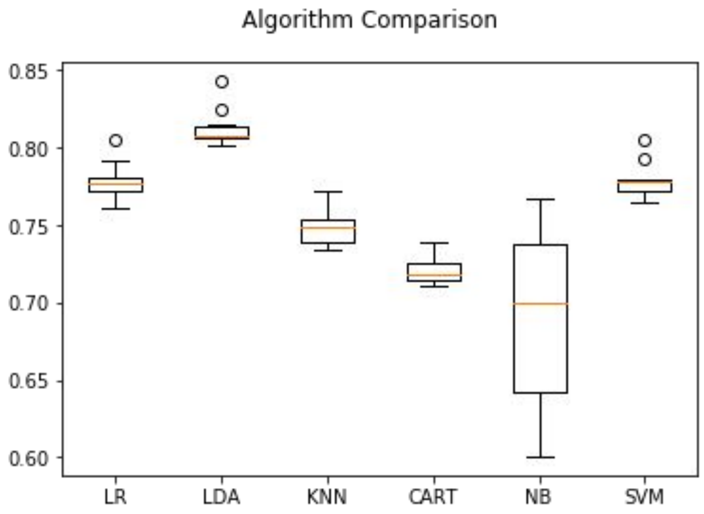
Mean number of consumers in the dataset who defaulted on payment in the next month is 22.1%.

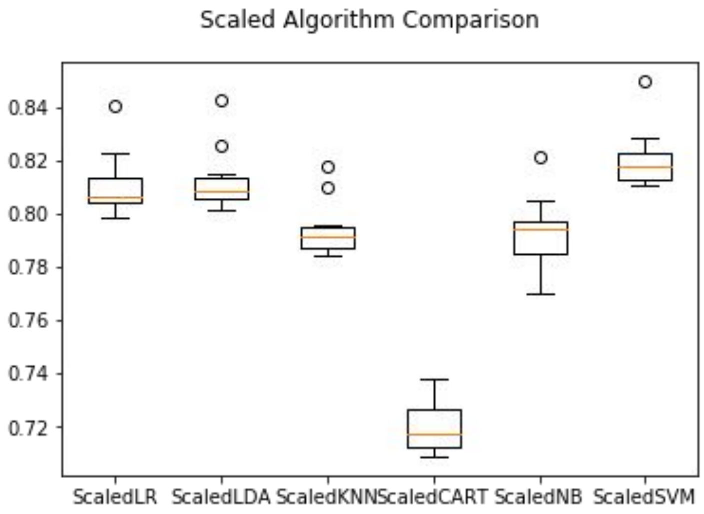
Evaluate Algorithms (All Features)





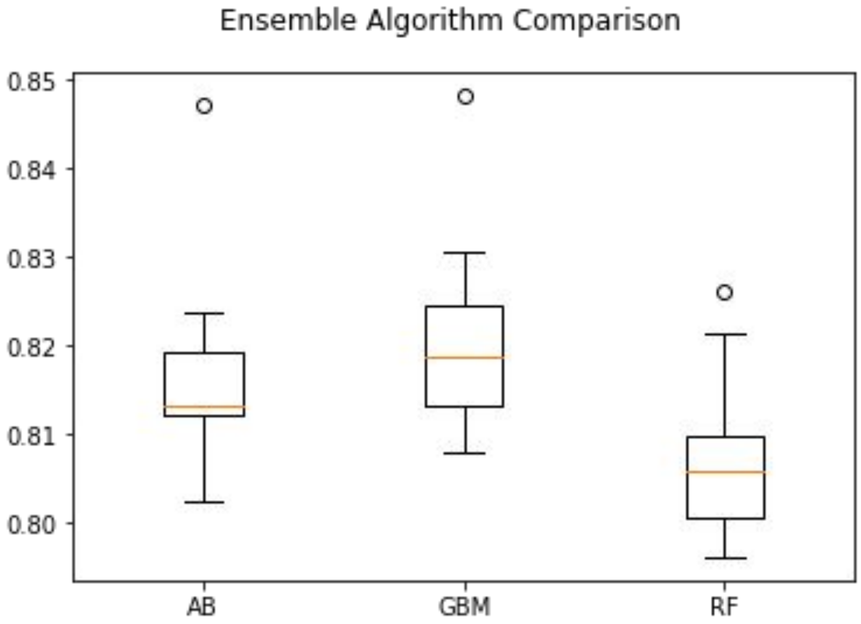
Evaluate Algorithms w/ Feature Reduction



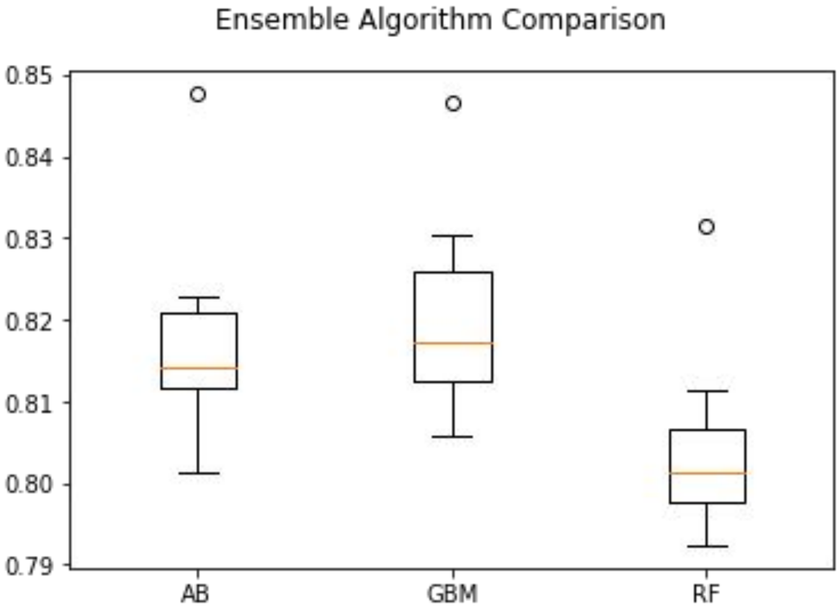


Evaluate Ensemble Methods

All Features



Feature Reduction



Observations

Mean number of consumers in the dataset who defaulted on payment in the next month is 22.1%. Of 6 algorithms and 3 ensemble methods, mostly performed not much better than chance.

At issue are 2 unbalanced classes with a ratio of 4:1 (that is, a model could just predict ‘client will not default’ always and have 78% accuracy).

After feature reduction, scaling the data and tuning models, best models only predicted ‘default’ with slightly better results.

Accuracy metric alone does not indicate which model will predict ‘default’. We need to consider Kappa and other metrics as well.

Predictions on Validation Dataset

Top 3 Performers of 9 algorithms evaluated

	Accuracy	Kappa
GradientBoostingMachines	0.820	0.374
LinearDiscriminantAnalysis	0.810	0.289
SupportVectorMachines	0.813	0.344

Conclusions

Based on the data on hand, we cannot ensure that customers can & will pay their loans, though predictions are over 30% better than chance.

We also cannot approve customers with the same certainty we have in the past due to skyrocketing defaults.

We need to greatly constrain limits on lending approvals and advise our banking clients to increase their provisions for loan losses.

CreditOne