

Bank Marketing Campaign Study

A Data-Driven Approach to Predict the Success of Bank
Telemarketing

The Business Aspect

Company

Banco Português de Investimento

- A Portuguese commercial bank
- Core services include providing term deposit

Context

From 2008-10,
Information is collected
from three main aspects

- Client
- Telemarketing
- Social and economic

Problem statement

Understand attributes'
meaning in a business
context

Predict if the client will
subscribe to a bank term
deposit

Maximize potential
business values

The Modeling Aspect

Data Dimension

Training set

- 4119 x 21

Test set

- 41188 x 21

Attributes

Client data

- Job, age, education

Telemarketing data

- Pdays, previous

Social/economic data

- emp.var.rate

Modeling

Exploratory analysis

Tree-based and
Kernel-based Modeling
and their Learning curve

Cost analysis

Exploratory Analysis

Understand attributes' meaning
in a business context

What defines a dimension?

What is the relationship between
dimensions?

How do we use these dimensions
to interpret the current business
situation?

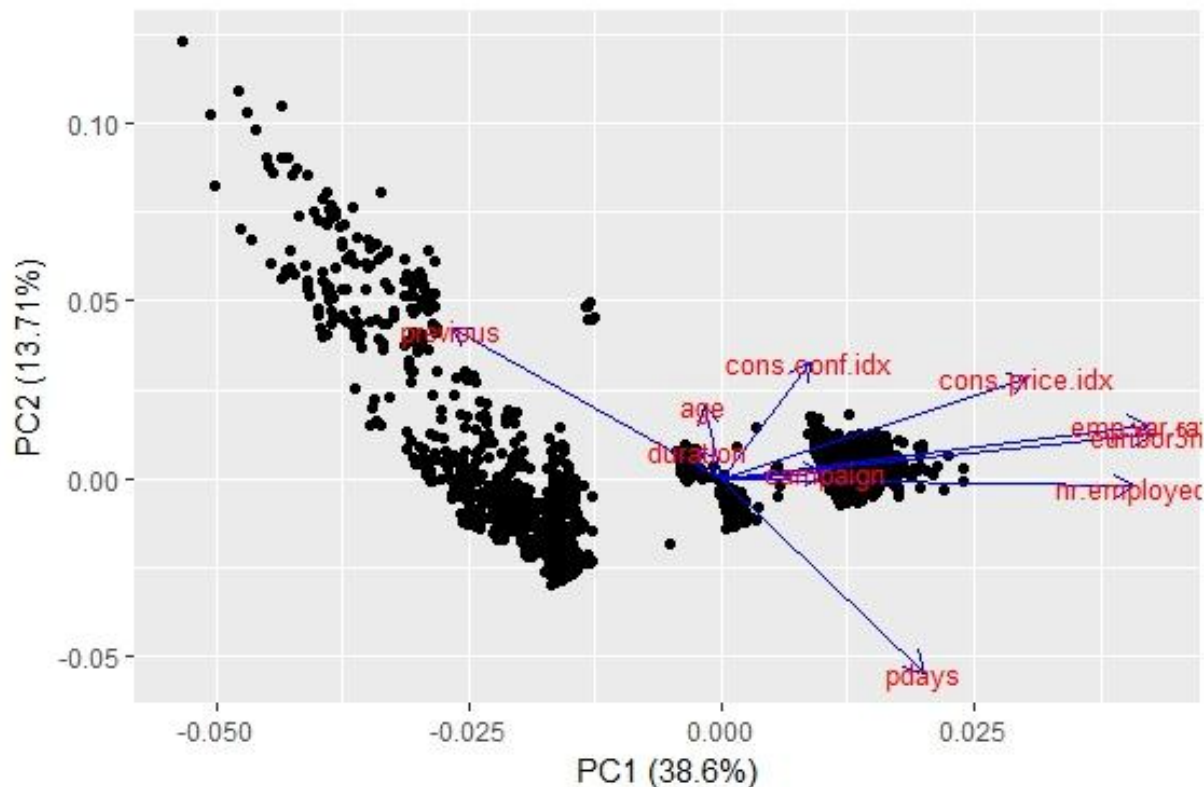
Principal Component Analysis

Notable variables from PC1

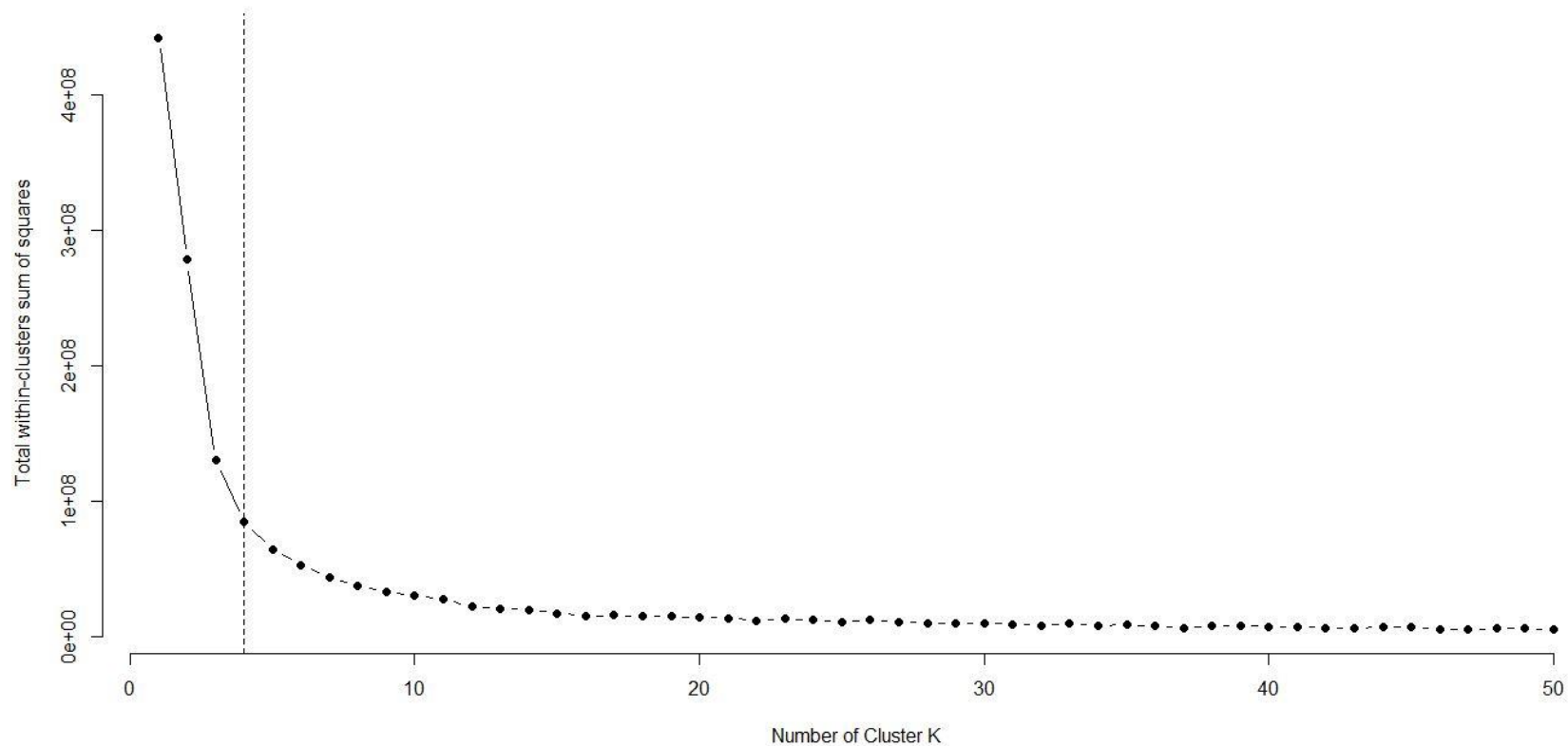
- Emp.var.rate
- Euribor3m
- Nr.employed
- Previous

Notable variables from PC2

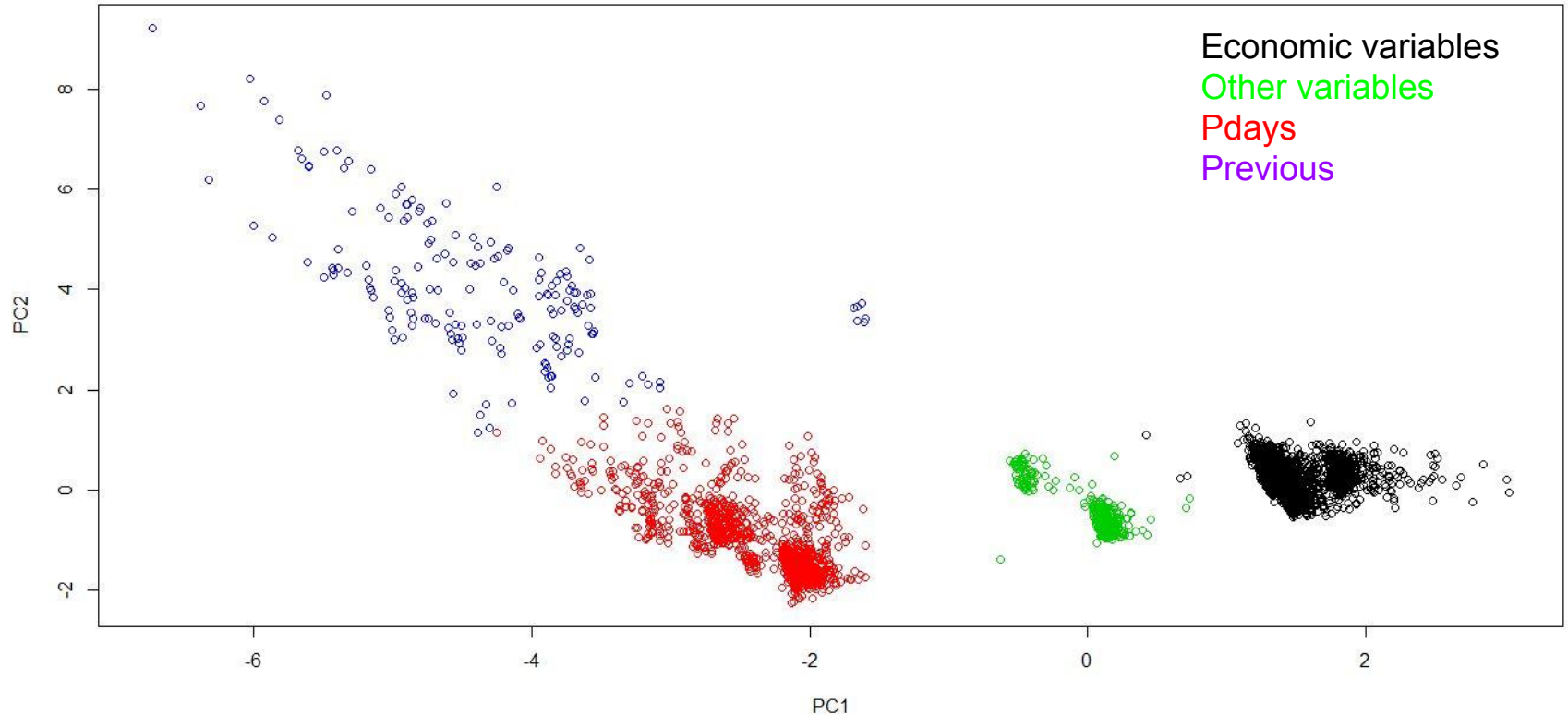
- Previous
- pdays



K-Means on PCA



K-Means on PCA



Modeling

Predict if the client will subscribe a bank term deposit

How to identify overfitting?

What are the learning curves?

Tree-Based Modeling

- Cross-validated error rate is used for both tuning parameters and accessing model accuracy
- Class ratio within the response variable
 - yes/no: 12.29%
- Parameters
 - Classification decision tree: minsplit, minbucket
 - Random forest: mtry
- Results
 - Classification decision tree:

	elapsed	
198.34000000	0.09328464	
 - Random forest:

	elapsed	
461.26000000	0.08760639	

Kernel-Based Modeling

- Parameters
 - SVC: cost
 - SVM - Polynomial: degree, cost, gamma
 - SVM - Radial: cost, gamma

- Results

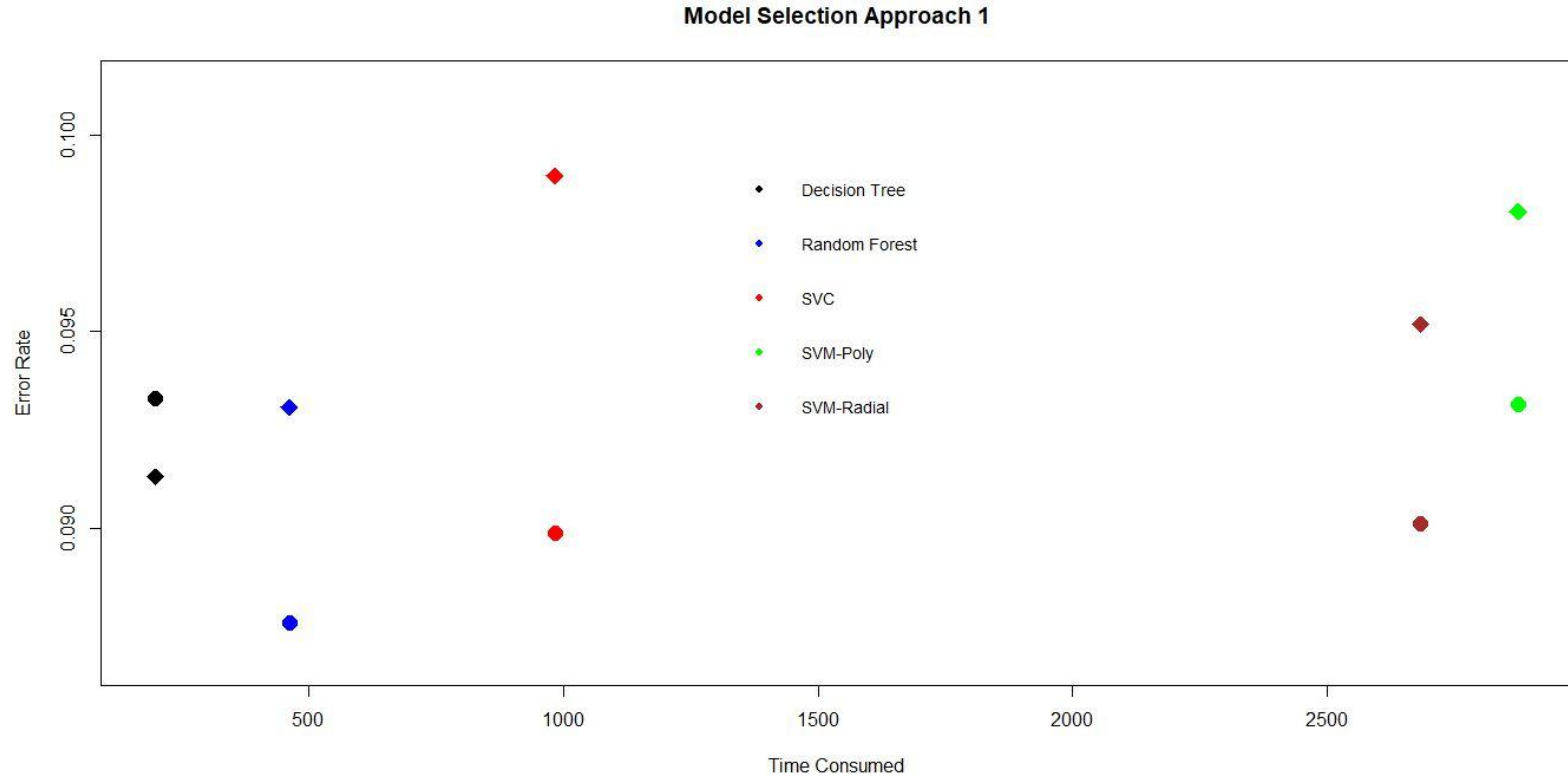
- SVC:

elapsed
983.3000000 0.08988952
- SVM - Polynomial:

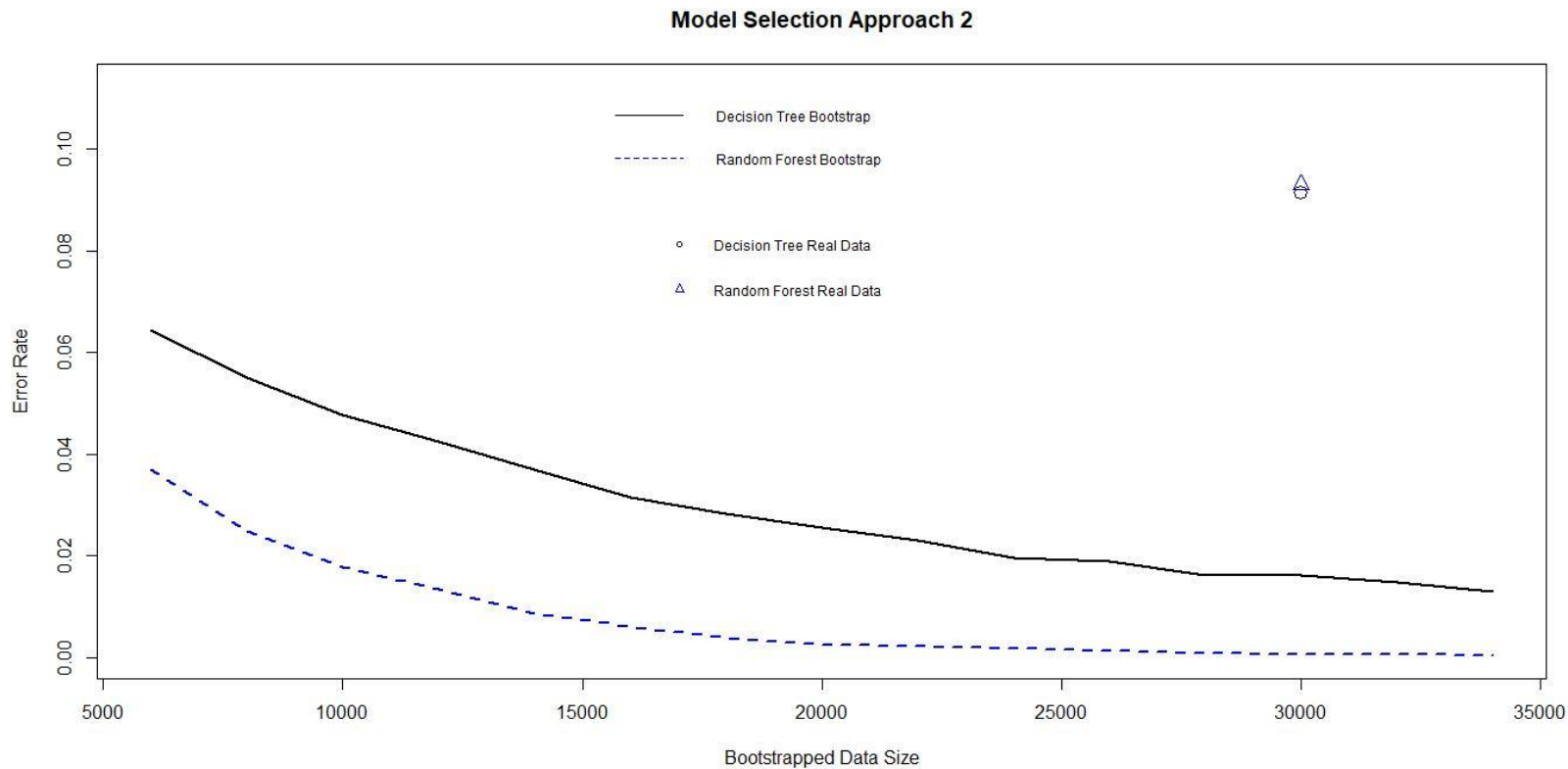
elapsed
2876.2200000 0.09316021
- SVM - Radial:

elapsed
2683.8800000 0.09011496

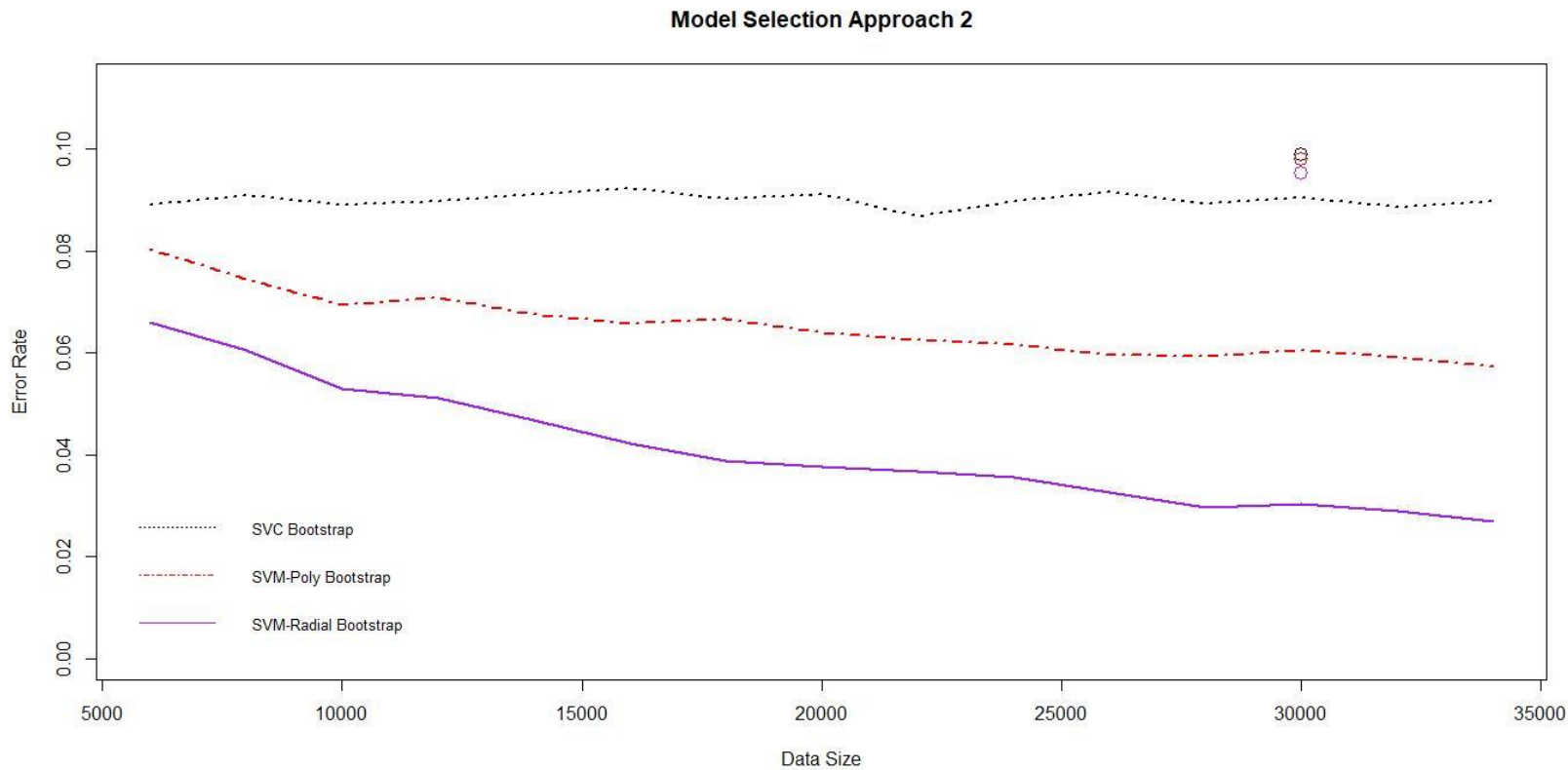
Modeling Results



Potential Overfitting Issue



Potential Overfitting Issue



Final Modeling Results

- Classification decision tree

TP	0.33434015
TN	0.97271516
FP	0.02728484
FN	0.66565985
Err	0.09479637

- SVM - Polynomial

TP	0.1977323
TN	0.9880065
FP	0.0119935
FN	0.8022677
Err	0.1024857

Cost Analysis

Maximize potential business
values from the model

What assumptions are made?

Assumptions

- The current campaign is to promote a specific type of certificate deposit.
- The bank promises to return 2% of annual cumulative amount to buyers. The bank itself will earn 5%.
- 40% of all clients are term deposits subscribers. While all clients are likely to purchase such certificate deposit, term deposit clients are the targeted customers for this campaign.
- When a normal client became a term deposit client for the first time, we promote our certificate deposit to them. The probability of them accepting our promotion is 30%.

Assumptions

- If the client did not buy the promoted product, we will follow up to these clients in two weeks to promote the product once again. There is a 30% chance for the client to buy the promoted product.
- The probabilities of a client buying the promoted product after the first contact is independent from the probabilities of buying the product after the second contact.
- There is 27% of chance for the clients to get irritated and cancel the term deposit subscription.
- The certificate deposit requires clients to deposit 2000 USD each month.

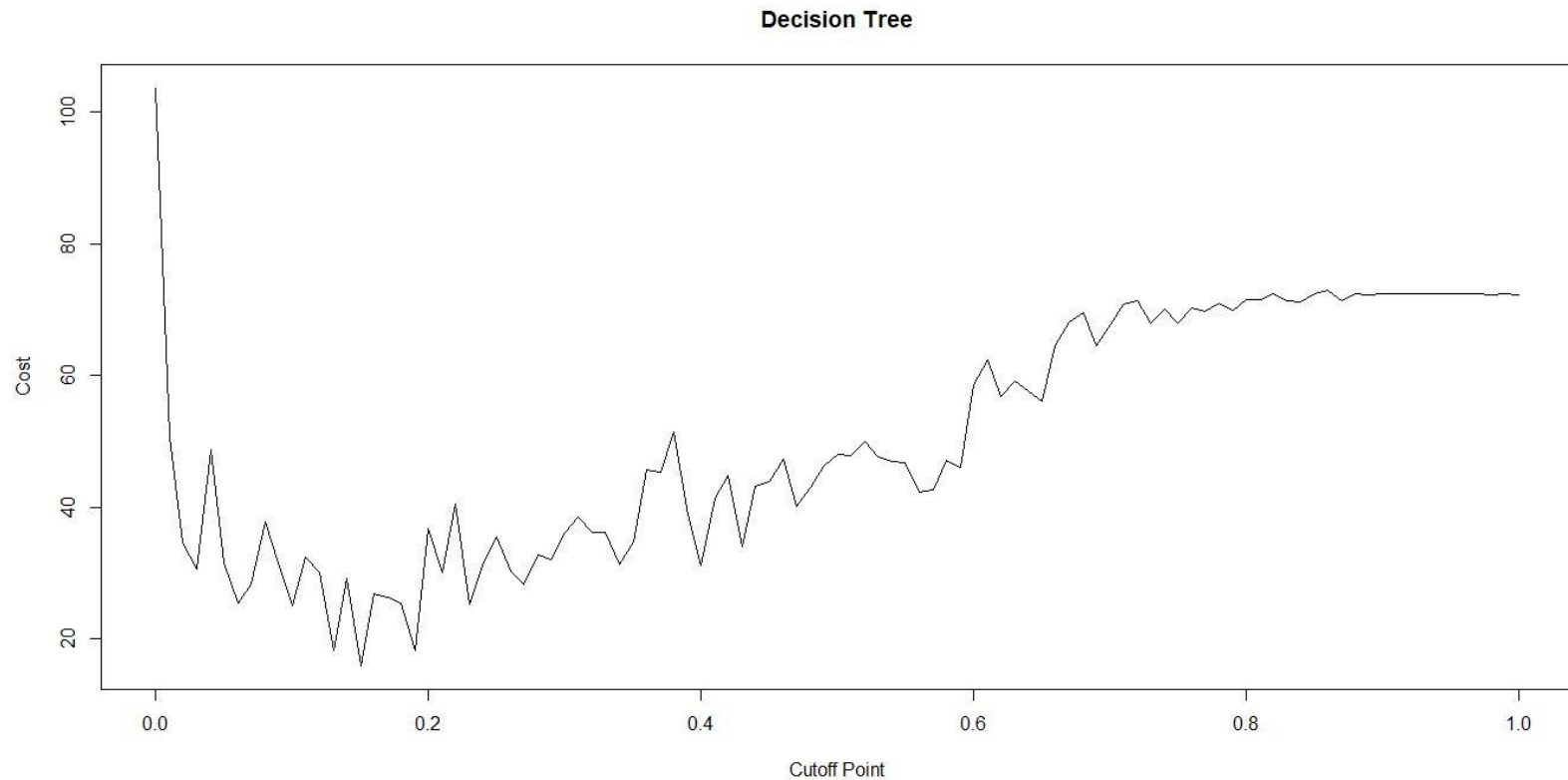
Cost Calculations

- 1 year cumulative amount: $\$2,000 * 12 = \$24,000$
 - Bank's earning based on 5%: $\$24,000 * 5\% = \$1,200$
 - Client's earning based on 2%: $\$24,000 * 2\% = \480
-
- 2nd cycle telemarketing cost: $-\$20$
 - Term deposit cancellation cost: $-\$20 - (\$1,200 * 27\%) = -\$344$
 - Opportunity cost for not promoting: $-\$50$

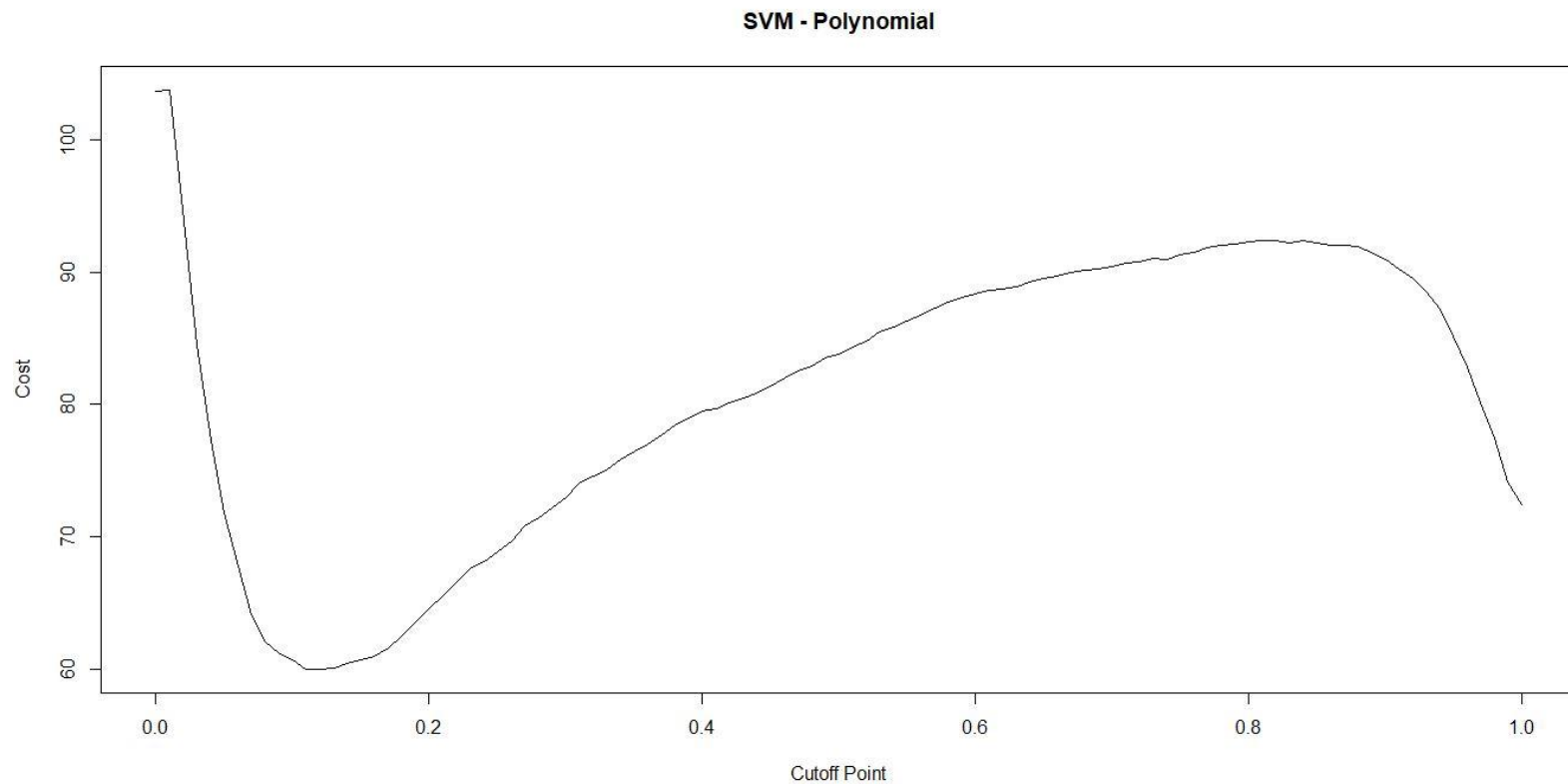
Comprehensive Probability Combinations

TD?	Offer?	Accept?	Probability	Net Cost
Y	Y	Y	$P.\text{yes} * P.\text{buy} * TP$	\$700
Y	Y	N	$P.\text{yes} * P.\text{still} * TP$	-\$344
Y	N	Y	$P.\text{yes} * P.\text{buy} * \text{Type 2}$	\$720
Y	N	N	$P.\text{yes} * P.\text{still} * \text{Type 2}$	-\$50
N	Y	Y	$P.\text{no} * P.\text{buy} * \text{Type 1}$	\$700
N	Y	N	$P.\text{no} * P.\text{still} * \text{Type 1}$	-\$20
N	N	Y	0%	
N	N	N		\$0

Results



Results



Comparison of two models

Decision Tree

This model has more variability, but significantly lower cost.

Option for larger banks with more room for risk

SVM-Polynomial

Higher costs, but more predictable.

More stable cost planning for banks