

A close-up photograph of a hand holding a black and gold pen over a white form titled "PETITION TO FILE FOR BANKRUPTCY". The form is on a wooden surface. The text "PETITION TO FILE FOR BANKRUPTCY" is printed in large, bold, black letters. Below the title, there are several lines of text, including "Creditors holding secured claims", "(First)", and "(City)". The hand is positioned over the form, ready to write.

PETITION TO FILE FOR BANKRUPTCY

Can a VC predict if a start up will become bankrupt ?

Group A2

Agenda

- Data Preprocessing
- Methods Used
 - Classification Trees - Model Rejected
 - Naïve Bayes Classification - Used for Prediction
 - K-NN - Used for Prediction
 - Logistic Regression - Used for Prediction
 - Discriminant Analysis – Used for Prediction
- Insights

Data Preprocessing

- Over 25 years of data from 250,000 PE funding
- Missing Data Handling
 - Delete Blank rows from the data
 - Replacement of blank cells by appropriate values
- Reducing the number of columns
 - predominantly qualitative data , PCA not used
- Data Partitioning

Classification Trees

Training Data

Classification Confusion Matrix		
	Predicted Class	
Actual Class	No	Yes
No	5254	1
Yes	14	33

Error Report			
Class	# Cases	# Errors	% Error
No	5255	1	0.02
Yes	47	14	29.79
Overall	5302	15	0.28

Validation Data

Classification Confusion Matrix		
	Predicted Class	
Actual Class	No	Yes
No	3148	2
Yes	22	10

Error Report			
Class	# Cases	# Errors	% Error
No	3150	2	0.06
Yes	32	22	68.75
Overall	3182	24	0.75

Test Data

Classification Confusion Matrix		
	Predicted Class	
Actual Class	No	Yes
No	2107	4
Yes	6	4

Error Report			
Class	# Cases	# Errors	% Error
No	2111	4	0.19
Yes	10	6	60.00
Overall	2121	10	0.47

Input Variables not continuous , binary variables ; NOT ACCURATE

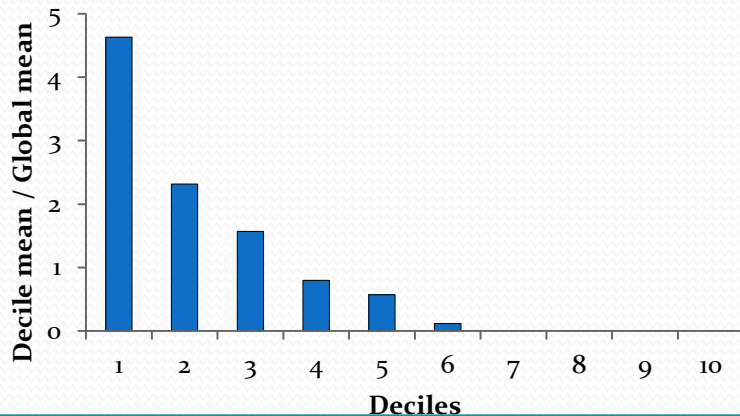
Naïve Bayes Classification

Cut off Prob.Val. for Success (Updatable)	0.85	(Updating the value here will NOT update value in detailed report)
---	-------------	--

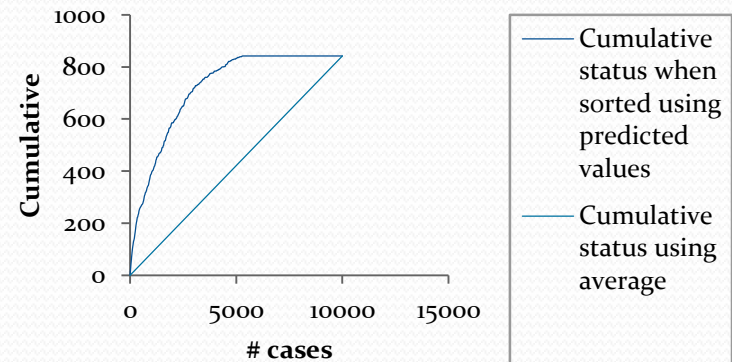
Classification Confusion Matrix		
	Predicted Class	
Actual Class	Yes	No
Yes	498	344
No	1069	8089

Error Report			
Class	# Cases	# Errors	% Error
Yes	842	344	40.86
No	9158	1069	11.67
Overall	10000	1413	14.13

Decile-wise lift chart (training dataset)



Lift chart (training dataset)



With an error of about 14 %, a PE fund will be able to identify if the investment could turn defunct.

Classification K-Nearest Neighbours

We basically had to set two variables to finalize the model

1. K
2. Cut off probability

Cut off Probability

First performed the KNN with the default value of cut off prob = 0.5

Validation Data scoring - Summary Report (for k=4)

Cut off Prob.Val. for Success (Updatable)	0.5
---	------------

Classification Confusion Matrix		
	Predicted Class	
Actual Class	1	0
1	0	32
0	0	3150

Error Report			
Class	# Cases	# Errors	% Error
1	32	32	100.00
0	3150	0	0.00
Overall	3182	32	1.01

Test Data scoring - Summary Report (for k=4)

Cut off Prob.Val. for Success (Updatable)	0.5
---	------------

Classification Confusion Matrix		
	Predicted Class	
Actual Class	1	0
1	0	10
0	0	2111

Error Report			
Class	# Cases	# Errors	% Error
1	10	10	100.00
0	2111	0	0.00
Overall	2121	10	0.47

We observe that the error in identifying bankruptcies was 100%. Also from the confusion matrix we observe that none of the actual bankruptcies were predicted

Classification K-Nearest Neighbours

Cut off Probability

The Scoring report suggested prob of success(bankruptcy) was for most cases

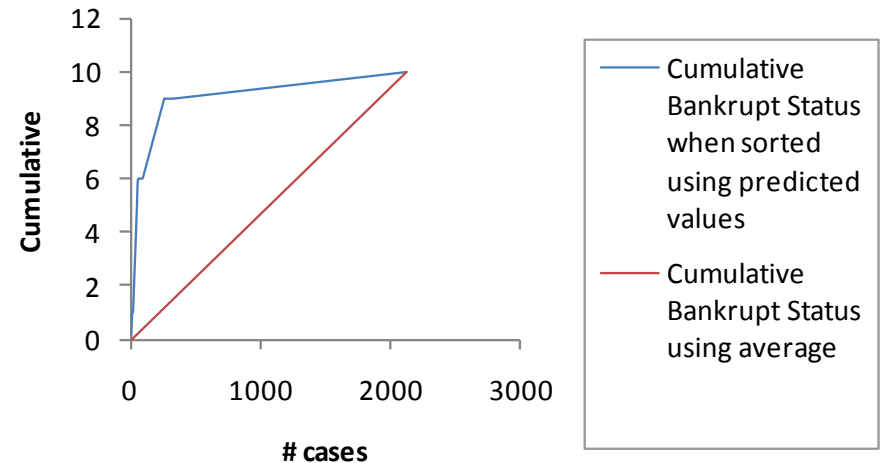
Test Data scoring - Summary Report (for k=4)

Cut off Prob.Val. for Success (Updatable)	0.03
---	-------------

Classification Confusion Matrix		
	Predicted Class	
Actual Class	1	0
1	9	1
0	255	1856

Error Report			
Class	# Cases	# Errors	% Error
1	10	1	10.00
0	2111	255	12.08
Overall	2121	256	12.07

Lift chart (test dataset)



We observed the phenomenal decrease in the error rate and from the confusion matrix of the test data we observe that we are able to correctly predict 9/10 times in case of bankruptcies

Classification K-Nearest Neighbours

Determination of optimum value of k

We run the KNN algorithm with different values of $k = 3, 4, 5$

Test Data scoring - Summary Report (for k=3)

Cut off Prob.Val. for Success (Updatable) **0.03**

Classification Confusion Matrix		
	Predicted Class	
Actual Class	1	0
1	9	1
0	269	1842

Error Report			
Class	# Cases	# Errors	% Error
1	10	1	10.00
0	2111	269	12.74
Overall	2121	270	12.73

Test Data scoring - Summary Report (for k=5)

Cut off Prob.Val. for Success (Updatable) **0.03**

Classification Confusion Matrix		
	Predicted Class	
Actual Class	1	0
1	9	1
0	259	1852

Error Report			
Class	# Cases	# Errors	% Error
1	10	1	10.00
0	2111	259	12.27
Overall	2121	260	12.26

Test Data scoring - Summary Report (for k=4)

Cut off Prob.Val. for Success (Updatable) **0.03**

Classification Confusion Matrix		
	Predicted Class	
Actual Class	1	0
1	9	1
0	255	1856

Error Report			
Class	# Cases	# Errors	% Error
1	10	1	10.00
0	2111	255	12.08
Overall	2121	256	12.07

We observe that lowest overall error is for $k=4$
Hence the final model is with $k=4$ and
Cutoff prob = 0.03

Logistic Regression

	Test Data	Training Data	Validation Data
Error in Prediction	13.39	13.20	13.58

Input variables	Coefficient	Std. Error	p-value
Constant term	12.623849	1.5338316	0
pubstatus_Private	-4.399322	0.8566406	2.8E-07
pubstatus_Public	-2.7558889	0.8938866	0.002049
company_veic6a_Medical/Health/Life Science	2.7605255	0.7901464	0.0004764
firm_stage_pref_Balanced	-2.3108971	1.1445354	0.0434799
firm_stage_pref_Control-block Purchases	-4.9802408	1.3637654	0.0002604
firm_stage_pref_Distressed Debt	-4.2062979	1.5973855	0.0084575
firm_stage_pref_Later Stage	-3.1204503	1.2228576	0.0107178
firm_stage_pref_Seed	-3.1980567	1.1030098	0.0037389
Firmage	-0.1054764	0.0152176	0

Error Report for Test Data			
Class	# Cases	# Errors	% Error
No	2105	282	13.40
Yes	16	2	12.50
Overall	2121	284	13.39

Following factors are important in predicting whether the company that the VC firm decides to fund will go bankrupt:

- Company's current public Status
- Company VE Primary Industry Class
- Firm Preferred Investment Stage
- Age of Venture Capital firms

Linear Discriminant Analysis

Output Variable: Bankruptcy Status(Yes/No)
Cut off Probability : 0.98

Training Data Scoring:

Error Report			
Class	# Cases	# Errors	% Error
No	5258	321	6.10
Yes	43	6	13.95
Overall	5301	327	6.17

Validation Data Scoring:

Error Report			
Class	# Cases	# Errors	% Error
No	3151	202	6.41
Yes	30	4	13.33
Overall	3181	206	6.48

Test Data Scoring:

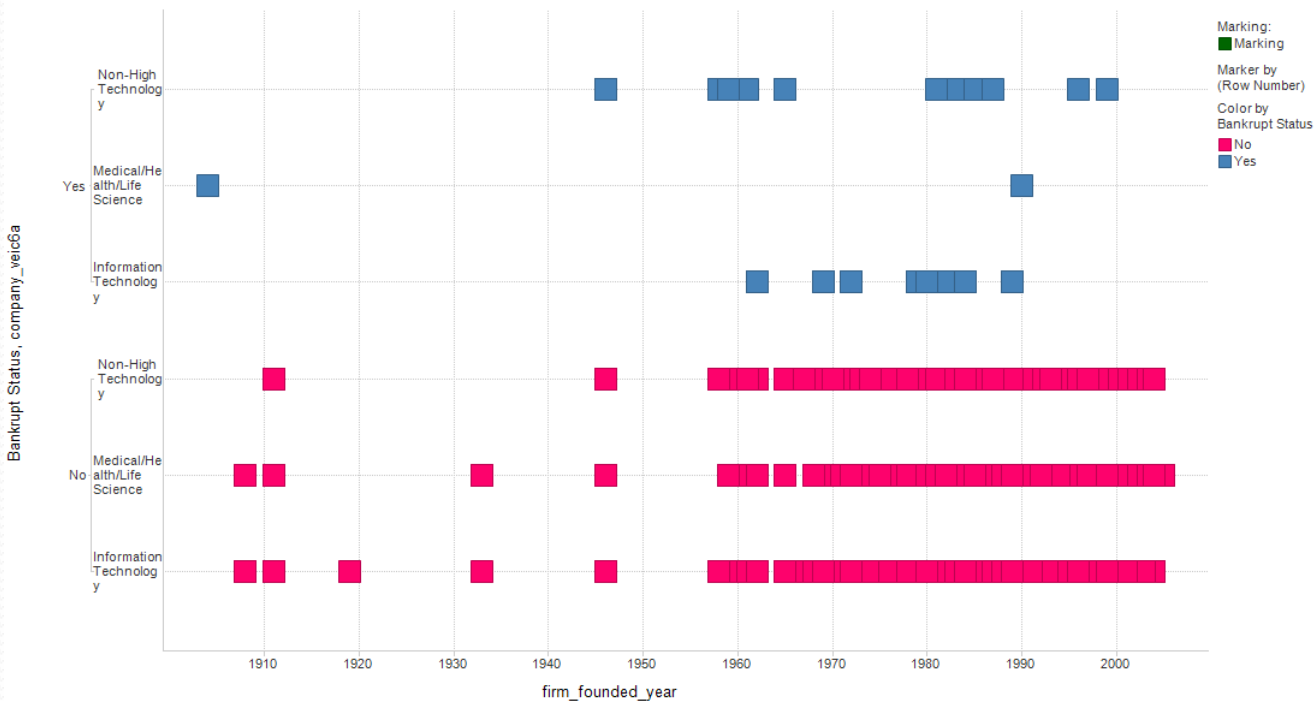
Error Report			
Class	# Cases	# Errors	% Error
No	2034	138	6.78
Yes	16	3	18.75
Overall	2120	141	6.65

Major factors determining whether a particular investment will result in Bankruptcy or not is dependent on mainly Firm_Stage_Pref and Firm_Type and Company VE Primary Industry Class

General Insights through Spotfire

- United States Companies in medical/health industry/life sciences industry have less chance of going bankrupt.
- Older VC firms because of the experience effect generally pick the correct firms to invest in.

Scatter Plot





Thank You