Bankruptcy Forecasting: A Predictive Analysis with SAS Enterprise Miner

Fall 2023 – MGMT 57100

Team: Data Jedi

Abhishek Krovvidi

Sai Teja Devalla

Sathwik Kanukuntla

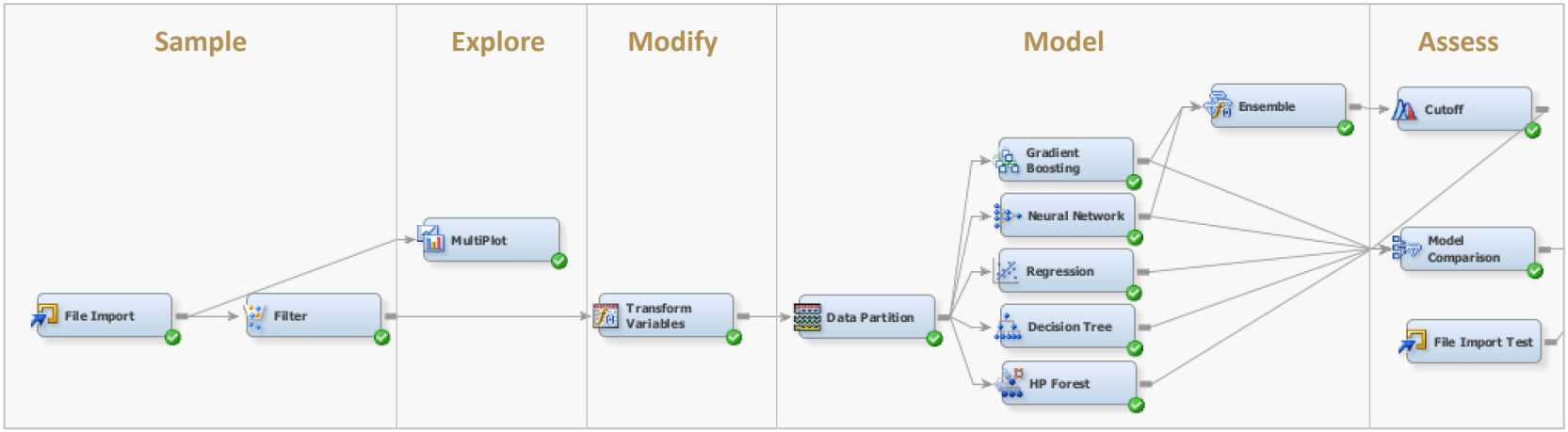
PURDUE UNIVERSITY

Problem Objective

Project Overview:	Introduction of a predictive model aimed at forecasting **bankruptcy** risks in companies.
Key Features:	Utilization of 64 financial indicators.
	Focus on profitability, liabilities, and asset management metrics.
Objective:	To offer a critical decision-making tool for stakeholders and financial analysts.
Impact:	Enhancing the identification of **potential financial distress**.
	Assessing **long-term viability** of companies in the current economic environment.

Our approach to evaluate different models

SEMMA



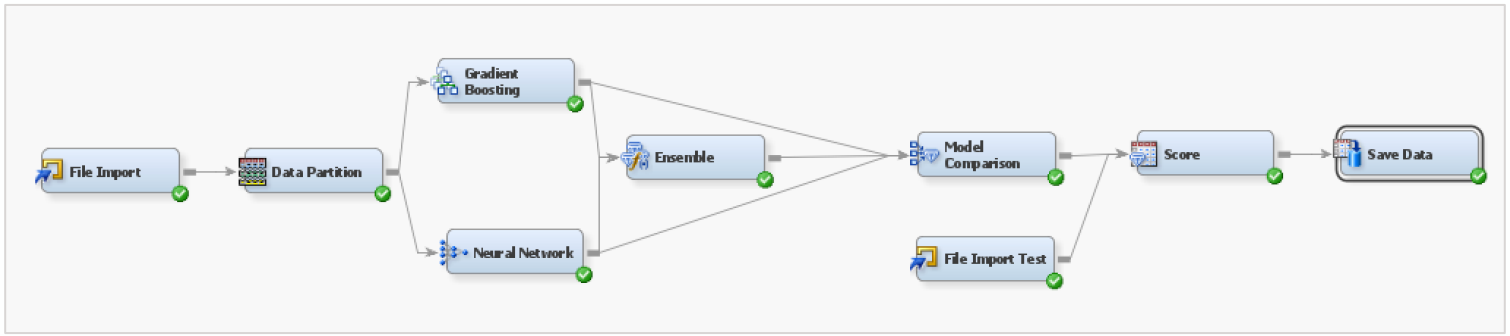
Fit Statistics

Selected Model	Model Node	Target Variable	Selection Criterion: Valid: Roc Index	Valid: Misclassification Rate	Valid: Cumulative Lift	Valid: Gini Coefficient
Y	Ensmbl	class	0.942	0.014845	7.211335	0.884
	Neural	class	0.935	0.016194	6.787139	0.871
	Reg	class	0.927	0.019343	7.635531	0.854
	Boost	class	0.915	0.015295	7.423433	0.83
	HPDMForest	class	0.849	0.021143	5.95053	0.697
	Tree	class	0.723	0.019343	5.155114	0.447

Modelling techniques

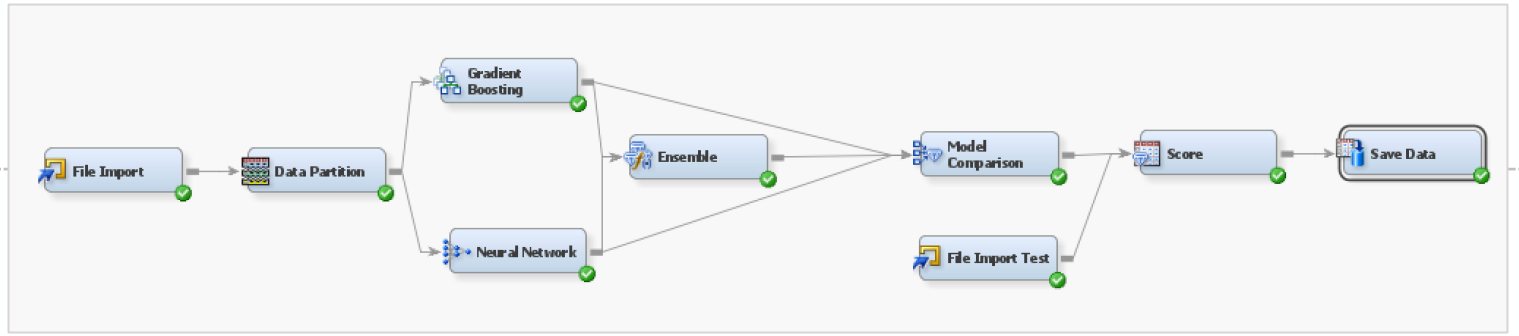
- Filtering outliers after exploring data through Multiplot
- Variable Transformation based on Skewness
- Model comparison of different models
- Tried Ensemble models with various combos
- Assessed through various Cutoff values
- Tuning model specific parameters across all the trials

What is the best model we identified?



"Ensemble of Gradient Boosting and Neural Network models is our best predictive model"

What is the best model we identified?



Gradient Boosting

Series Options	
N Iterations	183
Seed	12345
Shrinkage	0.05
Train Proportion	70
Splitting Rule	
Huber M-Regression	No
Maximum Branch	2
Maximum Depth	4
Minimum Categorical Size	5
Reuse Variable	1
Categorical Bins	30
Interval Bins	100
Missing Values	Use in search
Performance	Disk
Node	
Leaf Fraction	0.001
Number of Surrogate Rules	4
Split Size	20

Neural Network

Train	
Variables	
Continue Training	No
Network	
Optimization	
Initialization Seed	12345
Model Selection Criterion	Misclassification
Suppress Output	No

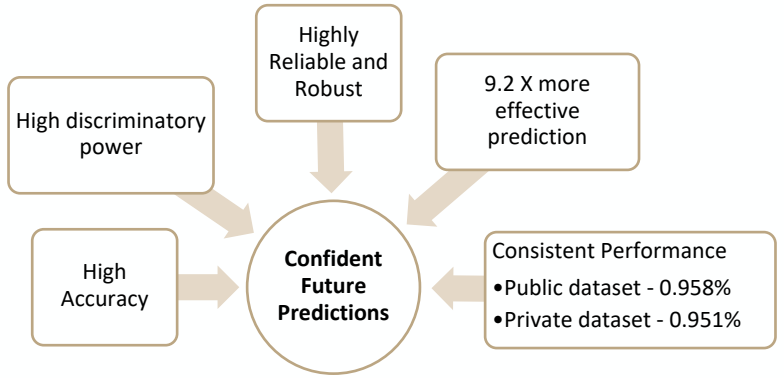
Network

Property	Value
Architecture	Multilayer Perceptron
Direct Connection	No
Number of Hidden Units	3

Fit Statistics

Selected Model	Model Node	Model Description	Target Variable	Selection Criterion: Valid: Roc Index	Valid: Misclassification Rate	Valid: Cumulative Lift	Valid: Gini Coefficient
Y	Ensmbl	Ensemble	class	0.969	0.012795	9.212133	0.939
	Neural	Neural Netw...	class	0.954	0.013595	9.02413	0.908
	Boost	Gradient Bo...	class	0.92	0.018792	6.580095	0.84

Model Highlights



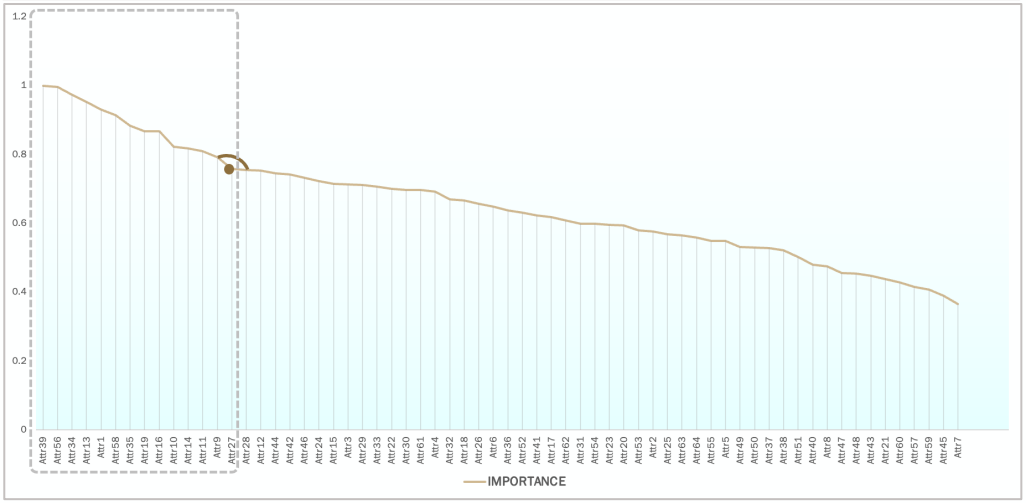
How do we interpret the output?

Gradient Boost - Model Output (Sample)

Variable Importance

Obs	NAME	LABEL	NRULES	NSURROGATES	IMPORTANCE	VIMPORTANCE	RATIO
1	Attr39		43	146	1.00000	0.89964	0.89964
2	Attr56		57	124	0.99703	0.94686	0.94969
3	Attr34		182	102	0.97453	0.94376	0.96843
4	Attr13		39	170	0.95288	0.82651	0.86738
5	Attr1		6	196	0.93132	0.83963	0.90155
6	Attr58		54	96	0.91379	0.83728	0.91627
7	Attr35		33	106	0.88333	0.82193	0.93050
8	Attr19		4	176	0.86840	0.66135	0.76157
9	Attr16		19	148	0.86782	0.87056	1.00316
10	Attr10		13	211	0.82272	0.73226	0.89004
11	Attr14		2	179	0.81764	0.67403	0.82436
12	Attr11		18	128	0.80910	0.63034	0.77907
13	Attr9		59	198	0.79202	0.75616	0.95472
14	Attr27		67	158	0.75903	0.85841	1.13093
15	Attr28		30	201	0.75500	0.92047	1.21917
16	Attr12		5	139	0.75381	0.65077	0.86330
17	Attr44		44	57	0.74628	1.00000	1.33997
18	Attr42		15	182	0.74162	0.53121	0.71628
19	Attr46		88	88	0.73295	0.85650	1.16856
20	Attr24		79	42	0.72262	0.73001	1.01022
21	Attr15		74	99	0.71454	0.73566	1.02956
22	Attr3		16	207	0.71311	0.81051	1.13659

Scree Plot – Interpret import variables



Variable Importance >= 0.9 (Top 6)

- **Sales Profitability** - Profit on sales / sales - Attr39
- **Sales Efficiency** - (Sales - cost of products sold) / sales - Attr56
- **Operational Leverage** - Operating expenses / total liabilities - Attr34
- **Profitability Ratio** - (Gross profit + depreciation) / sales - Attr13
- **Asset Profitability** - Net profit / total assets - Attr1
- **Cost to Sales Ratio** - Total costs / total sales - Attr58

Event Classification Table

Model Node	Model Description	Data Role	Target	Target Label	False Negative	True Negative	False Positive	True Positive
Neural	Neural Network	TRAIN	class		71	7332	8	88
Neural	Neural Network	VALIDATE	class		26	2440	8	27
Boost	Gradient Boosting	TRAIN	class		56	7340	0	103
Boost	Gradient Boosting	VALIDATE	class		47	2448	0	6
Ensmbl	Ensemble	TRAIN	class		41	7332	8	118
Ensmbl	Ensemble	VALIDATE	class		24	2440	8	29

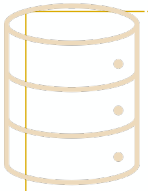
Other Scope

- Adjust Cutoff values to minimize False positives
- Try using more models in ensemble for better interpretability

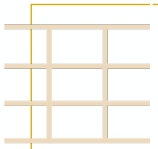
PURDUE UNIVERSITY

Objective | Approach | Best Model | **Interpret** | Learnings

What have we learnt from this project?

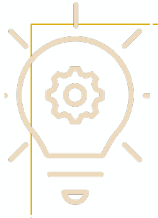


Experiential learning of SEMMA approach in building Predictive models

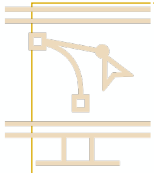


Various Data Pre-processing techniques

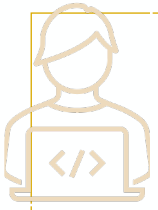
Building predictive models and Identifying the best model through Model comparison



Interpretation of different kinds of models



Hands-on Experience on SAS Enterprise Miner



Exposure to Kaggle Data competitions

Thank you!

PURDUE UNIVERSITY®