EUROPE 2022 DISCOVERY SUMMIT

Yield loss issue for an RFFE product : data analytics and machine learning for root cause searching (1/4)

Corinne Bergès, Martin Knotter, Maresa Labellarte

NXP Semiconductors

Introduction

- For a new WLAN product designed by NXP for the sixth Wifi generation, a yield loss issue is observed at the final test step, when the dies are packaged and ready to be shipped to the customers. Product quality is not at stake, but a project is launched to find the issue root causes and to improve yield.
- The main quality tool that is used for this study is a Fault Tree Analysis that allows to dig into each potential failure mode, without excluding some failure possibilities, nor jumping directly into an a priori or a first conclusion.
- NXP has been using Machine Learning methodology and algorithms from some years now, and Machine Learning was implemented in this yield loss case, too, in parallel to typical data analytics or univariate analysis.
- JMP PRO was used to build data analytics and machine learning analysis (partition, boosted trees or bootstrap models): the analysis mainly dealt about the difference between the unit test and the final product test, and on the difference in the laminate models.
- The poster that is proposed will present the case study and will detail the analysis, but also, will formalize the Machine Learning approach for a deployed Machine Learning usage.

Issue description and root cause searching

- The final product is constituted by two dies, mounted on a laminate: no issue was observed at the die test step after die manufacturing and before its packaging. So, a first study dealt with the potential packaging effect on yield.
- The yield loss issue was observed on the product model manufactured only on one type of laminates, which can lead to hypothesis on the laminate type impact. This topic was the second type of root cause searching ones.
- Finally, one study aimed the yield difference between the two dies per product that could have been due to a lightly different grounding of the two dies.

Main studies

- The root cause searching for this issue was structured by a Fault Tree Analysis that allowed a comprehensive study of all the possible assumptions, around the three topics previously listed and around still many more. A machine learning approach in JMP was used as much as possible, for its strong capabilities in RCPS.

Case study and input data

Case study

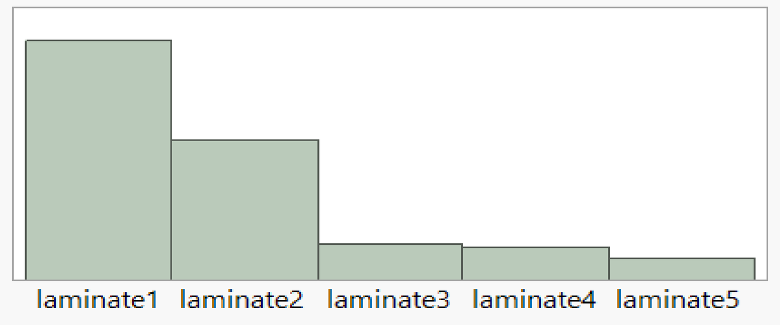
- This poster is presenting the machine learning analysis performed to study yield loss difference observed according to the laminate types.

Input data:

- 271460 dies manufactured with 5 different laminate types;
- For these 271460 dies, test values for 15 tests, at the Final Test step, when the dies are packaged and ready to be shipped to customers.

Distributions

laminate



Frequencies

Level	Count	Prob
laminate1	140696	0,51829
laminate2	81720	0,30104
laminate3	19240	0,07088
laminate4	17906	0,06596
laminate5	11898	0,04383
Total	271460	1,00000

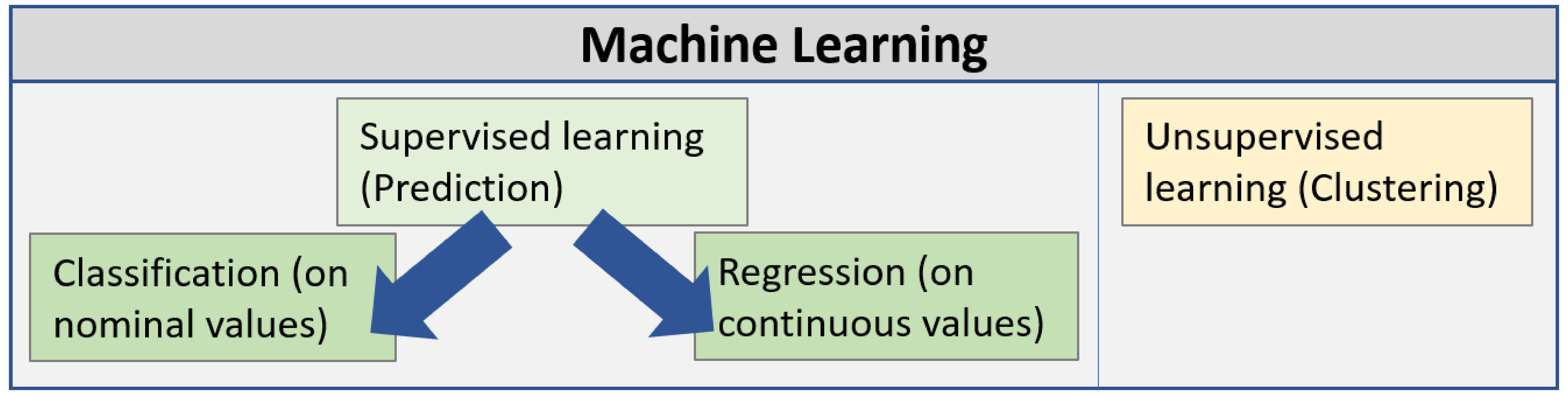
N Missing 0

5 Levels

Laminate distributions _ Laminate1 is more widely represented in the data (data unbalance): that means that stratified sampling is needed.

Prediction vs clustering _ Classification versus regression analysis

- In order to analyze laminate impact on yield loss, test values for 271460 dies are used, but for each of these dies, laminate is known: this information about the laminate clearly distinguishes an unsupervised analysis (clustering) to a supervised learning (prediction case). The study will try to see if it is possible to predict the laminate type from test values for the dies.
- Furthermore, laminate type is a nominal data: thus, machine learning approach will speak about a classification analysis.



The case study is fitting with a classification analysis used, typically in a prediction context, here for root cause searching.

Data preparation

- Some data preparing operations are often needed for some analysis in term of data formatting, but some of them are really also contributing to analysis efficiency: they are:
  - Removal of the missing values
  - Removal of the most correlated tests: in this analysis, this correlation analysis performed with a 90% correlation threshold, led to remove 7 tests among the 15 ones

Multivariate

Correlations

	test1	test2	test3	test4	test6	test8	test9	test15
test1	1,0000	0,8576	0,1843	0,6461	0,7964	0,2227	0,8293	0,3805
test2	0,8576	1,0000	0,1343	0,5365	0,6048	0,2693	0,7735	0,3508
test3	0,1843	0,1343	1,0000	0,3477	0,3687	0,1047	0,1532	0,8513
test4	0,6461	0,5365	0,3477	1,0000	0,8310	0,1684	0,7979	0,3434
test6	0,7964	0,6048	0,3687	0,8310	1,0000	0,1808	0,6698	0,3703
test8	0,2227	0,2693	0,1047	0,1684	0,1808	1,0000	0,2693	0,1013
test9	0,8293	0,7735	0,1532	0,7979	0,6698	0,2693	1,0000	0,2863
test15	0,3805	0,3508	0,8513	0,3434	0,3703	0,1013	0,2863	1,0000

The correlations are estimated by Row-wise method.

Correlation analysis

Train-test-validation data

- Data that is fitting with the case study, is split in three different datasets that analytics methods will use for three different purposes:
  - about 80% of the data are used to train the model (train data);
  - about 10% allow to determine the best parameters for the model (hyperparameters) (validation data);
  - lastly, evaluation metrics are computed on the remaining data to assess the model and its performance (about 20%) (test data).
- JMP provides a utility to perform this splitting that can stratify the 3 datasets as the stratification observed in the whole data volume.
- The K-fold cross-validation is another method to perform validation and test of machine learning models.

Make Validation Column

Stratified Validation Column

Randomly partitions the rows into training, validation and test sets while attempting to evenly distribute across levels of the stratification variable(s). Use this option when you want a balanced representation of a column's levels in each of the training, validation and test sets.

Stratification Columns: laminate

Specify rates or relative rates

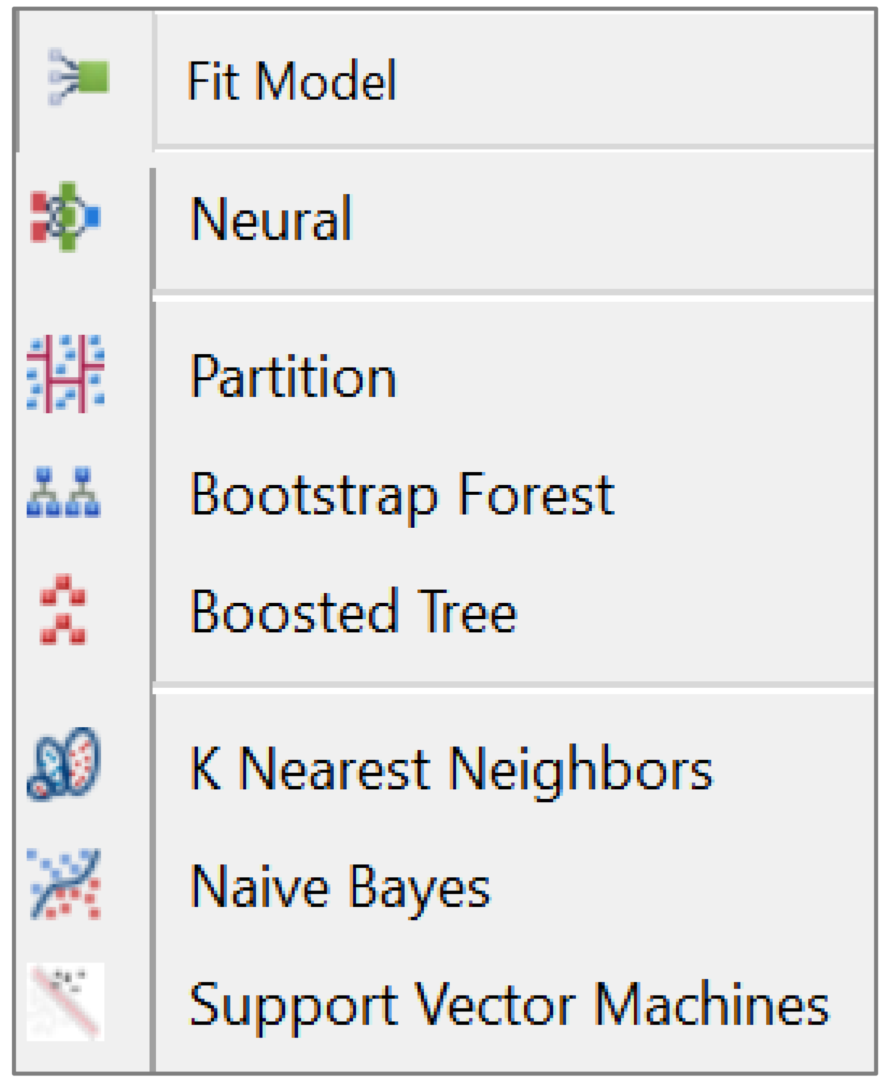
		Adjusted Rates	Row Counts
Training Set	0,7	0,7	124452
Validation Set	0,1	0,1	17779
Test Set	0,2	0,2	35557
Excluded Rows			0
Total Rows			177788

Train-test-validation data

Yield loss issue for an RFFE product : data analytics and machine learning for root cause searching

Machine Learning models implemented here

- Among all the different machine learning analysis that are possible in JMP, the following ones were performed for this issue:
  - Nominal Logistic in Fit Model menu
  - Partition or Tree
  - Bootstrap Forest (that is an ensemble model) (100 trees implemented here)
  - K Nearest Neighbors
  - Naives Bayes
  - Support Vectors Machines
  - Lastly, 2 neural models with different



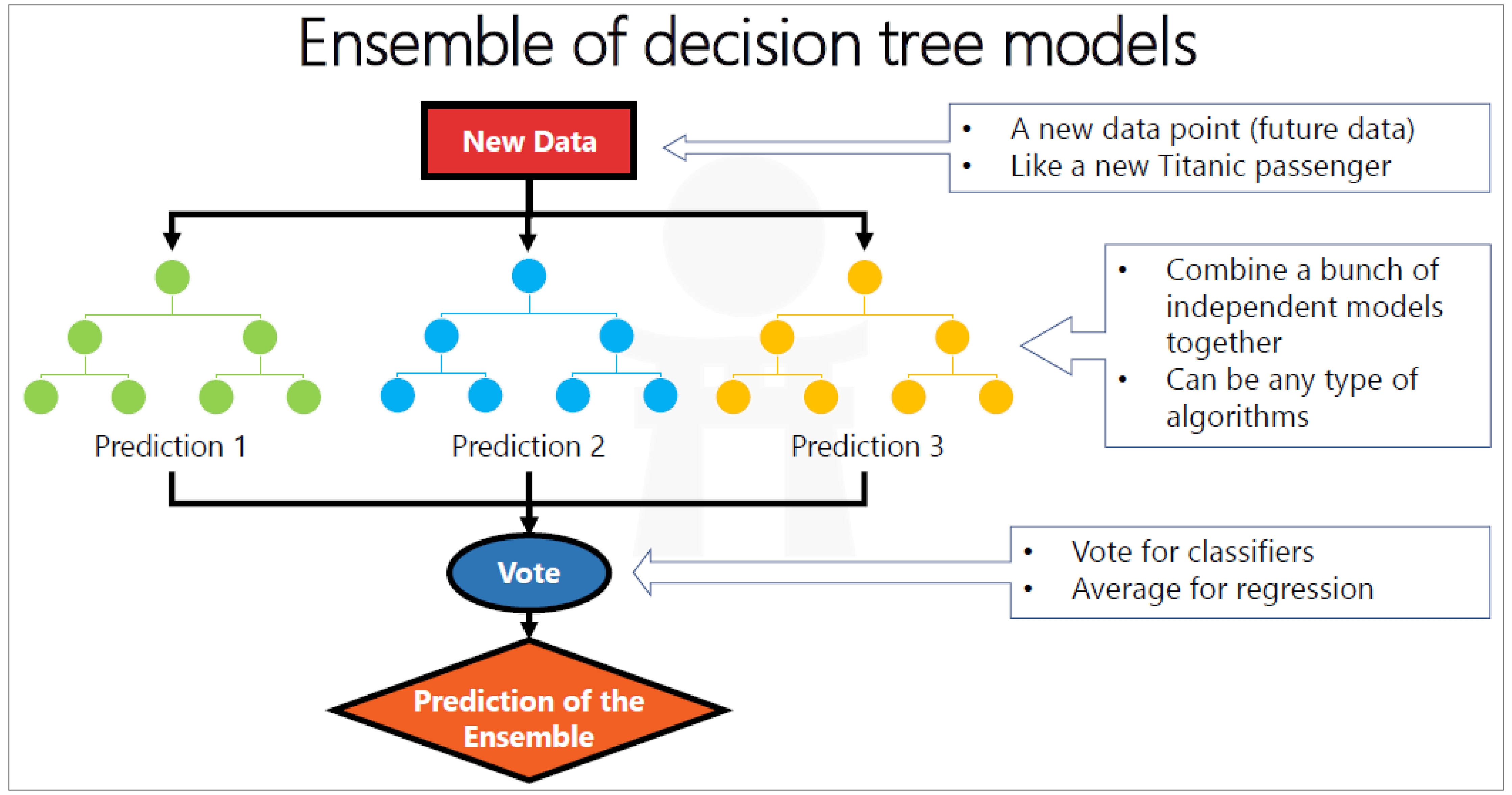
Each model in some words…

- Nominal Logistic: this model minimizes a specific cost function (called logit or sigmoid function), which makes it appropriate for classification.
- Partition or Tree: a decision tree classification is a simple algorithm which builds a decision tree. Each node of the decision tree includes a condition on one of the input features.
- A Random Forest is made of many decision trees. Each tree in the forest predicts a record, and each tree "votes" for the final answer of the forest.
- K Nearest Neighbor classification makes predictions for a sample by finding the k nearest samples and assigning the most represented class among them. This algorithm requires storing the entire training data into the model. This will lead to a very large model if the data is larger than a few hundred lines. Predictions may also be slow.
- Naive Bayes classifiers are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong (naïve) independence assumptions between the features.
- Support Vector Machine is a powerful 'black-box' algorithm for classification. Through the use of kernel functions, it can learn complex non-linear decision boundaries (ie, when it is not possible to compute the target as a linear combination of input features). SVM is effective with large number of features.
- Neural Networks are a class of parametric models which are inspired by the functioning of neurons. They consist of several "hidden" layers of neurons, which receive inputs and transmit them to the next layer, mixing the inputs and applying non-linearities, allowing for a complex decision function.

An ensemble model: what is it ?

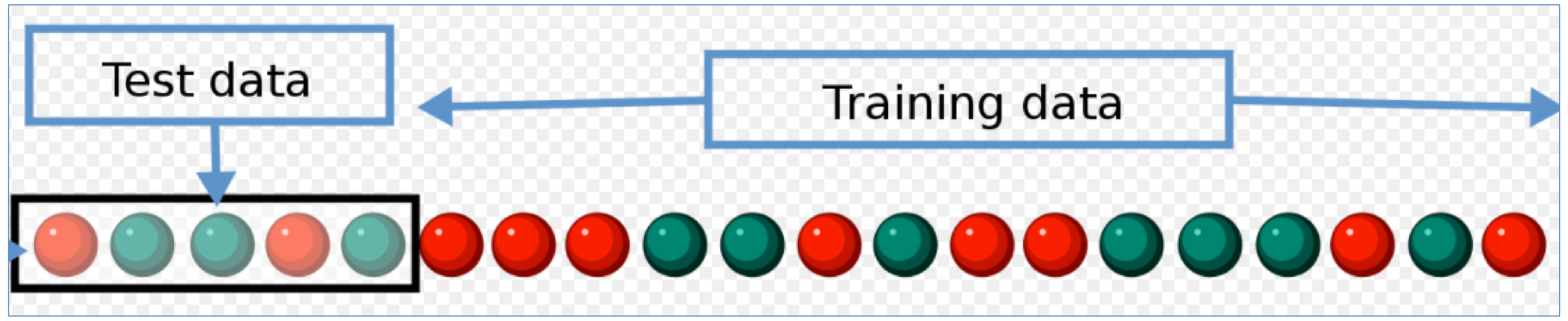
Ensemble models: ex: Boostrap Forest

- Improve model performance by combining multiple models
- Ensembles can be of any learning algorithm, including both classification and regression



Bootstrapping to reduce overfitting…

- Same model but not same training data … Bootstrap aggregating or bagging: bagging is a technique of building many decision trees at a time by randomly sampling with replacement, or bootstrapping, from the original dataset. This ensures variety in the trees, which helps to reduce the amount of overfitting.



Yield loss issue for an RFFE product : data analytics and machine learning for root cause searching

Metrics and model evaluation methods in machine learning

Metrics:

- Accuracy = (TP + TN)/(TP + TN + FP + FN)
- Precision : Out of all the examples the classifier labeled as positive, what fraction were correct? Precision = TP/(TP + FP)
- Recall : Out of all the positive examples, what fraction did the classifier pick up?
  Recall = TP/(TP + FN)
- Precision is more important for us than Recall when getting a False Positive is very costly
- Recall is more important for us than Precision if classifying any true positive into negative reaches to an apocalypse
- Tweaking a classifier is a matter of balancing what is more important for us: precision or recall.
- It is possible to get both in a single measure : the F-Score (or F1-Score or F-Measure)
  F1 is the harmonic mean of precision and recall, when the "average" of ratios (percentages) is needed.
  F1 = 2*P*R/(P+R)

Confusion matrix and ROC curve

- **Confusion matrix:**

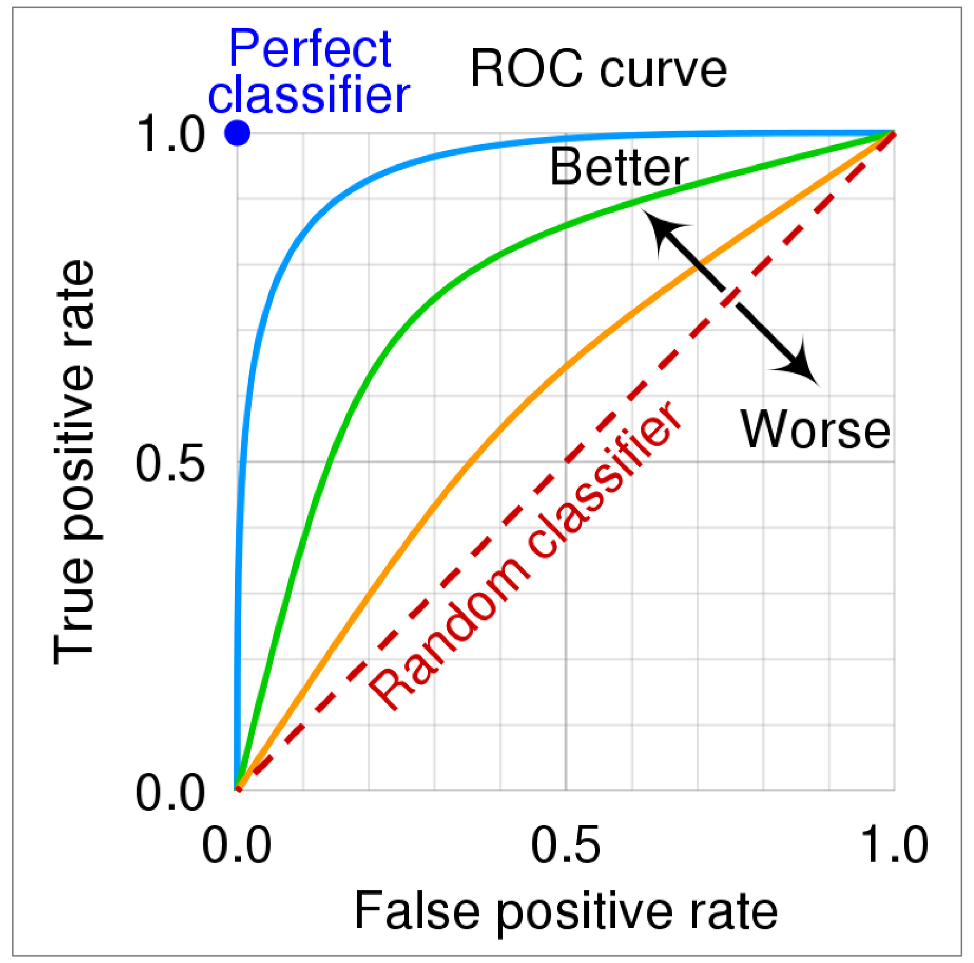
In predictive analytics for binary classification, a confusion matrix is a table with two rows and two columns that reports the number of false positives, false negatives, true positives, and true negatives. This allows more detailed analysis than mere proportion of correct classifications (accuracy). In a multiclassification problem, confusion matrices can have more categories.

		Predicted condition	
	Total population = P + N	Positive (PP)	Negative (PN)
Actual condition	Positive (P)	True positive (TP),	False negative (FN),
	Negative (N)	False positive (FP),	True negative (TN),

Confusion matrix (Source: Wikipedia)

- **ROC curve:**

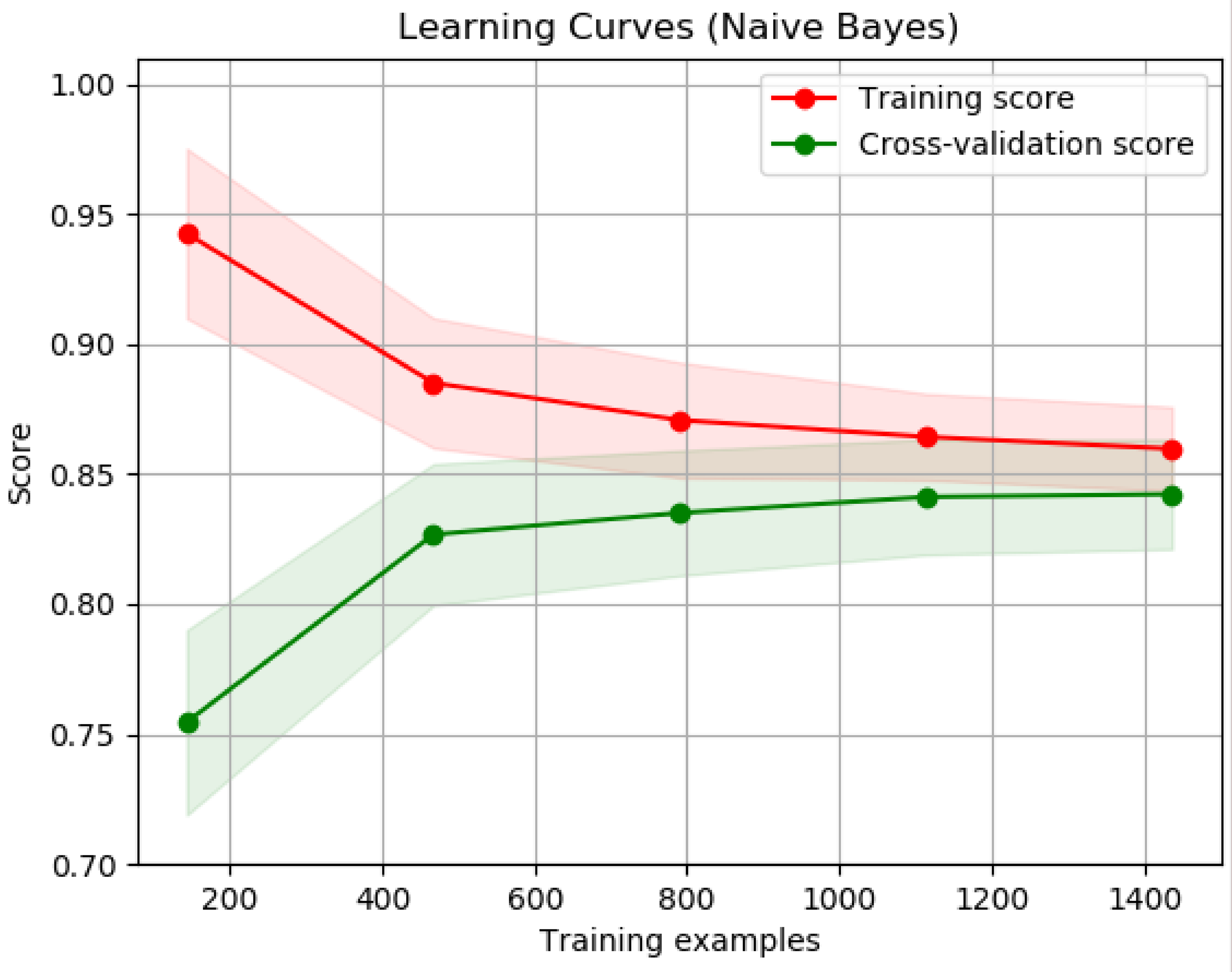
A receiver operating characteristic curve, or ROC curve, is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The true-positive rate is also known as sensitivity, recall or probability of detection. The false-positive rate is also known as probability of false alarm. (Source: Wikipedia)



ROC curve (Source: Wikipedia)

Learning curve

- In machine learning, **a learning curve (or training curve)** plots the optimal value of a model's loss function for a training set against this loss function evaluated on a validation data set with same parameters as produced the optimal function. It is a tool to find out how much a machine model benefits from adding more training data and whether the estimator suffers more from a variance error or a bias error. If both the validation score and the training score converge to a value that is too low with increasing size of the training set, it will not benefit much from more training data. (Source: Wikipedia)

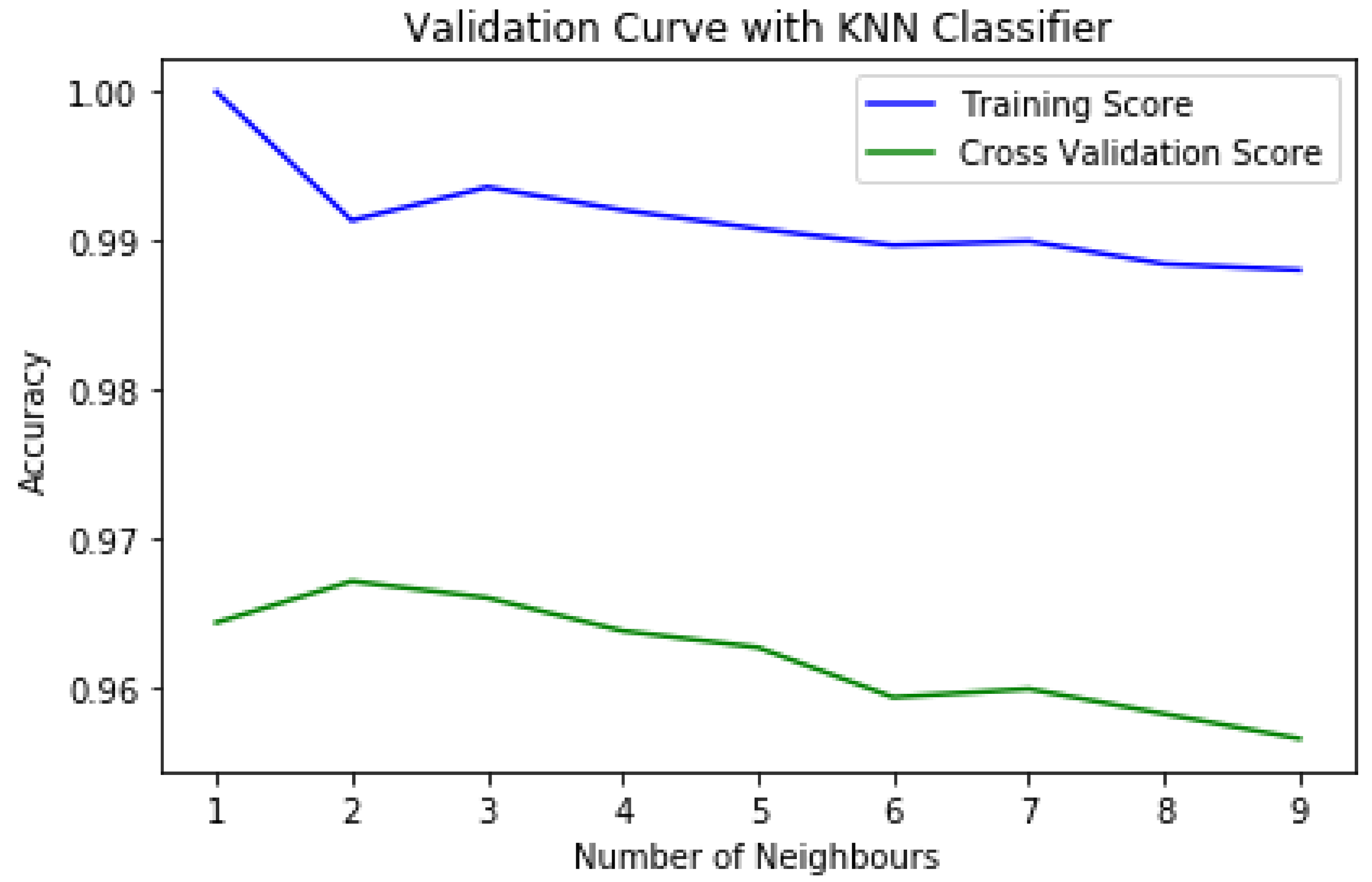


Learning curves (Source: Wikipedia)

Validation curve

- A **validation curve** is an important diagnostic tool that shows the sensitivity between to changes in a Machine Learning model's accuracy with change in some parameter of the model.
- A **validation curve** is used to evaluate an existing model based on hyper-parameters and is not used to tune a model.

(Source: Wikipedia)



Validation curves (Source: Wikipedia)

Variable contribution / importance

- Variable contribution table and variable importance plot allow to interpretate the models and to explain the issue by highlighting the key features with the highest contributions.
- Unfortunately, all the models do not offer this interpretability: for example, a neural network is more a black box.

Column Contributions

Term	Number of Splits	G^2	Portion
test3	1013	20420,3723	0,2691
test15	1178	16830,9452	0,2218
test1	991	10174,6029	0,1341
test2	813	9425,11545	0,1242
test9	911	7497,68621	0,0988
test6	1054	6655,53679	0,0877
test4	880	3343,95582	0,0441
test8	721	1528,74431	0,0201

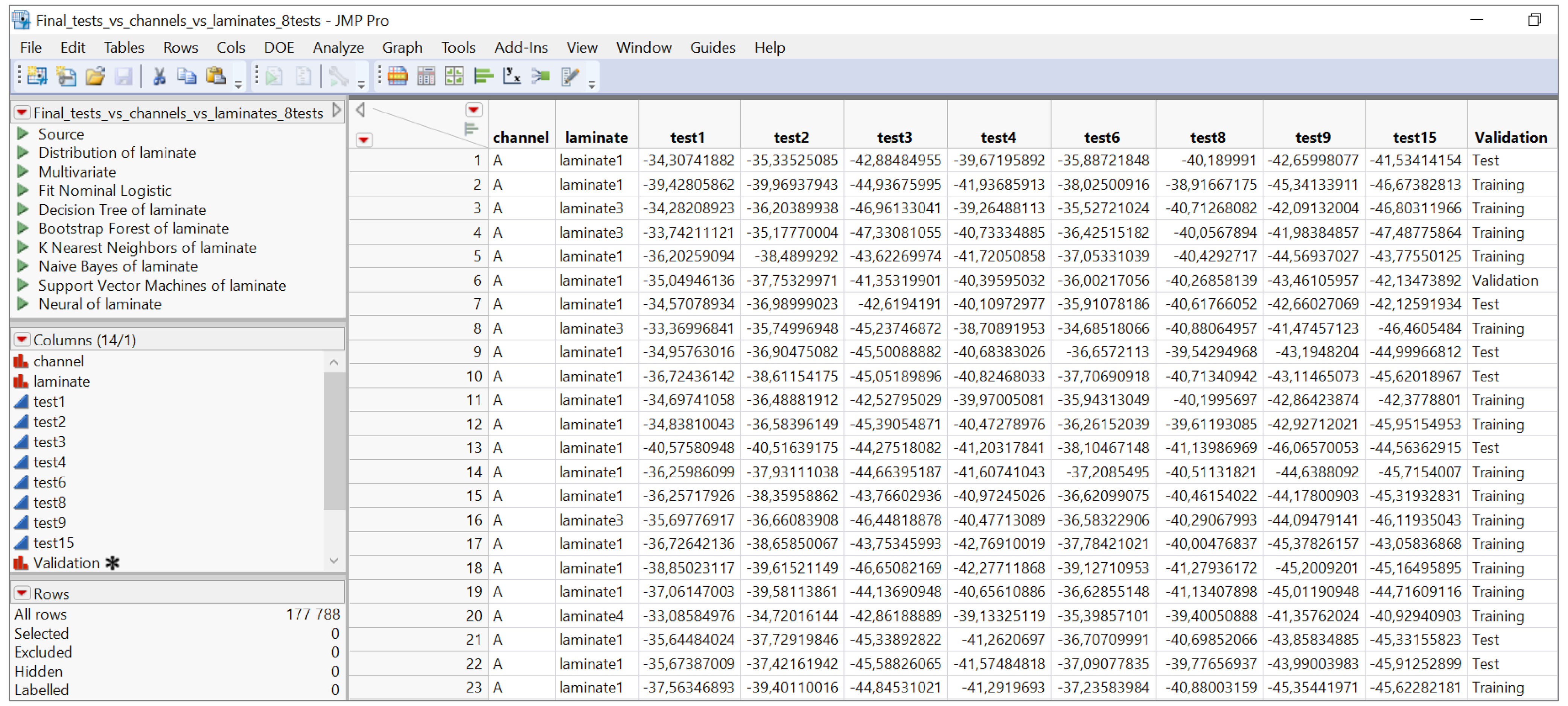
Test contribution for the Bootstrap Random forest in JMP

EUROPE 2022 DISCOVERY SUMMIT

Yield loss issue for an RFFE product : data analytics and machine learning for root cause searching

JMP file

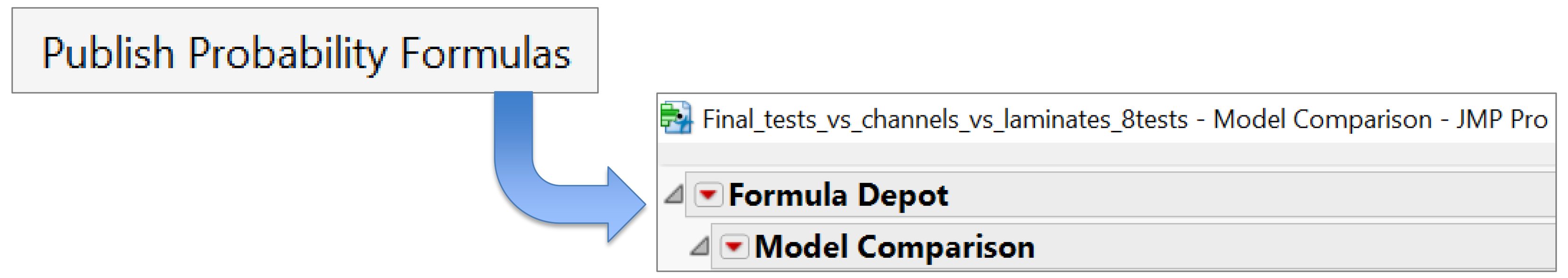
- Input table is saved in the JMP file:
  'Final_tests_vs_channels_vs_laminates_8tests.jmp'



	channel	laminate	test1	test2	test3	test4	test6	test8	test9	test15	Validation
1	A	laminate1	-34,30741882	-35,33525085	-42,88484955	-39,67195892	-35,88721848	-40,189991	-42,65998077	-41,53414154	Test
2	A	laminate1	-39,42805862	-39,96937943	-44,93675995	-41,93685913	-38,02500916	-38,91667175	-45,34133911	-46,67382813	Training
3	A	laminate3	-34,28208923	-36,20389938	-46,96133041	-39,26488113	-35,52721024	-40,71268082	-42,09132004	-46,80311966	Training
4	A	laminate3	-33,74211121	-35,17770004	-47,33081055	-40,73334885	-36,42515182	-40,0567894	-41,98384857	-47,48775864	Training
5	A	laminate1	-36,20259094	-38,4899292	-43,62269974	-41,72050858	-37,05331039	-40,4292717	-44,56937027	-43,77550125	Training
6	A	laminate1	-35,04946136	-37,75329971	-41,35319901	-40,39595032	-36,00217056	-40,26858139	-43,46105957	-42,13473892	Validation
7	A	laminate1	-34,57078934	-36,98999023	-42,6194191	-40,10972977	-35,91078186	-40,61766052	-42,66027069	-42,12591934	Test
8	A	laminate3	-33,36996841	-35,74996948	-45,23746872	-38,70891953	-34,68518066	-40,88064957	-41,47457123	-46,4605484	Training
9	A	laminate1	-34,95763016	-36,90475082	-45,50088882	-40,68383026	-36,6572113	-39,54294968	-43,1948204	-44,99966812	Test
10	A	laminate1	-36,72436142	-38,61154175	-45,05189896	-40,82468033	-37,70690918	-40,71340942	-43,11465073	-45,62018967	Test
11	A	laminate1	-34,69741058	-36,48881912	-42,52795029	-39,97005081	-35,94313049	-40,1995697	-42,86423874	-42,3778801	Training
12	A	laminate1	-34,83810043	-36,58396149	-45,39054871	-40,47278976	-36,26152039	-39,61193085	-42,92712021	-45,95154953	Training
13	A	laminate1	-40,57580948	-40,51639175	-44,27518082	-41,20317841	-38,10467148	-41,13986969	-46,06570053	-44,56362915	Test
14	A	laminate1	-36,25986099	-37,93111038	-44,66395187	-41,60741043	-37,2085495	-40,51131821	-44,6388092	-45,7154007	Training
15	A	laminate1	-36,25717926	-38,35958862	-43,76602936	-40,97245026	-36,62099075	-40,46154022	-44,17800903	-45,31932831	Training
16	A	laminate3	-35,69776917	-36,66083908	-46,44818878	-40,47713089	-36,58322906	-40,29067993	-44,09479141	-46,11935043	Training
17	A	laminate1	-36,72642136	-38,65850067	-43,75345993	-42,76910019	-37,78421021	-40,00476837	-45,37826157	-43,05836868	Training
18	A	laminate1	-38,85023117	-39,61521149	-46,65082169	-42,27711868	-39,12710953	-41,27936172	-45,2009201	-45,16495895	Training
19	A	laminate1	-37,06147003	-39,58113861	-44,13690948	-40,65610886	-36,62855148	-41,13407898	-45,01190948	-44,71609116	Training
20	A	laminate4	-33,08584976	-34,72016144	-42,86188889	-39,13325119	-35,39857101	-39,40050888	-41,35762024	-40,92940903	Training
21	A	laminate1	-35,64484024	-37,72919846	-45,33892822	-41,2620697	-36,70709991	-40,69852066	-43,85834885	-45,33155823	Test
22	A	laminate1	-35,67387009	-37,42161942	-45,58826065	-41,57484818	-37,09077835	-39,77656937	-43,99003983	-45,91252899	Test
23	A	laminate1	-37,56346893	-39,40110016	-44,84531021	-41,2919693	-37,23583984	-40,88003159	-45,35441971	-45,62282181	Training

Model comparison

- The 8 models are separately built on the input table
- Then, all the models are compared by JMP comparison capability through the 'Formula Depot' platform.



- If Neural models are excluded because of their low interpretability in this root cause problem searching case for which an understanding of the key factors is highly looked for in order to react against them, it appears that Nominal Logistic model is providing the best fitting.

Measures of Fit for laminate

Creator	.2 .4 .6 .8	Entropy RSquare	Generalized RSquare	Mean -Log p	RMSE	Mean Abs Dev	Misclassification Rate	N	AUC
Fit Nominal Logistic		0,7763	0,8744	0,147	0,1941	0,0793	0,0480	177788	0,9834
Partition		0,2115	0,3319	0,5181	0,3949	0,2962	0,2089	177788	0,7123
Bootstrap Forest		0,7684	0,8693	0,1522	0,2080	0,1075	0,0527	177788	0,9886
K Nearest Neighbors		.	.	.	.	.	0,0230	177788	.
Naive Bayes		0,0109	0,0195	0,6499	0,3373	0,1423	0,1336	177781	0,9470
Support Vector Machines		.	.	.	.	.	0,0286	177788	.
Neural		0,8298	0,9079	0,1119	0,1754	0,0619	0,0398	177788	0,9912
Neural		0,8913	0,9436	0,0714	0,1408	0,0396	0,0260	177788	0,9966

Nominal Logistic model

- Nominal Logistic model is resulting to a 4.80% mean misclassification rate

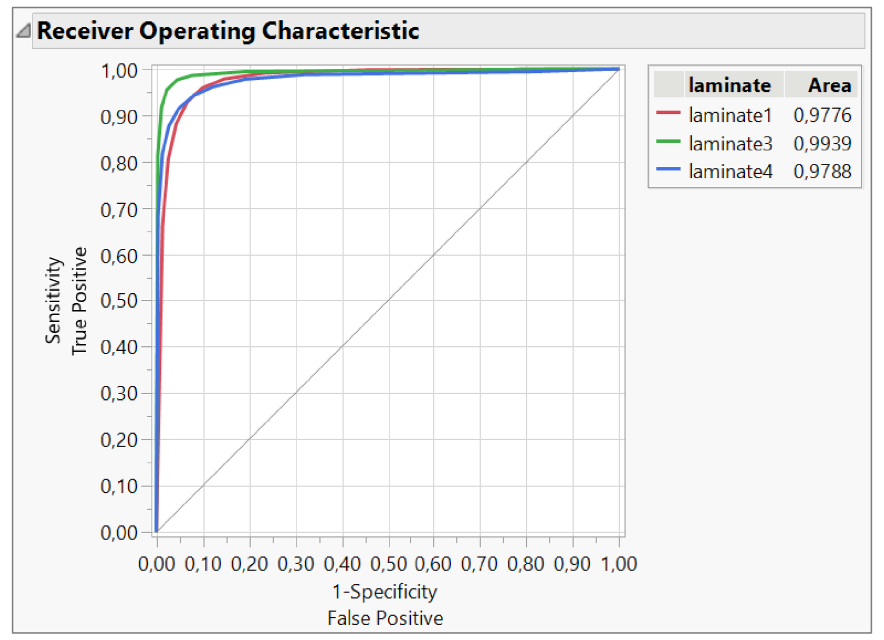
Confusion Matrix

Training

Actual laminate	Predicted Count laminate1	laminate3	laminate4
laminate1	137409	1453	1794
laminate3	1777	17436	15
laminate4	3476	21	14407

Confusion matrix in JMP

- ROC curve proves model goodness



ROC curve in JMP

- Parameter estimates offer good clues to understand on which tests laminate type change is observed.

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq	Lower 95%	Upper 95%
Intercept	5,13683343	1,1733209	19,17	<,0001*	3,00220052	7,57673795
test1	-3,269566	0,0754842	1876,2	<,0001*	-3,4185427	-3,1228732
test2	0,99459321	0,0359775	764,24	<,0001*	0,92435921	1,06538624
test3	-0,2398378	0,0233117	105,85	<,0001*	-0,285525	-0,1942618
test4	-1,4775098	0,0682754	468,31	<,0001*	-1,6111563	-1,3438638
test6	5,56149618	0,0857967	4201,9	<,0001*	5,39468842	5,73083962
test8	0,60332345	0,0282744	455,32	<,0001*	0,54952678	0,66027077
test9	-0,6190732	0,0483781	163,75	<,0001*	-0,7139058	-0,524794
test15	-1,1406458	0,0223613	2602,0	<,0001*	-1,1845816	-1,0969684

Parameters estimates and their confidence intervals in JMP for the Nominal Logistic model

Neural networks

- Here is the most complex neural network among the 2 ones implemented in this case study
- Modeling result is very high in term of confusion rate and variability explainability
- But interpretability is too small in this RCPS case, and another model has to be preferred.

Hidden Layer Structure

Number of nodes of each activation type

Activation Layer	Sigmoid TanH	Identity Linear	Radial Gaussian
First	3	3	3
Second	3	3	3

Model NTanH(3)NLinear(3)NGaussian(3)NTanH2(3)NLinear2(3)NGaussian2(3)

Training — laminate

Measures	Value
Generalized RSquare	0,9447912
Entropy RSquare	0,8934772
RMSE	0,1395525
Mean Abs Dev	0,039187
Misclassification Rate	0,0256967
-LogLikelihood	8711,0991
Sum Freq	124452

Confusion Matrix

Actual laminate	Predicted Count laminate1	laminate3	laminate4
laminate1	97125	670	663
laminate3	728	12707	26
laminate4	1085	26	11422

Confusion Rates

Actual laminate	Predicted Rate laminate1	laminate3	laminate4
laminate1	0,986	0,007	0,007
laminate3	0,054	0,944	0,002
laminate4	0,087	0,002	0,911

Validation — laminate

Measures	Value
Generalized RSquare	0,9421999
Entropy RSquare	0,8888186
RMSE	0,1432204
Mean Abs Dev	0,0401415
Misclassification Rate	0,0268857
-LogLikelihood	1299,0013
Sum Freq	17779

Confusion Matrix

Actual laminate	Predicted Count laminate1	laminate3	laminate4
laminate1	13865	101	99
laminate3	116	1804	3
laminate4	155	4	1632

Confusion Rates

Actual laminate	Predicted Rate laminate1	laminate3	laminate4
laminate1	0,986	0,007	0,007
laminate3	0,060	0,938	0,002
laminate4	0,087	0,002	0,911

Test — laminate

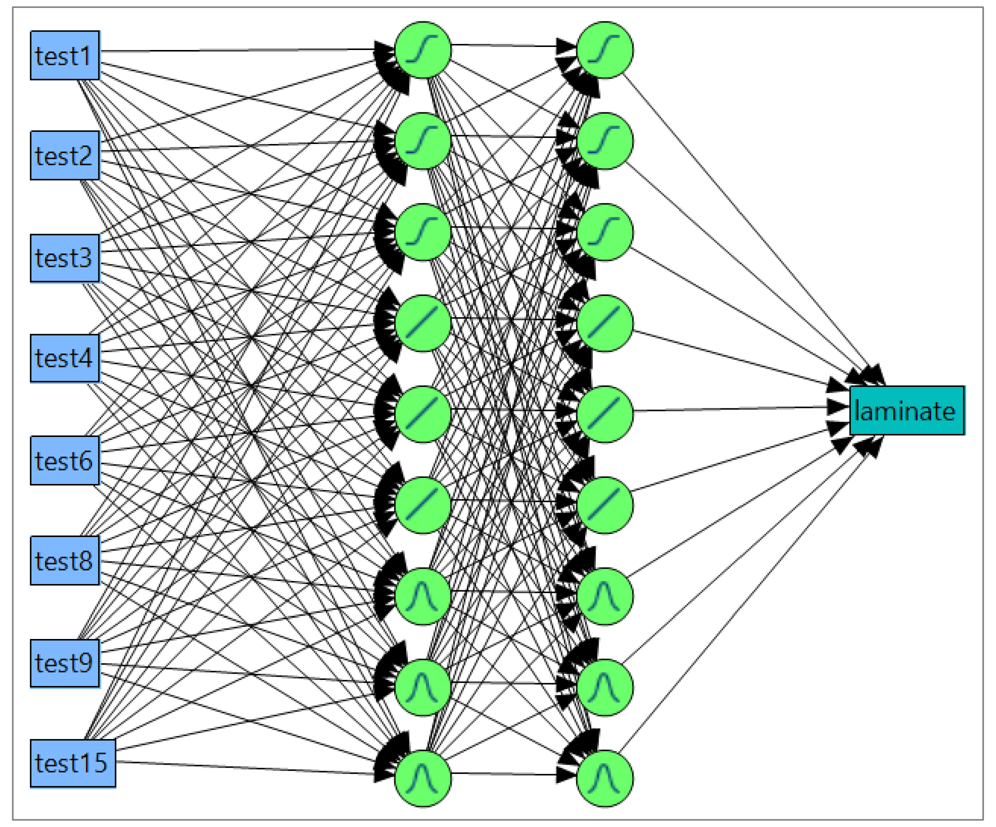
Measures	Value
Generalized RSquare	0,9399804
Entropy RSquare	0,8848822
RMSE	0,1437781
Mean Abs Dev	0,0405802
Misclassification Rate	0,0268583
-LogLikelihood	2689,026
Sum Freq	35557

Confusion Matrix

Actual laminate	Predicted Count laminate1	laminate3	laminate4
laminate1	27732	190	211
laminate3	212	3624	8
laminate4	326	8	3246

Confusion Rates

Actual laminate	Predicted Rate laminate1	laminate3	laminate4
laminate1	0,986	0,007	0,008
laminate3	0,055	0,943	0,002
laminate4	0,091	0,002	0,907



Conclusion

- JMP offers a strong platform in machine learning modeling, model evaluation and comparison, very comparable to Python capabilities, without the need from a user, to know how to code in Python
- Furthermore, JMP provides a utility to convert a machine learning JMP script into a Python code that may be used to deploy a model elsewhere in a Python environment
- In this case study of a yield loss on a WLAN product, machine learning in JMP succeeded to greatly help the engineers in different steps in the Root Cause Problem Solving.

Acknowledgements and References

- The authors thank the whole team of engineers involved in this yield loss case, for the data they provided, their expertise and their sharing, to move forward in the issue understanding.

- References: JMP documentation

Yield loss issue for an RFFE product :
data analytics and machine learning for root cause searching

The authors

Author 1:

D. Corinne Bergès

Six Sigma Master Black Belt for NXP Structured Problem Solving Organization, PhD

NXP Semiconductors

Six Sigma Master Black Belt, committed in innovative statistical analysis for design, quality or test, Corinne Bergès is involved in problem solving for manufacturing, sub-contractors and suppliers, and in machine learning projects. Previously, she was a Reliability and Failure Analysis engineer for Motorola and Freescale, a project leader in CNES spatial agency, and a design leader for spatial and seismic prospection. Engineer in electrotechnics, automatics and electronics, she has a PhD in electronic systems.

Author 2:

D. Martin Knotter

Title: Six Sigma Master Black Belt, Technical Director, Product Diagnostic Center (PDC), NXP Semiconductors

Six Sigma Black Belt, involved in problem solving on manufacturing issues throughout the complete chain from supplier to end-customer. Technical staff member PDC that encompasses FA and material analysis within NXP. Expertise is impact of contamination on semiconductor process technology. Previously, he was researcher at Philips Research on surface science and semiconductor ultraclean process technology. Prior to that he got a PhD in Chemistry.

Author 3:

Maresa Labellarte

Product Engineering, NXP Semiconductors

Product Engineer, involved in New Product Introduction, from product definition stage through industrial testing and validation till product release and volume ramp-up. Perform data analysis in order to evaluate the product performances compared to specifications and to enhance the total yield. She has a Master degree in Electronic Engineering.