Credit Scoring and Credit Control XV

Automated documentation creation for a streamlined, compliant score build process

Jonathan Clarke
Manager
UK Analytics

Matt Pead
Senior Statistical Modeller
UK Analytics

LexisNexis® RISK SOLUTIONS

Contents

- Who are LexisNexis Risk Solutions?
- Typical challenges when documenting a model build.
- Introducing an end-to-end model build process.
- Example outputs achievable using R Markdown.

Who are LexisNexis Risk Solutions?

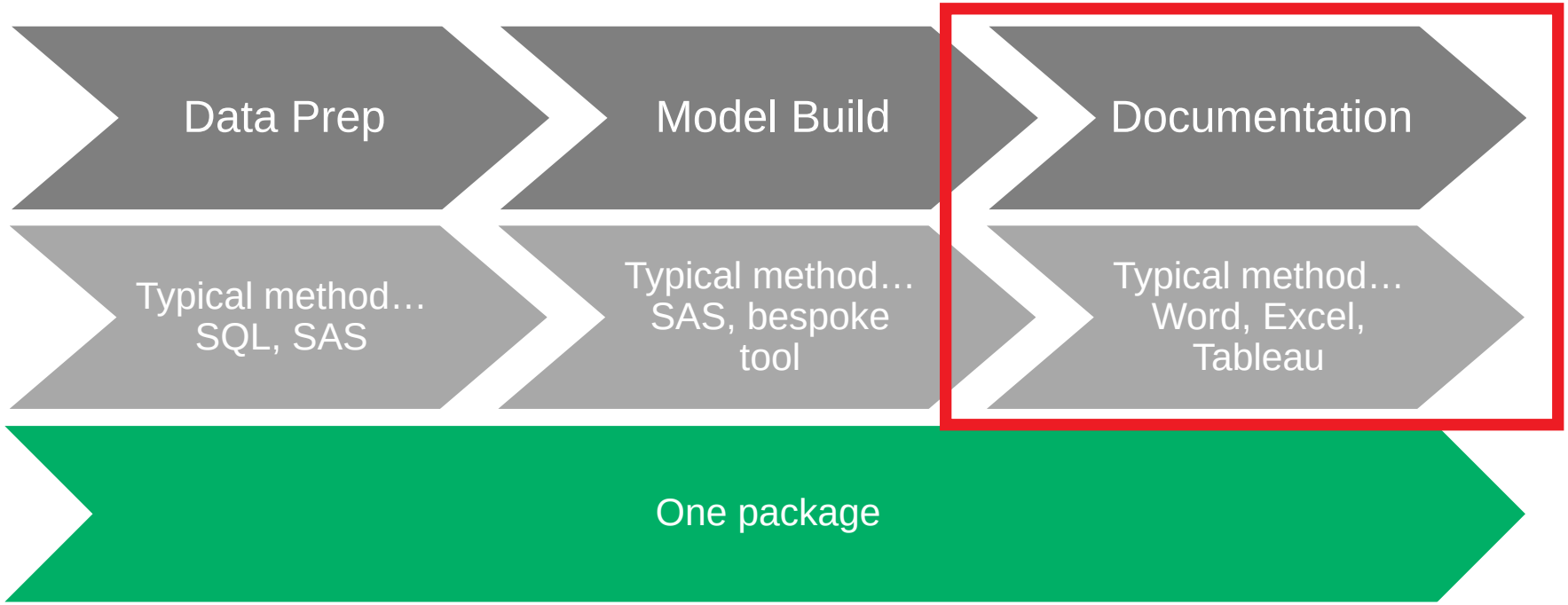
Our vision: Be the essential partner in the assessment of relationship risk.

- LexisNexis Risk Solutions is a leader in providing essential information that helps customers across all industries and government assess, predict, and manage risk.

- Combining cutting-edge technology, unique data and advanced analytics, we provide products and services that address evolving client needs in the risk sector while upholding the highest standards of security and privacy.

- LexisNexis Risk Solutions is part of RELX Group plc, a world-leading provider of information and analytics for professional and business customers across industries.

Achieving this vision with an efficient and compliant process

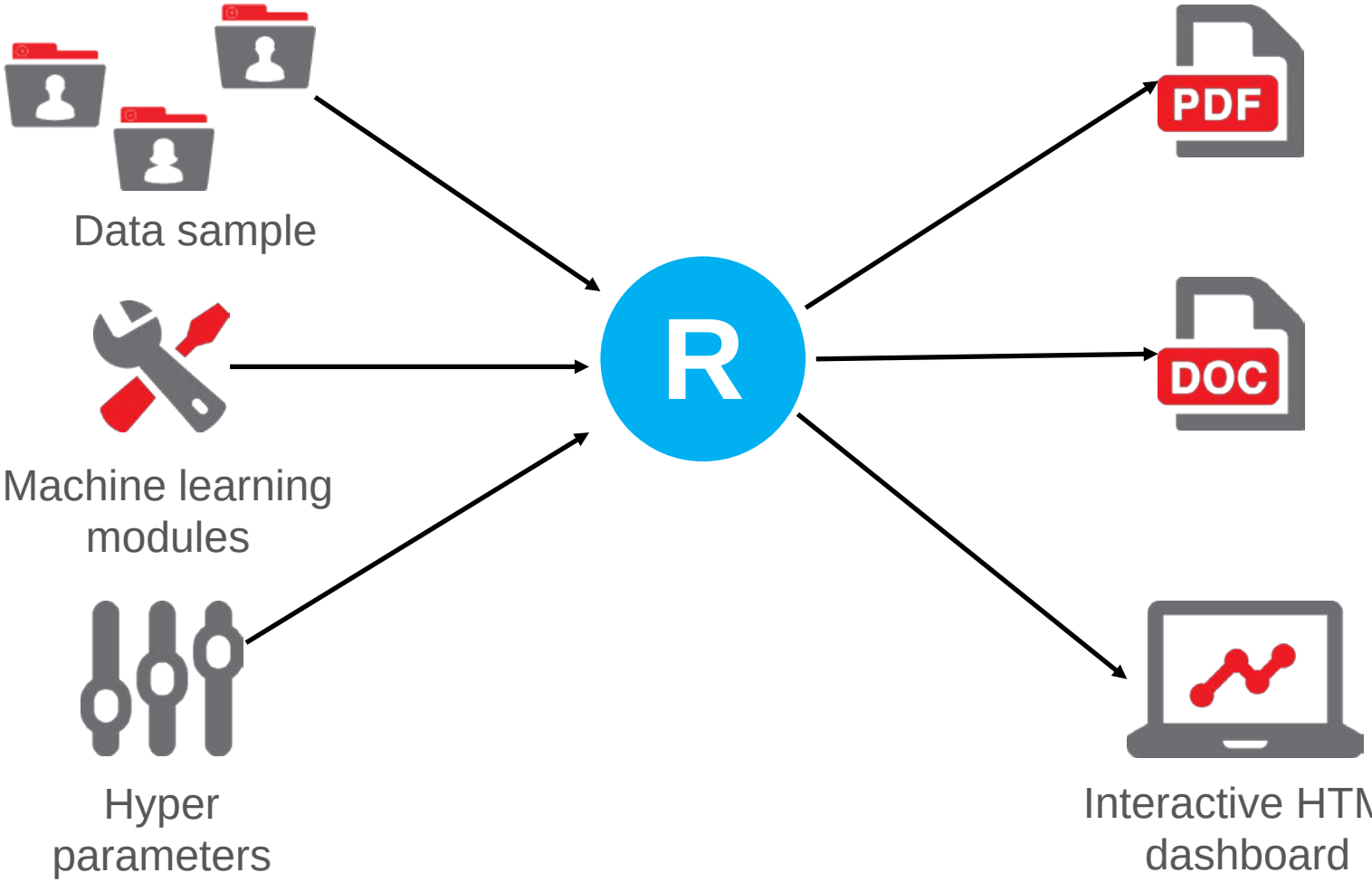


- We have an automated modelling process for many use cases and outcome definitions.
- We have the ability to easily leverage different machine learning techniques.
- We have the ability to produce documentation efficiently for peer review and internal audit / industry compliance (FCA, GDPR).

Typical challenges when documenting a model build

- There is a disconnect between the model build process and the documentation describing it.
- Multiple touch points may exist across different software (e.g. SAS to Excel to Word).
- Errors can be made when ‘tuning out’ during repeated editing of the document.
- Bugs/errors are found during documentation and peer review, requiring the model build process to be re-run.

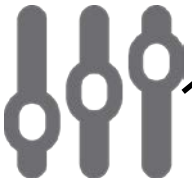
R Markdown - data prep, model building and documentation



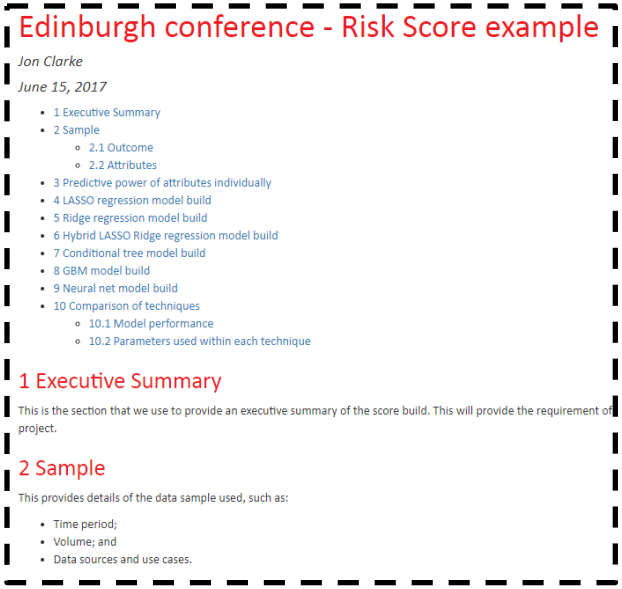
Example – predicting whether an individual will receive a CCJ

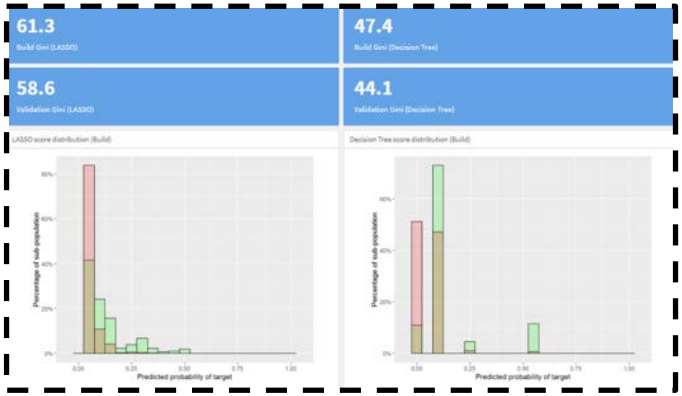
10,000 individuals.
78 LexisNexis Risk Solutions attributes.

LASSO regression, decision tree, gradient boosting, neural network



Hyper parameters for each technique

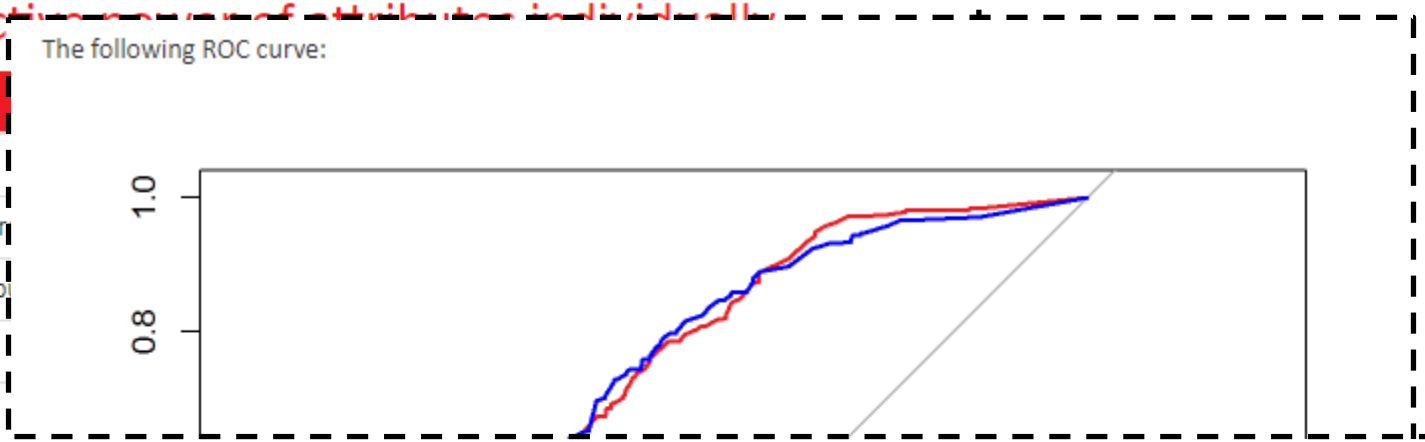




3 minutes

Detailed output capability within formal documentation

3 Predic



8 Comparison of techniques

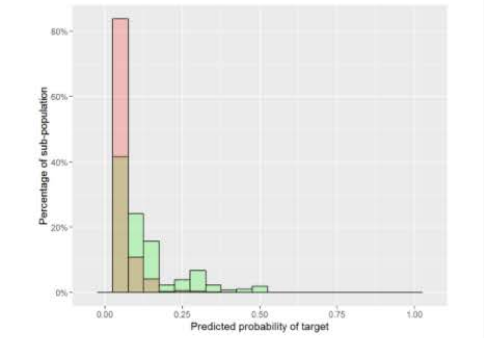
8.1 Model performance

Technique	Build_Gini	Validation_Gini
GBM	69.6	67.2
LASSO regression	61.3	58.6
Neural Net	63.0	56.8
Decision Tree	47.4	44.1

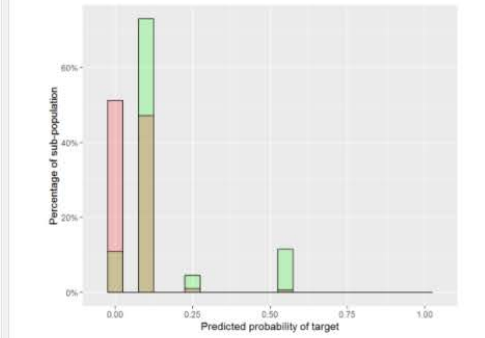
Headlines capability within HTML dashboard for quick review

61.3	47.4	69.6	63.0
Build Gini (LASSO)	Build Gini (Decision Tree)	Build Gini (GBM)	Build Gini (Neural Net)
58.6	**44.1**	**67.2**	**56.8**
Validation Gini (LASSO)	Validation Gini (Decision Tree)	Validation Gini (GBM)	Validation Gini (Neural Net)

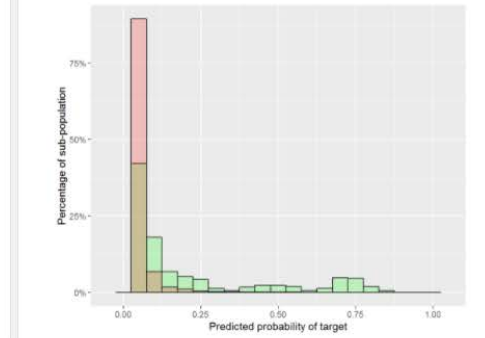
LASSO score distribution (Build)



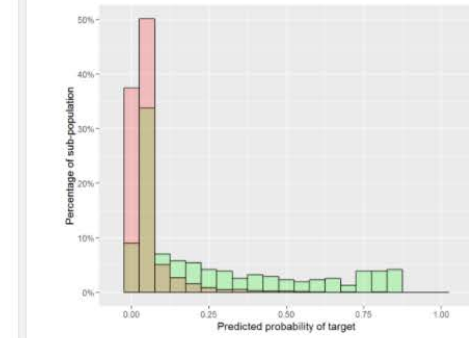
Decision Tree score distribution (Build)



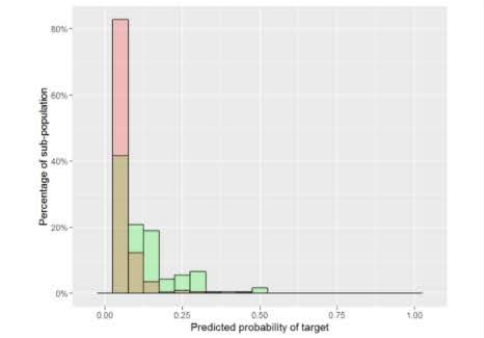
GBM score distribution (Build)



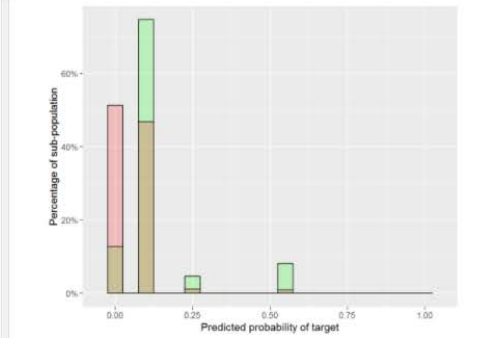
Neural network score distribution (Build)



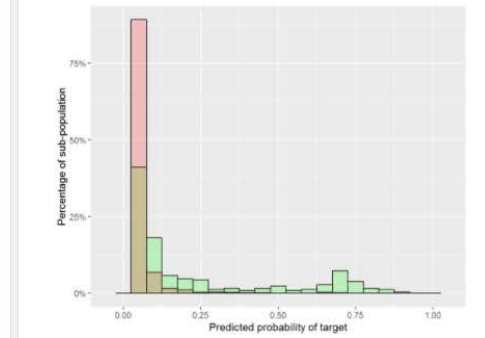
LASSO score distribution (Validation)



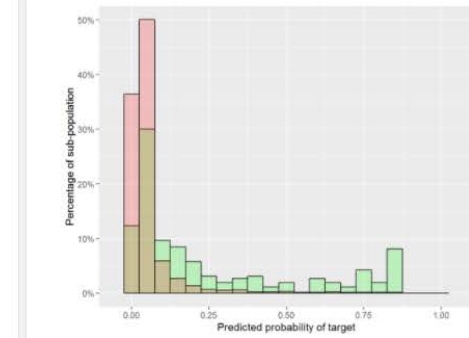
Decision Tree score distribution (Validation)



GBM score distribution (Validation)



Neural network score distribution (VSalidation)



Summary

- It is possible to develop an end-to-end model building process within RStudio, including the model documentation stage.
- Such a process dramatically reduces the time taken to build and document a predictive model.
- It also reduces the risk of human error when producing documentation for peer review and internal audit / industry compliance (FCA, GDPR).
- The process can be future proof in that extra metrics can be easily added, as can modules for new modelling techniques.
- This process has been adopted for developing models in LexisNexis Risk Solutions UK, for internal use and to build custom scores for clients.


LexisNexis®
RISK SOLUTIONS