

01. Introduction

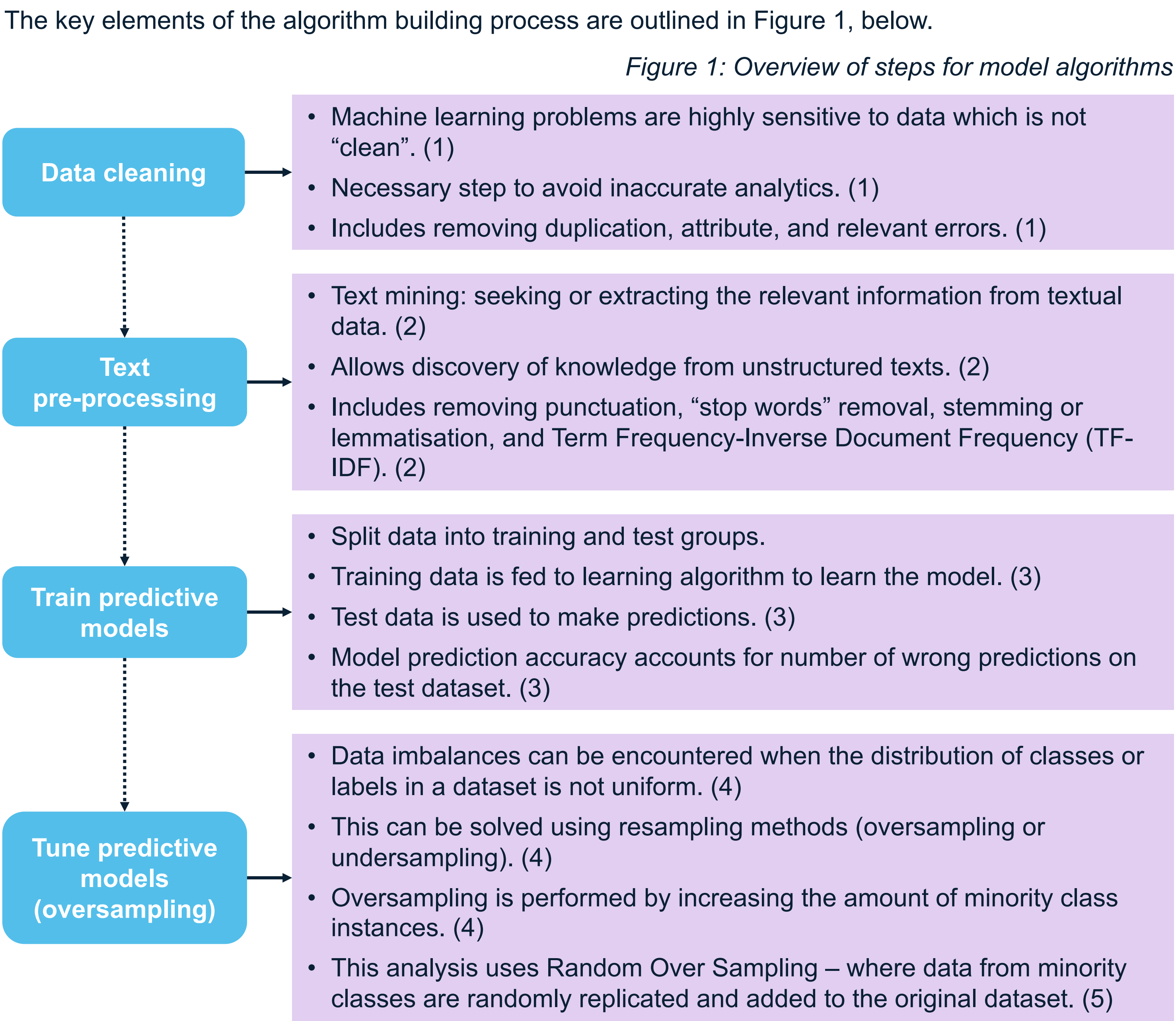
Health Technology Assessments (HTA) of new medicines are conducted in the United Kingdom (UK) and Ireland by the National Institute for Health and Care Excellence (NICE), the Scottish Medicines Consortium (SMC) [UK] and by the National Centre for Pharmacoeconomics (NCPE) [Ireland]. Each HTA agency issue reimbursement recommendations and summary reports.

Natural Language Processing (NLP) is a branch of artificial intelligence that involves the design and implementation of systems and algorithms able to interact through human language. It allows users to work with unstructured text data, gives computers the ability to understand text and spoken words, and ultimately to gain insight on sentiments from text data.

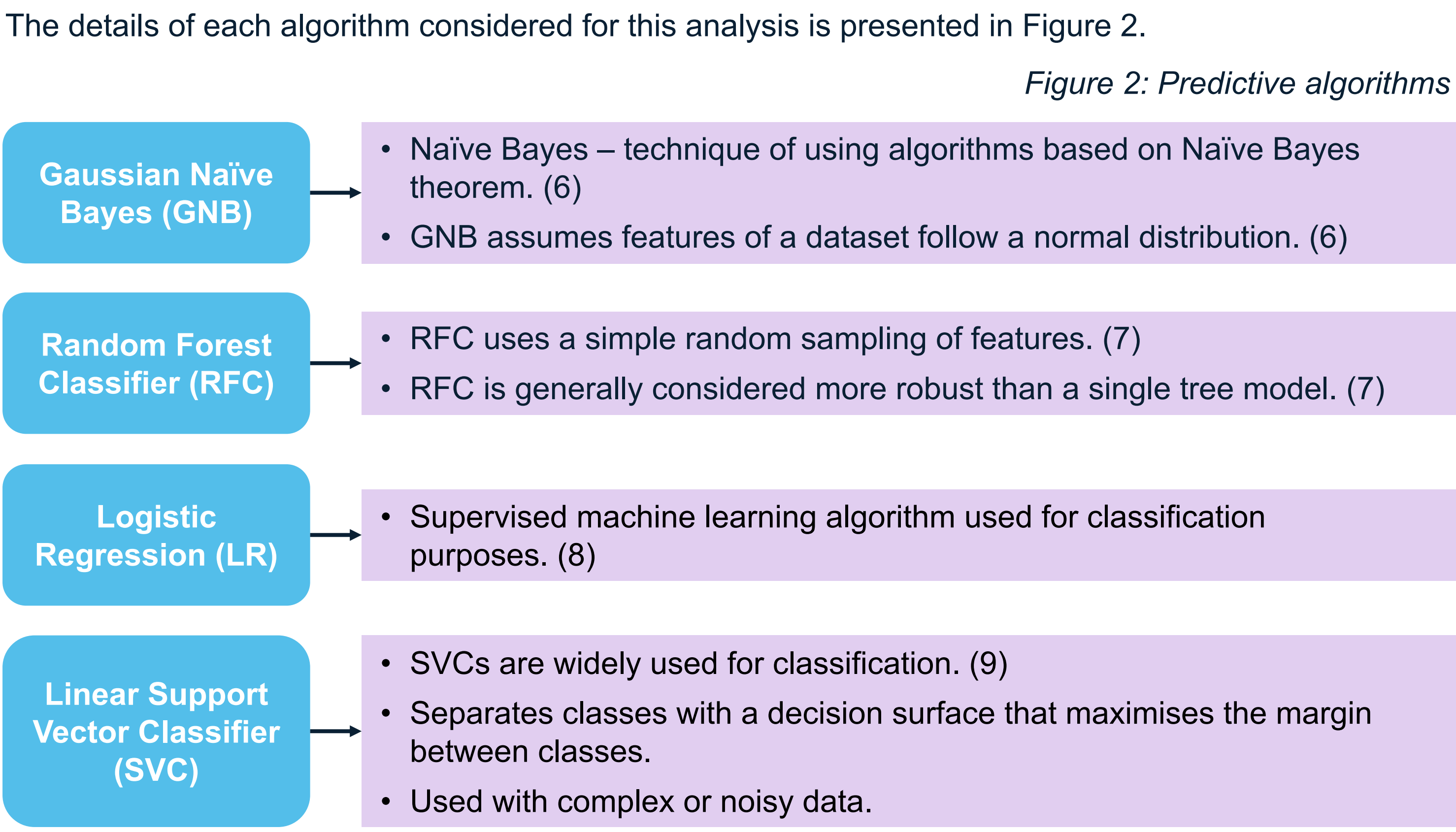
The key objectives of this research are:

1. Build algorithms that predict HTA outcome using the text in HTA agency summary reports.
2. Assess which algorithms perform the best for each jurisdiction.
3. Assess the feature importance of combinations of words within HTA agency summary reports across the three jurisdictions.
4. Discuss possible methods for future improvements in the predictive algorithms.

02. Machine Learning Process



03. Predictive Algorithms



06. Conclusions

The SMC dataset provided the most valuable insights of the three jurisdictions as to the features considered to affect HTA outcome.

Based on the analysis, budget related factors, Patient Access Schemes (PAS) and proposed positioning are among important considerations for HTAs in Scotland.

Similarly, cost effectiveness, Managed Access Protocols (MAPs) and calculations are likely to be important considerations for HTAs in Ireland, based on this analysis assessing the word combinations with the most influence for predicting HTA outcome.

Several challenges were encountered with the NICE data, including the balance between removal of noise from the data and keeping phrases together when extracting text from the PDFs.

Further research is required to continue to improve the predictive algorithms and gather more valuable insights on HTA outcome across the UK and Ireland. These include, but are not limited to:

- Increasing sample size**
To improve model training, evaluation, and performance.
- Researching alternative methods for cleaning the data**
Remove more noise from the data and provide more valuable insights - particularly important for NICE dataset.
- Adjusting the N-gram count** (number of words considered in a singular phrase)
Algorithms can consider longer word sequences when assessing feature importance
- Adjusting training and test split**
To prevent overfitting and accurately evaluate the algorithm’s generalisation ability on unseen data.

04. Methods

Summary reports across the three jurisdictions from 2021 – 2025 were selected and collated in Microsoft Excel® from the NICE (n=299), SMC (n=250) and NCPE (n=110) websites. Text from the reports were extracted from the PDFs ,pre-processed in Python®, with text matrices produced, and four model algorithms created for each jurisdiction.

The process for preparing the data from each jurisdiction is outlined in Figure 3 below.



05. Results

Model Accuracy

The overall accuracy of the models is outlined in Table 1. The RFC algorithm performed had the highest overall accuracy across all three jurisdictions, when considering the original samples. Precision (0.96; 0.79; 0.35) and recall (0.96; 0.84; 0.33) was also higher for RFC than the other algorithms across all three jurisdictions.

When considering precision and recall for each HTA outcome, values for the less frequent outcomes (i.e., reimbursement not recommended) were low. This signalled a class imbalance affected the initial results.

Oversampled Model Accuracy

Oversampling (using the Random Over Sampling method) improved overall accuracy for each model in all three jurisdictions.

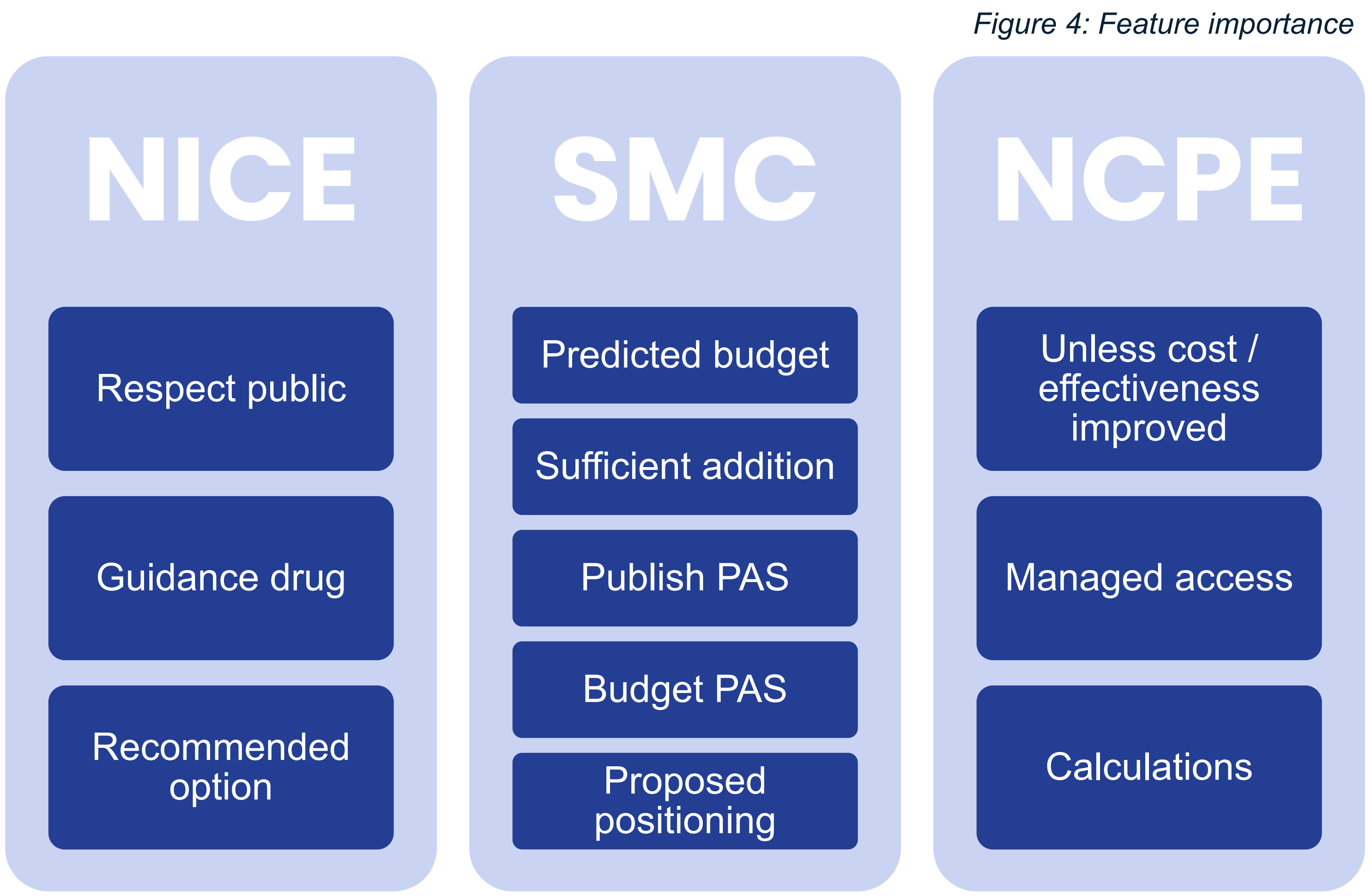
Precision and recall also improved across each model. The RFC remained the algorithm with the highest accuracy for all three jurisdictions.

Table 1: Model accuracy

	NICE	SMC	NCPE
Accuracy - Initial Models			
GNB	0.517	0.60	0.506
RFC	0.928	0.835	0.716
LR	0.574	0.612	0.584
SVC	0.617	0.624	0.611
Accuracy – Oversampled models			
GNB	0.828	0.840	0.903
RFC	0.961	0.90	0.967
LR	0.794	0.771	0.875
SVC	0.843	0.81	0.918

Feature importance

The oversampled algorithm with the highest accuracy (RFC) was utilised to assess feature importance in NICE, SMC and NCPE. The top 20 features (1 to 2-word combinations) with the most influence on predicting HTA outcome were assessed and are depicted below in Figure 4.



References

References for this poster can be found by scanning the QR code below:

