Decision making in Ireland; Utilising machine learning to build predictive algorithms that assess words contributing to NCPE decision making on reimbursement

Gribbon, Emer[1]; Dooley, Brenda[1]
[1]AXIS Healthcare Consulting Ltd, Ireland & UK

01. Introduction

In Ireland, the National Centre for Pharmacoeconomics (NCPE) conduct Health Technology Assessments (HTAs) on certain new medicines on behalf of the national payer, the Health Service Executive (HSE), issuing reimbursement recommendations and a short technical summary report (TSR).

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.[1]

This research aims to build predictive algorithms that predict the outcome of HTAs using the text in TSRs and assess the feature importance of phrases within TSRs, using NLP and predictive machine learning (ML) models.

02. Methods

A database was created using TSRs from the NCPE website. A sample of 115 recently assessed HTA submissions between 2020 and 2024 were selected and collated into an Excel file to train algorithms that predict the four possible NCPE recommendations:

1. **Considered for reimbursement**
2. **Considered for reimbursement if cost-effectiveness is improved**
3. **Not considered for reimbursement unless cost-effectiveness is improved**
4. **Not considered for reimbursement**

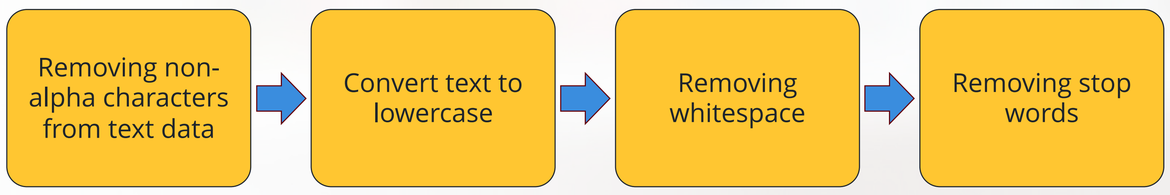
Table 1 provides an overview of the source and extracted data utilised for the analysis, which was subsequently converted into a data frame and analysed in Python®.

Table 1: Overview of source and extracted data

Source	Extracted data points
NCPE Website – Technical summary reports	HTA ID NCPE recommendation Document text

Text Preprocessing

Figure 1: Text preprocessing overview



Sentiment Analysis & Text Matrix

Polarity (the degree of positivity or negativity) was assigned to the text using a sentiment analyser function and the text was normalised using lemmatisation (reducing words based on their context) .

To convert the text data into features, a Term Frequency - Inverse Document Frequency (TF – IDF) was created.

TF – IDF aims to convert text documents into vector models based on the occurrence of words without considering the exact ordering.

- Term Frequency is the count of a term in a document
- Inverse Document Frequency is the logarithm of ratio of total documents available in the corpus and number of documents containing the term.
- TF – IDF formula gives the relative importance of a term in a list of documents.

Predictive Algorithms

To train the predictive algorithms, the variables were sorted into X (text) and Y (HTA outcome) variables. These variables were split into training and test set, with an initial training and test ratio of 0.7:0.3. The optimal training test split varies depending on the dataset, and trial and error to obtain the optimal training test split for a given dataset is often required.

Four predictive algorithms were created, and each respective accuracy assessed in order to obtain the algorithm with the highest accuracy:

1. **Gaussian Naïve Bayes** (GNB)
2. **Random Forest Classifier** (RFC)
3. **Logistic Regression**
4. **Linear Support Vector Classification** (SVC)

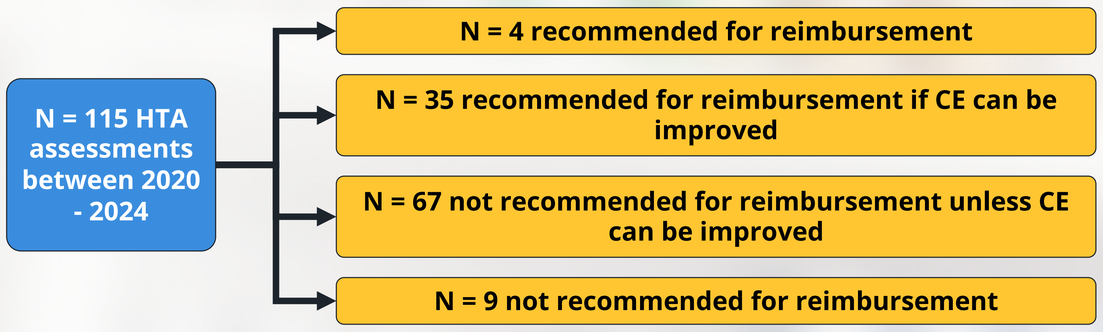
Following initial results, an oversampling technique known as Adaptive Synthetic Sampling was implemented to improve the accuracy of the predictive algorithms.

Adaptive Synthetic Sampling is an algorithm that generates synthetic data to balance datasets.

03. Results

Of the 115 NCPE HTA assessments with identified recommendations between 2020 and 2024, 3.5% (n=4) were recommended for reimbursement, 30.4% (n=35) were recommended for reimbursement if cost-effectiveness could be improved, 58.3% (n=67) were not recommended for reimbursement unless cost effectiveness could be improved, and 7.8% (n=9) were not recommended for reimbursement.

Figure 2: NCPE HTA Outcomes

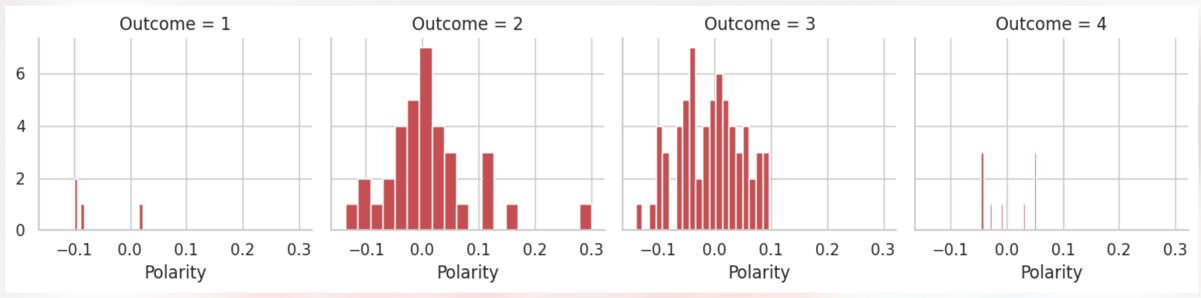


Polarity

Figure 3 below details the polarity in the text across the four different outcomes of a NCPE HTA. The similarities between the degrees of positivity and negativity of the text in Outcome 2 and Outcome 3 signals that it may be difficult to achieve a robust predictive model.

The imbalance in number of assessments with Outcome 1 and 4 compared to those with Outcome 2 and 3 also highlights the challenges that are likely to be encountered when training a predictive model with this dataset.

Figure 3: Visualisation of polarity between NCPE HTA outcomes



Predictive Model Accuracy

Table 2 details the accuracy of the initial predictive models that were created using the 115 HTA assessments between 2020 and 2024, as well as the precision for each outcome. These initial results show that the Logistic Regression model performs best in terms of overall accuracy,.

Table 2: Accuracy of predictive models

Model	GNB	RFC	Logistic regression	Linear SVC
Precision outcome 1	0.25	0.00	0.00	0.00
Precision outcome 2	0.04	0.08	0.00	0.04
Precision outcome 3	0.84	0.77	0.95	0.75
Precision outcome 4	0.14	0.00	0.00	0.00
Model accuracy	0.506	0.493	0.532	0.430

Oversampling Techniques

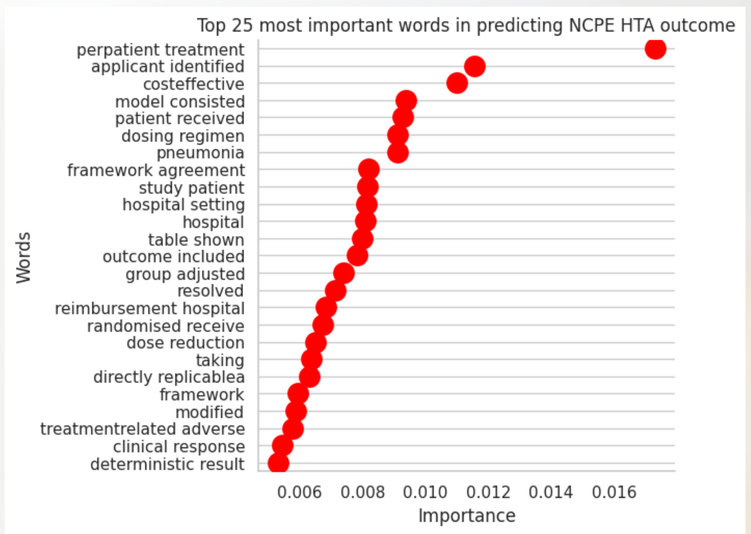
As expected, the results are affected by a class imbalance between the number of assessments for each outcome. The updated results, having implemented adaptive synthetic sampling are detailed in Table 3 below.

Table 3: Accuracy of predictive models with oversampling techniques

Model	GNB	RFC	Logistic regression	Linear SVC
Precision outcome 1	1.00	0.95	0.93	0.93
Precision outcome 2	0.73	0.80	0.82	0.84
Precision outcome 3	0.76	0.69	0.51	0.62
Precision outcome 4	1.00	0.98	1.00	1.00
Model accuracy	0.862	0.820	0.815	0.846

Feature Importance

Figure 4: Feature importance



Using the Gaussian Naïve Bayes predictive model algorithm, Figure 4 details the top 25 most important words that appear in technical summary reports when predicting the outcome of an NCPE HTA assessment.

From the sample utilised, "per-patient treatment", "applicant identified", and "cost-effective" appear to be the strongest predictors of NCPE assessment outcome, when considering the Gaussian Naïve Bayes predictive model algorithm.

04. Conclusions and Recommendations

The analysis of the 115 HTA assessments shows that most HTAs in Ireland require improvement of cost-effectiveness results with over 88% of NCPE HTA assessments between 2020 and 2024 requiring cost-effectiveness to be improved.

As expected, cost-effectiveness, dosing regimens, economic models, treatment related adverse events, and clinical results are likely to influence outcomes, ultimately impacting the final reimbursement decision and acceptable price to the payer. Interestingly, cost-effectiveness and economic models appear to be significantly more influential in outcome compared to clinical response and adverse events.

Although the initial predictive model results were low due to class imbalances, oversampling techniques were able to minimise this impact and significantly improve the model accuracies, as well as the precision of each outcome. Based on the results detailed, the Gaussian Naïve Bayes algorithm has the best performance of the four algorithms and should be considered for further model training.

However, it should be noted that the sample size is relatively low (n=115), and a larger pool of data would allow for more robust predictive models to be trained.

References: 1. Prakash M Nadkarni, Lucila Ohno-Machado, Wendy W Chapman, Natural language processing: an introduction, Journal of the American Medical Informatics Association, Volume 18, Issue 5, September 2011, Pages 544–551;

AXIS Healthcare Consulting Ltd
22 Lower Baggot Street, Dublin 2, D02 X658
85, Great Portland Street, First Floor, London, W1W 7LT, United Kingdom

axishealthcareconsulting.com
+353 (1) 529 2768; +44 7768740344
info@axishealthcareconsulting.com