Reimbursement Decisions in Ireland; Utilising Text Mining to Assess Factors Contributing to Decision Making

Gribbon, Emer[1]; Hernon, Marian[1]; Dooley, Brenda[1]

[1]AXIS Healthcare Consulting Ltd, Dublin, Ireland

01. Introduction

In Ireland, as part of the Health Service Executive's (HSE) process for determining reimbursement, medicines frequently undergo price negotiations with the Corporate Pharmaceutical Unit (CPU) followed by assessment by the HSE Drugs Group.

Data mining techniques can be utilised to extract relevant insights from Drugs Group meeting minutes and used to identify the most common words appearing to help infer what words are associated with a positive or negative recommendation from the HSE.

The aim of this research is to assess the proportion of medicines that are recommended for reimbursement at HSE Drugs Group, analyse the words that appear most frequently when medicines are either recommended for reimbursement or not recommended for reimbursement and identify which Machine Learning models are appropriate to obtain further insights.

02. NLP Overview – Preprocessing techniques

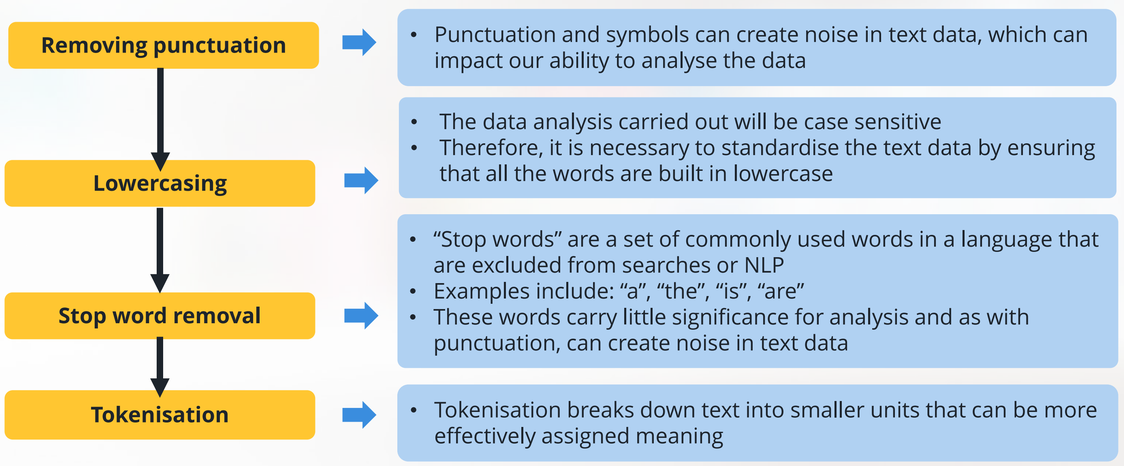
Natural Language Processing (NLP) is a branch of artificial intelligence (AI) which allows users to work with unstructured text data, gives computers the ability to understand text and spoken words, and ultimately to gain insight on sentiments (attitudes and emotions) from text.

There are several necessary steps to prepare text for NLP, these include:

- Removing punctuation.
- Lowercasing words
- "Stop word" removal.
- Tokenising the text.

These steps are outlined in Figure 1 below.

Figure 1: Overview of preprocessing techniques



03. Methods

A database was created using HSE Drugs Group meeting minutes between the period of January 2021 and April 2023.

As the meeting minutes are scanned PDFs, the data was extracted using RStudio®.

- The PDFs were converted to images
- Text was extracted from the images
- The text was saved as a .txt file

The text files were inputted into a centralised database in Microsoft Excel®. The specific data points extracted are detailed in Table 1.

Table 1: Overview of source and extracted data

Source	HSE website – Drugs Group meeting minutes
Extracted data points	• HTA ID • Drug name • Date of Drugs Group meeting • Summary of deliberations • Recommendation from Drugs Group (Reimbursed / Not reimbursed)

Analysis Plan – Frequency of words in Drugs Group deliberations

Using Python®, this data was converted to a data frame. Medicines which did not include a definitive recommendation were excluded from the analysis.

From this, two data frames were created for medicines which were recommended for reimbursement and also medicines not recommended for reimbursement.

In order to analyse the text in the meeting minutes, text cleaning techniques were utilised. The meeting text was extracted from the data frames and the relevant steps in Figure 1 were implemented.

The counts were then obtained for the occurrence of each word for the "recommended for reimbursement" data and "not recommended for reimbursement" data, respectively.

Analysis Plan – Predictive models to be considered for future analysis

The data of 105 medicines considered by the HSE Drugs Group between January 2021 and April 2023 was randomly split into a training and test set, from which 5 preliminary predictive models were built in Python® to infer which models could be used for sentiment analysis in the future.

- 75% of the medicines were randomly assigned to the training set.
- 25% of the medicines were randomly assigned to the test set.

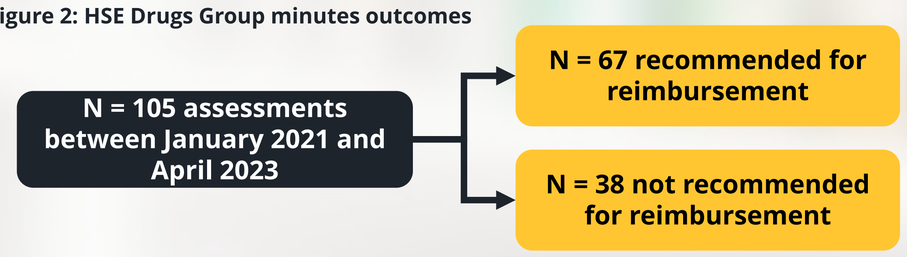
The observations in the training set were used to build models that predict the outcome of the deliberations, and the observations in the test set were used to obtain the accuracy of the models.

5 Predictive models were created: Linear Regression, Logistic Regression, Decision Tree, Random Forest and K-Nearest Neighbours.

04. Results

Of the 105 assessments with identified recommendations between January 2021 and April 2023, 63.8% (N=67) were recommended for reimbursement at drugs group, while 36.2% (N=38) were not recommended, shown in Figure 2 below.

Figure 2: HSE Drugs Group minutes outcomes



Frequency of words

The word frequency analysis shows the patterns in the most frequent words between medicines recommended for reimbursement and those that do not receive recommendation. A breakdown of the most frequent words throughout the data is detailed in the word clouds in Figure 3 and 4 below. Table 3 also details the frequency of some of the most common words per assessment, for each outcome at the HSE Drugs Group.

Figure 3: Most frequent words in positive outcomes

Figure 4: Most frequent words in negative outcomes

Table 2: Frequency of words per Drugs Group assessment

	Frequency (number of times per Drugs Group assessment)	
	Recommended for reimbursement	Not recommended for reimbursement
Evidence	73 (1.09)	43 (1.13)
Clinical	63 (0.94)	55 (1.45)
Treatment	58 (0.87)	43 (1.13)
Noted	45 (0.67)	25 (0.66)
Patients	44 (0.66)	29 (0.76)
Reviewed	36 (0.54)	22 (0.58)
Survival	34 (0.51)	8 (0.21)
Cost effectiveness	34 (0.51)	19 (0.50)
Chemotherapy	6 (0.09)	17 (0.45)
Commercial	33 (0.49)	16 (0.42)
Economic	18 (0.27)	15 (0.39)
Overall	28 (0.42)	14 (0.37)
Trial	26 (0.39)	14 (0.37)

Predictive Model Accuracy

Table 3 details the accuracy of the predictive models that were built using the sample of 105 Drugs Group assessments between January 2021 and April 2023.

Table 3: Accuracy of predictive models

Model	Accuracy
Linear regression	**-0.74**
Logistic Regression	**0.70**
Decision Tree	**0.78**
Random Forest	**0.74**
K-Nearest Neighbours	**0.63**

05. Conclusions and Recommendations

The ability for a medicine to be brought to price negotiations can often result in a medicine receiving a positive recommendation from the HSE Drugs Group for reimbursement, with over two thirds of medicines assessed by the Drugs Group between January 2021 and April 2023 resulting in this outcome.

Several word frequencies are relatively consistent across medicines, regardless of the recommendation (**Evidence, Cost effectiveness, Commercial, Trial**).

It appears that "**clinical**" is relatively more frequent among medicines not recommended for reimbursement (1.45 times per medicine vs 0.94), which may confirm the importance of clinical evidence, when the HSE Drugs Group make decisions.

"**Survival**" is relatively more frequent among medicines which are recommended for reimbursement (0.51 times per medicine vs 0.21), which may confirm the perceived value of survival statistics at HSE Drugs Group deliberations.

The accuracy of the predictive models provide insight as to which models should be considered for further analysis. Based on the data utilised, linear regression would not be an appropriate form of analysis, whereas Decision Trees and Random forests may be more insightful.

However, it should be noted that the sample of data is low (n=105), and a larger pool of data would provide more insight as to the best model to create. Furthermore, other metrics should be investigated such as precision and sensitivity to obtain a clearer picture.

AXIS Healthcare Consulting Ltd
22 Lower Baggot Street,
Dublin 2, D02 X658

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