NLP and machine learning to automate identification of suspected medication errors from real-world unstructured narratives

Jeffery L Painter,[1] François Haguinet,[2] Carolyn Cranfield,[3] Andrew Bate[3]

[1]GSK, Durham, NC, USA; [2]GSK, Wavre, Belgium; [3]GSK, London, UK

Presenter: Jeffery L Painter
E-mail: jeffery.l.painter@gsk.com
Poster code: MSR20

Background and aims

Natural language processing (NLP) allows for the transformation of text into numeric feature vectors which can be used in machine learning (ML) applications.[1]

We determined whether or not unstructured text or narratives (e.g., safety reports) contain mention of a medication error (MedError) by using NLP and ML.

In case a MedError was present, we aimed to sub-classify it as

- error with stated adverse drug reaction (ADR)
- error without harm
- intercepted error
- potential error

This classification is usually rules-based and requires manual review by trained safety scientists.

Main aim: to build ML models for automating this process, first identifying if narratives contain MedErrors or not, then identifying the correct sub-classification.

Methods

Safety reports were extracted from internal GSK safety database between January–November 2021 for training our models.

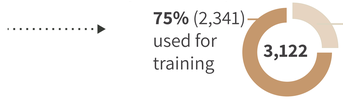
1. Data extracted
22,745 case narratives

2. Manually labeled* by safety scientists

3. Data cleaning
Narratives were processed using text vectorization (TF-IDF[2]), and all product names were masked to reduce potential bias.

4. A **Bernoulli naïve Bayes (binary) classifier**[&] was trained.

Split of data (RW unstructured case narratives)



5. A **multinomial naïve Bayes (MNB) classifier**[3] was built on 1,744 case narratives containing manually annotated sub-classifications.

Performance was evaluated using cross-validation.

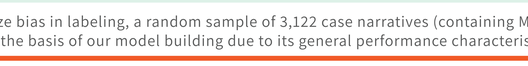
Concordance of labeling (based on a rules-based approach) was compared to an unsupervised clustering method using k-means for narratives not sub-classified.

Results

Binary classifier performance metrics

	Precision	Recall	F1-score	Support
Yes	0.80	0.92	0.86	419
No	0.88	0.73	0.80	362
Accuracy			0.83	781
Macro avg	0.84	0.82	0.83	781
Weighted avg	0.84	0.83	0.83	781

MedError classification



F1-score
for the detection of MedErrors

86% binary classifier

66% MNB classifier (averaged across sub-classes)

Concordance between rules-based and unsupervised clustering had kappa of **0.05**.

Conclusions

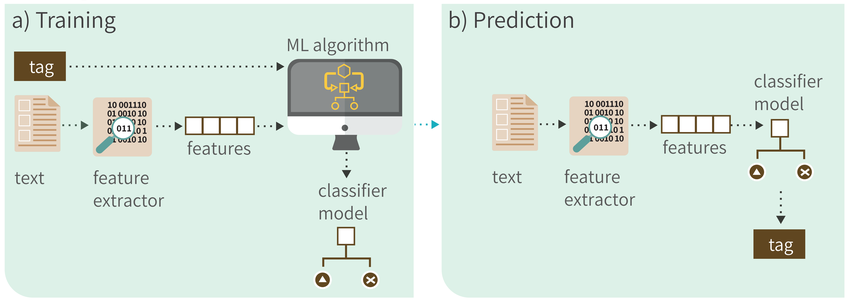
- Binary classification for the detection of MedErrors showed good performance.
- Sub-categorization with this data set was low, except for one sub-class (MedErrors with stated ADR).
- Given the adjudged importance of accurate sub-categorization, the current value added is uncertain. More training data will be required in order to consider its potential use.

PDF Interactive version available

Binary classification could be employed for screening unstructured text to determine whether

MedError is
▲ present
✕ not present

Building the classifier model



Steps to building an NLP and ML model for MedErrors identification

STEP 1

- Data acquisition and formatting case narrative data for use in model building

Data extraction from the ARGUS Oracle database using SQL to develop training and test data sets.

STEP 2

- Data labeling

Narratives were labeled according to their binary class: case or control (not a MedError). For a case, a secondary label indicating the sub-class of MedError found (e.g., with ADR, without harm) was applied.

STEP 3

- Data cleaning

NLP methods generally perform better after "cleaning" the data. A first test showed that product names weighted too much in the multi-label classification. Therefore, the product names were masked for the model building.

STEP 4

- Building a binary classifier

(1) Narrative represents a MedError
or
(2) Narrative does not represent a MedError

STEP 5

- Building a multi-class identifier classifier (conditioned to the binary classifier [step 4])

If a narrative was identified as MedError in step 4, next step was to attempt further sub-classification into the EMA categories (e.g., with ADR, without harm).

RW, real-world; **TF-IDF**, term frequency-inverse document frequency; **avg**, average; **SQL**, structured query language; **EMA**, European Medicines Agency.

*To minimize bias in labeling, a random sample of 3,122 case narratives (containing MedErrors) for ML training and 1,611 controls was collected.
&Chosen as the basis of our model building due to its general performance characteristics[4].

ISPOR 2023, 7–10 May 2023, Boston, MA, USA

Funding: GlaxoSmithKline Biologicals SA. **Acknowledgments**: Medical writing (Lucia Adina Truta), design and coordination support were provided by Akkodis Belgium c/o GSK. **Disclosures, References** available via the QR code.