

OD11 Application Of Natural Language Processing To Predict Final Recommendation Of Brazilian Health Technology Assessment Reports

Marilia Mastrocolla de Almeida Cardoso (marilia.cardoso@unesp.br), Juliana Machado-Rugolo, Lehana Thabane, Naila Camila da Rocha, Abner Barbosa, Juliana Tereza Almeida, Daniel da Silva Pereira Curado, Denis Satoshi Komoda, Silke Anna Theresa Weber and Luis Gustavo Modelli de Andrade

Introduction: Several countries established health technology assessment (HTA) processes to support decision-making. Considering the high volume of submissions processed by HTA agencies, approaches to determine factors associated with the approval would be beneficial. This study aimed to predict the final recommendation of the National Committee for Health Technology Incorporation (Conitec) using a natural language processing (NLP) algorithm for text extraction.

Methods: Conitec's 2012 to 2022 reports (n=389) were split into 75 percent training and 25 percent testing data. Tokenization enabled NLP models: Least Absolute Shrinkage and Selection Operator (LASSO), logistic regression, support vector machine (SVM), random forest, neural network, and Extreme Gradient Boosting (XGBOOST). Evaluation criteria included accuracy, area under the receiver operating characteristic curve (ROC AUC) score, precision, and recall. Cluster analysis with k-modes identified two clusters (group 0 = approved, group 1 = rejected).

Results: The neural network model demonstrated the best accuracy metrics with a precision of 0.815, accuracy of 0.769, ROC AUC of 0.871, and a recall of 0.746. Some tokenization identified that linguistic markers could contribute to the prediction of incorporation decision by the Brazilian HTA Committee, such as international HTA agencies' experience and the government as the main requester. Cluster and XGBOOST analysis identified similar results with approved technologies with a predominance of drugs assessment, mainly requested by the government, and not approved mostly assessing drugs, the industry as the main requester.

Conclusions: The NLP model could identify predictors for the final decision process on the incorporation of health technologies in Brazil's Unified Health System, opening paths for future work using HTA reports coming from other agencies. This model could

potentially improve the throughput of HTA systems by supporting experts with prediction/factors/criteria for approval or nonapproval as an earlier step.

OD12 Large Language Models As Powerful Tools In Health Technology Assessment

Joan Segur-Ferrer (joan.segur@gencat.cat), Carolina Moltó-Puigmartí, Berta Mestre Lleixà, Anna Godo Pla, Roland Pastells Peiró, Laura Llinàs-Mallol, Maria-Dolors Estrada Sabadell and Rosa Maria Vivanco-Hidalgo

Introduction: In recent years, large language models (LLMs) have emerged as powerful tools in natural language processing, presenting diverse opportunities across various sectors. In this context, the Agency for Health Quality and Assessment of Catalonia (AQuAS) is actively investigating how LLMs can enhance the development of health technology assessment (HTA) reports.

Methods: To assess the potential of LLMs in the development of HTA reports, our initial step involved a comprehensive review of technical literature to understand the functionalities of existing LLM tools. This effort was followed by a systematic identification of specific HTA report development tasks that these models could potentially facilitate. We then rigorously evaluated the performance of these tools in executing the delineated tasks, with a focus on meticulously recording the applied prompts and areas of high and low effectiveness, alongside any encountered errors and identified risks.

Results: Among a range of commercial tools, ChatGPT and Microsoft Copilot were selected for exploration due to their accessibility, ease of use, and versatility. Both tools demonstrated acceptable performance in certain tasks (e.g., data extraction), while their efficacy was notably lower in others (e.g., information search). Additionally, a range of risks requiring mitigation were identified (e.g., hallucination or authority bias). Consequently, we drafted a proposal for an internal guide with directives for the technical staff of the HTA unit on utilizing these two tools. Additionally, we constituted a working group.

Conclusions: LLMs have emerged as promising tools in the field of HTA. Over the past months, AQuAS has been investigating the potential of these models to improve the HTA report development process, targeting enhanced efficiency and improved quality. This exploration has led to the identification of numerous opportunities and associated risks within this innovative application.