International Journal of Population Data Science

Journal Website: www.ijpds.org

ATC-ICD: enabling domain experts to explore and evaluate machine learning models estimating diagnoses from filled predictions

Endel, F^{1*} and Weibrecht, N²

¹TU Wien, dexhelpp

²dexhelpp

Introduction

Administrative and reimbursement data from the Austrian health care system is linked and utilized for research and to support policy makers. Lacking standardized, reliable and systematic coding of diagnoses in the outpatient sector, statistical and machine learning models are developed to estimate individual diagnoses coded as ICD-10 based on filled prescriptions (ATC codes), hence called “ATC->ICD models”.

Evaluating the performance of such models, presenting predictions on a global as well as individual level, comparing different technological approaches and establishing trust by providing an intuitive insight into results for non-technical users are the aim of this project

reports.

Conclusion

Presenting ATC->ICD models renders them accessible to data scientists and domain experts. It allows us to collect valuable feedback and gain trust in complex, hard to understand methodologies and results.

Method

ATC->ICD models are presented utilizing interactive web interfaces based on the R shiny package. As one size does not fit all, customized applications are required for different models and points of view. Applying modularization of reoccurring functionality and retaining design principles like a common dashboard layout facilitates the development and training of users. Software containers and centralized infrastructure providing e.g. backup, encryption and authentication enables efficient deployment of new application and their maintenance.

Results

We developed interactive web-based dashboards enabling experts to explore the prediction of single ATC->ICD models and compare the output of different approaches. The possibility to export and annotate results allows us to collect expert opinions, enhance understanding and gain acceptance conveniently. The combination of various dynamic controls, e.g. to filter, search, sort and cluster results, provides flexible access to complex models and large datasets. Linked and interactive graphs and tables help to understand valid and identify erroneous results much faster than with raw output and printed

*Corresponding Author:

Email Address: florian.endel@tuwien.ac.at (F Endel)