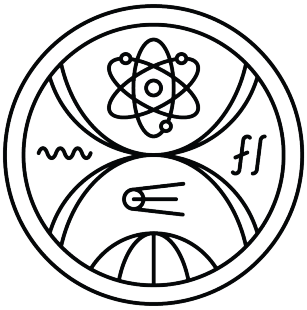
COMENIUS UNIVERSITY IN BRATISLAVA
FACULTY OF MATHEMATICS PHYSICS AND INFORMATICS

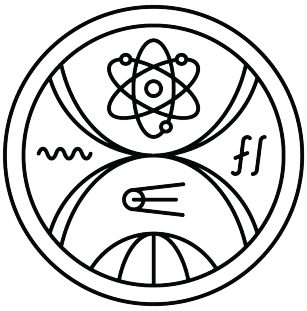


PREDICTION OF HEALTH STATUS DETERIORATION

Master thesis

2025 Bc. Marián Kravec

COMENIUS UNIVERSITY IN BRATISLAVA
FACULTY OF MATHEMATICS PHYSICS AND INFORMATICS



PREDICTION OF HEALTH STATUS DETERIORATION

Master thesis

Study program:	Applied informatics
Branch of study:	Applied informatics
Department:	Department of Applied Informatics
Supervisor:	MSc. František Dráček
Consultant:	

Bratislava, 2025 Bc. Marián Kravec

ZADANIE ZÁVEREČNEJ PRÁCE

Typ záverečnej práce: diplomová
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický

Názov: Predikcia zhoršenia zdravotného stavu
Prediction of Health Status Deterioration

Anotácia: V súčastnosti sa sektor zdravotníctva na Slovensku vyznačuje nizkou mierou využita dostupnych zdravotníckych dát. V rámci tejto prace je cieľom ukázať, že z existujúcjich dát je možné predikovať vyvoj dalšieho zdravotného stavu pacienta, poprípade odhadnúť vývoj budúcich nákladov za účelom lepšieho plánovania prerozdelenia financí v rámci sektoru.

Cieľ: Práca bude rozdelená na dve časti, v prvej študent urobí teoretické zhrnutie existujúchich metód spracovania dát a metód strojového učenia, ktoré sa budú dať potenciálne aplikovať na daný problém. V druhej časti navrhne a aplikuje predičkný model.

Literatúra: T. Sk, L. M. G, L. R. K and R. R. J, "Health Status Prediction using ML Techniques," 2022 6th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2022, pp. 1191-1196, doi: 10.1109/ICCMC53470.2022.9753766.

Jödicke, A.M., Zellweger, U., Tomka, I.T. et al. Prediction of health care expenditure increase: how does pharmacotherapy contribute?. BMC Health Serv Res 19, 953 (2019). https://doi.org/10.1186/s12913-019-4616-x

Vedúci: MSc. František Dráček
Konzultant: Ing. Lukáš Palaj
Katedra: FMFI.KAI - Katedra aplikovanej informatiky
Vedúci katedry: doc. RNDr. Tatiana Jajcayová, PhD.

Spôsob sprístupnenia elektronickej verzie práce:
bez obmedzenia

Dátum zadania: 05.10.2023

Dátum schválenia: prof. RNDr. Roman Ďurikovič, PhD.
garant študijného programu

.....
študent

.....
vedúci práce

I hereby declare that I have written this thesis by myself, only with help of referenced literature, under the careful supervision of my thesis advisor.

.....

Bratislava, 2025 Bc. Marián Kravec

Acknowledgment

WRITE ACKNOWLEDGMENT

Abstract

ABSTRACT EN

Keywords: TODO

Abstrakt

ABSTRACT SK

Kľúčové slová: TODO

Contents

List of Figures

List of Tables

Terminology

Terms

Abbreviations

- **CPT** - Current Procedural Terminology.
- **EHR** - Electronic Health Records.
- **LaBSE** - Language-agnostic BERT sentence embedding model.

Motivation

Chapter 1

Introduction

Chapter 2

Similar studies

One of sub-task for prediction of patient future is to group medical procedures into clusters because there are many procedures that even thought have different codes they are essentially same or similar enough that leaving them separate would only cause issue for predicting model.

For this task Lorenzi et al. from Duke University in Durham developed novel algorithm called Predictive Hierarchical Clustering [2]. This algorithm was developed for agglomerative clustering of surgical CPT codes. This algorithm uses one-pass bottom-up approach where they utilize EHR, more precisely using 317 predictors like lab values and patients history, excluding CPT information for 3,723,252 patients and 3,132 CPT codes where each patient have one main surgical CPT code. For each CPT code then they create tree containing patients with that code. Then at each iteration, the algorithm considers merging all pairs of existing trees. To compare two trees they utilize two hypothesis, first hypothesis say that data in both trees are generated from same model, while second say data in each tree is generated from models with different parameters. Final value is weighted average of probabilities of these two hypothesis considering data in trees, where weigth is probability of first hypothesis 2.1.

$$p(D_k|T_k) = p(H_1^k)p(D_k|H_1^k) + (1 - p(H_1^k))p(D_i|T_i)p(D_j|T_j) \tag{2.1}$$

Where D_k is set of data in merged tree (merged T_i and T_j), T_k is merged tree, H_1^k is first hypothesis, D_i and D_j are data in trees T_i and T_j.

Chapter 3

Medical data

Chapter 4

Proposed method

Chapter 5

Software design

Chapter 6

Implementation

Chapter 7

Research

Chapter 8

Results

Conclusion

REFERENCE SHOWCASE: 3

Bibliography

[1] Vicent Caballer-Tarazona, Natividad Guadalajara-Olmeda, and David Vivas-Consuelo. Predicting healthcare expenditure by multimorbidity groups. *Health Policy*, 123(4):427–434, 2019.

[2] Elizabeth C Lorenzi, Stephanie L Brown, Zhifei Sun, and Katherine Heller. Predictive hierarchical clustering: Learning clusters of cpt codes for improving surgical outcomes. In *Machine Learning for Healthcare Conference*, pages 231–242. PMLR, 2017.

[3] Riccardo Miotto, Li Li, and Joel T Dudley. Deep learning to predict patient future diseases from the electronic health records. In *Advances in Information Retrieval: 38th European Conference on IR Research, ECIR 2016, Padua, Italy, March 20–23, 2016. Proceedings 38*, pages 768–774. Springer, 2016.