- \* - \* - \* - \* - \* - \* - \* - \* - \* - \* - \* - \* - \* - \* - \* - \* - \* - \* -

**AlgorEAThm – a guide**

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

AlgorEAThm © 2021.v1

Table of Contents

* Introduction
* Zip Contents
* Actions
* Demonstrations
  + Detect Malnutrition
  + Label Notes
  + Pre-Process Notes
  + Update Model
  + Review Model
  + Quit
* Contact

Introduction

AlgorEAThm is a model used for detecting the chance of malnutrition for notes from Electronic Health Records. It uses classification into P (chance of malnutrition is present), A (chance of malnutrition is absent) and o (chance of malnutrition cannot be determined).

Graphical user interface, text, application

Description automatically generated

Zip Contents

* AlgorEAThm\_guide: this guide, holds the explanation of AlgorEAThm
* READ\_ME.txt: how to use AlgorEAThm
* algoreathm.py: the script holding the user interface
* <step>\_functions: python scripts holding the functions for AlgorEAThm
  + preprocessing\_functions: functions for pre-processing
  + labelling\_functions: functions for manual labelling
  + vectorization\_functions: functions for vectorization of notes
  + random\_forest\_functions: functions for creating random forest models
  + classification\_functions: functions for using random forest models
* BASE\_<file>: default settings for the model
  + BASE\_gne\_ste\_sg.model: base Word2Vec model
  + BASE\_100131\_processed\_notes.npy: base dictionary with 100131 processed notes
  + BASE\_gne\_ste\_sg\_vectors\_dict.npy: base dictionary with the words from BASE\_100131\_processed\_notes and the corresponding vectors as derived from BASE\_gne\_ste\_sg.model
  + BASE\_gne\_ste\_sg\_model.sav: base Classification Random Forest Model
* NOTEEVENTS\_50notes.csv: a small trial file holding textual notes from EHRs

Actions

When starting AlgorEAThm, the user gets 6 options.

Text

Description automatically generated

‘**Detect Malnutrition**’ performs pre-processing, vectorization and classification of the provided csv file. The results are stored in a txt file but can also be shown on the screen if the user wants that.

‘**Label Notes**’ checks for the provided csv file whether the notes are already labelled. If (some) notes are already labelled, the user is asked whether restarting the labelling or continuing the labelling is wanted. It then shows for each note the original text and asks the user what the chance of malnutrition is, if detectable. The results are stored for further use. The user can quit the labelling whilst unfinished, the results will be stored and the labelling can be continued later.

‘**Pre-Process Notes**’ performs pre-processing, which might be handy when a large dataset is handled because then the pre-processing step is separately executed.

‘**Update Model**’ trains a new Classification Model (Random Forest), calculates its accuracy, shows the accuracies per labelling thresholds over 5 runs and asks the user whether to replace the old model with the new model.

‘**Review Model**’ shows the settings of the model and allows for a manual replace of the model. The user should provide a sav file for the model and manually specify the labelling threshold and (if wanted) accuracy.

‘**Quit**’ quits the model and returns to the command prompt base.

Demonstration

Call AlgorEAThm in the following way:

Text

Description automatically generated

The loading of AlgorEAThm can take a while, because AlgorEAThm needs to load its default settings (models and vectors).

***Detect Malnutrition***

Text

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

***Label Notes***

Text

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

***Pre-Process Notes***

Text

Description automatically generated

Text

Description automatically generated

***Update Model***

This action requires the most care. For example, all the notes in the csv file should have been labelled before updating the model with that file. Also, 5 times plots are brought up and these plots should be (stored and) clicked away before continuing.

Chart

Description automatically generated

Text

Description automatically generated

How to close the graph:

Graphical user interface

Description automatically generated with medium confidence

If the user now reviews the model, the model has changed.

Text

Description automatically generated

***Review Model***

The default settings that show when the user asks to review the model without having applied changes are as follows:

A black screen with white text

Description automatically generated with low confidence

This means that the model uses Google News Embeddings extended with Self-Trained Embeddings using the Skip-Gram algorithm. The labelling threshold is 0.225 and the accuracy of the predictions 71.1%.

The Model can also be changed to a .sav and .npy file in the map algoreathm.

Text

Description automatically generated

***Quit***

Text

Description automatically generated

Contact

Eva Timmer

[*eva.timmer@wur.nl*](mailto:eva.timmer@wur.nl)

*BSc student at Wageningen University & Research*