

Treball final de màster

Estudi: Màster en Enginyeria Industrial

Títol: Estudi de l'efecte nocebo en pacients amb TDAH

Document: 0. Resum

Alumna: Mireia Porta Regué

Tutor/a: Dra. Beatriz López Ibáñez i Dr. Xavier Castells Cervelló

Departament: Enginyeria Elèctrica, Electrònica i Automàtica / Ciències Mèdiques

Àrea: Enginyeria De Sistemes i Automàtica / Farmacologia

Convocatòria (mes/any): Juny/2021

1. INTRODUCCIÓ

La Intel·ligència Artificial (IA) cada cop arriba a més camps, un d'ells és la Medicina, on s'ha demostrat com una eina indispensable per a donar suport en la presa de decisions de molts processos clínics. En particular, l'àrea d'estudi d'Aprenentatge Automàtic de la IA resulta clau per poder extreure coneixement a partir de les dades disponibles.

En aquest projecte es proposa utilitzar tècniques d'aprenentatge automàtic per analitzar les dades d'assajos clínics controlats amb placebo que investiguen l'eficàcia del tractament farmacològic per a malalts amb Trastorn per Dèficit d'Atenció amb Hiperactivitat (TDAH). Concretament se centra en l'estudi de l'efecte nocebo, els esdeveniments adversos en pacients als quals s'ha administrat placebo: una substància inert.

L'objecte d'aquest estudi és investigar per mitjà del mètode d'aprenentatge automàtic MetaForest les causes de la variabilitat de la resposta al nocebo entre els assaigs clínics que han investigat l'eficàcia i la seguretat del tractament farmacològic del TDAH, i determinar la fiabilitat de MetaForest en la predicció de l'efecte nocebo.

L'abast del projecte és:

- Estudiar les variables per tal de determinar quines tenen més importància sobre l'efecte nocebo.
- Predir, a partir de les dades analitzades, l'efecte nocebo en assaigs clínics amb malalts amb TDAH.
- Determinar el rendiment del mètode MetaForest en la predicció dels efectes adversos al placebo.

2. METAFORREST

El mètode MetaForest és una tècnica capaç d'explorar l'heterogeneïtat entre estudis i realitzar una selecció de variables, identificant moderadors rellevants dins d'un conjunt més ampli de candidats, sense provocar *overfitting* en el model. És una adaptació de l'algoritme Random Forest per a metanàlisis (Van Lissa, 2017).

Per tal d'utilitzar el mètode MetaForest cal aplicar un esquema de pesos al mostreig bootstrap, que significa que els estudis més precisos tenen més influència en l'etapa de construcció del model (Van Lissa, 2017).

Quan s'utilitza MetaForest es proporcionen dues mesures d'estimació del rendiment predictiu, R^2_{oob} i R^2_{cv} (Breiman, 2001).

Amb aquest mètode també s'obté la importància de les variables. El seu avantatge és que els valors d'importància les variables, en lloc de mesurar la relació entre cada moderador i la sortida de manera lineal, univariant i parcial, mesuren la contribució al poder predictiu final capturant també les relacions no lineals.

Per últim, també es pot analitzar la contribució de cada moderador a la mida de l'efecte mitjançant gràfiques de dependència parcial.

3. APLICACIÓ PRÀCTICA I CONCLUSIONS

3.1. Aplicació pràctica

L'estudi consisteix en dues etapes: una primera etapa de preparació de dades per poder utilitzar el mètode MetaForest per predir l'efecte nocebo. I una segona etapa on es construeix el model de predicció. A la Figura 1 s'observen els passos a seguir separats en les dues etapes mencionades.

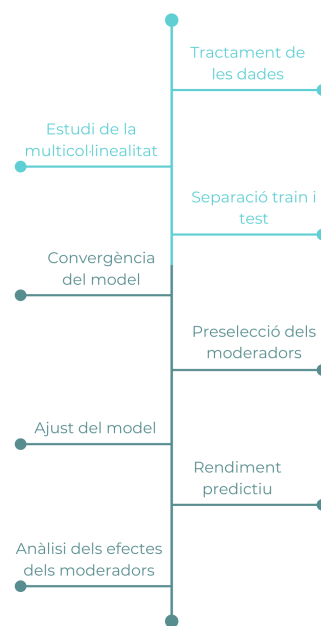


Figura 1. Diagrama del procés

Es realitza, en primer lloc, el tractament de les dades per tal de poder aplicar el mètode d'aprenentatge automàtic MetaForest. Començant per l'estudi de les variables i seguit de l'estudi de la multicol·linealitat, s'observa que cap variable presenta multicol·linealitat i no és necessari descartar-ne cap per poder aplicar el mètode. Tot seguit se separen les dades en dos conjunts, d'entrenament i de prova, per poder comprovar el rendiment del model un cop aplicat l'algoritme.

En la segona etapa, per a la construcció del model de predicció mitjançant MetaForest s'usa la metodologia descrita per Van Lissa (Van Lissa, 2018). Es comprova si el model convergeix per trobar el nombre òptim d'arbres de decisió i així accelerar el procés computacional, el seu comportament es pot observar a la Figura 2. La convergència s'analitza mitjançant

l'estabilització de la mitjana acumulada al quadrat de l'error de predicció out-of-bag en funció del nombre d'arbres del model.

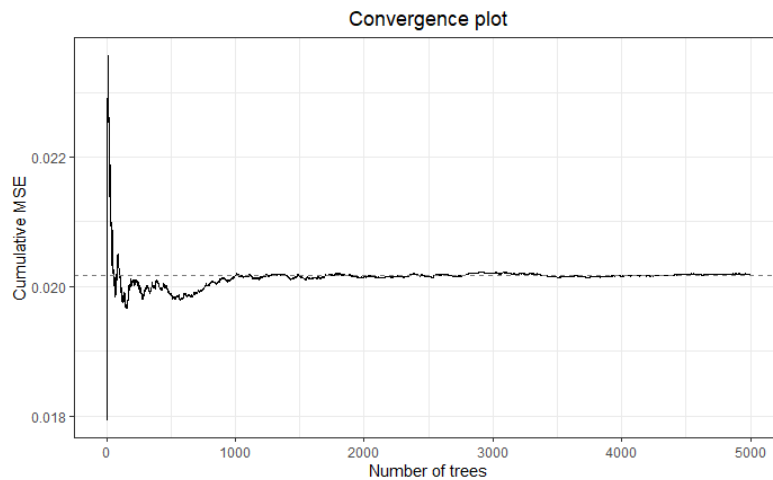


Figura 2. Gràfica de convergència del conjunt de dades d'entrenament

A continuació, com s'observa a la Figura 3, es realitza una preselecció de moderadors per tal d'evitar l'*overfitting* i es descarten aquells que no aporten valor a la predicció. La preselecció es basa a replicar l'anàlisi, visualitzar la distribució de la importància dels moderadors i filtrar els moderadors que tenen una importància negativa en la majoria de les rèpliques.

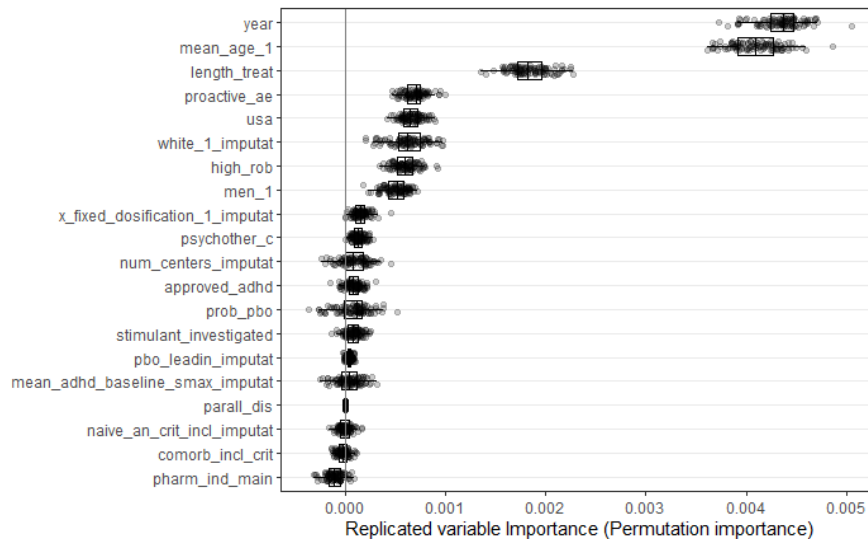


Figura 3. Importància de les variables del conjunt d'entrenament

Tot seguit s'ajusta el model mitjançant el paquet de R CARET (Kuhn, 2008). Com a paràmetres d'ajust es consideren els tres tipus de pesos (uniformes, fixes i aleatoris), el nombre de variables candidates a cada divisió de l'arbre (de 2 a 10) i la mida mínima del grup posterior (de 2 a 10).

Com a model final se selecciona el model amb un valor RMSE més baix, basat en una validació creuada de k-folds. Un cop obtingut el millor model pel conjunt d'entrenament, es comprova amb el conjunt de prova per tal de saber quin és el seu rendiment predictiu.

Finalment, com s'observa a la Figura 4, es procedeix a aplicar el model a tot el conjunt de dades i s'elabora l'anàlisi dels efectes dels moderadors. S'examina la importància de les variables i les gràfiques de dependència parcial, aquestes no mostren gaire variació en la mida de l'efecte.

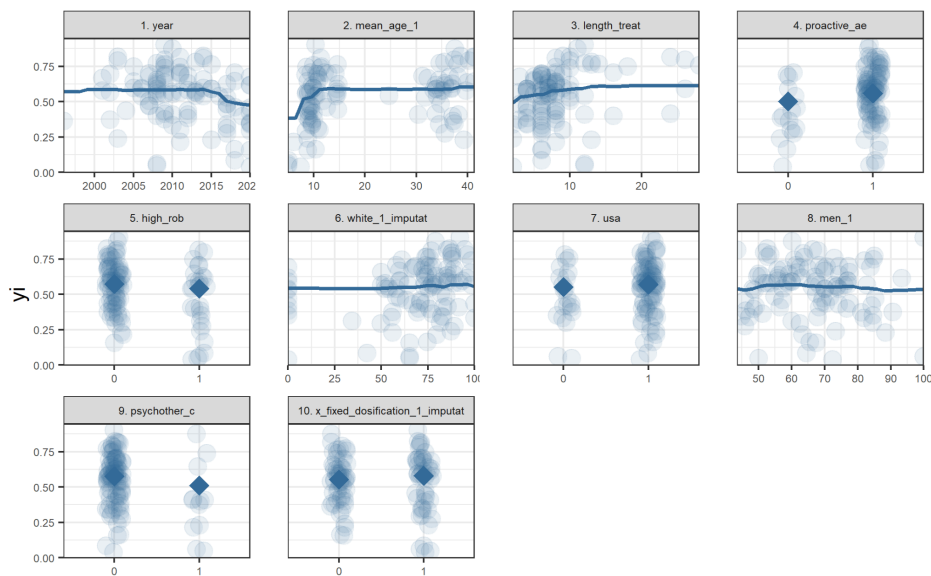


Figura 4. Gràfiques de dependència parcial

3.2. Conclusions

L'estudi de l'efecte nocebo en pacients amb TDAH conclou amb la detecció de les possibles causes de la variabilitat de la resposta al nocebo entre els assaigs clínics analitzats. Tot i que la mesura d'importància de les variables indica que no tenen uns efectes molt grans sobre el resultat, les variables més importants són: la mitjana d'edat, l'any de publicació, la durada del tractament i la proactivitat.

Malauradament els resultats de les prediccions del mètode MetaForest no són molt satisfactoris, ja que el rendiment del model és baix, per sota del 50%. Aquest comportament és normal en estudis del camp de la medicina, ja que el comportament humà és imprevisible. En un futur es podria estudiar la implementació d'altres algorismes i ampliar la base de dades per arribar a un rendiment predictiu més elevat.