

Presentació del conjunt de dades escollit:

En un estudi realitzat fa 10 anys, els autors van analitzar més de 100 000 pacients de diversos hospitals dels USA per trobar patrons de futures re-hospitalitzacions (Strack et al., 2014). L'article es pot trobar al següent [link](#). Les dades originals es poden descarregar al material suplementari del mateix article.

Un segon set de dades similars, encara que limitat a 25 000 pacients i simplificades, va ser presentat per una competició a [DataCamp](#). Aquest és el conjunt de dades que es proposa per fer el projecte final

Característiques generals:

El set de dades escollit s'ha extret de l'àmbit hospitalari. Estan centrades en les hospitalitzacions i re-hospitalitzacions de pacients. Es consideren diferents paràmetres com ara el número de procediments aplicats o el número de test de laboratori.

Una part important de les dades es fixa en la diagnosi de diabetis. Una determinació de laboratori molt important és la valoració de la concentració de glucosa en sang. Si aquest valor és alt ja indica que el pacient és diabètic.

Per pacients ja diagnosticats amb diabetis, és molt important fer el test HbA1c. Aquest test mesura la hemoglobina glicosilada, un paràmetre que descriu els nivells de glucosa en sang als últims 3 mesos. Aquest valor és independent del moment del dia en que es realitza el test de laboratori, a diferència del que passa amb les determinacions de glucosa en sang, que es requereix que el pacient estigui en dejú.

Les dades han estat extretes de la pàgina DataCamp.com. Es van fer servir per una competició, i s'han publicat els resultats obtinguts per altres usuaris d'aquesta pàgina web. El treball que es farà es basarà en tot els continguts apresos durant l'itinerari, i no en els treballs d'altres.

Definició de les variables:

La definició de totes les variables incloses al data set es detalla a la següent taula:

VARIABLE	TIPUS	DESCRIPCIÓ DE LA VARIABLE
age	object	grup d'edat del pacient
time_in_hospital	int64	en dies (de 1 a 14)
n_procedures	int64	número de procediments realitzats al pacient durant la seva estada a l'hospital
n_lab_procedures	int64	número de procediments de laboratori realitzats al pacient durant la seva estada a l'hospital
n_medications	int64	número de medicacions administrats al pacient durant la seva estada a l'hospital
n_outpatient	int64	número de visites ambulatories durant l'any anterior a l'hospitalització
n_inpatient	int64	número de visites d'hospitalització durant l'any anterior a l'hospitalització
n_emergency	int64	número de visites d'urgències durant l'any anterior a l'hospitalització
medical_specialty	object	especialitat mèdica per la qual es va ingressar el pacient
diag_1	object	diagnòstic primari
diag_2	object	diagnòstic secundari
diag_3	object	diagnòstic secundari addicional
glucose_test	object	resultat de anàlisi de glucosa al sèrum (alt (>200), normal, not performed)
A1Ctest	object	resultat del test A1C (alt (>7), normal, not performed)
change	object	mostra si hi ha hagut un canvi a la medicació ('yes' or 'no')
diabetes_med	object	mostra si es va receptar medicació per la diabetis ('yes' or 'no')
readmitted	object	mostra si el pacient va ser readmès a l'hospital ('yes' or 'no')

Taula 1: Informació del dataset

A la taula 2 es presenta les 5 primeres entrades del data set

	age	time_in_hospital	n_lab_procedures	n_procedures	n_medications	n_outpatient	n_inpatient	n_emergency	medical_specialty	diag_1	diag_2	diag_3	glucose_test	A1Ctest	change	diabetes_med	readmitted
0	(70-80)	8	72	1	18	2	0	0	Missing	Circulatory	Respiratory	Other	no	no	no	yes	no
1	(70-80)	3	34	2	13	0	0	0	Other	Other	Other	Other	no	no	no	yes	no
2	(50-60)	5	45	0	18	0	0	0	Missing	Circulatory	Circulatory	Circulatory	no	no	yes	yes	yes
3	(70-80)	2	36	0	12	1	0	0	Missing	Circulatory	Other	Diabetes	no	no	yes	yes	yes
4	(60-70)	1	42	0	7	0	0	0	InternalMedicine	Other	Circulatory	Respiratory	no	no	no	yes	no

Taula 2: heading del dataset

Presentació dels objectius:

L'objectiu principal del projecte es el de intentar predir els episodis de re-hospitalització. El principal factor que es pot considerar a partir d'aquest dataset es el del diagnòstic de diabetis, tant a nivell de test (glucosa o HbAc1, com de medicació. També s'ha de considerar totes les dades referents a l'any anterior, la seva durada i el seu cost en intervencions i proves de laboratori.

Per aquest projecte els objectius específics seran els següents:

- Determinar si hi ha diferencia significatives entre el pacients re-hospitalitzats i no pel que fa a totes les diferents variables.
- Fer una classificació del pacients segons les dades prèvies i determinar fins quin nivell es vàlid per predir la re-hospitalització
- Determinar fins quin punt el diagnòstic de diabetis pot ser determinant per predir la re-hospitalització futura dels pacients

Referències

Strack, B. et al. (2014) "Impact of HbA1c measurement on hospital readmission rates: Analysis of 70,000 clinical database patient records," BioMed Research International, 2014, pp. 1–11. Available at: <https://doi.org/10.1155/2014/781670>.