Project Description Doc

A Data Driven Tool to Assess Clinical Outcome Risk

Team 62

María Paula Álvarez, Cristian Rodríguez, Luis Daniel Chavarría,
Jeyson Guzmán, Juan Barrios and Luis Serna.

Correlation One | MinTIC
Data Science for All Colombia 6.0
Week 4 Project Submission
May 2022

In this document we present the first approximation to our project description.

Problem overview

Medical treatment has had a great evolution throughout history, from the prevention until the treatment of diseases that previously had no cure. Nevertheless, the time it takes for doctors to identify diseases and begin the right treatment has been a race against time since a long time ago. Developed countries which have more resources have more capacity to identify these problems earlier than developing countries where a huge gap between the quality and the time of response can be seen.

Nowadays plurpathologic patient cohort management is done reactively instead of proactively. This means that people have to schedule their controls themselves and doctors have to usually fix damage that could’ve been prevented.

Problem definition

With this project we attempt to construct a data-product able to assess the risk of different clinical outcomes such as hospitalization or death risk. This product will provide reliable feedback on a cohort of patients and identify who has the highest risk of a particular clinical outcome.

Medical outcomes assessment is a current research problem in public health focused on the estimation of clinical results or effects that different treatments or interventions can have in a patient population. The early determination of these outcomes can help to bring an adequate response for each patient, better focused on its symptoms, functional status, and quality of life, as well as reduce healthcare costs by taking a preventive approach and drawing out possible risk factors.

In this context, we intend to answer the question: what are the medical outcome predictors in a sample Colombian patient population based on previously gathered clinical information?

Importance of the problem

Medicine has come a long way since its inception and its practices have had drastic changes throughout history. In ancient times their applications were mainly rituals or involved the usage of plants but nowadays thanks to technology, investigation and artificial intelligence, etc. we are witnesses of the medical advance and new techniques that contribute more efficiently to the protection of health, the cure of diseases, also to improve the quality of health services and prevention of health risk.

The quality of health services is one of the main concerns of governments around the world. For the Ministry of Health in Colombia, it is a priority to guarantee access and quality care in which the minimum conditions for a service as important as this one are met. The Colombian law 1431 of 2011 states “The General System of Social Security in Health will be aimed at generating conditions that protect the health of Colombians, being the well-being of the user the central axis and articulating nucleus of health policies”.

Considering the previous description, we can develop a predictive model to measure the attention scales, hospitalization time and reduce the probability of having patients unfortunately die.

Data sets to use

We have been provided anonymized pluripathologic patient data for a 1 year follow-up program from a healthcare provider in Colombia. Talks to connect to the hospital databases and register a formal project in the institution are currently on-going, but meanwhile we were allowed to use the anonymized data.

- **Data files:**

Anonymized pluri-pathologic patient data: Contains data for 5511 patients of a health provider institution in the entire year 2020. 1 row per patient. There are 137 variables for each one of them, some functional, some related to the use of resources and some other about chronic pathologies. No data dictionary has been provided, and there is some missing data for some patients. The name of the columns is self-explanatory for some of the variables.

```
[1] "edad"
[2] "sexo"
```

```
 [3] "n_urg"
 [4] "n_hosp"
 [5] "n_cx"
 [6] "n_ce_med_general"
 [7] "n_ce_med_especializada"
 [8] "estancia_sala"
 [9] "estancia_uce"
[10] "estancia_uci"
[11] "pancreatobiliar"
[12] "glaucoma"
[13] "hipertension"
[14] "cerebrovascular"
[15] "vertigo_y_alteraciones_auditivas"
[16] "bradicardias_y_enfermedades_de_la_conduccion"
[17] "otras_genitourinarias"
[18] "depresion_y_alteraciones_del_animo"
[19] "epoc"
[20] "enfermedad_renal_cronica"
[21] "enfermedades_de_la_tiroides"
[22] "otras_enfermedades_digestivas"
[23] "hematologicas"
[24] "alteraciones_otorrinolaringologicas"
[25] "obesidad"
[26] "enfermedades_de_la_prostata"
[27] "esquizofrenia"
[28] "vascular_periferica"
[29] "alteraciones_de_la_agudeza_visual"
[30] "cromosomicas"
[31] "osteoporosis"
[32] "infecciones_cronicas"
[33] "anemia"
[34] "somatomorfos"
[35] "otras_enfermedades_oculares"
[36] "colitis_y_gastrointestinales_inferiores"
[37] "artrosis"
[38] "enfermedad_isquemica_cardiaca"
[39] "otras_respiratorias"
[40] "cataratas"
[41] "hepatopatia_cronica"
[42] "otras_neurologicas"
[43] "otras_psiquiatricas"
[44] "valvulares"
[45] "otras_dermatologicas"
[46] "autoinmunes"
[47] "trastornos_del_sueño"
[48] "enfermedad_venosa_y_linfatica"
[49] "migraña_y_sindromes_faciales_dolorosos"
[50] "dislipidemia"
[51] "esclerosis_multiple"
[52] "demencias"
[53] "alergia"
```

```
 [54] "asma"
 [55] "ulceras"
 [56] "enfermedad_inflamatoria_intestinal"
 [57] "neuropatias_perifericas"
 [58] "epilepsia"
 [59] "artropatías_inflamatorias"
 [60] "lumbalgia_cronica"
 [61] "otras_enfermedades_cardiovasculares"
 [62] "parkinson"
 [63] "cancer"
 [64] "fibrilacion_auricular"
 [65] "diabetes"
 [66] "falla_cardiaca"
 [67] "gastrointestinales_superiores"
 [68] "fact_otros"
 [69] "fact_estancias"
 [70] "fact_ayudas_dx"
 [71] "fact_interconsultas"
 [72] "fact_medicamentos_insumos"
 [73] "fact_otros_conceptos"
 [74] "fact_procedimientos_paquetes_qx"
 [75] "fact_rondas"
 [76] "fact_banco_sangre"
 [77] "fact_planta_oxigeno"
 [78] "fact_consulta_externa"
 [79] "fact_salud_oral"
 [80] "fact_total"
 [81] "peso"
 [82] "talla"
 [83] "saturacion_oxigeno"
 [84] "perímetro_muslo"
 [85] "perímetro_cintura"
 [86] "pliegue_triceps"
 [87] "pliegue_abdomen"
 [88] "pliegue_muslo"
 [89] "presión_arterial_sistólica"
 [90] "presión_arterial_diastólica"
 [91] "frecuencia_cardíaca_en_reposo"
 [92] "auto-calificacion_nivel_de_ejercicio"
 [93] "constantes"
 [94] "mets_-índice_metabólico"
 [95] "vo2_-_máxima_cantidad_de_oxígeno"
 [96] "índice_de_fragilidad_groningen"
 [97] "calificacion_indicefragilidad"
 [98] "tiempo_segundos_monopodal"
 [99] "calificacion_apoyo_monopodal"
[100] "tiempo_segundos_5metros"
[101] "calificación_velocidad"
[102] "indice_tobillo_brazo"
[103] "diabetes_mellitus"
[104] "tipo_diabetes_mellitus"
```

```
[105] "es_insulinorequiriente"
[106] "tiempo_con_el_diagnóstico"
[107] "glicemia"
[108] "hemoglobina_glicada"
[109] "control_diabetes"
[110] "tiene_hta"
[111] "control_hta"
[112] "tiempo_con_el_diagnóstico2"
[113] "epoc_bodex"
[114] "enfermedad_coronaria"
[115] "insuficiencia_cardíaca"
[116] "valvulopatía"
[117] "arritmia_o_paciente_con_dispositivo"
[118] "sufre_de_alguna_enfermedad_cardiovascular"
[119] "tabaquismo"
[120] "cuantos_cigarrillos_día"
[121] "años_de_consumo"
[122] "lipoproteina"
[123] "hdl"
[124] "colesterol_total"
[125] "trigliceridos"
[126] "clasificación_de_framinghan"
[127] "creatinina_1_consulta"
[128] "tasa_de_filtración_glomerular_tfg"
[129] "estadio_de_la_enfermedad_renal"
[130] "microalbuminuria"
[131] "tsh"
[132] "clase_funcional"
[133] "creatinina_2_consulta"
[134] "tasa_de_filtración_glomerular_tfg3"
[135] "cambio_de_tfg"
[136] "úlcera_de_pie_diabético"
[137] "estado_vital"
```