**Diabetes en adultos**

**Mayores de Tierra del Fuego**

**1. Resumen**

Los consensos internacionales respecto del tratamiento de la Diabetes demuestran la conveniencia de enfoques intensificados y un abordaje temprano de los mismos.Sin embargo, en los pacientes ancianos y con menor expectativa de vida los riesgos a corto plazo de los tratamientos pueden pesar más que el posible beneficio a largo plazo.

Como no es posible determinar la expectativa de vida de pacientes individuales, se trata de desarrollar y validar un modelo matemático que pronostique índices de mortalidad a 5 y 10 años para adultos mayores con diabetes, con o comorbilidades.

**2. Objetivos del modelo**

Este modelo busca encontrar qué variables socioeconómicas y de salud afectan

más fuertemente esta tasa de mortalidad.

El principal beneficio recae sobre el sistema de salud, ya que reduciría los costos al

evitar aplicar tratamientos más caros y complejos a pacientes que no van a obtener

el beneficio esperado dada su reducida esperanza de vida.

Por otro lado, esos pacientes no estarían sometidos al riesgo adicional de aplicar

esas alternativas terapéuticas, que en ningún caso son inocuas.

**3. Descripción de los datos**

Se trabaja con un dataset obtenido de una clínica de Río Grande, Tierra del Fuego que tiene sede en Ushusaia y Tolhuin , que recoge

información sobre veteranos de guerra de 65 o más años de edad al 01/01/2006.

Este dataset incluyó pacientes que cumplieron los siguientes criterios:

● Los pacientes tenían 65 años o más de edad al 1/1/2006.

● Ingresados al programa de atención de salud a Veteranos de las Fuerzas

Armadas.

● Todos tenían diabetes diagnosticada clínicamente al momento de su

inclusión en el ensayo.

● Al menos una visita al médico con controles registrados para tres parámetros

típicos de la diabetes (Tensión arterial, BMI y Hemoglobina A1c) en los 24

meses anteriores al estudio.

Por otro lado, se aplicaron las siguientes consideraciones al recabar los datos:

2

● El sexo sólo se considera binario Hombre/Mujer.

● El estado civil sólo se considera Casado, Soltero (nunca casado o divorciado)

o Viudo.

● Se definieron ocho grupos de prioridad para la inclusión en el estudio

teniendo en cuenta niveles de discapacidad física y dificultades

socioeconómicas.

A continuación se describen las variables que lo componen:

● SEX: 0=Mujer, 1=Hombre

● AGE: Edad en años

● RACE:

○ 1=Blanca

○ 2=Negra

○ 3=Otra

● MARRIED: Estado civil

○ MARRIED=Casado

○ SINGLE=Soltero (nunca casado o divorciado)

○ WIDOWED=viudo

● PRIORITY: Grupo de prioridad para la inclusión en el estudio.

○ 1: Discapacidad de 50% o mayor relacionada con el servicio.

○ 2: Discapacidad de 30 - 40% relacionada con el servicio.

○ 3: Discapacidad de 10 - 20% relacionada con el servicio.

○ 4: Discapacitados catastróficos - permanencia obligada en su hogar.

○ 5: Dificultades económicas.

○ 6: Discapacidad no relacionada con el servicio, otros.

○ 7: Sin discapacidad relacionada con el servicio ni dificultades

económicas

○ 8: Sin discapacidad relacionada con el servicio ni dificultades

Económicas

● N\_IP: Número de días de internación.

● N\_OP: Cantidad de consultas en consultorios externos.

● DEATH\_5: Muerto dentro de los 5 años siguientes a la entrevista.

● DEATH\_10: Muerto dentro de los 10 años siguientes a la entrevista.

● FRAILTY: Índice de fragilidad. Indica la presencia o ausencia de treinta

problemas de salud relacionados con la edad1.

● BMI: Índice de Masa Corporal (Peso / (altura en metros)2).

● SYSTOLIC: Presión arterial sistólica (máxima)

● DIASTOLIC: Presión arterial diastólica (mínima)

● TRI: Triglicéridos en sangre.

● LDL: Colesterol LDL en sangre.

● HDL: Colesterol HDL en sangre.

● A1C: Concentración de Hemoglobina A1c en sangre. Indica el control a largo

plazo de la diabetes

● MICROALB: Microalbuminuria en orina. Evalúa el estado de los riñones.

● SERUMALB: Albúmina en sangre.

● SERUMCRE: Creatinina en sangre.

● INSULIN: Prescripción de insulina: 0=no, 1=sí.

● ALPHA: Prescripción de inhibidor de la alfa-glucosidasa: 0=no, 1=sí.

● TZD: Presicripción de Tiazolidinedionas: 0=no, 1=sí.

● BIGUAN: Prescripción de Biguanidas: 0=no, 1=sí.

● SULF: Prescripción de Sulfonilureas: 0=no, 1=sí.

● OTHER\_MED: Prescripción de otros medicamentos contra la diabetes: 0=no,

1=sí.

● CHF: Enfermedad cardíaca congestiva: 0=no, 1=sí.

● ARRHYTHMIA: Arritmias cardíacas: 0=no, 1=sí.

● VALVULAR: Enfermedad valvular: 0=no, 1=sí.

● PHTN: Problemas en la circulación pulmonar: 0=no, 1=sí.

● PVD: Enfermedad vascular periférica: 0=no, 1=sí.

● HTN: Hipertensión arterial no complicada: 0=no, 1=sí.

● HTNCX: Hipertensión arterial complicada: 0=no, 1=sí.

● PARALYSIS: Parálisis: 0=no, 1=sí.

● NEUROOTHER: Otros problemas neurológicos: 0=no, 1=sí.

● PULMONARY: Enfermedad pulmonar crónica: 0=no, 1=sí.

● DMCX: Diabetes complicada: 0=no, 1=sí.

● HYPOTHYROID: Hipotiroidismo: 0=no, 1=sí.

● RENAL: Enfermedad renal: 0=no, 1=sí.

● LIVER: Enfermedad hepática: 0=no, 1=sí.

● PUD: Úlcera gástrica excluyendo casos con sangrado: 0=no, 1=sí.

● HIV: HIV/SIDA: 0=no, 1=sí.

● LYMPHOMA: Linfoma: 0=no, 1=sí.

● METS: Cáncer metastásico: 0=no, 1=sí.

● TUMOR: Tumor sólido sin metástasis: 0=no, 1=sí.

● RHEUMATIC: Artritis reumatoidea / Enfermedad en el colágeno vascular:

0=no, 1=sí.

● COAG: Coagulopatía: 0=no, 1=sí.

● OBESITY: Obesidad: 0=no, 1=sí.

● WEIGHTLOSS: Bajo peso: 0=no, 1=sí.

● ANEMIA: Anemia por deficiencias: 0=no, 1=sí.

Tratamiento de la diabetes en adultos mayores - ¿Más intenso es mejor?

● FLUIDSLYTES: Desórdenes electrolíticos o de fluidos: 0=no, 1=sí.

● BLOODLOSS: Anemia por pérdida de sangre: 0=no, 1=sí.

● ALCOHOL: Alcoholismo: 0=no, 1=sí.

● DRUGS: Drogadicción: 0=no, 1=sí.

● PSYCHOSES: Psicosis: 0=no, 1=sí.

● DEPRESSION: Depresión clínica: 0=no, 1=sí.

● SEVERE\_DEP: Depresión severa: 0=no, 1=sí.

● CAD: Enfermedad arterial coronaria: 0=no, 1=sí.

● AMI: Infarto agudo de miocardio: 0=no, 1=sí.

● RETINOPATHY: Retinopatía: 0=no, 1=sí.

● HYPERG: Hiperglucemia: 0=no, 1=sí.

● AMPUTATION: Amputación de miembro inferior: 0=no, 1=sí.

● FEET: Infecciones por pié diabético: 0=no, 1=sí.

● SMOKER: Tabaquismo: 0=no, 1=sí.

● RETSCREEN: Bajo controles por posible retinopatía: 0=no, 1=sí.

● PCI: Cateterismo coronario: 0=no, 1=sí.

● ABI: Índice tobillo - brazo2: 0=no, 1=sí.

● ESLD: Etapa final de cirrosis o hígado graso: 0=no, 1=sí.

● CABG: Cirugía de by-pass coronario: 0=no, 1=sí.

● BP\_RX: Prescripción de medicamentos anti hipertensivos: 0=no, 1=sí.

El dataset se compone de 275190 filas y 70 columnas.

Las variables se distribuyen de la siguiente forma:

● Variables numéricas: 14

● Variables categóricas con valores múltiples: 3

● Variables categóricas binarias: 53

**4. Análisis exploratorio**

En el análisis exploratorio se estudiaron las distribuciones de las variables

numéricas y categóricas, así como la presencia de nulos, valores duplicados e

incongruentes.

Al analizar la distribución de las variables numéricas, se descubrió que presentaban

gran dispersión en los valores por lo que en principio parecía existir una gran

cantidad de outliers. Al verificar los valores se encontró que en realidad estos

valores son plausibles desde el punto de vista médico y tendrían significancia como

indicadores de comorbilidades, así que se los dejó incluidos para revisarlos en una

etapa posterior.

Se encontró que las variables MICROALB y SERUMALB tienen una elevada proporción

de valores faltantes. Esto se explica en el primer caso porque la albuminuria es un

hallazgo patológico indicador de daño renal, y en el segundo caso porque no es un

análisis de rutina, sino que el médico tiene que tener motivos para solicitarlo.

El dataset en general presentó un 1,69% de registros duplicados, que fueron

eliminados.

Respecto de la variable Sexo, el dataset está muy desbalanceado, con únicamente

un 1,14% de mujeres registradas.

Se encontró que las variables OBESITY y WEIGHTLOSS (obesidad y bajo peso,

respectivamente) tenían valores incoherentes, ya que en muchos casos no se

relacionaban con los valores de BMI registrados (que son los marcadores clínicos de

obesidad o bajo peso) e incluso en varios casos se presentaban ambos marcadores

en la misma fila. Debido a todo esto se decidió eliminar estas columnas del dataset.

**5. Conversión de variables**

Considerando la dispersión encontrada en las variables numéricas y que los valores

nulos en las variables MICROALB y SERUMALB no podían ser excluidos del análisis

porque tienen significancia médica, se realizó una conversión de variables que

permitió aprovechar para el análisis toda la información disponible.

● Las variables numéricas se agruparon en función de su significación médica

en valores enteros (desde 1 a 2, 3, 4 o 5 según la variable).

● Todas las variables binarias se convirtieron a valores distintos de 0 (1=Falso,

2=Verdadero).

● Todos los valores nulos se convirtieron a 0.

**6. Análisis post-conversión**

**6.1. Correlaciones entre variables**

Utilizando el paquete pandas\_profiling se analizaron las correlaciones entre las

variables convertidas y se encontró que los siguientes grupos presentaban alta

correlación entre sí:

● AGE/RATE\_5/RATE\_10: La mortalidad se correlaciona muy fuertemente con

el grupo etario, ya que los mayores de 89 quedan todos juntos en el mismo

grupo.

● SERUMCRE/RENAL: Los valores anormalmente altos de creatinina sérica son

uno de los síntomas relacionados con enfermedades renales.

6

Tratamiento de la diabetes en adultos mayores - ¿Más intenso es mejor?

● DEATH\_5/DEATH\_10: Dado que indican si el paciente falleció dentro de los 5

o 10 años del estudio, claramente se correlacionan.

● HTN/BP\_RX: La detección de hipertensión arterial se correlaciona con la

prescripción de medicamentos para tratarla.

● HTNCX/RENAL: Una de las complicaciones de la hipertensión arterial es el

daño renal.

● DMCX/RETINOPATHY: Una de las complicaciones de la diabetes es la

retinopatía.

● LIVER/ESLD: La variable ESLD indica la etapa final de cirrosis o hígado graso,

ambas enfermedades hepáticas.

● DEPRESSION/SEVERE\_DEP: Todos los casos de depresión severa se indican

también como depresión, por lo que los valores se correlacionan.

● HDL/TRI: **Éste es el único caso que a priori no resulta claramente explicable.**

Se tomó la decisión de que, si en pasos futuros se consideraba eliminar alguna de

estas variables de alta correlación, se elegiría retener las siguientes:

● HDL/TRI: Se eligió TRI porque tiene menos nulos.

● SERUMCRE/RENAL: Se eligió RENAL

● HTN/BP\_RX: Se eligió BP\_RX

● HTNCX/RENAL: Se eligió RENAL para no sumar correlación.

● DMCX/RETINOPATHY: Se eligió DCMX.

● LIVER/ESLD: Se emplea LIVER

● DEPRESSION/SEVERE\_DEP: Se eligió DEPRESSION

Como se puede observar, en los casos en que la correlación se debió a variables que

indican grados de afectación o una clase general y otra particular, se mantuvo la de

menor grado o mayor generalidad.

**6.2. Análisis de frecuencia de combinaciones**

Se realizó un análisis de la frecuencia de combinaciones de valores para algunos

pares de variables que podían tener relación médica:

● SYSTOLIC - BMI

● DIASTOLIC - BMI

● TRI - BMI

● HDL - BMI

● LDL - BMI

● FRAILTY - BMI

● MICROALB - N\_OP (Número de visitas a consultorios externos)

● FRAILTY - N\_OP

● BMI - N\_OP

7

Tratamiento de la diabetes en adultos mayores - ¿Más intenso es mejor?

El análisis demostró que hay combinaciones de valores que son más frecuentes que

otras, pero no mostró ninguna asociación clara.

**6.3. Valores cero**

El análisis de pandas\_profiling reporta que las siguientes variables presentan

gran número de valores cero:

● N\_IP (número de internaciones): Esto resulta razonable teniendo en cuenta

que la gran mayoría de los enfermos de diabetes, a pesar de que tengan

alguna complicación, no suelen requerir internación.

● FRAILTY: El valor cero indica que el paciente no tiene una salud frágil;

considerando que estos es razonable considerar que aproximadamente un

tercio de los pacientes tengan una salud robusta a pesar de padecer una

enfermedad crónica. En todo caso sería necesario tener a disposición la

metodología de cálculo del índice de fragilidad para poder efectuar un análisis

similar al realizado con la obesidad y el bajo peso.

**6.4. Detección de outliers**

Ya teniendo las variables convertidas, se realizó la evaluación de la cantidad de

outliers por separado para las correlaciones a 5 y 10 años.

En ambos casos se trabajó haciendo 10 rondas de entrenamiento y selección de

outliers con los métodos Isolation Forest, Minimum Covariance

Determinant, Local Outlier Factor y One-Class SVM. En cada ronda se

registró el error medio absoluto de la regresión con los datos crudos y los datos sin

los outliers detectados por cada método. Al final se revisó si algún método era

consistentemente mejor para eliminar los outliers.

Los resultados fueron interesantes en el sentido de que el método más eficaz para

detectar outliers en la correlación para predicción a 5 años (OneClassSVM) no fue el

mismo que para la predicción a 10 años (IsolationForest). Además, aunque la

cantidad de outliers detectados fue prácticamente igual (2678 vs 2679) la reducción

en el MAE obtenida terminó siendo muy diferente para ambas correlaciones (>3.5%

para la predicción a 5 años contra <0,03% para la predicción a 10 años).

De todas maneras se eliminaron los outliers respectivos de ambas correlaciones.

**7. Cálculo de las tasas de mortalidad**

Una vez eliminados los outliers, se calculó la tasa de mortalidad para cada grupo etario en general y únicamente para los pacientes sin complicaciones:

Claramente se puede apreciar que la tasa de mortalidad para los pacientes sin

complicaciones es menor que para la población en general.