

ANALÍTICA INTERMEDIA Y AVANZADA: INTRODUCCIÓN A LA INTELIGENCIA ARTIFICIAL

PROYECTO FINAL

Predicción de Enfermedad Renal Crónica (ERC)

Elaborado por: Área de Analítica de T.I.

1. PLANTEAMIENTO DEL PROBLEMA

Se desea conocer qué pacientes tendrán Enfermedad Renal Crónica (ERC), en el próximo año. Para ello, se hace necesario calibrar un modelo de clasificación, que ayude a encontrar qué variables resultan ser factores de riesgo (*aumenta la probabilidad de tener la enfermedad*), y cuáles son factores protectores (*disminuyen la probabilidad de ocurrencia del evento*). De esta forma, respecto a este desenlace se estará haciendo una correcta gestión del riesgo en salud, al poder anticipar el suceso, con un intervalo de un año, para diseñar modelos de atención correctivos y diseñar mecanismos de impacto por persona.

2. DESCRIPCIÓN DE LOS DATOS

Después de una revisión bibliográfica estructural y un consenso entre especialistas de la enfermedad en consideración, se plantearon 39 variables (Antecedentes Familiares, Antecedentes Personales, Control, Hábitos, Resultados Clínicos y Sociodemográficas), como posibles predictores del desenlace (ver tabla 2, al final del documento).

En torno al diseño del experimento, se tomaron todas las medidas necesarias para generar criterios de inclusión y exclusión de las poblaciones en consideración. Así mismo, la forma de definición de las poblaciones y la extracción de las variables independientes, están acordes al problema, bajo un enfoque de datos transversales. Por tales razones, no se ofrece mayor detalle de estos aspectos en esta propuesta, además que no son necesarios para culminar el ejercicio.

3. ENUNCIADOS PROPUESTOS

Los siguientes numerales, condensan los lineamientos mínimos propuestos para solucionar el problema planteado previamente. Adicionalmente, se espera que el estudiante sea autónomo en decidir mejorar el ejercicio, articulando pasos adicionales, o profundizando en los actuales, en el desarrollo metodológico y analítico.

- a. Realizar análisis descriptivos condicionales (respecto a la variable de clasificación o variable dependiente) e incondicionales, con el objetivo de caracterizar las variables, observar su comportamiento frente al ejercicio de clasificación y empezar a plantear posibles hipótesis de causalidad.

| Paso | Objetivo | Tipo de variables | |
|---|---|-------------------|-----------|
| | | Categóricas | Numéricas |
| Análisis descriptivo incondicional | Compleitud de registros | X | X |
| | Análisis de frecuencias | X | |
| | Momentos centrales alrededor de la media | X | |
| | Gráficos descriptivos | X | X |
| Análisis descriptivo condicional | Conteo de registros | X | X |
| | Análisis de frecuencias | X | |
| | Diferencia de proporciones | X | |
| | Momentos centrales condicionales | | X |
| | Pruebas de diferencias entre momentos centrales | | X |
| | Gráficos descriptivos | X | X |

Tabla 1. Detalle de los objetivos del análisis descriptivo condicional e incondicional. *Elaboración propia.*

- b. Seleccionar los modelos de clasificación que se van a considerar
- c. Depurar las variables y hacer las transformaciones pertinentes para aplicar los algoritmos considerados
- d. Realizar el diseño de un proceso de validación cruzada, para poder hacer inferencia sobre el error de generalización del modelo
- e. Estimar cada uno de los modelos
- f. Analizar curvas de aprendizaje y curvas de validación de cada modelo
- g. Analizar el comportamiento de los rendimientos en las muestras de validación, para seleccionar el mejor modelo.
- h. Poner a prueba el mejor modelo en la muestra de pronóstico y evaluar el comportamiento de la matriz de confusión y curva ROC (si es del tipo de modelos probabilísticos)

4. CONDICIONES DE PRESENTACIÓN

Se requiere que el desarrollo del proceso se realice en Python o R. Es indispensable que los códigos utilizados estén suficientemente comentados y claros, como para replicar fácilmente el ejercicio. Adicionalmente a éstos, se requiere presentar un informe con los principales resultados del ejercicio, con normas ICONTEC 2019 o APA 2019. Si dentro de la argumentación de los resultados, requiere hacer uso de bibliografía, ésta debe quedar correctamente citada.

| Nombre Variable | Grupo Variable | Tipo Variable | Resultados | Descripción |
|--|-------------------------|-------------------|---|---|
| Historia Familiar de Diabetes | Antecedentes Familiares | Categórica | - No (0) - Sí (1) | Variable que mide la presencia actual o pasada de Diabetes en un integrante familiar |
| Historia Familiar de Enfermedad Cardiovascular | Antecedentes Familiares | Categórica | - No (0) - Sí (1) | Variable que mide la presencia actual o pasada de alguna Enfermedad Cardiovascular en un integrante familiar |
| Historia Familiar de ERC | Antecedentes Familiares | Categórica | - No (0) - Sí (1) | Variable que mide la presencia actual o pasada de Enfermedad Renal Crónica en un integrante familiar |
| Anemia | Antecedentes Personales | Categórica | - No (0) - Sí (1) | Variable categórica para determinar la presencia del diagnóstico de Anemia, como antecedente personal |
| Artritis Reumatoide | Antecedentes Personales | Categórica | - No (0) - Sí (1) | Variable categórica para determinar la presencia del diagnóstico de Artritis Reumatoide, como antecedente personal |
| Diabetes Mellitus | Antecedentes Personales | Categórica | - No (0) - Tipo I (1) - Tipo II (2) | Variable categórica para determinar la presencia del diagnóstico de Diabetes Mellitus como antecedente personal |
| Dislipidemia | Antecedentes Personales | Categórica | - No (0) - Sí (1) | Variable categórica para determinar la presencia del diagnóstico de Dislipidemia como antecedente personal |
| Duración de la Diabetes | Antecedentes Personales | Numérica Discreta | Número entero positivo | Tiempo transcurrido en días desde el diagnóstico de Diabetes, hasta t-1 |
| Duración de la Hipertensión | Antecedentes Personales | Numérica Discreta | Número entero positivo | Tiempo transcurrido en días desde el diagnóstico de Hipertensión, hasta t-1 |
| Enfermedad Cardiovascular | Antecedentes Personales | Categórica | - No (0) - Sí (1) | Variable categórica para determinar la presencia del diagnóstico de Enfermedad Cardiovascular como antecedente personal |

| Nombre Variable | Grupo Variable | Tipo Variable | Resultados | Descripción |
|--------------------------------|-------------------------|---------------|----------------------|--|
| Enfermedad Cerebrovascular | Antecedentes Personales | Categórica | - No (0) - Sí (1) | Variable categórica para determinar la presencia del diagnóstico de Enfermedad Cerebrovascular como antecedente personal |
| Enfermedad Vascular Periférica | Antecedentes Personales | Categórica | - No (0) - Sí (1) | Variable categórica para determinar la presencia del diagnóstico de Enfermedad Vascular Periférica como antecedente personal |
| Gestante | Antecedentes Personales | Categórica | - No (0) - Sí (1) | Variable categórica para determinar la presencia del diagnóstico de Gestante como antecedente personal |
| Hiperplasia Prostática | Antecedentes Personales | Categórica | - No (0) - Sí (1) | Variable categórica para determinar la presencia del diagnóstico de Hiperplasia Prostática como antecedente personal |
| Hipertensión Arterial | Antecedentes Personales | Categórica | - No (0) - Sí (1) | Variable categórica para determinar la presencia del diagnóstico de Hipertensión Arterial como antecedente personal |
| Hiperuricemia | Antecedentes Personales | Categórica | - No (0) - Sí (1) | Variable categórica para determinar la presencia del diagnóstico de Hiperuricemia como antecedente personal |
| Litiasis Renal | Antecedentes Personales | Categórica | - No (0) - Sí (1) | Variable categórica para determinar la presencia del diagnóstico de Litiasis Renal como antecedente personal |
| Lupus | Antecedentes Personales | Categórica | - No (0) - Sí (1) | Variable categórica para determinar la presencia del diagnóstico de Lupus como antecedente personal |
| Retinopatía Diabética | Antecedentes Personales | Categórica | - No (0) - Sí (1) | Variable categórica para determinar la presencia del diagnóstico de Retinopatía Diabética como antecedente personal |

| Nombre Variable | Grupo Variable | Tipo Variable | Resultados | Descripción |
|----------------------------------|---------------------|-------------------|---|--|
| Tratamiento para la Hipertensión | Control | Categórica | - No (0) - Sí (1) | Variable categórica para determinar la exposición continua de mínimo tres meses a medicamentos para el tratamiento de la hipertensión |
| Uso de AINES | Control | Categórica | - No (0) - Sí (1) | Variable categórica para determinar la exposición continua de mínimo tres meses a medicamentos AINES |
| Cigarrillo | Hábitos | Categórica | - Nunca (0) - A veces (1) - Siempre (2) | Variable categórica para determinar la frecuencia de reporte del hábito de cigarrillo |
| Ejercicio | Hábitos | Categórica | - Nunca (0) - A veces (1) - Siempre (2) | Variable categórica para determinar la frecuencia de reporte del hábito de ejercicio |
| Ácido Úrico | Resultados Clínicos | Numérica Continua | Número real positivo | Resultado del examen clínico de Ácido Úrico , medido en MG/DL |
| Colesterol HDL | Resultados Clínicos | Numérica Continua | Número real positivo | Resultado del examen clínico de Colesterol HDL, medido en MG/DL |
| Glicemia Post-Prandial | Resultados Clínicos | Numérica Continua | Número real positivo | Resultado del examen clínico de Glicemia Post-Prandial, medido en MG/DL |
| Hemoglobina | Resultados Clínicos | Numérica Continua | Número real positivo | Resultado del examen clínico de Hemoglobina, medido en G/DL |
| Hemoglobina Glicosilada | Resultados Clínicos | Numérica Continua | Número real positivo | Resultado del examen clínico de Hemoglobina Glicosilada, medido en % |
| Índice de Masa Corporal | Resultados Clínicos | Numérica Continua | Número real positivo | Resultado del Índice de Masa Corporal, medido en KG/M2 |
| Perímetro Abdominal | Resultados Clínicos | Categórica | - Sin Dato (0) - Normal (1) - Anormal (2) | Variable categórica para determinar la presencia de un resultado normal (sin riesgo), o anormal (con riesgo), entorno al examen clínico de Perímetro Abdominal |
| Presión Arterial Diastólica | Resultados Clínicos | Numérica Continua | Número real positivo | Resultado de la Presión Arterial Diastólica |

| Nombre Variable | Grupo Variable | Tipo Variable | Resultados | Descripción |
|-------------------------------|---------------------|-------------------|--|--|
| Presión Arterial Sistólica | Resultados Clínicos | Numérica Continua | Número real positivo | Resultado de la Presión Arterial Sistólica |
| Proteína C Reactiva | Resultados Clínicos | Numérica Continua | Número real positivo | Resultado del examen clínico de Proteína C Reactiva, medido en MG/DL |
| Proteinuria | Resultados Clínicos | Categórica | - Sin Dato (0) - Normal (1) - Anormal (2) | Variable categórica para determinar la presencia de un resultado normal (sin riesgo), o anormal (con riesgo), entorno al examen clínico de Proteinuria |
| Tasa de Filtración Glomerular | Resultados Clínicos | Numérica Continua | Número real positivo | Resultado de la Tasa de Filtración Glomerular, calculada a partir de la metodología: MDRD |
| Trigliceridos | Resultados Clínicos | Numérica Continua | Número real positivo | Resultado del examen clínico de Trigliceridos, medido en MG/DL |
| Edad | Sociodemográficas | Numérica Continua | Número entero positivo | Edad del paciente |
| Raza | Sociodemográficas | Categórica | - Sin Informacion Desde La Fuente (0) - Afroamericano (1) - Blanco (2) - Indígena (3) - Mestizo (4) - Mulato (5) - Zambo (6) | Raza del paciente |
| Sexo | Sociodemográficas | Categórica | - M (0) - F (1) | Sexo del paciente |

Tabla 2. Documentación de las variables explicativas para el desenlace de Enfermedad Renal Crónica (ERC). *Elaboración propia.*