



Instituto Tecnológico y de Estudios Superiores de Monterrey
Campus Monterrey
Ing. en Ciencia de Datos y Matemáticas

Modelación del aprendizaje con inteligencia artificial
Aprendizaje supervisado

Profesor

Jose Eduardo Ferrer Cruz

Estudiantes

Leonardo Laureles Olmedo A01659241

Mariana Rincón Flores A01654973

Carlos Mateos Pérez A01654085

Daniel Núñez López A01654137

Situación Problema: Identificar enfermedades y diagnóstico

Explicación del problema

El diagnóstico precoz engloba una serie de medidas para determinar la presencia de una enfermedad potencialmente grave. Las ventajas son muchas, disminuyen la tasa de mortalidad de estas enfermedades, la probabilidad de éxito de los tratamientos aumenta, las complicaciones se reducen, así como las secuelas de la enfermedad y del mismo tratamiento, de igual manera se podrían llegar a reducir los costos en caso de que la enfermedad se llegue a agravar.

Las pruebas de detección son exámenes que los doctores usan para detectar enfermedades y problemas de salud antes de que existan signos o causen síntomas. Estas pruebas sirven para detectar los problemas a tiempo, cuando pueden ser más fáciles de tratar. La prueba de detección dependerá del problema que se quiera detectar.

La actual pandemia SARS-CoV-2 plantea retos sanitarios, sociales y económicos de enorme magnitud. El diagnóstico correcto y rápido de la infección por SARS-CoV-2 es crítico, tanto desde el punto de vista epidemiológico, dado que muchas personas infectadas están asintomáticas, como desde el punto de vista clínico, para identificar y tratar a los pacientes cuanto antes.

Las técnicas de aprendizaje supervisado pueden ser de ayuda para predecir qué tan probable es que un paciente haya sido infectado con COVID-19 con base en sus comorbilidades habiendo estado en contacto con algún contacto positivo y con una gran cantidad de datos.

El objetivo del agente será recabar información sobre pacientes que fueron contagiados previamente, junto con la lista de comorbilidades que presente, para entrenarse y así utilizar un modelo de clasificación para dar la probabilidad de estar contagiado de COVID-19.

Descripción del Ambiente

La información se obtendrá de los sensores que se utilizan para detectar las enfermedades, para detectar la neumonía, se utiliza una radiografía de tórax, para la diabetes son los análisis de sangre que muestran si el nivel de glucosa en la sangre es demasiado alto. Para detectar el asma se realizan espirometrías, la cuál es una prueba que mide cuánto aire se inhala y se exhala de los pulmones, los sensores que se utilizan para detectar la hipertensión

son el esfigmomanómetro, que es el instrumento que se coloca en el brazo y se infla, para detectar las enfermedades cardiovasculares se utilizan electrocardiogramas como sensores, para detectar la obesidad se utiliza la báscula para medir altura y peso, de ahí se procede a realizar el cálculo del índice de masa corporal, para la insuficiencia renal crónica se realizan exámenes de sangre y orina, para detectar el tabaquismo se realiza un análisis de sangre y pruebas de nicotina.

El ambiente es completamente accesible ya que los sensores detectan todos los aspectos relevantes a la elección de una acción, ya que si los sensores de un agente le permiten tener accesos al estado total del ambiente se dice que es accesible. También es considerado semi-dinámico ya que el ambiente no cambia con el paso del tiempo pero sí va modificando la calificación asignada al desempeño del agente, el ambiente es discreto porque existe una cantidad limitada de percepciones y acciones y finalmente es episódico porque la respuesta que se da a una pregunta no afectará a la siguiente, es decir cada vez que se le pregunte algo, la estrategia se va a replantear.

Observaciones a Recolectar

Edad, sexo, si ha padecido alguna enfermedad antes como neumonia, diabetes, asma, hipertension, enfermedades cardiovasculares, obesidad, insuficiencia renal cronica, tabaquismo. Todas estas acciones serán tomadas por los sensores ya antes mencionados, y con una muestra para individuo es más que suficiente.

Se deben recolectar tantas observaciones como preguntas que haya en el cuestionario, pero al momento que se realice cada pregunta el agente deberá recolectar una sola observación a la vez.

Toma de Decisiones

El agente decidirá al final con base en los resultados de las observaciones previas, hasta no tener toda la información no podrá dar una predicción.

No tendrá rangos de tiempo, será con base en los valores de cada observación de las variables predictoras, es decir si la persona tiene más de 50 años, neumonía, enfermedades cardiovasculares, enfermedades renales, diabetes, asma, obesidad e hipertensión es más probable que sea infectado por COVID-19 que un adolescente sin ninguna de estas comorbilidades

Cuando se cumpla alguna condición, tendrá rangos y será con base en la cantidad de comorbilidades que tenga el paciente respecto a su edad y sexo.

Acciones Disponibles

Una vez que se tome la decisión se dirá si la persona tiene más probabilidades que sea infectado al haber estado en contacto con algún contagiado, esto quiere decir que la acción será discreta, ya que será un sí o no únicamente.

Igualmente, después de dar la predicción, se le preguntará a la persona si tuvo o no COVID-19 para mejorar el modelo y tener mejores predicciones. Las acciones del agente serán discretas ya que no cambian, al ser una cantidad limitada y estar definidas, no hay un cambio en el proceso del agente.

Recompensas

El agente tendrá una puntuación general que se determinará con base en sus predicciones, recibirá puntos positivos por cada predicción correcta y se le restará el doble de puntos por cada predicción incorrecta, por ejemplo, si el agente responde de manera correcta y se confirma su predicción, se le darán 5 puntos, pero si el agente no responde de manera correcta se le restará 10 puntos. Si el agente tiene una puntuación negativa significa que debe esforzarse más por tener una puntuación positiva, y si tiene una puntuación positiva, deberá de seguir realizando predicciones que le aporten recompensas. Por último, entre mayor sea la puntuación general del agente, significa que es más confiable.

Representación Gráfica del Modelo

El árbol de decisión lo realizamos con una muestra de cien registros de un dataset de los datos clínicos del COVID-19 en México con ayuda de python.

Primero cargamos el dataset en formato csv, luego elegimos las variables que mencionamos anteriormente, ya que son las comorbilidades que se mencionan en el dataset, posteriormente, quitamos las filas con valores nulos y para fines prácticos, solamente utilizamos las cien primeras filas y no todas las filas que contenía el dataset, ya que el árbol quedaría de mayor dimensión y se complicaría su interpretación y visualización para este caso. Por último, cambiamos los valores de 1 y 2 en el dataset por ceros y unos para evitar confusión de si los datos tenían un orden jerárquico o no. Una vez realizada la limpieza de datos, dividimos nuestro dataset en X como las variables predictoras y Y como nuestra variable target, que en este caso es el resultado de la prueba PCR. de covid.

Para realizar el arbol de decision utilizamos un algoritmo de clasificación llamado Decision Tree Classifier encontrado en la librería de scikit-learn, luego simplemente graficamos el árbol y el resultado es el siguiente:

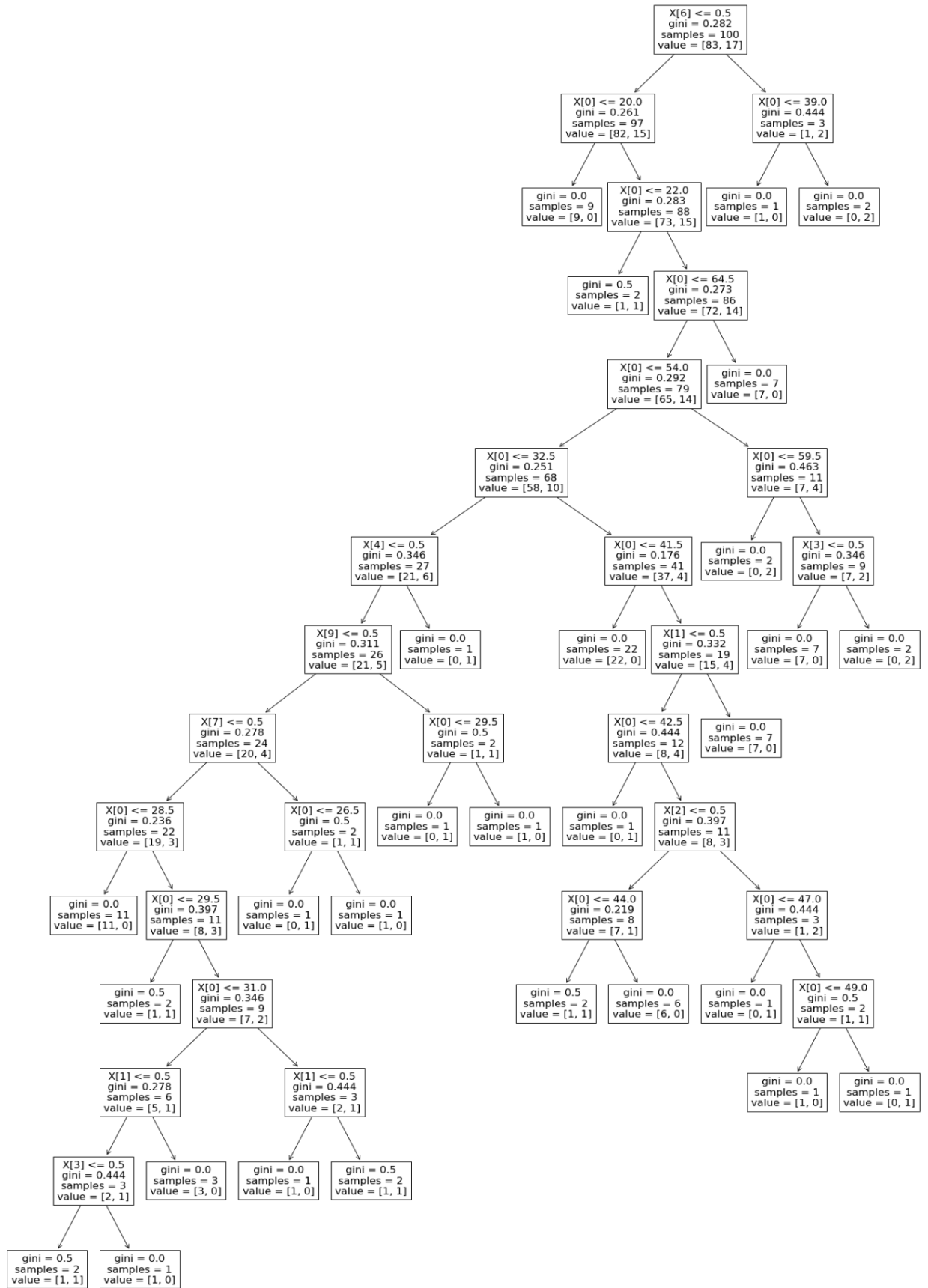


Figura 1. Árbol de decisión.

Para interpretar el árbol, si observamos el primer recuadro, podemos leer que dice $X[6] \leq 0.5$, esto significa que se ocupa la variable predictora número 6, o enfermedad cardiovascular, y que si esta es cero, es decir, menor a 0.5, entonces se pasa a la siguiente pregunta o condición hasta así llegar a una conclusión.

0. EDAD
1. SEXO
2. NEUMONÍA
3. DIABETES
4. ASMA
5. HIPERTENSIÓN
6. CARDIOVASCULAR
7. OBESIDAD
8. RENAL CRÓNICA
9. TABAQUISMO
10. RESULTADO

Referencias

- Kaggle. (2020, 5 junio). *México COVID-19 clinical data*. Recuperado 14 de junio de 2022, de <https://www.kaggle.com/datasets/marianarfranklin/mexico-covid19-clinical-data/meta-data>
- NIH. (2021, 20 junio). *Características clínicas y comorbilidades asociadas a mortalidad en pacientes con COVID-19 en México*. National Library of Medicine. Recuperado 14 de junio de 2022, de <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7816557/>