

Introducción a la inteligencia artificial

Nombre: Roberto Alvarado

Fecha: 04 de Mayo del 2025

1. Caso de estudio

La problemática que abordaré en este documento es la que seleccioné para mi proyecto de tesis. Me centraré en la implementación de un producto mínimo viable para predecir, según datos, la probabilidad de que un paciente pueda llegar a sufrir ERC. Citaré el documento de la propuesta para entender de que se trata la enfermedad.

”La insuficiencia renal, también conocida como enfermedad renal crónica (ERC), es una condición médica en la que los riñones pierden progresivamente su capacidad para filtrar desechos y exceso de líquidos de la sangre, lo que puede llevar a la acumulación de sustancias nocivas en el cuerpo y afectar otras funciones orgánicas. Entre las principales causas se encuentran la diabetes, la hipertensión, enfermedades auto inmunes, obstrucciones prolongadas en el tracto urinario debido a infecciones, enfermedades hereditarias, uso prolongado de medicamentos y riesgo cardiovascular” (Jiménez, 2025)

El problema es muy directo para ser resuelto por varios métodos de inteligencia artificial, porque tenemos un conjunto de datos, y queremos hacer un modelo de predicción, sin embargo, hay cuatro puntos a considerar que serán los que se presentarán como problemáticas de la investigación

- **Selección de atributos:**

Dentro de la base de datos que tenemos, existen varios atributos o características para cada entrada de información que tenemos. Muchos atributos consideramos que no son importantes. A que me refiero, tenemos por ejemplo varios atributos como lugar de nacimiento y lugar de residencia, estos aunque pueden ser interesantes de analizar, dentro de varios otros estudios del mismo tipo, no se consideran al hacer el análisis. Entonces uno de los primeros problemas que tenemos es filtrar que atributos vamos a seleccionar para entrenar el modelo

- **Búsqueda de base de datos pública**

Aunque la base de datos para la investigación es extensa, sabemos que la cantidad de datos no es la suficiente, por lo que necesitamos hacer una búsqueda de bases de datos públicas que tenemos

- **Selección del modelo**

Aunque sea un problema ”simple”, la selección del modelo de inteligencia artificial necesitará un análisis exhaustivo

- **Implementación de la herramienta**

Después de tener el modelo, la finalidad del proyecto es realizar una herramienta útil para ser utilizada por médicos, por lo que la planificación del proceso de implementación tendrá que considerar las herramientas actuales

2. Datos

En esta sección presentaré las opciones que tenemos para solucionar los 2 primeros puntos, ya que se relacionan mucho entre si. Primeramente, la selección de atributos, para esto tenemos dos opciones, uno un análisis de la literatura que existe ahora para los atributos a trabajar, para esto se han encontrado investigaciones que funcionaran como la base de la investigación. Hasta el momento, estos son algunos de los artículos que se leerán

- **Kidney Failure Prediction Models: A Comprehensive External Validation Study in Patients with Advanced CKD**

Link: <https://pmc.ncbi.nlm.nih.gov/articles/PMC8259669/>

- **Mortality Risk Prediction Models for People With Kidney Failure**

Link: <https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2828649>

Sin embargo, hay otra opción, que se podría considerar que se conoce como teoría de la información y nos podríamos centrar en testores para hacer una selección de los atributos a utilizar, sin embargo por temor al tiempo no es la solución adecuada (Alvarado-Moreira y Ibarra Fiallo, 2024)
Ahora para la búsqueda de bases de datos, tenemos algunas opciones

- United States Renal Data System
Link: <https://www.niddk.nih.gov/about-niddk/strategic-plans-reports/usrds>
- Chronic KIdney Disease dataset
Link: <https://www.kaggle.com/datasets/mansoordaku/ckdisease>

3. Selección del modelo

Para este caso todo se basará en experimentación, pero para generalizar, tenemos una problemática que requiere un aprendizaje supervisado para lograr una regresión de los datos, para eso tenemos múltiples opciones, pero consideraré tres

3.1. SVM

Las máquinas de vectores de soporte (SVM), son una idea interesante, donde su funcionalidad se puede simplificar en un ejemplo, tengo un número de puntos dentro de un campo, estos pueden pertenecer a dos categorías, un SVM, nos puede ayudar a encontrar la probabilidad que un punto se encuentre en una categoría o otra. Para nuestra investigación, tenemos dos categorías, sufrió la enfermedad renal crónica o no, entonces nuestro modelo nos puede ayudar a predecir si una persona se encuentra en la categoría de sufrió ERC. (Steinwart y Christmann, 2008)

Para ser más exacto, lo que vamos a necesitar es una regresión de vectores de soporte, estas son aquellas que nos facilitarán encontrar un valor para la probabilidad que un objeto de estudio pueda o no sufrir la enfermedad

3.2. Redes Neuronales

Las redes neuronales son un algoritmo matemático que se encarga de extender y de usar la idea general de una neurona en inteligencia artificial, estos buscan simular hacer lo que una neurona medicamente hace, una neurona esta definida en IA como una estructura. Las redes neuronales agrupan y codifican un grupo de ellas para lograr un resultado.

En general, una neurona es una estructura que generaliza la idea de tener un input, procesarlo, comparar con lo que se espera, y finalmente hacer un tuning de nuestro proceso para acercarnos más al resultado esperado. El conjunto de estas permite que este proceso sea un proceso de regresión y sea uno de los algoritmos más utilizados en la actualidad

Los modelos de redes neuronales son una generalización de lo que espero utilizar, siendo específico espero primero probar con una red neuronal recurrente o RNN (Schmidt, 2019) estoy seguro que es un buen primer paso, pero en la actualidad con tantas nuevas implementaciones tendré que hacer una investigación para encontrar la adecuada.

3.3. Random Forest

La idea base de la regresión con random forest (o bosque aleatorio), es crear árboles de decisión cada uno entrenado con un subset de la base de datos, cada uno de ellos tendrá su propia configuración, por lo que se hace un número muy grande de árboles de decisión, y se hace un promedio (generalmente) de los resultados, de esta forma se puede implementar una regresión utilizando random forest. (Breiman, 2001)

4. Implementación de la herramienta

Ahora, cuando el modelo este funcionando con los datos y resultados esperados, tenemos que implementar la herramientas, para esto tenemos dos opciones, por motivo de tiempo la herramienta se implementará como una herramienta web. Entre las opciones en como se implementará puede ser, Angular o Vue, que son aquellas que tengo conocimiento. Esta problemática aún no la tengo muy decidida.

5. Conclusiones

El caso de estudio, aunque parece que es un simple problema de regresión trae consigo un gran número de problemáticas que tienen que sobrellevarse para que esta investigación sea fortuita. El problema de la obtención de los datos y la selección de los atributos para entrar los modelos es algo nuevo que me tengo que enfrentar, hasta el momento la literatura sobre la problemática es basta por lo que estoy seguro que se pueden sacar conclusiones desde ahí. Si el tiempo lo permite, intentaré probar con más herramientas que permitan reducir la dimensionalidad del problema sin la necesidad de simplemente citar la literatura, pero será una problemática a presentarse en un futuro.

La inteligencia artificial será uso como ya se dijo en el análisis de datos, para esto este documento me ayudo a identificar el tipo de algoritmo que necesito, un modelo de regresión que se base en aprendizaje supervisado. Las opciones que cite aquí, son las primeras que probaré, sin embargo, estaré al tanto de nuevos posibles algoritmos que puedan enriquecer los resultados de los datos.

Algo que quería mencionar es que este trabajo me ayudo también es a saber que es lo que no debería, centrarme, por ejemplo, una red neuronal convolucional no tiene sentido de utilizar en atributos tan poco relacionados, o el uso de transformers que aunque sean interesantes, no tiene sentido hacer una implementación de los mismos en un problema de regresión con simplemente dos categorías.

Referencias

- Alvarado-Moreira, R., y Ibarra Fiallo, J. (2024, 07). New approach to facial expression recognition and classification using typical testors. En (p. 406-414).
- Breiman, L. (2001, 01 de Oct). Random forests. *Machine Learning*, 45(1), 5-32. Descargado de <https://doi.org/10.1023/A:1010933404324> doi: 10.1023/A:1010933404324
- Jiménez, Y. (2025). *Propuesta de trabajo de titulación*. UTPL.
- Schmidt, R. M. (2019). Recurrent neural networks (rnns): A gentle introduction and overview. *CoRR*, abs/1912.05911. Descargado de <http://arxiv.org/abs/1912.05911>
- Steinwart, I., y Christmann, A. (2008). *Support vector machines*. Springer New York. Descargado de <https://books.google.cl/books?id=HUnqnrpYt4IC>