

PREDICCIÓN DE SUPERVIVENCIA ANTE LA INSUFICIENCIA CARDIACA

Prediction of Survival in Heart Failure

Choque Forra Daniela Patricia

dpcfjrc@gmail.com

**Instituto de Investigaciones en Ciencia y Tecnología,
Universidad La Salle-Bolivia**

Mamani Mamani Jessyca Liset

jessycamamani20@gmail.com

**Instituto de Investigaciones en Ciencia y Tecnología,
Universidad La Salle-Bolivia**

Resumen

Justificación: En visto que actualmente las enfermedades son ya comunes entre nosotros, una que tiene un amplio conjunto es la Insuficiencia Cardiaca, que puede surgir por varias razones y estas aplicar en que tan mortal puede llegar a ser. El objetivo es predecir la supervivencia ante la insuficiencia cardiaca, contando con un dataset que contiene registros clínicos de pacientes con insuficiencia cardiaca. Se utilizó diferentes algoritmos de Aprendizaje Automático de los cuales se seleccionó el mejor modelo para luego utilizarlo en una interfaz gráfica. También se realizó el análisis de datos para observar la influencia que tienen respecto a la Insuficiencia Cardiaca. La metodología empleada fue cuantitativa de diseño exploratorio. Se logró presentar un análisis de cada factor en la insuficiencia cardiaca, tomando puntos como el

género, o algún otro tipo de enfermedad.

Palabras clave: Corazón, insuficiencia, predicción.

Abstract

Justification: Given that currently diseases are already common among us, one that has a wide range is Heart Failure, which can arise for various reasons and these apply to how deadly it can be. The objective is to predict survival in heart failure, counting on a dataset that contains clinical records of patients with heart failure. Different Machine Learning algorithms were used from which the best model was selected and then used in a graphical interface. Data analysis was also carried out to observe the influence they have regarding Heart Failure. The methodology used was quantitative with an exploratory design. An analysis of each factor in heart failure was presented, taking points such as gender, or some other type of disease.

Key words: Heart, failure, prediction.

Introducción

Muchas veces, no es posible saber cuándo puede una persona tener insuficiencia cardíaca, y a medida que avanza el tiempo, esto se hace cada vez más difícil de saber o confundir ciertos síntomas que se pueden producir por este problema.

El proyecto trata sobre la supervivencia ante la insuficiencia cardíaca y la aplicación de modelos de aprendizaje automático para predecir la supervivencia ante la Insuficiencia Cardíaca. Ante ello conoceremos que es la Insuficiencia Cardíaca y que es Aprendizaje Automático.

Insuficiencia Cardíaca

La insuficiencia cardíaca (IC) es un síndrome de disfunción ventricular. La

insuficiencia ventricular izquierda causa disnea y fatiga, mientras que la insuficiencia ventricular derecha promueve la acumulación de líquido en los tejidos periféricos y el abdomen. Los ventrículos pueden verse involucrados en forma conjunta o por separado. El diagnóstico inicial se basa en la evaluación clínica y se confirma con radiografías de tórax, ecocardiografía y medición de las concentraciones plasmáticas de péptido natri uréticos. El tratamiento consiste en la educación del paciente, diuréticos, inhibidores de la enzima convertidor de la angiotensina (ECA), bloqueantes de los receptores de angiotensina II, betabloqueantes, antagonistas de la aldosterona, inhibidores de la neprilisina, marcapasos/desfibriladores implantables y otros dispositivos especializados y la corrección de la causa o las causas del síndrome de IC. (Howlett, 2020).

Causas

La insuficiencia cardíaca casi siempre es una afección prolongada (crónica), pero se puede presentar repentinamente. Puede ser causada por muchos problemas diferentes del corazón.

La enfermedad puede afectar únicamente el lado derecho o el lado izquierdo del corazón. Más frecuentemente, ambos lados del corazón resultan comprometidos. (Allen LA, 2019).

La insuficiencia cardíaca ocurre cuando:

- Su miocardio no puede bombear (expulsar) la sangre del corazón muy bien. Esto se denomina insuficiencia cardíaca sistólica o insuficiencia cardíaca con una fracción de eyección reducida (HFrEF, por sus siglas en inglés). (Allen LA, 2019).
- El miocardio está rígido y no se llena de sangre fácilmente. Esto se denomina insuficiencia cardíaca diastólica o insuficiencia cardíaca con una eyección preservada (HFpEF, por sus siglas en inglés). (Allen LA, 2019).

80 Referentes Conceptuales

Aprendizaje Automático

El proceso de aprendizaje automático es similar al de la minería de datos. Ambos sistemas buscan entre los datos para encontrar patrones. Sin embargo, en lugar de extraer los datos para la comprensión humana –como es el caso de las aplicaciones de minería de datos– el aprendizaje automático utiliza esos datos para detectar patrones en los datos y ajustar las acciones del programa en consecuencia. Los algoritmos del aprendizaje automático se clasifican a menudo como supervisados o no supervisados. Los algoritmos supervisados pueden aplicar lo que se ha aprendido en el pasado a nuevos datos. Los algoritmos no supervisados pueden extraer inferencias de conjuntos de datos. (Rouse, 2017).

Conociendo ambos conceptos se puede empezar con el objetivo del proyecto.

Descripción de datos

El presente proyecto para la “Predicción de Supervivencia ante la Insuficiencia Cardíaca” utilizó un dataset que contiene registros clínicos de pacientes con insuficiencia cardíaca. Este dataset cuenta con los siguientes datos según Kaggle-Serpro:

- **Age:** La edad del paciente.
- **Anaemia:** Si el paciente tiene anemia.
- **Creatinine_phosphokinase:** El número de Creatinina fosfoquinasa que tiene el paciente.
- **Diabetes:** Si el paciente tiene Diabetes.
- **Ejection_fraction:** La fracción de eyección del corazón del paciente.
- **High_blood_pressure:** Si el paciente tiene hipertensión.
- **Platelets:** El número de plaquetas del paciente.
- **Serum_creatinine:** El nivel de Suero de creatinina en el paciente.
- **Serum_sodium:** El nivel de Sodio sérico en el paciente.

- **Sex:** El género del paciente.
- **Smoking:** Si el paciente fuma.
- **Time:** Días de tratamiento del paciente.
- **DEATH_EVENT:** Si el paciente sobrevive o no.

Análisis de datos

Se importó el Dataset utilizado en este proyecto el cual es “Registros clínicos de insuficiencia cardíaca. Nuestro dataset cuenta con 299 registros clínicos con los datos ya mencionados.

División aleatoria del Dataset al 90%

Se realizó una separación de datos donde se tomará aleatoriamente el 90% de los datos para el análisis y entrenamiento, el 10% será utilizado para prueba.

Comienzo del Análisis de datos

Se inició a analizar el dataset tomando algunos factores para observar su influencia en la predicción de la supervivencia ante la Insuficiencia cardiaca. Se acompañará con gráficos.

La edad y el género un factor importante para predecir la muerte ante IC

En este primer caso se tomó en cuenta la edad y género del paciente para observar si es un factor importante para predecir la muerte ante IC.

Metodología empleada

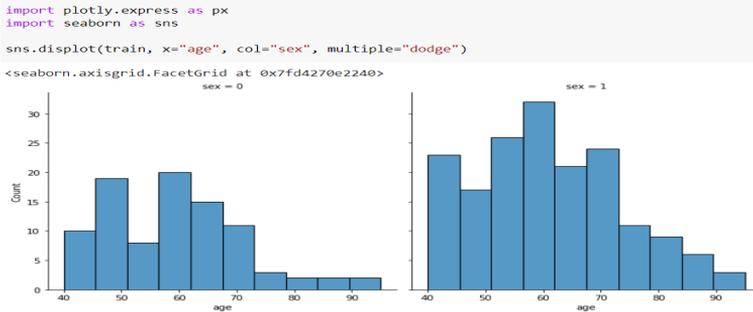
La metodología empleada fue cuantitativa de diseño exploratorio, es un tipo de investigación para estudiar el problema respecto algunas fases. Para obtener nuestros los datos de estudio realizamos un ‘Grupo Focal’, donde

expusimos lo investigado y llegamos a tomar los puntos más relevantes. Así mismo investigaciones en línea y bibliográfica. Identificando el problema, y proponiendo una herramienta de ayuda.

Para esto empleamos 4 modelos de aprendizaje KNN, arboles de decisión, regresión lineal, SVM y Random forest, estos nos ayudaran a determinar la predicción. Todo fue desarrollado en Python y sincrónicamente en Google Colab, usando librerías como Pandas, NumPy, SKLearn y otros.

Resultados Hallados

Fig 1: Análisis de comparación en población entre hombres y mujeres



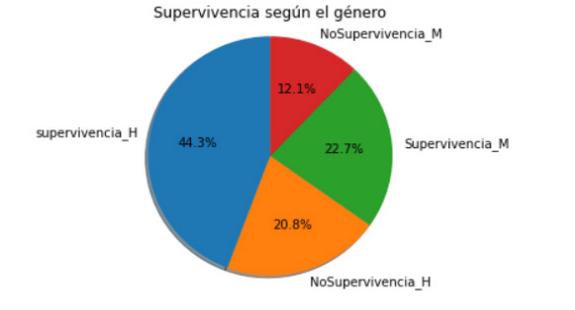
Fuente: Elaboración Propia

Resultados: Según el grafico se pudo observar que se tienen más pacientes entre 50 a 72 años Se observa que se tiene más pacientes hombres. Lo cual indica que tienden a padecer IC.

Análisis de porcentaje de Supervivencia según el género

Se selecciona el género y sexo. Se le asigna a cada uno el evento de muerte para determinar cuántos hombres y mujeres sobreviven o no sobreviven ante la IC.

Fig. 2: Análisis de Supervivencia entre Hombres y Mujeres



Fuente: Elaboración Propia

Resultados: Según el gráfico se puede observar que:

- El 22.7% de las mujeres sobreviven ante la IC.
- El 12.1% de las mujeres no sobreviven ante la IC.
- El 44.3% de los hombres sobreviven ante la IC.
- El 20.8% de los hombres no sobreviven ante la IC.

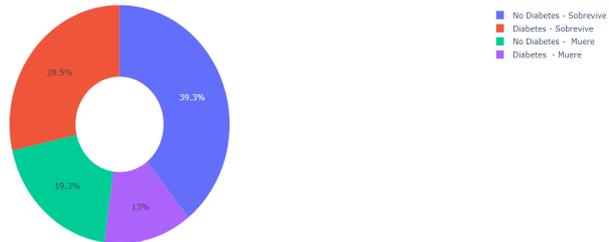
Probabilidad de una persona con diabetes o sin diabetes sobrevive ante la IC

La diabetes está estrechamente asociada a la insuficiencia cardíaca y, de hecho, se estima que cerca del 40% de los pacientes con insuficiencia cardíaca son diabéticos. A su vez, la diabetes acelera la evolución de esta enfermedad cardíaca. (prnoticias, 2016).

Se selecciona los datos donde una persona con diabetes o sin diabetes sobrevive o no sobrevive ante la IC.

Fig.3: Análisis de Probabilidad de Diabetes

Implicación de Diabetes antes no Supervivencia de IC

**Fuente:** Elaboración Propia**Resultados:** Según el grafico se puede observar que:

- El 39.3% de personas que no tienen diabetes sobreviven ante la IC.
- El 28.5% de personas que tienen diabetes sobreviven ante la IC.
- El 19.3% de personas que no tienen diabetes no sobreviven ante la IC.
- El 20.8% de personas que tienen diabetes no sobreviven ante la IC.

Probabilidad de que una persona con un bajo porcentaje de sangre que sale del corazón en cada contracción no sobreviva ante la IC

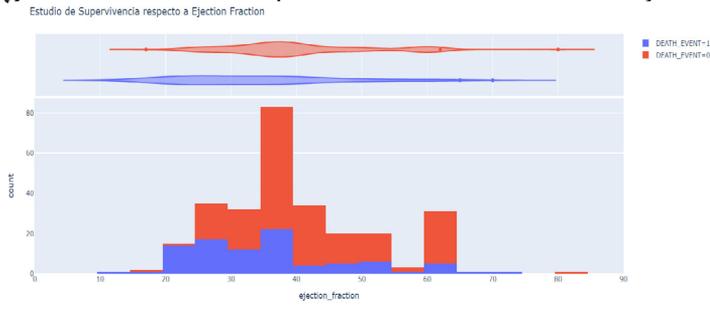
La fracción de eyección es una medida que puede medir la salud del corazón. Un número bajo de fracción de eyección puede ser un indicador de insuficiencia cardíaca. (Mankad, 2019).

Números de fracción de eyección:

- 50 a 70%: función cardíaca normal
- 40 a 55%: función cardíaca por debajo de lo normal. Puede indicar daño cardíaco previo por ataque cardíaco o miocardiopatía.
- Más del 75%: puede indicar una afección cardíaca como la miocardiopatía hipertrófica, una causa común de paro cardíaco repentino.
- Menos del 40%: puede confirmar el diagnóstico de insuficiencia cardíaca.

Se selecciona el dato de fracción de eyección y se lo relaciona con el dato de evento de muerte.

Fig. 4: Análisis entre poblaciones de factor de eyección



Fuente: Elaboración Propia

Resultados: Se puede observar que una mayoría se encuentra debajo de lo normal. Y a causa de que un porcentaje que está debajo de lo normal no han sobrevivido ante la IC.

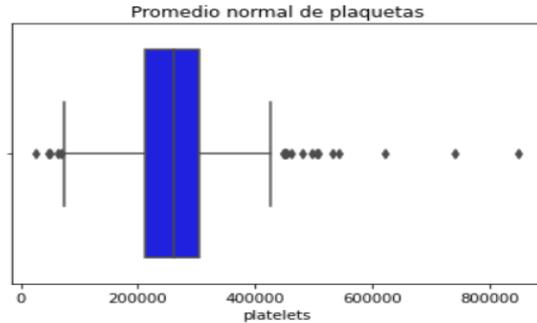
El desbalance en el promedio normal de plaquetas una razón para determinar la Supervivencia ante IC

Las plaquetas, son células sanguíneas producidas en la médula ósea y que son responsables por el proceso de coagulación sanguínea. La cantidad normal de plaquetas en la sangre es de 150,000 a 450,000 por micro litro (mcL) o 150 a $400 \times 10^9/L$. Los rangos de los valores normales pueden variar ligeramente.

Se selecciona el dato de número de plaquetas en la sangre para observar si los pacientes se encuentran o no dentro de lo normal. (AEAL, 2017).

Fig. 5: Análisis del promedio normal de plaquetas

```
sns.boxplot(x = train.platelets, color = 'blue')
plt.title("Promedio normal de plaquetas")
plt.show()
```



Fuente: Elaboración Propia

Resultados: Observamos que una mayoría se encuentra dentro de lo normal en el número de plaquetas en la sangre y se encuentran pocos valores atípicos.

La hipertensión un factor importante ante la supervivencia de IC

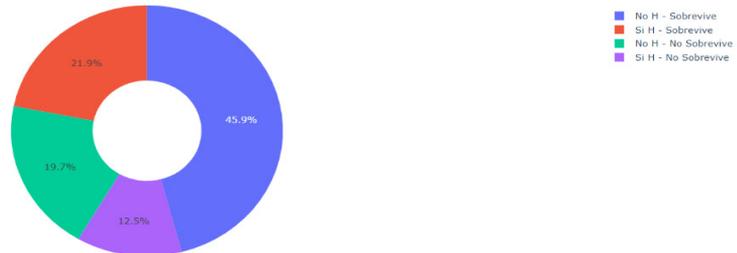
El corazón es el órgano encargado de bombear la sangre oxigenada a través de las arterias hacia todo el organismo. Al avanzar, la sangre ejerce una presión contra las paredes de las arterias, que se mide como presión arterial.

La hipertensión arterial se define por la detección de promedios de la presión arterial sistólica (“máxima”) y/o diastólica (“mínima”) por encima de los límites establecidos como normales. Dicho límite es de 140 mmHg para la sistólica y de 90 mmHg para la diastólica.

Se selecciona los datos donde una persona con hipertensión o sin hipertensión sobrevive o no sobrevive ante la IC (Universitario, 2018).

Fig. 6: Análisis de Supervivencia según la hipertensión

Análisis de Supervivencia según la Hipertensión



Fuente: Elaboración Propia

Resultados: Según el grafico se puede observar que:

- El 45.9% de personas que no tienen hipertensión sobreviven ante la IC.
- El 21.9% de personas que tienen hipertensión sobreviven ante la IC.
- El 19.7% de personas que no tienen hipertensión no sobreviven ante la IC.
- El 12.5% de personas que tienen hipertensión no sobreviven ante la IC.

El alto nivel de enzimas un factor para la No Supervivencia ante la IC

Las enzimas son proteínas complejas que producen un cambio químico específico en todas las partes del cuerpo. Por ejemplo, pueden ayudar a descomponer los alimentos que consumimos para que el cuerpo los pueda usar. La coagulación de la sangre es otro ejemplo del trabajo de las enzimas.

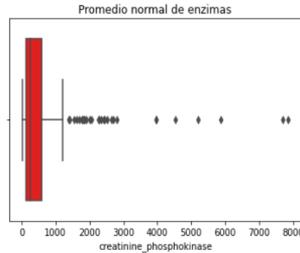
Normalmente se encuentran niveles bajos de estas proteínas y enzimas en la sangre, pero si el músculo cardíaco está lesionado, como por un ataque cardíaco, las proteínas y enzimas se escapan de las células dañadas del músculo cardíaco y aumentan sus niveles en el torrente sanguíneo.

Valores normales de CPK total: de 10 a 120 microgramos por litro (mcg/L). (J., 2019).

Se selecciona el dato del nivel de encimas para observar si los pacientes se encuentran o no dentro de lo normal.

Fig. 7: Promedio normal de Enzimas

```
sns.boxplot(x = train.creatinine_phosphokinase, color = 'red')
plt.title("Promedio normal de enzimas")
plt.show()
```



Fuente: Elaboración Propia

Resultados: Según la gráfica observamos que una mayoría se encuentra dentro de lo normal en el nivel de enzimas y se encuentran pocos valores atípicos.

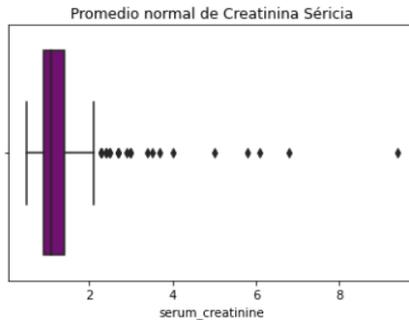
El alto nivel de creatinina sérica un factor para la No Supervivencia ante la IC

La creatinina es un producto de desecho generado por los músculos como parte de la actividad diaria. Normalmente, los riñones filtran la creatinina de la sangre y la expulsan del cuerpo por la orina.

Un resultado normal de nivel de creatinina sérica es de 0.7 a 1.3 mg/dL (de 61.9 a 114.9 $\mu\text{mol/L}$) para los hombres y de 0.6 a 1.1 mg/dL (de 53 a 97.2 $\mu\text{mol/L}$) para las mujeres. Las mujeres con frecuencia tienen niveles de creatinina más bajos que los hombres. Esto se debe a que ellas frecuentemente tienen menor masa muscular. (Hinkle J, 2014).

Se selecciona el dato del nivel de creatinina sérica para observar si los pacientes se encuentran o no dentro de lo normal.

Fig. 8: Promedio normal de Creatina sérica



Fuente: Elaboración Propia

Resultados: Según la gráfica observamos que una mayoría se encuentra dentro de lo normal en el nivel de creatinina sérica y se encuentran pocos valores atípicos.

Algoritmo de Predicción

En el presente proyecto se utilizará 5 algoritmos de predicción y se identificará cual es más preciso para la predicción ante la IC.

Entrenamiento de Datos para definir la Supervivencia de un Paciente con IC
En primer lugar, se importará las librerías que utilizaremos como:

- **Pandas:** Es una biblioteca de software escrita como extensión de NumPy para manipulación y análisis de datos para el lenguaje de programación Python. En particular, ofrece estructuras de datos y operaciones para manipular tablas numéricas y series temporales.
- **Numpy:** Es una biblioteca para el lenguaje de programación Python que da soporte para crear vectores y matrices grandes multidimensionales, junto con una gran colección de funciones matemáticas de alto nivel para operar con ellas.
- **Matplotlib:** Es una biblioteca para la generación de gráficos a partir de datos contenidos en listas o arrays en el lenguaje de programación

Python y su extensión matemática NumPy. Proporciona una API, pylab, diseñada para recordar a la de MATLAB.

También se importará el dataset que contiene los registros clínicos de personas que tiene IC.

Datos seleccionados para el entrenamiento

Se seleccionarán los siguientes datos para realizar nuestro entrenamiento ya que sé que se consideran importantes.

- **Age:** La edad, ya que se tienen más pacientes de la tercera edad.
- **Sex:** El sexo ya que se observó según el análisis de datos que hay diferencia de sobrevivir ante la IC entre hombres y mujeres.
- **Creatinine_phosphokinase:** El nivel de enzimas más alto de lo normal, podría significar que tiene inflamación del músculo cardíaco (miocardio) o que está teniendo o ha tenido recientemente un ataque al corazón.
- **Ejection_fraction:** Una medición de la fracción de eyección inferior al 40% puede ser evidencia de insuficiencia cardíaca o miocardiopatía.
- **High_blood_pressure:** La hipertensión supone una mayor resistencia para el corazón, que responde aumentando su masa muscular (hipertrofia ventricular izquierda). A la larga, está hipertrofia patológica (al contrario de la hipertrofia que se produce con el ejercicio) acaba siendo perjudicial, pudiendo producir angina, arritmias e insuficiencia cardíaca. (Torres, 2017).
- **Platelets:** Ya que tener un numero fuera de lo normal de plaquetas en la sangre puede producir un ataque cardiaco.
- **Serum_creatinine:** Un nivel alto de creatinina puede ser a causa de que un paciente tiene hipertensión.
- **Time:** Los días de tratamiento son importantes ya que a menores días de tratamiento es muy probable que el paciente no sobreviva, mientras que a mayores días de tratamiento es más probable que el paciente si sobreviva.

Importación de los módulos y división aleatoria del Dataset al 80%

Se importa los diferentes módulos que utilizaremos de la librería sklearn.

Scikit-learn es la librería más útil para Machine Learning en Python, proporciona algoritmos de aprendizaje supervisados y no supervisados.

Aquí tomaremos el conjunto del dataset dividido al 80% de los datos donde el 20% serán los datos de prueba, es decir, son aquellos datos que no vio el algoritmo de predicción.

Se seleccionan los datos que tomaremos para el entrenamiento del dataset. Se utiliza el `train_test_split`. La función `train_test_split` nos permite dividir un dataset en dos bloques, típicamente bloques destinados al entrenamiento y validación del modelo (llamemos a estos bloques “bloque de entrenamiento” y “bloque de pruebas”).

En estos modelos se utilizará la función `Accuracy_score` y una matriz de confusión

- **Accuracy score:** Es la exactitud (accuracy) mide el porcentaje de casos que el modelo ha acertado.
- **Matriz de confusión:** La matriz de confusión es una herramienta muy útil para valorar cómo de bueno es un modelo clasificación basado en aprendizaje automático. En particular, sirve para mostrar de forma explícita cuándo una clase es confundida con otra, lo cual nos, permite trabajar de forma separada con distintos tipos de error.

Fig. 9: Matriz de confusión

| | | Actual Values | |
|------------------|-----|----------------|----------------|
| | | Yes | No |
| Predicted Values | Yes | True Positive | False Positive |
| | No | False Negative | True Negative |

Fuente: Elaboración Propia

Modo de uso de cada modelo

Datos de entrenamiento para cada modelo

- `x_train` serán los datos de entrada.
- `y_train` los de salida.

Se crea una instancia del modelo

Ejemplo:

- `reg_log = LogisticRegression ()`

Se entrena el modelo con los datos de entrenamiento.

Ejemplo:

- `reg_log.fit(x_train,y_train)`

Se usa el modelo entrenado para obtener las predicciones con datos nuevos.

Ejemplo:

- `reg_log_pred = reg_log.predict(x_test)`

Se almacena el porcentaje de exactitud del modelo.

Ejemplo:

- `accuracy_score(y_test, reg_log_pred)`

Se pasa el porcentaje de exactitud del modelo y la predicción para graficar su matriz de confusión.

Ejemplo:

- `cm = confusión_matrix(y_test, reg_log_pred)`

Modelo de Regresión Logística

La regresión logística es una técnica de aprendizaje supervisado para clasificación. Es muy usada en muchas industrias debido a su escalabilidad y explicabilidad. (Martinez, 2020).

Utilizando este modelo, su respectiva matriz de confusión nos dio como resultado los siguientes datos:

- 40 verdaderos positivos.
- 3 falsos positivos.
- 5 falsos negativos.
- 12 verdaderos negativos.

Modelo de KNN

K-Nearest-Neighbor es un algoritmo basado en instancia de tipo supervisado de Machine Learning. Puede usarse para clasificar nuevas muestras (valores discretos) o para predecir (regresión, valores continuos). Al ser un método sencillo, es ideal para introducirse en el mundo del Aprendizaje Automático. Sirve esencialmente para clasificar valores buscando los puntos de datos “más similares” (por cercanía) aprendidos en la etapa de entrenamiento (ver 7 pasos para crear tu ML) y haciendo conjeturas de nuevos puntos basado en

esa clasificación. (Na8, 2018)

Utilizando este modelo, su respectiva matriz de confusión nos dio como resultado los siguientes datos:

- 36 verdaderos positivos.
- 7 falsos positivos.
- 13 falsos negativos.
- 4 verdaderos negativos.

Modelo de Arboles de Decisión

Según Ferero (2020), son un tipo de algoritmos de aprendizaje supervisado (i.e., existe una variable objetivo predefinida).

Principalmente usados en problemas de clasificación.

Las variables de entrada y salida pueden ser categóricas o continuas.

Divide el espacio de predictores (variables independientes) en regiones distintas y no superpuestas.

Parámetros de Árbol de decisión:

- **max_leaf_nodes:** Los mejores nodos se definen como una reducción relativa de la impureza.
- **random_state:** Controla la aleatoriedad del estimador. Las características siempre se permutan aleatoriamente en cada división.
- **criterion:** La función para medir la calidad de una división. Los criterios admitidos son “gini” para la impureza de Gini y “entropía” para la ganancia de información. En nuestro caso utilizaremos la entropía.

Utilizando este modelo, su respectiva matriz de confusión nos dio como resultado los siguientes datos:

- 43 verdaderos positivos.
- 0 falsos positivos.
- 6 falsos negativos.
- 11 verdaderos negativos.

Soporte de Máquina Vectorial SVM

En el aprendizaje automático, las máquinas de vectores de soporte (SVM, también redes de vectores de soporte) son modelos de aprendizaje supervisado con algoritmos de aprendizaje asociados que analizan los datos utilizados para el análisis de clasificación y regresión.

Las SVM se pueden utilizar para resolver varios problemas del mundo real:

- Las SVM son útiles en la categorización de texto e hipertexto, ya que su aplicación puede reducir significativamente la necesidad de instancias de entrenamiento etiquetadas tanto en la configuración estándar inductiva como transductiva.
- La clasificación de imágenes también se puede realizar utilizando SVM.
- Clasificación de datos satelitales como datos SAR utilizando SVM supervisado.
- Los caracteres escritos a mano se pueden reconocer utilizando SVM.
- El algoritmo SVM se ha aplicado ampliamente en las ciencias biológicas y otras.

Utilizando este modelo, su respectiva matriz de confusión nos dio como resultado los siguientes datos:

- 43 verdaderos positivos.
- 0 falsos positivos.
- 17 falsos negativos.
- 0 verdaderos negativos.

96 Random Forest

Un modelo Random Forest está formado por un conjunto de árboles de decisión individuales, cada uno entrenado con una muestra ligeramente distinta de los datos de entrenamiento generada mediante bootstrapping). La predicción de una nueva observación se obtiene agregando las predicciones de todos los árboles individuales que forman el modelo.

Parámetros de Random Forest:

- **max_features:** La cantidad de características que se deben considerar al buscar la mejor división:
 - Si es int, considere las max_features características en cada división.
 - Si es flotante, entonces max_features una fracción y las características se consideran en cada división.int (max_features * n_features)
 - Si es “automático”, entonces max_features=sqrt(n_features).
 - Si es “sqrt”, entonces max_features=sqrt(n_features)(igual que “auto”).
 - Si es “log2”, entonces max_features=log2(n_features).
 - Si es Ninguno, entonces max_features=n_features.
- **max_depth:** La profundidad máxima del árbol. Si es None, los nodos se expanden hasta que todas las hojas sean puras o hasta que todas las hojas contengan menos de min_samples_split muestras.
- **random_state:** Controla tanto la aleatoriedad

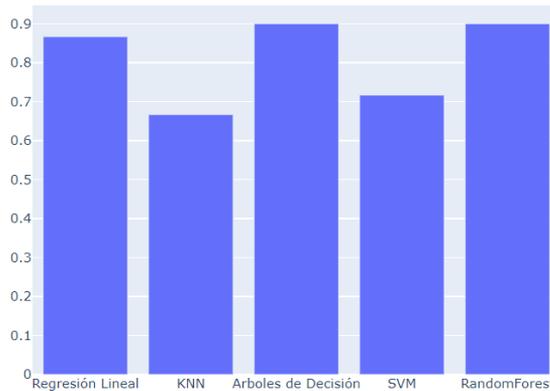
Utilizando este modelo, su respectiva matriz de confusión nos dio como resultado los siguientes datos:

- 40 verdaderos positivos.
- 3 falsos positivos.
- 3 falsos negativos.
- 14 verdaderos negativos.

Evaluación de exactitud de los modelos

Se realiza un gráfico de barras y se observa que modelo tiene más precisión según su porcentaje.

Fig. 10: Gráfico según exactitud de cada modelo
Porcentaje de Exactitud de cada modelo



Fuente: Elaboración Propia

- **Regresión Lineal:** 87% de precisión.
- **KNN:** 67% de precisión.
- **Árboles de Decisión:** 90% de precisión.
- **SVM:** 72% de precisión.
- **Random Forest:** 90% de precisión.

Se puede observar que el modelo de Árboles de Decisión y el modelo Random Forest tienen el mismo porcentaje de precisión, pero gracias a la matriz de confusión de cada uno podemos identificar cual es el mejor.

En la matriz de confusión del modelo de Árboles de decisión se tiene 0 falsos positivos mientras que en la matriz de confusión de Random Forest se tiene 3 falsos positivos, entonces podemos determinar que el mejor modelo es el de

Arboles de decisión. Ya que tener falsos positivos nos trae peligros.

Almacenamiento del mejor modelo

Ya que se ha obtenido el mejor modelo de predicción para nuestro caso se lo almacena, para utilizarlo en una interfaz gráfica.

Interfaz Grafica

Se cuenta con una interfaz gráfica para ingresar los datos de prueba y este nos indique el evento de muerte.

Interfaz

Fig. 11: Datos ingresados de un paciente que no sobrevivió a la IC

Registration Form

Supervivencia ante la insuficiencia cardíaca

Edad:

Género: Hombre Mujer

Nivel de enzima

Ejection Fraction

Hipertensión Si No

Núm. plaquetas en la sangre

Nivel de Creatina sérica en sangre (mg/dL)

Dias de Tratamiento

Respuesta:

No sobrevive

Fuente: Elaboración Propia

Fig. 12: Datos ingresados de un paciente que si sobrevivió a la IC

The screenshot shows a web form titled "Supervivencia ante la insuficiencia cardiaca" with the following fields and values:

- Edad: 55
- Género: Hombre Mujer
- Nivel de enzima: 835
- Ejection Fraction: 40
- Hipertensión: Si No
- Núm. plaquetas en la sangre: 27000
- Nivel de Creatina sérica en sangre (mg/dL): 1
- Dias de Tratamiento: 147

At the bottom of the form is a red "Enviar" button. A small modal window titled "Respuesta:" is open, showing a blue circle with the number "1" and the text "Sobrevive", with an "Aceptar" button below it.

Fuente: Elaboración Propia

Conclusión

Como conclusión se obtiene que gracias a los algoritmos de Machine Learning se puede llegar a predecir si una persona con insuficiencia cardíaca puede o no sobrevivir. Y esto puede se puede aplicar en el mundo real lo cual sería de gran ayuda para hospitales que traten estas enfermedades.

Utilizar Machine Learning de manera adecuada puede llegar a ser muy beneficiosa para diferentes áreas del mundo real.

En base al objetivo podemos demostrar que varios factores no son letales ante la IC, pero factores básicos que afectan están entre la edad y el género. El cual comprobando datos ya existentes da una correcta predicción , con un 90% de exactitud por Random forest.

100 Enlace del Proyecto

El proyecto se lo puede encontrar en GitHub:

<https://github.com/jessie4242/ProyectoAprendizajeAutomaticoInsuficienciaCardiaca>

Referencias

- AEAL. (20 de Febrero de 2017). *AEAL*. Obtenido de AEAL: <http://www.aeal.es/leucemia-mieloide-cronica-espana/1-la-sangre-y-la-medula-osea-normales/>
- Allen LA, S. L. (2019). Management of patients with cardiovascular disease approaching end of life. En B. E. Tomaselli GF, *Braunwald's Heart Disease: A Textbook of Cardiovascular Medicine*. (pág. 31). Philadelphia.
- Ferrero, R. (18 de Agosto de 2020). *Maxima Información*. Obtenido de Maxima Información: <https://www.maximaformacion.es/blog-dat/que-son-los-arboles-de-decision-y-para-que-sirven/>
- Hinkle J, C. K. (28 de Agosto de 2014). *MedlinePlus*. Obtenido de MedlinePlus: <https://medlineplus.gov/spanish/pruebas-de-laboratorio/prueba-de-creatinina/>
- Howlett, J. G. (Febrero de 2020). *MANUAL MSD*. Obtenido de MANUAL MSD Versión para profesionales: <https://www.msmanuals.com/es/professional/trastornos-cardiovasculares/insuficiencia-card%C3%ADaca/insuficiencia-card%C3%ADaca-ic>
- J., V. L. (2 de Julio de 2019). *MedlinePlus*. Obtenido de MedlinePlus: <https://medlineplus.gov/spanish/ency/article/002353.htm#:~:text=Las%20enzimas%20son%20prote%C3%ADnas%20complejas,del%20trabajo%20de%20las%20enzimas.>
- Mankad, R. (14 de Septiembre de 2019). *Mayo Clinic*. Obtenido de Mayo Clinic org: <https://www.mayoclinic.org/es-es/ejection-fraction/expert-answers/faq-20058286>
- Martinez, J. (2 de Septiembre de 2020). *IArtificial*. Obtenido de IArtificial.net: <https://www.iartificial.net/como-usar-regresion-logistica-en->

- python/
Na8. (10 de Julio de 2018). *Aprende Machine Learning*. Obtenido de Aprende Machine Learning: <https://www.aprendemachinelearning.com/clasificar-con-k-nearest-neighbor-ejemplo-en-python/>
- Prnoticias, R. (21 de Abril de 2016). *prnoticias*. Obtenido de prnoticias: <https://historico.prnoticias.com/salud/sala-de-prensa-pr-salud/20151837-diabetes-e-insuficiencia-cardiaca-una-relacion-bidireccional#inline-auto1611>
- Rouse, M. (Enero de 2017). *Search Datacenter*. Obtenido de Search Datacenter Tech Target: <https://searchdatacenter.techtarget.com/es/definicion/Aprendizaje-automatico-machine-learning>
- Torres, F. (12 de Abril de 2017). *HC marbella*. Obtenido de HC marbella: <https://www.hcmarbella.com/es/hipertension-que-es-y-como-nos-afecta/>
- Universitario, S. d. (17 de Mayo de 2018). Hospital Universitario. Obtenido de Hospital Universitario: <http://hospital.uncuyo.edu.ar/la-importancia-de-prevenir-y-tratar-la-hipertension-arterial>

Artículo recibido: 06-09-2020

Artículo Aceptado: 05-11-2020