# Caso de estudio 3

## Predicción de enfermedad de corazón

En la actualidad, las enfermedades cardíacas son la principal casusa de muerte a nivel mundial[1] . La enfermedad isquémica del corazón fue responsable del 16% de las muertes globales en el 2019. En los Estados Unidos, la principal causa de muerte es otra enfermedad cardíaca llamada enfermedad de la arteria coronaria[2].

Además de esto, entre 3 a 5% de las admisiones a hospitales se deben a pacientes con insuficiencia cardíaca. Un 2% de los costos médicos de los países desarrollados son resultado de dicha condición.[3]

Debido a este problema mundial, una de las disciplinas en las que se aborda el tema con la intención de mitigar los riesgos y costos, mejorando así la calidad de vida de las personas es Machine Learning. Utilizando diferentes técnicas de aprendizaje automático, se puede estimar con cierto grado de confianza la presencia, el subtipo, y la severidad de una enfermedad cardíaca. Más aún, se pueden predecir eventos como desestabilizaciones, re-hospitalizaciones y mortalidad.

**2- Demostrar conocimiento detallado de los conjuntos de datos correspondientes**

**3- Demostrar capacidad en la preparación previa de los datos**

Selección de atributos

Decido descartar los que tengan más de 33% de valores faltantes. También descarto name e id.

Columnas descartadas:

Name, Id, ccf, pncaden, cigs, years, dm, famhist, smoke, thaltime, slope, rldv5, ca, restckm, exerckm, restef, restwm, exeref, exerwm, thal, thalsev, thalpul, earlobe, ramus, om2, cathef, junk.

**4- Algoritmos y modelos**

En cuanto al algoritmo de aprendizaje para generar el modelo de predicción de enfermedades del corazón, claramente debemos considerar los algoritmos supervisados de clasificación. Dentro de estos, se podría justificar no utilizar knn por un consumo alto de memoria debido a que dataset de entrenamiento es uno grande, pero con los bajos costos de memoria de hoy en día, este no debería ser un impedimento para un conjunto de datos que pese un par de megas. Utilizaría Naive Bayes al principio como estándar o benchmark para comparar con otros modelos más complejos. Luego de obtener las métricas para el modelo de Naive Bayes, utilizaré difentes tipos de ensambles inicialmente Random Forest y alguna variante de Boosted Trees como XGBoost debido a su alta performance para casos de la industria estudiada. Además, como otra justificación para utilizar estos métodos, no existe un requerimiento de que las predicciones tengan que ser extremadamente rápidas, por lo que a utilización de un modelo de Random Forest con miles de árboles no sería un problema. Sería interesante también estudiar la performance de SVMs con distintos kernels para este caso particular.

Utilizando los estudios Citados por el paper "Heart Failure: Diagnosis, Severity Estimation and Prediction of Adverse

Events Through Machine Learning Techniques" (Evanthia E. Tripoliti, Theofilos G. Papadopoulos, Georgia S. Karanasiou, Katerina K. Naka, Dimitrios I. Fotiadis), Diversos grupos de investigadores utilizaron diferentes algoritmos de ML, por lo que elegir uno a priori sin probar una gran variedad de ellos sería un enfoque bastante ingenuo y sesgado en mi opinión. Debido a esto, utilizaré una variedad de algoritmos y elegiré el que mejor se adecúe al dataset elegido.

Como este es mi último caso de uso, decido utilizar toda la batería de conceptos aprendidos en el curso para generar modelos robustos con una validación correcta con la menor contaminación posible. También, se tomarán en cuenta las distintas métricas que podemos obtener de cada modelo para hacer las comparaciones de los modelos obtenidos

Primeros resultados con random forest: 84.2% de acc. Interfaz de usuario gráfica, Tabla

Descripción generada automáticamente

Resultados con gradientBoostedTrees

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

**5- Conclusiones generales sobre el caso abordado y la viabilidad u oportunidad de aplicación de técnicas de machine learning en el mismo.**

Bibliografía

[1] <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>

[2] Application of Machine Learning Algorithms to Predict Coronary Artery Calcification With a Sibship-Based Design Yan V. Sun1,\* , Lawrence F. Bielak1, Patricia A. Peyser1, Stephen T. Turner2, Patrick F. Sheedy II3, Eric Boerwinkle4, and Sharon L.R. Kardia (2008)

[3] Heart Failure: Diagnosis, Severity Estimation and Prediction of Adverse Events Through Machine Learning Techniques Evanthia E. Tripoliti a,b , Theofilos G. Papadopoulos a , Georgia S. Karanasiou a,b , Katerina K. Naka c,d , Dimitrios I. Fotiadis a,b, (2016)