

1. Descripción general

El objetivo de este proyecto es desarrollar un producto de analítica de datos para la predicción de enfermedades cardíacas. Se han identificado dos usuarios finales interesados en este producto:

1. Médicos que quieran incorporar un apoyo de analítica de datos en el proceso de evaluación de pacientes y la toma de decisiones asociada (solicitud de exámenes, chequeos y otros procedimientos).
2. Pacientes que quieran realizar una prueba sencilla en casa y así obtener información adicional que les acompañe en el proceso de decisión de visitar a un profesional médico.

Seleccione uno de los dos usuarios finales y diseñe su producto pensando en ese usuario. El nivel esperado de desarrollo de este producto es de **prototipo funcional** que deberá resultar de **iterar** sobre el producto desarrollado en el **proyecto 1**. Se espera por lo tanto que el producto sea una **versión mejorada** del obtenido en el proyecto 1.

2. Datos

Como en el proyecto 1, la fuente de datos es la disponible en el repositorio de la Universidad de California en Irvine: <https://archive-beta.ics.uci.edu/dataset/45/heart+disease>. Utilice el mismo conjunto de datos que empleó en el proyecto 1.

Tarea 1

Como parte del proyecto 1, Ud realizó un gran número de visualizaciones de datos. Seleccione **tres** de las **visualizaciones** que considere más importantes para incorporarlas en su tablero. Incluya las visualizaciones en uno o varios tabs diferentes al del modelo, o de una manera similar que considere apropiada. **Las visualizaciones deben tomar los datos de una base de datos independiente al servidor de aplicación y web.**

3. Modelos

Como parte del proyecto 1 Ud desarrolló un modelo de Redes Bayesianas basado en evidencia científica existente.

Tarea 2

Para el proyecto 2 deberá:

1. Desarrollar un método que permita evaluar el desempeño del modelo empleando datos de prueba diferentes a los datos de entrenamiento. Teniendo en cuenta que el modelo genera una predicción binaria sobre si el paciente tiene una enfermedad o no, calcule el % de verdaderos y falsos positivos y negativos.

2. Generar un nuevo modelo empleando métodos de aprendizaje de estructura y tomando como base la evidencia usada para construir el modelo del proyecto 1.
3. Evaluar su modelo del proyecto 1, el nuevo modelo y un modelo que le comparta otro grupo. Emplee los verdaderos y falsos positivos y negativos, así como puntajes estadísticos para comparar los modelos. Recuerde que para hacer válida la comparación debe emplear los mismos datos de prueba para todos los modelos.
4. Presentar y discutir sus resultados en el reporte.

4. Producto

El producto del proyecto 2 debe ser una nueva versión del tablero en dash presentado en el proyecto 1. El tablero debe quedar desplegado en AWS en una máquina de EC2, y emplear una base de datos en AWS RDS como soporte para las visualizaciones. Además, los modelos que se desplieguen deben estar serializados.

Tarea 3

Rediseñe y desarrolle el tablero teniendo en cuenta todo lo anterior, así como otros aprendizajes que hayan surgido durante el proyecto 1.

5. Entregables

Como resultado de las tareas anteriores deberá entregar los siguientes resultados y soportes:

1. **(30 puntos)** Resultado 1: reporte de máximo 7 páginas con los resultados del entrenamiento, evaluación y comparación de modelos.
2. **(25 puntos)** Resultado 2: presentación de no más de 10 minutos con los resultados principales de la modelización y el tablero desarrollado.
3. **(40 puntos)** Resultado 3: tablero desarrollado en Dash y desplegado en la nube empleando EC2 y RDS.
4. **(5 puntos)** Reporte de trabajo en equipo: incluya un pequeño reporte de cómo se dividieron las tareas entre los miembros del equipo.
5. Soporte 1: fuentes de análisis (cuadernos de jupyter o archivos .py con la modelización desarrollada, así como la evaluación de los modelos).
6. Soporte 2: fuentes del tablero (archivos .py del tablero desarrollado y .sql de la base de datos).
7. Soporte 3: repositorio Git en Github, con un historial de commits que claramente refleje el aporte de cada miembro del grupo.

Nota: si bien el trabajo es en grupo (de 2 o 3 personas), la nota es individual, luego es necesario que cada miembro del equipo demuestre su contribución al proyecto, tanto a través de los commits en el repositorio, como a través del reporte de trabajo en equipo y la sustentación.

6. Recomendaciones

1. El objetivo del proyecto es lograr un buen producto, bien soportado y claramente desarrollado. Justifique adecuadamente sus decisiones, observaciones y conclusiones.
2. Sea conciso y eficiente con el espacio. Ni el reporte ni la presentación deben ser largos. Al contrario, en un buen reporte cada gráfica y afirmación importa, y en una buena presentación cada diapositiva cuenta.
3. Es un trabajo en equipo. Defina los ítems de trabajo, asígneles entre los miembros del equipo, defina fechas de entrega y revisión interna. Discuta los resultados, observaciones y conclusiones. Priorice tareas y resultados a incluir.
4. Empiece a trabajar prontamente y discuta con el instructor su avance y resultados.

Fecha de entrega: jueves 4 de mayo, hasta las 15:30.