



Plan de Trabajo

"Predicción de la evolución de pacientes tras daño cerebral causado por trauma."

Universidad: UNIVERSIDAD POLITÉCNICA DE MADRID

Campus: MONTEGANCEDO

Centro: ESCUELA TÉCNICA SUPERIOR DE INGENIEROS INFORMÁTICOS

Identificación del Máster: MÁSTER UNIVERSITARIO EN INGENIERÍA INFORMÁTICA

Asignatura: TRABAJO FIN DE MÁSTER

Nombres del alumno:

ABEL DE ANDRÉS GÓMEZ

Fecha de Entrega: 24/02/2018

Descripción general del trabajo y lista de objetivos.

Con esta propuesta de Trabajo de Fin de Máster se persigue analizar una serie de datos provenientes de 10008 pacientes de 239 hospitales situados en 49 países que han sufrido una lesión cerebral traumática (TBI).

El objeto de análisis de los datos es poder crear un modelo idóneo que nos permita predecir la evolución de los pacientes en los próximos seis meses, pudiendo también determinar su estado.

Esta predicción se realizará a partir del conjunto de datos dado, entre los que se encuentran variables tan importantes como la edad, la forma en la que se produjo la lesión y los resultados del paciente habiendo evaluado su estado sobre la escala GSW (escala de Glasgow).

Para conseguir una buena predicción, se probarán varios modelos, teniendo en cuenta ciertas variables y descartando otras. Para la predicción se utilizará el análisis de regresión logística y se tendrá en cuenta la curva ROC, para detectar si el modelo utilizado se ajusta o no, es decir, si el modelo es bueno para realizar predicciones.

La lista de objetivos que se han definido para el plan de trabajo son los siguientes:

- Estudio del estado del arte y familiarización con el conjunto de datos con el que se va a trabajar.
- Limpieza y preparación del conjunto de datos.
- Estudio y comparación de modelos de aprendizaje sobre el conjunto de datos.
- Validación final del modelo utilizado.
- Análisis y documentación de los resultados.

Lista de tareas

El listado de tareas que se realizaran a lo largo de este TFM viene determinado en su mayor parte por las **etapas del análisis de datos** y son las siguientes:

Tarea	Duración (Semanas)	Fecha Inicio	Fecha Fin	Duración (Horas)
Estudio del estado del arte, investigación y familiarización con el conjunto de datos.	2	05/02/2018	15/02/2018	40
Preparación del conjunto de datos (limpieza y pre-procesado)	5	15/02/2018	26/03/2018	120
Análisis comparativo de distintos modelos de aprendizaje automático	12	26/03/2018	13/06/2018	256
Validación final del modelo y análisis de resultados.	2	13/06/2018	26/06/2018	40
Elaboración de la memoria del trabajo.	1	26/06/2018	02/07/2018	20
Preparación de la presentación.	1	02/07/2018	04/07/2018	9
Defensa del trabajo.	-	14/07/2018	14/07/2018	1
TOTAL HORAS				486










El periodo de estas tareas estará comprendido entre el **5 de febrero de 2018** (comienzo del semestre) y el **14 de julio de 2018** (fecha de entrega de la memoria final del TFM). Este periodo contiene aproximadamente **23 semanas**.

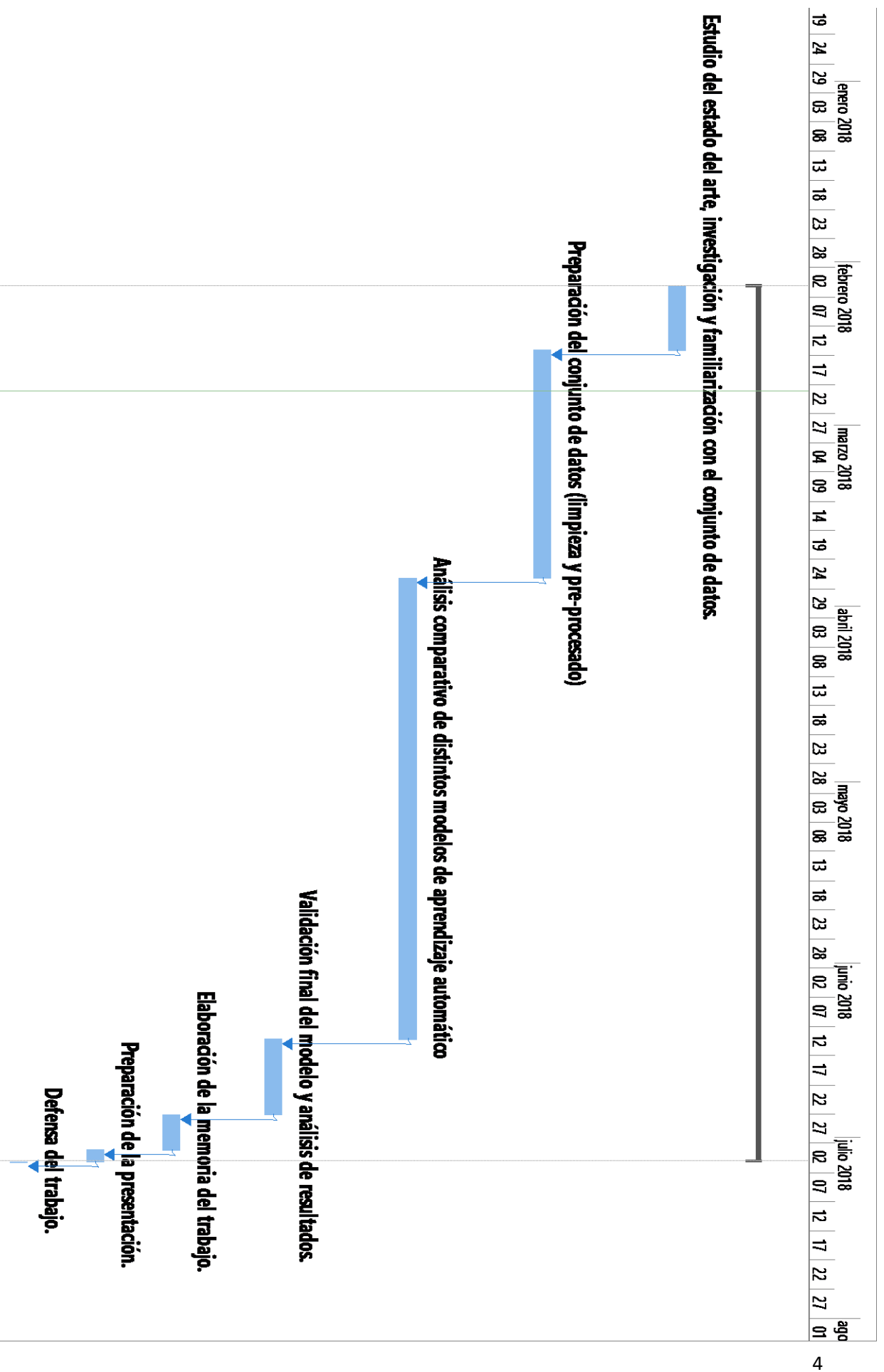
Teniendo en cuenta el número total de horas y el número de semanas, podemos determinar el número de horas que se van a dedicar a realizar el TFM, y que van a ser de **4,5 horas al día**.

Es necesario destacar que, en la tarea de “*Preparación del conjunto de datos*”, existen una serie de sub tareas cuya planificación no se ha incluido debido a que se trata de un proceso iterativo y se repetirán tantas veces como sea necesario para obtener un resultado óptimo del conjunto de datos.

Diagrama de Gantt

A continuación, se muestran las tareas junto con el diagrama de Gantt.

Id		Modo de tarea	Nombre de tarea	Duración	Comienzo	Fin	Predecesoras
1			1 TFM	486 horas	lun 05/02/18	mié 04/07/18	
2			1.1 Estudio del estado del arte, investigación y familiarización con	40 horas	lun 05/02/18	jue 15/02/18	
3			1.2 Preparación del conjunto de datos (limpieza y pre-procesado)	120 horas	jue 15/02/18	lun 26/03/18	2
4			1.3 Análisis comparativo de distintos modelos de aprendizaje	256 horas	lun 26/03/18	mié 13/06/18	3
5			1.4 Validación final del modelo y análisis de	40 horas	mié 13/06/18	mar 26/06/18	4
6			1.5 Elaboración de la memoria del	20 horas	mar 26/06/18	lun 02/07/18	5
7			1.6 Preparación de la presentación.	9 horas	lun 02/07/18	mié 04/07/18	6
8			1.7 Defensa del trabajo.	1 hora	mié 04/07/18	mié 04/07/18	7



Copia de la propuesta de trabajo escrito por el tutor

1. Título del trabajo

Predicción de la evolución de pacientes tras daño cerebral causado por trauma

2. Nombre del tutor

Antonio Latorre de la Fuente

3. Resumen general del trabajo

El objetivo de este trabajo es analizar datos provenientes de pacientes que han sufrido un traumatismo en el cerebro con distinto nivel de daño y crear un modelo predictivo que ayude a identificar la evolución del paciente durante los siguientes seis meses (efectos adversos leves, discapacidad severa o muerte). El conjunto de datos con el que se va a trabajar contendrá información de más de 10000 pacientes y contiene información personal, del accidente y clínica de diversa índole.

El estudiante, durante el trabajo, deberá ser capaz de procesar el conjunto de datos, preparándolo para el proceso de construcción del modelo (limpiando los datos, seleccionando las variables de mayor importancia, etc.). Una vez completado este paso, se procederá a realizar un estudio comparativo de distintos métodos de aprendizaje automático para llevar a cabo la predicción anteriormente citada. Dichas predicciones serán validadas con un conjunto de test previamente definido a partir del conjunto global de datos.

4. Lista de objetivos concretos

1. Familiarización del alumno con el conjunto de datos con el que se va a trabajar.
2. Limpieza del conjunto de datos:
 - a. Identificación y tratamiento de valores nulos.
 - b. Identificación de valores atípicos y tratamiento de estos (si procede).
 - c. Normalización de variables numéricas y codificación de variables categóricas.
 - d. Identificación de las variables más relevantes y reducción de dimensionalidad (si procede).
 - e. División del conjunto de datos en conjunto de aprendizaje y de test.
 - f. Análisis del balanceo de las distintas clases y tratamiento de este (si procede).
3. Estudio comparativo de distintos modelos de aprendizaje sobre el conjunto de datos de aprendizaje.
 - a. Ajuste de parámetros de los algoritmos por medio de técnicas de validación cruzada.
4. Validación final del modelo obtenido y análisis de los resultados.

5. Desglose de la dedicación total del trabajo en horas

Actividad	Horas
Estudio del estado del arte, investigación y familiarización con el conjunto de datos.	40 horas
Preparación del conjunto de datos (limpieza y pre-procesado).	120 horas
Análisis comparativo de distintos modelos de aprendizaje automático.	256 horas
Validación final del modelo y análisis de resultados.	40 horas
Elaboración de la memoria del trabajo.	20 horas
Preparación de la presentación.	9 horas
Defensa del trabajo.	1 hora

6. Conocimientos previos recomendados para hacer el trabajo

- Conocimientos en técnicas de análisis de datos.
- Conocimientos básicos del dominio médico.
- Programación en R
- Programación en Python