POLITÉCNICA
"Ingeniamos el futuro"

CAMPUS
DE EXCELENCIA
INTERNACIONAL

Máster Universitario en Ingeniería Informática

Universidad Politécnica de Madrid

Escuela Técnica Superior de
Ingenieros Informáticos

TRABAJO FIN DE MÁSTER

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

Autor: Abel de Andrés Gómez

Director: Antonio Latorre

MADRID, JULIO 2018

Agradecimientos

Resumen

Abstract

Tabla de Contenidos

Listado de Figuras

Listado de Tablas

1. INTRODUCCIÓN Y OBJETIVOS

1.1 Introducción

Las lesiones traumáticas cerebrales (**T**raumatic **B**rain **I**njury) conocidas también como lesiones cerebrales o de la cabeza ocurren cuando un golpe, impacto, sacudida u otra lesionen la cabeza causa daño al cerebro. Son principalmente el resultado de accidentes vehiculares, caídas, actos de violencia y lesiones deportivas.

Se estima que aproximadamente 2 millones de personas sufren anualmente de lesiones cerebrales, el índice de incidencia estimado es de 100 cada 100000 personas. Cada año se producen alrededor de 500000 lesiones cerebrales lo suficientemente graves como para exigir la hospitalización, llegando a causar 52000 fallecimientos anuales.

Las lesiones más comunes se dan entre los varones de 15 y 24 años, pudiendo ocurrir con cualquier edad. Muchas de estas lesiones son benignas y los síntomas desaparecen con el tiempo si reciben la atención adecuada, en otros casos el daño es más grave y puede provocar una incapacidad permanente e incluso la muerte.

Las consecuencias negativas tras un TBI dependen de muchos factores entre los que se pueden destacar la rapidez del diagnóstico y el tratamiento adecuado, que puede contribuir a aliviar algunas consecuencias de las lesiones. Por lo general es complicado predecir las consecuencias de una TBI en las primeras horas e incluso en los primeros meses ya que las consecuencias pueden permanecer ocultas durante muchos meses después.

En este contexto se hace necesario realizar un estudio de predicción que nos permita saber *"a posteriori"* las consecuencias de las lesiones en los próximos 6 meses justo después de haber tenido una lesión traumática cerebral.

1.2 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 meta a la que se pretende llegar con este trabajo es, en definitiva, obtener un patrón a partir de las características de los datos que nos permita realizar predicciones tempranas a partir de la información de un paciente que haya sufrido un daño cerebral traumático.

La lista de objetivos que se han definido para este 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.

1.3 Estructura

La estructura que va a definir el Trabajo Fin de Máster será la siguiente:

Capítulo 1. Introducción y Objetivos: En el primer capítulo se definen las necesidades por las que se va a realizar este trabajo. También se van a definir los objetivos que se persiguen con la realización de este. Por último, se presentará la estructura que tendrá el presente documento.

Capítulo 2. Estado del Arte: El objetivo que se pretende conseguir en este capítulo es introducir y situar al lector en el ámbito de Big Data, sus fases y las herramientas que se utilizan para lograr los objetivos deseados.

Capítulo 3. Desarrollo: En este capítulo se realiza una explicación detallada de todos los pasos que se van a seguir para conseguir obtener información de nuestros datos.

Capítulo 4. Resultado: En esta sección se valorarán los resultados que se han obtenido tras el uso de los distintos modelos analizados.

Capítulo 5. Líneas futuras: Este capítulo se centrará en las propuestas de líneas de trabajo futuro en este contexto.

2. ESTADO DEL ARTE

2.1 Big Data

Cuando hablamos de Big Data nos referimos a conjuntos de datos o combinaciones de conjuntos de datos cuyo tamaño (volumen), complejidad (variabilidad) y velocidad de crecimiento (velocidad) dificultan su captura, gestión, procesamiento o análisis mediante tecnologías y herramientas convencionales, como por ejemplo bases de datos relacionales, estadísticas convencionales o paquetes de visualización que nos permitan observar estos volúmenes de datos de forma inmediata.

En estas últimas décadas, el termino de Big Data ha sido y es bastante revolucionario, puesto que, aunque antes ya se utilizaba este concepto (cuando en un artículo de la revista Harper's Magazine se habla por primera vez de Big Data en 1989), no ha sido hasta comienzos del siglo XXI cuando se han empezado a plantear los problemas sobre el uso y almacenamiento de grandes volúmenes de datos. Ya en 2005 nace Hadoop, un entorno de trabajo de Big Data. Pero no es hasta 2016 cuando la palabra "Big Data" se convierte en la palabra de moda.

Centrándonos en el gran volumen de datos y de acuerdo al IDC (International Data Corporation), el volumen total de datos en 2013 fue de 2.8ZB. Los seres humanos estamos

creando y almacenando información constantemente y cada vez más en cantidades astronómicas y se espera que en 2020 se alcancen los 40ZB, unos 5247GB por persona.

A su vez, se hace necesario el análisis de estas grandes cantidades de datos con el objetivo -entre otros- de facilitar a las organizaciones el aprovechamiento de estos datos para identificar nuevas oportunidades. Este aprovechamiento se consigue mediante la recopilación y la búsqueda de tendencia dentro de estos.

2.2 Principales fases de Big Data

Las principales fases que se deben realizar a la hora de extraer información a partir de un conjunto de datos son las siguientes:

1. **Obtención de datos:** En esta primera fase lo que prima es la búsqueda y la obtención de los datos a partir de una fuente u origen. Los datos podrán ser estructurados, no estructurados, etc. Existen una serie de fuentes de donde se pueden sacar los datos, como, por ejemplo:
    a. GitHub
    b. Amazon
    c. Facebook
    d. Twitter
    e. Google (datos de mercado)

2. **Procesado de datos:** En esta fase lo que se persigue es la separación, agrupación y filtrado de datos, con el objetivo de producir que la información sea lo más significativa posible.

3. **Limpieza de datos:** Posteriormente, es necesario realizar una limpieza de los datos, puesto que muchas veces existen duplicaciones que son necesarias eliminar e incluso errores en los propios datos. Un ejemplo de estos errores podría ser los datos que se salen fuera de un intervalo cualitativo o cuantitativo determinado.

4. **Análisis exploratorio de datos:** Después de realizar la limpieza de datos estos se someterán a un tratamiento estadístico. Mediante este tratamiento estadístico se buscarán tendencias, se obtendrán histogramas para detectar grupos y se visualizarán distintos gráficos (medias, modas, desviaciones, máximos y mínimos, correlación, normalidad, etc) con el fin de identificar el modelo teórico más adecuado para la representación de estos datos.

5. **Modelado y algoritmos:** Se utilizan los datos estadísticos obtenidos en la fase anterior con el objetivo de buscar el modelo que se adapte mejor a nuestros datos y que pueda proporcionarnos.

6. **Producto de datos**: Se utiliza aplicaciones como PowerBI, Pentaho, QlikView, PeriscopeData e incluso documentos Excel, con el objetivo de visualizar y obtener resultados dinámicos.
7. **Comunicación / visualización de datos:** Se realizan informes por audiencia (comerciales, marketing, estrategia, dirección, técnicos, etc).

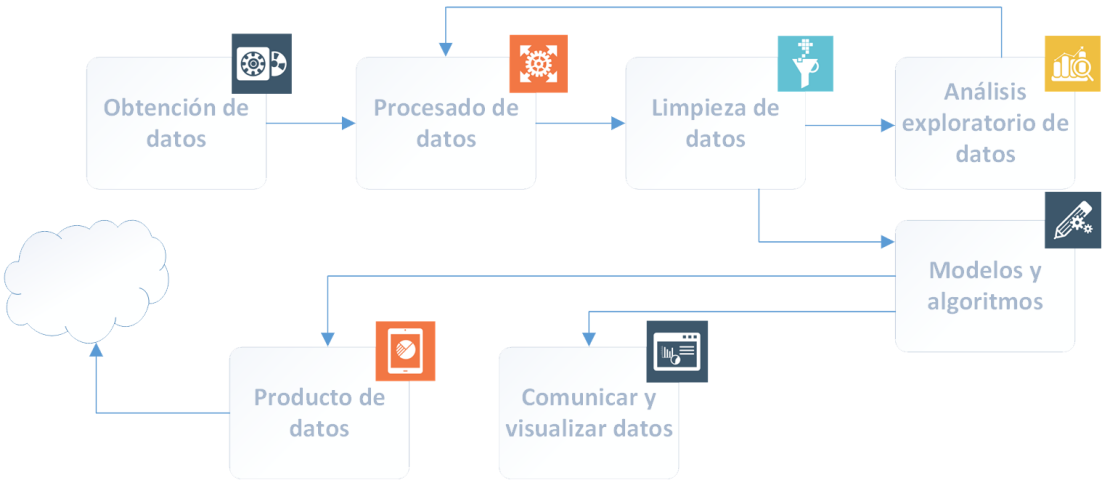

Ilustración 1.Fases de Big Data

2.3 Herramientas usadas en Big Data

A continuación, se van a describir algunas de las herramientas que se utilizan para realizar los procesos de Big Data:

1. **Hadoop**

   Es una una herramienta Big Data Open Source. Se considera el *"framework"* estándar para el almacenamiento de grandes volúmenes de datos. También se utiliza para el análisis y procesado de datos.

   Hadoop utiliza modelos de programación simple (aislando a los desarrolladores de las dificultades de la programación paralela) para el almacenamiento y el procesamiento distribuido. Hadoop distribuye por tanto el gran volumen de datos en nodos. Dispone, por consiguiente, de un sistema de archivos distribuido en cada nodo del cluster: HDFS (Hadoop Distributed File System) y se basa en el proceso de MapReduce.

2. **Apache Spark**

   Se trata de un motor de procesamiento de datos de código abierto. Realiza también una programación distribuida que consiste en distribuir el trabajo entre un conjunto de "clusters" que desde un punto de vista abstracto actúa como un solo ente que realiza el procesado. Se puede programar usando distintos lenguajes como Java,

Scala, Python o R. Es bastante más rápido en memoria y en disco que Hadoop MapReduce.

3. **Apache Storm**
   Es un sistema de computación distribuida en tiempo real orientado a procesar flujos constantes de datos, por ejemplo, datos que provienen de Twitter, pudiendo realizar estudios sobre *"trending topics"* al momento.

4. **Lenguaje R.**
   Es un lenguaje y un entorno de software frecuentemente usado para el cálculo estadístico y la visualización de gráficos. Es utilizado para la minería de datos, la investigación bioinformática y las matemáticas financieras.
   R se asemeja más a un lenguaje matemático más que a un lenguaje de programación, por lo que puede ser un inconveniente para los programadores para realizar análisis de Big Data. Su punto fuerte es el gran número de librerías creadas por la comunidad entre otras herramientas.

5. **Python**
   Es un lenguaje avanzado cuya ventaja a otros lenguajes es su uso relativamente fácil para usuarios que no están familiarizados con la programación, pero que necesitan trabajar con análisis de datos.
   También dispone de una gran comunidad detrás de este lenguaje que proporcionan un gran número de librerías, haciendo de Python un lenguaje muy eficiente para realizar Big Data.

3. DESARROLLO

3.1 PREPARACIÓN DE LOS DATOS

3.1.1 Clasificación entre: ALIVE, DEATH, NO-DATA y MD/GR

- En esta fase, se ha realizado una clasificación de los datos dados según 4 resultados finales:
  - Fallecidos o con discapacidades severas (SD-D).
  - Con discapacidad moderada o buena recuperación (MR-GR).
  - Vivos (pero sin resultados finales).
  - Sin datos.
- Para este procesado se han tenido en cuenta principalmente las siguientes variables:

- EO_Outcome
- EO_Symptoms
- TH_Outcome
- TH_Symptoms
- GOS5
- GOS8

Cuando las variables de GOS5 y GOS8 tienen datos

- Si las filas ya contenían datos en las columnas de GOS5 y GOS8, directamente se han clasificado -según estas variables-. De lo contrario, se ha tenido que analizar las otras variables.

```
head(datos.modelo[,c(17,18,27,28,29,30)])

##   EO_Outcome EO_Symptoms TH_Outcome TH_Symptoms GOS5 GOS8
## 1          4           1         NA          NA <NA> <NA>
## 2          4           3         NA          NA <NA>  MD+
## 3          4           2         NA          NA  SD* <NA>
## 4          4           2         NA          NA <NA>  GR+
## 5          4           1         NA          NA <NA> <NA>
## 6          4           2         NA          NA <NA>  SD-
```

Cuando las variables de GOS5 y GOS8 no tienen datos

- Si las variables de *"outcome"* contenían el valor de 1 (fallecimiento) o las variables de *"Symptoms"* contenían el valor de 6, directamente esas filas del conjunto de datos pasaban a clasificarse como fallecidos.

```
##    EO_Outcome EO_Symptoms TH_Outcome TH_Symptoms GOS5 GOS8 TH_Cause
## 22          1           6         NA          NA <NA> <NA>       NA
## 38          1           6         NA          NA <NA> <NA>       NA
## 50          1           6         NA          NA <NA> <NA>       NA
## 55          1           6         NA          NA <NA> <NA>       NA
## 61          1           6         NA          NA <NA> <NA>       NA
## 85          1           6         NA          NA <NA> <NA>       NA
```

- Si las variables de *"outcome"* contenían el valor de 4 (alta) y las de "*Symptoms"* el valor de 1, entonces se han clasificado como "Vivos (pero sin resultados finales)".

```
##    EO_Outcome EO_Symptoms TH_Outcome TH_Symptoms GOS5 GOS8 TH_Cause
## 1           4           1         NA          NA <NA> <NA>       NA
## 5           4           1         NA          NA <NA> <NA>       NA
```

```
## 18          4           1          NA          NA <NA> <NA>       NA
## 20          4           1          NA          NA <NA> <NA>       NA
## 36          4           1          NA          NA <NA> <NA>       NA
## 51          4           1          NA          NA <NA> <NA>       NA
```

- Se clasificarán como "Sin datos" todas aquellas filas que no contengan valores ni en las columnas de "*Symptoms"*. Se tienen en cuenta los transferidos a otros hospitales.

```
##      EO_Outcome EO_Symptoms TH_Outcome TH_Symptoms GOS5 GOS8 TH_Cause
## 384          NA          NA          NA          NA <NA> <NA>       NA
## 417          NA          NA          NA          NA <NA> <NA>       NA
## 985          NA          NA          NA          NA <NA> <NA>       NA
## 997          NA          NA          NA          NA <NA> <NA>       NA
## 2270         NA          NA          NA          NA <NA> <NA>       NA
## 2292         NA          NA          NA          NA <NA> <NA>       NA
```

- Si los "*Symptoms*" son de 4 o de 5 (Discapacidad Severa), entonces se clasificarán como "Fallecidos o con discapacidades severas".

```
##     EO_Outcome EO_Symptoms TH_Outcome TH_Symptoms GOS5 GOS8 TH_Cause
## 160          5           5          NA          NA <NA> <NA>       NA
## 241          5           5          NA          NA <NA> <NA>       NA
## 317          5           5          NA          NA <NA> <NA>       NA
## 336          5           5          NA          NA <NA> <NA>       NA
## 357          5           5          NA          NA <NA> <NA>       NA
## 573          5           5          NA          NA <NA> <NA>       NA
```

- Así mismo, si las variables de *"outcome"* contenían el valor de 4 y las de *"Symptoms"* el valor de 9, significa que el paciente ha sido dado de alta, pero no se tiene ningún dato sobre el estado final, por lo tanto, se han incluido en la clasificación de "Vivos (pero sin resultados finales)".

```
##      EO_Outcome EO_Symptoms TH_Outcome TH_Symptoms GOS5 GOS8 TH_Cause
## 409          4           9          NA          NA <NA> <NA>       NA
## 1000         4           9          NA          NA <NA> <NA>       NA
## 4859         4           9          NA          NA <NA> <NA>       NA
```

- Se han visto 3 elementos de "NODATA", cuyos pacientes obtienen un estado de *"Symptoms"* de 4, por lo que se envía a estado de fallecido, son datos anómalos.

```
##      EO_Outcome EO_Symptoms TH_Outcome TH_Symptoms GOS5 GOS8 TH_Cause
## 52           2           4          NA           4 <NA> <NA>        3
## 699          2           4          NA           4 <NA> <NA>       NA
## 3025         2           4          NA           4 <NA> <NA>        1
```

DATOS FINALES:

- Fallecidos o con discapacidades severas*:* 3559

- Con discapacidad moderada o buena recuperación*:* 5997
- Vivos (pero sin resultados finales)*:* 127
- Sin datos*:* 86
- Con NA*:* 239

3.1.2 Clasificación entre: ESCANEADOS y NO ESCANEADOS

- En primer lugar, se han encontrado ciertos datos anómalos, en los que aparecen datos escaneados (1) y no tienen los datos del escáner, entonces deberíamos ponerlo como no escaneado (2).

```
##      EO_Head.CT.scan EO_1.or.more.PH EO_Subarachnoid.bleed
## 201                1              NA                    NA
## 314                1              NA                    NA
## 1277               1              NA                    NA
## 3234               1              NA                    NA
## 3687               1              NA                    NA
## 4256               1              NA                    NA
##      EO_Obliteration.3rdVorBC EO_Midline.shift..5mm EO_Non.evac.haem
## 201                        NA                    NA               NA
## 314                        NA                    NA               NA
## 1277                       NA                    NA               NA
## 3234                       NA                    NA               NA
## 3687                       NA                    NA               NA
## 4256                       NA                    NA               NA
##      EO_Evac.haem
## 201            NA
## 314            NA
## 1277           NA
## 3234           NA
## 3687           NA
## 4256           NA
```

- A continuación, se van a clasificar los datos como:
    - Escaneados
    - No escaneados
    - En analisis

- Si el *"Outcome"* es 2 (el paciente se ha transferido a otro hospital), se ha escaneado en dicho hospital ("TH_SCAN") y no se tiene ninguna información en los escáneres, se clasificarán como "En análisis".

```
##     EO_Outcome TH_Head.CT.scan TH_1.or.more.PH TH_Subarachnoid.bleed
## 52           2            <NA>              NA                    NA
## 128          2            <NA>              NA                    NA
## 135          2            <NA>              NA                    NA
## 188          2            <NA>              NA                    NA
```

```
## 193          2            <NA>              NA                    NA
## 207          2            <NA>              NA                    NA
```

- Sobre el dataset **NO ESCANEADO**: Si el *"Outcome"* es 2 (el paciente se ha transferido a otro hospital) y no se ha realizado ningún escáner, pero si contiene datos en el escáner, entonces se clasificará como "Escaneado".

```
##      EO_Outcome TH_Head.CT.scan TH_1.or.more.PH TH_Subarachnoid.bleed
## 201           2               1               2                     2
## 217           2               1               2                     1
## 257           2               1               1                     2
## 314           2               1               1                     2
## 318           2               1               2                     2
## 1184          2               1               2                     2
```

- Sobre el dataset **NO ESCANEADO**: Nos hemos dado cuenta que existen datos anómalos, que contienen varios escáneres, pero, sin embargo, no se indica como escaneado, son los registros: 2628,3276,3279,8469,8655, etc. (En total son 12)

```
##      EO_Head.CT.scan EO_1.or.more.PH EO_Subarachnoid.bleed
## 2628               2               2                     2
## 3276               2               2                     2
## 3279               2               2                     2
## 3720               2               2                     2
## 7286               2               2                     2
## 8469               2               2                     2
```

- Sobre el dataset **EN ANALISIS**: Nos hemos dado cuenta de que existen datos anómalos. Para las variables de los pacientes que se han transferido a otro hospital (TH), existen variables de escáner *("TH_Head.CT.scan")* que se encuentran vacías, junto con el resto de variables del escáner en particular. Por lo tanto, se ha asignado el valor de 2 a la variable de escáner *("TH_Head.CT.scan")* y se han incluido en los escaneados, puesto que en todos ellos, en la variable *"EO_Head.CT.scan"* sí que existe un valor de 1 (escaneados) y no se han encontrado más anomalías en dichos datos.

```
##      EO_Head.CT.scan EO_1.or.more.PH EO_Outcome TH_Head.CT.scan
## 681                1               2          2            <NA>
## 1639               1               2          2            <NA>
## 5743               1               2          2            <NA>
## 8434               1               2          2            <NA>
## 8972               1               2          2            <NA>
##      TH_1.or.more.PH TH_Subarachnoid.bleed
## 681               NA                    NA
## 1639              NA                    NA
## 5743              NA                    NA
## 8434              NA                    NA
## 8972              NA                    NA
```

- Sobre el dataset **ESCANEADO**: Se van a eliminar todas las filas que no tengan información en el *"TH_Major.EC.injury"* y en el *"EO_Major.EC.injury"*.

```
##     EO_Outcome TH_Major.EC.injury
## 76           2                  NA
## 90           2                  NA
## 315          2                  NA
## 361          2                  NA
## 510          2                  NA
## 565          2                  NA
```

- Sobre el dataset **ESCANEADO**: Comprobamos que las variables: *"EO_Cause"* y *"EO_Symptoms"*, no contengan valores nulos.

```
##     EO_Cause EO_Major.EC.injury
## 177        3                  2
## 211       NA                  1
## 242       NA                  2
## 255        2                  2
## 293       NA                  2
## 321       NA                  1
```

- Sobre el dataset **ESCANEADO**: Comprobamos que la variable: *"EO_Outcome"* no se encuentre nula. (En total son 2 registros).

```
##      EO_Cause EO_Outcome
## 9036        3         NA
## 9333        3         NA
```

- Sobre el dataset **ESCANEADO**: Comprobamos que existe un valor anómalo (que se sale del rango) en un registro en la columna de *"EO_Major.EC.Injury"*. Este valor lo cambiaremos a positivo -> 1.

```
##      EO_Cause EO_Major.EC.injury
## 3862        2                 -1
```

DATOS FINALES:

- Vivos y escaneados: 4157
- Vivos y no escaneados: 1535
- Vivos en análisis: 305
- Fallecidos y escaneados: 2829
- Fallecidos y no escaneados: 439
- Fallecidos en análisis: 291

4.1.3 Eliminación y centralización de variables

- Se va a centralizar las variables de *"PUPIL_REACT_LEFT"* y *"PUPIL_REACT_RIGHT"*.

```
##   PUPIL_REACT_LEFT PUPIL_REACT_RIGHT ESTADOESCANER
## 1                1                 1      SCANEADO
## 2                1                 1      SCANEADO
## 3                1                 1      SCANEADO
## 4                1                 1      SCANEADO
## 5                1                 1      SCANEADO
## 6                1                 1      SCANEADO
```

  - Both reactive: 5662
  - No response unilateral: 497
  - No response: 634
  - Unable to assess: 193

- Ahora vamos a ver si podemos prescindir o aunar las variables de *"EO_Cause"* y *"TH_Cause"*. Para ello veremos en que caso, ambas variables difieren:

```
##      EO_Cause TH_Cause
## 2307        1        3
## 2813        3        1
## 3285        2        3
## 4021        3        1
```

- Como se puede observar, podríamos prescindir de la variable *"TH_Cause"*, puesto que recoge la misma información que *"EO_Cause"*.

- A continuación, vamos a aunar todas las variables del escáner. Si un paciente ha sido transferido a otro hospital y se han realizado los escáneres en dicho hospital, entonces, se mantendrán los últimos valores del escáner. En caso contrario, se usarán los primeros resultados de escáner obtenidos en el primer. Además, eliminaremos todas las variables que no se utilicen.

```
##   sex age cause ec eye motor verbal pupils phm sah oblt mdls hmt outco
me
## 1   0  11     1  1   1     5      1      1   2   2    2    2   2    MD
GR
## 2   0  14     1  2   1     2      1      1   1   2    2    2   1
D
## 3   0  14     1  2   2     5      1      1   2   2    2    2   1
D
## 4   0  14     1  2   2     5      2      1   2   2    2    2   2    MD
GR
## 5   0  14     3  2   4     6      4      1   2   1    2    2   2    MD
GR
## 6   0  15     1  2   1     5      1      1   2   2    2    2   2
D
```

4. RESULTADOS

Sección pendiente de desarrollo

5. CONCLUSIONES

Sección pendiente de desarrollo

6. LÍNEAS FUTURAS

Sección pendiente de desarrollo

7. BIBLIOGRAFÍA

[Manual abreviado de Análisis Estadístico Multivariante. Jesús Montanero Fernández] http://matematicas.unex.es Recuperado el 28 de marzo de 2018 de: https://ignsl.es/historia-del-big-data/

[Análisis Multivariente, usando R. José Carlos Vega Vilca] http://cicia.uprrp.edu Recuperado el 28 de marzo de 2018 de: http://cicia.uprrp.edu/publicaciones/Papers/ManualESTA5503.pdf

[Ambrosio Torres] https://www.r-bloggers.com. Recuperado el 2 de abril de 2018 de: https://www.r-bloggers.com/lang/spanish/940

[Selva Prabhakaran] http://r-statistics.co. Recuperado el 2 de abril de 2018 de: http://r-statistics.co/Outlier-Treatment-With-R.html

[Tema 3. Contraste de la normalidad multivariante. César A. Sanchez Sellero] http://eio.usc.es. Recuperado el 7 de abril de 2018 de: http://eio.usc.es/eipc1/base/BASEMASTER/FORMULARIOS-PHP/MATERIALESMASTER/Mat_142400_mmulti1011tema3.pdf

[Contrading. Victor A. Rico] http://www.cotradingclub.com. Recuperado el 7 de abril de 2018 de: http://www.cotradingclub.com/2017/05/25/prueba-de-normalidad-en-modelos-de-prediccion/

[Un análisis con R. Datos Multivariantes. Francesc Carmona] http://www.ub.edu. Recuperado el 9 de abril de 2018 de: http://www.ub.edu/stat/docencia/EADB/Ejemplo.pdf

[Javier Seoane, Carlos P. Carmona, Rocío Tarjuelo y Aimara Planillo] http://www.uam.es. Recuperado el 11 de abril de 2018 de: http://www.uam.es/personal_pdi/ciencias/jspinill/CFCUAM2014/RF_BRT-CFCUAM2014.html

[Manuel Sigüeñas Gonzales] https://rpubs.com Recuperado el 15 de abril de 2018 de: https://rpubs.com/MSiguenas/122473

[Analytics Vidhya Content Team] https://www.analyticsvidhya.com. Recuperado el 17 de abril de 2018 de: https://www.analyticsvidhya.com/blog/2016/03/practical-guide-principal-component-analysis-python/

[Grupo IGN] https://ignsl.es. Recuperado el 25 de abril de 2018 de: https://ignsl.es/historia-del-big-data/

[Instituto de Ingeniería del Conocimiento] http://www.iic.uam.es. Recuperado el 28 de abril de 2018 de: http://www.iic.uam.es/innovacion/herramientas-big-data-para-empresa/