



UNIVERSIDAD DE BURGOS  
ESCUELA POLITÉCNICA SUPERIOR  
Grado en Ingeniería en Informática



TFG del Grado en Ingeniería Informática  
Predicción de apuestas deportivas



Presentado por Nuño Basurto Hornillos  
en Universidad de Burgos — 23/01/2017  
Tutor: Alvar Arnaiz González  
Cristobal José Carmona del Jesús

---

# Índice general

---

<b>Índice general</b>	<b>I</b>
<b>Índice de figuras</b>	<b>III</b>
<b>Apéndice A Plan de Proyecto Software</b>	<b>1</b>
A.1. Introducción . . . . .	1
A.2. Planificación temporal . . . . .	1
A.3. Estudio de viabilidad . . . . .	4
<b>Apéndice B Especificación de Requisitos</b>	<b>5</b>
B.1. Introducción . . . . .	5
B.2. Objetivos generales . . . . .	5
B.3. Catalogo de requisitos . . . . .	6
B.4. Especificación de requisitos . . . . .	6
<b>Apéndice C Especificación de diseño</b>	<b>7</b>
C.1. Introducción . . . . .	7
C.2. Diseño de datos . . . . .	7
C.3. Diseño procedimental . . . . .	8
<b>Apéndice D Documentación técnica de programación</b>	<b>9</b>
D.1. Introducción . . . . .	9
D.2. Estructura de directorios . . . . .	9
D.3. Manual del programador . . . . .	9
D.4. Compilación, instalación y ejecución del proyecto . . . . .	9
D.5. Pruebas del sistema . . . . .	9
<b>Apéndice E Documentación de usuario</b>	<b>10</b>
E.1. Introducción . . . . .	10
E.2. Requisitos de usuarios . . . . .	10

<i>ÍNDICE GENERAL</i>	II
E.3. Instalación . . . . .	10
E.4. Manual del usuario . . . . .	10
<b>Bibliografía</b>	<b>11</b>

---

## Índice de figuras

---

## *Apéndice A*

---

# Plan de Proyecto Software

---

### **A.1. Introducción**

Vamos a ver cómo ha ido evolucionando el proyecto durante su realización, así como un estudio de viabilidad donde podemos ver si podemos llevarlo a cabo en el futuro.

### **A.2. Planificación temporal**

El proyecto se ha llevado a cabo utilizando metodologías ágiles, más concretamente SCRUM, realizando un numero de reuniones donde concretábamos los objetivos que íbamos a llevar acabo en cada sprint, no solo nos comunicábamos en estas reuniones, también utilizábamos el gestor de proyectos, Trello para resolver aquellas dudas que iban surgiendo durante la realización del sprint, una vez se acababa en el sprint en la reunión observábamos el trabajo realizado. Tanto en Trello como en GitHub podemos ver como se han ido desarrollando cada uno de los objetivos de los sprints.

#### **Sprint 1 (20/09/2016 – 04/10/2016)**

Es una toma de contacto con la gran cantidad de herramientas con las que vamos a trabajar. Lo primero y fundamental es la creación de una máquina virtual Ubuntu 14.04, sobre ella se va a instalar un servidor Apache, se va a seguir un tutorial que ha sido cedido por el tutor. Tras analizar algunas herramientas de gestión de proyectos como Git, Bitbucket o GitHub, se decide trabajar con GitHub como gestor de versiones y con Trello como gestor de proyectos. Tras ello vamos a montar Drupal sobre el servidor Apache, durante la instalación de Drupal se han ido encontrado algunos problemas para enlazarlo correctamente con la base de datos que se ha creado durante la instalación

de Apache. Por último se empieza a investigar sobre el scraping y se crea un sencillo script con el que extraer algún dato.

### **Sprint 2 (04/10/2016 – 21/10/2016)**

Lo primero es crear una estructura en la base de datos en la cual poder almacenar la información, que vamos a ir recopilando mediante los algoritmos de web scraping. El primer algoritmo de scraping que vamos a crear es el que extraiga los resultados de los partidos y con ello todas las estadísticas. Es importante la selección de la página web sobre la que realizar el scraping ya que se necesitan una uniformidad en la URL. La primera página web que elegimos es la de Marca.com ya que la URL de los partidos es válida, el problema surge al extraer varios partidos ya que en algunos partidos directamente no hay estadísticas. Tras una larga búsqueda se encuentra otra página web, resultados-futbol en ella no nos encontramos con los problemas de Marca y la URL es válida. Todos los datos que extraemos mediante el algoritmo de scraping hay que cargarlos en la base de datos, esto lo vamos a hacer utilizando una sintaxis de PHP especial para el acceso a MySQL a través de PHP. Inicialmente está habiendo muchos problemas para la utilización de algunas funciones de carga de datos. Finalmente, se intentan añadir algunos temas a Drupal como por ejemplo el de Trello para poder trabajar directamente sobre Drupal. Pero nos encontramos con un error de autenticación, que indagando por foros vemos que no tiene solución por lo cual descartamos estos y trabajamos con Trello independientemente desde el navegador.

### **Sprint 3 (21/10/2016 – 08/11/2016)**

En este sprint tenemos que implementar el algoritmo de backpropagation, esto en un principio es un problema ya que mi conocimiento de PHP no es demasiado grande y el tiempo que se va a tardar en llevarlo a cabo puede extenderse. Valoramos diferentes algoritmos que vamos encontrando por internet, para tomarlos como base para la creación de nuestro algoritmo, finalmente nos quedamos con uno en Python y procedemos a su traducción a PHP, resulta costosa dado que este algoritmo en Python tiene clases y es así como lo vamos implementando en PHP. Creamos un algoritmo de scraping que extrae, de la misma página utilizada anteriormente para la extracción de resultados, las cuotas de las casas de apuestas, lo óptimo es la ejecución de este script con la menor anterioridad a la jornada, ya que no todas las cuotas se encuentran en cualquier momento de la semana y estas van variando.

### **Sprint 4 (08/11/2016 – 15/11/2016)**

Dado que el algoritmo no ha sido terminado correctamente decidimos prolongar su implementación a este sprint y tratamos de ver su respuesta ante una base de datos sencilla como Iris y más adelante probarle con otras bases

de datos como balance, pima y wine. Tras varios ajustes en el código conseguimos una buena ejecución con Iris. Para ir preparando la entrada de datos a la red neuronal del algoritmo de backpropagation, se lleva a cabo un script que calcule las rachas de los equipos, además se realiza un tercer algoritmo de scraping para la recolección de los datos de cada equipo en la clasificación, este cambio conlleva un cambio en la base de datos. Creamos las instancias para darlas de entrada en el algoritmo de backpropagation pero aún hay que hacer algunos ajustes ya que los datos no son del todo correctos.

### **Sprint 5 (15/11/2016 – 23/11/2016)**

Lo primero en este Sprint es terminar las instancias que vamos a insertar en el algoritmo de backpropagation, en el anterior sprint se cargó la clasificación del equipo y la idea en un primer momento era hacer un update por cada jornada, en vez de esto vamos a cargar las clasificación de cada equipo en cada jornada, las rachas y la clasificación se unirán en la base de datos, teniendo la información de cada equipo en cada una de las jornadas. Tras varias semanas copiando y pegando el código de Sublime Text en el nodo de Drupal se encuentra una manera de vincularlo y así ahorrar bastante tiempo.

### **Sprint 6 (23/11/2016 – 06/12/2016)**

Con las instancias ya definidas vamos a intentar que la ejecución nos de valores con los que poder empezar a trabajar, estos valores cuestan obtenerlos ya que hay que hacer varios ajustes en la red neuronal, como el control del número de iteraciones o la cantidad de neuronas que deseamos utilizar, las pruebas las vamos almacenando poco a poco para luego compararlas tranquilamente.

### **Sprint 7 (06/12/2016 - 22/12/2016)**

En este sprint vamos a dejar terminada la interfaz, los informes que muestren resultados van a quedar acabados de cara al usuario, distinguiendo tres informes, el de toda la temporada, el que muestre lo obtenido en pasadas jornadas y el de la próxima jornada. También vamos a normalizar el nombre de los nodos de Drupal y de los scripts, el objetivo es una mayor homogeneidad. Las direcciones URL también van a ser normalizadas y se va a poner una pantalla de Inicio para el usuario.

### **Sprint 8 (22/12/2016 - 09/01/2017)**

En este último sprint vamos a dejar toda la documentación cerrada y algún pequeño retoque en la web. Los scripts van a ser reforzados con control de errores fuera de ejecución para cerrar completamente el proyecto.

### A.3. Estudio de viabilidad

En esta sección vamos a analizar si se puede apostar o no por este proyecto en un futuro, vamos a tener en cuenta diferentes puntos de vista, como la viabilidad económica y la legal. Sin duda ambos factores son determinantes para seguir con el proyecto o al menos para que podamos lucrarnos de ello.

#### Viabilidad económica

La viabilidad en términos económicos podemos enfocarla de dos formas, ya que depende del uso del programa o de la venta a una empresa del mismo.

- Uso del programa: Los beneficios y pérdidas que se han ido logrando, podemos observarlos en el balance general, lo normal es que cuanta más información posea la red neuronal, mejores resultados nos de.Imagen
- Compra de una empresa: Sin duda esta no es la finalidad del proyecto, es difícil que una gran empresa apueste por un proyecto como este dado que hay servicios de multinacionales que nos ofrecen una mayor capacidad de computación que el nuestro. Este proyecto lo veo algo más destinado al pequeño consumidor que quiera consultar resultados y opciones de apuestas en un momento dado.

#### Viabilidad legal

En el aspecto legal la única técnica cuestionable es el web scraping. En cuánto a su uso, se considera ilegal si puede generar un riesgo de asociación o comporte un aprovechamiento indebido de la reputación o el esfuerzo ajeno, es decir, es desleal sin supone un obstáculo a la afirmación de esa empresa en el mercado. El Tribunal Supremo dicta que las técnicas de scraping constituyen una técnica legal si se cumplen determinadas condiciones y supuestos.[referencia]

En resumidas cuentas, no depende del scraping en sí mismo, sino del uso que se le dan a los datos extraídos con este mismo. Si el uso de estos datos resultan competencias desleal nos encontraremos ante un acto ilegal.



---

## Especificación de Requisitos

---

### B.1. Introducción

Vamos a analizar profundamente todo lo referente a los objetivos del proyecto, si las metas, acordadas en un principio han sido alcanzadas y los requisitos que hemos establecido para guiarnos a lo largo del proyecto.

### B.2. Objetivos generales

Desde el comienzo del proyecto se propusieron unos objetivos sobre los cuales construir el proyecto.

- Como bien vemos en el nombre del proyecto, predicción de apuestas deportivas, este es sin duda la piedra angular del proyecto sobre la que elaborar el resto. Hay que obtener el resultado final de partidos de fútbol, esto lo lograremos utilizando redes neuronales.
- Una vez tenemos los resultados de los partidos, estos hay que mostrarlos a los usuarios, lo realizaremos utilizando una interfaz sencilla y atractiva, la cual pueda utilizar el usuario, dado que estamos tratando con las apuestas deportivas se van a mostrar algunas cuotas de las casas de apuestas.
- Dado que no se puede estar a diario pesando que algoritmos de scraping ejecutar o que datos hay que mostrar es necesario automatizar todo, de esta manera el usuario tendrá toda la información actualizada y no esperar la intervención humana.

### **B.3. Catalogo de requisitos**

En este apartado vamos a ver todas las características que posee nuestro proyecto y que se han implementado.

### **B.4. Especificación de requisitos**

---

## Especificación de diseño

---

### C.1. Introducción

En este apéndice vamos a tratar de todo aquello relacionado a la organización del proyecto.

### C.2. Diseño de datos

A lo largo de este proyecto la cantidad de datos con la que hemos trabajado es muy grande, dada la necesidad de datos de una red neuronal para un buen pronóstico. En la sección vamos a ver como hemos almacenado estos datos y que uso hemos hecho de estos.

#### Almacenamiento de datos

Los datos han sido almacenados en el sistema de gestión de bases de datos relacionales MySQL, estala hemos administrado utilizando PHPMyAdmin. Dada la gran cantidad de datos y su diversidad ha sido necesario trabajar con diferentes tablas. A continuación se van a explicar cada una de las tablas utilizadas:

- **equipos** En esta tabla encontramos los 20 equipos de primera división, se han dispuesto alfabéticamente y esta tabla posee tres campos, entre ellos está el nombre, este campo es muy importante a la hora de realizar el scraing, ya que es así como identifica la web a cada equipo, el tercer y último campo es «nombreCompleto», utilizado para mostrar el nombre de dicho equipo en la interfaz.
- **fecha\_jornada** Esta tabla contiene cada partido de cada jornada de la liga, con algunos campos como la fecha, la jornada o los equipos contendientes.

- **clasificacion\_jornada** Aquí podemos observar la posición en la clasificación de cada uno de los equipos en cada jornada, así como las últimas rachas de los equipos. Esta tabla resulta fundamental a la hora de insertar datos en la red neuronal.
- **partidos** Mediante las técnicas de scraping extraemos datos de cada partido, estos datos son almacenados aquí y contiene todas las estadísticas de ambos equipos. Los datos son utilizados para la entrada en la red neuronal.
- **pronosticos** Esta tabla contiene la salida de datos de la red neuronal y el resultado final obtenido por esta para cada uno de los partidos.
- **apuestas** Organiza todas las cuotas de las casas de apuestas para cada partido, registrando todos los identificadores.
- **cuotas** Posee las cuotas y sus valores dependiendo de la victoria local, empate o victoria visitante.
- **balance\_general** Almacena el balance de cada casa de apuestas en cada jornada, esta tabla se utiliza posteriormente para mostrarla en la interfaz.

### C.3. Diseño procedimental

En esta sección se va a poder observar el funcionamiento del proyecto.

#### Diagrama de flujo

En esta imagen ?? podemos ver la ejecución del programa y de su dependencia del día en el que nos encontremos. Los campos *fecha<sub>a</sub>ntes*

---

## **Documentación técnica de programación**

---

- D.1. Introducción
- D.2. Estructura de directorios
- D.3. Manual del programador
- D.4. Compilación, instalación y ejecución del proyecto
- D.5. Pruebas del sistema

---

## **Documentación de usuario**

---

- E.1. Introducción**
- E.2. Requisitos de usuarios**
- E.3. Instalación**
- E.4. Manual del usuario**

---

## **Bibliografía**

---