



**UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO**

---

---

**FACULTAD DE INGENIERÍA**

**MÉTODO QUETZAL: SISTEMA PROBABILÍSTICO  
DE APUESTAS DEPORTIVAS CON AVERSIÓN AL  
RIESGO**

**TESIS PROFESIONAL**

**QUE PARA OBTENER EL TÍTULO DE:  
INGENIERO EN COMPUTACIÓN**

**PRESENTA :**

**LUIS JESÚS SALAZAR VELÁZQUEZ**



**DIRECTOR DE TESIS:  
Ing. Manuel Enrique Castañeda Castañeda**

**Ciudad Universitaria, México, 2014**

## Índice

ÍNDICE	1
ÍNDICE DE TABLAS	4
ÍNDICE DE ILUSTRACIONES	5
1- INTRODUCCIÓN - EL FUTBOL Y LAS APUESTAS	6
1.1- Fútbol	6
1.2- Mercado de apuestas	6
1.2.1- Quiniela	6
1.2.2- Por resultado	7
1.2.3 Mercados por resultado	8
1.3- Objetivos	10
2- MARCO TEÓRICO	11
2.1- Metodología	11
2.1.1- Pasos	11
2.2- Aversión al riesgo	12
2.2.1- Riesgo de impago.	12
2.2.2- Riesgo de modelo	13
2.2.3- Riesgo por el azar	13
2.2.4- Conclusiones	13
2.3- ACP – Análisis de Componentes Principales	14
2.3.1- Pasos	14
2.3.2- Problema de dimensionalidad	14
2.3.3- Ejemplo	14
2.3.4- Conclusiones	16
2.4- Clasificador Bayesiano	16
2.4.1- Ejemplo	17
2.4.2- Conclusiones	18
2.5- Regresión Logística Multinomial	18

<b>2.6- Red Neuronal Artificial</b>	<b>19</b>
2.6.1- Regresión Logística en una Red Neuronal Artificial Perceptrón	20
2.6.2- Conclusiones	24
<b>3- ESTADO DEL ARTE</b>	<b>25</b>
<b>3.1- Métodos cualitativos</b>	<b>25</b>
<b>3.2- Rating Systems</b>	<b>26</b>
3.2.1- Rateform	26
3.2.2- Goals Superiority	28
<b>4- ANÁLISIS</b>	<b>30</b>
<b>4.1- Aprovisionamiento</b>	<b>30</b>
<b>4.2- Pre procesamiento</b>	<b>32</b>
4.2.1- ACP	33
<b>4.3- Modelo matemático</b>	<b>33</b>
4.3.1- Clasificador Bayesiano	33
4.3.2- Regresión Logística Trinomial- RNA	36
<b>5- DISEÑO E IMPLEMENTACIÓN</b>	<b>37</b>
<b>5.1- Filosofía</b>	<b>37</b>
<b>5.2- Definición Funcional: Aprovisionamiento</b>	<b>37</b>
<b>5.3- Definición Funcional: Preprocesamiento</b>	<b>37</b>
<b>5.4- Definición Funcional: Clasificador Bayesiano</b>	<b>39</b>
<b>5.5- Definición Funcional: Regresión Logística</b>	<b>40</b>
<b>6- EVALUACIÓN</b>	<b>42</b>
<b>6.1- Diseño de pruebas</b>	<b>42</b>
<b>6.2- Resultados</b>	<b>43</b>
<b>7- MÉTODO QUETZAL EN INTERNET</b>	<b>46</b>
<b>7.1- Servicio</b>	<b>46</b>

<b>7.2- Promoción</b>	<b>47</b>
<b>7.3- Futuro</b>	<b>48</b>
<b>8- CONCLUSIONES</b>	<b>49</b>
<b>ANEXO A – ESTADÍSTICAS DEL SET DE DATOS</b>	<b>50</b>
<b>Promedio de Goles a Favor de Local</b>	<b>50</b>
<b>Promedio de Goles en Contra de Local</b>	<b>52</b>
<b>Promedio de Goles a Favor de Visitante</b>	<b>53</b>
<b>Promedio de Goles en Contra de Visitante</b>	<b>55</b>
<b>ANEXO B – EVIDENCIAS</b>	<b>57</b>
<b>BIBLIOGRAFÍA</b>	<b>63</b>

## Índice de Tablas

Tabla 1 Quinielas en hispanoamérica .....	7
Tabla 2 Datos para ejemplo de ACP.....	14
Tabla 3 Datos expresado en la transformación lineal ACP .....	15
Tabla 4 Resultado del Clasificador Bayesiano .....	17
Tabla 5 Datos para ejemplo de RNA .....	20
Tabla 6 Proceso iterativo de RNA .....	23
Tabla 7 Resultado del entrenamiento de la RNA .....	23
Tabla 8 Ejemplo de actualización en el modelo Rateform.....	27
Tabla 9 Lista de torneos que abarca el Método Quetzal .....	30
Tabla 10 Resultados del ACP .....	33
Tabla 11 Resultados del test Anderson-Darlington .....	35
Tabla 12 Comparativa en la clasificación .....	44
Tabla 13 Comparativa en la aversión al riesgo.....	45
Tabla 14 Goles a Favor de Local .....	50
Tabla 15 Goles en Contra de Local.....	52
Tabla 16 Goles a Favor de Visitante.....	53
Tabla 17 Goles en Contra de Visitante.....	55

## Índice de Ilustraciones

Ilustración 1 Gráfica ACP .....	16
Ilustración 2 Ejemplo Clasificador Bayesiano .....	17
Ilustración 3 Neurona.....	19
Ilustración 4 Perceptron.....	20
Ilustración 5 Datos para ejemplo RNA .....	21
Ilustración 6 Modelo de RNA para Regresión Logística .....	21
Ilustración 7 Resultado del entrenamiento de la RNA .....	24
Ilustración 8 Ejemplo de pantalla donde se obtienen los datos de entrada .....	32
Ilustración 9 Modelo de la RNA para el Método Quetzal .....	36
Ilustración 10 Goles a Favor de Local.....	52
Ilustración 11 Goles en Contra de Local.....	53
Ilustración 12 Goles a Favor de Visitante .....	55
Ilustración 13 Goles en Contra de Visitante .....	56
Ilustración 14 Pantalla de Youtube .....	57
Ilustración 15 metodoquetzal.net .....	57
Ilustración 16 metodoquetzal.com.....	58
Ilustración 17 Versión gratuita .....	58
Ilustración 18 Apuesta 12/01 .....	59
Ilustración 19 Apuesta 26/01 .....	59
Ilustración 20 Apuesta 31/01 .....	60
Ilustración 21 Apuesta 8/02 .....	61
Ilustración 22 Apuesta 15/02 .....	62

## 1- Introducción - El futbol y las apuestas

### 1.1- Futbol

Deporte de conjunto, en el que participan dos equipos de 11 jugadores, cuya finalidad es anotar el mayor número de goles en la portería del contrario, a su vez evitando que el equipo contrario marque en la propia. Se considera ganador al equipo que más goles haya anotado, dándose también el caso en el que ambas escuadras hayan anotado el mismo número de goles, lo que se considerará un empate.

Toda acción en el partido ocurrirá en 2 tiempos de 45 minutos más minutos de compensación que reciben el nombre de tiempo reglamentario, separados por un periodo de descanso. En caso de empate durante una eliminatoria donde siempre debe haber un ganador se recurre a tiempos extras, o a penales.

### 1.2- Mercado de apuestas

Es el nombre que recibe la oferta y demanda por parte de casas de apuestas y/o estado, de multiplicadores de inversión. De manera general existen 2 formas de apuestas deportivas, la quiniela y por resultado.

#### 1.2.1- Quiniela

En esta forma, la casa de apuestas genera un listado con un número determinado de partidos que se jugaran en una fecha determinada. El objetivo es tener el mayor número de aciertos. Tiene una escala de premios, donde a menores aciertos, menor es el dinero recibido.

La característica principal de la quiniela es un bajo costo de participación y la capacidad de jugar más de una combinación. Esta forma de juego es popular en países de habla hispana.

Tabla 1 Quinielas en hispanoamérica

Quiniela	País	Sector
Progol	México	Público
Progol media semana	México	Público
Ganagol Televisa	México	Privado
Ganagol	Perú	Privado
La Quiniela	España	Público

Adicional a las casas de apuesta, se juega quinielas “locales” entre grupos de conocidos en una región determinada, por ejemplo una quiniela en la que solo participen los locatarios de un mercado popular.

### 1.2.2- Por resultado

En este tipo de apuesta, la casa de apuesta asigna a cada mercado un multiplicador que es conocido como momio<sup>1</sup> u Odd. La forma en que se calcula este multiplicador cambia según el mercado, tomando como ejemplo el mercado de “resultado final”:

$$Odd = \frac{1}{p(w = W_i | (a, b)) * k}$$

Donde:

$w$  Es la elección de la persona

$W = \{Local, Empate, Visitante\}$  Es el universo del mercado en cuestión

$(a, b)$  Es el encuentro en cuestión donde a es el equipo Local y b es el equipo Visitante.

$p(w = W_i | (a, b))$  Es la probabilidad de que un apostador seleccione un resultado dado que jugaran el equipo a y b

<sup>1</sup> En ingles se determina como Odd, que no debe ser confundido con Odds ratio. En apuestas deportivas se refiere al multiplicador de la inversión, mientras que Odd ratio se refiere a la proporción de éxitos y fracasos en términos probabilísticos.



$k$  Es la constante de ganancia de la casa de apuestas. Esta constante siempre es mayor a uno y por motivos de mercado no supera el 1.2.

### 1.2.2.1 Ejemplo

En un partido de la primera división española, se enfrentan los equipos Real Madrid y Getafe. El partido se realizará en la casa del Real Madrid.

Si el 80% de los apostadores están con el Real Madrid, entonces el Odd de una casa de apuestas con una constante de 1.10 será la siguiente:

$$p(w = Local | (Real Madrid, Getafe)) = 0.80$$

$$Odd = \frac{1}{0.80 * 1.1} = 1.13$$

Este momio de 1.13 nos indica que por cada 100 pesos invertidos en caso de ganar la apuesta recibiremos 13 pesos de ganancia.

### 1.2.3 Mercados por resultado

Los mercados más populares son los siguientes:

- Resultado Final: Se elige una de tres opciones
  - Local
  - Empate
  - Visitante.
- Doble oportunidad: Se elige una de tres opciones
  - Local o Empate
  - Empate o Visitante
  - Local o Visitante
- Marcador Correcto: Se tiene que acertar el número de goles de cada equipo.

- Descanso, final: Resultado final tanto al medio tiempo, como al final del partido.
- Línea de Gol: Son varias opciones, en la que se fija un número y se elige si habrá más o menos goles.
- Empate, apuesta no valida: Dos opciones, en caso de empate, se regresa el dinero<sup>2</sup>
  - Local
  - Visitante

---

<sup>2</sup> Se dice que toma momio 1.00

### 1.3- Objetivos

Generar un sistema que permita obtener el mayor rendimiento posible ante el menor riesgo en el mercado de apuestas del futbol.

Existen dos problemas principales para generar esta situación.

1. En la mayoría de los deportes los resultados son binomiales: Gana o Pierde. En el futbol encontramos que el resultado es trinomial: Local, Empate y Visitante.
2. Cada partido de futbol es único, por lo que no se puede calcular una probabilidad real para cada resultado de manera frecuentista.

Si bien no se puede calcular una probabilidad frecuentista, esta se puede obtener comparando equipos que previo al partido compartan características comunes. La característica más sobresaliente es el número de goles anotados. Se debe considerar que el método debe ser numérico y que admita el carácter trinomial del resultado.

Se realizará a partir de lo anterior los siguientes pasos.

1. El estado del equipo se medirá a partir de los goles anotados y recibidos tanto de local como de visitante.
2. Se revisará si es posible reducir las variables mediante ACP<sup>3</sup>.
3. El cálculo de probabilidades será mediante 2 formas, clasificación bayesiana y regresión logística múltiple.
4. Con base al principio de Pareto, se buscará una región en la cual el 20% de los partidos tenga una certidumbre en su resultado del 80% para cada técnica
5. Se comparará el rendimiento de ambas formas, adicionalmente se incluirá en la comparación el Rating System Goals Superiority<sup>4</sup>.

---

<sup>3</sup> Análisis de Componentes Principales del cual se hablará más adelante.

<sup>4</sup> Sistema con finalidad semejante al aquí desarrollado, se detallará más adelante

## 2- Marco Teórico

### 2.1- Metodología

La metodología de esta tesis no es única, es una composición de 2 metodologías ágiles: Lean Startup y Getting Real. La primera está enfocada en la generación de propuestas de valor ya sea en servicios o en productos para empresas recién creadas; mientras la segunda se enfoca en la producción de software a medida.

Ambas enfatizan en la eliminación de todo aquello que no genere un valor real. Considera que el producto perfecto no existe sea este material o lógico, por lo que se realizan iteraciones, partiendo de generar un producto funcional y lanzarlo al mercado, donde los consumidores lo validarán; acto seguido se realizarán ajustes al producto y será nuevamente puesto a prueba, reiniciando de esta manera el ciclo.

La elección de esta metodología se fundamenta en que el sistema, su explotación y su comercialización son inseparables como modelo de negocio, por lo que cualquier cambio en cualquiera de sus partes afecta a todas. La validación ante resultados, y la validación ante clientes son el motor de cambio, y este es constante.

#### 2.1.1- Pasos

1. Imagen General: A partir del resultado esperado, dividir el proceso en subprocesos. De esta manera se asegura que cada proceso sea lo más sencillo posible, otorgando una modularidad tanto en producción, como en pruebas.
2. Abstracción: En cada subproceso, definir lo esencial, aquello por lo que ese subproceso existe
  - a. Si no hay nada esencial, no es un subproceso, por lo que se eliminará. Evita cargas al producto final, eliminando procesos basura.
  - b. Lo esencial será inamovible. Evita de esta manera la introducción de “nuevas ideas” en una etapa temprana de desarrollo que puedan retrasar la finalización del producto.
3. Calidad: Al ser modular, cada subproceso se puede probar de manera independiente y con pautas definidas para el mismo.

4. Comercializar: Vender el producto y obtener seguimiento del cliente
5. Pivote: Discernir sugerencias, buscando aquellas que coincidan con la visión del producto y que sean de interés general, regresando al punto 1.

## 2.2- Aversión al riesgo

El riesgo es definido como: Posibilidad de que ocurra una desgracia o un contratiempo<sup>5</sup>.

Esta definición implica 2 cosas.

1. Acepta que puede ocurrir una pérdida, sin importar el tipo de esta.
2. Que existe en un ámbito de incertidumbre, mencionado como posibilidad y medido como probabilidad.

Es la aversión al riesgo lo que nos lleva a formular modelos, la necesidad ante la ignorancia de obtener información.

En el mercado de las apuestas existen 3 tipos de riesgos.

### 2.2.1- Riesgo de impago.

Independiente del resultado existe un riesgo, el riesgo de que la entidad con la que se apostó no pague. En el entorno, se asume que las apuestas ante la casa oficial de gobierno son seguras, ya que el mismo país es garante del gobierno, sin embargo un país puede llegar a la insolvencia.

Los casinos son considerados seguros en menor medida que el gobierno pero más que ante un particular. Esto se fundamenta en la cantidad de dinero que maneja un casino, asegurando que no llega a la insolvencia, sin embargo el casino puede cerrar sus puertas debido a una orden gubernamental no necesariamente ligada a una razón derivada de las apuestas. El simple hecho de no seguir las normas de protección civil podría causar su clausura.

---

<sup>5</sup> *Diccionario de Uso del Español*. María Moliner

Finalmente el riesgo de particulares se considera alto ya que puede ocurrir o la insolvencia, o el fraude<sup>6</sup>. En este caso, en una apuesta barrial el encargado de llevar la vaquita<sup>7</sup> puede fugarse con el dinero acumulado.

Este riesgo no puede ser calculado de manera directa y recae en el apostador dar o no un voto de confianza a la entidad con la que se apuesta.

#### 2.2.2- Riesgo de modelo

Un modelo es necesario para calcular la posibilidad de un resultado, este cálculo no necesariamente es representado por una probabilidad. Por ejemplo los métodos cualitativos generan confianza con base a un análisis y su resultado no es en base a una ecuación pero ciertamente es un modelo.

El riesgo de modelo radica en la mala interpretación del modelo o en la mala aplicación del modelo. Para este caso el riesgo es medible con base a los resultados del modelo.

#### 2.2.3- Riesgo por el azar

Debido a la naturaleza binomial de la apuesta, donde se gana o se pierde, aunque el modelo debidamente establecido nos genere un 99% de ganar, siempre existe ese 1% de perder.

Este riesgo si bien es medible es difícilmente pronosticable y se debe aceptar como tal.

#### 2.2.4- Conclusiones

Si bien el riesgo de impago no puede ser controlado, el riesgo de modelo y el riesgo del azar sí. Para esto es necesario que el sistema a desarrollar contemple un correcto análisis y una correcta implementación, así como la obligación de generar pruebas que lo validen.

A su vez el sistema cuenta a manera de objetivo con una política de riesgo aceptable, que se ubica en un nivel de confianza del 80%.

---

<sup>6</sup> Esta es la razón por la cual las quinielas de barrio se consideran ilegales. No por afectar a el pago de impuestos, sino, por no existir maneras de proteger al usuario.

<sup>7</sup> Forma coloquial de llamar a la bolsa de la quiniela, es decir el monto acumulado y máximo a ganar.

### 2.3- ACP – Análisis de Componentes Principales

En un experimento del cual no se conoce el modelo, se mide el mayor número de variables posibles. Estas variables pueden estar o no correlacionadas entre sí, por lo que es posible expresar estas variables en términos de otras. En algebra lineal podemos considerar que estas variables son linealmente dependientes y nuestro deseo es expresarlas de manera que sean linealmente independientes.

La técnica para lograr esta independencia lineal es mediante el análisis de componentes principales.

#### 2.3.1- Pasos

1. Calcular la matriz de varianza-covarianza.
2. Calcular el polinomio característico y con él, obtener los eigenvalores.
3. Determinar los eigenvectores.

#### 2.3.2- Problema de dimensionalidad

Un problema al estar trabajando con un alto número de dimensiones es su coste computacional, entonces es tanto deseo como requerimiento el trabajar con el menor número de dimensiones posibles. El ACP se vuelve una herramienta necesaria en este contexto. Los eigenvalores representan la importancia de su vector asociado, por lo que es posible obtener su ponderación en la explicación de los datos.

#### 2.3.3- Ejemplo

Tenemos el siguiente set de datos al cual aplicaremos ACP.

*Tabla 2 Datos para ejemplo de ACP*

x	y
1	1.1
2	1.9
3	3.1
4	3.9
5	5

Primero obtenemos la matriz de varianza-covarianza

$$\Sigma = \begin{bmatrix} 2.5 & 2.45 \\ 2.45 & 2.41 \end{bmatrix}$$

Su polinomio característico

$$\lambda^2 - 4.9\lambda + 0.0225$$

Y sus eigenvectores, eigenvalores y pesos respectivos

<i>eigenvector</i>	<i>eigenvalor</i>	<i>peso (%)</i>
$e_1(0.713, 0.7)$	4.895	99.9
$e_2(-0.7, 0.713)$	0.005	0.1

Al reexpresar el set original en términos de los eigenvectores

Tabla 3 Datos expresado en la transformación lineal ACP

$e_1$	$e_2$
1.483	0.0843
2.756	-0.0453
4.309	0.1103
5.582	-0.0193
7.065	0.065

Graficando ambas representaciones



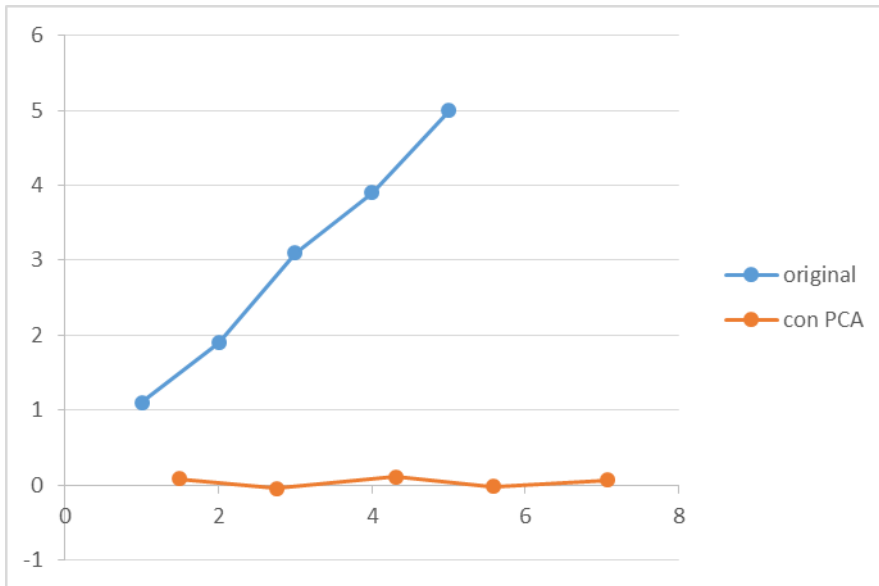


Ilustración 1 Gráfica ACP

#### 2.3.4- Conclusiones

Como se puede ver en la gráfica anterior ACP es útil en la reducción de dimensiones, el primer eigenvector acumula el 99.9 por ciento de la información según los cálculos siendo representado por el eje "X". En el eje "Y" las variaciones son mínimas por lo que se puede despreciar.

#### 2.4- Clasificador Bayesiano

En un problema de clasificación, tenemos una serie de casos a los que tenemos que asignar una clase con base a sus características. Un método para realizar esta tarea es la clasificación bayesiana, en este método de clasificación se le asigna una probabilidad a cada caso, donde la mayor probabilidad es la clase de ese caso.

Partimos del teorema de Bayes para la probabilidad condicional

$$P(w_i|x) = \frac{P(x|w_i)P(w_i)}{P(x)}$$

Donde  $w_i$   $i = 1, 2, \dots, N$  son las clases en las que se desea clasificar.

Sabemos que la probabilidad total debe ser igual a 1, por lo que reescribiendo

$$P(w_i | x) = \frac{P(x|w_i)P(w_i)}{\sum P(x | w_k)P(w_k)}$$

$P(w_i)$  se obtiene por medio de los datos de manera frecuentista, y asumimos que  $P(x|w_i)$  se comporta de manera normal

#### 2.4.1- Ejemplo

Suponemos dos clases, que se comportan de manera normal. La clase 1 tiene una media de 1 y una desviación estándar de 0.4, la clase 2 tiene una media de 2 y una desviación estándar de 0.3. La probabilidad de que algún caso sea de la clase 1 o la clase 2 es la misma.

Se presentan tres casos, con los valores de 0.5, 1.5 y 2, se desea clasificar estos casos por medio de la clasificación bayesiana. Gráficamente el problema se representa así.

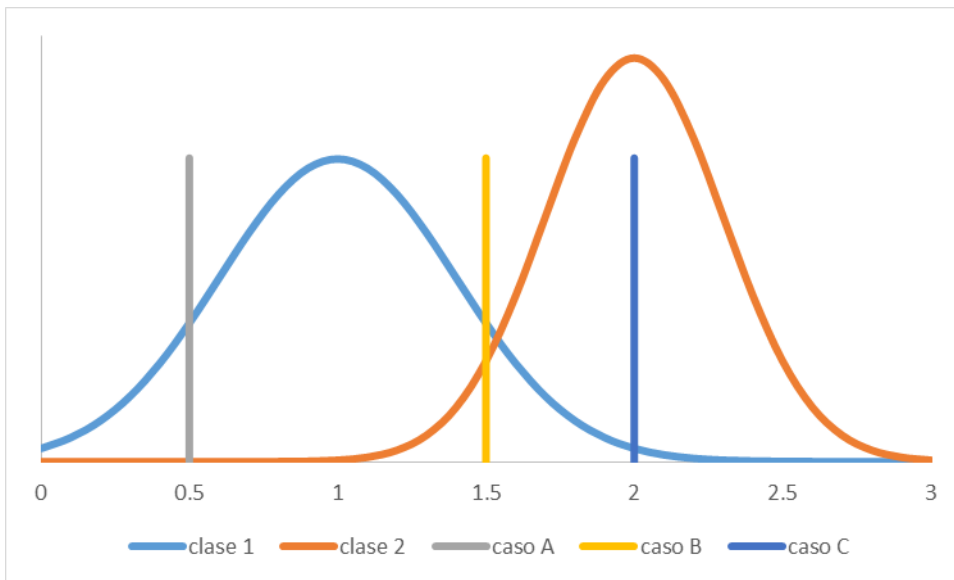


Ilustración 2 Ejemplo Clasificador Bayesiano

Al realizar los cálculos se obtienen los siguientes resultados

Tabla 4 Resultado del Clasificador Bayesiano

Casos	Punto	P(x) Clase 1	P(x) Clase 2	Clase
caso A	0.5	0.999989	0.000011	Clase 1
caso B	1.5	0.579314	0.420686	Clase 1

caso C	2	0.031901	0.968099	Clase 2
--------	---	----------	----------	---------

#### 2.4.2- Conclusiones

Con base en el ejemplo se puede ver que el caso B es un caso difícil de clasificación, sin embargo para fines del tema en esta tesis, es suficiente obtener las probabilidades.

#### 2.5- Regresión Logística Multinomial

La función logística es una función con crecimiento y decrecimiento limitado. Su dominio es de  $(-\infty, \infty)$  y su imagen  $(0,1)$ . Su aplicación es en el ámbito de clasificación, donde nos otorga una probabilidad de que pertenezca o no a una clase

La función logística se expresa de la siguiente manera:

$$P(a_i) = \frac{e^{f(a_i)}}{1 + e^{f(a_i)}}$$

Dado que la función anterior sirve para clasificar entre 2 clases, es necesario generalizarla.

Partiendo de la regla de Bayes igualamos con la función logística

$$P(w_i|x) = \frac{P(x|w_i)P(w_i)}{\sum P(x|w_k)P(w_k)} = \frac{e^{f(a_i)}}{1 + e^{f(a_i)}}$$

Reordenando términos

$$P(w_i|x) = \frac{P(x|w_i)P(w_i)}{\sum_{k \neq i} P(x|w_k)P(w_k) + P(x|w_i)P(w_i)} = \frac{e^{f(a_i)}}{1 + e^{f(a_i)}}$$

Por lo que es posible expresar como

$$P(w_i|x) = \frac{e^{f(a_i)}}{\sum e^{f(a_k)}}$$

La función  $f(a)$  es una combinación lineal de parámetros  $\beta$ , como se ilustra en la siguiente ecuación

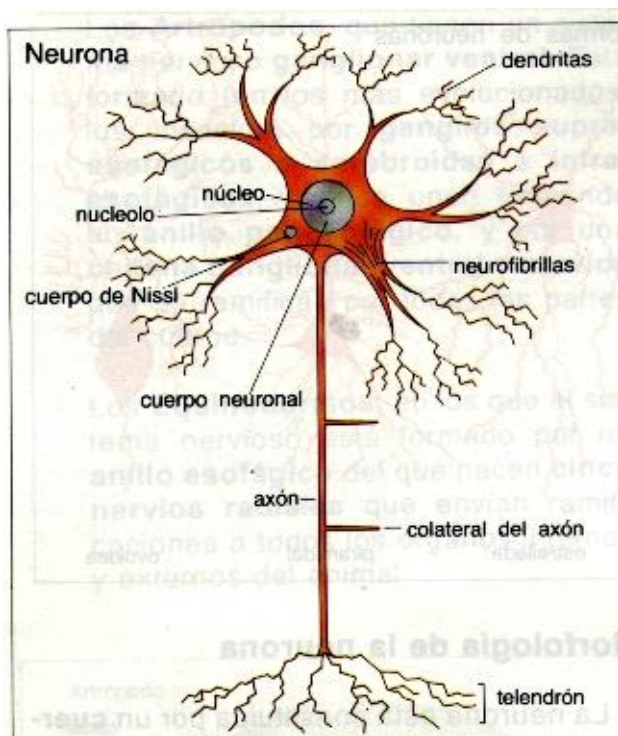
$$f(a_i) = \sum x_j \beta_{ij} + \beta_{i0}$$

Es importante notar que a diferencia de la regresión lineal no es posible obtener estas betas mediante algebra matricial, y que los únicos métodos que funcionan eficientemente son iterativos. El método elegido en esta tesis para el cálculo de las betas es la de red neuronal artificial.

## 2.6- Red Neuronal Artificial

La red neuronal artificial (RNA) se inspira en la forma en que se interconectan las neuronas en los sistemas nerviosos de seres vivos.

Ilustración 3 Neurona<sup>8</sup>



Se distinguen dos tipos de RNA, aquellas de aprendizaje supervisado y de aprendizaje no supervisado. Dentro de las RNA una de las más importantes y usadas es el Perceptrón.

En el Perceptrón, cada neurona tiene las siguientes características:

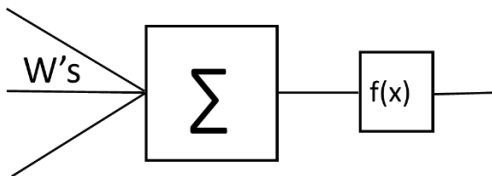
- Conexiones de entrada, con pesos asociados que indican la importancia de cada dato de entrada.

---

<sup>8</sup> Julia Anavel Pintado Cordova CC 3.0

- Sumador, el cual recoge los datos multiplicados por el peso de cada entrada, y otorga una única salida
- Función de transferencia, la cual convierte la salida del sumador en otra, según la función.

Ilustración 4 Perceptron



El aprendizaje consiste en la actualización de los pesos de las conexiones de entrada por medio del algoritmo general de “Backpropagation”, el cual se realiza de manera iterativa ya sea de manera on-line o en batch.

- Actualización On-Line. Por cada set de entrada, se realiza una actualización de peso.
- Actualización Batch. En cada set de entrada se realiza el cálculo de la actualización, pero no se actualiza hasta que hayan pasado todos los sets.

#### 2.6.1- Regresión Logística en una Red Neuronal Artificial Perceptrón

Tenemos el siguiente set de datos

Tabla 5 Datos para ejemplo de RNA

x	clase 1	clase 2
-4.5	1	0
-3.5	1	0
-2.5	1	0
-1.5	1	0
-0.5	0	1
0.5	1	0
1.5	0	1

2.5	0	1
3.5	0	1
4.5	0	1

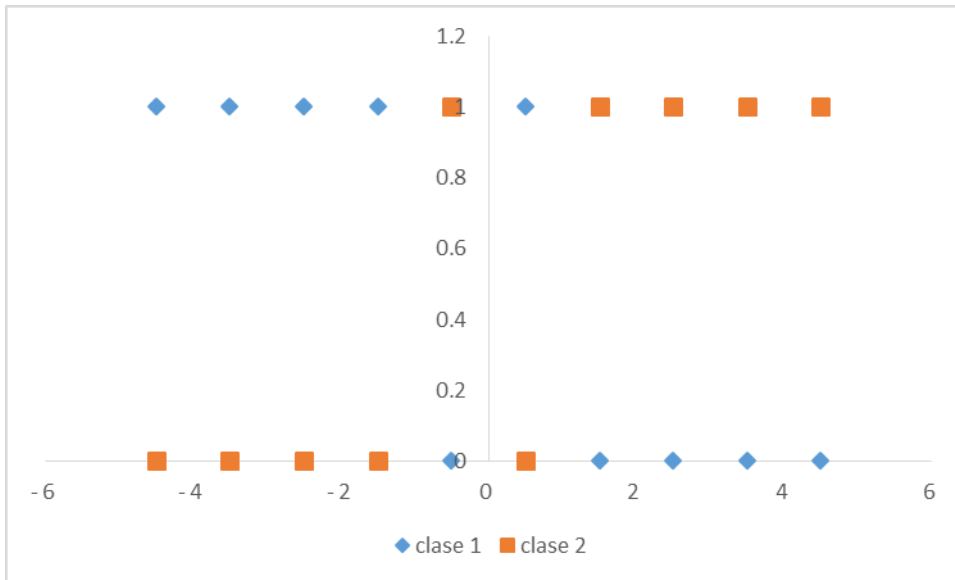


Ilustración 5 Datos para ejemplo RNA

Se puede ver claramente por la imagen que en el valor cero podemos establecer una frontera que dividiera ambas clases, sin embargo la clasificación de los puntos -0.5 y 0.5 estarían incorrectamente clasificados. Realizaremos entonces una clasificación bayesiana que nos dará las probabilidades de que un punto sea de una clase u otra. Para establecer los parámetros usaremos una RNA con la siguiente arquitectura.

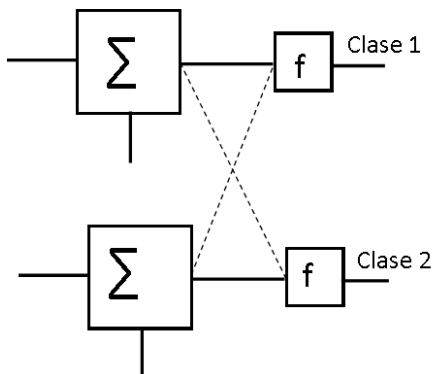


Ilustración 6 Modelo de RNA para Regresión Logística

La función de transferencia para la clase 1 y 2 se desprenden de la regresión logística multinomial, por lo que se expresan de la siguiente manera:

$$f(n_i) = P(w_i|x) = \frac{e^{f(a_i)}}{\sum e^{f(a_k)}}$$

La derivada de esta función está dada en términos de la misma función.

$$\frac{\partial f(n_i)}{\partial x} = f(n_i)(1 - f(n_i))$$

Con lo cual podemos construir las ecuaciones de backpropagation.

$$w(k + 1) = w(k) - \alpha (-2f(n_i)(1 - f(n_i))(t - a))(x)^T$$

$$b(k + 1) = b(k) - \alpha (-2f(n_i)(1 - f(n_i))(t - a))$$

Donde:

$w(k)$  y  $w(k+1)$  es el peso asociado a la conexión de entrada, en el estado actual y el estado siguiente.

$b(k)$  y  $b(k+1)$  es la constante de tendencia en el estado actual y el estado siguiente.

$\alpha$  es la constante de aprendizaje, generalmente un valor pequeña.

$t$  es el valor objetivo.

$a$  es el valor actual de la salida.

Para medir el error de la red, se recurre al error cuadrático medio

$$MSE = \sqrt{\frac{1}{n} \sum_n (t - a)^2}$$

Donde  $n$  la totalidad de los puntos del set de datos.

### 2.6.1.1- Resultados

Tabla 6 Proceso iterativo de RNA

época	Clase 1 ( $\beta_1, \beta_0$ )	clase 2 ( $\beta_1, \beta_0$ )	MSE
0	(0,0)	(0,0)	0.50000
1	(-0.318,0)	(0.318,0)	0.18136
2	(-0.46,0)	(0.46,0)	0.17049
3	(-0.462,0)	(0.462,0)	0.17049

En la tabla aparecen los pesos de las neuronas que son las betas del modelo de la Regresión Logística Multinomial, en cada época de actualización batch. Se terminó el proceso en la tercera iteración debido al nulo cambio en el MSE.

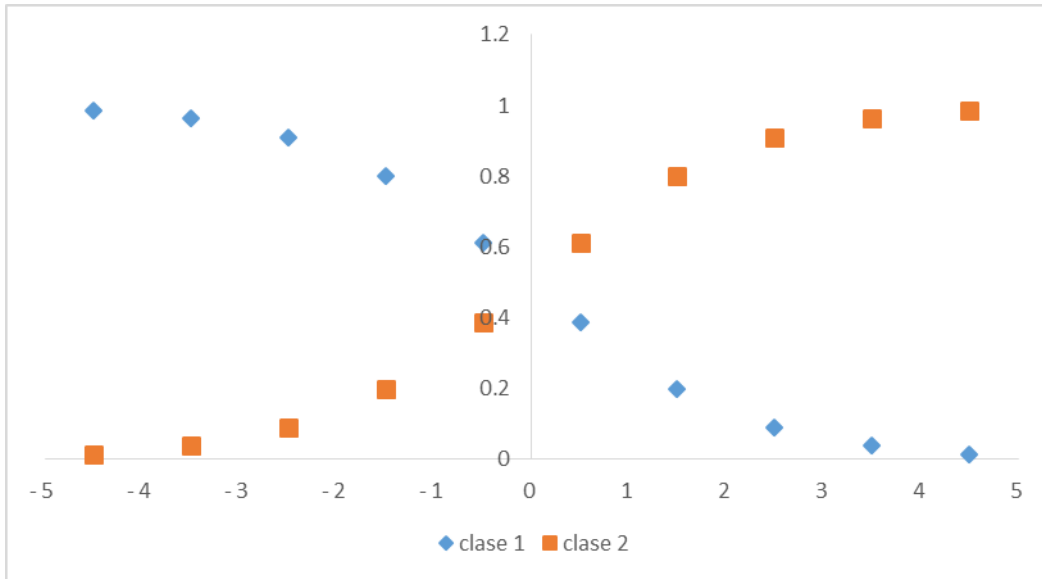
Las probabilidades calculadas para cada punto en el set de datos son las siguientes.

Tabla 7 Resultado del entrenamiento de la RNA

x	clase 1	clase 2
-4.5	0.984626	0.015374
-3.5	0.962139	0.037861
-2.5	0.909774	0.090226
-1.5	0.800037	0.199963
-0.5	0.61353	0.38647
0.5	0.38647	0.61353
1.5	0.199963	0.800037
2.5	0.090226	0.909774
3.5	0.037861	0.962139
4.5	0.015374	0.984626



Ilustración 7 Resultado del entrenamiento de la RNA



### 2.6.2- Conclusiones

El hecho de calcular las probabilidades en lugar de directamente clasificar nos permite tener en cuenta los casos donde la clasificación tiene dudas. En el ejemplo manejado, se puede apreciar que para los valores de  $x$  de  $-0.5$  y  $0.5$  las probabilidades calculadas no permiten ejercer un juicio contundente, contrastado con nuestros puntos originales se observa que son estos puntos donde la clasificación pura tendría un error.

Según la naturaleza del sistema, estos casos podrían ser dirigidos a un experto que realice la clasificación, o bien podrían ser desechados al no poder emitir un juicio acertado.

## 3- Estado del Arte

### 3.1- Métodos cualitativos

La forma más usual y tradicional de realizar pronósticos deportivos es por medio de métodos cualitativos. En estos métodos, el apostador de manera subjetiva predice el resultado del encuentro tomando uno o algunos de los siguientes criterios:

- Posición en la tabla general: Se espera que el equipo que ganó muchos partidos siga ganado, mientras que el que perdió siga perdiendo.
- Resultados de los últimos encuentros: También conocido como racha, está relacionado con el punto anterior, pero la escala de tiempo es menor.
- Valor de la localía: Se sabe que los locales tienen mayor oportunidad de ganar.
- Necesidad de puntuar: Tanto para calificar a otro torneo<sup>9</sup> o para evitar el descenso<sup>10</sup>. Aquellos equipos que no tienen necesidad pueden dejarse perder como estrategia.
- Jugadores en banca: Se tiene en cuenta la existencia de jugadores claves que por lesión no puedan jugar, o por estrategia se decida descansar.
- Estrategia: Cuando un equipo juega en dos torneos, por ejemplo liga y copa, se asume que el entrenador buscará un bien mayor según sus posibilidades en ambos torneos. Podría buscar una derrota o empate en uno de ellos, para asegurar un mejor desempeño en el otro.

Estos métodos presentan la siguiente ventaja:

- El apostador tiene más acercamiento con los equipo, por lo que puede detectar de una manera más rápida el descenso de rendimiento del equipo.

También las siguientes desventajas:

---

<sup>9</sup> En los torneos americanos es usual que los 8 equipos con mayor cantidad de puntos jueguen un segundo torneo llamado liguilla. En los torneos europeos de primera división las primeras posiciones clasifican a la champions league o europe league.

<sup>10</sup> Para mantener la competitividad de las primeras divisiones, se opta a que los peores puntuados al final del torneo descienda a la segunda división, a su vez las primeras posiciones de la segunda división ascienden a la primera.

- El apostador puede presentar un sesgo de empatía hacia algunos equipos. Por ejemplo, en los llamados clásicos, aunque un apostador no tenga favorito, si puede tener desprecio por uno de ellos.
- Maneja variables que no son medibles, ni comparables, por ejemplo: 2 equipos se van a enfrentar, el primero busca clasificar a liguilla y el segundo busca no descender, ¿Cuál de ellos tiene más necesidad?
- El método no es corregible. Si el apostador se equivoca en su pronóstico, ¿Cómo sabe que el error fue por azar o por método?
- Para su mejor rendimiento requieren de una gran experiencia y de un gran compromiso en el seguimiento de partidos.

### 3.2- Rating Systems

Dentro de esta clasificación están los métodos que siguen estos pasos:

1. Asignar comparativamente una puntuación que represente el rendimiento del equipo.
2. A partir de los últimos resultados poder actualizar la métrica anterior.
3. Calcular la probabilidad de obtener cada uno de los resultados con base a la puntuación de los equipos.

Destacan 2 métodos: *Rateform* y *Goals Superiority*.

#### 3.2.1- Rateform

Presentado en 1987, se deriva del sistema ELO para ajedrez, presenta las siguientes reglas:

1. Cada escuadra tiene asignada una puntuación que representa su rendimiento.
2. El promedio de puntuaciones de un torneo es de 1000
3. En cada encuentro, ambos equipos juegan una determinada cantidad de puntos. El equipo local juega el 7% de su puntuación, mientras que el visitante el 5%.
4. El ganador del encuentro se lleva el total de los puntos jugados, en caso de empate se reparten 50%-50%.

Este sistema mide la calidad de los encuentros, es decir si un equipo superior derrota a un equipo inferior, la ganancia puntual es mínima, mientras que de ser al revés, el equipo inferior gana una gran cantidad de puntos.

A partir de la puntuación de ambos equipos se calcula la probabilidad de que un equipo **no pierda** de manera análoga a la función logística

$$P(w = V_i | (a, b)) = \frac{R_i}{\sum R_j}$$

### 3.2.1.1- Ejemplo

Suponemos dos equipos, el equipo A jugará de local y tiene 1300 puntos, el equipo B juega de visitante y tiene 700 puntos. La siguiente tabla contiene las probabilidades, y los posibles escenarios

Tabla 8 Ejemplo de actualización en el modelo Rateform

Descripciones	Local	Visitante
Equipos	A	B
Puntuación	1300	700
Probabilidades	0.65	0.35
Monto jugado	91	35
Puntaje posterior		
Si Gana Equipo A	1335	665
Si Gana Equipo B	1209	791
Si Empatán	1272	728

### 3.2.1.2- Conclusiones

La ventaja de este método radica en la mejor aproximación del estado actual de equipo, la actualización establece un mejor sistema de premios y castigos. En el ejemplo se observa que cuando gana el equipo A, la mejora de puntaje es mínima, contrastando con el intercambio de puntos cuando gana el equipo B. Lo que nos lleva a decir que para el

equipo A ganar no le implica una ventaja real, mientras que ganar para el equipo B representa un aumento en su rendimiento.

La desventaja principal es el carácter binomial del cálculo de probabilidades, cuando el resultado de un partido es de carácter trinomial. Es decir, este método no calcula la probabilidad del empate.

Una segunda desventaja aparece con el ascenso/descenso, no hay criterio que determine que puntaje asignarle al nuevo equipo en el torneo.

### 3.2.2- Goals Superiority

Observar que un partido fue ganado por local, fue empate o ganó el visitante está en función del número de goles que anotó cada equipo. Partiendo de este punto, es razonable pensar que un equipo que ganó con una diferencia de 4 goles tiene mejor rendimiento que otro equipo que también gano, pero con una diferencia de un solo gol.

Este método para hacer su cálculo de probabilidades solo requiere de la diferencia de goles de los últimos 6 partidos del equipo que jugara de local, para ello hace uso de las siguientes fórmulas<sup>11</sup>.

$$P(w = L|a) = 0.0156DG_a + 0.4647$$

$$P(w = E|a) = 0.0003DG_a^2 - 0.0127DG_a + 0.2365$$

$$P(w = V|a) = -0.0003DG_a^2 - 0.0029DG_a + 0.2948$$

#### 3.2.2.1- Conclusiones

Este método tiene por ventaja el tener en cuenta la calidad de la victoria o de la derrota en función de los goles anotados. Establece con fórmulas sencillas el cálculo de probabilidades permitiendo una rápida implementación tanto a nivel manual como en un sistema de cómputo.

---

<sup>11</sup> Fueron establecidas a partir de un conjunto de 14,002 partidos de la primera, segunda y tercera división de Inglaterra, jugados entre 1993 y 2001

Al establecer un modelo lineal y dos cuadráticos, tiene limitantes con respecto al dominio posible, siendo este de  $[-29,26]$ . Anterior o posterior a este presenta probabilidades negativas o probabilidades mayores a uno, que por definición son imposibles.

Finalmente, al tener solo en cuenta al equipo local el poder de este método es limitado en el aspecto de no diferenciar un encuentro con un visitante fuerte de uno débil.

## 4- Análisis

### 4.1- Aprovechamiento

Dada la importancia a nivel mundial y su popularidad a nivel nacional, se toma la decisión en una primera instancia de solo tener 6 países dentro del Método Quetzal. Los torneos involucrados por lo tanto son:

*Tabla 9 Lista de torneos que abarca el Método Quetzal*

<b>Torneo</b>	<b>País</b>
Liga MX	México
Ascenso MX	México
Copa MX	México
Premier League	Inglaterra
Championship	Inglaterra
FA Cup	Inglaterra
League Cup	Inglaterra
Serie A	Italia
Serie B	Italia
Coppa Italia	Italia
Supercoppa	Italia
Primera División	España
Segunda División	España
Copa del Rey	España
Super Copa	España
Bundesliga	Alemania
2. Bundesliga	Alemania
DFB Pokal	Alemania
DFL-Supercup	Alemania
Ligue 1	Francia

Ligue 2	Francia
Coupe de France	Francia
Coupe de la Ligue	Francia
Trophée des champions	Francia

Para estos torneos, se toma el marcador final de cada partido<sup>12</sup>. El lugar de donde se obtienen dichos datos es la página <http://int.soccerway.com/>; de esta página se obtiene tanto el histórico como los datos de la semana actual. La página fue elegida por su amplio repertorio de ligas y por ser soportada por la empresa PERFORM Group PLC, la cual se dedica a la distribución de resultados de varios deportes tanto a casinos, como a equipos de la liga inglesa.

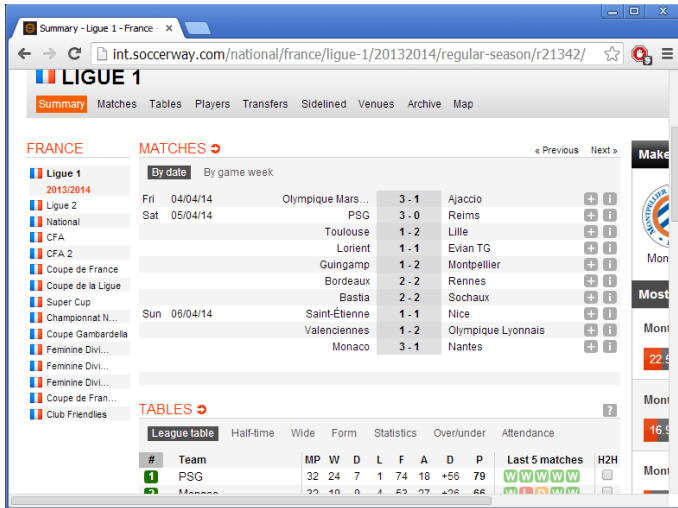
El aprovisionamiento de los datos se realiza de manera semiautomatizada, requiriendo de una gran carga manual. Dado que la empresa se dedica a la distribución de datos y que tiene un plan de pago, es ilegal obtenerlos de manera automática de su página sin una previa autorización, por lo que la búsqueda de resultados y su copiado son meramente manuales.

---

<sup>12</sup> Se planteó la posibilidad de obtener también el número de tiros de esquina, número de tarjetas y minutos adicionales agregados como tiempo de compensación al partido; sin embargo el hecho de que la base se actualiza 2 veces por semana, y que en una de estas actualizaciones ocurren más de 100 encuentros, provocó que se desechara esa idea dada la carga de trabajo que implica.



Ilustración 8 Ejemplo de pantalla donde se obtienen los datos de entrada



Se tomaron como histórico 21,628 partidos realizados entre el 1º de enero de 2010 y el 31 de diciembre de 2014 pertenecientes a los torneos ya mencionados.

#### 4.2- Pre procesamiento

Similar al Goals Superiority, también se considera que el gol es el mejor indicador; pero a diferencia de este, se considera que un empate a cero tiene un valor distinto que un empate a dos goles. También se considera como en Rateform que una victoria de visitante es más importante que una victoria de local.

Siguiendo la tónica de los Rating Systems, primeramente necesitamos una escala de puntos comparativa que represente el estado actual del equipo. A diferencia de estos el Método Quetzal establece 4 parámetros que determinan el estado actual del equipo:

- Promedio de goles anotados en los últimos seis partidos como local.
- Promedio de goles recibidos en los últimos seis partidos como local.
- Promedio de goles anotados en los últimos seis partidos como visitante.
- Promedio de goles recibidos en los últimos seis partidos como visitante.

Requiriendo que los 4 parámetros anteriores sean calculables en ambos equipos, el número de encuentros del histórico que satisfacen dicho requerimiento es de 17,506.

#### 4.2.1- ACP

Números: 4 parámetros por equipo, 2 equipos, un total de 8 dimensiones. Queriendo facilitar la tarea aplicamos ACP con el fin de reducir dimensiones. Dado que 4 dimensiones representan el estado actual del equipo, es sobre estas 4 donde se buscó la reducción.

Tabla 10 Resultados del ACP

lambda	valor	Porcentaje
1	0.441907	38.0780%
2	0.285909	24.6361%
3	0.223723	19.2777%
4	0.208991	18.0082%

Al obtener el valor de las lambdas y el porcentaje de información que contienen se concluye:

1. La lambda predominante no puede representar de manera única el estado del equipo
2. Se podrían usar las primeras tres y desprejar la cuarta al contener más del 80% de información, sin embargo la diferencia entre la lambda 3 y la lambda 4 no es lo suficientemente grande como para tomar este criterio.
3. Dada la poca varianza con respecto al porcentaje promedio, se tomaron las 4 dimensiones como estado del equipo, y se asume que ACP no nos permitió tomar un criterio para reducir dimensiones.

#### 4.3- Modelo matemático

##### 4.3.1- Clasificador Bayesiano

Partiendo de la ecuación ya mencionada en el marco teórico

$$P(w_i | x) = \frac{P(x|w_i)P(w_i)}{\sum P(x | w_k)P(w_k)}$$

Anteriormente se presentó que  $P(w_i)$  se obtiene de manera frecuentista y que  $P(x|w_i)$  se comporta de manera normal.

Teniendo en cuenta que contamos con ocho parámetros iniciales, debemos hacer uso de la normal multivariante.

$$P(x|w_i) = \frac{1}{(2\pi)^{d/2} |\Sigma_i|^{1/2}} e^{(-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1} (x-\mu_i))}$$

Donde  $d$  es el número de dimensiones, en este caso 8;  $\Sigma_i$  es la matriz de varianza-covarianza y  $\mu_i$  es el vector de promedios.

Para poder usar la normal multivariante es necesario:

1. Que los datos de cada dimensión se comporten de manera normal
2. Que la correlación entre dimensiones sea baja, más no necesariamente igual a cero.

La característica 2 puede ser deducida del ACP. La naturaleza del ACP busca reducir las correlaciones y de esta forma reducir dimensiones, al no poder reducir estas últimas se puede asumir que la correlación entre ellas es baja.

Para la característica 1 es necesario realizar un test, en este caso se usó Anderson-Darling

#### 4.3.1.1- Anderson-Darling

Según sus autores es mejor realizarlo con no más de 40 muestras, ordenadas de menor a mayor y previamente estandarizadas. Dado que tenemos 17,506 vectores, se realizarán 10 muestreos aleatorios de 40 muestras sin remplazo cada uno, a los que se les aplicara el test por cada dimensión.

La ecuación del test es la siguiente.

$$W_n^2 = -n - \frac{1}{n} \sum_{j=1}^n (2j-1) [\log(u_j) + \log(1-u_{n-j+1})]$$

$$H_0: W_n^2_{0.05} < 2.492$$

Donde  $u_j$  es la función acumulativa de la distribución normal valuada en ese punto.

Se obtienen estos resultados

Tabla 11 Resultados del test Anderson-Darlington

muestreo	dim 1	dim 2	dim 3	dim 4	dim 5	dim 6	dim 7	dim 8
1	1.534	1.536	1.637	1.053	0.626	1.908	1.441	1.529
2	0.352	0.591	1.123	2.524	0.834	2.206	1.072	1.708
3	0.590	3.198	1.057	1.484	1.056	2.781	0.987	0.612
4	0.617	1.048	2.342	0.796	0.798	1.019	0.957	2.129
5	0.964	0.616	2.871	1.058	0.439	2.030	4.029	1.086
6	0.525	0.888	1.038	0.944	1.022	0.733	0.797	0.601
7	1.326	0.850	0.797	2.783	0.542	2.857	0.529	1.124
8	0.452	0.582	0.407	1.065	2.062	1.189	0.790	0.753
9	0.339	1.526	0.982	1.546	0.577	0.509	1.028	0.576
10	0.774	0.785	2.333	1.215	1.432	0.642	1.112	0.810
<b>Promedio</b>	<b>0.747</b>	<b>1.162</b>	<b>1.459</b>	<b>1.447</b>	<b>0.939</b>	<b>1.587</b>	<b>1.274</b>	<b>1.093</b>

En rojo aparecen los test que no pasaron la prueba, ¿Por qué sucede esto? La respuesta está en la naturaleza de los datos. Por ejemplo, el promedio de promedios para una dimensión es de 1.5, es posible que exista un equipo que en un momento dado tenga un promedio mayor a 3 goles. Ya que la distribución normal es simétrica se esperaría que también un equipo tuviera un promedio menor a 0 goles, y ahí es donde radica el problema, no hay goles negativos.

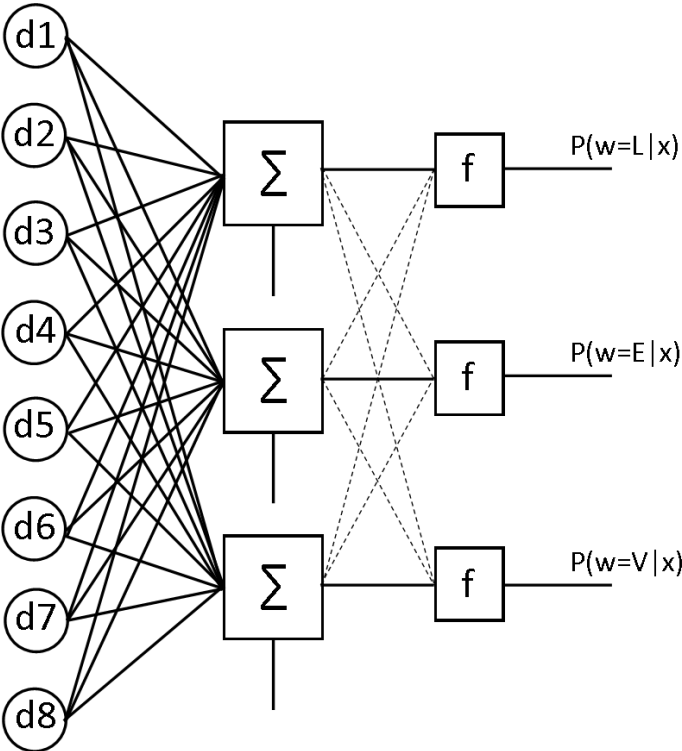
A pesar de que pareciera que el hecho de no tener una de las colas de la normal afectara los cálculos, se observa en la tabla que solo 7 casos de 80 pruebas presentan este

problema. Adicionalmente es necesario aclarar que la cola alargada solo representa entre el 1% y 3% de los datos según la dimensión<sup>13</sup>.

4.3.2- Regresión Logística Trinomial- RNA

Al quedar planteado en el marco teórico la necesidad de implementar una RNA que permita obtener los coeficientes de la Regresión Logística se determinó la siguiente arquitectura:

Ilustración 9 Modelo de la RNA para el Método Quetzal



Se eligió al actualización Batch pues esta presenta un descenso al óptimo más suave, con un  $\alpha$  de aprendizaje igual a 1. El proceso iterativo dura 1000 épocas.

Para alcanzar más rápidamente un óptimo es necesario que las entradas sean tanto positivas como negativas, por lo que previo a entrar a la red neuronal son estandarizadas con respecto a su media y varianza muestral.

<sup>13</sup> Para un mayor detalle del conjunto muestral ir al anexo A

## 5- Diseño e implementación

### 5.1- Filosofía

La implementación está basada en las metodologías ágiles de Lean Startup y Getting Real, por lo que se siguieron los siguientes lineamientos:

1. Cada parte del análisis constituye una entidad separada, esta parte representa el proceso mínimo necesario en el que se puede dividir el proyecto.
2. Cada entidad es codificada de manera independiente, con sus propios parámetros de calidad.
3. Se elige como lenguaje de programación al lenguaje C, al ser el lenguaje sobre el que tengo mayor dominio.

### 5.2- Definición Funcional: Aprovisionamiento

1. Se entra a la página web y se selecciona el torneo del que se requieren los resultados.
2. Aprovechando el formato de tabla se copian los datos en una hoja de cálculo.
3. Se realiza una limpieza de los marcadores mediante 2 expresiones formales:

$$[a - z] * * \rightarrow$$
$$- \rightarrow \backslash t$$

La primera expresión elimina cualquier letra y/o espacio, mientras la segunda transforma el guion separador en una tabulación.

### 5.3- Definición Funcional: Preprocesamiento

1. Se listan los partidos en orden cronológico
2. Por cada encuentro de la lista, se buscará en los partidos anteriores aquellos en los que coincida ya sea el local o el visitante.
  - a. La fecha máxima en la que se pueda buscar coincidentes será un año atrás de la fecha del partido en cuestión.
  - b. El número máximo de coincidentes es:

- i. 6 partidos donde el equipo local del encuentro en cuestión coincida en el equipo local.
  - ii. 6 partidos donde el equipo local del encuentro en cuestión coincida en el equipo visitante.
  - iii. 6 partidos donde el equipo visitante del encuentro en cuestión coincida en el equipo local.
  - iv. 6 partidos donde el equipo visitante del encuentro en cuestión coincida en el equipo visitante.
3. Por cada encuentro coincidente:
- a. Se toma el número de goles del encuentro
    - i. Si el equipo local del encuentro en cuestión coincide en el equipo local, el número de goles del equipo local se tomarán en cuenta para la métrica **Promedio de goles del local anotados como local** mientras que los goles del equipo visitante serán para la métrica **Promedio de goles del local recibidos como local**.
    - ii. Si el equipo local del encuentro en cuestión coincide en el equipo visitante, el número de goles del equipo local se tomarán en cuenta para la métrica **Promedio de goles del local recibidos como visitante** mientras que los goles del equipo visitante serán para la métrica **Promedio de goles del local anotados como visitante**.
    - iii. Si el equipo visitante del encuentro en cuestión coincide en el equipo local, el número de goles del equipo local se tomarán en cuenta para la métrica **Promedio de goles del visitante anotados como local** mientras que los goles del equipo visitante serán para la métrica **Promedio de goles del visitante recibidos como local**.
    - iv. Si el equipo visitante del encuentro en cuestión coincide en el equipo visitante, el número de goles del equipo local se tomarán en cuenta para la métrica **Promedio de goles del visitante recibidos como visitante** mientras que los goles del equipo visitante serán

para la métrica **Promedio de goles del visitante anotados como visitante.**

4. En caso de que algún encuentro no cumpla con los 24 partidos coincidentes, se considera que las métricas están incompletas y el encuentro no será tomado en cuenta.

#### 5.4- Definición Funcional: Clasificador Bayesiano

1. Se lista cada encuentro en el que las métricas estén completas, junto con la clasificación correcta: Local, Empate o Visitante. Cada métrica representa una dimensión.
2. Para cada clase:
  - a. Para cada dimensión se obtendrá el promedio
  - b. Se calcula la matriz de varianza-covarianza, el determinante y la inversa de esta matriz.
  - c. Se calcula una constante asociada a la clase, esta constante se obtiene multiplicando:
    - i. El número de encuentros de esa clase entre el número de encuentros totales.
    - ii. El inverso de dos veces  $\pi$  elevado a la cuarta potencia.
    - iii. El inverso de la raíz cuadrada del determinante de la matriz de varianza-covarianza.
3. Para calcular la probabilidad de cada clase para cualquier encuentro:
  - a. Para cada clase
    - i. A cada métrica se le restará su promedio, que será multiplicado transponiéndolo por la matriz de varianza-covarianza inversa y nuevamente por la métrica menos el promedio.
    - ii. Al resultado anterior se multiplicará por -0.5 y se utilizará como argumento en la función exponencial.
    - iii. A este nuevo resultado se le multiplicará la constante asociada a la clase.



- b. A cada resultado del paso anterior se dividirá entre la suma del total de resultados de cada clase. Al realizar esa división se asegura que la suma de probabilidades de las clases es igual a 1.

### 5.5- Definición Funcional: Regresión Logística

Es importante para el mejor entendimiento de la definición funcional tener en cuenta la imagen que representa la red neuronal.

1. Se establece una entidad conocida como RNA, que será representada por cuatro matrices, matriz de pesos y matriz de tendencias que son propias de la RNA; y las matrices de actualización para pesos y tendencias, usadas en el proceso de aprendizaje.
2. Todas las matrices de la RNA son iniciadas en ceros.
3. Se lista cada encuentro en el que las métricas estén completas las cuales serán presentadas estandarizadas, junto con la clasificación correcta: Local, Empate o Visitante. La clasificación del encuentro tiene que ser expresada con variables *dummy*; local se expresa como 1 0 0, empate como 0 1 0 y visitante como 0 0 1. Cada métrica representa una dimensión. A cada encuentro adicionalmente se incluyen tres matrices: matriz de salida previo a la función, matriz de salida de la función y matriz de sensibilidad
4. Se realiza una iteración de 1000 épocas, en cada de ellas se realizan los siguientes pasos:
  - a. Por cada encuentro de la lista:
    - i. La matriz de salida previo a la función es igual a la multiplicación de las métricas de entrada por la matriz de pesos, a esto se le suma la matriz de tendencias.
    - ii. Cada componente de la matriz de salida de la función es igual al componente de la matriz de salida previo a la función dividido entre la suma de los componentes de la matriz de salida de la función.
    - iii. Cada componente de la matriz de sensibilidad es igual a la multiplicación de la componente de la matriz de salida de la

función, uno menos componente de la matriz de salida de la función, la componente de la clasificación menos la componente de la matriz de salida de la función y el escalar -2.

- b. La matriz de actualización de pesos es igual al promedio de la multiplicación de los parámetros y la matriz de sensibilidad de cada encuentro de la lista.
- c. La matriz de actualización de tendencia es igual al promedio de la matriz de sensibilidad.
- d. Finalizando cada época se realiza la actualización de las matrices a modo de aprendizaje, esto se logra sumando a las matrices de la RNA las matrices de actualización.

## 6- Evaluación

Dada la naturaleza trinomial de la clasificación se requiere de una métrica especial para medir el rendimiento. La primera métrica de evaluación se basa en el hecho de que las tres probabilidades no son iguales, por lo que es posible ordenarlas de menor a mayor, una vez ordenadas determinar en cuál de ellas es el acierto. Si el acierto está en la de mayor probabilidad, implica que la clasificación fue correcta, mientras que si está en la de menor probabilidad nos dice que la clasificación fue incorrecta. Si está en el valor medio se considera que esta variación fue azarosa.

Una segunda métrica de evaluación proviene del objetivo de la tesis, encontrar una zona, en la que los partidos tengan un porcentaje del 80% o más de ocurrir. Para este caso se recurrirá a un mercado de apuestas doble, por lo que los encuentros candidatos deberán tener en su menor probabilidad una probabilidad no mayor al 20%. Se medirá en cuantos aciertos y cuantos errores hubo para este subconjunto. Adicionalmente se contrastará con simulaciones Montecarlo para probar que el rendimiento no depende del azar.

Ya que los métodos de Clasificación bayesiana y Regresión Logística dependen de un conjunto de entrenamiento para estimar sus parámetros, y que se busca un óptimo en esta tarea, es necesario crear más de un conjunto de entrenamiento y evaluación.

### 6.1- Diseño de pruebas

1. Un programa en C realiza muestreo aleatorio sin remplazo, con un número de 8753 muestras, que corresponde a la mitad del set de datos disponible.
2. Este programa corre en dos ocasiones, la primera con una semilla 0 y la segunda con una semilla 1.
3. Periodo de pruebas:
  - a. En el primer periodo, se toma a los elegidos por el muestreo con semilla 0 como set de entrenamiento y a los no elegidos como set de evaluación.
  - b. En el segundo periodo, se toma a los no elegidos por el muestreo con semilla 0 como set de entrenamiento y a los elegidos como set de evaluación.

- c. En el tercer periodo, se toma a los elegidos por el muestreo con semilla 1 como set de entrenamiento y a los no elegidos como set de evaluación.
    - d. En el cuarto periodo, se toma a los no elegidos por el muestreo con semilla 1 como set de entrenamiento y a los elegidos como set de evaluación<sup>14</sup>.
  4. Cada periodo servirá para entrenar, evaluar y comparar a los métodos de clasificación bayesiana y regresión logística, así como la comparación con el rating system Goal Superiority.
  5. La simulación de Montecarlo seguirá esta configuración.
    - a. Ya que el porcentaje esperado de encuentros con una probabilidad mayor al 80% es del 20%, un primer número aleatorio entre el cero y el uno será generado, si este es menor o igual a 0.2, ese encuentro será tomado como candidato.
    - b. A partir de los datos estadísticos del set de datos<sup>15</sup> tendremos los siguientes límites para un nuevo número aleatorio. Si es menor o igual a 0.375 se tomará como LE, si es mayor a 0.375 y menor o igual a 0.75 será un LV y si es mayor a 0.75 será un EV. Este segundo número aleatorio solo es generado en los encuentros que hayan sido marcados como candidatos en el paso anterior.
    - c. Se correrán 10 simulaciones por cada periodo de pruebas con la finalidad de obtener la segunda métrica de evaluación. Estas diez simulaciones tendrán por semilla del 0 al 9.

## 6.2- Resultados

Con respecto a la capacidad de clasificar correctamente se obtuvieron los siguientes resultados.

---

<sup>14</sup> Al conjunto del primer y segundo periodo, así como al del tercero y cuarto, se le llama validación cruzada.

<sup>15</sup> Ver anexo A

Tabla 12 Comparativa en la clasificación

	<b>Clasificación</b>								
	<b>Goal Superiority</b>			<b>Clasificación Bayesiana</b>			<b>Regresión Logística</b>		
<b>periodo</b>	<b>Correcto</b>	<b>Azar</b>	<b>Incorrecto</b>	<b>Correcto</b>	<b>Azar</b>	<b>Incorrecto</b>	<b>Correcto</b>	<b>Azar</b>	<b>Incorrecto</b>
<b>1</b>	45.493%	26.402%	28.105%	47.115%	28.390%	24.494%	45.721%	27.602%	26.677%
<b>2</b>	45.185%	27.019%	27.796%	46.361%	27.488%	26.151%	45.550%	27.819%	26.631%
<b>promedio 1,2</b>	45.339%	26.711%	27.950%	46.738%	27.939%	25.323%	45.636%	27.710%	26.654%
<b>3</b>	45.333%	26.745%	27.922%	46.521%	28.813%	24.666%	45.744%	27.613%	26.642%
<b>4</b>	45.344%	26.677%	27.979%	46.841%	28.173%	24.986%	45.539%	27.705%	26.757%
<b>promedio 3,4</b>	45.339%	26.711%	27.950%	46.681%	28.493%	24.826%	45.641%	27.659%	26.699%
<b>promedio General</b>	45.339%	26.711%	27.950%	46.710%	28.216%	25.074%	45.639%	27.685%	26.677%

De la tabla anterior llegamos a las siguientes conclusiones:

1. Como cualquier juego de azar, en el fútbol no es posible predecir un resultado. En el sentido de clasificación ningún modelo se puede considerar satisfactorio.
2. A pesar del punto anterior, la clasificación realizada por el Clasificador Bayesiano presenta un mejor rendimiento comparado con Goal Superiority y con la Regresión Logística. El Clasificador Bayesiano presenta un mayor porcentaje de correctos y un menor porcentaje de incorrectos a lo largo de los 4 periodos de pruebas y en sus promedios.

Obteniendo el subconjunto que presente un mínimo 80% de probabilidad se obtienen los siguientes resultados.

Tabla 13 Comparativa en la aversión al riesgo

<b>Aversión al riesgo</b>								
<b>periodo</b>	<b>Goal Superiority</b>		<b>Clasificación Bayesiana</b>		<b>Regresión Logística</b>		<b>Montecarlo Promedio</b>	
	<b>Aceptados</b>	<b>Rendimiento</b>	<b>Aceptados</b>	<b>Rendimiento</b>	<b>Aceptados</b>	<b>Rendimiento</b>	<b>Aceptados</b>	<b>Rendimiento</b>
<b>1</b>	22.095%	75.957%	26.928%	81.120%	4.524%	84.596%	19.925%	68.155%
<b>2</b>	21.490%	76.608%	29.875%	78.967%	2.673%	90.598%	19.925%	68.089%
<b>promedio 1,2</b>	21.793%	76.282%	28.402%	80.044%	3.599%	87.597%	19.925%	68.122%
<b>3</b>	21.878%	76.501%	27.191%	80.420%	2.433%	86.854%	19.925%	68.070%
<b>4</b>	21.707%	76.053%	30.709%	79.464%	4.935%	85.648%	19.925%	68.376%
<b>promedio 3,4</b>	21.793%	76.277%	28.950%	79.942%	3.684%	86.251%	19.925%	68.223%
<b>promedio General</b>	21.793%	76.280%	28.676%	79.993%	3.642%	86.924%	19.925%	68.172%

En esta métrica relacionada con la aversión al riesgo se puede observar la superioridad de la clasificación Bayesiana; se esperaba tener un 20% de encuentros aceptados y se obtuvo un promedio de 28.676% sin que se viera afectado el rendimiento. Se acepta entonces la clasificación Bayesiana como motor de cálculo del Método Quetzal.

## 7- Método Quetzal en Internet

En este capítulo se describe la experiencia del proceso de venta derivado del Método Quetzal, para evidencias se debe consultar el anexo B.

### 7.1- Servicio

La naturaleza de la metodología usada permite y obliga a exponer el producto ante mercado. Desde el mes de Enero del año 2014 aparece listado en el portal MercadoLibre.com.mx el servicio de momios Método Quetzal.

En un primer acercamiento decidí tener como público objetivo perfiles similares al mío:

1. Quinielistas y/o apostadores de casino.
2. Que tengan una tendencia de aversión al riesgo y que vean en las apuestas un tipo de sistema de inversión.
3. Que tengan acceso a internet.

Conociendo el tiempo que se puede llegar a dedicar al análisis de partidos el servicio de momios Método Quetzal consiste de:

1. Dos entregas semanales durante 30 días (1 mes), cada una listando los partidos de la semana a partir de la fecha de entrega. La primera entrega se realiza los días lunes y la segunda los días jueves. Estas fechas de entrega son elegida con base a 2 eventos principales en las quinielas, la quiniela de fin de semana y la de media semana.
2. Cada uno de los partidos presenta 2 secciones
  - a. Una primer sección de estadísticas, que incluye el promedio de goles a favor y en contra, promedio de partidos ganados, empatados y perdidos de los últimos 6 partidos para ambos equipos.
  - b. Una segunda sección que incluye la estimación de probabilidades generadas por el motor de Cálculo, marcando aquellas que tengan un porcentaje menor al 20% en su porcentaje menor, estas son las apuestas recomendadas para el mercado de doble oportunidad.

El costo nominal del servicio corresponde a \$80.00 MXN, el precio fue considerando los siguientes puntos:

- El costo debe representar un estatus de calidad, un precio bajo conllevaría a que el servicio no es de calidad
- Los apostadores de quiniela juegan o el fin de semana o la media semana, raramente las 2. Teniendo en cuenta que el servicio dura 30 días y las quinielas son cada 7, se espera que se jueguen en ese periodo 4 quinielas. Ya que cada quiniela tiene un costo de \$15.00 MXN, el apostador promedio invierte \$60.00 MXN en un mes, por lo que el costo del servicio debe ser similar. En caso de ser mayor provocará que sea más caro el costo de información que el costo de oportunidad, con lo que llevaría al apostador a buscar una opción más barata.

## 7.2- Promoción

La estrategia de promoción se basó en la generación de contenidos declinando la opción de compra de publicidad.

- Se realizó la compra de 2 dominios: metodoquetzal.com y metodoquetzal.net. El primer domino es un blog, mientras que el segundo es una página de aterrizaje.
- Se implementó un método de pago electrónico en la página de aterrizaje por medio de la plataforma de MercadoPago de MercadoLibre.
- Cada lunes se sube un video a Youtube a mi canal personal, donde se realiza el análisis de los partidos de la Liga MX, en este canal al momento se tiene 113 suscriptores.
- Por medio de una lista de correos se distribuye el día miércoles una versión reducida del servicio de momios Método Quetzal, esta incluye únicamente los partidos de fin de semana de México. Al momento se tiene 97 suscriptores en esta lista.
- Se tiene participación activa en varios grupos de quinielistas y apostadores dentro de la red social Facebook. También por este medio se distribuye la versión reducida, sin embargo no es posible contar el número de impactos.



### 7.3- Futuro

A partir de necesidades de mercado, se está realizando los siguientes desarrollos para en un futuro implementar las siguientes capacidades:

- Aumento de países para los que sea válido el modelo del Método Quetzal, se busca en una segunda instancia agregar Países Bajos, Portugal y Estados Unidos de América; en una tercera Grecia, Argentina y Brasil.
- El conteo de goles a lo largo de las temporadas se plantea como una mixtura de distribuciones de Poisson. Ya que la duración de los partidos es la misma, se busca obtener una aproximación de la lambda de Poisson que permita obtener para cada encuentro probabilidades para el mercado **línea de gol**.
- El Método Quetzal en su estado actual corresponde a una visión de racha, se espera poder implementar una modificación al Rating System Rateform que complemente con una visión de trayectoria.
- Ya que los momios de pago son distintos en partidos con probabilidad semejante, se puede considerar que un momio de pago se comporta como un tipo de activo financiero, por lo que es posible construir portafolios de inversión. Un reto futuro es encontrar un método que permita construir, comparar y elegir dichos portafolios.
- La implementación de una App para IOS y Windows Phone que facilite el método de suscripción, permita una mejor presentación de las estadísticas y sugerencias, y establezca un canal más directo con los consumidores.

## 8- Conclusiones

El Método Quetzal presenta un rendimiento ligeramente superior al ya existente Goal Superiority en la forma de clasificar el resultado final de los partidos. Su fortaleza en la política de aversión al riesgo, la cual permite aumentar el rendimiento en la clasificación a costa de ignorar partidos, siguiendo el principio de Pareto.

Dada las pruebas estadísticas y posterior a ellas las pruebas ante apuestas reales se concluye que los objetivos fueron cubiertos satisfactoriamente.

Concordando con lo expresado en la sección 7.3 aún quedan cosas por intentar por lo que la investigación continuará con una visión del futuro alentadora.

## Anexo A – Estadísticas del set de datos

Las siguientes tablas y gráficas contienen los valores del set de datos considerando el estado actual de ambos equipos. Ya que el set de datos es de 17,506, hay entonces 35,012 estados de los equipos, la mitad del local y la mitad del visitante.

Se puede observar en las gráficas el gran parecido que hay con la distribución normal, se debe notar que si bien las gráficas no son simétricas, el valor de la cola larga no representa más del 3%.

Se pudo eliminar aquellos registros que sobrepasan la parte simétrica con el fin de que los datos ajustaran mejor a la gaussiana, sin embargo se decide no hacerlo para conservar un set de datos real, tal y como aparecen en el fenómeno estudiado, en este caso los resultados del fútbol.

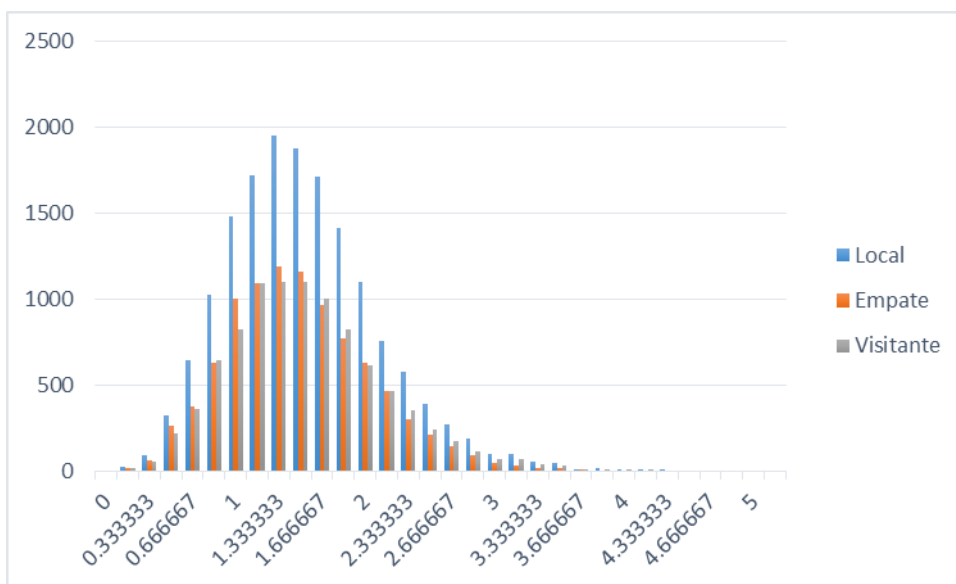
### Promedio de Goles a Favor de Local

Tabla 14 Goles a Favor de Local

Goles a Favor de Local	Local	Empate	Visitante
0	4	2	2
0.166667	23	15	18
0.333333	92	64	57
0.5	325	267	219
0.666667	644	380	365
0.833333	1029	633	644
1	1480	1004	826
1.166667	1718	1091	1096
1.333333	1952	1191	1102
1.5	1874	1161	1100
1.666667	1709	968	1007

1.833333	1417	773	823
2	1100	634	618
2.166667	759	466	468
2.333333	579	304	358
2.5	395	215	239
2.666667	274	148	178
2.833333	194	93	115
3	101	49	68
3.166667	101	33	70
3.333333	57	19	42
3.5	45	16	32
3.666667	10	10	11
3.833333	20	4	13
4	11	6	8
4.166667	10	4	9
4.333333	9	1	5
4.5	7	0	1
4.666667	4	3	5
4.833333	2	0	3
5	2	1	5
5.166667	1	1	1

Ilustración 10 Goles a Favor de Local



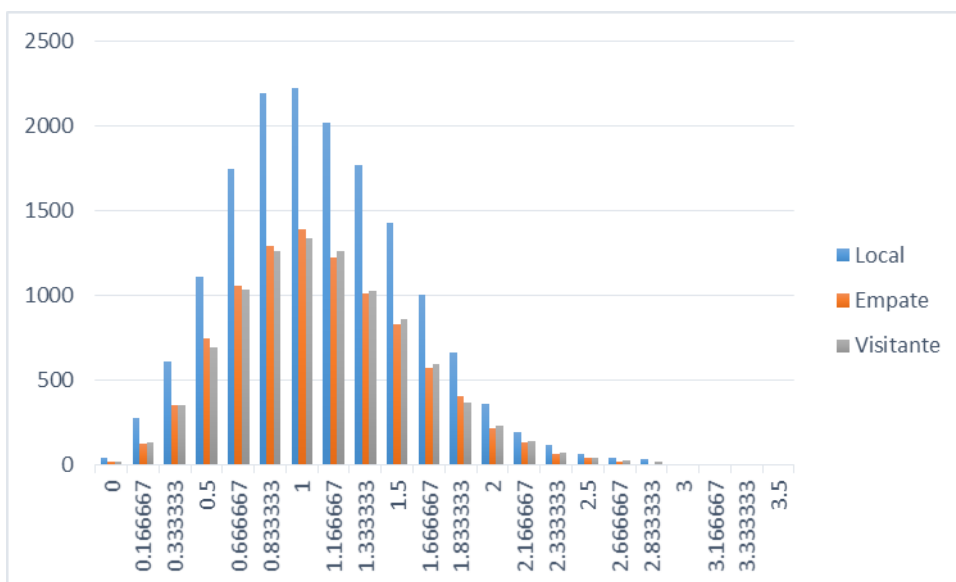
## Promedio de Goles en Contra de Local

Tabla 15 Goles en Contra de Local

Goles en Contra de Local	Local	Empate	Visitante
0	46	19	25
0.166667	275	130	134
0.333333	615	358	357
0.5	1114	746	693
0.666667	1746	1059	1032
0.833333	2195	1295	1263
1	2225	1389	1335
1.166667	2022	1223	1261
1.333333	1772	1015	1024
1.5	1428	833	865
1.666667	1004	575	597
1.833333	665	410	371

2	359	215	231
2.166667	198	137	146
2.333333	117	67	76
2.5	69	46	43
2.666667	45	21	27
2.833333	38	8	19
3	7	3	3
3.166667	6	3	4
3.333333	1	3	0
3.5	1	1	2

Ilustración 11 Goles en Contra de Local



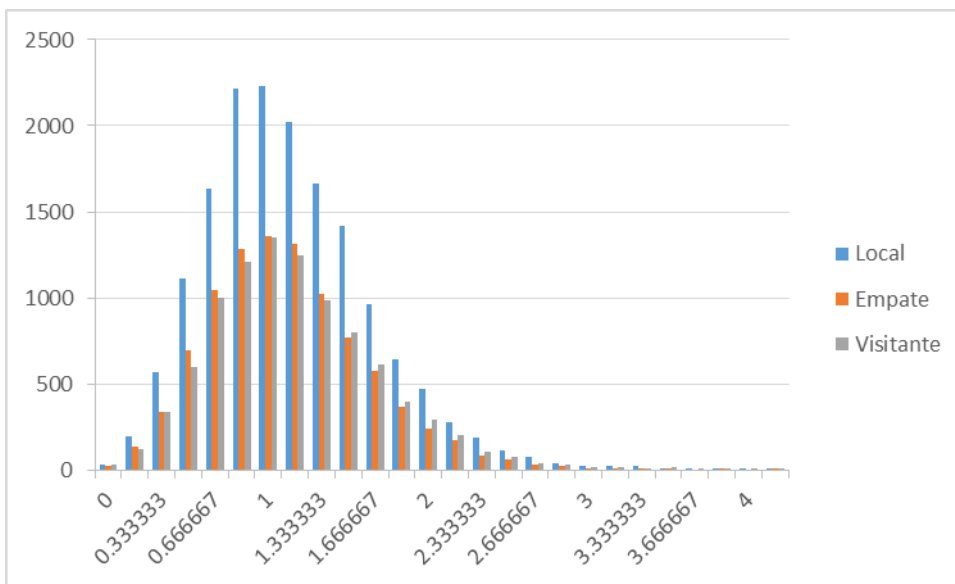
### Promedio de Goles a Favor de Visitante

Tabla 16 Goles a Favor de Visitante

Goles a Favor de Visitante	Local	Empate	Visitante
----------------------------	-------	--------	-----------

0	34	21	30
0.166667	193	137	118
0.333333	568	339	338
0.5	1109	692	601
0.666667	1637	1043	1003
0.833333	2217	1283	1208
1	2228	1355	1354
1.166667	2019	1316	1247
1.333333	1667	1023	984
1.5	1417	768	796
1.666667	964	576	616
1.833333	640	369	397
2	473	240	292
2.166667	275	170	200
2.333333	190	82	103
2.5	112	60	76
2.666667	72	28	40
2.833333	36	21	32
3	23	7	18
3.166667	23	9	19
3.333333	21	6	8
3.5	9	4	12
3.666667	8	0	6
3.833333	4	6	7
4	7	0	2
4.166667	2	1	1

Ilustración 12 Goles a Favor de Visitante



### Promedio de Goles en Contra de Visitante

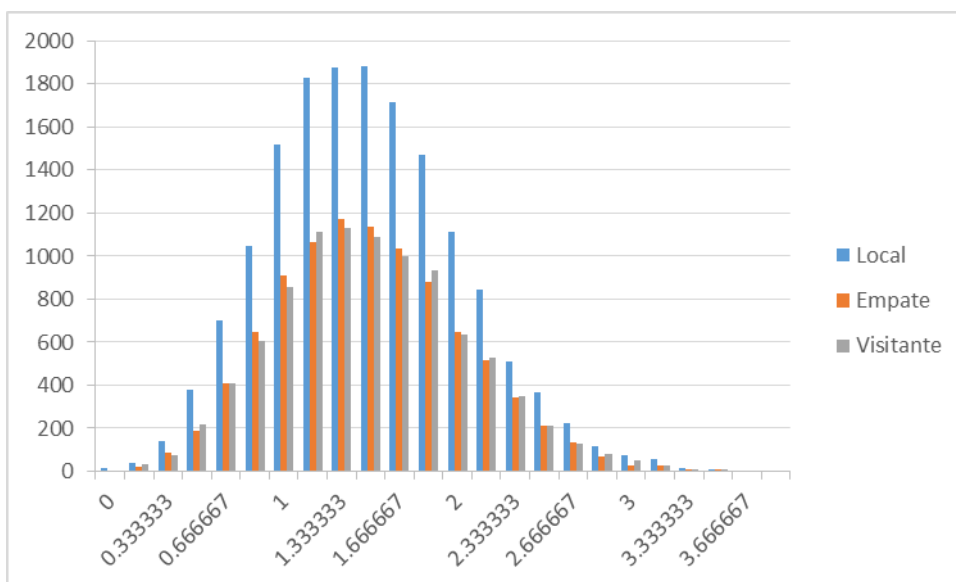
Tabla 17 Goles en Contra de Visitante

Goles en Contra de Visitante	Local	Empate	Visitante
0	12	1	6
0.166667	40	19	31
0.333333	140	86	75
0.5	379	191	219
0.666667	702	406	408
0.833333	1048	648	606
1	1521	912	855
1.166667	1828	1068	1112
1.333333	1876	1171	1134
1.5	1882	1136	1089
1.666667	1718	1035	998
1.833333	1470	883	933



2	1116	650	635
2.166667	842	516	530
2.333333	509	346	348
2.5	366	210	213
2.666667	221	135	129
2.833333	116	68	82
3	76	26	51
3.166667	56	24	26
3.333333	13	9	11
3.5	9	8	9
3.666667	4	6	6
3.833333	4	2	2

Ilustración 13 Goles en Contra de Visitante



## Anexo B – Evidencias

### Lista de videos en Youtube.

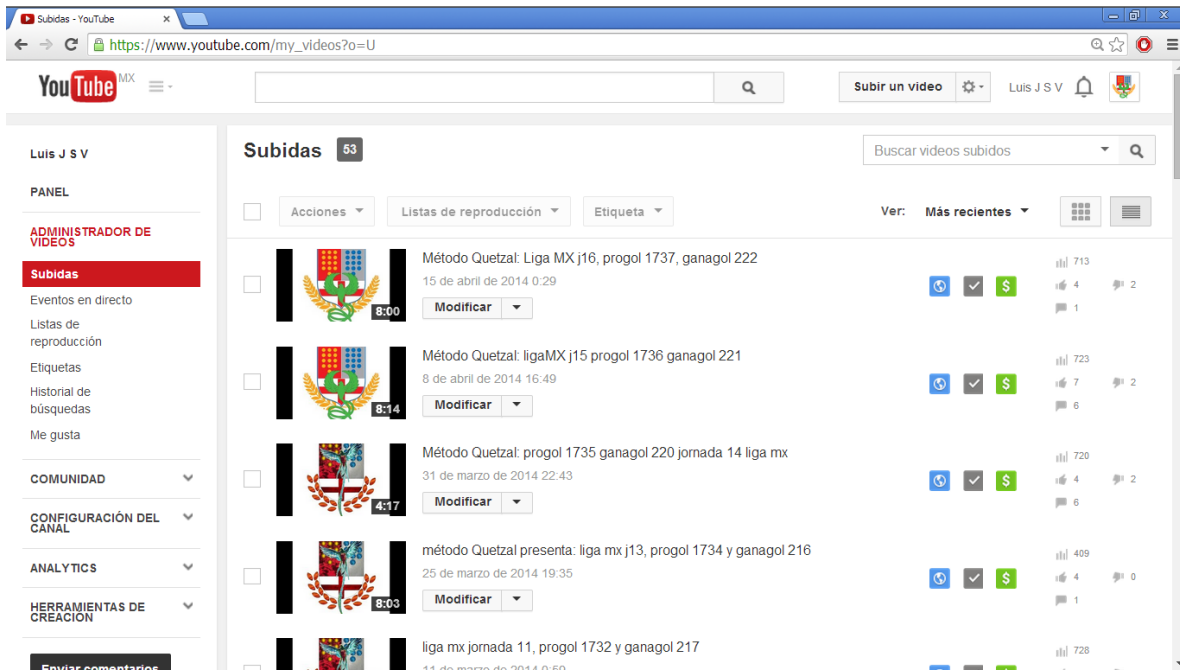


Ilustración 14 Pantalla de Youtube

### Páginas oficiales.

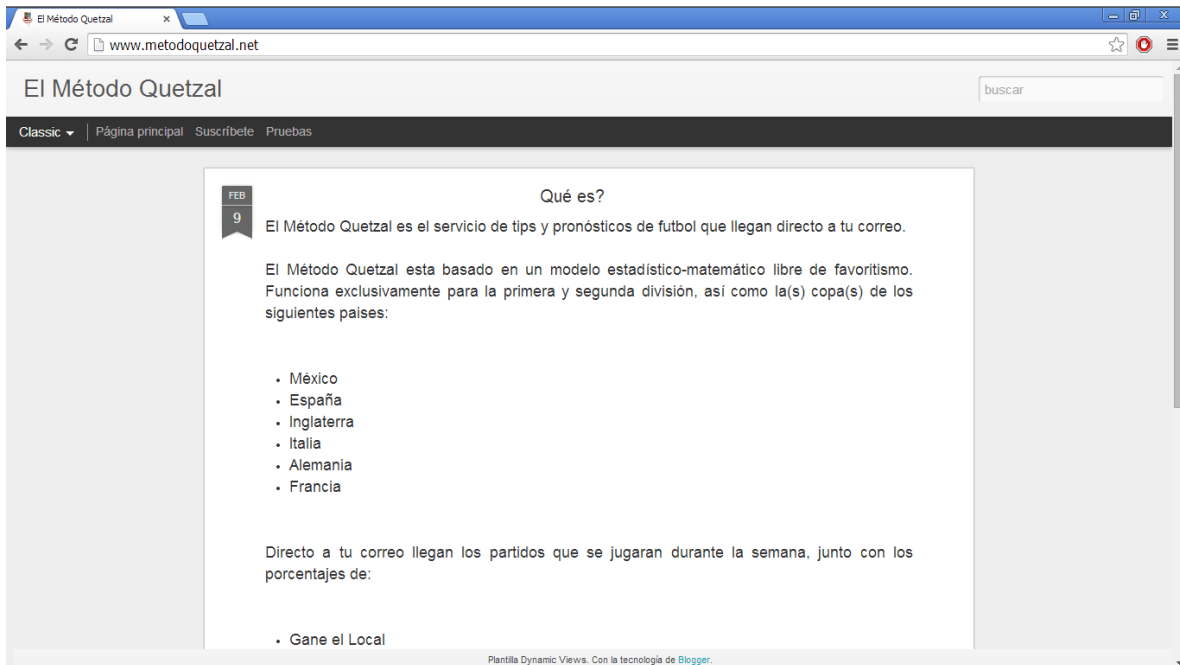


Ilustración 15 metodoquetzal.net



Ilustración 16 metodoquetzal.com

## Versión gratuita del servicio de momios Método Quetzal.

The screenshot shows an Excel spreadsheet with columns for dates, teams, and betting odds. The data is organized into 'Local' and 'Visitante' sections. The 'Local' section includes teams like Cruz Azul, Santos Laguna, América, Pachuca, Monterrey, León, Toluca, Atlante, and Pumas UNAM. The 'Visitante' section includes teams like Cruz Azul, Santos Laguna, América, Pachuca, Monterrey, León, Toluca, Atlante, and Pumas UNAM. The spreadsheet also includes columns for 'GF', 'GC', 'G', 'E', 'P', and 'V' (Victories). The bottom of the spreadsheet shows the status 'READY' and a zoom level of 100%.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1																				
2																				
3																				
4	18/04/2014	Morelia	19:30	Cruz Azul	MX		GF	GC	G	E	P		GF	GC	G	E	P			
5	18/04/2014	Santos Laguna	21:30	Chiapas	MX		2.666667	1.666667	0.833333	0	0.166667		1	1.833333	0.166667	0.166667	0.666667			
6	19/04/2014	América	17:00	Querétaro	MX		0.833333	1.666667	0.166667	0.166667	0.666667		1.333333	1.833333	0.166667	0.166667	0.666667			
7	19/04/2014	Pachuca	19:00	Puebla	MX		1.666667	0.666667	0.666667	0	0.333333		0.166667	2.5	0	0.333333	0.666667			
8	19/04/2014	Monterrey	19:00	Veracruz	MX		1.5	1.833333	0.166667	0.333333	0.5		0.833333	1	0.333333	0.5	0.166667			
9	19/04/2014	León	20:05	Tigres UANL	MX		1.166667	1.5	0.166667	0.333333	0.5		0.833333	0.5	0.333333	0.5	0.166667			
10	19/04/2014	Atlas	21:00	Toluca	MX		1	1.166667	0.166667	0.5	0.333333		1.166667	1.333333	0.333333	0	0.666667			
11	19/04/2014	Atlante	21:00	Tijuana	MX		1.666667	1.333333	0.5	0.333333	0.166667		1	2	0	0.333333	0.666667			
12	20/04/2014	Pumas UNAM	12:00	Guadalajara	MX		2	0.833333	0.5	0.333333	0.166667		1	1	0.333333	0.333333	0.333333			
13																				
14																				
15	18/04/2014	Morelia	19:30	Cruz Azul	MX															
16	18/04/2014	Santos Laguna	21:30	Chiapas	MX															
17	19/04/2014	América	17:00	Querétaro	MX															
18	19/04/2014	Pachuca	19:00	Puebla	MX															
19	19/04/2014	Monterrey	19:00	Veracruz	MX															
20	19/04/2014	León	20:05	Tigres UANL	MX															
21	19/04/2014	Atlas	21:00	Toluca	MX															
22	19/04/2014	Atlante	21:00	Tijuana	MX															
23	20/04/2014	Pumas UNAM	12:00	Guadalajara	MX															
24																				
25																				

Ilustración 17 Versión gratuita

Para dar confianza y probar la efectividad del Método Quetzal se realizaron apuestas en el casino on-line Bet365.

No.	Selecciones	Evento	Fecha	Condiciones de G/C	Cuota	Resultado
1	Toluca	Toluca v Monarcas Morelia (Empate, apuesta no válida)	12/01/2014	Ninguna	1.53	Push
2	Veracruz	Veracruz v Atlante (Empate, apuesta no válida)	11/01/2014	Ninguna	1.50	Ganada
3	Querétaro	Querétaro v Pumas de la UNAM (Empate, apuesta no válida)	10/01/2014	Ninguna	1.72	Ganada

Ilustración 18 Apuesta 12/01

No.	Selecciones	Evento	Fecha	Condiciones de G/C	Cuota	Resultado
1	Bolton	Bolton v Cardiff (Empate, apuesta no válida)	25/01/2014	Ninguna	1.80	Perdida
2	Celta de Vigo	Celta de Vigo v Real Betis (Empate, apuesta no válida)	24/01/2014	Ninguna	1.44	Ganada
3	Almería	Almería v Getafe (Empate, apuesta no válida)	26/01/2014	Ninguna	1.61	Ganada
4	Córdoba	Córdoba v Deportivo La Coruña (Empate, apuesta no válida)	25/01/2014	Ninguna	1.66	Perdida
5	Puebla	Puebla v Chiapas (Empate, apuesta no válida)	26/01/2014	Ninguna	1.61	Perdida
6	Sporting de Gijón	Sporting de Gijón v Recreativo de Huelva (Empate, apuesta no válida)	26/01/2014	Ninguna	1.33	Push
7	Torino	Torino v Atalanta (Empate, apuesta no válida)	26/01/2014	Ninguna	1.25	Ganada

Ilustración 19 Apuesta 26/01

No.	Selecciones	Evento	Fecha	Condiciones de G/C	Cuota	Resultado
1	Monarcas Morelia o Empate	Monarcas Morelia v León (Doble oportunidad)	31/01/2014	Ninguna	1.53	Ganada
2	Granada o Empate	Granada v Celta de Vigo (Doble oportunidad)	31/01/2014	Ninguna	1.29	Perdida
3	Empate o Borussia Dortmund	Eintracht Braunschweig v Borussia Dortmund (Doble oportunidad)	31/01/2014	Ninguna	1.071	Ganada
4	Paris St. Germain o Empate	Paris St. Germain v Bordeaux (Doble oportunidad)	31/01/2014	Ninguna	1.05	Ganada
5	Chateauroux o Empate	Chateauroux v Auxerre (Doble oportunidad)	31/01/2014	Ninguna	1.44	Ganada
6	Clermont Foot o Empate	Clermont Foot v Arles (Doble oportunidad)	31/01/2014	Ninguna	1.29	Ganada
7	Dijon o Empate	Dijon v Brest (Doble oportunidad)	31/01/2014	Ninguna	1.20	Ganada

*Ilustración 20 Apuesta 31/01*

No.	Selecciones	Evento	Fecha	Condiciones de G/C	Cuota	Resultado
1	Liverpool o Empate	Liverpool v Arsenal (Doble oportunidad)	08/02/2014	Ninguna	1.33	Ganada
2	Manchester City	Norwich v Manchester City (Resultado final)	08/02/2014	Ninguna	1.36	Perdida
3	Aston Villa o Empate	Aston Villa v West Ham (Doble oportunidad)	08/02/2014	Ninguna	1.28	Perdida
4	Swansea o Empate	Swansea v Cardiff (Doble oportunidad)	08/02/2014	Ninguna	1.20	Ganada
5	Empate o Nottingham Forest	Blackpool v Nottingham Forest (Doble oportunidad)	08/02/2014	Ninguna	1.25	Ganada
6	Empate o TSG Hoffenheim	SC Friburgo v TSG Hoffenheim (Doble oportunidad)	08/02/2014	Ninguna	1.42	Ganada
7	Wolfsburgo o Empate	Wolfsburgo v Mainz (Doble oportunidad)	08/02/2014	Ninguna	1.22	Ganada
8	Kaiserslautern o Empate	Kaiserslautern v Greuther Fürth (Doble oportunidad)	08/02/2014	Ninguna	1.20	Ganada
9	Napoli o Empate	Napoli v AC Milan (Doble oportunidad)	08/02/2014	Ninguna	1.28	Ganada
10	Carpi o Empate	Carpi v Spezia (Doble oportunidad)	08/02/2014	Ninguna	1.28	Ganada
11	Trapani o Empate	Trapani v Empoli (Doble oportunidad)	08/02/2014	Ninguna	1.57	Ganada

Ilustración 21 Apuesta 8/02

No.	Selecciones	Evento	Fecha	Condiciones de G/C	Cuota	Resultado
1	Pachuca o Empate	Pachuca v Atlante (Doble oportunidad)	15/02/2014	Ninguna	1.20	Perdida
2	Dorados o Empate	Dorados v Altamira (Doble oportunidad)	15/02/2014	Ninguna	1.36	Ganada
3	Club Celaya o Empate	Club Celaya v Zacatepec 1948 (Doble oportunidad)	15/02/2014	Ninguna	1.16	Perdida
4	Ipswich o Empate	Ipswich v Blackpool (Doble oportunidad)	15/02/2014	Ninguna	1.18	Ganada
5	Sheffield Wednesday o Empate	Sheffield Wednesday v Charlton (Doble oportunidad)	15/02/2014	Ninguna	1.22	Anulada
6	Manchester City o Empate	Manchester City v Chelsea (Doble oportunidad)	15/02/2014	Ninguna	1.29	Ganada
7	Florentina o Empate	Florentina v Inter de Milán (Doble oportunidad)	15/02/2014	Ninguna	1.33	Perdida
8	Latina o Empate	Latina v Cittadella (Doble oportunidad)	15/02/2014	Ninguna	1.22	Ganada
9	Siena o Empate	Siena v Juve Stabia (Doble oportunidad)	15/02/2014	Ninguna	1.20	Ganada
10	Pescara o Empate	Pescara v Varese (Doble oportunidad)	15/02/2014	Ninguna	1.22	Perdida

Ilustración 22 Apuesta 15/02

## Bibliografía

- Duda R.O., Hart P.E., Stork D.G. *Pattern Clasification*, 2ª ed., EE. UU. , Willey, 2000.
- Bishop C. M. *Pattern Recognition and Machine Learning*, 1ª ed., EE. UU., Springer, 2006.
- Scott Long J. *Regression Models for Categorical and limited dependent variables*, 1ª ed., EE. UU., Sage, 1997.
- Hagan M.T., Demuth H.B., Beale M. *Neural Network Design*, 1ª ed., EE. UU., PWS Publishing Company, 1996.
- De Veaux R.D., Velleman P. F., Bock D. E., *Stats Data and Models*, 3<sup>er</sup> ed., EE. UU., Addison-Wesley, 2011.
- Anderson T.W., Darling D.A. "A test of Goodness of Fit" *Journal of the American Statistical Association*, Vol. 49, No. 268, 1954, pp. 765-769.
- Chapra S.C., Canale R.P. *Métodos Numéricos Para Ingenieros* (trad.), 1ª ed., México, McGraw-Hill, 1988.
- Pita C. *Álgebra Lineal*, 1ª ed., México, McGraw-Hill, 1991.
- Buchdahl J. *Fixed Odds Sports Betting*, 1ª ed., Londres Ing., High Stakes, 2003.
- Drapkin T., Forsyth R. *The Punter's Revenge*, Londres Ing., Chapman & Hall, 1986.
- 37 Signals, *Getting Real*, 1ª ed., EE. UU., 37 Signals, 2006
- Ries E. *The Lean Startup*, 1ª ed., EE. UU., Crown Business, 2011.
- Taleb N. N. *¿Existe la suerte?* (trad.), 2ª ed., México, Paidos, 2009.
- Moliner M. *Diccionario de Uso del Español*, 3ª ed., España, Gredos, 2007. Vía web: [http://www.diclib.com/cgi-bin/d.cgi?p=riesgo&page=search&l=es&base=&prefbase=&newinput=1&st=&diff\\_examples=1&category=cat4#.U1YAFvmSxK8#ixzz2zagZWmOf](http://www.diclib.com/cgi-bin/d.cgi?p=riesgo&page=search&l=es&base=&prefbase=&newinput=1&st=&diff_examples=1&category=cat4#.U1YAFvmSxK8#ixzz2zagZWmOf)