



Universidad Tecnológica Metropolitana  
Facultad de Ingeniería  
Escuela de Informática y Computación

# Emparejamientos en partidas oficiales de Counter-Strike: Global Offensive

---

*Minería de Datos*

**INTEGRANTE**

Jonathan Cartagena

**ASIGNATURA**

Minería de Datos

**PROFESOR**

Pablo Figueroa

SANTIAGO, DICIEMBRE 2017

## Indice

<b>1. Introducción</b>	<b>3</b>
<b>2. Marco Teorico</b>	<b>3</b>
2.1 Proceso de Extracción del Conocimiento (KDD)	3
2.2 Metodología con KDD	4
2.2.1. Selección de datos	4
2.2.2. Pre-procesamiento	4
2.2.3. Transformación	4
2.2.4. Data Mining	4
2.2.5. Interpretacion y Evaluacion	4
2.3 Medidas de dispersion	4
2.4 Análisis factorial	5
2.5 Herramientas	5
2.5.1 BigML	5
2.5.2 IBM SPSS	5
<b>3. Hipotesis</b>	<b>5</b>
<b>4. Dataset</b>	<b>6</b>
<b>5. Variables</b>	<b>6</b>
5.1 Variables de Entrada	6
5.2 Variables de Salida	7
5.3 Limpiado de Variables	7
<b>6. Estadísticos Descriptivos</b>	<b>16</b>
<b>7. Análisis Factorial</b>	<b>16</b>
7.1 Test de Bartlett y KMO	17
7.2 Observaciones	23
<b>8. Transformación</b>	<b>23</b>
8.1 Entrenamiento de datos	23
8.2 Modelo de Naive Bayes	26
<b>9. Conclusiones</b>	<b>27</b>

## 1. Introducción

Los videojuegos son un área rica para la examinación de datos debido a su naturaleza digital. Ejemplos notables como la compleja economía en línea de EVE, el incidente de sangre corrupto de World of Warcraft e incluso los autos automovilísticos Grand Theft Auto nos dicen que la ficción está más cerca de la realidad de lo que realmente creemos. Se puede obtener una idea de la lógica y la toma de decisiones que enfrentan los jugadores cuando se colocan en escenarios hipotéticos y virtuales.

En este conjunto de datos se toma poco más de 600 partidos de emparejamiento competitivos del juego de Valve Counter-strike: Global Offensive (CS: GO). Los datos se extrajeron de repeticiones competitivas de emparejamientos enviadas a csgo-stats.

## 2. Marco Teorico

### 2.1 Proceso de Extracción del Conocimiento (KDD)

Proceso no trivial de identificar patrones válidos, novedosos y potencialmente útiles y en última instancia, comprensible a partir de los datos. Este proceso también es conocido por diferentes nombres que podrían ser sinónimos del mismo, entre los cuales se encuentran *Data Archeology*, *Dependency Function Analysis*, *Information Recollect*, *Pattern Data Analysis* o *Knowledge Fishing*.

KDD también supone la convergencia de distintas disciplinas de investigación; podemos nombrar algunas tales como el aprendizaje automático, estadística, inteligencia artificial, sistemas de gestión de base de datos, técnicas de visualización de datos, los sistemas para el apoyo a la toma de decisión (DSS) o la recuperación de información, entre otras.

### 2.2 Metodología con KDD

#### 2.2.1. Selección de datos

En esta etapa se determinan las fuentes de datos y el tipo de información a utilizar. Es la etapa donde los datos relevantes para el análisis son extraídos desde la o las fuentes de datos.

---

### *2.2.2. Pre-procesamiento*

Esta etapa consiste en la preparación y limpieza de los datos extraídos desde las distintas fuentes de datos en una forma manejable, necesaria para las fases posteriores. En esta etapa se utilizan diversas estrategias para manejar datos faltantes o en blanco, datos inconsistentes o que están fuera de rango, obteniéndose al final una estructura de datos adecuada para su posterior transformación.

---

### *2.2.3. Transformación*

Consiste en el tratamiento preliminar de los datos, transformación y generación de nuevas variables a partir de las ya existentes con una estructura de datos apropiada. Aquí se realizan operaciones de agregación o normalización, consolidando los datos de una forma necesaria para la fase siguiente.

---

### *2.2.4. Data Mining*

Es la fase de modelamiento propiamente tal, en donde métodos inteligentes son aplicados con el objetivo de extraer patrones previamente desconocidos, válidos, nuevos, potencialmente útiles y comprensibles y que están contenidos u “ocultos” en los datos.

---

### *2.2.5. Interpretación y Evaluación*

Se identifican los patrones obtenidos y que son realmente interesantes, basándose en algunas medidas y se realiza una evaluación de los resultados obtenidos.

## 2.3 Medidas de dispersión

Las medidas de dispersión nos informan sobre cuánto se alejan del centro los valores de la distribución.

- Rango o recorrido: El rango es la diferencia entre el mayor y el menor de los datos de una distribución estadística.
- Desviación media: La desviación media es la media aritmética de los valores absolutos de las desviaciones respecto a la media.
- Varianza: La varianza es la media aritmética del cuadrado de las desviaciones respecto a la media.
- Desviación típica: La desviación típica es la raíz cuadrada de la varianza.

## 2.4 Análisis factorial

Es una técnica estadística de reducción de datos usada para explicar las correlaciones entre las variables observadas en términos de un número menor de variables no observadas llamadas factores. Las variables observadas se modelan como combinaciones lineales de factores más expresiones de error. El análisis factorial se originó en psicometría, y se usa en las ciencias del comportamiento tales como ciencias sociales, marketing, gestión de productos, investigación operativa, y otras ciencias aplicadas que tratan con grandes cantidades de datos.

## 2.5 Herramientas

### 2.5.1 BigML

Herramienta en la nube, para la modelización de datos y el desarrollo de modelos de inteligencia artificial (*machine learning*) a partir de los mismos.

### 2.5.2 IBM SPSS

Es el software de análisis predictivo que ofrece técnicas avanzadas en un paquete fácil de usar que ayuda a encontrar nuevas oportunidades, mejorar la eficiencia y minimizar el riesgo. Así mismo, proporciona informes y análisis estadísticos, minería de datos y análisis de big data.

## 3. Hipotesis

Las partidas de emparejamiento no demuestran la mayoría de las veces el nivel de los jugadores en distintas ligas, esto debido al mal juego colectivo con otro jugadores y sistemas de puntuación fijados.

Se busca refutar o rechazar este sistema de puntuación actual mediante un algoritmo el cual permita evaluar de una forma más precisa el rendimiento de los distintos jugadores dentro de las partidas teniendo en cuenta distintos factores los cuales pueden afectar a dicha problemática como lo son: daño total causado, posición de daño, mapa disputado, tipo de ronda.

## 4. Dataset

El dataset que se obtiene contiene 33 columnas, de los cuales solo 10 serán descritos como las variables dado que estos presentan relevancia a la hipótesis, por ejemplo se eliminan variables como nombre de partida, tiempo de respuesta del servidor, posición exacta del impacto (coordenadas X, Y y Z), lado de bando, numeros de identificacion de usuario y otras irrelevantes para el estudio. Las variables filtradas sirven para la obtención de las salidas como se muestra en la Figura 4.1.

Figura 4.1: Cabecera de dataset

mapa	ata_side	vic_side	hp_dmg	arm_dmg	bomba_sitio	hitbox	eq_ganador	ct_eq_val	t_eq_val
de_dust2	Terrorist	CounterTerrorist	18	0	A	RightLeg	Terrorist	12400	4700
de_dust2	Terrorist	CounterTerrorist	22	1	A	Chest	Terrorist	12400	4700
de_dust2	Terrorist	CounterTerrorist	29	1	A	Stomach	Terrorist	12400	4700

Fuente: Elaboración Propia

Este conjunto de datos dentro de las 600 partidas proporciona cada entrada exitosa de duelos (o batalla) que tuvieron lugar para un jugador. Es decir, cada fila documenta un evento cuando un jugador es lastimado por otro jugador (daño por caída).

Total de datos: 64018 entradas.

Datos nulos: 0 entradas.

## 5. Variables

### 5.1 Variables de Entrada

1. mapa: El mapa oficial en el que se jugó el partido.
2. ata\_side: El equipo en el que está atacando el jugador que causó daño a la víctima. Puede ser terrorista o antiterrorista.
3. vic\_side: El lado en el que estaba la víctima. Puede ser terrorista o antiterrorista.
4. hp\_dmg: El daño total se reparte en ese duelo a la víctima. Cada jugador comienza la ronda con 100 hp máx.
5. arm\_dmg: El daño total se aplicó a kevlar. Tres cosas a tener en cuenta: 1. Kevlar es un artículo opcional que los jugadores eligen comprar 2. Kevlar solo protege el área del pecho y 3. El daño a Kevlar ya se contabiliza en hp\_dmg, es decir, si hp\_dmg = 50 y arm\_dmg = 50, el jugador solo ha perdido 50 CV y aún está vivo.
6. bomb\_sitio: El sitio donde se coloca la bomba (solo A o B) y vacío si bomba\_plantada es falso.
7. hitbox: el área del cuerpo en la que se golpeó a la víctima.
8. eq\_ganador: El equipo que ganó al final de esa ronda.
9. ct\_eq\_val: Valor total del equipamiento del equipo contra-terrorista (arma + granadas + armadura + utilidades) después del tiempo de compra.
10. t\_eq\_val: Valor del equipo total del equipo terrorista (arma + granadas + armadura + utilidades) después del tiempo de compra.

## 5.2 Variables de Salida

Esta variable contiene dos posibles valores 1 en el caso en que efectivamente el jugador tenga un bonus de rango y 2 en caso de ser una ejecución normal el cual no posee beneficio.

1. bonus: Valor que se asignara al jugador después de la ronda jugada, puede ser 1 o
- 2.

## 5.3 Limpiado de Variables

Tabla 5.1: Definición de variables

Variable	Tipo	Valor
mapa	Nominal	de_cache: 1 de_dust2: 2 de_mirage: 3 de_inferno, de_cbble, de_overpass, de_train: 4
ata_side	Nominal	CounterTerrorist: 1 Terrorist: 2
vic_side	Nominal	CounterTerrorist: 1 Terrorist 2
hp_dmg	Ordinal	Entre 1-9: 1 Entre 29-34: 2 Entre 35-58: 3 Entre 59-100: 4
arm_dmg	Ordinal	0 daño: 1 Entre 1-7: 2
bomba_sitio	Nominal	No plantada: 1 Plantada en A: 2 Plantada en B: 3
hitbox	Nominal	Chest: 1 Generic: 2 Head: 3 LeftArm, RightArm: 4 LeftLeg, RightLeg: 5 Stomach: 6
eq_ganador	Nominal	CounterTerrorist: 1 Terrorist: 2

Variable	Tipo	Valor
tipo_ronda	Nominal	ECO: 1 FORCEBUY, PISTOL_ROUND, SEMIECO: 2 NORMAL: 3

Fuente: Elaboracion propia

Tabla 5.1: Cabecera dataset limpiado

mapa	ata_side	vic_side	hp_dmg	arm_dmg	bomba_sitio	hitbox	eq_ganador	tipo_ronda
3	1	2	5	1	1	3	1	4
3	1	2	3	1	1	5	1	4
3	2	1	2	1	1	5	1	4

Fuente: Elaboración propia

En la figura 5.1 se encuentran las variables de entrada:

Figura 5.1: Variables de entrada

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	mapa	Numérico	1	0		Ninguno	Ninguno	12	Derecha	Nominal	Entrada
2	ata_side	Numérico	1	0		Ninguno	Ninguno	12	Derecha	Nominal	Entrada
3	vic_side	Numérico	1	0		Ninguno	Ninguno	12	Derecha	Nominal	Entrada
4	hp_dmg	Numérico	1	0		Ninguno	Ninguno	12	Derecha	Ordinal	Entrada
5	arm_dmg	Numérico	1	0		Ninguno	Ninguno	12	Derecha	Ordinal	Entrada
6	bomba_sitio	Numérico	1	0		Ninguno	Ninguno	12	Derecha	Nominal	Entrada
7	hitbox	Numérico	1	0		Ninguno	Ninguno	12	Derecha	Nominal	Entrada
8	eq_ganador	Numérico	1	0		Ninguno	Ninguno	12	Derecha	Nominal	Entrada
9	tipo_ronda	Numérico	1	0		Ninguno	Ninguno	12	Derecha	Nominal	Entrada
10	bonus	Numérico	1	0		Ninguno	Ninguno	12	Derecha	Nominal	Entrada

Fuente: Elaboración propia

A continuación se muestran las distintas variables con sus respectivos porcentajes y frecuencias:

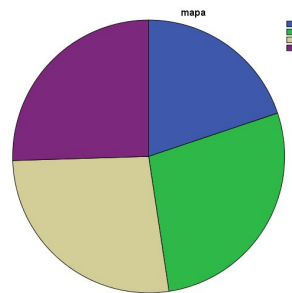
Tabla 5.2: mapa

mapa					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	1	4049	19,9	19,9	19,9
	2	5657	27,7	27,7	47,6
	3	5486	26,9	26,9	74,5
	4	5195	25,5	25,5	100,0
	Total	20387	100,0	100,0	

Fuente: Elaboración propia



Figura 5.3:: Gráfico mapa



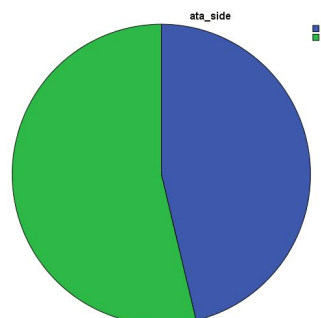
Fuente: Elaboración propia

Tabla 5.3 : ata\_side

ata_side					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	1	9430	46,3	46,3	46,3
	2	10957	53,7	53,7	100,0
	Total	20387	100,0	100,0	

Fuente: Elaboración propia

Figura 5.4: Grafico ata\_side



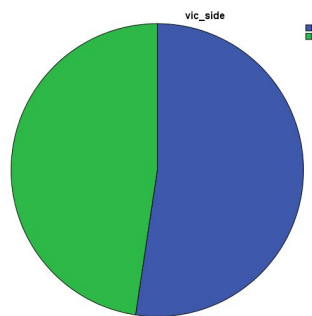
Fuente: Elaboración propia

Tabla 5.4: vic\_side

vic_side					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	1	10678	52,4	52,4	52,4
	2	9709	47,6	47,6	100,0
	Total	20387	100,0	100,0	

Fuente: Elaboración propia

Figura 5.4: vic\_side



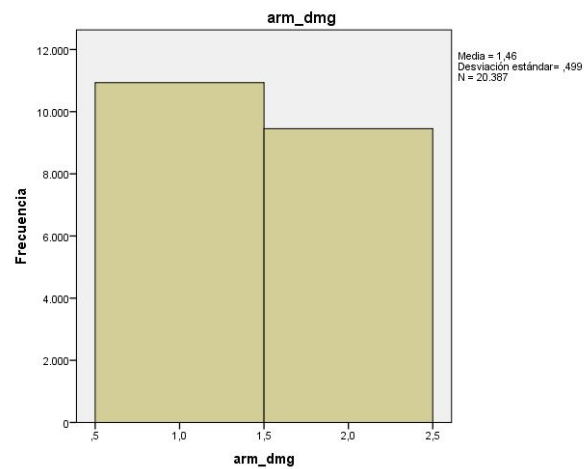
Fuente: Elaboración propia

Tabla 5.5: arm\_dmg

arm_dmg					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	1	10935	53,6	53,6	53,6
	2	9452	46,4	46,4	100,0
	Total	20387	100,0	100,0	

Fuente: Elaboración propia

Figura 5.6: Histograma arm\_dmg



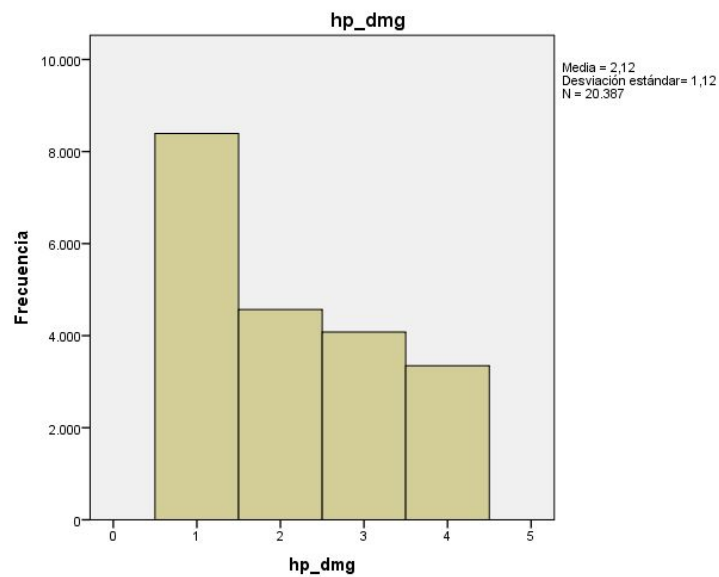
Fuente: Elaboración propia

Tabla 5.6: hp\_dmg

hp_dmg					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	1	8393	41,2	41,2	41,2
	2	4569	22,4	22,4	63,6
	3	4078	20,0	20,0	83,6
	4	3347	16,4	16,4	100,0
	Total	20387	100,0	100,0	

Fuente: Elaboración propia

Figura 5.7: Histograma de hp\_dmg



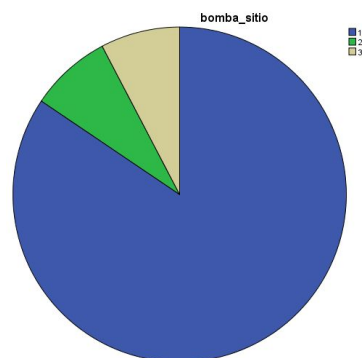
Fuente: Elaboración propia

Tabla 5.7: bomba\_sitio

bomba_sitio					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	1	17212	84,4	84,4	84,4
	2	1607	7,9	7,9	92,3
	3	1568	7,7	7,7	100,0
	Total	20387	100,0	100,0	

Fuente: Elaboración propia

Figura 5.7: Grafico bomba\_sitio



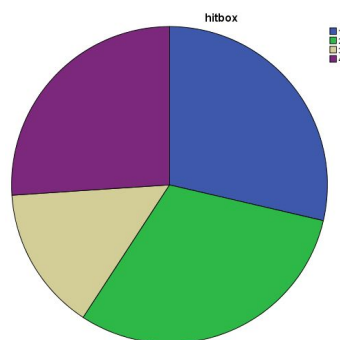
Fuente: Elaboración propia

Tabla 5.8: hitbox

hitbox					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	1	5836	28,6	28,6	28,6
	2	6236	30,6	30,6	59,2
	3	3006	14,7	14,7	74,0
	4	5309	26,0	26,0	100,0
	Total	20387	100,0	100,0	

Fuente: Elaboración propia

Figura 5.8: Histograma hitbox



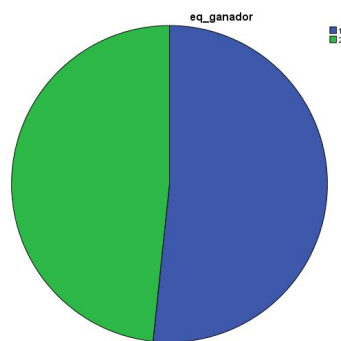
Fuente: Elaboración propia

Tabla 5.9: eq\_ganador

eq_ganador					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	1	10533	51,7	51,7	51,7
	2	9854	48,3	48,3	100,0
	Total	20387	100,0	100,0	

Fuente: Elaboración propia

Figura 5.9: eq\_ganador



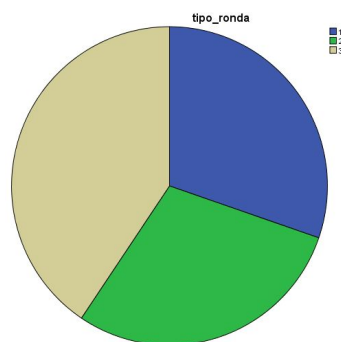
Fuente: Elaboración propia

Tabla 5.9: tipo\_ronda

tipo_ronda					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	1	6174	30,3	30,3	30,3
	2	5939	29,1	29,1	59,4
	3	8274	40,6	40,6	100,0
	Total	20387	100,0	100,0	

Fuente: Elaboración propia

Figura 5.10: tipo\_ronda



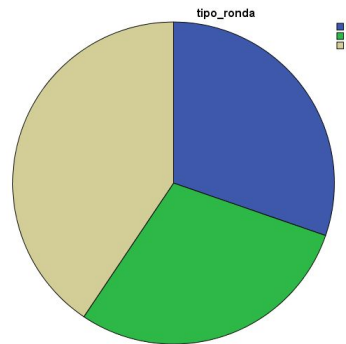
Fuente: Elaboración propia

Tabla 5.11: bonus

bonus					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	1	5357	26,3	26,3	26,3
	2	15030	73,7	73,7	100,0
	Total	20387	100,0	100,0	

Fuente: Elaboración propia

Figura 5.11: bonus



Fuente Elaboración propia

## 6. Estadísticos Descriptivos

Figura 6.1: Estadísticos Descriptivos

Estadísticos descriptivos											
	N	Rango	Mínimo	Máximo	Suma	Media	Error estándar	Desviación estándar	Varianza	Curtosis	Error estándar
	Estadístico	Estadístico	Estadístico	Estadístico	Estadístico	Estadístico		Estadístico	Estadístico	Estadístico	
mapa	20387	3	1	4	52601	2,58	,008	1,073	1,150	-1,252	,034
ata_side	20387	1	1	2	31344	1,54	,003	,499	,249	-1,978	,034
vic_side	20387	1	1	2	30096	1,48	,003	,499	,249	-1,991	,034
hp_dmg	20387	3	1	4	43153	2,12	,008	1,120	1,255	-1,204	,034
arm_dmg	20387	1	1	2	29839	1,46	,003	,499	,249	-1,979	,034
bomba_sitio	20387	2	1	3	25130	1,23	,004	,577	,332	4,065	,034
hitbox	20387	3	1	4	48562	2,38	,008	1,153	1,329	-1,384	,034
eq_ganador	20387	1	1	2	30241	1,48	,003	,500	,250	-1,996	,034
tipo_ronda	20387	2	1	3	42874	2,10	,006	,836	,698	-1,541	,034
bonus	20387	1	1	2	35417	1,74	,003	,440	,194	-,838	,034
N válido (por lista)	20387										

Fuente: Elaboración propia

## 7. Análisis Factorial

Matriz de correlaciones <sup>a</sup>											
		mapa	ata_side	vic_side	hp_dmg	arm_dmg	bomba_sitio	hitbox	eq_ganador	tipo_ronda	bonus
Correlación	mapa	1,000	-,009	,010	-,044	-,034	-,017	,014	-,027	-,052	,001
	ata_side	-,009	1,000	-,876	,055	,003	,050	,091	,058	,009	-,177
	vic_side	,010	-,876	1,000	-,074	,009	-,030	-,106	-,054	-,002	,167
	hp_dmg	-,044	,055	-,074	1,000	,082	-,034	,158	,015	-,092	-,402
	arm_dmg	-,034	,003	,009	,082	1,000	,011	,133	-,024	,248	,139
	bomba_sitio	-,017	,050	-,030	-,034	,011	1,000	,007	,030	-,014	-,108
	hitbox	,014	,091	-,106	,158	,133	,007	1,000	-,003	,050	-,111
	eq_ganador	-,027	,058	-,054	,015	-,024	,030	-,003	1,000	-,062	-,068
	tipo_ronda	-,052	,009	-,002	-,092	,248	-,014	,050	-,062	1,000	,392
	bonus	,001	-,177	,167	-,402	,139	-,108	-,111	-,068	,392	1,000
Sig. (unilateral)	mapa		,101	,084	,000	,000	,008	,020	,000	,000	,437
	ata_side	,101		,000	,000	,336	,000	,000	,000	,088	,000
	vic_side	,084	,000		,000	,100	,000	,000	,000	,387	,000
	hp_dmg	,000	,000	,000		,000	,000	,000	,016	,000	,000
	arm_dmg	,000	,336	,100	,000		,051	,000	,000	,000	,000
	bomba_sitio	,008	,000	,000	,000	,051		,163	,000	,021	,000
	hitbox	,020	,000	,000	,000	,000	,163		,352	,000	,000
	eq_ganador	,000	,000	,000	,016	,000	,000	,352		,000	,000
	tipo_ronda	,000	,088	,387	,000	,000	,021	,000	,000		,000
	bonus	,437	,000	,000	,000	,000	,000	,000	,000	,000	

a. Determinante = ,131

La matriz de correlaciones tiene un determinante de 0,131

Inversión de matriz de correlaciones											
	mapa	ata_side	vic_side	hp_dmg	arm_dmg	bomba_sitio	hitbox	eq_ganador	tipo_ronda	bonus	
mapa	1,007	-,003	-,009	,051	,021	,018	-,029	,029	,057	-,004	
ata_side	-,003	4,336	3,766	,122	-,067	-,077	,021	-,041	-,099	,225	
vic_side	-,009	3,766	4,315	,099	-,070	-,058	,107	,012	-,008	,003	
hp_dmg	,051	,122	,099	1,252	-,151	,102	-,117	,008	-,053	,548	
arm_dmg	,021	-,067	-,070	-,151	1,107	-,034	-,130	,007	-,223	-,145	
bomba_sitio	,018	-,077	-,058	,102	-,034	1,024	,002	-,022	-,032	,164	
hitbox	-,029	,021	,107	-,117	-,130	,002	1,061	,008	-,074	,104	
eq_ganador	,029	-,041	,012	,008	,007	-,022	,008	1,010	,046	,042	
tipo_ronda	,057	-,099	-,008	-,053	-,223	-,032	-,074	,046	1,260	-,509	
bonus	-,004	,225	,003	,548	-,145	,164	,104	,042	-,509	1,511	

### 7.1 Test de Bartlett y KMO

A continuación se realizó el test de Bartlett el cual da como resultado un p-valor igual a 0.000 y ya que si p-valor es menor a 0.05 aceptamos la hipótesis nula, lo que por consiguiente nos dice aplicar el análisis factorial. El análisis KMO nos da como resultado 0.529 lo que nos indica que las variables están correlacionadas.



Figura 7.1: Test de Bartlett y KMO

Prueba de KMO y Bartlett		
Medida Kaiser-Meyer-Olkin de adecuación de muestreo		,529
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado	41395,270
	gl	45
	Sig.	,000

Fuente: Elaboración propia

En la prueba de Bartlett como la sig.  $<0.05$ , se rechaza la  $H_0$  (Hipótesis nula). También se puede observar que la prueba de KMO es mayor a 0.5, que indica no rechazar la  $H_0$ , teniendo sentido realizar un Análisis Factorial.

Matrices anti-imagen											
		mapa	ata_side	vic_side	hp_dmg	arm_dmg	bomba_sitio	hitbox	eq_ganador	tipo_ronda	bonus
Covarianza anti-imagen	mapa	,993	-,001	-,002	,040	,019	,018	-,028	,028	,045	-,003
	ata_side	-,001	,231	,201	,022	-,014	-,017	,005	-,009	-,018	,034
	vic_side	-,002	,201	,232	,018	-,015	-,013	,023	,003	-,001	,000
	hp_dmg	,040	,022	,018	,798	-,109	,079	-,088	,006	-,033	,290
	arm_dmg	,019	-,014	-,015	-,109	,903	-,030	-,111	,006	-,160	-,087
	bomba_sitio	,018	-,017	-,013	,079	-,030	,976	,002	-,021	-,025	,106
	hitbox	-,028	,005	,023	-,088	-,111	,002	,943	,007	-,055	,065
	eq_ganador	,028	-,009	,003	,006	,006	-,021	,007	,990	,036	,027
	tipo_ronda	,045	-,018	-,001	-,033	-,160	-,025	-,055	,036	,794	-,267
	bonus	-,003	,034	,000	,290	-,087	,106	,065	,027	-,267	,662
Correlación anti-imagen	mapa	,507 <sup>a</sup>	-,001	-,004	,045	,020	,018	-,028	,029	,050	-,003
	ata_side	-,001	,513 <sup>a</sup>	,871	,052	-,030	-,037	,010	-,020	-,042	,088
	vic_side	-,004	,871	,516 <sup>a</sup>	,042	-,032	-,028	,050	,006	-,003	,001
	hp_dmg	,045	,052	,042	,514 <sup>a</sup>	-,128	,090	-,102	,007	-,042	,398
	arm_dmg	,020	-,030	-,032	-,128	,565 <sup>a</sup>	-,032	-,120	,007	-,189	-,112
	bomba_sitio	,018	-,037	-,028	,090	-,032	,372 <sup>a</sup>	,002	-,021	-,028	,132
	hitbox	-,028	,010	,050	-,102	-,120	,002	,665 <sup>a</sup>	,007	-,064	,082
	eq_ganador	,029	-,020	,006	,007	,007	-,021	,007	,785 <sup>a</sup>	,041	,034
	tipo_ronda	,050	-,042	-,003	-,042	-,189	-,028	-,064	,041	,558 <sup>a</sup>	-,369
	bonus	-,003	,088	,001	,398	-,112	,132	,082	,034	-,369	,554 <sup>a</sup>
a. Medidas de adecuación de muestreo (MSA)											

Fuente: Elaboración propia

Varianza total explicada									
Componente	Autovalores iniciales			Sumas de extracción de cargas al cuadrado			Sumas de rotación de cargas al cuadrado		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	2,061	20,607	20,607	2,061	20,607	20,607	1,896	18,962	18,962
2	1,566	15,662	36,269	1,566	15,662	36,269	1,532	15,315	34,278
3	1,283	12,830	49,100	1,283	12,830	49,100	1,459	14,586	48,864
4	1,045	10,453	59,553	1,045	10,453	59,553	1,069	10,689	59,553
5	,996	9,960	69,513						
6	,961	9,613	79,126						
7	,820	8,201	87,327						
8	,691	6,913	94,240						
9	,453	4,526	98,766						
10	,123	1,234	100,000						

Método de extracción: análisis de componentes principales.

Fuente: Elaboración propia

En el porcentaje acumulado de los factores es de 59%, por lo tanto se debe considerar más variables. Se agrega 3 componentes a este para completar un 80% obteniéndose:

Varianza total explicada									
Componente	Autovalores iniciales			Sumas de extracción de cargas al cuadrado			Sumas de rotación de cargas al cuadrado		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	2,061	20,607	20,607	2,061	20,607	20,607	1,892	18,923	18,923
2	1,566	15,662	36,269	1,566	15,662	36,269	1,480	14,802	33,724
3	1,283	12,830	49,100	1,283	12,830	49,100	1,332	13,324	47,048
4	1,045	10,453	59,553	1,045	10,453	59,553	1,023	10,232	57,280
5	,996	9,960	69,513	,996	9,960	69,513	1,004	10,041	67,322
6	,961	9,613	79,126	,961	9,613	79,126	1,003	10,033	77,355
7	,820	8,201	87,327	,820	8,201	87,327	,997	9,972	87,327
8	,691	6,913	94,240						
9	,453	4,526	98,766						
10	,123	1,234	100,000						

Método de extracción: análisis de componentes principales.

Fuente: Elaboración propia

Comunalidades		
	Inicial	Extracción
mapa	1,000	,991
ata_side	1,000	,937
vic_side	1,000	,935
hp_dmg	1,000	,773
arm_dmg	1,000	,747
bomba_sitio	1,000	,980
hitbox	1,000	,997
eq_ganador	1,000	,996
tipo_ronda	1,000	,630
bonus	1,000	,747
Método de extracción: análisis de componentes principales.		

Fuente: Elaboración propia

Varianza total explicada									
Componente	Autovalores iniciales			Sumas de extracción de cargas al cuadrado			Sumas de rotación de cargas al cuadrado		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	2,061	20,607	20,607	2,061	20,607	20,607	1,892	18,923	18,923
2	1,566	15,662	36,269	1,566	15,662	36,269	1,480	14,802	33,724
3	1,283	12,830	49,100	1,283	12,830	49,100	1,332	13,324	47,048
4	1,045	10,453	59,553	1,045	10,453	59,553	1,023	10,232	57,280
5	,996	9,960	69,513	,996	9,960	69,513	1,004	10,041	67,322
6	,961	9,613	79,126	,961	9,613	79,126	1,003	10,033	77,355
7	,820	8,201	87,327	,820	8,201	87,327	,997	9,972	87,327
8	,691	6,913	94,240						
9	,453	4,526	98,766						
10	,123	1,234	100,000						
Método de extracción: análisis de componentes principales.									

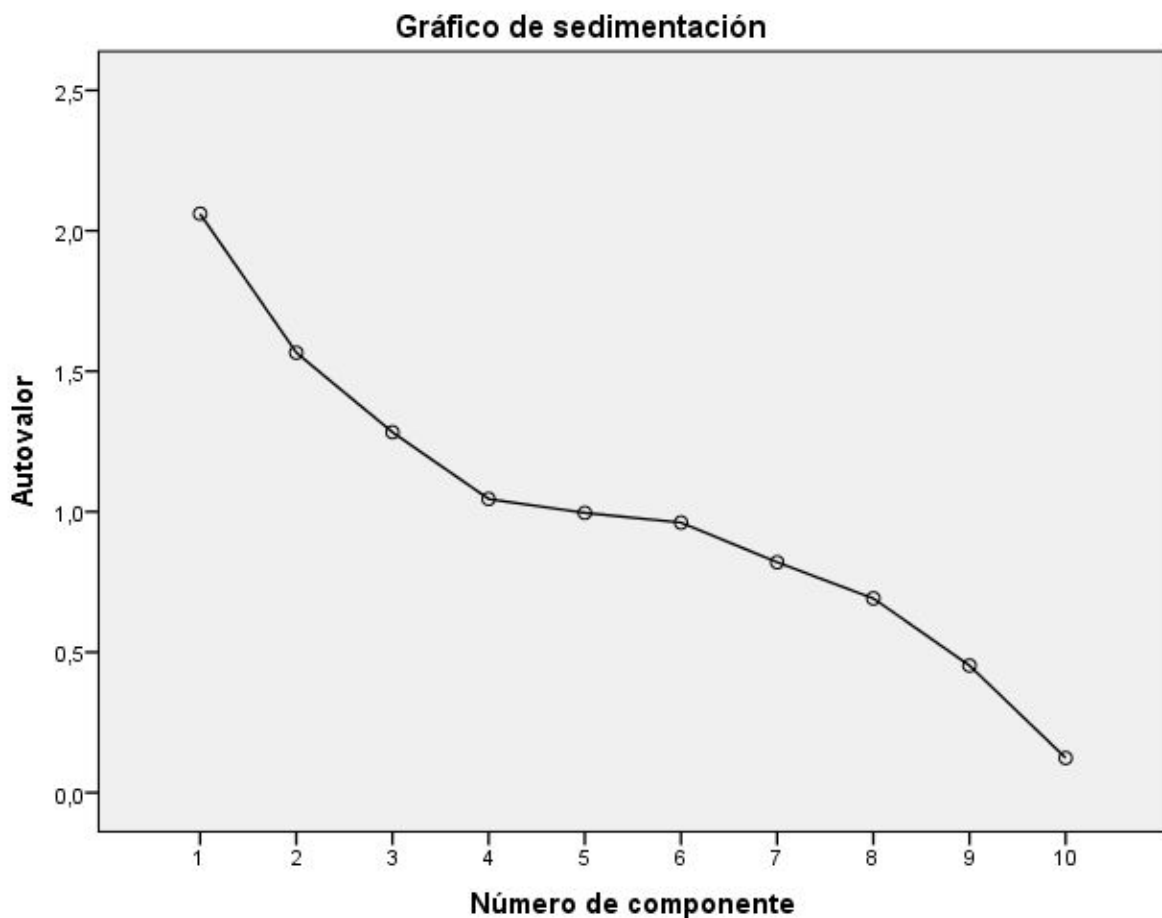
Fuente: Elaboración propia

Se distingue de la tabla los factores:

- Factor 1:
  - 20,6 % poder explicativo de varianza total.
  - de rotación varimax.
- Factor 2:
  - 15,6% poder explicativo de varianza total.
  - 12,6% de rotación varimax.
- Factor 3:
  - 12,93% poder explicativo de varianza total.
  - 10,6% de rotación varimax.

- Factor 4:
  - 10,453% poder explicativo de varianza total.
  - 10,3% de rotación varimax.
- Factor 5:
  - 9,983% poder explicativo de varianza total.
  - 10,21%% de rotación varimax.
- Factor 6:
  - 9,613% poder explicativo de varianza total.
  - 9.97% de rotación varimax.
- Factor 7:
  - 9,381% poder explicativo de varianza total.
  - 9.62% de rotación varimax.

Gráfico de sedimentación el cual distingue una presentación gráfica donde los factores están en el eje de abscisas y los valores propios



Según el gráfico de sedimentación los factores 1 y 2 están apartados hacia la izquierda, teniendo una diferenciación más marcada en el estudio.

En la figura se muestra la matriz de componente. Para el cálculo se ha elegido el método Varimax (método de rotación ortogonal que minimiza el número de variables que tienen saturaciones altas en cada factor).

	Componente						
	1	2	3	4	5	6	7
mapa	,019	-,065	-,156	-,638	,569	,418	,235
ata_side	-,848	,411	-,208	-,042	-,029	-,038	,041
vic_side	,850	-,405	,198	,063	,049	,035	-,026
hp_dmg	-,368	-,371	,606	-,055	-,165	-,042	,317
arm_dmg	,078	,431	,605	,095	,106	,100	,400
bomba_sitio	-,118	-,050	-,039	,574	,757	-,244	,020
hitbox	-,250	,080	,586	-,152	,198	,179	-,700
eq_ganador	-,150	-,099	-,105	,512	-,128	,821	,017
tipo_ronda	,247	,710	,238	,067	,002	,005	,061
bonus	,564	,618	-,175	-,027	-,048	,080	-,083

Método de extracción: análisis de componentes principales.

a. 7 componentes extraídos.

Matriz de Componentes Rotados							
	Componentes						
	1	2	3	4	5	6	7
mapa	-0,003	0,029	-0,025	-0,013	0,994	-0,016	0,012
ata_side	0,965	-0,053	0,005	0,029	-0,001	0,024	0,027
vic_side	-0,964	0,059	0,003	-0,001	0,004	-0,019	-0,043
hp_dmg	0,011	-0,850	0,176	-0,110	-0,053	-0,018	0,069
aim_dmg	-0,026	-0,159	0,846	0,051	0,041	0,025	0,036
bomba_sitio	0,023	-0,005	0,016	0,989	-0,013	0,011	0,002
hitbox	0,058	-0,097	0,075	0,003	0,012	-0,001	0,989
eq_ganador	0,036	-0,016	-0,024	0,012	-0,016	0,997	-0,001
tipo_ronda	0,053	0,389	0,674	-0,045	-0,102	-0,074	0,070
bonus	-0,151	0,752	0,354	-0,163	-0,014	-0,043	-0,072
Método de extracción: análisis de componentes principales.							
Método de rotación: Varimax con normalización Kaiser.							
La rotación ha convergido en 5 iteraciones.							

Se destaca de la matriz rotada las variables que tiene una mayor diferencia en su varianza total en factores que son:

1. Factor 1:
  - a. ata\_side representa un 96,5% del total del espacio de los factores.
2. Factor 2:
  - a. hp\_dmg representa un 85% del total del espacio de los factores
3. Factor 3:
  - a. hitbox representa un 84% del total del espacio de los factores
4. Factor 4:
  - a. tipo\_ronda representa un 98,9% del total del espacio de los factores
5. Factor 5:
  - a. bomba\_sitio representa un 99,4% del total del espacio de los factores
6. Factor 6:
  - a. eq\_ganador representa un 99,7% del total del espacio de los factores
7. Factor 7:
  - a. mapa representa un 98,9% del total del espacio de los factores

Al reducir los factores, y con el gráfico de sedimentación se puede observar que los primeros factores tienen diferencia con los otros, sin embargo, es necesario analizar detalladamente las variables de entrada para confirmar la reducción, teniendo en cuenta el efecto de cada variable en el estudio. arm\_dmg y vic\_side que no han arrojado una varianza explicativa significativa, son variables que se podrían considerar en un futuro tal vez, ya que para el objetivo del proyecto hay que determinar si efectivamente eliminó al enemigo son variables de entrada que pueden aportar un valor asociado a la solución de la problemática..

## 7.2 Observaciones

El estudio del análisis factorial genera alternativas según criterios generando distintos factores haciendo aceptable la cantidad a tomar pero se debe satisfacer la varianza cerca del 80% por esa razón se agregaron los 3 factores más cercanos para obtener un 87% de varianza explicada utilizando el método Varimax para la rotación.











## 8. Transformación

Tratamiento preliminar de los datos, transformación y generación de nuevas variables a partir de las ya existentes con una estructura de datos apropiada. Aquí se realizan operaciones de

agregación o normalización, consolidando los datos de una forma necesaria para la fase siguiente que es el modelado.

## 8.1 Entrenamiento de datos

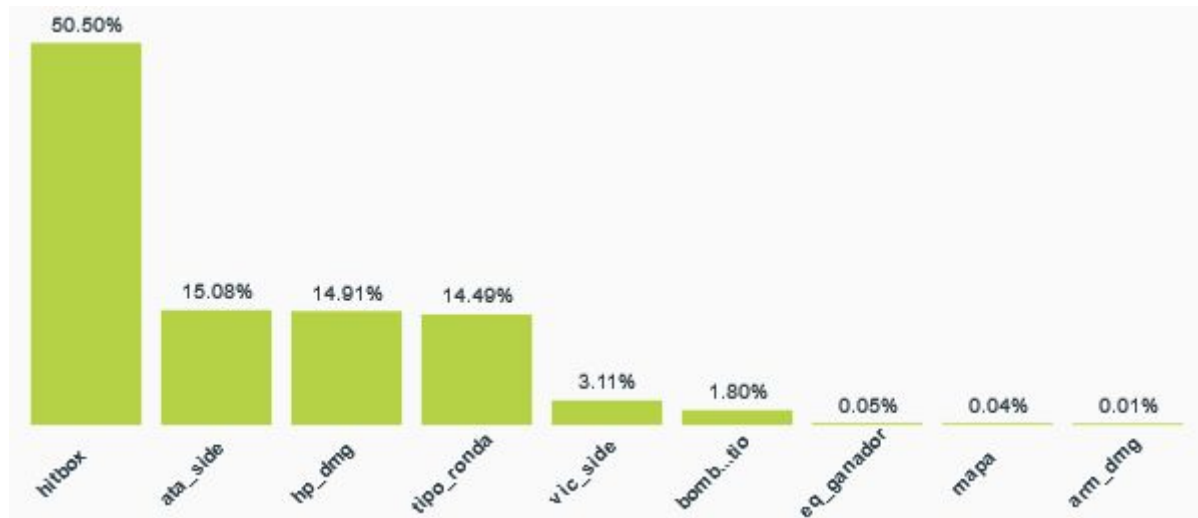
Entrenamiento aleatorio para los datos sin considerar el análisis factorial, considerando las 10 variables en estudio:

Name	Type	Count	Missing	Errors	Histogram
mapa	123	20,387	0	0	
ata_side	123	20,387	0	0	
vic_side	123	20,387	0	0	
hp_dmg	123	20,387	0	0	
arm_dmg	123	20,387	0	0	
bomba_sitio	123	20,387	0	0	
hitbox	123	20,387	0	0	
eq_ganador	123	20,387	0	0	
tipo_ronda	123	20,387	0	0	
bonus	123	20,387	0	0	

Fuente: Elaboración propia

Al obtener modelo el cual genera un árbol de decisión. En dónde se clasifica por las ramas más fuerte.

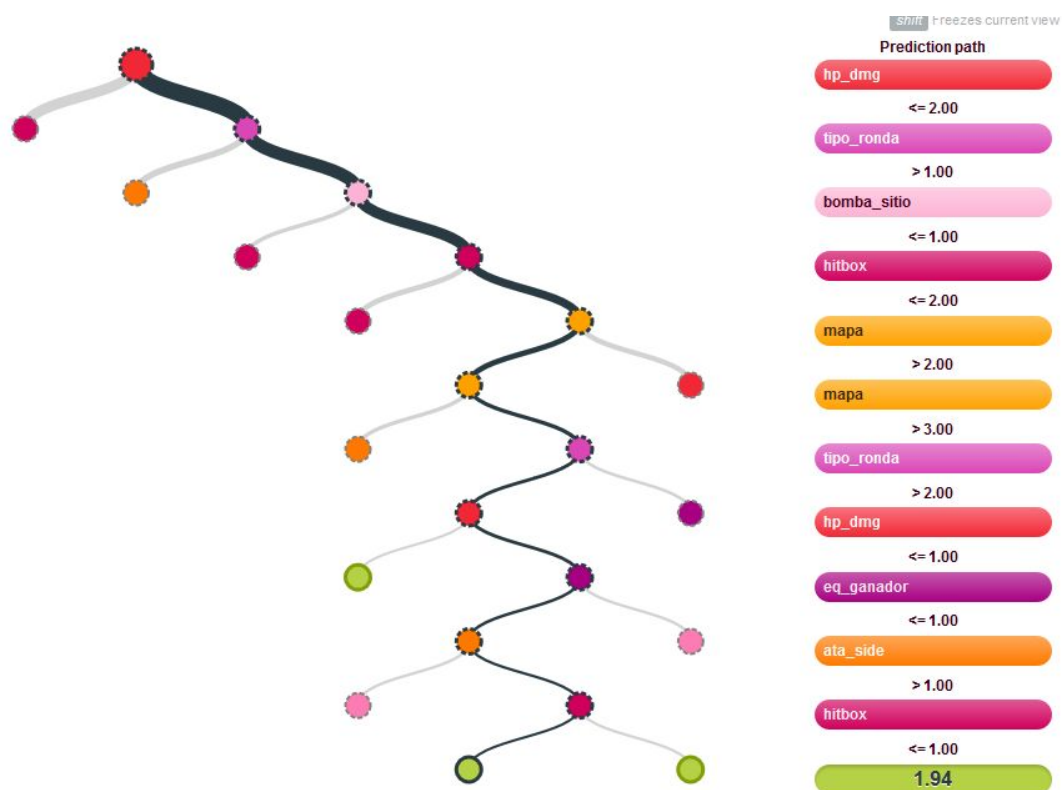




Fuente: Elaboración propia

Arbol de decisiones el cual predice la mejor ruta tomando como referencia los mayores porcentajes a la toma de una decisión.

Se puede observar las variables predictoras que tienen más pesos en la predicción. Como se puede observar la variable que tiene más peso en el modelo es la variable hp\_dmg con 15,48% y tipo\_ronda con 14,18% las cuales tienen mayor peso.



Fuente: Elaboración propia Bigml



El árbol de decisión tiene una confianza de 93% (dataset con análisis factorial). Tomando lo planteado en la hipótesis es posible establecer el bonus dependiendo de las variables

## 8.2 Modelo de Naive Bayes

En teoría la probabilidad y minería de datos, un clasificador bayesiano ingenuo es un clasificador probabilístico fundamenta en el teorema de Bayes y algunas hipótesis simplificadoras adicionales.

```
data = pd.read_csv("dataset_v6.csv",header=0)

data['split'] = np.random.randn(data.shape[0],1)

msk = np.random.rand(len(data)) <=0.8

train = data[msk]
test = data[~msk]
prediction_var=['mapa','ata_side','vic_side','hp_dmg','arm_dmg','bomba_sitio','hitbox','eq_ganador','tipo_ronda','bonus']

x_train = train[prediction_var]
y_train = train.Loan_Status

x_test = test[prediction_var]
y_test = test.Loan_Status

modelNB = GaussianNB()
modelNB_scores = cross_val_score(modelNB, x_train, y_train, cv=10, scoring='accuracy')
modelNB.fit(x_train, y_train)
print(modelNB_scores.mean())

predicted = modelNB.predict(x_test)

Output:

0.9125137762
```

Con el modelo Naive Bayes se obtiene una acertividad del 91%.

## 9. Conclusiones

El modelo que ha obtenido un mayor puntaje es el Decision Tree el cual tiene un 93% del puntaje en la predicción lo cual permite ver que es el mejor modelo para predecir.

Se concluye que el modelo más acertivo para este caso es el modelo de arboles obtenido en Bigml tanto que el sistema actual tiene un gran porcentaje de aceptación. Así también es mejorar el porcentaje asertivo generando diferentes iteraciones y mejorar el modelo limpiando los histogramas de manera correcta y probar nuevas interacciones.