

# Sistema de inferencia difusa: Asistente Gran DT

Emiliano David Gorr

G-4584/5

UNR

14 de marzo de 2014

## Resumen

El juego online Gran DT consiste en seleccionar 15 jugadores (11 titulares y 4 suplentes) de la primera división del fútbol profesional argentino. Estos jugadores obtienen un puntaje en función a su desempeño durante el partido, además suman y/o restan puntos extras según sus intervenciones en el partido, como pueden ser: convertir un gol (intervención positiva) o recibir tarjeta amarilla (intervención negativa). Finalizada cada fecha del torneo argentino, el usuario que más puntos sumo con sus jugadores, es el ganador de la fecha. Antes de cada fecha, a cada usuario se le permite realizar un máximo de 5 cambios en su equipo.

Este trabajo intenta, mediante un sistema de reglas de lógica difusa, estimar cuales son los jugadores que van a obtener un buen puntaje en la siguiente fecha del torneo, para incluirlos en el equipo.

Para realizarlo, teniendo las estadísticas de cada jugador hasta la fecha anterior (guardadas en una base de datos propia) obtiene las variables para cada uno (promedio del puntaje de clarín, promedio de gol, etc), calcula la pertenencia de las variables en los conjuntos borrosos correspondientes y realiza la inferencia. Posteriormente, mediante la defuzzificación se obtiene un índice para cada jugador. El que tenga mayor índice, es el jugador más recomendado para la siguiente fecha.

# 1. Motivación

El “Gran DT” es un juego online, donde cada usuario arma un equipo con jugadores del fútbol profesional, 11 titulares (1 arquero, 3 o 4 defensores, 3 o 4 medios y 2 o 3 delanteros) y 4 suplentes (uno por cada posición) que suman en caso de que alguno de los titulares no juegue. A partir de la 5ta fecha del torneo de primera división, empiezan a sumar puntos los jugadores del juego Gran DT.

Finalizada cada fecha, el usuario que sumo más puntos con sus 11 jugadores, es el ganador de la fecha y recibe un premio económico. Además, también hay un ranking general que acumula los puntos obtenidos durante todo el torneo (de la fecha 5 hasta la 19), y el usuario ganador recibe un premio económico mucho mayor.

¿Cuál es el puntaje de cada jugador?, durante una fecha cada jugador recibe un puntaje (de 0 a 10) según su desempeño en el partido, este puntaje es subjetivo y lo pone el periodista de clarín encargado de cubrir el partido. También puede sumar puntos extras por meter un gol (dependiendo de la posición, varía el puntaje de un gol), por ser figura (4 puntos), en caso de los defensores y arqueros pueden sumar 2 o 3 puntos respectivamente si durante el partido no les convierten ningún gol. Y también pueden restar puntos, por recibir amarilla (-2), roja (-4), por errar un penal (-4), por meterse un gol en contra (-2) y los arqueros, por cada gol recibido (-1).

Desde que comenzó el juego, participe ininterrumpidamente de todos los torneos (dos por año). No solo administrando mi equipo, sino también el de mis hermanas y el de mi mamá. Si bien es algo que me gusta, lleva mucho tiempo administrar un equipo y mucho más 4.

Siempre tuve la intención de realizar un programa que me ayude con esa tarea. Y en lo posible, que me ayude a mejorar mi desempeño en el juego.

Cuando dimos en la materia el tema Lógica Difusa, me di cuenta que era una herramienta que me podía servir mucho en la realización del asistente que quería hacer. Especialmente porque es perfecta para emular el razonamiento que yo hago cada vez que busco un cambio en mi equipo.

Por ejemplo, cuando busco un delantero pienso: “este delantero generalmente recibe un buen puntaje (6 o 7) y tiene buen promedio de gol, pero juega contra este equipo al le meten muy pocos goles, no me conviene, en cambio este otro delantero no suele obtener un gran puntaje (5 o 6) y tiene un promedio de gol medio o aceptable, pero juega contra un equipo que recibe muchos goles por partido, es más probable que meta un gol y así me sume más puntos. Me conviene este segundo delantero”

Los conjuntos difusos son la mejor manera para representar las variables que uso (por ejemplo, promedio de gol) para determinar si un jugador conviene para mi equipo o no.

Además, realizando un asistente con este sistema, permite que se aplique el razonamiento a todos los jugadores, cosa que manual o mentalmente no podría hacerlo una persona.

Todo lo anterior demuestra que la lógica difusa es una herramienta perfecta para hacer el asistente. Por eso, decidí realizar este trabajo.

## 2. Objetivo

El objetivo de este trabajo es realizar un sistema de inferencia difusa que permita estimar cuales son los jugadores que tienen más probabilidad de obtener un buen puntaje la siguiente fecha del Gran DT.

## 3. Métodos

### 3.1. Los datos

El primer punto del trabajo es el almacenamiento de los datos. Estos datos se obtienen de la ficha del partido.

La ficha del partido, es un recuadro que sale en el diario clarín al día siguiente del partido, y contiene los puntajes de cada jugador y las incidencias del partido: goles (quienes los hicieron), cambios, amonestados, expulsados y quien fue la figura.

Para almacenarlos decidí hacer una base de datos con varias tablas: Jugadores (contiene el id, nombre, equipo, posición, y por cada fecha el puntaje de 0 a 10 que le puso el periodista del diario clarín), Amonestados, Expulsados, Figuras, Partidos, Goles.

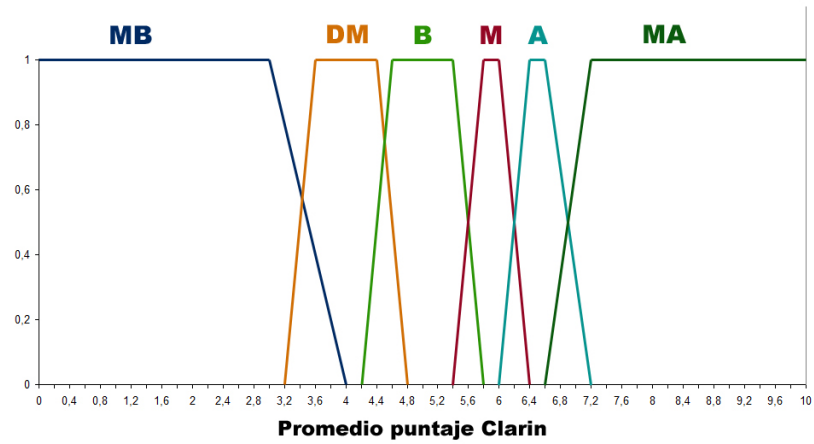
Al momento de la implementación (realice el programa en Python), una duda que surgió era como leer desde el programa esos datos. Una forma era mediante el acceso sql normal. La desventaja era que, si tenía la intención de distribuir el programa, necesitaba que esa base de datos este en un servidor publico y que todos tengan acceso a ella. Una mejor opción, fue exportar la base de datos a un archivo json, que es un formato de intercambio de datos, y después con un modulo en Python leer esos datos y guardarlos en arreglos para luego utilizarlos. La ventaja de usar este método, es que en caso de querer distribuir el programa, solo yo tengo acceso a la base de datos, y puedo colgar en la web el archivo json y hacer que el programa lea los datos contándose directamente a la url del archivo.

### 3.2. Conjuntos difusos

El sistema de inferencia difusa va a estar dividida según la posición de los jugadores. Claramente no es posible comparar jugadores de distinta posición, por un lado porque no es lógico comparar el promedio de gol de un delantero con el promedio de gol de un defensor, y por otro lado, porque los jugadores suman distinto según su posición: el gol de un delantero vale 4 mientras que el de un defensor vale 9, el defensor suma según su juego defensivo (valla invicta suma 2 puntos) mientras que los delanteros y medios solo suman por juego ofensivo.

Sin embargo, hay un conjunto difuso que es general a todos los jugadores. Es el promedio del puntaje clarín (este es el puntaje de 0 a 10 que pone el periodista del diario clarín en cada partido).

Así, tenemos el primer conjunto difuso: **PC (Promedio puntaje Clarín)**. Cuyas etiquetas son: DM = Demasiado Bajo; MB = Muy bajo; B = Bajo; M = Medio; A = Alto; MA = Muy alto

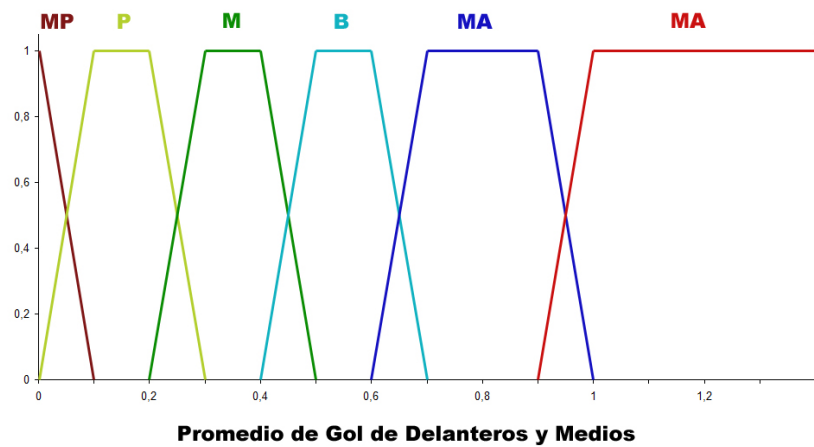


A continuación, los conjuntos difusos particulares para cada posición.

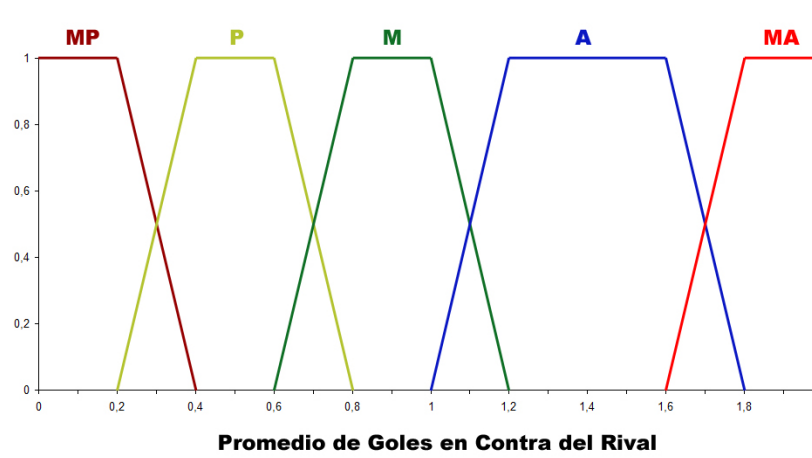
### 3.2.1. Conjuntos difusos para delanteros y medios

**Jug.PG (Promedio de Gol de Delantero y Medio)**

Etiquetas: MP = Muy poco; P = Poco; M = Medio; B = Bueno; A = Alto, MA = Muy Alto

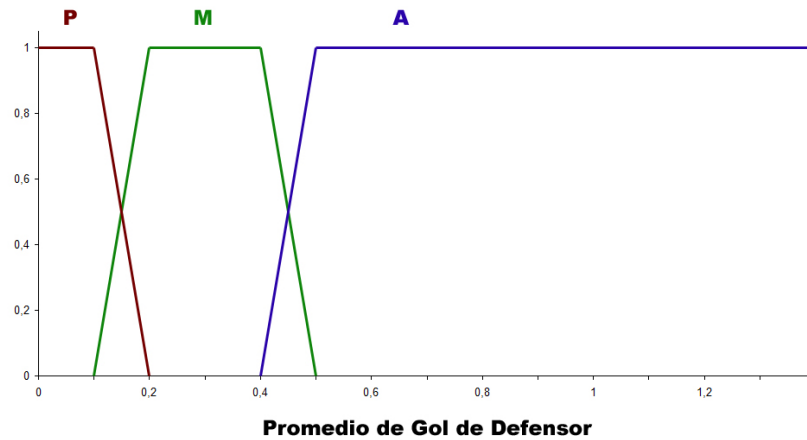


**PGCR (Promedio de Goles en Contra del Rival)** Es decir, el promedio de goles que recibe el próximo rival por partido MP = Muy poco; P = Poco; M = Medio; A = Alto; MA = Muy alto



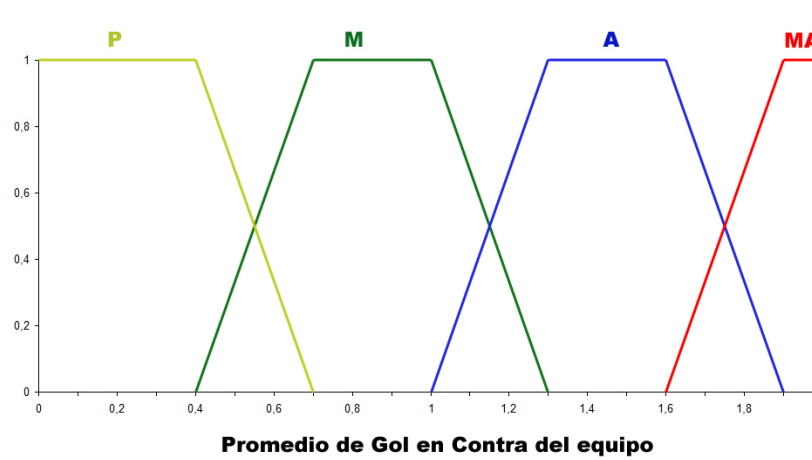
### 3.2.2. Conjuntos difusos para defensores y arqueros

**Def\_PG (Promedio de Gol de Defensor)** P = Poco; M= Medio; A = Alto

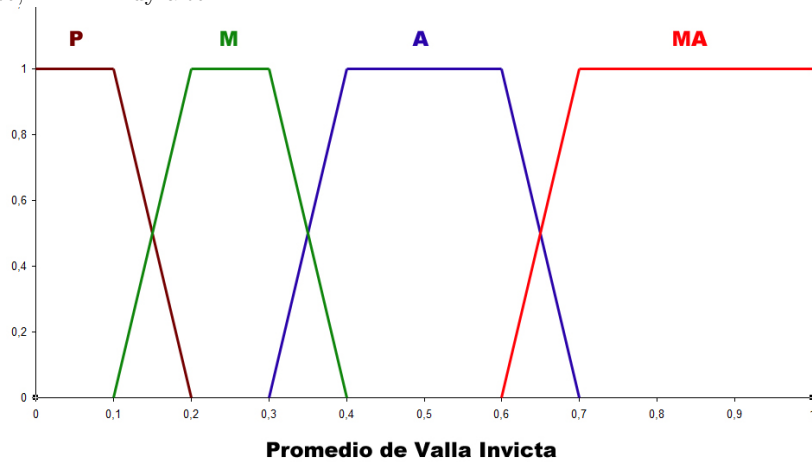


Este último conjunto difuso, es un claro ejemplo de porque la comparación se hace entre jugadores de la misma posición. No sería razonable comparar el promedio de gol de un defensor con el promedio de gol de un delantero.

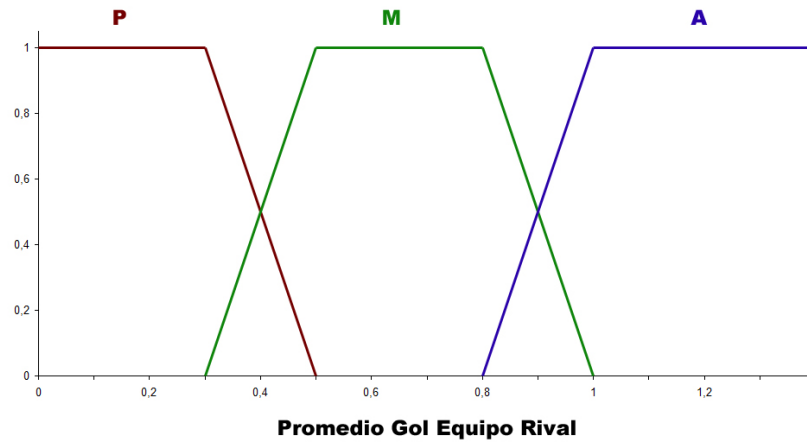
**PGC (Promedio de Gol en Contra del equipo)** Esta variable hace referencia al promedio de goles por partido que recibe su equipo, no a goles que marca en su propio arco.  
P = Poco; M = Medio; A = Alto; MA = Muy alto



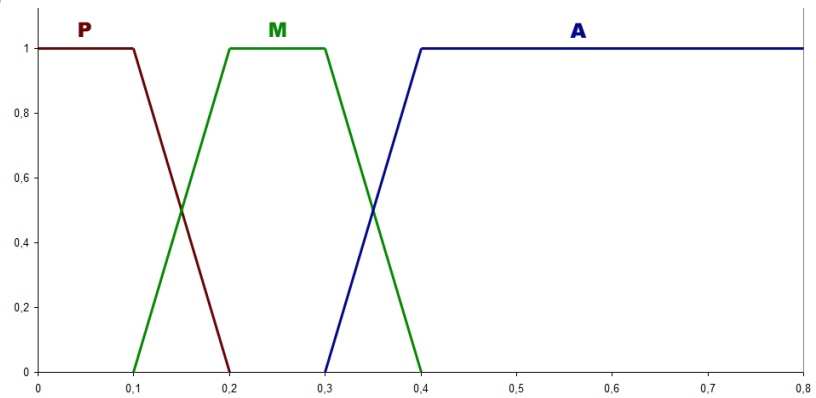
**PVI (Promedio de Valla Invicta)** P = Poco; M = Medio; A = Alto; MA = Muy alto



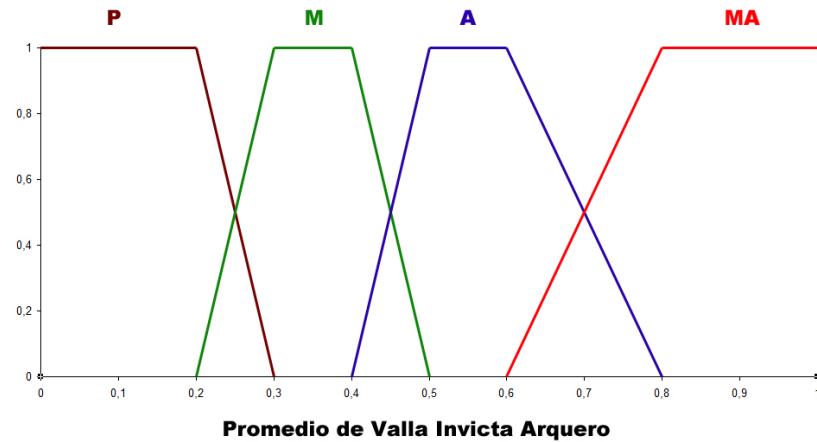
**PGER (Promedio Gol Equipo Rival)** P = Poco; M= Medio; A = Alto



**PTA (Promedio Tarjetas Amarillas)** P = Poco; M= Medio; A = Alto



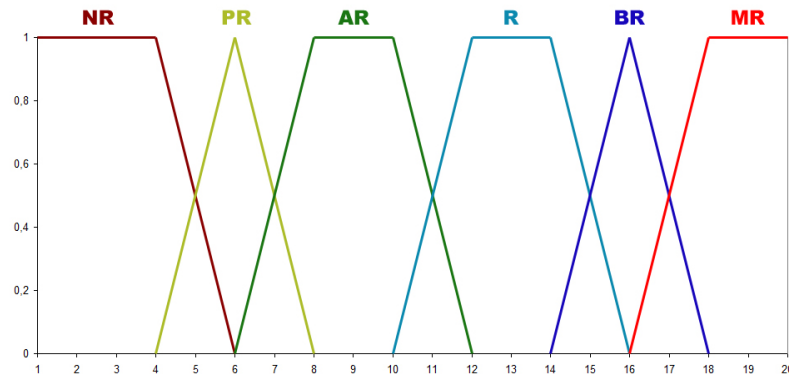
**PVIA (Promedio Valla Invicta Arquero)** P = Poco; M= Medio;  
A = Alto; MA = Muy Alto



### 3.2.3. Variable de Control

Finalizado el proceso de inferencia, voy a obtener como salida otro conjunto difuso:

**VC (Variable de Control)** NR = No recomendado, PR = Poco Recomendado, AR = Algo recomendado, R = Recomendado, BR = Bastante Recomendado, MR = Muy Recomendado



La posterior defuzzificación del resultado, me dará un índice de entre 0 y 20, que servirá para comparar los jugadores y así saber cual es el más recomendable.



### 3.2.4. Representación en Python

Para representar los conjuntos difusos en Python, elegí utilizar listas de tuplas de 5 elementos, de la forma (etiqueta,vii,vsi,vsd,vid) donde:

vii = vértice inferior izquierdo del trapecio. Coordenada( vii, 0)  
vsi = vértice superior izquierdo del trapecio. Coordenada( vsi, 1)  
vsd = vértice superior derecho del trapecio. Coordenada( vsd, 1)  
vid = vértice inferior derecho del trapecio. Coordenada( vid, 0)

Para ejemplificar, Promedio puntaje de Clarin, esta definido como:  
PC= [(DM,0,0,3,4),(MB,3.2,3.6,4.4,4.8),(B,4.2,4.6,5.4,5.8),(M,5.4,5.8,6.6,6.4),  
(A,6.6,6.6,7.2),(MA,6.6,7.2,10,10)]

## 3.3. Inferencia

### 3.3.1. Reglas

Por la cantidad de variables de entrada, el numero de reglas es elevado. Por lo tanto, en este informe voy a colocar solo algunas a modo de ejemplo.

Como mencione anteriormente, el puntaje de clarín que reciben los jugadores es un puntaje subjetivo. Es la calificación que el periodista del diario le otorga al jugador por el desempeño en el partido según su punto de vista.

Por mi experiencia en el juego, aprendí que hay jugadores que obtienen buen puntaje de forma regular. Por eso, el promedio del puntaje de clarin, tiene un peso bastante fuerte en el proceso de inferencia.

### Delanteros y Medios

Para determinar si un delantero o un medio es recomendable, me intereso en 3 cosas: Promedio de puntaje Clarin(PC), Promedio de Gol(PG), Promedio de Goles en Contra del Equipo Rival(PGCER).

La diferencia entre esas dos posiciones, es que en las reglas de delanteros, tiene más peso el promedio de gol que en las reglas de los medios. Mientras que el promedio de puntaje Clarin es más importante en los medios que en los delanteros.

Algunas reglas de ejemplo para delanteros:

Si el PC es Muy Alto(MA) y el PG es Alto(A) y el PGCER es Muy Alto (MA) o Alto(A) o Medio(M); entonces el delantero es Muy Recomendable(MR)

Si el PC es MA y el PG es Medio(M) y el PGCER es Poco(P); entonces el delantero es Bastante Recomendable(BR)

Si el PC es Alto(A) y el PG es Alto(A) y el PGCER es Poco(P); entonces el delantero es Bastante Recomendable (BR)

Si el PC es Alto(A) y el PG es Alto(A) y el PGCER es Muy Poco (MP); entonces el delantero es Recomendable (R)  
 Si el PC es Alto(A) y el PG es Medio y el PGCER es Muy Poco; entonces el delantero es Algo Recomendable (AR)  
 Si el PC es Alto(A) y el PG es Poco(P) y el PGCER es Muy Alto(MA); entonces el delantero es Bastante Recomendable (BR)  
 Si el PC es Bueno(B) y el PG es Poco(P) y el PGCER es Medio(M); entonces el delantero es Poco Recomendable(PR)  
 Si el PC es Bueno(B) y el PG es Poco(P) y el PGCER es (P); entonces el delantero es No Recomendable (NR)

Estas son solo algunas reglas a modo de ejemplo, la lista completa es muchísimo más amplia, no vale la pena incluirlas a todas en el informe porque serian varias hojas solo de reglas.

Al igual que para delanteros, solo voy a poner algunas reglas para los medios a modo de ejemplo:  
 Si el PC es MA y el PG es MA y el PGCER es A; entonces el mediocampista es Muy Recomendado  
 Si el PC es A y el PG es M y el PGCER es A; entonces el mediocampista es Recomendado  
 Si el PC es A y el PG es M y el PGCER es MA; entonces el mediocampista es Bastante Recomendado  
 Si el PC es MA y el PG es MA y el PGCER es A; entonces el mediocampista es Muy Recomendado  
 Si el PC es M y el PG es P y el PGCER es A; entonces el mediocampista es Algo Recomendado

### Defensores

El caso de los defensores es el más difícil de todos, tiene muchas variables de entrada: Promedio del puntaje de Clarín (PC), el Promedio de Gol (PG), el promedio de goles en contra que recibe por partido (PGC), el promedio de Valla Invicta (PVI), el promedio de goles a favor del equipo rival (PGER) y el promedio de tarjetas amarillas (PTA).  
 El defensor suma 9 puntos por cada gol que convierte, pero es muy poco frecuente que convierta un gol. Lo más probable es que sume puntos por valla invicta, es decir, que no le metan ningún gol durante el partido. Por lo tanto, en estas reglas, lo que se busca es que el promedio de valla invicta sea alto y el promedio de goles a favor del equipo rival sea poco.  
 Algunas reglas para ejemplificar:

Si el PC es MA y el PG es A y el PGC es P y el PVI es P y el PGER es M y el PTA es A; entonces el defensor es Muy Recomendado.  
 Si el PC es MA y el PG es A y el PGC es P y el PVI es P y el PGER es A y el PTA es P; entonces el defensor es Bastante Recomendado.  
 Si el PC es M y el PG es A y el PGC es M y el PVI es M y el PGER es P y el PTA es A; entonces el defensor es Bastante Recomendado.  
 Si el PC es A y el PG es A y el PGC es M y el PVI es M y el PGER es M y el PTA es P; entonces el defensor es Recomendado.

Si el PC es M y el PG es A y el PGC es A y el PVI es A y el PGER es M y el PTA es A; entonces el defensor es Algo Recomendado.

Al tener muchas más variables que en el caso de delanteros y medios, la cantidad de reglas para defensores es muchísimo mayor.

### Arqueros

Los Arqueros suman 3 puntos si no les convierten goles. Por lo tanto, para determinar cual es el arquero más recomendable me intereso en el Promedio del puntaje Clarin (PC), en el Promedio de Valla Invicta para Arquero (PVIA) y en el Promedio de Goles del Equipo Rival (PGER) Algunas reglas de ejemplo:

Si el PC es M y el PVIA es MA y el PGER es P; entonces el arquero es Muy Recomendado.

Si el PC es M y el PVIA es MA y el PGER es M; entonces el arquero es Bastante Recomendado.

Si el PC es M y el PVIA es MA y el PGER es A; entonces el arquero es Bastante Recomendado.

Si el PC es B y el PVIA es P y el PGER es P; entonces el arquero es Recomendado.

Si el PC es P y el PVIA es P y el PGER es M; entonces el arquero es Algo Recomendado.

## 3.4. El sistema en Python

Elegí hacer el programa en Python por la simplicidad que ofrece el lenguaje para programar.

Se puede trabajar fácilmente con listas, diccionarios, clases, funciones anónimas, etc.

Creé una clase **Jugador**, que contiene los datos que se usan para realizar el razonamiento. Esos datos, como mencione al principio del informe, los obtengo de un archivo json (resultado de exportar la base de datos a dicho formato). Con funciones sencillas calculo las variables necesarias, como son: promedio de puntaje clarín, promedio de gol, promedio de valla invicta, etc.

Luego, según la posición del jugador se aplica el razonamiento y se obtiene el índice de recomendación.

Para realizar la inferencia, con una función determino si un valor pertenece a un determinado conjunto difuso y cual es su nivel de pertenencia. Cuando determine el nivel de pertenencia de todos los valores de entrada, calculo cuales reglas se cumplen y en que grado.

Como T-norma utilice la intersección, es decir, cuando una regla se cumpla, el grado de pertenencia al conjunto difuso de la salida correspondiente será igual al mínimo entre todos los niveles de pertenencia de las variables de entrada.

Ejemplo para un Arquero:

Si el PC pertenece a la etiqueta M en 0,5, y el PVIA pertenece a la etiqueta MA en 1 y el PGER pertenece a la etiqueta A en 0,3; entonces, el arquero en Bastante Recomendable en grado 0,3. Este análisis se realiza para todas las reglas.

Una vez que se obtienen cuales son las reglas que se cumplen y en que grado lo hacen, se realiza la defuzzificación. Para la cual opte por el método Centros Promediados, que es un calculo parecido al de centro de masa, pero se ignora si hay solapamiento entre los conjuntos difusos, de esta manera el calculo es más sencillo y más rápido.

Finalmente, se muestran los resultados en una interfaz gráfica sencilla, que muestra cual es el equipo recomendado (1 arquero, 3 defensores, 4 medios y 3 delanteros). También se puede navegar por un menú para ver cuales son los 15 delanteros más recomendados, los 15 medios más recomendados, los 15 defensores más recomendados y los 5 arqueros más recomendados.

La interfaz gráfica la hice con el modulo TkInter de Python.

Equipo recomendado	Delanteros recomendados	Medios recomendados	Defensores recomendados	Arqueros recomendados
Equipo Recomendado:				
Jugador	POS	Equipo	Indice	Info
Marcelo Alberto Barovero	ARQ	River	17.1965252972	ver
Eder Alvarez Balanta	DEF	River	18.2981162981	ver
Paolo Duval Goltz	DEF	Lanus	14.8241201677	ver
Carlos Luciano Araujo	DEF	Lanus	13.8503937008	ver
Guido Hernán Pizarro	MED	Lanus	13.3735524949	ver
Fabian Vargas	MED	Independiente	13.0	ver
Matias Pérez García	MED	Tigre	13.0	ver
Federico Insúa	MED	Velez	13.0	ver
Ignacio Scocco	DEL	Newells	17.1965252972	ver
Silvio Ezequiel Romero	DEL	Lanus	14.5263157895	ver
Jonathan Yamil López	DEL	Rafaela	13.406779661	ver
Suplentes:				
Jugador	POS	Equipo	Indice	Info
Guillermo Sara	ARQ	Rafaela	16.0	ver
Carlos Roberto Izquierdoz	DEF	Lanus	12.8987341772	ver
Adrian Fernandez	MED	Independiente	13.0	ver
Mario Ignacio Regueiro	DEL	Lanus	11.3737288136	ver
Nota: La comparación de indices es entre jugadores de la misma posición, no usar el indice para comparar jugadores de distinta posición				

## 4. Discusión y Conclusiones

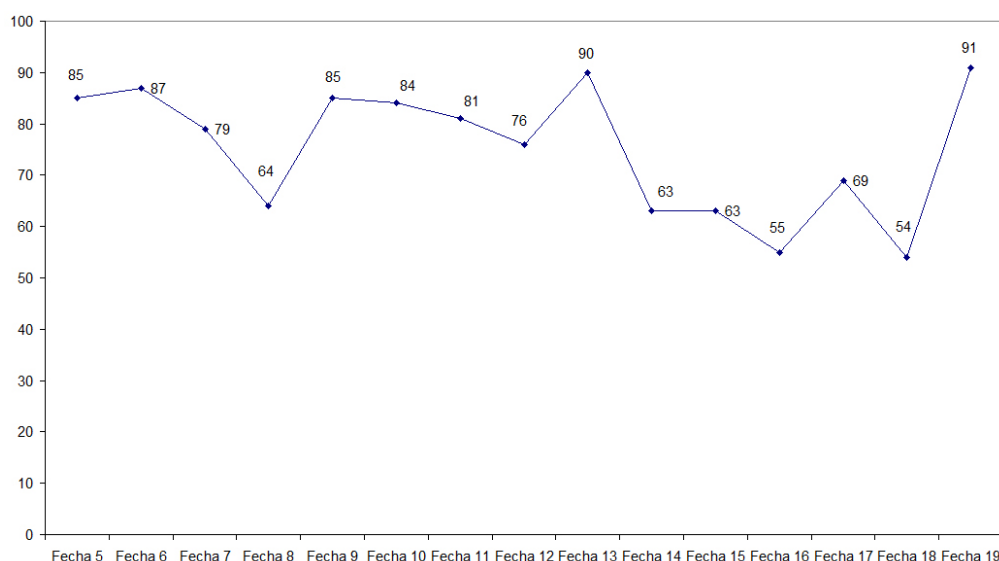
Antes de comenzar este trabajo, recolecte todos los datos del Torneo Final 2013 (correspondiente al torneo del primer semestre de 2013) para poder realizar un análisis del sistema.

Para realizar el análisis, calcule para cada fecha cuantos puntos hubiese sacado un equipo formado íntegramente por jugadores recomendados por el sistema.

El equipo lo forme de la siguiente forma: el arquero más recomendado, los 3 defensores más recomendados, los 4 medios más recomendados y los 3 delanteros más recomendados.

Muchas veces, hay jugadores recomendados por el sistema que luego no juegan, ya sea por lesión o por decisión técnica. Entonces, cada vez que un jugador del equipo en estudio no jugaba, lo remplazaba por el siguiente en el orden de recomendación.

Los puntos obtenidos fueron los siguientes:



Los usuarios ganadores de cada fecha, lo hicieron con un puntaje muy superior.

Estos fueron los puntajes ganadores:

Fecha 5	Fecha 6	Fecha 7	Fecha 8	Fecha 9	Fecha 10	Fecha 11	Fecha 12
136	141	132	145	151	128	152	142
Fecha 13	Fecha 14	Fecha 15	Fecha 16	Fecha 17	Fecha 18	Fecha 19	
142	140	156	126	150	134	144	

Podemos observar que los resultados obtenidos por el sistema difuso están lejos de ser los mejores de cada fecha.

Sin embargo, si observamos hasta la fecha 13, el sistema generaba equi-

pos que obtenían puntajes bastantes aceptables. El promedio de puntos de esas 9 fechas fue de: 81.2.

El ganador del torneo, es decir, el usuario que acumulo más puntos durante las 15 fechas (de la 5ta a la 19na) del torneo, obtuvo 1407 puntos que dan un promedio de 93.8 puntos por fecha.

Esto demuestra que el promedio de 81.2, que obtuvo el sistema en las primeras 9 fechas, era bastante bueno.

El problema surgió a partir de la fecha 14, quizás por ser la etapa definitoria del torneo hubo equipos que al no tener posibilidades de salir campeones jugaban sin motivación, o los jugadores que venían jugando bien les peso la presión, o hubo una merma física. O los equipos que jugaban instancias finales de algún otro torneo, como la copa Libertadores o la copa Argentina, decidieron rotar a sus mejores jugadores para que tengan descanso. O los equipos rivales, al conocer cuales eran los mejores jugadores del torneo, empezaron a diseñar tácticas para que esos jugadores no puedan desarrollar su juego.

Las alternativas son muchas, quizás el sistema no funciona correctamente y los buenos resultados de las primeras fechas fueron solo producto de la casualidad.

Para poder analizar correctamente el funcionamiento del sistema se necesitan evaluar varios torneos.

Por el momento, si bien los malos resultados de las ultimas fechas me desilusionan un poco, estoy conforme con el funcionamiento del sistema. Creo que pudo emular correctamente el razonamiento de un usuario medio.

## Referencias

Apunte *Lógica Borrosa*, Introducción Inteligencia Artificial, LCC, FCEIA, UNR