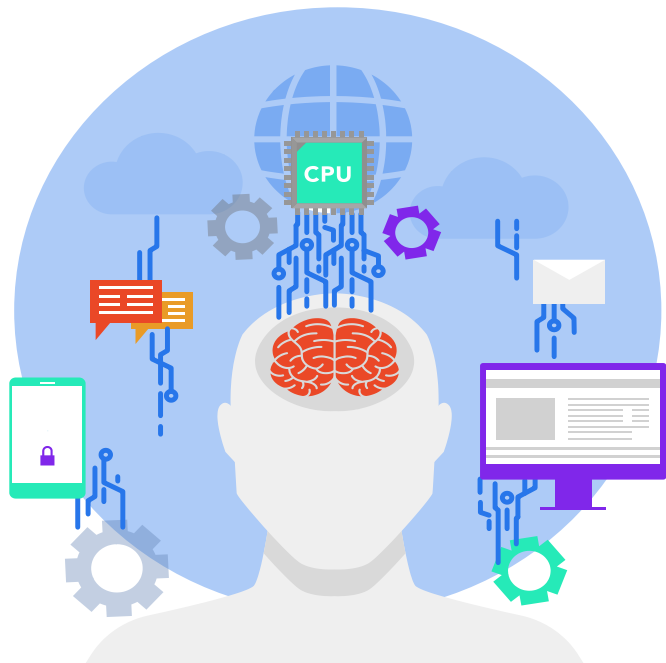
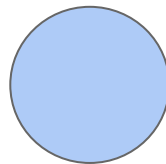




Universidad de Sevilla

**Máster Universitario en Ingeniería de Software: Cloud, Datos y
Gestión de TI**

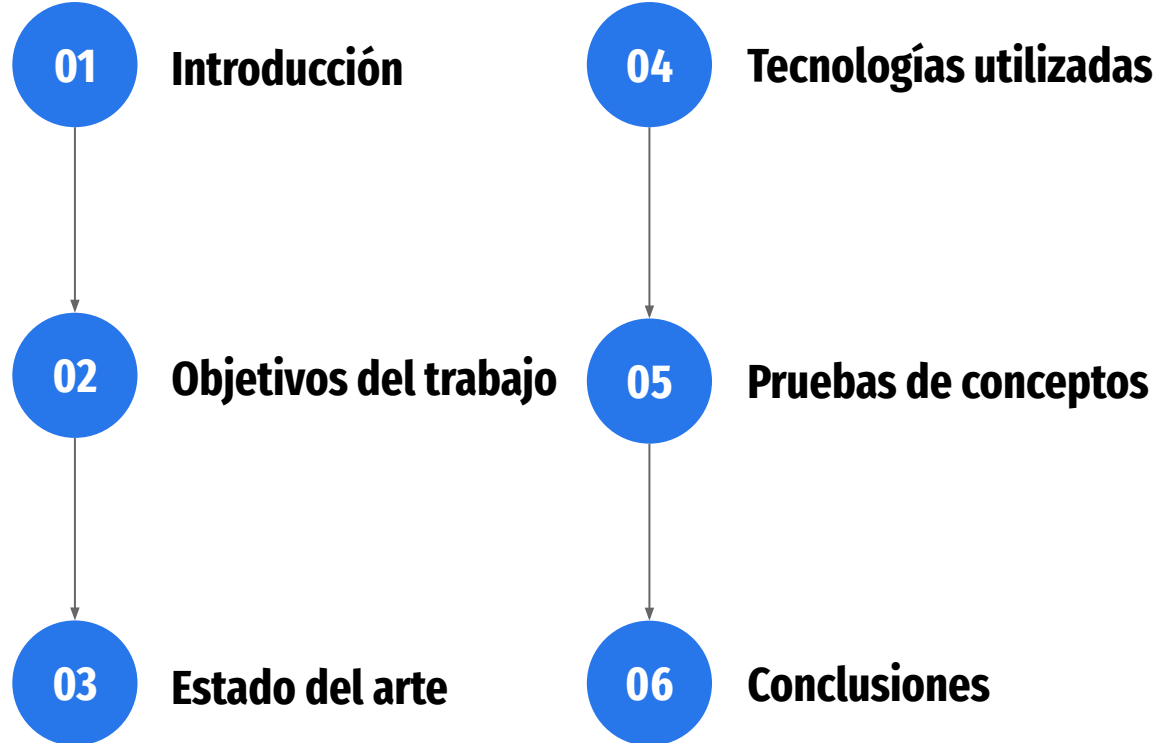
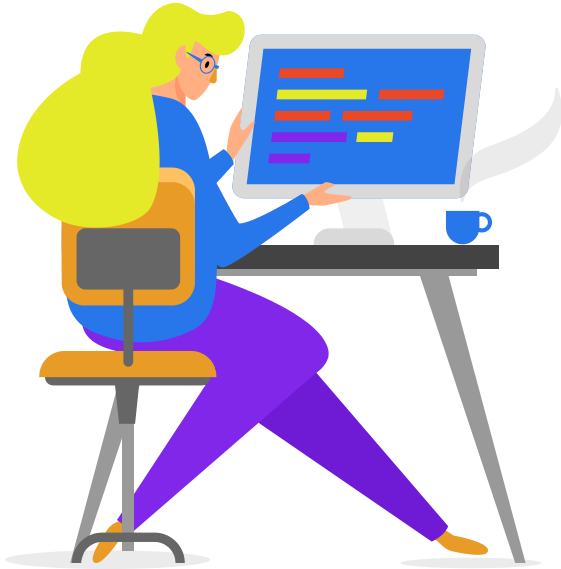


Análisis de datos deportivos con Machine Learning

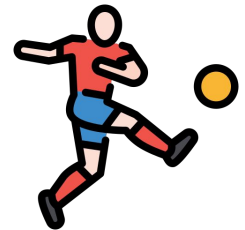
Realizado por: **Johnsiel Antonio Castaños Hernández**

Dirigido por: **José Antonio Troyano Jiménez**

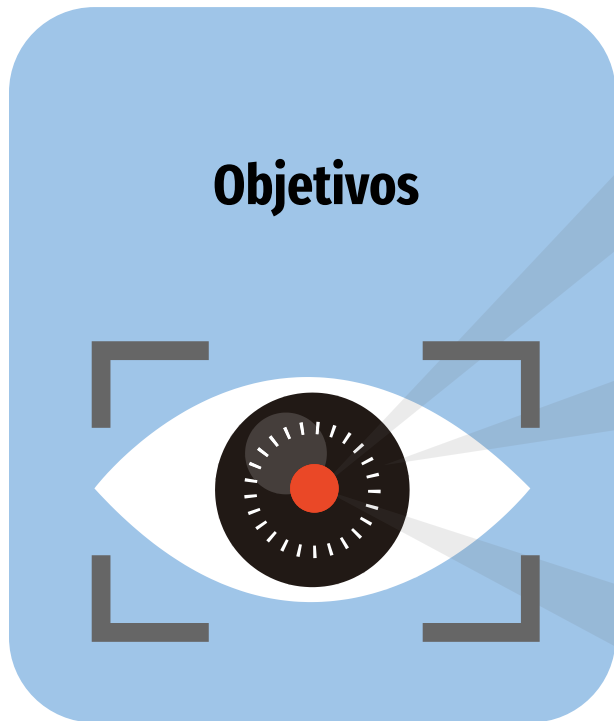
Tabla de contenido



Introducción



Objetivos del Trabajo de fin de máster

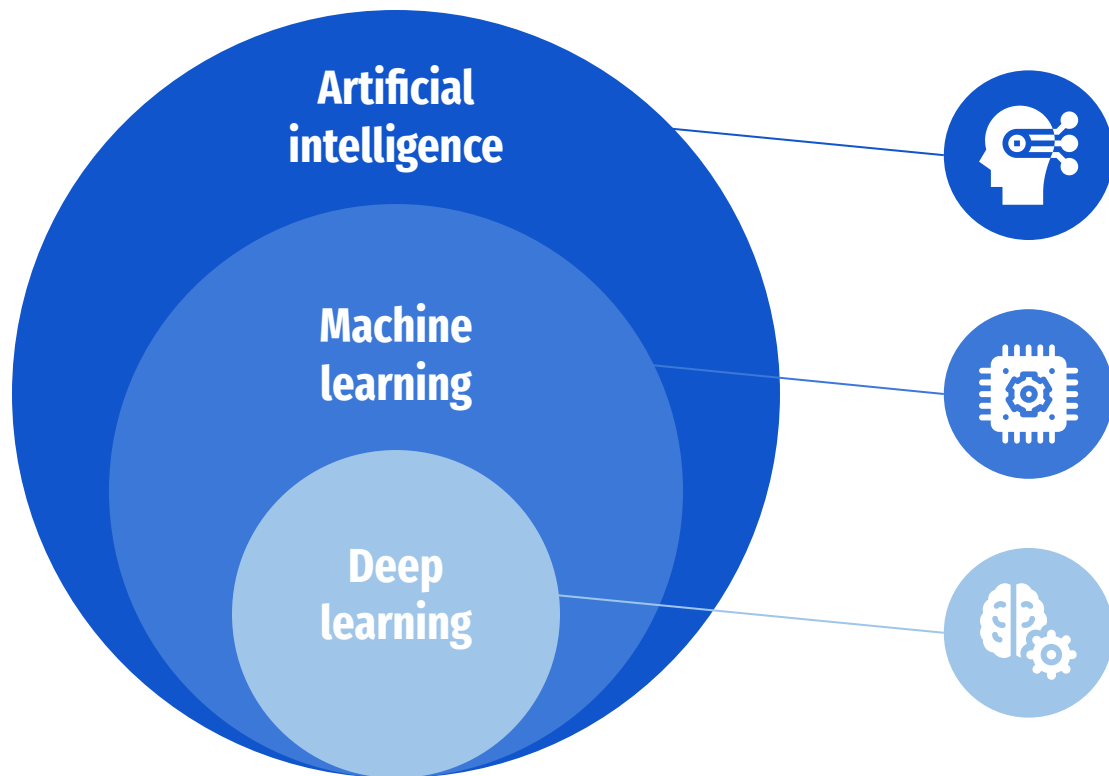


Conocer sobre fútbol

**Ampliar mis
conocimientos de ML**

**Implementar pruebas
de conceptos**

Estado del Arte



Casos de éxito

Sports Performance Platform

Características:

- Análisis deportivo.
- Visión del rendimiento.
- Predicción de lesiones.
- Y con los wearable: kilómetros recorridos, aceleración, etc.

Donde:

- Seattle Reign FC.
- Sport Lisboa.



Casos de éxito

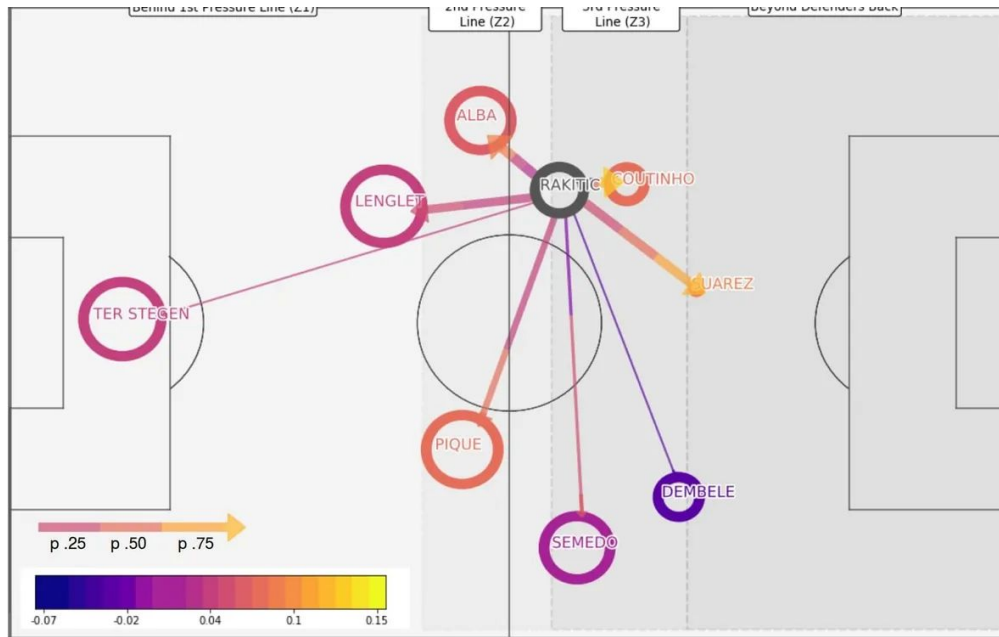
Principios de Johan Cruyff

Características:

- Calidad de la decisión tomada por los jugadores.
- Evaluación de riesgos.
- Beneficios de pases.

Donde:

- Barcelona FC.



Casos de éxito



Probabilidad de gol por video

Características:

- Análisis en tiempo real.
- Probabilidad de goles.

Donde:

- LaLiga Santander.

Tecnologías utilizadas

01



02



03



Pruebas de conceptos

01

Análisis de datos deportivos

02

Predicción de lesiones en atletas

03

Predicción de goles



Análisis de datos deportivos

Atributo	Descripción
id	Identificador único para el evento.
player_id	Identificador del jugador involucrado en el evento.
team_id	Identificador del equipo al que pertenece el jugador.
duration	Duración o tiempo de duración del evento.
timestamp	Marca de tiempo que indica cuándo ocurrió el evento.
type	Tipo de evento, como gol, pase, disparo, etc.
shot_type	Tipo de disparo, por ejemplo, con el pie o de cabeza.
shot_outcome	Resultado del disparo, como gol, bloqueado, desviado, etc.
shot_technique	Técnica utilizada para el disparo, como volea, cabezazo, etc.
second	Segundo en el que ocurrió el evento.
minute	Minuto en el que ocurrió el evento.
pass_type	Tipo de pase, como pase corto, largo, centro, etc.
pass_outcome	Resultado del pase, como exitoso, interceptado, etc.
location	Ubicación en el campo donde ocurrió el evento.
pass_end_location	Ubicación en el campo donde finalizó el pase.
pass_body_part	Parte del cuerpo utilizada para el pase, como pie, cabeza, etc.

Datos:

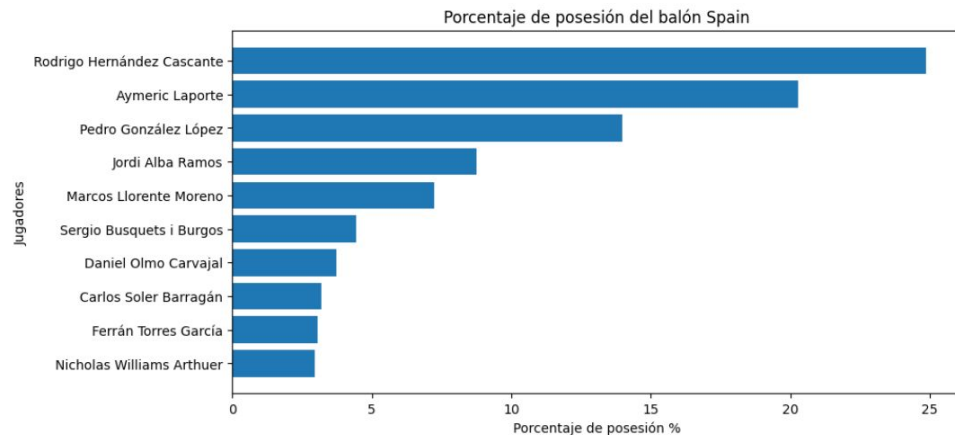
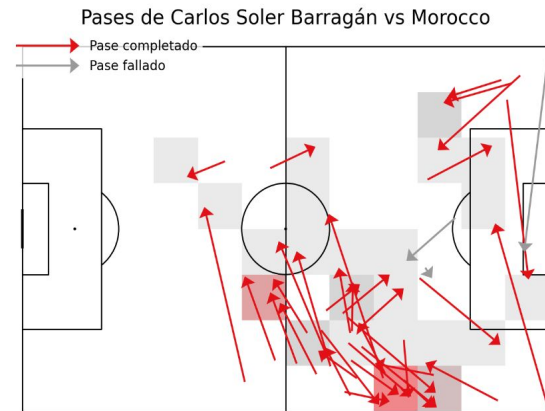
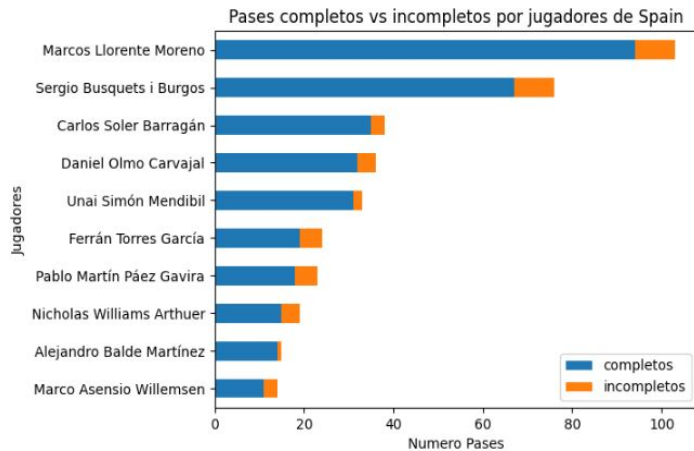
- StatsBombs.
- Cantidad de registros: 4750.
- Cantidad de columnas: 87.

Preprocesamiento:

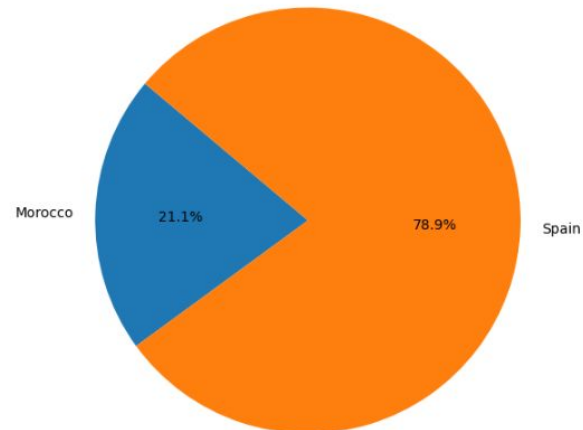
- Clasificación de los resultados de los pases.
- División de la columna de coordenadas en X y Y.
- Creación de nuevos atributos de coordenadas del fin del pases.

01

Análisis de datos deportivos

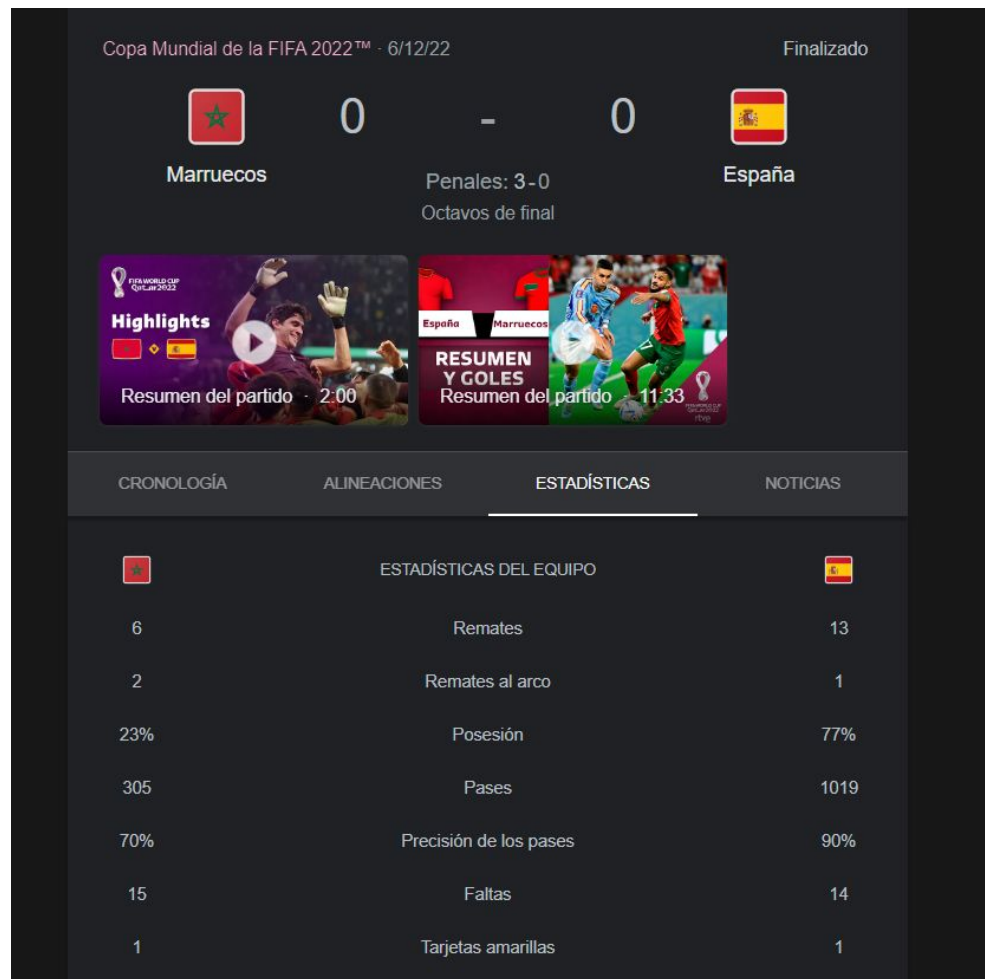


Porcentaje de Posesión del balón Spain vs Morocco



01

Análisis de datos deportivos



Predicción de goles

Datos:

- StatsBombs.
- Cantidad de registros: 16834.
- Cantidad de columnas: 14.
- 80% Entrenamiento.
- 20% Prueba.

Atributo
shotAerialWon
shotBodyPart
shotFirstTime
shotDeflected
shotOneOnOne
shotOpenGoal
shotOutcome
shotTechnique
shotType
playPattern
x
y
goal
Distance

Resultados de los modelos predictivos

Modelos:	Accuracy	recall	f1-score
K-Nearest Neighbors (KNN):	93%	93%	93%
Regresión Logística:	71%	71%	71%
Árbol de Decisiones:	94%	94,5%	94,5%
Soporte de Vectores (SVM):	75%	75,5%	75%

Preprocesamiento:

- Transformación de las columnas vacías a valores false.
- Aplicación de la codificación (encode) a las variables categóricas.
- Aplicación de oversampling.

03

Predicción de lesiones en atletas

Datos:

- Perfil de GitHub.
- Cantidad de registros: 2400.
- Cantidad de columnas: 7.
- 70% Entrenamiento y 30% Prueba.

Resultados de los modelos predictivos

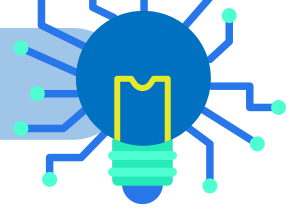
Modelos:	Accuracy	recall	f1-score
K-Nearest Neighbors (KNN):	96%	96.5%	96%
Regresión Logística:	70%	70.5%	70%
Árbol de Decisiones:	96%	95.5%	95.5%
Soporte de Vectores (SVM):	99%	99%	99%

Atributo
Injury
Athlete_ID
Date
Game Workload
Groin Squeeze
Hip Mobility
Rest Period

Preprocesamiento:

- Unión de diferentes datasets si es necesario.
- Creación de la columna de lesiones.
- Separación de las métricas..
- Cálculo de días de descanso de los atletas.
- Aplicación de oversampling.

Conclusiones

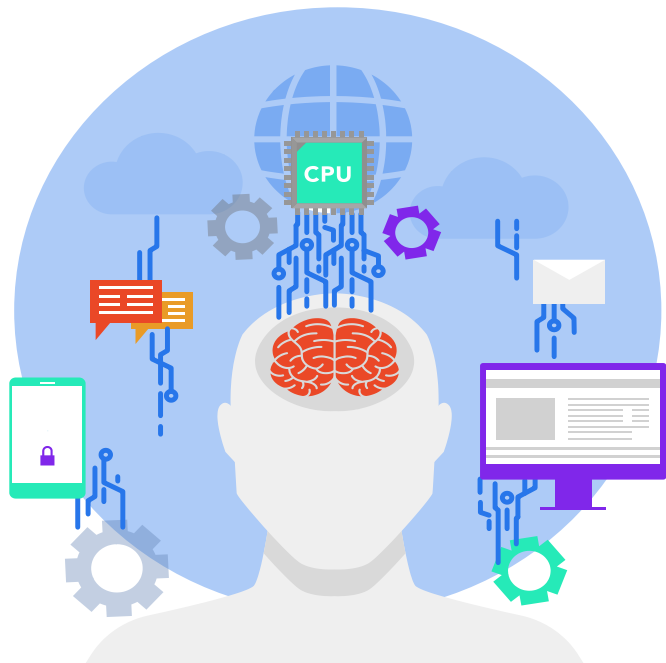
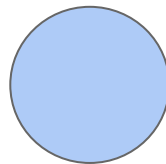


- En este trabajo se evidenció el alcance que puede tener el machine learning en el mundo de los deportes y específicamente del fútbol.
- Que el mundo del deporte está siendo potenciado por la ayuda de estas nuevas tecnologías que ayudan a la toma de decisiones.
- Que la elección correcta de modelos predictivo pueden ofrecernos resultados más eficientes que otros.



Universidad de Sevilla

**Máster Universitario en Ingeniería de Software: Cloud, Datos y
Gestión de TI**



Análisis de datos deportivos con Machine Learning

Realizado por: **Johnsiel Antonio Castaños Hernández**

Dirigido por: **José Antonio Troyano Jiménez**

Diagrama de Gantt

