



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

Clasificador de Aprendizaje Automático para determinar Acciones Clave en el
Desempeño Futbolístico de Jugadores

PROPUESTA DE TEMA DE MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL EN COMPUTACIÓN

CRISTIAN LILLO CIERO

MODALIDAD:
Doble Titulación con Magíster

PROFESOR GUÍA:
Javier Bustos

SANTIAGO DE CHILE
2024

1. Introducción

En la actualidad, el fútbol es un deporte que ha evolucionado en términos de tecnología y análisis de datos. Los clubes de fútbol han comenzado a utilizar herramientas tecnológicas para mejorar el rendimiento de sus jugadores y equipos. Sin embargo, en las etapas formativas, no se dispone de un modelo que permita evaluar el rendimiento de los futbolistas de manera objetiva. Esto conlleva a que los entrenadores deban evaluar el rendimiento de los jugadores en base a su experiencia y conocimiento del deporte, lo que puede llevar a sesgos y subjetividad.

En este contexto, el presente proyecto busca desarrollar un clasificador de aprendizaje automático que permita determinar las acciones clave en el desempeño futbolístico de los jugadores. Este clasificador permitirá a los entrenadores evaluar el rendimiento de manera objetiva y precisa. Para ello, se utilizarán técnicas de Machine Learning o Deep Learning para analizar los datos de los jugadores y determinar las acciones clave que influyen en su desempeño.

El desarrollo de este clasificador ayudará a identificar áreas de mejora y a tomar decisiones informadas en base a los datos. Además, se podrá utilizar en las etapas formativas de los clubes de fútbol para identificar y potenciar el talento desde edades tempranas, contribuyendo al desarrollo y formación de futbolistas.

2. Situación Actual

Guía (deshabilitar antes de entregar)

Discutir las soluciones o recursos existentes relacionados con el problema. Justificar por qué es necesario un trabajo novedoso.

(1 a 2 páginas)

Ejemplos de referencias:

- Conferencia: [1]
- Revista y Tesis: [2], [3]

3. Objetivos

Objetivo General

Se propone desarrollar un clasificador de aprendizaje automático que permita determinar las acciones clave en el desempeño futbolístico de los jugadores. Se utilizarán datos

provenientes de StatsBomb [4] para entrenar el clasificador y se evaluará comparándolo con las acciones que determinan el resultado de un partido de fútbol, como goles, asistencias, pases clave, entre otros. El objetivo es lograr un clasificador que sea capaz de identificar las acciones clave en el desempeño futbolístico de los jugadores con una precisión superior al 90%.

Objetivos Específicos

1. Obtener acciones realizadas por jugadores profesionales de fútbol. Estas deben estar en orden cronológico según la fecha del partido y el minuto de juego.
2. Entrenar distintos tipos de modelos, tanto de Machine Learning como de Deep Learning, para encontrar el que mejor se ajuste al problema y obtenga los mejores resultados.
3. Evaluar el clasificador con un conjunto de datos de prueba y verificar si se alcanza la precisión esperada.

Evaluación

Para evaluar el trabajo, se utilizará un conjunto de datos de prueba que no haya sido utilizado en el entrenamiento del clasificador. Se compararán las acciones identificadas por el clasificador con las acciones que influyen directamente el resultado de los partidos. La precisión del clasificador será la métrica principal para evaluar el trabajo, la cual deberá ser superior al 90% para considerar que el clasificador es efectivo.

4. Solución Propuesta

Se planea desarrollar el modelo en el lenguaje de Python, haciendo uso de librerías como scikit-learn, TensorFlow y Keras para el entrenamiento de los modelos de Machine Learning y Deep Learning. Como se mencionó previamente, una parte importante del trabajo consiste en determinar cuál tipo de modelo es el más óptimo para la resolución de la problemática. Para ello, se analizarán distintos modelos, como Random Forest (RF), Support Vector Machine (SVM), Redes Neuronales Convolucionales (CNN), entre otros, y se compararán los resultados obtenidos.

En relación a los datos, se dispone de una gran cantidad de archivos en formato JSON. Estos archivos están disponibles en el repositorio de GitHub de StatsBomb [4] y contienen información detallada de más de 3000 partidos. Los archivos se dividen en cuatro categorías: eventos, alineaciones, partidos y competiciones por temporada. Aunque el enfoque principal se centra en los eventos, se utilizarán los datos de las otras categorías para complementar la información y mejorar la calidad del clasificador.

Mediante el uso de librerías como Pandas y Matplotlib, se realizará un análisis exploratorio de los datos para identificar patrones y características relevantes. Se buscará deter-

minar objetivamente cuáles son las acciones que modifican el resultado de un partido de fútbol y que, por lo tanto, deben ser consideradas como acciones clave por parte del modelo. Este análisis permitirá seleccionar las variables más relevantes para el entrenamiento del clasificador y descartar aquellas que no aporten información significativa.

Se debe determinar además la mejor manera de representar los datos para el entrenamiento del modelo, pues debe ser capaz de identificar patrones y relaciones entre las acciones realizadas por los jugadores. Se considera que la representación de los datos es un aspecto fundamental para el éxito del clasificador, por lo que se dedicará especial atención a este aspecto.

5. Plan de Trabajo (Preliminar)

1. Definir la mejor alternativa para almacenar la masiva cantidad de datos de StatsBomb.
2. Realizar un procesamiento a los datos para obtener las acciones realizadas por cada jugador en todos sus partidos disputados.
3. Realizar un análisis exploratorio de los datos para identificar patrones y características relevantes.
4. Seleccionar las variables más relevantes para el entrenamiento del clasificador.
5. Definir la mejor manera de representar los datos para el entrenamiento del modelo.
6. Entrenar distintos modelos de Machine Learning y Deep Learning para determinar cuál es el más óptimo para el problema.
7. Evaluar el clasificador con un conjunto de datos de prueba y verificar si se alcanza la precisión esperada.
8. Escribir el informe final de la tesis.

Referencias

- [1] Robert M. Corless, David J. Jeffrey, y Donald E. Knuth, «A Sequence of Series for the Lambert W Function», en *International Symposium on Symbolic and Algebraic Computation*, pp. 197-204.
- [2] M. H. A. Newman y Alan M. Turing, «A Formal Theorem in Church's Theory of Types», *The Journal of Symbolic Logic*, vol. 7, pp. 28-33, [En línea]. Disponible en: <http://www.jstor.org/stable/2267552>
- [3] Alan M. Turing, «Systems of Logic Based on Ordinals».
- [4] «StatsBomb». [En línea]. Disponible en: <https://github.com/statsbomb/open-data>