



Introducción a Ciencia de Datos

Proyecto Aplicado:
Análisis de rendimiento del Campeonato Nacional de Fútbol
Chileno Profesional entre los años 2010-2020

Profesora:

Paula Aguirre

Ayudantes:

Vicente Agüero

Fernando de Diego

Alumnos:

Renato Aravena

Benjamín Jofré

Francisco Meza

Juan Manuel Zapata

CONTEXTO Y MOTIVACIÓN

La elección del tema de nuestro proyecto está ligada al fanatismo que tenemos cada uno de los integrantes por el fútbol, bajo esta premisa es que partimos discutiendo qué sería interesante analizar las múltiples estadísticas que se generan en torno al campeonato nacional de fútbol profesional chileno de primera división. Parte de nuestra investigación inicial sobre este tema nos llevó a buscar si la Asociación Nacional de Fútbol Profesional (ANFP) dentro de su página web proveía una base de datos relacionados a las estadísticas de los torneos, equipos, jugadores, etc, no obteniendo un resultado positivo ya que la única estadística que presentaba era la del torneo actual, sin desglose por equipos o jugadores. Nuestra segunda fuente que investigamos fue el Ministerio de Deporte (MINDEP), el cual no presentaba bases de datos al respecto.

Luego decidimos realizar una búsqueda en **Wikipedia** sobre cada año de torneo, fue aquí donde obtuvimos resultados positivos ya que este sitio presentaba varias tablas con información útil respecto al año de torneo que se requería y que podían ser obtenidas a través de web scraping. Como última fuente recurrimos al sitio web de **transfermarkt**, el cual provee información sobre valores de mercados de los equipos, valores de mercado de los jugadores, ingresos obtenidos por venta de jugadores de los equipos por año, costos obtenidos por la compra de jugadores de los equipos por año, etc. Una vez realizada la investigación inicial, nos dimos cuenta de que eran muchas las preguntas que eran posible intentar contestar a través del análisis de los datos encontrados, lo que nos reafirmó que la elección del tema fue acertada. El hecho de poder aplicar las herramientas y aprendizajes adquiridos en clase nos motivó a encontrar información relevante en este tema de gran interés para nosotros.

Creemos que el público que se beneficiaría de la información que puede entregar este proyecto son primeramente la ANFP, el MINDEP, los mismos equipos de fútbol y también los fanáticos del fútbol. El poder proveer información relevante sobre los

numerosos datos que genera el fútbol en Chile puede ser una motivación extra para la ANFP, por ejemplo, para generar una asociación con mayor transparencia.

OBJETIVOS

Los objetivos que nos propusimos son los siguientes:

1. Analizar y comprender las estadísticas descriptivas de desempeño por equipo por año de torneo.
2. Analizar y comprender la distribución por región de los equipos en permanencia en primera división.
3. Analizar la relación entre la popularidad de un equipo (medido en cantidad de hinchas) y su rendimiento en el torneo por año.
4. Analizar la relación del balance total por equipo en concepto de venta y compra de jugadores y su rendimiento en el torneo por año.
5. Analizar la relación entre ganancias netas por equipo y su rendimiento en el torneo por año.

Los beneficios que creemos podemos obtener mediante este proyecto es conocer las posibles correlaciones existentes entre los rendimientos de los equipos en los torneos por año con respecto a su popularidad, su balance total por concepto de venta y compra de jugadores y las ganancias netas obtenidas por conceptos de merchandising, auspicios, conceptos televisivos, etc.

Con esto, la audiencia objetivo sería como se especificó previamente la ANFP, MINDEP y los mismos equipos. Basados en estos objetivos es que intentaremos responder ciertas preguntas en base a los datos obtenidos.

DATOS

Los datos utilizados en este proyecto serán recolectados desde [Wikipedia](#) mediante web scraping para obtener las estadísticas de los torneos por año, los datos de los equipos, equipos por región, entre otros. Desde [transfermarkt](#) mediante recolección mecánica de los datos para la obtención de los balances totales por equipo en conceptos de compra y venta de jugadores y de los reportes financieros presentados por los equipos participantes de los torneos por año. Los datos a utilizar presentan variables numéricas discretas como lo son las estadísticas propias de los torneos, capacidad de los estadios, etc, también presentan variables numéricas continuas como lo son los reportes financieros de los equipos. También existirán variables categóricas como la agrupación de equipos por región.

PREGUNTAS DE INVESTIGACIÓN

1. ¿Existe correlación entre la popularidad de un equipo y su rendimiento en los torneos?
2. ¿Existe correlación entre las ganancias netas de un equipo y su rendimiento en los torneos?
3. ¿Cómo es la distribución de los equipos por región participantes de los torneos por año?
4. ¿Existe correlación entre ser un equipo de región y el rendimiento de dicho equipo?
5. ¿De qué regiones son los equipos con mejor rendimiento?
6. ¿Existe correlación entre el balance total en concepto de compra y venta de jugadores de los 3 mejores equipos y los 3 peores equipos de los torneos por años y su rendimiento en los torneos?
7. ¿Serán los equipos que más invierten los con mejor rendimiento?

DISEÑO TENTATIVO

Para el desarrollo de este proyecto se utilizara el lenguaje de programación Python y a través de las librerías pandas, geopandas, numpy, stats model, scikit-learn, matplotlib, seaborn, entre otras, junto con herramientas computacionales como el web scraping se espera poder responder las preguntas de investigación mediante gráficos y modelos estadísticos como la regresión lineal uni y multivariada. Se espera poder contar con el conocimiento teórico para poder desarrollar modelos predictivos simples que involucren aprendizaje supervisado para poder generar predicciones como por ejemplo, de que región serán los próximos campeones de los torneos que vienen.

REFERENCIAS

- Wikipedia : [Primera División de Chile 2010 - Wikipedia, la enciclopedia libre](#)
- Transfermarkt : [Primera División de Chile 2021 | Transfermarkt](#)