

Propuesta de Análisis Predictivo para la Optimización de la Distribución de Recursos en Programas de Apoyo a Deportistas de Alto Rendimiento

Paul Martin Ruiz Guardia
Data Engineer, Tecsup
Lima, Peru
paul.ruiz@tecsup.edu.pe

Abstract— Este artículo propone la construcción de un modelo analítico para mejorar la asignación de recursos en el Programa de Apoyo al Deportista (PAD) del Instituto Peruano del Deporte (IPD). A través del análisis de un vasto conjunto de datos históricos y actuales, se busca identificar las variables clave que influyen en el rendimiento de los deportistas de alto rendimiento. La hipótesis central es que un enfoque basado en datos puede optimizar el uso de fondos del PAD, lo que resulta en una mejor preparación y desempeño de los atletas. Se utilizarán técnicas avanzadas de visualización y machine learning para procesar y analizar los datos, con el fin de establecer patrones predictivos que guíen la distribución de los recursos. Se desarrollarán dos modelos de aprendizaje automático: uno para predecir el potencial de éxito de los deportistas y otro para optimizar la asignación de apoyos. Este enfoque no solo aumentará la transparencia y la efectividad del programa, sino que también permitirá un seguimiento más preciso del progreso de los deportistas.

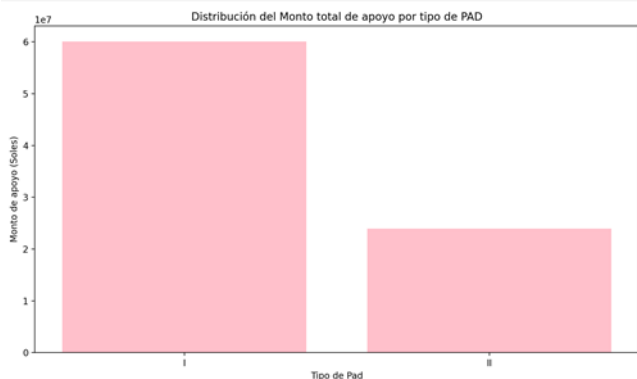
Keywords—Python, deportes, análisis de datos, estadística, machine learning, optimización de recursos, rendimiento deportivo.

I. INTRODUCCIÓN

En el universo del deporte peruano, donde el fervor competitivo y la aspiración al éxito definen el pulso de la nación, el Instituto Peruano del Deporte (IPD) se destaca como un faro de excelencia y apoyo. Su misión se materializa a través del Programa de Apoyo al Deportista (PAD), una red de soporte que entrelaza los hilos del esfuerzo con los del triunfo, apuntalando con sabiduría inversiones que alcanzan los 70 millones de soles. Dentro de esta suma, el PAD I absorbe la mayor parte con 50 millones, mientras que el PAD II recibe 20 millones, cifras que invocan tanto reconocimiento como interrogantes sobre su distribución y su impacto en el deporte. [1]

Figura 1

Análisis del Financiamiento por Tipo de Programa de Apoyo al Deportista (PAD)

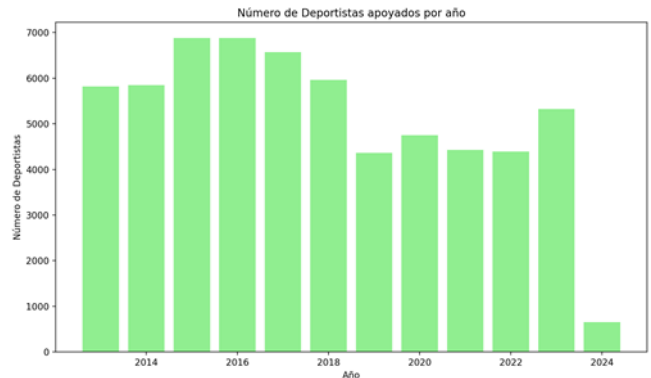


Fuente: Elaborado por el autor utilizando Python con las bibliotecas Matplotlib y Pandas, 2024. *Nota.* Elaboración propia (2024).

Los datos nos hablan de más de 6,500 atletas que han sentido el impulso de este programa en un solo año, una comunidad que se extiende y se diversifica con cada ciclo olímpico. No obstante, la magnitud de los números despierta una controversia sutil: ¿Se está fomentando la equidad entre todas las disciplinas deportivas? ¿Estamos potenciando talentos emergentes o reforzando los ya consolidados? Estas preguntas encuentran eco en los ecosistemas financieros del IPD y motivan un escrutinio meticuloso de cada sol invertido.[2]

Figura 2

Tendencias en el Apoyo a Deportistas por Año

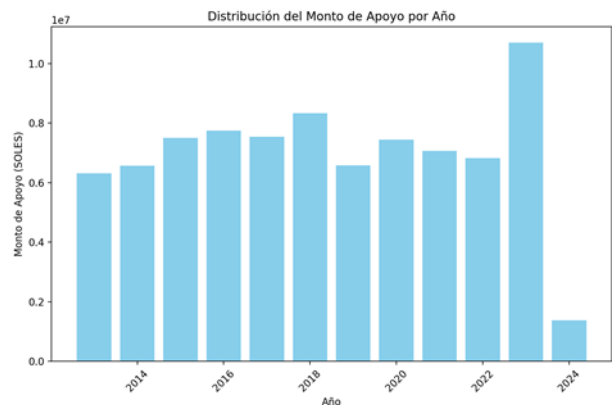


Fuente: Gráfico creado por el autor, basado en datos de los programas de apoyo al deportista. *Nota.* Elaborado con Python, usando Matplotlib y Pandas (2024).

La senda hacia la equidad y la eficiencia en el deporte es intrincada y sin garantías. La diversificación de las inversiones a lo largo de los años ha mostrado tendencias fluctuantes, con años como el 2019 marcando hitos de financiamiento superior a los 10 millones de soles, mientras otros, como el 2021, reflejan una contracción significativa a menos de la mitad de dicha cifra.

Figura 3

Distribución Anual del Monto de Apoyo Financiero a Deportistas.



Fuente: Este gráfico fue generado por el autor utilizando las bibliotecas Matplotlib y Pandas en Python. *Nota.* Basado en datos anuales del Programa de Apoyo a Deportistas (2024).

Es en este vaivén de cifras donde se plantea nuestra tesis: la ciencia de datos y los modelos predictivos poseen la llave maestra para desentrañar los secretos de una asignación de recursos óptima.

Con un análisis meticuloso de la data acumulada desde 2013 hasta la fecha, este estudio propone una evaluación transversal de los patrones de financiamiento y rendimiento. Se buscará no solo comprender la historia detrás de las cifras sino también proyectar el futuro de un deporte peruano más inclusivo y justo. Los resultados de esta investigación buscan empoderar a los tomadores de decisiones con la visión y la claridad necesarias para esculpir un legado de triunfo y orgullo nacional, uno que garantice que el próximo golpe, la próxima carrera, la próxima victoria, esté apoyada por un sistema que entiende de valor y no solo de valoraciones.

II. DEFINICIÓN DEL PROBLEMA

El problema a resolver radica en optimizar el Programa de Apoyo al Deportista (PAD) del Instituto Peruano del Deporte (IPD) para garantizar que la asignación de recursos económicos alcance su máximo potencial en el desarrollo y éxito de los deportistas de alto rendimiento. A través del análisis meticuloso de datos acumulados desde 2013, se busca descifrar el impacto real de las subvenciones otorgadas y las variables que definen el alto rendimiento. Se indagará sobre las cifras de financiación, buscando tendencias y correlaciones con el logro de medallas y la presencia en competiciones internacionales. Específicamente, se pretende analizar la eficacia del apoyo proporcionado, considerando factores como la federación deportiva, la disciplina practicada y la distribución geográfica de los atletas.

La investigación también profundizará en el perfil del apoyo económico del PAD, investigando si existen patrones que determinen la asignación de los montos: ¿Es la popularidad de una disciplina un factor determinante en la financiación que recibe? ¿Hay deportistas de alto rendimiento que, a pesar de recibir menos fondos, muestran resultados sobresalientes? De igual forma, examinaremos el fenómeno de los atletas que, si bien se benefician de las subvenciones, no logran sostener un desempeño constante, indagando en posibles causas y soluciones para mejorar la retención y continuidad.

Además, es crucial evaluar la estructura y la administración del PAD para determinar si el proceso es intuitivo y accesible para los atletas, buscando mejorar su experiencia y facilitar un acceso equitativo. La intención es crear una interfaz de apoyo que no solo sea efectiva, sino también amigable y estimulante para el desarrollo deportivo.

Por último, un componente esencial de esta iniciativa es fortalecer la formación en análisis de datos para los administradores del programa y las federaciones deportivas. Esto les permitirá comprender los patrones y tendencias dentro de los vastos datos del PAD, capacitándolos para tomar decisiones basadas en datos que benefician a los deportistas, optimicen los recursos y eleven el estándar de excelencia en el deporte peruano.

III. TRABAJOS RELACIONADOS

A. Proyecto del Laboratorio de Análisis y Estadísticas Deportivas de la Universidad de Virginia [3]

Desarrollo de un centro académico para la educación y la

investigación en analítica deportiva, que combina estadísticas avanzadas con análisis de rendimiento deportivo y prevención de lesiones.

B. Aplicaciones de IA en Análisis Deportivos de Moments Log [4]

Innovaciones en IA para análisis de rendimiento de jugadores, optimización de estrategias de equipo, reclutamiento y prevención de lesiones, transformando los deportes con análisis de datos de dispositivos portables y seguimiento en tiempo real.

C. Analítica Deportiva Avanzada de la Universidad de Berkeley [5]

El grupo de investigación de Berkeley ha desarrollado modelos avanzados de análisis deportivo que abarcan desde el rendimiento de jugadores hasta estrategias de equipo, utilizando la IA para predecir lesiones y optimizar los entrenamientos. Sus proyectos destacan la importancia de transformar grandes volúmenes de datos en insights accionables, críticos para la toma de decisiones en deportes de alto rendimiento.

IV. PROPUESTA DE DATA PRODUCT

A. Set de datos

El conjunto de datos para este estudio se compone de información detallada suministrada por el Instituto Peruano del Deporte (IPD), específicamente del Programa de Apoyo al Deportista (PAD). Esta compilación meticulosa incluye registros desde el año 2013 hasta marzo de 2024, abarcando múltiples dimensiones del apoyo deportivo: desde datos financieros como montos y periodicidad del soporte, hasta datos demográficos y de rendimiento de los atletas. Además, este dataset se enriquece con metadatos que proporcionan un contexto adicional sobre las federaciones, las categorías de deportes y el nivel de competición de cada atleta. Para lograr una integración efectiva y un análisis holístico, este set de datos ha sido importado y procesado utilizando herramientas avanzadas de ciencia de datos, asegurando una estructura de 'tidy data' para un manejo y visualización eficientes.

Complementando este corpus de información, se integrará un conjunto de dashboards interactivos desarrollados en Tableau, proporcionando un análisis visual comprensivo del impacto del PAD. Estos dashboards facilitarán la exploración de tendencias y patrones, y serán cruciales para evaluar la eficacia de la distribución de los fondos y el retorno de la inversión deportiva.

Sin embargo, existen dimensiones aún no capturadas en estos datos que son esenciales para un análisis exhaustivo. Información como la tasa de finalización de los programas de apoyo por atleta, el feedback directo de los beneficiarios a través de encuestas o entrevistas, y la participación y el compromiso de los atletas con las actividades y recursos del PAD, son aspectos que requieren ser incorporados. Estos podrían recolectarse a través de técnicas avanzadas de procesamiento de lenguaje natural (NLP) y encuestas digitales, y serían fundamentales para desarrollar predictores más sólidos del éxito deportivo.

Además, el seguimiento del progreso y la implicación de los deportistas en las etapas críticas de su desarrollo, así como la cuantificación del tiempo dedicado a entrenamientos y otras actividades relacionadas con su preparación, permitirá

ajustar los programas de apoyo para maximizar el rendimiento. Herramientas analíticas como Google Analytics [6] y software de mapas de calor se presentan como opciones viables para rastrear la interacción de los deportistas con los recursos online del PAD, otorgando así una comprensión más profunda de su comportamiento y necesidades.

La presente propuesta de set de datos se fundamenta en la convicción de que una comprensión amplia y multidimensional del apoyo deportivo no solo mejorará la asignación de recursos, sino que también reforzará el compromiso y la excelencia de los deportistas peruanos a nivel mundial. [7]

B. Modelamiento y/o Análisis

El análisis de este proyecto se centrará en aplicar técnicas descriptivas y predictivas de vanguardia utilizando Python y su robusto stack de librerías de ciencia de datos, incluyendo Pandas para la manipulación de datos, Matplotlib [8] y Seaborn para visualizaciones descriptivas, junto con Plotly para interactividad avanzada.

Inicialmente, se llevarán a cabo análisis exploratorios mediante la generación de visualizaciones que muestren la distribución y tendencias del apoyo financiero en el Programa de Apoyo al Deportista (PAD). Esto incluirá gráficos de líneas que detallan la evolución del apoyo a lo largo del tiempo, gráficos de barras resaltando las federaciones deportivas con mayores fondos y gráficos de dispersión que exploran la relación entre la inversión en los atletas y sus resultados deportivos. Adicionalmente, se integrará un dashboard interactivo en herramientas como Tableau o Power BI [9] para sintetizar estos insights y ofrecer una visión consolidada y accesible para los stakeholders.

La investigación se profundizará en la interacción de los deportistas con los programas del PAD, analizando cómo los atletas interactúan con los recursos y programas de entrenamiento. Se utilizarán datos de seguimiento de actividades y encuestas de retroalimentación para medir la efectividad del soporte proporcionado y adaptar las intervenciones según sea necesario.

Se desarrollarán modelos predictivos usando técnicas de machine learning de aprendizaje supervisado para predecir el éxito futuro de los deportistas, basándonos en una variedad de predictores desde el historial de rendimiento hasta las características demográficas y detalles de su participación en los programas del PAD. El objetivo será identificar tempranamente a aquellos atletas con riesgo de no cumplir sus metas y proporcionar apoyo personalizado.

Además, exploraremos el uso de la API de OpenAI [10] para desarrollar un sistema de recomendaciones personalizadas de salud y entrenamiento basado en las interacciones y necesidades específicas de los atletas. Este sistema utilizará técnicas avanzadas de procesamiento de lenguaje natural para analizar las interacciones de los deportistas y ofrecer consejos y recomendaciones en tiempo real.

Este enfoque se complementará con técnicas de aprendizaje no supervisado para descubrir agrupaciones y patrones en los datos. Utilizaremos algoritmos como K-Means y DBSCAN [11] para categorizar a los deportistas

por su perfil de rendimiento y características, permitiendo así intervenciones más personalizadas y efectivas.

Por último, integraremos modelos de series temporales para anticipar las necesidades futuras de financiamiento y los resultados deportivos, ajustando las estrategias de apoyo en tiempo real para maximizar la eficacia del programa.

En conjunto, nuestra aproximación analítica transformará los datos del PAD en una herramienta poderosa para optimizar los procesos, maximizar el impacto y elevar el potencial de nuestros atletas a un nivel sin precedentes, garantizando que cada inversión se traduzca en mejoras tangibles y sostenibles en el rendimiento deportivo.

V. BENEFICIOS

La integración del data product en el Programa de Apoyo al Deportista (PAD) del Instituto Peruano del Deporte (IPD) catalizará una transformación en la forma en que los recursos se asignan y se evalúa el rendimiento deportivo. Este sistema avanzado no solo optimizará la distribución de fondos basada en análisis predictivos y descriptivos, sino que también facilitará una respuesta personalizada a las necesidades de los atletas, garantizando que el soporte proporcionado sea el más adecuado para sus requerimientos individuales.

Al emplear modelos de machine learning [12] para identificar patrones y predecir el rendimiento de los deportistas, el PAD podrá anticipar y mitigar los riesgos de baja participación y fracaso en competencias. La personalización del apoyo basada en datos concretos y análisis predictivo permitirá dirigir recursos de manera eficaz hacia aquellos deportistas que, según los modelos, muestran mayor probabilidad de éxito o necesidad de intervención.

Además, la incorporación de la API de OpenAI para desarrollar un sistema de recomendaciones personalizadas transformará la interacción de los atletas con el programa. Este sistema utilizará técnicas avanzadas de procesamiento de lenguaje natural para analizar las consultas y feedback de los deportistas en tiempo real, proporcionando respuestas automáticas y personalizadas que no solo resolverán dudas de manera eficiente sino también ofrecerán recomendaciones proactivas basadas en sus patrones de entrenamiento y competencia. Este enfoque asegura una comunicación efectiva y un soporte continuo, mejorando significativamente la experiencia del atleta y su satisfacción con el programa.

El análisis detallado de las tendencias mediante herramientas de visualización avanzadas y el seguimiento en tiempo real de las interacciones y el progreso de los deportistas permitirá al IPD no solo ajustar los programas de entrenamiento y apoyo en tiempo real, sino también planificar estratégicamente las futuras iniciativas de desarrollo deportivo. Esto se traduce en una mejora continua del rendimiento deportivo y una mayor eficiencia en la gestión de los recursos.

La implementación de este data product no solo optimiza la asignación de recursos y mejora el rendimiento de los atletas, sino que también establece un precedente para la utilización de tecnologías avanzadas en la gestión deportiva nacional. Al proporcionar una base sólida para decisiones basadas en datos, el PAD fortalece su capacidad para construir un legado de éxito deportivo, promoviendo un impacto duradero y positivo en el deporte peruano.

VI. PLAN DE PROYECTO

El desarrollo del proyecto para el Programa de Apoyo al Deportista (PAD) del Instituto Peruano del Deporte (IPD) se guiará por la **metodología Agile Data Science**, [13] la cual permitirá un enfoque dinámico y adaptable para abordar tanto los desafíos iniciales identificados como aquellos que surjan durante la ejecución del proyecto.

Inicialmente, nos centraremos en el conocimiento profundo del negocio. A través de talleres con entrenadores, deportistas, y administradores del PAD, delinearemos claramente los problemas que el proyecto necesita resolver, como la optimización de recursos y la mejora del rendimiento deportivo. Se creará un Product Backlog que incluirá todas las tareas y funcionalidades deseadas, priorizadas según su importancia y urgencia. Esta fase de planificación inicial, que anticipamos tomará aproximadamente dos semanas, nos permitirá ajustar los objetivos y alinear las expectativas de todos los stakeholders.

Desarrollo Iterativo, cada iteración, o sprint, de nuestro proyecto comenzará con una planificación donde seleccionaremos las tareas más críticas del Product Backlog para ser desarrolladas. Cada sprint tendrá una duración de dos semanas y se enfocará en diferentes aspectos del proyecto:

Entendimiento de Datos y Preparación Preliminar, exploraremos y prepararemos los conjuntos de datos del IPD, evaluando la calidad y realizando una limpieza de datos. Esto nos permitirá tener una base sólida para los análisis y modelado posteriores.

Desarrollo de Modelos Predictivos y Herramientas de NLP, Aplicaremos técnicas de machine learning para modelar el rendimiento deportivo y desarrollaremos herramientas de NLP utilizando la API de OpenAI para analizar comentarios y retroalimentación de los deportistas.

Integración y Pruebas de Usabilidad, integraremos todos los modelos y herramientas en un sistema cohesivo y realizaremos pruebas con usuarios reales para asegurar que el sistema sea intuitivo y efectivo.

Evaluación y Ajustes, al finalizar los ciclos iniciales de desarrollo, llevaremos a cabo una evaluación completa del sistema en un entorno real para validar su efectividad. Esta fase permitirá realizar ajustes basados en el feedback directo de los deportistas y otros usuarios, garantizando que el sistema responda adecuadamente a sus necesidades y expectativas.

Implementación y Capacitación, una vez que el sistema haya sido ajustado y validado, procederemos a su implementación completa dentro del IPD. Organizaremos sesiones de capacitación para todos los usuarios finales, asegurando que estén plenamente capacitados para utilizar el sistema de manera eficaz y aprovechar todas sus funcionalidades para mejorar el rendimiento deportivo.

Monitoreo Continuo y Mejora, estableceremos un proceso de monitoreo continuo que nos permitirá evaluar el impacto del sistema en el rendimiento deportivo y la gestión

de recursos. Este enfoque de mejora continua asegurará que el sistema se mantenga relevante y efectivo ante las cambiantes demandas del deporte y las necesidades de los atletas.

REFERENCIAS

- [1] Instituto Peruano del Deporte. (2024). El Programa de Apoyo al Deportista (PAD) recibe incremento del IPD. Gobierno del Perú. <https://www.gob.pe/institucion/ipd/noticias/200959-el-programa-de-apoyo-al-deportista-pad-recibe-incremento-del-ipd>
- [2] Instituto Peruano del Deporte. (2024). IPD benefició a miles de personas a nivel nacional a través de sus distintos programas. Gobierno del Perú. <https://www.gob.pe/institucion/ipd/noticias/888651-ipd-beneficio-a-miles-de-personas-a-nivel-nacional-a-traves-de-sus-distintos-programas>
- [3] Laboratorio de Análisis de Estadísticas y Deportes. (s.f.). SASL. Universidad de Virginia. <https://uva.theopenscholar.com/sasl/>
- [4] Explorando la IA en el análisis de rendimiento y la optimización del rendimiento deportivo. (2023, 14 de agosto). Moments Log. <https://www.momentslog.com/development/architecture/exploring-ai-in-sports-analytics-and-performance-optimization>
- [5] Grupo de Análisis Deportivo de Berkeley. (s.f.). Proyectos. Universidad de Berkeley. Recuperado de <https://sportsanalytics.studentorg.berkeley.edu/projects.html>
- [6] Google Analytics Academy. (2023). [Google Analytics para Principiantes]. <https://analytics.google.com/analytics/academy/course/6?hl=es-419>
- [7] Instituto Peruano del Deporte. (2024). Deportistas beneficiados por el Programa de Apoyo al Deportista. Plataforma Nacional de Datos Abiertos. <https://www.datosabiertos.gob.pe/dataset/programa-apoyo-al-deportista-deportistas-beneficiados-instituto-peruano-del-deporte-ipd>
- [8] Zapata, J. R. (n.d.). Visualización de Datos con Python. <https://joserzapata.github.io/courses/python-ciencia-datos/visualizacion/>
- [9] DataScientest. (2023). Power BI vs. Tableau: ¿cuál es la mejor herramienta de Business Intelligence? <https://datascientest.com/es/power-bi-vs-tableau-es>
- [10] OpenAI. (2024, 11 de junio). OpenAI API. OpenAI. <https://openai.com/blog/openai-api>
- [11] Zapata, J. R. (2020). Visualización de Datos con Python. <https://joserzapata.github.io/courses/python-ciencia-datos/visualizacion/>
- [12] OpenWebinars. (2019, 15 de octubre). Modelos de Machine Learning. <https://openwebinars.net/blog/modelos-de-machine-learning/>
- [13] Journey, R. (2017). Agile Data Science 2.0: Building Full-Stack Data Analytics Applications with Spark. Amazon.com. https://www.amazon.com/-/es/Russell-Journey/dp/1491960116?language=es_US