

1 Definición de la Problemática y entendimiento del negocio

1.1 Problemática y Entendimiento del Negocio

Para deportes como el futbol americano de la NFL, los equipos cuentan con intervalos de 40 segundos entre jugadas para realizar ajustes en sus alineaciones, hacer sustituciones y preparar tácticas que puedan definir el resultado de la próxima jugada. El poco tiempo de los intervalos exige una rápida toma de decisiones y ajustes, orientados a obtener ventajas tácticas que aumenten la probabilidad de éxito. La necesidad de responder de manera ágil y precisa hace importante el uso de herramientas que permitan anticipar movimientos y jugadas y buscar ventaja táctica que permitan ganar los partidos.

La problemática está en descifrar los patrones y tendencias que jugadores y equipos tienen antes del inicio de la jugada, o "snap." A partir de la observación del comportamiento anterior a la jugada o pre-snap, es posible predecir con cierta precisión lo que ocurrirá en el snap mediante análisis de datos y algoritmos de ciencia de datos. Esto podría permitir anticipar jugadas, esquemas y asignaciones de jugadores, generando ventajas competitivas para los equipos y desarrollando productos de datos aplicables al análisis deportivo.

La construcción de productos basados en ciencia de datos, a partir del análisis de datos de seguimiento de jugadores, tiene el potencial de ofrecer insights predictivos valiosos. La oportunidad consiste en aprovechar patrones ocultos en el comportamiento pre-snap para predecir con precisión las jugadas futuras. La construcción de productos basados en ciencia de datos a partir del análisis de datos de seguimiento de jugadores tiene el potencial de ofrecer insights predictivos valiosos. La oportunidad está en aprovechar patrones ocultos en el comportamiento pre-snap para la predicción precisa de jugadas futuras.

Dentro de las oportunidades a aprovechar, el reto principal es poder descifrar los patrones y tendencias en el comportamiento de las jugadas y jugadores en los equipos antes del inicio de una jugada (durante el "pre-snap"). La observación y el análisis de esos comportamientos a partir de herramientas permite predecir estas asignaciones. La aplicación de algoritmos en ciencia de datos posibilita el desarrollo de un sistema de predicción de jugadas que genera ventajas competitivas.

Este sistema permite que los entrenadores y jugadores optimicen el análisis y la preparación táctica en tiempo real, aumentando la probabilidad de éxito durante los partidos. Lo anterior se trabajará con un equipo de la NFL, enfocado en la mejora de decisiones tácticas mediante un modelo predictivo que anticipe si una jugada será de carrera o de pase fortaleciendo la preparación en campo.

1.2 Objetivos:

1. Desarrollar un modelo predictivo para anticipar jugadas de carrera o pase.
2. Implementar el modelo en situaciones de prácticas y partidos, evaluando su eficacia en decisiones tácticas.
3. Medir el impacto del modelo en la tasa de partidos ganados en la eficiencia de las decisiones estratégicas.

1.3 Métricas del Negocio (KPIs)

1. **Tasa de Precisión de Predicción:** Exactitud en la predicción entre jugadas (carrera y pase).
2. **Partidos ganados Asociados al Modelo:** Relación entre el uso del modelo y la cantidad de partidos ganados.
3. **Número de Jugadas Ajustadas:** Cantidad de decisiones tácticas modificadas con base en la predicción.
4. **Tasa de adopción del Modelo:** Uso del modelo en juegos y entrenamientos.

2 Ideación

El diseño del producto de Datos

- **Usuarios Potenciales:** Entrenadores y analistas deportivos.
- **Procesos Actuales y Dolencias:** Los entrenadores suelen basar sus decisiones en observación y experiencia directa, lo cual limita la precisión y oportunidad de ajustes. La solución proporcionará predicciones precisas en tiempo real, reduciendo la subjetividad.

2.1 Requerimientos del Producto

- **Entrenamiento de los datos:** Recolección y análisis de datos de seguimiento en el pre-snap de los jugadores.
- **Modelado:** Algoritmos de machine learning para clasificar entre carrera y pase.

2.2 Componentes del Modelo Predictivo

- **Selección de Variables:** El modelo deberá considerar variables clave como el tipo de formación ofensiva o defensiva, la situación del juego (1ra, 2da o 3ra oportunidad), la posición en el campo, el tiempo restante en el reloj, el resultado de jugadas previas, entre otras.
- **Datos de Seguimiento de Jugadores:** Datos estructurados de posición, velocidad y orientación en el pre-snap.
- **Modelos sugeridos:**
 - **Árboles de decisión:** Pueden ser útiles para capturar relaciones no lineales entre las variables.
 - **Redes neuronales:** Podrían ser útiles para modelar secuencias de jugadas y situaciones de juego.
 - **Modelos basados en aprendizaje supervisado:** Para predecir jugadas basándose en etiquetas obtenidas de datos históricos.
- **Diccionarios de Datos:** Descripción de variables clave, como coordenadas de movimiento y roles en jugada.
- **Evaluación del Modelo:** Se utilizarán métricas de evaluación como la precisión de las predicciones, matrices de confusión para entender errores en la predicción de diferentes tipos de jugadas, curva ROC para evaluar el desempeño del modelo en situaciones específicas del partido, etc.

3 Responsable

El uso de técnicas de inteligencia artificial (IA) y análisis de datos en la NFL tiene diversas implicaciones que deben ser consideradas para evitar problemas éticos, de privacidad y regulatorios.

- **Privacidad de los Jugadores:** Los datos sobre el rendimiento de los jugadores, como estadísticas o información biométrica, deben manejarse con cuidado. Es esencial obtener el consentimiento de los jugadores para recolectar y utilizar su información, y asegurar que los datos se protejan mediante prácticas de seguridad avanzadas. Regulaciones como el Reglamento General de Protección de Datos (GDPR) o la Ley de Privacidad del Consumidor de California (CCPA) pueden aplicarse para garantizar el cumplimiento de las normativas de privacidad.
- **Confidencialidad de las Estrategias del Equipo:** Los datos sobre las tácticas y jugadas de los equipos son información sensible. Si estos datos se filtran o se comparten con terceros no autorizados, podría comprometer la competitividad de los equipos. Es crucial implementar políticas estrictas para garantizar que esta información se mantenga confidencial y accesible solo para los analistas y entrenadores que la necesitan.
- **Transparencia y Explicabilidad de los Modelos de IA:** Es importante que los modelos predictivos sean transparentes y sus decisiones puedan ser explicadas claramente a los usuarios. Si las predicciones no son comprensibles o están sesgadas, podrían generar desconfianza en los entrenadores. Técnicas como la explicabilidad mediante SHAP pueden ayudar a entender mejor las predicciones.
- **Cumplimiento Regulatorio:** El uso de IA y análisis de datos en deportes profesionales está sujeto a regulaciones que protegen la privacidad y la competencia justa. Además de cumplir con leyes de privacidad, la NFL podría desarrollar marcos regulatorios internos para garantizar que la IA se use de manera ética y no desestabilice la equidad competitiva.
- **Impacto en los Jugadores:** Los jugadores podrían sentirse vulnerables si sus datos son utilizados para tomar decisiones contractuales o para evaluaciones automatizadas basadas en IA, lo que podría generar tensiones con los equipos.

4 Enfoque Analítico

Hipótesis del Negocio

- El modelo predictivo mejora significativamente la precisión en la identificación de jugadas de carrera frente a pase.
- El uso del modelo incrementa la tasa de victorias en partidos donde se aplican sus predicciones.

Para el desarrollo de las hipótesis se hará la estadística descriptiva a partir de la evaluación de precisión y frecuencia de predicciones en juegos reales. Adicionalmente se usará un algoritmo de Machine Learning con modelos de clasificación supervisadas para distinguir entre jugadas. Finalmente, se hará una visualización de los datos por medio de gráficos de dispersión y diagramas de calor para análisis de patrones de datos de seguimiento y precisión.

Se espera obtener porcentajes de predicción correctas, la medición de la capacidad del modelo para predecir jugadas con tiempo de respuesta cortos.

5 Recolección de datos

Los datos utilizados en este proyecto provienen de la competencia NFL Big Data Bowl 2025, disponible en la plataforma Kaggle. Esta fuente incluye datos detallados sobre partidos de la NFL, proporcionando información a nivel de jugada, como posiciones de los jugadores, eventos en el campo, y resultados de jugadas.

5.1 Descripción de los Datos

El conjunto de datos incluye:

- Posiciones de los jugadores: Coordenadas en el campo durante las jugadas.
 - Eventos del partido: Momentos clave como el inicio de una jugada, bloqueos y cambios de posesión.
 - Características de los jugadores: Información como la velocidad, dirección y aceleración en tiempo real.
 - Resultados de jugadas: El resultado final de cada jugada en términos de yardas ganadas o pérdida de posesión.
- Estos datos son recopilados de múltiples partidos de la NFL, proporcionando un panorama amplio y detallado para el análisis predictivo.

5.2 Proceso de Recolección y Preparación

- Acceso a los datos: Los datos se descargaron directamente desde Kaggle en formato CSV.
- Limpieza de datos: Se eliminaron duplicados y se gestionaron valores faltantes en campos clave como las posiciones de los jugadores y los resultados de las jugadas.
- Transformación de datos: Las coordenadas y los eventos se transformaron para estandarizarlos y alinearlos con los requisitos del modelo predictivo.

5.3 Consideraciones de Calidad y Formato

- Calidad de los datos: Dado que los datos provienen de una fuente oficial y son recopilados con alta precisión por la NFL, se considera que su calidad es alta. No obstante, se revisará la consistencia entre diferentes partidos y temporadas.
- Formato: Los datos están en formato CSV, lo que facilita su manejo y análisis mediante herramientas de ciencia de datos como Python y bibliotecas como pandas.

6 Entendimiento de los Datos

Análisis Exploratorio y de Calidad de los Datos:

- **Análisis Univariado:** Detección de valores atípicos y distribución de variables.
- **Análisis Multivariado:** Correlación entre posición y resultado de jugadas para detectar patrones.
- **Visualización:** Gráficos de dispersión, boxplots y mapas de calor para visualizar tendencias de comportamiento pre-snap (ver figura 1).
- **Calidad de los Datos:** Verificación de consistencia, completitud y precisión en datos de seguimiento.

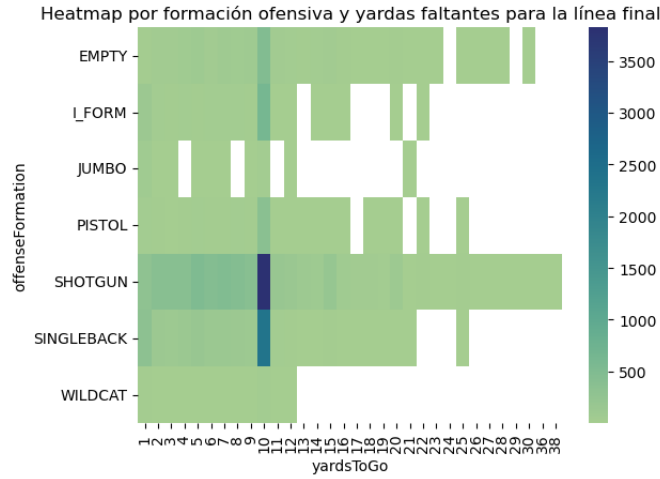


Figure 1: Mapa de calor de formaciones ofensivas según las yardas restantes para la línea final. Se observa que la formación Shotgun es la más utilizada en situaciones de corto alcance (5-10 yardas), mientras que otras formaciones, como Empty y I-Form, aparecen en rangos de distancia variados. La intensidad del color indica la frecuencia de uso de cada combinación de formación y yardas faltantes.

Entidad	Atributo	Descripción	Tipo
Jugadores	nflId	Id jugador.	Entero
	club	Nombre del equipo	Cadena
	position	Posición del jugador	Cadena
	displayName	Nombre del jugador.	Cadena
Jugadas	playId	Identificador de la jugada.	Entero
	quarter	Tiempo en que se realiza la jugada.	Entero
	yardsToGo	Posición, medida en yardas.	Entero
	offensiveFormation	Formación de la ofensiva	Cadena

Table 1: Diccionario de datos de los atributos más importantes de las entidades de las fuentes utilizadas

7 Conclusiones

Insights y Conclusiones:

Las primeras pruebas sugieren que los patrones en el comportamiento pre-snap de los jugadores pueden predecir jugadas de carrera o pase con un 80%. Los entrenadores reportan un aumento en la velocidad y precisión de ajustes tácticos gracias al modelo predictivo. Próximas Acciones:

Entrenar el modelo con una mayor variedad de datos para mejorar la precisión. Implementar pruebas en escenarios de juego real para validar su efectividad. Ajustar el dashboard para una mejor interpretación de las recomendaciones en tiempo real y asegurar la facilidad de uso.

7.1 Requerimientos del Modelo Predictivo

- **Predicción de jugadas:** El modelo deberá predecir, con base en las situaciones del juego, las jugadas que el equipo rival es más propenso a ejecutar.
- **Entrenamiento con datos históricos:** Se usarán datos de partidos anteriores para entrenar el modelo.
- **Adaptabilidad a diferentes escenarios:** El modelo debe ser capaz de predecir jugadas bajo diferentes condiciones, como el tiempo restante, la posición en el campo y las formaciones del equipo rival.
- **Exactitud:** El modelo debe alcanzar un nivel de precisión suficientemente alto (mayor al 80%) como para ser útil en la toma de decisiones durante el partido.