

PREDICCIÓN DE RENDIMIENTO OPTIMO EN CICLISMO: MACHINE LEARNING VERSUS POTENCIÓMETROS



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

CARLOS IGNACIO BERMÚDEZ SOLIS

PROFESOR GUÍA: DR. JUAN MANUEL FLOREZ URIBE

TESIS PARA OPTAR AL GRADO DE
LICENCIADO EN CIENCIAS MENCIÓN FÍSICA

DEPARTAMENTO DE FÍSICA

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA VALPARAÍSO - CHILE

DICIEMBRE - 2023

Esta tesis está dedicada a mis padres, ya que gracias a ellos estoy donde estoy y he
logrado lo que he logrado.

*“Para lograr realizar lo que quieres, aprende a sentirte cómodo en lo
incómodo.” ’*

Alejandro Jodorowsky.

AGRADECIMIENTOS

Quiero expresar mi más sincero agradecimiento a mis padres, quienes han sido mi apoyo incondicional a lo largo de mi vida. Su amor, dedicación y sacrificio han sido la fuerza motriz detrás de mis logros y éxitos. Agradezco también a mis amigos, quienes han compartido risas, desafíos y momentos inolvidables conmigo durante mi carrera. Su amistad ha sido invaluable, brindándome un círculo de apoyo vital. Además, quiero reconocer a las increíbles personas que he tenido el privilegio de conocer este año. Su positividad, inspiración y contribuciones han dejado una huella significativa en mi vida, enriqueciéndola de maneras que nunca imaginé. Estoy agradecido por el regalo de estas relaciones significativas.

RESUMEN

El ciclismo no es solo uno de los deportes más populares del planeta; la bicicleta se perfila como uno de los medios de transporte masivo del futuro por razones ecológicas, de salud y económicas. La optimización del rendimiento en todo tipo de ciclismo amateur/-competitivo se presenta como un requisito no solo para el avance a nivel deportivo, sino también para el diseño de nuevos medios en electromovilidad, el desarrollo de alimentación personalizada y el entendimiento de la biofísica humana.

La gran mayoría de métodos para el mejoramiento de la actividad ciclística en la actualidad están basados en el uso de una gran variedad de sensores, como GPS, cadencia, viento, frecuencia cardíaca, velocidad y potenciómetros. Estos últimos están altamente ligados al deporte de alto rendimiento y son ampliamente aceptados como quizás la mejor medida a ser utilizada como base para la optimización. Sin embargo, los potenciómetros son excesivamente costosos al ser dispositivos de alta complejidad electrónica y estar miniaturizados para reducir peso en su acoplamiento a la bicicleta. Además, se desarrollan mayormente para el ciclismo de carrera de alta competencia, lo que los hace realmente un accesorio solo al alcance de unos pocos.

En esta tesis, buscamos implementar un modelo de Machine Learning (ML) que se entrene con bases de datos de ciclistas amateur para luego predecir la potencia a partir de datos de fácil adquisición. Los modelos físicos analíticos son insuficientes para establecer una predicción confiable y reproducible que tome en cuenta todo el espacio de fase de parámetros medibles y sujetos a cada ciclista, tipo de terreno/vehículo y patrón de agotamiento/esfuerzo. Hasta el momento, hemos tenido éxito para propósitos generales.

Índice general

Agradecimientos	ii
Resumen	iii
Lista de figuras	vii
1. Introducción	1
1.1. Historia del ciclismo	1
1.2. Evolución de las estrategias de ciclismo	1
1.3. Ciclismo de competición	3
1.4. Desarrollo tecnológico en el ciclismo	4
1.5. Física del ciclismo	6
1.6. Problemáticas actuales del ciclismo	18
1.7. Machine learning: breve introducción en contexto	20
1.8. Retos actuales en física del ciclismo y sus aplicaciones	25
2. Datos y métodos	27
2.1. Obtención de datos	27
2.2. Datos experimentales - Primera recolección	29
2.3. Analisis estadístico - Primera recolección	32
2.4. Datos experimentales - Segunda recolección	33
2.5. Analisis estadístico - Segunda recolección	35
2.6. Machine learning	44
2.6.1. Primeros pasos.	44
2.6.2. Definir la red neuronal.	46
2.6.3. Tratamiento previo de los datos	47
2.6.4. Estructura del modelo	48
2.6.5. Hiperparametros	50

3. Resultados y conclusiones	54
3.1. Resultados	54
4. Conclusiones	66
Bibliografía	69

Índice de figuras

1.1.	Ejemplo de interfaz de usuario del software Strava.	7
1.2.	Ejemplo de interfaz de usuario del software Zwift.	8
1.3.	Gráfico de potencia real de una carrera.	12
1.4.	Cuatro posiciones diferentes utilizadas durante los descensos con el correspondiente coeficiente de presión estática $C_p = 2(P - P_0)/\rho V_0^2$ en superficies para ciclistas y bicicletas. Aquí, P es la presión estática, P_0 la presión atmosférica y V_0 la velocidad del centro de masa. El área frontal S y el área de arrastre SC_D también se indican para cuatro posiciones significativamente diferentes:(a) posición de Froome, (b) posición de Pantani, (c) top tube 4 y (d) posición de superman. Figura tomada de Cohen [10]. .	13
1.5.	Topología de flujo del sistema de vórtice a lo largo de la corriente para (a) 15° simétricos de baja resistencia régimen de flujo y 75° régimen de flujo asimétrico de alta resistencia visto desde (b) la izquierda lado y (c) el lado derecho. Los símbolos L, N, S y F denotan líneas principales, nodos, puntos de silla y focos, respectivamente, con subíndices S y A que indican si están asociados con la separación o el apego. Los números de subíndice se utilizan para identificar puntos críticos al abordar los resultados en apoyo de los patrones de fricción de la piel propuestos. Figura tomada de Crouch [11].	15
1.6.	Contornos de (a,c,e) velocidad media del viento y (b,d,f) coeficiente de presión media en un plano horizontal a 1 m por encima de la carretera, para las tres subconfiguraciones. Las figuras a,c,e también muestran los valores de las fuerzas de arrastre. Las figuras b, d, f también muestran la resistencia como porcentaje de la resistencia de un ciclista aislado a 15 m/s, que es 40,33 N.Figura tomada de Blocken [6].	16
1.7.	Arrastre de cada ciclista en el pelotón A como porcentaje del arrastre de un ciclista aislado que circula a la misma velocidad.	17

1.8. Arrastre de cada ciclista en el pelotón B como porcentaje del arrastre de un ciclista aislado que circula a la misma velocidad. Figura tomada de Blocken [5].	17
1.9. Ejemplo de 3 tipos de potenciómetros distintos. 1) Potenciómetro de biela. 2) Potenciómetro de buje. 3) Potenciómetro de pedal.	20
1.10. Ilustración de la estructura mas basica de una Red Neuronal	22
2.1. En esta foto se observa la primera reunión con el grupo de ciclismo 24BIKE, donde junto con Sergio, Hernán y el profesor guía Juan Manuel, realizamos una ruta en bicicleta bastante conocida por ellos por formar parte de las primeras rutas de iniciación para nuevos integrantes de su grupo. . .	28
2.2. Detalle de características con unidades de medida. Cada número corresponde a la etiqueta de la característica en los modelos de entrenamiento. .	29
2.3. Comparación de 3 rutas del mismo participante, con variaciones en el recorrido.	30
2.4. Aquí se observa cuáles son las variables que obtuvieron mayor correlación con la potencia para cada método estadístico.	33
2.5. Ejemplo de un ciclista entrenando mediante la simulación de un recorrido en vivo utilizando el software Zwift.	34
2.6. Aquí se observan las 4 rutas seleccionadas en Zwift para la segunda instancia de recolección de datos.	35
2.7. Comparación, mediante gráfico de violín, de la distribución de potencia de personas distintas en la misma ruta.	37
2.8. Comparación de perfil de potencia de un entrenamiento vs. perfil de potencia en un recorrido normal (en la misma ruta).	38
2.9. Representación tridimensional de Actividad 1.	39
2.10. Representación tridimensional de Actividad 2.	40
2.11. Representación tridimensional de Actividad 3.	41
2.12. Gráficos de interacción entre características.	42
2.13. Comparación entre potencia calculada según el modelo matemático y potencia real.	43
2.14. Esquema de referencia para cálculo de inclinación.	44
2.15. Gráficos de interacción entre características para rutas en subida.	45
2.16. Distintos tipos de configuraciones para redes neuronales. Figura tomada de Alvy [3]	46

2.17. En esta ilustración, se pueden apreciar las distintas partes que componen la estructura de una neurona. Figura tomada de DotCSV [1].	48
2.18. Estructura de la primera red neuronal para predicción de velocidad.	50
2.19. Estructura de la red neuronal final.	52
3.1. Resultados primera red neuronal para predecir velocidad.	55
3.2. Primer intento de graficar la fórmula con datos reales.	56
3.3. Segundo intento de graficar la fórmula (Actividad 1).	57
3.4. Segundo intento de graficar la fórmula (Actividad 2)	57
3.5. Segundo intento de graficar la fórmula (Actividad 3).	58
3.6. Cálculo de potencia mediante formula (Actividad 1)	58
3.7. Cálculo de potencia mediante formula (Actividad 2)	59
3.8. Cálculo de potencia mediante formula (Actividad 3)	59
3.9. Predicción de potencia mediante ML (Perfil Pearson)	60
3.10. Predicción de potencia mediante ML (Perfil Spearman)	60
3.11. Predicción de potencia mediante ML (Perfil Feature importance)	61
3.12. Predicción de potencia mediante ML (Perfil RFE)	61
3.13. Predicción de potencia mediante ML (Perfil PearsonR)	62
3.14. Predicción de potencia mediante ML (Actividad 1)	63
3.15. Predicción de potencia mediante ML (Actividad 2)	63
3.16. Predicción de potencia mediante ML (Actividad 3)	64
3.17. Comparación de 5 modelos con características distintas, incluyendo los errores del modelo matematico para las 3 actividades.	64
3.18. Comparación entre potencia calculada según el modelo matemático y potencia real.	65
3.19. Comparación entre potencia calculada según la red neuronal y potencia real.	65

Capítulo 1

Introducción

1.1. Historia del ciclismo

Desde la creación de las bicicletas aproximadamente en el siglo XVIII, el uso de este medio de transporte creció exponencialmente, por lo cual, a medida que se popularizaba obviamente surgieron mejoras de distintos tipos, pero una de las más significativas fue la propuesta por J.K. Starley en 1885 cuando propuso el sistema de cadena y engranaje, potenciando así a lo largo del tiempo, la evolución de las bicicletas hacia modelos cada vez más ligeros, más eficientes aerodinámicamente y más ergonómicos.

Posteriormente, en el año 1896, se dio lugar a la primera prueba de ciclismo en ruta en los juegos olímpicos que consistió en un recorrido de ida y vuelta entre Atenas y Maratón, donde la distancia fue aproximadamente de 87 km.

1.2. Evolución de las estrategias de ciclismo

Con el pasar de los años, se ha podido evidenciar una evolución clara en uno de los aspectos más importantes de los deportes en la actualidad, la estrategia. Si bien el ciclismo de ruta a nivel competitivo es un deporte muy caótico, esto debido a la cantidad de competidores que hay en una carrera y además por todas las decisiones diferentes que pueden tomar esos competidores, se ha logrado establecer cierto tipo de estrategias gracias al análisis de patrones y situaciones recurrentes. Tanto desde el análisis macro como micro, se obtienen múltiples patrones que se pueden estudiar y replicar para una mejora tanto individual como grupal de los competidores. Una de las estrategias grupales más vistosas

es la de recurrir al viaje grupal durante los primeros kilómetros de la carrera, esto significa que la mayoría de los corredores forman un grupo con una estructura de flecha (más o menos) con el fin de romper el viento que les llega de frente, esto obviamente beneficia en mayor medida a los corredores que están en las últimas posiciones de este grupo, y por lo mismo, el grupo va rotando constantemente. Mas adelante, en una parte más avanzada de la carrera, este gran grupo pierde velocidad en ciertos momentos, y los ciclistas mas habilidosos aprovechan esta ventana para segmentarse en grupos mas pequeños y aprovechar de sacar ventaja de tiempo.

Entonces se entiende que el trabajo en grupo es un aspecto crucial para los corredores. Hoenigman [19] menciona que la efectividad de este trabajo en grupo esta relacionada con el nivel de habilidad de cada ciclista, donde según sus resultados, los ciclistas mas débiles se benefician mas al trabajar de forma individual, en cambio aquellos ciclistas que son mas fuertes tienen mejores resultados trabajando de forma conjunta con sus compañeros. En estas situaciones, los compañeros de equipo que cooperan se unen a la escapada para brindar apoyo, aunque esto implica esfuerzos adicionales temporales que pueden llevarlos por encima de sus umbrales de resistencia.

Ya en este momento de la carrera lo que más se ve son estrategias “individuales”, que no son del todo individuales ya que generalmente en estos grupos más pequeños suelen haber 3 o 4 corredores que son del mismo equipo, y entre ellos, suelen tener solo a uno o dos que son considerados posibles ganadores y entrenan con ese objetivo, mientras que los otros dos solo son corredores de apoyo, y según eso también hay estrategias, por ejemplo si va 1 corredor solitario versus otros dos que son del mismo equipo, la estrategia del equipo va a ser que uno sea capaz de llevarse al solitario atrás, mientras que su compañero en algún momento ataque y se logre escapar por mucha distancia de los otros dos, en cambio para el solitario, la estrategia podría ser ir a rueda de cualquiera de los dos para ahorrar energía y lograr mantener el ritmo de ellos, para en algún momento poder atacar, y dejar a los dos atrás.

Obviamente, a un nivel elite las estrategias son muchísimo más complejas que esa y cada situación está muy estudiada, tanto las situaciones más generales como las más específicas.

1.3. Ciclismo de competición

El ciclismo como deporte ha ganado popularidad constantemente a lo largo de los años, y desde su creación hasta hoy en día ha tenido muchas ramificaciones que han convertido a la bicicleta en una herramienta deportiva muy versátil. Esto significa que al rededor del ciclismo se han extendido distintos tipos de competición, como las siguientes:

- El ciclismo de ruta es una de las ramas mas conocidas y populares del ciclismo de competición. Se caracteriza por tener recorridos largos y en distintos tipos de terreno. Esta competición se divide principalmente en cuatro formatos, las clásicas generalmente son recorridos muy largos y se realizan en un solo día. Por otro lado existen las carreras por etapas, donde se encuentran los eventos mas populares en ciclismo de competición como son, el Tour de Francia, la vuelta a España y el Giro de Italia. Las otras dos son las carreras contrarreloj que se dividen en contrarreloj individuales. En estas carreras, cada corredor se enfrenta al reloj y corre en solitario, sin la ayuda de otros ciclistas. Por otro lado, en las contrarreloj por equipos, un grupo de corredores trabaja en conjunto para intentar batir el tiempo establecido como equipo. En este tipo de contrarreloj, los corredores se turnan para liderar el grupo y aprovechan la aerodinámica y la colaboración para aumentar la velocidad y mejorar el tiempo del equipo.
- El ciclismo de pista es aquel que se realiza en velódromos, que son las pistas que tienen forma de ovalo. La principal característica de esta modalidad es la velocidad y resistencia de los ciclistas. Esta competición también tiene distintos formatos, como la persecución individual y por equipos, el keirin, la velocidad y la carrera por puntos. En la persecución individual, los ciclistas compiten en contrarreloj, tratando de registrar el mejor tiempo en un determinado número de vueltas. En la persecución por equipos, los equipos compiten en paralelo y tratan de atrapar o dejar atrás al equipo contrario. El keirin es una carrera en la que los ciclistas siguen a un vehículo motorizado durante las primeras vueltas y luego compiten entre sí para cruzar la línea de meta en primera posición. La velocidad es una carrera en la que dos ciclistas compiten en un sprint desde una posición de salida fija, tratando de superar al otro para llegar primero a la meta. La carrera por puntos es una competencia en la que los ciclistas acumulan puntos en sprints intermedios durante la carrera y el ciclista con más puntos al final de la carrera gana. La mayor exposición de ciclismo de pista son las olimpiadas.
- El ciclismo de montaña consiste en recorrer terrenos mucho mas complicados que el

asfalto, donde es necesario por parte de los ciclistas una serie de habilidades técnicas y destrezas para poder completar los circuitos. Obviamente hay un factor muy importante a tener en cuenta, y es que en este caso no se puede utilizar el mismo tipo de bicicleta que se usa para hacer ciclismo de ruta. La bicicleta que se suele utilizar es una MTB (Mountain Bike), y sus principales características son que es mas pesada para lograr una mayor estabilidad, las ruedas son mas anchas para mejorar el agarre con el suelo y evitar caídas por deslizamiento y además incluye amortiguación para que la rigidez no sea un problema al sortear las dificultades del terreno y permitir mayor movilidad de la bicicleta. Además, estas carreras también tienen distintos tipos, como lo son las campo través o Cross country, descenso, maratón y enduro. Una de las marcas que ha dado mas visibilidad a estas competiciones es Red Bull.

- El ciclo Cross logra combinar en una misma modalidad el ciclismo de carretera y de montaña, generalmente son circuitos cortos donde además de las dificultades naturales del terreno se añaden obstáculos que incluso pueden requerir que el ciclista tenga que bajarse de la bicicleta y superarlos a pie, llegando al punto de tener zonas tan difíciles de cruzar que tienen que ponerse la bici al hombro para superarlas. hay detalles que pueden considerarse mas obvios, como por ejemplo la selección de la bicicleta, dado que los ciclistas necesitan contar con mucha maniobrabilidad las bicicletas tienen que ser “livianas” aun que los diseños para esta competición varían mucho.

1.4. Desarrollo tecnológico en el ciclismo

Los primeros modelos de bicicleta, por muy impensable que sea, funcionaban con llantas de madera cubiertas de hierro. Obviamente estas bicicletas resultaban ser muy incómodas y “frágiles” dado que a grandes velocidades las ruedas de madera no resistían los impactos y se astillaban. Fue entonces en el año 1887 que por primera vez se reemplazó la rueda de madera por una rueda de caucho. Esto significó un avance en la sensación de comodidad al montar bicicleta.

Otra de las innovaciones más grandes del ciclismo es la implementación del pedal automático. Antes se utilizaban correas para sujetar los pies de los ciclistas al pedal, pero esto era un método muy peligroso, dado que cuando los ciclistas sufrían caídas no eran capaces de sacar sus pies de los pedales para apoyarse en el suelo, y las caídas eran más pe-

ligrosas. Por esto fue que se crearon los pedales automáticos, un mecanismo que permitía en primera instancia una mayor facilidad para colocar el zapato y quitarlo, y en segundo lugar permitía una mayor eficiencia en el uso de los pedales, esto porque al tener el pie sujeto al pedal puedes hacer un mejor uso de fuerza en los 360 grados de la vuelta que da el pedal, y no solo en los 180 que se empuja el pedal hacia abajo.

Por otro lado, uno de los aspectos mas importantes cuando se trata del ciclismo es el peso de la bicicleta, y una de las cosas que mas influía en esto es el material del que esta hecho el marco de la bicicleta. En un principio el marco de las bicicletas estaba hecho de acero, un metal que sin duda es bastante pesado, y eventualmente se logró sustituir este material por el aluminio. Un cambio notable para el peso de la bicicleta, considerando que aproximadamente una bicicleta con marco de acero pesaba entre 10 y 15 kg, y una bicicleta de aluminio pesaba entre 8 y 12 kg.

Sin embargo, en el año 1980 se comenzó a utilizar de forma experimental el material de fibra de carbono para crear marcos de bicicletas. Esto logró que el peso de la bicicleta volviera a bajar considerablemente a un rango de entre 6 y 10 kg. Luego de la gran inversión de un famoso empresario francés de la época, la marca pudo innovar en el desarrollo de estos marcos, creando así en 1986 la primera bicicleta de carbono de la historia que posteriormente ganaría el Tour de ese mismo año a manos del ciclista Greg Lemond.

Una de las innovaciones mas llamativas entre los apasionados del ciclismo y una de los motivos que me trae a desarrollar esta tesis, es el medidor de potencia que llegaba a resolver un problema bastante interesante del ciclismo, como dosificar y manejar el esfuerzo de manera concreta y no simplemente basándose en sensaciones que podían perfectamente ser muy imprecisas. Al poder saber la potencia en tiempo real que esta imprimiendo el ciclista, esto le permitía hacer cálculos y pruebas antes de la carrera para conocer sus umbrales de rendimiento en todos los aspectos de la carrera, lo cual al momento de correr le serviría de guía para optimizar de la mejor manera posible su desempeño. Una innovación que sin duda permite estar cada vez mas cerca del mejor desarrollo del performance del ciclista.

La nueva modalidad de e-racing [12] es una forma de competencia donde los ciclistas utilizan simuladores virtuales. En lugar de salir a la carretera, deben conectarse a través de su computador o "smart tv."^a una plataforma de e-racing y compiten siendo representados de manera virtual por un avatar digital. Esta forma de competencia ha ganado bastante

reconocimiento, tanto así que se han creado eventos como el Tour de Francia virtual y Campeonatos Mundiales de Ciclismo Esports.

Entre las ventajas del e-racing se encuentra la comodidad y la accesibilidad. Los ciclistas pueden competir desde casa, sin necesidad de salir o enfrentarse a las condiciones climáticas. Además, al ser un entorno virtual, permite conectarse y competir contra ciclistas de todo el mundo. Sin embargo, el e-racing también presenta problemas y desventajas. Uno de los problemas mas graves es la posibilidad de hacer trampas o manipular datos. Al tratarse de una competencia digital, existe la forma de utilizar bots o ingresar datos falsos para obtener ventajas injustas.

Dos de los software mas conocidos son Strava y Zwift. Strava se utiliza para el seguimiento y registro de actividades deportivas al aire libre, como correr, andar en bicicleta y nadar. Además, ofrece funciones sociales que permiten a los usuarios conectarse, seguir a otros atletas y participar en desafíos. Por otro lado, Zwift se especializa en el ciclismo indoor, proporcionando un entorno virtual donde los ciclistas pueden entrenar, competir y socializar en línea. Utiliza sensores en bicicletas estáticas o rodillos para traducir el esfuerzo del usuario en el juego, permitiendo participar en carreras virtuales y entrenamientos estructurados que buscan simular la realidad lo mas posible. Ambas plataformas tienen aplicaciones diferentes dentro del ámbito deportivo y pueden complementarse según los intereses y objetivos del usuario. En las figuras 1.1 y 1.2, se pueden observar las interfaces de usuario de Strava y Zwift respectivamente, donde además se puede observar como se da la interacción ciclista-simulador.

1.5. Física del ciclismo

Cuando se quiere hablar sobre la física en el ciclismo, existen una gran cantidad de factores a considerar, que van desde fenómenos sencillos hasta algunos realmente complejos para los cuales se dedican artículos completos. Algunos de los factores principales son los siguientes:

La dinámica y la cinemática, que permiten explicar el comportamiento de la bicicleta y su interacción con el ciclista. Por ejemplo, la distribución de peso a lo largo de la bicicleta y la ubicación de su centro de gravedad son aspectos muy importantes al momento de maniobrar la bicicleta y tener una buena estabilidad, dado que, si se tuviera un centro



Figura 1.1: Ejemplo de interfaz de usuario del software Strava.

de gravedad muy alto la bicicleta sería mas propensa a inclinarse hacia un lado u otro, pero si se tuviera un centro de gravedad mas bajo, entonces la bicicleta sería mucho mas estable y menos propensa a inclinarse. Por otro lado, el peso del ciclista también juega un papel muy importante en como se comporta la bicicleta, ya que si el ciclista es muy pesado esto genera una mayor fuerza de la bicicleta contra el suelo, o sea que las ruedas tendrán una mayor tracción con el suelo, lo cual puede ser una ventaja, pero también sucede que ese mismo peso luego le dificulta al ciclista poder imprimir una buena aceleración o subir cómodamente en segmentos donde aumenta la altitud.

Fricción en los componentes. Sabemos que uno de los componentes principales de la bicicleta son los rodamientos, que permiten hacer girar las ruedas de la bicicleta, los pedales y el manubrio en algunos casos. La fricción puede llegar a afectar de gran manera el desempeño del ciclista dada la perdida de energía que puede llegar a representar en forma de calor, dado que esta directamente relacionado a la transferencia de potencia del ciclista hacia la bicicleta. Por ende, resulta muy importante tener una buena lubricación y un buen material de rodamiento que minimice lo mas posible la perdida de energía por fricción.

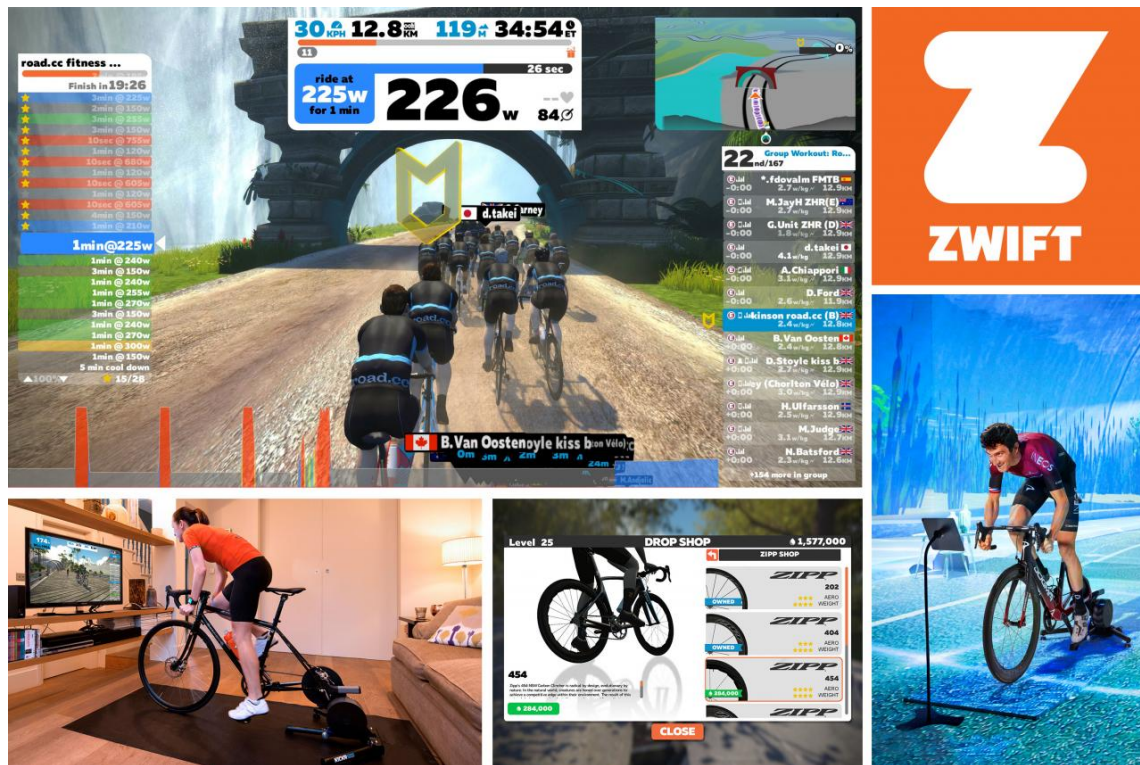


Figura 1.2: Ejemplo de interfaz de usuario del software Zwift.

Otro de los componentes que mayor pérdida de energía pueden representar por fricción son las ruedas y neumáticos, dado que las ruedas están en constante contacto con el suelo, lo que implica que tienen un impacto muy grande en la eficiencia de la bicicleta. Pero en este caso, mas que el material de la rueda, lo que mas impacto tiene es el neumático, que hace variar la resistencia que ejerce la rueda en el suelo dependiendo de la deformación que este produzca. A grandes rasgos lo que cambia es la cantidad de superficie de rueda que hace contacto con el suelo, mientras menos presión tengan los neumáticos, mayor será la superficie de contacto, por ende mayor será también la resistencia ejercida por la fricción, pero esto permitirá una mayor estabilidad de la bicicleta en zonas donde el suelo tenga imperfecciones, y en el caso contrario, mientras mayor sea la presión del neumático, menor será la superficie de contacto con el suelo y mas fácil será el desplazamiento.

Fuerza al pedalear y transmisión de potencia. Detrás de la mecánica propia de la bicicleta existen años de trabajo y mejoras que han permitido optimizar cada vez mas la eficiencia en la transmisión de trabajo del ciclista hacia la bicicleta. En primer lugar los pedales son aquellos que permiten transmitir la fuerza generada por las piernas hacia todo el mecanismo que conecta finalmente con las ruedas. Dentro de este mecanismo se en-

cuentran los pedales, las bielas y los platos, donde las bielas son aquellas que conectan los pedales con los platos. La longitud de las bielas y el tamaño de los platos tiene directa incidencia sobre la potencia generada y la velocidad de la bicicleta. Posteriormente, nos encontramos con la cadena, que debe estar correctamente lubricada para lograr una transferencia eficiente de la potencia minimizando lo mas posible la fricción, que puede darse tanto con los platos como de forma interna con la interacción entre todas las piezas que unen la cadena. Y por ultimo, pero no menos importante, tenemos los cambios de velocidades. Estos son uno de los aspectos mas importantes al momento de maximizar la transferencia de potencia hacia la bicicleta, dado que el estar en un cambio optimo permite al ciclista dosificar su esfuerzo de la mejor manera y mantener una velocidad constante durante los distintos segmentos de la ruta.

Resistencia aerodinámica, uno de los temas mas estudiados en la actualidad y uno de los aspectos que pueden llegar a ser mas determinantes en competición. En palabras simples, la resistencia aerodinámica es la fuerza que ejerce el aire oponiéndose al movimiento del ciclista. Esta puede dividirse en dos componentes principales, la resistencia frontal y de arrastre. La resistencia frontal es básicamente la resistencia que ejerce el aire sobre la parte frontal del ciclista y la bicicleta, o sea, toda la parte del ciclista que tiene en primera instancia, impacto directo con el aire. Por otra parte, la resistencia por arrastre se genera por el desplazamiento del ciclista dentro del fluido, que en este caso es el aire, y como esto depende de muchos factores, los ingenieros dedican mucho tiempo en mejorar todos los aspectos posibles, como la forma de los cascos, la forma de los marcos, e incluso el material y textura de la ropa que usan los ciclistas.

Considerando en términos generales todos estos factores, podemos acercarnos un poco mas a entender realmente de que manera todo esto afecta al ciclista. Primero intentaremos dejar claro un concepto muy importante y en el que se basa este trabajo, que es la potencia.

Basandose en el libro "Fundamentals of physics" [16], la potencia es la relación entre la cantidad de trabajo realizado y el tiempo empleado para hacerlo, y dentro de la misma explicación se utiliza el siguiente ejemplo, "Imagine dos modelos idénticos de un automóvil: uno con un motor de cuatro cilindros de precio base; y el otro con el motor opcional de mayor precio, un potente motor de ocho cilindros. A pesar de las diferencias de motores, los dos autos tienen la misma masa. Ambos suben una carretera cuesta arriba, pero el auto con el motor opcional tarda mucho menos en llegar a la cima. Ambos han realizado la misma cantidad de trabajo contra la gravedad, pero en diferentes períodos de

tiempo”. Gracias a esta definición podemos entender la potencia desde el punto de vista del ciclismo como, la cantidad de energía generada por el ciclista para impulsar la bicicleta superando las resistencias que se le presenten, como pueden ser la resistencia por gravedad que se menciona en el ejemplo, la resistencia por rodadura y la resistencia aerodinámica.

Hoy en día los ciclistas miden la potencia mediante unos medidores conocidos como potenciómetros, de los cuales se hablará en detalle mas adelante, pero básicamente estos dispositivos van instalados en la bicicleta y registran la cantidad de trabajo realizado por el ciclista en tiempo real. Suelen estar ubicados en diferentes partes de la bicicleta como el pedal, la biela o el pedalier, y utilizan distintos métodos para calcular la potencia generada.

Por otra parte, la potencia en el ciclismo depende de muchos factores como la intensidad del esfuerzo, la resistencia que ejerce el viento, el tipo de superficie y la inclinación del terreno. Para revisar esto mas en profundidad nos basaremos en el trabajo de Caroline Cohen, Emmanuel Brunet, Jérémy Roy y Christophe Clanet, en su artículo “Physics of road cycling and the three jerseys problem” [10], donde se embarcan de manera muy detallada en el entendimiento de estos factores. Primeramente, se presenta una ecuación que explica que la dinámica del ciclista se ve gobernada por el balance de energía.

$$\frac{d}{dt} \left(\frac{1}{2} MV^2 + Mgz \right) = P_m - P_f \quad (1.1)$$

V= Velocidad del centro de masa

z= Altura

g= Gravedad

M= Masa total (ciclista+bicicleta)

P_m = Potencia mecanica del ciclista

P_f = Potencia mecanica por fricción

Donde luego se define el termino P_f para despejarlo de la ecuación y definir de esa manera la potencia mecánica del ciclista que es lo que buscamos medir.

$$P_f = \frac{1}{2} \rho S C_D V^3 + \mu M g V, \quad (1.2)$$

ρ =densidad del aire

S =area frontal

C_d = coeficiente de drag

μ = coeficiente de resistencia de rodadura

Y aquí podemos ver como definen la potencia mecánica por fricción con dos atributos, la resistencia aerodinámica y la resistencia por fricción. Sin embargo, como se puede notar en la ecuación la contribución de la resistencia aerodinámica (dado V^3) es mucho mayor a partir de cierta velocidad, que la contribución por resistencia de rodadura. La velocidad para la cual ambas contribuciones son iguales la definen de esta manera:

$$V_\mu = \sqrt{2\mu Mg/\rho SC_D} \quad (1.3)$$

En el estudio se menciona que la contribución por rodadura resulta ser mucho menor que la aerodinámica para las velocidades que se manejan en el Tour de Francia, por lo cual simplemente descartan este termino. Dado que nuestro enfoque no es directamente el Tour de Francia, si tendremos en cuenta el termino de resistencia por rodadura y continuaremos con el procedimiento.

De esta manera, al despejar los términos de la primera ecuación podemos observar que la potencia mecánica del ciclista queda definida de la siguiente forma:

$$P_m = \frac{1}{2}\rho SC_D V^3 + Mg\alpha V + \frac{d}{dt} \left(\frac{1}{2}MV^2 \right) + \mu MgV \quad (1.4)$$

α =inclinación

La ecuación presentada (Ec. 1.4) modela la potencia mecánica (P_m) requerida por un ciclista para vencer las distintas fuerzas de resistencia durante su desplazamiento en una bicicleta. En este análisis, se desglosarán los términos de la ecuación para comprender detalladamente cada componente que contribuye a la potencia total.

En primer lugar, se encuentra la contribución de la resistencia aerodinámica ($\frac{1}{2}\rho SC_D V^3$). Esta componente refleja la energía necesaria para contrarrestar la resistencia del aire, don-

de ρ denota la densidad del aire, S representa el área frontal expuesta del ciclista y la bicicleta, C_D es el coeficiente de resistencia aerodinámica, y V indica la velocidad del ciclista. Es importante destacar que esta resistencia aumenta de manera cúbica con la velocidad.

La segunda contribución proviene de la resistencia gravitatoria ($Mg\alpha V$), que refleja la energía necesaria para superar las variaciones en la elevación del terreno. Aquí, M es la masa total del ciclista y la bicicleta, g es la aceleración debida a la gravedad, y α representa la inclinación de la superficie. Esta componente aborda los desafíos asociados con las pendientes ascendentes y descendentes.

El tercer término ($\frac{d}{dt} (\frac{1}{2}MV^2)$) describe la contribución debida a la aceleración. La derivada respecto al tiempo de la energía cinética proporciona la potencia necesaria para modificar la velocidad del ciclista, ya sea acelerando o desacelerando.

Finalmente, la resistencia por rodadura (μMgV) aporta a la ecuación, representando la energía disipada debido a la fricción entre las ruedas de la bicicleta y la superficie de la carretera. Aquí, μ denota el coeficiente de resistencia por rodadura.

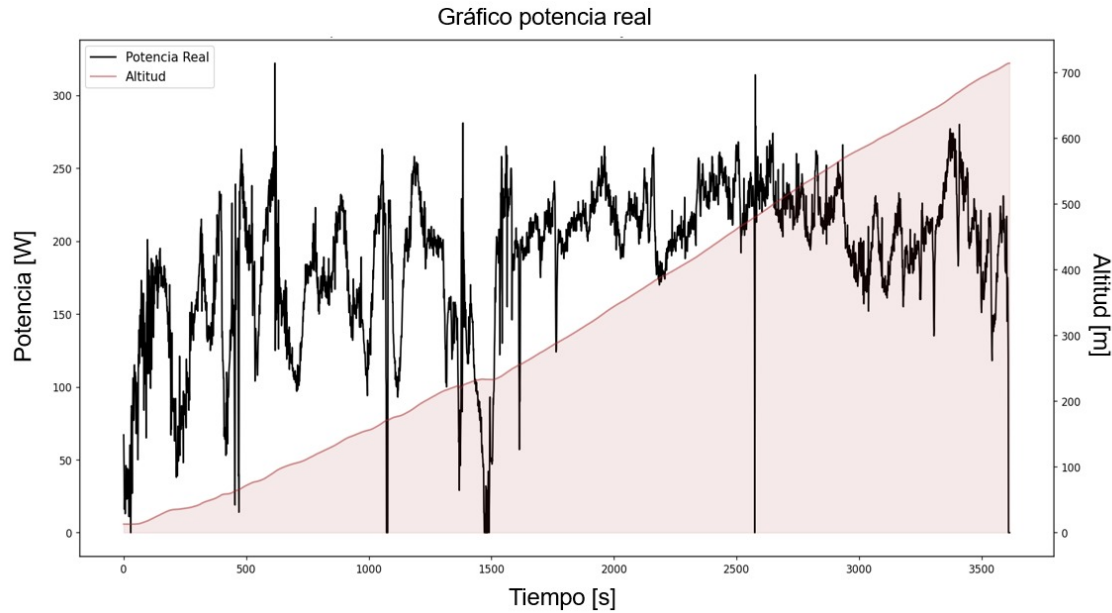


Figura 1.3: Gráfico de potencia real de una carrera.

En resumen, esta ecuación (Ec. 1.4) ofrece un análisis exhaustivo de los diversos factores físicos que influyen en la potencia mecánica necesaria para el desplazamiento de un

ciclista. La comprensión detallada de estos términos es esencial para evaluar el rendimiento y la eficiencia del ciclista en condiciones variables de ciclismo. Para ilustrar la relación entre estos factores y la potencia real generada por un ciclista durante una carrera, se presenta la figura 1.3, que muestra una representación gráfica de la curva de potencia real en función de las condiciones específicas de la ruta y el rendimiento del ciclista. Esta visualización proporciona una perspectiva adicional sobre cómo los componentes de la ecuación se manifiestan en la práctica y cómo la potencia varía a lo largo del tiempo en situaciones reales de ciclismo.

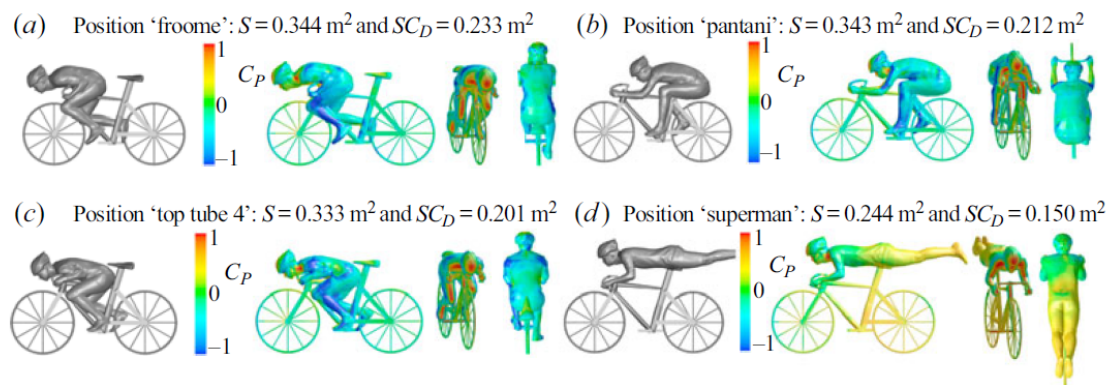


Figura 1.4: Cuatro posiciones diferentes utilizadas durante los descensos con el correspondiente coeficiente de presión estática $C_p = 2(P - P_0)/\rho V_0^2$ en superficies para ciclistas y bicicletas. Aquí, P es la presión estática, P_0 la presión atmosférica y V_0 la velocidad del centro de masa. El área frontal S y el área de arrastre SC_D también se indican para cuatro posiciones significativamente diferentes: (a) posición de Froome, (b) posición de Pantani, (c) top tube 4 y (d) posición de superman. Figura tomada de Cohen [10].

En un análisis más detallado sobre la aerodinámica en el ciclismo, uno de los aspectos cruciales es el efecto del viento sobre la superficie del ciclista, conocido como el área de resistencia frontal o "Frontal Drag Area" (DFA). Los ciclistas buscan minimizar este valor ajustando su posición para reducir la exposición al viento. Sin embargo, algunas posiciones extremas, como se ilustra en la figura 2.6, han sido prohibidas en competiciones debido a riesgos potenciales.

El comportamiento del aire como fluido al fluir sobre el ciclista es un tema complejo. Mientras la parte superior del ciclista suele ser simétrica, la región de la cadera hacia abajo experimenta variaciones en la simetría del tren inferior. En el estudio "Flow Topology in the Wake of a Cyclist and its Effect on Aerodynamic Drag" [11], los investigadores utili-

zan técnicas de túneles de viento para analizar cómo ciertas posiciones del ciclista afectan la resistencia aerodinámica. Se destaca que hay ángulos específicos del ciclo de la biela que no afectan el arrastre (dentro de los primeros 15 grados de rotación) y otros que generan máximos (en las posiciones de 75 y 255 grados).

El estudio también revela la formación de vórtices en la estela del ciclista, influenciados por la posición de las piernas. La altura y extensión de las piernas afectan la generación de vórtices, como se muestra en la figura 1.5. Esto evidencia que el fenómeno aerodinámico va más allá de la simple consideración del área frontal y el coeficiente de arrastre. Algunos factores, como la posición específica de las piernas, son más controlables, mientras que otros pueden ser determinantes y escapan al control del ciclista.

Como ya se comentó, el estudio del arrastre aerodinámico aplicado al ciclismo de manera individual es de vital importancia para comprender el gran trabajo que hay detrás de la optimización del rendimiento. Sin embargo, el análisis del drag en un grupo de ciclistas puede llegar a ser mucho más interesante e importante, dado que los ciclistas a menudo se benefician de la formación de pelotones y de estrategias en equipo. Cuando los ciclistas forman grupos o un gran pelotón, pueden llegar a reducir de manera significativa la resistencia del aire para el grupo. Al estar detrás se puede aprovechar la estela de aire que vimos anteriormente, si esto lo hacen muchos ciclistas al hilo, puede llegar a ser una ventaja brutal para el ciclista que va al final de la fila (no en todos los casos).

En este caso se revisa un estudio llamado 'Aerodynamic drag in cycling pelotons: New insights by CFD simulation and wind tunnel testing' [6], donde a grandes rasgos se estudia como influye el drag en dos configuraciones de pelotones de 121 ciclistas cada uno, esos en formación de carrera. En el pelotón A, las ruedas de los ciclistas de las filas subsiguientes se superponen parcialmente en vista lateral, en cambio, en el pelotón B las ruedas no se superponen, pero son tangenciales entre sí.

Para el experimento se utilizaron dos pruebas, la simulación por CFD (Computational Fluid Dynamics), donde se realiza un escaneo 3D de un ciclista en posición de carrera, y la simulación por tunel de viento, donde se utilizan modelos a una escala de un cuarto de un ciclista real. Antes de realizar las mediciones con las dos configuraciones principales (A y B), se realizaron pruebas previas con grupos más pequeños para validar la configuración y los parámetros de los dos pelotones más grandes.

Los resultados de estas pruebas fueron bastante interesantes. En la figura 1.6 se pueden

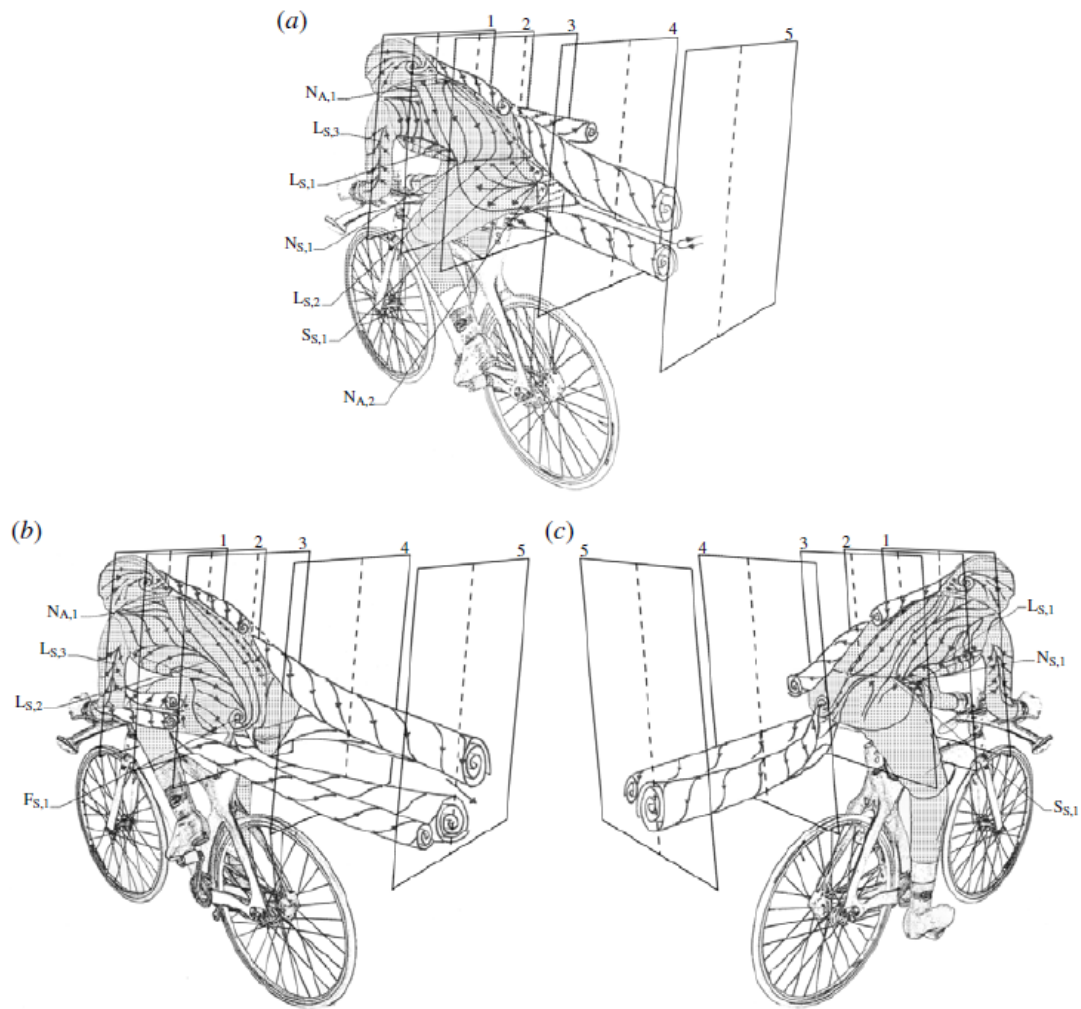


Figura 1.5: Topología de flujo del sistema de vórtice a lo largo de la corriente para (a) 15° simétricos de baja resistencia régimen de flujo y 75° régimen de flujo asimétrico de alta resistencia visto desde (b) la izquierda lado y (c) el lado derecho. Los símbolos L, N, S y F denotan líneas principales, nodos, puntos de silla y focos, respectivamente, con subíndices S y A que indican si están asociados con la separación o el apego. Los números de subíndice se utilizan para identificar puntos críticos al abordar los resultados en apoyo de los patrones de fricción de la piel propuestos. Figura tomada de Crouch [11].

observar dos tipos de grafica generados a partir de la simulación CFD, la de la izquierda representa la fuerza que ejerce el roce con el viento sobre cada configuración de ciclistas y la derecha es el porcentaje de roce con el viento que actúa sobre cada ciclista. Podemos notar que en las tres configuraciones, los primeros ciclistas reciben una cantidad brutalmente mayor que los últimos (casi un 50 % menos), e incluso aquellos que van detrás del

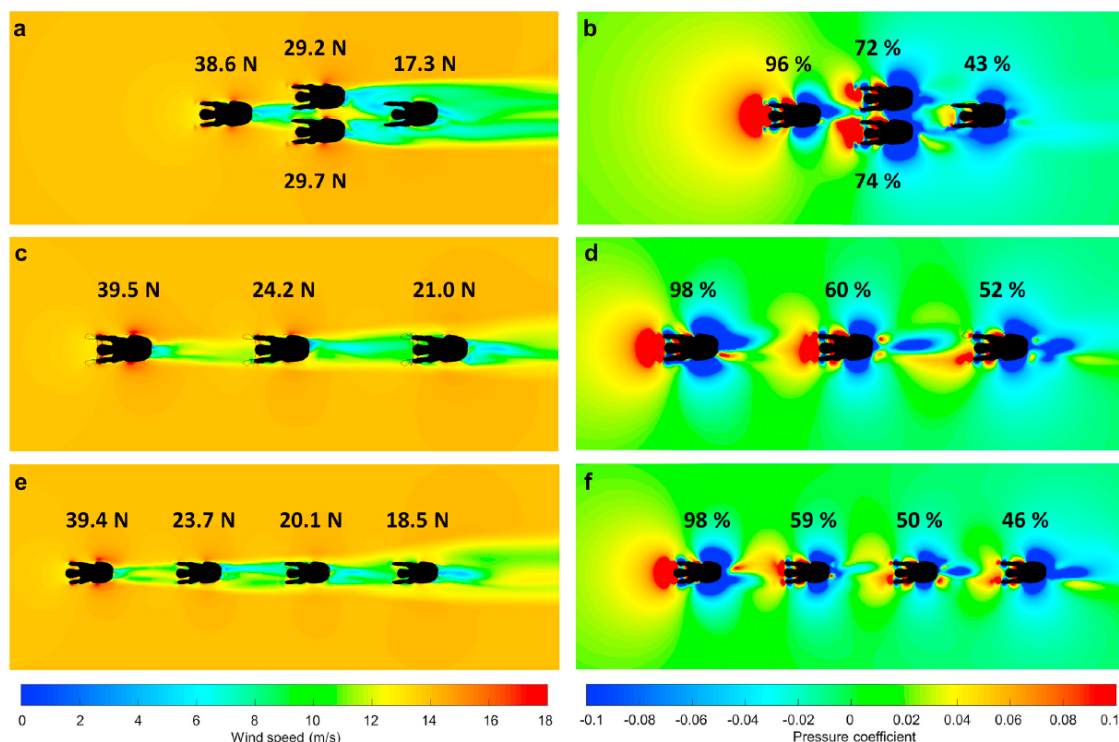


Figura 1.6: Contornos de (a,c,e) velocidad media del viento y (b,d,f) coeficiente de presión media en un plano horizontal a 1 m por encima de la carretera, para las tres subconfiguraciones. Las figuras a,c,e también muestran los valores de las fuerzas de arrastre. Las figuras b, d, f también muestran la resistencia como porcentaje de la resistencia de un ciclista aislado a 15 m/s, que es 40,33 N. Figura tomada de Blocken [6].

líder llegan a sentir un 30-40 % menos de drag que el líder. Con respecto a las pruebas realizadas en túnel de viento, se menciona que las variaciones respecto a la simulación son pequeñas y que se deben a factores experimentales como la vibración de los modelos. Por ende, es evidente que el simple hecho de ir en grupo, aunque sea de pocos ciclistas, puede generar una ventaja abrumadora por sobre aquellos que suelen rodar en solitario. Esto para casos donde los ciclistas se vean enfrentados a un roce muy alto con el aire, ya sea por altas velocidades o por la velocidad misma del viento, ya que por el contrario, en una escalada pronunciada no suele afectar de manera tan significativa el drag.

Posteriormente, proceden a realizar las pruebas en CFD y en túnel de viento para ambos pelotones (A y B). En las figuras 1.7 y 1.8 observamos el porcentaje de drag de cada uno de los 121 ciclistas en ambos pelotones. En estas graficas podemos observar muchos detalles interesantes sobre el comportamiento del drag en grupos masivos de ciclistas, en

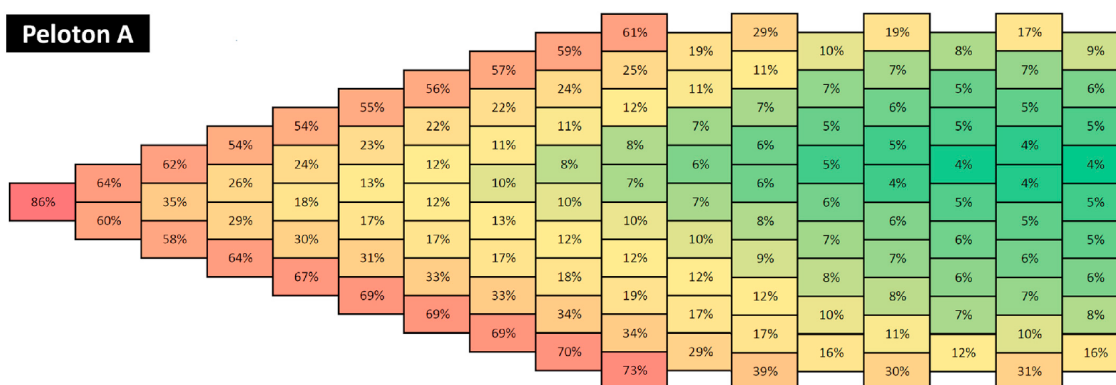


Figura 1.7: Arrastre de cada ciclista en el pelotón A como porcentaje del arrastre de un ciclista aislado que circula a la misma velocidad.

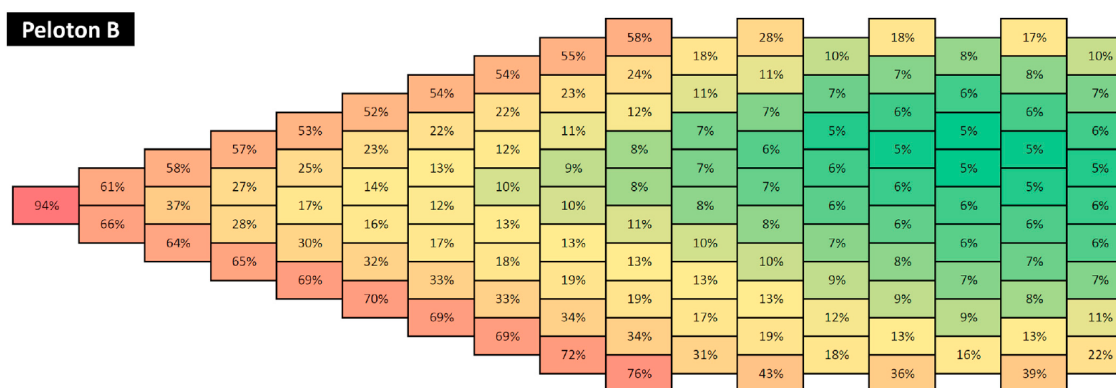


Figura 1.8: Arrastre de cada ciclista en el pelotón B como porcentaje del arrastre de un ciclista aislado que circula a la misma velocidad.Figura tomada de Blocken [5].

primer lugar podemos observar como se comporta el aire como fluido al tener contacto con el grupo, donde los ciclistas que se encuentran en la parte externa del grupo son los que actúan como una especie de caparazón, recibiendo todos (los que están en forma de flecha) no menos del 50 % de drag que recibiría un ciclista en solitario. Posterior a esta caparazón podemos ver que van disminuyendo los porcentajes a medida que nos acercamos al centro del pelotón, donde pueden llegar a recibir porcentajes impresionantemente pequeños de drag, como son 7-5 %, prácticamente van rodando casi sin roce del viento. Otro detalle que puede pasar por alto es la diferencia entre el porcentaje de los líderes de ambos pelotones. Principalmente, el hecho de que no estén sometidos al 100 % de drag que recibiría un ciclista individualmente es por un efecto que llaman "Subsonic upstream disturbance". La perturbación subsónica que se menciona se refiere al efecto que ocurre cuando un objeto en movimiento crea una zona de alta presión y turbulencia, en este caso

los ciclistas que están detrás del líder generan una estela turbulenta de aire debido a su presencia y movimiento, entonces cuando el líder del pelotón avanza, esta estela turbulenta actúa como una especie de 'colchón' de aire que reduce la resistencia aerodinámica que experimenta el líder del grupo [5]. Lógicamente, al tener a los dos ciclistas mas cerca del líder en el pelotón A, este efecto es mayor y disminuye un poco mas su porcentaje de drag. Como ultimo detalle, la desviación que se ve en la figura es debido a que los ciclistas están posicionados con una pierna mas arriba que la otra, lo cual afecta la simetría del problema y permite además observar cuando puede afectar la posición de las piernas en el flujo del aire.

De esta manera, se entiende mejor la complejidad del estudio de la fisica en el ciclismo, y la cantidad de fenómenos a tener en cuenta en las distintas situaciones a lo largo de una carrera. Incluso hay estudios que se enfocan en fenómenos mucho mas específicos [25], donde se discute que tanto afecta al drag la distancia que hay entre la rueda trasera del líder y la delantera del seguidor, o incluso la alineación entre los ciclistas.

1.6. Problemáticas actuales del ciclismo

Cuando abordamos el tema del rendimiento en el ciclismo, debemos considerar una combinación de aspectos físicos, técnicos, tácticos e incluso psicológicos. Gracias a los avances en el estudio del cuerpo y la mente humana, sabemos que el factor psicológico desempeña un papel crucial en el rendimiento deportivo. Una parte del nivel de rendimiento de un deportista puede depender puramente de sus entrenamientos, como su potencia, fuerza y capacidad aeróbica, áreas del deporte que han sido investigadas durante años y cuentan con mucho conocimiento de fácil acceso.

Sin embargo, existen otros aspectos del rendimiento que carecen de facilidad para su aplicación y conocimiento general debido a la falta de estudios y discusión. Ejemplos incluyen temas de nutrición y alimentación especializados para el deporte, así como técnicas especializadas para mejorar el rendimiento en entrenamientos estándar, como la utilización de medidores de rendimiento, como sensores de cadencia, pulsómetros, potenciómetros e incluso software de análisis de datos.

A lo largo del tiempo, algunos de estos dispositivos han llegado a estar al alcance de

muchas personas, como los pulsómetros, medidores de cadencia y ciclocomputadores. Sin embargo, productos como los potenciómetros son más comunes entre ciclistas de élite para mejorar el rendimiento. El problema principal radica en que estos productos suelen estar fuera del alcance económico de ciclistas amateurs, quienes buscan avanzar en su pasión por el ciclismo.

Inicialmente, un ciclista amateur buscará mejorar su bicicleta para acotar su desarrollo y desempeño, lo cual podría llevarlo a intentar armar su propia bicicleta, con un costo aproximado de 1000-2000 dólares para una bicicleta de buena calidad. Posteriormente, la adquisición de equipo deportivo acorde a las intenciones del deportista podría costar fácilmente 500 dólares más. Esto suma un total de 1500-2500 dólares solo para dar los primeros pasos en el ciclismo amateur, lo que ya representa un presupuesto considerable.

Por otro lado, un potenciómetro 1.9 en el contexto del ciclismo es un dispositivo que mide la potencia que un ciclista genera al pedalear. Este instrumento, crucial para el entrenamiento y rendimiento deportivo, puede instalarse en diversas partes de la bicicleta, como las bielas, el eje del pedal o el buje de la rueda trasera. Funciona midiendo la torsión, deformación o fuerza aplicada durante el pedaleo y luego convierte estos datos en una estimación precisa de la potencia generada en vatios. La lectura de potencia proporciona una medida objetiva y cuantitativa de la intensidad del esfuerzo, permitiendo un monitoreo en tiempo real, análisis detallado del rendimiento y una planificación más efectiva del entrenamiento para ciclistas de todos los niveles. El costo de estos aparatos suele oscilar entre 700-1200 dólares. Esto contribuye a que el ciclismo se perciba cada vez más como un deporte de élite, donde el acceso a ciertos dispositivos especializados puede representar un desafío financiero para los ciclistas aficionados.

Esto ha llevado a diferentes personas a buscar soluciones mas económicas para alcanzar un mejor rendimiento, y aquí es donde se abren muchas puertas en el área de investigación, donde se comienzan a buscar opciones, como por ejemplo intentar relacionar datos como, la cadencia, el ritmo cardiaco, la velocidad, y muchas variables mas, con el fin de obtener ese preciado dato que es la potencia. Durante los últimos años han habido variadas investigaciones donde se intentan comparar por ejemplo entrenamientos por zonas de frecuencia cardiaca y entrenamientos por zonas de potencia, y observar si la diferencia es realmente significativa. Si consideramos el entrenamiento basado en potencia y el entrenamiento basado en frecuencia cardiaca, los datos nos dicen que los dos métodos de entrenamiento dieron como resultado entrenamientos aproximadamente equivalentes [37].



Figura 1.9: Ejemplo de 3 tipos de potenciómetros distintos. 1) Potenciómetro de biela. 2) Potenciómetro de buje. 3) Potenciómetro de pedal.

Desde una perspectiva práctica, el estudio sugiere que, para ciclistas recreativos promedio, no hay una ventaja clara en el uso de PM en comparación con HR para mejorar el rendimiento y los cambios fisiológicos. Ambos métodos resultaron en cargas de entrenamiento equivalentes y efectos similares en el rendimiento. Aunque se elogian las ventajas teóricas de la PM en cuanto a precisión y respuesta temporal, el estudio no encontró evidencia empírica de su superioridad sobre el entrenamiento basado en HR.

1.7. Machine learning: breve introducción en contexto

Hoy en día el Machine learning o aprendizaje automático tiene cada vez mas reconocimiento, constantemente lo podemos ver en herramientas de procesamiento de texto, procesamiento de bases de datos enormes, resolución de problemas, identificación de patrones, entre muchas otras cosas. Y una de sus principales ventajas es la facilidad con la que puede manejar y modelar cantidades abrumadoras de datos. Esta tecnología se suele

basar en modelos estadísticos y algoritmos matemáticos que mediante una gran entrada de datos permiten a la maquina aprender y reconocer patrones para generar predicciones. Y cada vez mas podemos ver como se aplican este tipo de tecnologías a los deportes, como el futbol o el basquetbol, y en esta ocasión la intención es aplicar esta tecnología al campo del ciclismo con la finalidad de encontrar y utilizar la correlación entre distintos tipos de datos ciclisticos para la predicción a tiempo real de la variable de potencia, e incluso la predicción de posibles curvas que cambien a partir de eventos que puedan suceder en la carrera, como por ejemplo un sprint repentino o una separación del grupo, cosas que no están generalmente planeadas y que permitan al ciclista poder tomar este tipo de decisiones en medio de la carrera con la confianza de que se podrá estimar una nueva curva de potencia acorde al esfuerzo realizado en tiempo real, lo que le permita saber el esfuerzo mas o menos que debe hacer para terminar la carrera de la mejor manera.

Para entender un poco mejor la estructura que se debe seguir para aplicar una machine learning al problema que se quiere resolver, podemos basarnos en la estructura definida por Richter [36], donde este proceso consta de 4 pasos principales:

1. Captura y procesamiento de datos: Es importante recopilar un conjunto de datos lo suficientemente grande y relevante para la tarea en cuestión. La cantidad de datos es importante, pero la selección adecuada de características extraídas de esos datos es aún más crucial para el rendimiento del modelo.

2. Selección de características: Se debe realizar una selección cuidadosa de las características que se utilizarán como entrada para el modelo de Machine Learning. Es recomendable combinar enfoques basados en datos y conocimientos específicos del dominio para obtener un conjunto óptimo de características.

3. Modelado: Seleccionar la técnica de Machine Learning adecuada para el problema en cuestión. Esto implica elegir entre diferentes enfoques como regresión, árboles de decisión o redes neuronales. Además, es importante considerar la normalización de características y el equilibrio de los datos para obtener un rendimiento óptimo del modelo.

4. Evaluación: Se deben utilizar criterios de evaluación apropiados para estimar el rendimiento del modelo. La precisión es una medida comúnmente utilizada para problemas de clasificación, mientras que medidas como el error promedio o el coeficiente de determinación (r^2) son relevantes para modelos de predicción de magnitud. También es esencial realizar una evaluación adecuada utilizando conjuntos de datos separados para prueba y

validación cruzada, evitando la contaminación de datos que puede conducir a una sobreestimación del rendimiento del modelo.

La aplicación del aprendizaje automático (ML) en el ámbito deportivo ha abierto numerosas oportunidades y aplicaciones, tanto en deportes grupales como individuales. Acá se tienen cuatro artículos que permiten ilustrar la diversidad de modelos, enfoques de selección de características, tamaños de conjuntos de datos y técnicas de validación utilizadas en la aplicación del ML en el ámbito deportivo, lo que demuestra la flexibilidad y las múltiples formas de abordar el análisis de datos en el contexto deportivo.

- “The application of machine learning and deep learning in sport: predicting NBA players’ performance and popularity [31]”
- “The Application of Machine Learning Techniques for Predicting Match Results in Team Sport: A Review [9]”
- “Improving Sports Outcome Prediction Process Using Integrating Adaptive Weighted Features and Machine Learning Techniques [27]”
- “A game-predicting expert system using big data and machine learning [15]”

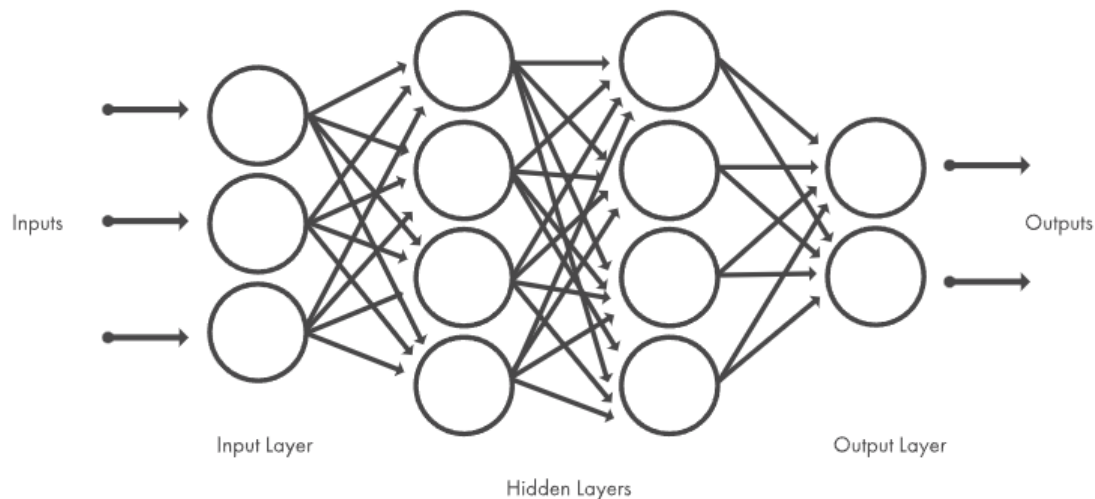


Figura 1.10: Ilustración de la estructura mas basica de una Red Neuronal

Como se mencionó anteriormente, en el ámbito del machine learning, hay diversos enfoques que se adaptan a diferentes tipos de problemas. Comencemos con el aprendizaje

supervisado: este enfoque implica entrenar modelos con datos etiquetados. Un ejemplo clásico es el reconocimiento de dígitos escritos a mano, donde el modelo se entrena con imágenes de números y sus correspondientes etiquetas. Durante el proceso de entrenamiento, el modelo examina las características visuales distintivas de cada dígito, como sus curvas y contornos, y ajusta sus parámetros internos para poder predecir la etiqueta de nuevos dígitos con alta precisión.

Pasando al aprendizaje no supervisado, este enfoque se centra en datos no etiquetados, buscando patrones y estructuras intrínsecas en los datos. Un ejemplo común es el clustering, utilizando algoritmos como k-means. Se tiene un conjunto de datos con características de diferentes tipos de frutas, pero sin etiquetas. Con el aprendizaje no supervisado, el algoritmo podría analizar las similitudes y diferencias entre las frutas y agrupar automáticamente aquellas que comparten características comunes. Este proceso proporciona una organización natural de los datos sin la necesidad de información previa sobre las categorías de las frutas.

En el aprendizaje por refuerzo, los modelos aprenden tomando decisiones y reciben retroalimentación en forma de recompensas o castigos. AlphaGo es un ejemplo destacado en este caso. Este programa de inteligencia artificial aprendió a jugar Go compitiendo contra sí mismo y ajustando su estrategia en función de las jugadas ganadoras. Cada movimiento exitoso es una recompensa”, lo que permite al modelo mejorar sus tácticas y estrategias a lo largo del tiempo.

Dentro del ámbito de la deep learning, las redes neuronales son fundamentales para tareas como el reconocimiento de patrones en secuencias temporales. Supongamos que se está trabajando con datos de series temporales, como el comportamiento del mercado de valores a lo largo del tiempo. Utilizando una red neuronal recurrente (RNN) o una red neuronal de memoria a largo plazo (LSTM), se puede entrenar el modelo para capturar patrones complejos y dependencias temporales en los datos. En este caso, la red aprende a reconocer correlaciones y patrones específicos en las fluctuaciones del mercado, permitiendo predicciones futuras basadas en esos patrones aprendidos.

Estos ejemplos ilustran solo algunas de las aplicaciones más conocidas de los diferentes enfoques en machine learning, mostrando cómo cada uno aborda problemas específicos y contribuye al desarrollo continuo de la inteligencia artificial.

En este caso particular, el enfoque más destacado para abordar el problema de reconocimiento de patrones en secuencias temporales es el uso de deep learning, específicamente mediante el empleo de redes neuronales. Dada la naturaleza secuencial de los datos, las redes neuronales recurrentes (RNN) o las redes neuronales de memoria a largo plazo (LSTM) resultan ser modelos ideales. La capacidad de las redes neuronales para

aprender dependencias temporales complejas las convierte en herramientas poderosas en la predicción y reconocimiento de patrones en series temporales. En este contexto, se iniciará explorando la forma más básica de una red neuronal para establecer los cimientos del modelo adaptado a la naturaleza secuencial de los datos.

Como se explica en Mathworks [22], una red neuronal en su forma mas simple es un modelo de computación que imita la estructura interconectada de las neuronas en el cerebro. Funciona combinando capas de procesamiento con nodos interconectados, y se inspira en sistemas nerviosos biológicos. Está compuesta por una capa de entrada, una o varias capas ocultas y una capa de salida, donde cada capa utiliza la salida de la capa anterior como entrada. Las redes neuronales pueden aprender de datos para reconocer patrones, clasificar datos y predecir eventos futuros, ajustando automáticamente las conexiones entre los nodos durante el entrenamiento.

Para entender esta estructura, nos basamos en la gráfica 1.10:

1. Entrada de Datos (Input Layer):

- Cada una de las neuronas en la capa de entrada representa una característica específica (Inputs), como velocidad, cadencia, inclinación y tiempo. Estos valores se toman en un instante de tiempo específico.

2. Capa de Entrada LSTM y capas Ocultas (Hidden Layers):

- Cada neurona de la capa de entrada LSTM tiene su propia celda de memoria. La velocidad, cadencia, inclinación y tiempo alimentan simultáneamente a todas estas celdas de memoria.
- Cada celda de memoria decide cómo retener o olvidar información pasada y cómo incorporar nueva información, permitiendo a la red manejar patrones temporales.

3. Capa de Salida (Output Layer):

- La capa de salida tiene una neurona, que predice el valor final de la potencia (Output). La información de las celdas de memoria LSTM y las capas ocultas anteriores se utiliza para calcular esta predicción.
- La relación entre las salidas de las capas ocultas y la neurona de salida se aprende durante el entrenamiento.

4. Entrenamiento:

- Durante el entrenamiento, la red LSTM ajusta automáticamente los pesos de todas las conexiones para minimizar la diferencia entre las predicciones y los valores reales de la potencia. Se utiliza una función de pérdida específica para problemas de regresión, como el Mean Squared Error.

Esta arquitectura demuestra la eficacia de las redes neuronales secuenciales, especialmente en la interpretación y predicción de datos temporales. Al permitir que la red aprenda las relaciones complejas entre las características temporales, como cambios en la velocidad o la inclinación, el modelo se adapta de manera óptima a la naturaleza secuencial de los datos, demostrando así la utilidad de este enfoque en la resolución de problemas de reconocimiento de patrones en secuencias temporales.

1.8. Retos actuales en física del ciclismo y sus aplicaciones

La física del ciclismo plantea desafíos fascinantes en la actualidad, desde los principios fundamentales de la física clásica hasta cuestiones más complejas, como la biomecánica. Uno de los desafíos más complejos que enfrenta este deporte es la resistencia del aire, una barrera principal para aumentar la velocidad que puede ser determinante en condiciones climáticas desfavorables. Reducir esta resistencia se ha convertido en un objetivo crucial, llevando a la búsqueda constante de implementaciones y equipos cada vez más aerodinámicos.

La biomecánica, por otro lado, despliega una importancia igualmente significativa. La forma en que un ciclista pedalea, su postura y su técnica pueden tener un impacto sustancial en el rendimiento. Optimizar estos aspectos se ha convertido en un desafío continuo en el ciclismo, ya que entender y mejorar la eficiencia biomecánica puede marcar la diferencia en competiciones de alto nivel.

Además, la transferencia de energía, a menudo subestimada, juega un papel crucial. Para que un ciclista aproveche al máximo su energía y avance de manera eficiente, se requieren componentes especializados capaces de transferir energía de manera óptima, minimizando las pérdidas desde los pies del ciclista hasta el suelo.

La aplicación de machine learning en este contexto agrega una dimensión innovadora. Al utilizar algoritmos y modelos, se pueden integrar y analizar datos complejos para comprender aún más las variables críticas que afectan el rendimiento del ciclista. Esto no solo permite una mejor optimización de los factores conocidos, como la resistencia del aire y la

biomecánica, sino que también abre la posibilidad de descubrir datos hasta ahora desconocidos y potencialmente más determinantes que la potencia, enriqueciendo la comprensión y mejora del rendimiento ciclístico.

Capítulo 2

Datos y métodos

2.1. Obtención de datos

Una de las principales cualidades que distinguen a esta tesis es su carácter experimental, lo que implica enfrentarse a un desafío crucial en el campo del ciclismo: la obtención de datos. De hecho, se reconoce ampliamente que los datos requeridos para llevar a cabo una investigación precisa y significativa en este ámbito son extremadamente difíciles de conseguir. Esto debido a diversas razones.

En primer lugar, el ciclismo es un deporte que se desarrolla en entornos variados. Los ciclistas suelen recorrer diferentes rutas, terrenos y condiciones climáticas, lo que implica una gran complejidad al momento de obtener datos precisos y coherentes. Además, el acceso a ciertos datos, como la frecuencia cardíaca, la cadencia o la potencia generada por el ciclista, requiere el uso de dispositivos medidores especializados. Si bien existen tecnologías avanzadas, como monitores de frecuencia cardíaca y medidores de potencia, no todos los ciclistas poseen estos dispositivos, lo que limita aún más la disponibilidad de datos para la investigación.

Otro obstáculo es la privacidad y la confidencialidad de los datos de los ciclistas. La recolección de datos biométricos y de rendimiento puede plantear preocupaciones éticas y legales, dado que un mal manejo de la privacidad de los datos puede perjudicar la integridad competitiva de un ciclista, lo que obliga a los investigadores a abordar estas cuestiones con responsabilidad y asegurar el anonimato y la protección de la información personal de los participantes.

Para llevar a cabo la recolección de datos, se exploraron diversas opciones como generar datos propios, contactar con algún profesional del deporte, con ciclistas amateur, buscar datos en internet, entre otras. Finalmente se decidió contactar con un grupo de ciclistas local conocido como 24BIKE con el propósito de solicitar su colaboración para proporcionar los datos necesarios para este estudio. Con este fin, se organizó una reunión virtual a través de Zoom, donde se presentó una visión general del proyecto de investigación. Durante la reunión, se les explicó a los ciclistas que se requería una cantidad significativa de datos por parte de ellos para avanzar en la investigación y al finalizar se realizó la propuesta de hacer un recorrido en bicicleta para conocernos mejor, y conversar detalles más puntuales. Un par de días después se concretó la propuesta y nos reunimos para realizar uno de los recorridos más comunes de ellos, que suelen utilizar para ciclistas principiantes.



Figura 2.1: En esta foto se observa la primera reunión con el grupo de ciclismo 24BIKE, donde junto con Sergio, Hernán y el profesor guía Juan Manuel, realizamos una ruta en bicicleta bastante conocida por ellos por formar parte de las primeras rutas de iniciación para nuevos integrantes de su grupo.

2.2. Datos experimentales - Primera recolección

Para la primera instancia de recolección se redactó un instructivo extenso, donde se indicaban todos los detalles importantes que se tenían en cuenta tanto para seleccionar a los participantes, como para seleccionar que datos serían interesantes y cuales no. Algunos de los criterios fueron los siguientes:

- 7-10 ciclistas por genero.
- 3 rutas distintas con 3 muestras por c/u.
- Datos con menos de 5 años de antigüedad.

Datos experimentales



Figura 2.2: Detalle de características con unidades de medida. Cada número corresponde a la etiqueta de la característica en los modelos de entrenamiento.

Teniendo esto en cuenta, la idea general era que los participantes buscaran rutas que hubieran realizado en común, para obtener varias muestras de la misma ruta, y que idealmente tuvieran un mínimo de tres muestras por cada ruta que eligieran. Para la extracción de estos datos, se utilizó una aplicación bastante conocida en el mundo del ciclismo llamada Strava. Desde la cual, se pueden exportar datos personales de recorridos ya realizados en dos formatos principales, TCX y GPX. Datos que posteriormente debían subir a una

carpeta de Dropbox creada especialmente para la recolección. Durante este periodo se resolvieron muchas dudas de los participantes y una vez finalizado el periodo de recolección, se descargan todos los datos recolectados, y se procesan mediante un archivo de Python creado especialmente para extraer los datos de los archivos en formato tanto GPX como TCX. A través de este código, se logra tanto la extracción como la generación de nuevos datos a partir de los originales, resultando en un total de 8 tipos de datos, como se muestra en la figura 2.2.

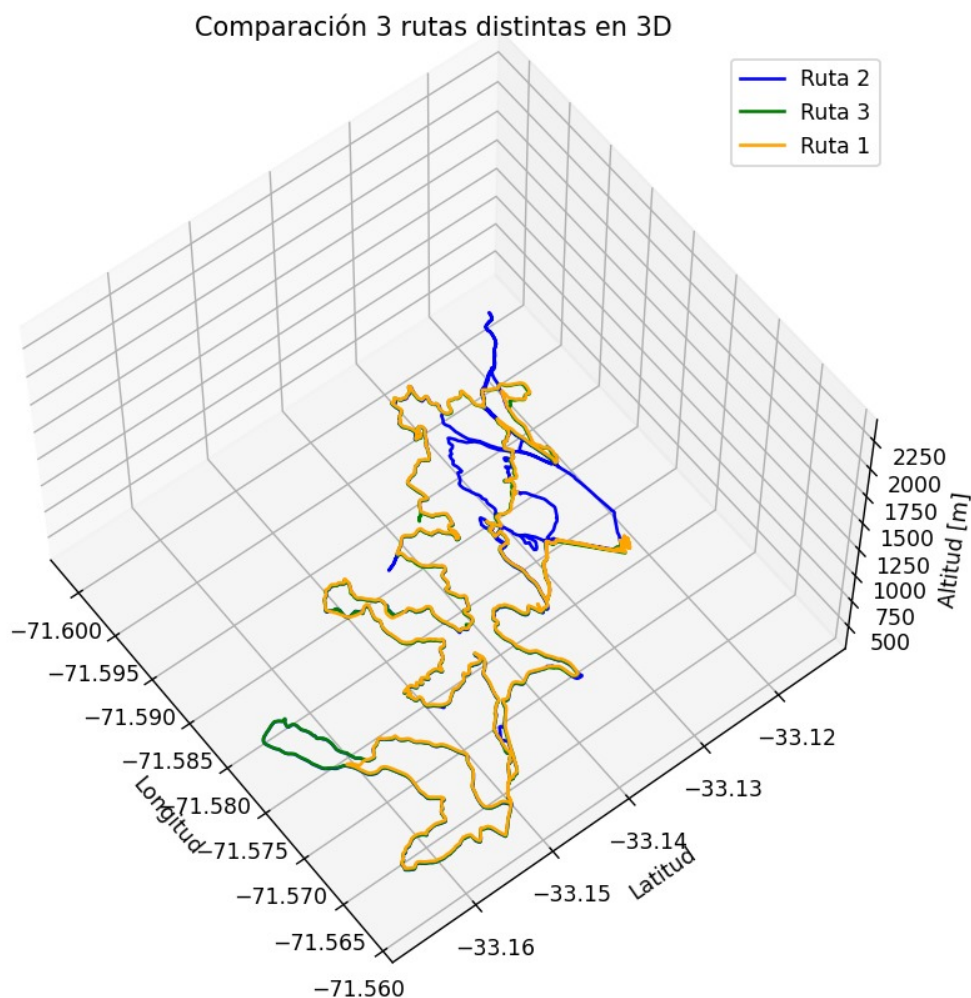


Figura 2.3: Comparación de 3 rutas del mismo participante, con variaciones en el recorrido.

En esta etapa inicial de recopilación de datos, se identifican oportunidades para mejorar la organización y consistencia en el manejo de la información. En primer lugar, se

observa la ausencia de carpetas designadas para clasificar los datos por rutas, lo que dificulta la identificación precisa en casos donde las rutas no están nombradas. Cabe destacar que esta falta de organización se debe a la decisión de permitir a los participantes la "libre elección" de las rutas comunes que podrían seguir, lo cual imposibilitó la creación anticipada de estas carpetas.

El principal desafío abordado para mitigar esta problemática fue la necesidad de ordenar visualmente las rutas. Para ello, se implementó un enfoque que consistió en imprimir las curvas de manera individual en un entorno tridimensional utilizando la biblioteca de Python, Matplotlib. Este proceso permitió examinar detenidamente cada ruta y identificar visualmente las variaciones. Posteriormente, se llevó a cabo una organización adicional de las rutas, agrupándolas según similitudes y asignándoles nombres comunes.

En segundo lugar, se detectan variaciones en las rutas a lo largo de la recopilación de datos. Se han identificado discrepancias en términos de longitud y trayectoria, donde algunas carreras comparten segmentos comunes hasta cierto punto, pero posteriormente divergen en direcciones, generando trayectos distintos.

Esta variabilidad en las rutas presenta un desafío significativo para el desarrollo de modelos de aprendizaje automático. Se consideraba crucial que las rutas de entrenamiento fueran consistentes, ya que se creía que esto proporcionaría un patrón más claro para que la machine learning aprendiera de manera efectiva. La uniformidad en las rutas se percibía como esencial en esta fase de investigación, permitiendo que los algoritmos de machine learning captaran patrones de manera más eficiente y contribuyeran así al éxito del proceso de entrenamiento.

Esta problemática se ilustra de manera destacada en la gráfica 3D 2.3, donde se muestra uno de los recorridos realizados por un participante. Aunque estos recorridos deberían ser idénticos, se observan variaciones bruscas en ciertas zonas específicas, generando ruido en los datos de entrenamiento utilizados para desarrollar modelos de machine learning.

Este proceso de ordenamiento y clasificación fue esencial para establecer una base de datos más coherente y consistente. La identificación y corrección de las variaciones en las rutas no solo contribuyó a mejorar la calidad de los datos de entrenamiento, sino que también facilitó el posterior análisis estadístico y permitió el entrenamiento de las primeras variantes más básicas de los modelos de machine learning. Este enfoque no solo fortaleció la integridad de los datos, sino que también sentó las bases para un desarrollo más efectivo de los modelos predictivos.

2.3. Analisis estadístico - Primera recolección

Una vez que los recorridos fueron organizados en carpetas, se procedió a realizar un análisis estadístico para obtener una comprensión más profunda de los datos recopilados. En una primera fase, se aplicó estadística básica utilizando la biblioteca Pandas [33], específicamente la función `describe`, para examinar características clave como la velocidad. Este análisis implicó la selección de rutas específicas, donde se construyó una matriz con todas las velocidades correspondientes. A partir de esta matriz, se calcularon estadísticas fundamentales como la media, desviación estándar, mediana, entre otros, mediante la función `describe` de Pandas. El objetivo principal de esta fase fue identificar posibles irregularidades, tales como datos faltantes o valores incoherentes, asegurando así la integridad de la información.

En el siguiente paso, se abordó el análisis de correlación entre las características y la potencia utilizando cinco métodos estadísticos distintos:

1. **Pearson:** Este método mide la correlación lineal entre dos variables continuas. Proporciona un valor entre -1 y 1, donde 1 indica una correlación positiva perfecta, -1 indica una correlación negativa perfecta, y 0 indica falta de correlación.
2. **Spearman:** A diferencia de Pearson, Spearman evalúa la correlación no lineal entre variables. Se basa en el ranking de las observaciones y proporciona un coeficiente de correlación de rango.
3. **Feature Importance:** Este método destaca la importancia relativa de cada característica en relación con la variable objetivo (en este caso, la potencia). Evalúa qué características tienen más peso en la predicción.
4. **RFE (Recursive Feature Elimination):** RFE es un enfoque que selecciona de manera recursiva un subconjunto de características y evalúa su impacto en el modelo. Elimina gradualmente las menos relevantes hasta alcanzar el número óptimo de características para la predicción.
5. **PearsonR:** Similar a Pearson, este método se centra en medir la correlación lineal, proporcionando una visión específica de la relación entre dos variables continuas.

Estos cinco métodos estadísticos fueron clave para iniciar las pruebas de machine learning, identificaron las características más relevantes para predecir la potencia en las carreras. Estos perfiles (resultados en la figura 3.9) proporcionaron una base lógica para

	PEARSON	SPEARMAN	F. IMPORTANCE	RFE	PEARSONR
CADENCIA	✓				✓
COORDENADAS					
VELOCIDAD				✓	
F. CARDÍACA			✓		✓
TIEMPO	✓	✓	✓	✓	
DISTANCIA					
ALTITUD					
INCLINACIÓN	✓	✓	✓	✓	✓

Figura 2.4: Aquí se observa cuáles son las variables que obtuvieron mayor correlación con la potencia para cada método estadístico.

seleccionar características y se tradujeron en un conjunto de datos de entrenamiento más eficiente.

2.4. Datos experimentales - Segunda recolección

En la segunda instancia de recolección, se simplificaron significativamente las indicaciones para los participantes, enfocándonos en dos elementos clave: el rodillo de entrenamiento y el medidor de frecuencia cardíaca. Todos los participantes fueron instruidos a utilizar un “Smart Trainer”, también conocido como Rodillo de transmisión directa inteligente, como se ilustra en la figura 2.5, junto con un medidor de frecuencia cardíaca. Estos Smart Trainers, en su mayoría, prescinden de la rueda trasera y se conectan directamente a la bicicleta mediante la cadena y los platos, mientras el marco de la bicicleta se acopla a la estructura del rodillo. Destaca su funcionamiento magnético basado en imanes de neodimio, caracterizándose por su notable silenciosidad.

Con el fin de establecer las rutas que los participantes seguirían durante sus simulaciones, se consideraron varios aspectos detallados a continuación. En primer lugar, se hizo hincapié en la utilización de la plataforma Zwift, reconocida como un simulador virtual de ciclismo. Zwift no solo proporciona una experiencia inmersiva de ciclismo en interior-



Figura 2.5: Ejemplo de un ciclista entrenando mediante la simulación de un recorrido en vivo utilizando el software Zwift.

res, sino que también sirve como una herramienta para medir y registrar datos importantes durante las sesiones de entrenamiento.

Es crucial señalar que Zwift presenta ciertas limitaciones temporales en cuanto a las rutas disponibles, lo que implicó una cuidadosa selección por parte del equipo de investigación. Para garantizar la diversidad de opciones en todo momento, se eligieron cuatro rutas principales basadas en criterios como la dificultad, el porcentaje de inclinación y la disponibilidad en la aplicación, como se detalla en la imagen 2.6.

La elección de estas rutas no solo se basó en su variedad y desafío, sino también en la consideración de la dinámica temporal de Zwift. La plataforma puede habilitar o deshabilitar rutas específicas en intervalos regulares, lo que influyó en la necesidad de seleccionar múltiples opciones. Se garantizó que, al menos, una de las cuatro rutas estuviera siempre disponible para los participantes.

En la segunda instancia de recolección, se constató que los datos mostraban una distribución normal. Con este punto de partida, el siguiente paso consistirá en realizar un

análisis estadístico detallado para explorar variabilidades, tendencias y posibles relaciones entre las variables recopiladas durante las simulaciones en el Smart Trainer.

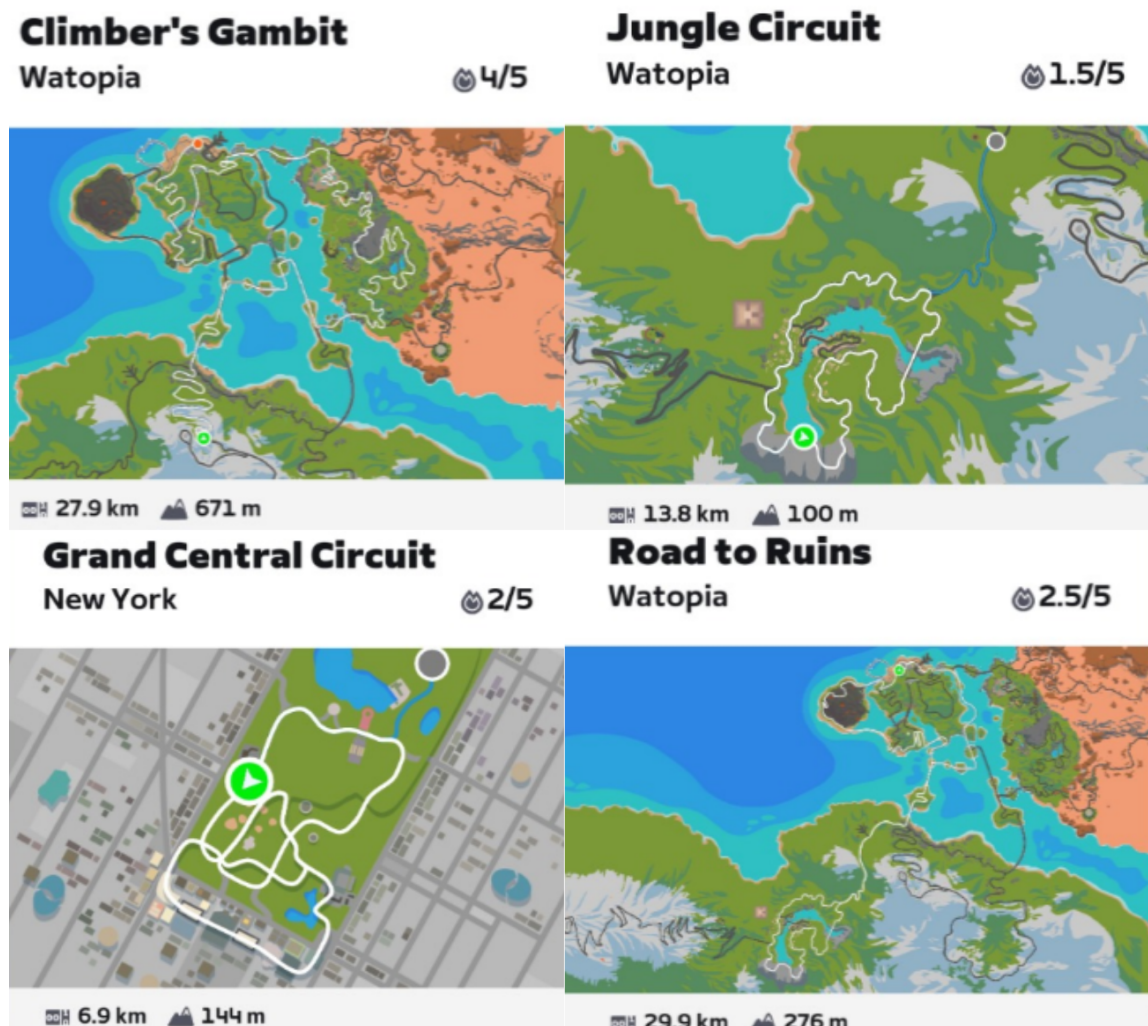


Figura 2.6: Aquí se observan las 4 rutas seleccionadas en Zwift para la segunda instancia de recolección de datos.

2.5. Analisis estadistico - Segunda recolección

Una vez finalizado el segundo proceso de recolección de datos, el siguiente paso será realizar el análisis de los datos para determinar si son todos útiles y en caso contrario descartarlos. Primero se subdividieron los datos en varias carpetas, para poder realizar un análisis mas ordenado; la carpeta principal llamada Ciclismo, luego dos subdivisiones llamadas Hombres y Mujeres, y luego en ambas carpetas se encontraban 5 carpetas mas con

los nombres de cada ruta y una llamada “Todas”, donde estaban todas las rutas que habían en las demás carpetas.

Cuando los datos están filtrados, se utiliza EDA o “Exploratory Data Analysis” para poder realizar la limpieza de los datos, donde se busca eliminar valores atípicos y datos erróneos o faltantes. Posteriormente, se procede a realizar un análisis estadístico descriptivo para obtener una comprensión básica de la distribución y características de los datos. Durante el EDA, se generan gráficos y visualizaciones que ayudan a identificar patrones, tendencias y relaciones entre variables.

El siguiente paso fue crear un cuaderno en Jupyter Notebook, que nos permite realizar de manera mucho más visual e intuitiva el análisis exploratorio de datos. Para esto, se utilizaron las siguientes librerías:

- ExtraccionCompleta (Creada para extraer los datos de los archivos de recolección)
- matplotlib (Para la visualización grafica de los datos)
- os (Para facilitar la búsqueda de datos entre carpetas)
- pandas (Para crear un dataframe con los datos seleccionados)
- ydataprofiling (Especializada para EDA)
- dtale (Especializada para EDA)

Luego, se crea un código “interactivo” para poder elegir de manera sencilla y eficaz los datos que se querían analizar; el nombre del ciclista, el recorrido, el genero y el peso, eran las opciones para seccionar los datos. Mediante gráficos conocidos como el grafico de Violín y el de Caja se logra determinar a simple vista comportamientos anómalos de dos características principales.

Al analizar el gráfico de Violín de la figura 2.7, destinado a visualizar la densidad de datos de la característica seleccionada (potencia), se detectó un comportamiento inusual en algunas mediciones. Se observó que, para un mismo recorrido, las formas de las gráficas eran notablemente distintas, incluso considerando que los recorridos del 2 al 6 pertenecían a la misma persona.

Una inspección más detallada reveló que la variabilidad se debía a la forma en que los participantes llevaron a cabo los recorridos. Se determinó que, en ciertos casos, los

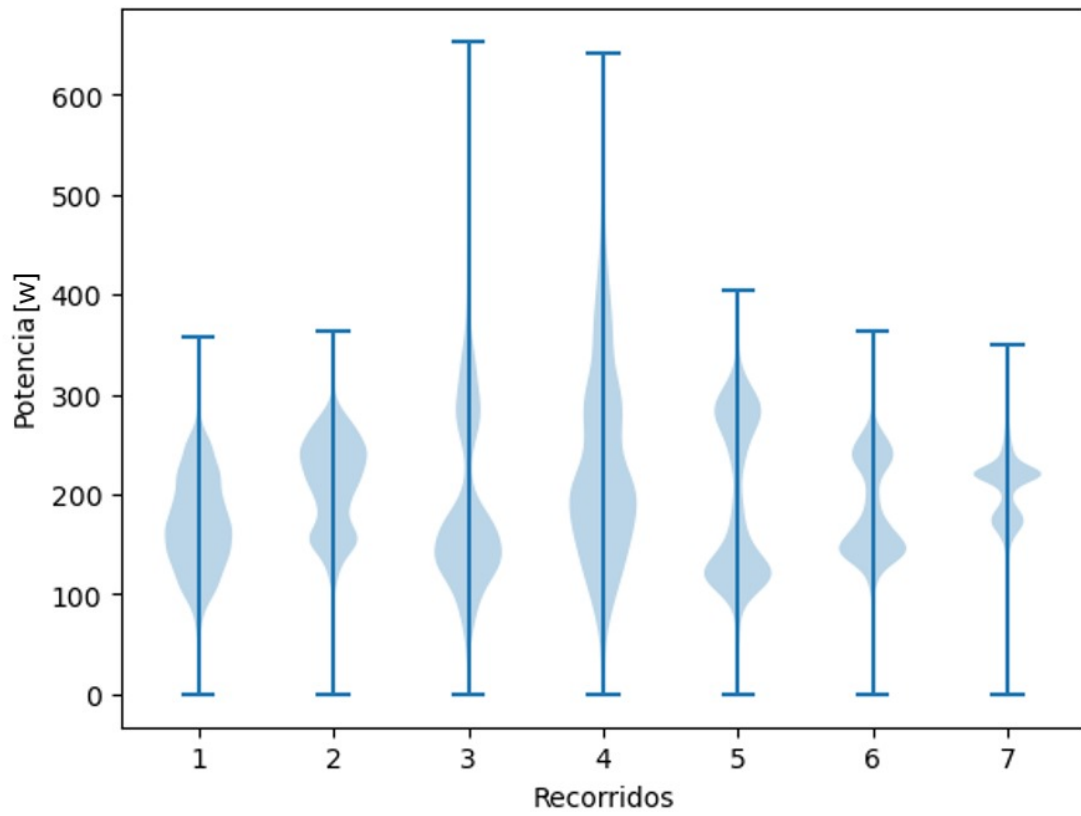


Figura 2.7: Comparación, mediante gráfico de violín, de la distribución de potencia de personas distintas en la misma ruta.

participantes realizaron los trayectos utilizando un formato de entrenamiento. Este formato implica la aplicación de variaciones intencionadas en la potencia durante la medición.

La figura 2.8 presenta un gráfico que compara una medición convencional del recorrido Climbers Gambit con otra realizada durante un entrenamiento. Esta comparación visualiza claramente cómo la introducción de variaciones intencionadas en la potencia durante el entrenamiento puede afectar significativamente la forma de las mediciones.

Con el objetivo de abordar el comportamiento inusual identificado en las curvas, se implementó un intento de descartar aquellas que presentaran variaciones significativas debido a entrenamientos intencionados en la potencia. Sin embargo, este enfoque demostró ser desafiante, ya que la cantidad de curvas que permanecían después de este filtrado era tan reducida que resultaba insuficiente para mantener la integridad del conjunto de datos.

Dadas estas circunstancias, se tomó la decisión de descartar todas las curvas recopiladas hasta ese momento y redefinir el objetivo del análisis. En consecuencia, se optó

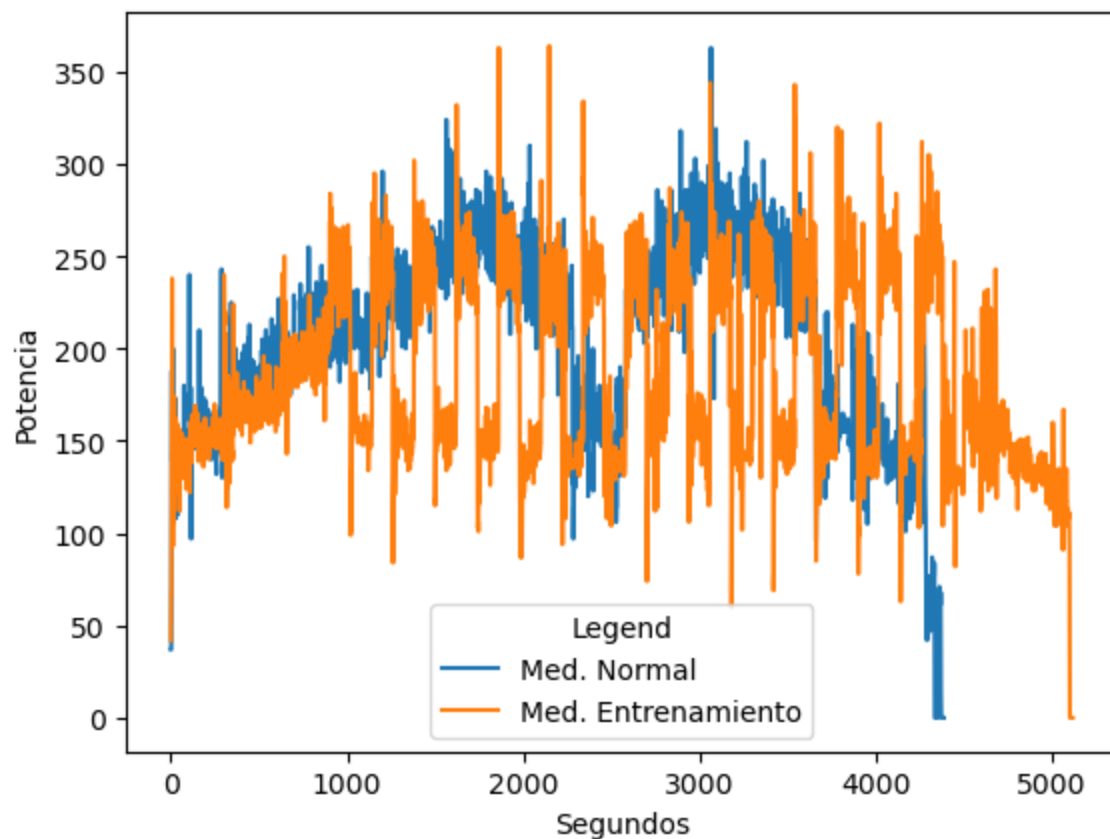


Figura 2.8: Comparación de perfil de potencia de un entrenamiento vs. perfil de potencia en un recorrido normal (en la misma ruta).

por utilizar exclusivamente los recorridos generados por un participante (ciclista amateur) externo al grupo de ciclistas que colaboraba en el proyecto. Esta elección se basó en la disponibilidad de datos más consistentes y en la capacidad de garantizar un conjunto de datos más homogéneo para el análisis estadístico y el entrenamiento de la machine learning.

Para llevar a cabo este análisis, se realizaron 30 recorridos que abarcaron diversas variaciones en el terreno, incluyendo segmentos de subidas, bajadas y terreno plano, con longitudes también variables. Además, se seleccionaron tres recorridos específicos identificados en las figuras 2.9, 2.10 y 2.11, denominados Actividad 1, 2 y 3, respectivamente. Estos recorridos fueron elegidos para la fase de predicción y no fueron utilizados durante el entrenamiento. Como se puede observar, todos poseen una característica común, y es que tienen inclinación positiva. Posteriormente, después de verificar la integridad de los datos, se llevó a cabo un análisis más detallado.

Al examinar el gráfico de interacciones en la figura 2.12, se observa una dispersión

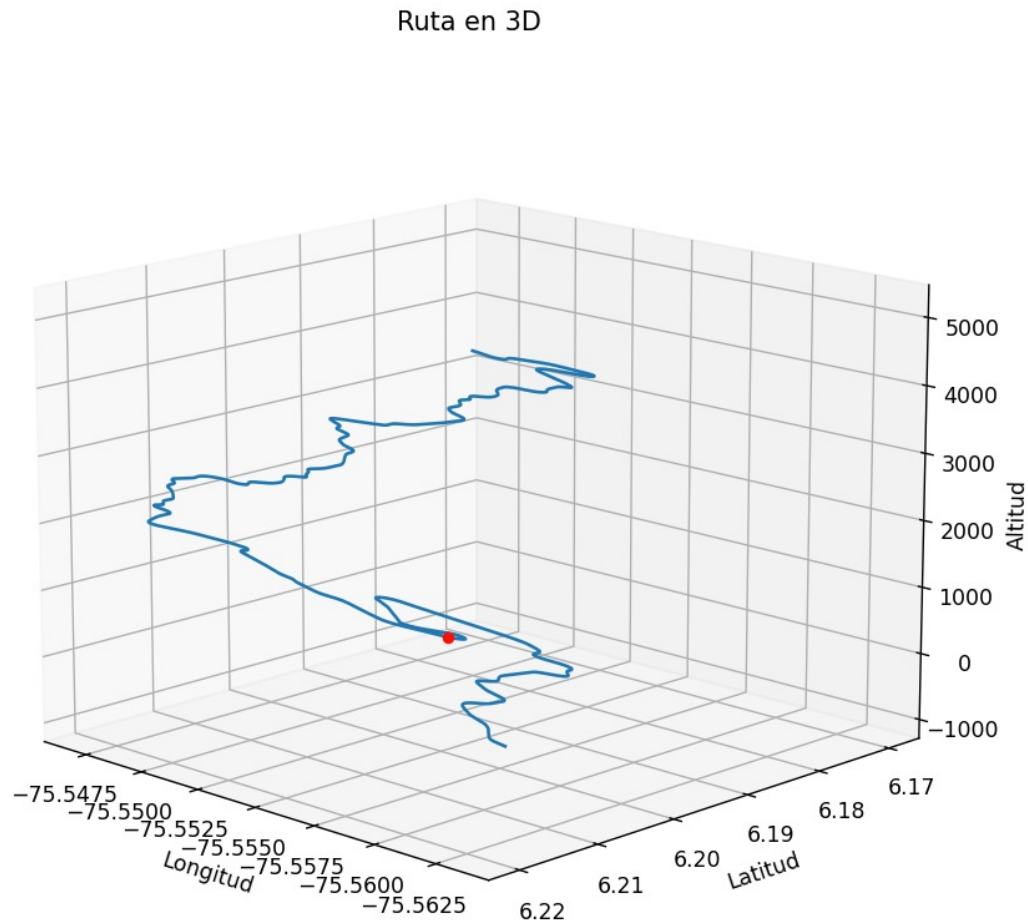


Figura 2.9: Representación tridimensional de Actividad 1.

significativa en todas las interacciones. Esta variabilidad se atribuye a la diversidad de características presentes en los recorridos, lo que dificulta la identificación de patrones o tendencias a simple vista. Por ejemplo, al analizar la relación entre la frecuencia cardíaca y la potencia, se observa una dispersión general entre las 120 y 160 pulsaciones por minuto y entre los 100 y 300 vatios.

En cuanto a la velocidad, se destaca una dispersión notoria durante los primeros 3000 segundos, luego de lo cual se estabiliza de manera más consistente. La potencia, por otro lado, muestra cierta estabilidad a partir de aproximadamente los 1000 segundos. Al analizar la interacción entre la potencia y la velocidad, se evidencia una dispersión sin una tendencia clara, con dos zonas de interacción más altas pero sin definirse claramente.

Esta dispersión general en las interacciones señala la complejidad inherente de los re-

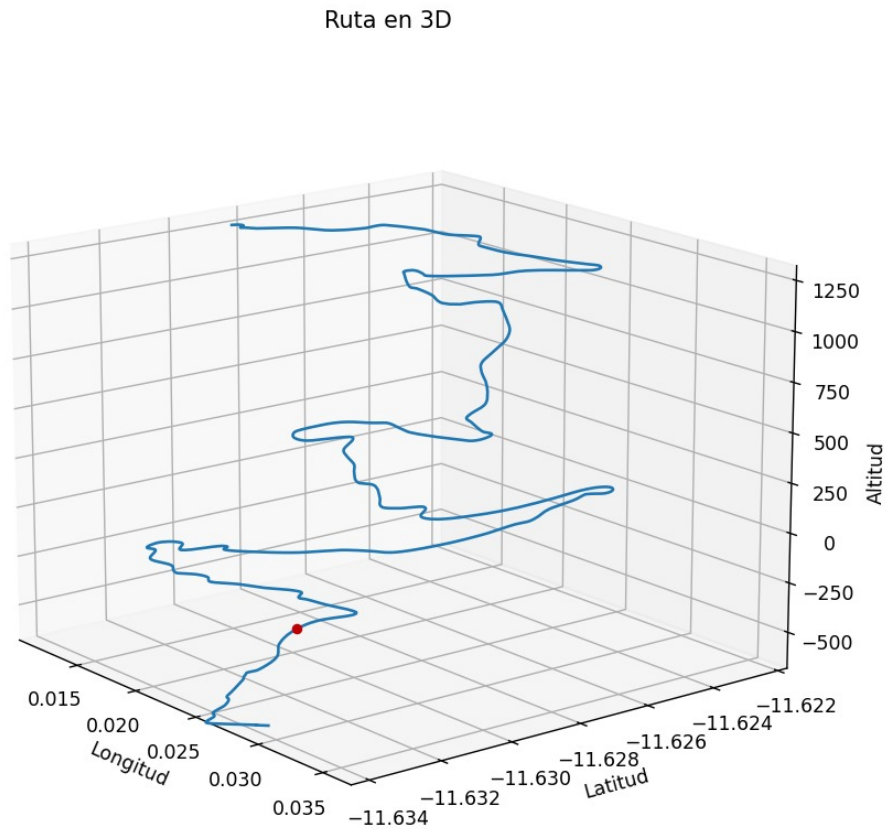


Figura 2.10: Representación tridimensional de Actividad 2.

corridos, donde diversas variables interactúan de manera dinámica.

Este análisis generó nuevas discusiones, incluida la idea de segmentar los recorridos en tramos, evaluando cada carrera individualmente y dividiéndola según si el tramo es en subida, plano o bajada. Esta propuesta se respaldó en el análisis paralelo del modelo matemático, donde se utilizaron algunas de las curvas para evaluar la precisión en el cálculo de la potencia según la fórmula.

Al graficar los resultados de la fórmula matemática, se notó una variabilidad en la precisión. Al profundizar en la discusión, se concluyó que la causa subyacente era la inclinación del terreno. Para respaldar esta conclusión, se graficó la potencia calculada en comparación con la altitud del terreno (véase en la figura 2.13, revelando una diferencia clara en los segmentos con inclinación positiva, es decir, en subida.

Ruta en 3D

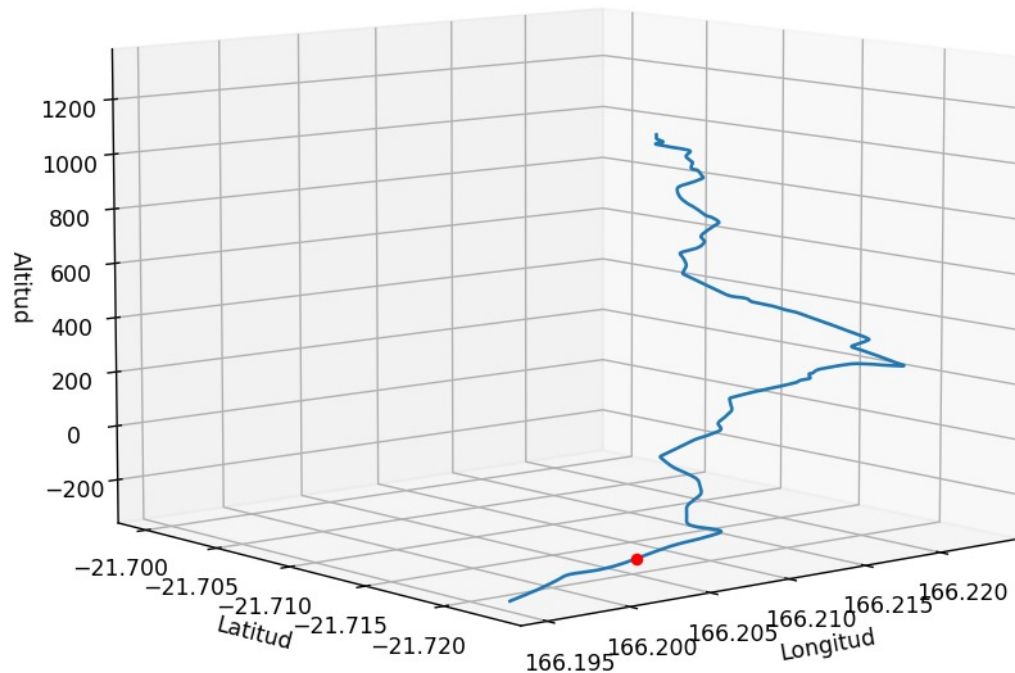


Figura 2.11: Representación tridimensional de Actividad 3.

Este fenómeno se entiende mejor al considerar a un ciclista descendiendo por una pendiente. En este escenario, la gravedad puede hacer que el ciclista no necesite pedalear para avanzar, afectando la precisión de los cálculos de potencia. Además, situaciones específicas, como descansar unos segundos o estar a rueda para minimizar el gasto energético, introducen complejidades adicionales en la medición de potencia, especialmente en terrenos planos o en bajada.

En resumen, los cálculos de potencia en situaciones de terreno plano o en bajada resultan menos confiables debido a factores externos difíciles de medir o controlar en este experimento. La fiabilidad aumenta significativamente en ascensos, donde la potencia del ciclista tiene un impacto más directo en el desplazamiento. Esta comprensión llevó a replantear el enfoque analítico, priorizando el estudio de los recorridos en ascenso para

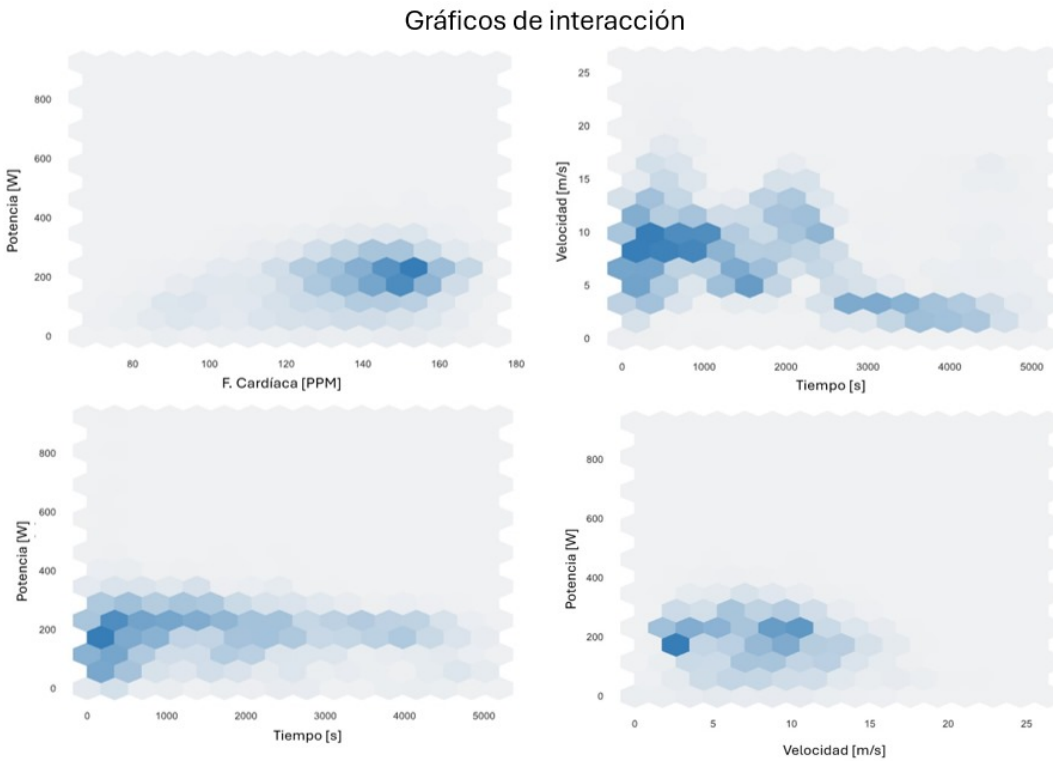


Figura 2.12: Gráficos de interacción entre características.

obtener resultados más precisos y significativos.

Es importante además definir el cálculo de una característica que se obtiene a partir de otras características, y en este caso, se trata de la inclinación (apoyo visual en figura 2.14). Esta se calculaba mediante dos datos: la altura y la distancia recorrida durante ese tramo, utilizando la siguiente fórmula:

$$\alpha = \frac{h_f - h_i}{d_h} \quad (2.1)$$

α = Inclinación

h = Altura

d_h = Distancia horizontal

Una vez divididos los recorridos, se comenzó a trabajar exclusivamente con los datos correspondientes a las subidas, además de un par de rutas previamente realizadas en terrenos con inclinación positiva, generando un total de 39 tramos de subida. Se aplicó

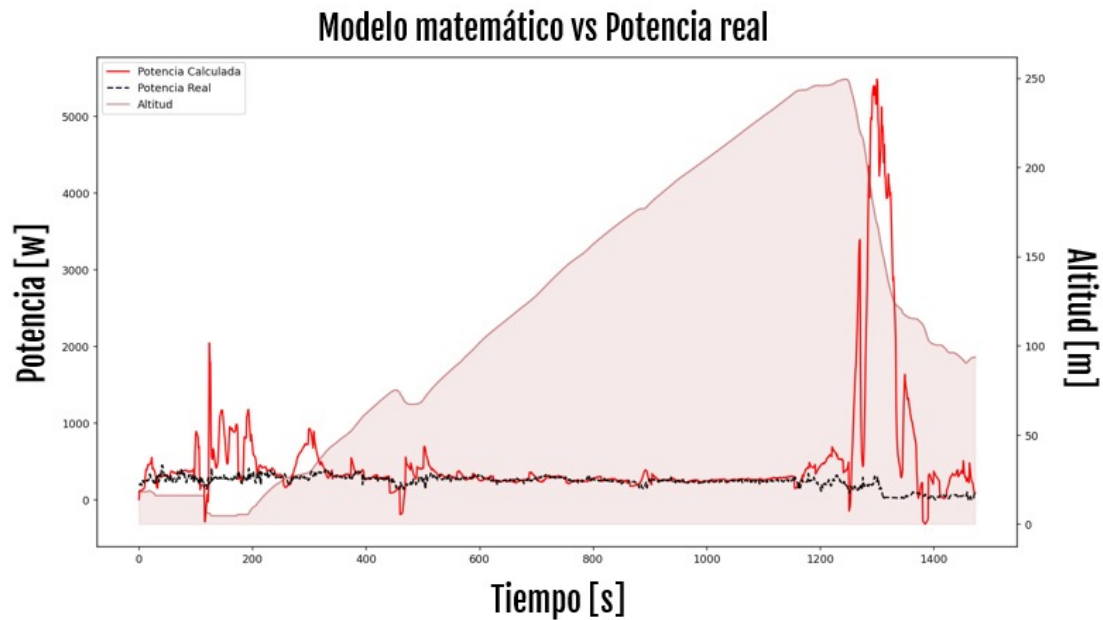


Figura 2.13: Comparación entre potencia calculada según el modelo matemático y potencia real.

la misma estadística utilizada anteriormente para comparar su comportamiento con los análisis previos. Como se puede observar en la figura 2.15, al comparar cada uno de los gráficos, se destacan cambios significativos. En general, las tendencias son mucho más claras.

En el gráfico de potencia vs frecuencia cardíaca, los datos muestran un agrupamiento más evidente en comparación con la figura 2.12. Respecto a la velocidad, se observa una estabilización mucho más temprana, aproximadamente en los 1000 segundos, en contraste con la figura anterior que presentaba esta estabilización alrededor de los 3000 segundos. Aunque no se aprecian cambios significativos en la potencia, en el último gráfico, la interacción entre la potencia y la velocidad muestra valores más estables en comparación con la figura anterior, que tenía entre dos y tres valores de interacción más altos.

Estos resultados indican que las tendencias en estos datos son más evidentes, lo cual sugiere que fortalecerán y facilitarán el proceso de aprendizaje de la machine learning.

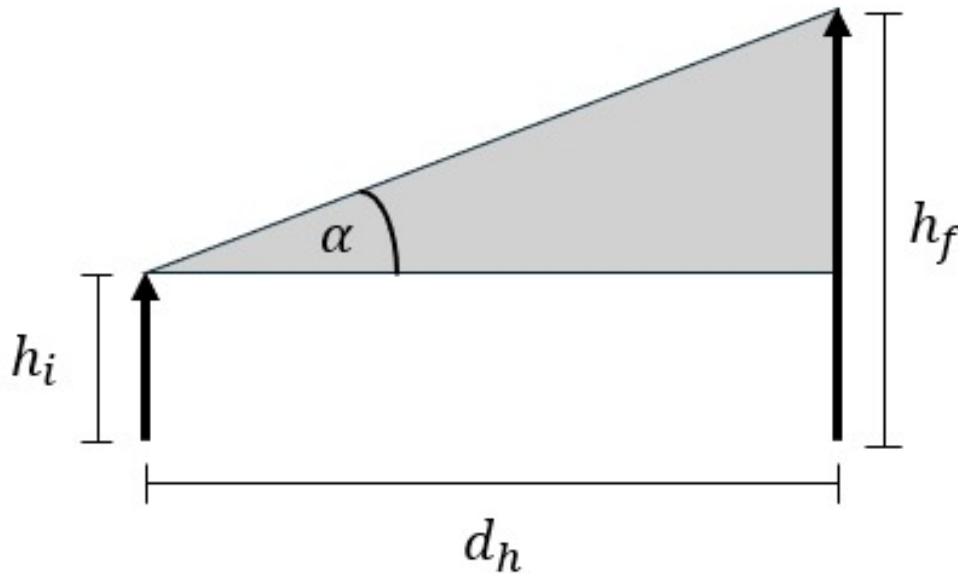


Figura 2.14: Esquema de referencia para cálculo de inclinación.

2.6. Machine learning

2.6.1. Primeros pasos.

Con los datos preparados, iniciamos las primeras pruebas con machine learning. Antes de sumergirnos en el análisis, es crucial comprender la naturaleza de la machine learning y por qué resulta tan valiosa para abordar fenómenos como la predicción de la potencia en el ciclismo.

La machine learning es un campo de la inteligencia artificial que se centra en el desarrollo de algoritmos capaces de aprender automáticamente a partir de datos y mejorar su rendimiento con la experiencia. A diferencia de la programación convencional, donde las instrucciones se codifican explícitamente, los modelos de machine learning pueden identificar patrones complejos y tomar decisiones sin intervención directa del programador.

En el estudio de la predicción de la potencia en el ciclismo, la machine learning aporta significativamente por varias razones:

1. **Captura de Patrones No Lineales:** Los modelos de machine learning son efectivos para identificar patrones y relaciones no lineales en conjuntos de datos complejos,

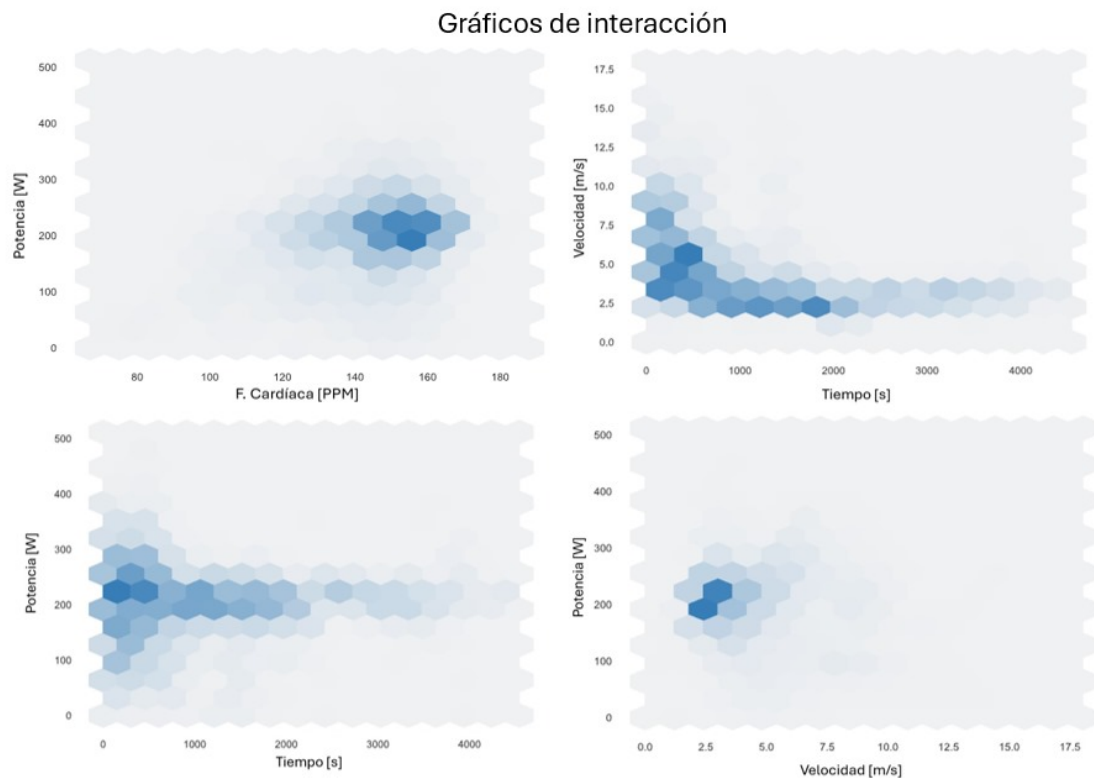


Figura 2.15: Gráficos de interacción entre características para rutas en subida.

como aquellos derivados del rendimiento ciclístico.

2. **Adaptabilidad Continua:** Estos modelos pueden adaptarse y mejorar con el tiempo a medida que se incorporan nuevos datos, lo cual es esencial en el ámbito del ciclismo donde el rendimiento está sujeto a diversas variables.
3. **Manejo de Variables Múltiples:** En el ciclismo, múltiples variables como velocidad, inclinación y frecuencia cardíaca afectan la potencia generada. Los modelos de machine learning pueden manejar eficientemente estas variables para ofrecer predicciones integrales.

Como mencionamos en la introducción, el enfoque óptimo para abordar este tipo de problema es el deep learning, específicamente las redes neuronales recurrentes (RNN), debido a la naturaleza temporal o secuencial de nuestro problema. La complejidad inherente a este tipo de datos hace que las RNN sean la elección más adecuada. No obstante, la creación de una red neuronal desde cero es un proceso no trivial. Por lo tanto, revisaremos paso a paso el proceso de construcción y mejora de la red neuronal.

2.6.2. Definir la red neuronal.

Elegir el tipo de red neuronal apropiado es crucial y depende de las características particulares de los datos y del problema que se está abordando. Actualmente, existen diversos tipos de redes neuronales, y en el esquema 2.16 se pueden observar algunas de ellas. Por ejemplo, las redes neuronales convolucionales (CNN) demuestran eficacia en tareas de visión por computadora, siendo especialmente útiles para la detección de patrones en imágenes. Por otro lado, las redes neuronales completamente conectadas (FNN) son versátiles y se aplican en una variedad de contextos, desde la clasificación de imágenes hasta el procesamiento de lenguaje natural. La elección del tipo de red debe alinearse con la naturaleza específica de los datos y los requisitos del problema en cuestión.

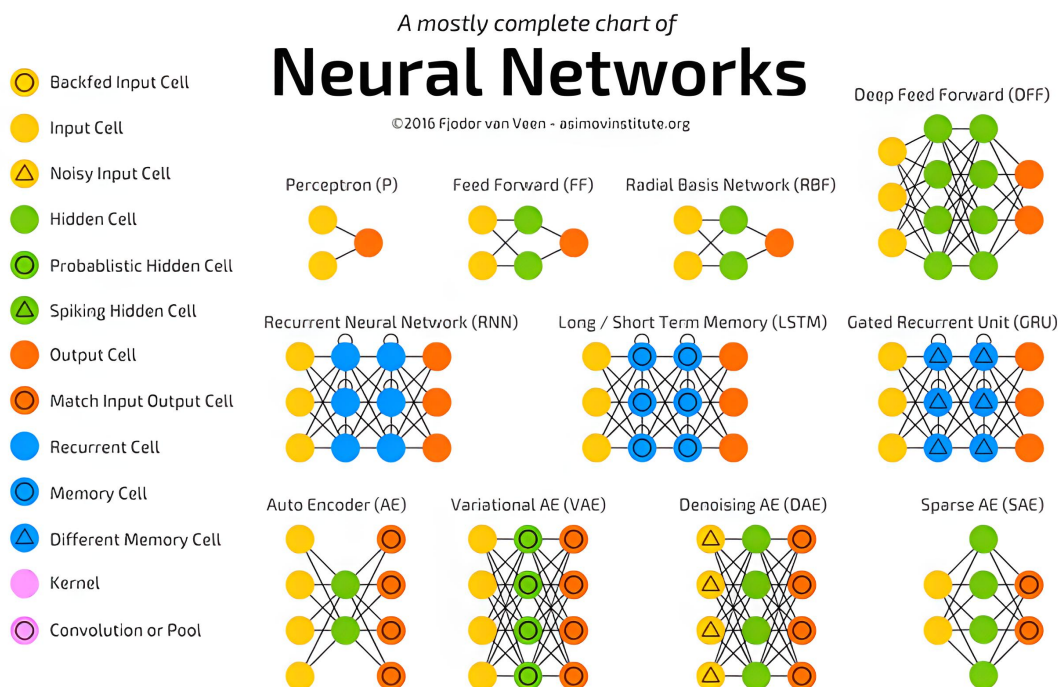


Figura 2.16: Distintos tipos de configuraciones para redes neuronales. Figura tomada de Alvy [3]

En este caso, donde se trabaja con datos en formato secuencial y temporal, se opta por utilizar una red neuronal recurrente (RNN). Las Redes Neuronales Recurrentes (RNN) son herramientas efectivas para el análisis de secuencias de datos al mantener una “memoria” interna que les permite recordar información de pasos de tiempo anteriores. Sin embargo, las RNN convencionales pueden enfrentar desafíos con los gradientes desvanecientes, lo que dificulta aprender relaciones a largo plazo en las secuencias. Para superar esta limi-

tación, se recurre a las Capas LSTM (Long Short-Term Memory). Las LSTMs son una variante especializada de las RNN que abordan el problema de los gradientes desvanecientes. Estas capas son particularmente útiles cuando se trata con secuencias extensas, ya que poseen una estructura de puertas que les permite controlar el flujo de información y capturar relaciones a largo plazo en los datos, convirtiéndolas en una elección ideal para el problema que se quiere resolver.

2.6.3. Tratamiento previo de los datos

Posterior a la elección del tipo de red, es necesario realizar un proceso crucial de tratamiento de los datos antes de introducirlos en la red neuronal recurrente (RNN). La normalización de los datos puede ser fundamental para asegurar que la red aprenda de manera efectiva y generalice bien a diferentes condiciones. Estos pasos de preprocesamiento son vitales para optimizar el rendimiento y la capacidad de aprendizaje de la red neuronal recurrente.

Para esto, se implementó la Normalización Z-score (estandarización) para el preprocesamiento de los datos. Esta técnica reescala los datos de manera que tengan una media de 0 y una desviación estándar de 1. La elección de esta normalización específica se debe a su eficacia para estandarizar los datos, facilitando así el entrenamiento y la convergencia de la red neuronal recurrente (RNN) de manera más eficiente.

Se calcula utilizando la siguiente fórmula:

$$X_{norm} = \frac{(X - \mu)}{\sigma} \quad (2.2)$$

μ = media de los datos

σ = desviación estándar

Es común que las carreras tengan longitudes variables. Para abordar esta variabilidad, se utiliza la técnica de “padding”, que consiste en agregar valores falsos, como ceros, al final de las secuencias más cortas hasta igualar la longitud de la secuencia más larga. Es importante destacar que, en el contexto de las capas LSTM, esta práctica no influye significativamente en la interpretación de los datos, ya que estas son bastante buenas en ignorar estos valores de “padding”.

2.6.4. Estructura del modelo

Ahora que se cuenta con las herramientas necesarias para preparar los datos y alimentarlos eficientemente al modelo, resta definir la estructura de dicho modelo y proporcionarle los datos. El proceso se inicia importando NumPy [17] para el manejo adecuado de los datos y utilizando la API de Keras de la librería TensorFlow. Se optó por utilizar TensorFlow [2] como librería principal debido a su prominencia en el ámbito del deep learning. Es ampliamente reconocida y utilizada en la comunidad de desarrollo, facilitando el aprendizaje y la implementación gracias a la abundancia de recursos y desarrollos de código abierto disponibles.

También se importa el modelo secuencial (Sequential) y las capas LSTM y Dense. En cuanto al modelo Sequential de Keras, su simplicidad lo convierte en una opción apropiada para esta aplicación. Este modelo es lineal y consta de una serie de capas dispuestas en secuencia, donde la salida de una capa se convierte en la entrada de la siguiente. Su estructura simple refleja la naturaleza secuencial de los datos, facilitando la construcción y comprensión del modelo en este contexto específico

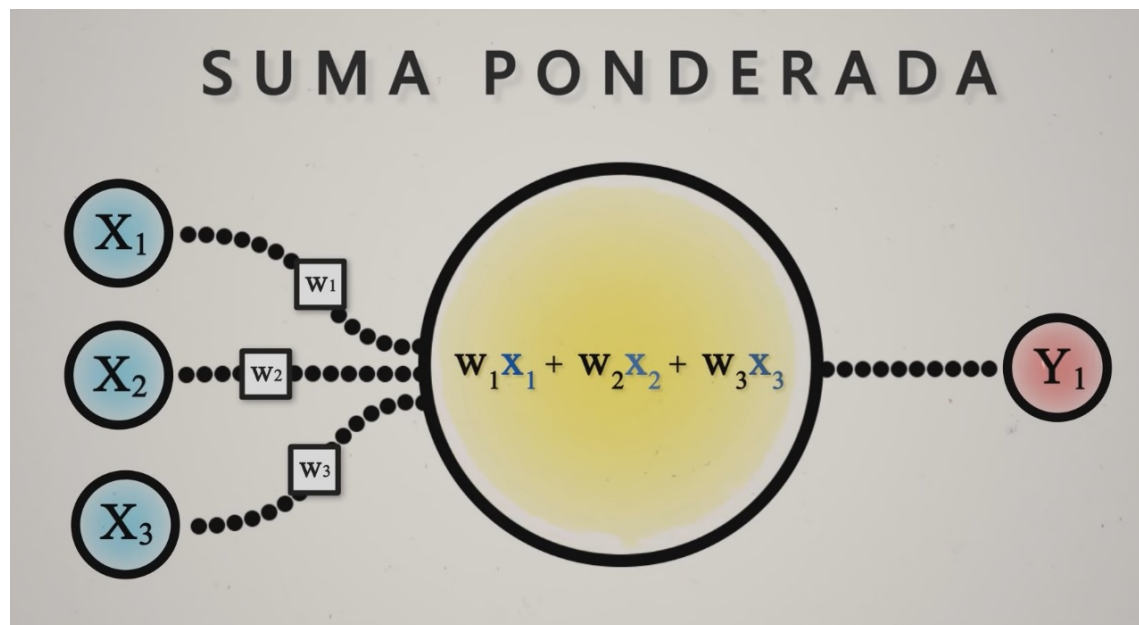


Figura 2.17: En esta ilustración, se pueden apreciar las distintas partes que componen la estructura de una neurona. Figura tomada de DotCSV [1].

En la Figura 2.17, se puede profundizar en el funcionamiento de una neurona, un com-

ponente esencial para comprender el proceso subyacente. En primer lugar, se observan los datos de entrada (X_n), los cuales establecen conexiones con la neurona a través de enlaces, cada uno asociado a su peso correspondiente (W_m). Dentro de la neurona, opera una función que, basándose en los datos de entrada y los pesos asignados a las conexiones, evalúa la relevancia de cada dato de entrada. Este proceso se completa con la generación de un resultado, almacenado en la neurona de salida (Y_1). En esencia, esta estructura permite que la neurona capture y procese la información de entrada, asignando ponderaciones a cada componente y generando una salida que refleja la influencia de los diferentes datos de entrada en el resultado final.

La elección de la arquitectura y los hiperparámetros de una red neuronal, como el número de neuronas y ciclos de entrenamiento, es un aspecto crucial para el rendimiento y la eficacia del modelo. Aquí hay algunas consideraciones clave que se tomaron en cuenta:

1. **Número de Neuronas:** El número de neuronas en una capa influye en la capacidad de la red para capturar patrones complejos en los datos. Demasiadas neuronas pueden llevar al sobreajuste, donde el modelo memoriza los datos de entrenamiento en lugar de aprender patrones generales. Pocas neuronas pueden resultar en un modelo insuficientemente complejo para capturar relaciones importantes. La elección óptima de neuronas depende del tamaño y la complejidad de los datos, y es a menudo un proceso de prueba y error.
2. **Profundidad de la Red:** La profundidad de la red se refiere a la cantidad de capas que contiene. Modelos más profundos pueden aprender representaciones jerárquicas más abstractas, pero también aumentan el riesgo de sobreajuste. Las LSTMs ya son una capa profunda por sí mismas, por lo que se debe considerar bien la cantidad adicional de capas que se desea agregar.
3. **Ciclos de Entrenamiento (Épocas):** El número de ciclos de entrenamiento (épocas) afecta la cantidad de veces que el modelo ve los datos de entrenamiento. Es importante encontrar un equilibrio, ya que demasiadas épocas pueden llevar al sobreajuste y demasiado pocas pueden dar como resultado un modelo subentrenado.
4. **Función de Activación:** La elección de la función de activación en una red neuronal, especialmente en capas LSTM, es crucial para introducir no linealidades en el modelo. La función de activación determina la salida de una neurona y su capacidad para aprender patrones complejos. Para las capas LSTM, se suele utilizar la

función de activación tangente hiperbólica ($\tanh$) o la unidad lineal rectificada (ReLU), dependiendo de la naturaleza del problema. La $\tanh$ es adecuada para manejar datos normalizados, mientras que ReLU puede ser más eficaz en situaciones donde se buscan representaciones más dispersas y dispersas de los datos. La elección de la función de activación puede influir significativamente en el rendimiento y la capacidad de aprendizaje de la red.

w

2.6.5. Hiperparametros

Inicialmente, se emprendió el desarrollo de una red neuronal básica con el propósito de predecir la velocidad a partir de las características de distancia y altitud, buscando comprender el comportamiento de los hiperparámetros. Tras varios intentos, se logró con éxito esta primera predicción. La estructura detallada del modelo se encuentra representada en el esquema 2.18.

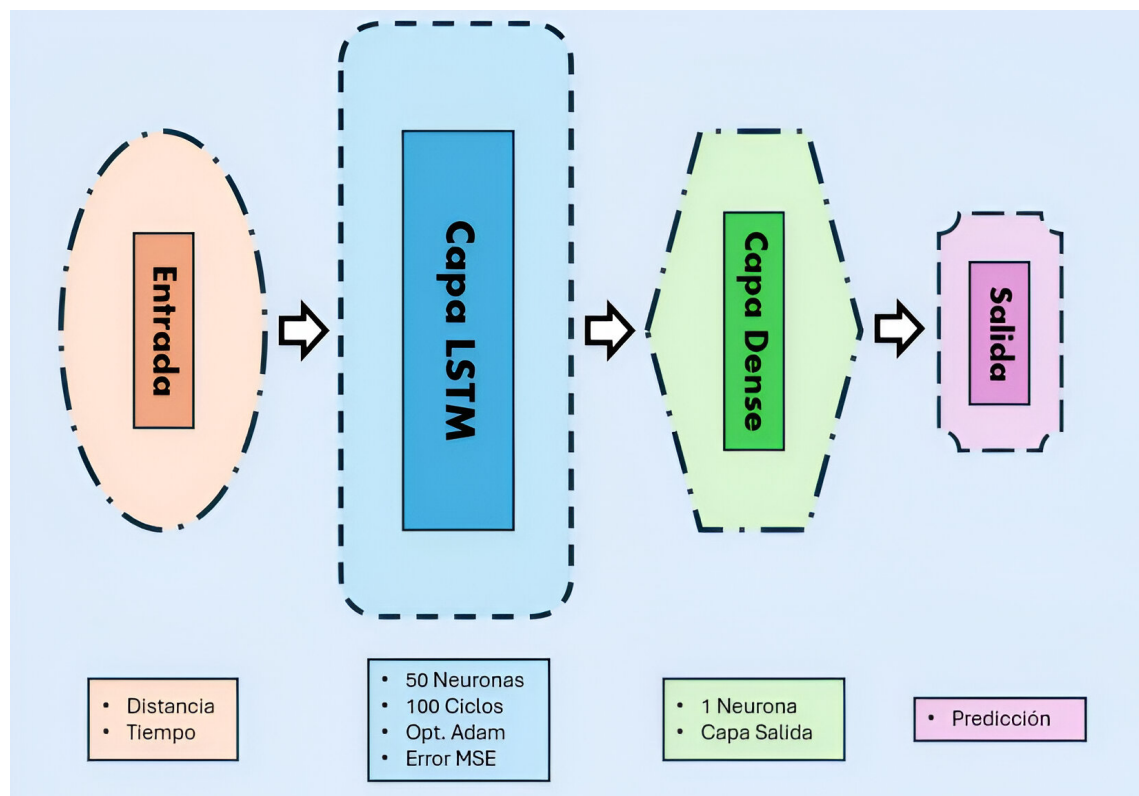


Figura 2.18: Estructura de la primera red neuronal para predicción de velocidad.

Con una comprensión más profunda de la estructura, se procedió a abordar la predicción de la potencia. Antes de iniciar las pruebas, se llevó a cabo una investigación para identificar los rangos óptimos de los parámetros de la red neuronal. Se destacó la importancia de reconocer que el valor óptimo de un hiperparámetro no puede conocerse de antemano para un problema específico. Por lo tanto, se consideró utilizar valores genéricos basados en el éxito previo en problemas similares o realizar pruebas mediante el método de prueba y error.

En este contexto, se optó por definir un rango para cada hiperparámetro y realizar pruebas utilizando valores aleatorios dentro de estos rangos. Se seleccionaron valores comúnmente utilizados en problemas similares de redes neuronales como punto de partida para los rangos.

Los parámetros sujetos a variación fueron los siguientes:

- Cantidad de neuronas por capa
- Cantidad de capas de la red
- Tipo de capa (LSTM o Dense)
- Error de entrenamiento (MSE o MAE)
- Ciclos

Como se señaló anteriormente, en estas primeras pruebas, se seleccionaron las características basándonos en la correlación que tenían con la característica objetivo (potencia). A pesar de varios intentos, logramos obtener resultados relativamente cercanos a la curva real de potencia, lo que brinda cierto optimismo sobre la posibilidad de utilizar el enfoque de machine learning para este propósito.

Después de extensas pruebas y análisis de los resultados obtenidos al variar los hiperparámetros, se llegó a la conclusión de que la mayoría de estos parámetros no sigue reglas predefinidas y, en su mayoría, se ajusta mediante pruebas para identificar los modelos más efectivos. Durante este proceso, se observó un patrón interesante: un conjunto específico de hiperparámetros mostró consistentemente un rendimiento destacado. Este conjunto en particular coincidió con combinaciones de hiperparámetros que ya se habían utilizado en investigaciones anteriores con redes neuronales aplicadas a problemas similares, específicamente el patrón de 128, 64 y 32 neuronas para las capas ocultas, sin considerar la primera capa de entrada.

Inspirados por esta observación, se centró la atención en el patrón específico de neuronas por capa (128, 64, 32) y se exploró su rendimiento al variar los demás hiperparámetros. Este enfoque tenía como objetivo ajustar los parámetros restantes para optimizar aún más el rendimiento del modelo. Los resultados obtenidos respaldaron la eficacia de este patrón en particular. Así, se identificó y utilizó con éxito una combinación de hiperparámetros que no solo demostró ser efectiva en investigaciones previas, sino que también se adaptó de manera sobresaliente a un conjunto de datos específico. Se puede visualizar la estructura de la red en la figura 2.19.

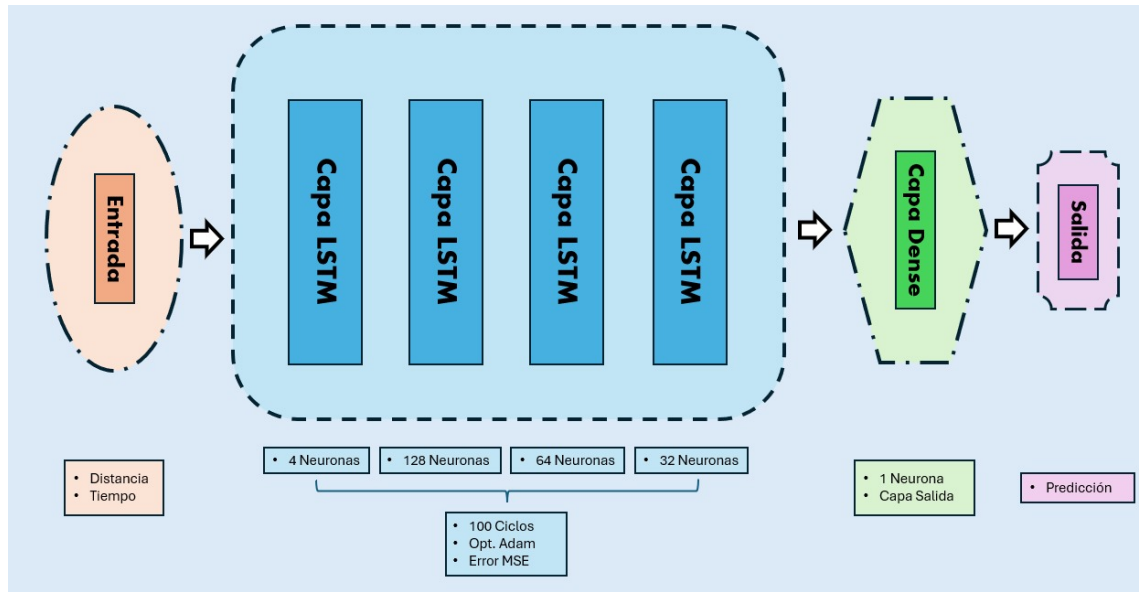


Figura 2.19: Estructura de la red neuronal final.

El error MSE es calculado de la siguiente manera:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.3)$$

n es el número de observaciones.

y_i es el valor real de la observación i .

$\hat{y}_i$ es el valor predicho para la observación i .

Este método tiene la característica de que penaliza en mayor medida los errores que son mas grandes.

Una vez que se tuvo el modelo final con el cual se trabajaría, solo restaba elegir las características que se utilizarían para entrenar el modelo y obtener el mejor resultado po-

sible. Se optó por explorar todas las posibles combinaciones de características, utilizando un código automatizado que generaba todas las combinaciones posibles entre estas características. Dado el considerable número de combinaciones, se decidió conservar solo aquellos modelos que obtenían un error MSE menor a 0.2, el cual se calcula incluso después de la normalización. A pesar de que este proceso demandó varios días de cálculos, al concluir se obtuvieron 62 modelos con un error MSE inferior a 0.2.

En este punto, solo quedaba identificar cuál era el mejor modelo para realizar la predicción. Además, se evaluó, basándose en el error, cuál era el modelo que proporcionaba los mejores resultados en las tres actividades. De esta manera, se seleccionaron 5 modelos con los cuales finalmente se llevaron a cabo todos los cálculos para la sección de resultados. La elección de estas combinaciones de características se basó en una evaluación exhaustiva de su desempeño en múltiples recorridos, considerando las 8 características disponibles y ampliando el enfoque más allá de los 5 perfiles proporcionados por los métodos estadísticos.

Cabe destacar que para mejorar la comprensión de los errores, estos se transformaron a errores porcentuales mediante la siguiente fórmula:

$$\text{MSE \%} = \left(\frac{\sigma^2}{\text{MSE}} \right) \times 100 \quad (2.4)$$

σ^2 es la varianza de los datos

Capítulo 3

Resultados y conclusiones

3.1. Resultados

Como se mencionó en el capítulo 2, uno de los primeros modelos desarrollados para probar la implementación de TensorFlow y la estructura básica de una red neuronal fue el de predicción de la velocidad a partir de la distancia y el tiempo. Durante estas pruebas, se varió principalmente el hiperparámetro correspondiente al número de ciclos de entrenamiento por capa. En la figura 3.1, se presentan cuatro ejemplos distintos, donde los ciclos de entrenamiento toman los siguientes valores respectivamente: 10, 20, 100 y 200. Al analizar los resultados en los gráficos, se observa que el primer intento con 10 ciclos proporciona una predicción bastante deficiente. Luego, con 20 ciclos, se evidencia una mejora significativa, aunque la precisión aún no es óptima. Posteriormente, al aumentar a 100 ciclos, la predicción se aproxima prácticamente a la exactitud. Sorprendentemente, al incrementar los ciclos a 200, no se observa una diferencia significativa respecto a los 100 ciclos. Esto sugiere que, a partir de cierto punto, la mejora en la predicción se estabiliza, independientemente de la cantidad adicional de ciclos de entrenamiento.

Posterior a estas primeras pruebas, con el objetivo de modelar la potencia a partir de características más accesibles para ciclistas de cualquier nivel, como la velocidad, cadencia, frecuencia cardíaca, entre otros, se llevó a cabo una evaluación inicial utilizando el modelo matemático-físico conocido como el modelo Cohen-2021 [10]. Esta evaluación se realizó mediante la aplicación del modelo a tres recorridos distintos realizados por una misma persona.

Cabe mencionar que, no abordado en el capítulo 2, se enfrentaron varios desafíos para lograr que el modelo matemático generara una curva coherente al compararla con otros

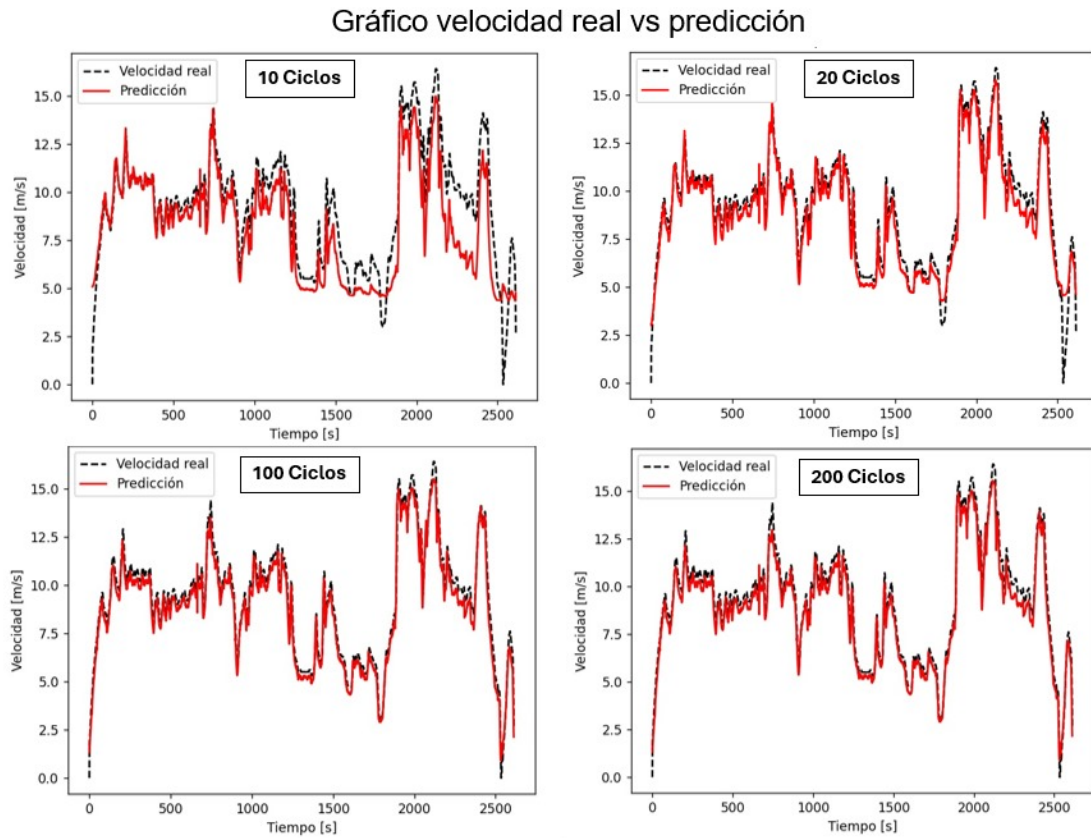


Figura 3.1: Resultados primera red neuronal para predecir velocidad.

trabajos previos, lo que implicó múltiples intentos y ajustes debido a la complejidad de este paso.

Para la evaluación del modelo, se implementó un código que recorre cada punto de un recorrido, calcula y guarda los datos de potencia obtenidos al introducirlos al modelo matemático.

Podemos observar en el gráfico 3.2, uno de los primeros intentos para graficar la potencia utilizando la fórmula. El error en el cálculo es evidente; en este caso, los primeros errores surgieron debido a un cálculo incorrecto del aporte de la energía cinética. Se probaron métodos básicos de derivación para la fórmula directa, pero los resultados no fueron satisfactorios. Posteriormente, se extrajeron los datos de velocidad del recorrido y se calculó, mediante un cuaderno de Mathematica, la contribución de la derivada temporal de la energía cinética mediante una aproximación numérica.

Una vez corregido el error mencionado, en una segunda instancia se obtuvieron los gráficos 3.3, 3.4 y 3.5. Aunque a simple vista podrían parecer peores, eran considerablemente mejores, ya que lograban replicar levemente la tendencia de la gráfica de potencia

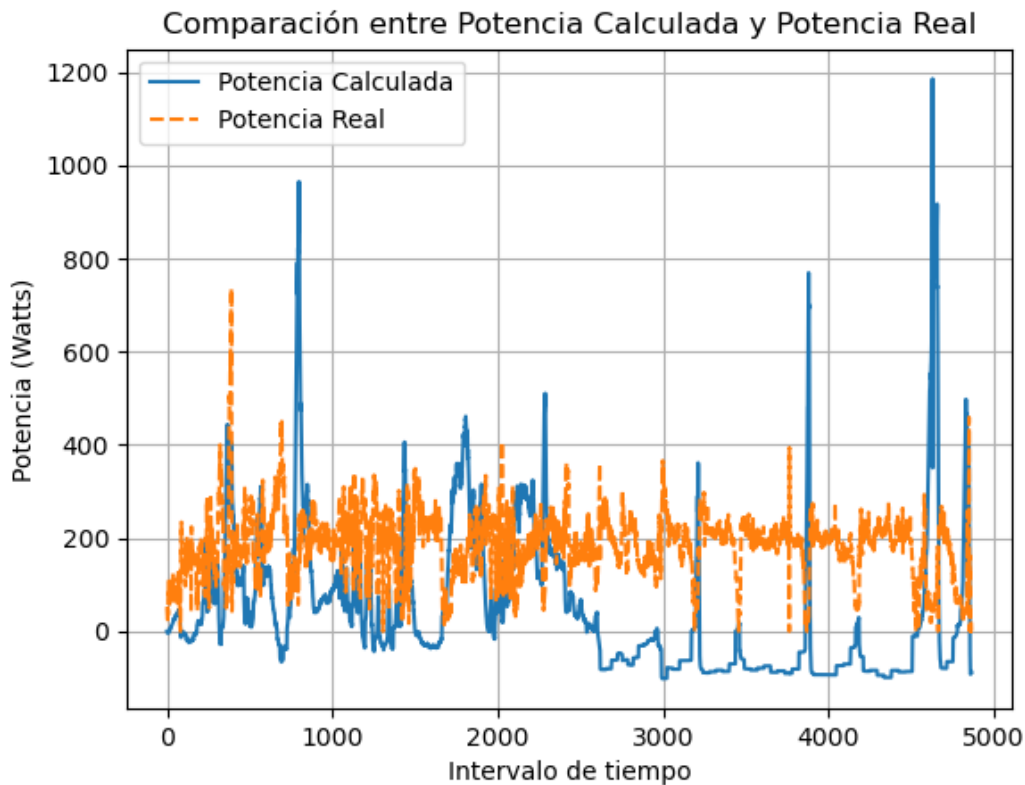


Figura 3.2: Primer intento de graficar la fórmula con datos reales.

real, representando un acercamiento valioso. Sin embargo, aún persistía un detalle que impedía que los gráficos proporcionaran una predicción precisa. Se identificó que el código contenía un error en la fórmula, donde una de las variables (la velocidad) no se multiplicaba con la parte gravitacional de la fórmula. Una vez corregido este aspecto, se lograron los resultados deseados.

Analizando los gráficos 3.6, 3.7 y 3.8, se destaca que la curva calculada a partir del modelo matemático muestra un notable parecido con la curva real. En ciertas partes, la similitud es considerable. Este resultado confirma la capacidad del modelo matemático de capturar patrones generales de comportamiento de la potencia en función de las características evaluadas, lo cual es coherente con hallazgos observados en otros trabajos.

No obstante, al enfocarnos en el objetivo específico de predecir la potencia a partir de otras características, se identifican irregularidades en los gráficos. En este caso, se observan discrepancias tanto por encima como por debajo de la curva real, incluso superando los 200 [W] en algunos casos. Esta observación resalta la necesidad de buscar una herra-

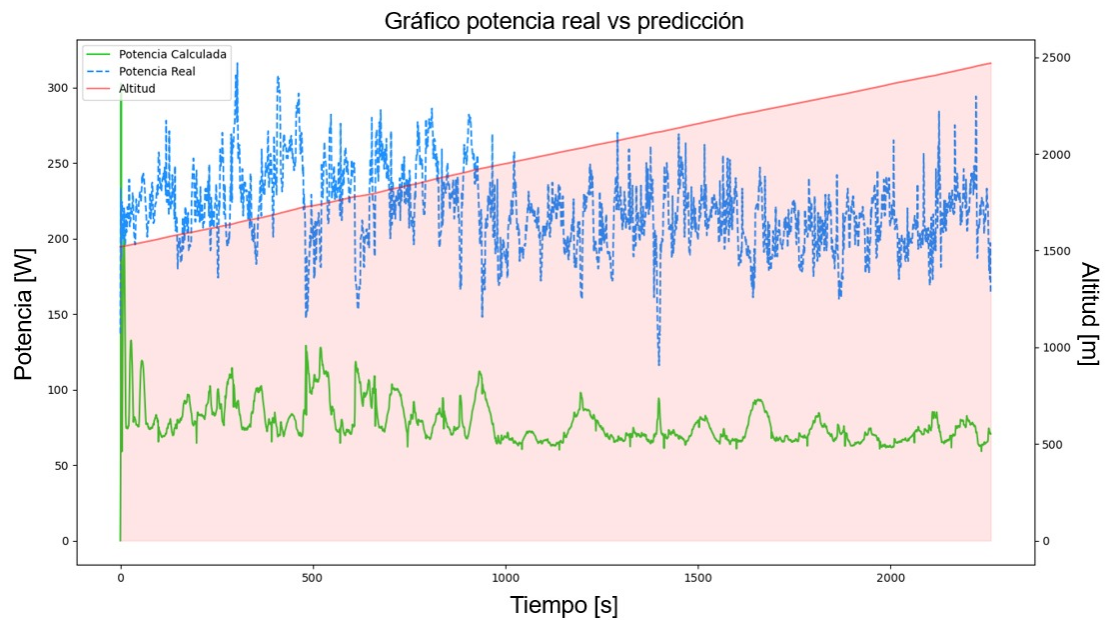


Figura 3.3: Segundo intento de graficar la fórmula (Actividad 1).

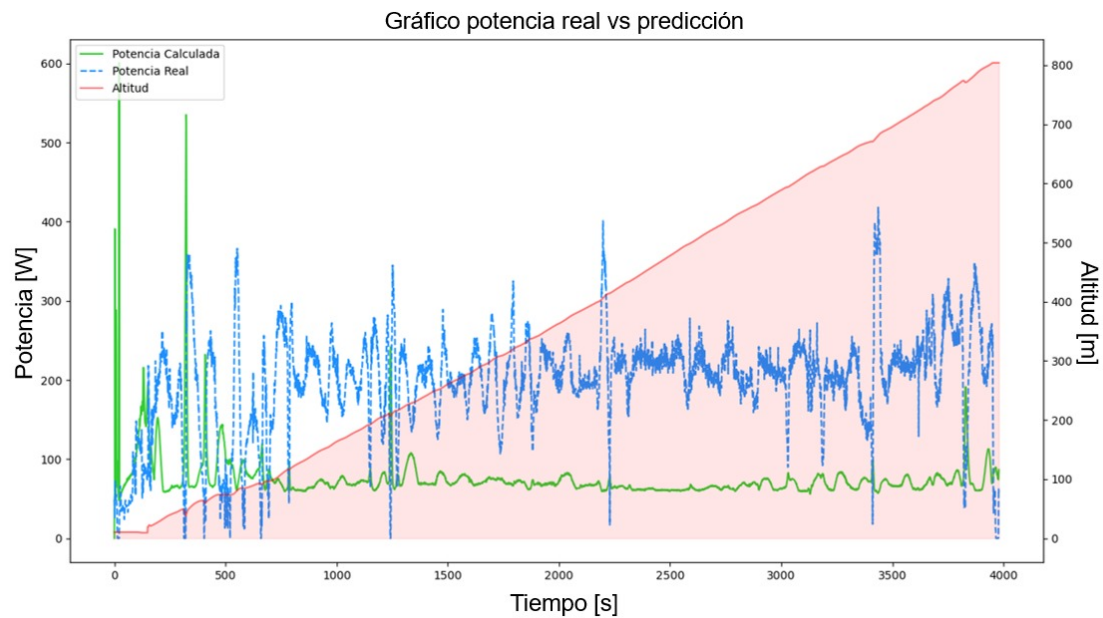


Figura 3.4: Segundo intento de graficar la fórmula (Actividad 2)

mienta más precisa y detallada para la predicción de la potencia, reforzando la importancia del desarrollo del modelo predictivo utilizando redes neuronales.

Una vez completado el desarrollo del modelo matemático, se procedió a entrenar la

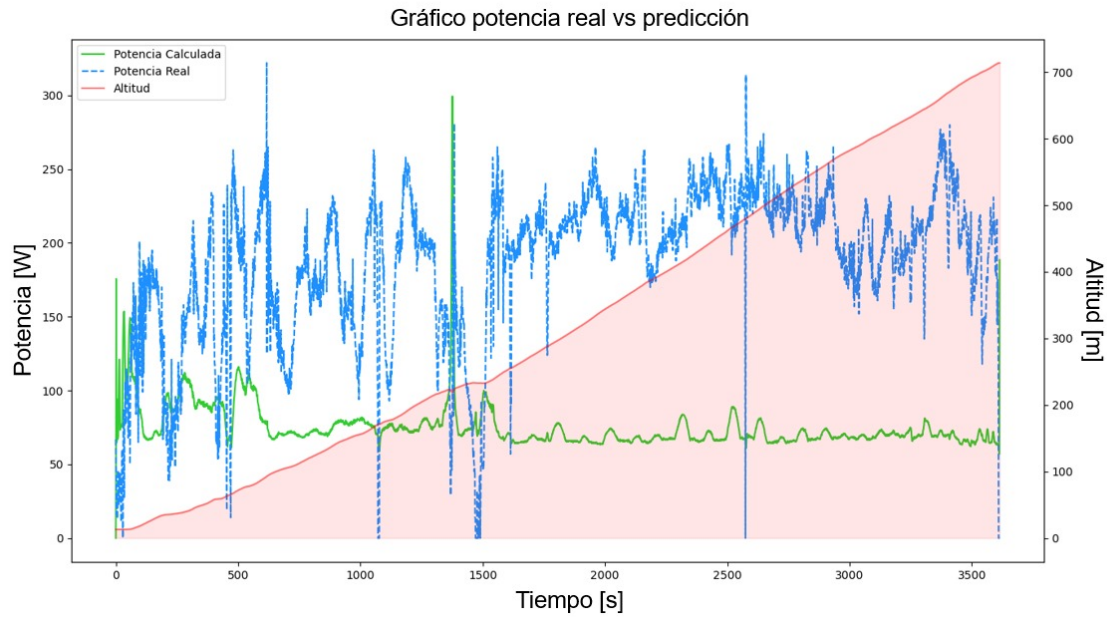


Figura 3.5: Segundo intento de graficar la fórmula (Actividad 3).

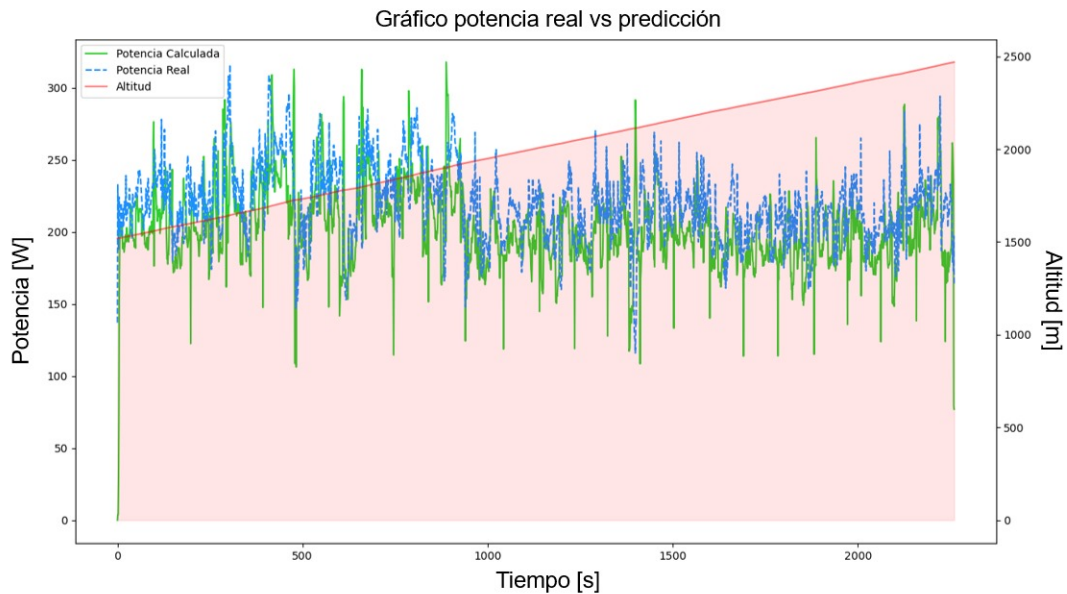


Figura 3.6: Cálculo de potencia mediante formula (Actividad 1)

red neuronal con el objetivo de igualar y eventualmente superar la precisión del modelo matemático. Durante la primera fase de recolección de datos, se generaron los primeros modelos basados en los perfiles obtenidos mediante los métodos estadísticos presentados en las figuras 3.9, 3.10, 3.11, 3.12 y 3.13.

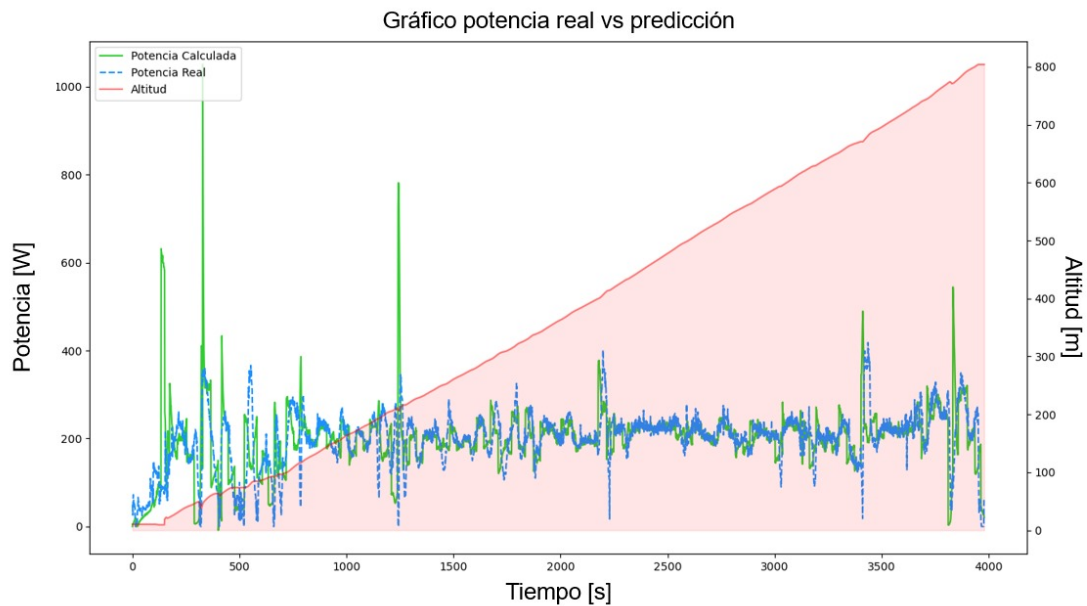


Figura 3.7: Cálculo de potencia mediante formula (Actividad 2)

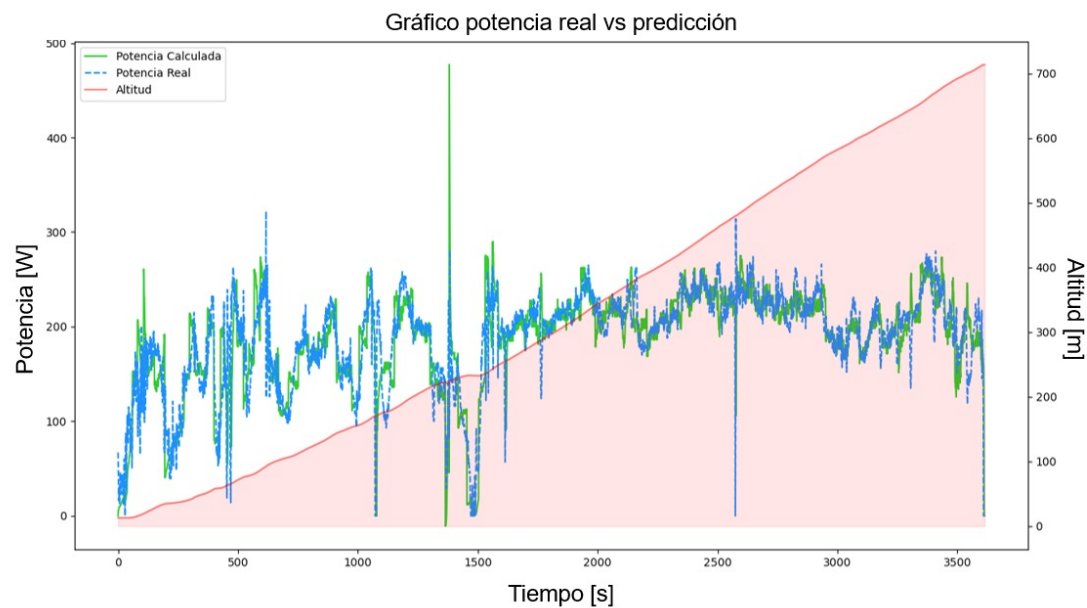


Figura 3.8: Cálculo de potencia mediante formula (Actividad 3)

Estos modelos, inicialmente contruidos con enfoques estadísticos, proporcionaron resultados que mostraron ciertas tendencias y patrones, aunque aún no alcanzaban la precisión deseada para predecir la potencia. A pesar de las limitaciones observadas en los

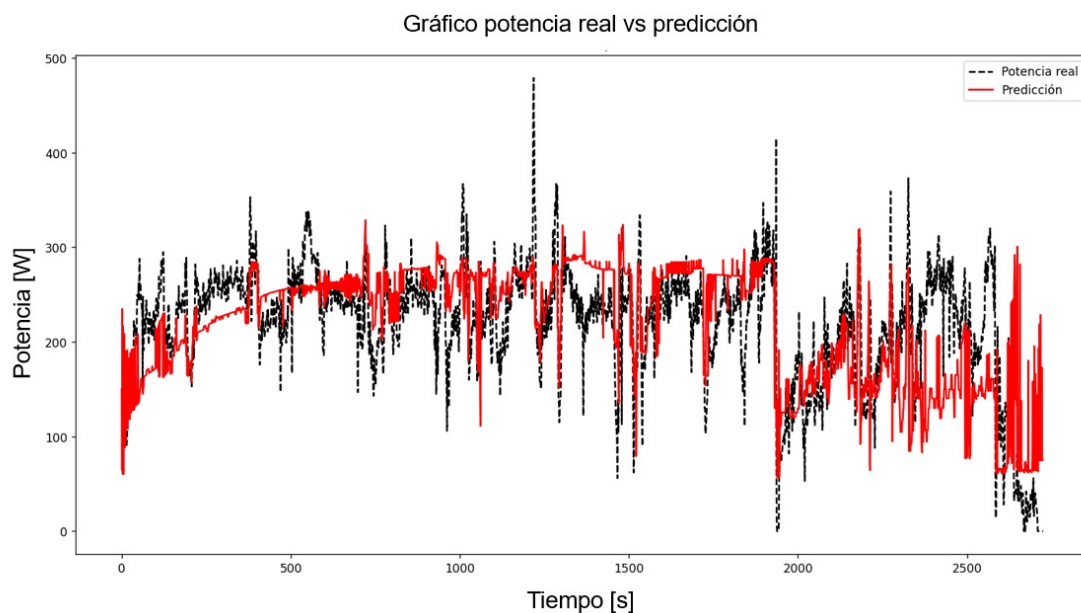


Figura 3.9: Predicción de potencia mediante ML (Perfil Pearson)

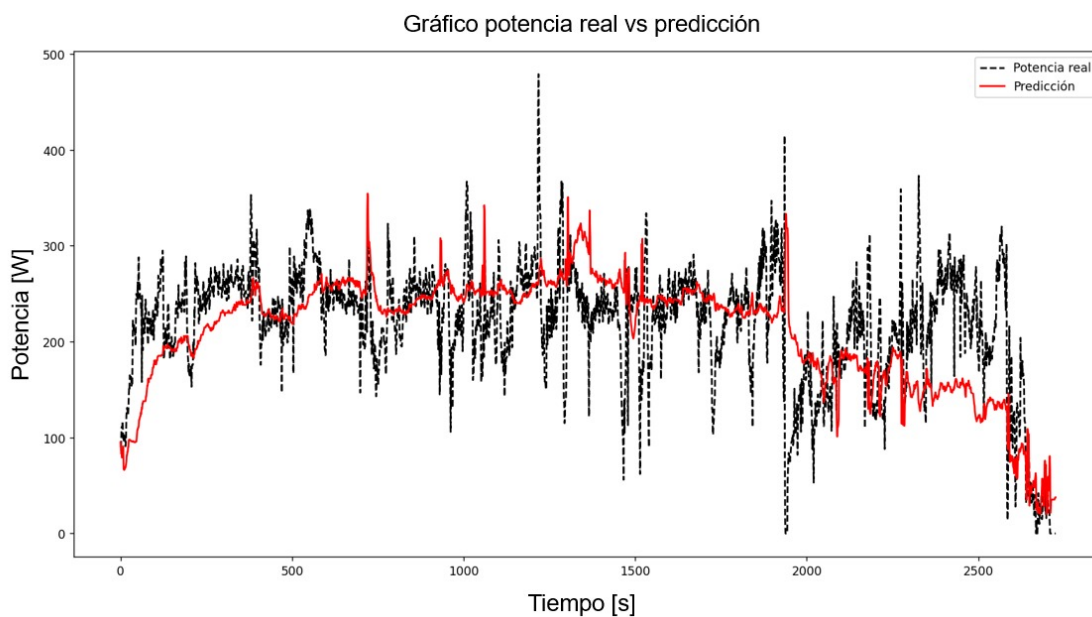


Figura 3.10: Predicción de potencia mediante ML (Perfil Spearman)

primeros intentos, estos resultados sirvieron como un punto de partida valioso para continuar con la mejora y optimización del modelo de red neuronal, destacando la importancia de explorar enfoques más avanzados y adaptados al problema específico de predicción de potencia en ciclismo.

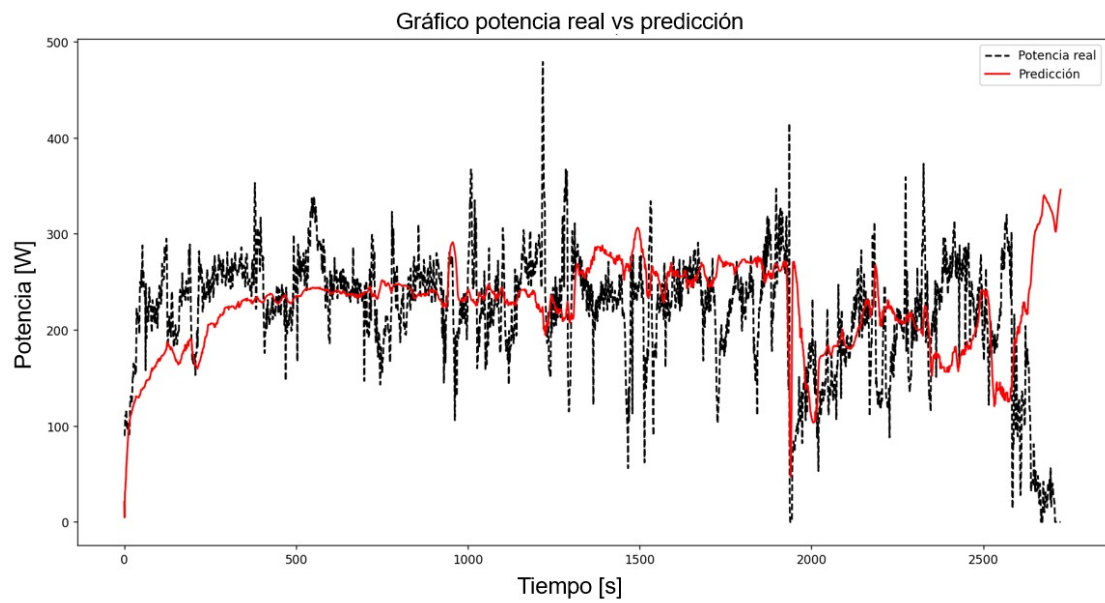


Figura 3.11: Predicción de potencia mediante ML (Perfil Feature importance)

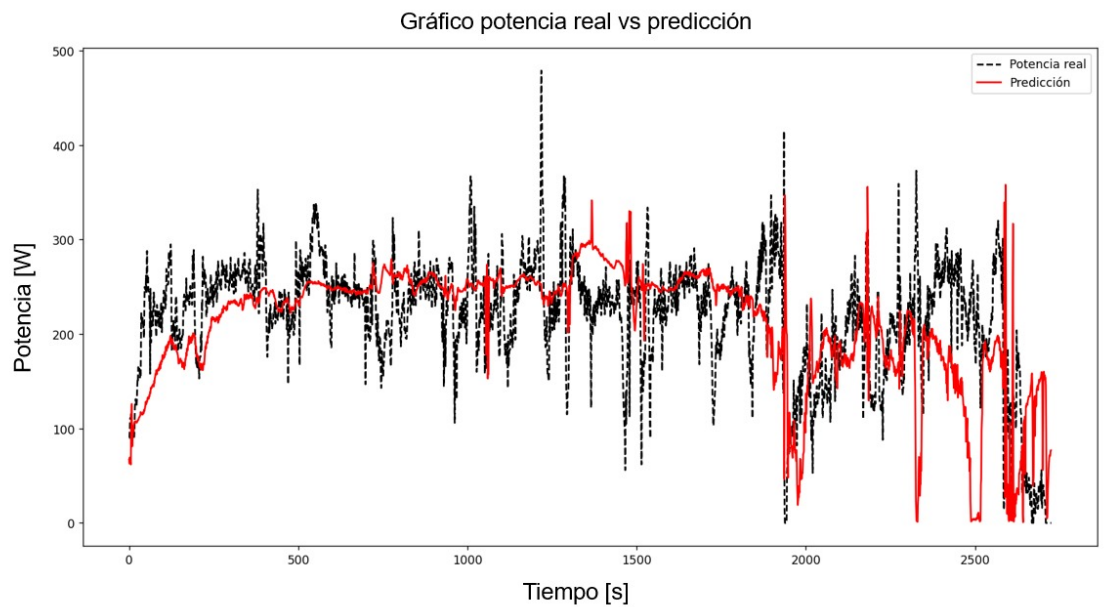


Figura 3.12: Predicción de potencia mediante ML (Perfil RFE)

n una fase posterior, como se detalló en el capítulo 2, se realizó un exhaustivo análisis de diversas combinaciones de modelos, registrando aquellos con un error inferior a 0.2 en el Error Cuadrático Medio (MSE). Un total de 62 modelos cumplieron con este criterio, y de entre ellos, se seleccionaron los cinco mejores para comparar sus predicciones con

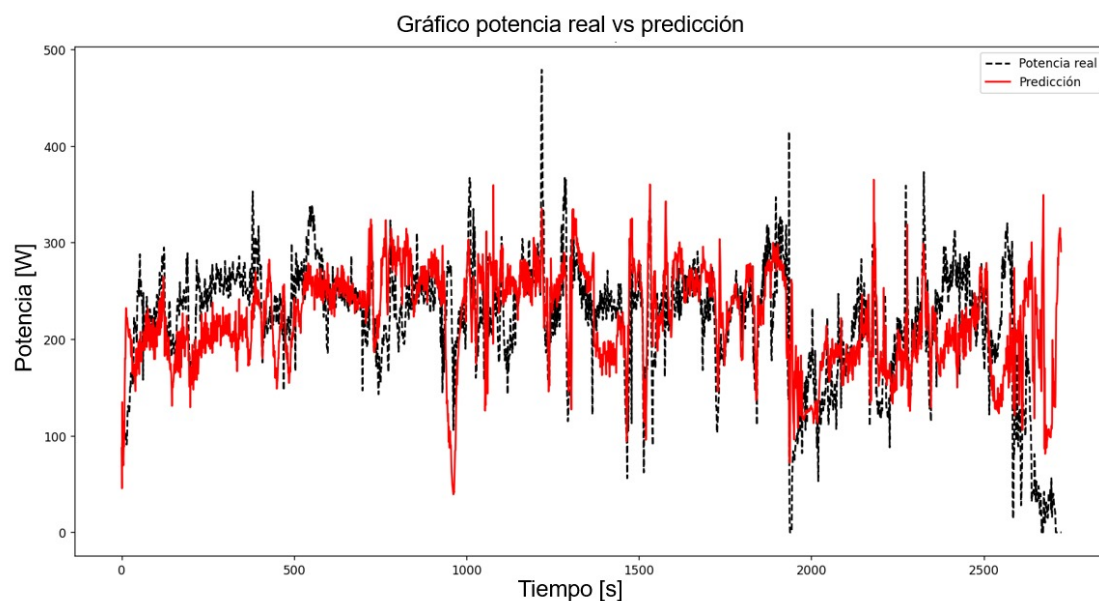


Figura 3.13: Predicción de potencia mediante ML (Perfil PearsonR)

las curvas utilizadas en el modelo matemático. La tabla 3.17 muestra la comparación de errores entre los cinco mejores modelos y el modelo matemático. En promedio, el modelo "2,3,4,7", que incluye las características "Tiempo, Velocidad, Inclinación, Altitud", demostró el mejor rendimiento.

Cabe destacar que cada uno de los cinco modelos de red neuronal superó notablemente al modelo matemático en términos de rendimiento promedio, subrayando la eficacia y superioridad de estos enfoques para predecir la potencia en ciclistas.

Las figuras 3.14, 3.15, 3.16, que corresponden a las actividades 1, 2 y 3, respectivamente, ilustran el uso del modelo "2,3,4,7". Este modelo, destacado por su eficacia en la predicción de potencia, exhibe resultados que se asemejan significativamente a la curva de potencia real. Cabe resaltar su capacidad para adaptarse a las secciones más suaves de la curva de potencia, así como para manejar elevaciones más pronunciadas, como se observa en la predicción de la actividad 3. Aunque se aleja ligeramente de la curva real después de la mitad, conserva una forma muy similar a esta. Este enfoque, que incluyó la exploración de múltiples combinaciones de características y la selección de modelos basada en el rendimiento, permitió obtener predicciones más precisas y coherentes con la realidad en comparación con los primeros intentos basados en perfiles estadísticos.

Finalmente, es crucial recordar que ambos modelos fueron sometidos a pruebas con datos que presentaban exclusivamente rutas en ascenso. Será interesante evaluar el desempeño de ambos modelos en situaciones donde las rutas tengan elevaciones e inclina-

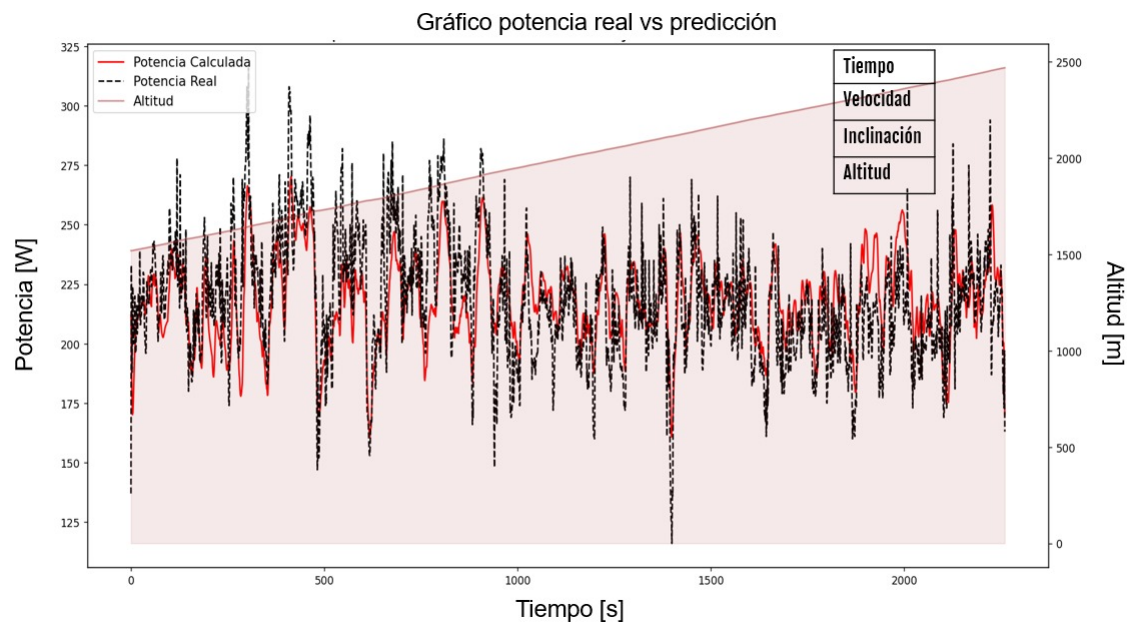


Figura 3.14: Predicción de potencia mediante ML (Actividad 1)

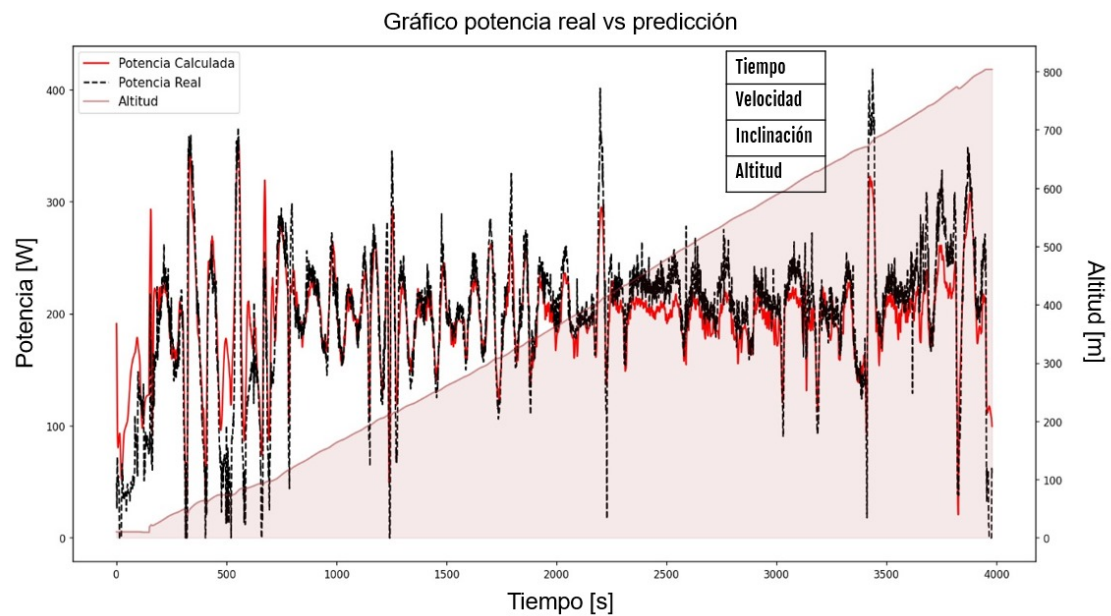


Figura 3.15: Predicción de potencia mediante ML (Actividad 2)

ciones más variadas. Al analizar el modelo matemático en la gráfica 3.18, se observa un rendimiento desfavorable en las secciones que representan terrenos planos y pendientes descendentes, lo cual era esperado según las explicaciones proporcionadas en el capítulo

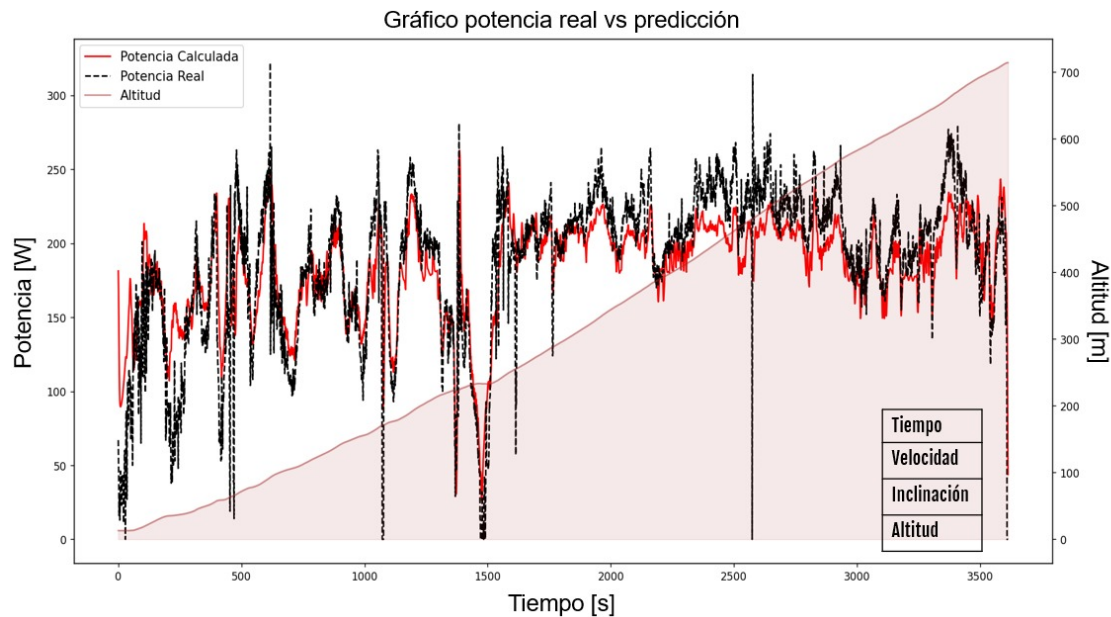


Figura 3.16: Predicción de potencia mediante ML (Actividad 3)

	Formula	M(2,3,4,7)	M(2,3,4,7,8)	M(2,3,4,7,9)	M(3,4,7)	M(3,4,7,8)
Actividad 1	178%	58%	65%	58%	60%	57%
Actividad 2	149%	28%	33%	28%	31%	30%
Actividad 3	36%	34%	41%	35%	35%	37%
Promedio	121%	40%	46%	40%	42%	42%
Prueba	10924%					

Figura 3.17: Comparación de 5 modelos con características distintas, incluyendo los errores del modelo matemático para las 3 actividades.

2.

En contraste, al examinar la gráfica 3.19, se aprecia un desempeño sobresaliente. A pesar de no haber sido entrenado con curvas de terrenos planos o en descenso, la predicción realizada es significativamente mejor que la del modelo matemático. La tabla de errores 3.17 refleja esta diferencia en el porcentaje de error en la fila denominada "Pruebas", destacando la abismal mejora del modelo de red neuronal en comparación con el modelo matemático.

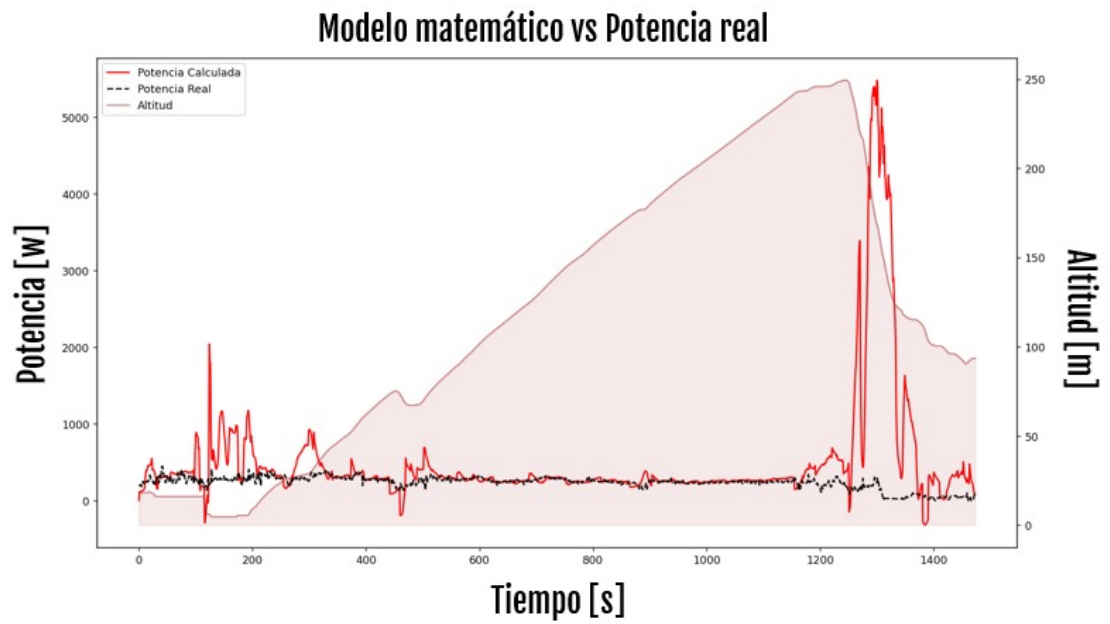


Figura 3.18: Comparación entre potencia calculada según el modelo matemático y potencia real.

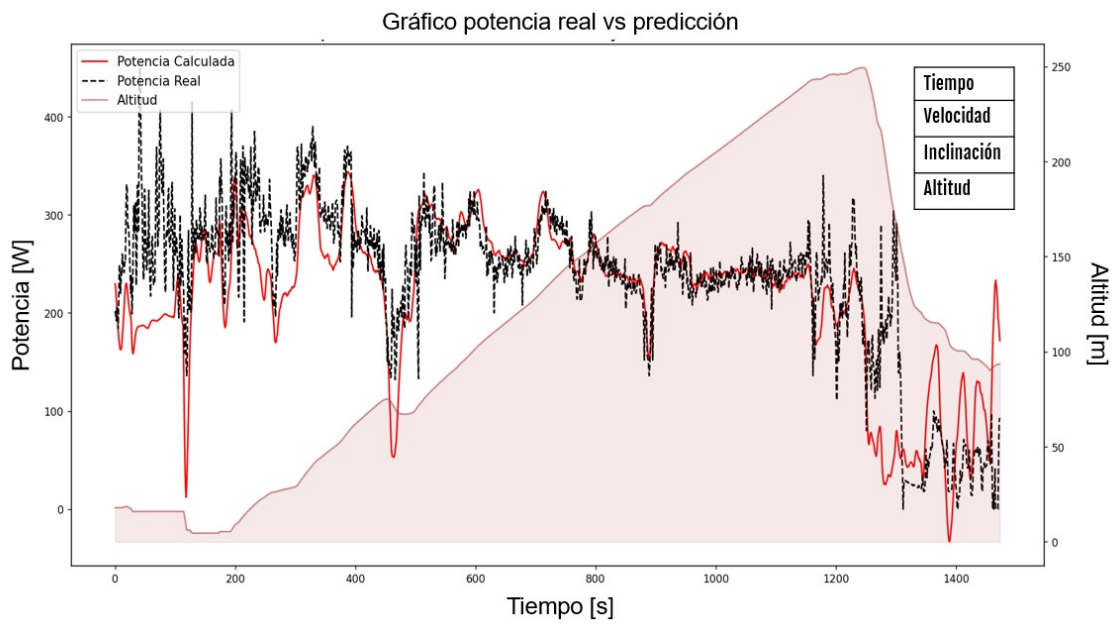


Figura 3.19: Comparación entre potencia calculada según la red neuronal y potencia real.

Capítulo 4

Conclusiones

Considerando lo obtenido por el método matemático, es interesante notar que, a pesar de que los valores de error fueron bastante grandes debido a la presencia de valores atípicos que se desviaban mucho de la curva real, la curva predicha no era tan deficiente. Esta fue capaz de predecir tanto bajadas como subidas, así como cambios bastante bruscos en la potencia. Dada la cantidad de características fuera de las consideradas que pueden afectar el rendimiento del ciclista, como la alimentación, el cansancio, el estrés, la temperatura y las condiciones climáticas, es lógico pensar desde un principio que una fórmula incapaz de considerar todos estos parámetros tendría dificultades para generar una curva con un bajo nivel de error.

Por otra parte, a pesar de que el modelo de machine learning (ML) recibe características aparentemente simples, como velocidad e inclinación, que también están presentes en la fórmula matemática, la ML demuestra su capacidad para identificar patrones que podrían ser prácticamente inobservables para un análisis estadístico convencional. En este sentido, se podría afirmar que, incluso cuando ciertas características no se especifican directamente, la ML puede interpretar y reconocer patrones implícitos en los demás datos.

Respecto a si se puede obtener un modelo capaz de predecir la potencia producida por un ciclista mediante características mucho más fáciles de medir, como la velocidad, la cadencia, la frecuencia cardíaca, entre otros, podemos afirmar que es completamente posible entrenar un modelo para cualquier persona que quiera realizar sus cálculos de potencia sin tener un potenciómetro, con una suficiente cantidad de datos de entrenamiento. Dado que durante el experimento se contó con una cantidad de datos relativamente pequeña pero suficiente para realizar los entrenamientos, podemos asumir que si incluimos el doble de datos o incluso más, podemos llegar a predicciones mucho más exactas de la poten-

cia. Además, como se observó en los resultados, los mejores modelos tienen parámetros que en principio deberían ser fáciles de conseguir con cualquier ciclocomputadora de bajo costo, como la velocidad, el tiempo y la altitud, y posteriormente la inclinación que sería calculada a partir de la altitud y la distancia recorrida.

La interpretación de los resultados generados por la machine learning puede resultar un desafío significativo. Por un lado, se puede argumentar que los resultados tienen coherencia física, ya que la velocidad, como se destacó anteriormente, desempeña un papel crucial en la predicción de la potencia. Además, la inclinación también es un factor importante para determinar el rendimiento del ciclista.

No obstante, la altitud podría presentar una relación menos directa con la potencia, y su interpretación puede variar en función de diversos factores. La complejidad radica en que la machine learning funciona como una “caja negra”, ejecutando miles de procesos por segundo y descubriendo patrones que podrían ser prácticamente invisibles sin su capacidad de análisis avanzado. Intentar atribuir un significado específico a cada uno de estos procesos podría ser una tarea desafiante y, en muchos casos, podría carecer de sentido.

Es importante tener en cuenta que la machine learning opera en base a los datos proporcionados, y su rendimiento podría mejorar significativamente con conjuntos de datos más extensos. En este sentido, la búsqueda de significado en cada aspecto del modelo podría no ser tan crucial como comprender la capacidad general del modelo para capturar patrones complejos y realizar predicciones precisas.

En resumen, se establece la factibilidad de que cualquier individuo desarrolle un estimador de su perfil de potencia para cualquier ruta, prescindiendo de la necesidad de un potenciómetro. La capacidad predictiva del modelo tiende a mejorar proporcionalmente al incremento del tamaño de la base de datos utilizada para su entrenamiento. Actualmente, las predicciones demuestran ser altamente precisas, especialmente en rutas con inclinaciones predominantemente positivas que facilitan una transmisión constante de energía del ciclista a los pedales.

Estos resultados abren nuevas posibilidades para explorar diversos enfoques y objetivos en modelos de machine learning. Además, se plantea la perspectiva de utilizar segmentos específicos del modelo actualmente desarrollado para evolucionarlo y abordar una variedad de interrogantes y desafíos adicionales en futuras investigaciones.

Bibliografía

- [1] ¿qué es una red neuronal? parte 1 : La neurona — dotcsv.
- [2] ABADI, M., AGARWAL, A., BARHAM, P., BREVDO, E., CHEN, Z., CITRO, C., CORRADO, G. S., DAVIS, A., DEAN, J., DEVIN, M., GHEMAWAT, S., GOODFELLOW, I., HARP, A., IRVING, G., ISARD, M., JIA, Y., JOZEFOWICZ, R., KAISER, L., KUDLUR, M., LEVENBERG, J., MANÉ, D., MONGA, R., MOORE, S., MURRAY, D., OLAH, C., SCHUSTER, M., SHLENS, J., STEINER, B., SUTSKEVER, I., TALWAR, K., TUCKER, P., VANHOUCKE, V., VASUDEVAN, V., VIÉGAS, F., VINYALS, O., WARDEN, P., WATTENBERG, M., WICKE, M., YU, Y., AND ZHENG, X. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
- [3] ALVY. El póster de las redes neuronales, Sep 2016.
- [4] ATKINSON, G., PEACOCK, O., ST CLAIR GIBSON, A., AND TUCKER, R. Distribution of power output during cycling: impact and mechanisms. *Sports Medicine* 37 (2007), 647–667.
- [5] BLOCKEN, B., DEFRAEYE, T., KONINCKX, E., CARMELIET, J., AND HESPEL, P. Cfd simulations of the aerodynamic drag of two drafting cyclists. *Computers & Fluids* 71 (2013), 435–445.
- [6] BLOCKEN, B., VAN DRUENEN, T., TOPARLAR, Y., MALIZIA, F., MANNION, P., ANDRIANNE, T., MARCHAL, T., MAAS, G.-J., AND DIEPENS, J. Aerodynamic drag in cycling pelotons: New insights by cfd simulation and wind tunnel testing. *Journal of Wind Engineering and Industrial Aerodynamics* 179 (2018), 319–337.
- [7] BOLLT, S. A. Equivalent speed: A new metric for cycling performance. *Medicine & Science in Sports & Exercise* 53, 8S (2021), 29.

- [8] BREFELD, U., DAVIS, J., VAN HAAREN, J., AND ZIMMERMANN, A. Machine learning and data mining for sports analytics. *Cham: Springer* (2018).
- [9] BUNKER, R., AND SUSNJAK, T. The application of machine learning techniques for predicting match results in team sport: A review. *Journal of Artificial Intelligence Research* 73 (2022), 1285–1322.
- [10] COHEN, C., BRUNET, E., ROY, J., AND CLANET, C. Physics of road cycling and the three jerseys problem. *Journal of Fluid Mechanics* 914 (2021), A38.
- [11] CROUCH, T. N., BURTON, D., BROWN, N. A. T., THOMPSON, M. C., AND SHERIDAN, J. Flow topology in the wake of a cyclist and its effect on aerodynamic drag. *Journal of Fluid Mechanics* 748 (2014), 5–35.
- [12] DYER, B. T. Cycle e-racing: Simulation or a new frontier in sports technology? *International Journal of Esports* 1, 1 (2020).
- [13] FENG, J., LIU, X., AND DENG, S. Optimal pacing strategy modeling of cycling individual time trials. In *Journal of Physics: Conference Series* (2022), vol. 2282, IOP Publishing, p. 012003.
- [14] GONZÁLEZ-HARO, C., BALLARINI, P. G., SORIA, M., DROBNIC, F., AND ESCANERO, J. F. Comparison of nine theoretical models for estimating the mechanical power output in cycling. *British Journal of Sports Medicine* 41, 8 (2007), 506–509.
- [15] GU, W., FOSTER, K., SHANG, J., AND WEI, L. A game-predicting expert system using big data and machine learning. *Expert Systems with Applications* 130 (2019), 293–305.
- [16] HALLIDAY, D., RESNICK, R., AND WALKER, J. *Fundamentals of physics*. John Wiley & Sons, 2013.
- [17] HARRIS, C. R., MILLMAN, K. J., VAN DER WALT, S. J., GOMMERS, R., VIRTANEN, P., COUNAPEAU, D., WIESER, E., TAYLOR, J., BERG, S., SMITH, N. J., KERN, R., PICUS, M., HOYER, S., VAN KERKWIJK, M. H., BRETT, M., HALDANE, A., DEL RÍO, J. F., WIEBE, M., PETERSON, P., GÉRARD-MARCHANT, P., SHEPPARD, K., REDDY, T., WECKESSER, W., ABBASI, H., GOHLKE, C., AND OLIPHANT, T. E. Array programming with NumPy. *Nature* 585, 7825 (Sept. 2020), 357–362.

-
- [18] HOCHREITER, S., AND SCHMIDHUBER, J. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.
- [19] HOENIGMAN, R., BRADLEY, E., AND LIM, A. Cooperation in bike racing—when to work together and when to go it alone. *Complexity* 17, 2 (2011), 39–44.
- [20] HORVAT, T., AND JOB, J. The use of machine learning in sport outcome prediction: A review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 10, 5 (2020), e1380.
- [21] HUNTER, J. D. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering* 9, 3 (2007), 90–95.
- [22] INC., T. M. Statistics and machine learning toolbox, 2022.
- [23] JOBSON, S. A., PASSFIELD, L., ATKINSON, G., BARTON, G., AND SCARF, P. The analysis and utilization of cycling training data. *Sports medicine* 39 (2009), 833–844.
- [24] KATAOKA, Y., AND GRAY, P. Real-time power performance prediction in tour de france. In *Machine Learning and Data Mining for Sports Analytics: 5th International Workshop, MLSA 2018, Co-located with ECML/PKDD 2018, Dublin, Ireland, September 10, 2018, Proceedings* 5 (2019), Springer, pp. 121–130.
- [25] KYLE, C. R. Reduction of wind resistance and power output of racing cyclists and runners travelling in groups. *Ergonomics* 22, 4 (1979), 387–397.
- [26] LOTFI, S., ET AL. Machine learning for sport results prediction using algorithms. *International Journal of Information Technology and Applied Sciences (IJITAS)* 3, 3 (2021), 148–155.
- [27] LU, C.-J., LEE, T.-S., WANG, C.-C., AND CHEN, W.-J. Improving sports outcome prediction process using integrating adaptive weighted features and machine learning techniques. *Processes* 9, 9 (2021), 1563.
- [28] MARTINS RATAMERO, E. Modelling peloton dynamics in competitive cycling: a quantitative approach. In *Sports Science Research and Technology Support: International Congress, icSPORTS 2013, Vilamoura, Algarve, Portugal, September 20-22, 2013. Revised Selected Papers* (2015), Springer, pp. 42–56.
- [29] MENASPA, P. Analysis of road sprint cycling performance.

- [30] NEE, P., AND HERTERICH, J. Modelling road cycling as motion on a curve. *Sports Engineering* 25, 1 (2022), 12.
- [31] NGUYEN, N. H., NGUYEN, D. T. A., MA, B., AND HU, J. The application of machine learning and deep learning in sport: predicting nba players' performance and popularity. *Journal of Information and Telecommunication* 6, 2 (2022), 217–235.
- [32] NVIDIA, VINGELMANN, P., AND FITZEK, F. H. Cuda, release: 10.2.89, 2020.
- [33] PANDAS DEVELOPMENT TEAM, T. pandas-dev/pandas: Pandas, Feb. 2020.
- [34] PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V., ET AL. Scikit-learn: Machine learning in python. *Journal of machine learning research* 12, Oct (2011), 2825–2830.
- [35] RATAMERO, E. M. Moped: an agent-based model for peloton dynamics in competitive cycling. In *International Congress on Sports Science Research and Technology Support* (2013), vol. 2, SCITEPRESS, pp. 134–143.
- [36] RICHTER, C., O'REILLY, M., AND DELAHUNT, E. Machine learning in sports science: challenges and opportunities. *Sports Biomechanics* (2021), 1–7.
- [37] ROBINSON, M. E., PLASSCHAERT, J., AND KISAALITA, N. R. Effects of high intensity training by heart rate or power in recreational cyclists. *Journal of sports science & medicine* 10, 3 (2011), 498.
- [38] SPRAGG, J., LEO, P., AND SWART, J. Predicting power outputs in a fatigued state: A pilot study. *Journal of Science and Cycling* 11, 2 (2022).
- [39] TRENCHARD, H. The complex dynamics of bicycle pelotons. *arXiv preprint arXiv:1206.0816* (2012).
- [40] TRENCHARD, H., RATAMERO, E., RICHARDSON, A., AND PERC, M. A deceleration model for bicycle peloton dynamics and group sorting. *Applied Mathematics and Computation* 251 (2015), 24–34.
- [41] VOGT, S., HEINRICH, L., SCHUMACHER, Y. O., BLUM, A., ROECKER, K., DICKHUTH, H.-H., AND SCHMID, A. Power output during stage racing in professional road cycling. *Medicine & Science in Sports & Exercise* 38, 1 (2006), 147–151.

-
- [42] VOGT, S., SCHUMACHER, Y. O., BLUM, A., ROECKER, K., DICKHUTH, H.-H., SCHMID, A., AND HEINRICH, L. Cycling power output produced during flat and mountain stages in the giro d'italia: a case study. *Journal of sports sciences* 25, 12 (2007), 1299–1305.