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[2mm] 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**

Lista de tablas **ix**

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. Implementación mas actual de la tecnología 4
1.5. Física del ciclismo 6
1.6. Potencia, ritmo cardiaco y rendimiento 16
1.7. Machine learning: breve introducción en contexto 17
1.8. Retos actuales en física del ciclismo y sus aplicaciones 19

2. Datos y métodos **1**

2.1. Obtención de datos de ciclismo 1
2.2. Datos experimentales: reales 2
2.3. Datos experimentales: simuladores 4
2.4. Estructura de datos en ciclismo y apps relacionados 4
2.5. Métodos estadísticos y de muestreo 6
2.6. Machine learning 9
2.6.1. Entender el problema y los datos. 9
2.6.1.1. Pearson. 11
2.6.1.2. Spearman. 12

2.6.1.3. Feature Importance. 13
2.6.1.4. Eliminación Recursiva de Características (RFE). 14
2.6.1.5. Prueba de correlación de PearsonR. 15
2.6.2. Elegir el tipo de red. 16
2.6.3. Preprocesamiento de los datos 16
2.6.4. Construir el modelo 18
2.6.5. Evaluar y ajustar el modelo 19
2.6.6. Hiperparametros 19

3. Resultados y conclusiones **1**
3.1. Resultados modelo matematico 1
3.2. Resultados Machine Learning 6
3.3. Conclusiones 7

Bibliografía **11**

Índice de figuras

1.1. 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. 10

1.2. 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. 12

1.3. 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. 14

1.4. Arrastre de cada ciclista en el pelotón A como porcentaje del arrastre de un ciclista aislado que circula a la misma velocidad. 15

1.5. Arrastre de cada ciclista en el pelotón B como porcentaje del arrastre de un ciclista aislado que circula a la misma velocidad. 15

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. . . 3
2.2. Aquí se observan las 4 rutas seleccionadas para la segunda instancia de recolección de datos. 5
2.3. Detalle de características con unidades de medida y número. 7
2.4. En este grafico de violín se observan 7 recorridos, y esto permite entender la distribución de los datos de potencia de los recorridos. 9
2.5. En esta figura se puede observar la comparación entre dos curvas de potencia, donde la curva mas suave (azul) se considera como un perfil de potencia normal. En cambio, la curva que posee muchos valles y crestas (naranja), se considera como ruido, ya que esta se consigue al registrar la curva de potencia de un entrenamiento basado en zonas de potencia o cardiacas. 10
2.6. Predicción de potencia de la primera versión de la Red Neuronal. 21

3.1. Cálculo de potencia mediante formula (Actividad 1) 2
3.2. Error en cálculo de formula (Actividad 1) 3
3.3. Cálculo de potencia mediante formula (Actividad 2) 3
3.4. Error en cálculo de formula (Actividad 1) 4
3.5. Cálculo de potencia mediante formula (Actividad 3) 4
3.6. Error en cálculo de formula (Actividad 1) 5
3.7. Predicción de potencia mediante ML (Actividad 1) 7
3.8. Predicción de potencia mediante ML (Actividad 2) 7
3.9. Predicción de potencia mediante ML (Actividad 3) 8
3.10. Comparación de 5 modelos con caracteristicas distintas, incluyendo los errores del modelo matematico para las 3 actividades. 8

Índice de tablas

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 [17] 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 club, 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 club, la estrategia de los dos del mismo club 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 su compañero y el solitario, 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, pero ese es un pequeño acercamiento a lo que ha evolucionado el ciclismo en este aspecto tan importante.

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. Implementación mas actual de la tecnología

Los primeros modelos de bicicleta, por muy impensable que sea, funcionaban con llantas de madera cubiertas de hierro. Obviamente estas bicicletas resultaban ser muy incomodas 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, como se imaginaran, 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 [10] 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 a una plataforma de e-racing y compiten siendo representados de manera virtual por un avatar digital. Esta forma de competencia ha ganado bastante reconocimien-

to, 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.

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 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 une 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. Otro de los componentes que mayor perdida 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 encuentran 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” [14], 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 articulo “Physics of

road cycling and the three jerseys problem" [8], 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.

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

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.

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

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:

$$V_\mu = \sqrt{2\mu M g / \rho S C_D} \tag{1.3}$$

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:

$$P_m = \frac{1}{2}\rho S C_D V^3 + Mg\alpha V + \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}MV^2\right) + \mu MgV \tag{1.4}$$

α=inclinación

Donde podemos ver a simple vista las contribuciones antes mencionadas. En primer lugar la contribución por resistencia aerodinámica, luego la gravitatoria, la aceleración y la resistencia por rodadura. Con lo cual somos capaces de calcular en primera instancia la potencia mecánica de un ciclista.

Figura 1.1: 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.

Ahora haremos un acercamiento un poco mas especifico a uno de los aspectos que

mas suelen afectar al rendimiento del ciclista, la aerodinámica, pero con una visión mas detallada. Principalmente lo que mas se estudia y conoce es el efecto del viento sobre la superficie de contacto que en este caso es el ciclista, ya que es donde mas se pierde energía, y por esto los ciclistas en su mayoría optan por intentar tener la mejor posición para tener el menor área de contacto directo con el viento, o sea buscan disminuir un valor que se conoce como Frontal drag área. En este ámbito, se han creado muchas técnicas en las que el frontal drag área disminuye bastante, pero hay posiciones que han sido creadas que llevan este concepto a un limite que puede llegar a ser incluso peligroso, como se puede ver en la figura 2.3, los ciclistas pueden llegar a adoptar posiciones que disminuyen drásticamente su área frontal, pero dado su grado de peligrosidad estas posiciones han tenido que ser prohibidas en el ciclismo de competición.
Un tema bastante complejo de estudiar es como se comporta el aire como fluido al estar fluyendo sobre el ciclista. Generalmente la parte superior del ciclista tiende a ser simétrica, lo que evita que se generen flujos de aire irregulares, pero de la cadera para abajo la cosa cambia, dado que constantemente se está variando la simetría del tren inferior. Para esto, T N. Crocuh, D. Burton y otros, en su estudio "Flow topology in the wake of a cyclist and its efect on aerodynamic drag" [9], desarrollan mediante técnicas de Tunel de Viento, cálculos para determinar el efecto del flujo del aire y ciertas posiciones del ciclista en la resistencia aerodinámica. Donde discuten que existen ciertos ángulos del ciclo de la biela (considerando el ángulo 0 cuando ambas bielas están orientadas de forma horizontal), donde el arrastre no se ve afectado (dentro de los primeros 15 grados de rotación) y otros donde se presentan máximos (en las posiciones de 75 y 255 grados). Cuando el maniquí se exponían a regímenes de alta resistencia, se observaba que se generaban grandes vórtices en su estela; vórtices generados por la separación del flujo de aire en ciertas partes del cuerpo. Evidentemente estos vórtices generaban una mayor resistencia al avance, por lo que la posición de las piernas juega un papel muy importante en la formación y ubicación de los vórtices. Cuando las piernas estaban en una posición elevada, se generaba un vórtice inferior cerca del centro del cuerpo, mientras que en una posición extendida se formaba un vórtice superior más alejado del centro del cuerpo. Se puede observar en la figura 1.2 como para diferentes posiciones de las piernas o distintos grados de la biela, el flujo de aire y los vórtices generados son completamente distintos y asimétricos, por lo que podemos entender que el fenómeno aerodinámico no es tan simple como calcular el área frontal y utilizar un coeficiente de arrastre para entenderlo realmente. Así como un ciclista puede contemplar estos factores que son hasta cierto punto mas controlables por si mismo y por su equipo, hay factores que incluso se escapan de su propio control, y que pueden llegar a ser incluso mas determinantes que este.

Figura 1.2: 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.

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 mas 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' [4], 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 mas pequeños para validar la configuración y los parámetros de los dos pelotones mas grandes.

Los resultados de estas pruebas fueron bastante interesantes. En la figura 1.3 se pueden 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 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 la fig 1.4 y 1.5 observamos el porcentaje de drag de cada uno

Figura 1.3: 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.

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 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 lideres 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

Figura 1.4: Arrastre de cada ciclista en el pelotón A como porcentaje del arrastre de un ciclista aislado que circula a la misma velocidad.

Figura 1.5: Arrastre de cada ciclista en el pelotón B como porcentaje del arrastre de un ciclista aislado que circula a la misma velocidad.

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 [3]. 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 [22], 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. Potencia, ritmo cardiaco y rendimiento

Cuando queremos hablar de 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 juega un rol muy importante en el rendimiento del deportista. Una parte del nivel al que rinda un deportista puede llegar a depender puramente de sus entrenamientos, como por ejemplo su potencia, su fuerza y su capacidad aeróbica, que en el fondo son áreas del deporte que se han investigado durante años y sobre las que hay mucho conocimiento a un alcance "fácil", sin embargo existen otros aspectos del rendimiento, que debido a la falta de estudios y discusión, carecen de una facilidad para su aplicación y conocimiento general, como por ejemplo temas de nutrición y alimentación especializados al deporte que se realiza, o técnicas especializadas en mejorar el rendimiento en entrenamientos estándar, como puede ser la utilización de medidores de rendimiento, como sensores de cadencia, pulsómetros, potenciómetros e incluso software de análisis de datos. Con el tiempo algunos de estos dispositivos ha logrado estar al alcance de muchas mas personas, como los pulsómetros y los medidores de cadencia, pero por otro lado, existen productos como los ciclo computadores y potenciómetros que son lo mas utilizado por ciclistas de elite para mejorar el rendimiento y el problema, es que este tipo de productos difícilmente suele estar al alcance de un ciclista amateur, alguien que le gusta mucho el ciclismo y le gustaría dar un paso mas en la búsqueda de conseguir su meta, pero que se las tiene que valer por si mismo para avanzar. De primeras, esta persona va a querer mejorar lo mas posible su bicicleta para poder acotar su desarrollo y desempeño lo mas posible a su rendimiento como ciclista, lo que probablemente lo llevará a intentar armar su propia bicicleta, y esto ya de por si, suele ronda los 1000-2000 dólares. Lo siguiente es comprar un equipo deportivo acorde a las intenciones que tenga el deportista donde se con un par de búsquedas se llegó a un aproximado de 500 dólares mas, luego un sensor de frecuencia cardiaca puede costar 50 dólares al igual que un medidor de cadencia, por lo que ya serían 100 dólares mas. Ya solo con esto van 1600-2600 dólares, por lo que nuestro presupuesto solo para

empezar avanzar en el ciclismo amateur ya son bastante alto, y aquí es cuando llega uno de los artefactos que mencionamos antes, el potenciómetro, que suele rondar los 700-800 dólares. En resumen, podemos decir que el ciclismo puede considerarse un deporte para gente con muchos recursos, que la media de la población simplemente no tiene.

Entonces 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 [34].

Con esto y otros estudios que se han hecho podemos concluir que realmente el entrenamiento por potencia no se diferencia tanto de los entrenamientos por potencia, por lo que podemos asumir que estos dos aspectos tienen una correlación bastante alta, hoy en día hasta existe una aplicación que es capaz de relacionar directamente los datos de frecuencia cardiaca con los de potencia, sacando un aproximado de la potencia que se acerca bastante a la realidad. Por ende, no es loco pensar que uno de los factores que mas a tener en cuenta sea la frecuencia cardiaca, si lo que se busca es lograr obtener los datos de potencia mediante la relación otro tipo de data.

1.7. Machine learning: breve introducción en contexto

Hoy en día el Machine larning 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 ciclísticos 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 [33], 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 (r2) 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 [28]
- The Application of Machine Learning Techniques for Predicting Match Results in Team Sport: A Review [7]
- Improving Sports Outcome Prediction Process Using Integrating Adaptive Weighted Features and Machine Learning Techniques [24]
- A game-predicting expert system using big data and machine learning [13]

Por otro lado, bansandose en la explicación que se da en (cita a matlab red neuronal), podemos entender las redes neuronales como una forma de imitar el funcionamiento de la estructura neuronal del cerebro. Cada neurona

1.8. Retos actuales en física del ciclismo y sus aplicaciones

La física del ciclismo presenta retos muy interesantes en la actualidad, desde aplicaciones básicas de física clásica hasta aplicaciones mas complejas como la biomecánica. Uno de los retos mas complejos a superar es la resistencia del aire, que suele ser la principal barrera para aumentar la velocidad e incluso puede llegar a ser determinante en días que el clima no sea muy favorable. Por ende uno de los desafíos mas importantes en este ámbito es lograr reducir la resistencia del ciclista y por tanto, buscar cada vez modelos de implementación y equipo mas aerodinámicos. Por otro lado, la biomecánica es algo igual de importante, dado que la forma en que un ciclista pedalea, su postura y su técnica pueden llegar a influir significativamente en su desempeño. Lograr entender esto, y poder optimizar cada uno de esos aspectos ha sido un reto constante en este deporte. Luego, tenemos la transferencia de energía que es una de las cosas que mas se pasa por alto, pero para que el ciclista sea capaz de aprovechar la mayor cantidad de energía para avanzar, se necesitan

componentes especializados en transferir energía de manera eficiente, para evitar la mayor cantidad de perdidas en el camino que recorre desde los pies del ciclista hasta el suelo.

Incluso se podría llegar a pensar en algún otro dato que no conocemos, que podría llegar a ser mas determinante que la potencia incluso, y que estemos pasando por alto.

Capítulo 2

Datos y métodos

2.1. Obtención de datos de ciclismo

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 realizo la propuesta de hacer un recorrido en bicicleta para conocernos mejor, y conversar detalles mas puntuales. Un par de días después se concreto la propuesta y nos reunimos para realizar uno de los recorridos mas comunes de ellos, que suelen utilizar para ïniciar.ª los nuevos ciclistas del grupo.

2.2. Datos experimentales: reales

Para la primera instancia de recolección se redacto 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 criterios eran los siguientes:

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

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 de sus datos. Una vez finalizado el periodo de recolección, se descargan todos los datos subidos por los participantes, y se procesan ingresándolos en un código que permite extraer los

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.

datos de los archivos en formato tanto GPX como TCX.
En este momento tenemos varios problemas con esta primera recolección, primero que todo, no establecimos carpetas para separar los datos por ruta, por lo cual, si las rutas no estaban nombradas era muy difícil identificar cual era cual. En segundo lugar, una de las cosas que necesitábamos es que las rutas fueran las mismas, pero al revisar cada archivo se observaron variaciones durante la carrera, como diferencias de longitud o en el recorrido, donde hacían el mismo recorrido hasta cierto punto pero eventualmente cambiaban de dirección y hacía que todas las carreras fueran distintas. Esto era un problema dado que en este punto de la investigación se consideraba muy importante que las rutas de entrenamiento fueran iguales para que fuera mas facil para la ML aprender.

Continuar...

2.3. Datos experimentales: simuladores

Para la segunda instancia de recolección, se presentaron a los participantes indicaciones mucho más simples que en la primera instancia. Estas instrucciones mencionaban que todos los participantes debían tener un rodillo de entrenamiento (Trainer) y un medidor de frecuencia cardíaca. En este caso, todos los participantes poseían un Smart Trainer o también conocido como Rodillo de transmisión directa inteligente. La mayoría de estos rodillos se utilizan sin rueda trasera, pero algunos si permiten tenerla puesta; se conectan mediante la cadena y los platos, y el marco de la bicicleta se encaja en la estructura. Este rodillo funciona mediante imanes de neodimio y a pesar de ser magnético es bastante silencioso.

Una vez que se plantearon los puntos importantes, se matizaron algunos detalles, como las rutas específicas que se eligieron revisando la aplicación de Zwift, conocida como una plataforma para simuladores virtuales de ciclismo. Además, se requirió que los participantes tuvieran una cuenta en Zwift, incluso si fuera una cuenta gratuita o en caso de no poder, que realizaran las mismas rutas indicadas en Strava. Considerando ciertos parámetros como la dificultad de la ruta, el porcentaje de inclinación y la disponibilidad en la aplicación, se seleccionaron cuatro rutas principales que podrían realizar los participantes que se pueden observar en .

2.4. Estructura de datos en ciclismo y apps relacionados

Los datos pueden venir en dos formatos como se menciona anteriormente, para esto se escribe un código que logra extraer los datos de cualquiera de los dos formatos de archivo que se muestran a continuación y los convierte en listas de datos.

TCX:

```
<Trackpoint>
  <Time>2023-02-20T01:23:34Z</Time>
  <Position>
   <LatitudeDegrees>-11.6396790</LatitudeDegrees>
   <LongitudeDegrees>166.9479510</LongitudeDegrees>
```

Figura 2.2: Aquí se observan las 4 rutas seleccionadas para la segunda instancia de recolección de datos.

```
</Position>
<AltitudeMeters>8.6</AltitudeMeters>
<DistanceMeters>1.9</DistanceMeters>
<HeartRateBpm>
 <Value>69</Value>
</HeartRateBpm>
<Cadence>28</Cadence>
<Extensions>
 <TPX xmlns="http://www.garmin.com/xmlschemas/ActivityExtension/
  <Speed>0.0</Speed>
```

```
      <Watts>48</Watts>
     </TPX>
    </Extensions>
  </Trackpoint>
```

GPX:

```
<trkpt lat="-11.6354570" lon="166.9522120">
 <ele>1.6</ele>
 <time>2023-07-03T14:31:58Z</time>
 <extensions>
  <power>96</power>
  <gpxtpx:TrackPointExtension>
   <gpxtpx:hr>71</gpxtpx:hr>
   <gpxtpx:cad>41</gpxtpx:cad>
  </gpxtpx:TrackPointExtension>
 </extensions>
</trkpt>
```

2.5. Métodos estadísticos y de muestreo

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 (Exploratory Data Analysis) para poder realizar en primer lugar 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

	Unidad	Número
Cadencia	RPM	1
Tiempo	s	2
Velocidad	m/s	3
Inclinación	Rad	4
HeartRate	f	5
Coordenadas	m	6
Altitud	m	7
Distancia	m	8
DistanciaMeta	m	9

Figura 2.3: Detalle de características con unidades de medida y número.

a identificar patrones, tendencias y relaciones entre variables. El EDA puede incluir técnicas como la creación de histogramas, diagramas de dispersión, gráficos de caja, gráficos de barras, mapas de calor y matrices de correlación, entre otros. Una vez que los datos han sido extraídos y preparados adecuadamente, están listos para ser utilizados en análisis más avanzados, modelos predictivos o cualquier otra aplicación que pueda ayudar a sacar el máximo provecho de la información recopilada durante el periodo de recolección. El proceso de extracción y procesamiento de datos es esencial para obtener conclusiones significativas y tomar decisiones basadas en evidencia a partir de los datos obtenidos.

El siguiente paso fue crear un cuaderno en Jupyter Notebook, que se suele recomendar para el EDA. Para comenzar, se llama a 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 ïnteractivo”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, primero con el grafico de Violín (que nos permite ver la densidad de datos de la característica seleccionada) de la figura 2.4, se observo un comportamiento no convencional en la característica de potencia de algunas mediciones. En principio, se noto que para un mismo recorrido, las formas de las graficas son en su mayoría bastante diferentes, incluso tomando en cuenta que desde el recorrido 2 al 6, son de la misma persona. Al revisar con mayor detalle, se determino que el causante de esto era la forma en que se habían hecho los recorridos por parte de los participantes, dado que en algunos casos ellos realizaron los recorridos con un formato de entrenamiento, lo que implica que durante la medición utilizaron métodos de entrenamiento que constan de realizar una variación intencionada de la potencia; se puede observar en la figura 2.5 un grafico donde se compara una medición normal del recorrido Climbers Gambit, versus una medición hecha durante un entrenamiento.

Por esto, pensando en el entrenamiento de la machine learning se determina en primera instancia descartar este tipo de datos, y dejarlos para una etapa posterior. Segundo, con el grafico de Caja se observaron datos atípicos en la característica de la inclinación. Que era calculada mediante dos datos, la altura y la distancia recorrida en un segundo con la siguiente formula:

$$Inclinación = \frac{Altura_F - Altura_I}{DistanciaHorizontal} \tag{2.1}$$

Esto era conveniente, dado que como se explico anteriormente cada cado era tomado con un segundo de diferencia, pero esto también puede significar ciertos datos atípicos dado que la frecuencia con la que calcula la inclinación a lo largo de la ruta varía constantemente. Para efectos de análisis, concluimos que la mejor opción era borrar estos datos atípicos, que eran muy pocos por cada ruta, poniendo un limite de inclinación en 0.35. Esto porque un 35 % de inclinación ya resulta incluso difícil para los mejores ciclistas

Figura 2.4: En este grafico de violín se observan 7 recorridos, y esto permite entender la distribución de los datos de potencia de los recorridos.

entonces datos mayores a esos simplemente generarían una variación mas complicada de asimilar por la machine learning.

2.6. Machine learning

Para el desarrollo de la machine learning se dividió el proceso en 5 pasos principales:

2.6.1. Entender el problema y los datos.

Para elegir el tipo de red necesitamos entender la naturaleza de nuestro problema. Para esto analizamos los datos que tenemos, en nuestro caso los datos de ciclismo están ordenados de manera temporal, de manera que cada segundo que pasa se registra un nuevo dato.

Figura 2.5: En esta figura se puede observar la comparación entre dos curvas de potencia, donde la curva mas suave (azul) se considera como un perfil de potencia normal. En cambio, la curva que posee muchos valles y crestas (naranja), se considera como ruido, ya que esta se consigue al registrar la curva de potencia de un entrenamiento basado en zonas de potencia o cardiacas.

Por ende, necesitamos redes neuronales que resuelvan un problema de series temporales. El principal objetivo es lograr relacionar de alguna manera los datos de entrada para obtener una predicción de la potencia a partir de esos datos. Para entender de mejor manera cuales son los datos realmente importantes y que características deberíamos tener mas en cuenta, haremos un análisis estadístico de los datos.

Primero que todo utilizaremos la librería Pandas (Cita a pandas) para analizar los datos con la función Describe.

```
                F. cardiaca     Altitud   Velocidad     Tiempo  \
count           2705.000000  2705.000000  2705.000000  2705.000000
```

mean	142.585952	29.498189	8.902181	1352.000000
std	11.975169	38.368560	2.801822	781.010563
min	69.000000	-6.200000	0.000000	0.000000
25%	139.000000	2.000000	6.900000	676.000000
50%	146.000000	10.400000	9.300000	1352.000000
75%	150.000000	52.000000	10.600000	2028.000000
max	163.000000	131.800000	16.000000	2704.000000

	Cadencia	Distancia	Inclinacion	Potencia
count	2705.000000	2705.000000	2705.000000	2705.000000
mean	68.627726	8.901442	0.005586	220.628096
std	13.642467	2.806362	0.028457	70.698036
min	0.000000	0.100000	-0.120000	0.000000
25%	64.000000	6.900000	0.000000	193.000000
50%	72.000000	9.300000	0.000000	230.000000
75%	77.000000	10.600000	0.020000	263.000000
max	95.000000	16.200000	0.130000	490.000000

La selección de características implica elegir las variables más relevantes para un análisis o modelo. En este caso, se utilizaron 5 métodos de correlación para ver cómo las características se relacionan con un resultado

Ahora veremos como se estudio la correlación entre variables:

2.6.1.1. Pearson.

	F.cardiaca	Altitud	Velocidad	Tiempo
Frecuencia cardiaca	1.000000	0.230920	-0.194604	0.215368
Altitud	0.230920	1.000000	-0.268846	0.431867
Velocidad	-0.194604	-0.268846	1.000000	-0.104211
Tiempo	0.215368	0.431867	-0.104211	1.000000
Cadencia	0.059970	-0.167787	0.422775	-0.371808
Distancia	-0.209096	-0.260506	0.995827	-0.111636
Inclinacion	0.370614	0.194006	-0.599647	0.000101
Potencia	0.279060	-0.141802	0.136391	-0.407232

	Cadencia	Distancia	Inclinacion	Potencia

Frecuencia cardiaca	0.059970	-0.209096	0.370614	0.279060
Altitud	-0.167787	-0.260506	0.194006	-0.141802
Velocidad	0.422775	0.995827	-0.599647	0.136391
Tiempo	-0.371808	-0.111636	0.000101	-0.407232
Cadencia	1.000000	0.429281	-0.176798	0.315605
Distancia	0.429281	1.000000	-0.638358	0.111503
Inclinacion	-0.176798	-0.638358	1.000000	0.393772
Potencia	0.315605	0.111503	0.393772	1.000000

Para entender esta tabla necesitamos saber lo siguiente:

- Correlación cercana a 1: Si dos características tienen una correlación cercana a 1, significa que existe una fuerte relación lineal positiva entre ellas. Esto significa que cuando una característica aumenta, la otra también tiende a aumentar. Un valor de 1 implica una correlación perfecta positiva.
- Correlación cercana a 0: Si dos características tienen una correlación cercana a 0, significa que no hay una relación lineal evidente entre ellas. Cuando una característica cambia, no podemos hacer una predicción confiable sobre cómo cambiará la otra característica basándonos en la correlación.
- Correlación cercana a -1: Si dos características tienen una correlación cercana a -1, significa que existe una fuerte relación lineal negativa entre ellas. Esto significa que cuando una característica aumenta, la otra tiende a disminuir. Un valor de -1 implica una correlación perfecta negativa.

Lista de correlación Pearson.

- Potencia - Tiempo (-0.4)
- Potencia - Cadencia (0.3)
- Potencia - Inclinación (0.39)

2.6.1.2. Spearman.

Probamos ahora el método de correlación de Spearman, esta correlación se basa en los rangos de tus datos, en lugar de sus valores reales. En otras palabras, cuando calculas la correlación de Spearman, primero conviertes los valores de tus datos en sus rangos (el número de orden si se ordenan), y luego calculas la correlación de Pearson en estos rangos.

	F.cardiaca	Altitud	Velocidad	Tiempo
Frecuencia cardiaca	1.000000	0.347139	-0.411253	0.162753
Altitud	0.347139	1.000000	-0.324346	0.533898
Velocidad	-0.411253	-0.324346	1.000000	-0.127870
Tiempo	0.162753	0.533898	-0.127870	1.000000
Cadencia	-0.309722	-0.344253	0.480970	-0.433055
Distancia	-0.414937	-0.316432	0.993505	-0.132628
Inclinacion	0.455630	0.205054	-0.587434	-0.004698
Potencia	0.177194	-0.250540	0.077643	-0.403112

	Cadencia	Distancia	Inclinacion	Potencia
Frecuencia cardiaca	-0.309722	-0.414937	0.455630	0.177194
Altitud	-0.344253	-0.316432	0.205054	-0.250540
Velocidad	0.480970	0.993505	-0.587434	0.077643
Tiempo	-0.433055	-0.132628	-0.004698	-0.403112
Cadencia	1.000000	0.481077	-0.279717	0.161421
Distancia	0.481077	1.000000	-0.629550	0.048483
Inclinacion	-0.279717	-0.629550	1.000000	0.386799
Potencia	0.161421	0.048483	0.386799	1.000000

- Potencia - Tiempo (-0.4)
- Potencia - Inclinación (0.38)

2.6.1.3. Feature Importance.

Este algoritmo de aprendizaje automático puede proporcionar un puntaje de importancia para cada característica, donde los puntajes más altos indican las características más importantes.

	feature	importance
7	Potencia	0.999822
6	Inclinacion	0.000033
3	Tiempo	0.000032
0	Frecuencia cardiaca	0.000028
5	Distancia	0.000024
2	Velocidad	0.000024

```
1              Altitud    0.000020
4             Cadencia    0.000016
```

Lista de correlación Feature Importance:

- potencia - inclinación (0.000033)
- potencia - Tiempo (0.000032)
- potencia - Frecuencia C. (0.000028)

2.6.1.4. Eliminación Recursiva de Características (RFE).

Es un método que encaja un modelo y elimina las características más débiles (las de menor importancia), luego repite el proceso con las características restantes, hasta un número especificado de características.

```
Características seleccionadas (True significa que la
característica fue seleccionada):
    Frecuencia cardiaca: False
    Altitud: False
    Velocidad: True
    Tiempo: True
    Cadencia: False
    Distancia: False
    Inclinacion: True

Ranking de las características (1 significa que la
característica es la más importante):
    Frecuencia cardiaca: 4
    Altitud: 2
    Velocidad: 1
    Tiempo: 1
    Cadencia: 3
    Distancia: 5
    Inclinacion: 1
```

Lista de correlación RFE.

- potencia - Velocidad
- potencia - Tiempo
- potencia - Inclinación

2.6.1.5. Prueba de correlación de PearsonR.

Evalúa la correlación entre cada característica y la variable objetivo. Un valor de p pequeño (generalmente menos de 0.05) indica que la característica está significativamente correlacionada con la variable objetivo.

Los resultados obtenidos son los siguientes:

1. La correlación de la Frecuencia cardíaca con la variable objetivo es 0.335 con un valor p muy pequeño (1.39e-69). Esto significa que hay una correlación positiva moderada entre la frecuencia cardíaca y tu variable objetivo, y que esta correlación es estadísticamente significativa. Es muy improbable que esta correlación se deba al azar.

2. La correlación de la Altitud con la variable objetivo es 0.021 con un valor p de 0.282. Esto indica una correlación muy baja, prácticamente nula, entre la altitud y tu variable objetivo, y el valor p no es lo suficientemente pequeño para considerar que esta correlación es estadísticamente significativa. En otras palabras, es bastante probable que cualquier correlación observada se deba al azar.

3. La correlación de la Velocidad con la variable objetivo es 0.142 con un valor p muy pequeño (2.87e-13). Esto indica una correlación positiva baja, pero estadísticamente significativa, entre la velocidad y tu variable objetivo.

4. La correlación del Tiempo con la variable objetivo es -0.211 con un valor p muy pequeño (1.23e-27). Esto indica una correlación negativa moderada entre el tiempo y tu variable objetivo, y que esta correlación es estadísticamente significativa. A medida que el tiempo aumenta, tu variable objetivo tiende a disminuir.

5. La correlación de la Cadencia con la variable objetivo es 0.349 con un valor p muy pequeño (9.97e-76). Esto indica una correlación positiva moderada entre la cadencia y tu variable objetivo, y que esta correlación es estadísticamente significativa.

6. La correlación de la Distancia con la variable objetivo es 0.121 con un valor p muy pequeño (5.97e-10). Esto indica una correlación positiva baja, pero estadísticamente significativa, entre la distancia y tu variable objetivo.

7. La correlación de la Inclinación con la variable objetivo es 0.419 con un valor p extremadamente pequeño (1.28e-111). Esto indica una correlación positiva moderada a alta entre la inclinación y tu variable objetivo, y que esta correlación es altamente estadísticamente significativa.

Todas las variables, excepto la altitud, tienen una correlación estadísticamente significativa con tu variable objetivo, aunque la fuerza de la correlación varía. La inclinación, la frecuencia cardíaca y la cadencia parecen ser las más correlacionadas con tu variable objetivo.

2.6.2. Elegir el tipo de red.

Elegir el tipo de red neural adecuada depende de la naturaleza de tus datos y el problema que estás tratando de resolver. En este caso, se está trabajando con datos tomados de forma secuencial y temporal para predecir la potencia. Esto sugiere que las mejores opciones son una red neuronal recurrente (RNN) y una capa LSTM debido a las características de secuencia en lo datos.

1. Red Neuronal Recurrente (RNN): Las RNN son aptas para tratar con secuencias de datos, ya que mantienen una ”memoriaïnterna que les permite recordar información de pasos de tiempo anteriores. Sin embargo, las RNN tradicionales pueden sufrir de problemas de desvanecimiento de gradientes, lo que dificulta el aprendizaje de relaciones a largo plazo en las secuencias.

2. Capa LSTM (Long Short-Term Memory): Las LSTMs son una variante de las RNN que abordan el problema de los gradientes desvanecientes y son especialmente útiles cuando se trabaja con secuencias largas. Las LSTMs pueden capturar relaciones a largo plazo en los datos y tienen una estructura de puertas que les permite controlar el flujo de información, lo que las hace ideales para problemas secuenciales.

2.6.3. Preprocesamiento de los datos

Los datos necesitan ser preprocesados antes de ser introducidos en la red. Esto puede implicar la normalización de los valores, la codificación de las categorías, el manejo de

valores faltantes, la transformación de los datos a la forma requerida por la red, etc. En este caso, necesitaremos normalizar los datos y posiblemente hacer ”padding”de las secuencias para que todas tengan la misma longitud.

Antes de alimentar tus datos a la red neuronal, es una buena práctica normalizarlos para que todos tengan la misma escala. Esto puede hacer que el modelo sea más fácil de entrenar. Y existen distintos tipos de normalización.

Una de las mas comunes es la Escala Min-Max, que reescala los datos para que todos los valores estén entre 0 y 1.
Se calcula utilizando la fórmula:

$$X_{norm} = \frac{(X - X_{min})}{(X_{max} - X_{min})} \tag{2.2}$$

La que se utilizó es la Normalización Z-score (estandarización) que reescala los datos para que tengan una media de 0 y una desviación estándar de 1.

Se calcula utilizando la fórmula:

$$X_{norm} = \frac{(X - mu)}{\sigma} \tag{2.3}$$

μ = media de lo datos
σ = desviación estándar

Como los datos son tomados de distintas rutas y ciclistas, es muy probable que las secuencias de entrada tengan longitudes variables, por lo que, se necesita hacer algo para que todas tengan la misma longitud antes de poder alimentarlas a la red neuronal. Una forma común de hacer esto es mediante el ”padding”, que implica agregar valores ”falsos”(como ceros) al final de las secuencias más cortas hasta que tengan la misma longitud que la secuencia más larga. Se puede llegar a pensar que el padding influye en como interpreta los datos la machine learning, pero en el caso de las capas LSTM no sucede así, ya que estas son bastante buenas en ïgnorar.[es]tos valores agregados.

2.6.4. Construir el modelo

Primero importaremos Numpy para manejar correctamente los datos. Además, utilizaremos la API de Keras de la librería Tensor Flow, donde importaremos primero el modelo que es de tipo secuencial (Sequential), luego importaremos los dos tipos de capas que utilizaremos, que serán la capa LSTM y la Dense. Y por ultimo, importaremos la función de Padding, para que todas las listas tengan la misma longitud. Importaremos también, la librería de matplotlib para visualizar el progreso de la ML.

Ahora que ya tenemos nuestras librerías podemos comenzar a trabajar con ellas. Pero antes, necesitamos entender la estructura de datos que tendremos y como recibe esa estructura nuestro modelo.

Basado en un estudio que habla sobre los ”sistemas multi agentes”se puede entender de mejor manera la primera parte del tratamiento del modelo:

1. Coordinación de capas LSTM: En una red neuronal LSTM, las capas LSTM pueden considerarse como agentes individuales que se comunican entre sí. Puedes explorar enfoques de coordinación entre estas capas para mejorar el rendimiento de tu red. Por ejemplo, puedes implementar una estructura de red neuronal en la que cada capa LSTM comparta información relevante con las capas vecinas a través de conexiones recurrentes, permitiendo una mejor transferencia de información a lo largo del tiempo.

2. Comunicación entre neuronas LSTM: Dentro de una capa LSTM, las neuronas también pueden considerarse como agentes que se comunican entre sí. Puedes utilizar técnicas de atención o mecanismos de comunicación entre neuronas LSTM para enfocar la atención en partes específicas de la secuencia de entrada. Por ejemplo, la capa de atención puede permitir que las neuronas LSTM se comuniquen entre sí para determinar qué partes de la secuencia de entrada son más relevantes para la tarea en cuestión.

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 tus 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 debes considerar la cantidad adicional de capas que deseas 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. La elección de las épocas también puede depender de la convergencia de la función de pérdida y del monitoreo del rendimiento en un conjunto de validación.

2.6.5. Evaluar y ajustar el modelo

Por ultimo, guardamos el modelo y creamos graficas que nos muestren como se comportan las perdidas en los dato de entrenamiento y los de validación. A base de prueba y error intentamos corregir los parámetros para intentar obtener cada vez datos mejor relacionados y mejores predicciones.

2.6.6. Hiperparametros

Una vez establecida la estructura del modelo, se llevaron a cabo diversas pruebas con el objetivo de encontrar un modelo que pudiera, al menos en cierta medida, asemejarse a la curva real de potencia. Inicialmente, se intentó desarrollar una red neuronal que, basándose en las características de distancia y altitud, pudiera predecir la velocidad. Tras algunos intentos, se logró este objetivo con éxito. Posteriormente, se exploró la predicción del

ritmo cardíaco mediante diversas combinaciones de características, obteniendo resultados moderadamente interesantes.

Con una comprensión más profunda del funcionamiento de la estructura, se procedió a los intentos por predecir la potencia. Para ello, se realizaron pruebas variando los parámetros de la red neuronal. Previo a esto, se llevó a cabo una breve investigación para identificar los rangos óptimos de estos parámetros, ya que se planeaba probarlos de manera aleatoria. A pesar de los esfuerzos iniciales, los resultados fueron decepcionantes, ya que las curvas no se ajustaban adecuadamente.

Los parámetros sujetos a variación fueron los siguientes:

- Cantidad de neuronas por capa
- Cantidad de capas de la red
- Tipo de capa
- Error de entrenamiento
- Ciclos

Estos constituyen los parámetros específicos de la red neuronal. Además, se consideraron las características que se emplearían para entrenar el modelo, que inicialmente eran:

- Velocidad
- Cadencia
- Tiempo
- Inclinación
- Altitud
- Frecuencia cardíaca
- Distancia

- Distancia

Como se señaló anteriormente, en estas primeras pruebas, se seleccionaron las características basándonos en la correlación que tenían con la característica objetivo (potencia).

Esto se hizo con el propósito de establecer un criterio que facilitara la organización de las pruebas.

Tras muchas pruebas se lograron obtener resultados muy prometedores y con predicciones relativamente cercanas a la curva real, esto se puede observar en al figura 2.6. Esto llevó a la conclusión de que la idea de realizar una predicción de la potencia mediante machine learning era factible.

Figura 2.6: Predicción de potencia de la primera versión de la Red Neuronal.

Una vez que obtuvimos mejores resultados, comenzamos a ajustar los hiperparámetros de la Machine Learning (ML). Investigando, llegamos a la conclusión de que la mayoría de estos parámetros no siguen reglas predefinidas y, en su mayoría, se ajustan mediante pruebas para identificar los modelos más efectivos. Por lo tanto, el siguiente paso fue automatizar el código, proporcionándole rangos en los que los hiperparámetros podrían variar. Luego, simplemente dejamos que el código calculara y revisamos los resultados. Se obtuvieron numerosos resultados interesantes, pero finalmente elegimos uno, en parte porque era de los mejores y, por otro lado, porque era una combinación de hiperparámetros

que se han utilizado en otras investigaciones con redes neuronales. Se consideró que estos parámetros podrían funcionar bien con nuestros datos.

- Optimizador: Adam
- Capas: 3 capas LTSM
- Neuronas: 128-64-32
- Ciclos: 120
- Función de activación: tanh

Una vez que se tuvo el modelo final con el que trabajaríamos, solo restaba elegir las características que utilizaríamos para entrenar el modelo y obtener el mejor resultado posible. Optamos por probar todas las posibles combinaciones de características, creando un código que automatizara este proceso al tomar todas las combinaciones posibles entre estas características. Además, dado el gran número de combinaciones, decidimos guardar solo aquellos modelos que obtuvieran un error MSE menor a 0.2, el cual se calcula aún estando normalizado. Este proceso tomó varios días de cálculos, pero una vez finalizado, obtuvimos 62 modelos con un error MSE menor a 0.2.

En este punto, solo quedaba identificar cuál era el mejor modelo para realizar la predicción. Medimos también, en base al error, cuál era el modelo que entregaba los mejores resultados en las tres actividades. Con esto, seleccionamos 5 modelos con los cuales finalmente realizamos todos los cálculos para la sección de resultados.

Capítulo 3

Resultados y conclusiones

3.1. Resultados modelo matematico

Uno de los primeros objetivos de este trabajo es lograr modelar la potencia a partir de otras características más accesibles para ciclistas de cualquier nivel, como la velocidad, cadencia, frecuencia cardíaca, entre otros. El primer paso consiste en entender qué tan efectivo es modelar esta característica a partir de los modelos matemático-físicos ya conocidos, en este caso, el modelo Cohen-2021 [8]. Para ello, probamos el modelo conocido con tres recorridos distintos de una misma persona, pero con una característica similar que se explicará más adelante.

Para evaluar este modelo, primero utilizamos un código que toma un recorrido completo y guarda uno por uno los datos calculados de potencia. Sin embargo, previamente necesitamos extraer del recorrido los datos de velocidad para calcular, mediante un cuaderno de Mathematica, la contribución de la derivada temporal de la energía cinética mediante el cálculo aproximado de la derivada numérica de estos datos.

Una vez obtenida esta derivada numérica y definidas todas las constantes pertenecientes al modelo (aproximadamente, ya que los datos del recorrido son virtuales), procedemos a calcular la potencia a partir del modelo matemático. Al tener tanto la potencia real como la potencia calculada, evaluamos esta última mediante el uso de un método estadístico de error conocido como MSE (Error Cuadrático Medio):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3.1}$$

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. Finalmente realizamos las graficas tanto de potencia como las del error.

Figura 3.1: Cálculo de potencia mediante formula (Actividad 1)

La característica común que se menciona antes es que, como se puede observar en los gráficos de potencia real versus potencia calculada, los tres recorridos muestran una constante subida. Esto se puede entender mejor si pensamos en un ciclista bajando por una pendiente. Como es lógico, el ciclista no necesariamente tendrá que pedalear para avanzar, ya que la gravedad hará su trabajo. Incluso si el ciclista intenta pedalear, habrá momentos donde la rueda por sí sola podría rotar más rápido de lo que el ciclista está pedaleando. Esto también puede ocurrir en un recorrido plano, pero en menor medida y de una manera más particular.

Primero, tenemos la situación más estándar, que es cuando los ciclistas dejan de pedalear durante un tiempo para descansar unos segundos. Por otro lado, también sucede que los ciclistas intentan ponerse a rueda durante la mayor cantidad de tiempo posible para minimizar el gasto energético, y esto tiene una implicancia difícil de notar a simple vista. Por ejemplo, al estar en una carrera solo o yendo el primero del grupo, se recibe el impacto

Figura 3.2: Error en cálculo de formula (Actividad 1)

Figura 3.3: Cálculo de potencia mediante formula (Actividad 2)

directo del aire. Por ende, cada vuelta del pedal es muy valiosa. En cambio, cuando se está a rueda, pasa algo muy particular. Al dar exactamente la misma vuelta de rueda que el ciclista que tenemos adelante, avanzamos un poco más. Esto sucede porque, al tener un porcentaje menor de choque directo con el viento, la rueda durante unos milisegundos o incluso menos gira un poco más que la rueda del ciclista que va primero. Este efecto es mucho más notorio para los ciclistas que van más atrás en la fila o pelotón. Por ende, los

Figura 3.4: Error en cálculo de formula (Actividad 1)

Figura 3.5: Cálculo de potencia mediante formula (Actividad 3)

ciclistas deben regular la velocidad a la que van para no chocar con el ciclista de adelante.

Todo esto, en resumen, hace que los cálculos de la potencia, tanto en plano como en bajada, tengan poca fiabilidad, ya que gran parte del tiempo el desplazamiento del ciclista depende de factores mucho menos controlables o medibles para este experimento. Por lo mismo, podemos entender que el caso donde los cálculos son más fiables es cuando el ci-

Figura 3.6: Error en cálculo de formula (Actividad 1)

clista está realizando una ruta ascendente, dado que la probabilidad de que ocurra alguno de los casos mencionados anteriormente disminuye drásticamente, y casi la totalidad del desplazamiento del ciclista depende de la potencia que este imprima en los pedales de la bicicleta.

Analizando puramente los gráficos, podemos notar a simple vista que, si bien la curva calculada a partir del modelo matemático tiene cierto parecido con la curva real y que incluso en ciertas partes se parece bastante, para el objetivo de predecir la potencia a partir de otras características, es bastante irregular. Esto se puede ver con mayor detalle en los gráficos que representan el error, donde un modelo ideal se vería como la línea punteada negra, es decir, sin errores. Sin embargo, en este caso, tenemos errores tanto por encima como por debajo, incluso superando los 200 [W] en algunos gráficos.

Al intentar recrear un modelo predictivo, tener errores en ciertas zonas que sean tan atípicos genera mucho ruido. Por ejemplo, la gráfica 2 3.4 de error sería un buen ejemplo de un modelo interesante, ya que no tiene errores tan grandes y más o menos mantiene una curva parecida a lo original durante toda la ruta. Pero aquí hay un detalle: para poder conseguir esa precisión con este modelo matemático, hay que ajustar un parámetro que es distinto para todas las curvas, y es la inclinación. Para lograr ese error en todos los cálculos, se tuvo que variar el rango en el que se medía la inclinación, por ejemplo, cada 180 metros. Es decir, durante 180 metros se toma el punto final e inicial y se calcula la

inclinación. Mediante prueba y error, se logró ajustar cada curva individualmente. Sin embargo, para poder escalar esto a bases de datos mucho más grandes, sería muy impreciso y caótico.

Teniendo en cuenta lo anterior, para poder lograr el objetivo de predecir la potencia a partir de otras características, nos vemos en la necesidad de utilizar otros métodos de predicción como el Machine Learning. El objetivo es acercarnos más a lograr una predicción más uniforme, reducir el error y que sea reproducible de manera sencilla.

3.2. Resultados Machine Learning

La única diferencia con los datos ingresados al modelo matemático es en cuanto a la inclinación. Para la red neuronal, se utilizó la característica de inclinación, pero calculada para cada intervalo de tiempo, es decir, se calcula entre cada dato tomado y no después de una distancia específica, como se explicó en los resultados del modelo matemático.

Como se comentó en el capítulo 2, se calcularon muchas combinaciones de modelos, guardando aquellas que tuvieran un error por debajo de 0.2 MSE. Se obtuvieron un total de 62 modelos con un error menor a 0.2 MSE, y de estos se seleccionaron los mejores 5 para probar sus predicciones en comparación con las mismas curvas que se ingresaron al modelo matemático. Como se puede observar en la tabla 3.10, se compararon los errores de los 5 mejores modelos, incluyendo los errores del modelo matemático, y se determinó que, en promedio, el que mejor rindió fue el modelo ”2,3,4,7”, cuyas características son ”Tiempo, Velocidad, Inclinación, Altitud”, respectivamente.

En las figuras 3.7, 3.8, 3.9 correspondientes a las actividades 1, 2 y 3 respectivamente, se utilizó el modelo ”2,3,4,7”, que obtuvo los mejores resultados para predecir la potencia, obteniendo resultados que se asemejan mucho a la curva de potencia real. También es importante destacar cómo el modelo es capaz de predecir desde las partes más suaves de la curva de potencia real hasta las elevaciones más altas, como se puede observar en la predicción de la actividad 3. En un principio, es capaz de ajustar la curva de manera muy eficiente, y aunque se aleja un poco de la curva real después de la mitad, sigue manteniendo una forma muy similar a esta.

Figura 3.7: Predicción de potencia mediante ML (Actividad 1)

Figura 3.8: Predicción de potencia mediante ML (Actividad 2)

3.3. 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

Figura 3.9: 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	1152,89	377,41	416,75	373,35	384,99	371,13
Actividad 2	5802,00	1080,17	1277,43	1083,46	1188,26	1171,26
Actividad 3	930,00	884,25	1055,63	915,10	913,10	967,24
Promedio	2628,30	780,61	916,60	790,64	828,78	836,54

Figura 3.10: Comparación de 5 modelos con caracteristicas distintas, incluyendo los errores del modelo matematico para las 3 actividades.

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 otro lado, aunque se podría pensar que a la ML se le están ingresando características simples como la velocidad y la inclinación, que también se ingresan en la fórmula, esta es capaz de encontrar patrones que quizás son imposibles de ver para cualquier persona normal mediante un simple análisis estadístico. Por lo tanto, se podría decir que incluso si estas características que no se consideran explícitamente pueden estar implícitas en los demás datos, la ML puede interpretarlas como tal.

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 potencia. 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.

Interpretar los resultados de la ML puede llegar incluso a ser un reto. Por un lado, se podría decir que los resultados obtenidos tienen sentido físico, dado que la velocidad es un parámetro que, como vimos antes, es muy importante para la predicción de la potencia. Luego, la inclinación es un parámetro también importante para determinar el rendimiento del ciclista. Sin embargo, por ejemplo, la altitud quizás no tiene una implicancia directa en la potencia y se puede interpretar de muchas formas qué relación puede tener. Pero la cuestión es que es como una caja negra, que puede realizar miles de procesos por segundo y encontrar cientos de patrones que serían imposibles de ver sin la ML, e intentar encontrarle significado a cada uno de los procesos que la machine learning realiza puede ser una tarea compleja e incluso sin sentido, considerando la posibilidad de que podría ser mucho mejor si tuviéramos más datos.

En conclusión, podemos decir que es posible para cualquier persona hacer un estimador de su perfil de potencia para cualquier ruta sin tener un potenciómetro, y mientras más grande sea la base de datos que se le entregue al modelo, mejor podría llegar a ser la predicción. De momento, estas predicciones son muy buenas, como comentamos antes, para rutas donde la mayor parte tiene una inclinación positiva que permita que la transmisión de energía del ciclista a los pedales sea constante. Esto abre las puertas a la posibilidad de explorar nuevos tipos de modelos de ML con distintos objetivos e incluso la posibilidad de recoger ciertas partes del modelo ya creado para poder evolucionarlo y responder muchas otras preguntas e incógnitas.

Bibliografía

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] BOLLT, S. A. Equivalent speed: A new metric for cycling performance. *Medicine & Science in Sports & Exercise 53*, 8S (2021), 29.

[6] BREFELD, U., DAVIS, J., VAN HAAREN, J., AND ZIMMERMANN, A. Machine learning and data mining for sports analytics. *Cham: Springer* (2018).

[7] 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.

[8] 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.

[9] 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.

[10] DYER, B. T. Cycle e-racing: Simulation or a new frontier in sports technology? *International Journal of Esports 1*, 1 (2020).

[11] 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.

[12] 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.

[13] 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.

[14] HALLIDAY, D., RESNICK, R., AND WALKER, J. *Fundamentals of physics*. John Wiley & Sons, 2013.

[15] HARRIS, C. R., MILLMAN, K. J., VAN DER WALT, S. J., GOMMERS, R., VIRTANEN, P., COURNAPEAU, 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.

[16] HOCHREITER, S., AND SCHMIDHUBER, J. Long short-term memory. *Neural computation 9*, 8 (1997), 1735–1780.

[17] 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.

[18] 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.

[19] HUNTER, J. D. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering 9*, 3 (2007), 90–95.

[20] 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.

[21] 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.

[22] KYLE, C. R. Reduction of wind resistance and power output of racing cyclists and runners travelling in groups. *Ergonomics 22*, 4 (1979), 387–397.

[23] 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.

[24] 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.

[25] 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.

[26] MENASPA, P. Analysis of road sprint cycling performance.

[27] NEE, P., AND HERTERICH, J. Modelling road cycling as motion on a curve. *Sports Engineering 25*, 1 (2022), 12.

[28] 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.

[29] NVIDIA, VINGELMANN, P., AND FITZEK, F. H. Cuda, release: 10.2.89, 2020.

[30] PANDAS DEVELOPMENT TEAM, T. pandas-dev/pandas: Pandas, Feb. 2020.

[31] 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.

[32] 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.

[33] RICHTER, C., O'REILLY, M., AND DELAHUNT, E. Machine learning in sports science: challenges and opportunities. *Sports Biomechanics* (2021), 1–7.

[34] 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.

[35] 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).

[36] TRENCHARD, H. The complex dynamics of bicycle pelotons. *arXiv preprint arXiv:1206.0816* (2012).

[37] 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.

[38] 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.

[39] 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.