

Análisis de simulaciones de Fórmula 1 con ayuda de inteligencia artificial

Autores :

Amanda Noris Hernández
Juan Miguel Pérez Martínez
Marcos Antonio Pérez Lorenzo

April 28, 2024



1 Introducción

1.1 ¿Qué es la Fórmula 1?

La Fórmula 1 es la principal competición de automovilismo internacional y el campeonato de motor más popular y prestigioso del mundo. Este deporte, que se ha convertido en un fenómeno global, se caracteriza por su intensidad, velocidad y la tecnología de vanguardia que emplean sus participantes. Con una historia rica y llena de emociones, la Fórmula 1 comenzó a tomar forma en Francia a partir de 1884, evolucionando rápidamente de simples carreras en caminos de un pueblo a otro a sofisticadas pruebas de resistencia. La Federación Internacional de Automovilismo (FIA) organizó el primer Campeonato Mundial de Pilotos en 1950, estableciendo las reglas que rigen este deporte hasta el día de hoy. Desde entonces, la Fórmula 1 ha sido testigo de la evolución de los vehículos, la tecnología y la competencia entre los pilotos, convirtiéndose en un referente en el mundo del automovilismo.

1.2 Simulaciones en los torneos reales

Los equipos de Fórmula 1 llevan a cabo simulaciones de carreras extensas y sofisticadas para optimizar su rendimiento y estrategia. Estas simulaciones, que a menudo implican el uso de tecnología de realidad virtual, permiten a los equipos probar nuevos diseños de coches y piezas, así como analizar los efectos de ajustar o cambiar todo lo posible sin romper los múltiples y complejos conjuntos de reglas y regulaciones de la F1. La instalación de simuladores de carreras de F1 es costosa, pero ahorra a los equipos los costos mucho mayores de realizar pruebas de vida precisas en instalaciones adicionales con autos y unidades reales. Un ejemplo destacado es el nuevo simulador de la Scuderia Ferrari, que se ha convertido en uno de los más avanzados del mundo. Este simulador, construido por Dynisma, ofrece una experiencia inmersiva y realista para los pilotos, permitiéndoles experimentar los movimientos del monoplaza sobre la pista de manera más precisa que nunca.

Sin embargo, los elementos físicos no son los únicos que determinan los resultados de una carrera, y teniendo en cuenta que en ellos son los que se centran la mayoría de las simulaciones avanzadas del mercado, el desarrollo de este trabajo se dirige más a los problemas de tomas de decisiones y funcionamiento interno de los equipos. La cooperación entre miembros de una

misma escudería juega un papel crucial en la optimización del desarrollo de la carrera y la toma de decisiones estratégicas en tiempo real. Esto incluye la predicción de la batalla en pista, la estrategia en boxes, y la ventana de parada, permitiendo a los equipos y a los fanáticos entender cómo las decisiones de los pilotos y sus equipos afectan a la carrera.

1.3 Descripción de una carrera

El desarrollo de una carrera de Fórmula 1 es un proceso complejo que comienza con la toma de decisiones estratégicas por parte del director del equipo. Estas decisiones pueden incluir la selección del piloto, la estrategia de carrera, y la configuración del coche. Una vez tomadas estas decisiones, el proceso se intensifica con la interacción entre el piloto y el equipo de ingenieros mecánicos de carrera. El ingeniero de carrera actúa como el enlace clave entre el piloto y el equipo, transmitiendo información sobre el estado del coche y las instrucciones del equipo al piloto, y viceversa. Esta comunicación es crucial para ajustar la estrategia de carrera en tiempo real, basándose en el rendimiento del coche y las condiciones de la pista. Además, el ingeniero mecánico de carrera debe mantener a los pilotos motivados y proporcionarles información sobre cómo mejorar su rendimiento. Después de la carrera, el ingeniero de carrera revisa los comentarios del piloto junto con los datos relevantes para identificar áreas de mejora para futuras carreras. Este ciclo de comunicación y adaptación es esencial para el éxito en la Fórmula 1, donde cada margen de tiempo y cada decisión pueden marcar la diferencia entre ganar y perder. Según la posición de la tabla final, el sistema actual otorga puntos a los diez primeros clasificados de un Gran Premio, con el ganador recibiendo 25 puntos, el segundo clasificado 18 puntos, el tercero 15 puntos, y así sucesivamente hasta el décimo clasificado, que recibe 1 punto.

1.4 Preguntas interesantes

Resulta atractivo hacerse varias preguntas sobre las carreras de Fórmula 1, como por ejemplo cuánto impacta la dificultad de la pista en el rendimiento de los pilotos, la comparación de la cantidad de pit stops según el clima, y las diferencias en el nivel de confianza en el equipo técnico entre pilotos ganadores y los que quedaron en último lugar. Primero, estas preguntas permiten explorar la complejidad y la variabilidad de las carreras de F1, donde factores como el clima, la dificultad de la pista, y la estrategia del

equipo pueden tener un impacto significativo en el resultado de la carrera y segundo, al analizar esas variables se puede obtener una comprensión más profunda de cómo los pilotos y los equipos manejan los desafíos únicos de cada carrera. Para responder a ellas se considero muy adecuado la realizacion de multiples simulaciones.

Además, otra motivación para el desarrollo de la simulación es hacer predicciones basadas en variables, a partir de las cuales se pueda desarrollar, por ejemplo, un modelo que pueda predecir el peor/mejor tiempo de una carrera, así como su tiempo promedio. Esto no solo es interesante desde un punto de vista académico, sino que también puede ser útil para los equipos de F1, ya que les permite planificar mejor sus estrategias y prepararse para las condiciones de carrera. Queda también el interés personal de analizar qué tan cercano a las carreras reales se pudiera modelar una simulacion.

2 Modelación del problema

2.1 Características del ambiente

1. **Accesible:** Los miembros de una escudería tienen acceso completo y actualizado a las condiciones del torneo, la carrera y la pista.
2. **No determinista:** Al estar compuesta la simulación por diversos agentes que compiten entre ellos, como los equipos y pilotos, realizar una misma acción, como por ejemplo acelerar para delantar a otro carro, puede resultar en diferentes resultados según el comportamiento del resto de los agentes implicados, pudiera ser que se logre rebasar al carro próximo, o que simplemente el carro próximo acelere también y queden en la misma posición.
3. **No episódico:** Los agentes pueden adaptar sus decisiones actuales a las consecuencias que las mismas tendrán en el futuro, como por ejemplo, un director decide que estrategia tomar en dependencia de si esta lo hará aumentar su posición en la tabla del torneo, y la elección de dicha estrategia es fundamental para el desempeño del equipo en la carrera.
4. **Dinámico:** Aunque un agente en específico no realice ninguna acción, el resto de los equipos compitiendo si lo harán, por lo que el medio se mantendrá en constante cambio.

5. **Continuo:** Los equipos y pilotos deben adaptarse constantemente a las condiciones cambiantes de la pista, el clima, y las estrategias de sus oponentes. Esta necesidad de adaptación y decisión rápida contribuye a la continuidad de la carrera, ya que cada decisión puede tener un impacto significativo en el resultado final.

2.2 Elementos del ambiente modelados

1. **Segmento de pista:** Los segmentos representan una sección de la pista, con atributos como longitud, grado de curvatura y grado de agarre. Estos atributos son fundamentales para determinar cómo un coche se comportará en ese segmento, afectando tanto la velocidad como la capacidad de manejo. La modelación de segmentos permite simular la variabilidad de las pistas de Fórmula 1, que pueden incluir rectas, curvas cerradas, y curvas abiertas, cada una con sus propias características.
2. **Pista:** Las pistas o circuitos representan la pista completa, compuesta por múltiples segmentos. La pista tiene un tipo de dificultad determinado por el promedio de los atributos de los segmentos, lo que permite simular la variabilidad de las pistas de Fórmula 1 en términos de dificultad. La ubicación de los pits también se determina de manera aleatoria, lo que añade otro nivel de complejidad a la simulación.
3. **Carrera:** La carrera representa una competencia en un circuito específico, e incluye el tiempo de inicio, la duración, el número de vueltas, de equipos, las condiciones climáticas y la pista. La carrera se actualiza en cada ciclo, considerando factores como las velocidades de los coches, el tiempo transcurrido, y las estrategias de los equipos. La simulación de la carrera también incluye la posibilidad de actualizar la tabla de posiciones basada en los resultados de la carrera, lo que refleja la competitividad y la estrategia en la Fórmula 1.
4. **Torneo:** El torneo para la carrera actúa como un entorno controlado para evaluar el rendimiento de los equipos y pilotos bajo diferentes condiciones. Al poseer una tabla de posiciones, se puede simular cómo los equipos y pilotos podrían terminar en diferentes posiciones en el campeonato mundial, basándose en su rendimiento en carreras individuales. Esto es crucial para entender cómo los factores como la estrate-

gia del equipo, la habilidad del piloto, y la suerte pueden influir en el resultado final del campeonato.

2.3 Otros elementos modelados

1. **Equipo:** La modelación de un equipo no solo se limita a la representación de los miembros del equipo (pilotos y mecánicos) y el coche y director, sino que también incluye la infraestructura y los sistemas que apoyan el rendimiento del equipo, como el sistema de mensajería. Este sistema de mensajería es esencial para la comunicación rápida y eficiente entre los miembros del equipo, lo cual es crucial en un deporte tan rápido y dinámico como la Fórmula 1.
2. **Estrategias de conducción:** Las estrategias de conducción se centran en cómo un piloto maneja el carro durante la carrera. Esto incluye aspectos como la aceleración, la cautela y la estrategia de frenado. La aceleración determina cuán rápido un piloto puede acelerar desde una parada, mientras que la cautela puede influir en cuán agresivamente el piloto toma las curvas. La estrategia de frenado puede variar desde frenar temprano para conservar la velocidad hasta frenar más tarde para ganar tiempo.
3. **Carros:** La modelación de los carros se centra en características como la velocidad máxima, la maniobrabilidad y el tipo de neumáticos, que son factores críticos en la competencia. La velocidad máxima determina cuán rápido un carro puede acelerar, mientras que la maniobrabilidad afecta cómo el carro puede manejar las curvas y cambios de dirección. Los tipos de neumáticos, como los neumáticos de carbono, pueden tener un impacto significativo en la tracción y el agarre, lo que a su vez afecta la capacidad del carro para mantenerse en la pista y competir en diferentes condiciones de pista.
4. **Neumáticos:** La modelación de neumáticos en la Fórmula 1 es crucial para maximizar el rendimiento del coche en diferentes condiciones de carrera, ya que los neumáticos pueden variar en rugosidad, resistencia, presión de aire, y rendimiento tanto en secas como mojadas. Los equipos deben elegir el tipo de neumático adecuado para cada situación, considerando factores como la dureza de los compuestos secos (C0 a C5)

para circuitos secos y los compuestos de pista mojada (como intermedios verdes y neumáticos de lluvia extrema azules) para condiciones húmedas. Esta elección estratégica permite a los equipos adaptarse a las condiciones cambiantes de la pista y a las estrategias de carrera, optimizando su velocidad y control en diferentes circunstancias.

3 Sistema multiagente

Los sistemas multiagentes son sistemas complejos que consisten en múltiples agentes que interactúan entre sí para alcanzar objetivos comunes o individuales. Estos sistemas son fundamentales en diversas áreas, incluyendo la simulación de la Fórmula 1, donde cada agente tiene roles y objetivos específicos que contribuyen al éxito del equipo. Los sistemas multiagentes pueden ser utilizados para resolver problemas complejos que son difíciles de manejar para un agente individual o un sistema monolítico. Estos sistemas se basan en la interacción entre múltiples agentes autónomos, cada uno con sus propios objetivos y capacidades, que trabajan juntos para alcanzar metas comunes o individuales. La combinación de cooperación, coordinación y negociación entre los agentes es crucial para el éxito de las interacciones y, por ende, del sistema en su conjunto.

En este caso se trabaja con un sistema compuesto por 3 tipos de agentes principales: pilotos, mecánicos y directores, donde cada uno cuenta con conocimientos, deseos y accionar diferentes y se encuentran presentes en cada equipo, donde un director, un mecánico y 2 pilotos interactúan y compiten de acuerdo a sus intereses.

3.1 Características de los agentes

1. **Reactivos:** Los agentes son reactivos debido a que operan basándose en sus percepciones actuales del entorno para tomar decisiones y actuar de manera oportuna. Cada agente (piloto, director, mecánico) percibe el estado actual de la carrera, incluyendo condiciones de la pista, clima, posición del equipo, y el rendimiento del coche. Esta percepción es crucial para tomar decisiones informadas, por ejemplo el piloto que este perdiendo la carrera puede llevar a cabo una estrategia de conducción más agresiva.

2. **Proactivos:** Son proactivos porque, en lugar de simplemente reaccionar a los cambios en el entorno, toman la iniciativa para anticipar y prepararse para posibles escenarios. Ellos tratan de controlar su destino al no limitarse a responder a eventos como ocurren. En cambio, toman medidas para influir en el resultado de la carrera, seleccionando estrategias y acciones que les permitan tener un mayor control sobre el desarrollo de la carrera, como por ejemplo intentar ganar al otro piloto del equipo aunque ese no haya sido el deseo inicial del director o sabotear la carrera al descubrir un engaño por parte del equipo.
3. **Sociables:** Los agentes reciben información unos de otros a partir del sistema de mensajería del equipo y a partir de esta realizan acciones que ayudan a lograr un objetivo común. Los agentes pilotos, director y mecánicos trabajan juntos como un equipo, donde cada uno tiene un rol específico pero todos contribuyen al objetivo común de ganar la carrera. La organización social se refleja en cómo los agentes se coordinan y se comunican entre sí para planificar y ejecutar estrategias. La coordinación entre los agentes es crucial para el éxito del equipo, así como la negociación, por ejemplo, un piloto puede decidir si confiar o hacer caso a la información que le brindan sus mecánicos o no.

3.2 Protocolo de comunicación

El protocolo de comunicación entre los agentes en la simulación se basa en un sistema de mensajería que utiliza estructuras con forma de mensajes para llevar a cabo el intercambio de información entre los agentes. Este sistema permite una comunicación eficiente y organizada entre los agentes pilotos, director y mecánicos, facilitando la coordinación y la toma de decisiones basadas en la información compartida. Los mensajes incluyen un identificador único, lo que asegura que cada mensaje sea fácilmente identificable. Esto es crucial para el seguimiento y la gestión de los mensajes dentro del sistema de mensajería. Además, cada mensaje tiene un remitente, un destinatario y un contenido, lo que permite una comunicación directa y específica entre los agentes. El sistema de mensajería mantiene una lista de todos los mensajes enviados. Este enfoque centralizado facilita la gestión y el seguimiento de todos los mensajes intercambiados entre los agentes. Para recibir mensajes, un agente especifica su identidad como destinatario ya que el sistema filtra la lista de mensajes para encontrar aquellos destinados al agente y los devuelve,

eliminándolos de la lista de mensajes del sistema para evitar duplicados y mantener la lista actualizada. Este proceso de envío y recepción de mensajes permite una comunicación fluida y eficiente entre los agentes, facilitando la coordinación y la toma de decisiones basadas en la información compartida.

3.3 Estructura de los agentes

La estructura BDI (Belief-Desire-Intention) es un modelo de agente inteligente que se utiliza en la simulación de sistemas complejos, como los de la Fórmula 1, debido a su capacidad para manejar múltiples tipos de agentes con diferentes roles y objetivos. Este modelo se basa en tres componentes principales:

- **Beliefs (Creencias):** Representan el conocimiento del agente sobre el mundo. En el contexto de la Fórmula 1, esto incluye información sobre el estado del coche, las condiciones de la pista, y las estrategias de conducción de los pilotos.
- **Desires (Deseos):** Son los objetivos que el agente quiere alcanzar. Para un piloto, podría ser ganar la carrera; para un mecánico, mantener el coche en buen estado; y para un director, asegurar el éxito del equipo.
- **Intentions (Intenciones):** Son los planes que el agente tiene para alcanzar sus deseos. Esto incluye las acciones específicas que el agente tomará, como elegir al mejor piloto para correr en una carrera o realizar un cambio de neumáticos en el carro.

La estructura BDI nos proporciona un marco robusto para modelar y gestionar la complejidad de los agentes en un entorno dinámico y competitivo por varias razones:

1. **Flexibilidad:** Permite modelar agentes con diferentes tipos de conocimientos, deseos y planes de acción. Esto es crucial en la Fórmula 1, donde cada agente tiene un rol específico pero interactúa con otros agentes con diferentes objetivos.
2. **Interacción entre agentes:** Facilita la interacción entre agentes con diferentes roles. Por ejemplo, un director puede comunicarse con un mecánico para informarle sobre el estado de la carrera, y un mecánico puede advertir al piloto sobre condiciones de la pista.

3. **Adaptabilidad:** Los agentes pueden adaptar sus creencias, deseos e intenciones en función de la información que reciben del entorno. Esto es esencial en la Fórmula 1, donde las condiciones cambian rápidamente y los agentes deben ser capaces de responder a estos cambios.
4. **Gestión de conflictos:** La estructura BDI permite manejar conflictos entre los deseos de diferentes agentes. Por ejemplo, un piloto puede querer acelerar, pero un mecánico puede tener la intención de realizar una parada de servicio para asegurar el coche. La estructura BDI permite que estos conflictos se resuelvan de manera efectiva.

3.4 Agente Director

La modelación de un agente directivo de equipo de Fórmula 1 en un sistema BDI se centra en la representación de las creencias, la generación de opciones basadas en estas creencias, la selección de intenciones y acciones basadas en estas opciones, y la actualización de las creencias a medida que el agente interactúa con el entorno. Este enfoque permite que el agente tome decisiones informadas y adaptativas basadas en su conocimiento actual y las condiciones del entorno, reflejando la lógica de toma de decisiones y la interacción con el entorno de una manera que imita la toma de decisiones humana. Las creencias actuales del agente incluyen información sobre el estado del equipo, la carrera, y el torneo, mientras que la función de revisión de creencias actualiza estas creencias basándose en nueva información. La función de generación de opciones evalúa las creencias actuales para determinar los deseos disponibles, y el conjunto de posibles intenciones representa las metas que el agente busca alcanzar. Los deseos del director, incluyen ganar la carrera, mantener una posición específica en el campeonato, o informar a los pilotos sobre la carrera y son evaluados utilizando la posición actual del equipo y la diferencia de puntos con los equipos adyacentes para decidir el deseo más adecuado. Se implementan diferentes estrategias de selección de pilotos basadas en los deseos del director. Por ejemplo, para el deseo de ganar la carrera y ascender en la tabla de posiciones del torneo, se seleccionan los 2 mejores pilotos para competir en la carrera, sin embargo, si la posición en la tabla no se ve afectada el director puede desear mantener su posición y enviar a competir a cualquier otro piloto para brindarle experiencia. Esta estrategia refleja la necesidad de equilibrar la experiencia y habilidades de los pilotos con las condiciones específicas de la carrera. El filtrado de intenciones selecciona la

intención más adecuada basándose en las creencias actuales, y la función de selección determina qué acción llevar a cabo para cumplir con la intención seleccionada. Este proceso de modelación refleja cómo el agente considera diferentes estrategias y objetivos basándose en su conocimiento actual y las condiciones del entorno, permitiendo que el agente tome decisiones informadas y adaptativas. Además, la selección de pilotos se realiza considerando factores como la experiencia, habilidades en la pista, habilidades en el clima, número de victorias, y el número de accidentes. Estos factores se evalúan mediante una función de fitness que calcula un valor de adecuación para cada piloto basado en estos criterios.

En resumen, la implementación del agente directivo de equipo de Fórmula 1 en un sistema BDI muestra cómo se seleccionan las estrategias y los pilotos a través de un proceso de toma de decisiones basado en la evaluación de las creencias actuales y la aplicación de heurísticas específicas para cada situación. Este enfoque permite al agente tomar decisiones informadas y adaptativas basadas en su conocimiento actual y las condiciones del entorno. La selección de estrategias y pilotos se basa en la evaluación de las creencias actuales del agente y en la aplicación de heurísticas específicas para cada situación. Por ejemplo, si el equipo está en la primera posición y la diferencia de puntos con el segundo equipo es suficiente, el deseo es mantener la posición uno.

3.5 Agente Mecánico

En la modelación del agente mecánico se sigue la misma estructura que en el anterior, incluyendo el conjunto de creencias actuales que captura la información sobre la carrera y el sistema de adaptación de neumáticos; la actualización periódica de las creencias basadas en nuevos datos del entorno; la función de generación de opciones, el conjunto de posibles intenciones, la función de filtrado, etc, por lo que es mas interesante hablar de 2 aspectos fundamentales de su comportamiento, el mecanismo de tomar la decision de que neumaticos poner al carro (que se vera en detalle mas adelante) y la veracidad, o no, de la informacion que brinda a los pilotos del equipo.

El sistema de adaptación de neumáticos juega un papel crucial en la toma de decisiones del agente mecánico. Este sistema utiliza información sobre las propiedades de los neumáticos y las condiciones del entorno, como la longitud del segmento de carrera, la curva, la rugosidad de la pista, y el clima, para determinar la adecuación de cada tipo de neumático. A través de un proceso

de simulación, el sistema evalúa la adecuación de cada opción de neumático y selecciona el que mejor se ajusta a las condiciones actuales, permitiendo al agente mecánico tomar decisiones basadas en la información más precisa y relevante. Este enfoque refleja la capacidad de la arquitectura BDI para integrar y procesar información compleja para facilitar la toma de decisiones efectivas en situaciones de alta incertidumbre y complejidad.

La inclusión de una variable aleatoria que decide si se debe informar de la verdad o no es un ejemplo de cómo se puede incorporar la incertidumbre y la toma de decisiones basada en probabilidades, que resulta muy interesante de modelar y que agrega una capa más de complejidad a la simulación. Esta variable aleatoria, junto con una probabilidad predefinida (en este caso, 0.95), determina si el mensaje enviado al piloto es verdadero o falso. La probabilidad de que el mensaje sea verdadero se basa en la generación de un número aleatorio y la comparación de este número con la probabilidad predefinida. Si el número aleatorio es menor o igual a la probabilidad predefinida, el mensaje es verdadero; de lo contrario, se genera un mensaje falso. Esta característica introduce un elemento de incertidumbre en la toma de decisiones del agente mecánico, lo que puede reflejar situaciones reales donde la información disponible puede ser incierta o incompleta y también afecta al agente piloto, que debe tener en cuenta el análisis de la información que se le provee.

3.6 Agente Piloto

En la modelación de un piloto de carreras se observa cómo cada componente del agente interactúa y se actualiza en función de su entorno y sus objetivos internos. Las creencias del piloto incluyen información sobre la confianza en el equipo, la posición actual en la carrera, y la confianza en el rendimiento del coche y del piloto en diferentes condiciones de pista y clima, así como sobre la posición de su compañero de equipo en la carrera. Estas creencias se actualizan periódicamente a lo largo del desarrollo de la carrera. La generación de deseos tiene en cuenta los conocimientos del piloto para una deliberación interna sobre que desea hacer, evalúa por ejemplo, la confianza en el equipo y decide si el piloto desea ganar la carrera o, en caso de baja confianza, prefiere sabotear la carrera, pues se siente traicionado. Las intenciones del piloto, se determinan en función de los deseos y la posición del piloto en el equipo, el piloto puede decidir si seguir manjeando como hasta el momento y mantener su posición, o si desea ganarle a su compañero de equipo por ambición, y en

ese caso realizaría la acción de acelerar a más no poder para tratar de ganar sea como sea.

El piloto decide confiar o no en su equipo en función de su posición actual en la carrera y las condiciones de la misma. Si el piloto descubre que una información dada por el mecánico no se corresponde con la realidad su confianza disminuye, en cambio, mientras vaya descubriendo que lo que le dicen es cierto su confianza aumenta y es mas propenso a seleccionar estrategias adecuadas para ganar la carrera y hacer lo mejor para el equipo. Este proceso refleja cómo las creencias del piloto sobre su equipo influyen en sus deseos y, por ende, en sus intenciones y acciones durante la carrera.

La selección de la mejor estrategia de conducción se realiza mediante un proceso de búsqueda genética . Este proceso comienza con una población inicial de estilos de conducción generados aleatoriamente. A través de un proceso de selección, cruce y mutación, la población se actualiza en cada generación para reflejar los estilos de conducción que mejor se ajustan a las condiciones específicas del segmento de la carrera en el que se encuentra el piloto. Este aspecto de abordará más a detalle en la próxima seccion de este informe.

4 Uso de la inteligencia artificial para aumentar el realismo de las simualciones

El uso de la inteligencia artificial (IA) en las simulaciones de Fórmula 1 ha revolucionado la forma en que los equipos y pilotos preparan y entrenan para las carreras. La IA no solo mejora la precisión y realismo de los simuladores, sino que también juega un papel crucial en la planificación estratégica de las carreras. En la modelacion planteada se hace uso de la inteligencia artificial sobre todo en los agentes inteligentes implementados y algunas de sus funcionalidades.

4.1 Conocimiento/razonamiento y lógica difusa en un sistema experto para seleccionar neumáticos

El sistema de adaptación de neumáticos representa un sistema experto diseñado para la adaptación de neumáticos en condiciones de carrera específicas, utilizando lógica difusa para evaluar la adecuación de diferentes tipos de neumáticos

basándose en una serie de condiciones de entrada. El conocimiento representado en el sistema experto para la adaptación de neumáticos en condiciones de carrera se basa en la experiencia y el conocimiento acumulado sobre cómo diferentes propiedades de los neumáticos y las condiciones de la carrera interactúan para determinar la adecuación de los neumáticos. Este conocimiento se encapsula en las reglas difusas definidas dentro del sistema, que son una forma de representar el conocimiento de manera que pueda ser utilizado por el sistema para tomar decisiones basadas en la entrada de datos. Las reglas difusas capturan la lógica subjetiva y la experiencia humana en la evaluación de la adecuación de los neumáticos. Por ejemplo, una regla puede establecer que un neumático con baja resistencia y baja rugosidad es adecuado para condiciones de pista lisa y soleado, reflejando la experiencia de que este tipo de neumático ofrece un buen equilibrio entre adherencia y durabilidad en estas condiciones. Otra regla puede indicar que un neumático con alta resistencia y alta rugosidad es más adecuado para pistas mojadas y curvas, reflejando la experiencia de que este tipo de neumático ofrece una mejor tracción y estabilidad en estas condiciones. Este conocimiento representado en el sistema experto es crucial para simular con precisión cómo diferentes tipos de neumáticos pueden adaptarse a una amplia gama de condiciones de carrera, permitiendo a los equipos de Fórmula 1, específicamente a los mecánicos, tomar decisiones informadas sobre qué tipo de neumáticos utilizar en cada segmento de la carrera. La capacidad del sistema para adaptarse a las condiciones cambiantes de la carrera y proporcionar recomendaciones basadas en el conocimiento humano acumulado es un aspecto clave de su eficacia.

La lógica difusa se utiliza en el sistema de adaptación de neumáticos para modelar la relación entre las propiedades de los neumáticos y las condiciones de la pista y del clima, permitiendo una evaluación más flexible y realista de la adecuación de los neumáticos para diferentes situaciones de carrera. Este enfoque difuso permite manejar la incertidumbre inherente a las condiciones de la pista y del clima, así como a las propiedades de los neumáticos, que pueden variar dentro de un rango continuo. El sistema de adaptación de neumáticos define conjuntos difusos para las propiedades de los neumáticos (como la rugosidad, resistencia, presión de aire, rendimiento seco y rendimiento mojado) y las condiciones de la pista y del clima (como la longitud, curva, rugosidad de la pista y clima). Estos conjuntos difusos representan la incertidumbre en las medidas de estas propiedades y condiciones, permitiendo una evaluación más precisa de la adecuación de los neumáticos para diferentes situaciones de carrera. Las reglas difusas se definen para

modelar la relación entre las propiedades de los neumáticos y las condiciones de la pista y del clima, y cómo estas relaciones afectan la adecuación de los neumáticos. Estas reglas difusas permiten que el sistema de adaptación de neumáticos tome decisiones basadas en la lógica difusa, lo que significa que puede manejar situaciones donde las condiciones de la pista y del clima no se ajustan exactamente a los conjuntos difusos definidos, pero aún así puede determinar la adecuación de los neumáticos. Por ejemplo, una regla difusa puede establecer que si el desempeño del neumático para climas húmedos es alto y el clima es lluvioso, entonces la adecuación del neumático es alta. Finalmente, el sistema de adaptación de neumáticos utiliza la lógica difusa para simular la adecuación de los neumáticos para diferentes segmentos de carrera y condiciones climáticas, permitiendo que los equipos de Fórmula 1 seleccionen el tipo de neumático más adecuado para cada situación. Este enfoque difuso mejora la precisión y el realismo de las simulaciones de Fórmula 1, permitiendo a los equipos y pilotos prepararse mejor para las carreras.

Este sistema es un ejemplo de cómo la inteligencia artificial y la lógica difusa pueden ser aplicadas para simular decisiones complejas que requieren considerar múltiples variables y condiciones.

4.2 Búsqueda de la mejor estrategia de conducción a través de una metaheurística implementada con algoritmo evolutivo

El problema de seleccionar la mejor estrategia de conducción para un segmento de la pista al que se enfrentan los pilotos en el contexto de una carrera de Fórmula 1, es no polinomial debido a la naturaleza compleja y multidimensional de las decisiones que se deben tomar. En este caso, la estrategia de conducción se determina en función de una serie de parámetros, como la aceleración, la cautela y la estrategia de frenado, que interactúan entre sí y con las condiciones específicas de cada segmento de la carrera. Estos parámetros pueden variar ampliamente y no siguen una relación lineal o polinomial simple, lo que hace que el problema sea intrínsecamente complejo. Además, el problema se complica aún más por la necesidad de adaptarse a las condiciones cambiantes de la carrera, como la rugosidad de la pista y la curva, que pueden cambiar rápidamente y requieren una evaluación continua y ajuste de la estrategia de conducción. Este tipo de problemas, donde se deben considerar múltiples variables y condiciones, y donde las soluciones

óptimas pueden depender de la interacción compleja entre estas variables, es recomendable atacarlos con metaheurísticas.

La metaheurística utilizada en este caso, la búsqueda genética, es una técnica adecuada para abordar este tipo de problemas. La búsqueda genética es un algoritmo de optimización inspirado en la teoría de la evolución natural, que pertenece a la clase más amplia de algoritmos evolutivos (EA). Se utiliza para generar soluciones de alta calidad para problemas de optimización y búsqueda mediante el uso de operadores biológicamente inspirados como la mutación, el cruce (crossover) y la selección. Además, este algoritmo se enfoca en la selección de estrategias óptimas para resolver problemas complejos, como la optimización de estilos de conducción en diferentes condiciones de pista.

El algoritmo funciona en una población de soluciones, donde cada solución es un individuo. Cada individuo tiene un cromosoma, que es un conjunto de parámetros (características) que definen al individuo. Cada cromosoma tiene un conjunto de genes, que pueden representarse de diversas maneras, como una cadena de 0s y 1s. En el contexto de la búsqueda genética aplicada a la optimización de estilos de conducción, estos genes pueden representar diferentes características de un estilo de conducción, como la aceleración, la cautela y la estrategia de frenado. Cada individuo tiene un valor de aptitud (fitness), que se calcula utilizando una función de aptitud. Esta función de aptitud representa la calidad de la solución, siendo mayor el valor de aptitud, mayor la calidad de la solución. La selección de los mejores individuos basada en su calidad se aplica para generar lo que se conoce como un "piscina de apareamiento" (mating pool), donde los individuos de mayor calidad tienen una mayor probabilidad de ser seleccionados en la piscina de apareamiento. En el caso de la optimización de estilos de conducción, la función de aptitud puede evaluar cómo un estilo de conducción se adapta a diferentes condiciones de pista, como la adherencia y la capacidad de maniobrabilidad. Los individuos en la piscina de apareamiento son llamados padres. Cada par de padres seleccionados de la piscina de apareamiento generará dos descendientes (hijos). Al aparearse entre individuos de alta calidad, se espera obtener descendientes de mejor calidad que sus padres. Esto ayuda a eliminar los individuos de baja calidad de la generación futura. Al seleccionar y aparearse entre individuos de alta calidad, hay mayores posibilidades de mantener las buenas propiedades de los individuos y descartar las malas. Finalmente, esto conduce a la solución óptima o aceptable deseada. Sin embargo, los descendientes generados utilizando los padres seleccionados solo tienen las

características de sus padres y no más, sin nuevas adiciones. Para superar este problema, se aplican cambios a cada descendiente para crear nuevos individuos. El conjunto de todos los nuevos individuos generados será la nueva población que reemplaza a la población anterior. Cada población creada se llama una generación. El proceso de reemplazar la población antigua por la nueva se llama reemplazo. En resumen, la búsqueda genética es un proceso iterativo que comienza con una población inicial de soluciones. A través de la selección, el cruce y la mutación, el algoritmo evoluciona la población para mejorar la calidad de las soluciones, con el objetivo de encontrar la solución óptima o aceptable para el problema dado. Este proceso se repite durante un número determinado de generaciones, o hasta que se cumpla algún criterio de parada.

En la resolución del problema de las estrategias, se utiliza un enfoque genético para optimizar estilos de conducción en diferentes segmentos de pista, considerando factores como la adherencia y la capacidad de maniobrabilidad. La selección de estrategias se realiza mediante la evaluación de la aptitud de cada estilo de conducción en función de estos factores, y se aplican operadores de cruce y mutación para generar nuevas estrategias que pueden mejorar la aptitud general de la población. Este proceso se repite hasta que se encuentra un estilo de conducción óptimo o se alcanza un número máximo de generaciones.

4.3 Procesamiento de lenguaje natural para la narración de la carrera con función de comentarista

LM Studio es una aplicación de escritorio que permite ejecutar modelos de lenguaje de gran tamaño (LLMs) localmente. Esta herramienta es especialmente relevante en el contexto de la inteligencia artificial y el aprendizaje automático, ya que ofrece una alternativa a los modelos de lenguaje basados en la nube, como ChatGPT, proporcionando mayor control y privacidad. Luego de un análisis de las opciones brindadas por LM Studio, se eligió trabajar con el modelo TheBloke's phi-2-GGUF debido principalmente a su descripción de 'small and fast', ya que se corresponde con la intención de que el modelo corra lo más parejo con el desarrollo de la simulación posible.

TheBloke's 'phi-2-GGUF' es un modelo de lenguaje de gran tamaño (LLM) disponible en el repositorio de Hugging Face. Este modelo está diseñado para la generación de texto y se basa en la arquitectura de trans-

formadores, lo que lo hace adecuado para una variedad de tareas de procesamiento de lenguaje natural (NLP). Es una adaptación del modelo Phi 2 de Microsoft, pero en formato GGUF. El formato GGUF es una representación específica de los modelos de lenguaje que facilita su uso y distribución. Este modelo está disponible en inglés y está orientado hacia la generación de texto, lo que lo hace útil para aplicaciones de chat, asistentes virtuales y para el caso específico de la simulación de Fórmula 1, para comentaristas deportivos. Cuenta con 2.7 billones de parámetros, un número extremadamente alto, lo que indica que el modelo tiene una capacidad de aprendizaje y representación muy amplia.

Para mejorar la respuesta del modelo de lenguaje (LM) se pueden modificar varios aspectos de su función `client.chat.completions.create()`. A continuación, se detallan algunos de los parámetros clave que se ajustaron para influir en la generación de texto:

temperature: Este parámetro controla la aleatoriedad de las respuestas del modelo. Un valor más bajo (por ejemplo, 0.5) hace que las respuestas sean más deterministas y enfocadas, mientras que un valor más alto (por ejemplo, 1.0) permite una mayor variabilidad y creatividad en las respuestas. Se estableció `temperature=0.7`, lo que ofrece un equilibrio entre creatividad y coherencia.

max tokens: Este parámetro limita la longitud de la respuesta generada por el modelo. Si se desea que las respuestas sean más concisas o detalladas, puedes ajustar este valor. Por ejemplo, `max tokens=100` limitará la respuesta a 100 tokens de texto.

n: Este parámetro especifica el número de respuestas generadas por el modelo. Se trabajó con `n=5` para obtener varias interpretaciones o ideas del modelo y escoger la más adecuada.

stop: Permite especificar una lista de tokens o cadenas de texto que, cuando aparezcan en la respuesta generada, harán que el modelo detenga la generación. Esto es útil para controlar el flujo de la conversación o para asegurarte de que el modelo no genere contenido no deseado.

top p: Este parámetro controla la distribución de probabilidad de los tokens generados. Un valor más bajo (por ejemplo, 0.1) hace que el modelo sea más conservador en sus elecciones, mientras que un valor más alto (por ejemplo, 0.9) permite una mayor diversidad en las respuestas. Ajustar este parámetro puede ayudar a controlar la calidad y relevancia de las respuestas generadas.

En cada paso de la carrera se realiza un log que permite guardar infor-

mación del momento, como por ejemplo la tabla de posiciones y las velocidades , y luego esta información se le pasa al LLM que actúa como comentarador devolviéndonos narraciones como la siguiente :

"Welcome to the most expected Formula 1 race of the year 2022 , the German Grand Prix!"

"The green flag has dropped and the race is underway! Walls is currently leading the field, but we can't count out Malon just yet - he's put in a phenomenal performance so far today. We're seeing some intense battles for position between the two Reddick drivers, with Dorothy Bayardo also looking to make his presence felt."

"The drivers are all ready for action, with Matthew Walls from Team Reddick in the lead, followed closely by Susan Welch from the same team. Next up is Dorothy Bayardo from Team Pontious, driving alongside Francis Dixon. Then we have Ivan Healy and John Burton from Team Moss, both aiming to take the top spot. But watch out! Preston Malone from Team Perry has a strong team of 10 drivers behind him."

"Our first driver, Matthew Walls, is off to an impressive start with a speed of 261.10615384615386. Next up is Susan Welch with a velocity of 273.5954871794872 and then Dorothy Bayardo at 278.6788817204301. Following her is Francis Dixon with 270.60021505376346, Ivan Healy with 307.337037037037, John Burton with 256.29925925925926, Stephen Aracena with 263.2514074074074 and finally Edward Kut with 210.17274074074072."

"Now let's talk about our veteran driver, Dorothy Bayardo, who's in the third spot with a velocity of 278.6788817204301. She's been on this track for years and knows her way around. However, we can't ignore our youngest driver, Francis Dixon, who's only 19 years old and has an impressive speed of 200.60021505376346. He's definitely making waves in the Formula 1 world and we can see that today."

"The first half of the race has been action-packed, and we've already seen some exciting overtakes and close battles for positions on the grid. With the second half of the race now underway, it's anyone's game - but I have a feeling that Walls is going to be able to maintain his lead."

"As the laps start to pile up, we're seeing more and more pressure being put on all drivers. We've already seen a few cars make pit stops for fresh tyres, and it'll be interesting to see how this affects their performance in the second half of the race."

"With just a few laps remaining, we're still in the midst of a thrilling battle between Walls and Welch. Both drivers are fighting tooth-and-nail for

position, with Sebastian Vettel also looking to make his presence felt.”

”Walls’s already leading by a comfortable margin over his closest rivals, but he can’t afford to relax now. There are still two laps left, and every second counts in Formula 1. The track conditions are wet and slippery, which adds an extra challenge for the drivers, especially when they have to make sharp turns at high speeds. Walls knows this, and he’s using his brilliant driving style to navigate the course with ease. He’s passing cars left and right, creating a gap that’s getting bigger by the second. He’s also showing his excellent overtaking skills, leaving his competitors in the dust behind him. He’s not letting anyone catch up to him, not even his teammate Susan Welch, who is trying his best to defend his position.”

”As we approach the final lap, tensions are running high. We’re seeing some incredible action on the track as cars jostle for position - it’s a real testament to the skill of these drivers!”

”And now, in the final moments of the race... Walls takes the checkered flag, securing his second victory of the season! What an incredible performance by all the drivers today. Congratulations, Matthew! This was truly a fantastic race to watch.”

”Ladies and gentlemen, the race is about to come to an end with a final lap of 755.7560205459595 seconds! The positions are tight with pilot Matthew Walls from Team Reddick in first place, followed by pilot Dorothy Bayardo from Team Pontious in second. Third place goes to pilot Susan Welch from Team Reddick and fourth place is Francis Dixon from Team Pontious. Fifth through ninth place are John Burton from Team Moss, Edward Kut from Team Netzer, Preston Malone from Team Perry, Juan Fisher from Team Perry, Stephen Aracena from Team Netzer, and Ivan Healy from Team Moss!”

5 Experimentos, análisis y resultados

Con el objetivo de recopilar información de interés de la simulación y realizar posteriormente análisis estadísticos con los datos se llevaron a cabo 200 simulaciones aleatorias, cuyos resultados se guardaron en el archivo simulacion.csv

Las variables extraídas y a partir de las cuales se trabajara a continuación son: De la carrera:

1. tiempo promedio : tiempo promedio que demoraron los pilotos en completar la carrera
2. clima : condición climática del día de la carrera
3. no vueltas : número de vueltas de la carrera
4. largo pista : largo en metros de la pista
5. dificultad pista : fácil, media o difícil según su curvatura y resbaladura principalmente
6. tiempo promedio vuelta : tiempo promedio en que se demora un piloto en hacer una vuelta a la pista

Del piloto peor posicionado:

1. dir lento deseo : estrategia del director
2. car lento velocidad max : velocidad máxima del carro
3. lento tiempo : tiempo de carrera
4. lento experiencia : años de experiencia
5. lento victorias : cantidad de victorias
6. lento paradas : cantidad de pit stops realizados
7. lento confianza : nivel de confianza en el equipo técnico
8. lento velocidad media : velocidad media con la que corrió

Del piloto mejor posicionado:

1. dir rapido deseo : estrategia del director
2. car rapido velocidad max : velocidad máxima del carro
3. rapido tiempo : tiempo de carrera
4. rapido experiencia : años de experiencia
5. rapido victorias : cantidad de victorias

6. rapido paradas : cantidad de pit stops realizados
7. rapido confianza : nivel de confianza en el equipo técnico
8. rapido velocidad media : velocidad media con la que corrió

5.1 Análisis inicial e hipótesis sobre los datos

Se comenzó el análisis describiendo el conjunto de datos con el objetivo de ver en que rango de valores trabajaremos. Estos fueron los resultados:

	tiempo_promedio	clima	no_vueltas	largo_pista \
count	199.000000	199.000000	199.000000	199.000000
mean	112.886717	2.527638	2.949749	16410.552764
std	53.334607	1.140614	1.358656	9103.792039
min	17.146301	1.000000	1.000000	1000.000000
25%	71.288097	2.000000	2.000000	9200.000000
50%	109.247136	2.000000	3.000000	15900.000000
75%	152.746201	4.000000	4.000000	23450.000000
max	267.718172	4.000000	5.000000	35900.000000

	dificultad_pista	dir_lento_deseo	car_lento_velocidad_max \
count	199.000000	199.000000	199.000000
mean	2.005025	1.301508	401.638191
std	0.235648	0.521795	59.796792
min	1.000000	1.000000	300.000000
25%	2.000000	1.000000	352.000000
50%	2.000000	1.000000	394.000000
75%	2.000000	2.000000	455.000000
max	3.000000	3.000000	500.000000

	rapido_tiempo	rapido_experiencia	rapido_victorias	rapido_paradas \
count	199.000000	199.000000	199.000000	199.000000
mean	113.423717	6.206030	4.969849	2.512563
std	53.728398	2.780299	3.146121	1.696280
min	17.146301	1.000000	0.000000	0.000000
25%	71.288097	4.000000	2.000000	1.000000
50%	108.106589	7.000000	5.000000	3.000000
75%	153.136301	9.000000	8.000000	4.000000
max	267.718172	10.000000	10.000000	5.000000

	rapido_confianza	rapido_velocidad_media
count	199.000000	199.000000
mean	4.085427	421.094226
std	2.561908	333.478323
min	0.000000	1.391284
25%	2.000000	164.921184
50%	4.000000	350.316560
75%	6.000000	586.365416
max	8.000000	1490.894376

	lento_velocidad_media	dir_rapido_deseo	car_rapido_velocidad_max	\
count	199.000000	199.000000	199.000000	
mean	421.094226	1.356784	407.140704	
std	333.478323	0.567060	58.866398	
min	1.391284	1.000000	301.000000	
25%	164.921184	1.000000	357.500000	
50%	350.316560	1.000000	413.000000	
75%	586.365416	2.000000	460.000000	
max	1490.894376	3.000000	499.000000	

	rapido_tiempo	rapido_experiencia	rapido_victorias	rapido_paradas	\
count	199.000000	199.000000	199.000000	199.000000	
mean	113.423717	6.206030	4.969849	2.512563	
std	53.728398	2.780299	3.146121	1.696280	
min	17.146301	1.000000	0.000000	0.000000	
25%	71.288097	4.000000	2.000000	1.000000	
50%	108.106589	7.000000	5.000000	3.000000	
75%	153.136301	9.000000	8.000000	4.000000	
max	267.718172	10.000000	10.000000	5.000000	

	lento_tiempo	lento_experiencia	lento_victorias	...	lento_confianza	\
count	199.000000	199.000000	199.000000	...	199.000000	
mean	113.423717	6.206030	4.969849	...	3.477387	
std	53.728398	2.780299	3.146121	...	2.641594	
min	17.146301	1.000000	0.000000	...	0.000000	
25%	71.288097	4.000000	2.000000	...	1.000000	
50%	108.106589	7.000000	5.000000	...	3.000000	
75%	153.136301	9.000000	8.000000	...	6.000000	
max	267.718172	10.000000	10.000000	...	8.000000	

	lento_velocidad_media	dir_rapido_deseo	car_rapido_velocidad_max	\
count	199.000000	199.000000	199.000000	
mean	421.094226	1.356784	407.140704	
std	333.478323	0.567060	58.866398	
min	1.391284	1.000000	301.000000	
25%	164.921184	1.000000	357.500000	
50%	350.316560	1.000000	413.000000	
75%	586.365416	2.000000	460.000000	
max	1490.894376	3.000000	499.000000	

5.1.1 Impacto de la dificultad de la pista en el rendimiento de los pilotos

Se analizará si la dificultad de la pista tiene un impacto significativo en el rendimiento de los pilotos, tanto en términos de tiempo promedio como en otras variables de rendimiento. La hipótesis sería que la dificultad de la pista afecta de manera diferente al rendimiento de los pilotos ganadores y los que quedaron en último lugar.

Statistics=0.0, p=1.0 para cada una de las dificultades.

El resultado de tu prueba de Kruskal-Wallis sugiere que no hay una diferencia significativa en los tiempos de carrera entre los pilotos que quedaron en último lugar y los que ganaron, considerando la dificultad de la pista. Esto significa que, basándonos en los datos y el nivel de significancia utilizado (generalmente 0.05), no podemos rechazar la hipótesis nula y concluir que la dificultad de la pista tiene un impacto significativo en el rendimiento de los pilotos en términos de tiempo de carrera.

5.1.2 Comparación de cantidad de pitstops según el clima

Analizar la cantidad de pitstops según el clima es de gran interés en el contexto de las carreras de Fórmula 1, ya que este puede tener un impacto significativo en el rendimiento de los pilotos y en la eficiencia de los equipos durante las paradas en boxes. La cantidad de pitstops puede revelar patrones y tendencias que pueden ayudar a los equipos a prepararse mejor para diferentes condiciones climáticas. Por ejemplo, si los equipos ganadores tienden a hacer más pitstops en días nublados o lluviosos que los perdedores, esto podría indicar que las paradas sistemáticas en estos climas para mejorar las condiciones técnicas del carro son importantes para el desenlace de la carrera.

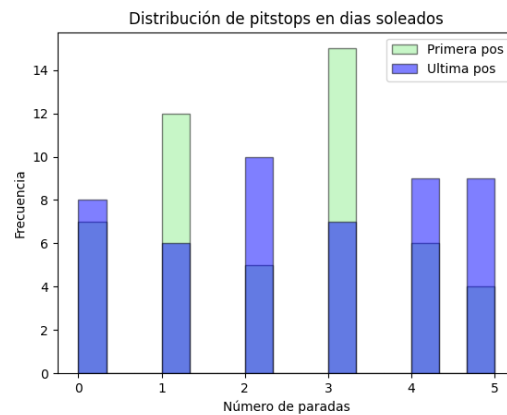


Figure 1: Soleado: Primera posición: 2.2653061224489797 Última posición: 2.6122448979591835

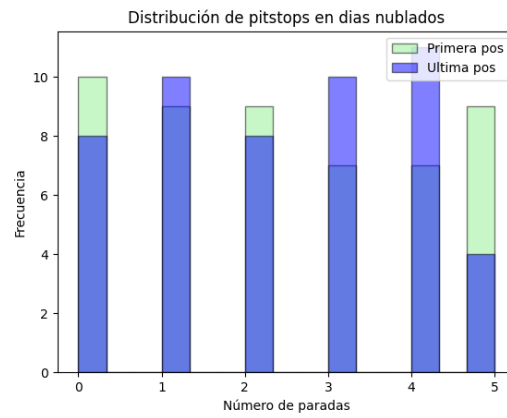


Figure 2: Nublado: Primera posición: 2.372549019607843 Última posición: 2.3529411764705883

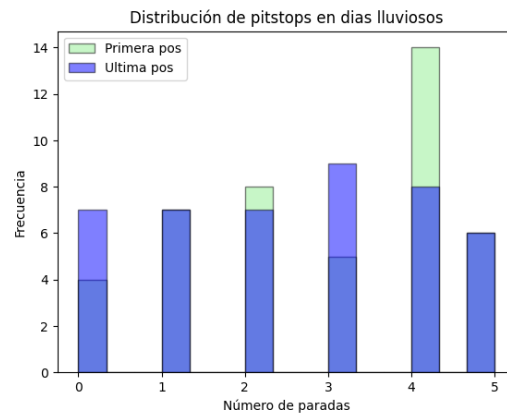


Figure 3: Lluvioso: Primera posición: 2.81818181818183 Última posición: 2.5

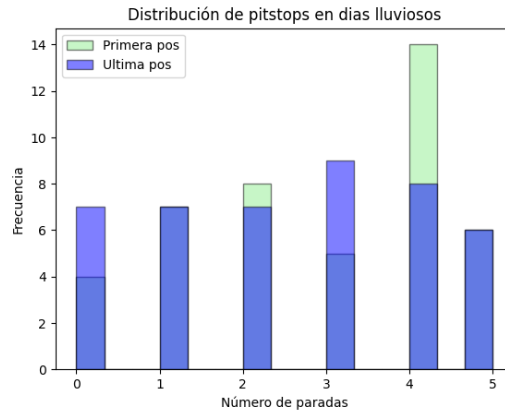


Figure 4: Tormentoso: Primera posición: 2.81818181818183 Última posición: 2.5

Como se especulaba al inicio, los gráficos y medias nos muestran que los equipos ganadores sí realizan más paradas de pit cuando las condiciones del clima no son favorables. Esto nos lleva a pensar que el intercambio entre mecánicos y pilotos es bastante importante para obtener una buena posición.

5.1.3 Diferencias en el nivel de confianza entre pilotos ganadores y los que quedaron en último lugar

Compararemos las medias de las variables que miden la confianza entre los pilotos que quedaron en último lugar y los que ganaron la carrera. La hipótesis nula es que no hay diferencias significativas entre las medianas de los grupos.

Se utilizara la prueba de Kruskal-Wallis ya que es una prueba no paramétrica que no asume que los datos sigan una distribución normal.

Statistics=5.446537840045688, $p=0.019607146344147716$

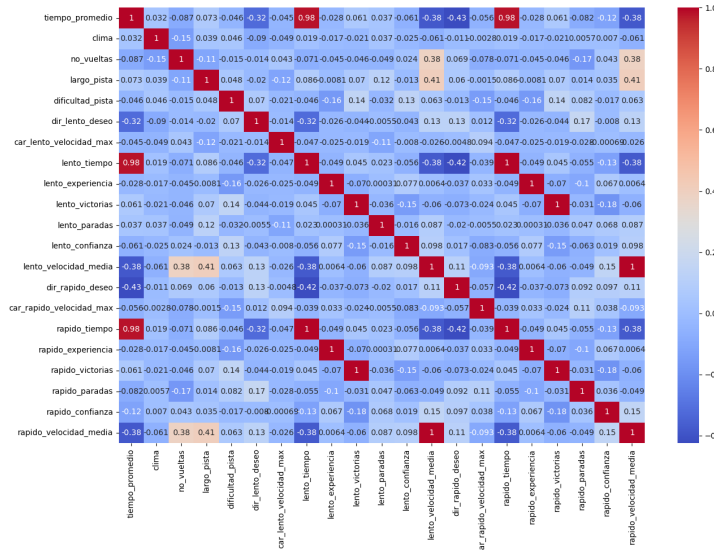
Los resultados de la prueba de Kruskal-Wallis, con ‘Statistics=5.446537840045688’ y ‘ $p=0.019607146344147716$ ’, indican que hay una diferencia estadísticamente significativa entre las medias de los grupos que se estan comparando. El valor p es 0.0196, que es menor que el nivel de significancia comúnmente utilizado de 0.05. Esto significa que se puede rechazar la hipótesis nula de que no hay diferencias significativas entre las medias de los grupos.

En el problema este es un resultado que se espera pues la confianza en el equipo de mecánicos es un factor clave para el desarrollo profesional de los pilotos. Los pilotos deben confiar en que su equipo de mecánicos les

proporcionará retroalimentación constructiva y oportuna, lo que es esencial para mejorar su rendimiento y aprender de sus errores. Si un piloto descubre que su equipo no le está proporcionando información actualizada/correcta es más probable que se enoje y planea sabotear su actuación en la carrera, haciendo que quede en última posición.

5.2 Correlación entre las variables

La matriz de correlación puede ser útil para identificar relaciones entre variables que podrían ser relevantes para la toma de decisiones, como el diseño de pistas de carreras, la planificación de horarios de conducción, o el desarrollo de estrategias de marketing dirigidas a diferentes segmentos de conductores.



En general la matriz obtenida no posee fuertes relaciones entre las variables, pero de igual manera se pueden apreciar algunas.

Variables fuertemente correlacionadas: Por ejemplo, tiempo promedio y lento tiempo tienen un coeficiente de correlación de 0.978858, lo que indica una fuerte correlación positiva. Esto es evidente ya que el tiempo más lento de la carrera también es utilizado para calcular el promedio de la misma. Lo mismo ocurre con el tiempo más rapido de la carrera.

Variables con correlación moderada: Por ejemplo, largo pista y rapido tiempo tienen un coeficiente de correlación de 0.086430, lo que indica una

correlación moderada. Esto sugiere que existe una relación entre el largo de la pista y el tiempo mas rápido en completar la carrera, pero no tan fuerte como en otros pares de variables. Entre tiempo promedio y largo pista: 0.073260. Aunque este coeficiente es positivo, su magnitud es baja. Esto indica que a medida que el largo de la pista tiende a aumentar, el tiempo promedio también.

5.3 Predicciones a partir de los datos

5.3.1 Predicción del peor tiempo

Predecir el tiempo de los perdedores en una carrera de Fórmula 1 es interesante porque permite a los equipos y a los analistas de datos anticipar los resultados de las carreras con mayor precisión. Esto es crucial para la toma de decisiones estratégicas, como la planificación de paradas en boxes y ajustes de estrategia durante la carrera, lo que puede marcar la diferencia entre ganar y perder.

Este problema se afrontará con una regresión lienal clásica.

OLS Regression Results						
Dep. Variable:	lento_tiempo		R-squared:	0.305		
Model:	OLS	Adj. R-squared:	0.292			
Method:	Least Squares		F-statistic:	22.69		
Date:	Sun, 28 Apr 2024	Prob (F-statistic):	3.11e-12			
Time:	00:51:24	Log-Likelihood:	-833.35			
No. Observations:	159	AIC:	1675.			
Df Residuals:	155	BIC:	1687.			
Df Model:	3					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	152.9327	11.805	12.955	0.000	129.613	176.252
largo_pista	0.0017	0.000	3.830	0.000	0.001	0.003
dir_lento_deseo	-29.3774	7.208	-4.076	0.000	-43.616	-15.139
lento_velocidad_media	-0.0740	0.012	-6.038	0.000	-0.098	-0.050
Omnibus:	0.062	Durbin-Watson:	2.167			
Prob(Omnibus):	0.969	Jarque-Bera (JB):	0.042			
Skew:	0.035	Prob(JB):	0.979			
Kurtosis:	2.963	Cond. No.	6.60e+04			

Figure 5: Summary del modelo

Error Cuadrático Medio: 1993.6648528951052 Coeficiente de Determinación R2: 0.11614268270564365

El modelo no nos brinda los mejores resultados y nos indica que pueden existir problemas numéricos en nuestros datos, esto es sin duda debido a la forma de generla los mismos a través de simulaciones aleatorias. Analicemos los supuestos:

Media de los residuos: $1.2101570619109127 \times 10^{-13}$ La media de los residuos es cero. Suma de los residuos: $1.9241497284383513 \times 10^{-11}$ La suma de los residuos es cero.

$D = 0.5131995498956742$, $p\text{-value} = 2.1298895774045553 \times 10^{-39}$ Los residuos no parecen estar normalmente distribuidos.

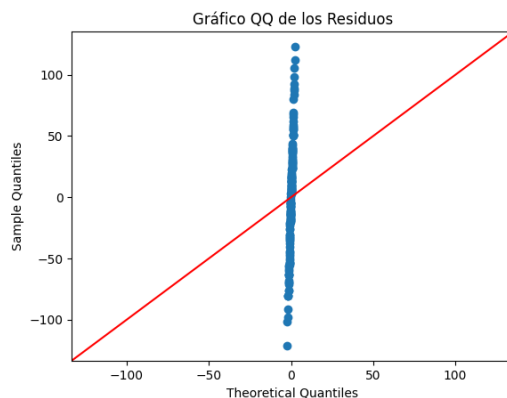


Figure 6: QQ plot de los residuos del modelo

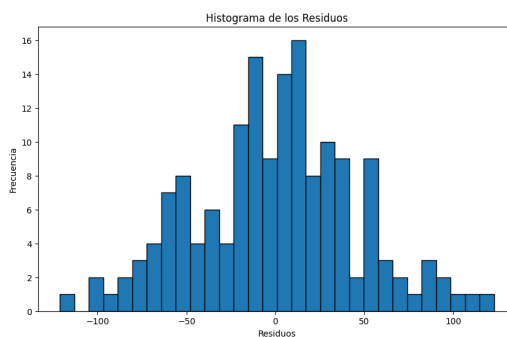
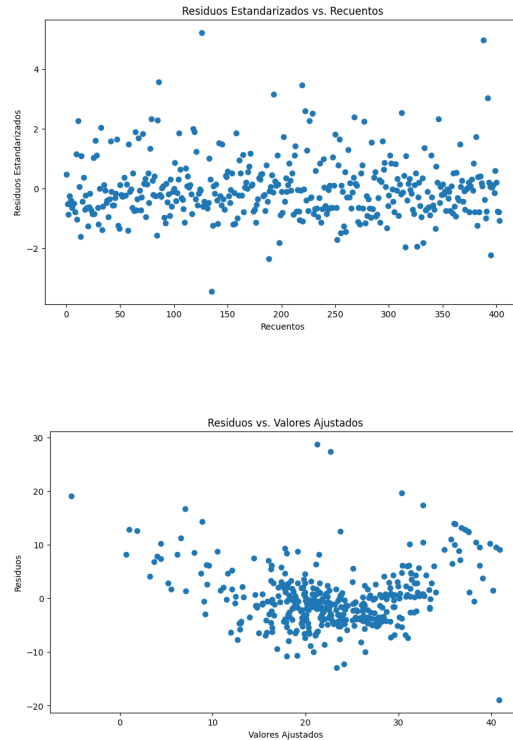


Figure 7: Histograma de los residuos del modelo

Los errores no poseen una distribución normal, por lo que no se cumplen los supuestos del modelo, pero se debe notar la linealidad que se observa en el gráfico QQ, la cual indica que es muy probable que lo que le falte al modelo para ser correcto es alguna variable que no se tuvo en cuenta a la hora de recopilar los datos. Esto se correspondería al alto valor del coeficiente de la constante que nos devuelve el modelo en su summary.

Estadística de Durbin-Watson: 2.167335454117047 No hay autocorrelación en los residuos.

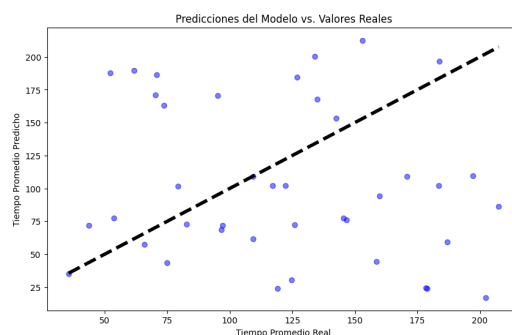
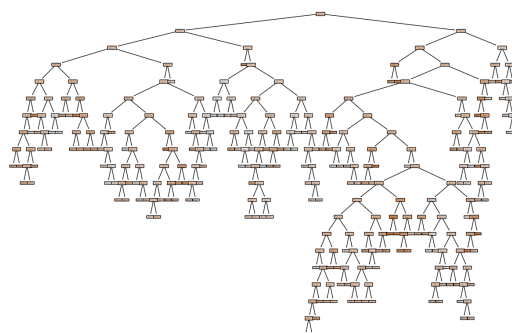


Test de Breusch-Pagan [(‘Lagrange multiplier statistic’, 0.976359557516906), (‘p-value’, 0.61374252548199), (‘f-value’, 0.4857285569333162), (‘f p-value’, 0.6156102998441584)] Los residuos son homocedasticos

5.3.2 Predicción del tiempo promedio de una carrera

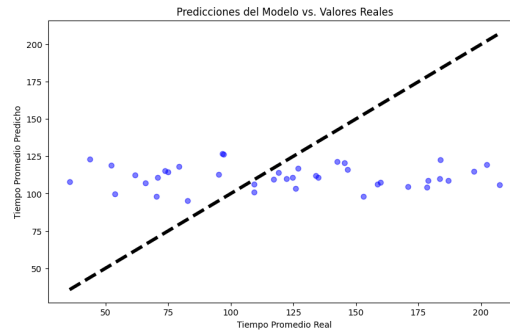
El interés en predecir el tiempo medio de una carrera a partir de variables como el clima, el número de vueltas, el largo de la pista y la dificultad de la pista radica en la capacidad de anticipar y optimizar el rendimiento deportivo. La predicción precisa del tiempo medio puede ser crucial para la planificación de entrenamientos, la asignación de recursos y la toma de decisiones estratégicas en competencias deportivas. Al entender cómo estas variables influyen en el tiempo de carrera, los directores y pilotos pueden ajustar sus estrategias para mejorar el rendimiento, ya sea reduciendo el tiempo de carrera o preparándose adecuadamente para condiciones específicas.

El uso de un árbol de decisión en la predicción del tiempo medio de una carrera a partir de variables como el clima, el número de vueltas, el largo de la pista y la dificultad de la pista es particularmente útil debido a su capacidad para manejar relaciones no lineales entre las variables predictivas y la variable objetivo. Los árboles de decisión son modelos de aprendizaje automático que pueden capturar complejidades y patrones no lineales en los datos, lo que los hace ideales para problemas donde las relaciones entre las variables no se pueden expresar fácilmente mediante ecuaciones lineales. Además, los árboles de decisión son fáciles de interpretar, lo que permite entender las decisiones tomadas por el modelo basadas en las características predictivas, lo cual es crucial en aplicaciones donde la explicabilidad del modelo es importante.



Este modelo nos devuelve resultados desastrosos para la predicción. Se puede observar un error cuadrático medio (MSE) de 6761.359826951349, lo que sugiere que hay una amplia discrepancia entre los tiempos promedio predichos por el modelo y los tiempos promedio reales, lo que indica que el modelo posee baja precisión. El gráfico nos ilustra este comportamiento.

Se pasa entonces a aplicar la técnica de regresión lineal, especialmente adecuada cuando se identifica una relación lineal entre las variables independientes (como las mencionadas) y la variable dependiente (el tiempo medio de la carrera). La regresión lineal permite modelar esta relación y hacer predicciones precisas sobre el tiempo medio de la carrera basándose en las condiciones específicas de cada carrera.



Al comparar los resultados de la regresión lineal y el árbol de decisión, se observa que el modelo de regresión lineal tiene un Error Cuadrático Medio (MSE) de 2463.44, mucho más bajo que el obtenido anteriormente. Esto indica que el modelo de regresión lineal tiene un rendimiento mejor en términos de cuán bien predice el tiempo medio de una carrera en comparación con el modelo de árbol de decisión. En general, un MSE más bajo sugiere que el modelo tiene un mejor ajuste a los datos, lo que se traduce en predicciones más precisas. En este contexto, el modelo de regresión lineal parece ser más efectivo para predecir el tiempo medio de una carrera basándose en las variables dadas, lo que podría ser debido a la relación lineal entre las variables predictivas y la variable objetivo, una característica que la regresión lineal puede modelar de manera más efectiva que el árbol de decisión, especialmente cuando la relación entre las variables es no lineal o cuando las características interactúan entre sí.

Sin embargo, no se considera ninguno de estos modelos lo suficientemente buenos para realizar predicciones a futuro.

5.4 Comparación de datos extraídos de la simulación y datos reales

Se optó por analizar el tiempo por vueltas en lugar de por carreras debido a la variabilidad inherente en las carreras de Fórmula 1. La variabilidad en la cantidad de vueltas, combinada con las diferencias en el tamaño y la longitud de los circuitos, introduce múltiples factores que pueden afectar significativamente los tiempos finales de la carrera. Al analizar el tiempo por vueltas, se busca una medida más consistente y comparable que refleje mejor las diferencias en el rendimiento entre las vueltas en diferentes carreras y entre los datos reales de la Fórmula 1 y los resultados de la simulación.

Los datos se obtuvieron del dataset Formula 1 World Championship (1950 - 2023), específicamente del archivo lap times.csv, descargado de Kaggle.com y que se muestra como el más completo en relación a datos históricos de F1.

```
Datos Reales
count      551742.000000
mean        95.706206
std         74.786046
min         55.404000
25%         81.993000
50%         90.673000
75%        102.243000
max         7507.547000
Name: milliseconds_en_segundos, dtype: float64

Datos simulados
count      199.000000
mean        90.638261
std         505.296442
min          4.843068
25%         21.946332
50%         40.787363
75%         71.746325
max        7154.676887
Name: tiempo_promedio_vuelta, dtype: float64
```

Media (mean): La media de los datos reales es 95.706206 segundos, mientras que la media de los simulados es 90.638261 segundos. Aunque una es ligeramente mayor, la diferencia es relativamente pequeña y puede estar dentro del rango de variabilidad esperado.

Desviación Estándar (std): La desviación estándar de milliseconds en segundos es 74.786046 segundos, y la de tiempo promedio vuelta es 505.296442 segundos. La desviación estándar de tiempo promedio vuelta es significativamente mayor, lo que indica una mayor variabilidad en los datos de la

simulación en comparación con los datos reales de la Fórmula 1. Esto puede deberse a la manera en la que son generados.

Valores Mínimos (min): El valor mínimo de milliseconds en segundos es 55.404000 segundos, mientras que el valor mínimo de tiempo promedio vuelta es 4.843068 segundos. La diferencia en los valores mínimos sugiere que los tiempos de vuelta en la simulación pueden ser significativamente más cortos que los tiempos de vuelta reales en la Fórmula 1.

Valores Máximos (max): El valor máximo de milliseconds en segundos es 7507.547000 segundos, y el valor máximo de tiempo promedio vuelta es 7154.676887 segundos. Aunque hay una diferencia significativa en los valores máximos, esto puede ser atribuido a la naturaleza de los datos y a la variabilidad inherente en los tiempos de vuelta.

Conclusión: La comparación valor a valor entre las estadísticas descriptivas de milliseconds en segundos y tiempo promedio vuelta revela diferencias significativas en la variabilidad y en los valores extremos. La desviación estándar y los valores mínimos y máximos sugieren que los datos de la simulación pueden no reflejar con precisión la variabilidad y los tiempos de vuelta en la Fórmula 1 real. Se debe tener en cuenta además la diferencia entre los tamaños de las muestras, quizás con un mayor número de simulaciones los valores se acerquen más a los reales. Aún así, estos valores resultan sorprendentemente cercanos a los observados a lo largo de la historia en las carreras de F1, sobre todo los valores de la media. Esta es una conclusión bastante alentadora con respecto a la correctitud de la simulación.

6 Conclusiones

La simulación de carreras de Fórmula 1 ha demostrado ser una herramienta valiosa para explorar la complejidad y variabilidad de estas competencias, permitiendo analizar cómo factores como el clima, la dificultad de la pista, las características de los pilotos y la estrategia del equipo pueden influir significativamente en el resultado de la carrera. Al realizar múltiples simulaciones, se ha logrado obtener una comprensión más profunda de cómo los pilotos y los equipos manejan los desafíos únicos de cada carrera, lo que ha sido fundamental para desarrollar modelos que pueden predecir el peor/mejor tiempo de una carrera, así como su tiempo promedio. Además, se llegaron a conclusiones sobre qué tan cerca a las carreras reales se puede modelar una simulación, lo que subraya la importancia de estas herramientas en el desar-

rollo y la preparación de los equipos de Fórmula 1. La simulación, tanto en el ámbito del piloto como en el informático, se ha convertido en una pieza clave del éxito de los equipos, permitiendo una preparación más detallada y una evaluación más precisa de las posibles configuraciones y estrategias a emplear en las carreras reales.

6.1 Recomendaciones para trabajos futuros

Para trabajos futuros, se recomienda simular más aspectos físicos de la carrera, como la aerodinámica de los coches, los materiales con los que están hechos y demás tecnologías que poseen, así como más interacciones entre los agentes, como la rivalidad histórica entre pilotos. Esto se debe a que estos factores pueden afectar significativamente las decisiones tomadas durante la carrera, influyendo en el rendimiento de los pilotos y en la estrategia de los equipos. La inclusión de estos elementos en las simulaciones puede proporcionar una comprensión más completa de los desafíos y oportunidades que enfrentan los participantes en la Fórmula 1, permitiendo desarrollar estrategias más efectivas y prepararse mejor para las condiciones de carrera reales. Además, la simulación de la rivalidad entre pilotos puede revelar patrones de comportamiento y tácticas que pueden ser explotados para obtener una ventaja competitiva, reflejando la complejidad y la dinámica del deporte.

6.2 Bibliografía

1. Temas de Simulación. Luciano García Garrido. Facultad de Matemática y Computación. Universidad de La Habana
2. Metaheuristics. From design to implementation. University of Lille – CNRS – INRI.
3. Formula One Race Analysis Using Machine Learning [link al paper](#)
4. Real-time decision making in motorsports: analytics for improving professional car race strategy [link al paper](#)