

Reto para el III Hackaton de la Facultad de Ciencias, 5-7 Febrero de 2025

Explorando el Plasma de Quarks y Gluones (QGP) con Aprendizaje Automático (ML):

El universo en sus primeros microsegundos de existencia era un lugar radicalmente diferente al que conocemos hoy. Una sopa extremadamente caliente y densa de partículas fundamentales llamada QGP dominaba el cosmos. Este estado de la materia, compuesto por quarks y gluones libres, es recreado en laboratorios modernos mediante colisiones de alta energía entre núcleos pesados, como los experimentos en el Gran Colisionador de Hadrones (LHC).

En estas colisiones, se generan partículas altamente energéticas llamadas partones (quarks y gluones). A medida que los partones atraviesan el QGP, experimentan un fenómeno conocido como pérdida de energía, que depende de varios parámetros físicos fundamentales:

1. **Constante de acoplamiento fuerte (α_s)** Representa la intensidad de las interacciones entre quarks y gluones. Valores más altos indican interacciones más fuertes.
2. **Escala de separación virtual (Q_0)**: Define el límite entre las fases de interacción de los partones en el QGP.
3. **Mecanismo de pérdida de energía**: Existen 2 modelos para describir este fenómeno:
 - **MATTER**: Interacciones iniciales del partón con el medio.
 - **MATTER-LBT**: Incorpora procesos de dispersión y radiación adicionales.

Los datos de estas colisiones se representan en imágenes de 32x32 píxeles, donde cada píxel codifica información sobre la energía y la dirección del flujo de partículas.

Planteamiento del problema:

La tarea es construir un modelo de aprendizaje automático (ML) que, a partir de las imágenes generadas por simulaciones de colisiones, sea capaz de predecir las 3 variables e:

1. **Mecanismo de pérdida de energía**: MATTER o MATTER-LBT (clasificación binaria).
2. **Constante de acoplamiento fuerte (α_s)**: Valores posibles: 0.2, 0.3 y 0.4 (multiclase).
3. **Escala de separación virtual(Q_0)**: Valores posibles: 1, 1.5, 2.0 y 2.5. (multiclase).

Para el desarrollo del problema se proporciona a cada grupo:

- El mismo dataframe con 1 millón de imágenes de colisiones, así como una tupla de las 3 variables a predecir.
- El mismo dataframe con 100.000 imágenes, sin variables objetivo. Para evaluar la precisión del modelo desarrollado de la manera más objetiva posible.

Como tarea adicional, se pide preparar un chatbot capaz de mostrar los resultados obtenidos y responder a preguntas que argumenten las explicaciones físicas que hay detrás de esos resultados en base a las 3 variables sin profundizar en las ideas

Tareas específicas:

1. Preprocesamiento de los datos:

- Las imágenes y los datos objetivo deberán tratarse adecuadamente por cada grupo para el correcto modelaje de los algoritmos.

2. Diseño del modelo:

- Implementa un algoritmo de ML (ej. red neuronal convolucional (CNN)) capaz de aprender características visuales relevantes de las imágenes.
- Diseña el algoritmo con las 3 salidas objetivo mencionadas.

3. Evaluación del modelo:

- La evaluación final se realizará sobre un dataset diferente al proporcionado para realizar el entrenamiento del modelo, asegurando objetividad entre grupos.
- Las métricas que se van a evaluar van a ser la precisión (accuracy) de cada una de las salidas y una media aritmética de las tres.
- Se deben analizar los resultados con una matriz de confusión para cada tarea.

4. Interpretación de resultados:

- Reflexiona sobre las predicciones del modelo. ¿Qué patrones visuales parecen ser más relevantes para predecir cada variable?
- Compara el desempeño del modelo en las tres tareas. ¿Qué tarea es más difícil de predecir? ¿Por qué?

5. Consulta de los resultados:

- Desarrolla un chatbot sencillo (desarrollado con o sin código), basado en IA generativa, que permita la consulta de los resultados.
- Incorpora argumentaciones sobre el porqué de las predicciones para respaldar los resultados con una base de conocimiento teórico.
- Se evaluará la capacidad de dar argumentaciones fundamentadas que no caigan en alucinaciones.

Entregables esperados:

1. Código fuente del modelo desarrollado:

- Debe incluir el preprocesamiento, la definición del modelo y el entrenamiento.
- Debe incluir también las métricas obtenidas, así como las matrices de confusión generadas y algún gráfico relevante.
- Fichero con las predicciones en formato tupla de las 3 variables indicadas.

2. Presentación en PowerPoint o similar:

- Incluye una breve introducción a los conceptos físicos del QGP.
- Explica cómo preprocesaste datos, diseñaste el modelo y evaluaste resultados.
- Presenta los resultados obtenidos, incluyendo las métricas y las visualizaciones de las matrices de confusión.
- Realiza un análisis de interpretabilidad de los datos respondiendo las preguntas propuestas.
- Realizar una conclusión con líneas de seguimiento y posibles aplicaciones futuras.

3. Gráficas y análisis:

- Muestra la evolución de la precisión y la pérdida durante el entrenamiento.
- Incluye una discusión sobre los desafíos y posibles mejoras.

4. Chatbot con IA generativa

- Responde dando los resultados obtenidos del modelo de predicción.
- Fundamenta con conocimiento teórico los resultados del modelo de predictivo.