



## Propuesta de Trabajo Fin de Grado del Doble Grado en Física y Matemáticas

*Responsable de tutorización:* Bruno Zamorano García

*Correo electrónico:* bzamorano@ugr.es

*Departamento:* Física Teórica y del Cosmos

*Área de conocimiento:* Física Teórica

*Responsable de cotutorización:* José Luis Romero Béjar

*Correo electrónico:* jlrbejar@ugr.es

*Departamento:* Estadística e Investigación Operativa

*Área de conocimiento:* Estadística e Investigación Operativa

*(Rellenar solo en caso de que la propuesta sea de un estudiante):*

*Estudiante que propone el trabajo:* Baltasar Miguel Fenoll

*Título:* Machine learning aplicado a la búsqueda de leptones neutros pesados en el experimento SBND

*Número de créditos:*  6 ECTS  12 ECTS

*Tipología del trabajo (marcar las casillas que correspondan):*

- 1. Revisiones y/o trabajos bibliográficos sobre el estado actual de aspectos específicos relacionados con la titulación
- 2. Estudio de casos, teóricos o prácticos, relacionados con la temática de la titulación, a partir de material disponible en los centros
- 3. Trabajos experimentales, de toma de datos de campo, de laboratorio, etc.
- 4. Elaboración de nuevas prácticas de laboratorio
- 5. Elaboración de un informe o un proyecto en el ámbito del grado de naturaleza profesional
- 6. Trabajos relacionados con las prácticas externas

*Descripción y resumen de contenidos:*

La identificación de patrones y la elaboración de predicciones son importantes en muchas aplicaciones de la física, tales como la astrofísica, la física de partículas, la mecánica cuántica, la termodinámica, la óptica, entre otras. En el aprendizaje supervisado o *machine learning*, se utiliza un conjunto de datos etiquetados para entrenar un modelo que pueda hacer predicciones o tomar decisiones en función de nuevas entradas no etiquetadas. Los algoritmos utilizados en el aprendizaje supervisado se basan en modelos probabilísticos, como la regresión lineal, la regresión logística, los árboles de decisión, las redes neuronales, entre otros. Para elegir una metodología de clasificación adecuada, es importante identificar el carácter cuantitativo o cualitativo de las variables de entrada, la distribución de probabilidad conjunta de las variables explicativas, y otras características de los datos. En función de estos factores, se pueden elegir diferentes técnicas de clasificación que sean más adecuadas para los datos en cuestión. Es importante destacar que el aprendizaje supervisado no es la única técnica de aprendizaje automático utilizada en física, ya que existen otras técnicas como el aprendizaje no supervisado, el aprendizaje por refuerzo, y otras técnicas de aprendizaje profundo. Cada una de estas técnicas tiene sus propias ventajas y desventajas, y su elección dependerá del tipo de problema que se esté abordando.

*(Continúa en página siguiente)*

Uno de los ámbitos en los que el aprendizaje supervisado es de uso habitual es el análisis de los datos registrados por detectores de Física de partículas. En este trabajo se pretende aplicar técnicas de aprendizaje supervisado y análisis multivariable para el análisis de los datos del experimento de neutrinos SBND. Este experimento, basado en el uso de cámaras de deriva de argón líquido, pretende confirmar o refutar la existencia de un cuarto neutrino estéril. Se trata de un detector muy sofisticado, que suministra una información muy granular y detallada acerca de las partículas secundarias que se producen tras la interacción del neutrino con el núcleo. Por todo ello, se trata de un contexto particularmente apropiado para el uso de técnicas de aprendizaje supervisado. Además del programa de neutrinos estériles, SBND posee un extenso programa de física, entre el que se incluye la búsqueda de partículas exóticas propuestas en modelos de extensión al Modelo Estándar de la física de partículas.

En este trabajo se propone que el alumno, a partir de una discusión general sobre la problemática de la clasificación estadística, profundice en el análisis del modelo más adecuado según el carácter de las variables de entrada, dando una visión completa y actualizada de sus aspectos fundamentales, su implementación, su validación y su aplicación a la búsqueda de leptones neutros masivos (*heavy neutral leptons*) en el experimento SBND.

*Actividades a desarrollar:*

- Revisión bibliográfica sobre metodologías consolidadas de clasificación estadística, con especial referencia a su estado actual, y exposición sintética del conocimiento desde una perspectiva global en el contexto del análisis estadístico de datos multivariantes.
- Revisión bibliográfica sobre la técnica de detección de las cámaras de deriva de argón líquido (LArTPCs), así como de los objetivos experimentales de SBND y del programa de física más allá del Modelo Estándar.
- Profundización en el análisis de datos según del modelo más adecuado a un problema concreto, según el carácter de los datos multivariantes usados como entrada, con una clara identificación de los elementos conceptuales, y exposición de sus fundamentos matemáticos y aspectos metodológicos.
- Aplicación para la clasificación de topologías de estado final mediante el uso y desarrollo eventual de procedimientos computacionales y gráficos relacionados.

*Objetivos planteados*

*Objetivo 1*

Identificación de los aspectos conceptuales inherentes a los enfoques de clasificación objeto de estudio y su formalización matemática.

Análisis pormenorizado de los fundamentos matemáticos que sustentan el desarrollo de la metodología objeto central del trabajo.

Desarrollo de un estudio aplicado e interpretación de resultados con referencia precisa a la fundamentación matemática de la metodología.

## ***Bibliografía***

- [1] C. C. AGGARWAL, *Data Mining: The Textbook*, (Springer, Switzerland, 2015).
- [2] G. D. GARSON, *Logistic Regression: Binary and Multinomial*, (Statistical Associates Publishing, Asheboro, 2014).
- [3] W. K. HARDLE, L. SIMAR, *Applied Multivariate Statistical Analysis*, (Springer, Berlin, 2015).
- [4] B. RATNER, *Statistical and Machine-Learning Data Mining. Techniques for Better Predictive Modelling and Analysis of Big Data*, (CRC Press, Boca Raton, 2011).
- [5] A. C. RENCHER, W. F. CHRISTENSEN, *Methods of Multivariate Analysis*, (John Wiley & Sons, New Jersey, 2012).
- [6] D. ZELTERMAN, *Applied Multivariate Statistics with R*, (Springer, New York, 2015).
- [7] <http://sbn-nd.fnal.gov/>
- [8] P. MACHADO, O. PALAMARA, D. SCHMITZ, *The Short-Baseline Neutrino Program at Fermilab*, *Ann.Rev.Nucl.Part.Sci.* 69 (2019) 363-387.
- [9] P. BALLETT, S. PASCOLI, M. ROSS-LONERGAN, *MeV-scale sterile neutrino decays at the Fermilab Short-Baseline Neutrino program*, arXiv:1610.08512
- [10] M. THOMSON, *Modern Particle Physics*, (Cambridge University Press, 2015)

*(Firmar solo en caso de trabajos propuestos por estudiantes)*

*Firma del estudiante*

*Firma del responsable de tutorización*

*Firma del responsable de cotutorización*

En Granada, a 16 de mayo de 2023.