

A continuación se discuten una serie de posibles ideas para el desarrollo de TFGs/TFMs:

Análisis de Datos con técnicas de Aprendizaje Automático Interpretable

Estamos en la era de los datos. Tecnologías como las redes inalámbricas (Wi-Fi, 4G/5G) y el IoT están permitiendo conectar a Internet cada vez más parcelas de nuestro mundo físico, creando avatares en el mundo digital que pueden ser medidos y analizados desde cualquier punto. Smart-home/city/energy/transportation/farming, e-health, image/text recognition son algunas de las áreas transformadas por este nuevo paradigma. Para responder a los nuevos retos y oportunidades, el aprendizaje automático ha recibido un gran impulso desde entidades públicas y privadas (con Google, Tesla, Amazon y otras compañías a la cabeza), generando tecnologías de alto rendimiento ante gigantescos volúmenes de datos.

El **aprendizaje automático interpretable** es aquel que, además de ofrecer funcionalidades típicas del aprendizaje automático (como clasificación, predicción, detección de anomalías) permite la interpretación del modelo aprendido, lo que lo hace especialmente útil en múltiples aplicaciones, como los servicios telemáticos con operador (gestión de red, detección de intrusiones en ciberseguridad) o aplicaciones telemáticas basadas en datos (e-health, redes de sensores & IoT, smart-lo-que-sea).

Si tienes interés en algún dominio concreto y quieres desarrollar un TFG/TFM donde el análisis de datos es pieza fundamental, o si el análisis de datos te interesa pero no tienes muy claro dónde o cómo aplicarlo, contacta conmigo. Algunos repositorios de datos extremadamente interesantes son:

- The Cancer Genome Atlas Program: <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>
Datos genómicos relacionados con el desarrollo de cáncer.
- Cawdad Repository: <http://cawdad.org/>
Datos/estadísticas medidas sobre redes inalámbricas con diferentes tecnologías.
- David Reich Lab: <https://reich.hms.harvard.edu/downloadable-genotypes-present-day-and-ancient-dna-data-compiled-published-papers>
Datos de ADN antiguo (Neandertales, humanos antiguos, etc.) para entender la evolución humana y la historia.
- CAIDA Repository: <https://www.caida.org/data/>
Datos de tráfico en Internet, topología, routing y similar.

¿Pre-requisitos del trabajo? Es recomendable un manejo adecuado de Matlab (idealmente) o, en su defecto, de Python o R.

¿Qué capacidades desarrollaré? Manejo de datos, programación de scripts, técnicas de aprendizaje automático, técnicas de investigación, redacción de documentos, desarrollo de proyectos.

Nivel de complejidad: Medio (configurable)

Titulación: Grado en Ingeniería Informática, Grado en Ingeniería de Tecnologías de Telecomunicación, Máster en Ingeniería de Telecomunicación.

Detección de Anomalías: PCA vs Autoencoders

En la era de los datos, la técnica de aprendizaje automático más extendida es sin duda las redes neuronales profundas. Cuando hablamos de estas redes en el contexto no supervisado, hablamos de Autoencoders. Se ha establecido en la comunidad científica la creencia de que la detección de anomalías basada en Autoencoders es superior a la basada en el Análisis por Componentes Principales (PCA). Sin embargo, no se ha realizado un estudio detallado en términos de la dimensionalidad de los datos (el número de características) ni de la interpretabilidad de los modelos. Este proyecto tiene por objetivo realizar dicho estudio. Los resultados pueden ser de relevancia en investigación (y por tanto publicables)

¿Pre-requisitos del trabajo? Es recomendable un manejo adecuado de Matlab y de Python.

¿Qué capacidades desarrollaré? Manejo de datos, programación de scripts, técnicas de aprendizaje automático, técnicas de investigación, redacción de documentos, desarrollo de proyectos.

Nivel de complejidad: Alto (configurable)

Titulación: Grado en Ingeniería Informática, Grado en Ingeniería de Tecnologías de Telecomunicación, Máster en Ingeniería de Telecomunicación.

Evaluación de dispositivos *lab_in_a_chip* para e-Health

El surgimiento del IoT es probablemente el paradigma llamado a cambiar de forma más evidente nuestra sociedad, al posibilitar multitud de nuevos servicios antes impensables. Un dominio donde esperamos avances significativos es el del e-Salud. Crisis como la del COVID19 han demostrado que estamos aún muy lejos de conocer en detalle el funcionamiento biológico de nuestro cuerpo. Sin embargo, los dispositivos IoT, como *wearables* de actividad, monitorización de salud o los laboratorios_en_un_chip, permiten recoger ingentes cantidades de datos en tiempo real sobre nuestros hábitos, nuestra biología y los distintos contaminantes a los que nos exponemos.

Aunque el desarrollo de nuevos dispositivos de actividad y monitorización de salud es una parte relevante de la I+D+d actual, se ha estudiado menos el uso de dispositivos *lab_in_a_chip*. El objetivo de este proyecto es realizar una revisión de estos dispositivos y, si se ve viable, realizar una prueba de concepto a partir de algún dispositivo comercial.

¿Pre-requisitos del trabajo? Ninguno.

¿Qué capacidades desarrollaré? Conocimiento de dispositivos y técnicas e-Health, configuración de entornos de pruebas, técnicas de investigación, redacción de documentos, desarrollo de proyectos.

Nivel de complejidad: Medio (configurable)

Titulación: Grado en Ingeniería Informática, Grado en Ingeniería de Tecnologías de Telecomunicación, Máster en Ingeniería de Telecomunicación.

Privacidad en AI: Federated Learning

Compañías como Google, Facebook, Microsoft, Netflix o Amazon utilizan ingentes cantidades de datos de clientes para particularizar y mejorar el rendimiento de sus servicios. El precio a pagar es una clara pérdida de privacidad. Esta pérdida de privacidad es más relevante al acercarse a nuestro mundo físico, en lugar de a una cuenta digital. Este es uno de los riesgos asociados al desarrollo de los entornos Smart-lo-que-sea, como smart-home, Smart-city o smart-health. Existe una necesidad acuciante del desarrollo de metodologías que permitan el uso de la Inteligencia Artificial sin necesidad de la compartición de datos sensibles. Una técnica prometedora en este ámbito es Federated Learning, donde los modelos de Machine Learning son calibrados de forma distribuida.

El objetivo de este proyecto es trabajar en una combinación de Federated Learning con técnicas de Análisis Multivariante para el análisis de datos privados.

¿Pre-requisitos del trabajo? Es recomendable un manejo adecuado de Matlab (idealmente) o, en su defecto, de Python o R.

¿Qué capacidades desarrollaré? Manejo de datos, programación de scripts, técnicas de aprendizaje automático, técnicas de investigación, redacción de documentos, desarrollo de proyectos.

Nivel de complejidad: Alto (configurable)

Titulación: Grado en Ingeniería Informática, Grado en Ingeniería de Tecnologías de Telecomunicación, Máster en Ingeniería de Telecomunicación.

Revisión de Sistemas de Recomendación

Netflix, Facebook, Spotify o Amazon son ejemplos de sistemas de recomendación: buscan patrones en los gustos de los clientes para poder recomendar nuevos ítems. Existe una extensa literatura en el tema.

El objetivo de este proyecto es realizar un estado de la técnica y una comparativa de un conjunto selecto de estrategias.

¿Pre-requisitos del trabajo? Es recomendable un manejo adecuado de Matlab, Python o R.

¿Qué capacidades desarrollaré? Manejo de datos, programación de scripts, técnicas de aprendizaje automático, técnicas de investigación, redacción de documentos, desarrollo de proyectos.

Nivel de complejidad: Alto (configurable)

Titulación: Grado en Ingeniería Informática, Grado en Ingeniería de Tecnologías de Telecomunicación, Máster en Ingeniería de Telecomunicación.

Desarrollo de herramientas open-software de Análisis de Datos

Dos proyectos de Github en los que tengo especial interés son:

- La MEDA Toolbox (<https://github.com/josecamachop/MEDA-Toolbox>): para el desarrollo de nuevos algoritmos de análisis de datos en Matlab.
- El FC Parser (<https://github.com/josecamachop/FCParser>) para el parseo de ingentes cantidades de datos en contadores.

Ambos proyectos requieren de colaboración, que puede ser encauzada en forma de TFG/TFM.

¿Pre-requisitos del trabajo? Es recomendable un manejo adecuado de Matlab o Python.

¿Qué capacidades desarrollaré? Manejo de datos, programación, técnicas de aprendizaje automático, redacción de documentos, desarrollo de proyectos.

Nivel de complejidad: Medio (configurable)

Titulación: Grado en Ingeniería Informática.

Optimización Matemática en Diseño de Red

Este proyecto se enmarca en el ámbito de la optimización matemática introducida en la asignatura Diseño y Dimensionado de Redes (DDR), de 4º del Grado en Ingeniería de Tecnologías de Telecomunicación.

¿Pre-requisitos del trabajo? Haber cursado o tener proyectado cursar DDR.

¿Qué capacidades desarrollaré? Técnicas de Optimización, diseño y dimensionado de redes, configuración de entornos de pruebas, técnicas de investigación, redacción de documentos, desarrollo de proyectos.

Nivel de complejidad: Alto (configurable)

Titulación: Grado en Ingeniería de Tecnologías de Telecomunicación, Máster en Ingeniería de Telecomunicación.

Diseño Estadístico de Experimentos en Red

Este proyecto se enmarca en el ámbito del diseño de experimentos introducido en la asignatura Planificación y Explotación de Redes y Servicios (PERS), de 1º del Máster en Ingeniería de Telecomunicación.

¿Pre-requisitos del trabajo? Haber cursado o tener proyectado cursar PERS.

¿Qué capacidades desarrollaré? Técnicas de Diseño de Experimentos, técnicas de análisis de varianza, diseño y dimensionado de redes, configuración de entornos de pruebas, técnicas de investigación, redacción de documentos, desarrollo de proyectos.

Nivel de complejidad: Medio (configurable)

Titulación: Máster en Ingeniería de Telecomunicación.