

MEMORIA DE PRÁCTICAS

EXTRACURRICULARES

Grado en Ciencia de Datos

Estudiante: José Milán Server

Tutor UPV: Joaquín Martínez Minaya

Universidad: Universitat Politècnica de València

Empresa: Valencian Research Institute for Artificial Intelligence (VRAIN)

Duración de las prácticas: 27/01/2026 – 30/06/2026

Índice

1. Introducción, descripción de la empresa y contexto de las prácticas	2
1.1. Descripción general de la empresa	2
1.2. Organización interna	2
1.3. Descripción general del proyecto desarrollado	2
2. Relación con el plan de estudios, tecnologías y competencias adquiridas	3
3. Diario de trabajo	4
3.1. Fase 1. Incorporación y preparación metodológica	4
3.2. Fase 2. Consolidación del marco teórico y definición del enfoque	5
3.3. Fase 3. Desarrollo del análisis bayesiano espacio-temporal	5
3.4. Fase 4. Preparación de la comunicación científica	6
3.5. Fase 5. Refinamiento de los modelos y ampliación del estudio	6
3.6. Fase 6. Desarrollo del Trabajo Fin de Grado, documentación y cierre del periodo	7
4. Aprendizaje	7
5. Evaluación de las prácticas	8

1. Introducción, descripción de la empresa y contexto de las prácticas

En este apartado se describe el contexto general de las prácticas extracurriculares, incluyendo una presentación de la empresa, su ámbito de actuación y el entorno en el que se ha desarrollado el proyecto.

1.1. Descripción general de la empresa

El Instituto Universitario de Investigación VRAIN (Valencian Research Institute for Artificial Intelligence), perteneciente a la Universitat Politècnica de València (UPV), es uno de los centros de investigación en Inteligencia Artificial más importantes de España. Con más de 30 años de trayectoria, VRAIN destaca por su excelencia científica y su capacidad de transferencia tecnológica, avalada por una producción científica de más de 2.000 artículos y la participación en más de 135 proyectos competitivos, principalmente de la Unión Europea y el Plan Nacional de Investigación.

La estructura de VRAIN es de carácter multidisciplinar, integrando a 176 investigadores distribuidos en ocho grupos de investigación de referencia nacional e internacional. Su actividad investigadora se despliega en doce áreas clave, que van desde el Machine Learning y Deep Learning hasta la Verificación de Software, pasando por el Procesamiento del Lenguaje Natural y la Ciencia de Datos Genómicos. Estos desarrollos se aplican en sectores estratégicos como la salud, la agricultura, la industria y la sostenibilidad medioambiental, conectando la vanguardia teórica con soluciones reales.

1.2. Organización interna

El proyecto se planteó de manera individual, lo que me permitió experimentar de primera mano la metodología de trabajo de un centro de investigación de élite: rigor en el tratamiento de datos, revisión constante de la literatura científica y autonomía en la toma de decisiones técnicas. La supervisión estuvo a cargo principalmente de Joaquín Martínez y Víctor Sánchez, quienes realizaron un seguimiento continuo del proyecto, aportando orientación técnica y metodológica en las distintas fases del mismo.

1.3. Descripción general del proyecto desarrollado

El proyecto asignado se enmarca en el ámbito de la modelización predictiva aplicada al sector salud, centrándose en el estudio de la demanda del Transporte No Asistido (TNA)

en la ciudad de Valencia.

El Transporte No Asistido es un servicio destinado al traslado de personas que, sin requerir atención médica urgente, necesitan un transporte especializado debido a su situación clínica. Este tipo de servicio incluye, entre otros, traslados por altas hospitalarias, consultas externas, tratamientos periódicos como hemodiálisis o quimioterapia, rehabilitación y pruebas complementarias. A pesar de su importancia operativa y económica, existe una escasez de estudios centrados en analizar las causas de la variabilidad de la demanda de TNA entre distintas zonas urbanas.

El objetivo principal del proyecto fue la implementación de estrategias descentralizadas mediante agentes inteligentes capaces de anticipar la demanda de este servicio, identificando los factores más influyentes en su variabilidad espacial y temporal. Para dar soporte a esta arquitectura multi-agente, se llevó a cabo un análisis exhaustivo de los datos disponibles, prestando especial atención a su calidad, estructura y adecuación para la modelización y el entrenamiento de los agentes.

El trabajo desarrollado durante estas prácticas constituyó, además, la base de mi Trabajo Fin de Grado, centrado en el análisis y la predicción de la demanda diaria de transporte sanitario no asistido a nivel de barrio. De este modo, las prácticas y el TFG se desarrollaron de manera coordinada: el entorno de VRAIN proporcionó el marco investigador, los datos, la supervisión técnica y los recursos necesarios, mientras que el TFG permitió estructurar, documentar y evaluar académicamente los resultados obtenidos.

2. Relación con el plan de estudios, tecnologías y competencias adquiridas

Las tareas desarrolladas durante las prácticas guardan una relación directa con los contenidos y competencias adquiridos a lo largo del Grado en Ciencia de Datos. La formación académica recibida ha sido especialmente útil para abordar actividades como el análisis exploratorio de datos, análisis estadístico y la evaluación de la viabilidad de aplicar técnicas de modelización en un contexto real.

En el ámbito de la visualización de datos, se aplicaron conocimientos adquiridos en la asignatura **Visualización (VIS)**, cursada durante el primer cuatrimestre del tercer curso. Esta formación permitió representar la información de forma clara y comprensible, facilitando la identificación de patrones y tendencias relevantes.

Desde una perspectiva estadística e inferencial, los conocimientos adquiridos en las asignaturas **Métodos Estadísticos para la toma de decisiones I (MET I)** y **Modelos Descriptivos y Predictivos I (MDP I)** aportaron conocimientos clave en técnicas de

reducción de la dimensionalidad, como el Análisis de Componentes Principales (PCA), así como en el uso de herramientas estadísticas para simplificar y analizar conjuntos de datos complejos.

En relación con el análisis temporal, se aplicaron conceptos de series temporales trabajados en la asignatura **Economía Digital (EDIG)**, empleando el lenguaje R para el estudio de la evolución temporal de la demanda y la detección de patrones estacionales.

Asimismo, las asignaturas **Métodos Estadísticos para la Toma de Decisiones II (MET II)** y **Modelos Descriptivos y Predictivos II (MDP II)** resultaron fundamentales para la correcta aplicación de técnicas estadísticas y de aprendizaje automático. En particular, en MET II se trabajaron métodos como el análisis de la varianza (ANOVA) y distintos tipos de regresión, mientras que en MDP II se abordaron métodos de clasificación, directamente relacionados con los análisis realizados.

De manera transversal, las asignaturas de **Proyectos I, II, III y Gestión de Proyectos (GPR)** contribuyeron al desarrollo de competencias en organización del trabajo, toma de decisiones, seguimiento de hitos y comunicación de resultados, habilidades que han resultado fundamentales durante el desarrollo de las prácticas.

3. Diario de trabajo

El desarrollo de las prácticas extracurriculares se organizó en diferentes fases de trabajo, definidas de acuerdo con la evolución del proyecto y con las principales tareas realizadas en cada periodo. Debido a la duración de las prácticas y a la continuidad existente entre algunas actividades, se ha optado por una organización temática y temporal más amplia, en lugar de realizar una división estrictamente bisemanal.

3.1. Fase 1. Incorporación y preparación metodológica

Mi incorporación a este nuevo periodo de prácticas se produjo con una base de trabajo ya consolidada. Durante las prácticas curriculares anteriores se habían realizado los principales procesos de limpieza, filtrado, integración y transformación de los datos en bruto. Por este motivo, el trabajo desarrollado en VRAIN durante las prácticas extracurriculares pudo comenzar desde una etapa de mayor madurez técnica.

En esta primera fase, el trabajo se centró principalmente en la preparación teórica y metodológica necesaria para abordar técnicas de modelización más avanzadas. Se llevó a cabo una revisión de arquitecturas basadas en sistemas multiagente y se profundizó en el estudio de modelos bayesianos aplicados al análisis espacio-temporal.

El objetivo era explorar soluciones más sofisticadas que permitieran mejorar los resultados obtenidos en las etapas anteriores del proyecto. Para ello, se prestó especial atención tanto a la capacidad predictiva de los modelos como a su interpretabilidad y a la posibilidad de cuantificar la incertidumbre asociada a sus estimaciones.

3.2. Fase 2. Consolidación del marco teórico y definición del enfoque

Una vez revisadas las principales alternativas metodológicas, el trabajo se orientó hacia la consolidación del marco teórico y la definición precisa del enfoque que se utilizaría durante el resto de las prácticas.

Al disponer de un conjunto de datos previamente procesado, fue posible centrar los esfuerzos en comprender con mayor profundidad la naturaleza del problema y su aplicación dentro del ámbito sanitario. También se estudió la posible arquitectura de agentes inteligentes que podría utilizarse para integrar los diferentes componentes desarrollados en el proyecto.

Durante esta fase se realizaron diversas reuniones de seguimiento con los tutores, que permitieron contextualizar los objetivos técnicos y relacionarlos con las necesidades reales del servicio de transporte sanitario. Este intercambio de ideas fue fundamental para concretar las siguientes líneas de trabajo y comenzar el desarrollo de modelos más complejos, superando los análisis descriptivos y predictivos básicos realizados inicialmente.

3.3. Fase 3. Desarrollo del análisis bayesiano espacio-temporal

La siguiente fase se centró en el desarrollo de un enfoque basado en inferencia bayesiana para analizar la demanda de transporte sanitario no asistido en la ciudad de Valencia.

Mediante esta metodología se estudió la evolución espacial y temporal de la demanda a nivel de barrio. Los modelos desarrollados permitieron incorporar variables demográficas, socioeconómicas, ambientales y relacionadas con la infraestructura sanitaria y de transporte. Además, se incluyeron componentes espaciales y temporales destinados a representar las relaciones existentes entre barrios próximos y la evolución de la demanda a lo largo del tiempo.

A diferencia de otros modelos predictivos utilizados previamente, el enfoque bayesiano permitió cuantificar explícitamente la incertidumbre de las estimaciones y obtener resultados con una interpretación epidemiológica. En particular, fue posible analizar el riesgo relativo de demanda de cada barrio e identificar zonas con una demanda superior o inferior

a la esperada.

Durante esta fase también se compararon distintas especificaciones espacio-temporales con el objetivo de seleccionar el modelo que ofreciera un equilibrio adecuado entre ajuste, complejidad e interpretabilidad.

3.4. Fase 4. Preparación de la comunicación científica

Como resultado del trabajo realizado, se me propuso participar en el III Congreso de Salud, Desastres y Desarrollo Sostenible (SDDS). Esta oportunidad supuso un hito relevante dentro de las prácticas, ya que requirió adaptar el estudio técnico a un formato de comunicación científica.

El trabajo de esta fase incluyó la selección y síntesis de los resultados más relevantes, la elaboración de gráficos y tablas, la descripción de la metodología utilizada y la redacción de la comunicación técnica. También fue necesario adaptar el contenido a las limitaciones de extensión y a los criterios establecidos por el congreso, manteniendo el rigor metodológico y facilitando la comprensión del estudio.

La preparación de esta comunicación permitió revisar de forma crítica el trabajo realizado, identificar los resultados con mayor interés científico y mejorar la capacidad para transmitir conceptos estadísticos y computacionales a un público especializado.

3.5. Fase 5. Refinamiento de los modelos y ampliación del estudio

Tras la preparación inicial de la comunicación científica, el trabajo continuó con el refinamiento de los modelos desarrollados y la ampliación del análisis. Se revisaron diferentes alternativas para representar las dependencias espaciales y temporales, así como distintas métricas para evaluar la calidad de los modelos.

También se analizaron los residuos, la capacidad predictiva y la incertidumbre de las estimaciones. Estos análisis permitieron detectar posibles limitaciones, comprobar si permanecían patrones espaciales o temporales sin explicar y valorar la robustez de los resultados obtenidos.

Paralelamente, se comparó el enfoque bayesiano con diferentes técnicas de aprendizaje automático. Esta comparación permitió estudiar las ventajas y limitaciones de cada familia de modelos. Mientras que los métodos de aprendizaje automático ofrecían una elevada capacidad predictiva, los modelos bayesianos proporcionaban una mayor interpretabilidad y una representación explícita de la incertidumbre y de la estructura espacio-temporal del problema.

3.6. Fase 6. Desarrollo del Trabajo Fin de Grado, documentación y cierre del periodo

Durante la fase final de las prácticas, una parte importante del trabajo se destinó al desarrollo y finalización de mi Trabajo Fin de Grado, elaborado a partir del proyecto realizado en VRain. Esto implicó organizar y documentar de manera estructurada todo el proceso seguido, desde la preparación de los datos y la definición metodológica hasta la evaluación de los modelos y la interpretación de los resultados.

La elaboración del TFG requirió revisar los procedimientos implementados, actualizar los resultados de los modelos, justificar las decisiones metodológicas adoptadas y comparar de manera crítica los enfoques bayesianos y de aprendizaje automático. También se prepararon las figuras, tablas y materiales necesarios para comunicar los resultados de forma clara y rigurosa.

Además de evaluar el rendimiento predictivo, se profundizó en la interpretación de las variables, la cuantificación de la incertidumbre y la identificación de patrones relevantes para la planificación del transporte sanitario. El análisis permitió estudiar las diferencias existentes entre barrios, los cambios temporales de la demanda y las zonas en las que existía una mayor probabilidad de observar valores superiores o inferiores a los esperados.

Finalmente, se llevó a cabo la redacción de la memoria académica y la preparación de la presentación para la defensa del TFG. Este proceso permitió sintetizar el trabajo realizado durante las prácticas y exponer de forma estructurada tanto las aportaciones metodológicas como los resultados y las principales conclusiones del proyecto.

Este periodo de prácticas extracurriculares permitió, por tanto, dar continuidad al trabajo iniciado durante las prácticas curriculares y desarrollar una propuesta metodológica considerablemente más avanzada, que se materializó tanto en el Trabajo Fin de Grado como en diferentes actividades de difusión científica. Asimismo, hizo posible participar en todas las etapas del proceso investigador, desde la definición del problema y la construcción de los modelos hasta su interpretación, documentación, evaluación y comunicación.

4. Aprendizaje

Las prácticas en VRain han supuesto un salto cualitativo en mi formación al permitirme experimentar un entorno de investigación de vanguardia. Esta estancia me ha servido para conocer desde dentro el funcionamiento de un centro de referencia en Inteligencia Artificial y colaborar con profesionales de gran valor. He comprendido que la investigación científica requiere un rigor metodológico y una profundidad teórica que van mucho más

allá de la ejecución de procesos técnicos, transformando mi visión sobre el papel de la Inteligencia Artificial en la sociedad.

Desde el punto de vista técnico, el mayor reto fue la implementación de la Inferencia Bayesiana, una metodología que no se abordó en profundidad durante el grado. Bajo la orientación de mis tutores, tuve que desarrollar una actitud proactiva para dominar este enfoque espacio-temporal. Este proceso de autoaprendizaje no solo me ha dotado de una herramienta analítica muy potente, sino que ha reforzado mi capacidad para resolver problemas complejos de manera autónoma, una competencia fundamental para cualquier científico de datos.

La elaboración del Trabajo Fin de Grado en el contexto de estas prácticas también supuso un aprendizaje especialmente relevante. Me permitió abordar un proyecto de Ciencia de Datos de principio a fin, tomando decisiones sobre la preparación de los datos, la selección de metodologías, la validación de los modelos y la presentación de los resultados. Asimismo, reforzó mi capacidad para redactar documentación académica, justificar decisiones técnicas y comunicar de manera comprensible resultados estadísticos complejos.

Finalmente, la oportunidad de participar en el III Congreso de Salud, Desastres y Desarrollo Sostenible ha sido el hito más enriquecedor de mis prácticas. Esta experiencia me ha permitido desarrollar habilidades transversales de comunicación científica y síntesis de resultados que no había ejercitado hasta la fecha. Presentar el trabajo ante una comunidad internacional ha sido la mejor manera de cerrar esta etapa, aportándome una madurez profesional y una capacidad de exposición que serán claves en mi futuro desempeño laboral.

5. Evaluación de las prácticas

La valoración global de este periodo de prácticas extracurriculares en VRAIN es muy positiva. A lo largo de los cinco meses de duración de las prácticas, pude profundizar en el proyecto iniciado durante el periodo curricular, desarrollar metodologías más avanzadas y aplicar los resultados obtenidos tanto en mi Trabajo Fin de Grado como en distintas actividades de difusión científica.

Uno de los aspectos más destacables es la calidad humana y profesional del equipo. Tanto los tutores como los investigadores con los que he interactuado han facilitado una integración rápida y acogedora. El proyecto de agentes inteligentes para el transporte sanitario resulta altamente motivador, no solo por su complejidad técnica y el uso de inferencia bayesiana, sino por el impacto social real que tiene la optimización de estos servicios de salud.

Por último, quiero expresar mi agradecimiento al instituto VRAIN y a mis tutores por la confianza depositada en mí desde el primer día y por brindarme la oportunidad de presentar nuestro trabajo en el III Congreso de Salud, Desastres y Desarrollo Sostenible (SDDS). Esta experiencia ha sido fundamental para reafirmar mi interés por desarrollar soluciones técnicas que tengan un impacto social tangible.