



Memoria final

Proyectos de Innovación Docente 2023-2024

1. Identificación del proyecto

Título: Propuesta cooperativa de trabajo para estudiantes de Ingeniería de Datos UZ , CESI Engineer-IT-Computer Science e Ingeniería y Ciencia de Datos URJC en minería de datos aplicada

Programa: PIET (Programa de Innovación Estratégica de Titulaciones)

Línea: PIET_3

Centro: Escuela Universitaria Politécnica de la Almunia de Doña Godina

Estudio: Ingeniería de Datos en Procesos Industriales

2. Coordinadores del proyecto

Coordinador Luis Mariano Esteban Escaño

Correo electrónico lmeste@unizar.es

Departamento

Centro Escuela de Doctorado, Instituto Universitario de Investigación de Biocomputación y Física de Sistemas Complejos (BIFI), Escuela Universitaria Politécnica de la Almunia de Doña Godina

3. Resumen del proyecto

El trabajo cooperativo entre alumnos de diversas universidades es una herramienta de motivación que refuerza los contenidos impartidos en clase y

sobre todo ayuda a la asimilación de las competencias asignadas a las materias. En este proyecto se pretende llevar a cabo el estudio de un caso práctico en minería de datos con la colaboración de estudiantes de la Universidad de Zaragoza, CESI y Universidad Rey Juan Carlos. Son estudiantes de Ingeniería y Ciencia de Datos, Ingeniería Biomédica e Informática. El caso práctico es un reto que tiene aplicación en los servicios de Obstetricia y Ginecología de cualquier hospital. Se generarán datos sintéticos a partir de una muestra de pesos fetales obtenidos por ecografía de 9585 embarazos. Dichos datos se depurarán, modelizarán y validarán analizando su capacidad de predicción de pesos bajos o grandes para la edad gestacional (percentil de peso inferior a 10 o superior 90 respectivamente). También se generarán Apps para su uso práctico. Para conseguir el reto se crearán grupos de trabajo que colaborarán de forma cooperativa entre las universidades y que presentarán los resultados en una jornada final.

4. Participantes en el proyecto

Nombre y apellidos	Correo electrónico	Departamento	Centro
Cristina Soguero Ruiz	Cristina.soguero@urjc.es		
José Felipe Ortega Soto	felipe.ortega@urjc.es		
María Etelvina Escorihuela Sahún	mescori@unizar.es		Escuela de Doctorado, Escuela Universitaria Politécnica de la Almunia de Doña Godina
Ricardo Mariano Savirón Cornudella	rsaviron@gmail.com	Departamento de Anatomía e Histología Humanas	Facultad de Medicina , Escuela de Doctorado

5. Rellene, de forma esquemática, los siguientes campos a modo de ficha-resumen del proyecto

Otras fuentes de financiación sin detallar cuantía

No existen ningún tipo de financiación para este proyecto

Tipo de proyecto (Experiencia, Estudio o Desarrollo)

Experiencia

Contexto de aplicación/Público objetivo (titulación, curso...)

El proyecto recoge la experiencia de aprendizaje basado en proyectos de forma cooperativa entre las universidades Cesi ÉCOLE D'INGÉNIEURS en el grado de Ciencia de Datos, Universidad Rey Juan Carlos en el grado de Ingeniería y Ciencia de Datos y Universidad de Zaragoza en el grado de Ingeniería de Datos en

Procesos Industriales. En el caso de de Cesi, el proyecto se centraba en el campus de Pau, y se ubicaba en una asignatura de experiencia profesional, donde el propósito es desarrollar un caso práctico real, acercando al alumnado a la ejecución de proyectos en empresa o casos aplicados. En la Universidad Rey Juan Carlos, se centraba en 2º curso de Ingeniería y Ciencia de Datos como un reto motivador para aplicación de metodología en casos aplicados. Por último, en el grado de Ingeniería de Datos en Procesos industriales el proyecto se centró en la asignatura de Minería de Datos, de 3º curso, formaba parte de las actividades del curso y en el proceso de evaluación de la asignatura tenía un peso de un 20%.

Curso académico en que se empezó a aplicar este proyecto

El proyecto se ha aplicado en el curso 2023/2024.

Interés y oportunidad para la institución/titulación

El mayor interés radica en establecer un contacto entre el alumnado y profesorado de las universidades implicadas en el proyecto, de forma que se contrasten y complementen diferentes perspectivas de abordar un estudio práctico.

Métodos/Técnicas/Actividades utilizadas

Para la realización del proyecto, a lo largo del curso se han introducido los conceptos teóricos para el estudio que se iba a llevar a cabo y se han ido aplicando en el caso práctico. El caso práctico radica en construir un modelo de crecimiento fetal a partir de datos sintéticos generados sobre datos reales de estimación del peso fetal para distintas edades gestacionales. Se aplican modelos de regresión, se estima la varianza semanal, y se construyen intervalos de confianza para el peso estimado en cada semana de edad gestacional. Estos modelos también dependen de características materno-fetales y son validados sobre los datos del estudio, aproximadamente 6.000 partos. Una vez validados los modelos, se pueden visualizar los resultados mediante gráfico de dispersión o creando Apps en shiny para su uso práctico.

Los resultados fueron expuestos en una jornada que se denominó “International meeting on Data mining 2024” celebrated on April 24, 2024 donde se confrontaron metodologías y resultados seguidos por el alumnado de cada universidad.

Tecnologías utilizadas

Se utilizó un soporte Moodle para centralizar los trabajos, y modelos de regresión lineal o regresión por cuantiles para el análisis de los datos. También se generaron Apps mediante el uso de shiny. Todos los análisis fueron realizados con el lenguaje de programación R version 4.2.2 (The R foundation for Statistical computing). Se impulso el trabajo colaborativo entre universidades mediante el análisis del caso práctico.

Tipo de innovación introducida: qué soluciones nuevas o creativas desarrolla

Fundamentalmente, la innovación consiste en crear una red de trabajo colaborativo entre las tres universidades. Sobre el alumnado, se ha propuesto una actividad motivacional, para el caso de la universidad francesa es algo que esta planteado en el propio plan de estudios, es necesario realizar actividades de carácter aplicado y donde establezcan cooperación internacional. Éste es un buen ejemplo a copiar, y en la práctica es lo que se ha hecho, incorporar en la asignatura de Minería de Datos del grado de Ingeniería de Datos en procesos industriales la resolución de un caso práctico basado en datos reales y con todos los problemas del análisis de datos reales, datos incoherentes o repetidos, modelización no perfecta o necesidad de validación. También es importante la realización de una jornada de puesta en común de resultados entre las universidades, ha resultado importante para reflexionar sobre los resultados y la posibilidad de afrontar el estudio desde otras perspectivas.

Impacto del proyecto

El proyecto ha tenido problemas en su realización porque no se ha podido realizar en todas las universidades en el mismo período temporal, y el grado de conocimientos teóricos no era el mismo en todas las universidades. En la Universidad Rey Juan Carlos, solo se ha desarrollado hasta el 2º curso la titulación de Ingeniería y Ciencia de Datos, esto ha reducido su aplicación a un reto práctico propuesto para estudiantes sin ubicarlo en una asignatura. En el caso de la Universidad CESI han participado 3 grupos, 9 participantes, que desarrollaron el proyecto en 2 meses entre marzo y abril. En la Universidad de Zaragoza, fueron 9 grupos y 22 participantes y el proyecto se desarrolló en el primer cuatrimestre.

Características que lo hacen sostenible

Se ha establecido un contacto entre las universidades que estimamos será sostenible en el tiempo, aunque el planteamiento es posible que se adapte para ser mas similar en las 3 universidades. El reto se focalizará mas al planteamiento

de la universidad francesa, un caso práctico más focalizado en el tiempo y relacionado con asignaturas relacionadas con la ciencia de datos.

Posible aplicación a otras áreas de conocimiento

Aunque no es descartable, es difícil establecer la extensión a otras áreas de conocimiento.

6. Contexto del proyecto

Necesidad a la que responde el proyecto, mejoras obtenidas respecto al estado del arte, conocimiento que se genera.

La primera necesidad a la que responde el proyecto es trabajar en entornos de aplicación real en la modelización de datos. Las bases de datos utilizadas con frecuencia corresponden a casos donde se han depurado previamente los datos y ya han sido analizado con diversos modelos cuyo funcionamiento es casi perfecto, y quizá alejado de los problemas que hay que afrontar en el tratamiento de datos real. Para acercarse a un problema real se proporciono datos generados sintéticamente a partir de datos de crecimiento fetal pero sin depurar y se proporciono información sobre la metodología a utilizar para el caso de estudio. También es importante que el caso práctico se afronte con una metodología de trabajo en equipo y con la responsabilidad de confrontar los resultados con otros equipos de trabajo. Aunque esto puede ser un ejercicio dentro de la clase fue importante hacerlo de una manera ciega, de forma que el alumnado recibía formación en las distintas universidades de forma independiente, y luego se ponía en común los resultados.

7. Objetivos iniciales del proyecto

Qué se pretendía obtener cuando se solicitó el proyecto.

Se pretendía el aprendizaje a través del estudio de un caso práctico, fomentando el trabajo en equipo y generando información de forma colaborativa. También era un objetivo exponer, confrontar y aprender de los resultados obtenidos por los distintos equipos de trabajo en entornos multilingües.

8. Métodos de estudio/experimentación y trabajo de campo

Métodos/técnicas utilizadas, características de la muestra, actividades realizadas por los estudiantes y el equipo, calendario de actividades.

Se generaron 10 bases de datos con tamaño muestral 6000 a partir de 9585 datos reales de crecimiento fetal. De forma independiente se crearon 12 grupos de

trabajo que tenían que construir un modelo de crecimiento, estimar los percentiles de crecimiento para diferentes edades gestacionales, y validar los resultados estimando los porcentajes reales que aparecían por encima o debajo de los percentiles estimados. En la Universidad de Zaragoza, el trabajo se realizó desde octubre a diciembre, en la Universidad CESI desde marzo a abril y en la Universidad Juan Carlos I se hizo la propuesta en mayo aunque a fecha de hoy está sin resolver. Los resultados se expusieron de forma conjunta entre el grupo de Ingeniería de Datos de UNIZAR y de Ciencia de Datos de CESI el 24 de abril del 2024 con la participación de 3 grupos de estudiantes de CESI y 4 grupos de UNIZAR. Tras la exposición de los resultados en inglés de cada uno de los grupos, se establecían preguntas y comentarios entre los participantes. Tras la realización del encuentro, todos los estudiantes recibieron un diploma de participación en el evento firmado por ambas entidades.

9. Conclusiones del proyecto

Conclusiones: lecciones aprendidas, impacto.

La eficacia del proyecto ha estado condicionado por la programación de los cursos en las tres entidades participantes, ha sido satisfactoria para los estudiantes de CESI y UNIZAR, y con menor éxito en URJC. Lo ideal sería haber podido participar conjuntamente de forma simultánea, aunque esto finalmente no ha sido posible. En cualquier caso, si ha existido un trabajo cooperativo entre los estudiantes de CESI y UNIZAR, el alumnado de UNIZAR quizá era más conocedor del caso práctico y ha podido explicar a los participantes de CESI como modelar y validar los datos, mientras que desde CESI se han construido aplicaciones prácticas o Apps que han sido explicadas con detalle en sus exposiciones. El principal impacto del proyecto es el desarrollo de competencias transversales, como el trabajo en entornos multilingües, trabajo en equipo, y la importancia de exposición de resultados y debate en común para la mejora continua.

10. Continuidad y Expansión

Transferibilidad (que sirva como modelo para otros contextos),

Sostenibilidad (que pueda mantenerse por sí mismo), Difusión realizada .

El proyecto esperamos que tenga continuidad el próximo curso, estableciendo el “International meeting on data mining” como un evento con continuidad. La transferibilidad es limitada aunque podría extenderse a otras materias dentro del campo de Ingeniería y Ciencia de datos.

11. Resultados del proyecto indicando si son acordes con los objetivos planteados en la propuesta y cómo se han comprobado

Método de evaluación, Resultados.

Los resultados pueden clasificarse como satisfactorios aunque mejorables. En el caso del grado de Ingeniería de datos de UNIZAR y Ciencia de Datos de CESI son plenamente satisfactorios, puesto que se ha alcanzado el objetivo de estudio de un caso práctico abordado por ambas entidades y con puesta en común por parte del alumnado. Todos los estudiantes participantes alcanzaron una calificación positiva en la evaluación del mismo que fue incorporada de distinta forma a la evaluación de las asignaturas de Minería de Datos o Ciencia de Datos. En el caso de la Universidad Rey Juan Carlos no pudieron participar de la puesta en común de resultados.

En cuanto a los resultados reales del estudio, todos los participantes modelizaron los datos de una forma adecuada, siendo un análisis de regresión por cuantiles el que alcanzo el mejor ajuste de los datos, un estudio que correspondió a alumnas de UNIZAR, obteniendo aproximadamente un 10% de datos por encima del percentil 90 o por debajo del percentil 10, pero todos los grupos obtuvieron buenos resultados en cuanto a capacidad de ajuste de los datos. En este sentido fue muy interesante la exposición que se hizo de los modelos, donde cada grupo detallo como había depurado los datos, estimado el modelo de crecimiento y realizado la validación, estableciéndose un interesante debate entre los grupos de ambas universidades.

En cualquier caso, todos los participantes fueron calificados con notas iguales o superiores a Notable en el análisis del caso práctico.