



Becas colaboración curso 2020/2021

Fecha: 19 Junio 2020

Vicerrectorado de Investigación, Innovación y Transferencia

Subcomisión de I+D+i

Propuesta del departamento ORGANIZACIÓN DE EMPRESAS

Núm Proyecto: 2020/28/00002

Responsable

Fernández Diego, Marta

E-mail

marferdi@omp.upv.es

Ext.

76850

Responsable

González Ladrón de Guevara, Fernando Raimundo

E-mail

fgonzal@omp.upv.es

Ext

76823

Título proyecto

Descripción y evaluación de técnicas de imputación para la mejora del desempeño de modelos de estimación de esfuerzo en proyectos software tradicionales y desarrollados con metodologías ágiles

Valoración proyecto

4

Descripción proyecto

Existe un creciente interés en ingeniería de software en utilizar técnicas y mecanismos de imputación que permitan mitigar las consecuencias del problema planteado por la existencia de datos perdidos, no observados, inválidos, desconocidos o nulos. Estos datos ausentes suponen una pérdida en la capacidad, calidad y desempeño de modelos de estimación basados en repositorios. Las técnicas de imputación de valores perdidos permiten mitigar la problemática que estos datos originan.

Entre las técnicas de imputación más utilizadas destacan aquellas basadas en los valores de la media aritmética y la mediana, así como los valores más próximos al valor faltante de acuerdo con la distancia euclídea (método de imputación K-nearest neighbours K-NN). La imputación múltiple es una alternativa a los procedimientos tradicionales de tratamiento de los valores perdidos que permite incorporar el máximo de información, sin perder representatividad de la base imputada respecto a la original y manteniendo el grado de variabilidad de todas las variables involucradas en la imputación.

El objetivo de este proyecto es realizar una comparación de dichas técnicas de imputación y una valoración de su impacto en algunos de los métodos de estimación de esfuerzo en ingeniería del software más utilizados (como son los métodos basados en técnicas de regresión empleados como *baseline*; para métodos basados en técnicas de *machine learning*). En particular, se quiere identificar las implementaciones existentes de imputación múltiple en Python y/o R para mejorar el desempeño de los métodos de estimación utilizando el repositorio de proyectos de software ISBSG (International Software Benchmarking Standards Group).



Becas colaboración curso 2020/2021

Fecha: 19 Junio 2020

Actividades a realizar por el alumno

- Descripción de técnicas de imputación utilizadas en repositorios públicos de proyectos de software;
- Comparación de dichas técnicas en modelos de estimación de esfuerzo;
- Identificación, análisis y valoración de implementaciones de imputación múltiple en Python y/o R;
- Aplicación de la imputación múltiple en la base de datos ISBSG para evaluar su desempeño en modelos de estimación de esfuerzo;
- Validación del procedimiento diseñado en modelos de estimación basados en machine learning; para proyectos desarrollados con metodologías ágiles.

Horario

Mañanas: 10:00-13:00h