

**FICHA SESIÓN TÉCNICA:****Uso de técnicas de machine learning para la producción de estadísticas oficiales****Fecha: 31 de marzo de 2022****Plazo presentación de instancias:** del 16 al 28 de marzo**Lugar de realización:** Avda. Manoteras, 52, planta 6 aula 06**Modalidad:** híbrida (presencial y online con plataforma webex)**Duración:** 2 horas**Horario:** 12:00 a 14:00**Contenido de la sesión:**

Se presentan en esta sesión algunas primeras experiencias pilotos en el uso de técnicas de *machine learning* para la ejecución de ciertas tareas del proceso de producción de estadísticas oficiales. Tras una breve introducción sobre cuestiones fundamentales relativas a la inferencia, se describirán dos casos de uso en operaciones estadísticas concretas producidas y difundidas por el Instituto Nacional de Estadística. Por una parte, se presenta el empleo de *random forests* para la depuración de la variable ocupación de la Encuesta Europea de Salud 2017, que ha permitido una detección más eficiente de errores de medida ajenos al muestreo. Por otra parte, se muestran los resultados preliminares del proyecto SBS-DIRCE, en el que se predicen los valores de las variables del cuestionario de la Estadística Estructural de Empresas de todas las unidades del DIRCE a partir de la información recibida de registros administrativos. Se referencian algunos otros proyectos piloto que indagan el uso de estas técnicas para modernizar e industrializar el proceso estadístico de producción.

Dirigido a:

Personal del INE y de los servicios de estadística de las administraciones públicas y usuarios interesados en la modernización del proceso de producción estadística.

Dinámica de la sesión:

La sesión comenzará con la exposición del tema, seguida de un debate e intercambio de opiniones entre ponentes y asistentes.

Ponentes:

David Salgado, Dpto. Metodología y Desarrollo de la Producción Estadística, INE

Luis Sanguiao, Dpto. Metodología y Desarrollo de la Producción Estadística, INE