

Obligatorio

Descripción del caso

La pobreza es un problema socioeconómico que afecta a millones de personas en todo el mundo, incluyendo a parte de la población de Uruguay. La comprensión de este fenómeno y la capacidad de identificar personas y hogares en situación de pobreza son aspectos fundamentales para diseñar políticas efectivas de reducción de la pobreza y promover la igualdad de oportunidades.

En nuestro país, la clasificación de pobreza se realiza mediante el método del ingreso ([ver INE, 2006](#)). Entre varios desafíos metodológicos, la aplicación de este método mediante encuestas trae aparejado el problema de la calidad de las respuestas sobre ingresos - y más en general, sobre variables monetarias, que incluyen también preguntas sobre gastos de consumo - debido a que las personas, mundialmente, muestran mayor reticencia a contestar acerca de esas variables, o tienden a la subdeclaración en caso de contestarlas.

Una alternativa para este problema es la aplicación de técnicas de machine learning de clasificación con el objetivo de identificar la situación de pobreza de las personas a partir de sus características socioeconómicas, sin considerar las variables correspondientes a ingresos y gastos. Planteado como un problema de aprendizaje supervisado, se cuenta con la clasificación de hogares y personas en situación de pobreza a partir del método del ingreso (variable objetivo o respuesta): se puede utilizar la Encuesta Continua de Hogares en Uruguay 2022 (ECH 2022) como fuente de datos para desarrollar un modelo de clasificación capaz de predecir la situación de pobreza de las personas en base a sus características socioeconómicas, exceptuando las variables correspondientes a ingresos y gastos (conjunto de atributos o variables independientes).

El conjunto de datos de la ECH 2022 incluye a la variable de respuesta (denominada “pobre”, descrita como “Pobreza según metodología 2006”) y variables correspondientes tanto a las personas como a los hogares que integran, las que se utilizarán en este trabajo como características de entrada para el modelo de clasificación. **Adjunto a este documento se entrega una copia del diccionario de variables de la ECH en la que se indica explícitamente cuáles variables NO se pueden utilizar en el desarrollo del modelo.**

El proceso de desarrollo del modelo incluirá pasos como la exploración y preparación de los datos, la selección de características relevantes, la construcción y entrenamiento del modelo utilizando técnicas de aprendizaje automático, y la evaluación del rendimiento del modelo utilizando las métricas de clasificación que considere pertinente.

El resultado final será un modelo capaz de predecir la situación de pobreza de las personas en Uruguay utilizando los datos disponibles en la Encuesta Continua de Hogares como una alternativa al método del ingreso, es decir, sin considerar variables de ingresos y gastos.

Objetivo

Desarrollo de un modelo de clasificación para predecir la **situación de pobreza de las personas** en Uruguay utilizando datos de la [Encuesta Continua de Hogares en Uruguay 2022](#). El objetivo es construir un modelo que pueda clasificar correctamente a las personas en dos categorías: pobre y no pobre, en función de las características socioeconómicas recopiladas en la encuesta.

Tareas

Para ello, deberá seguir los siguientes pasos:

1. Cargar los datos de la encuesta continua de hogares de Uruguay del año 2022 (**ECH_2022.csv**).
2. Realizar un análisis exploratorio de los datos, incluyendo la visualización de las variables y su distribución, matriz de correlación, la detección e imputación de valores faltantes, outliers y la identificación de posibles variables predictoras.
3. Dividir los datos en un conjunto de entrenamiento y un conjunto de prueba utilizando la técnica de holdout.
4. Implementar diferentes algoritmos de *machine learning* correspondientes a las técnicas presentadas en el teórico y/o el práctico durante el curso, para predecir la variable objetivo.
5. Investigar y aplicar la técnica de stacking para combinar los modelos y mejorar la precisión de la predicción.
6. Evaluar la precisión del modelo utilizando diferentes métricas correspondientes a problemas de clasificación
7. Presentar los resultados en un informe técnico, incluyendo la descripción del modelo, los resultados de las métricas y las conclusiones y recomendaciones. Se adjunta una guía con recomendaciones para la presentación del informe.

Entrega

La entrega consistirá de:

1. Un informe detallado de los resultados obtenidos y decisiones tomadas.
2. Una jupyter notebook con toda la implementación realizada siguiendo todas las tareas descritas anteriormente.

La evaluación del trabajo tendrá en cuenta el **cumplimiento de todos los puntos de la tarea**, la **exhaustividad del trabajo** realizado, la **calidad del informe** y el **desempeño obtenido** en los resultados de test.