

IA Aplicada: Machine Learning y Deep Learning para Decisiones

Estratégicas

Se entrega el siguiente set de datos correspondientes a una encuesta nacional de hogares y de salud familiar realizada por parte de una ONG asociada al gobierno de turno, que tenía como principal objetivo categorizar a los habitantes de acuerdo a su ingreso total para el estudio del indicador de % pobreza, **CensoPoblacion.xlsx**, en el cual encontramos información sociodemográfica y del nivel educativo de los habitantes tales como: la edad, el capital ganado y el capital perdido, las horas trabajadas a la semana y más.



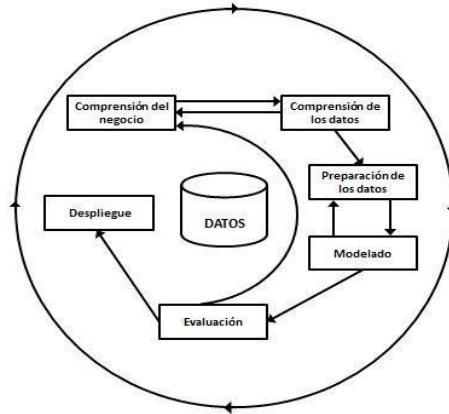
El principal objetivo de la siguiente problemática es tratar de encontrar poblaciones vulnerables para de cara a los próximos proyectos de inversión y políticas públicas priorizar ciertos grupos de mayor necesidad de acuerdo con sus características económicas/sociales y así mitigar en medida de lo posible el % de pobreza. Se requiere que los ayude a **PREDECIR LA PROPENSIÓN DE LOS NIVELES DE INGRESOS DE LOS HABITANTES**.

- Las variables que se disponibilizan son:

Variable	Descripción de la Variable
EDAD	Edad de la persona.
CAPGANADO CAPPERD	Ganancia de capital o disposición de valores, bonos o inmuebles.
HORASEMANA	Perdida de capital o disposición de valores, bonos o inmuebles.
EDUCACIONNUM	Horas a la semana trabajadas por la persona en su trabajo dependiente.
EDUCACION	Número de años dedicados al estudio o educación. Nivel educativo de la persona.
ESTADOCIV	Estado civil de la persona.
SEXO	Género o sexo de la persona.
INGRESO	Ingreso de la persona asociado a su trabajo dependiente e independiente.

Entregables del Laboratorio:

Bajo la metodología CRISP – DM, construir un modelo o prototipo de regresión lineal, respondiendo las siguientes preguntas:



- 1.- Definir el problema de la naturaleza que se tiene a continuación, además de los objetivos de negocio bien definidos. **(02 puntos)**
- 2.-¿Qué tipo de variables se utilizan en el problema de negocio? **(01 punto)**
- 3.- Realizar un análisis exploratorio de los datos para entender un poco las variables con las cuáles queremos resolver el problema. **(02 punto)**
- 4.- Para la información disponible, dividir los datos en muestras para entrenamiento y validación, con el objetivo de evaluar los resultados obtenidos. **(01 punto)**
- 5.- Con el objetivo de realizar la estimación o pronóstico de los niveles de ingresos de los usuarios o personas, construya modelos o prototipos de regresión lineal, comparándolos e interpretando los coeficientes del mejor modelo para la explicación a las áreas de negocios. **(10 puntos)**
- 6.- Validar con los datos de validación el modelo o prototipo de regresión construidos, que conclusiones podría dar respecto a las estimaciones observadas. Construya el % de acierto de los modelos de regresión a un nivel de control de calidad del 5%. **(04 puntos)**