

Introduction

Este trabajo fue realizado en el mes de octubre de 2020 sobre los datos de una serie de encuestas realizadas a través de la Dirección Nacional de Estadísticas Sociales y de Población y de la Dirección de Estadísticas Poblacionales en las localidades urbanas de 5.000 y más habitantes de todo el territorio nacional, las entrevistas fueron directas a través de dispositivos digitales.

Dicha base permite obtener estimaciones sobre la población con algún tipo de discapacidades a nivel nacional y tiene como finalidad servir a todas aquellas personas, instituciones gubernamentales, empresas o centros de investigación que requieran información sobre la temática abordada

Datasets

Las features principales que se consideraron fueron los siguientes:

- Cobertura de salud
- Tipo de beneficio que recibe
- Sabe leer y escribir
- Causa de la dificultad
- Máximo nivel educativo alcanzado
- Condición de actividad.
- Motivo principal por el que no busco trabajo
- Cantidad de personas en el hogar
- Cantidad de personas con dificultad en el hogar
- Tipo de hogar
- Grupos de edad
- Cantidad de dificultades
- Cantidad y tipo de dificultad

Metodos

Logistic Regression

Es un clasificador lineal compuesta por una regression lineal, precedida de una función activación sigmoide, por lo cual el output es binario y no continuo. La probabilidad de la clase Yi viene dada por la siguiente función:

$$P(y_i | x) = \sigma(wtx)$$

La función sigmoide $\sigma = \frac{1}{1 + \exp(-x)}$

K-Nearest neighbors

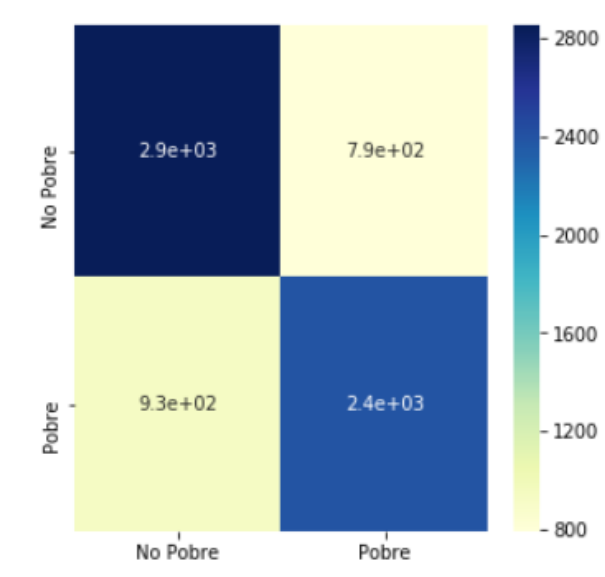
Es un modelo de clasificación en el cual un nuevo dato es agrupado según K vecinos más cerca de uno que del otro según la distancia:

SVM

Se trata de un clasificador lineal, el cual busca un hiperplano que máxima el margen entre las clases. En el caso de que las clases no sean linealmente separables se acude al Soft Margin, un penalizador de muestras mal clasificadas, las cuales se penalizan con un Costo seleccionado por el usuario.

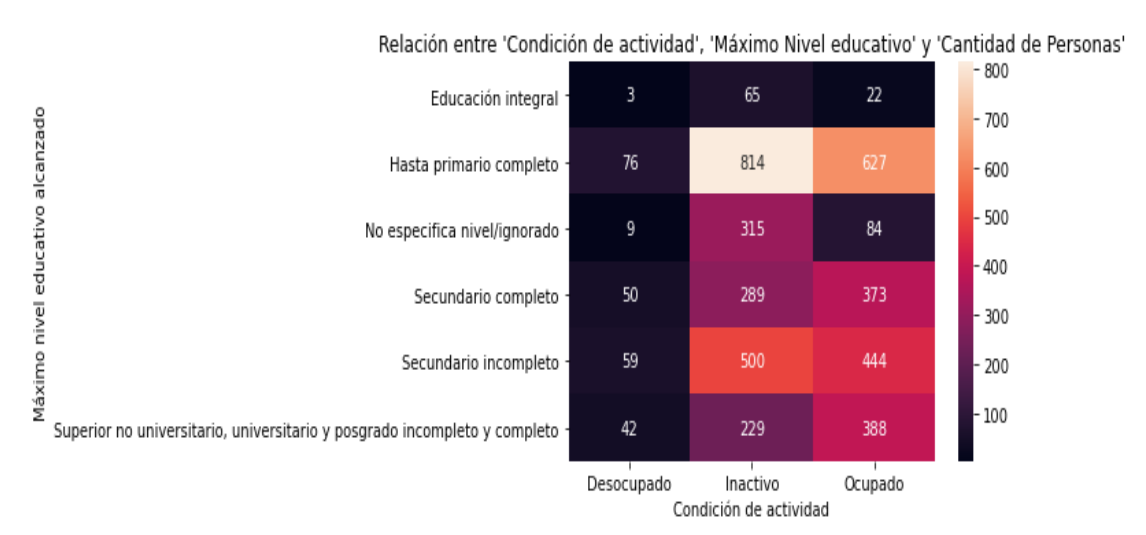
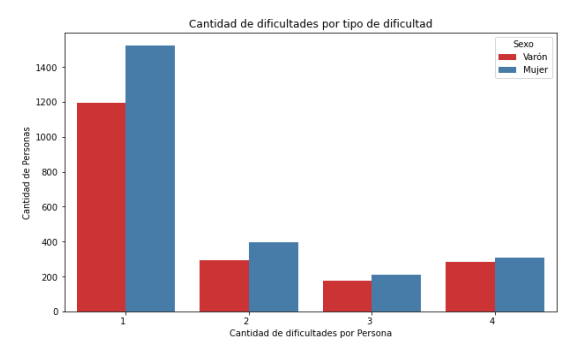
Resultados

Models	Metrics		
	Accuracy in %	Recall in %	Precision in %
LR	62,41%	71,92	74,99
KNN	73,32%	70,02	72,37
SVM	64,00%	67,57	76,98



Analysis Exploratorio de datos

En el preprocesamiento de dato, analizamos distintos factores que pueden llegar a afectar la condición de actividad laboral.



Conclusiones

Este trabajo nos permitió adquirir y profundizar el uso de herramientas como python.

A la vez nos permitió implementar conceptos de machine learning en aspectos de interés social.

Obtuvimos resultados concretos sobre el comportamiento de cada persona y sus problemáticas en la sociedad tomando en cuenta su ocupación, los niveles de estudio alcanzado, la vivienda, entre otros factores de básicos que determinan condiciones de vida.