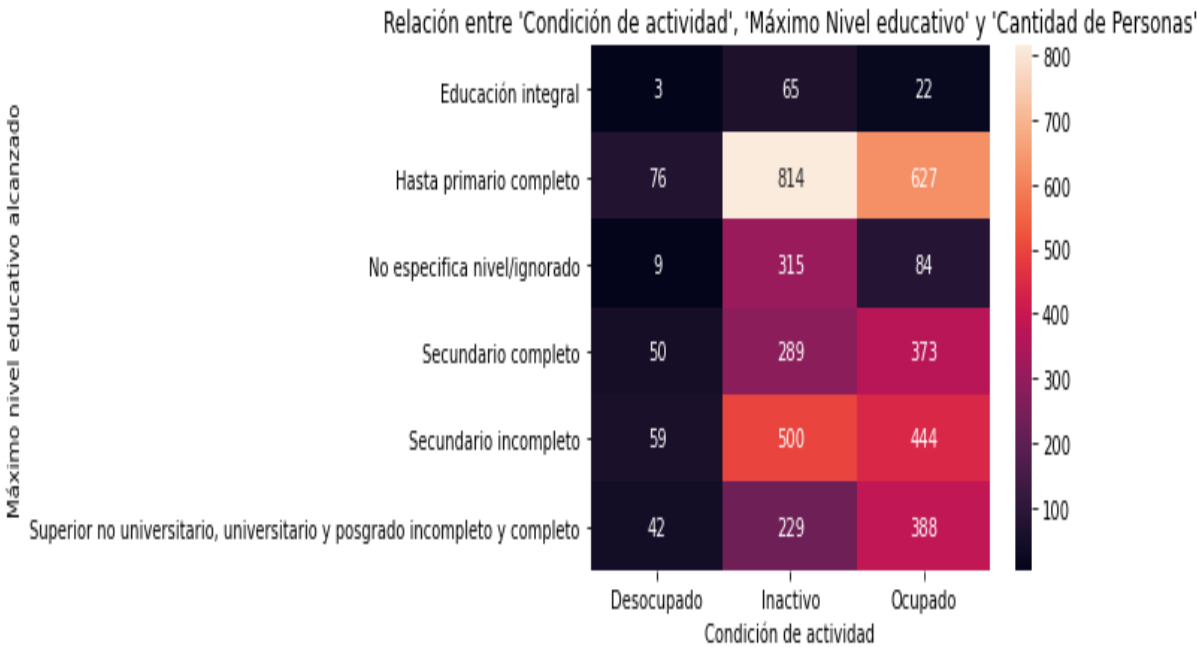


OBJETIVOS

Objetivo específico:

En el siguiente trabajo utilizamos distintos modelos de clasificación con datos obtenidos del INDEC con el fin de clasificar a las personas con falta de capacidad n su condición ocupacional.

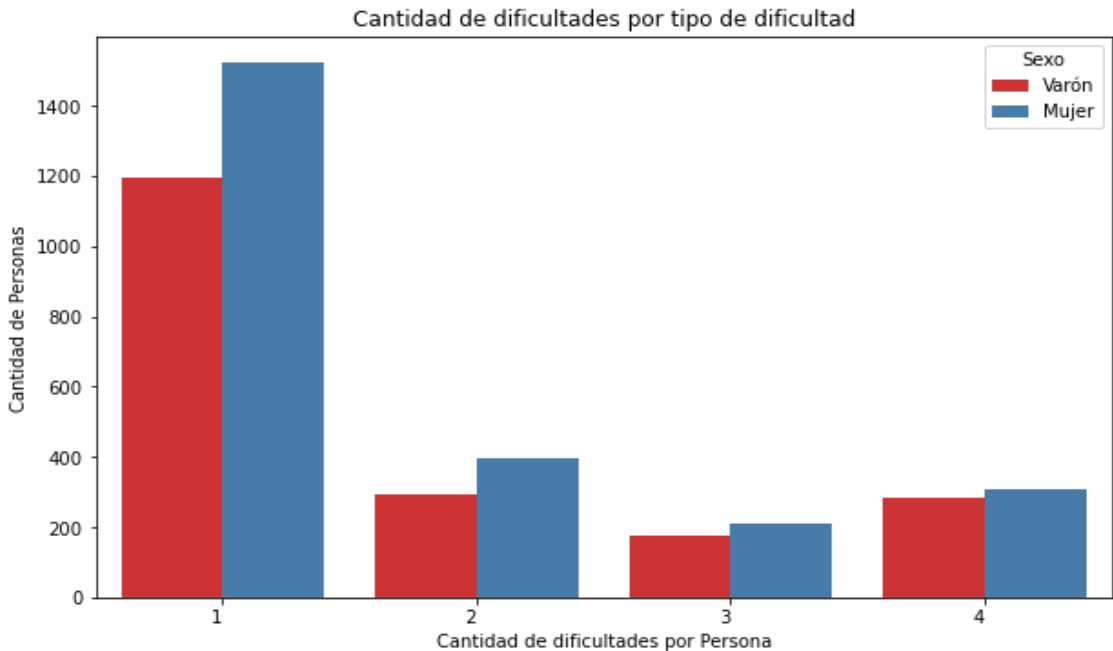


DATA & PREPROCESSING

Las features principales que se consideraron fueron los siguientes:

- 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
 - Edad de inicio de la dificultad
- 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

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



Modelos de Clasificación

Utilizamos tres modelos de clasificación supervisada distinta:

- 1) Logistic Regression
- 2) K-Nearest neighbors
- 3) 3)SVM Logistic R

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 Y_i 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:

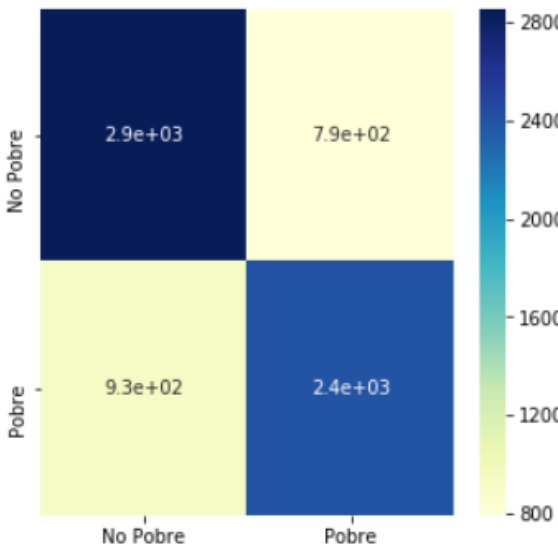
$$d_E(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

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

Confusion Matrix



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

Conclusión

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.

Futuras Mejoras

Incorporación de más samples con personas con distintas discapacidades, en especial en edades laboralmente activas, para la mejor categorización.