

## ¿Por qué el reto LUCA?

Hemos elegido el reto LUCA porque estaba íntimamente ligado con las áreas de conocimiento de los miembros del grupo. El hecho de ser un reto que requiere trabajo multidisciplinar nos motivó para tratar de lograrlo. Además, contamos con una compañera colombiana que conoce extraordinariamente la realidad de este país.

## ¿Por qué este modelo?

La visualización de los datos analizados sientan un precedente de cómo la población Colombiana desarrolla su actividad socio-económica en torno a su región, la realización de este panóptico descifra la capacidad productiva y de correlación frente aspectos que han influenciado en su habita, por lo tanto, deja al descubierto que en algunas regiones demandan una necesidad de contar servicios básicos ya sea por despoblación o por actividad laboral que se concentra en algún espacio del territorio nacional.

## ¿Cómo lo hemos hecho?

A través de datos abiertos de Colombia obtenidos de fuente locales como el DANE o internacionales como Data República hemos recopilado suficiente información como para ser capaces de estudiar la situación socioeconómica de Colombia. En primer lugar hemos realizado un preprocesamiento de los datos gracias a técnicas de *machine learning* de *dimensionality reduction* y *feature selection*. Gracias a esto hemos conseguido desarrollar unos índices que reflejan distintas aspectos socioeconómicos de Colombia.

Gracias a otras técnicas estadístico/matemáticas hemos sido capaces de estudiar las diferencias entre departamentos respecto a distintos aspectos socioeconómicos. La principal técnica usada ha sido el MANOVA. Esta herramienta nos permite extraer las diferencias significativas entre departamentos gracias a un complejo análisis de varianzas multivariante. Toda la información extraída se ha representado en mapas departamentales de Colombia para facilitar la visualización.

A través de los datos cedidos por Telefónica hemos elaborado un modelo de *machine learning* capaz de predecir el número de desplazamientos con origen y destino en los distintos departamentos colombianos. Este modelo posee la propiedad de ser dinámico y adaptativo. Significa que es capaz de adaptarse a *data sets* más completos pudiendo aprender de estos y por la tanto generar automáticamente sus propias decisiones y predicciones. Todo ello sin necesidad de ser explícitamente programado para ello.

El algoritmo utilizado es conocido como *Random Forest*. Este algoritmo genera una gran cantidad de *decision trees* individuales, cada uno de estos formados con una muestra aleatoria de variables y observaciones del *data set*. Cada *decision tree* genera independientemente su predicción y el algoritmo en su conjunto se queda con aquella predicción más elegida, teniendo en cuenta así distintas agrupaciones posibles de variables y de individuos.

Se ha realizado un estudio estadístico/matemático de los datos y se ha apreciado un patrón estacional en los desplazamientos entre departamentos. Se ha conseguido introducir este patrón gracias a la inclusión de retardos en los movimientos entre departamentos. La identificación de este tipo de patrones facilita enormemente la óptima modelización del proceso.

Se ha dividido el *data set* provisto por Telefónica en un *training set* con el que se ha entrenado al modelo y un *test set* con el que se ha evaluado el error cometido por el modelo a la hora de predecir los movimientos. A pesar de obtener buenos resultados, estos se pueden mejorar con una mayor cantidad de datos y con una potencia computacional mayor.

Como bien se ha dicho, la principal característica de este modelo es su adaptabilidad. Se puede mejorar enormemente alimentándole con más información, es decir, con una mayor cantidad de datos, ya que de esta forma es capaz de aprender mejor las relaciones que genera el proceso estudiado.

## ¿Es aplicable a la realidad?

Este análisis brindará una herramienta eficaz para ser utilizada como observatorio social para las nuevas proyecciones que se pretende alcanzar; Colombia está experimentando un cambio vertiginoso ya que ha apostado a la desmovilización de grupos al margen de la ley que se encontraban en algunos entes territoriales; con esta herramienta se permitirá ver el alcance de desarrollo como desempeño de la Industria, la agricultura, movilidad entre departamentos, datos demográficos después de este proceso político social.

## ¿Qué hemos descubierto?

El análisis de datos nos han arrojado resultados que descubren donde se puede acentuar la brecha digital y el uso que se hace de la telefonía móvil en algunos departamentos de la república de Colombia. Al ser estos tiempos un indicador

importante para observar factores como:

- . Población que demanda Infraestructura
- . Desarrollo de competencias digitales en la población
- . Cambio cultural frente al uso y fomento de la información
- . Necesidad de gestión, apropiación de procesos de implementación de recursos tecnológicos.