

# Análisis de la calidad de vida en Colombia

Juan Ángel Castellanos  
Iván Rodrigo Castillo  
Gabriela Fernanda Vega

Prof. Gustavo Garzón



Ingeniería de Sistemas  
03 de Aril de 2020

## Descripción

Las encuestas de calidad de vida surgen como respuesta a la necesidad de caracterizar la población en los diferentes aspectos involucrados en el bienestar de los hogares. Al conocer la situación actual de nuestro país Colombia y teniendo en cuenta la falta de información por parte de los medios periodísticos, se decidió hacer el análisis de la calidad de vida que presenta el país.

## Dataset

Usaremos el data set proporcionados por el DANE en la encuesta llamada "Encuesta nacional de calidad de vida 2018" de esta manera hacer una estimación de los servicios básicos con los que cuentan los hogares colombianos. El dataset esta compuesto por 35 variables y 88713 registros.

## Clasificación

Clasificamos los datos usando una lista de Ground truth para identificar los servicios con los que cuenta cada estrato dependiendo de las viviendas donde se habita y las diferentes características en cuanto a calidad de vida que representan la población colombiana.

## Tipos de Clasificadores

Los tipos de clasificadores y los parámetros experimentados son :

1. RFC (Random Forest Classifier)

$n_{estimators} = 100$

$max_{depth} = 50$

2. SVM (Support Vector Machine)

$kernel = 'rbf'$

3. DTC (Decision Tree Classifier)

$max_{depth} = 10$

## Estrato

El estrato de vivienda nos puede brindar un indicativo más claro de la calidad de vida que tienen los hogares colombianos, sus opciones son: Bajo - Bajo, Bajo, Medio - Bajo, Medio, Medio - Alto, Alto, Planta eléctrica, No conoce el estrato o no cuenta con recibo de pago, Recibos sin estrato o el servicio es pirata

## Tipo de vivienda

El tipo de vivienda nos proporciona información pertinente para el análisis, los valores de esta columna son: Casa, Apartamento, Cuarto(s), Vivienda tradicional indígena, Otro tipo de vivienda

## Material del techo

El material del techo de una vivienda nos permite determinar las condiciones en las que viven los ciudadanos, si son aptas para obtener una buena calidad de vida. Dentro de las opciones que nos brinda la columna tenemos: Plancha de concreto, Tejas de barro, Teja de asbesto - cemento, Teja metálica, Teja plástica, Paja, palma, Material de desecho

# PCA

PCA es una técnica de Extracción de Características que nos permite combinar las entradas del dataset de una manera específica y eliminar algunas de las variables (características) "menos importantes" manteniendo la parte más importante todas las variables

- \* Se aplica PCA a la característica con mejor clasificación que se obtuvo, el cual es tipo de Vivienda.

- \* Se ajusta el PCA con cada clasificador (RF, SV y DT)

```
-----RESULTADOS-----  
PCA CON 10 COMPONENTES  
  
Random Forest Classifier:  
TRAIN accuracy: 0.941  
TEST  accuracy: 0.800  
  
Support Vector Machine:  
TRAIN accuracy: 0.805  
TEST  accuracy: 0.800  
  
Decision Tree Classifier:  
TRAIN accuracy: 0.829  
TEST  accuracy: 0.801
```

# Resultados

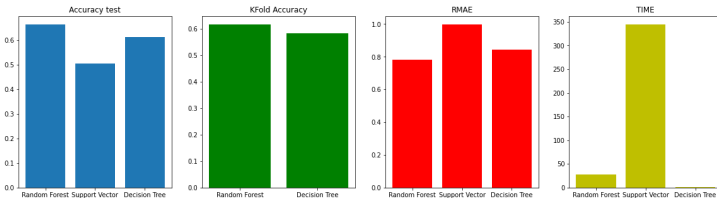


Figura 2: Resultados con Ground Truth: Estrato

Cuando se toma como ground truth el estrato, debido a los datos que contiene el dataset podemos observar que el accuracy se encuentra en un rango de 0.5 a 0.7 lo cual podemos considerarlo como un resultado aceptable para nuestro trabajo de clase, sin embargo, podemos notar que el clasificador que mejor se comporta en este caso es el Random Forest teniendo en cuenta su precisión y su poco tiempo de ejecución a comparación de los otros dos clasificadores.

# Resultados

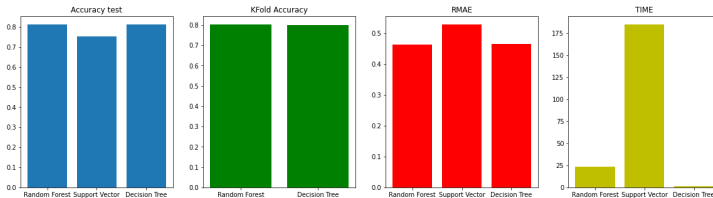


Figura 3: Resultados con Ground Truth: Tipo de vivienda

Cuando se toma como ground truth el tipo de vivienda podemos notar que el accuracy de los clasificadores ha aumentado de una manera significativa teniendo en cuenta que esta vez los resultados rondan el rango de 0.75 a 0.81 y notando que los tres clasificadores utilizados se comportaron de una manera muy similar esta vez siendo el ganador el Decision tree al tener el menor tiempo de ejecución y el mejor accuracy.



# Resultados

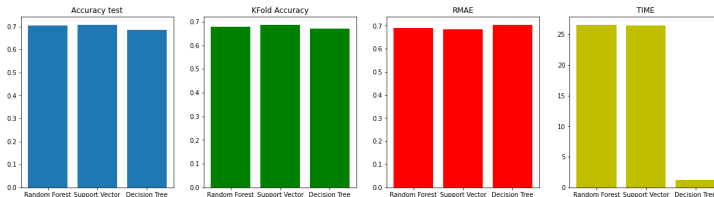


Figura 4: Resultados con Ground Truth: Material del techo

Por ultimo al tener como ground truth uno de los materiales de la vivienda en este caso el techo podemos ver que los tres clasificadores tienen resultados muy similares a excepción del tiempo de ejecución el cual podemos ver que el decision tree es mucho menor en comparación.

GRACIAS!!!