

Abstract

In this work, machine learning was applied: supervised and unsupervised, to a set of data on divorces in Ecuador. It began by defining what you want to predict and identifying who would contribute this result, then the process of data collection and preparation was carried out, joining the database of 2017 and 2018, forming a database with 49 variables of which only 12 variables were considered to work, the textual values were replaced by integer numerical values, in order to optimize them, likewise, the data was divided both for training and for the tests, randomly choosing 80% and 20%, finally the algorithms of Decision Tree and Random Forest were used for prediction, K-means and Cobweb to perform the grouping using RapidMiner, and WEKA, obtaining that adultery is the main cause of divorce in the country.

1. Problem



In Ecuador in recent years there has been a significant increase in the number of divorces, for this reason the interest is born to know: *¿What will be the main cause of divorce in Ecuador?*



**SUPERVISED
LEARNING**



**NOT
SUPERVISED
LEARNING**

BANCO DE DATOS ABIERTOS



DIVORCIOS

<http://aplicaciones3.ecuadorencifras.gob.ec/BIINEC-war/index.xhtml>

2. Data collection

Estadísticas de Divorcios			
Año: 2017	Se han realizado 361 descargas para el año seleccionado.		
Seleccione el período:	ANUAL		
Datos Abiertos CSV	158.74 KB	Descargar	
Presentación de Resultados		Descargar	
Formularios	2.24 KB	Descargar	
Sintaxis	1.04 MB	Descargar	
Metodología	17.32 MB	Descargar	
Gráficos		Descargar	

Diccionario_EDV_2017.xls
EDV 2017.csv
Metadatos_EDV_2017.xls

Estadísticas de Divorcios			
Año: 2018	Se han realizado 223 descargas para el año seleccionado.		
Seleccione el período:	ANUAL		
Datos Abiertos CSV	1.42 MB	Descargar	
Base de Datos SPSS	1.63 MB	Descargar	
Tabulados y series históricas		Descargar	
Diccionario de variables	456.4 KB	Descargar	

Diccionario_EDV_2018.xls
EDV 2018.csv
Metadatos_EDV_2018.xls

3. Data preparation

EDV 2017.csv U EDV 2018.csv

prov_insc	anio_insc	dur_mat	cau_div
Napo	2017	17	Mutuo acuerdo
Chimborazo	2017	12	El abandono injustificado de cualquiera de los cónyuges por más de seis meses ininterrumpidos
Cotopaxi	2017	18	El abandono injustificado de cualquiera de los cónyuges por más de seis meses ininterrumpidos
Tungurahua	2017	2	Mutuo acuerdo
Tungurahua	2017	10	Mutuo acuerdo
Chimborazo	2017	24	El abandono injustificado de cualquiera de los cónyuges por más de seis meses ininterrumpidos
Cotopaxi	2017	36	Mutuo acuerdo
Zamora Chín	2017	2	El estado habitual de falta de armonía de las dos voluntades en la vida matrimonial
Santo Domingo	2018	11	El abandono injustificado de cualquiera de los cónyuges por más de seis meses ininterrumpidos
Carchi	2018	58	El abandono injustificado de cualquiera de los cónyuges por más de seis meses ininterrumpidos
Pichincha	2018	5	Por mutuo consentimiento vía judicial
El Oro	2018	52	El abandono injustificado de cualquiera de los cónyuges por más de seis meses ininterrumpidos

Dictionary

Nombre del Campo (Encabezado Columna)	CATEGORIA	SIGNIFICADO
anio_insc	1	2017
	2	2018
(cau_div)		Mutuo acuerdo
		Por mutuo consentimiento vía judicial
	1	Por mutuo consentimiento vía notarial

anio_insc	prov_insc	cau_div	edad_hom	niv_insth	p_etnica	ho_area	resh	edad_muj	niv_instm	p_etnica	mu_area	resm	dur_mat
1	17	1	24	2	2	1	23	2	2	2	1	2	
1	17	1	31	2	2	1	25	2	2	2	1	1	
1	5	5	57	2	2	1	47	2	2	2	1	2	
1	4	1	64	2	2	1	58	3	2	2	1	3	
1	18	4	31	2	2	0	33	3	2	2	0		
2	22	5	38	2	2	1	0	2	0	0	1	1	
2	3	5	80	1	0	1	0	1	0	0	1	5	
2	16	1	0	0	0	1	28	1	2	2	0		
2	6	5	0	0	0	1	72	1	2	1	1	5	
2	9	5	0	0	0	1	74	1	1	1	1		

Transformation

4. Data splitting

Training 80%



Test 20%

RapidMiner

RANGE

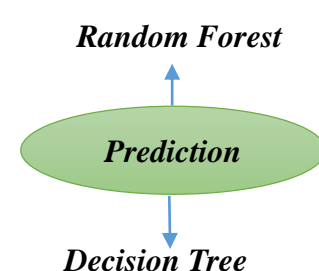
SAMPLE

sample percentage

Percentage: 80

APPLY

5. Training a model



Attribute	Full Data (54441.0)	0 (987.0)	1 (1287.0)	2 (1589.0)	3 (278.0)	4 (1908.0)
anio_insc	1.4715	2	2	2	2	1
prov_insc	11.0444	7.9504	17.6418	15.8615	9.554	16.0094
edad_hom	42.3265	40.9433	43.6131	41.2341	39.9101	40.1431
niv_insth	1.894	0.9392	0.9635	2.0554	1.3525	0.9156
p_etnica_hom	1.9685	1.9878	2.0373	1.9824	2.964	1.924
area_res	0.8727	1	1	0	1	1
edad_muj	39.3854	37.8521	40.3652	38.7231	37.277	36.9009
niv_instm	1.9142	2.2705	1.303	2.0673	1.3813	1.7794
p_etnica_muj	1.9652	1.9899	1.9899	1.9748	2.9892	2.0105
area_resm	0.8708	1	1	0	1	1
dur_mat	14.8157	12.9605	15.8081	14.4978	15.4317	12.7998
cau_div	3.1882	2.9524	2.5758	2.2731	3.6259	3.8056

6. Model validation

Predicción*100/ valor real

Decision Tree

Random Forest

prediction(c...	RENDIMIENTO	prediction(c...	RENDIMIENTO
3.227	322.6909536	3.240	64.80551049
3.227	80.67273841	3.240	324.0275524
3.227	64.53819072	3.240	324.0275524
3.227	322.6909536	3.240	81.00688811
3.227	322.6909536	3.240	81.00688811
3.227	64.53819072	3.240	81.00688811
3.227	322.6909536	3.240	81.00688811
3.227	64.53819072	3.240	81.00688811
3.227	322.6909536	3.240	81.00688811
3.227	322.6909536	3.240	81.00688811
3.227	80.67273841	3.240	81.00688811
3.227	80.67273841	3.240	81.00688811

Adultery of one of the spouses.

Conception or illegitimate birth.

3 Adultery

Repository link where all the work done to make the prediction of the main cause of divorces in Ecuador is located.

https://github.com/CecivelSolano/IA_Proyecto_Divorcios-2020.git