

Aprendizaje automático supervisado y no supervisado: caso de estudio "defunciones fetales" utilizando la base de datos abierta del "INEC - Ecuador".



Facultad de Energía, Carrera de Ingeniería en

Steeven Armijos-Bravo

Resumen

En el presente proyecto se aplicó 3 algoritmos de Aprendizaje Supervisado (Clasificación): RandomTree, RandomForest, J48, con el fin de predecir la principal causa que ocasiona las defunciones fetales en el año 2020, también se utilizó algoritmos de Aprendizaje No Supervisado: Clustering (SimpleKMeans), para agrupar a la enfermedad más probable que ocasiona las defunciones fetales de acuerdo a características comunes, y **Reglas de Asociación** (Apriori) para obtener reglas en base al conjunto de datos que ayuden a la toma de decisiones, por ejemplo en diagnósticos médicos. Para ello se realizó un proceso de definición del problema, recopilación de datos (obtenidos del Banco de Datos Abiertos del INEC, se descargó archivos .csv de años 2015-2018), **preparación de los datos** (se eliminó tildes, caracteres especiales, datos atípicos), **división de** datos (se aplicó el 80% para el entrenamiento y el 20% para el test), entrenamiento (se entrenó los 3 algoritmos de clasificación antes mencionados) y validación de los modelos (se recopiló los datos en una tabla con la precisión de predicción de cada algoritmo de clasificación). Finalmente se obtuvo que la principal causa de defunciones fetales es ocasionada por la enfermedad: Hipoxia intrauterina, no especificada.

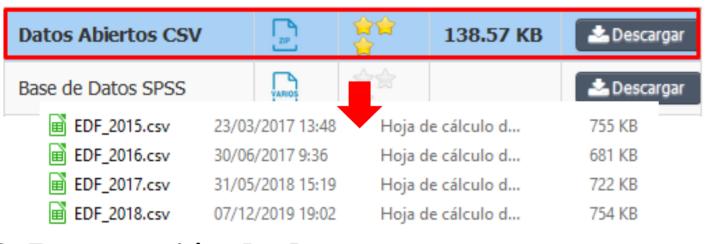
1. Problema

¿Cuál es la principal causa que ocasiona las defunciones fetales en el año 2020?

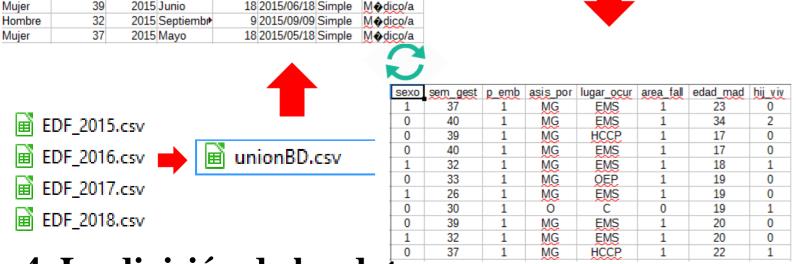


2. Recopilación de datos

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







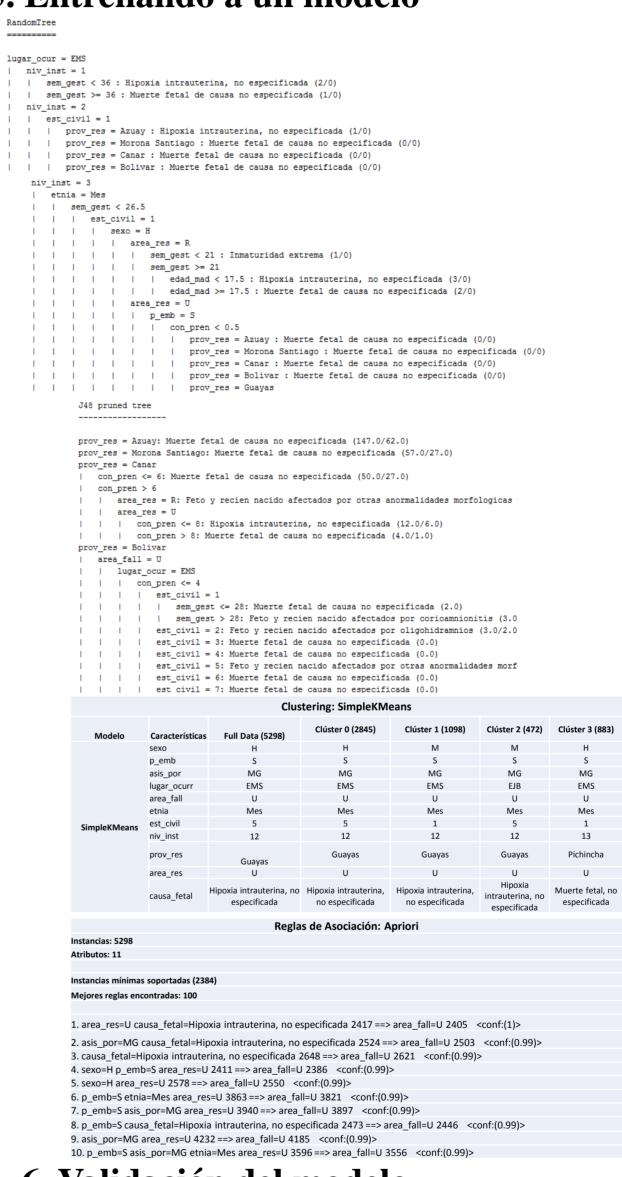
4. La división de los datos



Mujer

 Use training set Supplied test set More options.

5. Entrenando a un modelo





redictions on test spirt === RANOOMFOREC							
reductions on test split === Randomrorest		Correctos	Incorrectos	Correctos	Incorrectos	Correctos	Incorrectos
inst# actual predicted error prediction		563	497	461	599	539	521
1 3:'Hipoxia intrauterina, no especificada' 3:Hipoxia intrauterina, no especificada 0.755							
2 1:'Muerte fetal de causa no especificada' 1:Muerte fetal de causa no especificada 0.74		53.11%	46.89%	43.49%	56.51%	50.85%	49.15%
3 30:'Malformacion congenita, no especificada' 3:Hipoxia intrauterina, no especificada + 0.533							
4 3:'Hipoxia intrauterina, no especificada' 3:Hipoxia intrauterina, no especificada 0.92							
5 3:'Hipoxia intrauterina, no especificada' 3:Hipoxia intrauterina, no especificada 0.92							
Predictions on test split === RandomTree							
inst# actual predicted error prediction							
1 3:'Hipoxia intrauterina, no especificada' 3:Hipoxia intrauterina, no especificada	1						
2 1:'Muerte fetal de causa no especificada' 1:Muerte fetal de causa no especificada	1						
3 30:'Malformacion congenita, no especificada' 3:Hipoxia intrauterina, no especificada	+ 1						
4 3:'Hipoxia intrauterina, no especificada' 3:Hipoxia intrauterina, no especificada	1						
5 3: Hipoxia intrauterina, no especificada' 3: Hipoxia intrauterina, no especificada	1						
Predictions on test split === J48							
inst# actual predicted error prediction							
1 3:'Hipoxia intrauterina, no especificada' 3:Hipoxia intrauterina, no especificada	0.785						
2 1:'Muerte fetal de causa no especificada' 1:Muerte fetal de causa no especificada	0.612						
3 30:'Malformacion congenita, no especificada' 3:Hipoxia intrauterina, no especificada	+ 0.7	785					
4 3:'Hipoxia intrauterina, no especificada' 3:Hipoxia intrauterina, no especificada	0.785						
5 3:'Hipoxia intrauterina, no especificada' 3:Hipoxia intrauterina, no especificada	0.785						

RandomForest

Instancias = 1060: 80% de entrenamiento y 20% test