



# UNIVERSIDAD NACIONAL DE LOJA

## FACULTAD DE LA ENERGÍA, LAS INDUSTRIAS Y LOS RECURSOS NATURALES NO RENOVABLES

### CARRERA EN COMPUTACIÓN

#### FICHA INFORMATIVA DEL ALGORITMO

Carrera: Ingeniería en Sistemas "A"

Fecha: 6 de febrero de 2022

Materia: Inteligencia Artificial

Integrantes:

- Francisco Agreda Sánchez
- Israel Campoverde Peñaherrera
- Eduardo Leon Castillo
- Josué Ortega Jaramillo

**Cuadro Comparativo de las instancias correctamente clasificadas de las diferentes clasificaciones de WEKA..**

Algoritmos	Resultados			
Lazy (IBK)	Correctly Classified Instances	86	68.8	%
	Incorrectly Classified Instances	39	31.2	%
	Kappa statistic	-0.0016		
	Mean absolute error	0.0569		
	Root mean squared error	0.2095		
	Relative absolute error	100.017	%	
	Root relative squared error	134.8653	%	
	Total Number of Instances	125		
	Ignored Class Unknown Instances		41	
Trees(DecisionStump)	Correctly Classified Instances	104	83.2	%
	Incorrectly Classified Instances	21	16.8	%
	Kappa statistic	0.0724		
	Mean absolute error	0.0449		
	Root mean squared error	0.1545		
	Relative absolute error	78.8573	%	
	Root relative squared error	99.4817	%	
	Total Number of Instances	125		
	Ignored Class Unknown Instances		41	

Rules (OneR)	Correctly Classified Instances	75	100	%
	Incorrectly Classified Instances	0	0	%
	Kappa statistic	1		
	Mean absolute error	0		
	Root mean squared error	0		
	Relative absolute error	0	%	
	Root relative squared error	0	%	
	Total Number of Instances	75		
	Ignored Class Unknown Instances		91	
Bayes (BayesNet)	Correctly Classified Instances	96	76.8	%
	Incorrectly Classified Instances	29	23.2	%
	Kappa statistic	-0.0272		
	Mean absolute error	0.0495		
	Root mean squared error	0.1795		
	Relative absolute error	87.0363	%	
	Root relative squared error	115.5389	%	
	Total Number of Instances	125		
	Ignored Class Unknown Instances		41	

En la tabla anterior se puede comprobar cómo el algoritmo de Rules (OneR) tiene un porcentaje mayor de instancias correctamente clasificados. Por esta razón ha sido seleccionado para el posterior análisis de los datos.

### Algoritmo Machine Learning

OneR es un algoritmo de clasificación simple y muy efectivo frecuentemente usado en aplicaciones de aprendizaje de máquinas. A pesar de que es difícil mejorar OneR debido a su sencillez puede lograrse utilizando mejores métodos para manejar algunas de las excepciones. OneR viene de One Rule, es un algoritmo de clasificación que genera un árbol de decisión de un único nivel. OneR es capaz de inferir reglas de clasificación a partir de un conjunto de instancias. El algoritmo crea una regla para cada atributo en los datos de entrenamiento, luego escoge la regla con la 4 tasa de error1 más pequeño como su "one rule". Para crear una regla para cada atributo debe determinarse la clase más frecuente para cada valor del atributo.

El algoritmo OneR implementado en WEKA es muy efectivo deduciendo la "one rule" basado en un único atributo. Las desventajas que podemos advertir en este algoritmo son:

- El algoritmo trata todos los atributos numéricamente evaluados como continuos usa un método directo para dividir el rango de valores en intervalos disjuntos. Esto introduce un riesgo de "overfitting" en el caso de atributos evaluados de forma continua, por ejemplo: números de teléfono, etc...
- El "overfitting" de atributos nominales con valores unicos tales como nombres de personas, direcciones de correo electrónico, etc...
- Selección aleatoria de un atributo cuando las tasas de error son iguales.
- Selección aleatoria de una clase cuando dos o más clases dan la misma tasa de error con un atributo.

Sus resultados pueden ser bastantes buenos en comparación con algoritmos mucho más complejos. La idea es hacer reglas que prueban un solo par atributo-valor, probar todos los pares

atributo-valor y seleccionar el atributo que ocasione el menor número de errores, por tanto, sólo se utiliza un atributo para predecir la salida.

#### **Clasificador One:**

- RbatchSize: Si se realiza la agrupación de los sujetos por lotes, tamaño de cada lote.
- minBucketSize: Mínimo de categorías en las que se discretizan las variables numéricas incluidas en el análisis.
- numDecimalPlaces: Número de decimales presentados en los datos mostrados en el output

#### **Bibliografía.**

- [1] F. F. Mateos, "MINERÍA DE DATOS WEKA," pp. 1–23.
- [2] F. Martinez, "Aplicación de técnicas de minería de datos con software Weka".
- [3] A. Saavedra, "" Estimación del estado del flujo de tráfico mediante preprocesado y minería de datos. Aplicación de Dataset de posiciones GPS de taxis de Porto " TRABAJO FIN DE GRADO," *Universidad de Las Palmas de Gran Canaria*, vol. 1, p. 71, 2016.