

Influencia de las variables climatológicas sobre la conexión inalámbrica: Base Emisora-Receptora

(Influence of climatic variables on wireless: case study Base-Station Receiver)

Rodolfo Najarro Quintero¹, Eduardo Samaniego Mena¹, Freddy Fares Vagas¹, Amilkar Puris Cáceres^{1,2}

Resumen:

El presente trabajo de investigación se realiza con el objetivo de encontrar la relación que existe entre las condiciones climatológicas y la pérdida de la conexión inalámbrica en base terrena. Los datos son suministrados por un centro de meteorología radicado en la zona y una empresa de telecomunicación que opera en el mismo lugar. Se estudia directamente clasificadores basados en reglas difusas debido a lo fácil que resulta la interpretación de las reglas y manejo de los datos. Para este proceso se utilizó la aplicación informática Keel que ofrece técnicas de clasificación supervisada y aplicación Weka para el pre procesamiento de datos. Se estudiaron nueve clasificadores basados en reglas difusas donde el Furia-C fue el que mejores resultados obtuvo en cuanto a cantidad de reglas y a la precisión de clasificación. Algunas de las reglas obtenidas corroboran la influencia que tiene la lluvia fuerte sobre la pérdida de la señal, pero aparecen otras variables relacionadas con la pérdida de la señal, que incorporan nuevos conocimientos en el área, como por ejemplo la influencia del punto de rocío y la humedad relativa media.

Palabras clave: clasificación supervisada; climatología; atenuación de la señal; reglas difusas

Abstract:

The development of this research is done with the aim of finding the relationship between weather conditions and the loss of wireless connection. The data were obtained by a meteorological center of the area and a telecommunications company that operates in the same place. We studied different models based on fuzzy logic due to the easy interpretation the easy interpretation of the rules and data management. We used the Weka application that provides tools for pre-processing of data and Keel software tool for data classification. Nine classifiers based on fuzzy rules were applied, where the Furia-C was that better results obtained in order to quality and quantity of rules. In this scenario, a preprocessing of data were computed, where some techniques to improve the information was performed. Some of the obtained rulers corroborate the influence of heavy rain over the loss of the signal, but other relationships that incorporate new knowledge in the area, such as dew point and the average relative humidity appear.

Keywords: supervised classification; climatology; signal attenuation; fuzzy ruler

¹ Universidad Técnica Estatal de Quevedo, Quevedo– Ecuador ({rnajarro, esamaniego , ffares , apuris}@uteq.edu.ec)

² Universidad Estatal de Milagro, Milagro-Ecuador

1. Introducción

En la actualidad el desarrollo de las telecomunicaciones está orientado al uso de las tecnologías digitales, reduciendo en gran medida las transmisiones analógicas. Su evolución ha sido notable en los sistemas de radiocomunicación por satélite para radiodifusión digital directa de TV y audio, comunicaciones móviles de banda estrecha, y novedosos proyectos de comunicaciones fijas de banda ancha, utilizando tanto satélites geoestacionarios como no-geoestacionarios.

Existen algunos factores que afectan la transmisión inalámbrica, entre los cuales se encuentran; la atenuación, la distorsión, la pérdida del espacio libre, la absorción atmosférica, la refracción, el ruido entre otros. Estos factores han sido estudiados y se han definido algunos parámetros a tener en cuenta para disminuir su impacto en los servicios inalámbricos.

Específicamente la atenuación o pérdida de la potencia de la señal es una de las causas que más afecta la comunicación inalámbrica y ocurre principalmente por dos factores; la distancia entre el transmisor y el receptor de la señal, y por la influencia del clima. En ambos casos, la señal distorsionada provoca que el receptor no interprete bien la información enviada y por ende existan fallas en el servicio. Para disminuir la atenuación causada por la distancia, se han utilizado soluciones tecnológicas que abarca el uso de repetidores y amplificadores para mantener la potencia de la señal para transmisiones en larga distancia. En el caso de atenuación por condiciones climatológicas, solo se conoce de estudios que involucran la variable lluvia como factor fundamental en la pérdida de la señal (Moncada, D., 2006), (Fermín, J. R., & Simancas, M., 2010), no así la influencia de otras variables climatológicas.

Por tal motivo, en el presente trabajo se presenta un caso de estudio real, con información de diferentes variables climatológicas medidas durante 500 días, en la misma zona donde se tiene una base receptora que brinda servicios de internet a un grupo considerable de clientes. El objetivo es, utilizar esta información para encontrar las relaciones que existen entre las condiciones climatológicas con la pérdida de la señal. Para la construcción del modelo computacional se aplican técnicas de minería de datos específicamente modelos basados en reglas difusas que ayudan a encontrar relaciones y tendencias en la información.

2. Metodología

Para el desarrollo de la presente investigación se contó con datos de diferentes variables climatológicas tomadas en la estación experimental “Pichilingue” en el cantón Quevedo, donde existe una extensión del INAMHI³ (Servicio Meteorológico e Hidrológico Nacional del Ecuador). Además, se obtuvo el registro de una empresa de telecomunicaciones que brinda servicios en la zona y cuenta con una base terrena que recibe la señal de una estación primaria de forma

3 <http://www.serviciometeorologico.gob.ec/>

inalámbrica. Los datos suministrados por esta empresa contienen la hora y el día en que fue interrumpido el servicio por falta de señal.

La base de conocimiento se formó a partir de 1500 registros tomados en un período de 500 días (enero del 2014 a junio del 2015), de los cuales se proporcionaron 3 mediciones meteorológicas por día (6 am, 12 pm y 6 pm). El reporte de pérdida de la señal, se asoció al registro meteorológico más cercano a la hora del fallo, los restantes registros se completaron sin fallos.

Con esta información se aplicó un método analítico a partir de un estudio experimental en el cual se aplicaron diferentes técnicas de minería de datos, específicamente clasificadores basados en reglas difusas, que utilizan algoritmos evolutivos para encontrar las reglas (Gao, Q., & He, N. B., 2016). El objetivo de la aplicación de estos modelos es encontrar aquel que mejor relacione las variables climatológicas con la pérdida de la señal. Para llevar a cabo el proceso de descubrimiento se utilizó la herramienta Weka⁴ (Waikato Environment for Knowledge Analysis) para el análisis y preprocesamiento de los datos y el *software* KEEL⁵ (Knowledge Extraction based on Evolutionary Learning) para la aplicación de diferentes clasificadores basado en reglas difusas.

3. Resultados y Discusión

En este punto se muestra todo el proceso desarrollado para descubrir la relación entre las variables climatológicas y la pérdida de la señal con énfasis en la extracción de reglas difusas. La *Tabla 1* muestra los parámetros utilizados en el proceso, donde PS representa la variable dependiente o de clase para el caso de estudio.

Tabla 1. Variables utilizadas en el proceso de descubrimiento

Variable	Descripción	Variable	Descripción
TM	Temperatura media	TVA	Tensión de vapor de agua
TMAX	Temperatura máxima	PR	Punto de rocío
TMIN	Temperatura mínima	HE	Heliofania
TOSC	Oscilación térmica	EV	Evaporación del agua
HRM	Humedad relativa Media	LL	Cantidad de lluvia
HRMAX	Humedad relativa máxima	PS	Disponibilidad de la Conexión: 0 = pérdida 1 = conexión
HRMIN	Humedad relativa mínima		

4 <http://www.cs.waikato.ac.nz/ml/weka/>

5 <http://keel.es/>

3.1 Análisis y preparación de los datos

La calidad de los resultados obtenidos en el proceso de obtención de conocimientos no solo depende de los métodos de extracción, sino también, de cómo se haya conformado la base de datos y todo el preprocesamiento desarrollado para obtener una base de conocimiento lo más sólida posible. A continuación, se describen cada uno de los pasos realizados para el preprocesamiento de la información.

Paso1. Análisis de outliers: para llevar a cabo la verificación de los datos y detectar posibles ruidos o inconsistencias en los registros climatológicos, se procedió analizar los valores de rango permisibles de cada variable, tomando como referencia los datos suministrados por el sitio <http://www.iniap.gob.ec/> donde aparecen promedios anuales de cada variable en la zona. Como resultado del análisis gráfico se detectó que todas las variables estaban dentro de los estándares establecidos para la zona de estudio por lo cual no se aplicó ningún algoritmo para eliminar outliers.

Paso2. Análisis de desbalance de clases: la presencia de clases no balanceadas en bases de conocimientos es un problema frecuente en muchas aplicaciones de aprendizaje automático y cuyos efectos sobre el desempeño de los clasificadores son notables. El problema ocurre cuando el número de instancias perteneciente a cada clase es muy diferente provocando en los clasificadores un sobre aprendizaje de la clase dominante y poco aprendizaje de las clases con pocas instancias.

En este sentido la base de datos utilizada en esta investigación cuenta con una variable clase que representa el estado de la conexión inalámbrica y tiene dos valores; conexión o no conexión. En la *Figura 1 a)* se puede apreciar que aparece una dominancia de la clase conexión con 1155 casos mientras que la clase sin conexión solo tiene 345 representantes. Luego, la *Figura 1 b)* muestra los datos luego de aplicar el algoritmo SMOTE (Synthetic Minority Oversampling Technique) (Xie, Z., Jiang, L., Ye, T., & Li, X., 2015) que sigue una estrategia de sobremuestreo donde las nuevas instancias se crean a partir de la interpolación de los casos más cercanos a la clase minoritaria.

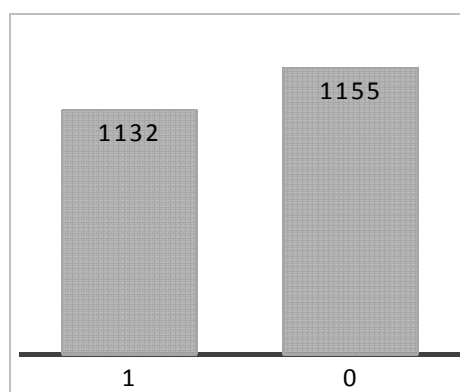


Figura 1. Análisis de la variable clase

Paso 3. Discretización de las variables: para poder obtener un conjunto de datos con rangos más asequibles para trabajar, se ejecutaron dos algoritmos en el Keel, el de frecuencia y amplitud uniforme. La *Figura 2* muestra el esquema seguido y los valores obtenidos para la formación de grupos para cada variable se muestran en la *Tabla 2*.

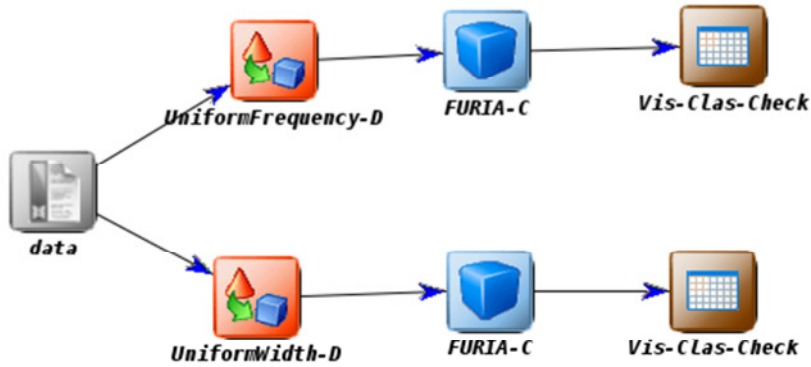


Figura 2. Diagrama para la discretización de variables en Keel.

Tabla 2. Variables discretizadas y sus grupos

Variable	AttributeIndices	Bins	Variable	AttributeIndices	Bins
TM	first-4,last	5	TVA	8-10,last	5
TMAX			PR		
TMIN			HE		
TOSC			EV	11,last	7
HRM	5-7,last	6	LL	12,last	3
HRMAX					
HRMIN					

Paso 4. Selección de atributos o variables: en este proceso se elimina información redundante que no es beneficiosa para el proceso de extracción de conocimiento, esto posibilita que los algoritmos funcionen de forma más rápida y precisa. La base de casos de este estudio presenta 12 atributos predictores, por lo que técnicamente no es necesario realizar selección de atributos. No obstante, se probaron 6 algoritmos de filtrado existentes en Weka como se muestra en la *Figura 3*, para seleccionar las variables más importantes.

En la *Tabla 3* se puede observar que la mayoría de los algoritmos obtienen el mismo subconjunto de 7 variables, solo el algoritmo de búsqueda aleatoria (RandomSearch) realiza una selección más simplificada, dejando fuera la variable humedad relativa media (HRM).

A continuación, en la *Figura 4* se presentan los valores de precisión y desviación estándar obtenidos por el clasificador FURIA-C (Hühn, J., & Hüllermeier, E., 2009) con cada subconjunto de variables. Se puede apreciar que la mayor precisión se obtiene utilizando el subconjunto de 7 variables, incluso sus resultados son muy similares a los alcanzados por el mismo clasificador si se utilizan todas las variables. Otro elemento a tener en cuenta es la cantidad de reglas que se

obtienen porque marcan la diferencia en cuanto a la eficiencia del clasificador. Para este indicador, el utilizar solo 7 variables en el proceso de descubrimiento produce 8 reglas, mientras que, si se utilizan todas las variables, el número de reglas asciende a 24 reglas.

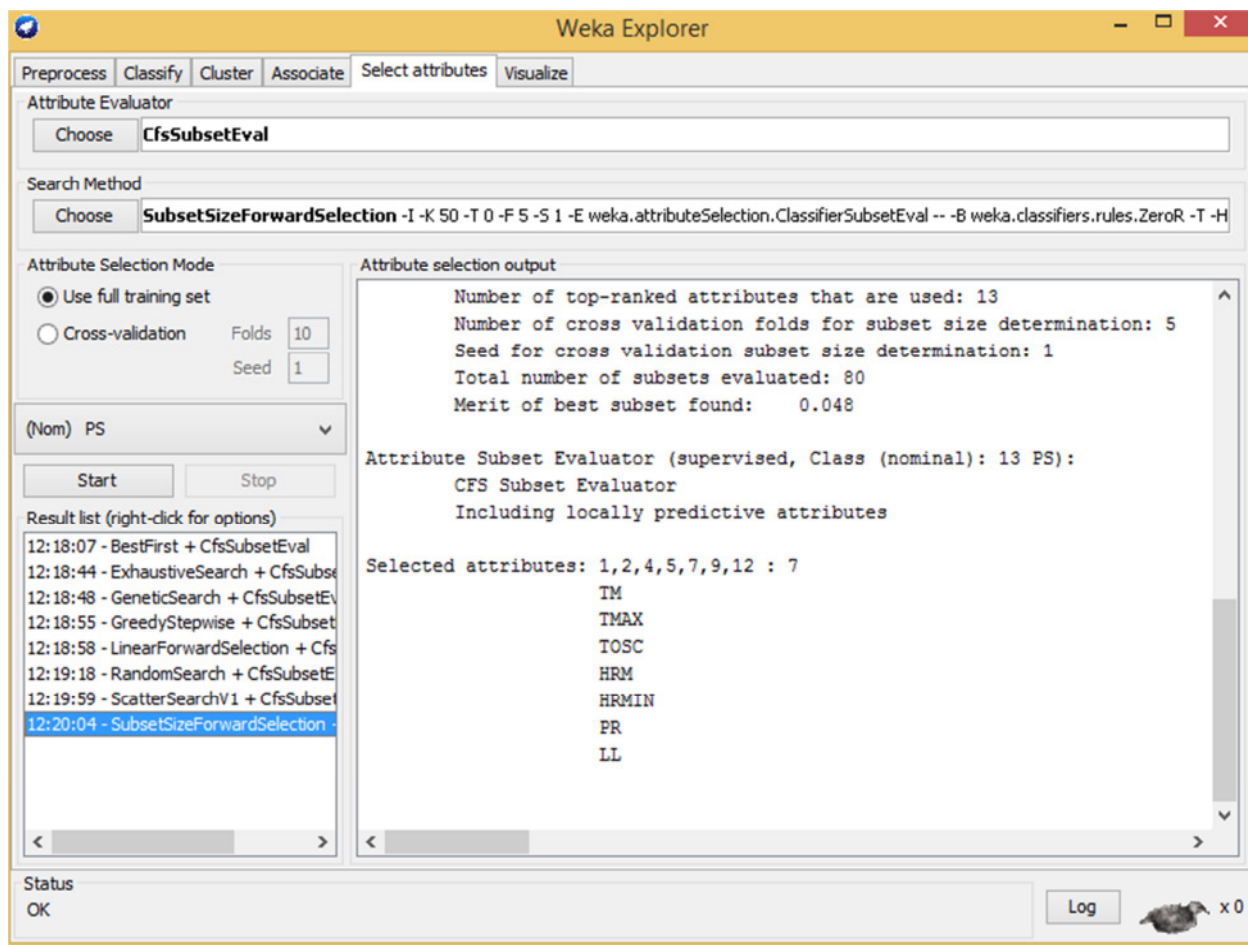


Figura 3. Selección de atributos utilizando Weka

Tabla 3. Resultados de la selección de variables

Algoritmo	Atributos
Best first	TM, TMAX, TOSC, HRM, HRMIN, PR, LL
SubsetSizeForwardSelection	TM, TMAX, TOSC, HRM, HRMIN, PR, LL
GeneticSearch	TM, TMAX, TOSC, HRM, HRMIN, PR, LL
GreedyStepwise	TM, TMAX, TOSC, HRM, HRMIN, PR, LL
ScatterSearchV1	TM, TMAX, TOSC, HRM, HRMIN, PR, LL
RandomeSearch	TM, TMAX, TOSC, HRMIN, PR, LL

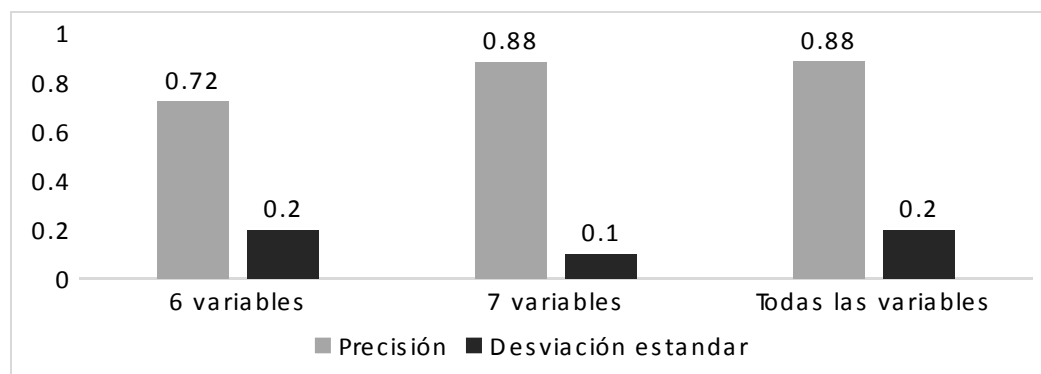


Figura 4. Resultados del proceso de clasificación para cada conjunto de datos

Con estos resultados se procedió a continuar el estudio con el subconjunto de 7 para encontrar la forma más descriptiva para la pérdida de la señal en el caso real de estudio.

3.2. Estudio comparativo e interpretación de las reglas

A continuación, se realiza una comparación entre 9 clasificadores basados en reglas difusas para seleccionar aquel que mejor represente los datos a partir de 3 indicadores: Precisión, Cantidad de reglas y Desviación estándar (delante de cada indicador, valor de orden en el *ranking*). La *Figura 5* presenta el modelo definido en Keel para la ejecución de los experimentos.

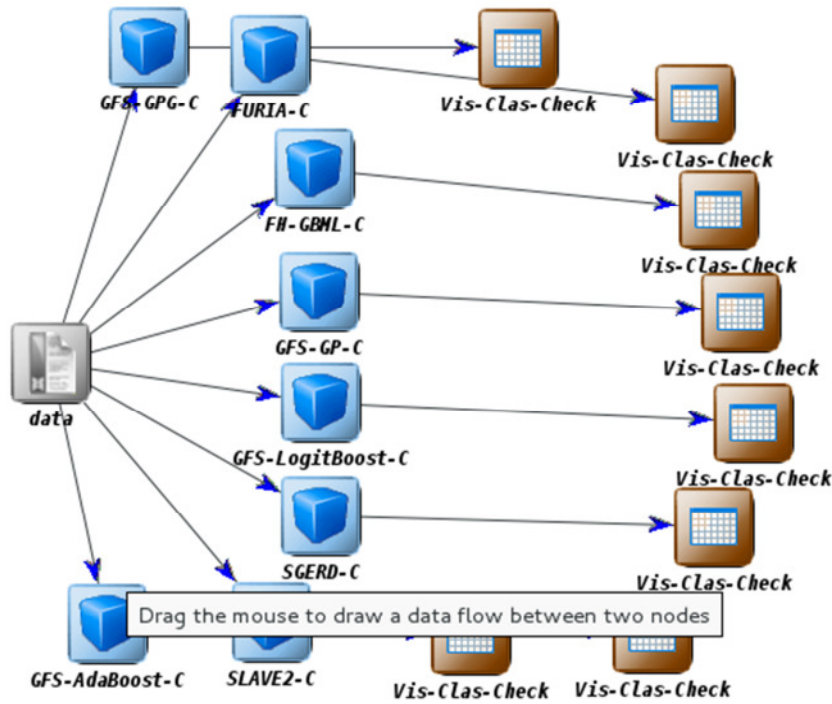


Figura 5. Modelo en Keel para el estudio comparativo entre clasificadores

La *Tabla 4* muestra que el clasificador FURIA-C es el que mejores valores obtiene para el indicador precisión superando en más de 0.5 al siguiente clasificador mejor situado (GFS-LogitBoost-C). Para el caso del indicador cantidad de reglas este clasificador obtuvo el segundo lugar con 8 reglas solo superado por SGERD-C que presentó una baja precisión. Y para el indicador desviación estándar el algoritmo FURIA-C se encuentra entre los 3 que mejores valores obtiene.

Según los resultados antes descritos se puede concluir que el algoritmo FURIA-C es el que mejores resultados alcanza de manera general, debido a que supera significativamente en precisión a los demás modelos y para los otros indicadores la diferencia con el mejor es mínima. A continuación, la *Tabla 5* describe la base de reglas obtenidas por el algoritmo.

Se puede observar que las 3 primeras reglas determinan que las variables más relacionadas con la pérdida de la señal son: Lluvia (LI), Punto de Rocío (PR), Temperatura de Oscilación (TOSC), Temperatura Media (TM) y la Humedad Relativa Media (HRM).

Tabla 4. Resultados del análisis comparativo de clasificadores difusos

Algoritmo	Precisión		Cantidad de reglas		Desviación estándar	
FH-GBML-C (Murcia, C., Bonilla, G., & Melgarejo, M., 2014)	5	0,7492	4	11	4	0,0203
FURIA-C (Huhn & Hullermeier., 2009)	1	0,8834	2	8	3	0,0199
Grammar-based GP algorithm (GFS-GP-C) (Sánchez & Couso, 2001)	4	0,8009	6	14	2	0,0122
AdaBoost-C (Del Jesús, H., Junco & Sánchez, 2004)	6	0,7188	4	11	7	0,0328
GFSMaxLogitBoost-C (González, A., Pérez, R., Caises, Y., & Leyva, E. 2012)	7	0,7144	5	13	5	0,0217
GFS-GPG-C (Sánchez & Couso, 2001)	3	0,8276	6	14	1	0,0101
GFS-LogitBoost-C (Bartczuk, L., Przybyl, A., & Cpałka, K., 2016).	2	0,8317	3	9	6	0,0239
SGERD-C (Mansoori, Zolghadri & Katebi, 2008)	8	0,6618	1	7	9	0,0477
SLAVE-C (Sudha, M. 2017)	9	0,6509	7	15	8	0,0368

Tabla 5. Reglas obtenidas por el clasificador FURIA-C

Regla	Interpretación	Precisión
R1	Si TMax mayor que 26,52 y LL mayor 104,72 entonces pérdida de señal.	0,94
R2	Si PR mayor que 20,2 y HRM mayor que 95 and TOSC entre 3,6 y 6,4 entonces pérdida de señal.	0,86
R3	Si TM entre 25,24 y 26,52 y TOSC menor que 3,6 y HRM entre 82,79 y 89,39 entonces pérdida de señal.	0,77
R4	Si TOSC menor que 3,6 y LL entre 52,36 y 78,53 entonces no hay pérdida de señal.	0,80
R5	Si PR mayor que 20,2 entonces no hay pérdida de señal.	0,80
R6	Si HRMIN entre 38,93 y 56 entonces no hay pérdida de señal.	0,79
R7	Si TM mayor que 26,52 entonces no hay pérdida de señal.	0,76
R8	Si TM 23,96 y 25,24 TOSC menor 3,6 entonces no hay pérdida de señal.	0,72

Específicamente la regla R1 relaciona la lluvia de gran intensidad (superior a los 104 mm) y la temperatura máxima alta (superior a los 26 grados), con la pérdida de la señal bajo una precisión del 94%. Esta información es totalmente consistente con estudios realizados por otros autores. Por su parte la regla R2 incorpora nuevo conocimiento en el área, esto se debe a que para un punto de rocío alto (mayor que 20,2), una humedad relativa alta (mayor al 95%) y una temperatura de oscilación normal (entre 3,6 y 6,4) existe una pérdida de la señal en el 88% de los casos en que se encuentra esta combinación de valores en las variables. De la misma forma ocurre con la regla R3, que cuando aparece relacionada una temperatura media normal (entre 25,24 y 26,53) con una temperatura de oscilación baja (menor 3,6) y una humedad relativa normal (entre el 82,79% y 89,39%) existe una probabilidad de 0,77 que se pierda la señal.

4. Conclusiones y Recomendaciones

La presente investigación se realizó para encontrar un modelo basado en reglas difusas que relacionen un conjunto de variables climatológicas con la pérdida de la señal inalámbrica. Para ello, se utilizó un escenario real de estudio con datos suministrados por una estación meteorológica, en la misma zona donde opera una empresa de telecomunicaciones que brinda servicios a la comunidad.

Dentro de las 12 variables climatológicas utilizadas en el proceso de extracción del conocimiento, solo 7 de ellas se relacionan directamente con la pérdida de la señal inalámbrica. Esto se obtuvo a través del proceso de selección de atributos donde se probaron diferentes algoritmos tipo filtro, conjuntamente con el clasificador FURIA-C y los mejores resultados se obtuvieron con el subconjunto formado por Temperatura Media, Temperatura Máxima, Humedad Relativa Media, Humedad Relativa Mínima, Oscilación Térmica, Punto de Rocío y Lluvia.

El modelo difuso obtenido responde a 8 reglas cuya interpretación se ajusta en parte a los conocimientos previos del fenómeno como es el caso, cuando la temperatura media es alta y la lluvia es fuerte entonces la señal se pierde. No obstante, se obtuvieron otras reglas no tan fáciles de deducir por los expertos como: cuando el punto de rocío es alto, la humedad relativa es alta y la oscilación térmica es media la señal también se pierde y presenta conocimiento nuevo en el ámbito de estudio para ser utilizados en posteriores investigaciones.

Para seguir profundizando en el estudio de la influencia de las variables climatológicas con la pérdida de la conectividad, se recomienda aplicar otros clasificadores para refinar los modelos obtenidos. Además, se puede replicar el estudio en otros escenarios de comunicación como es la comunicación satelital.

Bibliografía

- Bartczuk, L., Przybyl, A., & Cpalka, K. (2016). A new approach to nonlinear modelling of dynamic systems based on fuzzy rules. *International Journal of Applied Mathematics and Computer Science*, 26(3), 603-621.
- Del Jesús, M. J., Hoffmann, F., Navascués, L. J., & Sánchez, L. (2004). Induction of fuzzy-rule-based classifiers with evolutionary boosting algorithms. *IEEE Transactions on Fuzzy Systems*, 12(3), 296-308.
- Fermín, J. R., & Simancas, M. (2010). Pronóstico de la atenuación por lluvia en las comunicaciones satelitales mediante métodos de regresión. *Télématique: Revista Electrónica de Estudios Telemáticos*, 9(3), 23-35.
- Gao, Q., & He, N. B. (2016). Study on Fuzzy Classifier Based on Genetic Algorithm Optimization. In *Proceedings of the 5th International Conference on Electrical Engineering and Automatic Control* (pp. 725-731). Springer Berlin Heidelberg.
- González, A., Pérez, R., Caises, Y., & Leyva, E. (2012). An efficient inductive genetic learning algorithm for fuzzy relational rules. *International Journal of Computational Intelligence Systems*, 5(2), 212-230.
- Hühn, J., & Hüllermeier, E. (2009). FURIA: an algorithm for unordered fuzzy rule induction. *Data Mining and Knowledge Discovery*, 19(3), 293-319.

- Mansoori, E. G., Zolghadri, M. J., & Katebi, S. D. (2008). SGERD: A steady-state genetic algorithm for extracting fuzzy classification rules from data. *IEEE Transactions on Fuzzy Systems*, 16(4), 1061-1071.
- Moncada, D. (2006). Efectos de Atenuación por Lluvia en los Medios de Transmisión entre Sistemas Satelitales y Estaciones Terrestres. *Trabajo de grado para optar al Título de Magíster en Telemática. Universidad Rafael Bellosó Chacín. Maracaibo, Venezuela.*
- Murcia, C., Bonilla, G., & Melgarejo, M. (2014). Fuzzy Classifiers Tuning Through an Adaptive Memetic Algorithm. *IEEE Latin America Transactions*, 12(2), 197-204.
- Sánchez, L., Couso, I., & Corrales, J. A. (2001). Combining GP operators with SA search to evolve fuzzy rule based classifiers. *Information Sciences*, 136(1), 175-191.
- Sánchez, L., Couso, I., & Corrales, J. A. (2001). Combining GP operators with SA search to evolve fuzzy rule based classifiers. *Information Sciences*, 136(1), 175-191.
- Sudha, M. (2017). Intelligent decision support system based on rough set and fuzzy logic approach for efficacious precipitation forecast. *Decision Science Letters*, 6(1), 95-106.
- Xie, Z., Jiang, L., Ye, T., & Li, X. (2015). A synthetic minority oversampling method based on local densities in low-dimensional space for imbalanced learning. In *International Conference on Database Systems for Advanced Applications* (pp. 3-18). Springer International Publishing.