UNIVERSIDAD INTERNACIONAL SEK

FACULTAD DE CIENCIAS NATURALES Y AMBIENTALES

Trabajo de Fin de Carrera Titulado:

"ANÁLISIS DEL COMPORTAMIENTO DE LAS VARIABLES AMBIENTALES Y ENERGÉTICAS QUE INCIDEN EN LA OPERACIÓN DEL CATAMARÁN SOLAR "INER 1" EN LAS ISLAS GALÁPAGOS MEDIANTE SELECCIÓN DE ATRIBUTOS"

Realizado por:

MARCELO JAVIER MOYA CAJAS

Director del proyecto:

PhD. JAVIER MARTÍNEZ GÓMEZ.

Como requisito para la obtención del título de:

MASTER EN ECOEFICIENCIA INDUSTRIAL

Quito, 09 de septiembre de 2019

DECLARACIÓN JURAMENTADA

Yo, MARCELO JAVIER MOYA CAJAS, con cédula de identidad # 1721871307, declaro bajo juramento que el trabajo aquí desarrollado es de mi autoría, que no ha sido previamente presentado para ningún grado a calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración, cedo mis derechos de propiedad intelectual correspondientes a este trabajo, a la UNIVERSIDAD INTERNACIONAL SEK, según lo establecido por la Ley de Propiedad Intelectual, por su reglamento y por la normativa institucional vigente.

FIRMA

1721871307

DECLARATORIA

El presente trabajo de investigación titulado:

"ANÁLISIS DEL COMPORTAMIENTO DE LAS VARIABLES AMBIENTALES Y ENERGÉTICAS QUE INCIDEN EN LA OPERACIÓN DEL CATAMARÁN SOLAR "INER 1" EN LAS ISLAS GALÁPAGOS MEDIANTE SELECCIÓN DE ATRIBUTOS"

Realizado por:

MARCELO JAVIER MOYA CAJAS

como Requisito para la Obtención del Título de:

MASTER EN ECOEFICIENCIA INDUSTRIAL

ha sido dirigido por el profesor

PhD. JAVIER MARTINEZ GOMEZ

quien considera que constituye un trabajo original de su autor

FIRMA

LOS PROFESORES INFORMANTES

Los Profesores Informantes:

Después de revisar el trabajo presentado,

lo han calificado como apto para su defensa oral ante

el tribunal examinador

FIRMA
Mónica Delgado Yánez
C.I. 171358255-7

FIRMA
Edilberto Llanes Cedeño
C.I. 175713703-7

Quito, 09 de septiembre de 2019

DEDICATORIA

A Dios y la Virgen Santísima por todas sus bendiciones.

A ti amor que en el pre grado siempre te quedaste a mi lado entregando todo tu apoyo y amor como mi novia, ahora lo volviste hacer como mi esposa no tengo nada más que agradecimiento y amor por todo lo que has hecho por mí te amo.

Para ustedes Mamita y Abuelita que nunca se supieron rendir y depositaron toda su fe y amor en mí, esta es la culminación de su esfuerzo, las extraño mucho jamás las olvidare.

Padre esta alegría te la debía aún falta más por aprender, gracias por todo.

Ñaña sin tu cariño, amistad y comprensión mi vida hubiera sido muy solitaria gracias por estar siempre junto a mí.

AGRADECIMIENTO

Hermanos muchas gracias por su ejemplo y apoyo incondicional siempre los he escuchado.

A mis tíos, tías y primos que han sido el mejor ejemplo de lucha constante, gracias por haberme cobijado siempre, aunque las obligaciones ahora nos separen siempre están en mis oraciones.

A mis amigos tanto de la Universidad UTE como del IIGE muchas gracias por cubrirme la espalda mientras lograba finalizar la maestría.

09/09/2019

Para someter a:

To be submitted:

ANÁLISIS DEL COMPORTAMIENTO DE LAS VARIABLES AMBIENTALES Y ENERGÉTICAS QUE INCIDEN EN LA OPERACIÓN DEL CATAMARÁN SOLAR "INER 1" EN LAS ISLAS GALÁPAGOS MEDIANTE SELECCIÓN DE ATRIBUTOS.

Marcelo Moya[1,2,3], Javier Martínez Gómez [1,2], Gonzalo Guerrón [1,2].

[1] Universidad Internacional SEK, Facultad de Ciencias Naturales y Ambientales, Quito, Ecuador.

[2]Instituto de Investigación Geológico y Energético. IIGE.

[3]Universidad UTE, UTE.

*AUTOR DE CORRESPONTENCIA: Ing. Marcelo Moya Universidad Internacional SEK, Facultad de Ciencias Naturales y Ambientales, Quito, Ecuador.

Teléfono: +593-995584925; email: mjmoya.mee@uisek.edu.ec

Título corto o Running title: Eficiencia energética “INER 1”, variables que afectan su operación.

EFICIENCIA ENERGÉTICA INER 1 VARIABLES QUE AFECTAN SU OPERACIÓN

Resumen.

El transporte marítimo de pasajeros en las Islas Galápagos - Canal de Itabaca se basa en embarcaciones con motores de combustión, que consumen un promedio anual de 4 200 galones de combustible que producen alrededor de 38 toneladas de CO_2 por año. La operación del catamarán solar "INER 1" de propulsión eléctrica - fotovoltaica es un modelo sustentable y renovable para el transporte marítimo de pasajeros dentro de las Islas Galápagos. En este sentido, la variabilidad tanto de la radiación solar, el cambio brusco de las mareas junto a la cambio de la velocidad y dirección del viento y el incremento de los turistas de acuerdo a las temporadas seca y húmeda provocan un aumento en el consumo de energía de la embarcación generando necesidad de absorber energía adicional de la red eléctrica de las Islas. El objetivo del presente trabajo de titulación es formular un estudio que permita seleccionar los atributos más influyentes en la operación del catamarán solar "INER 1" en las Islas Galápagos. Para lo cual se utilizó la metodología KDD que permite realizar el descubrimiento de conocimiento en bases de datos y mediante la selección de atributos se identificó la combinación de variables ambientales, sociales y energéticas que afectan su eficiencia energética encontrando una relación directa con los siguientes atributos definidos en la investigación como: 1) Energ (Energía utilizada), 2) Tur (Turistas y residentes), 3) Fotov (Parque fotovoltaico), 4) Glrad (Radiación Global), 5) date (fecha y hora), 6) Term9 (Termoeléctrica 9). El uso de la metodología y los algoritmos propuestos permitió obtener un 4.95% en el parámetro error cuadrático medio y un 98.94% de exactitud en la clasificación y predicción de energía que requiere la embarcación en un periodo de tiempo.

Palabras clave: Islas Galápagos; biodiversidad; solar fotovoltaica; embarcación; algoritmo; inteligencia artificial.

Abstract.

Maritime passenger transport in the Galapagos Islands - Itabaca Channel is based on boats with combustion engines that consume an annual average of 4 200 gallons of fuel and produce about 38 tons of CO_2 per year. The operation of the solar catamaran "INER 1" electric propulsion - PV is a renewable and sustainable model for passenger shipping in the Galapagos Islands. In this regard, variability of solar radiation, the abrupt change of the tides by the change of speed and direction of the wind and increasing tourists according to both dry seasons and wet cause an increase in energy consumption of the vessel and need for additional generating power grid Islands. The aim of this study is to formulate a titration study to select the most influential attributes in the operation of the solar catamaran "INER 1" in the Galapagos Islands. KDD methodology that enables knowledge discovery in databases by selecting attributes the combination of environmental, social and energy variables affecting energy performance was identified was used. The energy consumption of the boat has a direct relationship with the attributes defined in research as: 1) Energ (Energy used), 2) Tur (tourists and residents), 3) Fotov (PV Park), 4) Glrad (Global Radiation), 5) date (date and time), 6) Term9 (Thermoelectric 9). Considering the 6 best attributes filtered by the proposed algorithms one 4.95% in the mean squared error parameter and a 98.94% accuracy in classification and prediction of energy that will search the boat was obtained.

Key words: Galapagos Islands; biodiversity; solar photovoltaic; boat; algorithm; artificial intelligence.

Índice de Contenidos

Página

Resumen. **2**
Abstract. **3**
Introducción. **7**
Minería de Datos. 11
Identificación de atributos. 14
Evaluador de atributos. 15
Pre-procesamiento de información. 16
Redes neuronales artificiales. 17
Catamarán solar "INER 1" 19
Parámetros meteorológicos. 21
Variables energéticas. 21
Variables sociales. 22
Alcance de la investigación 23
Justificación de la investigación 23
Objetivos 24
Materiales y Métodos. **25**
Resultados. **31**
Análisis de atributos. 31
Análisis de Energía 43
Discusión. **44**
Conclusiones. **46**
Agradecimientos **47**
Bibliografía. **48**

Índice de Figuras

Página

Figura 1. Catamarán Solar "INER 1". 10
Figura 2. Perceptrón una sola capa. 18
Figura 3. Perceptrón Multicapa. 19
Figura 4. Sistema de gestión energética de la embarcación solar "INER 1". 22
Figura 5. Ruta de ingreso a Santa Cruz – Islas Galápagos. 23
Figura 6. Representación gráfica del proceso descubrimiento de conocimiento en bases de datos (KDD) 26
Figura 7. Muestra de la base de datos en formato ARFF. 32
Figura 8. Tabla de atributos y datos estadísticos. 33
Figura 9. Configuración de parámetros para la identificación de atributos. 34
Figura 10. Importancia del atributo InfoGain vs atributos. 37
Figura 11. Importancia del atributo Jrip vs atributos. 39

Índice de Tablas

Página

Tabla 1. Información de las variables utilizadas en la investigación 28
Tabla 2. Ejemplo de los datos utilizados en el estudio 29
Tabla 3. Clasificación de los 25 atributos pre-procesados 35
Tabla 4. Configuración del primer evaluador de atributos 35
Tabla 5. Clasificación de atributos evaluador InfoGain 36
Tabla 6. Configuración del segundo evaluador de atributos 37
Tabla 7. Clasificación de atributos evaluador Classifier attribute Jrip 38
Tabla 8. Configuración del tercer evaluador de atributos 39
Tabla 9. Clasificación de atributos evaluador ReliefAttributeEval 40
Tabla 10. Atributos principales evaluador InfoGain 41
Tabla 11. Clasificación de los 6 atributos con el evaluador 41
Tabla 12. Atributos principales evaluador Classifier attribute Jrip 42
Tabla 13. Clasificación de los 6 atributos con el evaluador 42
Tabla 14. Generación y consumos eléctricos de la embarcación solar “INER 1” 43

Introducción.

El incremento de la calidad de vida y la longevidad en el ser humano ha provocado un crecimiento poblacional exponencial en los últimos 200 años, esto ha generado grandes beneficios a la humanidad pero también ha generado grandes afectaciones en los ecosistemas, recursos naturales y en general con el medio ambiente. Uno de los problemas más grandes son las actividades humanas que provocan una gran cantidad de emisiones de gases de efecto invernadero, químicos que afectan a la capa de ozono desechos que no pueden ser reutilizados (Helling, 2017).

En las últimas décadas el ser humano ha prestado una gran importancia al cuidado de los recursos energéticos renovables, no renovables y a la protección del medio ambiente tratando de identificar y mitigar los impactos ambientales generados por la producción y manufactura de tecnología (Helling, 2017). El desarrollo de metodologías de análisis y mitigación de impactos ambientales han permitido reducir el daño causado al medio ambiente, de tal manera que se han desarrollado varias metodologías para identificar y mitigar impactos ambientales una de las metodologías más utilizadas es el Análisis de Ciclo de Vida y se (LCA) debido a su estandarización y facilidad de aplicación (Beatriz Rivela, 2014).

De acuerdo a Ko et al. (2016), el sistema económico actual se encuentra basado en el uso de embarcaciones destinadas para el transporte de pasajeros, productos primarios, intermedios y productos acabados. Una embarcación presta sus servicios mientras es utilizable generando ingresos al propietario y al mismo tiempo genera impactos ambientales a través de las emisiones de la combustión y residuos de combustible al medio ambiente. En la fase final de utilización es decir, en el desmantelamiento, se producen impactos significativos en el medio ambiente y en las personas involucradas.

Las embarcaciones con sistemas de propulsión eléctricos han tenido un importante

desarrollo tecnológico en los últimos años, tal es el caso que en el año 2007 la embarcación denominada Sun21 arribo a Nueva York siendo el primer bote con propulsión eléctrica utilizando únicamente energía solar en cruzar el océano atlántico con un recorrido aproximado de 11 000 kilómetros ahorrando un aproximado de 4 000 litros de diésel (Gursu, 2016). Posteriormente el "MS Tûranor PlanetSolar" en 2012 concluyó un recorrido de 584 días alrededor del mundo alcanzando un nuevo record. Sin embargo el bote fue optimizado para romper sus propios records y cruzo el atlántico en 22 días en 2013 (Bleicher, 2013).

Según Lucas et al. (2013) un modelo de transporte basado en energía eléctrica generada a través de hidrogeno ya sea por electrólisis o mediante estaciones de abastecimiento permitiría reducir las emisiones de gases de efecto invernadero en un rango de 0.6 -2.2 gCo_2/MJ en Europa y entre 6.0-17.7 gCO_2/MJ en Estados Unidos, esto se debe a que los países que tienen una matriz energética basada en energía renovables tienen un mejor desempeño en una análisis de ciclo de vida propuesto con las mismas políticas en ambos escenarios.

De acuerdo a Ling-Chin et al. (2016) el análisis de ciclo de vida en embarcaciones permite identificar los impactos asociados a los sistemas de generación eléctrica y generadores auxiliares concluyendo que es posible mejorar los procesos de manufactura y operación de las embarcaciones siguiendo la metodología de análisis de ciclo de vida.

Las Islas Galápagos poseen una economía sensible que se encuentra basada en el turismo por lo tanto existe una gran cantidad de turistas nacionales y extranjeros que visitan las islas y requieren de transporte marítimo entre ellas (Schiller et al., 2015). De acuerdo a Maggí et al. (2018), en el año 2015 se contabilizaron 70 451 pasajeros nacionales y 154 304 pasajeros extranjeros que ingresaron a las Islas Galápagos. Del total de pasajeros que ingresaron 173 487 el 77 % ingreso por el aeropuerto de Baltra y los 51.116 pasajeros restantes el 23 % por la isla de San Cristóbal. Las visitas de los turistas a las Islas se

encuentran determinadas por dos estaciones climáticas la cálida y lluviosa que va desde Enero hasta Junio y la seca y fría que empieza en Julio y termina en Diciembre de cada año. De acuerdo a lo mencionado existen dos temporadas turísticas en las Islas que se las conoce como temporada alta que comprende los meses de Enero, Junio, Julio, Agosto, Diciembre con un promedio de 19 370 visitantes y la temporada baja Febrero, Marzo, Abril, Mayo, Septiembre, Octubre, Noviembre con un promedio de 18 022 visitantes (Parque-Nacional-Galápagos, 2017).

En junio de 2014, se inauguró en las Islas Galápagos la primera lancha eléctrica solar denominada Solaris, la cual fue desarrollada por la Dirección del Parque Nacional Galápagos (DPNG), en coordinación con el Fondo Mundial para la Naturaleza (WWF) (García et al., 2013). Fue dimensionada considerando un peso total de 1 800 kg con un uso de 1,5 horas al día y un recorrido de 7,6 km que representan un viaje ida y vuelta en el tour de las Tintoreras en la Isla Isabela. Su implementación muestra el interés del Ecuador y del mundo en desarrollar e implementar tecnologías amigables con el medio ambiente (Jara-Alvear et al., 2013).

El gobierno nacional del Ecuador se encuentra preocupado acerca de la viabilidad ecológica, económica y social respecto a las actividades diarias que realiza el ser humano en las Islas Galápagos (Jose Alvear, 2013). Una de ellas y la más problemática también es el transporte marítimo de pasajeros que amenaza al ecosistema marítimo y terrestre (Muñoz Barriga, 2015). Por ello las autoridades de Ecuador impulsan el camino hacia una movilidad marítima sustentable basada en la sustitución de fuentes de energía de origen fósil por fuentes de energías renovables (Cuvi et al., 2016).

El Instituto de Investigación Geológico y Energético empeñado en la implementación de tecnologías eficientes y amigables con el medio ambiente, desarrolló una embarcación prototipo tipo catamarán basado en energía solar denominado “INER 1” el cual se encuentra

en operación desde octubre de 2015. Para disminuir el daño ambiental que está realizando el ser humano se ha implementado un modelo sostenible y renovable para el transporte marítimo de pasajeros en las Islas reemplazando las barcazas convencionales que son contaminantes con un prototipo tipo catamarán propulsado por motores eléctricos alimentado por energía solar que se encarga de transportar a los pasajeros provenientes del aeropuerto de la Isla Baltra a la Isla Santa Cruz a través del canal de Itabaca. El catamarán solar INER 1 mostrado en la Figura 1 puede transportar un máximo de 40 pasajeros sentados y 2 en silla de ruedas con su respectivo equipaje en 8 recorridos de 0.6 km cada uno. Posee una potencia fotovoltaica instalada de 4,2 kWp con una capacidad de almacenamiento de 11,5 kWh @ 2h y un consumo máximo de 3 kWh por día de la red eléctrica de la isla Baltra la cual se basa en energía eólica. El canal de Itabaca donde la embarcación realiza su actividad de transporte de pasajeros posee una radiación solar directa promedio de 600 Wh/m^2 (Moya et al., 2015).

Figura 1. Catamarán Solar “INER 1”.

Fuente: Recuperado del Instituto de Investigación Geológico Energético agosto 2019.

Touati et al. (2017) en uno de sus trabajos menciona que uno de los mayores problemas de las energías renovables es la intermitencia del recurso lo cual impide la

penetración de dicha energía en los sistemas de potencia. Para contrarrestar esta restricción, se han desarrollado varios métodos de predicción analizados en términos de su rendimiento de acuerdo al tipo de tecnología en el que se requiera aplicar. Dichos sistemas de predicción permiten estimar la cantidad de energía que puede producir una fuente renovable en determinadas condiciones y escenarios de producción (Salcedo-Sanz et al., 2018).

De acuerdo a Hatwaambo et al. (2009) la obtención de parámetros climatológicos en lugares donde no se tiene acceso a una estación meteorológica como las Islas Galápagos, es posible mediante la plataforma Meteonorm ya que los cálculos realizados a través de ella se encuentran basados en promedios de bases de datos de alrededor de 8000 estaciones meteorológicas en el mundo y 5 satélites geoestacionarios en periodos mayores a 10 años. Para la radiación solar el sistema brinda valores máximos en condiciones en las que el cielo se encuentra despejado (Zhang et al., 2018). Según Meteonorm (2019), las comparaciones de los valores de radiación solar muestreadas para períodos más largos presentan una variación de menos del 2 % para todas las estaciones meteorológicas, pero los modelos matemáticos computacionales muestran un error menor que la variación en la radiación total medida entre un año y el otro.

Okello et al. (2015) en su trabajo menciona que los datos meteorológicos derivados de la plataforma Meteonorm poseen muy pequeñas desviaciones ya que sus modelos utilizan datos promedio de periodos entre (1985 - 2005), esto permitió realizar análisis de sistemas fotovoltaicos con una alta precisión ya que el rendimiento presentado en la simulación oscila entre los 5 754 y 5 771 kWh/año y los datos reales medidos por el inversor solar de la instalación fotovoltaica es 5 757 kWh/año con una variación del 3 % entre los datos medidos y los datos generados mediante la plataforma.

Minería de Datos.

En referencia al análisis del comportamiento de variables climatológicas, energéticas

y sociales se utiliza el método de selección de variables o conocido también como identificación de parámetros de entrada más importantes en un sistema. Este método de análisis permite mejorar la capacidad de predicción de sus variables de más peso, dicho problema ha sido investigado de forma amplia en los campos de la informática y aprendizaje automático, donde recibe el nombre de problema de selección de características (FSP) (Salcedo-Sanz et al., 2018).

La selección de atributos forma parte de todas las técnicas utilizadas en la minería de datos que como concepto general consiste en aplicar varias metodologías de análisis en grandes volúmenes de información para encontrar información que sea útil, aplicable y no aleatoria. Por lo tanto la minería de datos se encarga de buscar las relaciones y patrones entre toda la mayor cantidad de datos o series de datos que se disponga para analizar (Witten et al., 2016). Un patrón es la representación de una tendencia de datos que se repiten que son obtenidos de una base de datos o fuente de información, para que se lo defina de tal forma debe cumplir una serie de características tales como: coherencia con el tema de investigación, adaptabilidad a una gran cantidad de datos, no ser trivial, nuevo y comprensible. En este sentido lo que define a un patrón es su adaptabilidad con una gran cantidad de datos y su facilidad para relacionarse con ellos (Tan, 2018).

La minería de datos utiliza dos métodos los predictivos que buscan comprender el comportamiento de varias variables para predecir su futuro comportamiento dentro de los más conocidos se tiene los árboles de decisión y lo métodos basados en regresión matemática (Frank et al., 2009). Los métodos descriptivos que se enfocan en encontrar patrones que muestren la información que se requiere y en el hecho de que puedan ser entendidas por el ser humano para esto no precisan registros de datos y no dependen de patrones, reglas o correlaciones; el método descriptivo más conocido es el *clustering* o agrupamiento (Capozzoli et al., 2015).

El método de selección de atributos consiste en filtrar las características más relevantes del conjunto de datos principal que describa mejor el comportamiento de todos sus objetos, su función es reducir las dimensiones de la base de datos descartando así la información que no es relevante o redundante para la representación de la información, esto permite disminuir el costo computacional y una generalización del algoritmo de aprendizaje en su etapa inicial (Rodríguez et al., 2014). La selección de atributos permite utilizar diferentes combinaciones de características y observar cual es el comportamiento global del sistema. Es ampliamente utilizada en el aprendizaje de máquina, reconocimiento de imágenes, entre otros (Chandrashekar et al., 2014). Se forma de dos componentes principales que son la función de evaluación y el método de generación de subconjuntos basados en procesos de búsqueda (Rodríguez et al., 2014).

De acuerdo a Jain et al. (1997), existen varios algoritmos para la selección de atributos donde se considera la probabilidad del error como una función dependiente del tipo de clasificador usado, el tamaño de los datos de entrenamiento y los datos de prueba utilizados, en base a esto presenta la taxonomía disponible para los métodos de selección de algoritmos mostrados a continuación:

Algoritmos determinísticos de una sola solución, los métodos más comunes para la selección de atributos pueden ser divididos en dos categorías los algoritmos donde se inicia sin atributos y entre cada iteración se los puede ir incluyendo, y los que empiezan configurados con un conjunto lleno de atributos y se los va eliminando en cada iteración de acuerdo al error obtenido. Este tipo de algoritmos no garantiza que sus resultados sean los más óptimos para el sistema (Guyon et al., 2003).

Algoritmos determinísticos de múltiples soluciones, este método utiliza búsquedas iniciales en el conjunto de atributos manteniendo una secuencia ordenada de las posibles soluciones, este tipo de algoritmos no se encuentran descritos de forma clara en estudios

previos (Chandrashekar et al., 2014).

Algoritmos estocásticos de múltiple solución, este método introduce los algoritmos genéticos en la selección de atributos, un conjunto inicial de atributos se introduce representándolo como un arreglo binario, donde el número total de atributos se consideran cromosomas y cada uno de ellos es evaluado por su comportamiento de sobrevivir y adaptarse a la siguiente generación, los nuevos cromosomas son creados en base a los anteriores por procesos de cruce y mutación (Almaraashi, 2018).

Los algoritmos óptimos utilizan métodos basados en una técnica denominada *branch and bound* (Rama y limite) usado para buscar el conjunto de atributos óptimos mucho más rápido que una búsqueda exhaustiva. Este método resulta difícil de aplicar cuando se tiene conjuntos grandes de atributos debido a que la complejidad del algoritmo se vuelve exponencial (Dash et al., 1997).

Algoritmo tipo *Node Pruning* (nodo de poda), este método utiliza una red neuronal tipo *backpropagation* retroalimentada basada en un algoritmo de aprendizaje para clasificación por patrones el cual permite remover los atributos que no correspondan de su conjunto y optimizar la clasificación. Utiliza la función del error medio cuadrático en el entrenamiento de la red en cada iteración hasta que el error y el tamaño de la red sea despreciable (Leary et al., 2017).

Identificación de atributos.

Los algoritmos de minería de datos se encuentran en función de la calidad de la información de entrada, esto se debe a que si alguno de estos parámetros contiene información irrelevante, redundante o datos ruidosos el proceso de aprendizaje del algoritmo se volverá complejo y requerirá de una carga computacional elevada (Martínez, 2017). Debido a ello, las técnicas de selección de atributos permiten filtrar y eliminar la

información que sea irrelevante y redundante, tomando como punto de partida los datos de entrada en función de optimizar los algoritmos de minería de datos sobre un conjunto de datos (Aguilar–Ruiz et al., 2005).

De acuerdo a Lessmann et al. (2015) la selección de un clasificador de conjuntos homogéneos agrupan las predicciones de múltiplos modelos que sirven de base. En ese sentido, los desarrollos tanto empíricos como teóricos han mostrado que la combinación de varios modelos incrementa la precisión de la predicción, donde el aprendizaje crea conjuntos homogéneos de forma dependiente o independiente. Un ejemplo de algoritmos dependientes son los algoritmos de impulso que agregan de forma iterativa modelos de base que están entrenados para disminuir el error de predicción. El estudio se enfoca en el análisis y comparación de métodos simples de baja carga computacional que poseen resultados precisos comparados con métodos sofisticados y de alta capacidad computacional mismos que no aseguran obtener la precisión necesaria para tomar el resultado como válido, este tipo de algoritmos son desarrollados para aplicaciones específicas donde su rendimiento mejora de forma notable.

Evaluador de atributos.

De acuerdo al trabajo de investigación presentado por Mina (2018) donde analiza los algoritmos de clasificación y selección de atributos ordenados de acuerdo al área de investigación y a los resultados obtenidos, se considera que para el análisis de parámetros eléctricos se recomienda la utilización del algoritmo para la medición de la ganancia de la información de cada atributo para obtener el peso de cada uno de ellos y su influencia en el arreglo total de datos. El algoritmo de clasificación *InfoGainAttributeEval*, realiza la evaluación de los atributos midiendo la ganancia de la información que posee cada uno con respecto a la clase, el algoritmo de forma previa requiere discretizar los atributos numéricos

(Navarro et al., 2001). Para cumplir con los parámetros que requiere el modelo, se recomienda utilizar el modo de selección de atributos tipo *ranker* el cual se encarga de entregar un rango a cada uno de los atributos atendiendo a su importancia individual (Gutiérrez, 2016).

El método de búsqueda es el encargado de determinar el estilo en que la búsqueda es realizada (Arulmurugan et al., 2018). La selección de atributos de forma adicional puede ser realizada en dos formas:

- Usando todo el conjunto de datos como conjunto de entrenamiento (*training set*).
- Usando validación cruzada, donde el conjunto es dividido, y una parte es utilizado para el entrenamiento del sistema (*cross validation: CV*) (Salcedo-Sanz et al., 2018).

Pre-procesamiento de información.

En el trabajo presentado por Leary et al. (2017) la creación y gestión de una base de datos donde se almacenen todas la recopilación de variables tanto auxiliares como principales de forma histórica generadas con el proceso de identificación de parámetros, es uno de los procesos que demanda una mayor prioridad ya que desde ahí se generara los análisis estadísticos que se requiera. Las variables ordenadas y filtradas se deben tratar mediante un software especializado en la minería de datos, en este caso se propone el uso del software libre WEKA (Waikato Enviroment for Knowledge Analysis) desarrollado por la Universidad de Waikato Hamilton, New Zealand.

El software Weka posee un grupo de técnicas de filtrado, análisis, clasificación y agrupamiento que se pueden aplicar con éxito a múltiples campos, tales como el marketing, procesos de manufactura asistida, salud, generación y consumo de energía, finanzas, medicina, entre otros; para su aplicación se debe de tener en cuenta el tipo de tarea que se desea realizar. Para ello, la información debe pasar por una etapa de pre-procesamiento,

selección de atributos y visualización de los resultados obtenidos. Los métodos que más utilizados son los siguientes (Frank et al., 2009):

- Métodos de clasificación basados en redes neuronales.
- Manipulación de métodos numéricos sobre datos (Funciones estadísticas).
- Métodos de clasificación basados en máquinas de soporte vectorial.
- Metaclasificadores.
- Métodos de implementación de árboles de decisión.
- Implementación de agrupaciones.
- Métodos de estimación de probabilidades.
- Clases para implementar clasificadores bayesianos.

Redes neuronales artificiales.

De acuerdo a Gutiérrez (2016) un algoritmo de clasificación de datos permite ordenar y predecir el comportamiento general de un conjunto de datos al que pertenecen mediante el análisis de sus relaciones y la estimación de datos estadísticos como medias, error relativo, error absoluto y varianzas.

De acuerdo a Navarro et al. (2001), el estudio de las redes neuronales artificiales permite conocer que el peso asociado a una sinapsis o conexión entre dos neuronas se incrementa de acuerdo a cada activación que proviene de otra neurona mediante la misma sinapsis de acuerdo al entrenamiento que esta reciba hasta que el error medio cuadrático sea el menor. Un perceptrón es la unidad más simple de una red neuronal y se puede utilizar para la clasificación de clases o conceptos linealmente separables, su funcionamiento consiste en la suma ponderada (función de activación) de las entradas multiplicadas por un peso en cada una de ellas como se puede ver en la ecuación 1.

$$v(x)=\sum_{i=0}^{n} w_i x_i \quad (1)$$

Donde:

$V_{(X)}$= Salida de la red neuronal.

W_i= Peso sináptico.

X_i= Datos de entrada.

Σ=Función de activación.

En la Figura 2, se presenta el modelo conceptualizado de un perceptrón de In entradas, Wn pesos, su función de activación y una salida O. El aprendizaje del perceptrón consiste en generar los pesos de cada una de las entradas y el valor del peso umbral, para esto se utiliza la regla de entrenamiento del perceptrón que muestra la convergencia a una solución determinada de pasos si las clases son linealmente separables.

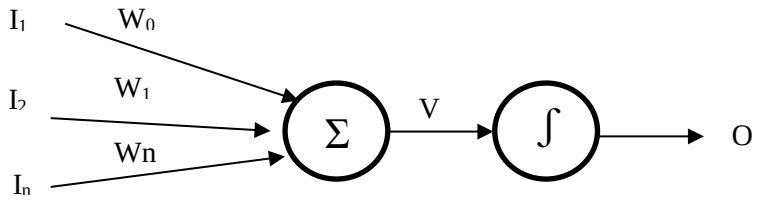

Figura 2. Perceptrón una sola capa.

El perceptrón simple es solucionable solo si las clases son linealmente separables, cuando una clase de datos en su clasificación no lo es se requiere la utilización de una configuración de red neuronal tipo perceptrón multicapa cuya estructura se presenta en la Figura 3.

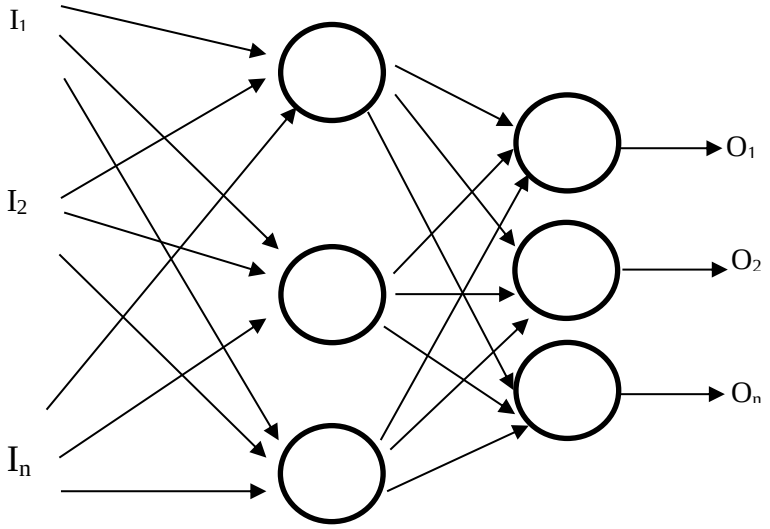

Figura 3. Perceptrón Multicapa.

El algoritmo de aprendizaje más utilizado en la red perceptrón multicapa es la retro-propagación del error o conocida como *Backpropagation,* la cual utiliza la suma de los errores cuadráticos de los valores de las muestras como el valor a optimizar como se muestra en la ecuación 2.

$$E(w)=\frac{1}{2}\sum(y-o)^2 \tag{2}$$

Donde:

o= Salida de la red en función de valores de prueba.

y= Salida esperada

W= Pesos sinápticos de la redneuronal.

Catamarán solar "INER 1"

El uso de las energías renovables es utilizado de forma extensa en varias aplicaciones una de ellas es los sistemas de transporte marítimos, a pesar sus beneficios en este caso particular de la energía solar – fotovoltaica, existen diferentes razones técnicas, ambientales,

energéticas y sociales que pueden llegar a ser una limitante en el uso de ella, uno de estos factores talvez el de más importancia, es la intermitencia del recurso solar que impide que los sistemas solares desempeñen un papel fundamental en el desarrollo sustentable en este caso del transporte marítimo de pasajeros (Zhang et al., 2018).

Dicha intermitencia en el recurso requiere del desarrollo de sistemas precisos de predicción que permitan estimar y gestionar de forma adecuada la energía renovable. En este contexto y considerando que la embarcación INER 1 requiere de la utilización de un punto de recarga que se encuentra conectado a la red eléctrica de las Islas Galápagos – Santa Cruz donde existen sistemas de generación renovable (solar y eólico) y generación termoeléctrica, junto a la variabilidad de parámetros climatológicos y ambientales como velocidad y dirección de viento, temperatura, humedad, radiación solar y mareas es importante desarrollar un estudio que permita identificar que variables son las de mayor peso y cuales afectan más a la eficiencia energética en la operación de la embarcación INER 1 en el Canal de Itabaca Isla Santa Cruz.

De acuerdo a la investigación realizada por Moya et al. (2015), existe un uso deficiente de la energía en su mayoría proveniente de combustibles de origen fósil en el transporte marítimo de pasajeros en el canal de Itabaca, esto debido a que en promedio las embarcaciones convencionales consumen anualmente 4 200 galones de combustible y producen una emanación de 38 toneladas de CO_2 por año, considerando que el combustible utilizado es transportado desde el continente hacia la provincia de Galápagos. Por lo tanto, una solución a este problema es el desarrollo de medios de transporte sustentables basados en energías renovables, de forma específica la energía solar; esto se debe a la gran cantidad de recurso existente en las Islas Galápagos (INER, 2014).

Según Khmaissia et al. (2018), existen varios modelos predictivos propuestos y analizados en base a su rendimiento para cada tecnología, en el estudio de los sistemas de

predicción una parte muy importante es la selección de las variables de entrada más influyentes en los resultados de un sistema de predicción.

Parámetros meteorológicos.

Según Jurado et al. (2015) las variables meteorológicas a considerar para un estudio de selección de atributos deben ser: radiación solar global, temperatura y humedad ambiental, velocidad y dirección del viento, datos que pueden generarse a través del software Meteonorm (Kumar et al., 2017), junto a la recopilación de tablas de mareas que se encuentran en el repositorio digital del Instituto Oceanográfico de la Armada (INOCAR, 2016). De acuerdo a del Álamo (2016) el método de los duodécimos permite obtener datos intermedios entre los valores de la pleamar y la bajamar utilizando el comportamiento sinusoidal del mar para aproximar sus valores a una medición horaria.

Variables energéticas.

Según (Zhang et al., 2018) los datos necesarios para analizar el desempeño energético de un instalación solar fotovoltaica requiere de: la generación de energía de la red eléctrica completa considerando todas las tecnologías posibles, la energía fotovoltaica generada por el sistema y absorbida por el banco de baterías de la instalación y la energía fotovoltaica entregada para consumo o sincronización hacia la red eléctrica estos parámetros deben ser medidos y registrados en intervalos de una hora. Además se debe considerar la información relevante respecto a la generación eléctrica de otros sistemas de generación como instalaciones eólicas, plantas termoeléctricas, centrales hidráulicas, entre las principales.

Respecto a la información energética de la embarcación “INER 1” , de acuerdo a Moya et al. (2015) la embarcación solar posee un sistema de gestión de energía el cual permite almacenar la información sobre su comportamiento tanto de su generación, consumo y almacenamiento mostrada en la Figura 4. Para el presente estudio se consideró la energía

consumida por el sistema de propulsión de la embarcación que se encuentra compuesto por dos sistemas eléctricos de impulsión y dirección de 11,8 kW cada uno.

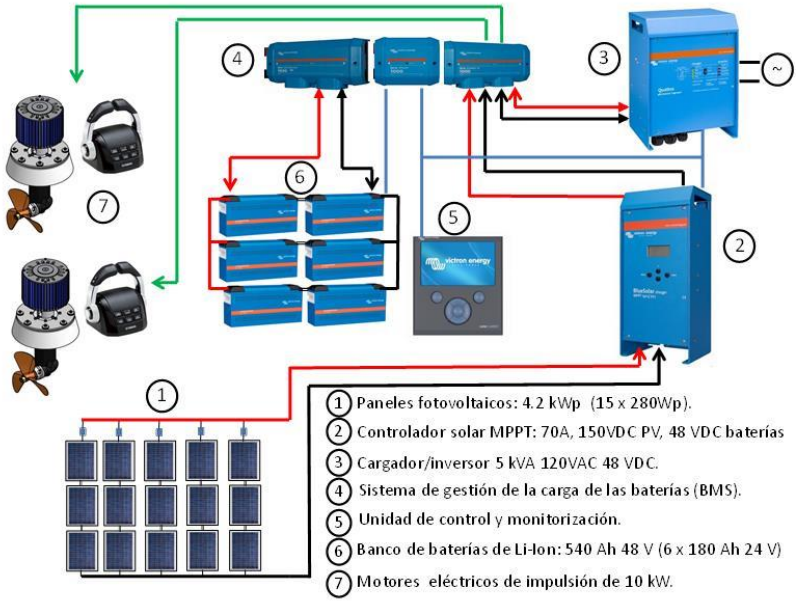

Figura 4. Sistema de gestión energética de la embarcación solar "INER 1".

Fuente: Moya et al. (2015).

Variables sociales.

Se consideran como variables sociales los movimientos humanos como migración y turismo interno o externo que puede afectar al comportamiento de la embarcación en un periodo determinado de tiempo (Saeteros et al., 2017). Esta información se encuentra disponible en la Dirección del Parque Nacional Galápagos en el informes de visitantes anuales publicado en el año 2017 con año base 2016 (Parque-Nacional-Galápagos, 2017) con un total de 218 365 arribos de turistas internos o externos y movimiento propio de la población de las Islas.

El ingreso las islas Galápagos se lo realiza vía aérea con 3 empresas (TAME, LAN, Avianca) que manejan las frecuencias de los vuelos tanto de llegadas como de partidas en el horario desde las 9:00 a 16:00 con aproximadamente un vuelo por hora, en la se presen la

ruta que realizan las personas que arriban por la Isla de Baltra para llegar a Santa Cruz.

Figura 5. Ruta de ingreso a Santa Cruz – Islas Galápagos.

Fuente: Recuperado de Google Maps agosto 2019.

Alcance de la investigación

La investigación se centrara en definir las variables ambientales, energéticas y sociales de mayor influencia en el rendimiento del sistema eléctrico del catamarán solar "INER 1" utilizando métodos para la selección de atributos, para ello se cuenta con datos de radiación solar, temperatura ambiental, tablas de mareas, generación fotovoltaica, capacidad de almacenamiento del banco de baterías, energía consumida desde el punto de recarga de la embarcación y energía consumida en el transporte de pasajeros, además de la cantidad de pasajeros transportados de acuerdo a las temporadas turísticas de las islas del periodo 2016 a 2017.

Justificación de la investigación

De acuerdo a la estacionalidad climática de las Islas Galápagos la mayor afluencia de turistas es en las temporadas seca y en la temporada húmeda, las cuales son las más apetecidas por el turismo local e internacional. Estas estaciones vienen acompañadas con la variación de la radiación solar, cambio brusco de las mareas junto a la variación de la velocidad y dirección del viento lo cual provoca un incremento en el consumo energético de la embarcación solar fotovoltaica INER 1. Se plantea la identificación de una o varias combinaciones de variables que afecten el desempeño energético de la embarcación y a futuro desarrollar técnicas de control sea operacional o una mejora tecnológica en los sistemas de control de la embarcación para mejorar el desempeño energético de la misma.

Objetivos

Para el desarrollo del presente trabajo de titulación se ha planteado el siguiente objetivo general: formular un estudio que permita seleccionar los atributos más influyentes en la operación del catamarán solar "INER 1" en las Islas Galápagos que permitan generar estrategias de control tanto electrónico como operacional para mejorar su eficiencia en el uso de la energía. El estudio será la base para futuras investigaciones sobre la gestión y predicción del recurso solar que es la principal fuente de energía de la embarcación. El cual contara con los siguientes objetivos específicos:

- Recopilar, analizar y tratar la información base necesaria para la investigación.
- Procesar el conjunto de atributos en base al método más óptimo seleccionado.
- Analizar los resultados obtenidos.
- Realizar las conclusiones que engloban a los atributos más importantes o influyentes en la operación de la embarcación en las Islas Galápagos

Materiales y Métodos.

La metodología seleccionada para el desarrollo de la investigación es la utilizada por Khmaissia et al. (2018) en su investigación, la cual se basa en el *Knowledge Discovery in Databases (KDD)* que es el concepto que recibe el proceso de múltiples pasos para identificar patrones en bases de datos que permitan identificar nueva información.

El proceso de descubrimiento de conocimiento en bases de datos o KDD consta de varios pasos mostrados en la Figura 6 que pueden repetirse para obtener los mejores conjuntos de información y requiere de la intervención del usuario experto enfatizando su naturaleza interactiva. Se compone de 6 pasos principales que se detallan a continuación (D'Oca et al., 2015):

1. Selección de Datos, corresponde a la creación o selección de un conjunto de datos o subconjunto de variables en el que se realizara el estudio.
2. Limpieza y procesamiento de información, trata las estrategias para completar campos de información faltantes o eliminación de ruido dentro de la base de datos.
3. Transformación de datos, encuentra características útiles para representar la información en función del objetivo de la investigación.
4. Minería de datos, trata de encontrar un método que permita buscar patrones en la base de datos.
5. Interpretación y evaluación de datos, decidir que parámetros o atributos pueden ser apropiados e interpretar los patrones minados.
6. Extracción del conocimiento, consisten en consolidar el conocimiento encontrado para realizar un análisis más detallado de su comportamiento.

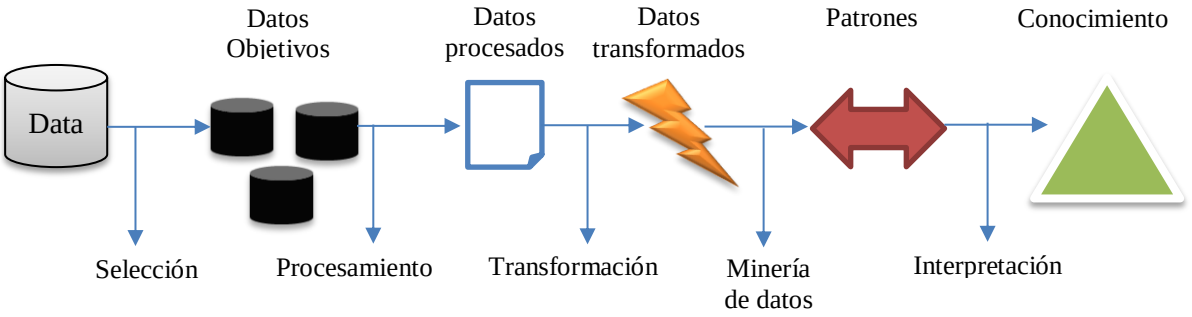

Figura 6. Representación gráfica del proceso descubrimiento de conocimiento en bases de datos (KDD)

De acuerdo a la metodología KDD, el primer paso a seguir es recopilar las series de datos experimentales considerando la disponibilidad y la calidad de la información además del uso de descriptores adecuados en función del tipo de información adquirida. En este caso se propone la interacción de parámetros de tiempo en relación con parámetros climatológicos, generación y consumos energéticos tanto de la embarcación como de la red eléctrica de Santa Cruz y Baltra además de la entrada de turistas y residentes a las Islas Galápagos por la terminal aérea de Baltra.

En este sentido y considerando los parámetros meteorológicos requeridos por el sistema, se generaron datos de temperatura ambiental, humedad relativa, radiación global, velocidad y dirección de viento y precipitación mediante la plataforma Meteonorm en el periodo desde el 01-01-2016 hasta el 31-12-2016 donde se han considerado intervalos de una hora entre cada medición considerado como punto central el canal de Itabaca en las Islas Galápagos. Además, se analizaron los datos existentes de las tablas de mareas que se encuentran en el repositorio digital del Instituto Oceanográfico de la Armada (INOCAR, 2016) y se realizó la aproximación a muestreo horario para tener todas las variables ambientales en un mismo formato.

La generación eléctrica de la red de Baltra Santa Cruz se encuentra compuesta por un sistema eólico de 2,25 MW instalado en la isla Baltra, un sistema fotovoltaico de 1,5 MWp instalado en Puerto Ayora - Santa Cruz y la central termoeléctrica de 3,4 MW instalada en Santa Cruz, esta información es gestionada por la empresa pública ELECGALÁPAGOS mediante la cual se ha recopilado la información de generación eléctrica en periodos de una hora.

De acuerdo a los datos presentados por la Dirección General del Parque Nacional Galápagos (DPNG, 2016), se registró que el 78 % que representa 170 255 personas del total que ingresaron a las Islas Galápagos lo realiza por el aeropuerto de la Isla Baltra y por lo tanto también cruzan en el mismo número de personas el canal de Itabaca. En este sentido, se debe considerar que respecto al año 2015 se registró un decrecimiento del 3 % que representa un total de 6 000 turistas siendo un año atípico en el comportamiento del turismo en las islas debido al terremoto del 16 de abril de 2016 en la costa Ecuatoriana y a la apreciación del dólar.

Según la metodología KDD planteada para el desarrollo de la investigación, el levantamiento y pre-procesamiento de la información permitirá obtener los atributos más influyentes en la operación de la embarcación solar INER 1 y analizar sus resultados para discutir y emitir conclusiones sobre el mismo. En este contexto, en la Tabla 1 se presenta las variables de entrada utilizadas (atributos), abreviaturas utilizadas en el procesamiento de la información junto a sus respectivas unidades de medida.

Tabla 1

Información de las Variables de Estudio

Variable de entrada	Abreviatura	Unidades
Fecha y hora	date	aaaa-dd-mm hh:mm:ss
Energía utilizada por la embarcación (red eléctrica Baltra - Santa Cruz, generación Fotovoltaica de la embarcación, banco de baterías)	Energ	kWh
Temperatura	Temp	°C
Humedad relativa	HR	%
Radiación global	GlRad	kWm^2/h
Dirección de viento	WD	° desde el norte
Velocidad de viento	WS	m/s
Precipitación	Prec	mm
Unidad Caterpilar 1	Term1	kWh
Unidad Caterpilar 3	Term3	kWh
Unidad Caterpilar 4	Term4	kWh
Unidad Caterpilar 5	Term5	kWh
Unidad Caterpilar 6	Term6	kWh
Unidad Caterpilar 7	Term7	kWh
Unidad Hyundai 8	Term8	kWh
Unidad Hyundai 9	Term9	kWh
Unidad Hyundai 10	Term10	kWh
Unidad Hyundai 11	Term11	kWh
Unidad Hyundai 12	Term12	kWh
Unidad Hyundai 13	Term13	kWh
Parque eólico Baltra	Eol	kWh
Parque Fotovoltaico Santa Cruz	Fotov	kWh
Altura del mar	Mar	m
Turistas	Tur	personas

La creación de la base de datos requiere de una configuración especial en formato ARFF que maneja el software WEKA (Frank et al., 2009). Para generar el archivo base se utilizó un fichero CSV *Comma-separated values* donde se encuentran los datos ordenados de forma horaria como se muestra en la Tabla 2.

Tabla 2

Ejemplo de los Datos Utilizados en el Estudio

Timestamp	Energ	Temp	HR	GlRad	WD	WS	Prec	Term 1	Term 3	Term 4	Term 5	Term 6	Term 7	Term 8	Term 9	Term 10	Term 11	Term 12	Term 13	Eol	Fotov	Mar	Tur	Class
'2016-01-01 00:00:00'	0.01	25,8	82	0	173	1.2	45	0	0	0	670	445	0	1229	1383	0	0	0	0	0	0	0.6	0	Apagado
'2016-01-01 03:00:00'	0.01	26,2	78	0	159	0.7	43	0	0	0	590	405	0	1188	1342	0	0	0	0	0	0	0.675	0	Apagado
'2016-01-01 06:00:00'	0.01	26	65	0	162	1.2	35	0	0	0	670	0	0	1408	1404	0	0	0	0	0	0	0.825	0	Apagado
'2016-01-01 07:00:00'	0.31	25,9	78	0	186	0.9	42	0	0	0	700	0	0	1280	1315	0	0	0	0	0	0	1.7	0	Apagado
'2016-01-01 08:00:00'	1.11	26	62	0.011	192	1.6	33	0	0	0	620	435	0	1100	1080	0	0	0	0	0	0	1.625	0	Bajo
'2016-01-01 09:00:00'	1.51	26,3	62	0.071	169	1.3	34	0	0	0	600	435	0	950	950	0	0	0	0	0	0	1.475	0	Bajo
'2016-01-01 10:00:00'	1.69	26,6	62	0.126	208	3	34	0	0	0	600	435	0	1050	1080	0	0	0	0	0	0	1.255	0	Bajo
'2016-01-01 11:00:00'	1.92	26,9	76	0.192	185	2.4	44	0	0	0	610	435	0	1080	970	0	0	0	0	0	0	1.035	0	Bajo
'2016-01-01 12:00:00'	2.6	27,4	59	0.307	226	2.2	34	0	0	0	550	440	0	1000	950	0	0	0	0	0	0	0.885	13	Bajo
'2016-01-01 13:00:00'	1.81	27,9	55	0.422	190	2.6	32	0	0	0	600	439	0	992	873	0	0	0	0	761	364	0.56	0	Bajo

Para el procesamiento de la información se planteó como atributo de salida del sistema la variable clase (class) la cual representa el consumo de energía de la embarcación mediante el uso de un atributo nominal donde se consideró los rangos bajo, medio bajo, medio alto y alto referentes al consumo energético de la embarcación. Como menciona Martínez (2017), para el planteamiento de la variable de salida de una selección de atributos se requiere plasmar la experiencia del observador que permita cuantificar de forma adecuada el sistema que se va a investigar.

Resultados.

La etapa de preclasificación de datos requiere de una estructuración en la base de datos con un formato llamado arff *Attribute-Relation File Format* el cual permite vincular el atributo su formato con los datos previamente parametrizados en el archivo csv. Para formar el archivo arff se debe ingresar un encabezado en la primera parte de la base de datos donde debe constar el nombre del programa, autor, relaciones y atributos con su nombre abreviado junto a su respectivo tipo de formato quedando de la siguiente forma:

```
% CATAMARAN SOLAR RELACION ENERGIA VARIABLES AMBIENTALES y SOCIALES
% Autor Marcelo Moya
@relation cat
@attribute date date 'yyyy-MM-dd"HH:mm'
@attribute Energ numeric
@attribute Temp numeric
@attribute HR numeric
@attribute GlRad numeric
@attribute WD numeric
@attribute WS numeric
@attribute Prec numeric
@attribute Term1 numeric
@attribute Term3 numeric
@attribute Term4 numeric
@attribute Term5 numeric
@attribute Term6 numeric
@attribute Term7 numeric
@attribute Term8 numeric
@attribute Term9 numeric
@attribute Term10 numeric
@attribute Term11 numeric
@attribute Term12 numeric
@attribute Term13 numeric
```

@attribute Eol numeric

@attribute Fotov numeric

@attribute Mar numeric

@attribute Tur numeric

@attribute class (Alto,MedAlto,MedBajo,Bajo,Apagado)

En la Figura 7, se muestra la estructuración final del archivo arff con el encabezado y a continuación los datos ingresados en el mismo orden de cada variable que se encuentra en el encabezado separados por una coma.

```
% CATAMARAN SOLAR RELACION ENERGIA VARIABLES AMBIENTALES y SOCIALES
% Autor Marcelo Moya
@relation cat
@attribute date date 'yyyy-MM-dd''HH:mm'
@attribute Energ numeric
@attribute Temp numeric
@attribute HR numeric
@attribute GlRad numeric
@attribute WD numeric
@attribute WS numeric
@attribute Prec numeric
@attribute Term1 numeric
@attribute Term3 numeric
@attribute Term4 numeric
@attribute Term5 numeric
@attribute Term6 numeric
@attribute Term7 numeric
@attribute Term8 numeric
@attribute Term9 numeric
@attribute Term10 numeric
@attribute Term11 numeric
@attribute Term12 numeric
@attribute Term13 numeric
@attribute Eol numeric
@attribute Fotov numeric
@attribute Mar numeric
@attribute Tur numeric
@attribute class {Alto,MedAlto,MedBajo,Bajo,Apagado}

@data
'2016-01-01 00:00:00',0.01,25.8,82,0,173,1.2,45,0,0,0,670,445,0,1229,1383,0,0,0,0,0,0.6,0,Apagado
'2016-01-01 03:00:00',0.01,26.2,78,0,159,0.7,43,0,0,0,590,405,0,1188,1342,0,0,0,0,0,0.675,0,Apagado
'2016-01-01 06:00:00',0.01,26,65,0,162,1.2,35,0,0,0,670,0,0,1408,1404,0,0,0,0,0,0.825,0,Apagado
'2016-01-01 07:00:00',0.31,25.9,78,0,186,0.9,42,0,0,0,700,0,0,1280,1315,0,0,0,0,0,1.7,0,Apagado
'2016-01-01 08:00:00',1.11,26,62,0.011,192,1.6,33,0,0,0,620,435,0,1100,1080,0,0,0,0,0,1.625,0,Bajo
'2016-01-01 09:00:00',1.51,26.3,62,0.071,169,1.3,34,0,0,0,600,435,0,950,950,0,0,0,0,0,1.475,0,Bajo
'2016-01-01 10:00:00',1.69,26.6,62,0.126,208,3,34,0,0,0,600,435,0,1050,1080,0,0,0,0,0,1.255,0,Bajo
'2016-01-01 11:00:00',1.92,26.9,76,0.192,185,2.4,44,0,0,0,610,435,0,1080,970,0,0,0,0,0,1.035,0,Bajo
'2016-01-01 12:00:00',2.6,27.4,59,0.307,226,2.2,34,0,0,0,550,440,0,1000,950,0,0,0,0,0,0.885,13,Bajo
'2016-01-01 13:00:00',1.81,27.9,55,0.422,190,2.6,32,0,0,0,600,439,0,992,873,0,0,0,0,761,364,0.56,0,Bajo
'2016-01-01 14:00:00',0.86,28.2,53,0.421,148,3,32,0,0,0,600,440,0,1241,1086,0,0,0,0,522,285,0.635,0,Apagado
'2016-01-01 15:00:00',1.77,28.3,54,0.269,190,2.4,33,0,0,0,570,439,0,1278,1129,0,0,0,0,389,392,0.785,0,Bajo
'2016-01-01 16:00:00',0.51,28.1,68,0.178,221,3.7,42,0,0,0,500,440,0,1099,1033,0,0,0,0,814,151,1.005,0,Apagado
```

Figura 7. Muestra de la base de datos en formato arff.

A continuación se debe importar el archivo arff dentro del programa Weka Explorer donde se deberán solucionar los errores que genere el programa debido a inconsistencias en la información, datos faltantes o fuera de rango y errores de formato en los atributos, esto representa la última etapa de filtrado de información que permitirá generar la tabla de estadísticas y atributos mostrados en la Figura 8.

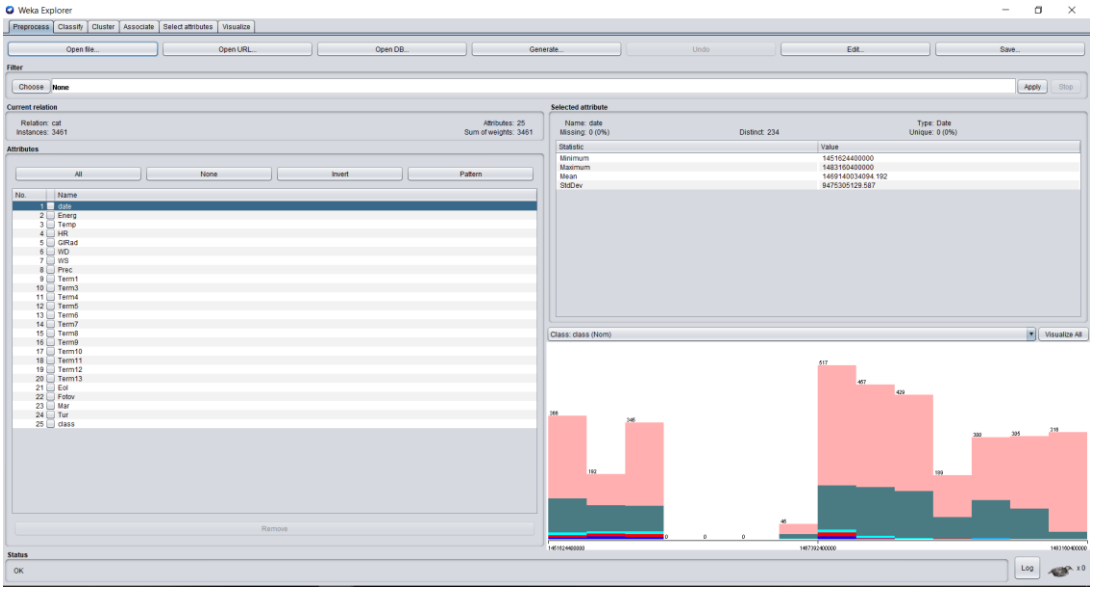

Figura 8. Tabla de atributos y datos estadísticos.

Luego de haber revisado a detalle y considerado todas las correcciones de los posibles errores en la información con la aplicación de filtros sea de relleno o de eliminación de datos, la etapa de pre-procesamiento de la información culmina con una base de datos adecuada para la investigación de selección de atributos.

En este sentido, y de acuerdo a lo mencionado en la introducción la configuración del evaluador de atributos se presenta en la Figura 9, donde se escoge el evaluador de atributos *(InfoGainAttributeEval)*, el método de búsqueda *(Ranker)* y el modo de selección *(Full training set)*.

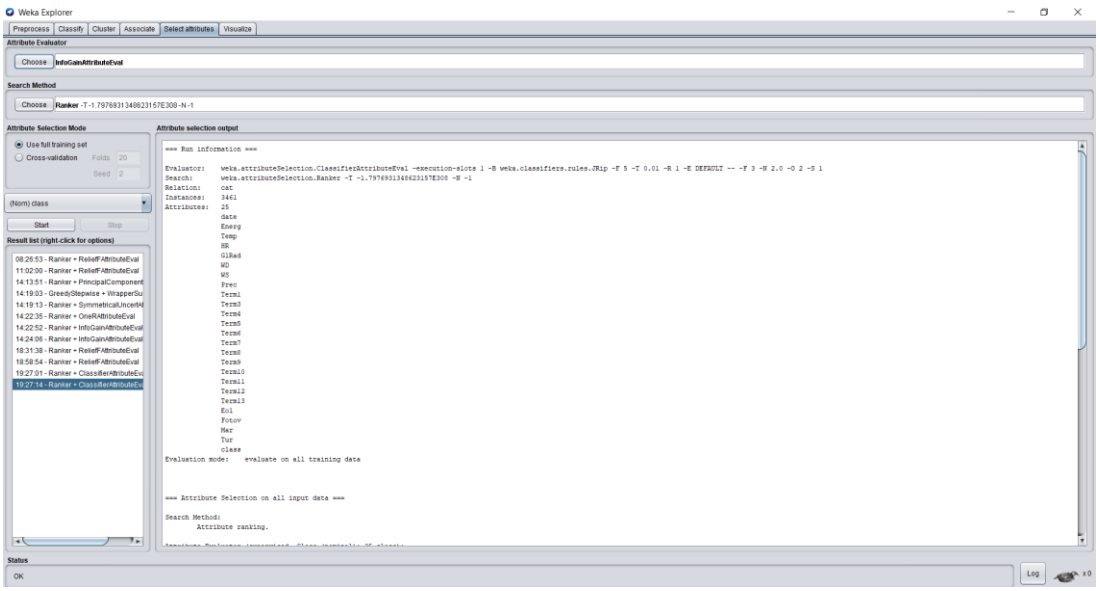

Figura 9. Configuración de parámetros para la identificación de atributos.

Luego de obtener el modelo entrenado y probado se continúa con la examinación a detalle de los atributos seleccionados y clasificados de acuerdo a su grado de importancia respecto al sistema. Las pruebas realizadas muestran que existen atributos que de forma espacial y temporal muestran un comportamiento independiente de la matriz de entrenamiento.

Análisis de atributos.

Para evaluar la calidad de los datos pre-procesados y sin discriminar ningún atributo se realiza una prueba de clasificación donde el objetivo es obtener el valor más alto de datos clasificados de forma correcta con el menor error. El algoritmo de clasificación utilizado fue una red neuronal tipo perceptrón multicapa con 29 capas ocultas que equivalen a la suma de las clases junto a los atributos y 5 nodos de salida que representan el consumo de energía en este caso son bajo, medio-bajo, medio-alto y alto.

Tabla 3

Clasificación de los 25 Atributos Pre-Procesados

Clasificador	Resultados clasificador		
MultiLayerPerceptron 29 capas ocultas Tipo de red *backpropagation* 5 nodos de salida	Instancias correctamente clasificadas	922	88,8247%
	Instancias mal clasificadas	116	11,1753%
	Estadística kappa	0,668	
	Error absoluto medio	0,0459	
	Error cuadrático medio	0,2081	
	Error absoluto relativo	26,4544%	
	Error cuadrático relativo	74,224%	
	Número total de instancias	1038	

Considerando que en la primera clasificación de la información pre-procesada con los 25 atributos se obtuvo un 88,8247% de valores correctamente clasificados se continúa con el proceso de selección de atributos para discriminar atributos que no tengan relación con la investigación.

Para el desarrollo de la discriminación de atributos se realizaron 60 experimentos en donde se utilizaron varias configuraciones y modelos de selección de atributos obteniendo en su mayoría resultados erróneos, no solucionables o clasificaciones donde la radiación solar y la energía eran excluidas de la selección. En este sentido y de acuerdo a la metodología y revisión bibliográfica establecida se analizaron a detalle los 3 evaluadores de atributos más utilizados. La configuración del primer evaluador se muestra en la Tabla 4.

Tabla 4

Configuración del Primer Evaluador de Atributos

Evaluador de atributos	**Modo de selección de atributos**	**Método de Búsqueda**	**Variable de estudio**
InfoGainAttributeEval	Full training set	Ranker	class

Como resultado del procesamiento del primer evaluador de atributos se obtuvo la información mostrada en la Tabla 5, donde se muestra el orden de clasificación de cada atributo y su peso de importancia en el conjunto de datos respecto a la variable de salida

class la cual representa el valor nominal del consumo de energía (bajo, medio ajo, medio alto, alto).

Tabla 5

Clasificación de Atributos Evaluador InfoGain.

Orden de clasificación	Atributo	InfoGain Merito
1	Energ	1,09798
2	Tur	0,57595
3	Fotov	0,18327
4	GlRad	0,10919
5	date	0,05443
6	Term9	0,02289
7	Term10	0,01495
8	Term11	0,01279
9	Term13	0,01208
10	Term3	0,01025
11	Prec	0,00955
12	Term5	0,00942
13	Term8	0,00848
14	Term12	0,0069
15	Mar	0,00671
16	Term4	0,00627
17	HR	0,00622
18	Term1	0
19	Eol	0
20	Temp	0
21	Term6	0
22	Term7	0
23	WS	0
23	WD	0

En la Figura 10, se muestra la gráfica de la relación que existe entre el peso asignado a la importancia de los atributos obtenidos por el método *InfoGain* y los atributos donde se puede discriminar los atributos que poseen una mayor influencia en el conjunto de datos.

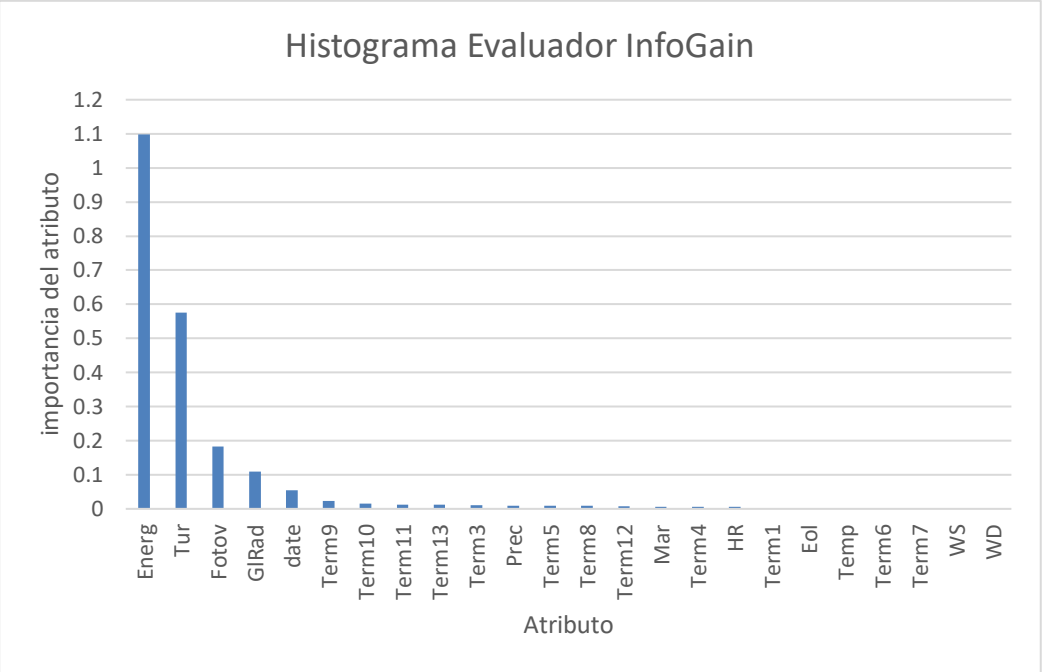

Figura 10. Importancia del atributo InfoGain vs atributos.

A continuación en la Tabla 6, se presenta la configuración utilizada en el segundo experimento de evaluación de atributos.

Tabla 6

Configuración del Segundo Evaluador de Atributos

Evaluador de atributos	**Modo de selección de atributos**	**Método de Búsqueda**	**Variable de estudio**
ClassifierAttibuteEval Jrip	Full training set	Ranker	class

El segundo evaluador de atributos genero la información mostrada en la Tabla 7, donde se encuentra el orden de clasificación de cada atributo y su peso de importancia en el conjunto de datos respecto a la variable de salida class la cual representa el valor nominal del consumo de energía (bajo, medio ajo, medio alto, alto).

Tabla 7

Clasificación de Atributos Evaluador Classifier Attribute Jrip

Orden de clasificación	Atributo	Classifier attribute Jrip Merito
1	Energ	0,3113551
2	Tur	0,182028316
3	Fotov	0,013002023
4	Term11	0,001059424
5	GlRad	0,000866802
6	date	0,000144467
7	WS	0
8	Term8	0
9	HR	0
10	Term1	-1E-16
11	Prec	-0,000144467
12	Term4	-0,000144467
13	Temp	-0,000433401
14	Term6	-0,000433401
15	Term7	-0,000433401
16	Term5	-0,000577868
17	Mar	-0,000722335
18	Eol	-0,000722335
19	WD	-0,000722335
20	Term9	-0,001300202
21	Term3	-0,001444669
22	Term13	-0,001589136
23	Term12	-0,001733603
23	Term10	-0,002455938

En la Figura 11, se muestra la gráfica de la relación que existe entre el peso asignado a la importancia de los atributos obtenidos por el método Classifier attribute Jrip y los atributos donde se puede discriminar los atributos que poseen una mayor influencia en el conjunto de datos.

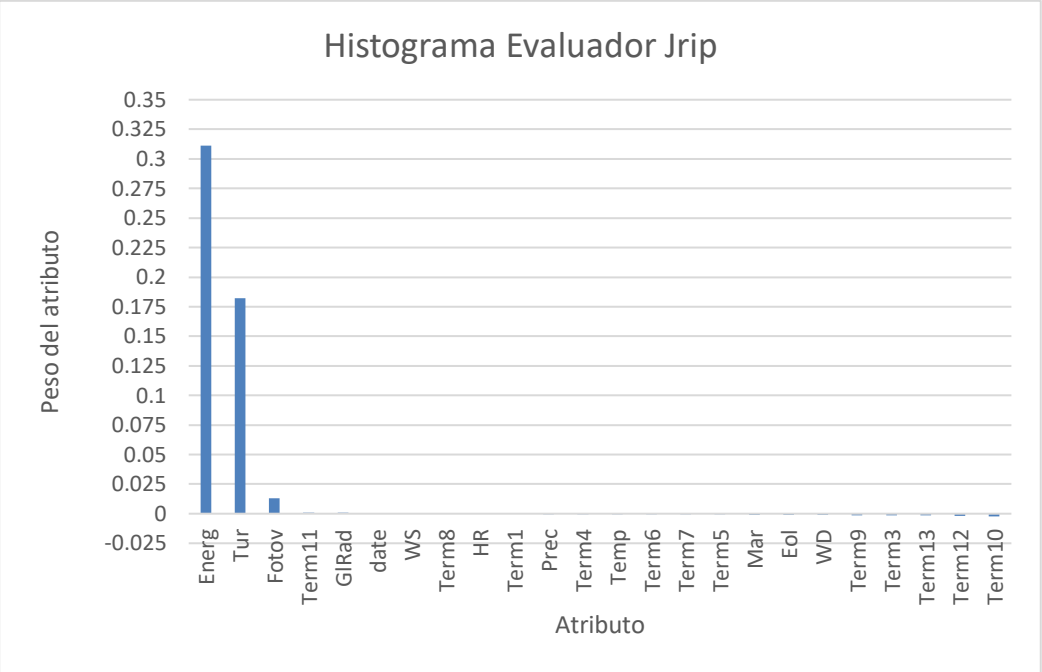

Figura 11. Importancia del atributo Jrip vs atributos.

Luego de haber realizado dos clasificaciones que se encuentran dentro de los parámetros deseados en función del comportamiento observado de la embarcación se continúa con un último evaluador que se recomienda en este tipo de estudios. Como se muestra en la Tabla 8 el evaluador *Relief* permite evaluar la información de los atributos pero considerando la influencia con los atributos de los conjuntos cercanos.

Tabla 8

Configuración del Tercer Evaluador de Atributos

Evaluador de atributos	Modo de selección de atributos	Método de Búsqueda	Variable de estudio
ReliefAttributeEval	Full training set	Ranker	class

Como resultado del procesamiento del tercer evaluador de atributos se obtuvo la información mostrada en la Tabla 9, donde se encuentra el orden de clasificación de cada atributo y su peso de importancia en el conjunto de datos en este caso en particular se

observa que el atributo fecha es discriminado de forma total de la clasificación lo cual de acuerdo a la experiencia del observador no es posible debido a que todos los parámetros meteorológicos, energéticos y el turismo son dependientes de forma directa de la fecha de su creación por lo tanto las mediciones diarias y horarias no pueden ser excluidas de la clasificación.

Tabla 9

Clasificación de Atributos Evaluador ReliefAttributeEval

Orden de clasificación	Atributo	ReliefAttributeEval Merito
1	Tur	0,135493
2	Energ	0,110779
3	Fotov	0,055994
4	GlRad	0,034984
5	Term3	0,028548
6	Eol	0,02564
7	Term9	0,024717
8	Mar	0,023221
9	HR	0,023154
10	Term8	0,019967
11	Term10	0,019596
12	Term11	0,018686
13	Temp	0,017447
14	Term4	0,017266
15	WS	0,016182
16	Term5	0,015636
17	Term12	0,014259
18	Term6	0,013598
19	Prec	0,012945
20	Term13	0,011907
21	WD	0,009496
22	Term1	0,006279
23	Term7	0,000108
24	date	0

A continuación, se debe evaluar la calidad de la selección de atributos para ello se considera utilizar los 6 atributos de mejor comportamiento en los dos primeros experimentos ya que el tercer experimento fue eliminado del estudio por falta de coherencia en el orden de selección de los atributos.

En la Tabla 10, se presenta los 6 atributos principales del evaluador *InfoGain* que van a ser utilizados en el experimento de clasificación para buscar incrementar los atributos clasificados de forma correcta.

Tabla 10

Atributos Principales Evaluador InfoGain

Orden de clasificación	Atributo	InfoGain Merito
1	Energ	1,09798
2	Tur	0,57595
3	Fotov	0,18327
4	GlRad	0,10919
5	date	0,05443
6	Term9	0,02289

Para evaluar la calidad de la selección de atributos se realiza una prueba de clasificación donde el objetivo es obtener el valor más alto de datos clasificados de forma correcta con el menor error. El algoritmo de clasificación mostrado en la Tabla 11, requirió de una red neuronal tipo perceptrón multicapa con 11 capas ocultas que equivalen a la suma de las clases junto a los atributos y 5 nodos de salida que representan el consumo de energía clasificado en 4 niveles bajo, medio-bajo, medio-alto y alto.

Tabla 11

Clasificación de los 6 Atributos con el Evaluador Infogain

Clasificador	Resultados clasificador		
	Instancias correctamente clasificadas	1027	98,9403%
	Instancias mal clasificadas	11	1,0597%
MultiLayerPerceptron	Estadística kappa	0,9725	
11 capas ocultas	Error absoluto medio	0,0057	
backpropagation	Error cuadrático medio	0,0495	
5 nodos de salida	Error absoluto relativo	3,2581%	
	Error cuadrático relativo	17,6365%	
	Número total de instancias	1038	

En la Tabla 12, se muestran los 6 atributos principales discriminados de acuerdo al evaluador *ClassifierAttributeEval Jrip* los cuales van a ser utilizados en el experimento de

clasificación para buscar incrementar el número de aciertos en la clasificación y disminuir el error al mínimo posible.

Tabla 12

Atributos Principales Evaluador Classifier attribute Jrip

Orden de clasificación	Atributo	Classifier attribute Jrip Merito
1	Energ	0,3113551
2	Tur	0,182028316
3	Fotov	0,013002023
4	Term11	0,001059424
5	GlRad	0,000866802
6	Term9	0,02289

La calidad del proceso de selección de atributos se la realiza mediante una prueba de clasificación mostrada en la Tabla 13, el algoritmo de clasificación utilizado fue una red neuronal tipo perceptrón multicapa con 11 capas ocultas que equivalen a la suma de las clases junto a los atributos y 5 nodos de salida que representan el consumo de energía clasificado en 4 niveles bajo, medio-bajo, medio-alto y alto.

Tabla 13

Clasificación de los 6 Atributos con el Evaluador Classifier Aattribute Jrip

Clasificador	Resultados clasificador		
	Instancias correctamente clasificadas	1000	96,3391%
	Instancias mal clasificadas	38	3,6609%
MultiLayerPerceptron	Estadística kappa	0,9725	
11 capas ocultas	Error absoluto medio	0,0057	
backpropagation	Error cuadrático medio	0,0495	
5 nodos de salida	Error absoluto relativo	3,2581%	
	Error cuadrático relativo	17,6365%	
	Número total de instancias	1038	

De acuerdo a los experimentos realizados y a los resultados obtenidos en la Tabla 11 usando el evaluador *InfoGain,* los atributos más influyentes en el comportamiento energético de la embarcación “INER 1” son: 1) Energ (Energía utilizada por la embarcación), 2) Tur (Turistas y residentes que ingresan o salen a las islas), 3) Fotov (Parque fotovoltaico Santa

Cruz), 4) Glrad (Radiación Global), 5) date (fecha y hora), 6) Term9 (Unidad termoeléctrica 9 de la central Santa Cruz).

Análisis de Energía

En la Tabla 14 se muestra un resumen del consumo y generación de energía de la embarcación solar "INER 1" en el año 2016 , se puede ver que la embarcación consume en su mayor parte energía solar sea para almacenar en baterías o para los servicios y movimiento de la embarcación en un total de 2 320,62 kWh considerando que la embarcación requiere 3 389,16 kWh, la energía restante se la absorbe de la red eléctrica de Baltra Santa Cruz por un total de 1 068,54 kWh de los cuales 788,62 kWh han sido para almacenamiento de energía en las baterías.

Tabla 14

Generación y Consumos Eléctricos de la Embarcación Solar "INER 1"

Red eléctrica a baterías kWh	Red eléctrica a consumidores kWh	Energía solar a baterías kWh	Energía solar a red eléctrica kWh	Energía solar a consumidores kWh	Energía total kWh
788,62	279,92	1451,64	0	868,98	3389,16

Discusión.

La presente investigación se encuentra dentro del campo general de la minería de datos y de forma específica en el ámbito de la selección de atributos, maneja la metodología KDD que permite realizar el descubrimiento de información y sus diferentes relaciones dentro de una base de datos, el uso experimental de una serie de combinaciones tanto de evaluadores como de clasificadores basados en la serie de revisiones bibliográficas realizadas consiguen obtener indicadores estadísticos con un rendimiento mucho mayor que estudios similares, en este sentido a continuación se presenta una comparación con estudios de relevancia. Almaraashi (2018) utilizó diferentes algoritmos de selección de atributos y redes neuronales artificiales para predecir la radiación solar diaria en 8 locaciones de Arabia Saudita, de acuerdo a los resultados experimentales obtenidos usando el evaluador tipo *Relief* el error cuadrático medio entre un algoritmo de predicción usando solo los atributos seleccionados es del 19,9 % comparado con un 52,1 % de un algoritmo de predicción usando todos los atributos del sistema, en este sentido la selección de atributos permite obtener resultados confiables de predicción del consumo de energía en un periodo de tiempo; por otro lado Leary et al. (2017) en su trabajo concluye que el uso de varias capas en una red neuronal artificial con varios escenarios de aprendizaje junto a una selección de los 4 principales atributos y un pre procesamiento de la información logra obtener una exactitud en la clasificación como predicción del 92,2 %. Como concluye Salcedo-Sanz et al. (2018) en su investigación el uso de selección de atributos y un pre-procesamiento de la información como primer paso para la predicción de energía permite incrementar el desempeño de un algoritmo de clasificación y recomienda el uso de varias combinaciones de técnicas que no requieren de una elevada carga computacional generando resultados adecuados al sistema que se encuentra analizando sea energía solar, eólica o mareomotriz. Respecto a la investigación realizada y considerando los mejores 6 atributos filtrados por el

algoritmo evaluador *InfoGain* se obtuvo un 4,95 % en el parámetro error cuadrático medio usando un algoritmo de clasificación y predicción *MultiLayerPerceptron* tipo *backpropagation con* 11 capas ocultas y un 98,94 % de exactitud en la clasificación y predicción, la combinación de técnicas desarrolladas han permitido obtener mejores resultados que las investigaciones mencionadas generando una base para desarrollar investigaciones sobre predicción del consumo de energía de la embarcación y control operacional respecto a la energía adicional de la red eléctrica que la embarcación requiere para cumplir su demanda anual no sea de origen fósil.

El resumen mostrado sobre el consumo y generación de energía eléctrica del sistema muestra que la embarcación requiere 1 068,54 kWh de energía adicional al año la cual proviene de la red eléctrica de Baltra Santa Cruz. Al momento la embarcación obtiene parte de esta energía de la unidad termoeléctrica 9 y del parque fotovoltaico de la central Santa Cruz. El porcentaje exacto de energía eléctrica proveniente de la unidad termoeléctrica 9 se lo puede obtener de un estudio de cargas y flujos eléctricos, esta mezcla de fuentes de energía disminuye le eficiencia energética global de la embarcación por usar una fuente de origen fósil.

Conclusiones.

La energía adicional requerida por la embarcación "INER 1" asciende a 1 068,54 kWh que provienen del parque fotovoltaico de Santa Cruz y de la unidad termoeléctrica 9 de la central de generación eléctrica de Santa Cruz; si se considera que el atributo de radiación solar es el cuarto más influyente sobre el sistema y si la radiación solar disminuye tanto en la embarcación como en el parque fotovoltaico el consumo de energía requerido vendrá en su mayoría de la central de generación termoeléctrica Santa Cruz.

El atributo Energía representa el consumo eléctrico de la embarcación que proviene de su generación fotovoltaica y de la red eléctrica tanto para almacenamiento de energía como para el consumo de los sistemas eléctricos de fuerza y control. La variable muestra el comportamiento de la embarcación ya que mantiene una relación directa con el atributo *date* que representa la fecha y hora y el atributo Tur que simboliza la cantidad de turistas que transporta la embarcación.

A medida que la cantidad de turistas y las frecuencias de navegación se incrementan el consumo de energía de la embarcación también aumenta, se requiere generar estrategias de control operacional que permitan a la embarcación utilizar la red eléctrica cuando el parque eólico de la isla Baltra se encuentre entregando energía eléctrica a la red ya que al momento el atributo viento y generación eólica no mantienen ninguna relación con los ciclos de carga y sus horas de uso.

El incremento o descenso de la marea no influye en el consumo de energía de la embarcación debido a que la variación en la altura se produce en centímetros por cada hora por lo tanto se observa que el atributo Mar no guarda ninguna relación con la energía. Cabe recalcar que en el estudio no se consideró la velocidad y dirección con la que se desplaza la corriente marina por la inexistencia de esa información.

El atributo radiación global Glrad guarda una relación directa con el comportamiento

energético de la embarcación considerando que la base de su generación energética es el sistema fotovoltaico de 4,2kWp que lleva en su cubierta superior junto a la central de generación fotovoltaica Santa Cruz y a la central termoeléctrica entregan la energía eléctrica con la que recarga la embarcación solar "INER 1".

Los atributos temperatura y humedad ambiental no mantienen ninguna relación con el comportamiento energético de la embarcación, su variación no afecta en el consumo de energía del sistema. Esto se debe a que la embarcación cuenta con sistemas de ventilación pasivos para disipar el calor generado en los motores y baterías.

Los atributos restantes que no fueron considerados en el estudio no mantienen ninguna relación con el consumo energético de la embarcación, de esta forma se logra obtener un conjunto de datos que servirán de base para implementar un modelo de predicción de consumo de energía considerando el incremento anual de turistas y la radiación solar que existe en la zona de operación de la embarcación.

Agradecimientos

La presente investigación forma parte del trabajo arduo y continuo del personal del Instituto de Investigación Geológico y Energético junto a sus direcciones de Transferencia Tecnológica y de Servicios Especializados que vieron nacer a la primera embarcación solar de Latinoamérica y la han mantenido a flote por de 5 años. También se agradece la valiosa aportación de información de la empresa pública CELEC EP encargada de la generación eléctrica y de la Dirección del Parque Nacional de Galápagos encargada de la contabilización si seguimiento de los turistas y residentes que ingresan a las islas.

Bibliografía.

Aguilar–Ruiz, J. S., & Díaz–Díaz, N. (2005). Selección de atributos relevantes basada en bootstrapping. In.

Almaraashi, M. (2018). Investigating the impact of feature selection on the prediction of solar radiation in different locations in Saudi Arabia. *Applied Soft Computing, 66*, 250-263. doi:https://doi.org/10.1016/j.asoc.2018.02.029

Arulmurugan, R., & Anandakumar, H. (2018). Early Detection of Lung Cancer Using Wavelet Feature Descriptor and Feed Forward Back Propagation Neural Networks Classifier. In *Computational Vision and Bio Inspired Computing* (pp. 103-110): Springer.

Beatriz Rivela, A. M., Byron Criollo, Diego Vaca, Jesús Lopez Villada. (2014). Enviromental Polices in Marine Transport: A Case Study of Solar Ship in Galapagos Islands. *EuroSun*.

Bleicher, A. (2013). Solar sailor [Dream Jobs 2013 - Renewables]. *IEEE Spectrum, 50*(2), 45-46. doi:10.1109/MSPEC.2013.6420144

Capozzoli, A., Lauro, F., & Khan, I. (2015). Fault detection analysis using data mining techniques for a cluster of smart office buildings. *Expert Systems with Applications, 42*(9), 4324-4338.

Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering, 40*(1), 16-28.

Cuvi, N., & Guijarro, D. (2016). ¿ Una trayectoria hacia la insustentabilidad? La movilidad terrestre en la isla Santa Cruz, Galápagos. *Revista Transporte y Territorio*(15), 216-240.

D'Oca, S., & Hong, T. (2015). Occupancy schedules learning process through a data mining framework. *Energy and Buildings, 88*, 395-408.

Dash, M., & Liu, H. (1997). Feature selection for classification. *Intelligent data analysis, 1*(3), 131-156.

del Álamo, L. M. (2016). *Relaciones entre mareas y otros fenómenos geodinámicos: estudios en la Península Ibérica y California*: Universidad Complutense de Madrid.

DPNG, D. d. P. N. G. (2016). Informe anual de visitantes a las áreas protegidas.

Frank, E., Hall, M., Holmes, G., Kirkby, R., Pfahringer, B., Witten, I. H., & Trigg, L. (2009). Weka-a machine learning workbench for data mining. In *Data mining and knowledge discovery handbook* (pp. 1269-1277): Springer.

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis, 38*(4), 367-378. doi:https://doi.org/10.1016/S0167-9473(01)00065-2

García, J., Rangel, E., & Farías, M. (2013). Informe Galápagos 2011-2012. DPNG, GCREG, FCD y GC. . In. Puerto Ayora.

Gursu, H. (2016). Solar And Wind Powered Concept Boats: The Example Of Volitan. *METU Journal of the Faculty of Architecture, 31*(2).

Gutiérrez, J. A. G. (2016). Comenzando con Weka: Filtrado y selección de subconjuntos de atributos basada en su relevancia descriptiva para la clase. *ETSI Informática, Universidad Nacional de Educación a Distancia*.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research, 3*(Mar), 1157-1182.

Hatwaambo, S., Jain, P. C., Perers, B., & Karlsson, B. (2009). Projected beam irradiation at low latitudes using Meteonorm database. *Renewable Energy, 34*(5), 1394-1398. doi:https://doi.org/10.1016/j.renene.2008.09.011

Helling, R. K. (2017). The Role of LCA in Sustainable Development. In M. A. Abraham (Ed.), *Encyclopedia of Sustainable Technologies* (pp. 237-242). Oxford: Elsevier.

INER. (2014). Informe del sistema fotovoltaico del catamarán solar de las Islas Galápagos. In.

Quito.
INOCAR, I. O. d. l. A. (2016). Tabla de Mareas In: INOCAR.
Jain, A., & Zongker, D. (1997). Feature selection: Evaluation, application, and small sample performance. *IEEE transactions on pattern analysis and machine intelligence, 19*(2), 153-158.
Jara-Alvear, J., Pastor, H., Garcia, J., Casafont, M., Araujo, E., & Calderon, E. (2013). *Embarcaciones solares, una evolución al transporte marino en las islas Galápagos, Ecuador.* Paper presented at the 1st International congress and scientific expo ISEREE.
Jose Alvear, H. P., Juan García. (2013). Embarcaciones Solares, una evolución al transporte marino en las Islas Galápagos, Ecuador. *ISEREE, 1*.
Jurado, S., Nebot, À., Mugica, F., & Avellana, N. (2015). Hybrid methodologies for electricity load forecasting: Entropy-based feature selection with machine learning and soft computing techniques. *Energy, 86*, 276-291. doi:https://doi.org/10.1016/j.energy.2015.04.039
Khmaissia, F., Frigui, H., Sunkara, M., Jasinski, J., Garcia, A. M., Pace, T., & Menon, M. (2018). Accelerating band gap prediction for solar materials using feature selection and regression techniques. *Computational Materials Science, 147*, 304-315. doi:https://doi.org/10.1016/j.commatsci.2018.02.012
Ko, N., & Gantner, J. (2016). Local added value and environmental impacts of ship scrapping in the context of a ship's life cycle. *Ocean Engineering, 122*, 317-321. doi:https://doi.org/10.1016/j.oceaneng.2016.05.026
Kumar, N. M., Kumar, M. R., Rejoice, P. R., & Mathew, M. (2017). Performance analysis of 100 kWp grid connected Si-poly photovoltaic system using PVsyst simulation tool. *Energy Procedia, 117*, 180-189. doi:https://doi.org/10.1016/j.egypro.2017.05.121
Leary, D., & Kubby, J. (2017). Feature Selection and ANN Solar Power Prediction. *Journal of Renewable Energy, 2017*, 7. doi:10.1155/2017/2437387
Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research, 247*(1), 124-136.
Ling-Chin, J., & Roskilly, A. P. (2016). A comparative life cycle assessment of marine power systems. *Energy Conversion and Management, 127*, 477-493. doi:https://doi.org/10.1016/j.enconman.2016.09.012
Lucas, A., Neto, R. C., & Silva, C. A. (2013). Energy supply infrastructure LCA model for electric and hydrogen transportation systems. *Energy, 56*, 70-80. doi:https://doi.org/10.1016/j.energy.2013.04.056
Maggí, P. D. E., Rodríguez, J. P. C., Litardo, J. E. T., & Caviedes, E. C. E. (2018). ANÁLISIS DEL MERCADO TURÍSTICO DE LAS ISLAS GALÁPAGOS: CASO CRUISING GALÁPAGOS. *Observatorio de la Economía Latinoamericana*. doi:http://www.eumed.net/2/rev/oel/2018/02/mercado-turistico-galapagos.html
Martínez, J. J. (2017). Análisis de minería de datos distribuida con Weka Parallel en computadoras con múltiples procesadores físicos y lógicos. *Economía y Administración (E&A), 6*(2), 155-166.
Meteonorm. (2019). Publication Page. Retrieved from www.meteonorm.com
Mina, M. A. E. (2018). WEKA, ÁREAS DE APLICACIÓN Y SUS ALGORITMOS: UNA REVISIÓN SISTEMÁTICA DE LITERATURA/WEKA, AREAS OF APPLICATION AND THEIR ALGORITHMS: A SYSTEMATIC REVIEW OF LITERATURE. *Revista Científica ECOCIENCIA, 2018*(Especial).
Moya, M., & Arroyo, D. (2015). Análisis de los resultados de la operación del catamarán "Génesis Solar" en el estrecho de Itabaca (Islas Galápagos). *Congreso Internacional I+D+I, 2*.
Muñoz Barriga, A. (2015). La contradicción del turismo en la conservación y el desarrollo en

Galápagos - Ecuador. *Estudios y perspectivas en turismo.*, *24*(2), 399-413. doi:http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1851-17322015000200012&lng=es&tlng=es.

Navarro, L., & Javier, J. (2001). *Selección de atributos en aprendizaje automático basada en teoría de la información*.

Okello, D., van Dyk, E. E., & Vorster, F. J. (2015). Analysis of measured and simulated performance data of a 3.2kWp grid-connected PV system in Port Elizabeth, South Africa. *Energy Conversion and Management, 100*, 10-15. doi:https://doi.org/10.1016/j.enconman.2015.04.064

Parque-Nacional-Galápagos. (2017). *Informe anual 2017 visitantes a las áreas protegidas de Galápagos.* Retrieved from http://www.galapagos.gob.ec/

Rodríguez, Y., Fernández, Y., Bello, R., & Caballero, Y. (2014). Selección de atributos relevantes aplicando algoritmos que combinan conjuntos aproximados y optimización en colonias de hormigas. *Revista Cubana de Ciencias Informáticas, 8*, 79-86.

Saeteros, A. M., Da Silva, E. V., & Calles, V. G. (2017). Análisis de la Demanda del Turismo para la Gestión Sustentable del Destino en las Islas Galápagos-Ecuador. *Anais Brasileiros de Estudos Turísticos-ABET, 7*(1), 78-88.

Salcedo-Sanz, S., & Cornejo-Bueno, L. (2018). Feature selection in machine learning prediction systems for renewable energy applications. *Renewable and Sustainable Energy Reviews, 90*, 728-741. doi:https://doi.org/10.1016/j.rser.2018.04.008

Schiller, L., Alava, J. J., Grove, J., Reck, G., & Pauly, D. (2015). The demise of Darwin's fishes: evidence of fishing down and illegal shark finning in the Galápagos Islands. *Aquatic Conservation: Marine and Freshwater Ecosystems, 25*(3), 431-446.

Tan, P.-N. (2018). *Introduction to data mining*: Pearson Education India.

Touati, F., Chowdhury, N. A., Benhmed, K., San Pedro Gonzales, A. J. R., Al-Hitmi, M. A., Benammar, M., . . . Ben-Brahim, L. (2017). Long-term performance analysis and power prediction of PV technology in the State of Qatar. *Renewable Energy, 113*, 952-965. doi:https://doi.org/10.1016/j.renene.2017.06.078

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*: Morgan Kaufmann.

Zhang, W., Dang, H., & Simoes, R. (2018). A new solar power output prediction based on hybrid forecast engine and decomposition model. *ISA Transactions*. doi:https://doi.org/10.1016/j.isatra.2018.06.004