

Estudio comparativo de técnicas de minería de datos para la predicción de rutas de huracanes
Comparative study of data mining techniques for the prediction of hurricane routes

Coronado Arjona, M.A.¹, Bianchi Rosado, V. M.²,
Vivas Burgos, J. A.³, Perera Collí, M.A.⁴
¹²³⁴ Instituto Tecnológico de Tizimín, Yucatán
Km. 3.5 Carretera Final Aeropuerto Cupul
¹mcorarj@yahoo.com.mx, ²bianchitkd@hotmail.com,
³jvivas80@hotmail.com, ⁴miguel_pc@hotmail.com

Fecha de recepción: 26 de marzo 2017

Fecha de aceptación: 25 de abril 2017

Resumen. Los huracanes son las tormentas más grandes y violentas que pueden existir sobre la tierra. Su peligrosidad radica en la velocidad que pueden alcanzar sus vientos, llegando a superar los 250 kilómetros por hora y desatando 9 billones de litros de lluvia al día, en consecuencia, sus efectos son a gran escala y con frecuencia muy destructivos en pérdidas humanas y materiales. A sabiendas de la inexactitud en las trayectorias, muchos habitantes esperan hasta el último momento antes de abandonar su hogar y pertenencias con la esperanza de que el fenómeno meteorológico cambie su curso. Es por esto que surge la necesidad de determinar la mejor técnica para predecir rutas de huracanes. El estudio consistió en entrenar los algoritmos de las técnicas de predicción, regresión lineal, k vecinos más cercanos y perceptrón multicapa, para obtener los modelos que permitan la comparación de datos predictivos con las trayectorias reales de huracanes y así determinar la exactitud de la predicción. Se encontró que la técnica de regresión lineal obtuvo los mejores resultados.

Palabras clave: CRISP-DM, K Vecinos más Cercanos, Minería de Datos, Perceptrón Multicapa, Regresión Lineal.

Summary. Hurricanes are the largest and most violent storms that exist on Earth. Their dangerousness lies in the speed that can reach their winds, reaching over 250 km per hour and unleashing 9 billion liters of rain a day, so their effects are large scale and very destructive. Due to the effects mentioned before, the number of human and material losses are high. This is because of the inaccuracy in trajectories and many inhabitants wait until the last moment to leave their home and belongings, in the hope that the weather phenomenon will change its course. In this way arises the need to find the best hurricane prediction technique. The study consisted in training the algorithms of prediction techniques, linear regression, k nearest neighbors and multilayer perceptron, to obtain the models that allow the comparison of predictive data with the actual hurricane trajectories and thus determine the accuracy of the prediction. It was found that the linear regression technique obtained the best results.

Keywords: CRISP-DM, K Nearest Neighbors, Data Mining, Multilayer Perceptron, Linear Regression.

1 Introducción

Cada año, de marzo a noviembre (y con casos especiales hasta mediados de diciembre), las costas mexicanas son amenazadas e impactadas en diferentes intensidades por poderosos huracanes que ponen en riesgo las vidas y el patrimonio de las personas que habitan esas zonas cada vez más densamente pobladas.

Afortunadamente, con el avance tecnológico se cuenta con sistemas de alerta que con bastante anticipación informan a las comunidades desde la formación del meteoro hasta su fuerza, velocidad de desplazamiento y dirección; incluso una aproximación de su trayectoria y posición lo cual ayuda a tomar precauciones al desalojar a los pobladores que se encuentren en la probable zona de afectación de los potentes vientos de este tipo de fenómeno meteorológico, lo cual reduce los riesgos y pérdidas.

Sin embargo, en la actualidad no hay nada más impredecible que el comportamiento del clima y los fenómenos meteorológicos, especialmente de los huracanes, que como han enseñado durante el paso de las décadas en que se registra su formación y accionar, dependen de variables que cambian en fracciones de segundo influenciados incluso por los propios efectos del fenómeno.

Es en este punto en que la información disponible relacionada con estos fenómenos meteorológicos (presión atmosférica, temperatura del aire, temperatura del mar, corrientes marinas, dirección del viento) y las herramientas para procesarla con eficiencia pueden ser de gran ayuda para reducir los riesgos y las pérdidas materiales y humanas en consecuencia. Se cuenta con registros históricos así como modernas tecnologías para recolectar datos abarcando grandes extensiones territoriales lo cual permite incrementar su efectividad, sin embargo, hasta el momento los modelos generados difieren muchísimo entre sus resultados y la ruta real del meteoro, por lo que es deseable contrastar los resultados de los algoritmos para determinar aquel que mejor se haya aproximado a la ruta real de los huracanes estudiados.

Debido a su gran potencial destructivo, los huracanes o tifones, por su duración, fuerza y recorrido son de especial interés para la república mexicana ya que cada año se presenta la amenaza de que su fuerza impacte en

alguno de los litorales del territorio nacional y se interne en el mismo provocando daños materiales y de vidas los cuales inciden en la economía del país.

Se compararon tres algoritmos de minería de datos para determinar cuál tiene mayor precisión en su predicción, lo cual puede redundar en una mejor preparación de la protección civil, y a su vez, de los bienes y la producción agrícola-pecuaria en la ruta proporcionada por el modelo.

2 Revisión de la literatura

2.1 Definición de minería de datos

La literatura remite a que “la minería de datos es entendida como el proceso de descubrir conocimientos interesantes, como patrones, asociaciones, cambios, anomalías y estructuras significativas a partir de grandes cantidades de datos almacenadas en bases de datos, data warehouses, o cualquier otro medio de almacenamiento de información” [1].

La aplicación de algoritmos de minería de datos requiere de actividades previas destinadas a preparar los datos de manera homogénea. Esta primera etapa es también conocida como ETL (Extract, Transform and Load) [1].

2.2 Tipos de modelos

La minería de datos tiene como objetivo obtener información útil a través del análisis de los datos. El conocimiento que se obtiene, puede ser en forma de relaciones, patrones o reglas inferidos de los datos y desconocidos, o bien, en forma de una descripción más concisa (es decir, un resumen de los mismos) [10].

Los modelos obtenidos a partir de los datos y de las técnicas de minería empleadas, pueden dar como resultado uno de dos tipos de modelos: predictivos o descriptivos. Con los primeros, se intenta determinar valores desconocidos de variables de interés utilizando para ello otras variables o atributos de la base de datos. Mientras tanto, los modelos descriptivos inspeccionan las propiedades del conjunto de datos para describir patrones, que explican o resumen al mismo conjunto, no para predecir nuevos valores.

2.3 Tareas de la minería de datos

Las tareas de la minería de datos pueden ser predictivas o descriptivas. Dentro de las tareas predictivas se encuentran la clasificación y la regresión, mientras que el agrupamiento (clustering), reglas de asociación, reglas de asociación secuenciales y las correlaciones son tareas descriptivas.

La clasificación es probablemente la tarea más utilizada y mejor entendida. Tiene las siguientes tres características [18]:

- El aprendizaje es supervisado.
- La variable dependiente es categórica.
- Hace énfasis en construir modelos capaces de asignar nuevas instancias a una de un conjunto de clases bien definidas.

La regresión es la técnica más a menudo utilizada para crear modelos de datos en estadística tradicional. Esta técnica de modelado estadístico examina la relación entre una variable dependiente y una o más variables independientes. Existen tres tipos de ésta: lineal, no lineal y logística [16]. La principal diferencia con respecto a la clasificación es que el valor a predecir es numérico.

El agrupamiento (clustering) es la tarea predictiva por excelencia y consiste en examinar datos para encontrar grupos de elementos que guardan similitudes y que a la vez son diferentes a los objetos pertenecientes a otros grupos [1]. A diferencia de la clasificación, en lugar de analizar los datos etiquetados con una clase, los analiza para generar la etiqueta.

Las correlaciones son una tarea descriptiva que se usa para examinar el grado de semejanza de los valores de dos variables numéricas [10]. Cuando el coeficiente de correlación r es positivo, las variables están perfectamente correlacionadas (tienen comportamiento similar). Un valor negativo en el coeficiente de correlación r establece que las variables están correlacionadas negativamente (una variable crece y la otra decrece). Mientras que si es 0, entonces no hay correlación.

Las reglas de asociación son también una tarea descriptiva, muy similar a las correlaciones que tiene como objetivo identificar relaciones no explícitas entre atributos categóricos. Las reglas de asociación no implican una relación causa-efecto [10].

2.4 Técnicas de minería de datos

Los algoritmos de minería de datos se clasifican en dos grandes categorías: supervisados o predictivos y no supervisados o de descubrimiento del conocimiento.

Los algoritmos supervisados o predictivos, como su nombre lo indica, predicen un dato desconocido a partir de un conjunto de datos conocidos previamente llamados descriptivos. A partir de datos con etiqueta conocida se induce un modelo que relaciona dicha etiqueta con los atributos descriptivos; tal relación sirve para realizar la predicción en datos cuya etiqueta es desconocida. Los algoritmos supervisados requieren de una fase de entrenamiento (construcción de un modelo usando un subconjunto de datos con etiqueta conocida) y prueba (prueba del modelo sobre el resto de los datos).

Por otro lado, los algoritmos no supervisados se caracterizan por descubrir modelos o características significativas a partir únicamente de los datos de entrada. Estos algoritmos realizan tareas descriptivas como el descubrimiento de patrones y tendencias en los datos actuales. El descubrimiento de estos patrones sirve para llevar a cabo acciones y obtener un beneficio científico o de negocio de ellas.

Tabla 1. Algoritmos de minería de datos

SUPERVISADOS	NO SUPERVISADOS
Árboles de decisión	Detección de desviaciones
Inducción neuronal	Segmentación
Regresión	Agrupamiento (clustering)
Series temporales	Reglas de asociación
	Patrones secuenciales

2.5 Metodología CRISP-DM

CRISP-DM es una metodología para orientar los trabajos de minería de datos. Incluye las etapas o fases de un proyecto, las tareas a desarrollar por cada fase, una explicación de las relaciones entre las tareas y los productos entregables por cada etapa [IBM].

La ventaja de CRISP-DM es su flexibilidad ya que es posible ajustarlo a las necesidades de una empresa y la secuencia de las actividades no es rigurosa dado que los proyectos pueden avanzar o retroceder fases como se requiera [IBM].

Las etapas de la metodología CRISP-DM se describen a continuación:

- **Comprensión del negocio o problema:** Esta primera etapa consiste en entender los objetivos y requerimientos del proyecto a fin de convertirlos a objetivos técnicos y en un plan del proyecto.
- **Comprensión de los datos:** Se recopilan y estudian los datos para encontrar posibles problemas de calidad en ellos.
- **Preparación de los datos:** Es la adecuación de los datos dependiendo de la técnica de minería de datos que se utilice en la siguiente etapa. La calidad de los datos obtenidos dependerá de la correcta preparación.
- **Modelado:** Consiste en aplicar alguna técnica de minería de datos dependiendo de la tarea que se quiera realizar, ya sea predictiva o descriptiva, para la obtención de un modelo que representa al conocimiento.
- **Evaluación:** Los expertos interpretan el modelo de acuerdo al conocimiento preexistente del dominio para garantizar que se alcanzan los objetivos para la toma de decisiones.
- **Despliegue del modelo:** Se integra el modelo obtenido a los procesos de toma de decisiones dentro de las organizaciones.

3 Metodología de la investigación

Para el desarrollo de este proyecto, se utilizó la metodología CRISP-DM (Cross Industry Standard Process for Data Mining) la cual generó las siguientes acciones:

Etapa 1. Comprensión del negocio o problema

En esta etapa se determinó cuál es el objetivo del proyecto (lo que se quiere hacer) y cómo alcanzarlo. Es decir, fue necesario definir cuál es el proceso a seguir para la obtención del modelo a partir de los datos históricos con las técnicas de minería de datos seleccionadas y para su posterior evaluación.

Se precisó un inventario de los recursos (datos, software, hardware) con los que se contaba, y que ayudarían para la obtención del modelo predictivo de rutas. Los datos obtenidos del sitio web de la Administración Nacional Oceánica y Atmosférica (NOAA) se asumieron como válidos.

Para finalizar, se diseñó un plan el cual sirvió como parámetro para medir el progreso del proyecto. Entre otros puntos, este plan incluyó: el objetivo previamente definido, la factibilidad del proyecto, los recursos con los que se cuenta para su realización y cronograma de actividades.

Etapa 2. Comprensión de los datos

La recolección de los datos fue a través del sitio web del NOAA. Se transformó el campo multivalorado “tiempo” en los nuevos campos “fecha” y “hora”. Además, se efectuó la definición de la estructura de los datos como sigue:

Tabla 2. Estructura de los datos

Campo	Tipo	Tamaño
Nombre	Alfanumérico	15
Año	Numérico	
ID	Numérico	
Latitud	Numérico	
Longitud	Numérico	
Hora	Numérico	
Fecha	Numérico	
Vientos	Numérico	
Presión	Numérico	
Status	Alfanumérico	3

La información registrada de los primeros huracanes se encuentra incompleta porque al inicio, no se incluía el parámetro “presión” sino hasta años posteriores. Tomando en consideración a fenómenos meteorológicos más recientes, se determinó que los datos adquiridos si satisfacen los requerimientos del proyecto.

Etapa 3. Preparación de los datos

Se determinó que todos los datos obtenidos del sitio web, a excepción del nombre del fenómeno, del año y de la fecha; servirían para la realización del proyecto pues son datos relevantes y necesarios que los meteorólogos utilizan para la predicción de las trayectorias de los huracanes.

Se eliminaron aquellos registros que no contienen información acerca de la presión y se tomaron en consideración sólo aquellas filas en donde sí se indicaba. Esta eliminación no afectó en ningún momento el proyecto porque siguen siendo suficientes los datos.

La herramienta de minería de datos (WEKA) que se utilizó no requiere que los datos estén en un orden específico o que sea necesario establecer un atributo clave. La única condición es que los datos a probar estén en el formato CSV (Delimitado por comas) o ARFF.

Etapa 4. Modelado

El tipo de tarea que se realizó para este proyecto es predictiva. Se compararon 3 algoritmos para determinar cuál era el que mejor se aproximaba a las trayectorias reales (latitud y longitud) de los huracanes. Los algoritmos que fueron probados son: K vecinos más próximos, perceptrón multicapa y regresión lineal.

Para ello, se creó un archivo (todos.arff) el cual contiene 583 registros correspondientes a 24 meteoros diferentes (huracán, depresión tropical, etc.) en diversos años como se muestra en la figura 1.

```

todos: Bloc de notas
Archivo Edición Formato Ver Ayuda
@relation todos
@attribute adv real
@attribute latitud real
@attribute longitud real
@attribute hora real
@attribute viento real
@attribute presion real
@attribute status {DT,TT,TET,DET,H1,H2,H3,H4,H5,DST}
@data
1,27.50,-79.00,0,25,1013,DT
2,28.50,-79.00,6,25,1013,DT
3,29.50,-79.00,12,25,1013,DT
4,30.50,-79.00,18,25,1013,DT
5,31.50,-78.80,0,25,1012,DT
6,32.40,-78.70,6,25,1012,DT
7,33.30,-78.00,12,25,1011,DT
8,34.00,-77.00,18,30,1006,DT
9,34.40,-75.80,0,35,1004,TT
10,34.00,-74.80,6,40,1002,TT
    
```

Figura 1. Archivo con los registros de 24 meteoros diferentes

Para cada uno de los algoritmos se buscó que:

- El coeficiente de correlación sea lo más cercano a 1, lo cual nos indica que la evolución de los atributos (latitud y longitud) es similar o parecida.
- El error cuadrático sea mínimo.

Los resultados obtenidos por cada técnica, pueden apreciarse en la tabla número 3.

Tabla 3. Resumen por técnica utilizada

	Error Cuadrático	Coefficiente de correlación
KNN		
Latitud	91.6321 %	0.633
Longitud	83.8459 %	0.7804
Regresión Lineal		
Latitud	67.8179 %	0.9402
Longitud	76.6586 %	0.5568
Perceptrón Multicapa		
Latitud	95.7574 %	0.848
Longitud	102.7428 %	0.0245

Los mejores resultados se obtuvieron con el método de “Regresión Lineal” dado que el error cuadrático fue mínimo (67.8179% y 76.6586%) y el coeficiente de correlación fue la más cercana a 1 (0.9402 y 0.5568).

Posteriormente, se tomó de manera aleatoria a un huracán que no haya sido incluido en el archivo “todos.arff”. Se procuró que el tiempo de vida del meteoro seleccionado sea superior al resto de los demás huracanes (mientras más datos mejor).

Tomando como referencia al meteoro Claudette (el cual, nunca llegó a la categoría de huracán) del año 1979, se creó un archivo con el nombre del huracán pero con extensión arff. De Claudette se tienen 57 registros pero se excluyeron los últimos 15 con el objeto de esos mismos fueran pronosticados.

```

@relation claudette

@attribute adv real
@attribute latitud real
@attribute longitud real
@attribute hora real
@attribute viento real
@attribute presion real
@attribute status {DT,TT,TET,DET,H1,H2,H3,H4,H5,DST}

@data
1,12.50,-46.30,12,20,1014,DT
2,12.80,-48.40,18,20,1014,DT
3,13.40,-50.40,0,25,1012,DT
4,14.60,-52.10,6,25,1012,DT
5,15.70,-53.80,12,30,1011,DT
6,16.50,-55.50,18,30,1011,DT
7,17.00,-57.20,0,30,1011,DT
8,17.50,-58.80,6,30,1011,DT
9,17.80,-60.30,12,35,1011,TT
10,18.00,-62.10,18,40,1010,TT
    
```

Figura 2. Datos del meteoro Claudette

Etapa 5. Evaluación

Los valores predichos que se obtuvieron a través de la técnica de regresión lineal son los siguientes:

Tabla 4. Comparativo entre coordenadas reales y predichas

Id	REALES		PREDICHOS	
	Latitud	longitud	latitud	Longitud
1	30.2	-95.3	23.32	-76.75
2	30.6	-95.1	23.33	-76.84
3	30.8	-95.4	23.68	-74.68
4	31.3	-96.3	23.7	-74.57
5	31.8	-96.6	23.8	-71.96
6	32.7	-96.4	23.8	-71.94
7	34.0	-95.9	23.81	-71.99
8	35.3	-95.3	23.84	-72.09
9	36.4	-94.6	23.6	-74.2
10	37.8	-93.4	23.6	-71.82
11	38.6	-91.0	23.9	-74.82
12	38.8	-88.0	23.9	-72.5
13	39.0	-85.2	23.97	-72.6
14	39.0	-82.8	23.97	-75.3
15	39.0	-80.2	24.32	-78.3

Los valores de la Latitud y Longitud predichos por WEKA no corresponden a los valores reales que siguió el meteoro en Julio de 1979. En la figura número 3 puede observarse la trayectoria real seguida por Claudette y en la figura número 4 los datos predichos por el software.



Figura 3. Trayectoria real seguida por el huracán Claudette

4 Análisis de resultados

Se graficaron únicamente los primeros nueve puntos (del total de 15) debido a que no tenía caso generar el resto de las coordenadas debido a la discrepancia con los valores reales. La figura 4 indica que no se está siguiendo un patrón idéntico a la trayectoria original del huracán Claudette (como se muestra en la figura 3) pese a que se utilizó el mejor método de predicción de las tres técnicas que fueron comparadas.



Figura 4. Trayectoria predicha con la técnica de regresión lineal

Tal vez, considerando a otras variables además de las utilizadas en este estudio, si se llegue a aproximar las rutas originales y predichas. Por ejemplo, en la base de datos descargada de NOAA se encuentran el año, fecha y tiempo del fenómeno, desde su formación hasta su desintegración. Es posible que estas variables puedan influir en los resultados finales ya que la temperatura del agua no es la misma dependiendo del mes y hora en que fueron tomados los registros.

5 Conclusiones

La metodología aplicada, CRISP-DM, ha demostrado ser una herramienta valiosa al proporcionar la estrategia a seguir en el proceso de encontrar un algoritmo que determine de forma más precisa, la trayectoria seguida por los huracanes que se forman en el Atlántico. Debido a que la metodología aplicada cuenta con múltiples etapas, algunas de ellas no fueron aplicadas en el proyecto por la reducida cantidad de variables empleadas y la calidad de los datos usados.

Se eligieron los siguientes tres algoritmos de minería de datos debido a su naturaleza predictiva: regresión lineal, K vecinos más cercanos y perceptrón multicapa. Tomando como referencia datos históricos de los huracanes como hora, viento, presión y el status (depresión, tormenta, huracán en alguna categoría, entre otros), se predijeron las coordenadas geográficas, latitud y longitud, para cada uno de los algoritmos.

Fue la técnica de regresión lineal en donde se obtuvo los mejores resultados para la función “error cuadrático” y del coeficiente de correlación, ya que se buscó que fuera mínima y lo más cercano a uno, respectivamente.

Una vez hallado el algoritmo más eficiente de los tres con base en el estudio comparativo, se trató de encontrar cuáles eran los últimos nueve puntos geográficos (latitud y longitud) de la trayectoria seguida por un huracán elegido aleatoriamente. Aunque el software y la técnica utilizados predijeron cada par de coordenadas, éstas seguían un comportamiento errático en comparación con la ruta original.

La regresión lineal prometía ser la técnica que predijera con exactitud la trayectoria seguida por el huracán que se seleccionó para la etapa de pruebas. Sin embargo, al hacer el estudio aplicando sólo ésta, las coordenadas predichas distan por mucho de las reales. Al unir los puntos correspondientes (ver figura 4) de cada par de coordenadas geográficas, se pudo observar que la ruta obtenida difiere de la original. Con lo anterior queda demostrado que, al menos para las tres técnicas de minería de datos seleccionadas y las variables estudiadas en este proyecto, ninguna de ellas puede ser utilizada como método confiable para predecir las rutas de huracanes.

Sin embargo, algunos sistemas de predicción tales como CLIPER (Climate and Persistence) que utiliza la misma técnica estadística pero que toma como referencia datos pasados y presentes del huracán, sí determinan con mayor precisión la ruta seguida por los mismos fenómenos meteorológicos.

Con este trabajo se sienta precedente para realizar estudios similares pero utilizando algoritmos de minería de datos distintos.

6 Trabajos futuros

Todo trabajo de investigación implica un esfuerzo que es importante sea reconocido como una inversión, de tal manera que es preciso identificar líneas de trabajo que permitan darle continuidad y de esta forma agregar valor al conocimiento generado. En consecuencia, a continuación se presenta una lista del trabajo futuro que es necesario realizar para continuar con el avance en el conocimiento de la minería de datos y su aprovechamiento para la predicción de fenómenos meteorológicos.

- Validar la relación de las variables ajenas al huracán para la predicción de su trayectoria usando minería de datos.
- Predicción del clima posterior al paso de un huracán utilizando minería de datos.
- Predicción de la producción agropecuaria en el municipio en relación con los factores climáticos utilizando minería de datos.
- Determinar si existe repetición de ciclos climáticos en una zona a partir de los registros en una estación meteorológica.
- Predicción mediante minería de datos, de las condiciones atmosféricas en una zona a partir de la repetición de ciclos climáticos.
- Predecir el riesgo de formación de huracanes y tormentas a partir de parámetros por configuraciones atmosféricas.
- Predecir de la deserción y reprobación estudiantil en el nivel superior mediante la minería de datos.

7 Agradecimientos

Al MEM. Mariano Matú Sansores por su invaluable apoyo en la revisión del abstract.

A nuestro amigo MVZ. Raúl Martín Fernández, responsable de la estación climatológica en nuestro Instituto, por los datos brindados para ésta y para futuras investigaciones relacionadas con fenómenos meteorológicos.

Al Instituto Tecnológico de Tizimín, por las facilidades brindadas para el uso de sus instalaciones e infraestructura de telecomunicaciones.

Referencias

1. Bramer, M.: *Clustering*. Mackie, I. (Ed): *Principles of Data Mining*. Springer, pp. 311-312 (2007)
2. Buranasing, A.: *Storm Intensity Prediction using Artificial Neural Network*. <http://www.interjournal.cmru.ac.th/science/Paper/FullPaper-S2O-005.pdf>. (n.d.). Accedido el 17 de enero de 2017
3. Cárdenas, R.: *¿Es posible predecir la trayectoria de un huracán?*. http://www.cienciorama.unam.mx/a/pdf/108_cienciorama.pdf. (n.d.). Accedido el 16 de enero de 2017
4. Castaño, J.; Baethgen, W.; Gimenez, A.; et al: *Evolución del clima observado durante el periodo 1931-2000 en la región sureste de América del Sur*. http://www.inia.org.uy/disciplinas/agroclima/publicaciones/ambiente/se_a_s_2007.pdf. (n.d.). Accedido el 18 de enero de 2017
5. Cazorla, A.; Olmo, F.; Alados, L.: *Estimación de la cubierta nubosa en imágenes de cielo mediante el algoritmo de clasificación KNN*. <http://dns2.aet.org.es/congresos/xi/ten76.pdf>. (2005). Accedido el 17 de enero de 2017.
6. CNNEspañol: *¿Cómo se forman los huracanes y por qué son tan peligrosos?*. <http://cnnespanol.cnn.com/2016/10/06/como-se-forman-los-huracanes-y-por-que-son-tan-peligrosos/#0>. (2016). Accedido el 16 de enero de 2017.
7. Federación Internacional de Sociedades de la Cruz Roja y de la Media Luna Roja: *Peligros meteorológicos: tormentas tropicales, huracanes, ciclones y tifones*. <http://www.ifrc.org/es/introduccion/disaster-management/sobre-desastres/definicion--de-peligro/tormentas-tropicales-huracanes-ciclones-y-tifones/>. (n.d.). Accedido el 16 de enero de 2017.
8. García, A.: *Aprendizaje*. Alfaomega (Ed): *Inteligencia Artificial. Fundamentos, práctica y aplicaciones*. Alfaomega, pp. 220-238 (2013)
9. Hernández, J.; Ramírez, M.; Ferri, C.: *¿Qué es la minería de datos?*. Fayerman, D. (Ed): *Introducción a la minería de datos*. Pearson, pp. 3-38 (2004)
10. Hayati, M.; Mohebi, Z.: *Application of Artificial Neural Networks for Temperature Forecasting*. <http://waset.org/publications/8486/application-of-artificial-neural-networks-for-temperature-forecasting>. (2007). Accedido el 17 de enero de 2017
11. Mora, J.; Morales, G.; Barrera, R.: *Evaluación del clasificador basado en los k vecinos más cercanos para la localización de la zona en falla en los sistemas de potencia*. <http://www.scielo.org.co/pdf/iei/v28n3/v28n3a11.pdf>. (2008). Accedido el 16 de enero de 2017
12. Nath, S.; Kotal, S.: *Seasonal prediction of tropical cyclone activity over the north Indian Ocean using the neural network model*. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0187-62362015000400006. (2015). Accedido el 16 de enero de 2017
13. National Geographic: *Huracanes*. <http://www.nationalgeographic.es/medio-ambiente/desastres-naturales/hurricane-profile>. (n.d.). Accedido el 16 de enero de 2017
14. Nettleton, D.: *Data Modeling*. Morgan Kaufmann (Ed): *Comercial Data Mining*. Elsevier, pp. 149-150 (2014)
15. Petisco, S.: *Método de regionalización basado en análogos*. http://www.aemet.es/documentos/es/idi/clima/escenarios_CC/Metodo_regionalizacion_temperatura.pdf. (2008). Accedido el 16 de enero de 2017
16. Roiger, R.: *Data Mining. A Closer Look*. Taylor & Francis Group (Ed): *Data Mining. A tutorial-based primer*. CRC Press, pp. 33-34 (2017)
17. ScienceDaily: *More accurate method for predicting hurricane activity*. <https://www.sciencedaily.com/releases/2012/09/120911103409.htm>. (2012). Accedido el 17 de enero de 2017

18. Sierra, B.: *Algoritmos de clasificación por vecindad*. Martín-Romo, M. (Ed): *Aprendizaje automático: conceptos básicos y avanzados. Aspectos prácticos utilizando el software WEKA*. Pearson, pp. 23-38 (2006)
19. Valero, S.; Salvador, A.; García, M.: *Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos*. <http://www.utim.edu.mx/~svalero/docs/e1.pdf>. (2005). Accedido el 18 de enero de 2017