
Modelo de predicción de derrames de petróleo utilizando Inteligencia Artificial

Oil spill prediction model using Artificial Intelligence

Alexander Inga Alva

<https://orcid.org/0000-0001-6302-0234>

alexander.inga@unmsm.edu.pe

Universidad Nacional Mayor de San Marcos, Facultad
Ingeniería de Sistemas e Informática, Unidad de
Posgrado. Lima, Perú

RECIBIDO: 23/07/2022 - ACEPTADO: 18/08/2022 - PUBLICADO: 20/09/2022

RESUMEN

La presente investigación se enfocó en el análisis de las variables que conllevaron a los derrames de petróleo ocurridos durante los años 1900 al 2019 en el estado de New York de Norteamérica. Según la ley y los reglamentos estatales de este país, los derrames que podrían contaminar las tierras o las aguas del estado deben ser informados por el causante del derrame (en algunos casos, por cualquier persona que tenga conocimiento de la contaminación). Cada registro de derrame incluye: información administrativa (región y código único de derrame), tipo de instalación, fecha/hora del derrame, ubicación, factor contribuyente, fuente y causa del derrame, tipo de material derramado, cantidad derramada y recuperada, cuerpos de agua superficiales afectados, fecha de cierre (actividad de limpieza terminada) y Costo Ambiental. Con dicha información se desarrollaron modelos de predicción de Inteligencia Artificial y se determinó, por su mayor exactitud, al modelo de Árbol de Decisión como el predictor más idóneo de futuros desastres ecológicos.

Palabras clave: Inteligencia Artificial; Derrames de Petróleo; Desastres Ecológicos.

ABSTRACT

The present investigation focused on the analysis of the variables that led to oil spills that occurred during the years 1900 to 2019 in the state of New York in North America. Under state law and regulations in this country, spills that could contaminate state lands or waters must be reported by the person responsible for the spill (and, in some cases, by anyone with knowledge of the contamination). Each spill record includes: administrative information (region and unique spill code), type of facility, date/time of spill, location, contributing factor, source and cause of spill, type of material spilled, quantity spilled and recovered, bodies of surface water affected, closure date (cleanup activity completed) and Environmental Cost. With this information, Artificial Intelligence prediction models were developed and, due to its greater precision, the Decision Tree model was determined as the most suitable predictor of future ecological disasters.

Keywords: Artificial Intelligence; Oil Spill; Environmental Disasters.

I. INTRODUCCIÓN

Desde tiempos inmemoriales el petróleo ha sido conocido y usado por los primeros hombres: “Y el valle de Sidim estaba lleno de pozos de asfalto.” (Génesis 14:10). En el imperio Incaico, se le conoció como “Copé” y se usaba como impermeabilizante de cántaros y vasijas de barro y, posteriormente, de las embarcaciones españolas. A nivel mundial, fue el coronel William Drake quien perforó el primer pozo con fines comerciales en 1859 en la localidad de Oil Creek, Titusville en el estado norteamericano de Pensilvania y en 1863. En el norte del Perú, en la zona llamada “La Brea” se perforó el primer pozo sudamericano (Inga, 2018). Dentro de los hidrocarburos, la contaminación por derrames de petróleo se ha convertido en un evento no solo trágico para el medio ambiente, sino que está acompañado de una gran visibilidad para la sociedad y, sobretodo, cuando este ocurre en el medio marino.

Antecedentes

A continuación, para poner en contexto nuestro estudio se indicarán los 10 principales derrames de petróleo tanto a nivel mundial como en Estados Unidos.

Principales derrames de petróleo en Estados Unidos

1. Deepwater Horizon (20/04/2010). La petrolera Deepwater Horizon de Transocean registró un derramamiento de 133.98 millones de galones de petróleo en el Golfo de México.
2. Hawaiian Patriot (1977). Sucedió en Hawai y se registró un derrame de 31 millones de galones de petróleo.
3. Epic Colocotronics (1975). Sucedió en Puerto Rico y se registró un derrame de 18 millones de galones de petróleo. Hawaiian Patriot (1977). Sucedió en Hawai y se registró un derrame de 31 millones de galones de petróleo.
4. Exxon Valdez (24/03/1989). En Alaska, Exxon Valdez registró un derrame de petróleo de más de 11000000 de gal. de petróleo.
5. Burmah Agate (1979). Sucedió en Texas y se registró un derrame de 10,7 millones de galones de petróleo.
6. Corinthos (1975). Sucedió en Filadelfia y se registró un derrame de 10,5 millones de galones de petróleo.
7. Texaco Oklahoma (1971). Sucedió en Carolina del Norte y se registró un derrame de 7,9 millones de galones de petróleo.
8. Huaracán Katrina (2005). Sucedió en Luisiana y se registró un derrame de 7 millones de galones de petróleo.
9. Mega Borg (1990). Sucedió en Texas y se registró un derrame de 5.1 millones de galones de petróleo.
10. Santa Bárbara (28/01/1969) Sucedió en California y se registró un derrame de 4 millones de galones de petróleo (Portal Ambiental, s.f.).

Algoritmos Propuestos

En esta sección describiremos los algoritmos que se usaron para realizar la predicción de un posible derrame de petróleo.

Naive Bayes

La bondad que tiene este algoritmo para realizar la clasificación se basa en el Teorema de Bayes, por lo tanto, el clasificador Naive Bayes considera que los efectos entre clases, producto de características específicas, son independientes, por lo tanto, se establece una independencia condicional de clase (AprendelA, 2019).

Decision Trees

Cuando nos referimos al método de aprendizaje supervisado, para realizar clasificaciones, el algoritmo de árbol de decisión crea un modelo predictor de una variable objetivo basado en el aprendizaje de reglas de decisión sencillas obtenidas de las características propias de los datos (Scikit Learn, s.f.).

K-Nearest Neighbors

Es un algoritmo de tipo supervisado que es útil para clasificar muestras que posean valores discretos y/o para hacer predicciones con valores continuos. Este algoritmo también puede ser usado para calcular la regresión. Para realizar su clasificación se basa en puntos de datos que tengan similitudes por cercanía, las cuales son reconocidas durante la fase de entrenamiento. Con esta retroalimentación el algoritmo puede generar predicciones de ubicación de nuevos puntos (Aprende Machine Learning, 2018).

Neural Networks

La simplificación del cerebro humano puede ser modelado a través de una red neuronal, el cual para funcionar requiere de una cantidad considerable de unidades de procesamiento que, trabajando de manera simultánea e interconectada, simulan a las neuronas humanas.

Estas unidades organizan en capas: entrada, proceso y salida, pero con la peculiaridad que existen varias capas ocultas dentro del procesamiento y las conexiones son ponderadas asignándoles pesos. De esta forma, la red aprende evaluando uno a uno los registros, generando predicciones para cada uno de ellos y haciendo los justos a las ponderaciones ante una inconsistencia. Dicho proceso se

itera hasta alcanzar un valor óptimo que satisfaga la necesidad de exactitud (IBM, s.f.).

II. METODOLOGÍA

Con el propósito de encontrar un modelo que prediga la eventualidad de un desastre ambiental recurrimos a examinar las variables que estuvieron presentes en los derrames de petróleo acontecidos entre los años 1900 al 2019 en Estados Unidos. El Dataset original de este artículo procede del sitio web Kaggle y contiene los datos de derrames de petróleo y derivados, así como de otros materiales peligrosos (Kaggle, s.f.). Con el fin de tener las variables representativas se realizó una limpieza de la data y se consideraron finalmente 18 variables (ver Tabla 1).

Tabla 1
Variables presentes en los derrames petroleros (1900-2019) en New York Estados Unidos.

Etiqueta de datos	Descripción de datos
Spill Number	Número de identificación de derrame específico del programa. Nombre
Program Facility Name	Nombre del programa del sitio o instalación.
Street1	Calle o ubicación de la dirección del derrame.
Locality	Ciudad, pueblo o aldea de la dirección del derrame.
County	Ubicación del condado del derrame.
DEC Región	NYSDEC Región donde se localiza el derrame: Región 1: (Long Island) condados de Nassau y Suffolk Región 2: (Ciudad de Nueva York) Brooklyn, Bronx, Manhattan, Queens y Staten Island Región 3: (Valle del Bajo Hudson) condados de Dutchess, Orange, Putnam, Rockland, Sullivan, Ulster y Westchester Región 4: (Región de la capital/Northern Catskills) Condados de Albany, Columbia, Delaware, Greene, Montgomery, Otsego, Rensselaer, Schenectady y Schoharie Región 5: (Este de Adirondacks/Lago Champlain) Condados de Clinton, Essex, Franklin, Fulton, Hamilton, Saratoga, Warren y Washington Región 6: (Oeste de Adirondacks/Eastern Lake Ontario) Condados de Herkimer, Jefferson, Lewis, Oneida y St. Lawrence Región 7: (Nueva York central) condados de Broome, Cayuga, Chenango, Cortland, Madison, Onondaga, Oswego, Tioga y Tompkins.
Spill Date	Fecha de derrame. (Fechas más antiguas a 1978 son estimadas.)
Received Date	Fecha de recepción de la información recibida.
Contributing Factor	Causa del derrame: Desconocido Error humano Accidente de Tráfico Falla de equipamiento Vandalismo Fallo en prueba de tanque Mantenimiento Deliberado Tambores abandonados Falla del tanque Tanque sobrecargado Tormenta Otro
Water Body	Nombre de la superficie de agua afectada por el derrame. Llenar cuando la fuente afectada fue una superficie de agua.

Tabla 1: Continuación

Etiqueta de datos	Descripción de datos	
Source	Fuente del derrame (Tipo de instalación): Desconocido Comercial / Industrial Institucional, Educativa, Gubernamental, Otro. Instalación Principal (MOSF) > 400,000 gal Instalación No Principal > 1,100 gal Estación de servicio u otra instalación PBS	
	Vehículo de pasajeros	
	Vehículo Comercial	
	Camión Cisterna	
	Vivienda Privada	
	Embarcación	
	Vagón de ferrocarril	
	Transformador	
	Almacén de productos químicos a granel	
	Close Date	Fecha de cierre de derrame
	Material Name	Nombre común de la sustancia /material derramado
Material Family	Representa la familia del material derramado: Petróleo, Oxigenados, Materiales Peligrosos, etc.	
Quantity	Cantidad de sustancia encontrada/derramada. (Si se desconoce: "0") Tenga en cuenta que la cantidad puede incluir la cantidad de suelo o agua contaminada que se combina con el material derramado.	
Units	Unidades para la cantidad de sustancia encontrada /derramada: Galones o libras (Este campo puede estar en blanco si no se reporta.)	
Recovered	Cantidad de material recuperado, si procede. "Recuperado" Las unidades del material son las mismas que las unidades de derrame de "Cantidad". (Si la cantidad recuperada es desconocida: "0"). Tenga en cuenta que la cantidad recuperada puede ser mayor que la cantidad derramada cuando el suelo o el agua contaminada se recogen con el material derramado.	
Costo Ambiental	Hace referencia a la contaminación producida en el medio ambiente y se encuentra relacionado a la recuperación del material contaminante vertido en el lugar del derrame.	

A continuación, se procedió a contrastar diferentes algoritmos de Inteligencia Artificial utilizando el programa de ciencia de datos Knime (Knime, 2006). Se revisó el dataset y se limpiaron los datos, eliminando registros o columnas incompletas y variables poco o nada representativas como, por ejemplo: los identificadores de posición de cada registro. Luego se corrieron los diversos modelos de evaluación y, con la ayuda de la Matriz de Confusión, se verificó la exactitud de cada algoritmo. Finalmente, se obtuvo el algoritmo predictor de mayor exactitud con respecto a la posibilidad de un Derrame de Petróleo (ver Figura 1).

RESULTADOS

Usando el lenguaje de programación Python en COLAB, establecimos las localidades donde se generaron más derrames (ver Figura 2).

Calculamos la cantidad y familia de materiales contaminantes (ver Figura 3).

Se determinó la Fuente de los derrames y sus causas (ver Figura 4).

Se determinaron los derrames de los derivados del petróleo (ver Figura 5).

Se determinó las fuentes de los derrames del petróleo (ver Figura 6).

Se determinaron las causas de los derrames del petróleo (ver Figura 7).

Y para establecer la elección del mejor método predictor se procedió a comparar los resultados de los algoritmos: MultiLayer Perceptron (ver Figura 8), Naive Bayes (ver Figura 9), K-Nearest Neighbor (ver Figura 10) y Decision Tree (ver Figura 11).

Los resultados se muestran en la Tabla de Matriz de Confusión (ver Tabla 2).

De acuerdo a la Matriz de Confusión el algoritmo Decision Tree alcanzó la exactitud de 100 %, es

Figura 1
Arquitectura del proceso.

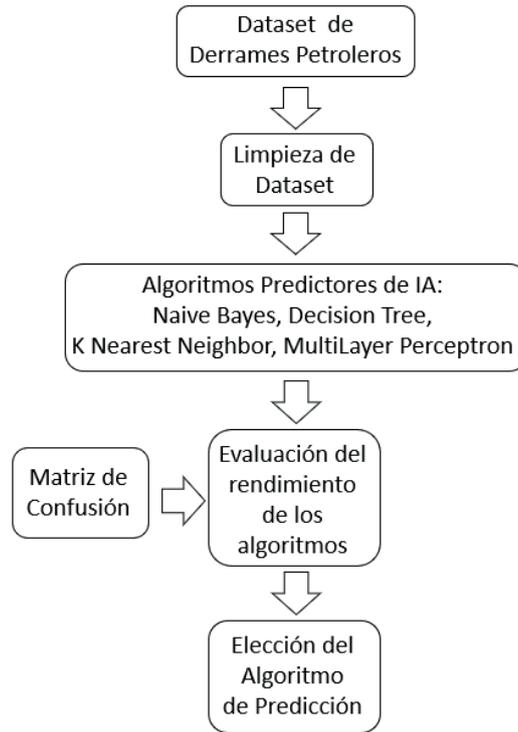


Figura 2
Localidades donde se generaron más derrames.

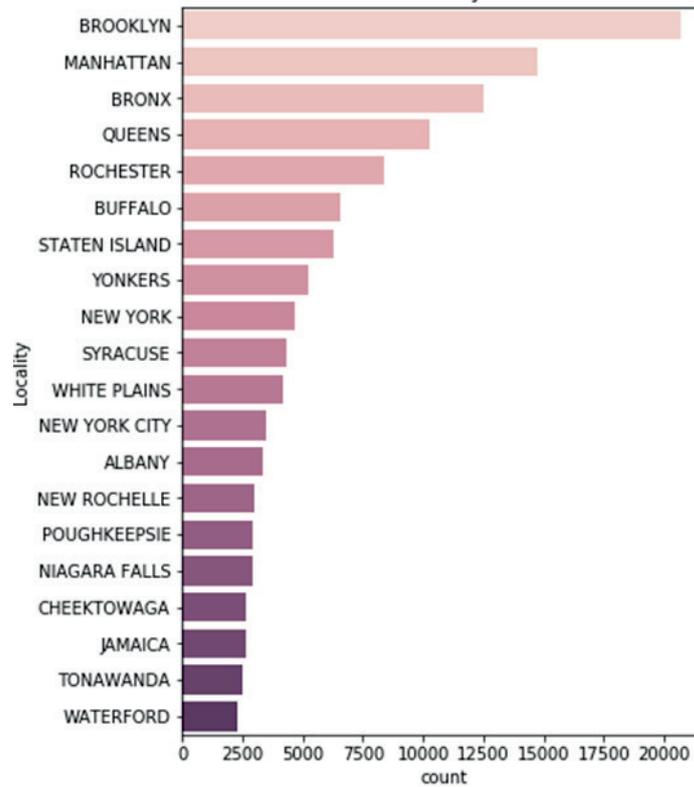


Figura 3
Familia de materiales contaminantes.

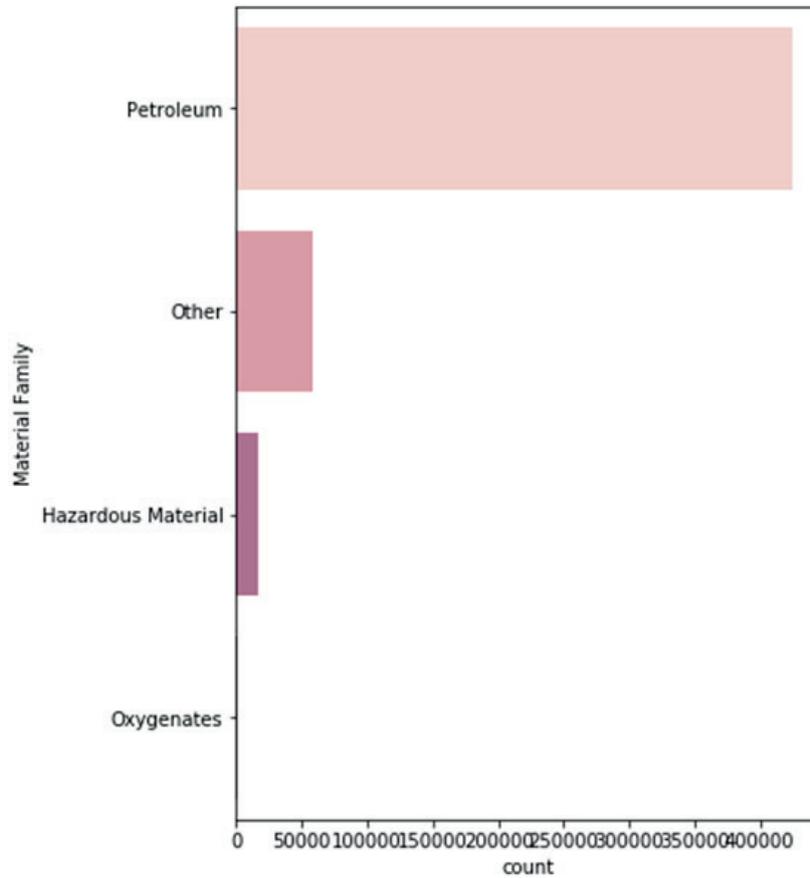


Figura 4
Mapa de Calor de Fuente de los derrames y sus causas.

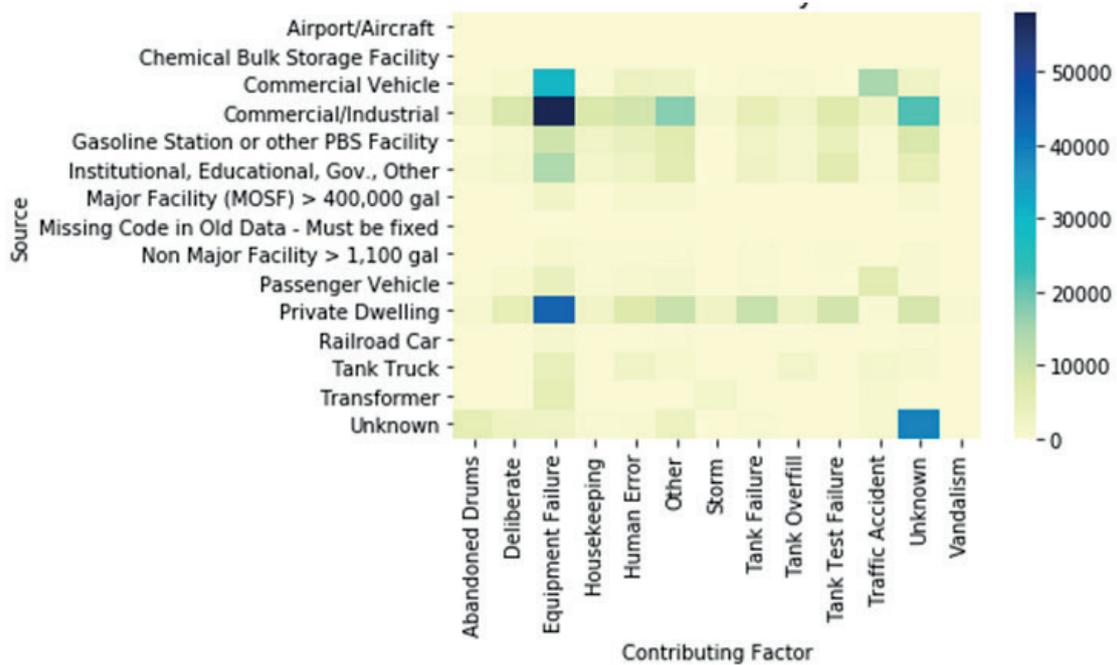


Figura 5

Derrames de los derivados del petróleo.

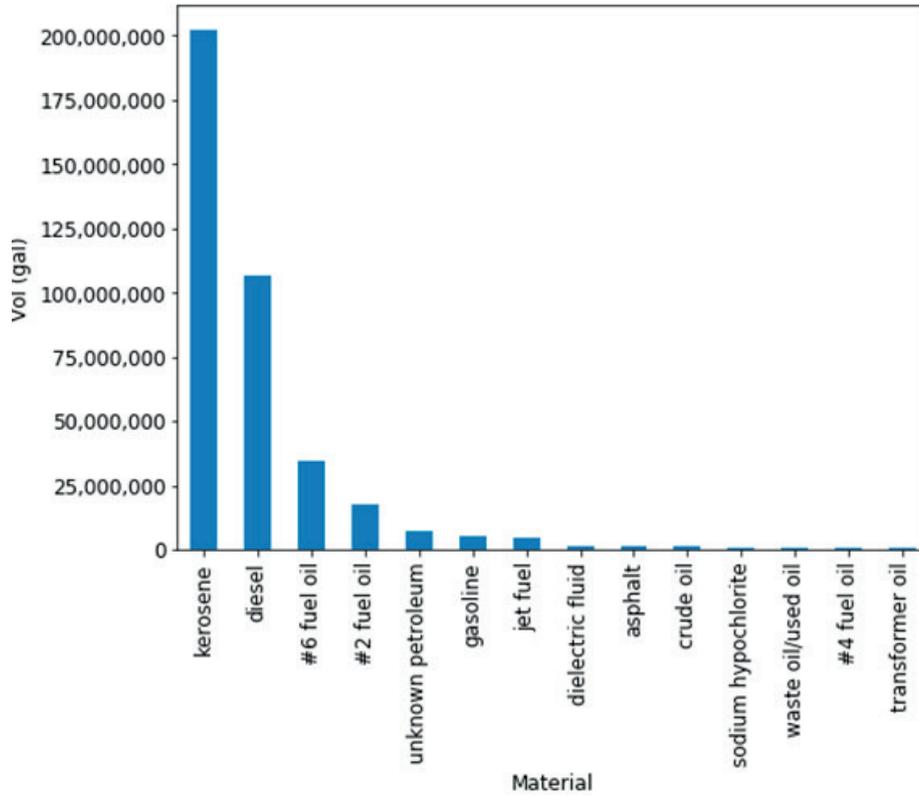


Figura 6

Fuentes de los derrames del petróleo.

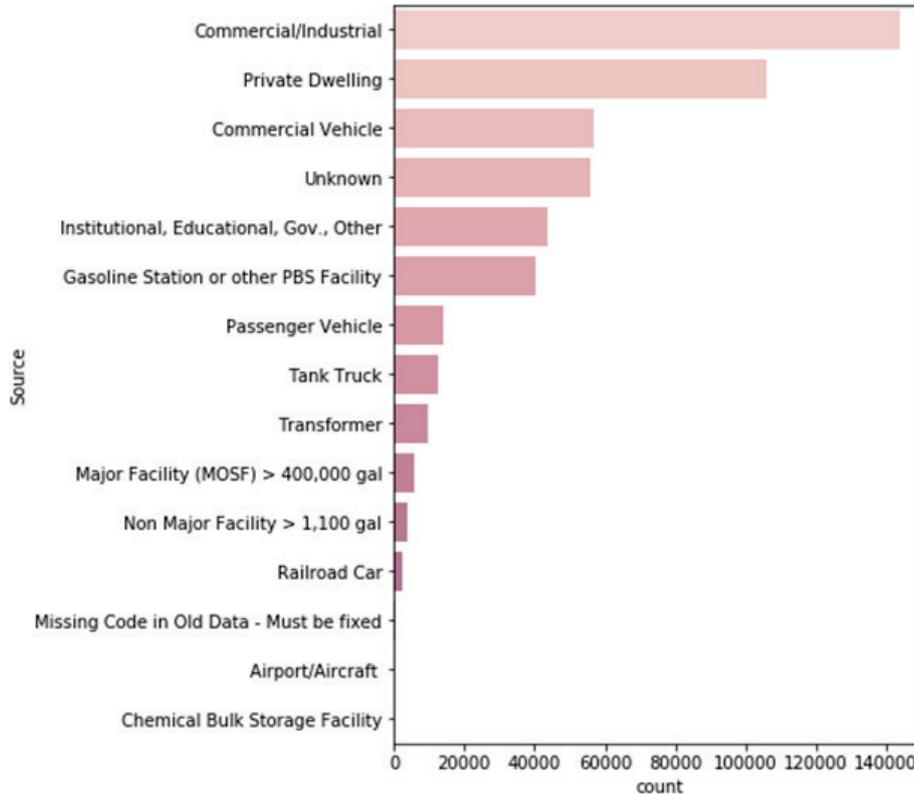


Figura 7
Causas de los derrames del petróleo.

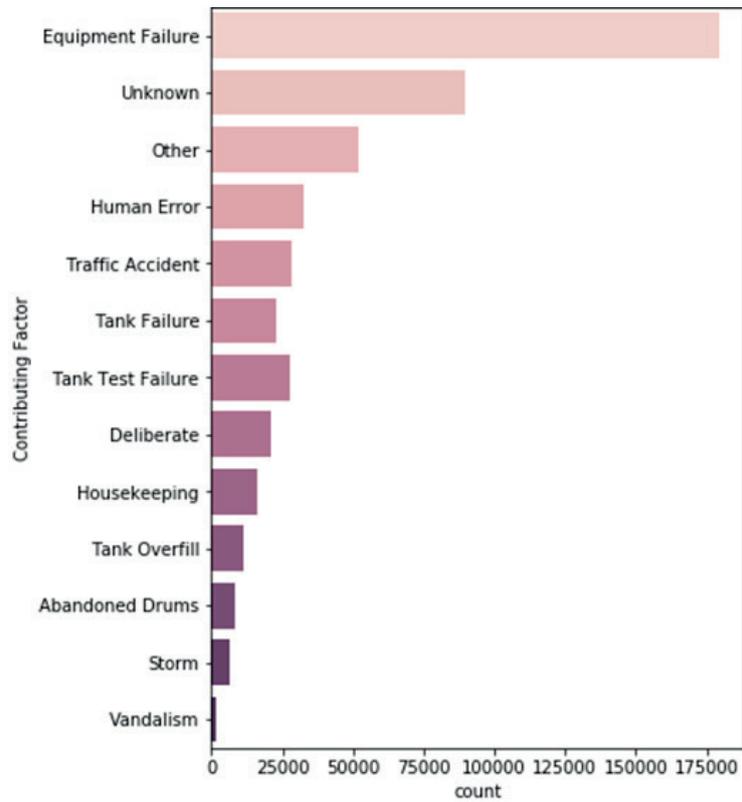


Figura 8
Algoritmo MultiLayer Perceptron

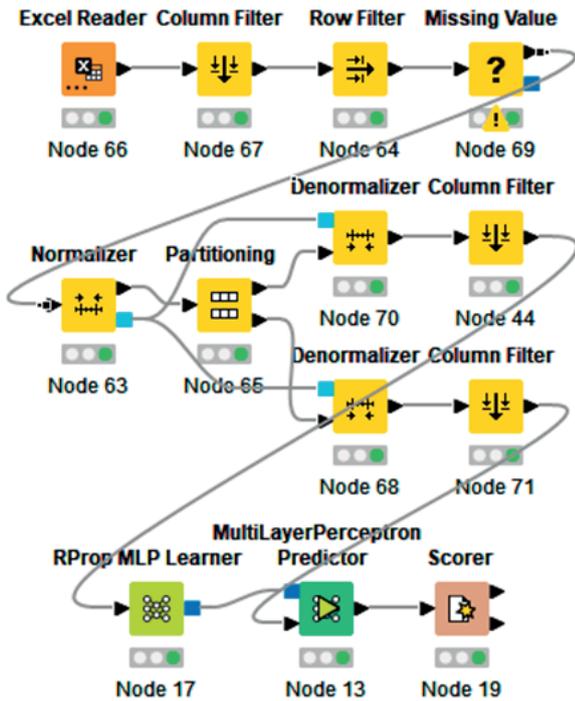


Figura 9
Algoritmo Naive Bayes

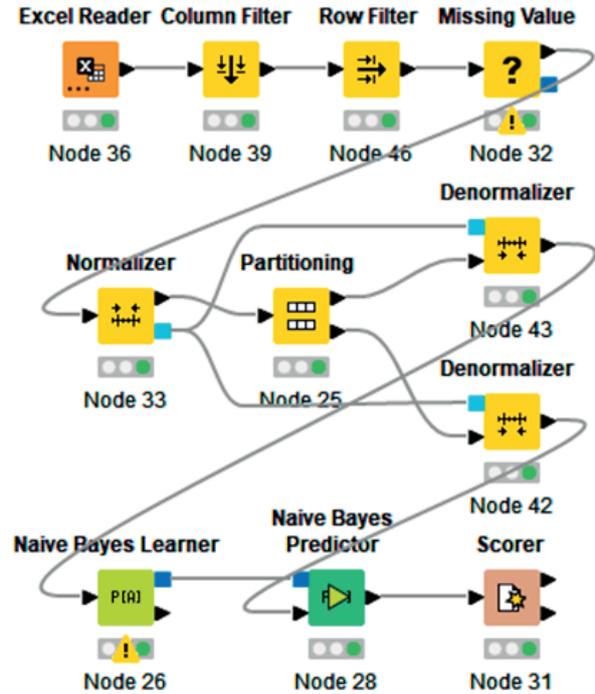


Figura 10

Algoritmo K-Nearest Neighbor

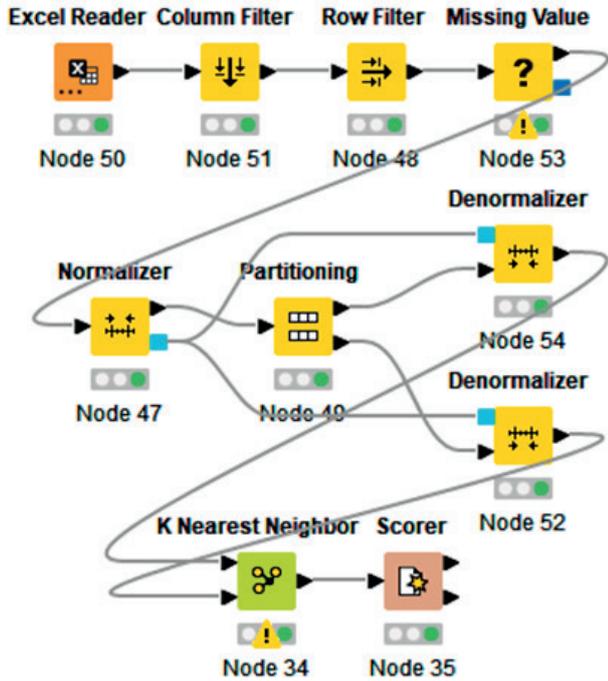


Figura 11

Algoritmo Decision Tree

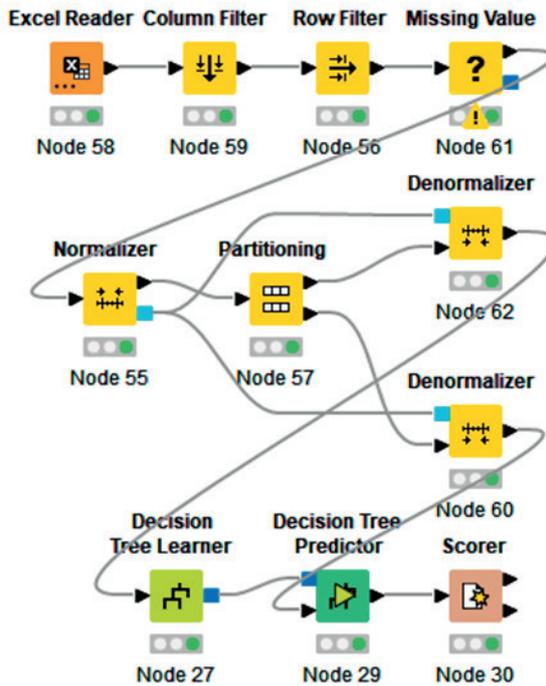


Tabla 2

Matriz de Confusión

Algoritmo	Clasificación Correcta	Exactitud	Clasificación Incorrecta	Error	Kappa Cohen (k)
MultiLayer Perceptron	110270	87.5%	15765	12.5%	0.000%
Naive Bayes	111970	88.8%	14065	11.2%	0.174%
K-Nearest Neighbors	112649	89.4%	13386	10.6%	0.367%
Decision Trees	126035	100.0%	0	0.0%	1.000%

decir, el máximo valor entre todos los algoritmos considerados en este estudio.

CONCLUSIONES

A pesar de que solo se utilizó data de derrames de petróleo del estado de New York de los Estados Unidos, las 18 variables que se consideraron, tras la limpieza de la data, fueron lo suficientemente representativas para evaluar con la Inteligencia Artificial el costo ambiental de dicho estado y, posicionó al modelo de Árbol de Decisión como el predictor más idóneo para determinar futuros desastres ambientales.

RECOMENDACIONES

Se recomienda ampliar el estudio a los demás estados de Norteamérica y posteriormente, considerando

variables similares a las evaluadas, ampliar el estudio a otros países del mundo.

También se deberían considerar variables que están intrínsecamente relacionadas a los factores ambientales y de salud y seguridad en el trabajo.

REFERENCIAS BIBLIOGRÁFICAS

[1] AprendeIA. (20/09/2019). *Naive Bayes - Teoría*. <https://aprendeia.com/naive-bayes-teoria-machine-learning/>

[2] Aprende Machine Learning. (10/07/2018). *Clasificar con K-Nearest-Neighbor ejemplo en Python*. <https://www.aprendemachinellearning.com/clasificar-con-k-nearest-neighbor-ejemplo-en-python/>

- [3] Cnnespanol. (04/10/2021). *Los peores derrames de petróleo de la historia*. <https://cnnespanol.cnn.com/2021/10/04/lpeores-derrames-petroleo-trax/>
- [4] IBM. (s.f.). *El modelo de redes neuronales*. <https://www.ibm.com/docs/es/spss-modeler/SaaS?topic=networks-neural-model>
- [5] Inga, A. (2018). Incertidumbre de datos geoespaciales en la exploración petrolera. *Revista Peruana de Computación y Sistemas 2018* 1(2):41-54, doi:10.15381/rpcs.v1i2.15378
- [6] Kaggle. (s.f.). *Oil and Gas*. <https://www.kaggle.com/datasets/raspberrypie/oil-and-gas>
- [7] Knime (Nº de versión 4.5.2). (2006). Open Source. Alemania: University of Konstanz.
- [8] Portal Ambiental. (s.f.). *Los mayores derrames de petróleo en Estados Unidos*. <https://www.portalam biental.com.mx/sabias-que/20211006/los-mayores-derrames-de-petroleo-en-estados-unidos>
- [9] Scikit Learn. (s.f.). *Decision Trees*. <https://scikit-learn.org/stable/modules/tree.html>

Fuentes de financiamiento:

Propia.

Conflictos de interés:

El autor declara no tener conflictos de interés.

Contribución del Autor

El autor en mención ha conceptualizado los objetivos, recopilando datos, definiendo metodologías y desarrollando la investigación basada en el software Python y sus bibliotecas especializadas en ciencias de datos.