



**TRABAJO DE GRADO
Opción Seminario-Diplomado.**

**ALGORITMO COMPUTACIONAL PARA EL ANÁLISIS Y TOMA DE DECISIONES
EN LA GENERACIÓN DE ENERGÍAS RENOVABLES, UTILIZANDO ESTRATEGIAS
DE MACHINE LEARNING**

Corporación Universitaria Remington.

Facultad de ciencias básicas e Ingeniería

Ingeniería de Sistemas

Estudiantes:

YEIMY MIREYA CUELLAR MEDINA

BRYAN STIWART SOTO ROJAS

Tutor: Juan Carlos Briñez de León

Opción de Trabajo de grado Seminario-Diplomado.

2024

Dedicatoria

A nuestras familias por ser un pilar de vital importancia en nuestra formación, por estar presentes en el proceso y por su apoyo incondicional llevando siempre nuestro camino por el sendero de la cultura, el conocimiento y la educación, punto en el que tanto énfasis tuvimos con la típica frase “la mejor herencia es la educación”. Gracias por insistir siempre porque a veces no veíamos con claridad lo que era mejor para nosotros.

Agradecimientos

A Dios primeramente, a nuestra casa de estudios UniRemington, a nuestros padres por la formación y acompañamiento en cada etapa del camino, a nuestros hermanos, a mi esposa a nuestros compañeros por muchas veces motivarnos y apoyarnos

Tabla de Contenidos

1	7	
1.1	9	
1.2	9	
		9
1.3	10	
1.4	10	
1.5	11	
1.6	12	
1.7	12	
1.8	12	
1.9	13	
1.10	15	
Objetivos:		14
1.11	15	
1.12	15	
1.12.1	16	
1.12.2	16	
1.12.3	17	
1.12.4	17	
1.12.5	18	
1.12.6	18	

	5
1.12.7	18
1.12.8	18
1.12.9	19
1.12.10	¡Error! Marcador no definido.
1.12.11	19
1.12.12	19
Conclusiones y trabajos futuros	22
Referencias	23

RESUMEN

La energía es la fuente de la vida, así como nuestro sistema nervioso central provee pequeñas descargas que permiten movimientos musculares y reacciones indispensables para el diario vivir. La ingeniería en su evolución descubrió la importancia de los mecanismos para facilitar las tareas diarias en la vida del hombre que posteriormente añadiéndoles una fuente de energía, se puede convertir un mecanismo simple en un mecanismo sofisticado con autonomía de movimiento y reacciones que ha sido de gran ayuda para el desarrollo del mundo moderno, como nuestros teléfonos celulares, ordenadores, vehículos y demás herramientas que complementan las tareas y que muchas veces sirven incluso como un asistente personal, permitiendo al hombre que sabe hacer uso de ellas, proyectar sus metas a objetivos más altos ya que mucha de su carga se aliviana gracias a la energía.

La energía es producida a través de movimientos mecánicos que se pueden dar por la quema de hidrocarburos, por la separación de átomos o incluso por el esfuerzo de la tracción humana o animal, sin embargo se puede notar que estos recursos son limitados y que en cualquier momento los niveles de producción pueden reducirse hasta su extinción. ¿Qué pasaría si logramos producir energía de una fuente autorenovable y autosustentable? En el presente documento se puede evidenciar como las energías renovables además de cumplir su propósito en las labores del hombre, conservan el cuidado del medio ambiente y reducen la necesidad de inversión, ya que, fuentes tan sencillas como un río con buen cauce es algo que se encuentra en la naturaleza y por lo que no se cobra. Mantener una relación simbiótica con los recursos naturales no solo debe ser una tarea u obligación, también debe ser considerado un lujo, ya que es más lo que se obtiene que su inversión,

Palabras clave

Energías renovables, Datasets, Análisis de datos, Machine learning, Clasificación, medio ambiente, energía solar, energía eólica, energía hidráulica

1 Marco conceptual y contextual

Al detenernos a pensar en cómo funciona nuestro mundo moderno, podemos evidenciar que todas las cosas y/u objetos realizan funciones gracias al uso de energía eléctrica en combinación con sistemas mecánicos simples pero concretos para llevar a cabo la tarea para la cual fue diseñada. cómo se mencionó en “Historia de la electricidad” (s.f.)

Aunque fue en 1646 la primera vez que apareció la palabra “elétrico” o “electricidad” (en una publicación en la obra Pseudodoxia Epidémica, del escritor Thomas Browne), la humanidad sabía desde mucho antes de las pequeñas descargas eléctricas que transmitían algunos peces. Incluso en textos del Antiguo Egipto, que datan del 2750 a.C, los autores se referían a estos peces como “los tronadores del Nilo”. Escritores antiguos describieron la sensación al tocar estos peces como un efecto de adormecimiento, que era propiciado por las descargas eléctricas que emitían estos peces y rayas eléctricas. Estos hechos conforman el inicio de lo que conocemos como historia de la electricidad. “Historia de la electricidad” (s.f.)

La energía eólica es uno de los sistemas utilizados, llevando así muchos años cumpliendo con esta función, aunque en ocasiones no sea la más apta debido a que como su nombre lo indica la energía que produce es por medio del viento y en ocasiones se escasea del mismo provocando que la producción baje, pero sigue siendo una de las energías renovables más utilizadas. Por otro lado, la energía hidráulica cuyo principal elemento utilizado para su proceso es el agua, siendo así un elemento fundamental para todo ser vivo, esta energía es utilizada desde hace mucho tiempo y al pasar de los años se ha venido implementando nuevas tecnologías que no

afecten tanto al medio ambiente, siendo este uno de los principales problemas la gran cantidad de herramientas utilizadas para realizar todo el proceso y obtener como producto final la energía. Es de vital importancia que las energías renovables sean utilizadas de la mejor manera para que así mismo no generen un impacto ambiental que se pueda ver afectado en un futuro no muy lejano.

1.1 Pregunta problema:

¿Cómo desarrollar una estrategia computacional para ayudar en la identificación temprana de generación de energías renovables, a partir de algoritmos de Machine Learning?

1.2 Acercamiento a los datos:

los datos para establecer el crecimiento y la distribución de energía renovable en todo el mundo fueron tomados de la plataforma kaggle, publicados por AHMED GAITAN con el nombre “**Producción mundial de energía renovable (2000-2023)**”, el cual trata de un estudio que surge de la recopilación de datos obtenidos por funcionarios del Cairo (Egipto) tomando un periodo desde el 2000 hasta el 2023. En el que se almacena información anual relevante como las lecturas de energías producidas por corrientes de agua, eólica, solar, entre otras fuentes de energía renovable.



1.3

1.4 Figura 1. imagen del conjunto de datos

1.5 Descripción de variables.

El dataset escogido para poder realizar el trabajo cuenta con 7 columnas divididas en 6 variables de entrada y una de salida que corresponde al total de energías renovables, la cual podemos observar en la siguiente figura, donde nos muestra la lista de variables contenidas en la base de datos.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 240 entries, 0 to 239
Data columns (total 7 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                -
0   Year                                  240 non-null   int64
1   Country                              240 non-null   object
2   SolarEnergy                          240 non-null   float64
3   WindEnergy                           240 non-null   float64
4   HydroEnergy                          240 non-null   float64
5   OtherRenewableEnergy                 240 non-null   float64
6   TotalRenewableEnergy                 240 non-null   float64
dtypes: float64(5), int64(1), object(1)
memory usage: 13.2+ KB
```

1.6 Figura 2. Lista de variables en el conjunto de datos

Donde se puede observar, que los datos de las variables son de tipo numérico, las cuales 5 de ellas son tipo decimal, una es de tipo entera y una es de tipo texto, en el caso de la variable `country` la cual es la que nos indica que ciudades utilizan los diferentes tipos de energías renovables, asimismo se puede evidenciar en qué año fue utilizada el tipo de energía, podemos ver también los tipos de energía renovables y el total de estas utilizadas en un año.

En la tabla se pueden observar siete variables de las cuales se encuentran seis numéricas (cinco decimales y una entera) y una de texto que más adelante transformaremos en numérica para realizar el análisis como se muestra en la figura 3. La variable “*Year*” nos muestra un periodo recopilado desde el 2000 al 2023, la variable “*Country*” muestra las lecturas tomadas en diferentes países y regiones del mundo, las variables “*SolarEnergy*, *WindEnergy* y *HydroEnergy*” muestran de manera discriminada los datos obtenidos de la producción de las energías Solar, Eólica e Hidráulica; como también, el total de valores por año que sería nuestro dato de salida.

```
[ ] #Mapeando todas la variables categóricas a numéricas
Reemplazo={'USA':0,'China':1,'India':2,'Germany':3,'UK':4,'France':5,'Brazil':6,'Canada':7,'Australia':8,'Japan':9}
Datos_Loan['Country']=Datos_Loan['Country'].map(Reemplazo)

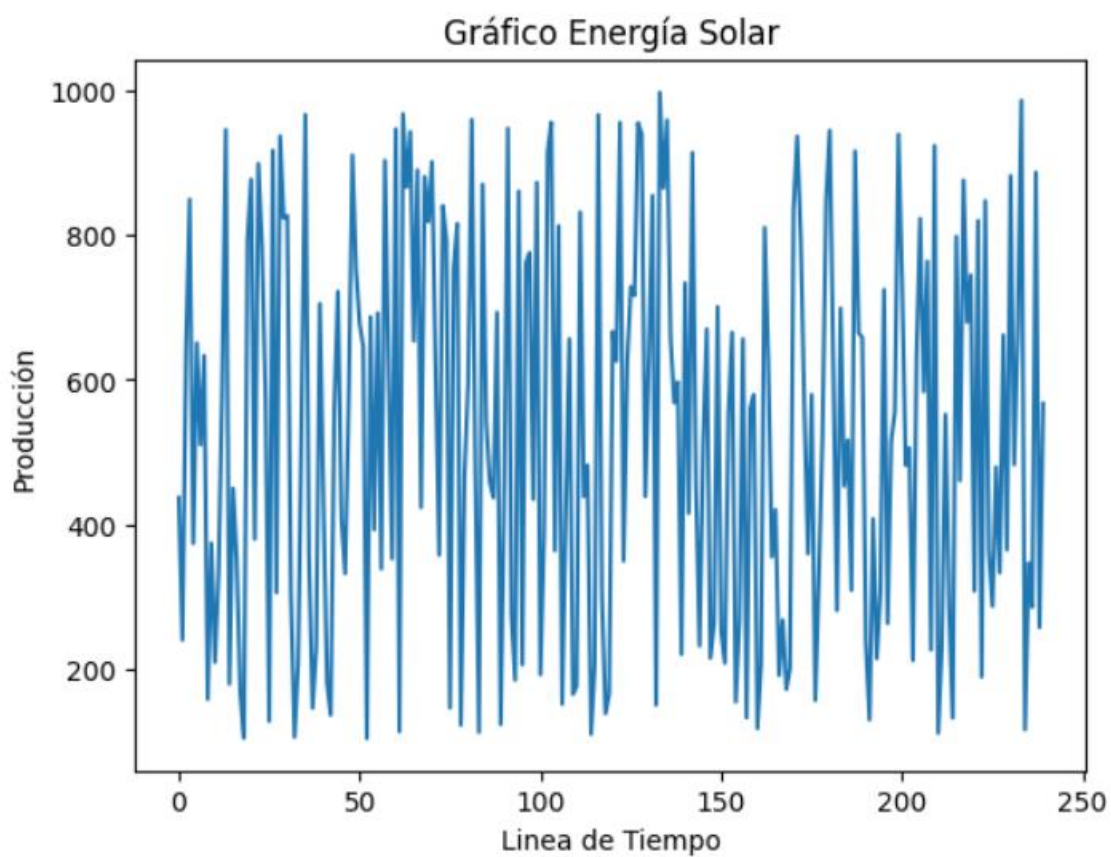
Datos_Loan.head()
```

	Year	Country	SolarEnergy	WindEnergy	HydroEnergy	OtherRenewableEnergy	TotalRenewableEnergy
0	2000	0	437.086107	1435.928598	1544.389701	319.396318	3736.800724
1	2001	0	240.416776	402.792876	398.742141	439.779266	1481.731059
2	2002	0	641.003511	1120.494351	334.993640	486.459433	2582.950935
3	2003	0	849.198377	476.040844	609.102444	132.532029	2066.873694
4	2004	0	373.818019	882.183361	1034.306532	181.053113	2471.361025

1.7 Figura 3. Cambios de tipo de variables

1.8 Aproximaciones con gráficos - analítica.

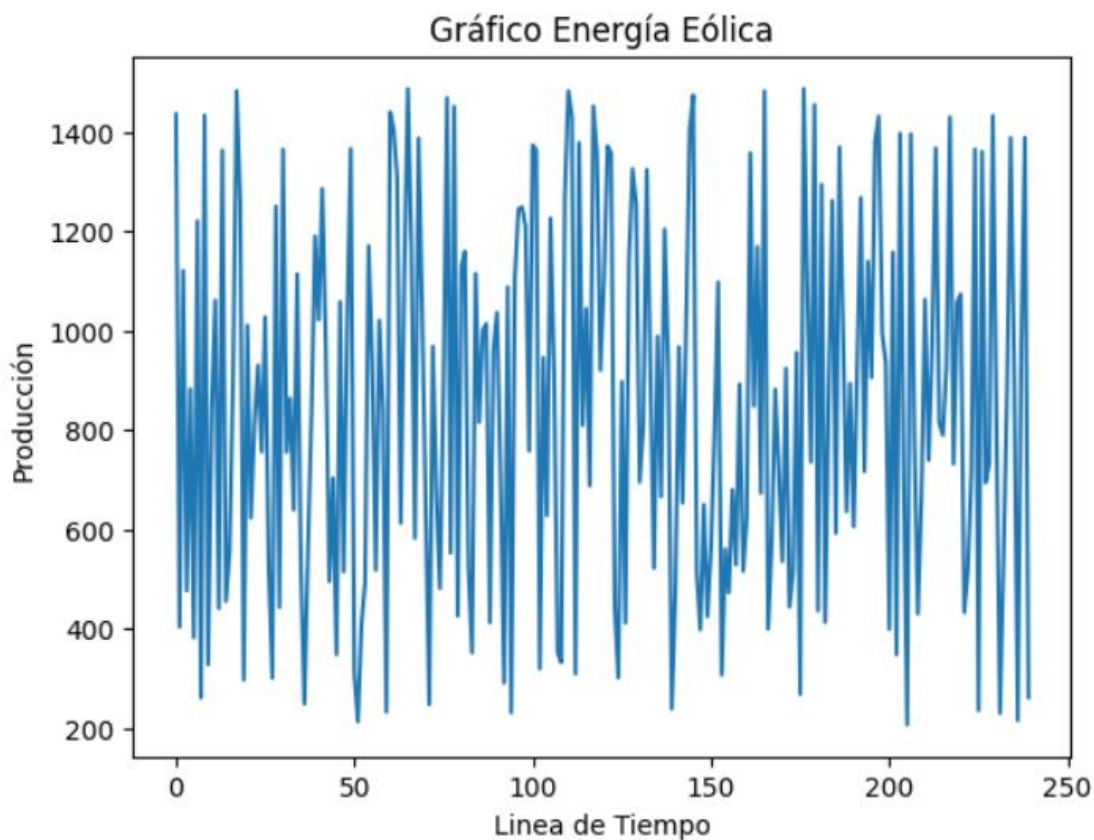
Según el estudio realizado por el análisis de los algoritmos obtenemos que el valor con frecuencia para la energía Solar es de 600 Gigavatios por hora como se muestra a continuación:



1.9 Figura 4. Producción de energía Solar

Otros aspectos relevantes son que el pico más bajo en temporadas de invierno o cuando más escasee la luz solar por presencia de cielo nublado se pueden producir alrededor de 200 GWh mientras que, en temporadas de verano la producción puede alcanzar los 1000 GWh.

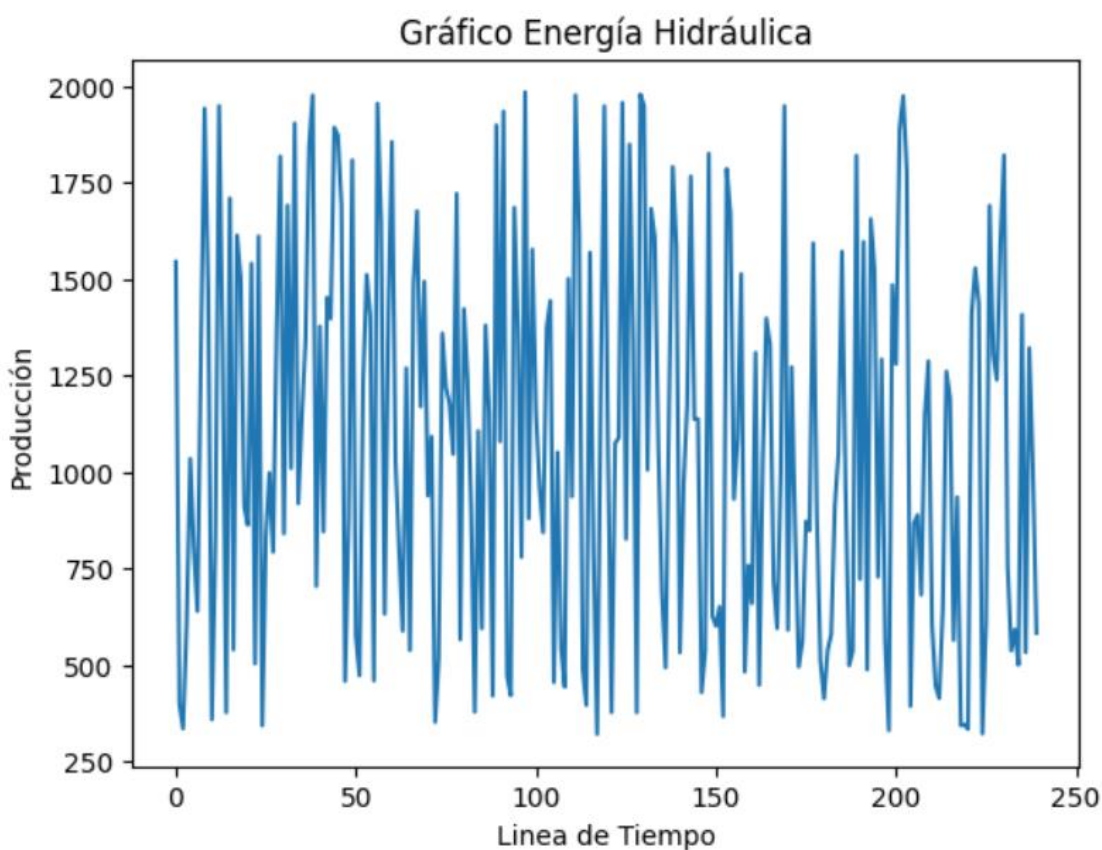
En el estudio de la energía Eólica podemos determinar que la media ronda entre los 900 GWh como se revela en la siguiente gráfica:



1.10 Figura 5. Producción de energía Eólica

De la gráfica anterior podemos concluir que, en la temporada más baja de viento se pueden producir hasta 550 GWh y como tope máximo nos encontramos una producción que asciende por arriba de los 1400 GWh

Para la energía Hidráulica tenemos un rango bastante interesante, enfocándonos en la fuente de producción podemos notar que, esta fuente es la más estable ya que si nos ubicamos en una zona fluvial las corrientes de agua, además de ser muy poco variables, también son muy poderosas a la hora de generar energías cinéticas y mecánicas que posteriormente convertiremos en Electricidad. Esta fuente de energía tiene una frecuencia ubicada en los 1375 GWh superando el tope máximo de las anteriores como se puede apreciar en la gráfica de la figura 6.



1.11 Figura 6. Producción de energía Eólica

Yendo más al detalle se puede notar que en los tiempos más extremos de sequía o cuando disminuyen los cauces de los ríos y represas la producción puede bajar hasta los 900 GWh (producción promedio de la energía eólica), mientras que en sus picos más altos puede fácilmente alcanzar los 2000 GWh, siendo nuestra fuente de energía top.

Para garantizar la seguridad de la producción de energía hidráulica sólo debemos conservar los espacios hídricos naturales, restringiendo el acceso en ciertos lugares, fomentando la cultura ciudadana para el cuidado de los ecosistemas y preservación del medio ambiente.

1.12 Objetivos:

1.13 Objetivo general.

Implementar un algoritmo computacional para el análisis y toma de decisiones a partir de datos de generación de energías renovables, utilizando estrategias de machine learning.

1.14 Objetivos específicos.

- Caracterizar y procesar los datos de interés, con miras a la toma de decisiones informadas.
- Implementar un algoritmo de Machine learning para la toma de decisiones a partir de los datos de interés.

- Evaluar y analizar el desempeño de los algoritmos implementados para la toma de decisiones.
- Validar el funcionamiento de toma de decisiones a partir de datos nuevos.

1.14.1 Desarrollo e implementación del aprendizaje

Inicialmente se toma el dataset y se evalúan superficialmente los datos con ayuda de Excel para determinar si cumple con los criterios de la búsqueda, se concluye que la lectura de datos obtenida del dataset cumple las expectativas y se procede a realizar algoritmos para organizar los datos de manera rápida y efectiva.

1.14.2 Preparación de los datos

Cargamos el dataset a nuestro IDE con ayuda del siguiente código

```
#Cargaremos los datos
import pandas as pd
from google.colab import files
uploaded = files.upload()
for filename in uploaded.keys():
    Datos_Loan = pd.read_csv(filename, sep=',')

Datos_Loan.head()
```

1.14.3 Figura 7. Cargue de datos

• `global_renewable_energy_production.csv`(text/csv) - 25124 bytes, last modified: 8/7/2024 - 100% done
 Saving `global_renewable_energy_production.csv` to `global_renewable_energy_production.csv`

	Year	Country	SolarEnergy	WindEnergy	HydroEnergy	OtherRenewableEnergy	TotalRenewableEnergy
0	2000	USA	437.086107	1435.928598	1544.389701	319.396318	3736.800724
1	2001	USA	240.416776	402.792876	398.742141	439.779266	1481.731059
2	2002	USA	641.003511	1120.494351	334.993640	486.459433	2582.950935
3	2003	USA	849.198377	476.040844	609.102444	132.532029	2066.873694
4	2004	USA	373.818019	882.183361	1034.306532	181.053113	2471.361025

1.14.4 Figura 8. Retorno rápido de los datos

Se evalúan variables, tipos de variables y cantidad de las mismas

```
#Evaluamos las variables
Datos_Loan.info()
```

1.14.5 Figura 9. Evaluación de variables

1.15

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 240 entries, 0 to 239
Data columns (total 7 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Year                   240 non-null   int64
1   Country                240 non-null   object
2   SolarEnergy            240 non-null   float64
3   WindEnergy             240 non-null   float64
4   HydroEnergy            240 non-null   float64
5   OtherRenewableEnergy  240 non-null   float64
6   TotalRenewableEnergy  240 non-null   float64
dtypes: float64(5), int64(1), object(1)
memory usage: 13.2+ KB
```

1.15.1 Figura 10. Retorno de análisis de variables

En este punto optimizaremos nuestra Database descartando datos nulos o columnas no deseadas.

```
#Eliminamos filas que tengan datos nulos para optimizar el cálculo
Datos_Loan=Datos_Loan.dropna()
Datos_Loan.head()
```

1.15.2 Figura 11. Algoritmo de descarte de nulos

	Year	Country	SolarEnergy	WindEnergy	HydroEnergy	OtherRenewableEnergy	TotalRenewableEnergy
0	2000	USA	437.086107	1435.928598	1544.389701	319.396318	3736.800724
1	2001	USA	240.416776	402.792876	398.742141	439.779266	1481.731059
2	2002	USA	641.003511	1120.494351	334.993640	486.459433	2582.950935
3	2003	USA	849.198377	476.040844	609.102444	132.532029	2066.873694
4	2004	USA	373.818019	882.183361	1034.306532	181.053113	2471.361025



1.15.3 Figura 12. Retorno de datos optimizados

Reemplazamos la variable de tipo texto (“Country”) para analizar la información según el país o región.

```
#Mapeando todas la variables categóricas a numéricas
Reemplazo={'USA':0,'China':1,'India':2,'Germany':3,'UK':4,'France':5,'Brazil':6,'Canada':7,'Australia':8,'Japan':9}
Datos_Loan['Country']=Datos_Loan['Country'].map(Reemplazo)

Datos_Loan.head()
```

1.15.4 Figura 13. Mapeo de la información



	Year	Country	SolarEnergy	WindEnergy	HydroEnergy	OtherRenewableEnergy	TotalRenewableEnergy
0	2000	0	437.086107	1435.928598	1544.389701	319.396318	3736.800724
1	2001	0	240.416776	402.792876	398.742141	439.779266	1481.731059
2	2002	0	641.003511	1120.494351	334.993640	486.459433	2582.950935
3	2003	0	849.198377	476.040844	609.102444	132.532029	2066.873694
4	2004	0	373.818019	882.183361	1034.306532	181.053113	2471.361025

1.15.5 Figura 14. Retorno de la información modificada a conveniencia

Realizando los pasos anteriores podremos obtener los análisis y resultados esperados como se evidencia en las figuras 4, 5 y 6.

1.15.6 Modelo de toma de decisiones

En este caso el modelo utilizado es el de regresiones ya que nos permite predecir el tipo de energía utilizada en el año 2024

1.15.7 Validación del modelo

Como funciona para datos nuevos, y como un usuario le da uso a los algoritmos

Figuras y tablas

Las tablas y figuras debes ponerlas en la página donde las mencionas por primera vez en el texto. Si tu tabla o figura es muy grande debes ponerla en una página separada.

Tablas y figuras deben ser puestas en páginas diferentes independientemente de su tamaño. No debes dejar espacios en blanco en las páginas de texto, pero es posible dejar espacio en blanco en páginas que solo contienen tablas y figuras.

También puedes poner las tablas y figuras en un capítulo llamado “Apéndice” al final del trabajo de grado. Si haces esto, debes indicar en el texto, que existe este Apéndice en tu trabajo de grado.

Los títulos de las tablas deben ser puestos sobre las mismas y en las figuras deben ponerse debajo. Todas las tablas deben tener al menos dos columnas y la fila de títulos.

Los títulos de las tablas deben ser puestos sobre las mismas y en las figuras deben ponerse debajo. Todas las tablas deben tener al menos dos columnas y la fila de títulos.

Tabla 1. El título debe ser breve y descriptivo.

<i>Columna x</i>	<i>Columna Y</i>
------------------	------------------

Datos

Datos

Datos

Datos

Conclusiones y trabajos futuros

En este trabajo se realizó la construcción de un algoritmo para la toma de decisiones para la predicción de energías renovables utilizando algoritmos de machine learning, mostrando así que la energía con más utilidad es la hidráulica.

La energía que se produce a través de fuentes limpias y renovables además de tener muy buen potencial productivo, tiene un gran impacto ecológico que a su vez fomenta la preservación de los espacios naturales, lo que garantiza a la humanidad muchos más recursos para la producción y desarrollo de grandes ideas que sigan innovando nuestro diario vivir.

Sería de gran ayuda para la población colombiana, la inversión y construcción de represas para la producción de energía hidráulica que abastezcan cada rincón del país y generen empleos y conciencia en la juventud.

Para trabajos futuros, se proyecta la implementación del modelo directamente en la industria, donde se pueda tener un comparativo en tiempo de real de la energías renovables cual es la más usada y lo ofrecido mediante el modelo de machine learning.

Referencias

Historia de la electricidad. (s.f.). Endesa.

Moragues, J., & Rapallini, A. (2003). *Energía eólica.* Buenos Aires: Instituto Argentino de la Energía General Mosconi.

Soria, E., Mosquera, P., & Ruesga, L. M. (2008). *Energía hidráulica.* Haya

Eólica, E. (2014). *Energía eólica.*

Solar, P. (2020). *Energía solar.* *Acesso em, 15(03), 202013-2014.*

Paredes, L. F. (2017). *Modelado matemático de la incertidumbre asociada a la generación de energías renovables* (Doctoral dissertation, Universidad Pública de Navarra).