

Concurso de Ingeniería Financiera IV

Proyecto:

Wizeline -US Energy Council 1

*Alfonso Aguilar Aguirre*

*Andrés García Ruvalcaba*

*José Antonio Quintero Rodríguez*

*Ana Paula Rangel Ochoa*

Índice

[Introducción: 3](#_Toc529474952)

[¿Cómo se abordó el problema? 3](#_Toc529474953)

[Teoría 4](#_Toc529474954)

[Desarrollo: primera visualización de los datos 4](#_Toc529474955)

[Desarrollo: tratamiento de datos en Python. 5](#_Toc529474956)

[Desarrollo: implementación de métodos de inteligencia artificial en Matlab 5](#_Toc529474957)

[Conclusiones 6](#_Toc529474958)

# Introducción:

Wizeline nos presentó una base de datos , obtenida de la US Energy Administration (EIA), que se puede encontrar en : <https://www.eia.gov/electricity/data/state>. Dicha base datos nos presenta un registro anual del megawattaje/hora de los recursos energéticos separados por estado. Se espera que después de hacer un correcto análisis de datos, podamos responder preguntas que la empresa nos haría, como:

En caso de crecimiento o decrecimiento de un tipo de fuente energía …

* ¿Qué tan relevante ese ese cambio en particular para las regiones afectadas?
* ¿Cuál podría ser una posible causa?
* ¿Sería redituable invertir en dicha fuente de energía?
* ¿Es posible construir una gráfica que muestre la sobreproducción y la subproducción de algún tipo de energía en ciertas regiones?
* ¿Es posible reconstruir dicha gráfica dinámicamente de manera que se muestra la información por productores, cambio en el crecimiento y por año?
* ¿Es posible construir una gráfica que muestre el consumo de energía por región?

# ¿Cómo se abordó el problema?

Dada la base de datos obtenida de la US Energy Administration (EIA), se procedió a analizar la información mediante el uso de tablas y gráficas dinámicas, que nos pudieran mostrar la información más relevante de una manera sencilla y visual para facilitar el desarrollo de respuestas a las preguntas hechas por Wizeline. Posteriormente, fue notorio el hecho de que era necesario hacer tratamiento de los datos, por medio de Python. Dichos datos se utilizaron para dos cosas:

1. Se exportó a Excel la información tratada para mejorar las tablas y gráficas dinámicas
2. Se importó el mismo Excel a Matlab y se utilizó un modelo Adaline que propone crear un pronóstico de producción de energía basado en los datos históricos de los cambios porcentuales de cada fuente de energía en cada estado del país.

# Teoría

* Un método de inteligencia artificial es una rama de la computación que relaciona un fenómeno natural con una analogía artificial a través de programas de computadora. La IA puede ser tomada como ciencia si se enfoca hacia la elaboración de programas basados en comparaciones con la eficiencia del hombre, contribuyendo a un mejor entendimiento del conocimiento humano.
* El adaline (de ADAptative LINear Element) es un tipo de [red neuronal artificial](https://es.wikipedia.org/wiki/Red_neuronal_artificial) desarrollada por el profesor [Bernard Widrow](https://es.wikipedia.org/w/index.php?title=Bernard_Widrow&action=edit&redlink=1) y su alumno [Ted Hoff](https://es.wikipedia.org/w/index.php?title=Ted_Hoff&action=edit&redlink=1) en la [Universidad de Stanford](https://es.wikipedia.org/wiki/Universidad_de_Stanford) en 1960. Generalmente se compone de una sola capa de n neuronas (por tanto n valores de salida ) con m entradas con las siguientes características:
  + Las m entradas representan un vector x de entrada que pertenece al espacio R m
  + Por cada neurona, existe un vector w de pesos sinápticos que indican la fuerza de conexión entre los valores de entrada y la neurona. En la práctica representan la [ponderación](https://es.wikipedia.org/w/index.php?title=Ponderaci%C3%B3n&action=edit&redlink=1) de cada entrada sobre la neurona.
  + Una constante θ.
  + La salida y de la neurona se representa por la función de activación, que se define como y = ∑ i = 1 n x i w i + θ

# Desarrollo: primera visualización de los datos

Este fue el arranque del proyecto en el que se decidió lograr una mejor visualización de los datos dándoles ordenes específicos de acuerdo a fuentes de energía, estado y años, esto permitió entender más a detalle la información y la evolución de las variables en el tiempo.

Para atender a la primera pregunta se tenía que tener claro el cambio en la producción de energía obtenida a partir de cada diferente fuente de energía para cada estado. Para este punto la visualización realizada con tablas y gráficas dinámicas no fue suficiente, por lo que se recurrió a los conocimientos de ciencia de datos, específicamente de limpieza y tratamiento, utilizando la herramienta que el lenguaje de programación Python brinda.

# Desarrollo: tratamiento de datos en Python.

Debido a las facilidades que la utilización de “data frames” con Python genera y a la implementación de matrices para realizar cálculos, fue posible obtener el cambio porcentual en la producción total de energía por estado a lo largo del tiempo y en la producción obtenida a partir de cada fuente de energía, también de cada estado. Lo que nos proporcionará información acerca de la importancia o el peso que tiene el crecimiento o disminución de la producción obtenida a partir de una fuente específica en la sobre o subproducción de energía total en cada estado desde 1990 hasta 2017.

# Desarrollo: implementación de métodos de inteligencia artificial en Matlab

En esta parte del proyecto se implementó el modelo Adaline para la serie de tiempo de cada estado, se logró obtener los pesos para las variables que conforman el modelo de predicción de producción de energía, lo que también significa encontrar la fuente de energía que tiene más importancia a la hora de calcular la producción total de energía por estado.

Variables del modelo: los cambios porcentuales a lo largo del tiempo de energía por cada fuente diferente para cada estado.

Variable de salida: cambio porcentual de producción total de energía por estado.

Supuestos:

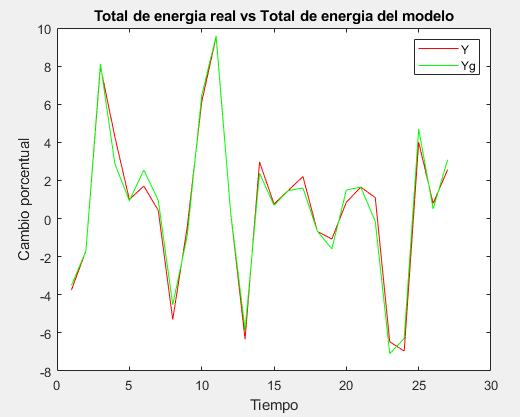
El modelo es de primer grado para que los pesos correspondan solo a los tipos de fuentes de energía y no a más variables creadas.

El porcentaje de crecimiento de un estado está dado por una ecuación lineal que tiene con o variables los cambios en la producción de energía de cada fuente diferente.

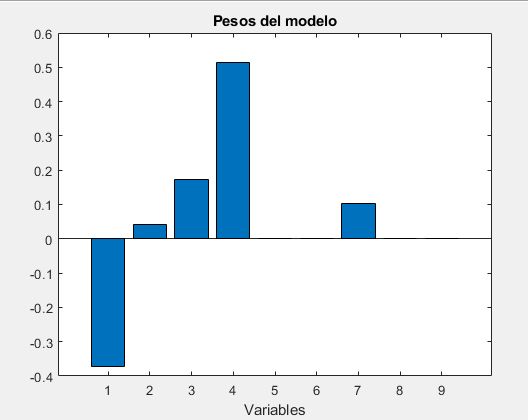
Para el modelo Adalina se utiliza el método de mínimos cuadrados, donde el error queda cuantificado.

# Conclusiones

Se puede concluir que el modelo utilizado para predecir el cambio en la producción de energía es correcto, dado que se compararon las gráficas de los datos reales contra los datos de las predicciones del modelo, el cual se ajusta bastante bien, puesto que el error es bajo. Para muestra, se corrió el modelo para el estado AK y en la gráfica de la fuente de energía se está graficando Coal.



A partir del gráfico de barras mostrado a continuación se puede conocer la variable que más afecta en el pronóstico, es decir, el cambio porcentual de la fuente de energía que más impacto genera en el cambio de la producción total del estado.



Para concluir, se hace una comparación entre el cambio en la producción de energía total del estado y el cambio en la producción total de la fuente específica.

