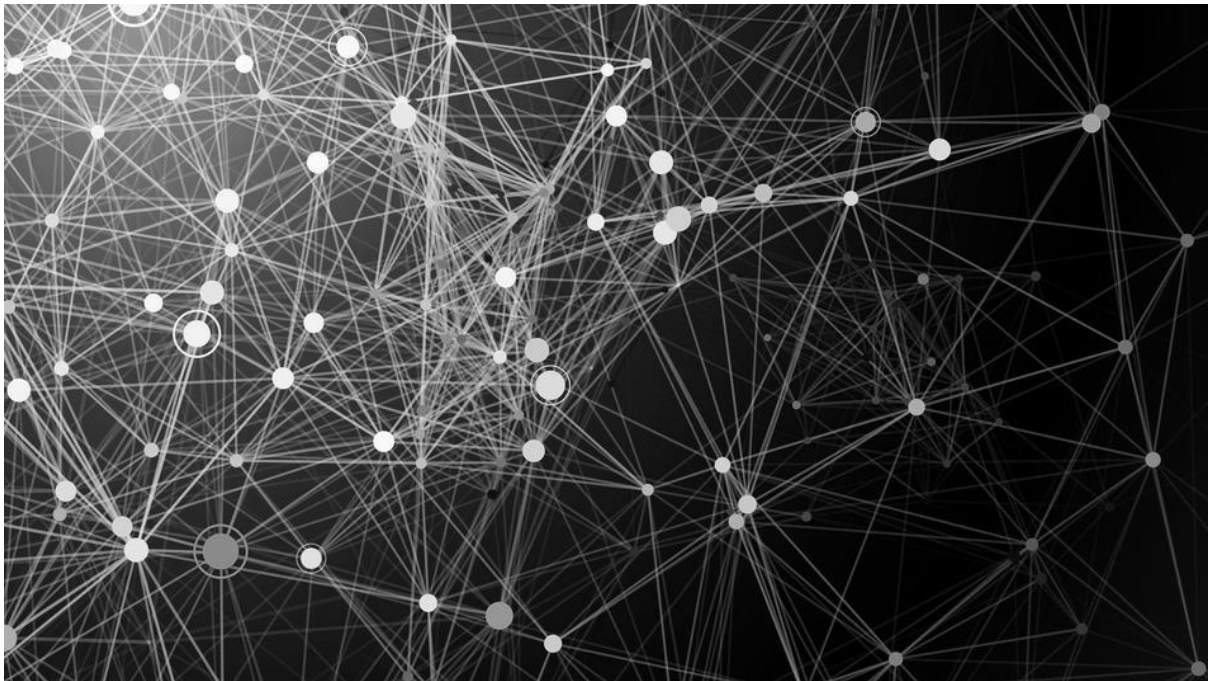


PROYECTO DE ANÁLISIS DE DATOS: “windPower”



Grado en Ingeniería Informática - Universidad de La Laguna

Itinerario: Computación - **Curso:** Tercero

Asignatura: Tratamiento Inteligente de Datos

Miembros del equipo: Adrián Rodríguez Bazaga, Eleazar Díaz Delgado, Rudolf Cicko, Ruymán Rodríguez Martín

Índice

1. Introducción	1
2. Datos a analizar	2
3. Objetivo del análisis	3
4. Variables	4
5. Técnicas que se emplearán	5
6. Fuente de los datos	6
7. Dificultades que pueden presentarse	7
8. Aplicaciones previas	8

Introducción

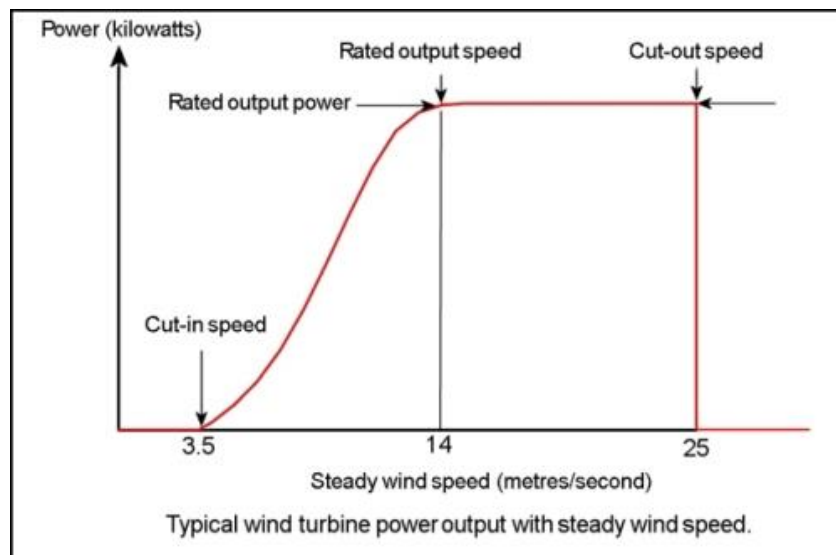
En este proyecto de análisis de datos realizado en la Universidad de La Laguna, en los estudios de Grado en Ingeniería Informática para la asignatura Tratamiento Inteligente de Datos trataremos de obtener a través de los datos que tenemos una estimación sobre si en determinados estados de Estados Unidos se podría generar mayor cantidad de energía renovable eólica que la que se genera actualmente (calculando el diferencial), estudiando además el posible ahorro en el uso de energía nuclear que se podría hacer (que emite residuos muy peligrosos), añadiendo también el porcentaje de energía que ya no necesitará dicho estado comprar a otro estado (en caso de que lo haga, esto ocurre cuando 'Consumo de energía > Producción de energía'), consiguiendo con esto un ahorro económico significativo (no será parte de este estudio el de estimar cuánto es dicho ahorro económico).

Datos a analizar

Los datos a analizar son los siguientes: Velocidad media del viento por ciudades en cada estado de Estados Unidos, producción anual de energía eólica por estado, extensión de terreno por estado y porcentaje de dicho terreno que es público (pertenece al gobierno), producción anual de energía nuclear y por último el consumo energético por cada estado.

A parte de estos datos tenemos ciertas constantes que son usadas para determinar si las condiciones del lugar en cuestión son óptimas, estas son:

- Velocidad del viento óptima para un aerogenerador: 14 metros por segundo



- Extensión media requerida de un parque eólico: 40 kilómetros cuadrados.

Por supuesto no será objetivo de nuestro estudio el de tener en cuenta permisos de construcción, licencias...etc.

En principio no harían falta más datos para ofrecer las conclusiones que estamos buscando.

Objetivo del análisis

A partir del análisis de los datos definidos en el apartado anterior se estimará si determinados estados de EEUU tienen la posibilidad de construir más parques eólicos, mostrando además que porcentaje de energía extra producirían, haciendo además un estudio de cuánta energía nuclear podrían ahorrarse (reducir producción en centrales nucleares o incluso cerrarlas si la producción de la energía eólica renovable es suficientemente grande), y también si gracias a esa cantidad extra de energía que pueden llegar a producir podrían ahorrar dinero en cuanto a energía comprada a otros estados se refiere (esto pasa cuando 'Energía consumida > Energía producida'), por supuesto nuestro estudio no determinará la cantidad de dinero, simplemente mostrará un porcentaje significativo de dicho 'posible' beneficio.

Como idea adicional, se pretende desarrollar una aplicación con interfaz gráfica en C++, usando Qt, para representar el estudio mediante gráficas e indicadores significativos.

Variables

Las variables a analizar, junto con sus identificadores, son las siguientes:

- **x(c)** = Velocidad media del viento en la ciudad c
- **y(e)** = Producción anual de energía eólica en el estado e
- **z(e)** = Extensión de tierra en el estado e
- **w(e)** = Porcentaje de terreno público en el estado e
- **i(e)** = Producción anual de energía nuclear en el estado e
- **j(e)** = Consumo energético en el estado e

Constantes:

- **v_opt** = Velocidad óptima del viento para la producción de energía por parte de un aerogenerador. Valor = 14 metros por segundo
- **e_pe** = Extensión media de un parque eólico. Valor = 40 kilómetros cuadrados

Relaciones:

- **(t_disp)**: Kilómetros cuadrados de terreno donde se podría construir un parque eólico en el estado e: $z(e) - (z(e) * w(e))$
- Producción media de un parque eólico en condiciones óptimas (**p_med**):

$$\mu = \frac{Power}{\frac{1}{2} \rho U^3 \frac{\pi d^2}{4}}$$

Resultado utilizando 'v_opt' = 62,55 KW por hora por aerogenerador

A la hora de analizar los datos se debe comprobar de que $x(c) = v_{opt}$, con un rango de error de [-1,+1].

Estimación media de producción por parque eólico: (Cantidad de aerogeneradores posibles en el parque * p_med) = $4930 * 62,55 = 308371,5$ **KW por hora** = 0,308371 TWh (TeraWattios por hora). → **p_epe**

- **(c_pe)**: Posible cantidad de parques eólicos instalables (sin contar actualmente ya instalados) = (t_disp / e_pe)
- **(pos_prod)**: Posible producción de energía eólica: $(c_pe * p_epe)$ MWh → Para calcular la posible producción anual multiplicar este resultado por 8760 (24 * 365)
- Energía eólica extra posible (**energia_posible**): $(pos_prod - y(e))$
- Posible ahorro en energía nuclear: $(i(e) - energia_posible)$

Técnicas que se emplearán

Ya que el proyecto cuenta con grandes volúmenes de datos y lo que queremos es extraer información útil y presentarlo de forma que pueda ser entendida por un humano, y por lo tanto sacar provecho (ya que los datos por si solos no sirven para nada), estaríamos hablando de **Data Mining**.

Se generarán **árboles de decisión** a partir de los datos utilizando el algoritmo **ID3**.

Fuente de los datos

- Velocidad media del viento por ciudades: <https://wind-speed.weatherdb.com/>
- Área de cada estado: https://en.wikipedia.org/wiki/List_of_U.S._states_and_territories_by_area
- Porcentaje de territorio público sobre el total: <http://www.summitpost.org/public-and-private-land-percentages-by-us-states/186111>
- Producción de energía nuclear por estado: <http://www.eia.gov/nuclear/generation/index.html>
- Producción de energía eólica por estado: http://apps2.eere.energy.gov/wind/windexchange/wind_installed_capacity.asp
- Consumo energético por estados: <http://www.eia.gov/state/>

Dificultades

Una de las dificultades es la de pasar todos los datos a un formato homogéneo para trabajarlo todo de la misma manera y la otra dificultad será la de poder guardar tal volumen de datos y acceder a él de forma eficiente.

Aplicaciones previas

Una aplicación remarcable de un proyecto parecido a este fue el de “Vestas”, un fabricante de aerogeneradores danés, cuyo proceso de análisis de datos utilizado para crear modelos de ubicación de aerogeneradores para sus clientes resultaba cada vez menos convincente; el proceso tardaba varias semanas en ejecutarse y no podía analizar ingentes conjuntos de datos que la empresa consideraba necesarios para predecir con precisión la posible ubicación de un aerogenerador y su potencia. Los ingenieros de Vestas deseaban comenzar a desarrollar sus propias previsiones utilizando datos reales registrados de aerogeneradores de clientes en lugar de modelos preinstalados; según los planes, el reto era incrementar los requisitos de capacidad de datos a seis petabytes.

Gracias al uso de una solución de Big Data en un superordenador, uno de los más grandes del mundo hasta la fecha, y de una solución de modelado diseñada para aprovechar la información de un amplio conjunto de factores, entre los que se incluyen datos estructurados y no estructurados, ahora la empresa puede ayudar a sus clientes a optimizar la ubicación del aerogenerador y, como resultado de ello, su rendimiento.

Este nuevo entorno de información permite a la empresa gestionar y analizar datos de la ubicación y del tiempo de formas que antes no eran posibles a fin de obtener conocimientos que puedan llevar a una mejor toma de decisiones sobre la ubicación de los aerogeneradores y las operaciones, así como previsiones más exactas de la producción energética. Los modelos detallados significan una mayor certeza del caso de negocio, resultados más rápidos y una capacidad de previsión y una fiabilidad mejoradas, lo que disminuye los costes para los clientes por kilovatio/hora producido y aumenta la precisión de los cálculos de rentabilidad de la inversión (ROI) del cliente. Estas tecnologías reducen aproximadamente en un 97% (de semanas a horas) el tiempo de respuesta a las consultas planteadas por los usuarios empresariales y mejoran en gran medida la eficacia de la ubicación de los aerogeneradores.