

PROYECTO PREDICCIÓN + P.U. PARA VILLAVICENCIO META

Sergio Andrés Vargas Mendez, seavargasme@unal.edu.co

Carlos Andrés Méndez Beltrán, camendezb@unal.edu.co

Juan José Vargas Trujillo, jjvargast@unal.edu.co

Resumen—El objetivo de este trabajo consiste en la elaboración de un sistema de predicción para corto plazo(1 día para la generación y 1 semana para la demanda) enfocado en la región de Villavicencio-Meta, basado en modelos estocásticos, modelos de aprendizaje con redes neuronales, modelos de predicción estadísticos y arboles de decisión, todo esto con el objeto de estimar y evaluar lo acertado de la predicción(discretizado en intervalos horarios) y a su vez, elaborar metodologías para la predicción de demanda eléctrica, predicción de generación con plantas solares de 100 MW, teniendo en cuenta variables como la irradiancia, la temperatura, los ángulos de incidencia y demás factores que puedan variar los resultados de manera significativa y predicción de generación de un parque eólico de 100 MW, adicional a esto, se integran los modelos predictivos en una aplicación usando el lenguaje de programación PYTHON y como paquete principal, streamlit.

Index Terms—Modelos predictivos, demanda energética, generación energética, streamlit, energía solar, energía eólica, irradiancia, temperatura.

I. INTRODUCCIÓN

LA predicción es una herramienta fundamental e indispensable a la hora de realizar y planear diferentes eventos futuros para aplicaciones específicas. La modelización de la generación y de la demanda energética son muy importantes ya que permiten al proveedor contar con información más confiable y precisa referente a los mercados y a los usuarios o clientes del sistema interconectado o involucrado, así como una mejor posición y seguridad teniendo en cuenta que se reduce la incertidumbre para la toma de decisiones y que dichas decisiones no afecten nuestros

modelos.

Los modelos de predicción nos aportan diferentes horizontes relevantes dependiendo de cuales sean las decisiones a las cuales queramos enfocar nuestro esquema, sin embargo las capacidades de análisis actuales nos permiten llegar mucho más allá, siendo cada vez más importante, realizar moldeamientos de dichos sistemas en la actualidad, no solo a nivel local sino también extrapolado a nivel mundial.

En este trabajo se plantea la elaboración de un sistema de predicción usando estrategias y modelos asistidos por computadora, basándolo a su vez, en modelos horarios para la predicción diaria y semanal según el caso que corresponda.

En este documento se analizan y se registran los resultados obtenidos de un sistema de predicción implementado para la generación y la demanda energética en el municipio de Villavicencio. Este sistema predictivo es evaluado con base en su exactitud y desempeño, se tiene en cuenta el rendimiento y la velocidad de respuesta a la predicción.

II. MARCO TEÓRICO

A. Redes Neuronales

Las redes neuronales artificiales son un modelo inspirado en el funcionamiento del cerebro humano. Esta formado por un conjunto de nodos conocidos como neuronas artificiales que están conectadas y transmiten señales entre sí. Estas señales se transmiten

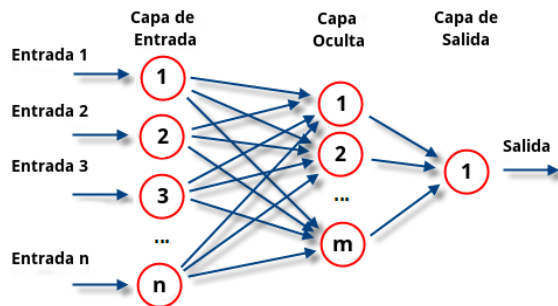


Figura 1: Esquema Red Neuronal

desde la entrada hasta generar una salida.

Dichas redes aprenden, usan regresiones lineales donde tenemos variables de entrada, se usa un algoritmo de aprendizaje denominado backpropagation que consiste en que la red puede aprender de sus errores, dentro de su proceso de entrenamientos, la red verifica cuanto se equivoca el modelo y trata de corregir o ajustar para obtener la salida más fiel a la realidad, también nos sirve para interpretación de imágenes y reconocimiento facial mediante caracterización por medio de la interconexión neuronal.

Pese a que en la actualidad se ha popularizado su uso, las redes neuronales existen desde la década de 1950. Sin embargo, la poca potencia de los equipos de aquella época y la inexistencia de algoritmos que permitiesen a las redes aprender de forma eficiente provocó que estas dejaran de usarse. Ha sido posteriormente, gracias a la creación del algoritmo de Backpropagation, al uso de GPUs que permiten realizar grandes optimizaciones para este tipo de cálculos y al mayor número de datos disponibles para entrenamiento, cuando las redes neuronales han vuelto a resurgir y a ganar protagonismo en diversos campos. Gracias a estas mejoras ha sido posible la aparición del Deep Learning, el cual se basa en el uso de redes neuronales profundas, es decir, redes formadas por un gran número de capas para tareas complejas.

El alcance de las funciones de las redes neuronales es muy amplio, debido a su

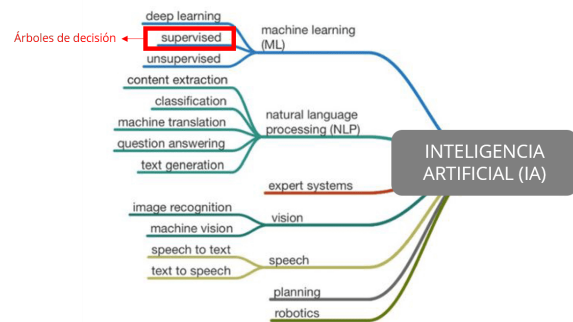


Figura 2: Diagrama de un árbol de decisión

funcionamiento, son capaces de aproximar cualquier función existente con el suficiente entrenamiento. Principalmente las redes neuronales son utilizadas para tareas de predicción y clasificación. Su rango de actuación es amplio y de gran utilidad hoy en día, no solo se utilizan para aplicaciones de Industria 4.0 (reconocimiento de piezas y defectos que no han sido introducidos previamente por ejemplo), si no que son utilizados en otras áreas como la economía, en la que pueden ayudar a predecir cuanto van a variar los precios a lo largo de los años, o incluso en medicina donde son de gran ayuda para diagnosticar diversos problemas de salud.

Las redes neuronales se han convertido en una pieza clave para el desarrollo de la Inteligencia Artificial, es uno de los principales campos de investigación y el que mas esta evolucionando con el tiempo, ofreciendo cada vez soluciones más complejas y eficientes.

III. MODELOS DE PREDICCIÓN

Árbol de decisión

Un árbol de decisión es un modelo predictivo que divide el espacio de los predictores agrupando observaciones con valores similares para la variable respuesta o dependiente.

Para dividir el espacio muestral en sub-regiones es preciso aplicar una serie de reglas o decisiones, para que cada sub-región contenga la mayor proporción posible de

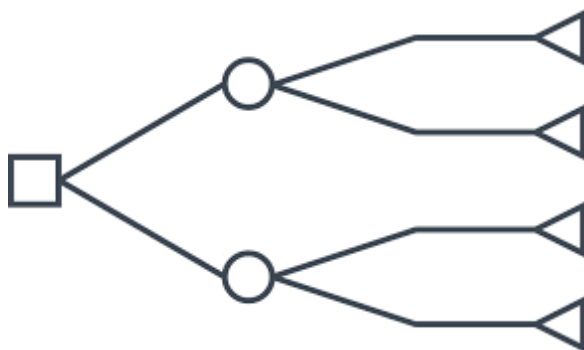


Figura 3: Estructura básica de un árbol de decisión

individuos de una de las poblaciones.

Si una sub-región contiene datos de diferentes clases, se subdivide en regiones más pequeñas hasta fragmentar el espacio en sub-regiones menores que integran datos de la misma clase.

Los algoritmos de aprendizaje automático se clasifican en dos tipos:

- Supervisados
- No supervisados

Un árbol de decisión es un algoritmo supervisado de aprendizaje automático porque para que aprenda el modelo necesitamos una variable dependiente en el conjunto de entrenamiento.

Ventajas y desventajas de los árboles de decisión Al hacer uso de esta herramienta surgen ventajas e inconvenientes.

Ventajas

- Son fáciles de construir, interpretar y visualizar. Selecciona las variables más importantes y en su creación no siempre se hace uso de todos los predictores.
- Si faltan datos no podremos recorrer el árbol hasta un nodo terminal, pero sí podemos hacer predicciones promediando las hojas del sub-árbol que alcancemos.
- No es preciso que se cumplan una serie de supuestos como en la regresión lineal (linealidad, normalidad de los residuos, homogeneidad de la varianza, etc.).
- Sirven tanto para variables dependientes cualitativas como cuantitativas, como para variables predictoras o in-

dependientes numéricas y categóricas. Además, no necesita variables dummies, aunque a veces mejoran el modelo.

- Permiten relaciones no lineales entre las variables explicativas y la variable dependiente. Nos podemos servir de ellos para categorizar variables numéricas.

Desventajas

- Tienden al sobreajuste u overfitting de los datos, por lo que el modelo al predecir nuevos casos no estima con el mismo índice de acierto.
- Se ven influenciadas por los outliers, creando árboles con ramas muy profundas que no predicen bien para nuevos casos. Se deben eliminar dichos outliers.
- No suelen ser muy eficientes con modelos de regresión.
- Crear árboles demasiado complejos puede conllevar que no se adapten bien a los nuevos datos. La complejidad resta capacidad de interpretación.
- Se pueden crear árboles sesgados si una de las clases es más numerosa que otra.
- Se pierde información cuando se utilizan para categorizar una variable numérica continua.

Random Forest

Un modelo Random Forest está formado por un conjunto de árboles de decisión individuales, cada uno entrenado con una muestra ligeramente distinta de los datos de entrenamiento generada mediante bootstrapping). La predicción de una nueva observación se obtiene agregando las predicciones de todos los árboles individuales que forman el modelo.

Muchos métodos predictivos generan modelos globales en los que la única ecuación se aplica a todo el espacio muestral. Los métodos estadísticos y de machine learning basados en árboles engloban a un conjunto de técnicas supervisadas no paramétricas que consiguen segmentar el espacio de los predictores en regiones simples, dentro de las cuales es más sencillo manejar las interacciones. Es

esta característica la que les proporciona gran parte de su potencial.

IV. PROGRAMAS

IV-A. *Conversión PU*

El programa de conversión Pu emplea diferentes métodos de entrada para determinar el valor en Pu de la impedancia de cada una de las líneas del circuito electo, permitiéndole al usuario modificar valores como las relaciones de transformación de los diferentes trafos, la potencia del sistema, la tensión de los segmentos y la impedancia de los diferentes componentes en 3 diferentes formatos: pu, ohm y porcentaje.

IV-B. *Aerogenerador*

El programa del aerogenerador, emplea datos extraídos de diferentes paginas oficiales para consecuentemente compararlos con la curva de potencia de un aerogenerador, calculando así la potencia total generada por el miso en el lapso de un día y de un año, te permite escoger el numero de aerogeneradores así como el modelo del mismo.

IV-C. *Predicción*

Te permite emplear diferentes modelos de predicción, para que a partir de datos extraídos de fuentes oficiales retro alimentar una inteligencia artificial de modelo "Árbol de decisión" para construir un modelo matemático capaz de predecir el comportamiento de una red de paneles solares con un lapso de hasta un mes, te permite elegir entre diferentes modelos de paneles así como una ciudad para elegir, aparte de esto, el programa te permite conocer el grado de error de forma gráfica al hacer una comparativa entre los datos reales con los predichos.

V. ANÁLISIS Y RESULTADOS

Dentro del desarrollo del programa y de su interfaz, usamos funciones para definir los estilos gráficos en css y funciones para llamar los archivos en Excel, también se usan funciones para convertir los archivos de petición en una serie temporal y archivos

para convertir la serie temporal horaria.

Se hizo un logotipo de la compañía cuyo nombre es Datafluids, como primera pantalla tenemos una breve descripción de ¿Quiénes somos? y el objetivo como tal de la compañía, esta parte es el inicio de la interfaz y la presentación de nuestra página.

Posteriormente usamos otro archivo css con colores del logotipo, donde encontramos una lista desplegable que nos muestra las predicciones de generación, demanda y un ítem donde podemos realizar las conversiones.

En el ítem predicciones se despliegan unos submenús de selección donde se encuentran opciones como Agentes, Explorador de datos, Predictor, Proyección solar, proyección eólica.

Según la sección que escojamos, así mismo vamos a obtener unas bases de datos y gráficas. Para el caso de los agentes tenemos una base de datos que comprende desde el 2015 hasta el 2021, nos carga los agentes que tenemos disponibles para poder realizar el modelo de predicción y la operación de los datos, solo los datos que cumplan ciertos criterios serán usados por el modelo.

Los resultados de agente que se seleccione muestran los datos de demanda por fecha, el dataframe se visualiza de modo completo, nos muestra el precio de bolsa nacional en una franja horaria y el volumen útil diario.

La herramienta usa datos base como las variables exógenas, tenemos una lista con todos los operadores de red que tenemos a nivel nacional, para el caso de Villavicencio-Meta se usa el agente correspondiente.

En la herramienta de predicción, se debe ingresar una fecha de inicio de predicción, el agente que vamos a predecir y la variable exógena que vamos a utilizar. Los tiempo que el modelo nos permite predecir están determinados en un día, una semana, dos semanas y hasta 1 mes después de la fecha

que escojamos (la unidad base son los días).

Para el caso de la exactitud del modelo, dependerá del tiempo de predicción y de la cantidad de cortes de suministro de energía que tenga dicho agente, para el caso de codensa la predicción de demanda sera sólida y concisa, pero para el caso del Meta tenemos al agente de red EMSA y dado que presenta más cortes de energía eléctrica, el modelo será menos preciso al momento de la predicción a largo plazo, para cortos plazos es más exacta la predicción.

EL programa toma la fecha que se defina como fecha inicial, el modelo escoje un año hacia atrás y el tiempo que escogí en el futuro, esto con el objetivo de tener suficientes datos, posterior a eso se carga el archivo en excel de los agentes e información exógena adicional.

El modelo siempre se entrena con un año hacia atrás de la fecha que se escoja para la predicción, la opción shuffle nos permite ver los resultados de manera organizada, también se usa de manera paralela la función del Random Forest Regressor, el cual nos permite automatizar la creación de los modelos para no tener que realizar ajustes posteriores.

Dentro de nuestro modelo también Se crean un vector de 0 a 23, para mostrar mejor los datos, se realizan unas gráficas por mes, lo primero que se necesita son las funciones encadenadas el la función postday pedimos la figura de la potencia en años como datos de entrada, el archivo de irradiación para Villavicencio, numero de paneles y la potencia instalada, todo esto para el cálculo del número de paneles. Nuestras funciones de supuestos nos agregan los datos del primer mes, el mes se valida para ver si el dato es el mínimo y para evitar errores la columna no puede dar cero, las horas de máxima irradiancia se determina el mínimo de potencia nominal para que en los meses siguientes no tengan problema.

En el id tenemos los datos de tensión y corriente los cuales se calculan con el operador de potencia que es una linealización de la curva que lo que hace es entrar potencia y nos devuelve irradiancia.

VI. CONCLUSIONES

- El pronóstico de la demanda es un insumo importante para la operación, planeación y expansión de los sistemas eléctricos de potencia; ya que puede ser un factor determinante para la toma de decisiones en el mercado eléctrico.
- Las características principales de la serie de demanda son la tendencia, estacionalidad, ciclos y la irregularidad, además dicha serie tiene una fuerte influencia de variables económicas y climáticas.
- El modelo de predicción por arboles de decisión plantea un sistema robusto que genera resultados bastante precisos y confiables en corto y mediano plazo.
- No hay un estudio sistemático y ordenado respecto a cuáles modelos son mejores. Para el caso colombiano, no hay un estudio consistente donde se indique cuál o cuáles modelos son los mejores de acuerdo con el pronóstico requerido. Más aún, existe una amplia variedad de modelos no lineales que aun no han sido estudiados.

REFERENCIAS

- [1] R. Ferrero. (2021, Feb 24). Blog-Dat [Online]. Available: <https://www.maximaformacion.es/>
- [2] Zaragoza. (2019, Oct 22). Industria 4.0, Inteligencia Artificial [Online]. Available: <https://www.atrainnovation.com/>
- [3] J. Amat. (2020, Oct). Random Forest con Python [Online]. Available: <https://www.cienciadedatos.net/>