**Universidad Autónoma de Tamaulipas**

Registro de obra ante el INSTITUTO NACIONAL DEL DERECHO DE AUTOR (INDAUTOR)

* Título de la obra: Aplicación para la predicción de energía eólica en la región Istmo, Oaxaca
* **Breve** descripción de la obra (Máximo 5 líneas):

Se presenta una aplicación en Python para predecir la energía generada (MWh) en las plantas eólicas Eurus, Bii-Hioxio y Eoliatec, de la región Istmo, Oaxaca. De acuerdo con el día y mes solicitado, se predicen hasta 24 horas de valores de energía, utilizando registros meteorológicos y tres modelos de regresión a elegir: lineal (LR), vectores de soporte (SVR) y red neuronal profunda (LSTM).

* Datos del autor o coautores:

Autor 1:

* + Nombre completo: Ángel Alonso Galarza Chávez
  + Fecha de nacimiento: 31 de Julio del 2000
  + lugar de nacimiento: Reynosa, Tamaulipas
  + Nacionalidad: mexicana
  + **Porcentaje** de participación de los coautores en la obra: 40 %
  + **RFC** del autor o coautores: GACA0007315N0
  + Correo electrónico: angel\_alonso1@outlook.es
  + Teléfonos: 899 465 2919
  + Domicilio particular (calle, número interior y/o exterior, colonia, Delegación/Municipio, País, entidad federativa y C.P.): Emiliano Zapata #09, col. Rancho Grande, Reynosa, Tamaulipas, CP. 88610
  + Lugar de adscripción en la UAT: Unidad Académica Multidisciplinaria Reynosa - Rodhe
  + Grado académico: Ingeniería

Autor 2:

* + Nombre completo: José Lázaro Martínez Rodríguez
  + Fecha de nacimiento: 22 de Julio de 1987
  + lugar de nacimiento: Cd. Victoria, Tamaulipas
  + Nacionalidad: mexicana
  + **Porcentaje** de participación de los coautores en la obra: 20 %
  + **RFC** del autor o coautores: MARL8707228w8
  + Correo electrónico: lazaro.martinez@uat.edu.mx
  + Teléfonos: 8341561145
  + Domicilio particular (calle, número interior y/o exterior, colonia, Delegación/Municipio, País, entidad federativa y C.P.): Calle Avetorillo #6310, Fraccionamiento Paseo de los Cisnes, Cd. Victoria, Tamaulipas, México, C.P. 87084
  + Lugar de adscripción en la UAT: Facultad de Ingeniería y Ciencias
  + Grado académico: Doctorado

Autor 3:

* + Nombre completo: René Fernando Domínguez Cruz
  + Fecha de nacimiento: 27 de abril de 1972
  + lugar de nacimiento: Perote, Veracruz
  + Nacionalidad: mexicana
  + **Porcentaje** de participación de los coautores en la obra: 15 %
  + **RFC** del autor o coautores: DOCR720427LL6
  + Correo electrónico: rfdominguez@docentes.uat.edu.mx
  + Teléfonos: (899) 160 2068. Oficina: (899) 921 3300 Ext. 8114
  + Domicilio particular (calle, número interior y/o exterior, colonia, Delegación/Municipio, País, entidad federativa y C.P.): Privada Amatista No. 574. Fracc. Vista Hermosa. Cd. Reynosa, Tamaulipas, México. CP. 88710.
  + Lugar de adscripción en la UAT: Unidad Académica Multidisciplinaria Reynosa - Rodhe
  + Grado académico: Doctorado

Autor 4:

* + Nombre completo: Esmeralda López Garza
  + Fecha de nacimiento: 9 de agosto de 1987
  + lugar de nacimiento: Reynosa, Tamaulipas
  + Nacionalidad: mexicana
  + **Porcentaje** de participación de los coautores en la obra: 15 %
  + **RFC** del autor o coautores: LOGE870809MMA
  + Correo electrónico: elgarza@docentes.uat.edu.mx
  + Teléfonos: (899) 3082825
  + Domicilio particular (calle, número interior y/o exterior, colonia, Delegación/Municipio, País, entidad federativa y C.P.): Calle Berlín #814 Col. Cañada, Cd. Reynosa Tamaulipas, México. C.P. 88700
  + Lugar de adscripción en la UAT: Unidad Académica Multidisciplinaria Reynosa - Rodhe
  + Grado académico: Doctorado

Autor 5:

* + Nombre completo: Ana Bertha Ríos Alvarado
  + Fecha de nacimiento: 24 de marzo de 1984
  + lugar de nacimiento: Naucalpan de Juárez, Estado de México
  + Nacionalidad: mexicana
  + **Porcentaje** de participación de los coautores en la obra: 10 %
  + **RFC** del autor o coautores: RIAA840324IX7
  + Correo electrónico: arios@docentes.uat.edu.mx
  + Teléfonos: 8341096881
  + Domicilio particular (calle, número interior y/o exterior, colonia, Delegación/Municipio, País, entidad federativa y C.P.): Calle Avetorillo #6310, Fraccionamiento Paseo de los Cisnes, Cd. Victoria, Tamaulipas, México, C.P. 87084
  + Lugar de adscripción en la UAT: Facultad de Ingeniería y Ciencias
  + Grado académico: Doctorado
* Se ha dado a conocer la obra, en caso afirmativo, **indicar la fecha: No se ha dado a conocer**

* Especificar si se trata de una obra primigenia o derivada. En caso de ser derivada, especificar de qué tipo (ampliación, traducción, arreglo, compendio, adaptación, paráfrasis, compilación, transformación o colección) y *mencionar el* *Título y Nombre del autor de la obra original*: **Obra primigenia**
* **INCLUIR :** registros, videos, documentos, y todos aquellos elementos utilizados y de **evidencia que justifican la creación primigenia de la obra** que en este acto transmiten: Se incluye Anexo 1 – **Descripción de la Propuesta**
* **Copia** credencial de elector ***vigente*** del autor o de los coautores. Se incluyen copias digitales junto a este documento.
* **Copia** de Credencial de la **UAT** (en caso de ser empleado de esta institución). Se incluyen copias digitales.
* Contrato de Transmisión de Derechos—(éste es elaborado por personal de la Coordinación de Transferencia del Conocimiento)
* **Remitir 1 ejemplar de la Obra en formato .Zip**
* **Para SOFTWARE:** Código Fuente completo o el Programa Ejecutable completo; impreso o en disco compacto.

Enviar la obra en formato .Zip y que éste no rebase los **300 MB**

Anexo 1

Descripción de la propuesta

# Metodología y propuesta

Se presenta una aplicación en Python para predecir la energía generada (MWh) en las plantas eólicas Eurus, Bii-Hioxio y Eoliatec, de la región Istmo, Oaxaca. De acuerdo con el día y mes solicitado, se predicen hasta 24 horas de valores de energía, utilizando registros meteorológicos y tres modelos de regresión a elegir: lineal (LR), vectores de soporte (SVR) y red neuronal profunda (LSTM). Para el desarrollo de la aplicación se propuso una metodología que se compone de varias etapas para el análisis y exploración de datos, preprocesamiento de datos, modelado, evaluación y pronóstico. La metodología de desarrollo se presenta en la Figura 1.

Imagen que contiene Forma

Descripción generada automáticamente

Figura 1 Metodología para el desarrollo de la aplicación

## Obtención de datos

El objetivo de esta etapa es la recolección de datos que se utilizarán para llevar a cabo el desarrollo del modelo. En este caso se utilizan datos que representen las series de tiempo, dichos datos se espera que sean capturados por un dispositivo de medición, el cual consiste en sensores que capturan las variables meteorológicas como la velocidad de viento en intervalos de una hora. Generalmente estos conjuntos de datos vienen en tres formatos diferentes: Valores separados por coma (CSV, por sus siglas en inglés); Bases de datos, las cuales contienen registros dentro de tablas y archivos de Excel. Los datos con los cuales se llevaron a cabo los distintos experimentos corresponden al conjunto de datos de Daily weather data from Nizanda, México (2006-2022) (Muñoz, y otros, 2023). Dicho conjunto de datos cuenta con un total de 56736 observaciones y contiene 26 características, donde las más relevantes son la velocidad del viento y el tiempo de medición de la observación, mismas que se usaron como entrada para la construcción del modelo de predicción.

## Análisis de datos

Dentro de esta etapa se considera un proceso de selección de características, es decir, conservar las características que aporten el mejor rendimiento para la construcción de un modelo de predicción. Para ello se debe de realizar un análisis del conjunto de datos, con el propósito de obtener más información acerca de la estructura de los datos, tipos de características, distribuciones, falta de valores o errores de mediciones. Se requiere de observación por medios visuales para identificar las características que presentan los datos.

## Preprocesamiento

Una vez que se han recolectado los datos a utilizar, es importante que, antes de llevarlos directamente al modelado, se requiere ejecutar algunas tareas de preprocesamiento y análisis. El propósito de esta etapa es preparar los datos para que, al momento de utilizarlos para la construcción del modelo de predicción, estos estén en un formato adecuado. Esto debido a que los conjuntos de datos generalmente vienen con valores nulos, tipos de datos no deseados, falta de integridad de los datos, etc.

## Modelado

Una vez que se haya obtenido el conjunto de datos analizado y preprocesado, se puede llevar a cabo el desarrollo del modelo de predicción de generación de energía. En esta etapa se realizan ciertas tareas para obtener un modelo para dicha predicción, donde se hace una separación del conjunto de datos. Una parte del conjunto de datos corresponde para el entrenamiento del modelo y la segunda parte es para su evaluación. El entrenamiento se realiza considerando diversos algoritmos de aprendizaje máquina, los cuales son evaluados para ajustar distintos hiper-parámetros que permitan obtener el mejor resultado de predicción

## Evaluación

Con el modelo ya entrenado, debemos de evaluar su comportamiento con el conjunto de datos de evaluación, además de realizar una serie de simulaciones para verificar que el desempeño del modelo de predicciones de energía es el adecuado. De acuerdo con el problema de regresión se utilizan métricas para estimar el rendimiento del modelo de predicción de energía generada tales como la raíz del error cuadrático medio (RMSE), error cuadrático medio (MAE) y r-cuadrado (R²). En caso contrario, cuando el modelo no obtiene un buen desempeño, se regresa a la etapa de modelado, donde se cambiarán los ajustes de los parámetros, este paso se repetirá hasta encontrar la configuración ideal para el modelo.

## Pronóstico

Como última etapa se cuenta con el pronóstico, con el objetivo de una vez desarrollado el modelo, un usuario ingrese una secuencia de valores de entrada y el modelo realice una secuencia de valores que representa el pronóstico y por medio de una función de conversión de velocidad del viento a energía generada despliegue los resultados. Así mismo se presentará una gráfica de serie de tiempo donde se pueda visualizar la varianza de la generación de energía, de este modo, el usuario obtenga la información del pronóstico y en base a esa información, realice la toma de decisiones.

# Descripción de la aplicación

La aplicación se compone de diversas herramientas que fueron integradas gracias a la aplicación de la metodología antes presentada. En este apartado se presentan las herramientas que se requieren para ejecutar la aplicación asi como también las opciones disponibles para su configuración.

## Herramientas utilizadas

Las herramientas utilizadas para la construcción de la aplicación se muestran en la Tabla 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Herramienta | Versión | Sitio web | Licencia | Descripción |
| Python | 3.10 | <https://www.python.org/> | MIT | Lenguaje de programación interpretado de alto nivel |
| Jupyter Lab | 3.5.3 | <https://jupyter.org/> | CCO | Entorno de desarrollo basado en web |
| Pandas | 1.5.2 | <https://pandas.pydata.org/> | BSD | Biblioteca de Python para análisis de datos |
| Numpy | 1.23.5 | <https://numpy.org/> | BSD 3 | Biblioteca de Python para creación de arreglos y funciones matemáticas |
| Sckit-Learn | 1.2.2 | <https://scikit-learn.org/stable/> | BSD 3 | Biblioteca de Python con conjunto de herramientas para aprendizaje maquina y minería de datos |
| Matplotlib | 3.6.2 | <https://matplotlib.org/> | BSD | Biblioteca de Python para crear figuras de calidad |
| Seasborn | 0.12.2 | <https://seaborn.pydata.org/> | BSD 3 | Biblioteca de Python para crear graficas estadísticas |
| Statsmodels | 0.14.0 | <https://www.statsmodels.org/stable/index.html> | BSD 3 | Biblioteca de evaluación y pruebas estadísticas |
| Neupy | 0.8.2 | <http://neupy.com/docs/tutorials.html> | MIT | Biblioteca que provee algoritmos de aprendizaje profundo |
| TensorFlow | 2.10 | <https://www.tensorflow.org/> | Apache-2.0 | Interfaz para crear, entrenar y evaluar redes neuronales |

Tabla 1. Descripción de las herramientas computacionales utilizadas.

## Configuración

Para ejecutar la aplicación se presentan algunos puntos por configurar, como por ejemplo las rutas de los datos, selección de fecha por predecir, asi como también el modelo de regresión a utilizar.

### Configuración de datos

La predicción requiere datos históricos como ejemplo para capturar datos futuros. Dichos datos se adjuntan a la aplicación y se debe indicar la ruta en que se encuentran. La Figura 2 presenta el código donde se debe indicar la ruta de los datos para obtener las observaciones

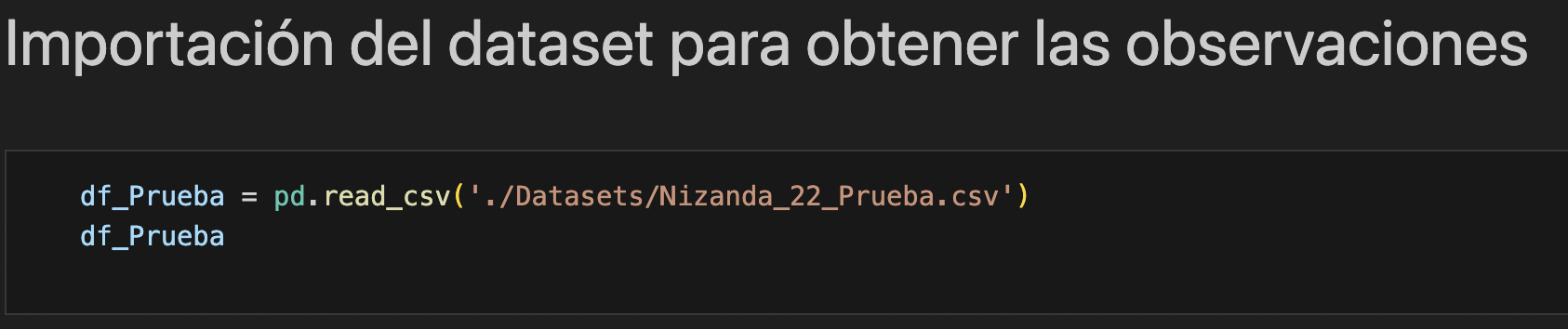


Figura 2 Datos para observaciones

### Ruta de modelos

De igual forma se debe configurar la ruta donde se encuentra cada modelo de regresión y modelo de escalado de datos (para normalizar la entrada). Por ejemplo, para la configuración de M-SVR se presenta la Figura 3 donde se debe indicar cada ruta indicada. Esta solicitud de rutas se presenta para cada uno de los tres modelos considerados para la aplicación (SRV, LR, LSTM). Cabe decir que solo se debe ejecutar el bloque del modelo que se quiere utilizar.



Figura 3 Configuración de modelo y escalador de datos para M-SVR

### Día solicitado

Se presentan algunos fragmentos de ejemplo para utilizar la aplicación. En la Figura 4 se presenta el código para solicitar las predicciones de energía para el día 15 de octubre. Cabe mencionar que se indican líneas de código adicionales que sirven de utilidad para calcular el margen de error que presentan las estimaciones.

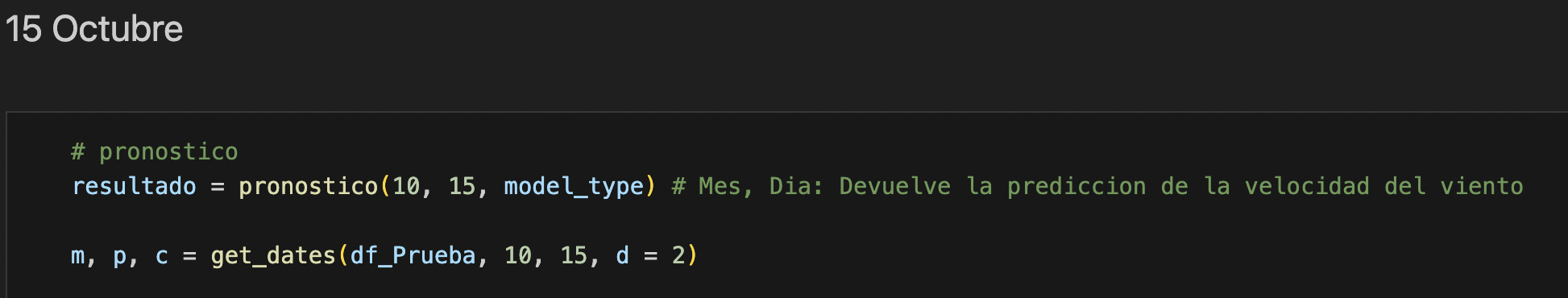


Figura 4 Fragmento de código para solicitar predicciones de un día y mes

# Evaluación y Resultados

Como parte de la metodología para desarrollar la aplicación, se tuvieron que realizar varios experimentos con la finalidad de configurar parámetros y ajustes que permitieran seleccionar el mejor modelo de regresión. Para ello se utilizaron principalmente dos métricas: RMSE y MAE (Spüler, Sarasola-Sanz, Birbaumer, Rosenstiel, & Ramos-Murguialday).

### Evaluación por días del año

Se realizaron predicciones de todos los días del año, las cuales fueron comparadas con respecto a valores reales para obtener una diferencia de error. El objetivo de este experimento es evaluar el rendimiento de los distintos modelos de regresión. La Figura 5 muestra el resultado de la evaluación de los tres modelos.

Gráfico

Descripción generada automáticamente

Figura 5 Comparativa de modelos de predicción utilizando la métrica RMSE

Asimismo se presentan los resultados utilizando la métrica MAE tal como se muestra en la Figura 6.

Gráfico

Descripción generada automáticamente con confianza media

Figura 6 Comparativa de modelos de predicción utilizando la métrica MAE

De acuerdo con la evaluación se obtuvieron resultados desde 0.25 a 0.05 o menor para los modelos regresión lineal, SVR y LSTM. De estos resultados se destacan los resultados de SVR donde se mantiene mas consistente en las distintas pruebas con respecto a la mediana. Sin embargo, los modelos demostraron un comportamiento hasta cierto punto similar, por lo que se decidió agregar los tres para que el usuario seleccione el que mas le convenga.

### Resultados de energía

Para los resultados de las predicciones se ofrecen tres aspectos: la distribución del viento con respecto a valores reales, la distribución de energía generada por cada planta eólica, y los resultados de energía generada.

La Figura 7 presenta una comparativa de la predicción de viento con respecto a los valores reales. La idea es visualizar cómo se comporta el modelo de regresión para determinar qué velocidad del viento se tendría en un momento determinado. Cabe decir que este análisis se realiza con los datos que se tienen de registros pasados por lo que es solo un indicador del rendimiento como referencia.

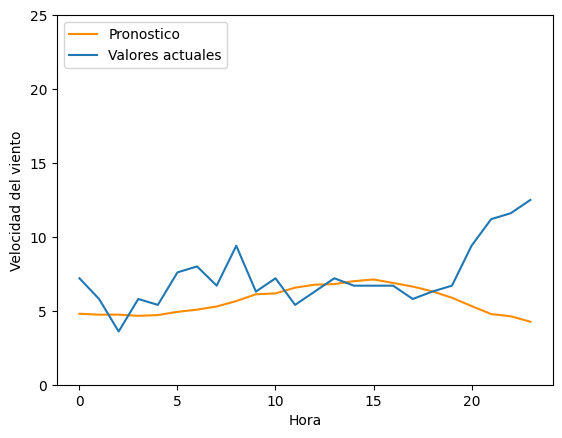


Figura 7 Distribución del viento con respecto a valores reales

Por otro lado, con dichos valores se realiza una gráfica que demuestra la energía generada en MWh para cada uno de los tres parques eólicos en un lapso de 24 horas, tal como se muestra en la Figura 8.

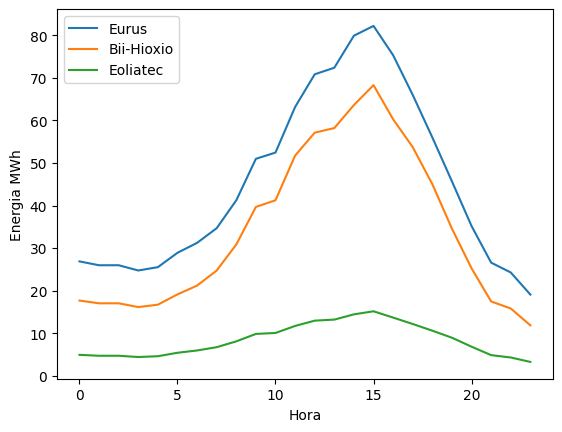


Figura 8 Distribución de energía generada por cada planta eólica

Finalmente, los resultados de las predicciones por día y por planta eólica son exportados a una hoja de cálculo para su posterior análisis o utilización. Un fragmento de las 10 primeras horas del resultado para el 15 de octubre, que se acompaña por un estimado del error de medición, se puede visualizar en la Figura 9.

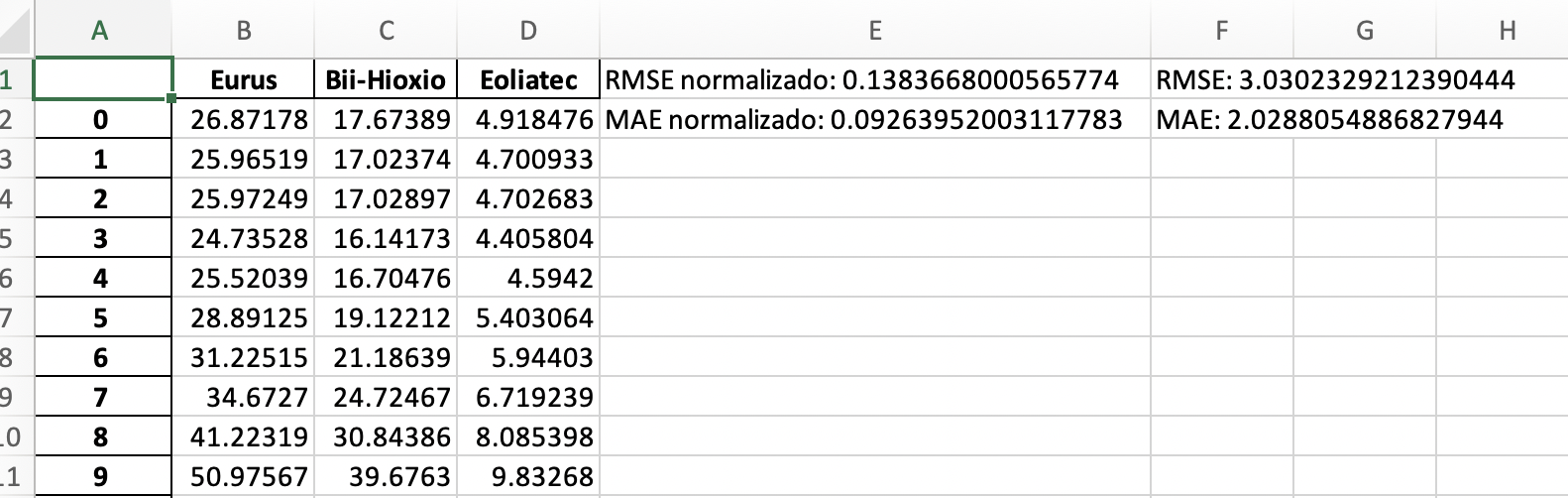


Figura 9 Resultados en hoja de cálculo

# Conclusión

Este documento presenta detalles de la metodología y configuración requeridos para el desarrollo de la aplicación Python para predecir la energía generada (MWh) en las plantas eólicas Eurus, Bii-Hioxio y Eoliatec, de la región Istmo, Oaxaca. De acuerdo con el día y mes solicitado, se predicen hasta 24 horas de valores de energía, utilizando registros meteorológicos y tres modelos de regresión a elegir: lineal (LR), vectores de soporte (SVR) y red neuronal profunda (LSTM).

# Bibliografía

Muñoz, R. a.-T.-C.-R. (Marzo de 2023). *Zenodo.* doi:10.5281/zenodo.7718709

Muñoz, R., Bongers, F., Lebrija-Trejos, E., Gallardo-Cruz, J. A., Enríquez, M., Romero-Romero, M. A., & Meave, J. A. (2023). *Daily weather data from Nizanda, Mexico (2006-2022)*. Obtenido de Zenodo: https://doi.org/10.5281/zenodo.7718709

Spüler, M., Sarasola-Sanz, A., Birbaumer, N., Rosenstiel, W., & Ramos-Murguialday, A. (s.f.). Comparing metrics to evaluate performance of regression methods for decoding of neural signals. *In 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (págs. 1083--1086). IEE.