

Informe Trabajo práctico No. 4 Programación 2024-1

David Aristizabal Giraldo – Ingeniería mecatrónica

Pedro Gómez – Ingeniería mecatrónica

Es importante antes de iniciar con la explicación de la metodología del proyecto plantear el ¿por qué es importante hacer este proyecto? Para un ingeniero mecatrónico, la realización de proyectos de analítica de datos en el ámbito de las energías renovables no solo mejora la eficiencia y la efectividad de los sistemas, sino que también tiene un impacto positivo en la economía, la sociedad y el medio ambiente. Estos proyectos permiten a los ingenieros mecatrónicos desempeñar un papel fundamental en la transición hacia un futuro más sostenible y tecnológicamente avanzado, integrando conocimientos multidisciplinarios para desarrollar soluciones innovadoras y eficientes.

Inicialmente es del caso recalcar que en medio de la búsqueda de la base de datos nos encontramos con bastantes obstáculos, siendo esta la parte del proyecto con la cual nos sentimos mas retados ya que fue un poco complicado encontrar algo que fuera realmente de nuestro interés y que adicional a ello cumpliera con los parámetros asignados por el docente, a su vez en la elaboración del código a medida que íbamos avanzando presentamos inconvenientes con la base de datos después de ser escogida, ya que en algunas ocasiones se omitían las columnas o simplemente nos arrojaba un error el cual no nos permitía visualizar la función que necesitábamos, todos estos problemas fueron resueltos con el conocimiento adquirido y con ayuda de las páginas de las librerías que utilizamos.

Metodológicamente hablando usamos el paso a paso que veníamos trabajando anteriormente en el cual inicialmente se cargaba un dataset el cual también debía ser muy bien inspeccionado, seguido a ello también debíamos hacer un proceso de “depuración”, en el cual estábamos en la obligación de hacer una serie de operaciones d empieza de datos que nos iban ayudar a tener un adelanto en el procesamiento de los datos que realmente si necesitábamos e íbamos a utilizar. Seguido a lo anterior debíamos de plantear dos soluciones mediante funciones o evaluaciones de curvas de las cuales hablare mas adelante ya que con base a estas se hizo una comparación

dentro de la cual utilizamos un análisis de concordancia para que mediante los modelos anteriormente implementados pudiéramos visualizar datos útiles y tangibles.

Mediante la utilización de los métodos anteriormente mencionados Utilizando el modelo random forest, los valores en las gráficas son bastante similar en él y (score), e incluso llegan a ser constantes en el caso del training score. Además, el training score y el cross validation score tienen valores bastantes lejanos los unos de los otros. En el modelo SVM, las gráficas son más volátiles, aunque estas llegan a tener más cercanía que el Random Forest, en los valores del eje y (score) en algunos puntos, utilizando la matriz de confusión para el SVM, muchos valores de la matriz fueron 0.

Finalmente decidimos sacar un listado de conclusiones enumerándolo convenientemente de acuerdo con los resultados anteriormente mencionados.

1. Diferencias en el Comportamiento de los Modelos:

- El modelo Random Forest mostró un ajuste excesivo a los datos de entrenamiento (overfitting), lo cual puede limitar su capacidad para generalizar en datos nuevos.
- El modelo SVM, aunque es más volátil, mostró una mejor proximidad entre los valores de score en ciertos puntos, sugiriendo una mejor capacidad de generalización.

2. Limitaciones y Áreas de Mejora:

- La discrepancia entre el training score y el cross-validation score en el modelo Random Forest indica la necesidad de técnicas adicionales para mitigar el overfitting, como la reducción de la complejidad del modelo o el aumento de la cantidad de datos de entrenamiento.
- La presencia de muchos valores de 0 en la matriz de confusión del SVM sugiere que es necesario ajustar los parámetros del modelo o explorar otros métodos de preprocesamiento de datos para mejorar la clasificación.

3. Implicaciones para el Campo de la Energía Renovable:

- La capacidad de los modelos para predecir y analizar la producción y consumo de energía, así como otros factores críticos, puede ser mejorada mediante la optimización de los modelos y la inclusión de más datos relevantes.
- Los hallazgos de este análisis pueden guiar futuras investigaciones y aplicaciones en el ámbito de las energías renovables, permitiendo un mejor diseño y operación de sistemas de energía sostenible.

Este proyecto destacó la importancia de seleccionar y ajustar adecuadamente los modelos de aprendizaje automático para el análisis de datos complejos en el campo de las energías renovables. Aunque ambos modelos presentaron desafíos específicos, sus análisis proporcionan información valiosa para mejorar la eficiencia y sostenibilidad de los sistemas de energía renovable. La continua optimización y validación de estos modelos es esencial para aprovechar plenamente el potencial de la analítica de datos en la promoción de un futuro energético más limpio y eficiente.