**TITULO PROYECTO: Diseño de indicador de eficiencia proyectada en paneles solares.**

**INTEGRANTES:**

**Juan Andres Sanchez Osorio**

**Cristian Perez Arango**

1. **ENTENDIMIENTO DEL NEGOCIO (Experto del Negocio)**
   1. DESCRIPCIÓN DEL NEGOCIO

* El proyecto Ecovilla – Smart Energy Center que surge de la necesidad de garantizar la sostenibilidad energética, ambiental y alimentaria de las ciudades se centra en desarrollar soluciones integrales a estos problemas. Mas información sobre el proyecto [aquí.](https://investigacion.upb.edu.co/es/projects/ecovilla-smart-energy-center-sec-construcci%C3%B3n-de-la-ecovilla-en-e)
  1. DESCRIPCIÓN DEL PROBLEMA
* El Laboratorio Nacional de Comunidades Energéticas (LNCE) se constituye como un espacio de innovación, análisis y transferencia de conocimiento para el impulso de soluciones energéticas sostenibles en Colombia. En el marco del desarrollo de la plataforma de monitoreo, este informe tiene como propósito describir la analítica de datos que soporta la visualización y toma de decisiones en comunidades energéticas, definiendo la bodega de datos, el ecosistema de información, los indicadores y los dashboards.

En este proyecto se busca proporcionar el indicador de **eficiencia proyectada** en energía solar para el LNCE, el cual tiene como objetivo estimar el rendimiento esperado bajo condiciones futuras del recurso, en nuestro caso, paneles solares.

* 1. OBJETIVOS DE LA MINERÍA
* Desarrollar un modelo predictivo tipo **regresión** usando redes neuronales (Perceptron, Backpropagation) para predecir la cantidad de energía que produce un panel solar en un periodo de tiempo bajo unas condiciones meteorológicas.
* **IMPORTANTE definir si este punto lo vamos a ejecutar:** Desarrollar un modelo predictivo tipo **regresión no lineal** usando SVR o Gradient Boosting **(Definir cual o ambos)**
  1. DISEÑO DE SOLUCIÓN

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Problema | Tipo de Minería | Tipo de aprendizaje | Requerimiento datos | Métodos | Evaluación |
|  |  | Clustering |  |  |  |
|  |  | Clasificación | \* Histórico  \* Variable obj  \* Relación entre predictoras y la objetivo | RN  Arbol  RF  Xgboost | Matriz de conf  P  R  ROC |

* Evaluación esperada

Línea base P=60%

* 1. RECURSOS PARA CREACIÓN DEL MODELO Y PARA **DESPLIEGUE**
* Hw
* Sw

1. **ENTENDIMIENTO DE LOS DATOS (Datos específicos del problema) (Experto TI)**
   1. CICLO DE LOS DATOS: Generación, Almacenamiento, Modificación (ruta), Periodicidad
   2. DICCIONARIO DE DATOS

|  |  |  |
| --- | --- | --- |
| Variable | Descripción | Tipo |
|  |  |  |

* 1. REGLAS DE CALIDAD DESDE EL NEGOCIO (No salen de los datos)

|  |  |
| --- | --- |
| Variable | Regla calidad (valores válidos) |
| Edad | 18-80 |
| Sexo | F,M |

1. **PREPARACIÓN DE DATOS**
   1. INTEGRACIÓN
   2. SELECCIÓN DE VARIABLES
   3. ESTADÍSTICA DESCRIPTIVA
   4. LIMPIEZA DE ATÍPICOS
   5. LIMPIEZA DE NULOS
   6. ANÁLISIS DE CORRELACIONES PARA REDUNDANCIA
   7. ANÁLISIS DE CORRELACIONES PARA IRRELEVANCIA (PREDICCIONES)
   8. BALANCEO (CLASIFICACIÓN)
   9. INGENIERÍA DE CARACTERÍSTICAS
      1. CREACIÓN DE NUEVAS VARIABLES
      2. REDUCCIÓN DE DIMENSIÓN (OPCIONAL EN PREDICCIONES)
      3. TRANSFORMACIONES
2. **MODELAMIENTO, EVALUACIÓN E INTERPRETACIÓN**
   1. CONFIGURACIÓN MÉTODOS DE MACHINE LEARNING
   2. ANALISIS DE MEDIDAS DE CALIDAD
   3. SELECCIÓN DEL MEJOR MODELO

* Comparación de calidad mediante pruebas estadística ANOVA, Tukey
* Tiempo computacional de creación y despliegue

1. **DESPLIEGUE**
   1. PREDICCIÓN DE DATOS FUTUROS: almacenar modelo, pipes para el despliegue, servicio web de desliegue
   2. MONITOREO
   3. CRONOGRAMA DE MANTENIMIENTO/RE-ENTRENAMIENTO