**Identificar los Hogares Vulnerables en La Guajira para Priorizar el Acceso a Energías Limpias mediante IA**

**Jaime Andrés Londoño**

**Geovanny Vergara**

**Claudia Cárdenas**

**Claudia Lorena Ramírez**

**Jheyson Eduardo Galvis Valencia - Ejecutor técnico**

**Jose Blanquicett - Mentor**

**Carlos Alberto Riascos Moreno – Monitor**

**Alexa Johana Marin – Ejecutor de Ingles**

**Judi Alejandra Salazar Ayala – Ejecutor de Proyecto**

**Mintic- U de Antioquia – Universidad de Caldas- Ubicua**

**Proyecto Bootcamp- Inteligencia Artificial – Nivel Exploratorio**

**Medellín**

**9 de diciembre 2024**

**Pregunta de Investigación Ajustada**

¿Cómo puede un modelo de inteligencia artificial basado en clasificación supervisada identificar hogares vulnerables en La Guajira para priorizar su acceso a soluciones de energía limpia y contribuir a la reducción de la pobreza energética en la región?

**1. Título del Proyecto**

**Identificar los Hogares Vulnerables en La Guajira para Priorizar el Acceso a Energías Limpias mediante IA**

**2. Resumen**

Este proyecto tiene como objetivo desarrollar un modelo de inteligencia artificial (IA) supervisado para clasificar hogares vulnerables en La Guajira y optimizar su acceso a soluciones energéticas limpias, específicamente solar y eólica. A través del procesamiento de datos socioeconómicos, energéticos y ambientales, obtenidos de fuentes públicas como Open Data Colombia y XM, el modelo empleará el algoritmo Random Forest para priorizar eficientemente la asignación de recursos energéticos. La implementación culminará con un programa interactivo en Python, desarrollado con el framework FastAPI, que incluirá un formulario para facilitar la entrada de datos. Este sistema determinará la elegibilidad de cada hogar para recibir energía renovable y especificará el tipo más apropiado según sus características particulares.

**3. Introducción**

La Guajira, ubicada en el extremo noreste de Colombia, abarca una extensión de 20,848 km² y alberga aproximadamente 1,015,909 habitantes. En términos económicos, ocupa el puesto 21 a nivel nacional, con un PIB nominal en 2023 de USD 13,182 millones y un PIB per cápita de USD 12,976. Su diversidad climatológica, caracterizada por temperaturas que oscilan entre 22 y 30 °C en promedio, con extremos desde 3 °C en zonas montañosas hasta 42 °C en las bajas, la convierte en un territorio con excepcional potencial para la generación de energías renovables.

A pesar de estas condiciones favorables, numerosas comunidades guajiras enfrentan una marcada pobreza energética. La ausencia de suministro eléctrico constante limita significativamente su desarrollo, profundizando la desigualdad social y económica. Esta realidad subraya la necesidad imperante de implementar soluciones sostenibles que aprovechen las capacidades naturales de la región para generar energías limpias y garantizar su acceso a los hogares más vulnerables.

El proyecto propone abordar esta problemática mediante un modelo de inteligencia artificial que identificará y clasificará hogares con mayor vulnerabilidad energética, facilitando que los tomadores de decisiones accedan a recomendaciones específicas y contribuyan a una transición energética equitativa.

**4. Objetivos:**

**General:** Clasificar los hogares vulnerables en La Guajira para optimar y priorizar su acceso a energías limpias mediante Inteligencia Artificial.

**Específicos:**

* + Entrenar y validar un modelo de Inteligencia Artificial basado en el algoritmo de clasificación Random Forest. Este modelo procesará los datos ingresados en la aplicación y proporcionará una sugerencia de priorización basada en los resultados obtenidos.
  + Presentar en un framework de Fastapi con un formulario interactivo con el usuario.

**5. Marco Teórico**

Para comprender cómo identificar y priorizar hogares vulnerables que necesitan acceso a energía limpia en La Guajira, necesitamos entender cuatro conceptos fundamentales que se entrelazan en esta investigación.

**Transición Energética Justa**

La transición energética justa constituye el marco más amplio de nuestra investigación. Este concepto va más allá del simple cambio de fuentes de energía contaminantes a limpias. Representa una transformación que debe beneficiar primero a quienes más lo necesitan.

En el contexto de La Guajira, la transición energética justa significa asegurar que las comunidades más vulnerables, sean los primeros beneficiarios del cambio hacia energías renovables. Esto es particularmente relevante dado que la región posee un alto potencial para energía solar y eólica, mientras paradójicamente muchas de sus comunidades carecen de acceso confiable a la electricidad.

La justicia en esta transición se manifiesta en tres aspectos principales: el acceso equitativo a la energía limpia, la participación de las comunidades en las decisiones, y la distribución justa de los beneficios.

**Pobreza Energética**

Este concepto es crucial para nuestra investigación porque nos proporciona los criterios iniciales para identificar hogares vulnerables. No se trata solo de la ausencia de conexión eléctrica, sino de una condición multidimensional que afecta la calidad de vida de las familias. En el caso de La Guajira, la pobreza energética se manifiesta cuando un hogar:

* No puede acceder a servicios energéticos básicos
* Gasta una proporción excesiva de sus ingresos en energía
* Recibe un servicio energético poco confiable o de mala calidad

**Clasificación mediante Inteligencia Artificial**

La inteligencia artificial supervisada es la herramienta que nos permitirá procesar grandes cantidades de datos para identificar patrones de vulnerabilidad energética. En nuestra investigación, emplearemos la siguiente técnica *Random Forest.*

Random Forest representa una evolución de los árboles de decisión, empleando un conjunto de múltiples árboles que votan para producir una predicción más robusta y precisa. Breiman (2001), su creador, lo define como "una combinación de árboles predictores donde cada árbol depende de los valores de un vector aleatorio muestreado independientemente".

Sus características fundamentales incluyen la construcción paralela de múltiples árboles de decisión, el muestreo aleatorio de variables para cada árbol, y la votación mayoritaria para la clasificación final.

En la aplicación para clasificación energética, Random Forest resulta particularmente útil por su capacidad de procesar múltiples variables socioeconómicas, su habilidad para capturar interacciones complejas entre variables, y la generación de rankings de importancia de factores de vulnerabilidad.

**Variables de Clasificación**

Las variables que utilizaremos para la clasificación surgen de la intersección entre la teoría de pobreza energética y las características específicas de La Guajira. Estas se agrupan en tres categorías principales: variables socioeconómicas, energéticas y ambientales.

**Proyecto Gemelo - IPSE: Un Referente para la Priorización Energética[[1]](#footnote-2)**

El proyecto desarrollado por el Instituto de Planificación y Promoción de Soluciones Energéticas (IPSE) en 2023 representa un caso significativo para nuestro marco teórico. Su objetivo fue desarrollar un sistema basado en inteligencia artificial para identificar y priorizar Zonas No Interconectadas (ZNI) que requerían intervenciones energéticas inmediatas. El proyecto logró analizar 1,710 comunidades, alcanzando una precisión del 84% en su clasificación.

* **Relevancia para Nuestra Investigación**

El proyecto IPSE comparte importantes similitudes con nuestra investigación en La Guajira, principalmente en la identificación de comunidades vulnerables para intervenciones energéticas, la utilización de técnicas de IA para procesar datos complejos, el trabajo con comunidades rurales y dispersas incluyendo poblaciones indígenas, y el manejo de datos en áreas con información limitada.

* **Aprendizajes Clave y Variables Críticas**

La experiencia del IPSE enriquece nuestro entendimiento de la vulnerabilidad energética en el contexto colombiano. Su metodología reveló que la vulnerabilidad debe evaluarse considerando múltiples dimensiones, incluyendo características técnicas como la distancia a infraestructura y potencial renovable; factores sociales como la organización comunitaria y servicios básicos existentes; y elementos económicos como la capacidad de pago y actividades productivas.

El proyecto demostró el valor de combinar el análisis computacional de datos cuantitativos con la validación cualitativa comunitaria y la adaptación a contextos culturales específicos.

* **Adaptaciones para La Guajira**

La experiencia del IPSE sugiere adaptaciones cruciales para nuestro estudio. En el ámbito metodológico, destaca la necesidad de integrar variables culturales específicas de la comunidad, adaptar los algoritmos al contexto geográfico y social de La Guajira, e implementar procesos de validación comunitaria. En términos prácticos, es fundamental establecer alianzas con actores locales, documentar el conocimiento tradicional y garantizar la participación comunitaria en todas las fases del proyecto.

* **Arquitectura de IA en el Proyecto IPSE**

El IPSE implementó una arquitectura de dos niveles que combinaba Redes Neuronales Convolucionales (CNN) para el procesamiento de imágenes satelitales y Redes Neuronales Profundas (DNN) para la integración de datos socioeconómicos. Este enfoque híbrido permitió alcanzar una precisión del 84% en la clasificación de zonas prioritarias.

La primera capa del modelo utilizaba una CNN procesando imágenes satelitales de 30m de resolución a través de tres capas convolucionales, capas de MaxPooling y una capa fully-connected, extrayendo características como patrones de asentamiento, accesibilidad y uso del suelo.

La segunda capa empleaba una DNN para procesar datos demográficos, económicos, indicadores de infraestructura y variables socioculturales, utilizando tres capas densas con dropout para prevenir overfitting.

El sistema de clasificación final combinaba los outputs mediante una capa de concatenación, dos capas densas finales y una capa de salida con función softmax para clasificación multiclase. Este diseño demostró ser particularmente efectivo para identificar distintos niveles de prioridad en las ZNI, ofreciendo una base sólida para adaptar a nuestro estudio en La Guajira.

**6. Metodología**

**Diseño de Investigación:**

* **Tipo de Estudio:** Investigación aplicada con enfoque cuantitativo.
* **Enfoque:** Diseño de un modelo de Inteligencia Artificial supervisada para clasificar hogares.

**Población y Muestra:**

* **Población:** Hogares del departamento de La Guajira.
* **Muestra:** Datos ingresados por los usuarios encuestadores para un conjunto representativo de hogares vulnerables de la Guajira con el objetivo de dar priorización a los hogares.

**Herramientas de Recolección de Datos:**

* **Fuentes de Datos (a través de encuesta)**
  + Usuario
* **Fuentes de datos de entrenamiento del modelo:**
  + **Open Data Colombia:** Datos socioeconómicos.
  + **XM:** Consumo eléctrico por región.
  + **IDEAM:** Radiación solar y datos ambientales.

**7. Análisis de Datos**

* **Preprocesamiento:**

***Variables de entrada:***

1. Variables Socioeconómicas

Estas variables permiten evaluar el nivel de vulnerabilidad social y económica de los hogares, lo cual es clave para identificar aquellos en situación de pobreza energética.

* **Ingresos mensuales del hogar**: Representa el nivel de ingresos de los habitantes y ayuda a determinar si pueden cubrir el costo de los servicios energéticos.
* **Estrato socioeconómico:** Los hogares de estrato bajo suelen estar más expuestos a la pobreza energética.

2. Variables Energéticas

Estas variables se centran en el consumo y costo energético actual, proporcionando información clave sobre la situación energética de los hogares.

* **Costo mensual de la electricidad:** Proporción del ingreso familiar destinado a cubrir el servicio eléctrico, lo que indica la carga económica de la energía.

3. Variables Ambientales

Estas variables son importantes para evaluar el potencial de energías renovables, especialmente solar y eólica, en la región de La Guajira.

* **Radiación solar promedio**: Medida en KWh/m² por día; ayuda a determinar el potencial de energía solar en el área.
* **Velocidad y estabilidad del viento**: Medida en m/s; indica la viabilidad de la energía eólica.

Se tuvo en cuenta la normalización de las variables numéricas (ej. ingresos, consumo eléctrico).

* **Entrenamiento del Modelo:**
  + División del dataset en conjunto de entrenamiento (80%) y validación (20%).
  + Uso de algoritmo supervisado:  Random Forest
* **Evaluación del Modelo:**
  + Métricas: accuracy\_score

**8. Resultados Esperados**

* **Hipótesis:** Los hogares con menores ingresos y mayor consumo eléctrico presentan mayor vulnerabilidad energética, siendo candidatos ideales para recibir acceso prioritario a energías limpias.
* **Supuestos:**
  + Los datos del entrenamiento del modelo fueron creados para su aprendizaje.
  + Los datos disponibles serán suficientes y relevantes para entrenar un modelo de IA eficaz.
  + Las comunidades podrían enfrentar otros desafíos energéticos; sin embargo, estaríamos destacando como viables los resultados del modelo entre solar y eólica.
* **Producto Final:** Un modelo entrenado y validado que clasifique los hogares según su nivel de vulnerabilidad energética, acompañado de recomendaciones para la implementación de soluciones de energía solar y eólica.

Ruta repositorio: https://github.com/jaime200617/ProyectoCincoNeuronas

**9. Referencias Bibliográficas**

* **IDEAM:** [https://www.ideam.gov.co](https://www.ideam.gov.co/)
* **Open Data Colombia:** [https://www.datos.gov.co](https://www.datos.gov.co/)
* **XM Colombia:** [https://www.xm.com.co](https://www.xm.com.co/)
* **IPSE:** <https://ipse.gov.co/>

1. IPSE. (2023). Sistema de priorización basado en inteligencia artificial para Zonas No Interconectadas en Colombia: Informe técnico final (Informe No. IPSE-2023-001). Instituto de Planificación y Promoción de Soluciones Energéticas para las Zonas No Interconectadas. https://www.ipse.gov.co/informes/proyecto-ia-zni-2023 [↑](#footnote-ref-2)