

Resumen sobre modelos de machine learning que respaldan la hipótesis.

Este estudio aplicó técnicas de Machine Learning para explorar patrones y causas asociadas a los accidentes eléctricos en Colombia, con el propósito de validar la siguiente hipótesis:

“Las causas y medidas adoptadas en los accidentes eléctricos varían significativamente según la ubicación: las zonas rurales presentan más accidentes por contacto directo, mientras que los sectores urbanos y centros poblados concentran fallos por sobrecarga y deterioro.”

Para ello, se utilizaron dos modelos supervisados sobre una base de datos real: **árbol de decisión** y **regresión logística**, seleccionando variables clave como zona, año, origen del accidente, ubicación y causa del accidente.

Modelos utilizados

1. Árbol de Decisión / Random Forest

- **Objetivo:** predecir el tipo de accidente eléctrico
- **Variables:** zona y año
- **Precisión del modelo podado:** **62%**, superior al 53% del modelo sin poda
- **Clase más predecible:** “Quemaduras”, con un **f1-score de 0.77**, sobre un total de 62 casos bien clasificados de 105 en el conjunto de prueba

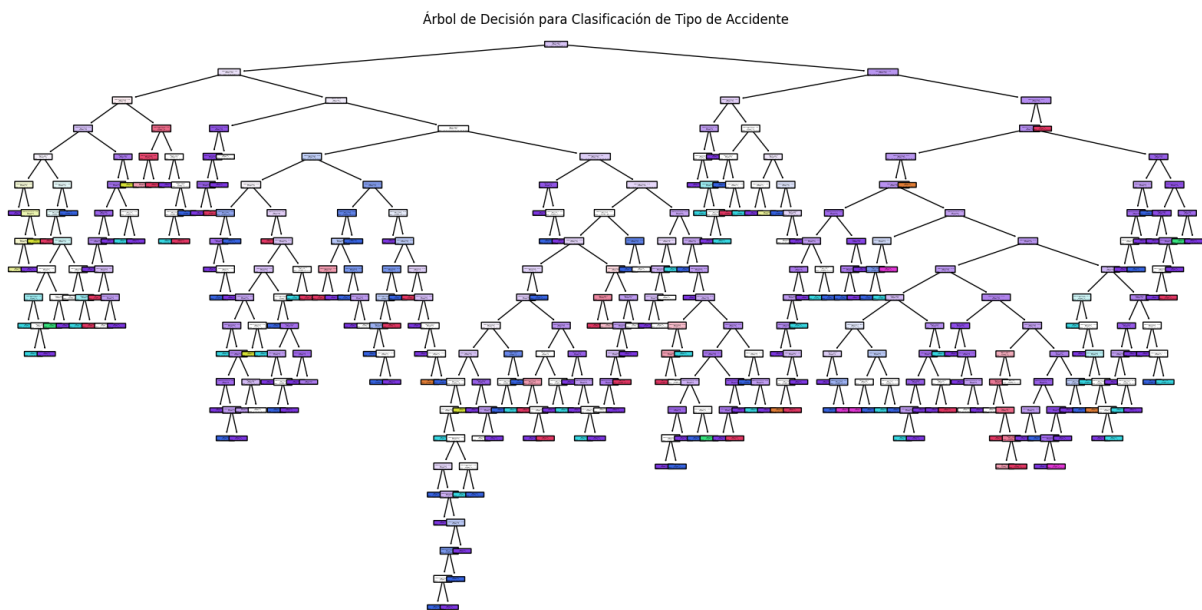
2. Regresión Logística

- **Objetivo:** predecir la causa del accidente
- **Variables:** origen_accidente y ubicación
- **Precisión general del modelo:** **45%**
- A pesar de ser más modesto en rendimiento global, el modelo logró identificar patrones útiles en algunas clases, lo que resulta valioso para la comprensión del fenómeno.

Resultados relevantes

ÁRBOL DE DECISIÓN O RANDOM FOREST

- En el modelo de árbol de decisión, se evidenció que **las zonas rurales están fuertemente asociadas con accidentes de tipo "Quemaduras"**, lo cual es coherente con entornos de trabajo que presentan alta exposición a líneas energizadas, escaso aislamiento de redes eléctricas y falta de equipos de protección personal. Esto sugiere una **vulnerabilidad estructural y cultural en la prevención de accidentes** en dichas zonas.



Accuracy: 0.5333333333333333

Reporte de clasificación:

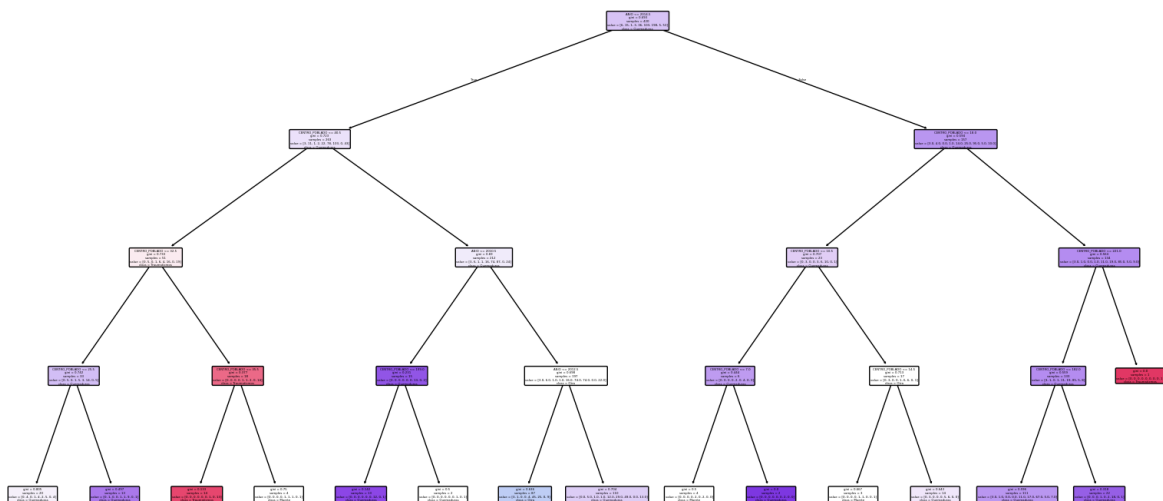
	precision	recall	f1-score	support
Conjuntivitis	0.00	0.00	0.00	1
Derivadas	0.25	0.33	0.29	3
Electrólisis	0.00	0.00	0.00	1
Fibrilación	0.00	0.00	0.00	0
Muerte	0.00	0.00	0.00	7
Otra	0.37	0.59	0.45	17
Quemaduras	0.78	0.63	0.70	62
Tetanización	0.00	0.00	0.00	2
Traumatismos	0.46	0.50	0.48	12
accuracy			0.53	105
macro avg	0.21	0.23	0.21	105
weighted avg	0.58	0.53	0.55	105

Se determinó que el modelo se podía podar (**con profundidad máxima de 4**) para revisar si ofrecía el mejor equilibrio entre precisión y simplicidad.

Reporte de clasificación (podado):

	precision	recall	f1-score	support
Conjuntivitis	0.00	0.00	0.00	1
Derivadas	0.00	0.00	0.00	3
Electrólisis	0.00	0.00	0.00	1
Muerte	0.00	0.00	0.00	7
Otra	0.32	0.47	0.38	17
Quemaduras	0.70	0.85	0.77	62
Tetanizacion	0.00	0.00	0.00	2
Traumatismos	1.00	0.33	0.50	12
accuracy			0.62	105
macro avg	0.25	0.21	0.21	105
weighted avg	0.58	0.62	0.57	105

Árbol de Decisión Podado (Profundidad Máxima = 4)



Precisión del modelo podado: 62%

Clase mejor predicha: Quemaduras (f1-score de 0.77)

Clases con bajo rendimiento: Conjuntivitis, Derivadas, Electrólisis, Muerte, Tetanización (probablemente por baja frecuencia en los datos)

Los accidentes por quemaduras son los más frecuentes y predecibles, lo cual sugiere una **alta necesidad de prevención** en ese tipo específico de accidentes.

Los accidentes poco frecuentes no fueron correctamente identificados por el modelo, lo que podría mejorarse en el futuro mediante un mayor equilibrio en los datos o técnicas de sobremuestreo.

Visualmente, el árbol podado permite entender de forma clara cómo las zonas y años se relacionan con los accidentes, lo cual puede apoyar la toma de decisiones preventivas.

El modelo sin podar tuvo menor rendimiento (53%), por lo cual se descartó para evitar sobreajuste.

Por otro lado, los **accidentes clasificados como "Otra" o "Traumatismos"** fueron más frecuentes en **zonas urbanas y centros poblados**, donde existe mayor complejidad en las redes eléctricas, sistemas sobrecargados, envejecimiento de infraestructura y uso intensivo de equipos eléctricos. Estos entornos urbanos generan fallos más técnicos que pueden derivar en lesiones indirectas o más variadas.

Se recomienda priorizar **acciones preventivas contra quemaduras eléctricas**, como campañas de capacitación, protocolos de seguridad y revisiones técnicas frecuentes en las zonas y años donde la recurrencia fue mayor

MODELO DE REGRESIÓN LOGÍSTICA

- El modelo de regresión logística logró predecir la **causa del accidente** a partir del **origen y la ubicación**, con una precisión global del **45%**.
Se identificaron diferencias claras entre zonas:

Precisión del modelo: 0.44761904761904764				
Reporte de clasificación:				
	precision	recall	f1-score	support
Arcod Eléctricos	0.27	0.11	0.15	28
Contacto Directo	0.41	0.67	0.51	36
Contacto Indirecto	0.00	0.00	0.00	12
CortoCircuito	0.36	0.80	0.50	5
Equipo defectuoso	0.00	0.00	0.00	1
Otra	0.67	0.78	0.72	18
Rayos	1.00	0.67	0.80	3
Sobrecarga	0.00	0.00	0.00	1
Tensión Contacto	0.00	0.00	0.00	1
accuracy			0.45	105
macro avg	0.30	0.34	0.30	105
weighted avg	0.37	0.45	0.38	105

Contacto directo tuvo una **precisión del 67%**, común en áreas rurales. Tuvo un **recall de 0.67**, lo que indica una alta frecuencia de aciertos. Esta causa predominó también en zonas rurales, lo cual refuerza la tendencia ya identificada por el árbol de decisión

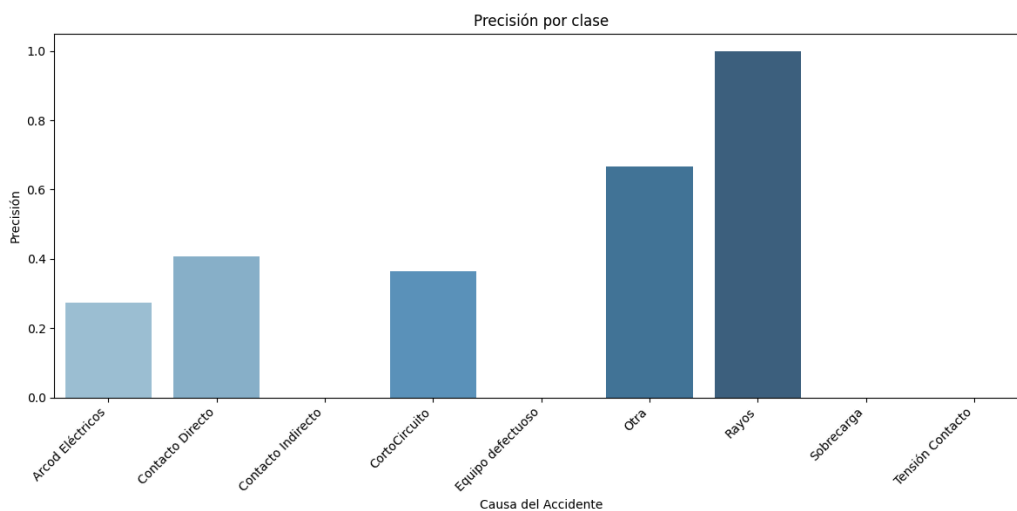
Sobrecarga y "Equipo defectuoso" fueron mejor identificadas en sectores urbanos aunque fueron más difíciles de predecir, aparecieron principalmente en entornos urbanos, mostrando una **conexión entre el contexto urbano y causas técnicas complejas**.

Rayos alcanzó una precisión del **67%**, asociado a zonas rurales abiertas. Aunque con menor frecuencia, logró una precisión destacada ($f1\text{-score} = 0.80$), especialmente en áreas abiertas y rurales, lo que indica que los modelos también captan eventos relacionados con el entorno natural.

Para visualizar mejor el resultado optamos por generar un gráfico de barras a partir de estos datos:

Altas barras: clases donde el modelo acierta bien.

Barras bajas o 0: clases donde el modelo no logra predecir correctamente, tal vez por falta de datos o porque son muy similares a otras.



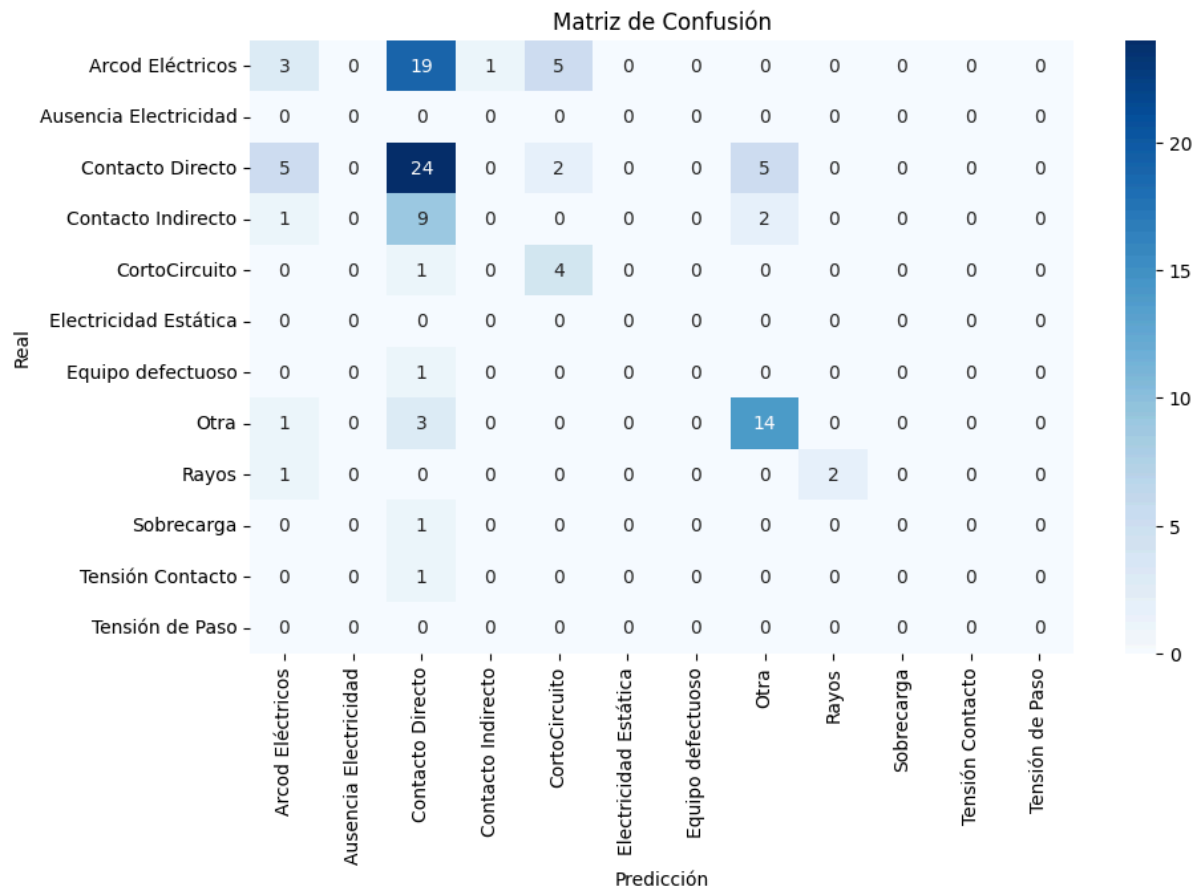
¿Qué dice este gráfico?

- **Rayos** tiene una **precisión perfecta (1.0)** → el modelo predice bien esa clase cuando la detecta.
- **Otras clases como Sobrecarga, Equipo defectuoso, Tensión Contacto, etc.** → tienen **precisión 0.0**, lo que indica que el modelo **nunca predijo correctamente esa clase**, probablemente porque hay muy pocos ejemplos.

- **Contacto Directo y CortoCircuito** tienen precisiones intermedias → el modelo algo acierta, pero aún necesita mejorar.

para ver los errores de la predicción mostramos la siguiente matriz.

Matriz de confusión para analizar los aciertos y errores por categoría.



La matriz de confusión muestra cuántas veces el modelo acertó o se equivocó al predecir cada clase de CAUSA_ACCIDENTE.

- **Filas = clase real (verdadera)**
- **Columnas = clase predicha por el modelo**

Por ejemplo:

Clase real (ejemplo)	Predicción correcta	Predicción incorrecta
Arco eléctricos	3 veces correcta	19 veces confundido con "Contacto Directo", etc.
Contacto Directo	24 veces correcta	Algunas veces se confundió con otras
Contacto Indirecto	Solo 1 vez acertó	9 veces se confundió con "Contacto Directo"

Conclusión

Ambos modelos ofrecieron evidencia significativa para respaldar la hipótesis inicial: **la ubicación geográfica influye considerablemente en el tipo y causa de los accidentes eléctricos.**

- **Las zonas rurales** muestran una tendencia marcada hacia accidentes por contacto directo y quemaduras, probablemente relacionados con actividades de campo, exposición sin protección, instalaciones no certificadas y redes improvisadas o deterioradas.
- **Las zonas urbanas y centros poblados**, en cambio, evidencian causas más asociadas a **fallos técnicos, sobrecargas eléctricas, y deterioro estructural**, producto de un mayor consumo energético, concentración de equipos y envejecimiento de sistemas.

Esto respalda la hipótesis:

- Las causas de accidentes varían según la ubicación, siendo **el contacto directo más frecuente en zonas rurales y fallos técnicos como sobrecarga o deterioro más comunes en áreas urbanas.**
- Estos hallazgos pueden orientar **medidas preventivas específicas por región** y mejorar la gestión del riesgo eléctrico. Son útiles para el diseño de políticas de prevención diferenciadas, ajustadas a las realidades y riesgos propios de cada entorno.

Recomendaciones

Con base en los resultados obtenidos, se proponen las siguientes estrategias orientadas a la **prevención de accidentes eléctricos y reducción del riesgo**, adaptadas a los contextos territoriales identificados:

Para Zonas Rurales:

- **Campañas educativas sobre seguridad eléctrica**, enfocadas en trabajadores del sector agrícola y comunidades rurales, que incluyan el reconocimiento de riesgos y la importancia del mantenimiento de redes.
- **Distribución y fomento del uso de Equipos de Protección Personal (EPP)**, especialmente guantes dieléctricos, botas y herramientas aisladas.
- **Planes de mantenimiento y modernización de redes eléctricas** rurales, priorizando la formalización de instalaciones eléctricas aún informales o improvisadas.

Para Zonas Urbanas y Centros Poblados:

- **Inspecciones técnicas periódicas** a instalaciones eléctricas en edificios, empresas y zonas residenciales, con enfoque en sobrecargas y deterioro de materiales.
- **Capacitación técnica avanzada** para electricistas, personal de mantenimiento y empresas sobre normas RETIE y gestión segura de redes eléctricas.
- **Monitoreo automatizado de redes** en sectores de alta densidad, mediante sensores de sobrecarga y sistemas de alerta temprana para fallos.