**Ensayo investigación del Modelo Prueba 2 Random Forest y XGBoost.**

Objetivo del modelo:

Predecir si un accidente eléctrico podría ser grave (es decir, con resultado de muerte), a partir de variables como edad, sexo, causa del accidente, ubicación, tipo de vinculación, entre otras.

¿Qué logra el proyecto?:

Identificar factores de riesgo que están correlacionados con accidentes mortales.

Generar alertas tempranas en situaciones similares a las que históricamente han terminado en muerte.

Apoyar la toma de decisiones en seguridad laboral, priorizando medidas preventivas más efectivas.

Optimizar recursos de prevención, dirigiendo la atención hacia los casos con mayor probabilidad de gravedad.

Aplicación práctica para Grupo Electra:

Focalizar capacitaciones, revisiones y equipos de seguridad donde más se necesita.

Justificar inversiones en prevención basadas en datos.

Reducir incidentes fatales mediante intervención temprana.

¿Qué hace este modelo que acabamos ?

Toma datos históricos de accidentes laborales ocurridos en la empresa, con variables como edad, tipo de vinculación, sexo, sección, etc.

Clasifica si un accidente será grave (muerte) o no:

La variable objetivo del modelo es GRAVE, construida con base en si TIPO\_LESION fue "muerte".

Se entrenaron dos modelos: RandomForest y XGBoost.

Se usó SMOTE y RandomOverSampler para balancear clases (porque los accidentes graves son poco frecuentes).

Se evaluó el desempeño con AUC, F1-score y curvas ROC.

Se identificó el umbral óptimo de predicción, para hacer el modelo más preciso.

Se exportaron resultados y métricas para uso en herramientas como Tableau.

Permite saber la probabilidad de que un accidente con ciertas características termine en muerte.

¿Cumple con lo solicitado?

Sí, el modelo cumple técnicamente con lo que te solicitan.

Porque:

Te permite predecir accidentes graves (variable GRAVE).

Puedes identificar situaciones de alto riesgo, lo que facilita priorizar estrategias de prevención.

La empresa puede usar estos resultados para actuar antes de que ocurra un accidente mortal.

¿Qué podrías hacer adicionalmente para mejorarlo?

Interpretabilidad:

Agrega un análisis de importancia de variables (feature importance), para saber qué factores están más asociados a los accidentes graves.

Ejemplo: si "CAUSA\_ACCIDENTE" o "SECCION\_EMPRESA" tienen alta importancia, podrías recomendar intervenciones específicas.

Dashboard o visualización:

Crear un dashboard en Tableau o Power BI con los CSV generados para mostrar:

Predicciones.

Riesgo por sección, edad, causa.

Comparación entre modelos.

Recomendaciones estratégicas:

A partir de los resultados, elaborar recomendaciones concretas para seguridad industrial (por ejemplo, reforzar capacitación en secciones con mayor riesgo).

Conclusión

Tu modelo es coherente con el objetivo del proyecto y responde a lo que te están solicitando. Lo tienes bien encaminado. Ahora puedes concentrarte en interpretar y presentar los resultados para que tengan impacto en la toma de decisiones.

¿Quieres que te ayude a escribir esa conclusión o recomendación final para el informe?

Perfecto. Aquí tienes un texto claro, formal y directo que puedes incluir en las conclusiones o recomendaciones finales del informe de tu proyecto:

Conclusión del Proyecto

Este proyecto logró desarrollar un modelo predictivo que permite anticipar la probabilidad de que un accidente laboral en la empresa Grupo Electra tenga un desenlace grave (específicamente, la muerte del trabajador). A través de la recolección, análisis y tratamiento de los datos históricos de accidentes, se entrenaron y evaluaron modelos de aprendizaje automático como Random Forest y XGBoost, utilizando técnicas de balanceo como SMOTE y RandomOverSampler, para abordar el desequilibrio natural en la ocurrencia de eventos fatales.

El modelo final muestra una capacidad aceptable de discriminación (AUC superior al 0.70), lo que permite identificar con anticipación escenarios de mayor riesgo. Esta capacidad predictiva representa una herramienta clave para priorizar acciones preventivas y estrategias de intervención, enfocando los esfuerzos de la empresa en los perfiles y condiciones que históricamente se asocian con mayor gravedad.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Recomendaciones Estratégicas

Implementar sistemas de monitoreo preventivo que, apoyados en los resultados del modelo, generen alertas para los casos de mayor riesgo (por ejemplo, trabajadores jóvenes o nuevos, ubicaciones críticas, causas específicas de accidentes).

Fortalecer las capacitaciones en las áreas más sensibles, según la sección empresarial o causa más frecuente asociada a accidentes fatales identificada por el modelo.

Integrar los resultados en un dashboard interactivo (por ejemplo, Tableau o Power BI), lo cual facilitará la visualización de patrones de riesgo y apoyará la toma de decisiones en tiempo real.

Reentrenar y actualizar el modelo periódicamente, incorporando nuevos datos para mantener su precisión y adaptabilidad a los cambios en condiciones laborales.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

¿Qué el modelo actualmente?

Analiza datos históricos de accidentes laborales con variables relevantes como:

Edad, tiempo de vinculación, causa del accidente, sección de la empresa, etc.

Equilibra los datos desbalanceados (pocas muertes vs. muchos no graves) usando técnicas como SMOTE y RandomOverSampler, lo cual es crítico para que el modelo no subestime los casos graves.

Predice la probabilidad de que un accidente sea grave (muerte) con modelos como:

Random Forest

XGBoost (optimizando umbrales con F1-score)

Evalúa el rendimiento de los modelos con métricas reales:

F1-score, matriz de confusión, AUC-ROC, umbral óptimo.

Genera salidas útiles:

Exporta resultados para Tableau (visualización).

Permite priorizar zonas, cargos o condiciones con mayor riesgo.

Brinda priorización basada en probabilidades reales, útil para que la empresa tome decisiones de prevención basadas en datos y evidencia.

Utiliza técnicas modernas de aprendizaje automático con interpretabilidad y métricas confiables.