**Informe Final: Análisis Predictivo de Accidentes Eléctricos en la Empresa**

**1. Introducción**

La prevención de accidentes eléctricos representa uno de los principales desafíos en la gestión de seguridad de empresas del sector eléctrico. El análisis sistemático de los datos históricos de accidentes permite identificar patrones, factores de riesgo y diseñar estrategias de intervención basadas en evidencia. Este informe resume el proceso y los hallazgos de un análisis predictivo realizado sobre la base de datos de accidentes eléctricos reportados por la empresa, con el objetivo de identificar los factores que más inciden en la gravedad de estos eventos.

**2. Objetivo**

Identificar los factores que permiten predecir la ocurrencia de accidentes eléctricos graves, utilizando técnicas estadísticas y de aprendizaje automático, y proponer recomendaciones para focalizar las estrategias preventivas.

**3. Hipótesis y Pregunta de Investigación**

**Hipótesis:**  
Ciertos factores (tipo de accidente, ubicación, tipo de vinculación laboral, causa reportada, entre otros) influyen significativamente en la gravedad de los accidentes eléctricos reportados en la empresa. La identificación de estos factores permitirá priorizar estrategias de prevención más efectivas.

**Pregunta-problema:**  
¿Cuáles son los principales factores asociados a la ocurrencia y gravedad de los accidentes eléctricos en la empresa, y cómo puede priorizarse la intervención preventiva para reducir su incidencia?

**4. Metodología**

El proceso analítico se desarrolló en cinco etapas, siguiendo las directrices establecidas para proyectos de análisis de datos:

1. **Limpieza y exploración de datos**: Revisión de la calidad y estructura de la base de datos.
2. **Análisis exploratorio (EDA)**: Estudio descriptivo de las variables, identificación de tendencias y patrones iniciales.
3. **Modelado predictivo**: Construcción de modelos para identificar variables asociadas a la gravedad de los accidentes.
4. **Visualización de resultados**: Elaboración de dashboards y gráficos para facilitar la interpretación y comunicación.
5. **Conclusiones y recomendaciones**: Redacción del informe final y propuesta de acciones preventivas.

**5. Análisis Exploratorio de Datos**

**5.1. Desbalance de clases**

* Accidentes no graves: 358 (aproximadamente 91%)
* Accidentes graves (muerte): 32 (aproximadamente 9%)
* Este fuerte desbalance representa una limitación para los modelos predictivos, ya que dificulta la detección automática de los eventos graves.

**5.2. Principales hallazgos descriptivos**

* **Tipo de vinculación:** La mayoría de los accidentes graves ocurrieron en trabajadores contratistas, lo que sugiere un posible mayor riesgo en este grupo.
* **Ubicación:** Ciertas ubicaciones presentan una concentración mayor de accidentes graves.
* **Antigüedad:** Se observa variabilidad en la antigüedad de las víctimas, pero no se encontró un patrón claro entre antigüedad y gravedad.
* **Causa del accidente:** Algunas causas específicas, como el contacto directo o la falta de uso de elementos de protección personal, aparecen asociadas a mayor riesgo, aunque la relación no es determinante.

**6. Modelado Predictivo**

**6.1. Técnicas aplicadas**

* Se emplearon modelos de regresión logística, Random Forest y XGBoost.
* Se aplicó **SMOTE** para balancear la clase minoritaria y mejorar la capacidad de los modelos para identificar accidentes graves.

**6.2. Resultados de los modelos**

* **Precisión global** (accuracy): hasta 0.92 antes de SMOTE, y 0.87 después del balanceo.
* **Recall para accidentes graves**: mejoró de 0.00 (sin balanceo) a 0.27 (con SMOTE).
* **AUC-ROC**: alrededor de 0.67, indicando una capacidad de discriminación moderada.

**6.3. Interpretación**

* Los modelos predictivos actuales son capaces de identificar tendencias generales y segmentos de mayor riesgo (por ejemplo, contratistas), pero su capacidad para predecir casos graves sigue siendo limitada.
* Las variables más influyentes fueron el tipo de vinculación y la ubicación, aunque ninguna variable por sí sola explica completamente la gravedad de los accidentes.

**7. Limitaciones del Análisis**

* **Tamaño y representatividad de la base de datos:** El número de casos graves es bajo, lo que limita el poder predictivo de los modelos.
* **Variables disponibles:** Faltan variables contextuales importantes (uso de EPP, condiciones del entorno, cumplimiento de protocolos, factores climáticos).
* **Desbalance de clases:** Aunque SMOTE ayuda, no resuelve por completo el problema de baja sensibilidad para los eventos críticos.

**8. Recomendaciones**

**8.1. Para la gestión y la toma de decisiones**

* Utilizar los modelos predictivos como **herramientas de apoyo** para focalizar acciones preventivas, no como criterio exclusivo.
* **Priorizar la formación y supervisión** en los grupos identificados como de mayor riesgo (contratistas, ciertas ubicaciones).
* **Refinar y ampliar la base de datos**: Implementar procedimientos para recolectar más y mejores datos, incluyendo nuevas variables relevantes.
* **Monitoreo continuo:** Actualizar el modelo y el análisis conforme se disponga de más información.

**8.2. Para la visualización y el análisis continuo (Tableau)**

* Exportar periódicamente la base de datos a archivos .csv para alimentar dashboards en Tableau.
* Visualizar indicadores clave: accidentes graves/no graves por tipo de vinculación, ubicación, causa, antigüedad, y otros.
* Utilizar filtros para monitorear tendencias en tiempo real y ajustar las estrategias preventivas en función de los datos actualizados.

**9. Guía breve para la visualización en Tableau**

1. **Exportar los datos en formato .csv desde Python.**
2. **Importar el archivo en Tableau** usando la opción "Archivo de texto".
3. **Crear los principales gráficos:**
   * Barras apiladas: accidentes graves/no graves por tipo de vinculación.
   * Mapas de calor: ubicación vs. gravedad.
   * Histograma/línea de tiempo: accidentes por año/mes.
   * Boxplot: antigüedad vs. gravedad.
4. **Construir dashboards** combinando varios gráficos e incorporando filtros.
5. **Interpretar resultados y comunicar hallazgos** al equipo directivo y de seguridad.

**10. Conclusiones**

* El análisis predictivo confirma la importancia del **tipo de vinculación** y la **ubicación** como factores asociados a la gravedad de los accidentes eléctricos.
* El modelo predictivo construido, aunque limitado por la cantidad y calidad de los datos, representa la mejor aproximación técnica posible en el contexto actual.
* Se recomienda fortalecer la cultura del reporte, la recolección de datos y la formación en seguridad para avanzar hacia una gestión más preventiva y basada en evidencia.

**11. Mensaje final para directivos y equipo de seguridad**

*“El camino hacia la reducción de accidentes graves en la empresa requiere la integración de análisis de datos, experiencia del personal y mejora continua en los procesos. La tecnología y la analítica predictiva deben ser aliadas, pero siempre en el marco de una gestión integral y colaborativa de la seguridad.”*