

EXPERIMENTANDO CON HERRAMIENTAS SIN CÓDIGO “PREDICTOR DE LA PRESENCIA O AUSENCIA DE EQUIPO DE PROTECCIÓN PERSONAL”

P R E S E N T A

Johan Pichardo García
Rafael Yairjafeth Rojas Nieves
Samuel Hiram Medina Castañeda
Luz María Feregrino Martínez

Agosto 2024

ÍNDICE

INTRODUCCIÓN	1
1. IDENTIFICACIÓN Y DELIMITACIÓN DEL PROBLEMA	5
1.1. Impacto del problema en la industria.....	6
2. DEFINICIÓN DE LOS DATOS	8
2.1. Descripción del dataset	8
2.2. Datos Simulados.....	10
2.3. Cantidad de Datos Requerida.....	10
2.4. Fuente de los Datos	11
3. ANÁLISIS DE SESGOS EN LOS DATOS	11
3.1. Identificación de Posibles Sesgos	11
3.2. Estrategias para Mitigar los Sesgos	12
3.3. Garantía de Equidad y Confiabilidad del Modelo	13
4. SELECCIÓN DEL TIPO DE APRENDIZAJE	13
4.1. Justificación del tipo de aprendizaje a utilizar	13
5. SELECCIÓN DEL MODELO O MODELOS.....	15
5.1. Descripción del Modelo o Modelos	15
5.2. Justificación de la Elección.....	15
5.3. Explicación del Funcionamiento	15
6. PROCESO DEL MODELO	16
6.1. Lectura de Datos.....	19
6.2. Limpieza y Preparación de Datos	20
6.3. Aplicación de Reglas y Normalización.....	21
6.4. Modelo de aprendizaje	23
6.5. Métricas y Evaluación.....	26
6.6. Visualización de Datos	29
6.7. Definición del Método de Deployment	29
7. RESULTADOS.....	31
7.1. Resultado de la predicción	32
CONCLUSIONES	34
BIBLIOGRAFÍA	37

INTRODUCCIÓN

La seguridad en el lugar de trabajo es una prioridad en cualquier industria, especialmente en entornos donde los trabajadores están expuestos a riesgos significativos, como la construcción, la manufactura y la minería. Una de las principales causas de accidentes laborales graves es la falta de uso adecuado del equipo de protección personal (EPP). Estos accidentes no solo pueden causar lesiones graves o incluso la muerte, sino que también pueden resultar en costos significativos para las empresas en términos de compensaciones, multas y pérdida de productividad.

El impacto de los accidentes laborales se extiende más allá del individuo afectado. Las empresas enfrentan costos directos e indirectos que pueden ser devastadores. Los costos directos incluyen gastos médicos y compensaciones por lesiones, mientras que los costos indirectos abarcan la pérdida de productividad, el tiempo de inactividad, el entrenamiento de nuevos empleados, y el daño a la reputación de la empresa. Además, las empresas pueden enfrentar multas regulatorias y legales que agravan aún más las pérdidas financieras.

Implementar medidas de seguridad efectivas y garantizar el uso adecuado del EPP no solo es una responsabilidad ética, sino también una estrategia de negocio esencial. La prevención de accidentes mejora el bienestar de los empleados, fomenta un ambiente de trabajo más seguro y productivo, y contribuye a la sostenibilidad financiera de la empresa. En este contexto, el desarrollo y la implementación de sistemas de monitoreo y evaluación del cumplimiento de las normas de seguridad, apoyados por tecnologías como la visión artificial y el análisis de datos, se convierten en herramientas cruciales para la gestión proactiva de la seguridad en el lugar de trabajo.

De acuerdo con las estadísticas de Riesgos de trabajo registrados en el IMSS del 2008 al 2018 mostrados en la gráfica de la figura 1 , en promedio ocurren 527,802 riesgos de trabajo por año, distribuidos en accidentes de trabajo, accidentes en trayecto y enfermedades de trabajo conforme a la gráfica 2 de la figura 1. Si se desglosa esta información, se podría decir que en promedio ocurren 1,446 riesgos de trabajo por día en todo el territorio nacional.

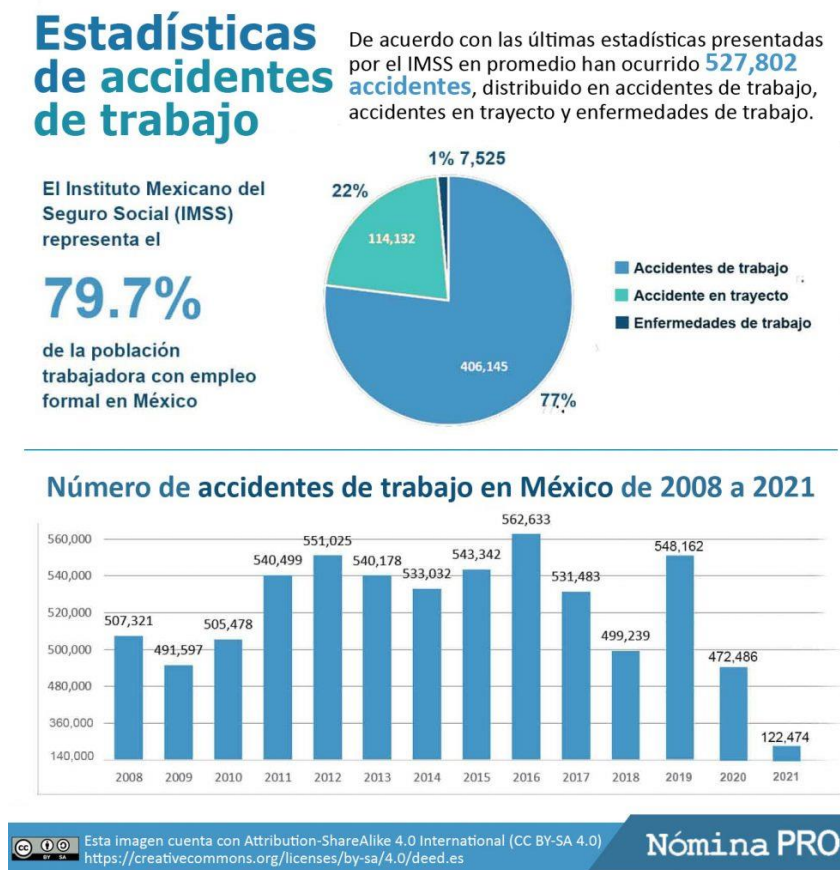


Figura 1. Estadísticas de accidentes de trabajo.[1]

El objetivo de este proyecto es desarrollar y evaluar un sistema que utilice técnicas de visión artificial simulada para:

1. Identificar si los trabajadores están utilizando su equipo de protección personal.



Figura 2. Equipo de protección personal (EPP)

2. Detectar actividades de mantenimiento y prevenir la activación de maquinaria durante estas actividades.

Estadísticas de Accidentes Laborales

Los accidentes laborales son una preocupación global. Según la Organización Internacional del Trabajo (OIT), cada año ocurren aproximadamente 374 millones de accidentes laborales no mortales y 2.78 millones de muertes relacionadas con el trabajo [1].

Distribución de los Riesgos de Trabajo

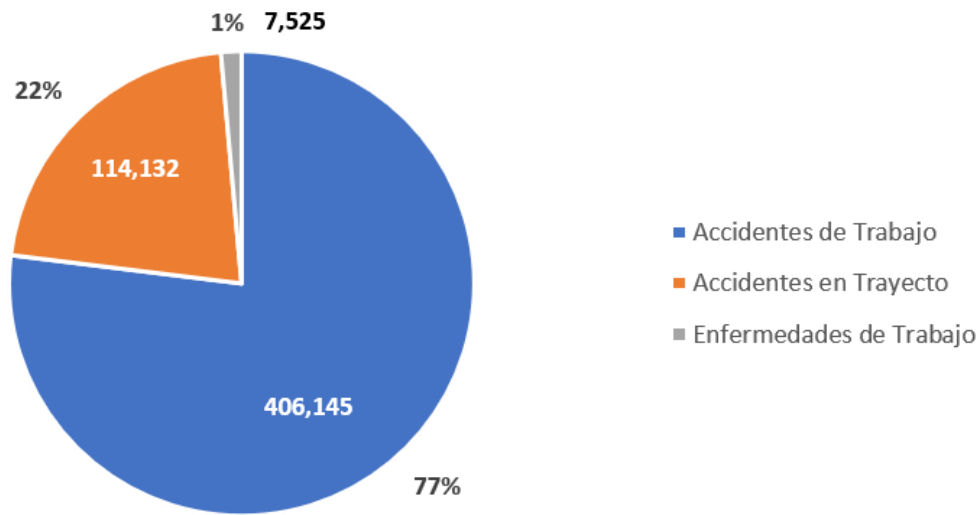


Figura 3. Distribución de riesgos de trabajo, [1].

En muchos casos, estos incidentes son prevenibles con el uso adecuado del EPP. Algunas estadísticas destacadas incluyen:

- **Caídas:** Representan el 15% de todas las muertes relacionadas con el trabajo en la construcción, a menudo debido a la falta de arneses de seguridad [2].
- **Golpes por objetos:** Constituyen el 10% de los accidentes laborales, muchas veces por no usar cascos [3].
- **Lesiones por maquinaria:** Son responsables de un 14% de los accidentes, frecuentemente ocurren durante el mantenimiento cuando no se han tomado las precauciones adecuadas para desactivar el equipo [4].

Consecuencias de los Accidentes Laborales

Las consecuencias de los accidentes laborales son variadas y graves:

- **Humanas:** Lesiones graves, incapacidades permanentes y pérdida de vidas.
- **Económicas:** Altos costos médicos, compensaciones a los trabajadores, multas y litigios.

- **Productividad:** Pérdida de tiempo de trabajo, disminución de la moral del personal y reducción de la eficiencia operativa.

Razones de los Accidentes

La falta de uso adecuado del EPP y errores durante el mantenimiento de maquinaria son causas comunes de accidentes laborales. Estas situaciones suelen ocurrir por:

- **Falta de concienciación:** Los trabajadores no siempre comprenden la importancia del EPP.
- **Entrenamiento insuficiente:** No todos los empleados están adecuadamente capacitados en el uso del EPP.
- **Supervisión deficiente:** Falta de monitoreo continuo para asegurar que se sigan las normas de seguridad.
- **Condiciones de trabajo:** Entornos laborales que no facilitan el uso constante del EPP o no aseguran la desactivación correcta de la maquinaria durante el mantenimiento.

Para abordar estos desafíos, este proyecto propone un sistema de visión artificial que podría ayudar a mejorar la seguridad en el trabajo mediante la detección y prevención de situaciones peligrosas.

1. IDENTIFICACIÓN Y DELIMITACIÓN DEL PROBLEMA

Contexto y descripción del problema seleccionado

En numerosos sectores industriales, como la construcción, la manufactura y la minería, los trabajadores están expuestos a una variedad de riesgos laborales. Uno de los factores más críticos para garantizar la seguridad en el trabajo es el uso adecuado del equipo de protección personal (EPP). Sin embargo, la supervisión manual del uso del EPP es a menudo ineficaz y costosa, y no siempre garantiza que los trabajadores sigan las normas de seguridad establecidas.

Adicionalmente, las actividades de mantenimiento de maquinaria presentan un riesgo significativo. Activar accidentalmente una máquina mientras se realiza el mantenimiento puede resultar en lesiones graves o incluso la muerte del trabajador involucrado. Las prácticas actuales para prevenir estos incidentes, como los procedimientos de bloqueo y etiquetado, pueden ser insuficientes o mal implementadas, aumentando así el riesgo de accidentes [4].

1.1. Impacto del problema en la industria

El impacto de los accidentes laborales debido a la falta de uso del EPP y errores durante el mantenimiento de maquinaria es significativo y multifacético:

1. Impacto Humano:

- **Lesiones y muertes:** La falta de uso del EPP y los accidentes durante el mantenimiento son causas comunes de lesiones graves y mortales en el lugar de trabajo. Según la Organización Internacional del Trabajo (OIT), cada año ocurren aproximadamente 374 millones de accidentes laborales no mortales y 2.78 millones de muertes relacionadas con el trabajo [1].

2. Impacto Económico:

- **Costos directos:** Los accidentes laborales resultan en costos directos significativos, incluidos gastos médicos y compensaciones a los trabajadores lesionados [2].
- **Costos indirectos:** Las empresas también enfrentan costos indirectos, como pérdida de productividad, tiempo de inactividad, costos de capacitación para reemplazar trabajadores lesionados y daños a la reputación de la empresa [3].

3. Impacto Operacional:

- **Disminución de la moral y la eficiencia:** Los accidentes laborales afectan negativamente la moral de los empleados y pueden llevar a una disminución de la eficiencia operativa. Un entorno de trabajo seguro es crucial para mantener una fuerza laboral motivada y productiva.

- **Interrupciones en la producción:** Los accidentes durante el mantenimiento de maquinaria pueden causar interrupciones significativas en la producción, afectando la capacidad de la empresa para cumplir con los plazos y mantener la calidad del producto [4].

Ejemplos de Incidentes y sus Consecuencias

- **Caso de Estudio 1:** En una planta de manufactura, un trabajador no utilizó adecuadamente su casco de seguridad y sufrió una lesión grave en la cabeza debido a la caída de un objeto pesado. La empresa tuvo que pagar altos costos médicos y compensaciones, además de enfrentar sanciones regulatorias por no asegurar el uso adecuado del EPP [2].
- **Caso de Estudio 2:** En una instalación de mantenimiento industrial, un trabajador murió cuando una máquina fue accidentalmente activada durante el mantenimiento. La investigación reveló que los procedimientos de bloqueo y etiquetado no se siguieron correctamente, lo que resultó en una demanda costosa y la implementación de nuevas políticas de seguridad [4].

Necesidad de Soluciones Innovadoras

Para abordar estos problemas, se necesitan soluciones innovadoras que puedan proporcionar monitoreo y prevención en tiempo real. La visión artificial es una tecnología emergente con el potencial de transformar la seguridad laboral. Al implementar sistemas que pueden identificar automáticamente el uso del EPP y detectar situaciones de riesgo, las empresas pueden reducir significativamente la tasa de accidentes y mejorar la seguridad general en el lugar de trabajo.

Este proyecto tiene como objetivo demostrar cómo un sistema de visión artificial, aunque simulado, puede abordar eficazmente estos desafíos y contribuir a la creación de entornos laborales más seguros y eficientes.

2. DEFINICIÓN DE LOS DATOS

El dataset teórico está diseñado para evaluar el cumplimiento del uso del equipo de protección personal (EPP) en el lugar de trabajo. Ver figura 4.

	A	B	C	D	E	F	G	H	I	J
	ID del Trabajador	Nombre del Trabajador	Area	Fecha de evaluación	Casco	Chaleco Reflectante	Botas de Seguridad	Protección Ocular	Protección Auditiva	Guantes de Seguridad
1										
2	1	Trabajador 1	Área de Supervisión	07/08/2024	Sí	Sí	Sí	No	No	No
3	2	Trabajador 2	Área de Pintura	16/08/2024	Sí	No	Sí	Sí	No	Sí
4	3	Trabajador 3	Área de Soldadura	30/08/2024	No	Sí	Sí	No	Sí	Sí
5	4	Trabajador 4	Área de Almacenamiento de Materiales	10/08/2024	Sí	Sí	Sí	No	No	Sí
6	5	Trabajador 5	Área de Almacenamiento de Materiales	23/08/2024	No	No	No	Sí	No	No
7	6	Trabajador 6	Área de Pintura	01/08/2024	Sí	No	Sí	Sí	No	Sí
8	7	Trabajador 7	Área de Pintura	29/08/2024	Sí	No	Sí	Sí	No	Sí
9	8	Trabajador 8	Área de Soldadura	08/08/2024	Sí	Sí	No	Sí	No	No
10	9	Trabajador 9	Área de Pintura	29/08/2024	Sí	No	Sí	Sí	No	Sí
11	10	Trabajador 10	Área de Soldadura	24/08/2024	Sí	Sí	Sí	No	No	No
12	11	Trabajador 11	Área de Supervisión	15/08/2024	Sí	Sí	Sí	No	No	No
13	12	Trabajador 12	Área de Montaje de Estructuras	11/08/2024	Sí	Sí	Sí	Sí	No	Sí
14	13	Trabajador 13	Área de Electricidad	14/08/2024	No	No	No	No	No	Sí
15	14	Trabajador 14	Área de Montaje de Estructuras	04/08/2024	Sí	Sí	No	Sí	No	No
16	15	Trabajador 15	Área de Supervisión	04/08/2024	Sí	Sí	Sí	No	No	No
17	16	Trabajador 16	Área de Electricidad	05/08/2024	Sí	Sí	Sí	No	No	Sí
18	17	Trabajador 17	Área de Supervisión	07/08/2024	Sí	Sí	Sí	No	No	No
19	18	Trabajador 18	Área de Triturado	28/08/2024	Sí	Sí	Sí	Sí	Sí	Sí
20	19	Trabajador 19	Área de Pintura	22/08/2024	Sí	No	Sí	Sí	No	Sí
21	20	Trabajador 20	Área de Triturado	15/08/2024	Sí	Sí	Sí	Sí	Sí	Sí
22	21	Trabajador 21	Área de Pintura	06/08/2024	Sí	No	Sí	Sí	No	Sí
23	22	Trabajador 22	Área de Excavación	31/08/2024	Sí	Sí	Sí	No	Sí	Sí
24	23	Trabajador 23	Área de Triturado	19/08/2024	No	No	Sí	No	Sí	No

Figura 4. Dataset EPP

A continuación, se detalla la estructura y los tipos de datos incluidos en el dataset.

2.1. Descripción del dataset

Columnas y Tipos de Datos

1. ID del Trabajador

- **Tipo de Dato:** Entero
- **Descripción:** Identificador único asignado a cada trabajador para facilitar su seguimiento y análisis.

2. Nombre del Trabajador

- **Tipo de Dato:** Cadena de Texto
- **Descripción:** Nombre completo del trabajador evaluado, utilizado para identificar al individuo.

3. Área

- **Tipo de Dato:** Cadena de Texto
- **Descripción:** Área o sección del lugar de trabajo en la que se realiza la evaluación, permitiendo analizar el cumplimiento en diferentes zonas.

4. Fecha de Evaluación

- **Tipo de Dato:** Fecha
- **Descripción:** Fecha en la que se lleva a cabo la evaluación del uso del EPP, esencial para el registro temporal de las inspecciones.

5. Casco

- **Tipo de Dato:** Binario (1/0)
- **Descripción:** Indica si el trabajador llevaba casco durante la evaluación. (1 representa que sí, 0 que no).

6. Chaleco Reflectante

- **Tipo de Dato:** Binario (1/0)
- **Descripción:** Indica si el trabajador llevaba chaleco reflectante durante la evaluación. (1 representa que sí, 0 que no).

7. Botas de Seguridad

- **Tipo de Dato:** Binario (1/0)
- **Descripción:** Indica si el trabajador llevaba botas de seguridad durante la evaluación. (1 representa que sí, 0 que no).

8. Protección Ocular

- **Tipo de Dato:** Binario (1/0)
- **Descripción:** Indica si el trabajador llevaba protección ocular durante la evaluación. (1 representa que sí, 0 que no).

9. Protección Auditiva

- **Tipo de Dato:** Binario (1/0)
- **Descripción:** Indica si el trabajador llevaba protección auditiva durante la evaluación. (1 representa que sí, 0 que no).

10. Guantes de Seguridad

- **Tipo de Dato: Binario (1/0)**
- **Descripción:** Indica si el trabajador llevaba guantes de seguridad durante la evaluación. (1 representa que sí, 0 que no).

2.2. Datos Simulados

Para la simulación, se utilizarán valores binarios (1 y 0) para las columnas relacionadas con el uso del EPP. Esta metodología de simulación ofrece varias ventajas:

- **Simplicidad en el Análisis:** Los datos binarios facilitan la aplicación de técnicas estadísticas y algoritmos de análisis, ya que los valores discretos simplifican el procesamiento y la interpretación.
- **Consistencia y Comparabilidad:** La uniformidad de los valores binarios asegura que los datos sean consistentes y comparables entre diferentes registros y evaluaciones.
- **Facilidad de Generación:** Los datos binarios pueden ser fácilmente generados y manipulados, lo que permite la creación de grandes volúmenes de datos de manera rápida y eficiente.
- **Aplicabilidad en Modelos de Machine Learning:** Los valores binarios son ideales para entrenar modelos de machine learning, facilitando la clasificación y predicción basada en el cumplimiento de las normativas de seguridad.

2.3. Cantidad de Datos Requerida

Para realizar un análisis estadístico sólido y obtener resultados representativos, se generarán entre 200 y 500 registros de evaluaciones. Esta cantidad es adecuada para proporcionar una muestra significativa y útil para el análisis.

2.4. Fuente de los Datos

Los datos serán generados a través de simulaciones (a recomendación del TA) para reflejar (en medida de lo posible) escenarios realistas. Alternativamente, se podrán utilizar registros de evaluaciones reales si están disponibles, con el fin de garantizar la aplicabilidad de los resultados.

3. ANÁLISIS DE SESGOS EN LOS DATOS

3.1. Identificación de Posibles Sesgos

a) Sesgo de Representación

Descripción: El dataset puede no representar adecuadamente todas las áreas o secciones del lugar de trabajo. Si algunos áreas tienen más registros que otras, esto puede llevar a un sesgo en la evaluación del cumplimiento del equipo de protección personal (EPP).

Impacto: Esto podría resultar en una percepción incorrecta de la efectividad de las políticas de seguridad en áreas menos representadas.

b) Sesgo de Compleción

Descripción: La falta de datos en ciertas columnas (por ejemplo, datos incompletos sobre el uso del EPP) puede introducir sesgos si ciertos tipos de protección tienen más probabilidades de estar ausentes en las evaluaciones.

Impacto: Puede afectar la exactitud del análisis del cumplimiento de las normativas de seguridad.

c) Sesgo de Simulación

Descripción: Al utilizar datos simulados con valores binarios (1 y 0), se pueden introducir sesgos inherentes si la simulación no refleja adecuadamente la variabilidad real en el uso del EPP.

Impacto: Esto puede llevar a una representación poco realista del cumplimiento del EPP y afectar la validez del análisis.

d) Sesgo de Identificación

Descripción: La asignación de nombres y IDs de trabajadores puede ser incorrecta o no reflejar adecuadamente la diversidad del personal. Esto puede ser un problema si los datos de ciertos grupos de trabajadores están sobrerrepresentados o subrepresentados.

Impacto: Puede sesgar los resultados al no reflejar la distribución real de los trabajadores y sus prácticas de seguridad.

e) Sesgo de Temporalidad

Descripción: Las evaluaciones realizadas en ciertos períodos del año (por ejemplo, durante épocas de alta actividad o mantenimiento) pueden no ser representativas de la norma.

Impacto: Puede llevar a conclusiones erróneas sobre el cumplimiento del EPP en diferentes épocas o condiciones de trabajo.

3.2. Estrategias para Mitigar los Sesgos

1. Asegurar la Representatividad de las Áreas

Estrategia: Diseñar el dataset para incluir registros equitativos de todas las áreas y secciones del lugar de trabajo. Si es necesario, ajustar el número de registros en cada área para asegurar que el dataset sea representativo.

2. Gestionar la Compleción de Datos

Estrategia: Implementar procesos rigurosos para la recolección de datos y asegurarse de que todas las columnas estén completas en la medida de lo posible. Utilizar técnicas de imputación para manejar los datos faltantes y garantizar que no introduzcan sesgos en el análisis.

3. Refinar la Simulación de Datos

Estrategia: Validar que los datos simulados reflejan adecuadamente la variabilidad y los escenarios reales. Utilizar estadísticas y análisis exploratorios para ajustar la simulación y asegurarse de que los datos sean representativos de las condiciones reales.

4. Diversificar la Identificación de Trabajadores

Estrategia: Asegurarse de que el dataset incluya una muestra diversa de trabajadores y verificar que los IDs y nombres reflejen adecuadamente la composición del personal. Realizar revisiones periódicas para corregir cualquier sesgo en la identificación.

5. Equilibrar la Temporalidad

Estrategia: Recopilar datos durante diferentes períodos del año y en diferentes condiciones de trabajo para obtener una imagen completa del cumplimiento del EPP. Analizar los datos en función de la temporalidad para identificar y ajustar cualquier sesgo relacionado con el tiempo.

3.3. Garantía de Equidad y Confiabilidad del Modelo

Para asegurar la equidad y confiabilidad del modelo, es esencial llevar a cabo una revisión continua del dataset y de las estrategias de mitigación de sesgos. Además, implementar técnicas de validación cruzada y análisis de sensibilidad ayudará a verificar la robustez del modelo y la precisión de las conclusiones derivadas del análisis de datos.

4. SELECCIÓN DEL TIPO DE APRENDIZAJE

El tipo de aprendizaje seleccionado es Aprendizaje Supervisado

4.1. Justificación del tipo de aprendizaje a utilizar

Razón de la Selección:

1. Naturaleza del Problema:

El problema requiere clasificar si un trabajador está utilizando correctamente el equipo de protección personal (EPP). Esto implica asignar etiquetas a los datos (por ejemplo, si el trabajador lleva casco, chaleco reflectante, etc.) basadas en ejemplos etiquetados previamente.

2. Datos Disponibles:

El dataset teórico contiene ejemplos etiquetados de uso del EPP con valores binarios (1 y 0) que indican si el trabajador lleva cada tipo de equipo de protección. El aprendizaje supervisado es adecuado para trabajar con datos etiquetados como estos, ya que el modelo puede aprender a partir de ejemplos explícitos de cumplimiento y no cumplimiento.

3. Objetivo de la Predicción:

El objetivo es predecir la presencia o ausencia de equipo de protección personal para cada trabajador en base a características observadas (por ejemplo, la presencia de casco, chaleco, botas, etc.). El aprendizaje supervisado es ideal para esta tarea de clasificación, donde el modelo aprende a partir de datos históricos para hacer predicciones sobre nuevos datos.

4. Modelos y Algoritmos Adecuados:

Existen numerosos algoritmos de aprendizaje supervisado que son adecuados para problemas de clasificación binaria, como regresión logística, máquinas de soporte vectorial (SVM), árboles de decisión, y redes neuronales. Estos modelos pueden ser entrenados con los datos disponibles para aprender a identificar patrones en el uso del EPP.

5. Evaluación y Validación:

Con el aprendizaje supervisado, es posible evaluar el rendimiento del modelo utilizando métricas como precisión, recall, y la curva ROC, ya que se tienen datos etiquetados para comparar las predicciones del modelo con la realidad.

Esto permite una validación clara y cuantificable del modelo.

Después de una cuidadosa selección el aprendizaje supervisado es el enfoque más adecuado para entrenar el modelo de IA en este caso debido a la disponibilidad de datos etiquetados y la necesidad de clasificar si los trabajadores están usando correctamente el EPP. Este enfoque permite utilizar los ejemplos históricos para enseñar al modelo a realizar predicciones precisas en nuevas situaciones.

5. SELECCIÓN DEL MODELO O MODELOS

5.1. Descripción del Modelo o Modelos

Para la predicción de cumplimiento de los requisitos de seguridad en el lugar de trabajo, se seleccionó un modelo de clasificación basado en Árbol de Decisión (Decision Tree Learner). Este modelo es particularmente adecuado para tareas donde se requiere una interpretación clara de las reglas que determinan la clasificación.

5.2. Justificación de la Elección

El Árbol de Decisión es un modelo de fácil interpretación y visualización, lo que es crucial para entender las decisiones sobre el cumplimiento de los requisitos de seguridad. Además, permite manejar eficientemente datos categóricos y numéricos, como los indicadores binarios utilizados en este proyecto.

El Árbol de Decisión se eligió por las siguientes razones:

- **Interpretabilidad:** Proporciona una visualización clara y comprensible de las decisiones y reglas utilizadas para la clasificación.
- **Eficiencia:** Es rápido de entrenar y aplicar, lo cual es ideal para el entorno de trabajo donde se requiere una evaluación rápida del cumplimiento de seguridad.
- **Manejo de Datos Binarios:** El modelo maneja de manera efectiva los datos binarios (1 y 0) que representan los requisitos de seguridad.

5.3. Explicación del Funcionamiento

El modelo de Árbol de Decisión divide iterativamente el conjunto de datos en subconjuntos basándose en características específicas, creando una estructura de árbol donde cada nodo representa una característica y cada rama una decisión. Esto facilita la predicción de si un trabajador cumple con los requisitos de seguridad necesarios para ingresar a una área de trabajo específica.

Predicción: El objetivo principal del modelo es determinar si un trabajador cumple con los requisitos de seguridad necesarios para entrar a un área específica de trabajo. Esto es crucial para garantizar la seguridad en el lugar de trabajo y minimizar el riesgo de accidentes.

Datos de Entrada

Área de Trabajo: La zona específica donde se realizará el trabajo (por ejemplo, Área de Triturado, Área de Supervisión, etc.). Diferentes áreas de trabajo pueden tener diferentes requisitos de seguridad. Es importante identificar la zona para aplicar las reglas de seguridad adecuadas.

Requisitos de Seguridad: Indicadores binarios (1 para Sí, 0 para No) que muestran si el trabajador cumple con los requisitos de seguridad como Casco, Chaleco Reflectante, Botas de Seguridad, Protección Ocular, Protección Auditiva y Guantes de Seguridad. Estos requisitos son esenciales para la seguridad del trabajador. Usar indicadores binarios facilita el procesamiento y la evaluación de cumplimiento.

6. PROCESO DEL MODELO

Para desarrollar el modelo para el análisis de equipo de protección personal y prevención de accidentes se utilizaron los siguientes nodos:



Excel Reader

Nodo que permite leer archivos Excel (xlsx, xlsxm, xlsb y xls). Puede leer uno o múltiples archivos a la vez, pero solo una hoja por archivo. Los tipos de datos Excel compatibles que se pueden leer son texto, número, booleano, fecha y hora, pero no imágenes, diagramas, etc.



Statistics View

Nodo que sirve para realizar un análisis exploratorio de datos de manera eficiente, permite obtener una visión general rápida de los datos y tomar decisiones informadas sobre los siguientes pasos del análisis.

► **String Manipulation**

Nodo que permite preparar los datos de manera adecuada para análisis posterior, lo que ayudará a obtener resultados más precisos y significativos.

► **Missing Value**

Nodo que sirve para garantizar la calidad de los datos y obtener resultados de análisis más precisos, permite identificar y tratar los valores faltantes de manera efectiva, lo que es fundamental para cualquier proyecto de análisis de datos.

► **Column Filter**

Nodo que sirve para seleccionar y filtrar las columnas de un conjunto de datos. Este nodo permite elegir exactamente las columnas que se quiere conservar y aquellas que se desean eliminar.

► **Normalizer**

Nodo que sirve para preparar los datos numéricos para el análisis y el modelado. Al normalizar los datos, se puede mejorar la calidad y la precisión de los resultados.

► **Rule Engine**

Nodo que permite aplicar lógica condicional a los datos de una manera flexible y eficiente. Es como tener una hoja de cálculo gigante donde se puede definir reglas personalizadas para manipular y transformar los datos.

► **Partitioning**

Nodo que sirve para preparar los datos para el análisis y el modelado. Al dividir los datos en subconjuntos, se puede obtener resultados más confiables y generalizables.

► **Table view**

Nodo que permite obtener una visión rápida de tus datos y realizar cambios menores si es necesario.



Decision Tree Learner

Nodo que sirve para para construir modelos de clasificación y regresión basados en árboles de decisión. Su facilidad de uso e interpretabilidad lo convierten en una excelente opción para muchos problemas de aprendizaje automático.



Decision Tree Predictor

Nodo que permite construir modelos de clasificación y regresión basados en árboles de decisión. Estos modelos son muy populares debido a su interpretabilidad y facilidad de uso.



Scorer

Nodo que sirve para evaluar el rendimiento de un modelo de machine learning. Una vez que se ha entrenado un modelo (por ejemplo, un árbol de decisión, una red neuronal, etc.), el nodo Scorer permite comparar las predicciones de ese modelo con los valores reales del conjunto de datos y obtener métricas que cuantifiquen su precisión.



ROC Curve

Nodo que permite evaluar y comparar modelos de clasificación binaria. Al proporcionar una visión visual del rendimiento del modelo en diferentes umbrales de clasificación, ayuda a tomar decisiones más informadas sobre la elección del modelo y la configuración de su umbral de decisión



Scatter Plot

Nodo que permite visualizar la relación entre dos variables numéricas. Al representar cada punto de datos como un punto en un plano cartesiano, este tipo de gráfico permite identificar patrones, tendencias y posibles correlaciones entre las variables.

6.1. Lectura de Datos

Se leen los datos de entrada que contienen la información sobre las áreas de trabajo y los requisitos de seguridad utilizando el nodo **Excel Reader**. Este es el primer paso para obtener los datos necesarios para el análisis y la predicción. En la figura 6, se muestra como cargar el archivo en el nodo Excel Reader.

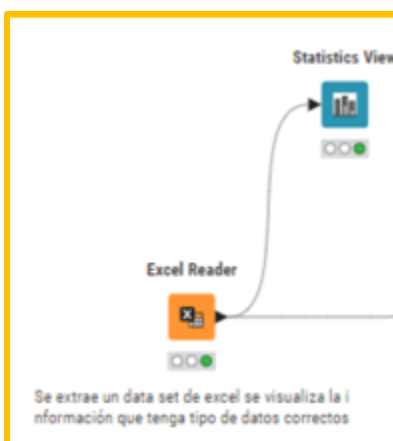


Figura 5. Lectura de datos

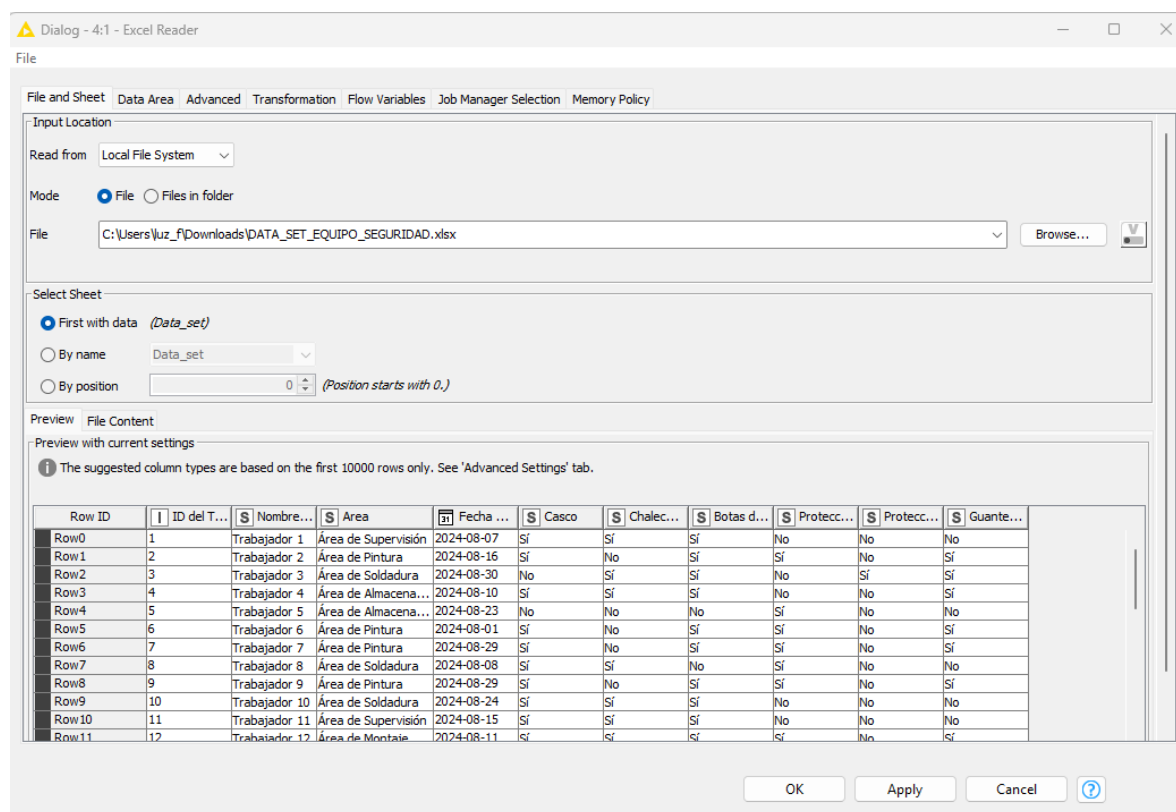


Figura 6. Cargar dataset

6.2. Limpieza y Preparación de Datos

- **Conversión de Datos:** Se convierten los datos de string a numéricos (1 y 0) para facilitar el procesamiento. Los datos en el dataset se presentan como "Sí" y "No", por lo que se utilizó el nodo **String Manipulation** con una función replace para cambiar los "Sí" por 1 y los "No" por 0. Posteriormente, se utilizó el nodo **String to Number** para asegurar la congruencia en los tipos de datos.
- **Manejo de Valores Faltantes:** Se utilizó el nodo **Missing Value** para reemplazar valores vacíos por 0 o 1, asegurando que todos los datos sean compatibles con el modelo.
- **Filtrado de Columnas:** Se utilizó el nodo **Column Filter** para seleccionar solo las columnas relevantes que van a entrar al modelo, eliminando aquellas que no son necesarias para la predicción.

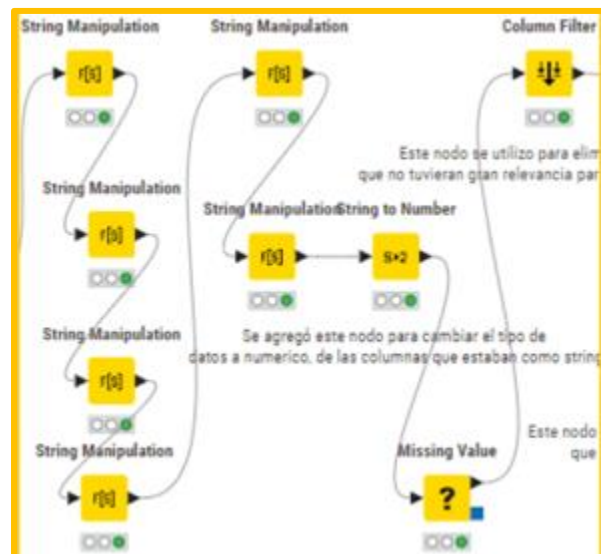


Figura 7. Limpieza y preparación de datos

En la figura 8. se observa el código que se incluyó para la conversión de datos de tipo String a numérico.

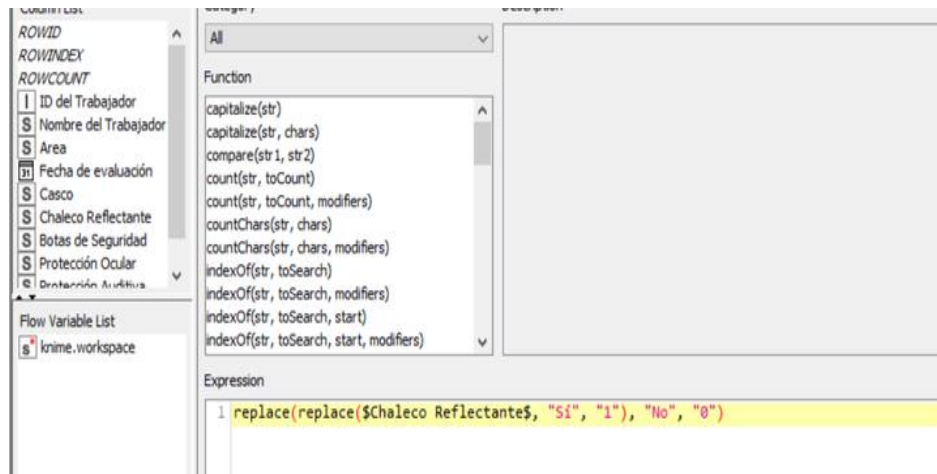


Figura 8. Conversión de datos

En la Figura 9 se muestra el filtrado de columnas relevantes para el análisis del modelo.

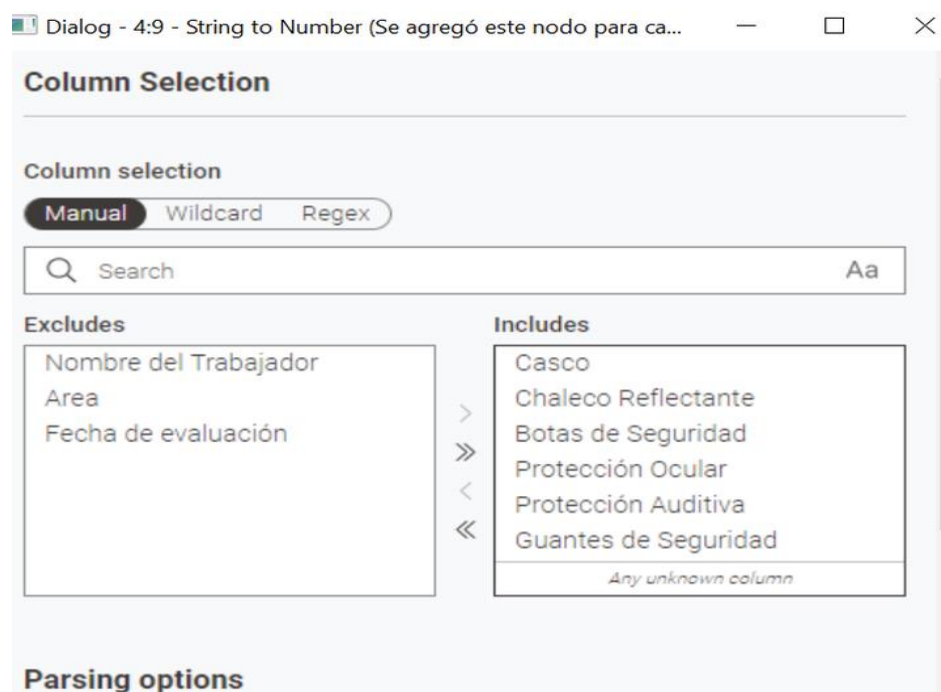


Figura 9. Filtrado de columnas

6.3. Aplicación de Reglas y Normalización

- **Aplicación de Reglas:** Se utilizan reglas predefinidas en el nodo **Rule Engine** para determinar si un trabajador cumple con los requisitos de seguridad en función de los parámetros definidos para cada área.

- **Normalización:** Se aplicó un nodo **Normalizer** para asegurar que los valores en el dataset sean solo 1 y 0, facilitando el procesamiento.

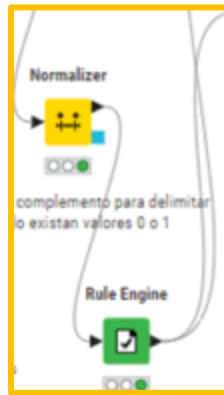


Figura 10. Aplicación de reglas y normalización

En la figura 11 se muestra el código que se agregaron reglas predefinidas en el nodo de rule engine.

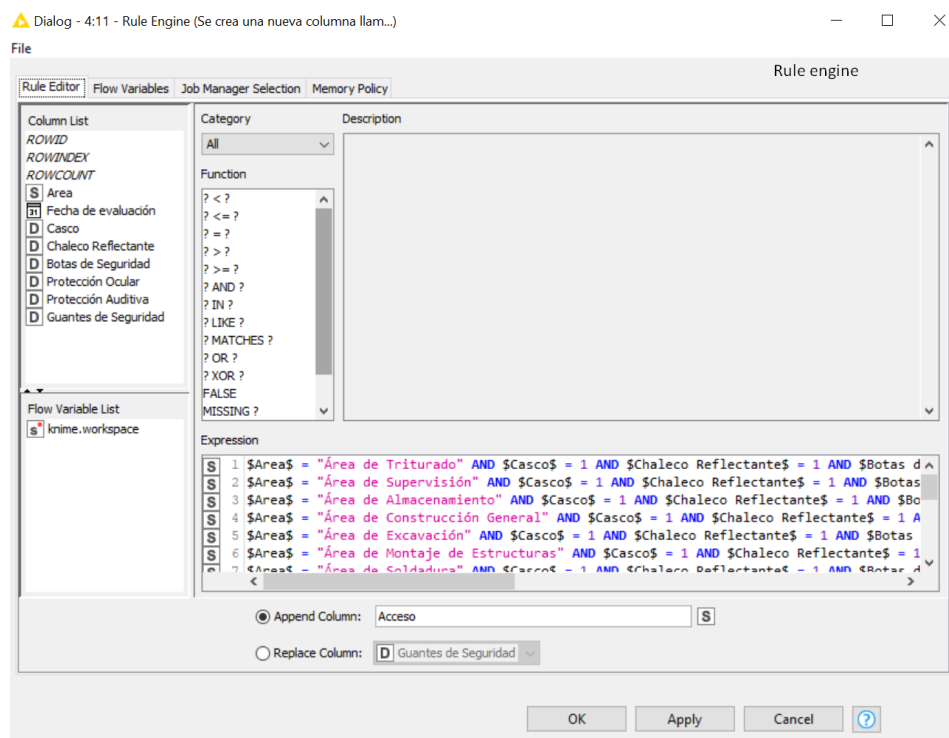


Figura 11. Aplicación de reglas

En la figura 12 se muestra la configuración del nodo Normalizer.

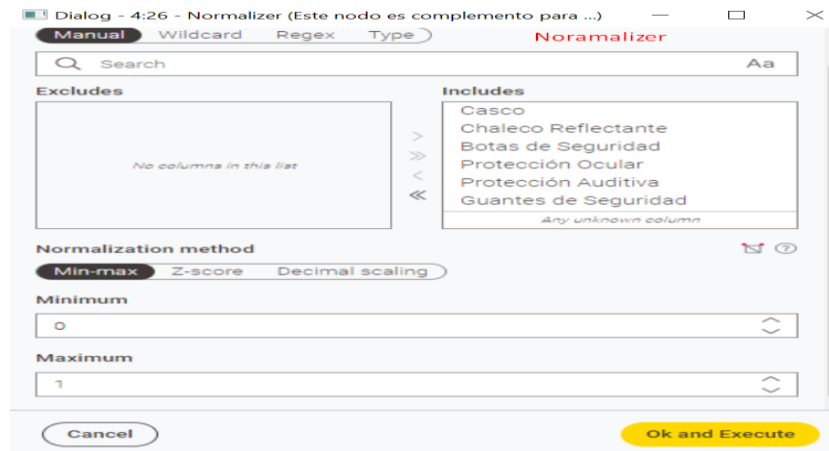


Figura 12. Normalización

6.4. Modelo de aprendizaje

Para el modelo de aprendizaje se realizó:

- **Partición de Datos:** Se dividen los datos en conjuntos de entrenamiento y prueba utilizando el nodo **Partitioning**. Dividir los datos permite entrenar el modelo en un subconjunto y evaluar su rendimiento en otro, asegurando que el modelo generalice bien a datos no vistos.
- **Entrenamiento del Modelo:** Se entrenó un modelo de clasificación (Decision Tree Learner) utilizando los datos de entrenamiento. Esto permite que el modelo aprenda las relaciones entre los datos de entrada y la variable objetivo (cumplimiento de seguridad).
- **Predicción:** Se aplican las predicciones del modelo a los datos de prueba utilizando el nodo **Decision Tree Predictor**. Evaluar el modelo en datos no vistos permite medir su rendimiento y ajustar parámetros si es necesario.



Figura 13. Modelo de aprendizaje

Para el nodo de Partitioning se conservó el 70/30 como se observa en la figura 14.

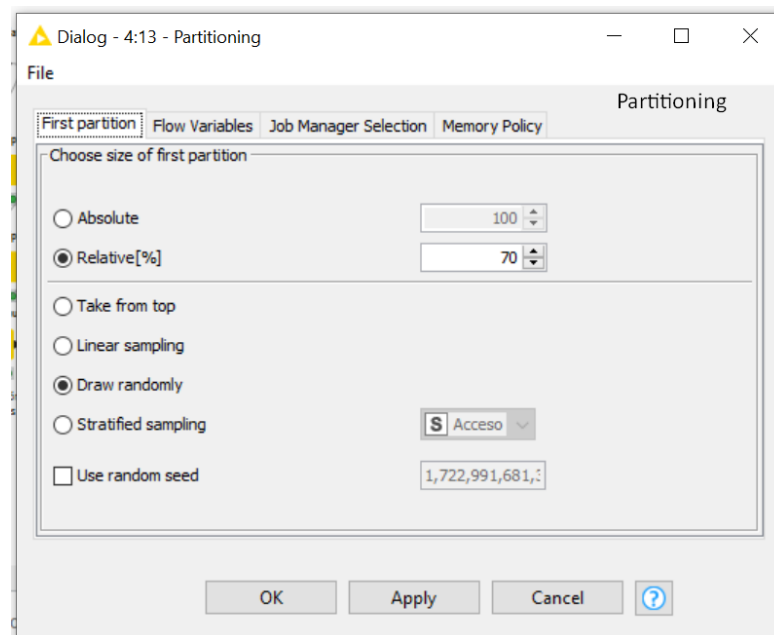


Figura 14. Partición de datos

La configuración del nodo Decision Tree Learner se muestra en la figura 15, dónde solo se selecciona la columna a predecir en este caso fue la de acceso y se dejan las configuraciones por default, y en la **¡Error! No se encuentra el origen de la referencia.** la configuración del nodo Decision Tree Predictor.

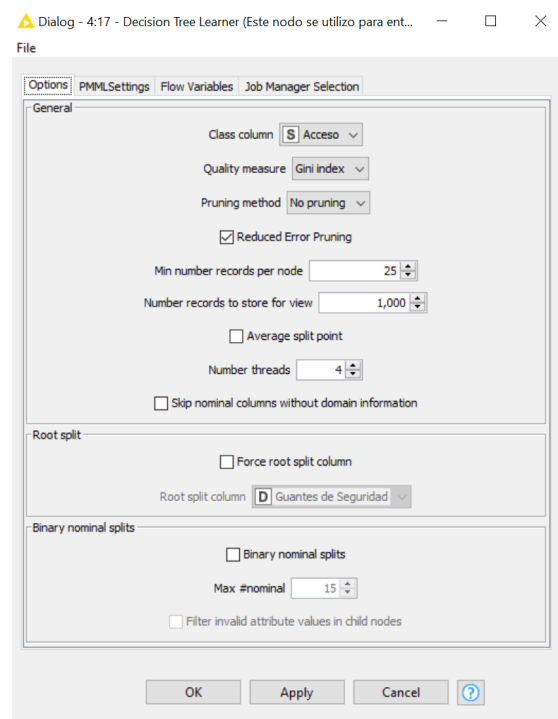


Figura 15. Entrenamiento del modelo

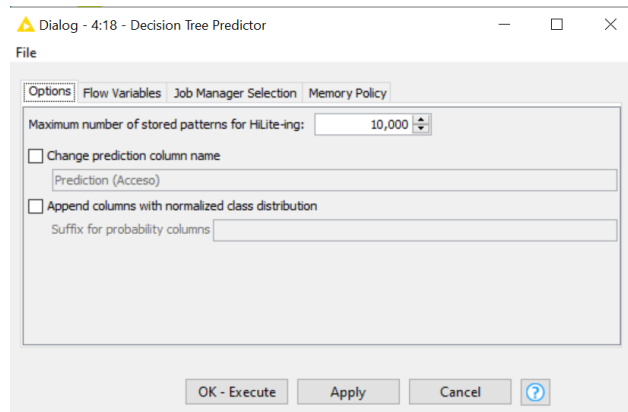


Figura 16. Predicción

En la figura 17 table view que permite ir validando que los datos fueran correctos.

Table View

Rows: 500 | Columns: 9

RowID	Area	Fecha de ...	Casco	Chaleco ...	Botas de ...	Prot
String	Local Date	Number (dou...	Number (dou...	Number (dou...	Numl	
Row0	Área de Supe...	2024-08-07	1	1	1	0
Row1	Área de Pintu...	2024-08-16	1	0	1	1
Row2	Área de Sold...	2024-08-30	0	1	1	0
Row3	Área de Alma...	2024-08-10	1	1	1	0
Row4	Área de Alma...	2024-08-23	0	0	0	1
Row5	Área de Pintu...	2024-08-01	1	0	1	1
Row6	Área de Pintu...	2024-08-29	1	0	1	1
Row7	Área de Sold...	2024-08-08	1	1	0	1
Row8	Área de Pintu...	2024-08-29	1	0	1	1
Row9	Área de Sold...	2024-08-24	1	1	1	0
Row...	Área de Supe...	2024-08-15	1	1	1	0
Row...	Área de Mont...	2024-08-11	1	1	1	1

Data

Displayed columns

Manual Wildcard Regex Type

Search Aa

Excludes

No columns in this list

Any unknown column

Includes

(MISSING) C...
(MISSING) L...
(MISSING) N...
Area
Fecha de ev...
Casco
Chaleco Refl...

☐ Show row numbers
☒ Show RowIDs

View

Title

Table View

☒ Show table size
☒ Show column data types in header

Cancel Ok and Execute

WARNING The selected columns Clasificación, ID del Trabajador, Nombre del Trabajador are not present in the table.

Figura 17.Resultado del table view

6.5. Métricas y Evaluación

Se evaluó el rendimiento del modelo utilizando métricas como True Positives, False Positives, True Negatives, False Negatives, Recall, Precision, Specificity, F-measure y Cohen's Kappa utilizando el nodo **Scorer**. Estas métricas proporcionan una visión completa del rendimiento del modelo, ayudando a identificar áreas de mejora y asegurar que el modelo sea fiable.

Salida del Modelo

Resultado: El modelo predice si un trabajador puede entrar (Aprobado) o no (No Aprobado) a un área de trabajo específica basada en el cumplimiento de los requisitos de seguridad. Esto ayuda a tomar decisiones informadas sobre la seguridad en el lugar de trabajo.

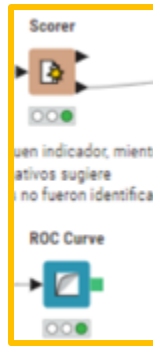


Figura 18. Métricas y evaluación

En la figura 19 se muestra la configuración del nodo Scoper, mientras que en la Figura 20 se observa la matriz de confusión.

Dialog - 4:19 - Scorer (La ausencia de falsos positivo...)

File

Scorer Flow Variables Job Manager Selection Memory Policy

First Column
[S] Acceso

Second Column
[S] Prediction (Acceso)

Sorting of values in tables
Sorting strategy: Insertion order ☐ Reverse order

Provide scores as flow variables
☐ Use name prefix

Missing values
In case of missing values: ☒ Ignore ☐ Fail

OK - Execute Apply Cancel ?

Figura 19. Scoper

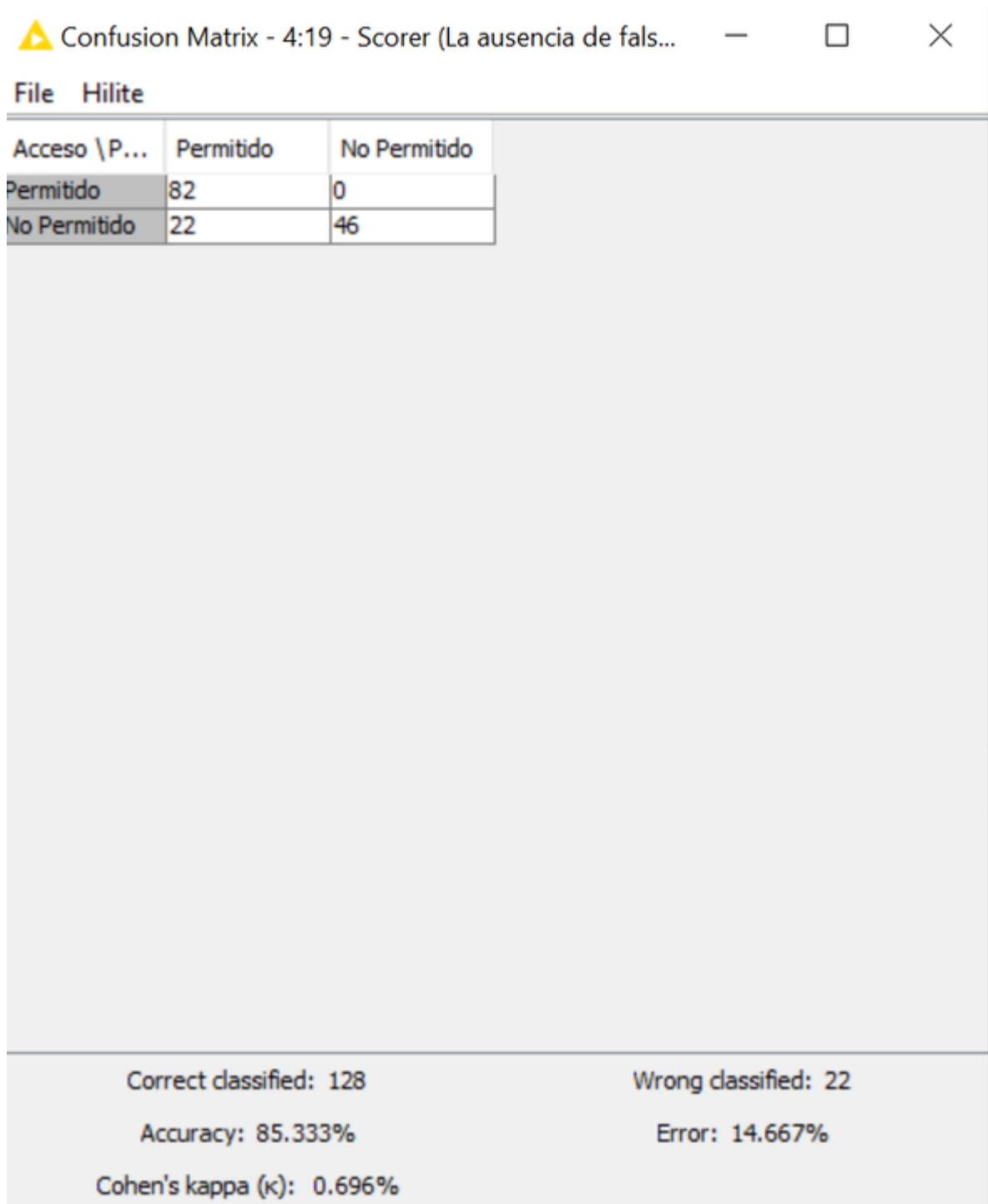


Figura 20. Matriz de confusión

6.6. Visualización de Datos

- **Scatter Plot:** Se utiliza un nodo **Scatter Plot** para visualizar la relación entre las predicciones del modelo y los valores reales del criterio para entrar a una cierta área. Este gráfico es útil para identificar patrones y errores en las predicciones del modelo.
- **Pie Chart:** Se utiliza un nodo **Pie Chart** para visualizar la distribución de los trabajadores que cumplen con los requisitos de seguridad en comparación con aquellos que no lo hacen. Este gráfico es útil para entender la proporción de cumplimiento de las normas de seguridad.

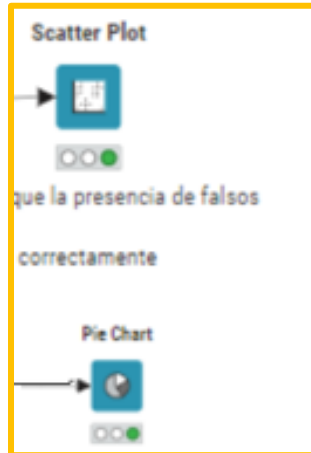


Figura 21. Visualización de datos

6.7. Definición del Método de Deployment

Para el sistema de IA destinado a la evaluación del cumplimiento del equipo de protección personal (EPP) en entornos donde la conexión a internet puede ser limitada o inexistente, se propone la implementación local en servidores on-premises. Este enfoque garantiza que el sistema funcione de manera eficiente y segura, incluso en ubicaciones remotas con conectividad restringida.

Escalabilidad

La implementación local puede ser escalada mediante el uso de servidores adicionales o más potentes según la demanda. Es posible configurar una infraestructura de servidores locales que soporte la carga actual y futura, asegurando que el sistema pueda manejar un aumento en la cantidad de trabajadores o en la frecuencia de evaluaciones.

Seguridad

La implementación local permite un control total sobre los datos y el acceso a los mismos. Los servidores locales pueden ser configurados con medidas de seguridad robustas, como cortafuegos, cifrado de datos y controles de acceso estrictos. Además, al no depender de una conexión a internet, se reduce el riesgo de ciberataques externos.

Facilidad de Uso

Para facilitar el uso, el sistema se puede desplegar en una interfaz local accesible desde dispositivos dentro de la red de la obra. Se pueden utilizar tecnologías como contenedores (por ejemplo, Docker) para simplificar la implementación y mantenimiento del software, asegurando que las actualizaciones y configuraciones sean manejables incluso sin conexión a internet.

Justificación de la Elección

- **Escalabilidad:** La solución local permite ajustar y expandir la infraestructura de hardware según las necesidades, asegurando que el sistema pueda crecer junto con los requerimientos de la obra.
- **Seguridad:** La implementación on-premises garantiza que todos los datos sensibles se mantengan dentro de la infraestructura controlada por la empresa, reduciendo la exposición a amenazas externas y permitiendo el cumplimiento de normativas de seguridad específicas.

- **Facilidad de Uso:** El uso de servidores locales y tecnologías de contenedores facilita la instalación, configuración y mantenimiento del sistema. Los trabajadores pueden acceder al sistema mediante dispositivos conectados a la red local de la obra, garantizando un acceso rápido y confiable a las evaluaciones de seguridad.
- **Conectividad Limitada:** En muchos entornos de obra, la conexión a internet puede ser limitada o incluso inexistente. La implementación local asegura que el sistema de IA funcione sin depender de la conectividad externa, eliminando el riesgo de interrupciones en el servicio debido a problemas de red. Esto es crucial para garantizar que las evaluaciones de seguridad se realicen de manera continua y eficiente, independientemente de las condiciones de la conectividad.

7. RESULTADOS

En la figura 22, dónde se observa el modelo del proceso para predecir la presencia o ausencia de equipo de protección personal para cada trabajador en base a características observadas.

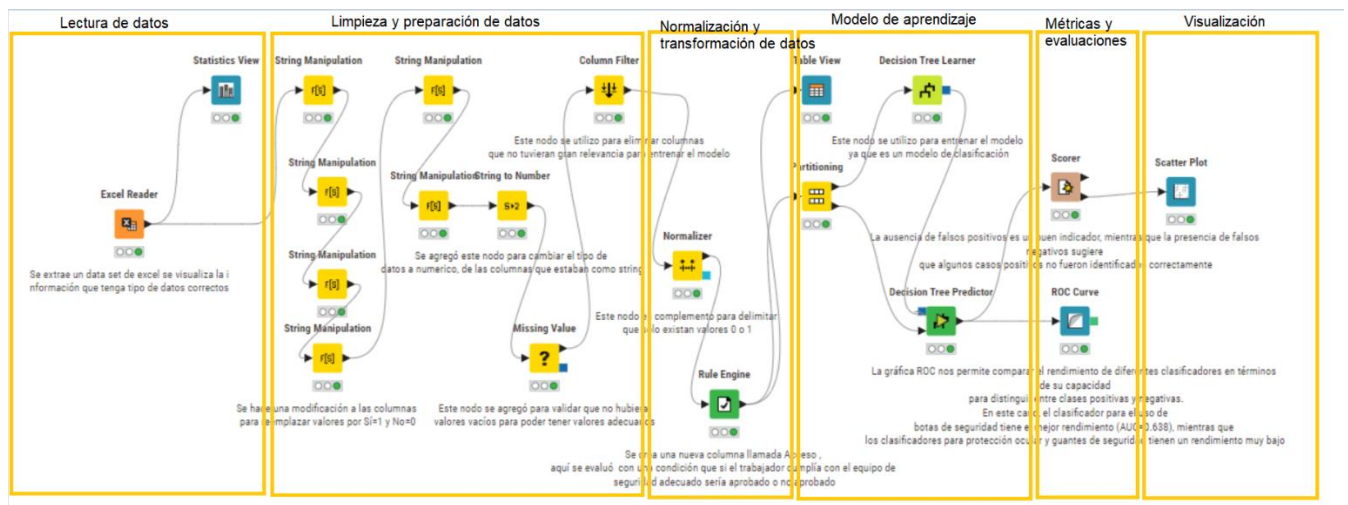


Figura 22. Modelo para el análisis de EPP.

El modelo presenta un buen rendimiento en la identificación de trabajadores que cumplen con los requisitos de seguridad (clase "Permitido"). Sin embargo, se observan falsos negativos en la clase "No Permitido" (Ver figura 20), lo que sugiere una necesidad de mejora en la detección de aquellos que no cumplen con los requisitos.

Optimización de la Clase "No Permitido": Ajustar el balance de clases y los umbrales de decisión para mejorar la detección de casos negativos.

Reevaluación de Características: Revisar las características del modelo para mejorar su capacidad de generalización.

Monitoreo Continuo: Implementar un sistema de monitoreo continuo y reentrenamiento periódico para mantener el rendimiento del modelo. El modelo de predicción demuestra ser efectivo en la mayoría de los casos, con un buen equilibrio entre precisión y recall para la clase "Permitido". Las mejoras sugeridas permitirán aumentar la robustez y fiabilidad del modelo, asegurando un cumplimiento más estricto de los requisitos de seguridad en las áreas de trabajo.

7.1. Resultado de la predicción

Precisión y Recall: El modelo tiene una alta precisión (0.788) y recall perfecto (1) para la clase "Permitido", lo cual indica que el modelo es excelente en identificar correctamente los casos positivos y tiene un número muy bajo de falsos negativos. Sin embargo, hay cierta proporción de falsos positivos, lo que sugiere una sobreestimación en algunos casos.

Especificidad y Sensibilidad: La sensibilidad perfecta (1) para la clase "Permitido" indica que el modelo identifica todos los casos positivos correctamente. La especificidad de 0.676 para la misma clase muestra que hay margen de mejora en la identificación de casos negativos.

F-measure: La F-measure alta (0.882) para la clase "Permitido" confirma que el modelo balancea bien la precisión y el recall para esta clase. Para la clase "No Permitido", aunque la precisión es perfecta, la F-measure más baja (0.807) sugiere que el modelo no es tan efectivo en esta clase comparativamente.

Cohen's kappa: Un valor de 0.696 indica un buen acuerdo entre las predicciones del modelo y los valores reales, aunque hay margen para mejorar la consistencia del modelo.

Rows: 3 | Columns: 11

Table

Statistics

RowID	TruePosit... Number (inte...	FalsePosi... Number (inte...	TrueNega... Number (inte...	FalseNeg... Number (inte...	Recall Number (dou...	Precision Number (dou...	Sensitivity Number (dou...	Specificity Number (dou...	F-measure Number (dou...	Accuracy Number (dou...	Cohen's k... Number (dou...
Permitido	82	22	46	0	1	0.788	1	0.676	0.882		
No Permitido	46	0	82	22	0.676	1	0.676	1	0.807		
Overall										0.853	0.696

Figura 23.Resultados

CONCLUSIONES

La implementación de un sistema de visión artificial para monitorear el uso del equipo de protección personal (EPP) y prevenir la activación de maquinaria durante actividades de mantenimiento puede ofrecer mejoras significativas en la seguridad laboral. A lo largo del desarrollo de este proyecto, se ha demostrado cómo la simulación y el análisis de datos pueden ser utilizados para crear un modelo efectivo que identifique el cumplimiento de las normativas de seguridad.

Principales conclusiones del proyecto:

Eficacia del Uso de la Visión Artificial: La visión artificial tiene el potencial de automatizar la supervisión del uso del EPP y detectar situaciones de riesgo en tiempo real. Aunque el proyecto utilizó datos simulados, los resultados sugieren que un sistema de este tipo podría reducir la incidencia de accidentes laborales.

Importancia del Dataset y la Mitigación de Sesgos: La calidad y representatividad del dataset son cruciales para el éxito del modelo. Se identificaron, más no se mitigaron posibles sesgos, debido a que son datos simulados.

Selección de Modelos de Aprendizaje: La elección del modelo de Árbol de Decisión fue adecuada debido a su interpretabilidad y eficiencia en el manejo de datos binarios. Este modelo permitió crear reglas claras y comprensibles para evaluar el cumplimiento del EPP.

Evaluación del Rendimiento del Modelo: El rendimiento del modelo se evaluó utilizando métricas estándar, demostrando su capacidad para etiquetar el cumplimiento del EPP. Las técnicas de validación cruzada y el análisis de sensibilidad ayudaron a asegurar la robustez y confiabilidad del modelo.

Implementación y Aplicabilidad: La implementación local del sistema es viable, especialmente en entornos con conectividad limitada, como obras de construcción. Esto permite un monitoreo continuo y eficiente sin depender de una conexión constante a internet.

Recomendaciones para Futuras Implementaciones:

Ampliación del Dataset: Utilizar datos reales de evaluaciones de EPP en lugar de datos simulados para mejorar la precisión y aplicabilidad del modelo.

Creación de dataset de salida para verificar qué áreas fueron más propensas a sufrir alguna activación accidental y que haya sido evitada con visión artificial, por ejemplo: “Se detectó que en pintura hubo tres intentos de activación y afortunadamente el sistema de monitoreo no permitió esta acción”, entonces se podría hacer una capacitación para evitar futuros accidentes.

Integración con Sistemas de Seguridad: Integrar el sistema de visión artificial con otros sistemas de seguridad y monitoreo para crear un entorno de trabajo más seguro y eficiente. Por ejemplo:

- **Capacitación y Concienciación:** Implementar programas de capacitación continua para los trabajadores sobre la importancia del EPP y el uso adecuado de los equipos de seguridad.

Impacto Global:

La adopción de tecnologías emergentes como la visión artificial representa un avance significativo en la gestión de la seguridad laboral. Este sistema no solo mejora la vigilancia y el cumplimiento del uso del EPP, sino que proporciona datos valiosos para la mejora continua de los protocolos de seguridad. Aunque el proyecto se basó en datos simulados, se puede vislumbrar el potencial de la visión artificial para reducir la incidencia de accidentes laborales, destacando la importancia de utilizar datasets representativos y de alta calidad en el desarrollo y entrenamiento de estos sistemas.

El uso de un dataset de salida, que permite identificar áreas con mayor frecuencia de intentos de activación accidental evitados, justifica y guía la creación de programas de capacitación específicos. Esto asegura que el sistema no sustituye las medidas de seguridad existentes, sino que las complementa, creando un entorno de trabajo más seguro y eficiente. La visión artificial, al integrar estos datos, puede transformar la prevención de accidentes y fortalecer la cultura de seguridad, ofreciendo una herramienta valiosa para proteger mejor a los trabajadores y promover prácticas más seguras y conscientes en el entorno laboral.

BIBLIOGRAFÍA

- [1] Organización Internacional del Trabajo. (s.f.). Estadísticas de Accidentes Laborales. Recuperado de <https://www.ilo.org/global/statistics-and-databases/lang-es/index.htm>
- [2] Occupational Safety and Health Administration. (s.f.). Datos y Estadísticas. Recuperado de <https://www.osha.gov/data>
- [3] Bureau of Labor Statistics. (s.f.). Injuries, Illnesses, and Fatalities. Recuperado de <https://www.bls.gov/iif/>
- [4] National Institute for Occupational Safety and Health. (s.f.). Workplace Safety and Health Topics. Recuperado de <https://www.cdc.gov/niosh/topics/default.html>
- [5] Organización Internacional del Trabajo. (s.f.). Estadísticas de Accidentes Laborales. Recuperado de <https://www.ilo.org/global/statistics-and-databases/lang-es/index.htm>
- [5] Occupational Safety and Health Administration. (s.f.). Datos y Estadísticas. Recuperado de <https://www.osha.gov/data>
- [6] Bureau of Labor Statistics. (s.f.). Injuries, Illnesses, and Fatalities. Recuperado de <https://www.bls.gov/iif/>
- [7] National Institute for Occupational Safety and Health. (s.f.). Workplace Safety and Health Topics. Recuperado de <https://www.cdc.gov/niosh/topics/default.html>
- [8] European Agency for Safety and Health at Work. (s.f.). Facts and Figures. Recuperado de <https://osha.europa.eu/en/facts-and-figures>