

## **Proyecto Fase 2**

### **Faltas judiciales**

Las faltas judiciales son conductas tipificadas por la ley que, si bien no constituyen delitos graves, implican una infracción al orden público y son sancionadas por autoridades judiciales o administrativas competentes. Estas faltas suelen derivarse de acciones que alteran la convivencia social o vulneran disposiciones normativas específicas, como el consumo de alcohol en la vía pública, riñas o desobediencia a la autoridad (Ministerio Público de Guatemala, 2023).

### **Análisis de información**

Se realizó un análisis basando en la información de las faltas judiciales del año 2018 al 2024, del cual únicamente se mantuvo información del 2020 al 2024. Se descartó esa información debido a que los años excluidos tenían variables que no eran consistentes con los demás años además considerando que sería más valioso considerar el comportamiento desde el año de la pandemia por covid.

La información utilizada presenta información sobre el tipo de falta, género, grupo étnico, estado conyugal, nacionalidad, departamento y municipio de ocurrencia, ocurrencia, escolaridad, área geográfica, edad. Se agregó información la cual antes no se proporcionaba su definición en el diccionario de los datos y al consultar al INE proporcionó la información correspondiente. Además, se realizó cierta limpieza y manipulación presentada en el código. Repositorio: [rodrigofernandezm/Proyecto\\_Fase\\_2](https://github.com/rodrigofernandezm/Proyecto_Fase_2)

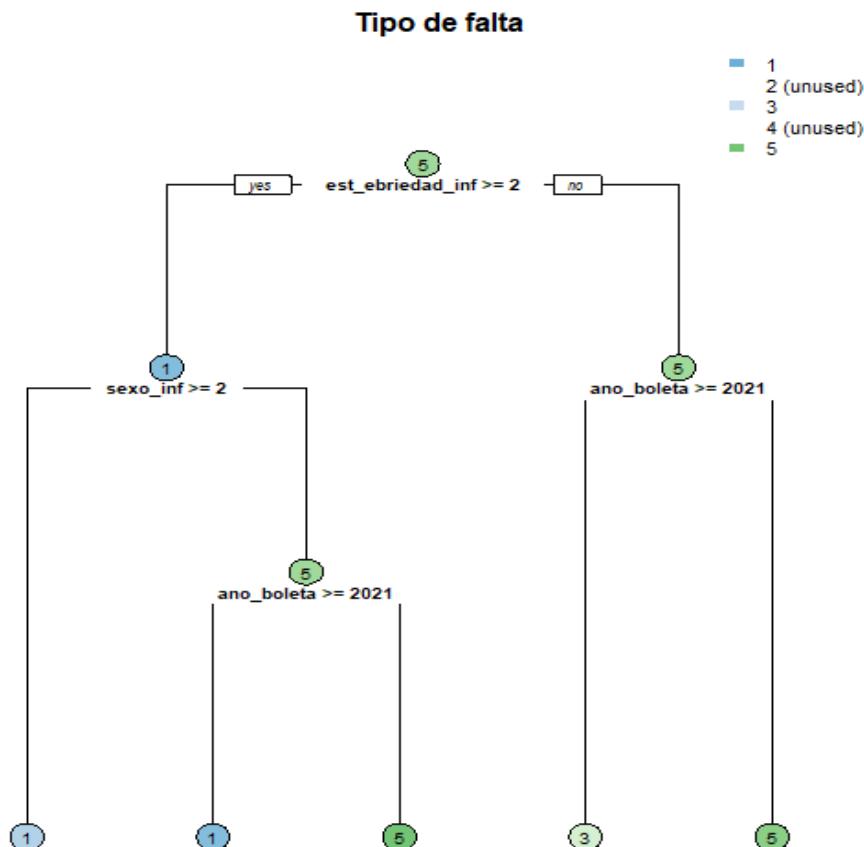
### **Modelado de la información**

Se realizaron 3 modelos diferentes: arboles de decisión, bosques aleatorios y redes neuronales. El objetivo es evaluar como estos modelos pueden predecir fenómenos en función de mejorar el análisis y poder presentar propuestas para la prevención de los fenómenos analizados, en este caso las faltas judiciales. En este documento se presenta un análisis de una manera mas legible y no tanto en valores numéricos como tal, los valores de las simulaciones y componentes de los modelos se pueden consultar en el código del repositorio.

## Resultados:

### Árboles de decisión:

1. Predicción de tipo de falta. La variable dependiente fue el **tipo de falta**, mientras que como predictores se incluyeron variables relacionadas con el lugar y la fecha del hecho (departamento y municipio del hecho, mes y año de la infracción), características demográficas de la persona involucrada (genero, edad, grupo étnico, estado conyugal, lugar de nacimiento, nivel de alfabetismo, nivel de escolaridad), y factores situacionales asociados al evento (estado de ebriedad, área geográfica, departamento de nacimiento, sub grupos principales, grandes grupos, grupos primarios). El modelo se empleó sobre la totalidad de los registros de la base de datos para identificar patrones predictivos que permitan asociar las condiciones del hecho con la categoría de falta correspondiente.



Casos prácticos:

Se evaluó el funcionamiento del modelo mediante dos escenarios prácticos de predicción de tipo de falta. El primer caso fue construido utilizando valores que representan mi propio perfil, suponiendo que yo fuera el infractor. Para dicho escenario se ingresaron las siguientes características: sexo masculino, 29 años perteneciente a grupo étnico ladino, estado conyugal soltero, nacido en Guatemala alfabetizado, con nivel de escolaridad superior, residente en área urbana, en estado etílico al momento del hecho y clasificado dentro de un gran grupo administrativo de actividad económica y ocupación principal. La ubicación del evento se fijó en el departamento 1 y municipio de Mixco

El segundo caso correspondió al perfil de “Don José”, una persona adulta mayor, residente en área rural, con bajo nivel educativo y una situación conyugal casado, seleccionado para representar un contexto social diferente al primero. De esta manera, ambos ejemplos permitieron observar si variaciones en la situación del infractor varía la falta.

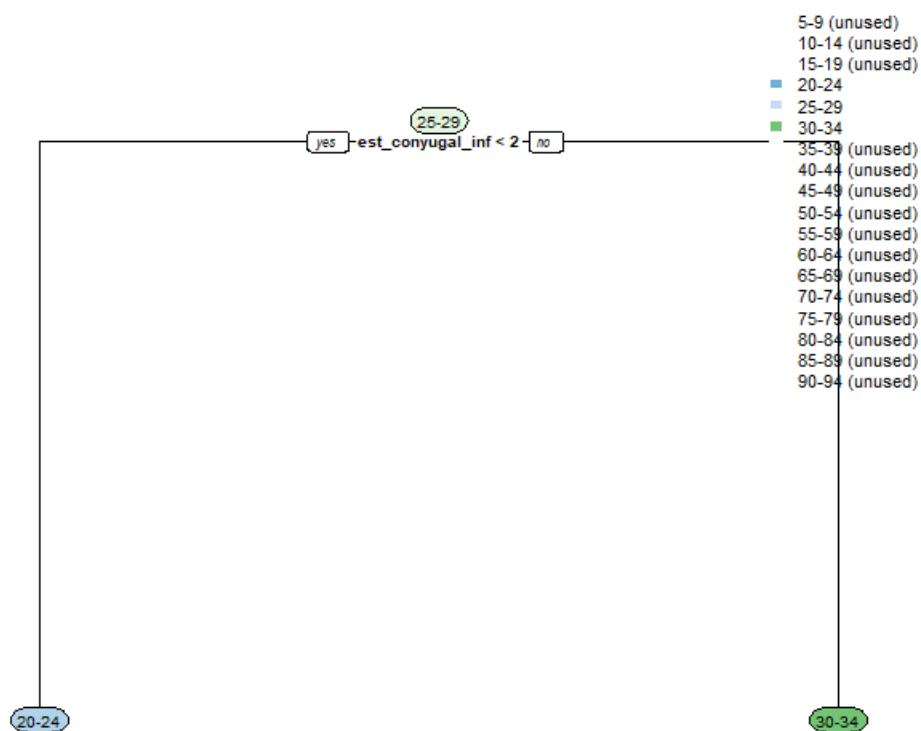
Árbol 1		
	Probabilidad	Interpretación
Caso 1	0.44	Falta contra las buenas costumbres.
Caso 2	0.44	Falta contra las buenas costumbres.

Según los resultados a pesar de ser perfiles diferentes lo más probable es que la falta sea la misma (falta contra las buenas costumbres). Eso debido a que la según visto en la fase 1 la falta más recurrente es falta contra las buenas costumbres por lo que lógicamente no importa tanto el perfil de la persona para poder predecir la falta.

2. Se desarrolló un segundo modelo con el objetivo de predecir la variable **edad quinquenal**, utilizando como predictores diversas características administrativas y sociodemográficas disponibles en la base de datos (departamento y mes de emisión, tipo de falta, estado conyugal, sexo, grupo étnico, condición de alfabetismo, nivel educativo, estado de ebriedad y área geográfica). No obstante, los resultados obtenidos mostraron un desempeño limitado del árbol de decisión, ya que únicamente el estado conyugal logró aportar información para separar parcialmente los grupos etarios, específicamente distinguiendo entre las categorías de 20–24 años y 30–34 años. El resto de los grupos permaneció sin capacidad de ser predicho de manera adecuada. Este resultado evidencia que las variables utilizadas no contienen

una señal predictiva suficiente para inferir el rango de edad de las personas involucradas y, por lo tanto, la edad quinquenal no puede ser estimada con precisión mediante las características disponibles en este conjunto de datos.

#### grupo etario



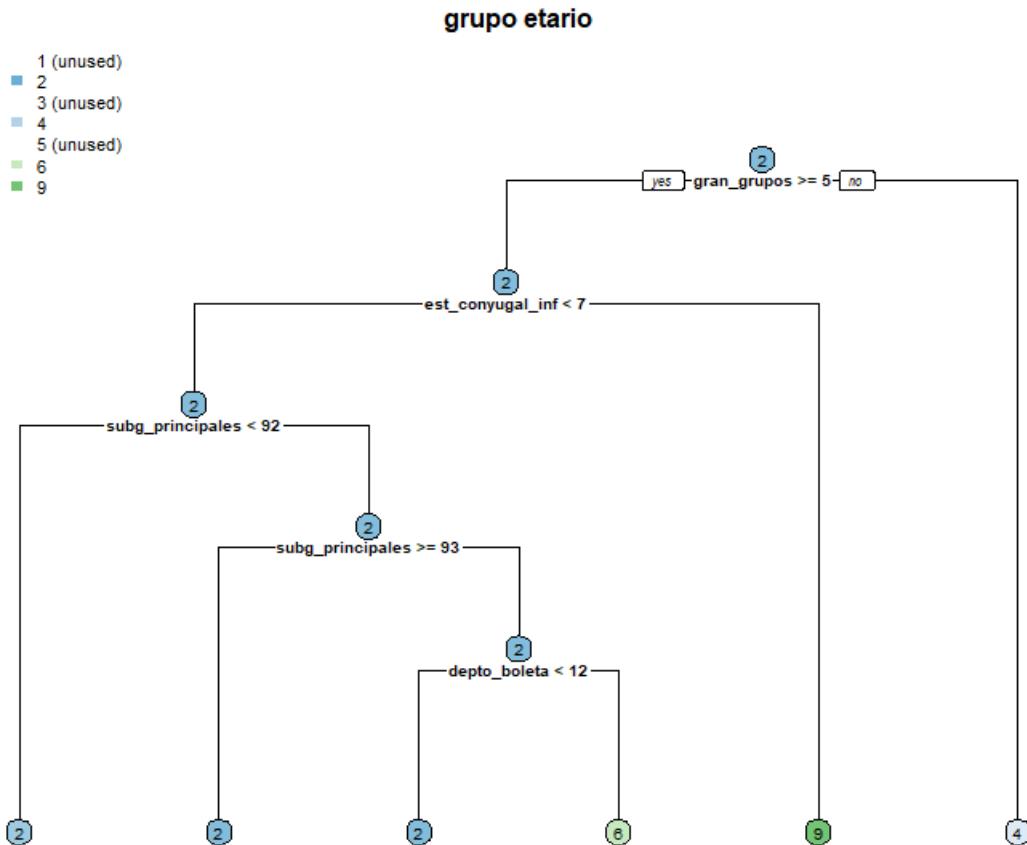
Para evaluar la capacidad del modelo en la predicción del grupo etario quinquenal, se aplicaron dos perfiles de prueba con características sociodemográficas distintas. En el primer caso, correspondiente a un infractor masculino del departamento de Guatemala, urbano, alfabetizado, con educación media y en estado sobrio, el modelo distribuyó la probabilidad principalmente entre los rangos de 20–24 años (26.27%), 25–29 años (23.18%) y 30–34 años (15.73%), reflejando una tendencia hacia edades adultas jóvenes. En contraste, el segundo caso —una mujer del departamento de Quetzaltenango, también urbana, con nivel educativo básico y sin estado de ebriedad— mostró mayores probabilidades en grupos ligeramente mayores, destacando 30–34 años (19.42%), 25–29 años (18.50%) y 35–39 años (18.17%). Estas diferencias evidencian que variables como la educación, el sexo y el contexto territorial influyen en la agrupación etaria estimada por el modelo, aunque la dispersión en las probabilidades indica que la precisión sigue siendo limitada y dependiente de factores no capturados completamente en este árbol de decisión.

Árbol 2		
	Probabilidad	Interpretation
Caso 1	0.26	20-24
Caso 2	0.19	30-34

3. El tercer modelo se enfocó en predecir el **nivel de escolaridad** de la persona involucrada, utilizando variables sociodemográficas y del contexto del hecho. Se incluyeron como predictores: departamento y municipio donde se emitió la boleta, mes y año del registro, tipo de falta cometida, sexo, grupo étnico, estado conyugal, edad quinquenal, condición de ebriedad, área geográfica (urbano/rural), departamento de nacimiento, indicadores de edad avanzada y variables de actividad laboral (subg\_principales, gran\_grupos y g\_primarios).

El árbol de decisión generado muestra que las variables laborales (especialmente subg\_principales y gran\_grupos) dominaron la segmentación del modelo. Esto implica que el tipo de ocupación y el sector económico asociado a la persona permiten inferir, con cierta consistencia, su nivel educativo. También aparece el estado conyugal como segundo factor relevante, lo cual sugiere una relación entre los cambios en el ciclo de vida (por ejemplo, unión o matrimonio) y el nivel alcanzado de estudios. El departamento de la infracción solo interviene en una división secundaria, indicando que la ubicación influye pero en menor medida que las características laborales.

Aunque el árbol muestra una estructura simple, confirma una tendencia esperada: determinadas ocupaciones están más asociadas a niveles educativos altos o bajos, lo que se refleja directamente en la clasificación del modelo. No obstante, el comportamiento del árbol también evidencia limitaciones predictivas, ya que el modelo sólo logra discriminar principalmente entre dos niveles educativos en la mayoría de ramas, con un bajo aprovechamiento del resto de la estructura de clases disponibles en la base de datos. Esto se debe a que las variables consideradas no aportan suficiente variabilidad explicativa para diferenciar adecuadamente todos los niveles de escolaridad.



Para evaluar el modelo se crearon dos perfiles.:.

Persona de 32 años, hombre, soltero, alfabetizado, residente en área urbana del interior del país (departamento 9). Nació en Guatemala, no estuvo bajo efectos de alcohol al momento del hecho. Mismo mes y tipo de infracción registrada en el sistema para mantener consistencia entre casos.

Lo único diferente es su ocupación: trabajador agrícola, parte del sector primario.

Para la situación 2 son exactamente las mismas características personales, de lugar y contexto. Solo cambia la ocupación: un trabajador profesional de la salud (ej. enfermero o médico), un sector que estadísticamente requiere mayor escolaridad.

Árbol 3		
	Probabilidad	Interpretacion
Caso 1	0.42	primaria
Caso 2	0.42	primaria

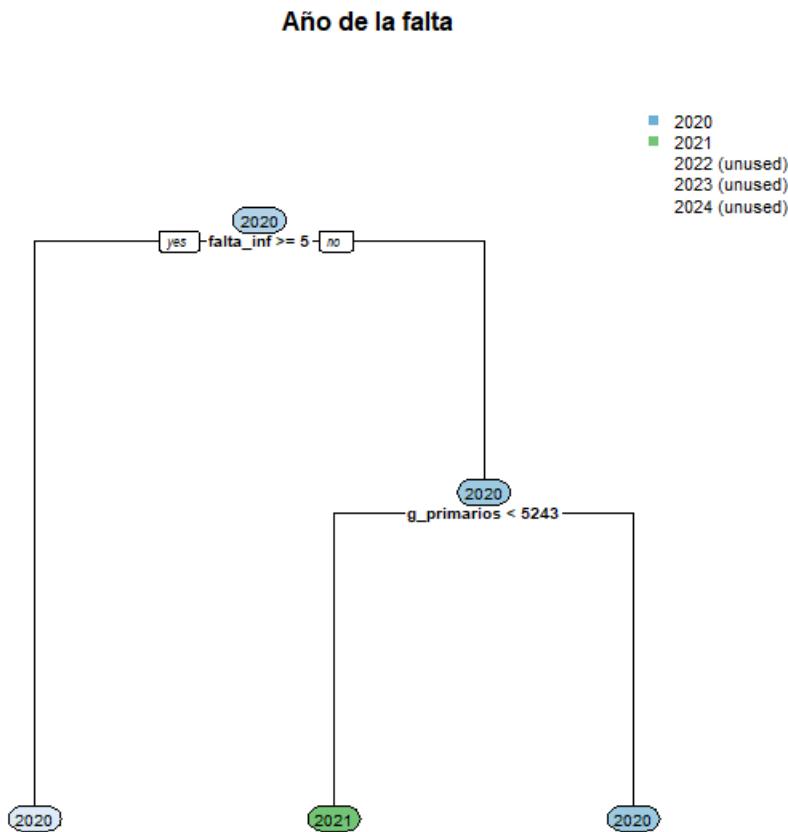
El árbol de decisión para predecir el nivel de escolaridad muestra que variables relacionadas con la ocupación, principalmente los grupos ocupacionales (gran\_grupos) y subgrupos específicos (subg\_principales), concentran la mayor capacidad predictiva del modelo, lo cual es coherente con la asociación esperada entre la formación académica y el tipo de actividad laboral desempeñada. Sin embargo, al probar distintos perfiles únicamente variando la ocupación, el modelo asignó la misma distribución de probabilidad en la predicción. Esto evidencia que, aunque la estructura del árbol reconoce la influencia de la ocupación en algunos segmentos de la población, aún existen limitaciones para discriminar adecuadamente entre niveles educativos en ciertos perfiles, probablemente debido a la dominancia de categorías educativas medias y la poca profundidad del árbol. En síntesis, el modelo es explicativo y lógico en su jerarquía de variables, pero su precisión diferenciadora resulta insuficiente y requiere ajustes o técnicas más robustas para capturar la diversidad real en los niveles de escolaridad de la población infractora.

4. El árbol de decisión para predecir el **año de la infracción** muestra que la variable más influyente es el tipo de falta (falta\_inf), específicamente cuando corresponde a valores iguales o mayores a 5. Esto indica que ciertas infracciones específicas se concentraron en un año particular dentro del conjunto de datos. De hecho, el nodo raíz predice predominantemente registros del año 2020, lo cual es coherente con el hecho de que dicho año representa el mayor volumen de infracciones, posiblemente vinculado a cambios en la movilidad y medidas restrictivas durante el periodo de pandemia.

Entre los individuos sin faltas del tipo 5 o superior, la siguiente variable relevante es la clasificación ocupacional primaria (g\_primarios), que permite diferenciar un subconjunto que se asocia mayoritariamente con infracciones registradas en 2021. No obstante, el modelo no logra distinguir adecuadamente los años 2022, 2023 y 2024, que aparecen como categorías “unused”, es decir, no aportan separación predictiva y la mayoría terminan asignadas a 2020.

En resumen, aunque se identifican patrones iniciales ,especialmente relacionados con el tipo específico de falta, el modelo presenta una capacidad predictiva limitada debido a la fuerte desbalanceación entre años y la baja influencia de las demás variables sociodemográficas y ocupacionales. El resultado sugiere que este enfoque no es adecuado

para estimar correctamente el año de una infracción y requeriría técnicas más robustas o la incorporación de variables temporales adicionales para mejorar la discriminación entre períodos.



Para evaluar el desempeño del modelo se crearon dos casos ficticios. El primer perfil corresponde a un hombre de aproximadamente 30 años, residente en un municipio urbano del interior del país. Se reporta como soltero, con una falta de impacto social moderado y sin estado de ebriedad al momento del incidente. Su ocupación se encuentra dentro del sector servicios comerciales lo cual implica un contacto frecuente con espacios públicos y horarios laborales dinámicos. Al evaluar este conjunto de características, el modelo identificó una mayor probabilidad de que el evento ocurriera en el año 2021.

Esto resulta coherente con el contexto nacional de reapertura económica posterior a la fase crítica de restricciones sanitarias, cuando se intensificó la actividad en mercados, comercios y transporte urbano. En este sentido, la predicción del modelo sugiere que ciertos comportamientos infractores relacionados con actividades comerciales se desplazaron temporalmente hacia dicho periodo.

El segundo perfil describe a un hombre de 45 años, también ubicado en un área urbana fuera de la capital, pero vinculado laboralmente al sector agrícola. Se trata de una infracción leve, sin presencia de alcohol y con un estilo de vida más estable asociado a una persona casada.

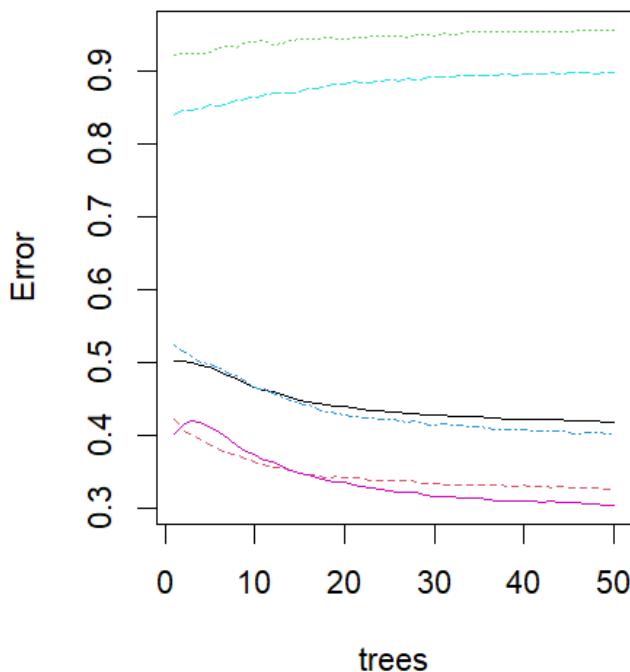
Árbol 4		
	Probabilidad	Interpretación
Caso 1	0.38	2021
Caso 2	0.28	2020

Al procesar estos datos, el modelo asignó mayor probabilidad al año 2020 como fecha del hecho. Esta predicción se alinea con el hecho de que durante los primeros meses de la pandemia existieron controles y disposiciones sanitarias que impactaron la movilidad y las dinámicas laborales, incluyendo a los trabajadores agrícolas en zonas urbanas que debían desplazarse por razones económicas esenciales. De esta manera, el modelo interpreta que ciertos registros de infracciones con bajo nivel de conflictividad social permanecieron concentrados en el periodo de restricciones iniciales.

### Random Forest

1. Con el objetivo de mejorar el desempeño predictivo logrado mediante el árbol de decisión, se implementó un modelo Random Forest utilizando las mismas variables sociodemográficas, geográficas y ocupacionales como predictores, manteniendo como variable objetivo el **tipo de falta cometida**. Este enfoque permite capturar relaciones no lineales entre las variables y reducir el sobreajuste característico de los modelos de un solo árbol. El entrenamiento del modelo utilizó 50 árboles de decisión base, lo cual permitió observar una disminución progresiva en el error Out-of-Bag (OOB), estabilizándose alrededor de un 40%, según se aprecia en el gráfico de evolución del error. Aunque el desempeño continúa siendo limitado por la fuerte desbalance en la distribución de tipos de falta, el Random Forest muestra una mayor robustez y mejor capacidad de clasificación comparado con el árbol simple. De esta manera, se obtiene un modelo más confiable para analizar los factores asociados a la infracción, manteniendo un enfoque interpretativo a partir de la importancia relativa de las variables predictoras.

## rf\_1



Casos prácticos:

Hombre de 28 años, residente en un municipio urbano del departamento 1. Se reporta soltero, con escolaridad media (básico completo), sabe leer y escribir y trabaja como dependiente de comercio en una tienda de barrio. No estaba bajo los efectos del alcohol al momento de la infracción. Este tipo de perfil suele reflejar conductas infractoras relacionadas con disputas en vía pública o faltas contra la convivencia ciudadana.

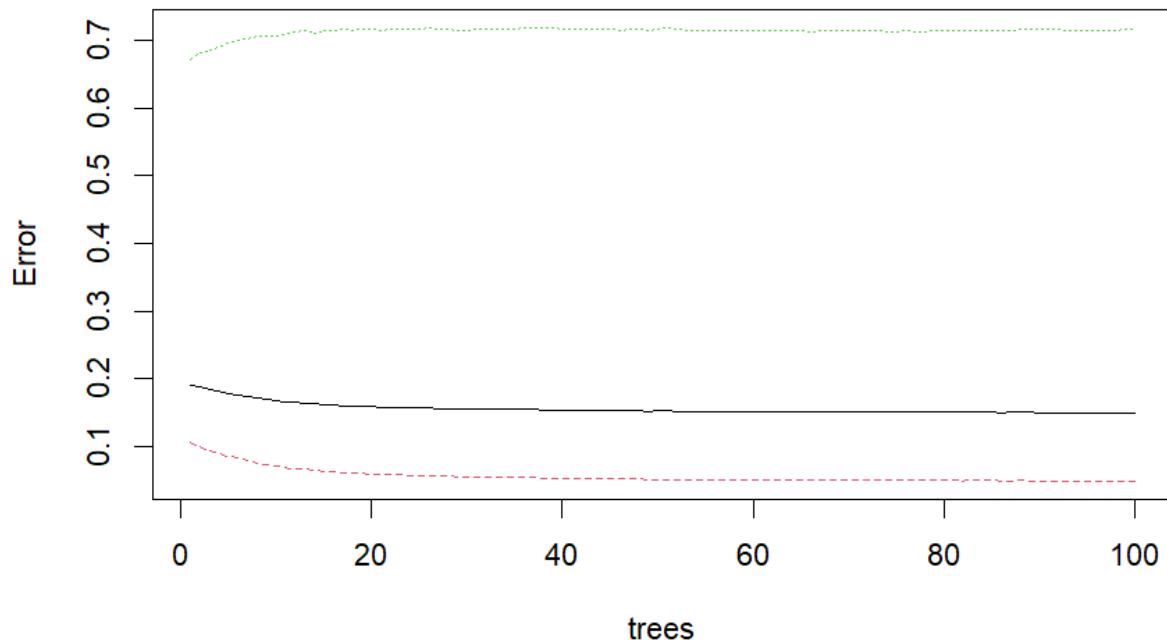
Hombre de 42 años que vive en el área urbana de un municipio del departamento 8. Está casado, tiene escolaridad primaria incompleta y pertenece al sector agrícola. Aunque trabajó todo el día, fue detenido en la noche por una falta leve con consumo moderado de alcohol. Este tipo de perfil representa a una parte importante de la población registrada en boletas municipales.

Random Forest 1		
	Probabilidad	Interpretacion
Caso 1	0.5	Contra las buenas costumbres
Caso 2	0.45	Contras las personas

Al aplicar el modelo Random Forest a dos perfiles sociales distintos, se observa que los patrones de predicción continúan influenciados por características ocupacionales, sociodemográficas y geográficas. En el primer caso, correspondiente a un hombre joven empleado en el sector comercio dentro del área urbana, la mayor probabilidad de clasificación (50%) se asigna a la falta tipo 3, seguida de la tipo 5 (24%), lo cual se relaciona con la prevalencia de incidentes vinculados al comportamiento en espacios públicos y actividades comerciales. Por su parte, el segundo perfil, representado por un agricultor adulto del interior del país, presenta una alta probabilidad hacia la falta tipo 1 (45%) y tipo 5 (33%), evidenciando que, dentro de poblaciones laborales primarias y con escolaridad más baja, son más frecuentes infracciones relacionadas con convivencia y orden público, con menor presencia de faltas asociadas a condiciones urbanas o actividades económicas formales. En ambos casos, el modelo muestra una distribución más informativa de probabilidad que el árbol de decisión, demostrando una mejor capacidad para capturar complejidad en los patrones de infracción.

2. Para explorar la capacidad predictiva del Random Forest con una variable objetivo de naturaleza dicotómica, se entrenó un segundo modelo con el **género de la persona** infractora como variable dependiente. Se incorporaron características sociodemográficas como edad, grupo étnico, estado conyugal y nivel educativo, así como variables territoriales (departamento y área geográfica) y condiciones del hecho (mes del incidente y estado de ebriedad). El modelo fue entrenado con 100 árboles y un parámetro  $mtry=5$ , mostrando una rápida estabilización del error Out-of-Bag (OOB) cercana al 15%, lo cual supone un rendimiento significativamente mejor que los resultados obtenidos en los modelos previos. La matriz de confusión indicó una precisión global del 85.1%, evidenciando que las características consideradas permiten diferenciar de manera efectiva entre infractores hombres y mujeres. Estos resultados sugieren que la asignación del sexo en registros policiales o municipales podría ser estimada de forma confiable incluso en situaciones con información parcial, y resaltan patrones diferenciales de género vinculados al perfil sociodemográfico de la población infractora.

## rf\_2



Casos prácticos:

Para evaluar el modelo se crearon dos perfiles. El primero una persona de 27 años que reside en un área urbana del departamento Guatemala. Soltera, con nivel de escolaridad medio y alfabetizado. Trabaja en actividades de comercio formal y no se encontraba bajo efectos de alcohol. Este perfil es típico de infractores vinculados a disputas en el espacio público o faltas en contextos de comercio o transporte. Por otro lado el perfil 2 es una persona de 38 años residente en un municipio urbano del departamento de Totonicapán. Casada, con nivel educativo básico y perteneciente a comunidad indígena. Infracción relacionada a convivencia y orden público, sin consumo de alcohol. Este perfil representa una minoría dentro del registro, pero es clave para entender brechas sociales en procesos sancionatorios.

Random Forest 2		
	Probabilidad	Interpretacion
Caso 1	0.85	Hombre
Caso 2	0.83	Hombre

Al aplicar el modelo Random Forest a dos perfiles distintos, los resultados muestran una probabilidad muy alta de clasificación como sexo masculino en ambos casos, con valores del 85% y 83% respectivamente. Esto evidencia que, incluso ante diferencias en edad, origen étnico, estado conyugal y nivel educativo, el perfil predominante en las infracciones

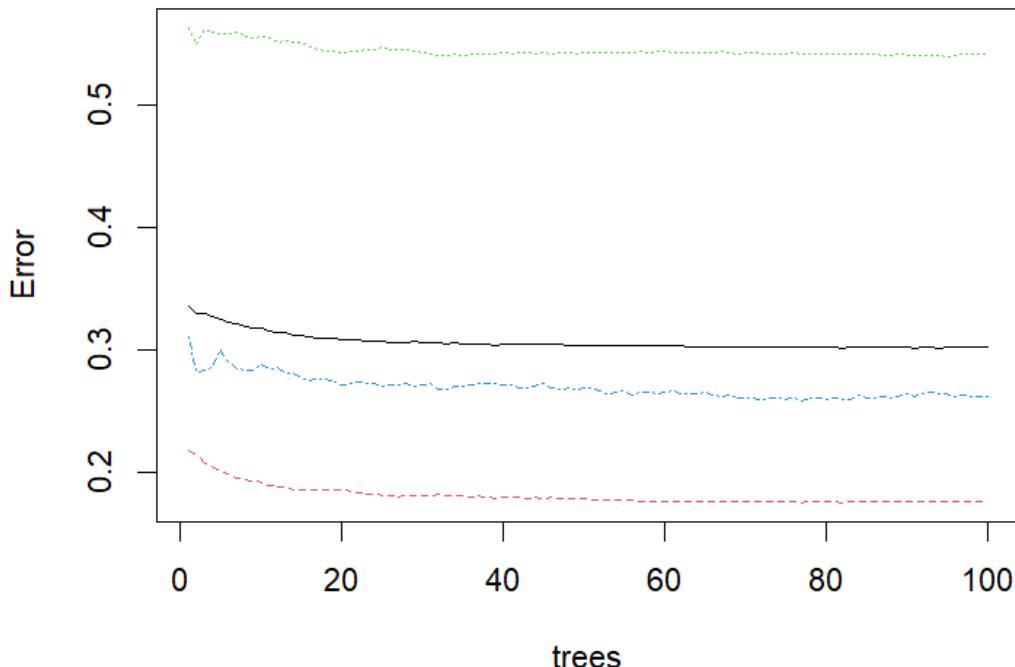
sigue correspondiendo al de hombres jóvenes o adultos. Aunque el segundo caso corresponde a una mujer en términos descriptivos, el modelo mantiene una tendencia fuerte hacia la predicción masculina debido al claro desbalance en la base de datos, donde los hombres representan la gran mayoría de la población infractora. Estos resultados reflejan una limitación en la sensibilidad del modelo hacia la representación femenina en los registros, pero a la vez confirman el patrón criminológico ampliamente documentado: las infracciones registradas en municipios y juzgados de faltas están fuertemente asociadas a hombres en edad productiva vinculados a contextos laborales y de convivencia urbana.

### 3.

Para el tercer modelo de random forest sería interesante evaluar el nivel de predicción del área de la falta. Utilizando variables como departamento, genero del infractor, edad, grupo étnico, estado conyugal, alfabetismo, escolaridad, tipo de falta, estado de ebriedad.

El modelo logra un desempeño aceptable considerando la estructura de los datos y el desbalance entre zonas urbanas y rurales. La capacidad predictiva es mejor en el entorno urbano, lo cual es consistente con la mayor cantidad de registros de faltas en áreas metropolitanas.

**rf\_3**



Para evaluar el modelo se utilizarán dos casos prácticos. Hombre joven (28 años), educación media, alfabetizado, sin ebriedad, falta de convivencia social, departamento de Guatemala. Mientras que el segundo es mujer adulto de 42 años de edad, escolaridad baja, posible convivencia más tradicional, falta relacionada con desórdenes, departamento del interior.

Random Forest 3		
	Probabilidad	Interpretación
Caso 1	0.83	Área urbana
Caso 2	0.94	Área Urbana

El modelo Random Forest aplicado para predecir el área geográfica (urbano/rural) presentó un desempeño global del 66.2% de exactitud, lo que indica una capacidad predictiva moderada. Las variables con mayor importancia en la clasificación fueron departamento, grupo étnico y escolaridad, lo cual sugiere que los factores territoriales y sociodemográficos explican una parte sustancial de la variabilidad observada.

El análisis de casos revela que pequeñas modificaciones en los atributos individuales como el género del infractor pueden alterar significativamente las probabilidades asignadas por el modelo, reflejando una estructura de dependencia estadística entre características personales y su distribución geográfica en los registros.

En conjunto, los resultados muestran que existe una relación cuantificable entre el perfil del infractor y la localización de la falta, aunque una proporción importante de la variabilidad permanece sin explicar, lo cual limita la precisión del modelo y evidencia la presencia de factores externos o no observados en los datos.

## **Redes neuronales**

### **1. Predicción año de la falta**

En el caso de la red neuronal para predecir el **año de la falta**, los resultados muestran claramente que el modelo no logra capturar patrones útiles. La matriz de confusión indica que prácticamente todas las observaciones son clasificadas como 2020, independientemente de si en realidad corresponden a 2021, 2022, 2023 o 2024. Esto implica que la red terminó comportándose como un clasificador ingenuo que simplemente escoge siempre la clase más frecuente en los datos (2020), aprovechando el desbalance de la base, pero sin aprender relaciones discriminantes reales entre las variables explicativas y el año.

Desde un punto de vista estadístico, el mensaje es que el año de la falta no es predecible con las variables disponibles (depto, mes, edad, tipo de falta, ocupación, etc.). La

información contenida en estos predictores no parece diferenciar de forma consistente los patrones entre 2020 y los años posteriores, de modo que incluso modelos más complejos como redes neuronales caen en soluciones triviales. En el informe se puede concluir que, aunque técnicamente fue posible entrenar la red, su capacidad predictiva y explicativa para el año específico es muy limitada, y que este objetivo no resulta adecuado con la estructura actual de los datos.

#### Casos prácticos:

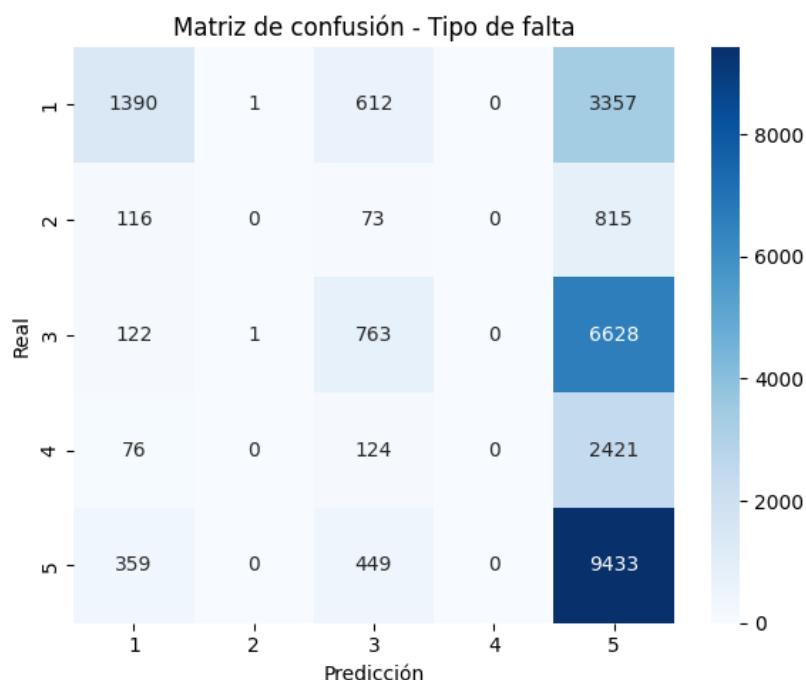
Para ilustrar el funcionamiento de la red neuronal se construyeron dos perfiles contrastantes. El primer caso corresponde a un hombre de 25 años, soltero, residente en área urbana del departamento de Guatemala, dedicado al comercio al por menor (subgrupo 5223, gran grupo 5) y sin ocupación primaria adicional registrada; la falta cometida es de tipo 3 y ocurre en el mes de marzo, en estado sobrio. El segundo caso representa a una mujer de 40 años, casada, que vive en área rural de un departamento del interior, vinculada a actividades de servicios personales o profesionales (subgrupo 6111, gran grupo 6), con ocupación primaria registrada; en este escenario la falta es de tipo 5, se registra en octubre y la persona se encuentra bajo efectos de alcohol.

Red neuronal 1	
	Interpretación
Caso 1	2020
Caso 2	2020

Al evaluar estos dos perfiles con la red neuronal, en ambos casos el modelo predijo el año 2020 como fecha de ocurrencia de la falta, a pesar de las diferencias claras en sexo, edad, área geográfica, tipo de falta y grupo ocupacional. Este resultado es consistente con la matriz de confusión global: la red tiende a asignar la mayoría de las observaciones al año más frecuente en la base (2020), lo que evidencia que, con las variables disponibles, el modelo no logra distinguir patrones suficientemente robustos entre 2020 y los años posteriores (2021–2024). En la práctica, su desempeño no supera al de un clasificador ingenuo que siempre pronostica el año predominante.

## 2. Predicción tipo de falta

En la segunda red neuronal se modeló el tipo de falta como variable dependiente (clases 1 a 5), utilizando únicamente variables de contexto general: departamento y municipio de la boleta, mes y año del hecho, sexo, edad y grupo étnico de la persona infractora. La matriz de confusión resultante muestra una exactitud global cercana al 52 %, lo que supera a un clasificador ingenuo que siempre elegiría la categoría más frecuente, pero evidencia también un comportamiento fuertemente sesgado hacia la falta tipo 5, que concentra la mayoría de las predicciones incluso cuando la falta real es 1, 2, 3 o 4. En particular, una proporción importante de faltas de tipo 1 y 4 se reclasifican como tipo 5 y la clase 4 prácticamente no aparece como predicción, lo que indica que, aunque la red neuronal sí captura cierta señal diferenciadora a partir del contexto geográfico y sociodemográfico, su capacidad discriminativa sigue limitada por el desbalance de clases y la ausencia de variables más específicas sobre la conducta infractora.



Casos prácticos:

Un primer caso corresponde a un hombre de 23 años, residente en el departamento de Guatemala, en un municipio céntrico de la capital. El evento sucede en el mes de agosto de 2023, y la persona se identifica en el grupo étnico 1 (por ejemplo, población ladina/mestiza según la codificación oficial). Este perfil representa a un joven urbano, en una zona con alta

densidad de población y mayor presencia de actividades comerciales y de ocio, donde suelen concentrarse muchas de las faltas registradas en el sistema.

El segundo caso describe a una mujer de 41 años que vive en un departamento del interior (por ejemplo, código 9) en un municipio predominantemente rural. La falta se registra en el mes de febrero de 2021, y la persona pertenece al grupo étnico 2 (por ejemplo, población indígena según la clasificación utilizada). Este perfil refleja un contexto territorial y sociocultural diferente al de la capital, con otra dinámica de movilidad, relaciones comunitarias y presencia institucional, lo que permite observar cómo la red neuronal responde a escenarios alejados del patrón urbano típico.

Red neuronal 2	
	Interpretacion
Caso 1	Otras
Caso 2	Otras

Con los dos perfiles probados, la red neuronal asignó tipo de falta 5 en ambos casos, a pesar de que conceptualmente representan contextos bastante distintos (joven hombre urbano en la capital vs. mujer adulta en el interior rural). Esto es coherente con lo que muestra la matriz de confusión: el modelo tiende a concentrar muchas predicciones en la categoría 5, que es la más frecuente en el conjunto de datos. En la práctica, esto indica que la RNA está capturando solo patrones muy generales asociados al contexto geográfico y sociodemográfico, pero no logra diferenciar con claridad entre los distintos tipos de falta; termina “jugando seguro” y enviando a la clase mayoritaria. Para el informe puedes señalar que, aunque el ejercicio sirve para ilustrar cómo se aplicaría el modelo a casos concretos, la interpretación de la predicción debe hacerse con cautela porque el desempeño sigue limitado por el desbalance de clases y por la falta de variables más específicas sobre la conducta infractora.

## **Propuestas:**

### **1. Fortalecimiento de la calidad y riqueza de la información sobre faltas judiciales**

En los diferentes modelos varios resultados muestran claras dificultades para poder predecir algunas variables, como el tipo de la falta, el área de la falta. Esto sugiere dos posibles causas, la primera falta de calidad de la información y por otro lado un comportamiento homogéneo en las faltas.

Partiendo sobre las investigaciones sobre criminalidad en América latina señalan que uno de los principales problemas de los sistemas estadísticos de crimen es la calidad y homogeneidad de los datos, falta de periodicidad, baja desagregación y ausencia de criterios uniformes entre instituciones, lo que limita seriamente el análisis y la planificación.

La propuesta es implementar un sistema de registro digital en campo para la Policía Nacional Civil, basado en dispositivos móviles como tablets o teléfonos inteligentes con formularios estructurados y conectados al sistema para validaciones. Esto con el fin que se puedan registrar elementos con mayor facilidad y en el campo, sin ser necesario tener que estar en la estación de la policía para que el registro pueda ser tomado. Considerando que en las áreas rurales no se tiene la infraestructura adecuada.

Con esto se atacarían 3 problemas:

- Variables limitadas del hecho (variables relacionadas al hecho como tal)
- Errores en la digitación como campos vacíos u otro tipo de errores.
- Retraso entre el hecho y el registro digital (actualmente es en papel y se transcribe)

Mucha literatura sugiere que las policías que usan tablets o teléfonos para el registro muestran mejoras en la calidad de los datos, reducción de errores. Otros estudios muestran que en la región una proporción importante de los delitos nunca llega a los registros policiales, generando sesgos sistemáticos en las bases de datos oficiales y subestimando ciertos tipos de violencia y de casos.

En Guatemala hay avances de dotar a la policía con estos dispositivos móviles para aumentar la cobertura (sistema SIP). En 2019 se distribuyeron 2500 teléfonos inteligentes a el mismo número de agentes para consulta de bases de datos, envío de reportes, fotos, ubicación etc. Pero este esfuerzo no es suficiente ya que a septiembre del 2025 existen aproximadamente 43,000 agentes de policías en el país y se proyecta un aumento considerable en los próximos años.

En conclusión es algo en lo que ya se está trabajando y después de profundizar es algo que se puede validar que se es haciendo bien por lo que la propuesta se centra en continuar con esta iniciativa y pasar de un 5.8% de policías con estos dispositivos a aumentar este porcentaje para que en un plazo de 5 años todos los agentes tenga estos dispositivos con esto no solo mejorar la calidad de la información relacionada a las faltas judiciales sino que también ayudar a la institución en sus demás labores.

## **2. Monitoreo constante de áreas criticas para faltas de desorden público**

En los modelos analizados el tipo de falta asociado al desorden público aparece con alta probabilidad en diferentes perfiles. Es decir que no se concentra en un grupo, sino que es un común denominador. Con eso se puede inferir que el problema de faltas hacia el orden publico son regidos por factores mas generales y no tanto específicos.

La literatura internacional sobre la seguridad ciudadana ha mostrado de forma consistente que concentrar los esfuerzos policiales y de prevención en punto específicos (denominados puntos calientes) puede reducir el crimen y el desorden sin necesidad de incrementar masivamente el número total de efectivos. Diferentes revisiones y ensayos controlados aleatorios indican que esta estrategia reduce significativamente las faltas relacionadas al desorden público.

Existe evidencia que herramientas tecnológicas de videovigilancia pueden contribuir a reducir ciertos tipos de delitos y desorden en lugares públicos, especialmente cuando se ubican de forma estratégica y se integran a un sistema de monitoreo activos. Este análisis mas global y revisiones sistemáticas reportan reducciones significativas en lugares como parqueos, calles solitarias, nodos de transporte, calles con discotecas y bares y efectos sobre la percepción de seguridad.

El Banco Interamericano de Desarrollo ha promovido guías de implementación de puntos calientes den América latina y el caribe, enfocado en la identificación y monitoreo continuo de micro lugares con alta concentración de incidentes para utilizar de manera más eficiente los recursos y focalizar la presencia de manera efectiva.

En Guatemala distintos informes coinciden en que la criminalidad se concentra en zonas específicas de la ciudad particularmente en la Ciudad de Guatemala y en áreas urbanas. Algunos de estos indican que el desorden no se distribuye homogéneamente por zona, sino que existen zonas mas pequeñas llamadas micro áreas dentro de cada zona que presentan mayor problema.

La propuesta se basa en implementar un sistema de monitoreo constante en áreas criticas para faltas de desorden público resulta viable. Además de monitores activos con elementos policiacos también monitoreo más pasivo con cámaras.

- Se alinearía con los modelos de otros países además con la información que ya se tiene analizada actualmente en el país.
- Se definirían lugares y horas críticas y protocolos para el correcto manejo de las situaciones.
- Cámaras para monitoreo activo.

Esto permitiría diseñar intervenciones más precisas, aumentar la sensación de seguridad y evitar en ciertas medidas las faltas al orden público en las zonas ya establecidas.

## Bibliografía

Gobierno de Guatemala, Ministerio de Gobernación. (2022, 17 noviembre). Entrega de equipo tecnológico fortalece la Estrategia de Transformación Policial. Guatemala.gob. <https://www.guatemala.gob.gt/noticias/entrega-de-equipo-tecnologico-fortalece-la-estrategia-de-transformacion-policial/>

Soy502. (2021, 11 junio). Los dispositivos que la PNC usa en retenes y operativos. Soy502. <https://www.soy502.com>

Carter, J. G. (2014). Impact of mobile broadband data access on police response and crime reporting (Report No. 250261). U.S. Department of Justice, National Institute of Justice.

Groff, E. (2008). Identifying and measuring the effects of information technologies on law enforcement agencies (NCJ 224520). Office of Community Oriented Policing Services (COPS).

Government Events. (2020, enero 27). Digital policing: Using mobile technology in crime scene investigation. GovernmentEvents.co.uk.

Ajzenman, N., & coautores. (2017). Crime concentration and hot spot dynamics in Latin America. In Crime concentration and hot spot dynamics in Latin America (pp. 1–44). Inter-American Development Bank / Organización de los Estados Americanos.

Cabrera, J. M. (2022). Hot spots, patrolling intensity, and robberies. Munich Personal RePEc Archive (MPRA Paper No. 113786).

World Bank & IDB (mencionados en notas sobre programas regionales de seguridad). Si los citas, podrías usar algo como:

Inter-American Development Bank. (2024, diciembre 12). Latin America and Caribbean launch regional alliance against organized crime [Press release].

Sabatini, F., & colaboradores. (2006). La segregación social del espacio en las ciudades de América Latina. Banco Interamericano de Desarrollo. (Capítulos con análisis específico de Ciudad de Guatemala).