



Universidad de San Carlos de Guatemala

Facultad de Ingeniería

Escuela de Estudios de Postgrado

Introducción a la Minería de Datos

**ANÁLISIS DE PATRONES DE LESIONADOS Y FALLECIDOS EN ACCIDENTES DE
TRÁNSITO EN GUATEMALA (2015-2023) MEDIANTE MINERÍA DE DATOS**

Brandon René Portillo González

Carnet: 999011994

Repositorio: <https://github.com/usac201612398/PFinalF2>

Guatemala, noviembre 2025

INDICE DE CONTENIDO

1.	RESUMEN.....	1
2.	INTRODUCCION	2
3.	METODOLOGIA DE LA INVESTIGACION.....	3
3.1.	Datos y Preparación.....	3
3.2.	Técnicas de análisis	3
4.	PRESENTACION DE RESULTADOS.....	5
4.1.	Árboles de Decisión	5
4.1.1.	Predicción de tipo de evento	5
4.1.2.	Predicción de tipo de vehículo.....	6
4.1.3.	Predicción de jornada	7
4.1.4.	Predicción de departamento	8
4.2.	Bosques Aleatorios.....	9
4.2.1.	Predicción de tipo de evento	10
4.2.2.	Predicción de tipo de vehículo.....	11
4.2.3.	Predicción de jornada	12
4.2.4.	Predicción de sexo de la persona.....	14
4.2.5.	Predicción de departamento	15
4.3.	Redes Neuronales.....	16
4.3.1.	Predicción de tipo de evento	17
4.3.2.	Predicción de tipo de vehículo.....	17
4.3.3.	Predicción de jornada	17
4.3.4.	Predicción de sexo de la persona.....	18
4.3.5.	Predicción de departamento	18
5.	DISCUSION DE RESULTADOS	19
5.1.	Árboles de Decisión	19
5.2.	Bosques Aleatorios.....	20
5.3.	Redes Neuronales.....	21
5.4.	Comparación entre modelos.....	21
5.4.1.	Tipo de evento (tipo_eve)	21

5.4.2. Tipo de vehículo (tipo_veh)	22
5.5. Comparación por variable en común.....	22
6. CONCLUSIONES.....	24
7. REFERENCIAS.....	25

INDICE DE TABLAS

Tabla 1.	Resumen – Árboles de decisión	5
Tabla 2.	Casos de predicción para árboles	6
Tabla 3.	Árbol - Predicciones para el tipo de evento	6
Tabla 4.	Árbol - Predicciones para el tipo de evento	7
Tabla 5.	Árbol - Predicciones para el horario o jornada de ocurrencia	8
Tabla 6.	Árbol - Predicciones para el departamento	9
Tabla 7.	Resumen – Bosques Aleatorios	9
Tabla 8.	Matriz de confusión – Tipo de evento	10
Tabla 9.	RF - Predicciones de tipo de evento.....	11
Tabla 10.	Matriz de confusión – Tipo de Vehículo.....	11
Tabla 11.	RF - Predicciones de tipo de vehículo.....	12
Tabla 12.	Matriz de confusión – Horario o jornada	12
Tabla 13.	RF - Predicciones de jornada	13
Tabla 14.	Matriz de confusión – Sexo de la persona	14
Tabla 15.	RF - Predicciones de sexo de la persona.....	14
Tabla 16.	Matriz de confusión – Departamento	15
Tabla 17.	RF - Predicciones de departamento	16
Tabla 18.	Resumen – Redes Neuronales.....	16
Tabla 19.	RN - Predicciones de tipo de evento.....	17
Tabla 20.	RN - Predicciones de tipo de vehículo	17
Tabla 21.	RN - Predicciones de jornada	17
Tabla 22.	RF - Predicciones de sexo de la persona.....	18
Tabla 23.	RF - Predicciones de departamento	18

INDICE DE FIGURAS

Figura 1.	Representación gráfica de árbol para tipo de evento.....	5
Figura 2.	Representación gráfica de árbol para tipo de vehículo	6
Figura 3.	Representación gráfica de árbol para jornada	7
Figura 4.	Representación gráfica de árbol para departamento	8
Figura 5.	Grafico de error de bosque aleatorio para tipo de evento	10
Figura 6.	Gráfico de error de bosque aleatorio para tipo de vehículo.....	12
Figura 7.	Gráfico de error de bosque aleatorio para la jornada	13
Figura 8.	Gráfico de error de bosque aleatorio para sexo de la persona.....	14

Figura 9. Gráfico de error de bosque aleatorio para departamento 15

1. RESUMEN

En la segunda fase del proyecto se aplican modelos predictivos sobre el conjunto consolidado de accidentes de tránsito en Guatemala (2015–2023), con el objetivo de estimar la probabilidad de ocurrencia de escenarios críticos previamente identificados en la primera fase.

Se considera la predicción de variables clave como gravedad del accidente, tipo del vehículo, tipo de evento, horario de ocurrencia, departamento de los hechos, entre otras que se pueden analizar si son sometidas a cierta segmentación para obtener mejores resultados debido al desbalance del *dataset*.

Se desarrollaron cuatro árboles de decisión, tres bosques aleatorios y tres redes neuronales para determinar el comportamiento de los incidentes de transito en el país, de cual se observa que: Predominan los casos de lesionados, hombres, mayores de edad que colisionan generalmente en motocicletas y con mayor frecuencia en los meses de marzo y diciembre.

Este análisis complementa la fase exploratoria y proporciona herramientas predictivas para la toma de decisiones en seguridad vial en Guatemala.

2. INTRODUCCION

La segunda fase del proyecto tiene el objetivo de construir modelos que estiman la probabilidad de ocurrencia de eventos graves y apoyar la toma de decisiones en seguridad vial.

Para ello, se implementan tres enfoques complementarios. Primero, los árboles de decisión ofrecen interpretabilidad y reglas claras para comprender los factores que influyen en la gravedad del accidente. Seguidamente, los bosques aleatorios mejoran la precisión y reducen la varianza mediante el uso de múltiples árboles, lo que permite una evaluación más robusta de las variables predictoras. Finalmente, las redes neuronales capturan relaciones complejas y no lineales entre las variables, aportando una perspectiva más detallada en la predicción de riesgos.

Esta fase no solo busca comparar el desempeño de los modelos, sino también generar propuestas concretas basadas en los hallazgos predictivos. Dichas propuestas están orientadas a reducir la incidencia de accidentes y optimizar estrategias de prevención en el contexto guatemalteco.

3. METODOLOGIA DE LA INVESTIGACION

3.1. Datos y Preparación

Se consolidaron datos de 8 archivos Excel (2015-2020, 2022-2023) y 1 archivo SPSS (2021), totalizando más de 75,000 registros. Las variables seleccionadas incluyeron: año, mes, día, departamento, sexo, edad, condición (lesionado/fallecido), tipo de vehículo y tipo de evento. Se eliminaron registros con valores desconocidos (ej.: edad = 999, sexo = 9, tipov_veh=99) para reducir ruido.

3.2. Técnicas de análisis

El objetivo es comparar el desempeño de los modelos y generar información útil para la toma de decisiones en seguridad vial.

- Arboles de Decisión:**

Se implementaron cinco modelos con el algoritmo rpart, cada uno con un objetivo distinto: Tipo de evento (tipo_eve), tipo de vehículo (tipo_veh), horario (día/noche) y departamento de ocurrencia (depto_ocu).

- Random Forest:**

Se desarrollaron tres modelos Random Forest con el paquete randomForest, replicando los objetivos de los árboles y utilizando de 50 a 100 árboles por predicción. Además, se evaluó el error OOB y se generaron gráficos de error.

- **Redes Neuronales:**

Se entrenaron tres redes neuronales utilizando el paquete tensorflow, replicando los mismos objetivos de los árboles para predecir tipo de vehículo, tipo de evento y departamento de ocurrencia.

- **Visualización:**

Gráficos de árboles, error, barras y redes para interpretar resultados.

4. PRESENTACION DE RESULTADOS

4.1. Árboles de Decisión

Tabla 1.

Resumen – Árboles de decisión

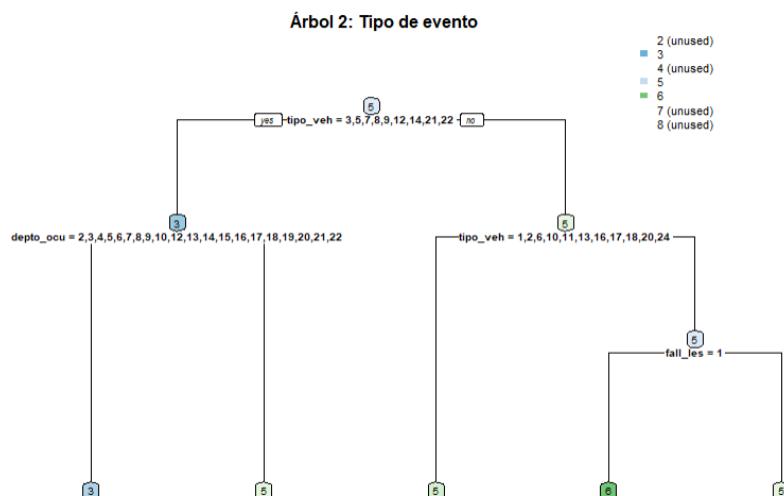
Árbol	Predictoras	Target	cex	Variables importantes
1	6	tipo_eve	0.5	tipo_veh, depto_ocu, fall_les, mayor_menor
2	6	tipo_veh	0.5	tipo_veh, depto_ocu, fall_les, mayor_menor
3	6	horario	0.5	tipo_veh, tipo_eve, dep_ocu
4	5	depto_ocu	0.5	tipo_veh, tipo_eve

Nota. Elaborado en Word 365 y obtenido de procesamiento mediante el lenguaje R versión 4.5.1.

4.1.1. Predicción de tipo de evento

Figura 1.

Representación gráfica de árbol para tipo de evento



Nota. Elaboración propia. Obtenido mediante biblioteca rpart.plot procesando mediante el lenguaje R versión 4.5.1.

Tabla 2.

Casos de predicción para árboles

caso	sexo_per	horario	mayor_menor	tipo_veh	depto_ocu	fall_les	tipo_eve
1	Mujer	Día	Mayor	Motocicleta	Guatemala	Fallecido	Choque
2	Hombre	Noche	Menor	Automóvil	Santa Rosa	Fallecido	Atropello
3	Mujer	Día	Menor	Pick up	Quiché	Lesionado	Derrape
4	Hombre	Día	Mayor	Camioneta sport o blazer	Izabal	Lesionado	Colisión
5	Mujer	Noche	Mayor	Camión	Jutiapa	Lesionado	Vuelco

Nota. Elaborado en Word 365 y obtenido de procesamiento mediante el lenguaje R versión 4.5.1.

Tabla 3.

Árbol - Predicciones para el tipo de evento

	1	2	3	4	5	6	7	8
1	0.319	0.017	0.083	0.176	0.401	0.005	0.001	0.319
2	0.202	0.095	0.022	0.653	0.008	0.019	0.001	0.202
3	0.213	0.371	0.103	0.240	0.012	0.060	0.000	0.213
4	0.202	0.095	0.022	0.653	0.008	0.019	0.001	0.202
5	0.213	0.371	0.103	0.240	0.012	0.060	0.000	0.213

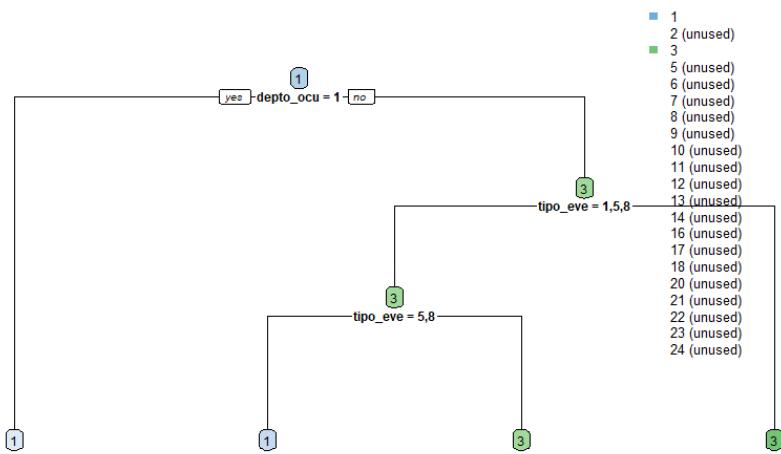
Nota. Elaborado en Word 365 y obtenido de procesamiento mediante el lenguaje R versión 4.5.1.

4.1.2. Predicción de tipo de vehículo

Figura 2.

Representación gráfica de árbol para tipo de vehículo

Árbol 3: Tipo de vehículo



Nota. Elaboración propia. Obtenido mediante biblioteca rpart.plot procesando mediante el lenguaje R versión 4.5.1.

Tabla 4.

Árbol - Predicciones para el tipo de vehículo

	1	2	3	5	6	7	9	10	11	12	14	18	22
1	0.51	0.09	0.12	0.04	0.02	0.02	0.04	0.01	0.01	0.11	0.02	0.01	0.00
2	0.36	0.10	0.24	0.10	0.04	0.01	0.04	0.00	0.01	0.04	0.03	0.01	0.01
3	0.10	0.08	0.36	0.09	0.02	0.05	0.09	0.00	0.01	0.18	0.03	0.00	0.01
4	0.20	0.11	0.27	0.06	0.02	0.02	0.14	0.00	0.01	0.08	0.06	0.01	0.01
5	0.10	0.08	0.36	0.09	0.02	0.05	0.09	0.00	0.01	0.18	0.03	0.00	0.01

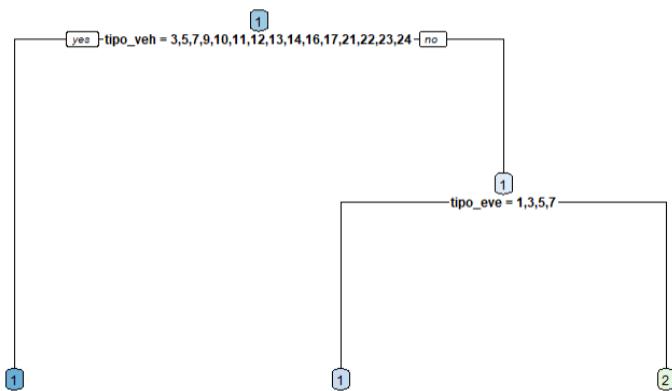
Nota. Elaborado en Word 365 y obtenido de procesamiento mediante el lenguaje R versión 4.5.1.

4.1.3. Predicción de jornada

Figura 3.

Representación gráfica de árbol para jornada

Árbol 4: Horario



Nota. Elaboración propia. Obtenido mediante biblioteca rpart.plot procesando mediante el lenguaje R versión 4.5.1.

Tabla 5.

Árbol - Predicciones para el horario o jornada de ocurrencia

	1	2
1	0.40	0.60
2	0.55	0.45
3	0.69	0.31
4	0.55	0.45
5	0.69	0.31

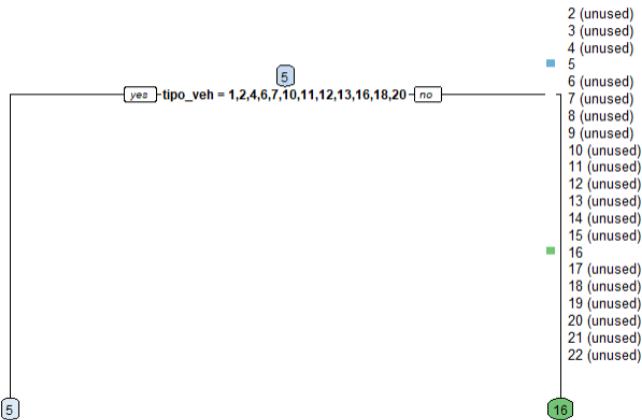
Nota. Elaborado en Word 365 y obtenido de procesamiento mediante el lenguaje R versión 4.5.1.

4.1.4. Predicción de departamento

Figura 4.

Representación gráfica de árbol para departamento

Árbol 5: Departamento de ocurrencia



Nota. Elaboración propia. Obtenido mediante biblioteca rpart.plot procesando mediante el lenguaje R versión 4.5.1.

Tabla 6.

Árbol - Predicciones para el departamento

	2	3	4	5	6	9	10	11	13	14	16	17	18	22
1	0.04	0.06	0.06	0.16	0.06	0.05	0.05	0.05	0.03	0.03	0.07	0.05	0.06	0.05
2	0.04	0.06	0.06	0.16	0.06	0.05	0.05	0.05	0.03	0.03	0.07	0.05	0.06	0.05
3	0.05	0.03	0.06	0.07	0.05	0.05	0.05	0.03	0.05	0.07	0.12	0.05	0.04	0.05
4	0.04	0.06	0.06	0.16	0.06	0.05	0.05	0.05	0.03	0.03	0.07	0.05	0.06	0.05
5	0.05	0.03	0.06	0.07	0.05	0.05	0.05	0.03	0.05	0.07	0.12	0.05	0.04	0.05

Nota. Elaborado en Word 365 y obtenido de procesamiento mediante el lenguaje R versión 4.5.1.

4.2. Bosques Aleatorios

Tabla 7.

Resumen – Bosques Aleatorios

	Predictoras	Target	Arboles	mtry	Test %	Train %	OOB%	Accuracy
1	6	tipo_eve	50	3	0.3	0.7	46.75	0.54
2	6	tipo_veh	100	4	0.3	0.7	59.93	0.40
3	6	horario	100	4	0.3	0.7	38.2	0.61
4	6	sexo_per	50	3	0.3	0.7	28.95	0.71

5	5	depto_ocu	50	3	0.3	0.7	84.69	0.16
----------	---	-----------	----	---	-----	-----	-------	------

Nota. Elaborado en Word 365 y obtenido de procesamiento mediante el lenguaje R versión 4.5.1.

4.2.1. Predicción de tipo de evento

Tabla 8.

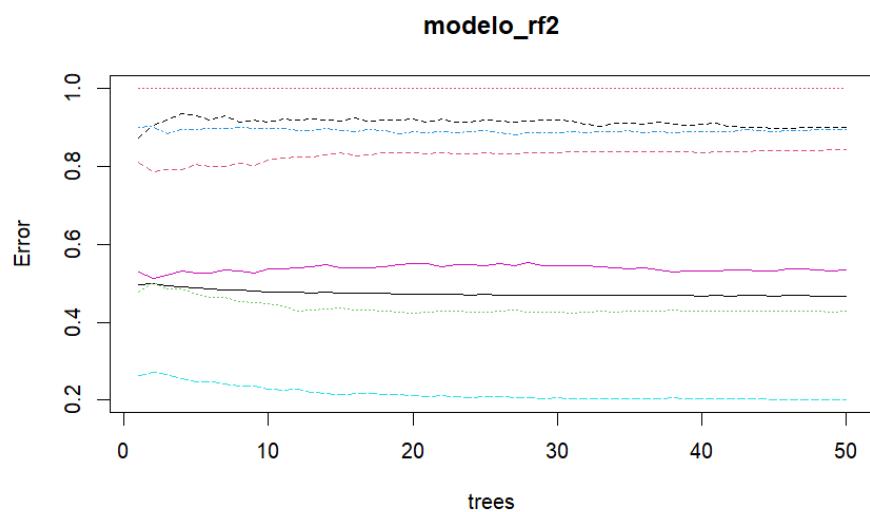
Matriz de confusión – Tipo de evento

	2	3	4	5	6	7	8	class.error
2	854	888	63	2613	934	25	0	0.8411754
3	432	2411	99	1192	56	24	0	0.4278595
4	133	436	190	715	269	32	0	0.8929577
5	542	923	79	9280	748	35	0	0.2004825
6	88	48	6	1399	1345	0	0	0.533957
7	74	290	21	238	8	71	0	0.8988604
8	2	1	0	8	2	0	0	1

Nota. Elaborado en Word 365 y obtenido de procesamiento mediante el lenguaje R versión 4.5.1.

Figura 5.

Grafico de error de bosque aleatorio para tipo de evento



Nota. Elaboración propia. Obtenido mediante biblioteca randomForest y plot, procesando mediante el lenguaje R versión 4.5.1.

Tabla 9.

RF - Predicciones de tipo de evento

	2	3	4	5	6	7	8
1	0	0	0	0.94	0.06	0	0
2	0.08	0.28	0.02	0.62	0	0	0
3	0.12	0.84	0.04	0	0	0	0
4	0.62	0.12	0	0.26	0	0	0
5	0.08	0.74	0.06	0.12	0	0	0

Nota. Elaborado en Word 365 y obtenido de procesamiento mediante el lenguaje R versión 4.5.1.

4.2.2. Predicción de tipo de vehículo

Tabla 10.

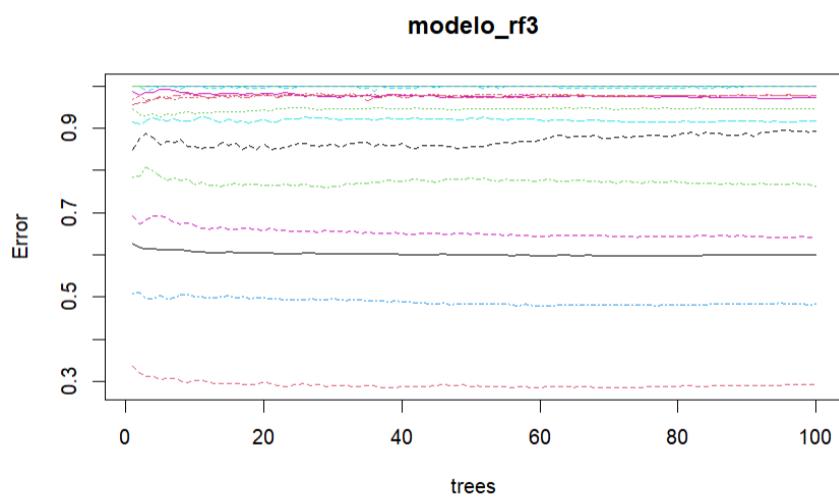
Matriz de confusión – Tipo de Vehículo

	1	2	3	5	6	7	8	9	10	11	12	13	14	16	17	18	20	21	22	23	24	class.error
1	5555	101	1512	91	20	20	1	271	1	5	248	0	14	0	0	1	0	0	4	0	0	0.29
2	1242	148	899	45	4	17	0	164	0	1	151	0	12	0	0	0	0	0	3	0	0	0.94
3	2131	166	3700	187	14	24	0	497	0	1	412	1	24	0	0	1	0	0	8	0	0	0.48
5	735	36	797	161	17	5	0	103	0	1	100	0	3	0	0	1	0	0	1	0	0	0.92
6	370	15	192	19	18	4	0	14	1	0	27	0	0	0	0	0	0	0	0	0	0	0.97
7	174	16	226	1	2	75	0	52	0	0	152	0	0	0	0	1	0	0	0	0	0	0.89
8	32	2	22	3	0	0	0	5	0	0	5	0	0	0	0	0	0	0	0	0	0	1.00
9	602	23	1013	43	0	15	0	582	0	2	169	0	6	0	0	1	0	0	0	0	0	0.76
10	50	0	6	0	1	0	0	2	0	0	2	0	0	0	0	0	0	0	0	0	0	1.00
11	147	2	82	3	0	1	0	20	0	0	23	0	4	0	0	0	0	0	0	0	0	1.00
12	888	27	786	37	1	32	0	177	0	0	1094	0	12	0	0	0	0	0	0	0	0	0.64
13	15	0	8	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.00
14	323	14	433	6	0	2	0	121	0	0	79	0	22	0	0	0	0	0	0	0	0	0.98
16	6	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.00
17	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.00
18	92	2	49	4	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	1.00
20	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1.00
21	0	0	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.00
22	57	5	76	4	0	1	0	15	0	0	10	0	4	0	0	0	0	0	4	0	0	0.98
23	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.00
24	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.00

Nota. Elaborado en Word 365 y obtenido de procesamiento mediante el lenguaje R versión 4.5.1.

Figura 6.

Gráfico de error de bosque aleatorio para tipo de vehículo



Nota. Elaboración propia. Obtenido mediante biblioteca randomForest y plot, procesando mediante el lenguaje R versión 4.5.1.

Tabla 11.

RF - Predicciones de tipo de vehículo

	1	2	3	5	6	7	9	12	14	18
1	0.64	0.04	0	0.06	0.03	0	0.04	0.19	0	0
2	0.38	0.19	0.25	0.07	0.01	0.01	0.03	0.02	0.01	0.03
3	0	0	1	0	0	0	0	0	0	0
4	0	0	0.99	0	0	0	0.01	0	0	0
5	0.1	0.02	0.23	0	0	0.01	0.07	0	0.57	0

Nota. Elaborado en Word 365 y obtenido de procesamiento mediante el lenguaje R versión 4.5.1.

4.2.3. Predicción de jornada

Tabla 12.

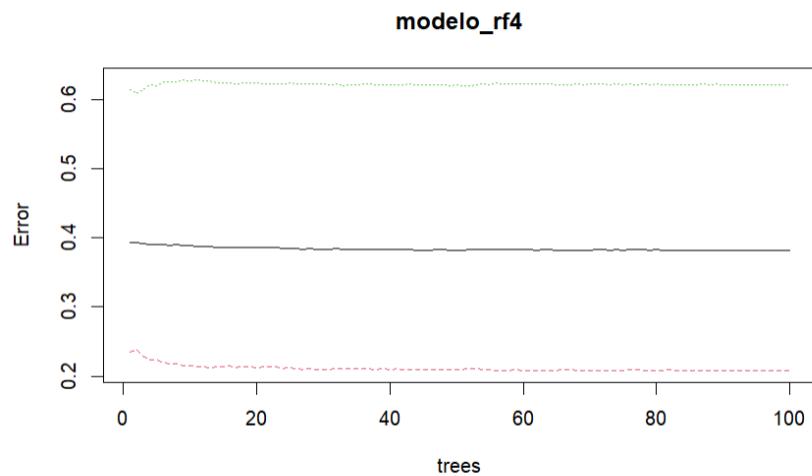
Matriz de confusión – Horario o jornada

	1	2	class.error
1	25597	6717	0.2078666
2	14594	8884	0.6216032

Nota. Elaborado en Word 365 y obtenido de procesamiento mediante el lenguaje R versión 4.5.1.

Figura 7.

Gráfico de error de bosque aleatorio para jornada



Nota. Elaboración propia. Obtenido mediante biblioteca randomForest y plot, procesando mediante el lenguaje R versión 4.5.1.

Tabla 13.

RF - Predicciones de jornada

	1	2
1	0.15	0.85
2	0.6	0.4
3	0.93	0.07
4	0.7	0.3
5	0.42	0.58

Nota. Elaborado en Word 365 y obtenido de procesamiento mediante el lenguaje R versión 4.5.1.

4.2.4. Predicción de sexo de la persona

Tabla 14.

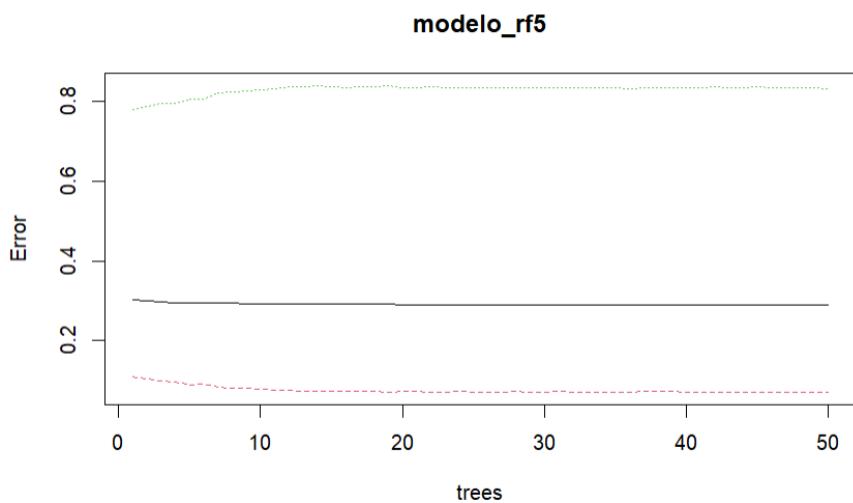
Matriz de confusión – Sexo de la persona

	1	2	class.error
1	36954	2868	0.0720049
2	13285	2685	0.8318723

Nota. Elaborado en Word 365 y obtenido de procesamiento mediante el lenguaje R versión 4.5.1.

Figura 8.

Gráfico de error de bosque aleatorio para sexo de la persona



Nota. Elaboración propia. Obtenido mediante biblioteca randomForest y plot, procesando mediante el lenguaje R versión 4.5.1.

Tabla 15.

RF - Predicciones de sexo de la persona

	1	2
1	1	0
2	0.42	0.58
3	0.84	0.16
4	0.72	0.28

5	0.86	0.14
---	------	------

Nota. Elaborado en Word 365 y obtenido de procesamiento mediante el lenguaje R versión 4.5.1.

4.2.5. Predicción de departamento

Tabla 16.

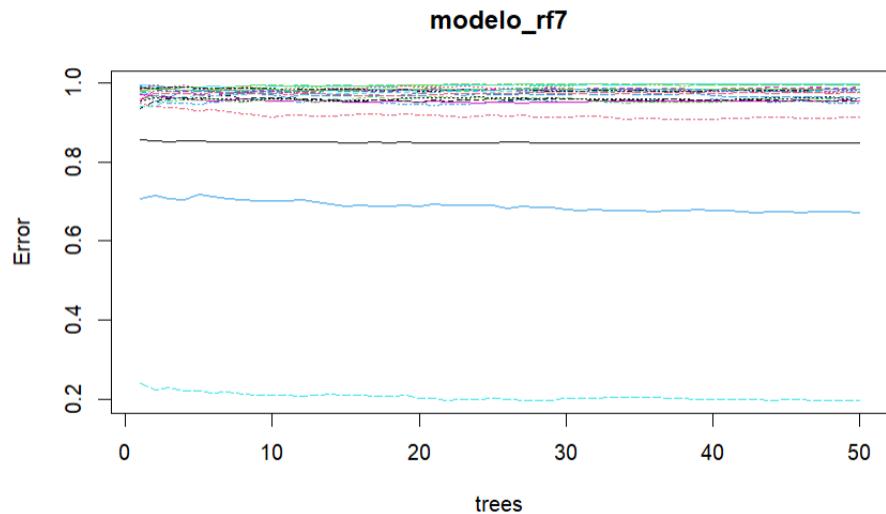
Matriz de confusión – Departamento

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	class.error
2	20	4	43	960	19	9	2	27	49	6	5	12	19	1	258	13	9	5	9	0	6	0.9864499
3	3	11	59	1437	30	2	1	47	27	5	12	15	19	0	257	14	9	13	3	0	7	0.9944191
4	14	8	116	1456	27	4	0	62	40	2	10	6	49	0	364	14	8	6	8	1	5	0.9472727
5	19	14	124	3802	21	6	0	88	35	15	21	8	21	4	439	33	34	19	7	0	29	0.197721
6	11	8	66	1345	95	15	2	38	36	4	9	20	42	1	295	17	6	6	4	0	16	0.9533399
7	7	6	46	498	22	48	4	26	43	1	10	14	34	0	264	8	7	0	6	2	11	0.9545885
8	4	1	27	253	4	7	13	23	11	3	5	4	19	0	125	5	1	0	3	0	6	0.9747082
9	9	7	67	1148	32	8	0	70	34	3	6	8	22	1	298	24	8	10	6	0	13	0.9605411
10	24	10	48	1220	7	4	0	53	69	5	7	3	27	0	267	14	17	8	5	2	11	0.961688
11	16	2	45	1146	3	3	0	20	32	4	3	5	11	0	198	7	8	8	3	1	0	0.9973597
12	4	13	55	818	14	10	0	23	31	3	17	10	33	1	267	23	11	7	10	4	5	0.9874908
13	5	11	45	622	37	15	3	17	16	3	8	52	44	7	323	16	8	1	4	1	16	0.9585327
14	8	12	66	647	19	7	3	32	32	2	14	9	130	1	428	17	13	2	29	3	15	0.9126931
15	1	5	36	738	22	3	1	9	10	2	3	13	37	7	307	11	8	6	9	0	7	0.994332
16	18	26	72	1509	24	17	8	55	58	7	21	27	93	7	992	38	12	10	15	5	17	0.6727153
17	9	7	86	1363	16	7	0	28	26	3	14	9	28	1	307	43	15	6	4	4	6	0.9783047
18	8	3	49	1344	4	6	1	21	26	5	9	5	27	3	249	27	39	18	8	2	9	0.979066
19	9	1	47	1091	4	2	1	23	20	3	7	1	29	2	217	18	13	25	4	0	6	0.983585
20	7	1	28	874	20	2	0	21	15	0	8	7	58	0	189	11	10	9	34	0	5	0.973826
21	12	1	25	592	14	4	3	6	18	2	3	12	30	1	200	14	2	2	3	17	10	0.9824923
22	11	1	58	1146	27	6	0	35	40	4	3	22	34	2	236	19	12	9	4	3	28	0.9835294

Nota. Elaborado en Word 365 y obtenido de procesamiento mediante el lenguaje R versión 4.5.1.

Figura 9.

Gráfico de error de bosque aleatorio para departamento



Nota. Elaboración propia. Obtenido mediante biblioteca randomForest y plot, procesando mediante el lenguaje R versión 4.5.1.

Tabla 17.

RF - Predicciones de departamento

	3	4	5	6	7	8	9	10	12	13	14	15	16	17	19	21	22
1	0	0	0.8	0.04	0	0	0	0.02	0.04	0	0	0	0.02	0.06	0	0	0.02
2	0.04	0.2	0.16	0	0.02	0	0.04	0.2	0.06	0	0.02	0	0.16	0.04	0	0.02	0.04
3	0	0.02	0	0	0	0	0	0	0	0.38	0.12	0	0.48	0	0	0	0
4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0.02	0	0.16	0.02	0	0.16	0	0	0	0.08	0.02	0.08	0.36	0.08	0.02	0	0

Nota. Elaborado en Word 365 y obtenido de procesamiento mediante el lenguaje R versión 4.5.1.

4.3. Redes Neuronales

Tabla 18.

Resumen – Redes Neuronales

Predictoras	Target	Salida	Capas	Neu			Epochs	Loss	Accuracy	Test (%)
				1	2	3				
1	6	tipo_eve	8	3	8	8	50	1.29	0.47	20
2	6	tipo_veh	24	3	16	16	24	50	1.85	0.37
3	6	horario	2	3	8	4	2	50	0.67	0.59
4	6	sexo_per	2	3	8	4	2	50	0.56	0.72
5	5	depto_ocu	22	3	16	16	22	50	2.9	0.14

Nota. Elaborado en Word 365 y obtenido de procesamiento mediante el lenguaje R versión 4.5.1.

4.3.1. Predicción de tipo de evento

Tabla 19.

RN - Predicciones de tipo de evento

	2	3	4	5	6	7	8
1	0.16	0.03	0.04	0.51	0.27	0.00	0.00
2	0.17	0.08	0.01	0.73	0.00	0.02	0.00
3	0.17	0.31	0.07	0.38	0.01	0.06	0.00
4	0.22	0.26	0.05	0.41	0.01	0.05	0.00
5	0.27	0.15	0.13	0.34	0.10	0.02	0.00

Nota. Elaborado en Word 365 y obtenido de procesamiento mediante el lenguaje python y tensorflow.

4.3.2. Predicción de tipo de vehículo

Tabla 20.

RN - Predicciones de tipo de vehículo

	1	2	3	5	6	7	8	9	17	18	22
1	0.30	0.10	0.19	0.05	0.01	0.04	0.00	0.05	0.00	0.00	0.01
2	0.47	0.13	0.19	0.08	0.05	0.01	0.00	0.01	0.00	0.01	0.00
3	0.03	0.06	0.52	0.05	0.00	0.03	0.00	0.09	0.00	0.00	0.01
4	0.16	0.11	0.30	0.07	0.02	0.02	0.01	0.17	0.00	0.01	0.01
5	0.36	0.10	0.32	0.08	0.02	0.01	0.00	0.06	0.00	0.00	0.00

Nota. Elaborado en Word 365 y obtenido de procesamiento mediante el lenguaje python tensorflow.

4.3.3. Predicción de jornada

Tabla 21.

RN - Predicciones de jornada

	1	2
1	0.55	0.45
2	0.57	0.43
3	0.76	0.24
4	0.60	0.40
5	0.65	0.35

Nota. Elaborado en Word 365 y obtenido de procesamiento mediante el lenguaje python y tensorflow.

4.3.4. Predicción de sexo de la persona

Tabla 22.

RF - Predicciones de sexo de la persona

	1	2
1	0.90	0.10
2	0.75	0.25
3	0.53	0.47
4	0.57	0.43
5	0.74	0.26

Nota. Elaborado en Word 365 y obtenido de procesamiento mediante el lenguaje Python y tensorflow.

4.3.5. Predicción de departamento

Tabla 23.

RF - Predicciones de departamento

	2	3	4	5	6	7	8	9	10	13	14	15	16	17	18	19	20	21	22
1	0.03	0.04	0.05	0.16	0.06	0.02	0.01	0.04	0.04	0.03	0.03	0.04	0.08	0.09	0.07	0.04	0.04	0.03	0.05
2	0.04	0.06	0.08	0.12	0.06	0.05	0.03	0.06	0.03	0.03	0.03	0.03	0.07	0.04	0.04	0.04	0.04	0.02	0.06
3	0.04	0.03	0.06	0.06	0.05	0.04	0.02	0.05	0.04	0.07	0.09	0.04	0.11	0.05	0.04	0.04	0.04	0.03	0.05
4	0.05	0.05	0.06	0.12	0.06	0.03	0.02	0.07	0.04	0.03	0.03	0.03	0.05	0.06	0.05	0.05	0.04	0.02	0.06
5	0.02	0.06	0.05	0.13	0.03	0.03	0.02	0.04	0.04	0.03	0.03	0.04	0.17	0.03	0.07	0.04	0.02	0.06	0.05

Nota. Elaborado en Word 365 y obtenido de procesamiento mediante el lenguaje Python y tensorflow.

5. DISCUSIÓN DE RESULTADOS

5.1. Árboles de Decisión

Se analizó cuatro variables para predecir cinco casos desconocidos con los cuales se encontraron hechos interesantes de estudio. Al analizar la consecuencia del accidente, el mes en el que ocurrió, el sexo de la persona y si era mayor o menor de edad no se obtienen ramificaciones para los árboles correspondientes. No obstante, se deja constancia en el R Notebook haciendo un balanceo de datos para la variable fall_les, debido a que la mayoría de los casos concluyen en lesiones ocasionadas por alguna colisión mientras se conduce en motocicleta y el afectado es mayor de edad, con mayor probabilidad de incidencia en los meses de diciembre y marzo, así también se observa que se presentan menores casos en junio.

Resalta un caso de lesionado con 65% de probabilidad que sea provocada por una caída de una camioneta en el departamento de Izabal si la persona es hombre mayor de edad. Pero de manera general las probabilidades mas altas se las llevan los eventos que suceden a través de choques, atropellos y caídas.

Exceptuando las motocicletas se observa que el segundo transporte más probable de accidentes ya sea en fallecimientos o lesionados es el automóvil con probabilidades de 51% que fallezca alguien si es en Guatemala y un 36% que concluya en lesiones para el interior del país, esto para los casos en los que una mujer es la afectada.

Para los cinco casos predichos se observa que los incidentes suceden con mayor frecuencia de día con probabilidad superiores al 50% de las veces en el interior del país, con excepción del departamento de Guatemala donde se tiende con un 60% de las veces a que suceda más en jornada nocturna.

Exceptuando al departamento de Guatemala se analiza la predicción de accidentes por departamento para estos cinco casos desconocidos y se observa que tienen a concentrarse más en Escuintla en caso de colisiones, choques o atropellos y Alta Verapaz si se vuelca o derrapa en automóvil. No obstante, en todos los departamentos existen probabilidades que se den estos casos.

5.2. Bosques Aleatorios

Al someter a predicción los cinco casos y analizar el tipo de evento se ve que los choques, atropellos y vuelcos son los más probables. Específicamente en Guatemala suceden atropellos en motocicletas que concluyen en muertes en un 94% de las veces mientras que los choques y vuelcos están con más incidencia en el interior del país con más énfasis en las mujeres menores de edad.

Exceptuando la motocicleta se observa que el automóvil resalta más en las predicciones de estos cinco escenarios. No obstante, se obtiene que un accidente puede ocurrir en Izabal con un 99% de probabilidad si se conducen en camioneta y el afectado es hombre mayor de edad con lesiones. Además, se aprecia que el 64% de las veces que una mujer se conduce en automóvil en Guatemala fallece debido a conducir en jornada diurna a través de un choque.

Al igual que el modelo de árboles de decisión para los casos en los que la persona afectada sufre un accidente en el departamento de Guatemala se obtiene alta probabilidad de que suceda de noche y se trate de una mujer

mayor de edad, mientras que en los otros casos hay tendencia a que suceda de noche, pues ya depende de los medios de transporte y lugar de ocurrencia.

Tal y como se muestra en la tabla 15 los hechos en Guatemala que manejan motocicleta están inclinados a que son reportes de hombres mayores de edad mientras que para el caso en Santa Rosa, por ejemplo, la probabilidad es el 58% de que la afectada sea menor de edad conduciéndose en un automóvil en jornada nocturna.

Al igual que en el modelo de árboles de decisión acá resaltan más las probabilidades de ocurrencia de accidentes en el departamento de Escuintla, Alta Verapaz, pero también Huehuetenango.

5.3. Redes Neuronales

Se obtuvo el mismo patrón de que los casos suceden en jornada diaria, donde el afectado es hombre que se conduce a través de automóvil o pick-up. Además, con base en los cinco escenarios predichos sobre casos desconocidos también se refleja mayor probabilidad de ocurrencia q4e 36s accidentes se den en los departamentos de Escuintla, Alta Verapaz y Huehuetenango.

5.4. Comparación entre modelos

Para los cinco escenarios definidos y manteniendo las mismas variables predictoras (sexo, horario, mayor/menor, tipo de vehículo, departamento y, cuando aplica, tipo de evento), se contrastaron las probabilidades y clases predichas por Árboles de decisión, Bosques Aleatorios y Redes Neuronales.

5.4.1. Tipo de evento (tipo_eve)

Con motocicleta y jóvenes, las clases Colisión/Choque/Vuelco concentran mayor probabilidad en los tres modelos. RF y RN capturan mejor la combinación de predictores (tipo_veh + horario + depto_ocu) y tienden a elevar la probabilidad de Colisión/Choque en escenarios de capital y noche.

El Árbol es presenta reglas explícitas, pero puede perder precisión cuando hay muchas clases y categorías. Estos resultados son coherentes con los patrones de fase 1, donde colisión/choque/vuelco emergieron como eventos dominantes.

5.4.2. Tipo de vehículo (tipo_veh)

Los tres modelos asignan mayor probabilidad a vehículos comunes (automóvil, camioneta, pick-up, moto, camión) según el escenario. La red neuronal puede captar mejores interacciones no lineales, como el efecto del horario nocturno sea distinto según departamento, mientras que RF conserva estabilidad y el árbol preserva la rigidez de las reglas.

La alta concentración de motocicletas en escenarios de riesgo (jóvenes, noche, capital/interior con alta circulación) se confirma de forma consistente, en línea con la Fase 1.

5.5. Comparación por variable en común

- **sexo_per:** Los tres modelos muestran que el sexo impacta menos que tipo_eve y tipo_veh; el efecto se vuelve significativo cuando coincide con edad joven y moto.

- **horario:** El horario nocturno incrementa la probabilidad de incidentes graves en RF y RN; el Árbol lo refleja en reglas superiores cuando combina con moto y choque/collision.
- **mayor_menor:** La condición menor refuerza el riesgo en combinación con moto y eventos críticos; RF/RN ajustan mejor este efecto conjunto.
- **tipo_veh:** Motocicleta es el predictor más determinante; los tres modelos elevan su probabilidad de severidad/eventos de alto riesgo.
- **depto_ocu:** Departamentos con alta densidad vehicular concentran mayor probabilidad de eventos críticos; RF/RN ajustan mejor el patrón espacial, mientras el Árbol aporta reglas claras para comunicar el riesgo.

6. CONCLUSIONES

- 1)** Los árboles muestran reglas claras que permiten identificar factores críticos. Variables como tipo_eve (colisión, choque, vuelco) y tipo_veh (motocicleta) aparecen en los nodos superiores, indicando su alta relevancia en la gravedad del accidente.
- 2)** Los bosques reducen la varianza y mejoran la estabilidad frente a los árboles individuales. Las variables más importantes fueron tipo_eve, tipo_veh y horario, confirmando patrones detectados en la Fase 1. Las probabilidades predichas para los escenarios fueron más consistentes que en los árboles.
- 3)** Las redes neuronales logran capturar relaciones no lineales entre las variables, ofreciendo una precisión superior en algunos casos (accuracy <= 75%). Sin embargo, su interpretabilidad es limitada en comparación con los árboles y bosques. Las métricas de validación reflejan buen ajuste, aunque requieren mayor cuidado en el preprocesamiento (codificación y normalización).
- 4)** En general, los tres enfoques coinciden en que los accidentes graves están asociados a automóviles y vehículos típicos (exceptuando la motocicleta), eventos de choque, atropello y vuelco en horarios nocturnos en ciudad y diurnos en el interior del país, especialmente en departamentos como Santa Rosa, Escuintla, Baja Verapaz y Huehuetenango.

7. REFERENCIAS

Organización Mundial de la Salud (OMS). (2018). Informe sobre la situación mundial de la seguridad vial.

Wickham, H., & Grolemund, G. (2016). R for Data Science. O'Reilly Media.

Guatemala, I. N. (5 de Noviembre de 2025). *INE*. Obtenido de INE: <https://www.ine.gob.gt/bases-de-datos/accidentes-de-transito/>

González, B. R. (7 de 11 de 2025). *GitHub*. Obtenido de usac201612398/DataMining: <https://github.com/usac201612398/DataMining>