

ANÁLISIS DE DATOS

DESASTRES NATURALES

IMPACTO HUMANO Y ECONÓMICO

**Soluciones aplicadas de Análisis de Datos
y Aprendizaje Automático para entender y
prever la magnitud de fenómenos
naturales, con el fin de mitigar su impacto
nivel de destrucción.**

Equipo 2

Alejandra Eng
Ángel Bardon
Elena Blanco
Rafael De Marco

Fecha:

31/07/2023

Curso

Análisis de Datos
Escuela Digital
Fundación Adecco

OBJETIVOS

Académicos o de Aprendizaje:

Adquirir conocimientos y habilidades de análisis de datos en un contexto real, mediante metodología CRISP-DM en el tema sobre desastres naturales.

Aplicación Práctica:

Conocer la magnitud y el impacto socio-económico para evaluar riesgos y diseñar soluciones prácticas como consultoría de inteligencia de negocio, diseño de dashboards y modelos predictivos.



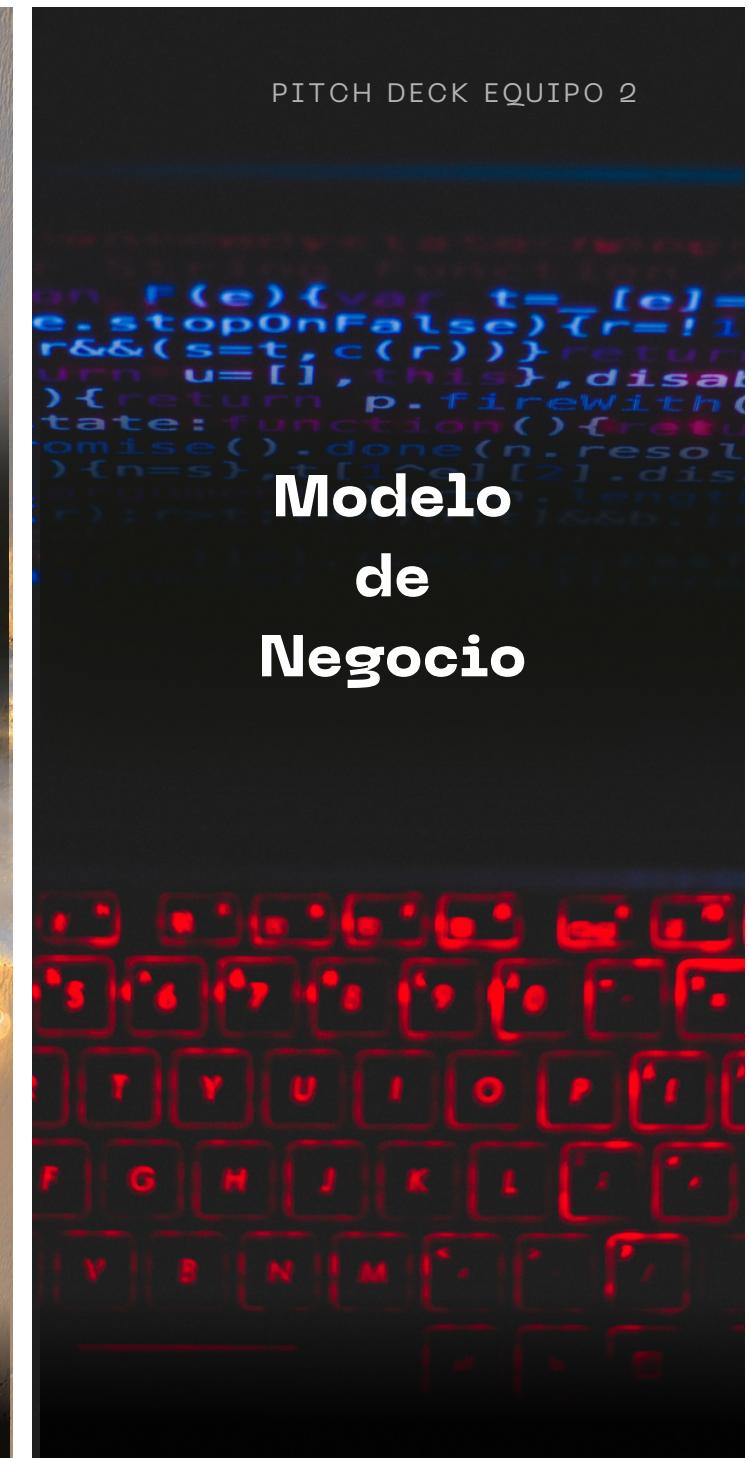
EDA:

Terremotos
Sequías
Tormentas
Inundaciones



Modelo Predictivo:

Nivel de Destrucción



Modelo de Negocio

CONTENIDOS



Metodología y
dataset em-dat
Técnicas



EDA
Sequías



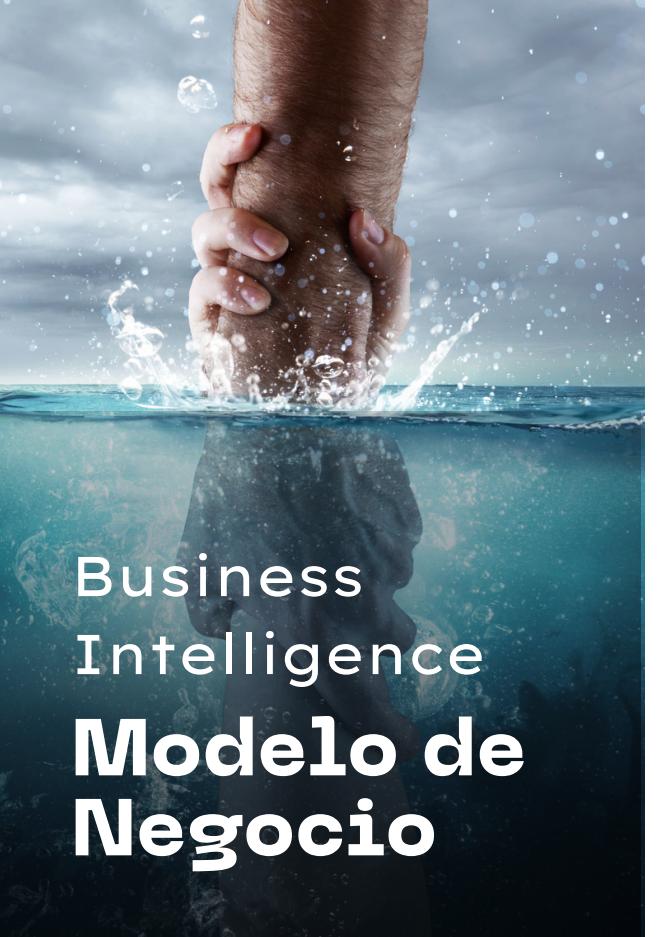
EDA
Tormentas



EDA
Terremotos



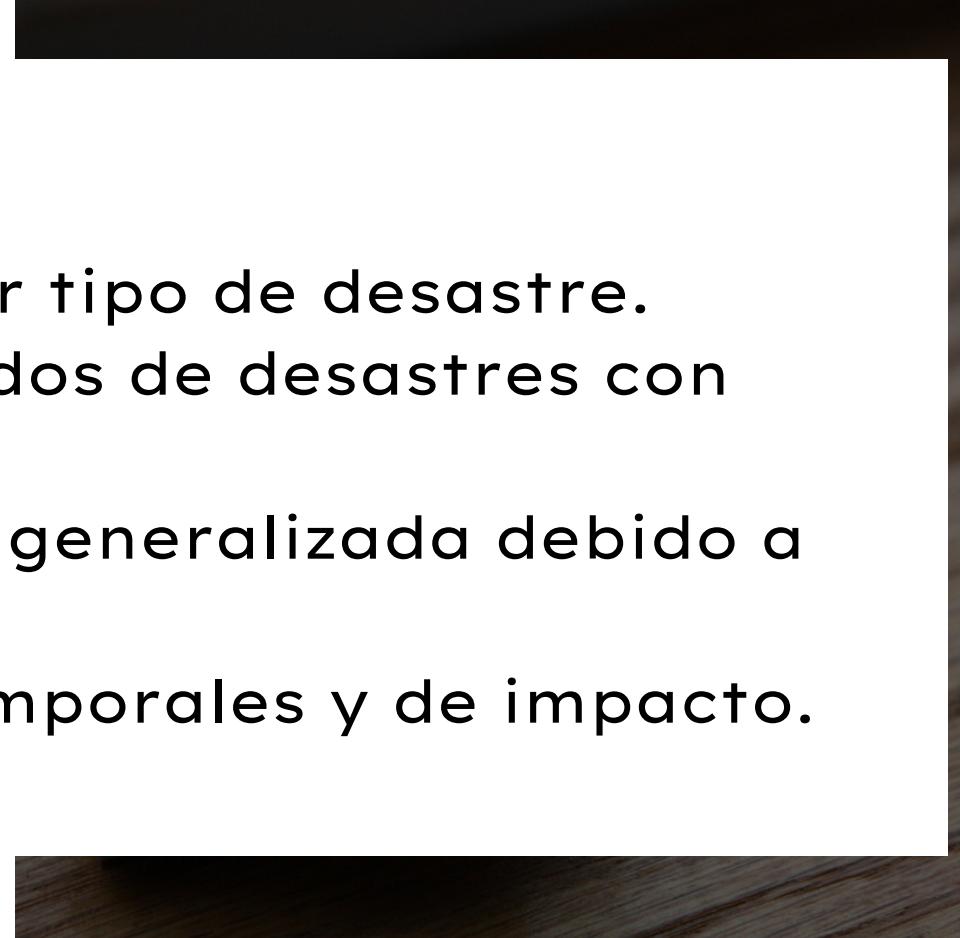
EDA y Modelo
Predictivo
Inundaciones



Business
Intelligence
**Modelo de
Negocio**

PROCESOS GENERALES EDA

- Selección de 4 tipos de desastres.
- Limpieza y segmentación con filtro por tipo de desastre.
- Outliers no se eliminan: son datos válidos de desastres con grandes impactos.
- Limpieza y relleno de valores nulos no generalizada debido a naturaleza del dataset.
- Análisis de categorías geográficas, temporales y de impacto.

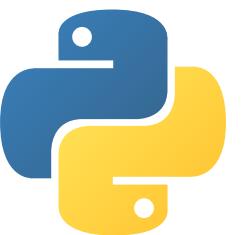


MACHINE LEARNING

Modelo predictivo de Nivel de Destrucción



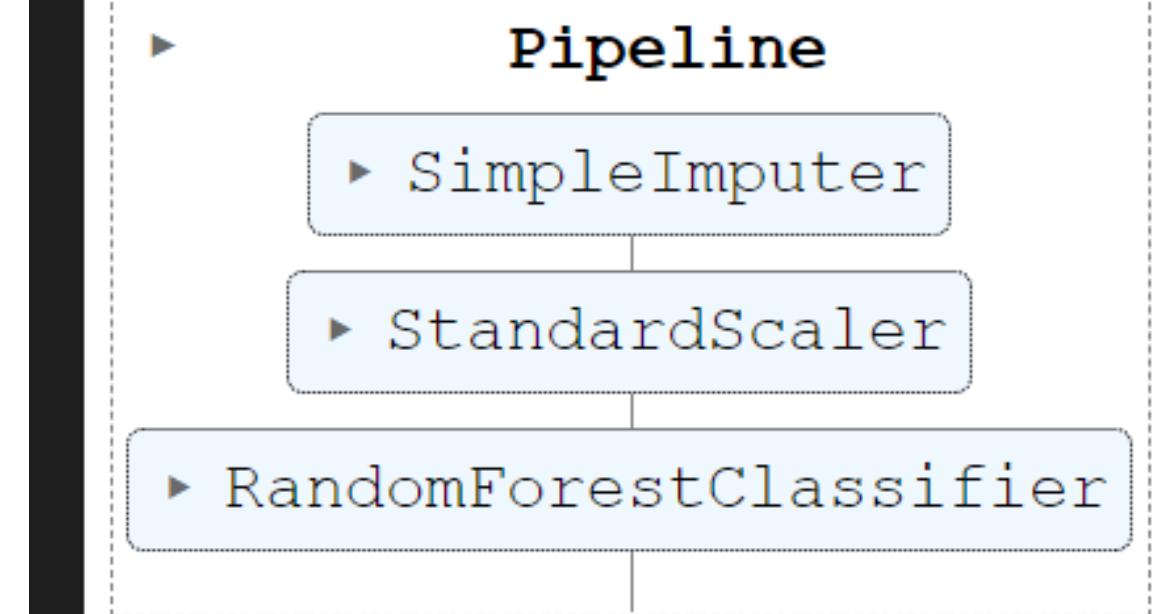
Herramientas



PROCESO DEL MODELO ML PREDICTIVO

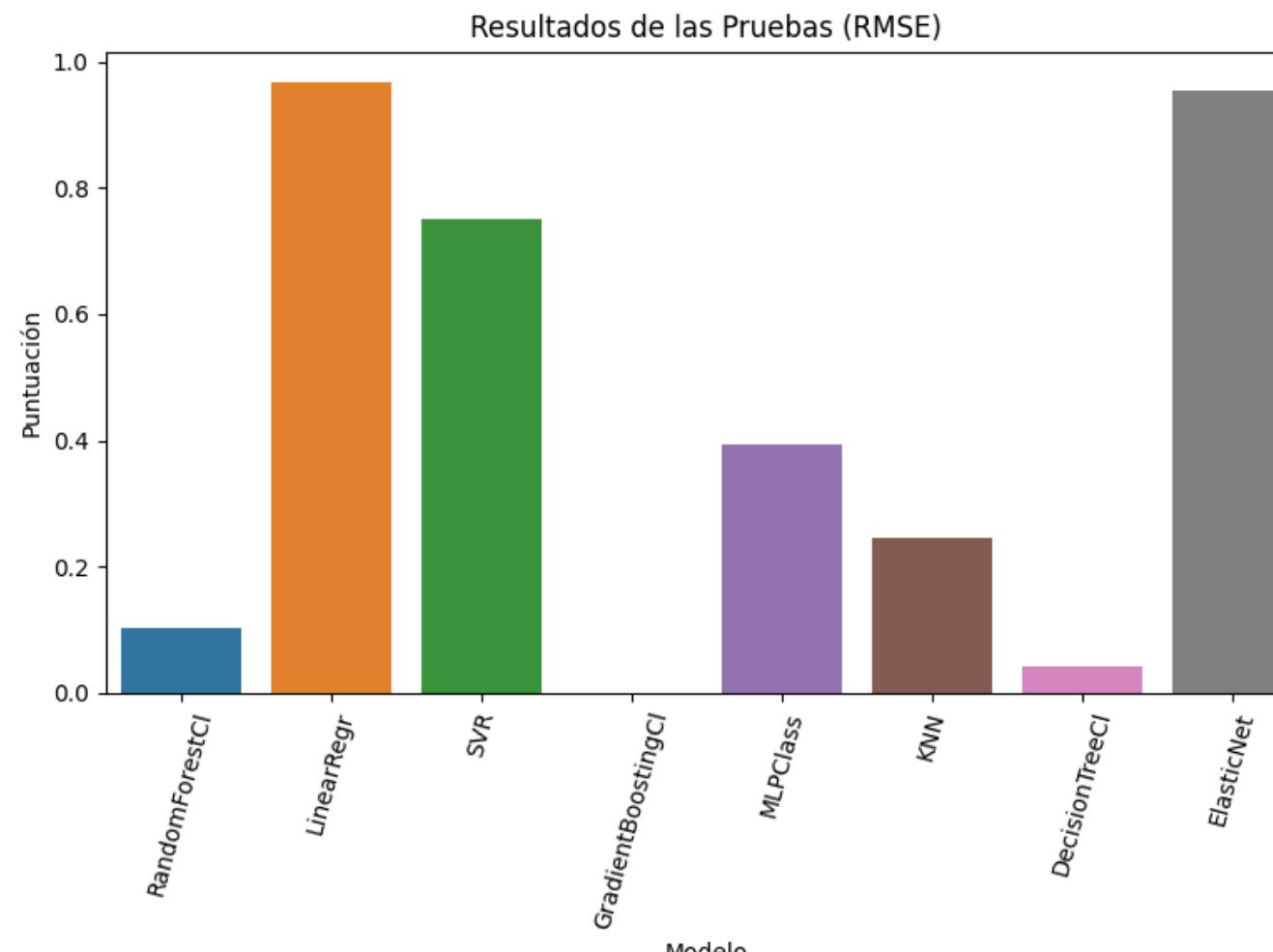
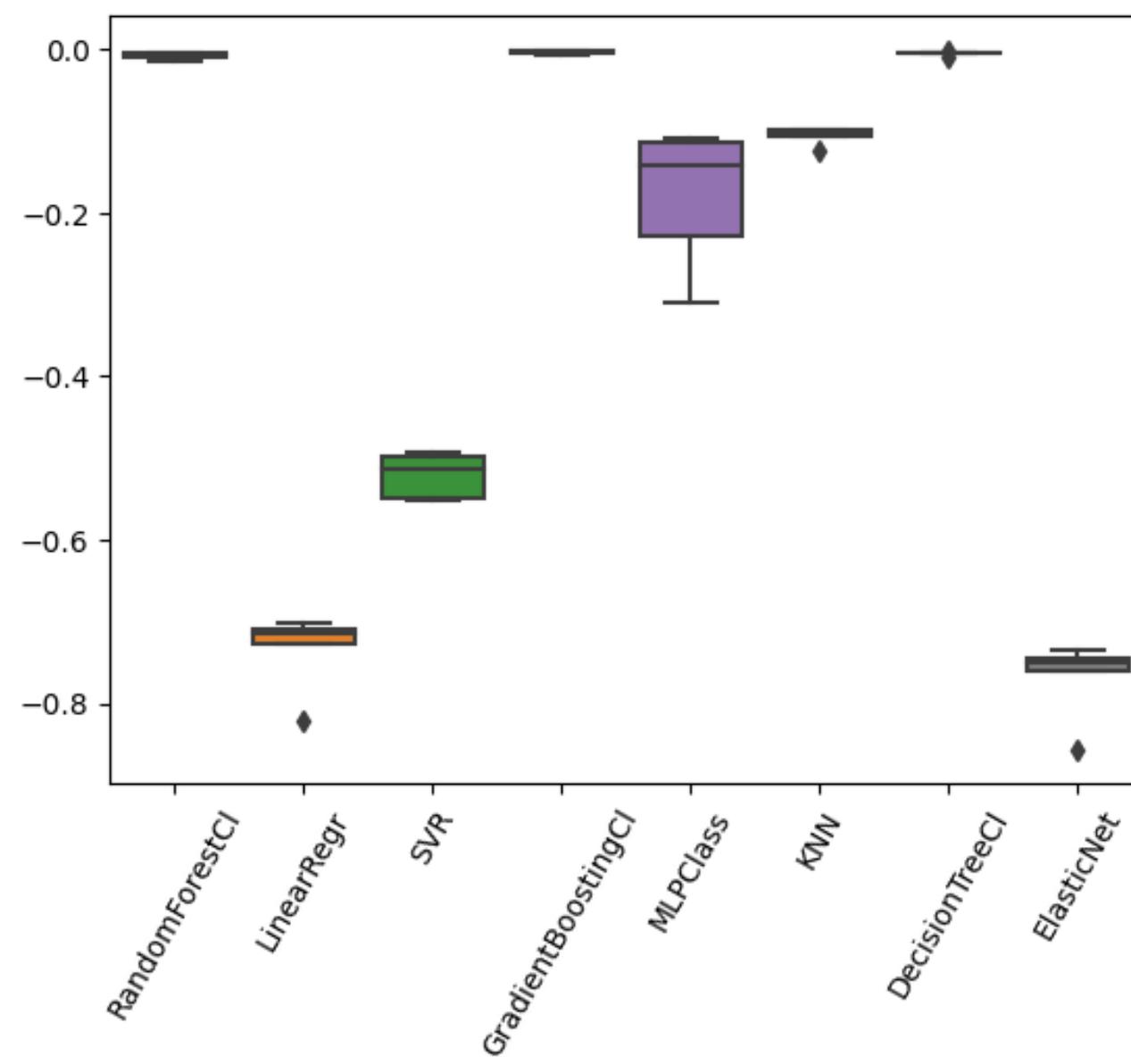
Prototipado de un modelo predictivo para anticipar el Nivel de Destrucción, con base en Valor de Magnitud de inundaciones (km2) y definición de clasificaciones del 1-4 por umbrales estadísticos basados en percentiles. Posteriormente, se asigna a cada clase una combinación de máximos de cada tipo de impacto: fallecidos, afectados y daños.

```
pipeline.fit(X_train, y_train)
```

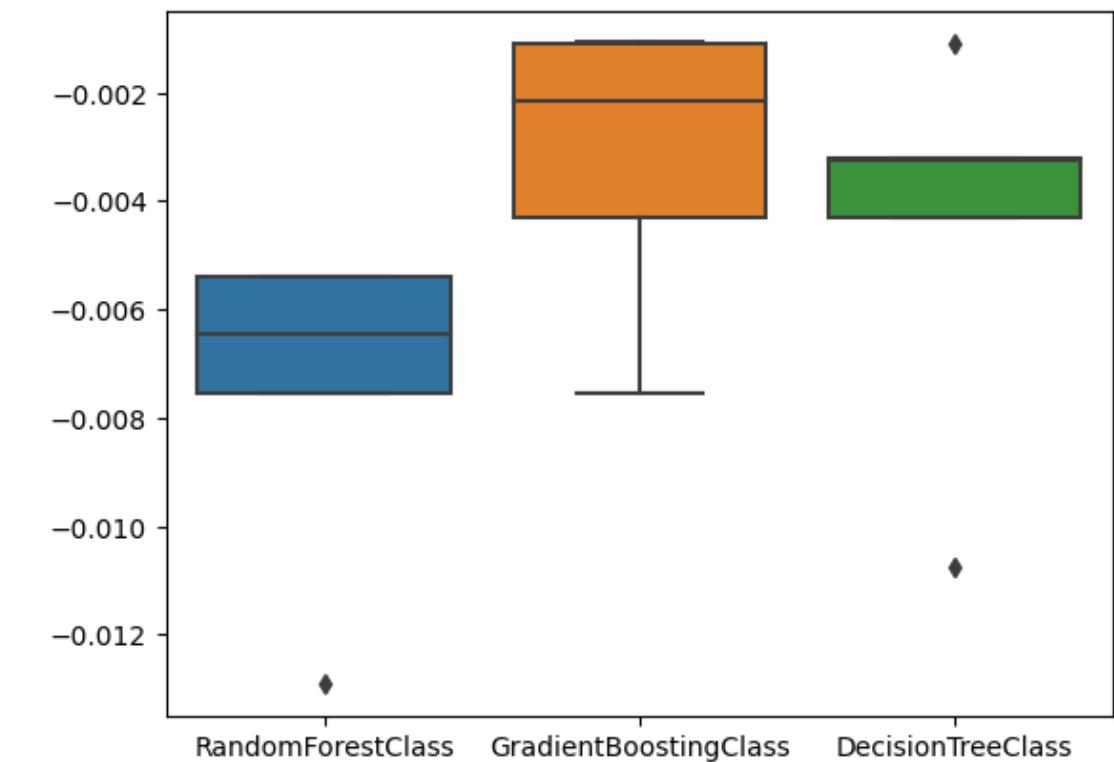


VALIDACIÓN DE ALGORITMOS

Análisis comparativo



Modelos finalistas:
 RandomForestClassifier
 GradientBoostingClassifier
 DecisionTrees



PREDICCIÓN DE NIVEL DE DESTRUCCIÓN

Modelo ML con Random Forest Classifier

```
# Evaluar desempeño del dataset
print("Reporte de clasificación:")
print(classification_report(y_pred, y_pred))
```

Reporte de clasificación:

	precision	recall	f1-score	support
1.0	1.00	1.00	1.00	33
2.0	1.00	1.00	1.00	114
3.0	1.00	1.00	1.00	98
4.0	1.00	1.00	1.00	111
accuracy			1.00	356
macro avg	1.00	1.00	1.00	356
weighted avg	1.00	1.00	1.00	356

RESULTADOS DE CLASSIFICATION REPORT

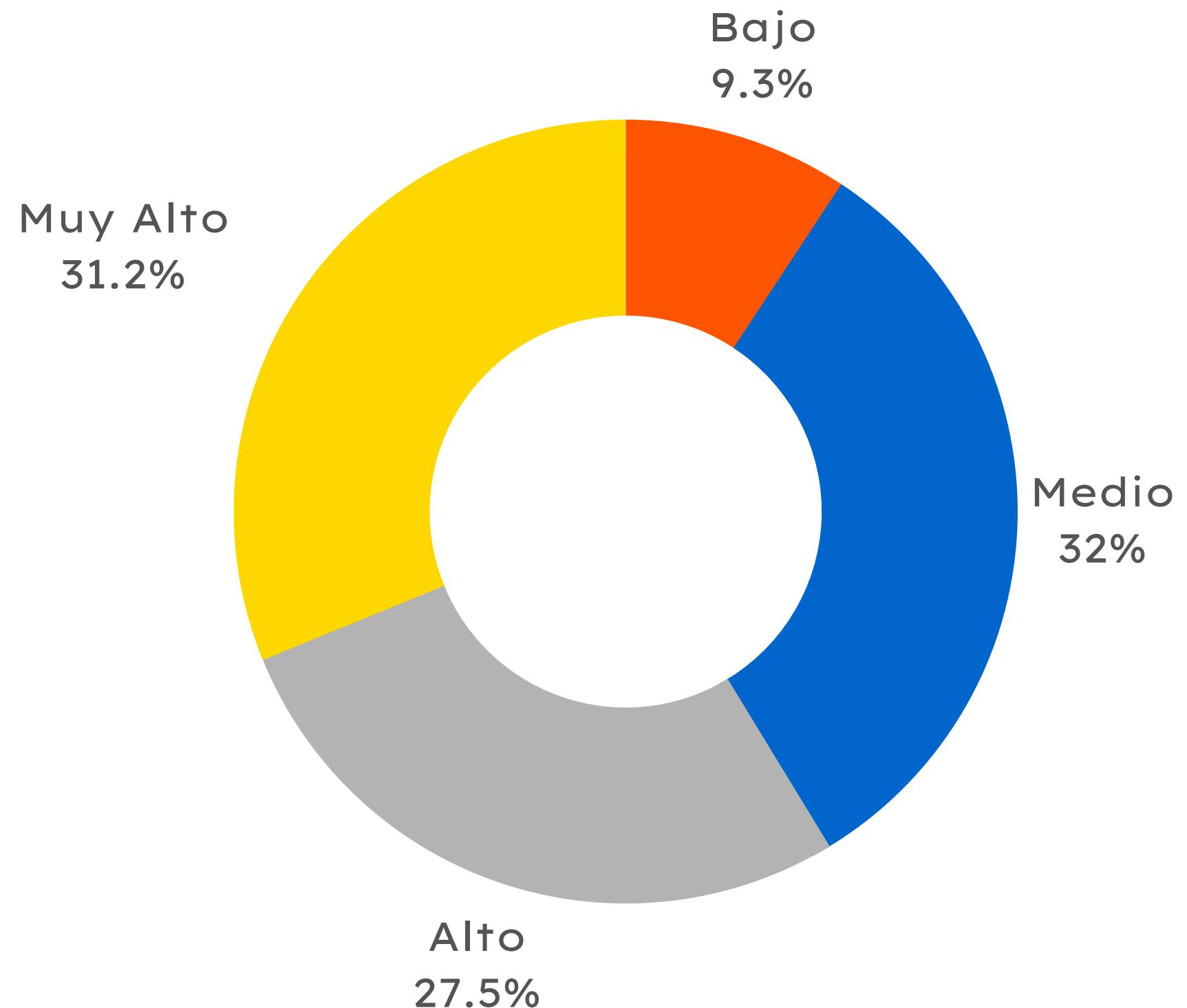
Métricas colectivas de nivel de precisión y capacidad predictiva:

MUY ALTA

Diagnóstico:

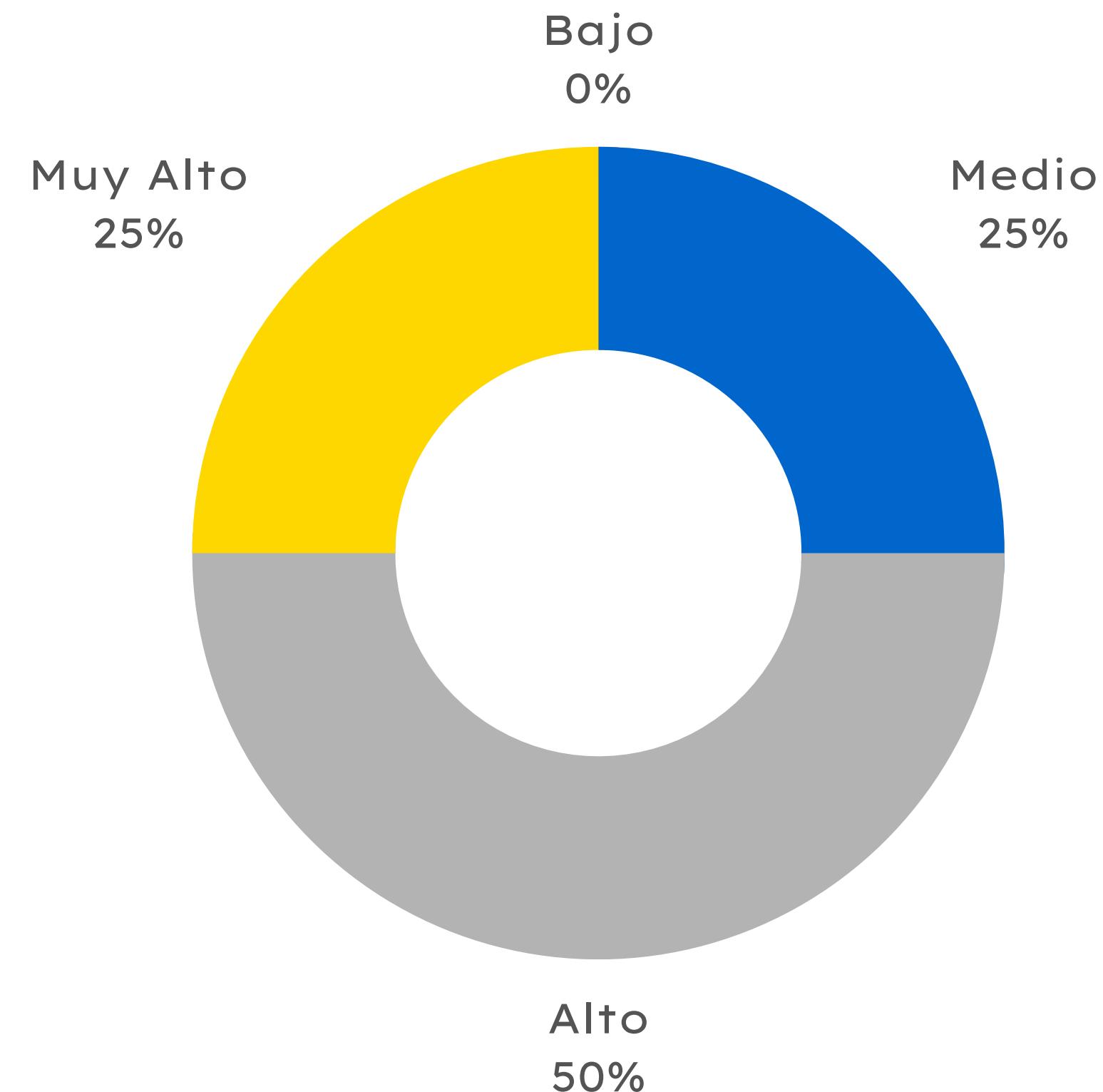
- 1) Posible overfitting.
- 2) Escasez de datos para niveles Bajo y Alto.
- 3) Entrenar con menos columnas.
- 4) Necesario ajustar umbrales con percentiles también en columnas de impacto.
- 5) Explorar otra configuración para umbrales no basada en estadística.

PREDICCIÓN DEL NIVEL DE DESTRUCCIÓN



INUNDACIONES PREDICCIÓN

PREDICCIÓN DEL NIVEL DE DESTRUCCIÓN



DESASTRES NATURALES

MODELO PREDICTIVO: TESTS

Test A: Controlado

```
X_destruction_controlled= pd.DataFrame({'Dis Mag Value': [10, 4500, 26000, 90000],  
                                         'Start Year': [2024, 2025, 2026, 2027],  
                                         'Start Month': [1, 5, 8, 12],  
                                         'Start Day': [6, 16, 24, 30],  
                                         'End Year': [2024, 2025, 2026, 2027],  
                                         'End Month': [1, 5, 9, 1],  
                                         'End Day': [8, 21, 4, 28],  
                                         'Total Deaths': [70, 250, 800, 10000],  
                                         'Total Affected': [5000, 25000, 300000, 1000000],  
                                         'Total Damages Adj': [200000, 5000000, 12000000, 101000000]  
                                         })  
  
X_destruction_controlled['Destruction Level'] = pipeline.predict(X_destruction_controlled)  
X_destruction_controlled
```

	Dis Mag Value	Start Year	Start Month	Start Day	End Year	End Month	End Day	Total Deaths	Total Affected	Total Damages Adj	Destruction Level
0	10	2024	1	6	2024	1	8	70	5000	200000	2.0
1	4500	2025	5	16	2025	5	21	250	25000	5000000	3.0
2	26000	2026	8	24	2026	9	4	800	300000	12000000	3.0
3	2	2027	12	30	2027	1	28	10	10	10	2.0

Dataset Controlado o Test A:

	precision	recall	f1-score	support
2	1.00	1.00	1.00	1
3	1.00	1.00	1.00	2
4	1.00	1.00	1.00	1
accuracy				1.00
macro avg	1.00	1.00	1.00	4
weighted avg	1.00	1.00	1.00	4

Métricas de nivel de precisión y capacidad predictiva en Classification Report:

MUY ALTA

- Generaliza bien con datos desconocidos, demostrando precisión y recall del 100% en el conjunto de prueba. Resultados de test controlado sugieren su utilidad para clasificar niveles de destrucción en inundaciones en situaciones reales.
- Recomendable validar en mayor volumen de conjuntos adicionales.
- El modelo se comporta adecuadamente y podría ser una herramienta valiosa para predicciones precisas con nuevos datos.

DESASTRES NATURALES

MODELO PREDICTIVO: TESTS

Test B: Outliers

```
X_destruction_outliers = pd.DataFrame({'Dis Mag Value': [10, 4500, 26000, 2], # se cambió  
                                         'Start Year': [2024, 2025, 2026, 2027],  
                                         'Start Month': [1, 5, 8, 12],  
                                         'Start Day': [6, 16, 24, 30],  
                                         'End Year': [2024, 2025, 2026, 2027],  
                                         'End Month': [1, 5, 9, 1],  
                                         'End Day': [8, 21, 4, 28],  
                                         'Total Deaths': [70, 250, 800, 10], # se quitaron 0s  
                                         'Total Affected': [5000, 25000, 300000, 10], # se quitaron 0s  
                                         'Total Damages Adj': [200000, 5000000, 12000000, 10] # se qui  
                                         })  
  
X_destruction_outliers['Destruction Level'] = pipeline.predict(X_destruction_outliers)  
X_destruction_outliers
```

	Dis Mag Value	Start Year	Start Month	Start Day	End Year	End Month	End Day	Total Deaths	Total Affected	Total Damages Adj	Destruction Level
0	10	2024	1	6	2024	1	8	70	5000	200000	2.0
1	4500	2025	5	16	2025	5	21	250	25000	5000000	3.0
2	26000	2026	8	24	2026	9	4	800	300000	12000000	3.0
3	2	2027	12	30	2027	1	28	10	10	10	2.0

Outliers Dataset o Test B:

	precision	recall	f1-score	support
1	0.00	0.00	0.00	1
2	0.50	1.00	0.67	1
3	1.00	1.00	1.00	2
accuracy			0.75	4
macro avg	0.50	0.67	0.56	4
weighted avg	0.62	0.75	0.67	4

Métricas de nivel de precisión y capacidad predictiva en Classification Report:

MEDIA

- Dificultades en clasificar la clase 1, obteniendo precisión y recall de 0 para esta.
- Buena capacidad predictiva para las clases 2 y 3, con precisión y recall del 50% y 100% respectivamente. Precisión promedio para Test B con outliers es del 62%, indicando que el modelo se desempeña de manera aceptable en general, pero tiene dificultades con datos atípicos o fuera de rango.
- Es importante tener en cuenta que, en los desastres naturales, pueden presentarse casos poco comunes o excepcionales.

DESASTRES NATURALES

MODELO PREDICTIVO: TESTS

Test C: Low

```
X_destruction_low = pd.DataFrame({
    'Dis Mag Value': [5, 3000, 1500, 1], # Valores 'Dis Mag Value' reducidos
    'Start Year': [2024, 2025, 2026, 2027],
    'Start Month': [1, 5, 8, 12],
    'Start Day': [6, 16, 24, 30],
    'End Year': [2024, 2025, 2026, 2027],
    'End Month': [1, 5, 9, 1],
    'End Day': [8, 21, 4, 28],
    'Total Deaths': [35, 150, 2, 5], # Valores 'Total Deaths' reducidos
    'Total Affected': [2500, 12000, 8, 5], # Valores 'Total Affected' reducidos
    'Total Damages Adj': [100000, 3000000, 6, 5] # Valores 'Total Damages Adj' reducidos
})
X_destruction_low['Destruction Level'] = pipeline.predict(X_destruction_low)
X_destruction_low
```

	Dis Mag Value	Start Year	Start Month	Start Day	End Year	End Month	End Day	Total Deaths	Total Affected	Total Damages Adj	Destruction Level
0	5	2024	1	6	2024	1	8	35	2500	100000	2.0
1	3000	2025	5	16	2025	5	21	150	12000	3000000	2.0
2	1500	2026	8	24	2026	9	4	2	8	6	1.0
3	1	2027	12	30	2027	1	28	5	5	5	2.0

Low Dataset o Test C:

	precision	recall	f1-score	support
1	0.00	0.00	0.00	2
2	0.00	0.00	0.00	0
3	1.00	1.00	1.00	1
4	0.00	0.00	0.00	1
accuracy			0.25	4
macro avg	0.25	0.25	0.25	4
weighted avg	0.25	0.25	0.25	4

Métricas de nivel de precisión y capacidad predictiva en Classification Report:

MEDIA

- Dificultades en clasificar la clase 1, obteniendo precisión y recall de 0 para esta.
- Buena capacidad predictiva para las clases 2 y 3, con precisión y recall del 50% y 100% respectivamente. Precisión promedio para Test B con outliers es del 62%, indicando que el modelo se desempeña de manera aceptable en general, pero tiene dificultades con datos atípicos o fuera de rango.
- Es importante tener en cuenta que, en los desastres naturales, pueden presentarse casos poco comunes o excepcionales.

RECOMENDACIONES

Mejora de resultados de predicciones

- Investigación experimental con más pruebas con datos nuevos de mayor volumen.
- Preprocesado con menos columnas
- Pipeline con modelos finalistas validados con desempeño alto como DecisionTrees y GradientBoostingClassifier.
- Nuevos ajustes a hiperparámetros y composición reducida de X.
- Examinar curvas de aprendizaje para obtener una visión completa del rendimiento del modelo.

BUSINESS INTELLIGENCE

Gestión del socorro en casos de desastre



DISASTER RELIEF
BI Dashboard



Herramientas





DIRIGIDO A:

GOBIERNOS Y
AGENCIAS DE GESTIÓN
DE DESASTRES

EMPRESAS DE
SEGUROS

EMPRESAS DE
CONSTRUCCIÓN E
INFRAESTRUCTURA

ORGANIZACIONES DE
AYUDA HUMANITARIA

COMUNIDADES
VULNERABLES



Estrategia

Mejorar estrategias de prevención, mitigación y respuesta.

Especificidad sobre riesgos

Integrar en procesos de suscripción y evaluación de riesgos para ofrecer pólizas más precisas y ajustadas a riesgos específicos de cada región.

Resiliencia

Identificar zonas de alto riesgo y diseñar edificaciones e infraestructuras resistentes.

Gestión de ayuda

Anticipar y preparar recursos y dirigir la ayuda hacia las áreas más afectadas.

Protección

Preparación ante el riesgo y toma de decisiones informadas sobre la protección y evacuación.

PITCH DECK

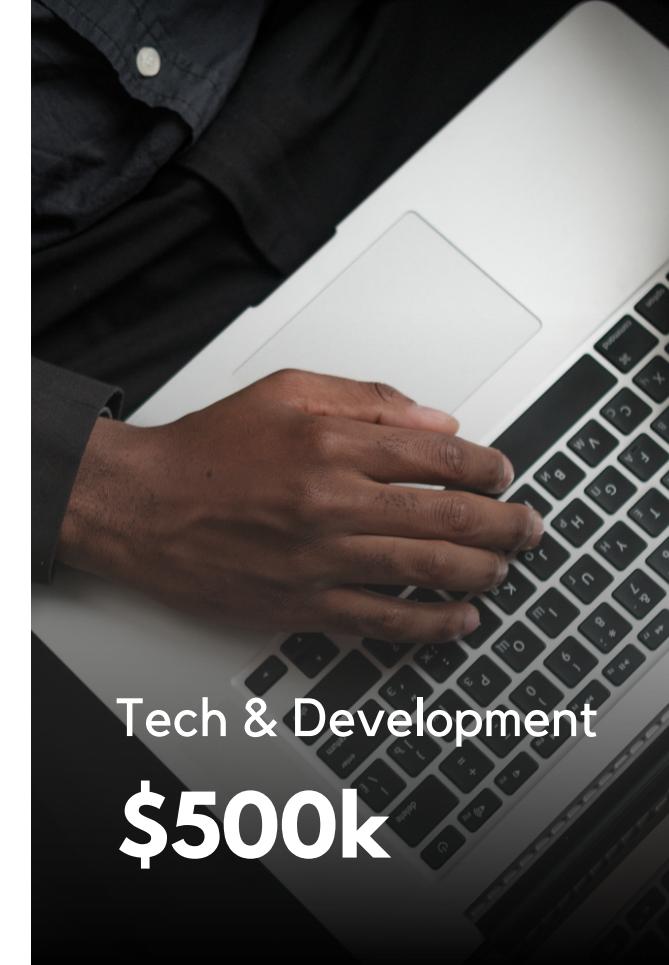
FUTURE

DESASTRES NATURALES
CONCLUSIONES

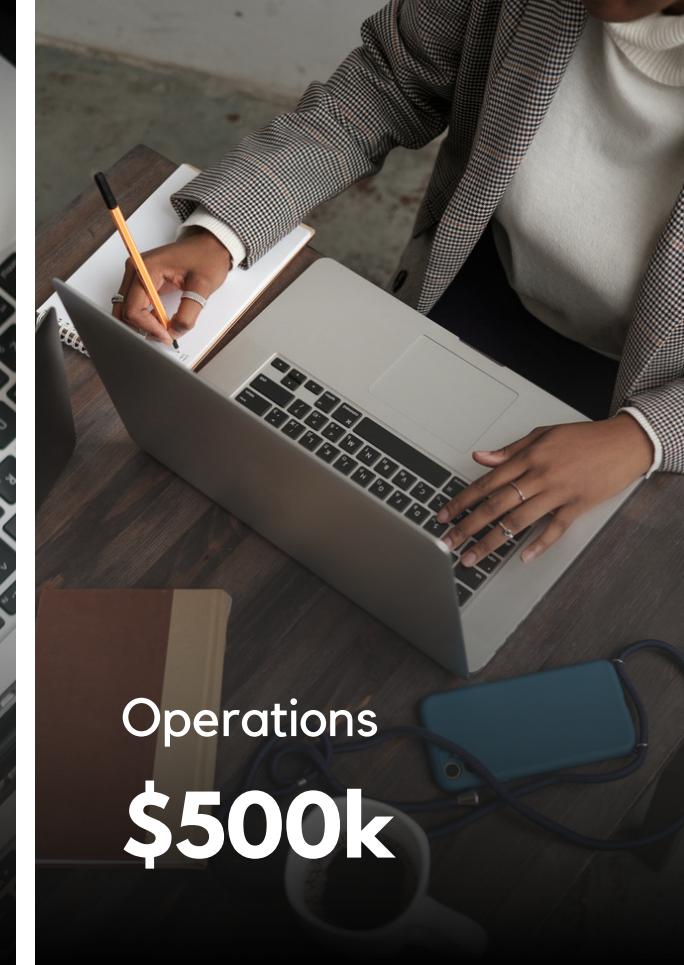
CONCLUSIONES



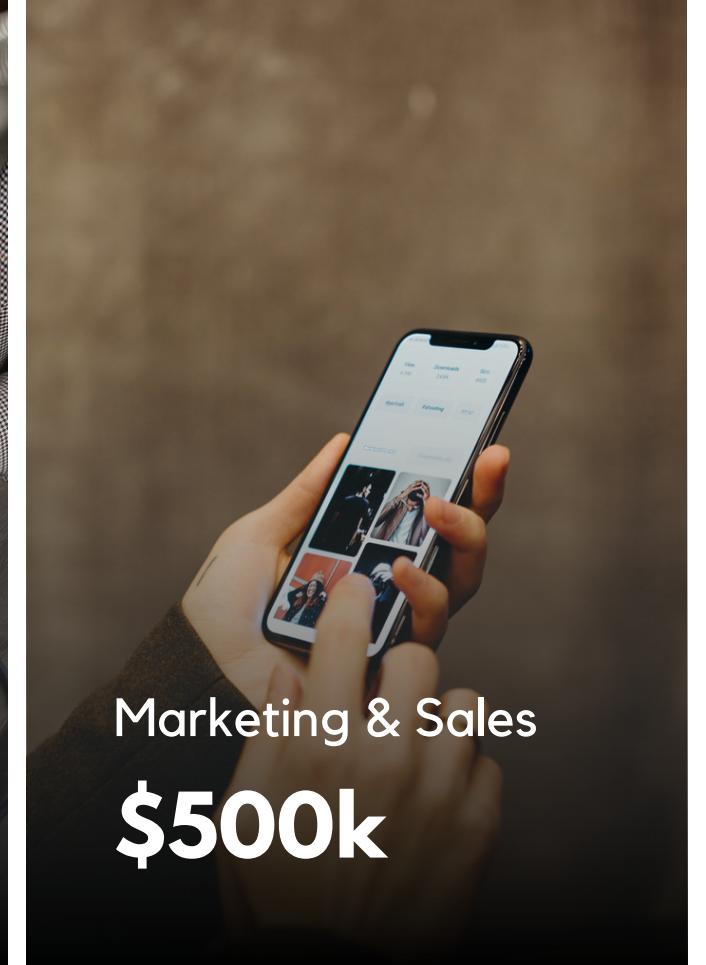
Seeking
\$1.5M



Tech & Development
\$500k



Operations
\$500k



Marketing & Sales
\$500k

GRACIAS

CONTACTA CON EL EQUIPO

Elena Blanco

blancoelena11@gmail.com

Alejandra Eng

criptolmin@gmail.com

LinkedIn:
[@alejandraengbroca](https://www.linkedin.com/in/alejandraengbroca)

Rafael De Marco

rdemarcoz@outlook.com

LinkedIn:
[@rafael-de-marco-z-53097014](https://www.linkedin.com/in/rafael-de-marco-z-53097014)

Ángel Bardon

bardon.angel@gmail.com