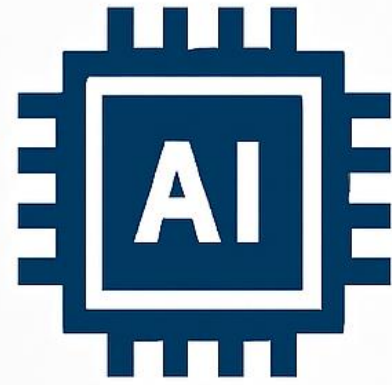


***Predicción de
Severidad en
Accidentes de
Tráfico: Analisis
predictivo con ML +
MLOps***



Índice de Contenido



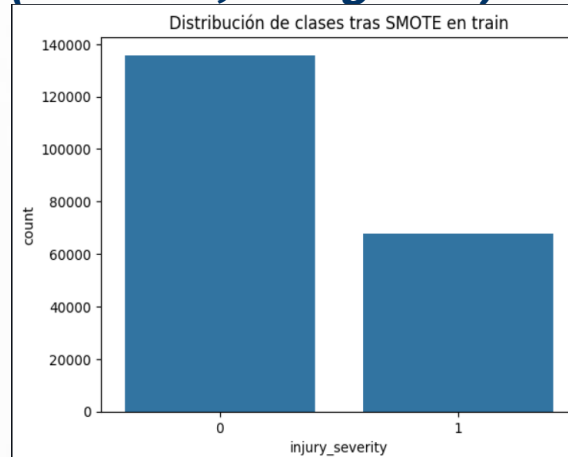
- *Contextualización del problema: costes significativos*
- *Introducción a los datos.*
- *Metodología General: EDA, variable objetivo, preprocesamiento*
- *Métricas de Evaluación*
- *Modelos utilizados*
- *Resultados: Modelo Mejor Desempeño*
- *Despliegue*
- *Conclusiones*



Contexto

- *La seguridad vial es un reto global.*
- *Impacto: reducir muertes y lesiones graves.*
- *Variables contextuales y su relevancia*
- *Conexión con los objetivos*
- *Mayor problemática: Desbalance*
- *Objetivo: Modelo binario*

(0 = leve, 1 = grave).



📌 “Dada la información del accidente (clima, estado de la vía, alumbrado, hora, día, mes, causa, etc.),
¿qué tan probable es que ocurra una lesión leve o grave?”

💡 Visualización tras
técnica de balanceo
SMOTE

Metodología General



Preprocesamiento:

- No se eliminaron outliers
- Normalizaron variables categóricas: **Target Encoding**
- **4 técnicas:** **One-hot Encoding**
Label Encoding
Ordinal Encoding
- Procesar variables temporales (día, mes, año)

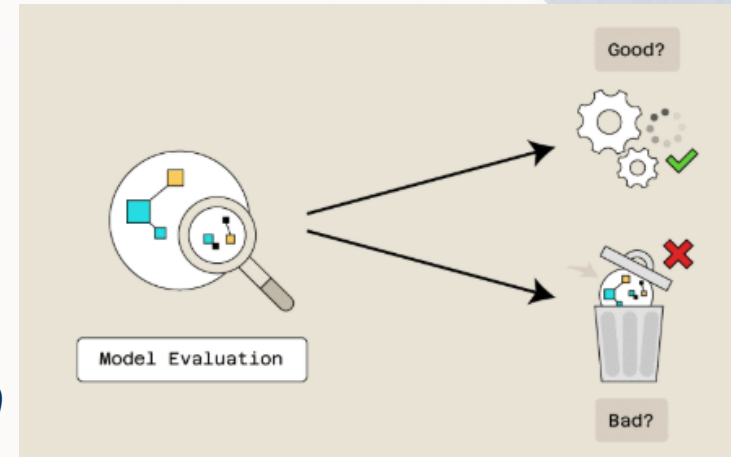
Análisis Exploratorio de Datos (EDA)

- Comprender la estructura del dataset
- Detectar posibles problemas de calidad de datos:
- Evaluar correlaciones y asociaciones

- ✖ La elección de la codificación de variables categóricas depende de dos factores clave:
- **Cardinalidad de la variable** (número de categorías distintas).
 - **Orden o jerarquía inherente** (si hay un orden lógico entre categorías).

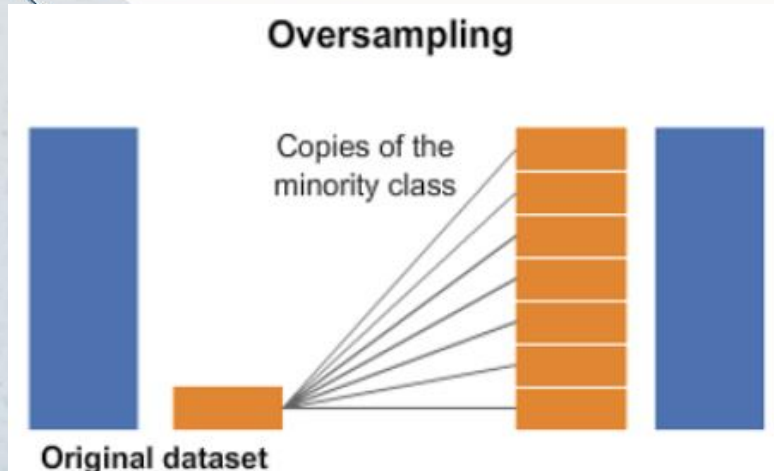
Métricas de Evaluación

- **Accuracy** (global)
- **Precisión** (FP costosos)
- **Recall** (FN críticos)
- **F1** (balance)
- **ROC-AUC** (discriminación)



Desbalanceo:

- **SMOTE:** Técnica de sobremuestreo que, genera nuevas muestras sintéticas seleccionando un ejemplo minoritario y un vecino más cercano de la misma clase.
- Algoritmos específicos



Modelos Utilizados

- *Decision Tree*
- *Random Forest*
- *Gradient Boosting*
- *XGBoost*
- *Logistic Regression (class_weight, SMOTE)*
- *Balanced RF*
- *Easy Ensemble*

Primeros insights:

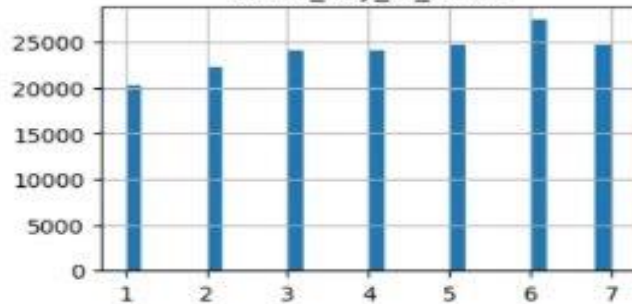
Insights pre-testing:

Dia de la semana: Sabado ❌

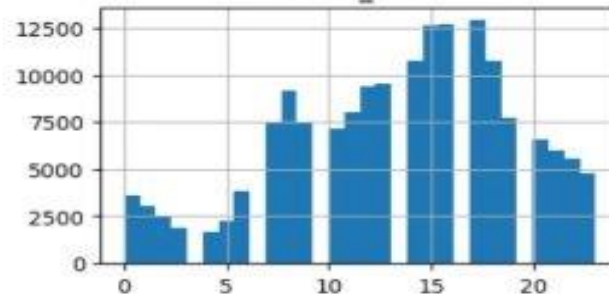
Hora: 15h – 18h ❌

Mes: Octubre ✅

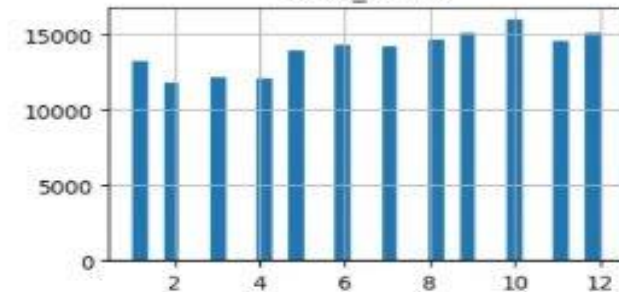
crash_day_of_week



crash_hour



crash_month



Insights post-testing:

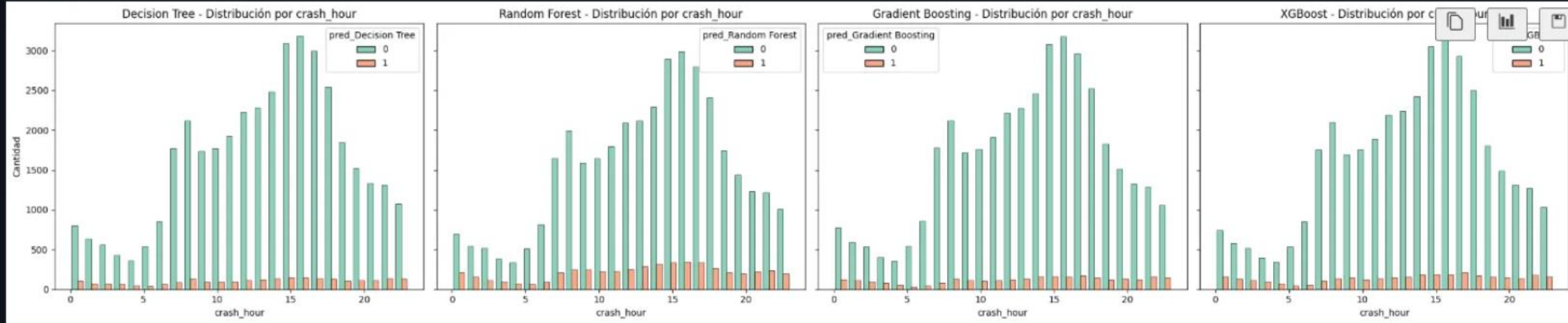
Dia de la semana: Viernes ✅

Hora: 16h ✅

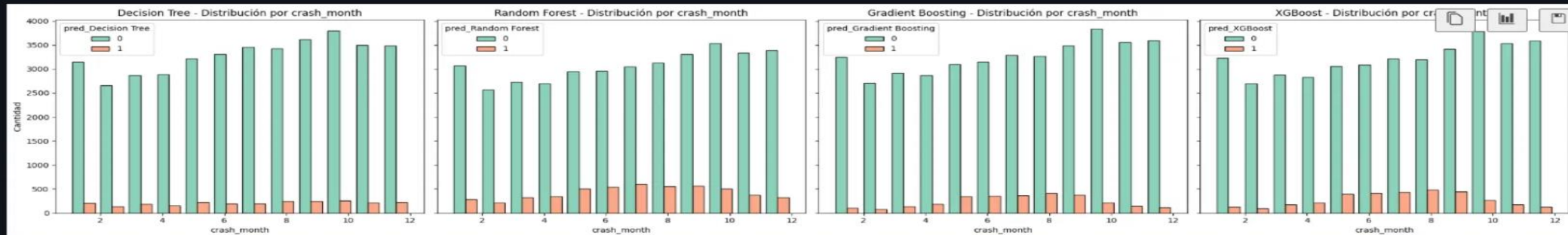
Mes: Octubre ✅

Insights post-testing

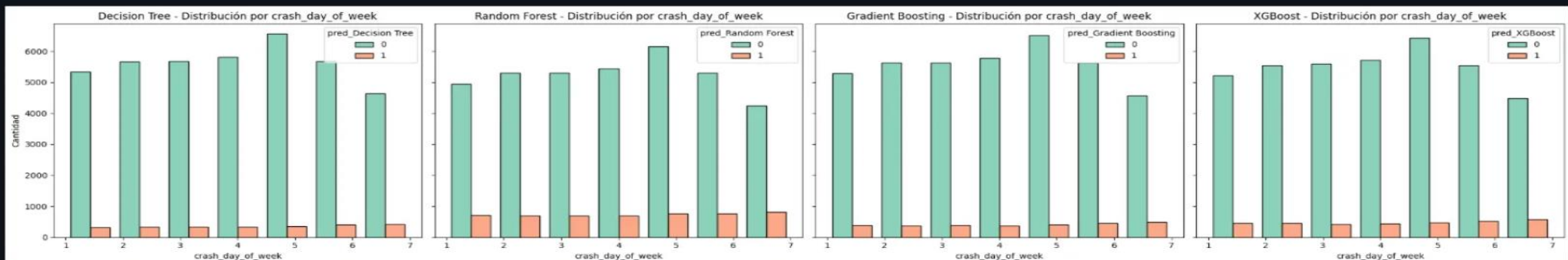
Histogramas de crash_hour por modelo



Histogramas de crash_month por modelo



Histogramas de crash_day_of_week por modelo



Interpretación de resultados:



- Modelos predicen bien la clase leve, con alta precisión y recall.
- Clase grave, difícil de predecir, con recall bajo incluso con ensembles.
- La accuracy global no es representativa del desempeño real, debido al desbalance de clases.



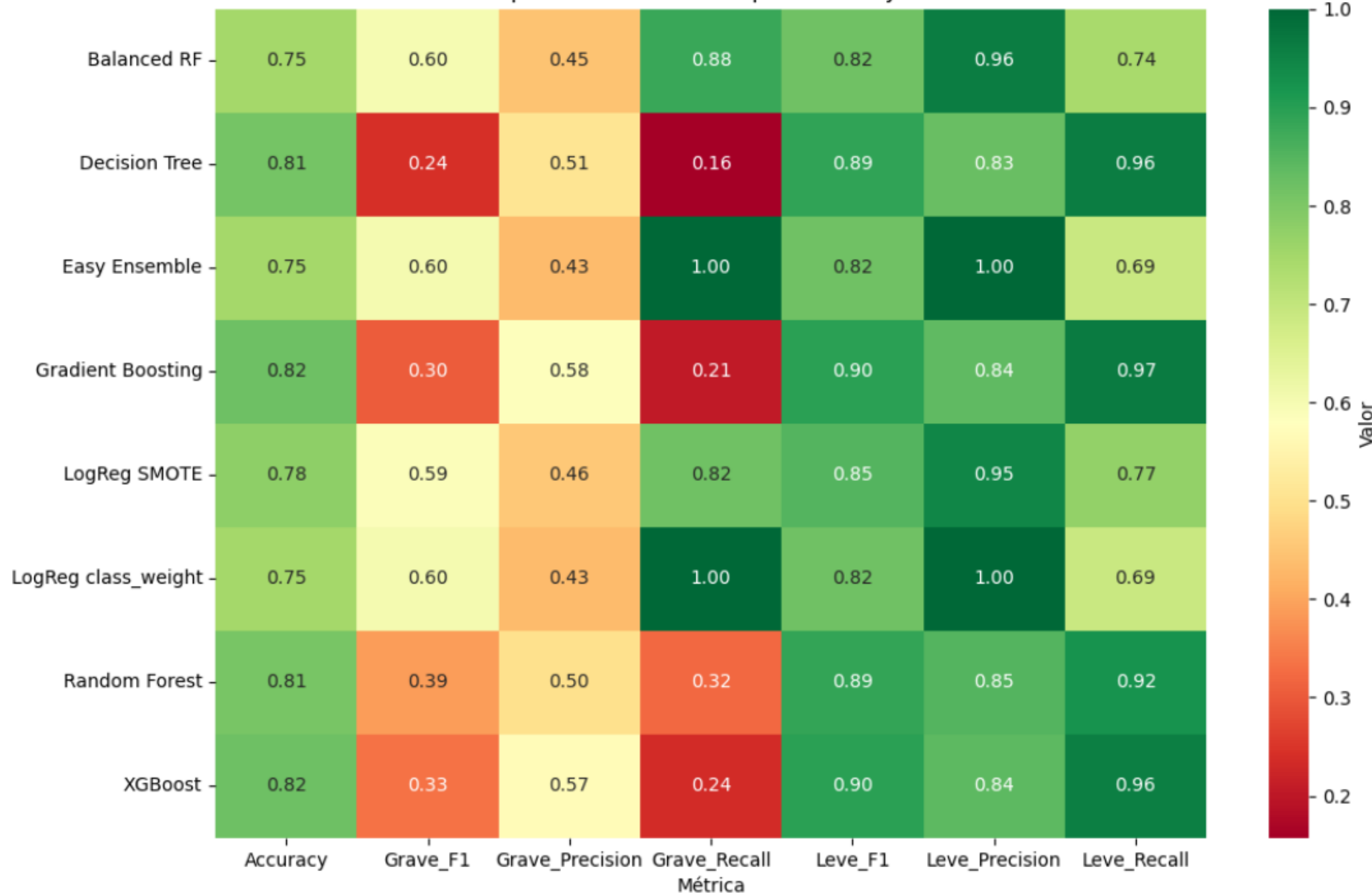
Validación cruzada

La evaluación se realizó con validación cruzada:

- **GridSearchCV** para optimizar hiperparámetros en modelos como Regresión Logística y Random Forest, y
- **StratifiedKFold** para evaluar estabilidad en Balanced Random Forest y Easy Ensemble.

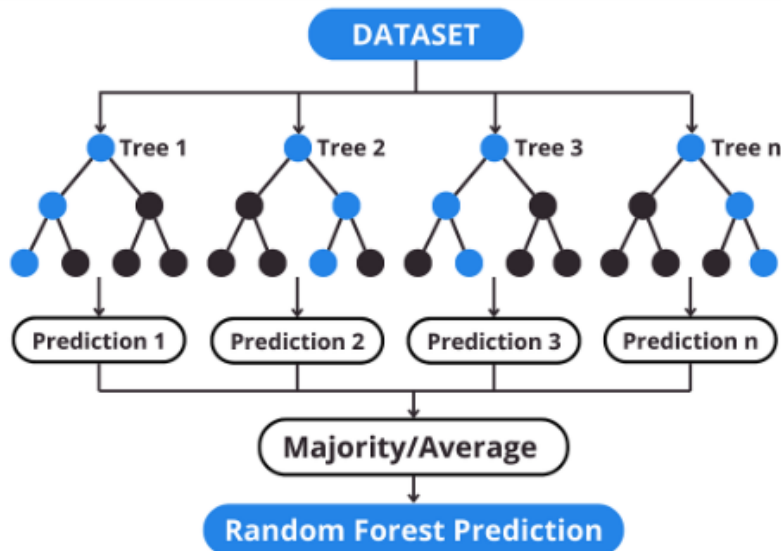


Comparación de métricas por modelo y clase

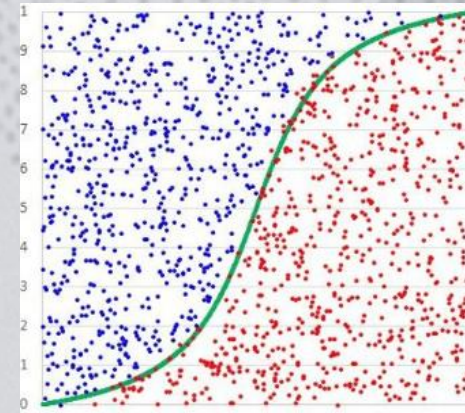


Mejor Modelo: Balance Random Forest

- *Mejor equilibrio entre detectar leves y graves. gracias a su enfoque ensemble.*
- *Enfoque ensemble mejora el recall.*
- *Metricas mas realistas*



🏆 **La Regresión Logística con balanceo de pesos sorprendió: alcanzó un recall cercano al 100% en la clase grave.**



Despliegue del Modelo

¿cómo llevar esto a un entorno real?



Crash Injury Severity API 0.1.0 OAS 3.1

/openapi.json

default

GET

/ Home

POST

/predict Predict Severity

- *API desarrollada en FastAPI.*
- *Pipeline preparado para producción (MLOps).*

Response body

```
{  
  "prediction": 1,  
  "probabilidad_grave": 0.7728082551713007  
}
```



Conclusiones

"Este proyecto no solo nos permitió aplicar técnicas avanzadas de machine learning, sino que demuestra cómo el análisis de datos y modelos predictivos pueden tener un impacto real en la seguridad vial, optimizando recursos y salvando vidas. Y lo más importante: nos enseñó la relevancia de evaluar métricas por clase y adaptar los modelos a los desafíos del mundo real."





Fin de la Presentación

*Gracias por su atención. Preguntas y comentarios son
bienvenidos.*

