



# Análisis de Error en Modelos

# About me





## Métrica por tarea (<http://hf.co/metrics>)

Task/tarea	Metrica	Descripción
Clasificación (de texto, de token, de audio, etc)	F1, precision, accuracy	Proporciones de TP, FP, TN, FN
Question Answering extractivo	Exact match	
Traduccion	BLEU, SACREBLEU	Precision basada en n-grams
Resumen (summarization)	ROUGE	Recall basado en n-grams
Generacion de texto	Perplexity	Distribucion probabilisticca de predicciones
Speech Recognition	WER, CER	



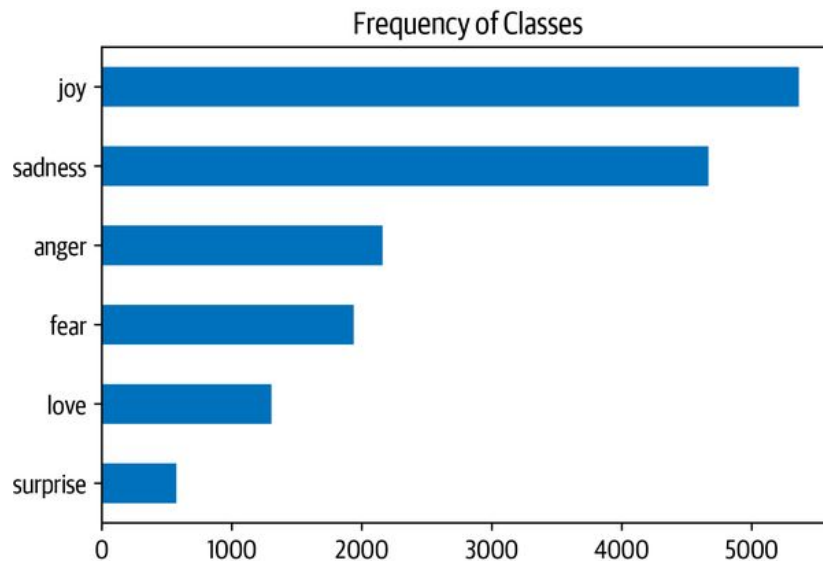
Que pasa si tienen buenas métricas?

# Dataset no balanceado



```
import matplotlib.pyplot as plt

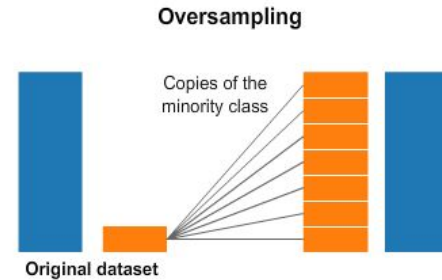
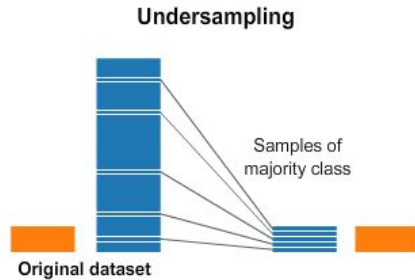
df["label_name"].value_counts(ascending=True).plot.barh()
plt.title("Frequency of Classes")
plt.show()
```





# Dataset no balanceado

- Downsample o upsample
- Generación de datos sintéticos
- Weighted loss
- Boosting





# Otros problemas potenciales

- Error en el cálculo de métricas
- Dataset esta roto
  - E.j. label consistency, malos labels
- Data leakage (i.e. entrenar con datos de prueba)



Que pasa si tienen malas métricas?





# Malas metricas

- Lo mejor es encontrar ejemplos para ver, como ejemplos:
  - Cerca de un threshold
  - El label y el score del modelo sean muy diferentes, y el modelo tenga alta confianza
  - Highest loss
- Errores en tokenización
- Clusters probabilisticos

# Samples con mayor loss



```
df_test.sort_values("loss", ascending=False).head(10)
```

text	label	predicted_label	loss
i feel that he was being overshadowed by the supporting characters	love	sadness	5.704531
i called myself pro life and voted for perry without knowing this information i would feel betrayed but moreover i would feel that i had betrayed god by supporting a man who mandated a barely year old vaccine for little girls putting them in danger to financially support people close to him	joy	sadness	5.484461
i guess i feel betrayed because i admired him so much and for someone to do this to his wife and kids just goes beyond the pale	joy	sadness	5.434768



Haciendo debugging con transformers

[https://github.com/TheHackerLlama/charlas/blob/main/debugging\\_transformers.ipynb](https://github.com/TheHackerLlama/charlas/blob/main/debugging_transformers.ipynb)



# Errores frecuentes en pipelines

- Estas usando un modelo válido?
- El modelo tiene un archivo de configuración
- Hay algo que sobresalga en el archivo de configuración

# Tips



- Mira los datos antes de entrenar
  - Haz encoding y decoding para asegurarte que todo funciona bien
- Siempre haz un dummy training primero
- Mira la distribución de clases, estas de acuerdo con los labels?
- Asegurate que los labels hagan sentido
- No hagas hyperparameter tuning hasta tener un baseline
- No ver una sola metrica



# Errores de GPU

- CUDA out of memory
- El código puede estar bien, pero esto significa que tienes mucho en memoria
- Como usar menos GPU?
  - No tener varios modelos a la vez
  - Reducir el batch size
  - Considera una versión mas pequeña