

# Transformers aplicado a la detección y análisis de agresividad en redes sociales

Victor Manuel Gómez Espinosa

Dr. Victor Muñoz S / Dr. A. Pastor López M

Centro de investigación en matemáticas, Unidad Monterrey

victor.gomez@cimat.mx



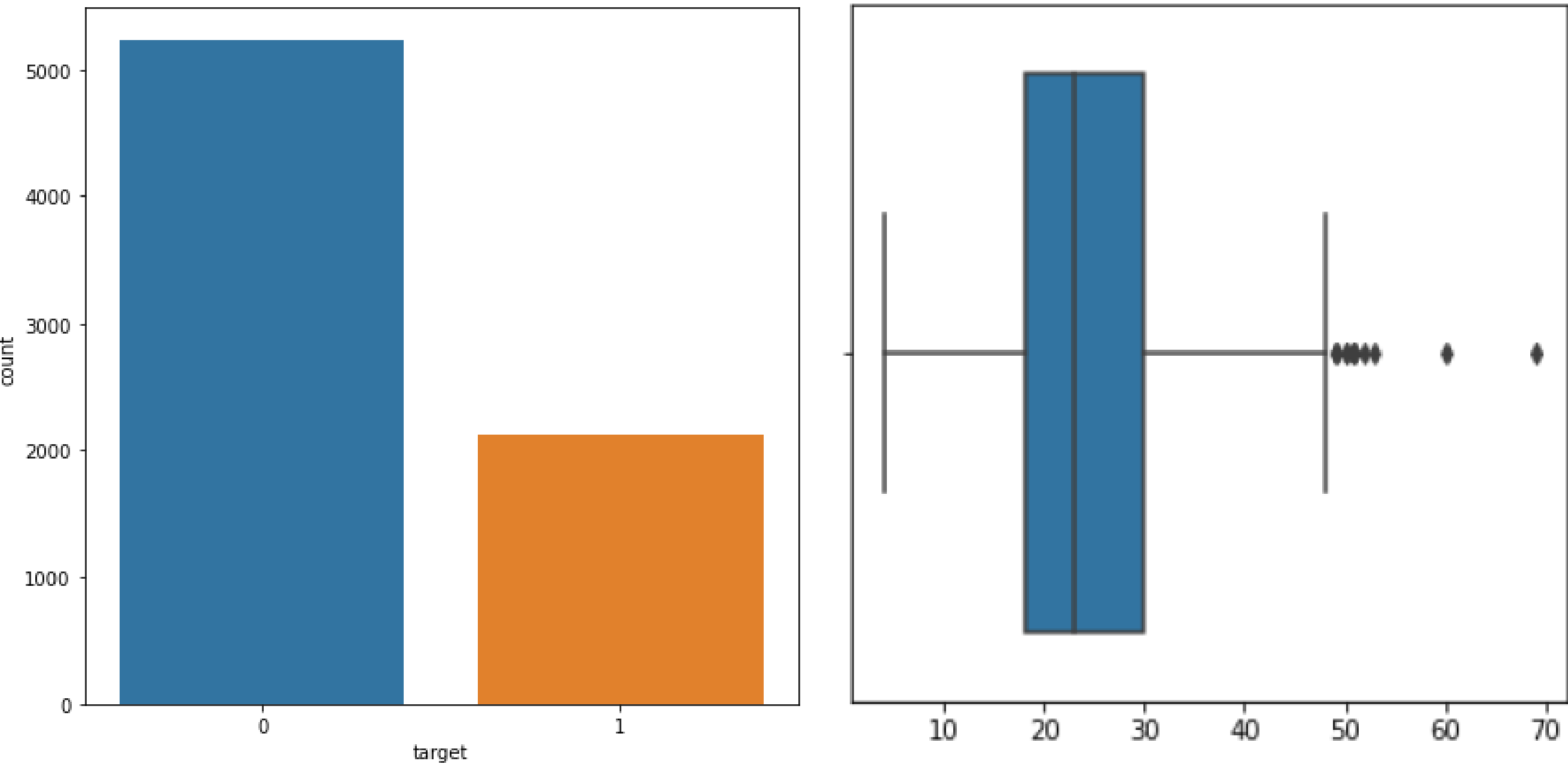
## Introducción

La agresividad se define como aquello que busca causar daño y puede incitar a la violencia, lo cual es a lo que se exponen los usuarios en redes sociales, y estos pueden causar daños a largo plazo, en algunas ocasiones llevando al suicidio, de ahí la importancia de buscar formas de identificar este tipo de manifestaciones.

## Objetivo

Identificar la agresividad en tweets en español de México mediante la adaptación de un modelo basado en la arquitectura Transformer, adicionalmente, se busca poder interpretar estos modelos.

## Datos: Corpus MEX-A3T 2020



## Métodos

### Pre-proceso

#YordiEnEXA Inche maricon 🤔🤔🤔 Ahora entiendo porque eres Ro  
yordi en ex a inche maricon cara llorando de risa cara llorando de risa

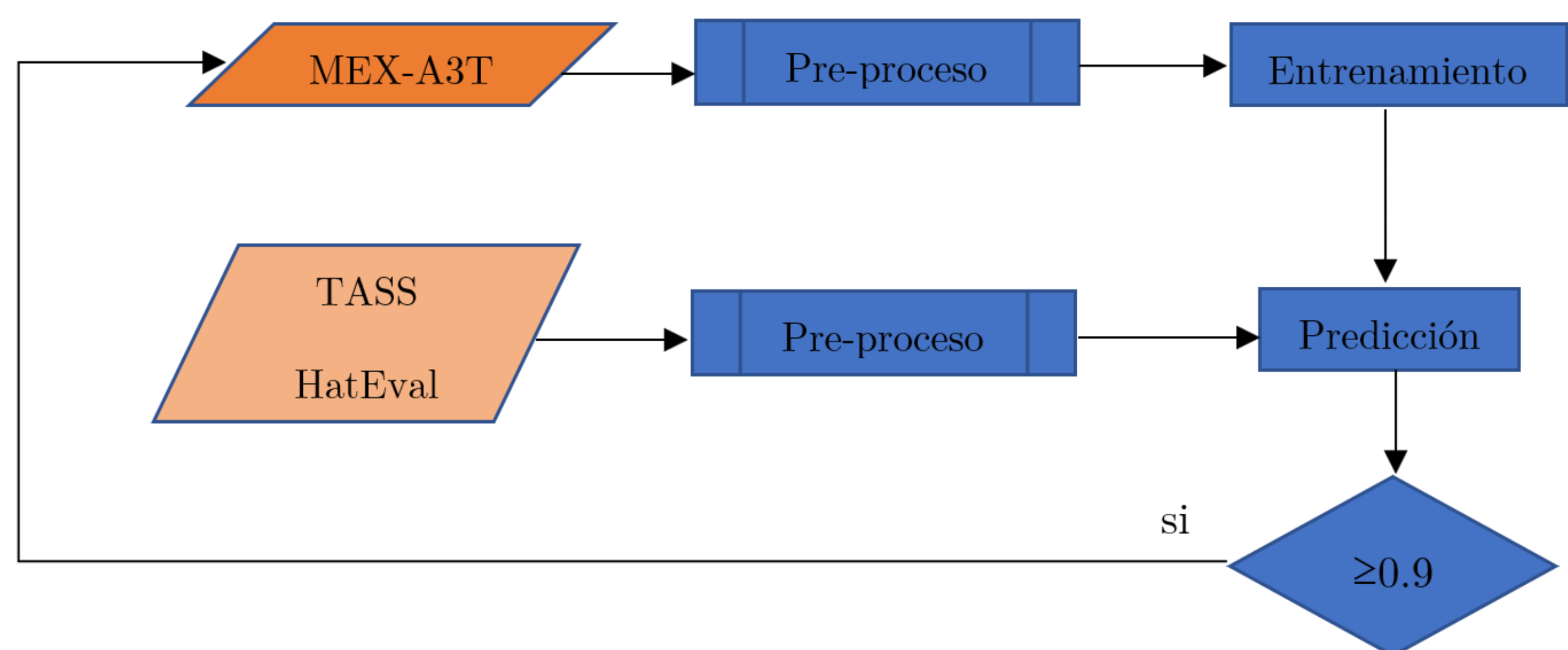
Anduve como loca buscando y no me acorde de ahí:( gracias !!

anduve como loca buscando y no me acorde de ahí tristeza gracias !!

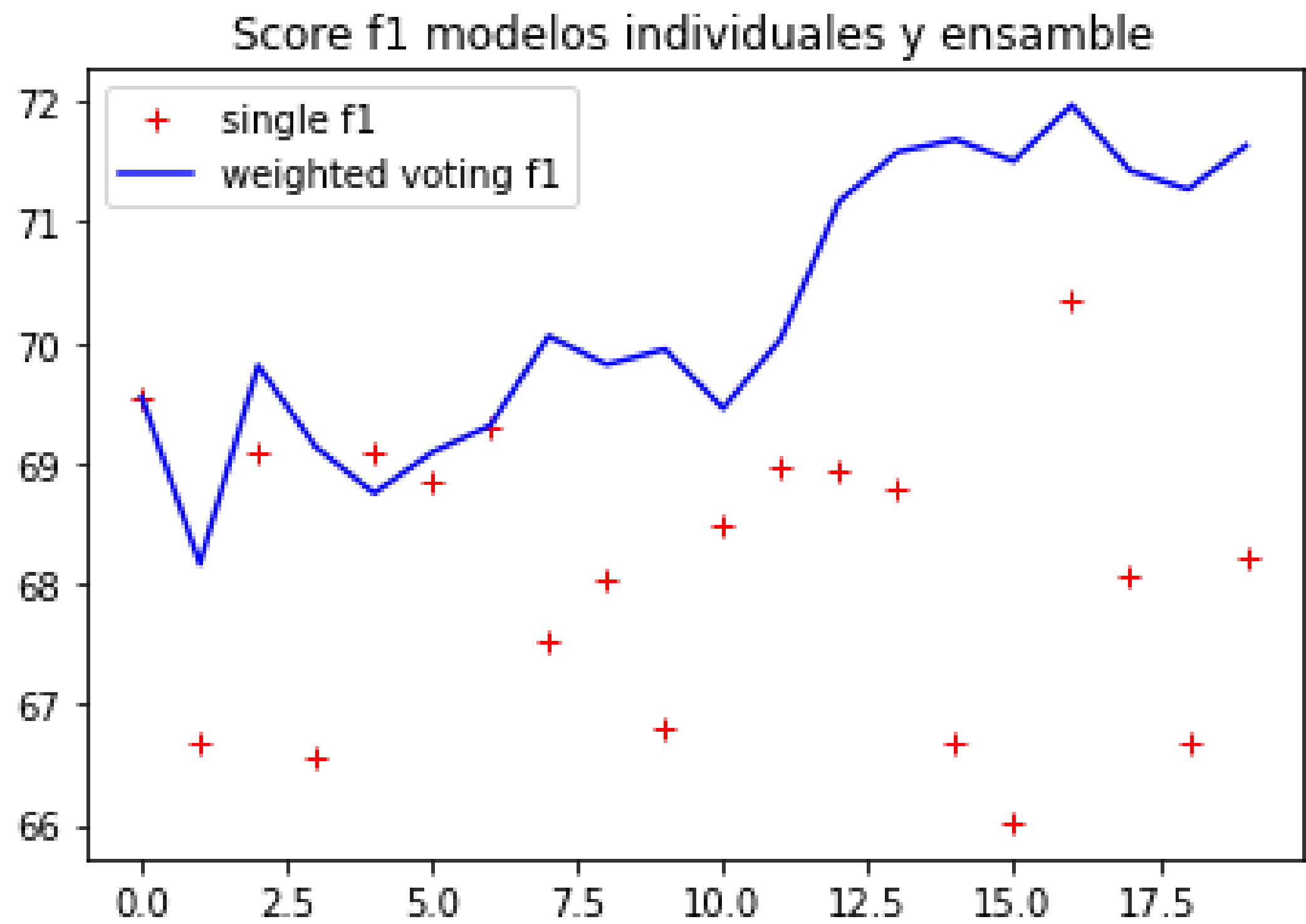
¿Por que verga mis vecinos son tan perros asquerosos? Alv

¿por que verga mis vecinos son tan perros asquerosos? a la verga

### Ampliación del corpus



## Ensamble de BERTs



## Resultados

<i>MEX-A3T</i>	<i>F1</i>	<i>OffendMex</i>	<i>F1</i>
<i>CIMAT</i>	<b>0.8004</b>	<i>CIMAT</i>	<b>0.7026</b>
<i>CIMAT-1</i>	0.7998	<i>NLP-CIC</i>	0.6931
<i>CIMAT-2</i>	0.7971	<i>DCCD INFOTEC</i>	0.6847
<i>UPB-2</i>	0.7969	<i>cimat-gto</i>	0.6791
<i>UACH-2</i>	0.7720	<i>UMUTeam</i>	0.6705

## Visualización de la atención



## Conclusiones y trabajo futuro

- Se nota una mejora de realizar un pre-proceso y ampliar el corpus de entrenamiento.
- Se observó que la clasificación de la agresividad pone atención a una acción de causar daño.
- La metodología propuesta puede utilizarse efectivamente en otras tareas (1er lugar en MeOffendEs 2021 subtask 3).
- Probar un modelo multimodal.

## Referencias

- [1] Aragón ME, Jarquín-Vasquez H, Montes-Y-Gómez M, et al (2020) Overview of mex-a3t at iberlef 2020: Fake news and aggressiveness analysis in mexican Spanish. CEUR Workshop Proc 2664:222–235
- [2] Guzman-Silverio M, Balderas-Paredes Á, López-Monroy AP (2020) Transformers and data augmentation for aggressiveness detection in mexican Spanish. CEUR Workshop Proc 2664:293–302
- [3] Devlin J, Chang M-W, Lee K, et al (2019) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding