

# INTERPRETABILITY ON SENTIMENT ANALYSIS OF CONTENT RELATED TO PERCEPTION OF SECURITY

Luisa Fernanda Chaparro, Ph.D.

This work was funded by Bank of National Investment  
Programs and Projects, National Planning Department,  
Government of Colombia (BPIN: 2016000100036).

Facultad de Ciencias  
Sede Bogotá



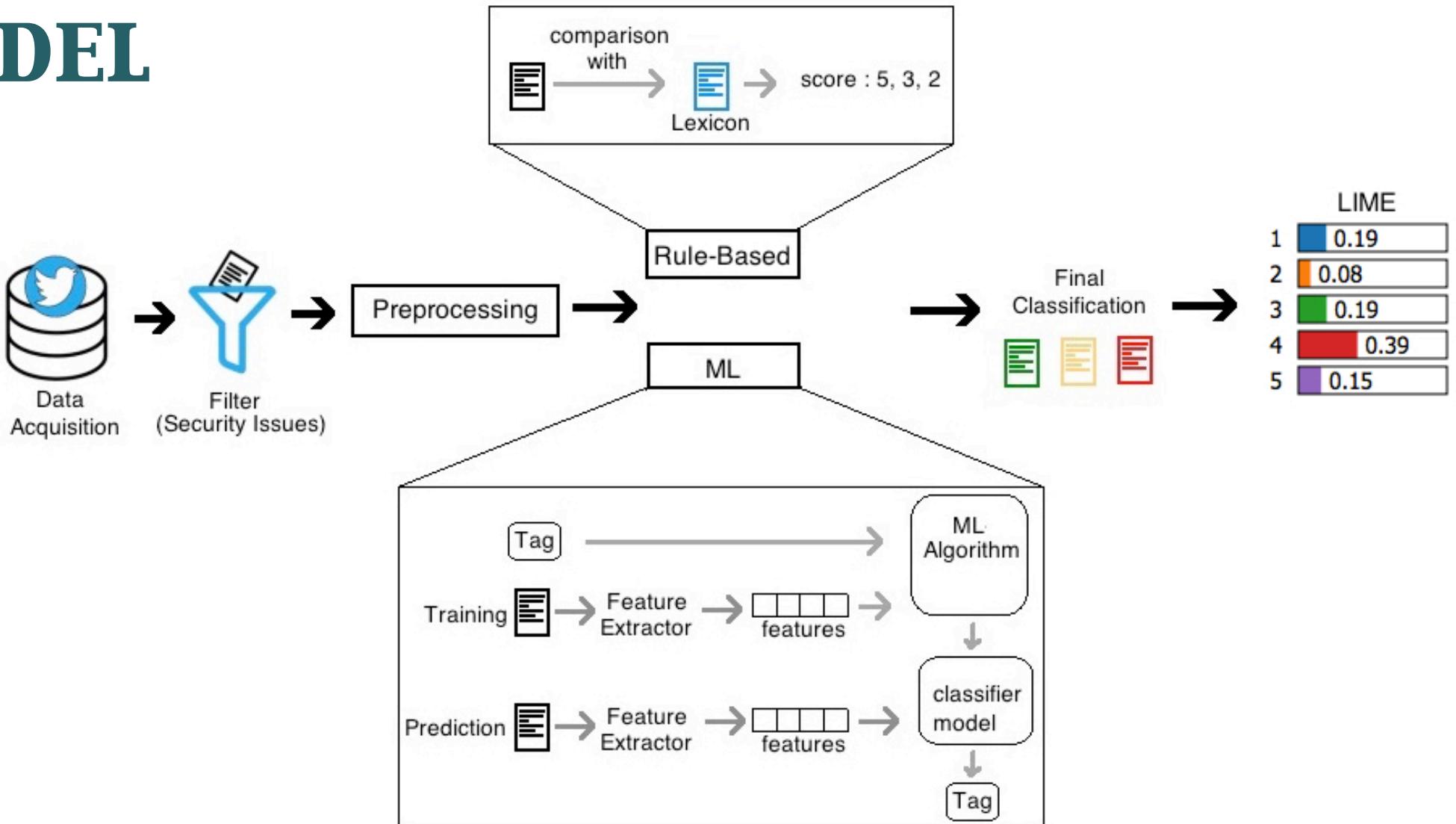
UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA

# MOTIVATION

The constant opinion monitoring on social networks content like Twitter, related to security issues, helps to understand the perception of security (fear of crime)

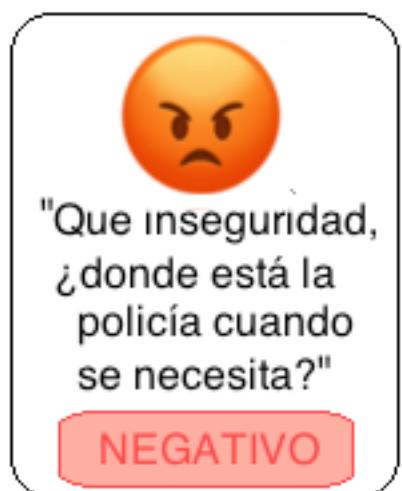
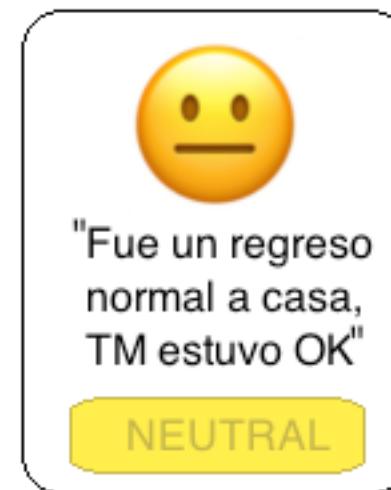
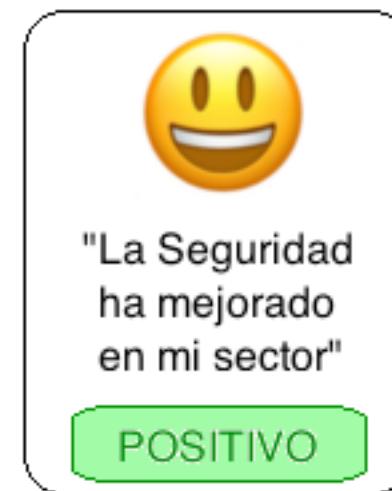


# MODEL



# SENTIMENT ANALYSIS

- Technique that detects polarity within text, whether a whole document, paragraph, sentence, or clause.
- Fine-grained Sentiment Analysis:
  - Very positive
  - Positive
  - Neutral
  - Negative
  - Very negative
- The main types of algorithms used include:
  - **Rule-based** systems that perform sentiment analysis based on a set of manually crafted rules.
  - **Automatic** systems that rely on machine learning techniques to learn from data.



# DATA & PREPROCESSING

Tweet	Score
Comparto nuestro aliado estrategico para la Escuela de Paz: CORPORACION CASA DE LA MUJER DE SUBA. Apoyo Eje tematico:Comunidad, Paz y Equidad de Genero. Alli llevo siendo tallerista voluntaria, por la erradicacion de todo tipo de violencia hacia la mujer desde el año 2017	5
@FREDYGUERRAHERR @consigliererojo @ClaudiaLopez @TransMilenio @NoticiasCaracol @PoliciaBogota Si se permite esto, se acaba la seguridad en Bogota, no es tema de izquierda ni derecha, es algo que nos afecta a todos por igual, asi que es momento de pensar en las consecuencias.	3.2
@FREDYGUERRAHERR @adelve3 @ClaudiaLopez @TransMilenio @NoticiasCaracol @PoliciaBogota Que hpta desorden de ciudad, ya aqui todos hacen lo que se les viene en gana y ni hablar en las calles: vendedores ambulantes, expendedores de drogas, desaseo, inseguridad, un asco de ciudad y todo empezo desde el gobierno de Luis Eduardo Garzon y los siguientes, no hicieron ni m	1.1
@luciacbastidasu @TransMilenio Pero esto esta prohibido normas de convivencia fumar marihuana en espacio cerrado es un delito, pero como la policia brilla por su ausencia y como ciudadanos no denunciamos si no permitimos Esto ellos hacen lo que quieran	1.3
Lo del bus de TM y la marihuana no esta nada bien, pero ojala tuviera el mismo nivel de indignacion cuando roban, acosan o violentan a las personas en TM, lo que esta mal no se puede categorizar. Todo esta mal, no puede ser que unas cosas las denunciamos y otras las normalicemos	2



## Data:

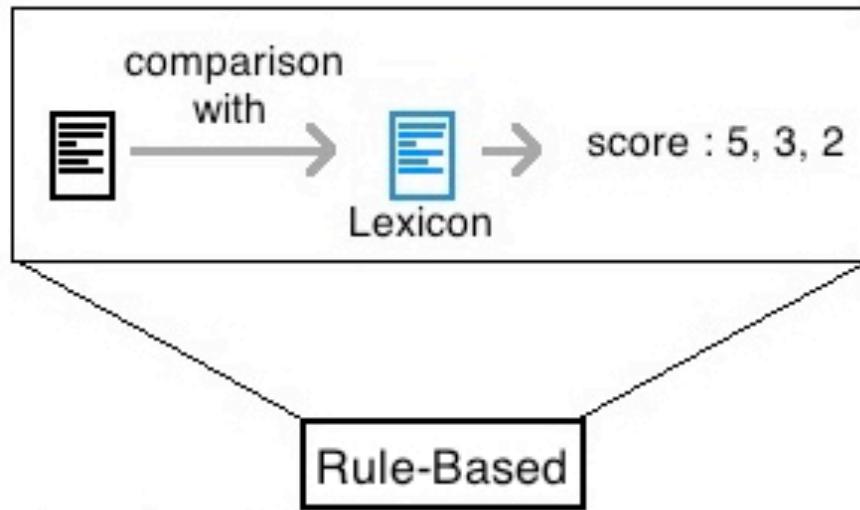
- ~2000 Tweets collected from Dec 2019- Apr 2020
- Previously tagged as Security Issues content.

## Preprocessing:

- Remove links, mentions, and hashtags.
- Remove stopwords and accents
- Stemming
- Tokenization

# RULE-BASED APPROACH

- Defines two lists of polarized words → Lexicon
- Counts the number of positive and negative words that appear in a given tweet.
- Compute the Score:

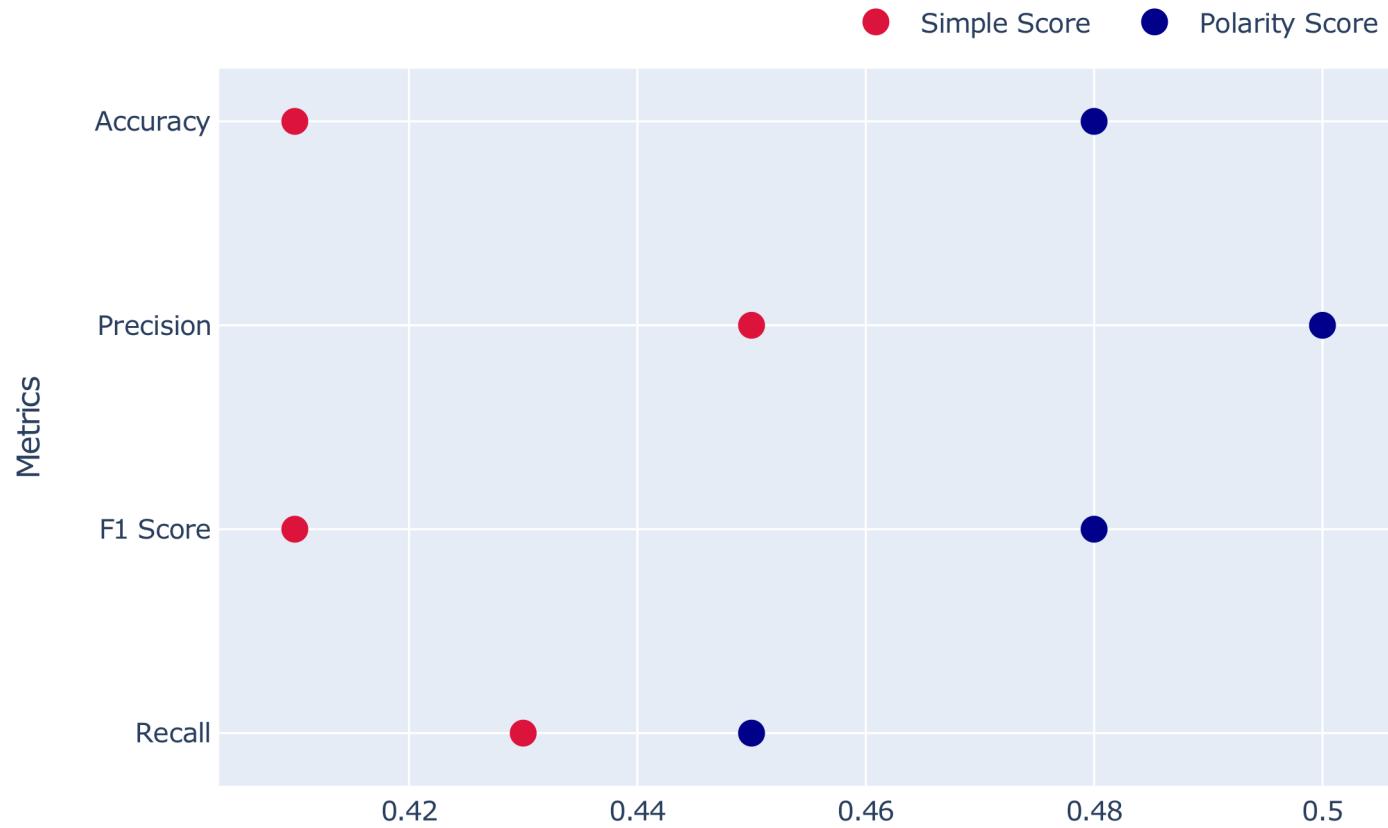


$$\text{Simple Score}_{(\text{Tweet})} = \frac{(\# \text{ Positive words} + \# \text{ Negative words})}{\# \text{ Total words}}$$

$$\text{Polarity Score}_{(\text{Tweet})} = \frac{\sum_{w \in \text{words}} (\text{Score Pos } w) - (\text{Score Neg } w)}{\# \text{ Total words}}$$

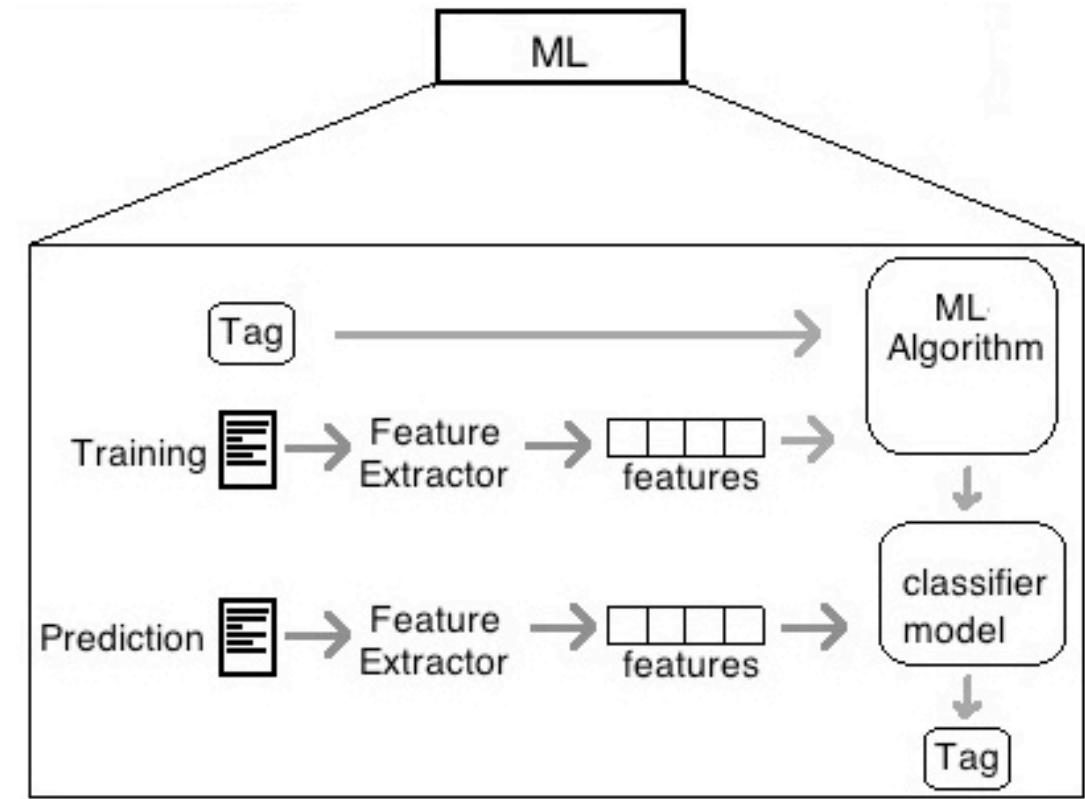
# RULE-BASED APPROACH

Metrics for Rule-Based Method



# ML APPROACH

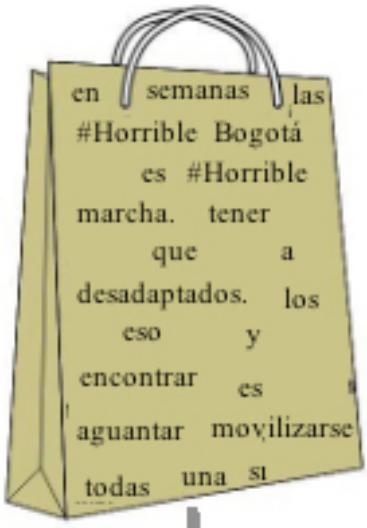
- Is usually modeled as a classification problem, whereby a classifier is fed a text and returns a category.
- Training process:
  - The model learns to associate a particular input to the corresponding output based on the test samples used for training.
  - The feature extractor transfers the text input into a feature vector.
  - Pairs of feature vectors and tags are fed into the machine learning algorithm to generate a model.
- Prediction process:
  - the feature extractor is used to transform unseen text inputs into feature vectors.
  - These feature vectors are then fed into the model, which generates predicted tags.



# FEATURE EXTRACTION

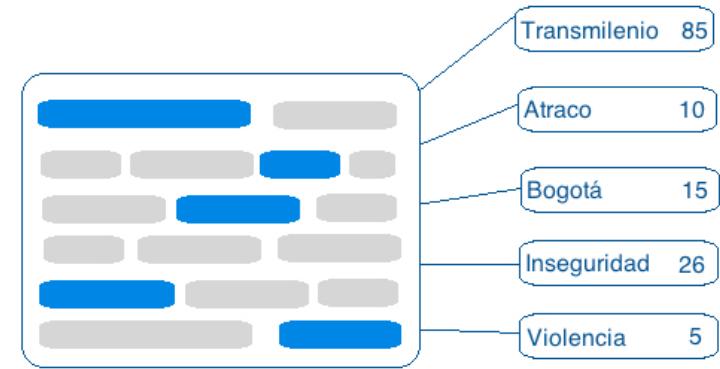
## Bag-of-words model(BoW )

#Horrible es movilizarse  
en Bogotá y encontrar  
todas las semanas una  
marcha. Eso si es  
#Horrible tener  
que aguantar a  
los desadaptados.



#Horrible	2	una	1
es	2	marcha	1
movilizarse	1	Eso	1
en	1	si	1
Bogotá	1	tener	1
y	1	que	1
encontrar	1	aguantar	1
todas	1	a	1
las	1	los	1
semanas	1	desacatados	1

## TF-IDF model

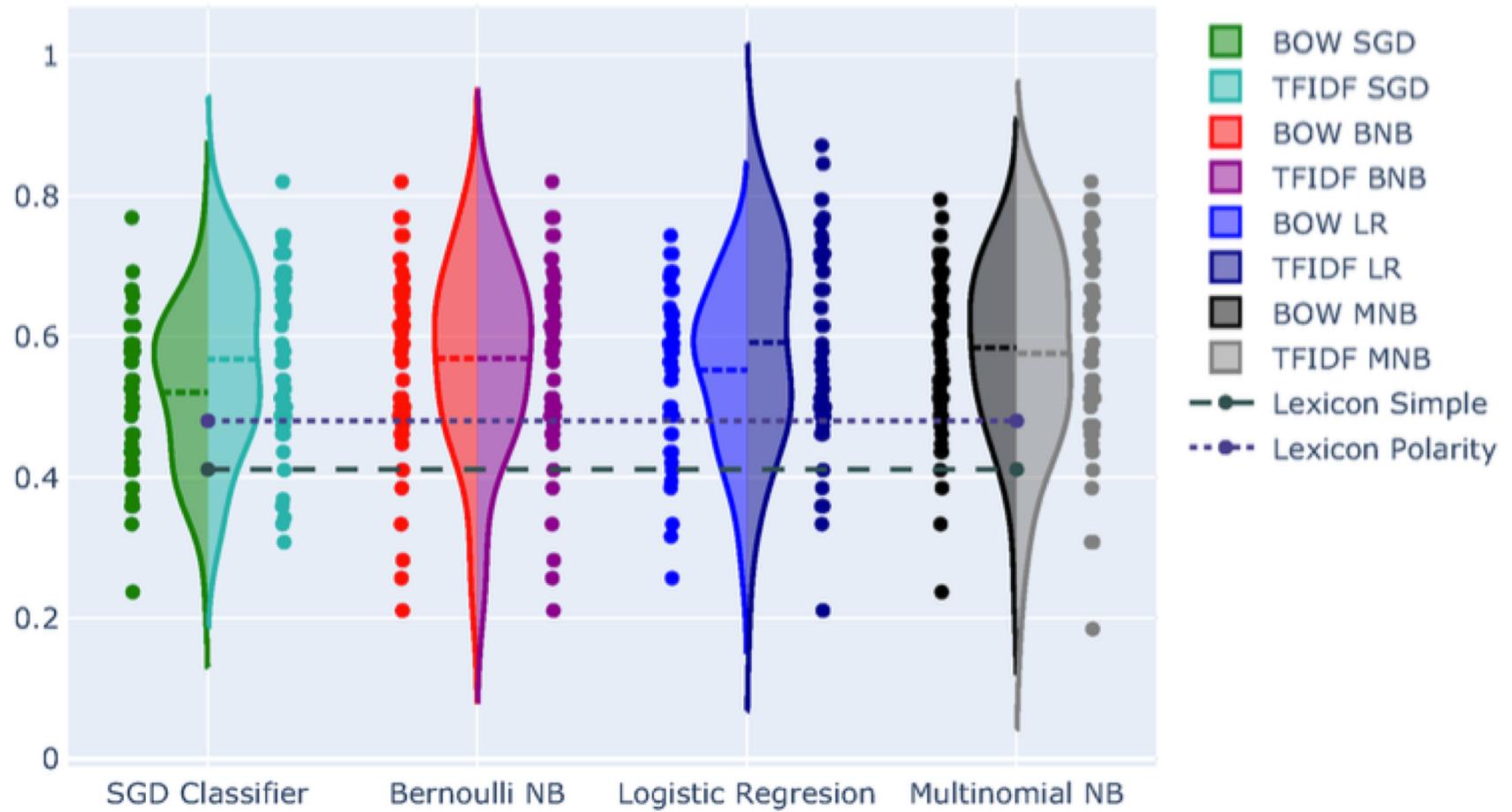


TF-IDF evaluates how relevant a word is to a document in a collection of documents.

- This is done by multiplying two metrics:
  - how many times a word appears in a document,
  - and the inverse document frequency of the word across a set of documents.

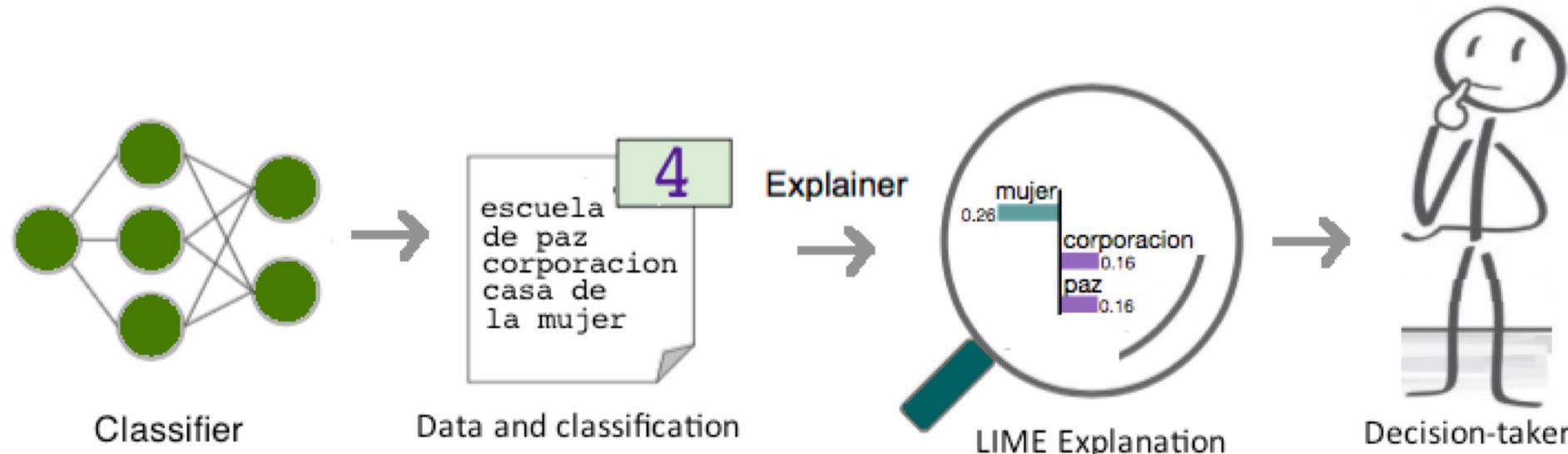
# CLASSIFIER MODELS

Accuracy



The texts are invariant to the classifiers' methods.  
Which one could explain better the Perception of Security?

# INTERPRETABILITY - LIME



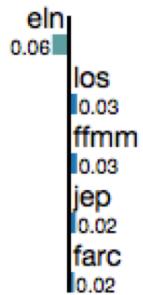
- Explained black box machine learning models.
- Interpretability model locally around the prediction
- Determining trust in individual predictions is an important problem when the model is used for decision making
- provide qualitative understanding between the input variables and the response

# LIME

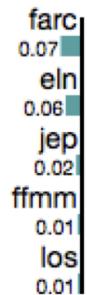
Prediction probabilities

1	0.07
2	0.81
3	0.02
4	0.08
5	0.03

NOT 1

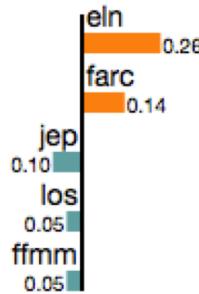


NOT 3



NOT 2

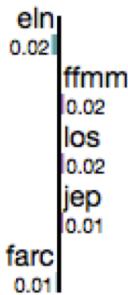
2



1

NOT 5

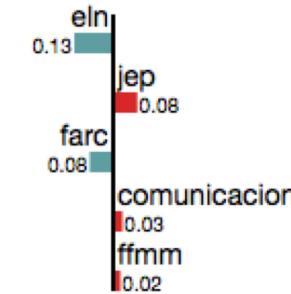
5



3

NOT 4

4

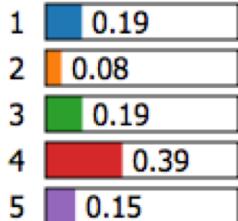


## Text with highlighted words

la campana de des prestigio a las ffmm emprendida por activistas de izquierda en complicidad con la jep medios de comunicacion y universidades cooptadas buscan trasgredir la verdad y culpar a militares y policias de los asesinatos de farc eln y otros grupos narcotraficantes

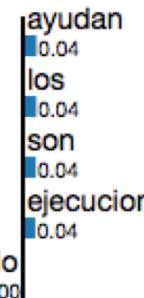
# LIME

Prediction probabilities



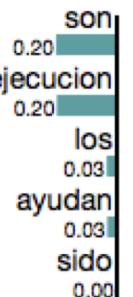
NOT 1

1



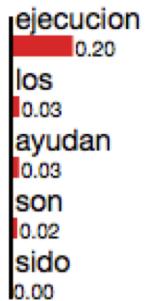
NOT 2

2



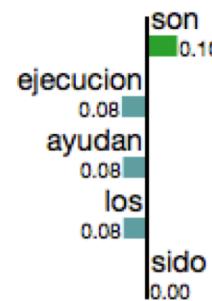
NOT 4

4



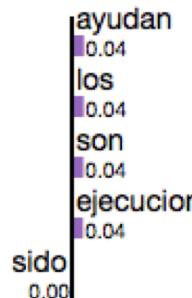
NOT 5

5



NOT 3

3



## Text with highlighted words

las cosas como son los extremos tampoco le ayudan al debate llevar este caso a ejecucion es exagerado lo que no quita que si haya sido un asesinato

# MULTINOMIAL NAIIVE BAYES - LIME

**Level 1**

rural	hpta
dan	john
alcaldes	asociacion
human	cartel
man	hach
mil	plop
criminal	lideres
capital	desorden
accion	vih
educacion	asco

— +

**Level 2**

caption	cool
disminucion	saque
inpec	militar
vial	ir
son	eln
judicial	interior
carnaval	par
comparacion	virus
rol	dic
simon	armas

**Level 3**

bull	disminucion
saque	caption
virus	carnaval
militar	simon
irmilitar	caos
col	son
hampon	rol
interior	sensibilizacion
ningun	caen
armas	agil

**Level 4**

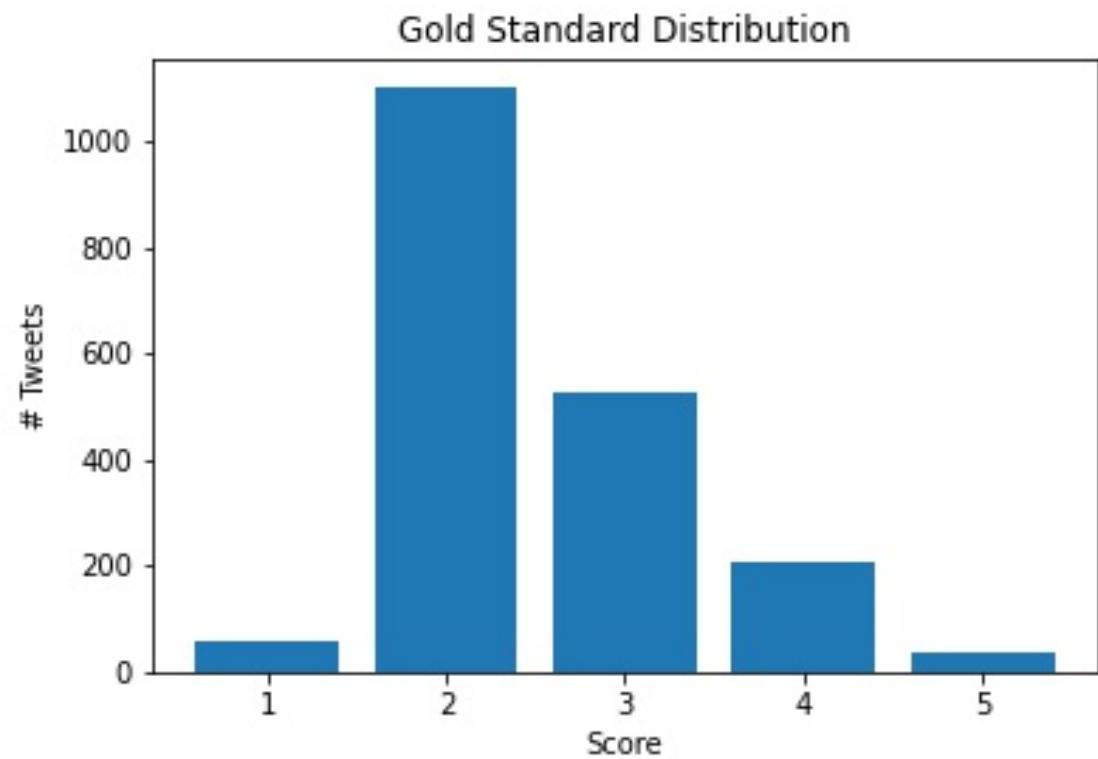
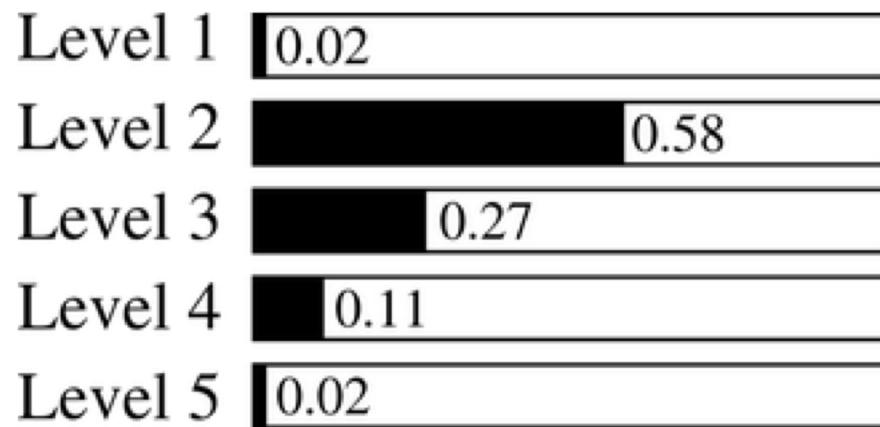
sensibilizacion	actor
usos	coronel
sur	comparacion
va	iredencionr
ojo	inclusion
señor	municipal
caos	alzar
eln	reunion
educacion	migracion
esmad	judicial

**Level 5**

alcaldes	individual
hab	documental
mil	aplicacion
asi	persecucion
par	ojos
sitp	ipes
peor	internacional
bien	cc
dan	inpec
simpueson	crimen



# MULTINOMIAL NAIIVE BAYES - LIME



# DISCUSSION & FUTURE WORK

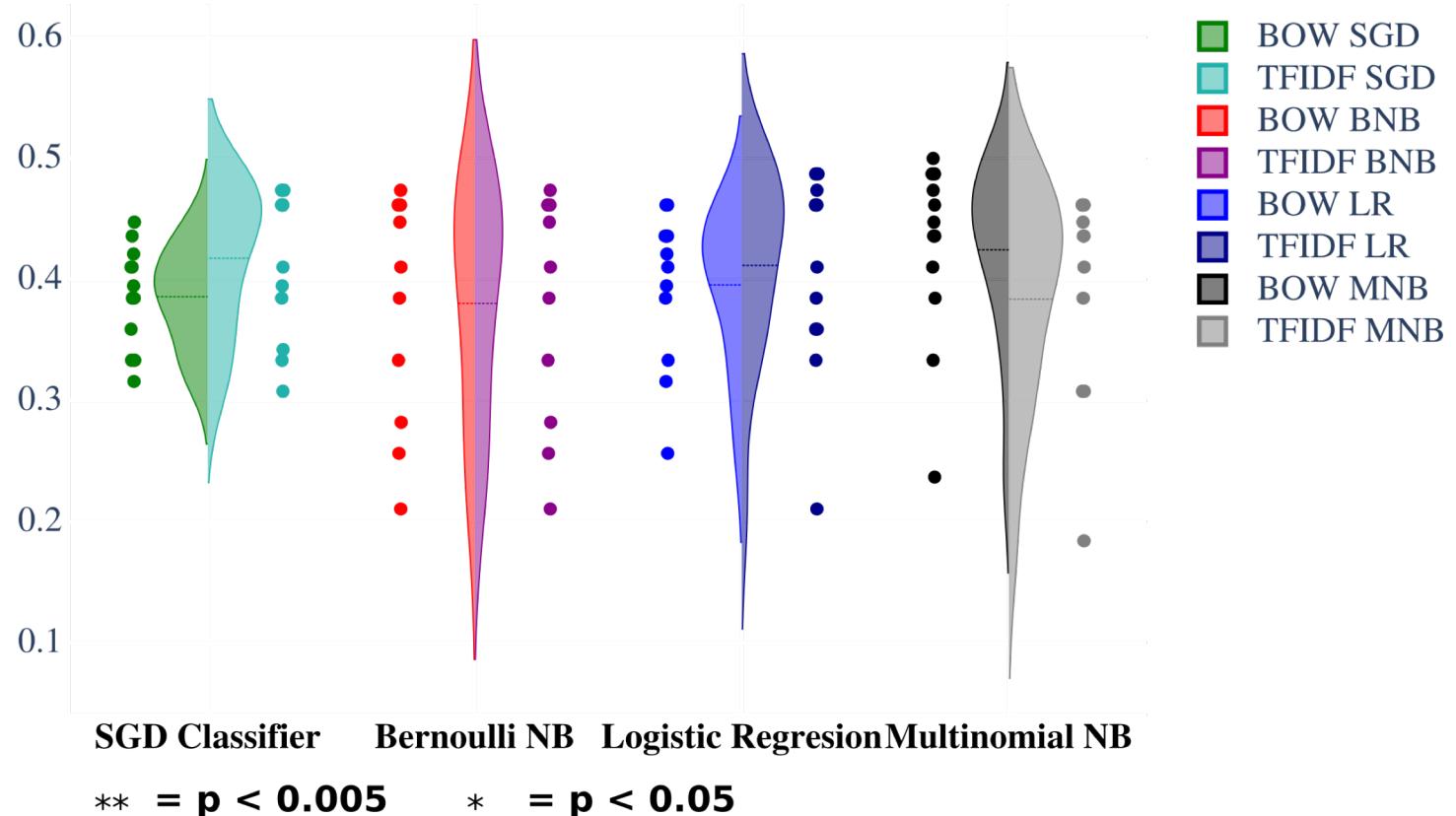
- LIME shows a good and reasonable explanation non-only for the local or the individual analysis of the tweets, where the weights given to the principal features are as expected.
- Also, the general probability of the explanation that gives the idea of the quantification of the database, in other words, give the quantification of the perception of security is quite similar to the one made by the experts at the time to tag each tweet based in the polarity of the texts.
- Analyze and interpret the results of all the texts for others classifiers with similar performance, to determine which of them allows us to better understand and adequately quantify the perception of security.
- Obtain the quantification of the Perception of security in time.

# THANKS!

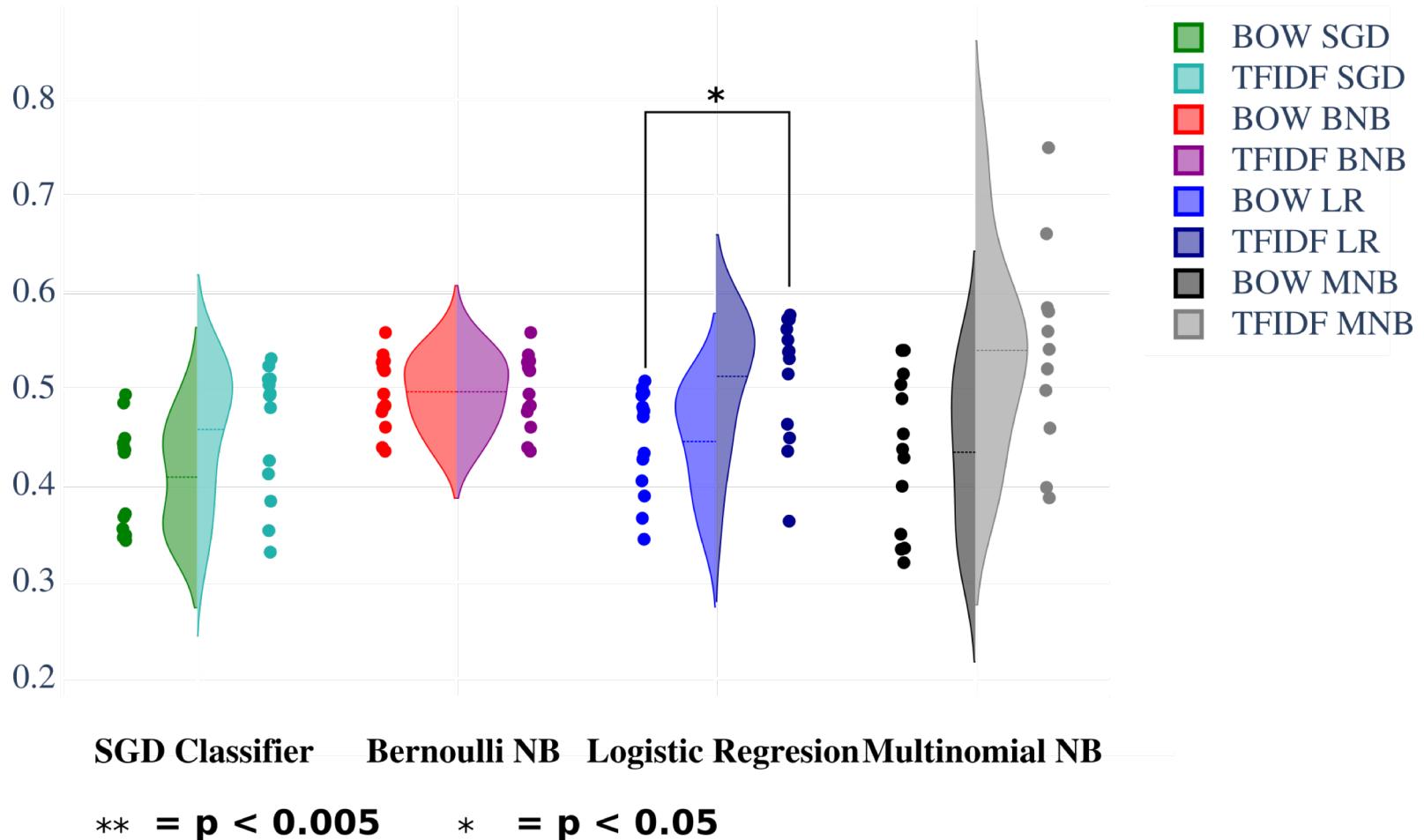
# BACKUP

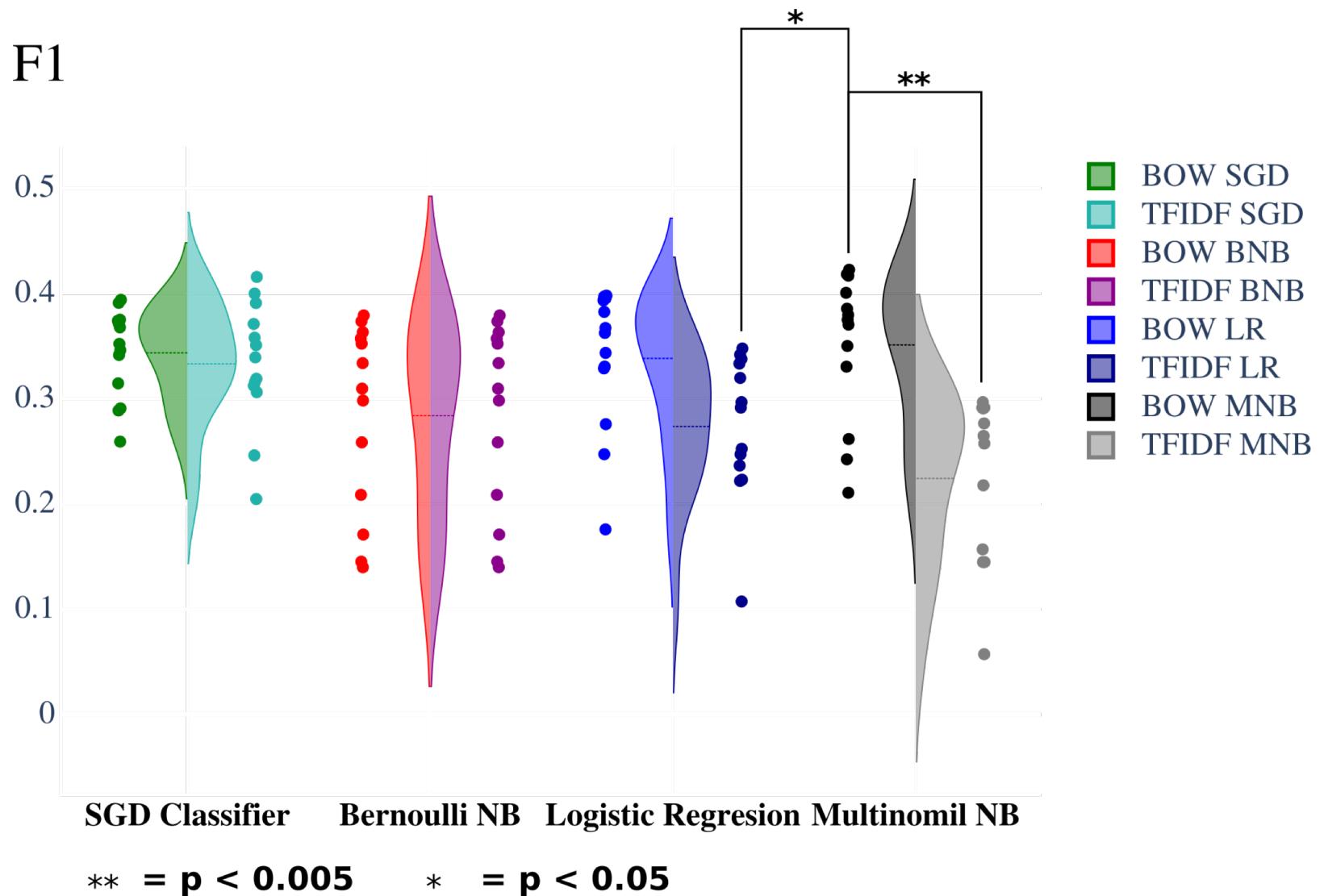


# Accuracy



# Precision





# Recall

