UNAM iimas INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y EN SISTEMAS

INSTITUTO DE INGENIERÍA UNAM

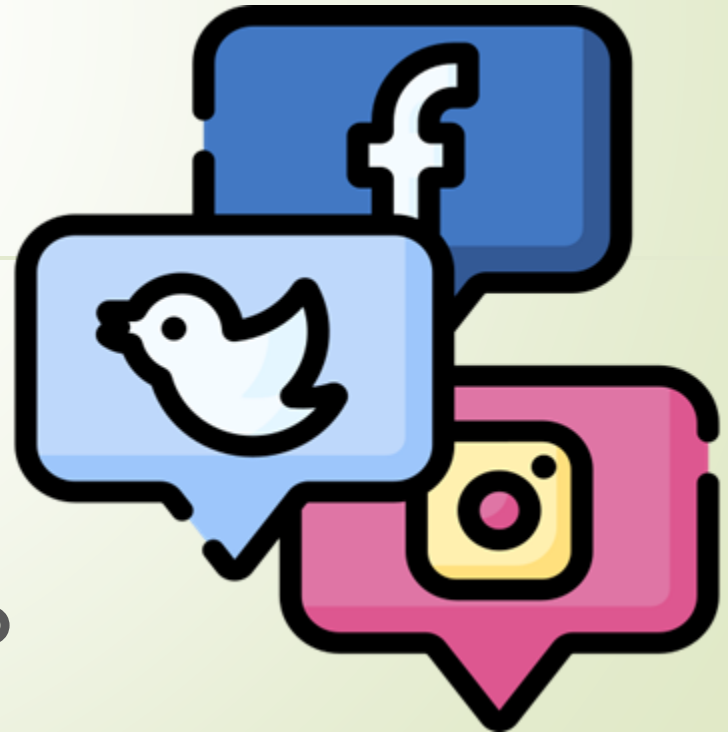
PCIC

Exploring the effectiveness of combined lightweight models in the detection of mental health disorders

Roberto Huerta Ponce

Dra. Gemma Bel-Enguix

Dra. Helena Montserrat Gómez Adorno



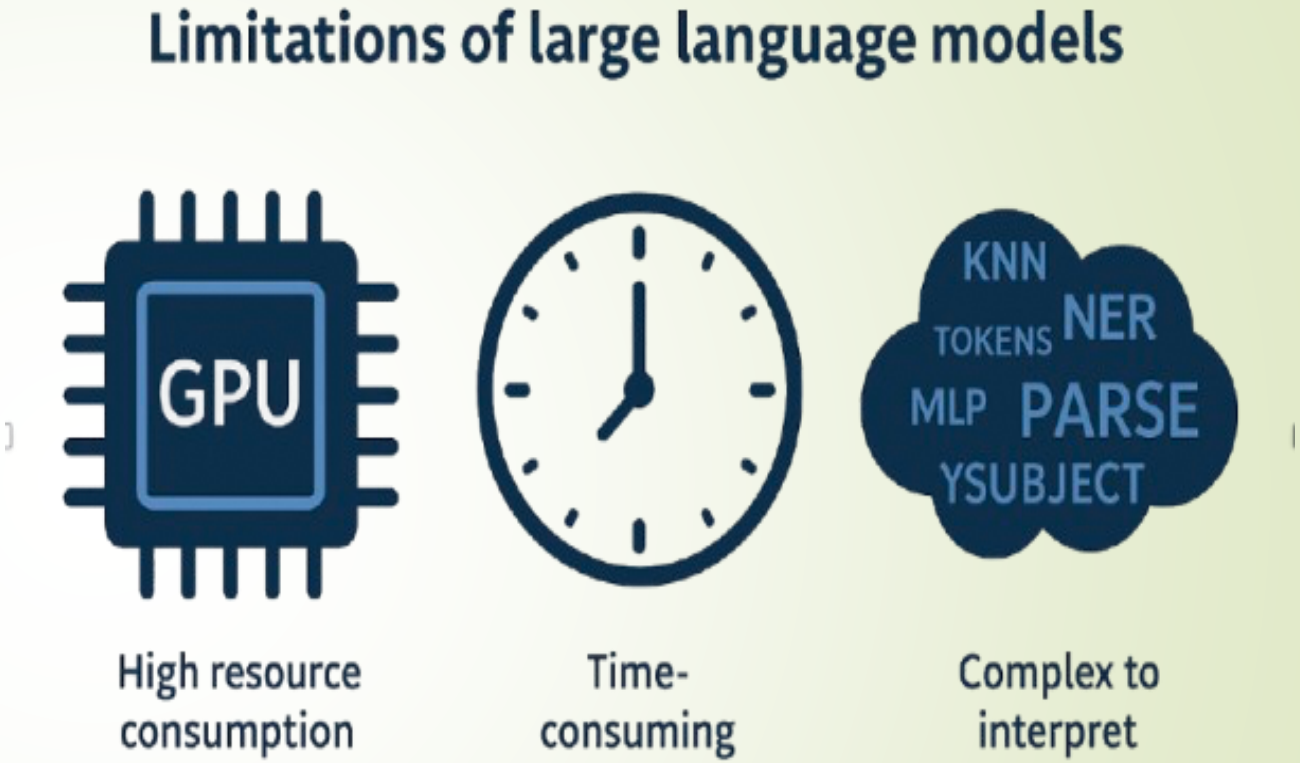
The importance of early detection

Social media contains valuable traces of people's emotional and behavioral expressions.

Analyzing these posts may help recognize patterns potentially related to mental health risks.

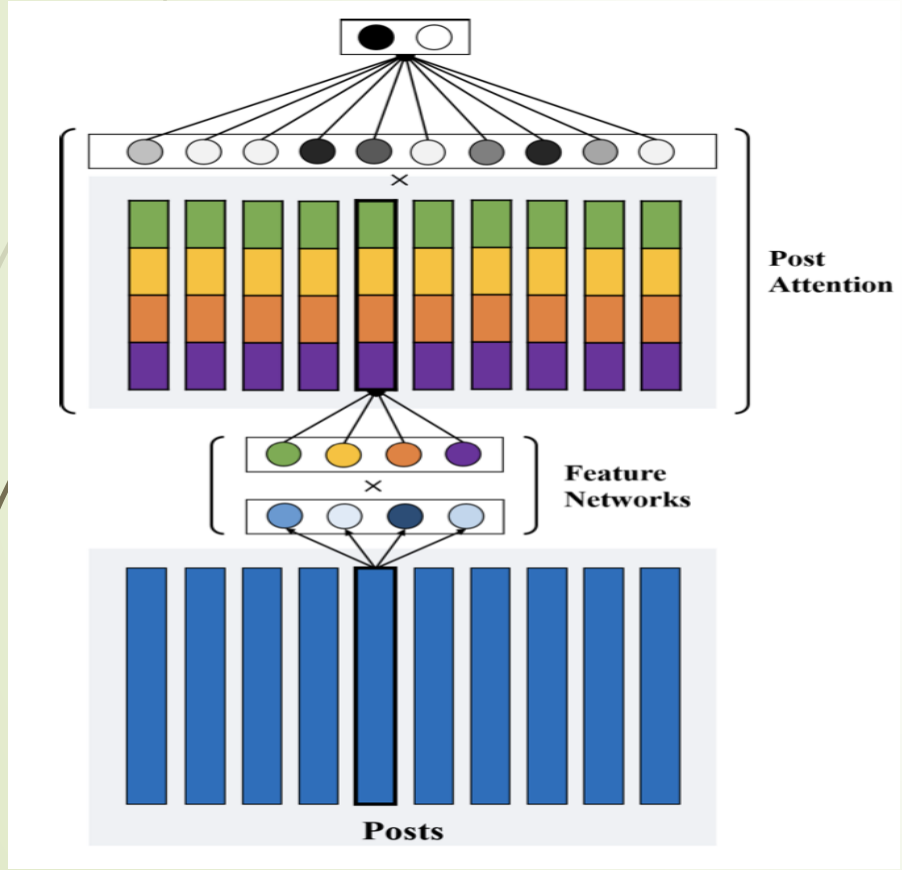
Limitations of current models

Several works have adopted BERT-based models for the detection of mental health conditions.

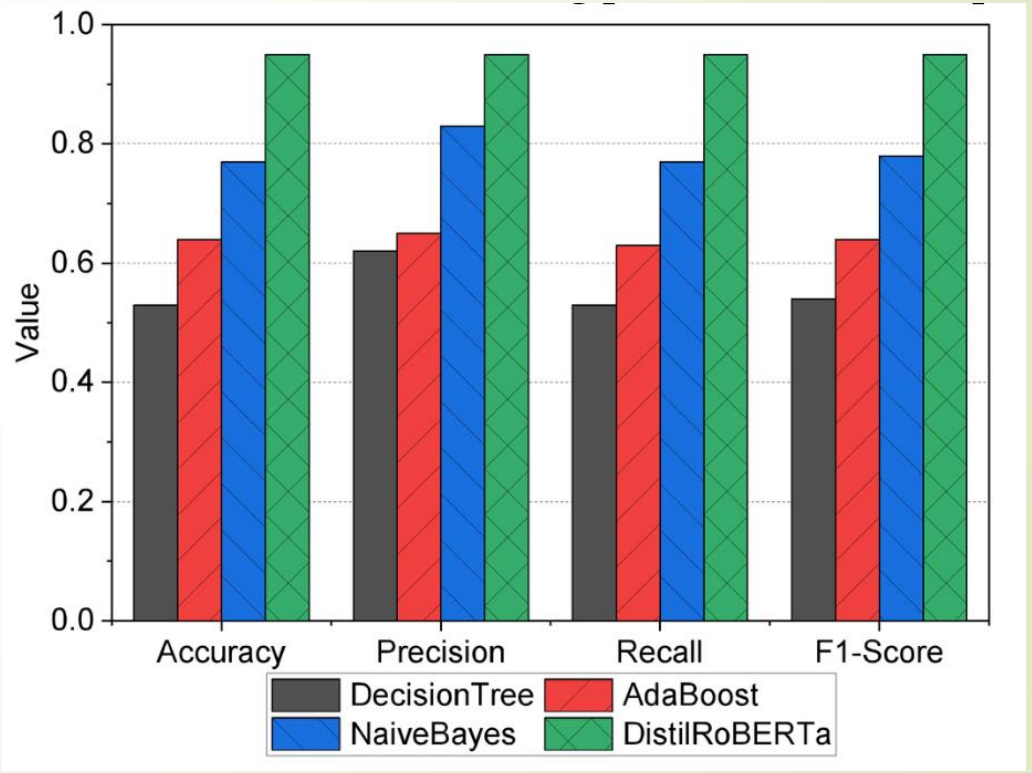


Related Work

Feature Attention Network (FAN) (Song et al., 2018):

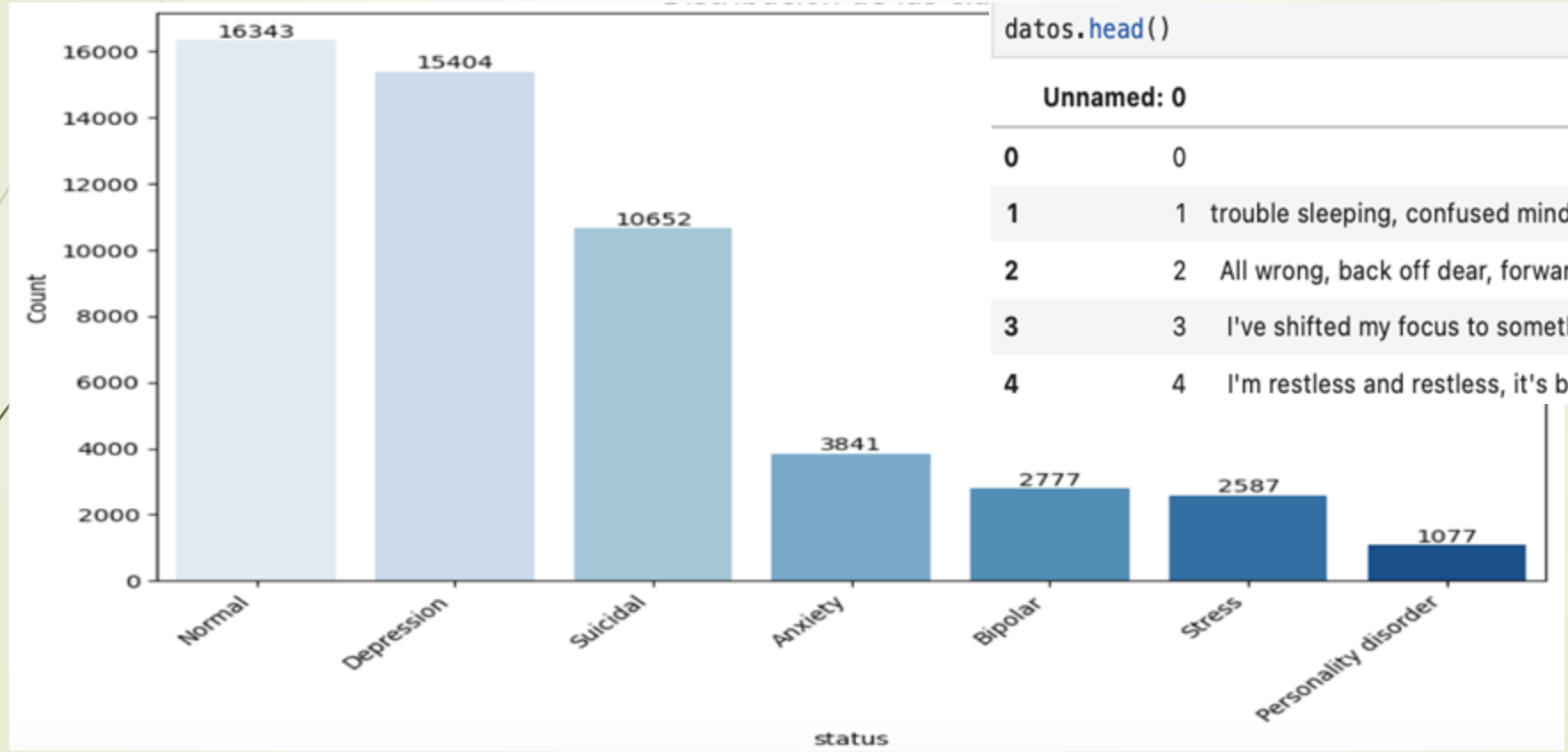


Transformer-based models (DistilRoBERTa) (Zhao et al., 2024):



Our proposal

Dataset used



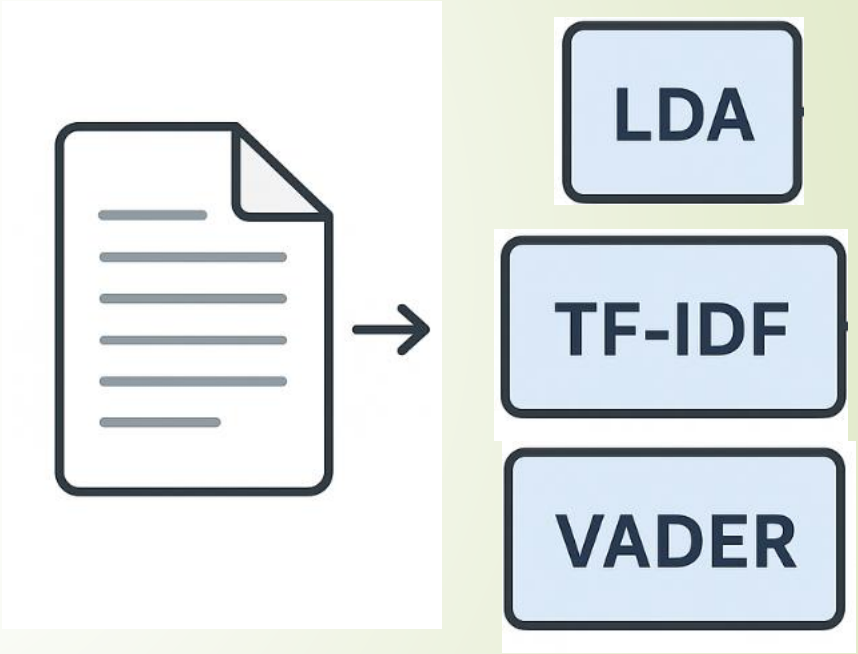

```
datos.head()
```

	Unnamed: 0	statement	status
0	0	oh my gosh	Anxiety
1	1	trouble sleeping, confused mind, restless hear...	Anxiety
2	2	All wrong, back off dear, forward doubt. Stay ...	Anxiety
3	3	I've shifted my focus to something else but I'...	Anxiety
4	4	I'm restless and restless, it's been a month n...	Anxiety

Suchintika Sarkar. **Sentiment Analysis for Mental Health Dataset*** Kaggle. Disponible en: [https://www.kaggle.com/datasets/suchintikasarkar/sentiment-analysis-for-mental-health/data]

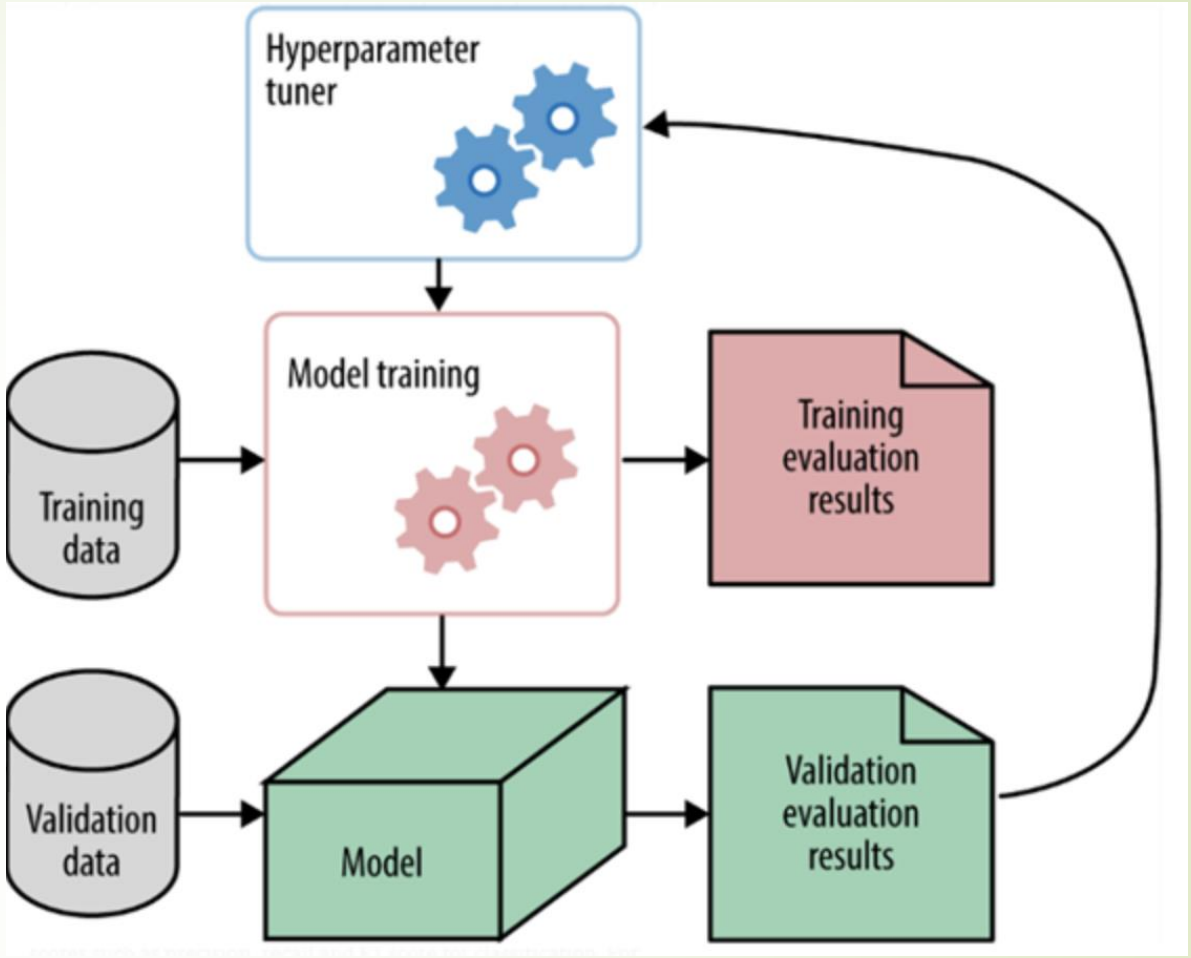
Feature extraction

- **VADER**: Analyzes sentiment.
- **TF-IDF**: Captures word relevance in each document's context.
- **LDA**: Extracts latent topics present in the text.



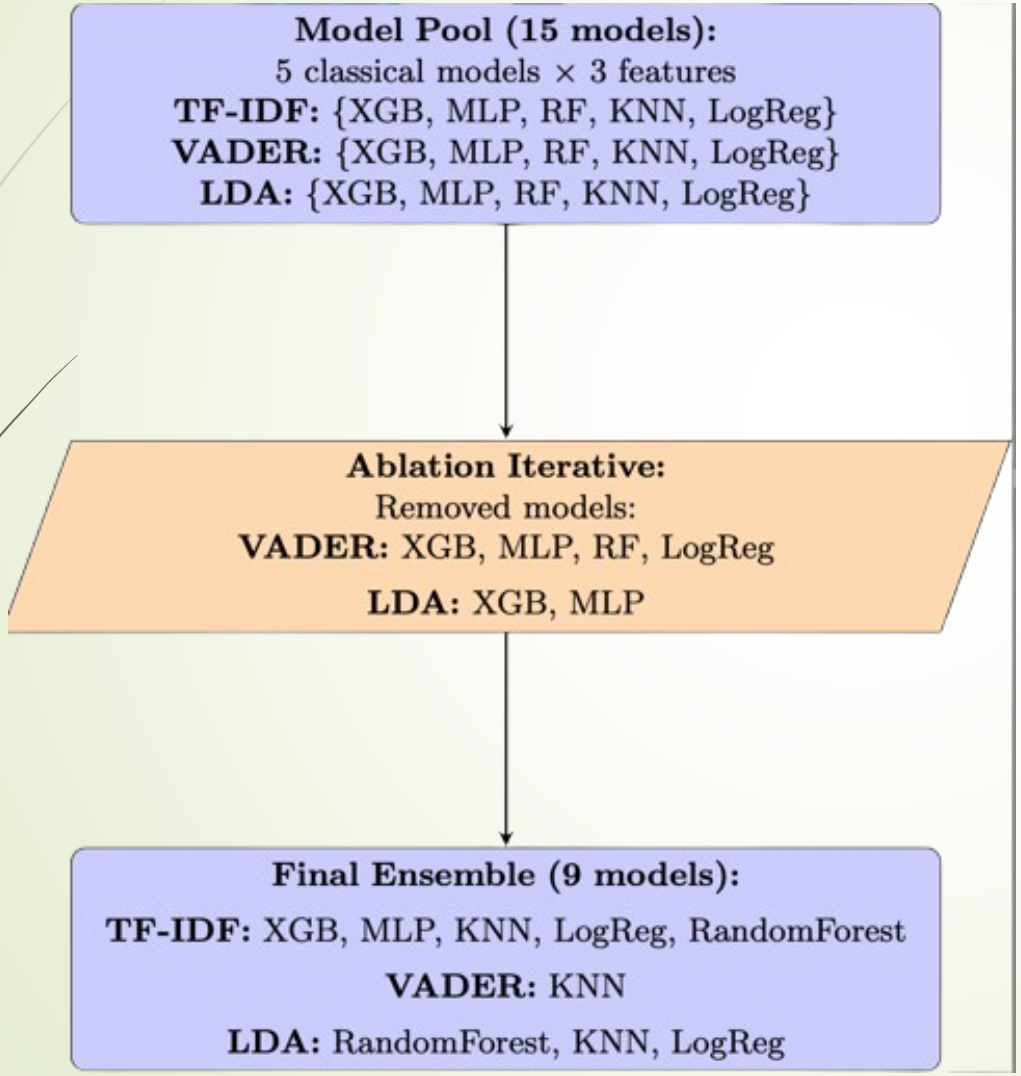
Base models

- Logistic Regression
- K-Nearest Neighbors (KNN)
- Random Forest
- Gradient Boosting
- Multilayer Perceptron (MLP)



Each model "sees" the text differently.

Ablation and Weighted Model Combination



Model	Weight
XGB_TFIDF	0.2913
MLP_TFIDF	0.2372
RandomForest_LDA	0.194
KNN_VADER	0.0855
KNN_LDA	0.0849
KNN_TFIDF	0.0618
LogReg_TFIDF	0.0439
RandomForest_TFIDF	0.001
LogReg_LDA	0.0005

Ensemble performance evaluation

Metrics used:

Model	F1 Macro (avg)	Time (min)	Memory (MB)	Processing
DistilRoBERTa	**78.32 %**	25.6	2,608.63	GPU
Our proposal	70.3 %	6.02	501.92	CPU
Base proposal	64.6 %	**3.5**	**182.21**	CPU

Platform	CPU Model	Physical Cores	Logical Cores	Frequency	GPU
Colab (with CPU)	Intel Xeon virtualized	1	2	2.2 GHz	No
Colab (with GPU)	Intel Xeon @ 2.2GHz	1	2	2.2 GHz	Tesla T4

torch.cuda.max_memory_allocated() for GPU (VRAM)
psutil.Process().memory_info().rss for CPU (RAM)

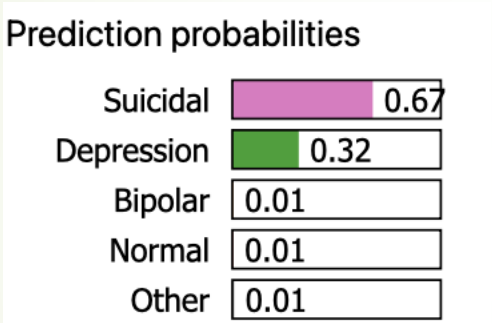
Not all models see the same

Models are complementary because each captures different aspects.

LIME evidence — both correct

LIME shows that each model focuses on different words or topics.
Correct label: Suicidal

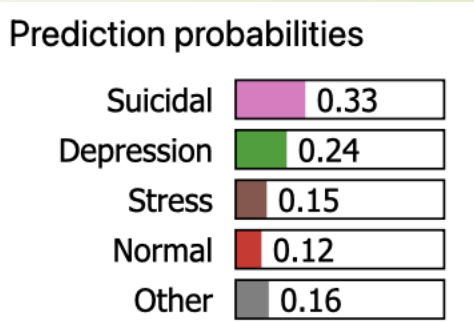
mlp + lda => Suicidal



Text with highlighted words

i am so fucking tired i have gotten diagnosed with bpd a few months ago and its made how i feel and why i feel that way a lot clearer to me but it does not make me want to end it any less i have been failing in life and loosing all of my friends and i just cannot take it anymore i really want to fucking kill myself i am so tired thinking of ending it

knn + vader => Suicidal



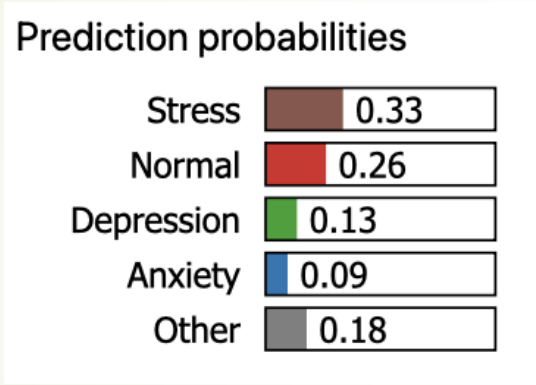
Text with highlighted words

i am so fucking tired i have gotten diagnosed with bpd a few months ago and its made how i feel and why i feel that way a lot clearer to me but it does not make me want to end it any less i have been failing in life and loosing all of my friends and i just cannot take it anymore i really want to fucking kill myself i am so tired thinking of ending it

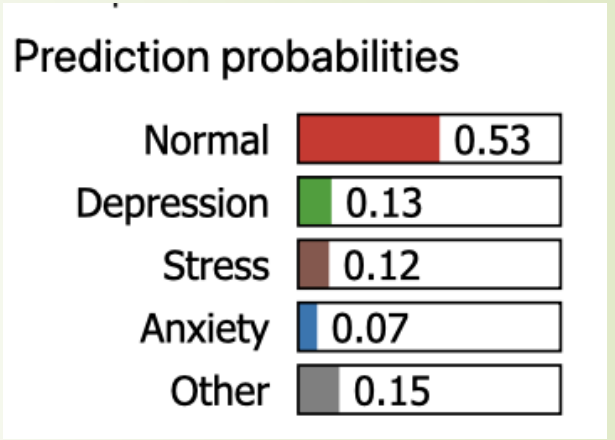
LIME evidence — only one correct

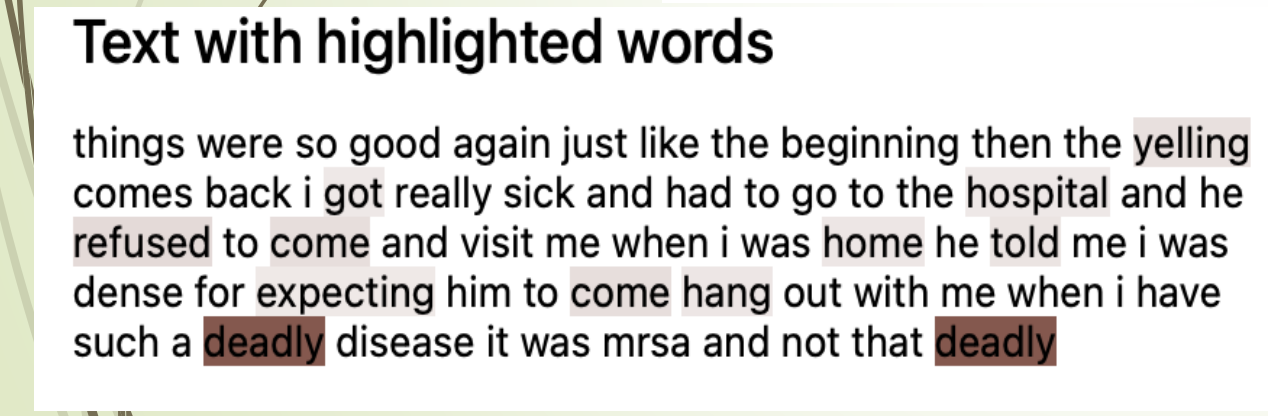
Even when models focus on similar words, they may classify differently.
Correct label: Normal

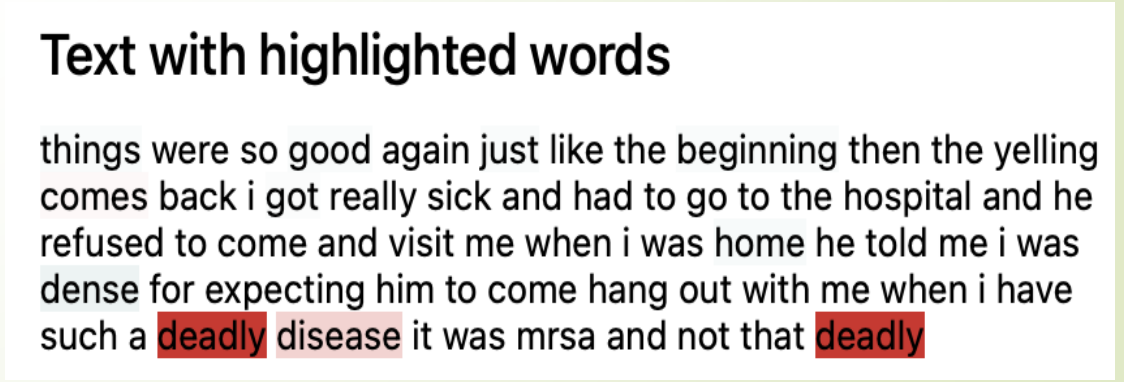
mlp + lda => Stress



logistic regression + lda => Normal



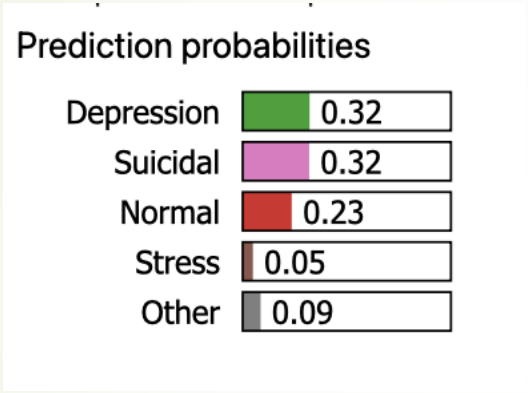




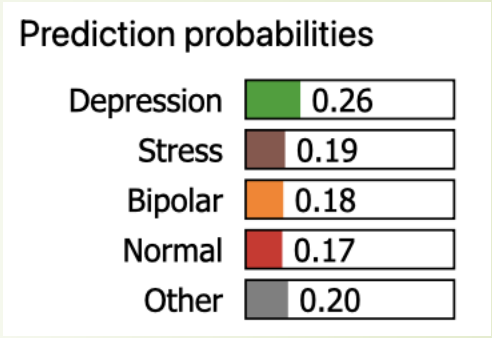
LIME evidence — both incorrect

Even when models focus on similar words, they may still fail.
Correct label: Suicidal

Random forest + tfidf => Depression



knn + vader => Depression



Text with highlighted words

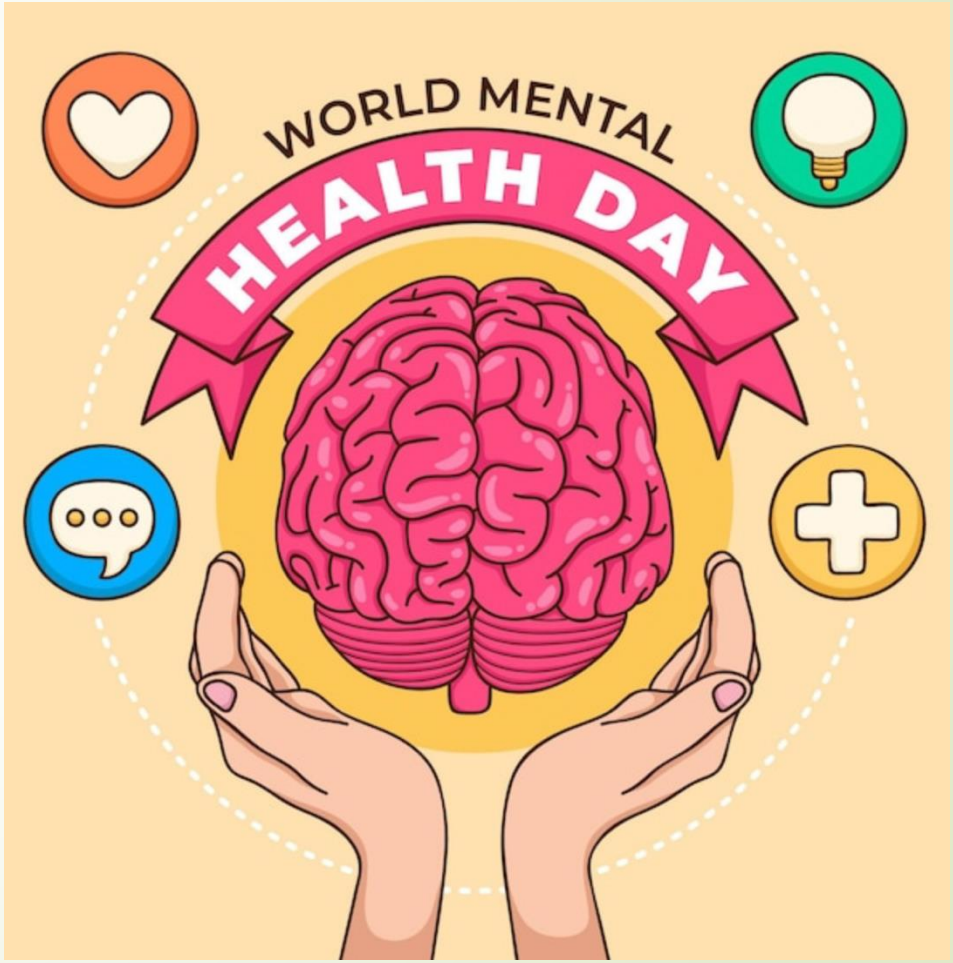
i have been sat here with a knife for the past minutes debating if i should go through with it or not nobody likes me and the rest of my life is not worth living anymore i do not know what to do

Text with highlighted words

i have been sat here with a knife for the past minutes debating if i should go through with it or not nobody likes me and the rest of my life is not worth living anymore i do not know what to do

What we learned from this work

- Lightweight models capture complementary linguistic features.
- Ensemble reduces time and computational cost.
- Interpretability remains fully accessible.
- Combining multiple linguistic techniques improves performance



Next steps and research directions

- Incorporate additional linguistic features (domain-specific lexicons)
- Explore alternative ensemble methods
- Evaluate the approach on larger and diverse datasets

Before closing…

Source: https://www.reddit.com/r/CourageTheCowardlyDog/comments/19b24g4/there_is_no_such_thing_as_perfect_youre_beautiful/?utm_source=share&utm_medium=web3x&utm_name=web3xcss&utm_term=1&utm_content=share_button

Thank you for your attention.

Project materials available online:

https://github.com/Roberhp/SCAI_presentation

Any aditional questions?

rhuertap@comunidad.unam.mx

References

1. *World Health Organization. (n.d.). Depression. World Health Organization. Retrieved May 26, 2024, from https://www.who.int/es/health-topics/depression*
2. *Song, H., You, J., Chung, J.W., & Park, J.C. (2018). Feature Attention Network: Interpretable Depression Detection from Social Media. ACL Anthology*
3. *Zhao, Z., & Wang, J. (2024). Exploring the Potential of Large Language Model in Predictive Mental Health Diagnosis of Athletes. Advances in Education, Humanities and Social Science Research.*
4. *Suchintika Sarkar. Sentiment Analysis for Mental Health Dataset. Kaggle. https://www.kaggle.com/datasets/suchintikasarkar/sentiment-analysis-for-mental-health/data*