Analyzing Mental Health: Text Classification for Mental Health Conditions

Course Lecturer: Dr. Alexander(Sasha) Apartsin

Shahar Saadon | 206560526
Dudi Saadia | 318970944
Shanel Asulin | 205616014

Introduction

Problem & Objectives

Problem Description

We aim to classify a person's mental or emotional state based on a short text message (like a post or chat).

Why is it important?

It can help detect mental health issues early and support professionals in understanding emotional states.

Why is it challenging?

The texts are short, vague, and often written in informal or inconsistent language.

Introduction

Problem & Objectives

Project Goals

- ❑ Explore if machine learning can classify emotional states from text.
- ❑ Compare several models to find the best one.
- ❑ Understand which words are most important for each emotion.

Literature Review

Key Contribution	Best Result	Methods	Dataset & Task	Study
TF-IDF text features, simple yet effective	77% (LightGBM)	Naive Bayes, MLP, LightGBM	10K Reddit posts (6 conditions)	Nova (2023)
Multi-model benchmark with deep/transfer learning	83% (RoBERTa)	Traditional ML, DL, RoBERTa	17K Reddit posts (5 classes)	Ameer et al. (2022)
Detection from general text, not only support forums	81% (RoBERTa)	BERT, XLNet, RoBERTa	100K Reddit posts (9 DSM-5 conditions)	Dinu & Moldovan (2021)
Comparison across ML, BERT, and LLMs in Russian corpora	88% (LLM finetuned)	AutoML, RuBERT, Vikhr, LLaMA3	5 datasets (DE/DSM/AL/AD/AC - Depression & Anxiety, in Russian)	Kuzmin et al. (2024)
Zero-shot ChatGPT (gpt-3.5) on 3 mental health tasks	86% (depression)	ChatGPT Zero-shot	Reddit/blog posts – depression, stress (binary), suicidality (5-class)	Lamichhane (2023)
Instruction-finetuned LLMs outperform GPT-4 in accuracy	+4.8% (over GPT-4)	Zero-/Few-shot, Instruction Finetuning	Reddit (mental health labels) – multi-task setup	Xu et al. (2024)

Dataset & Preparation

Data Sources

Main dataset: Over 53,000 text statements labeled from Kaggle with 7 mental health states.
Additional dataset: Statements labeled by depression levels (Mild, Severe, Normal)

Merging Strategy

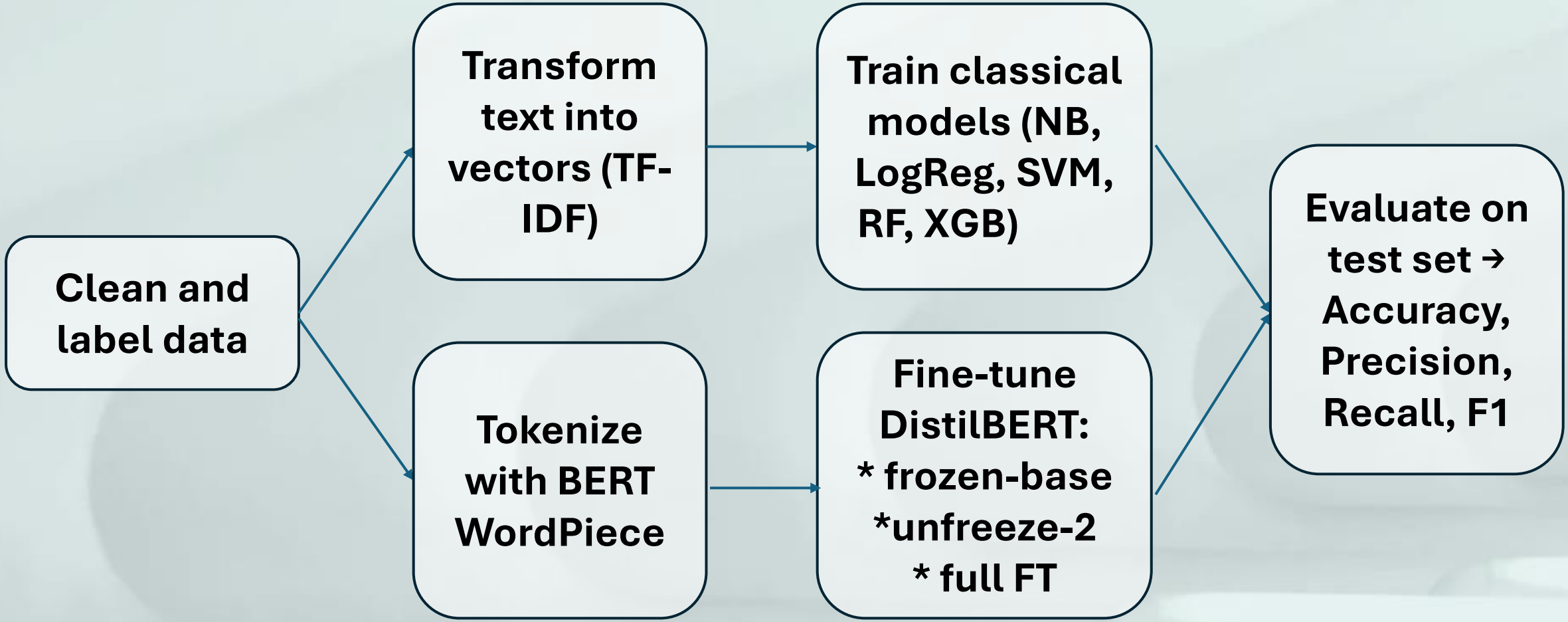
We added the new data to increase the number of examples.
>>All "Mild" and "Severe" were labeled as Depression
>>All "Normal" statements were labeled as Normal

Final Format

Text: short English statements from social media

Label: one of 7 categories (e.g., Anxiety, Depression, Stress, etc.)

Models and Processing Pipelines



What we used:

Text preprocessing: TF-IDF vectorization with and without n-grams.

Models:

- Naive Bayes
- Logistic Regression
- SVM
- Random Forest
- XGBoost
- BERT
- DistilBERT

Training details:

* Train/Test split: 80/20.
*Balanced classes: using stratified split.
* Platform: local Jupyter Notebook.
* Batch size & epochs: default values for classical models.

For DistilBERT:

4 epochs, 2x4 batch.

(4×2 = batch 4, grad-acc 2)

How We Used the Metrics

Metrics Overview

- Accuracy – overall correctness
- Precision – false-alarm control
- Recall – catch positives
- F1 – balance P/R

During training/testing –

- Training: monitor val-loss & F1, stop if plateau.
- Testing: rank models by macro-F1.

Why we used multiple metrics - Accuracy can mislead on imbalanced data, so we rely on macro-F1 for fair comparison

Intermediate & Baseline Results

	Model	Accuracy	Precision	Recall	F1 Score
0	Naive Bayes	0.700	0.710	0.690	0.70
1	Logistic Regression (TF-IDF)	0.740	0.740	0.740	0.73
2	Logistic Regression (n-grams)	0.750	0.740	0.740	0.74
3	SVM (n-grams)	0.730	0.730	0.730	0.73
4	Random Forest	0.690	0.720	0.680	0.66
5	XGBoost	0.750	0.750	0.750	0.75
6	DistilBERT 2 K / 3 E	0.666	0.670	0.680	0.69
7	DistilBERT 10 K (frozen)	0.664	0.660	0.670	0.68
8	DistilBERT 10 K (unfreeze-2)	0.719	0.720	0.710	0.72
9	DistilBERT 55 K (full FT)	0.793	0.767	0.785	0.78

DstilBERT Results

🔍 Predictions on new text samples:

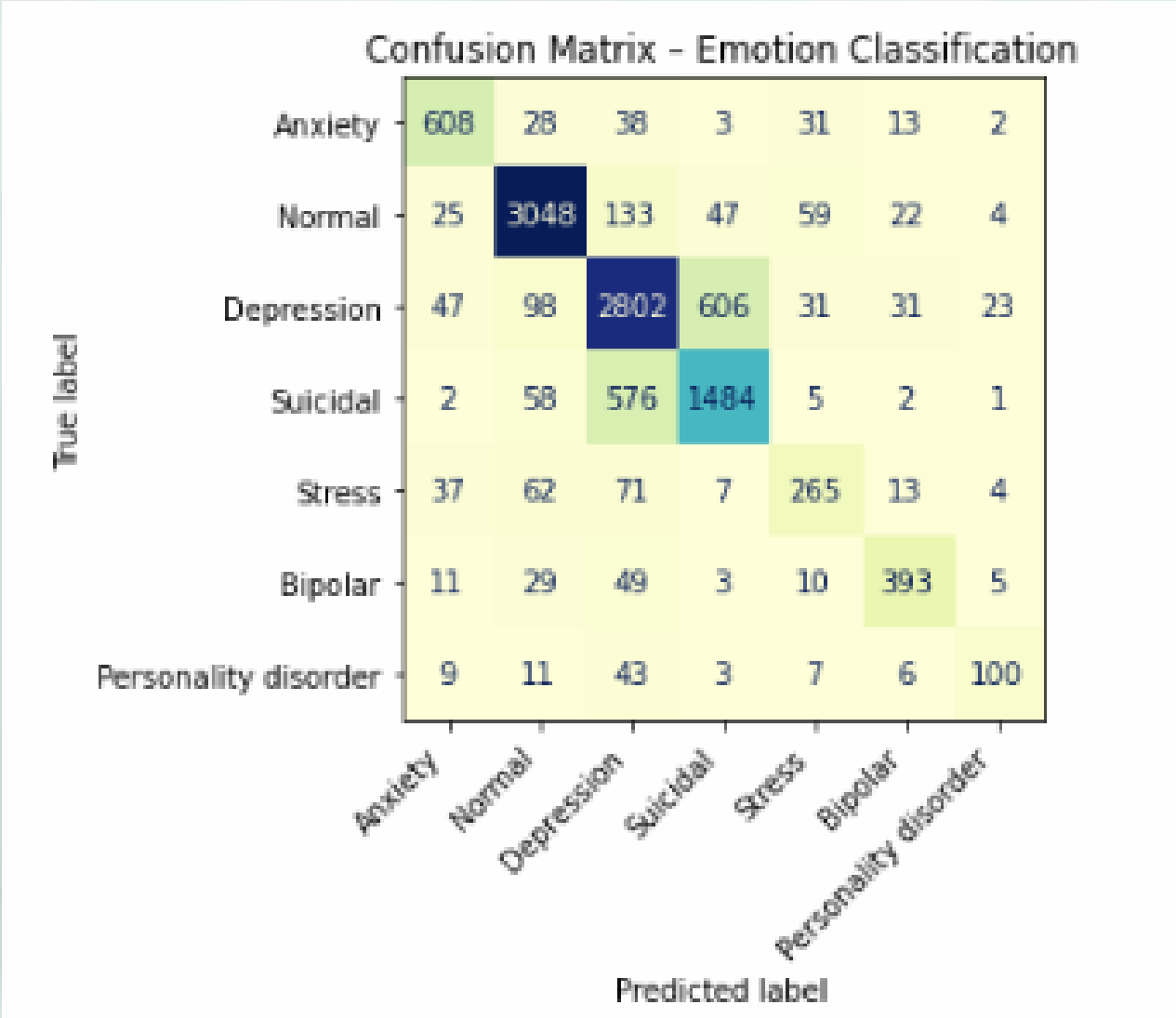
Text: I feel amazing and productive today!
Predicted state: Normal

Text: Nothing makes sense anymore.
Predicted state: Depression

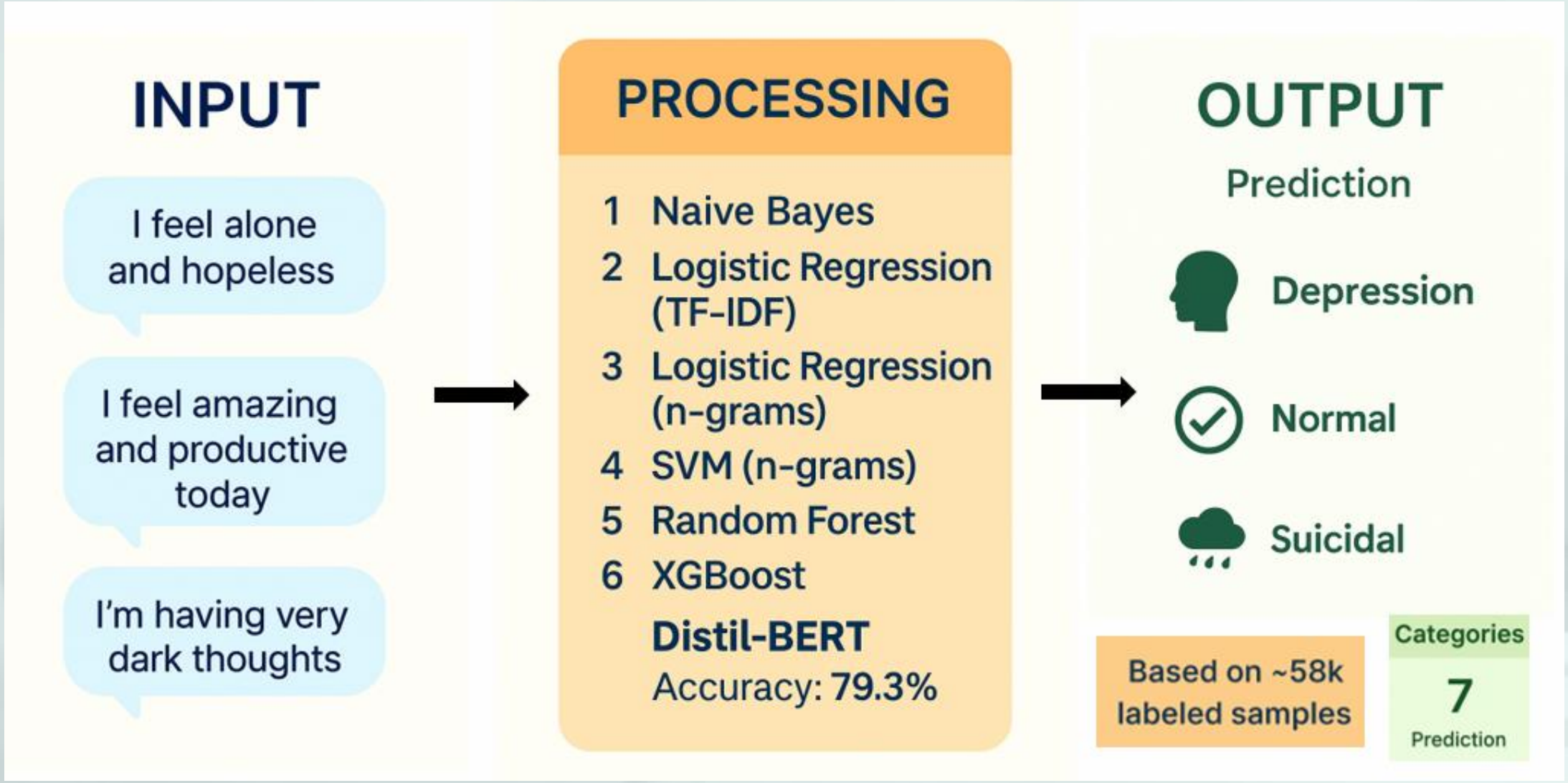
Text: I'm anxious about everything I do.
Predicted state: Anxiety

Text: I want to die
Predicted state: Suicidal

Text: im very hungry
Predicted state: Normal



Graphical abstract



Thank You ☺