

Аналіз текстових повідомлень із використанням методів обробки природної мови та машинного навчання для виявлення симптомів депресії

Доповідач: Дехтяренко Мар'яна Сергіївна, КН-4

Керівник: Борозенний Сергій Олександрович, ст.в.

Мета й завдання дослідження

Мета роботи : впровадження системи автоматизованого аналізу текстового контенту щоб забезпечити можливість виявлення симптомів депресії та інших психічних розладів за допомогою методів машинного навчання та NLP.

Об'єктом дослідження є процес автоматизованого оброблення текстових даних для виявлення депресивних ознак.

Предметом дослідження є алгоритмічні та програмні засоби попереднього оброблення тексту, векторизації, побудови та порівняння класифікаційних моделей, а також їхнє впровадження в прикладну систему скринінгу депресії.

Мета й завдання дослідження

Для досягнення поставленої мети виконано такі завдання:

- проведено стандартизоване оброблення тексту
- реалізовано чотири методи векторизації та виконано порівняльний аналіз їхнього впливу на ефективність класифікаційних алгоритмів.
- розроблено та порівняно роботу трьох класифікаторів
- оцінено продуктивність моделей за допомогою метрик accuracy, precision, recall, F1-score та матриці невідповідностей
- створено клієнт-серверний вебзастосунок, який надає користувачам можливість оперативно отримувати попередню оцінку свого ментального стану.

ДАНІ Й ПОПЕРЕДНЯ ОБРОБКА

Використані збірки даних для навчання моделей та основні етапи попередньої обробки тексту

«Reddit Mental Health Data»

0 = Stress

1 = Depression

2 = Bipolar disorder

3 = Personality disorder

4 = Anxiety

«Student-Depression-Text»

0 = normal mental state

1 = depression

МЕТОДИ ВЕКТОРИЗАЦІЇ ТЕКСТУ

- **Bag-of-Words**(фіксує частоту кожного слова в документі, але не фіксує контекст і порядок слів).
- **N-грами**(враховують послідовності з n слів (біграми, триграми тощо), що дозволяє вловлювати локальний контекст і стійкі словосполучення).
- **TF-IDF**(враховує не лише частоту слова у тексті, а й рідкість слова в усій колекції, підсилюючи вплив унікальних та важливих термінів)
- **Word2Vec**(Перетворює слова на багатовимірні вектори так, щоб слова зі схожим змістом розташовувалися близько одне до одного у векторному просторі; враховує семантику контексту).

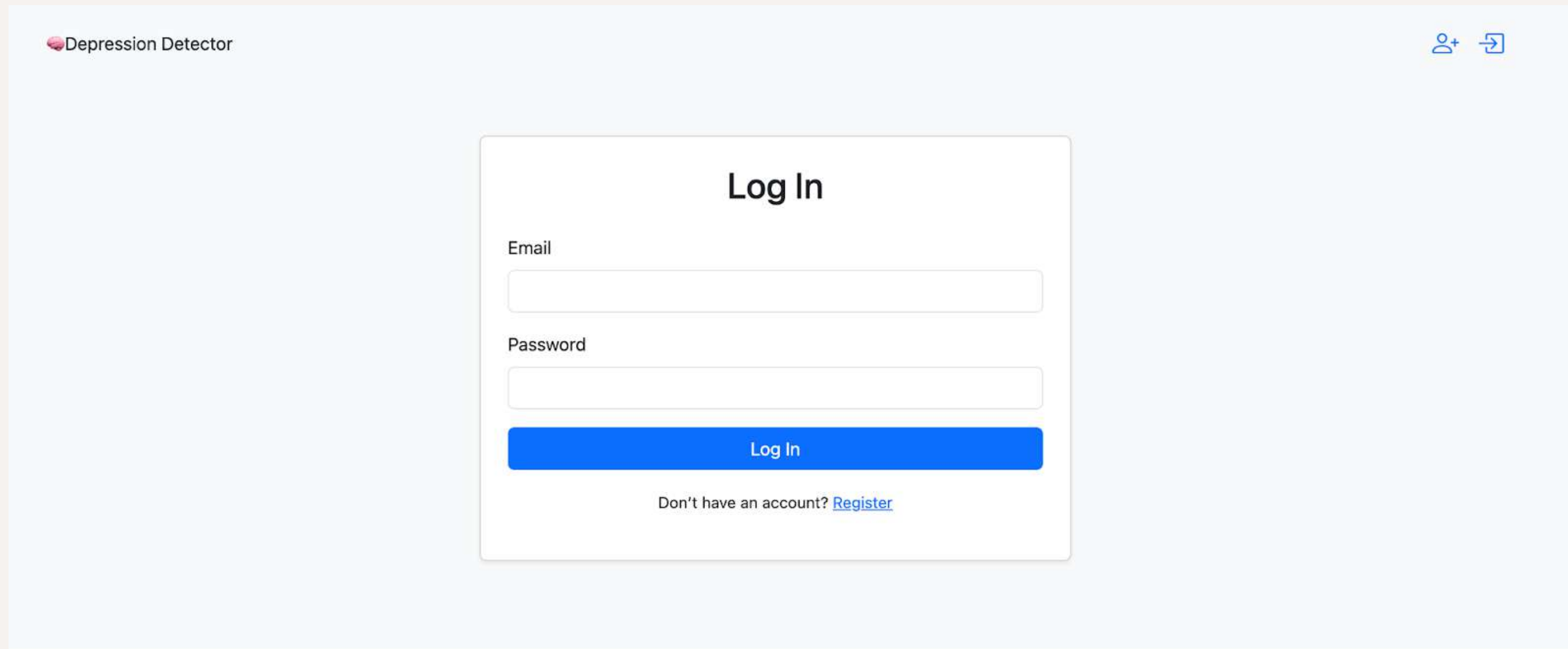
МОДЕЛІ МАШИННОГО НАВЧАННЯ

- **Support Vector Machine (SVM)**- шукає оптимальну гіперплощину для розділення класів, використовуючи “kernel trick” для роботи з нелінійними залежностями.
- **Naive Bayes**- класифікує тексти на основі ймовірностей і припущення незалежності слів; особливо ефективний із розрідженими даними.
- **Random Forest**- формує ансамбль дерев рішень, кожне з яких навчається на різних підмножинах даних і ознак; фінальний прогноз базується на голосуванні дерев, що забезпечує високу точність і стійкість до шуму.

ОЦІНКА ЯКОСТІ МОДЕЛЕЙ

Модель	Векторизація	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Naive Bayes	Bag-of-Words	71.73	73.25	71.73	71.95
	N-грами	75.42	77.58	75.42	75.75
	TF-IDF	68.20	72.21	68.20	68.71
	Word2Vec	30.96	45.43	30.96	27.68
SVM	Bag-of-Words	66.69	69.31	66.69	66.73
	N-грами	68.88	72.57	68.88	68.97
	TF-IDF	59.14	72.33	59.14	59.95
	Word2Vec	36.24	44.24	36.24	31.81
Random Forest	Bag-of-Words	75.42	77.74	75.42	75.77
	N-грами	73.07	76.74	73.07	73.54
	TF-IDF	77.85	79.54	77.85	78.19
	Word2Vec	58.98	59.39	58.98	58.90

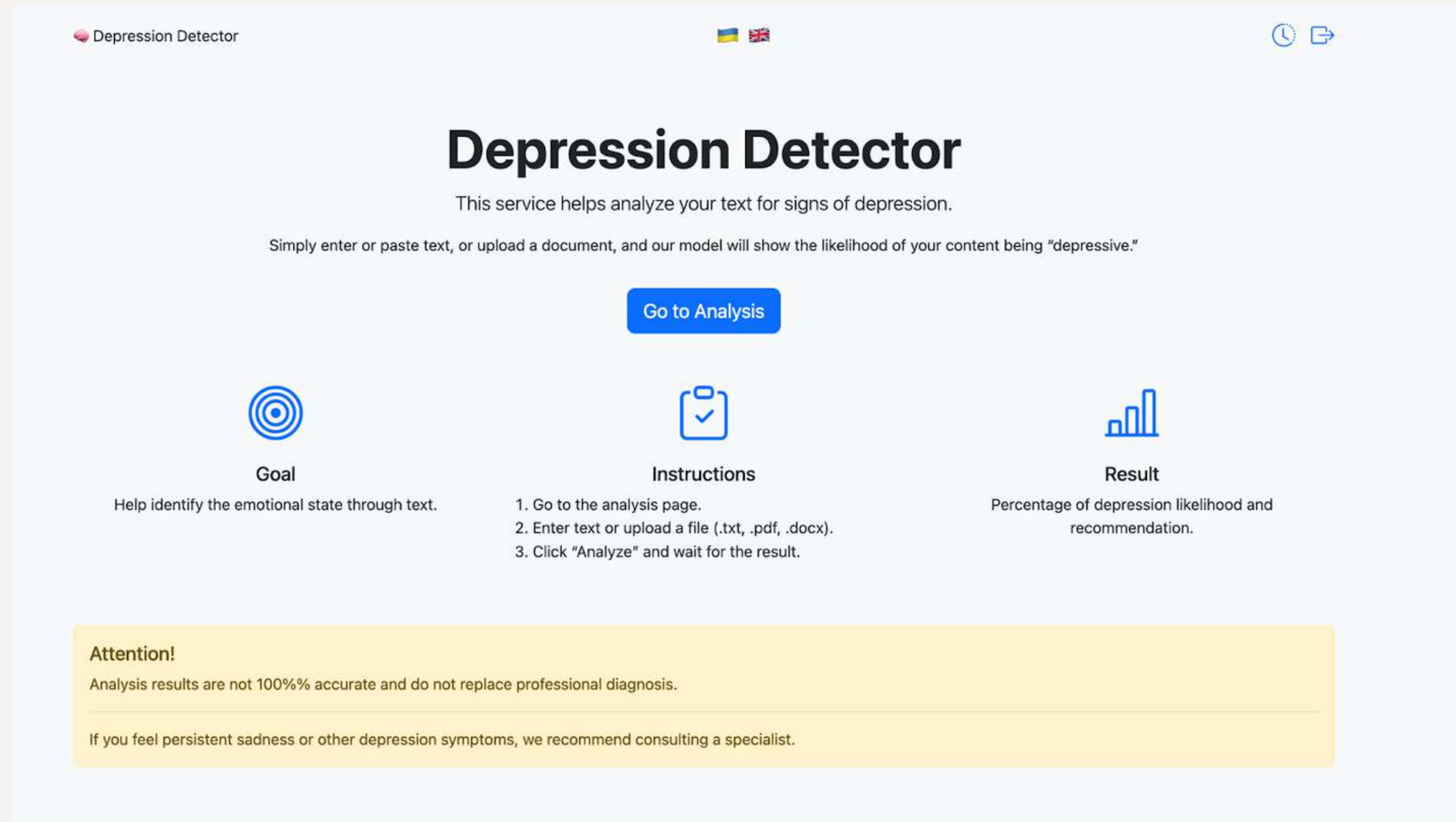
РОЗРОБЛЕННЯ ВЕБ-ЗАСТОСУНОКУ



The screenshot shows a web application interface for 'Depression Detector'. In the top left corner, there is a logo consisting of a brain icon and the text 'Depression Detector'. In the top right corner, there are two icons: a person icon with a plus sign and a square icon with an arrow pointing out. The main content is a white rectangular box with rounded corners, titled 'Log In'. Inside this box, there are two input fields: the first is labeled 'Email' and the second is labeled 'Password'. Below these fields is a prominent blue button with the text 'Log In'. At the bottom of the box, there is a link that says 'Don't have an account? [Register](#)'.

Вікно входу у застосунок

РОЗРОБЛЕННЯ ВЕБ-ЗАСТОСУНОКУ



Головна сторінка

РОЗРОБЛЕННЯ ВЕБ-ЗАСТОСУНОКУ

Depression Detector 🇺🇦 🇬🇧 🕒 ↗

Text Analysis

[← Home](#)

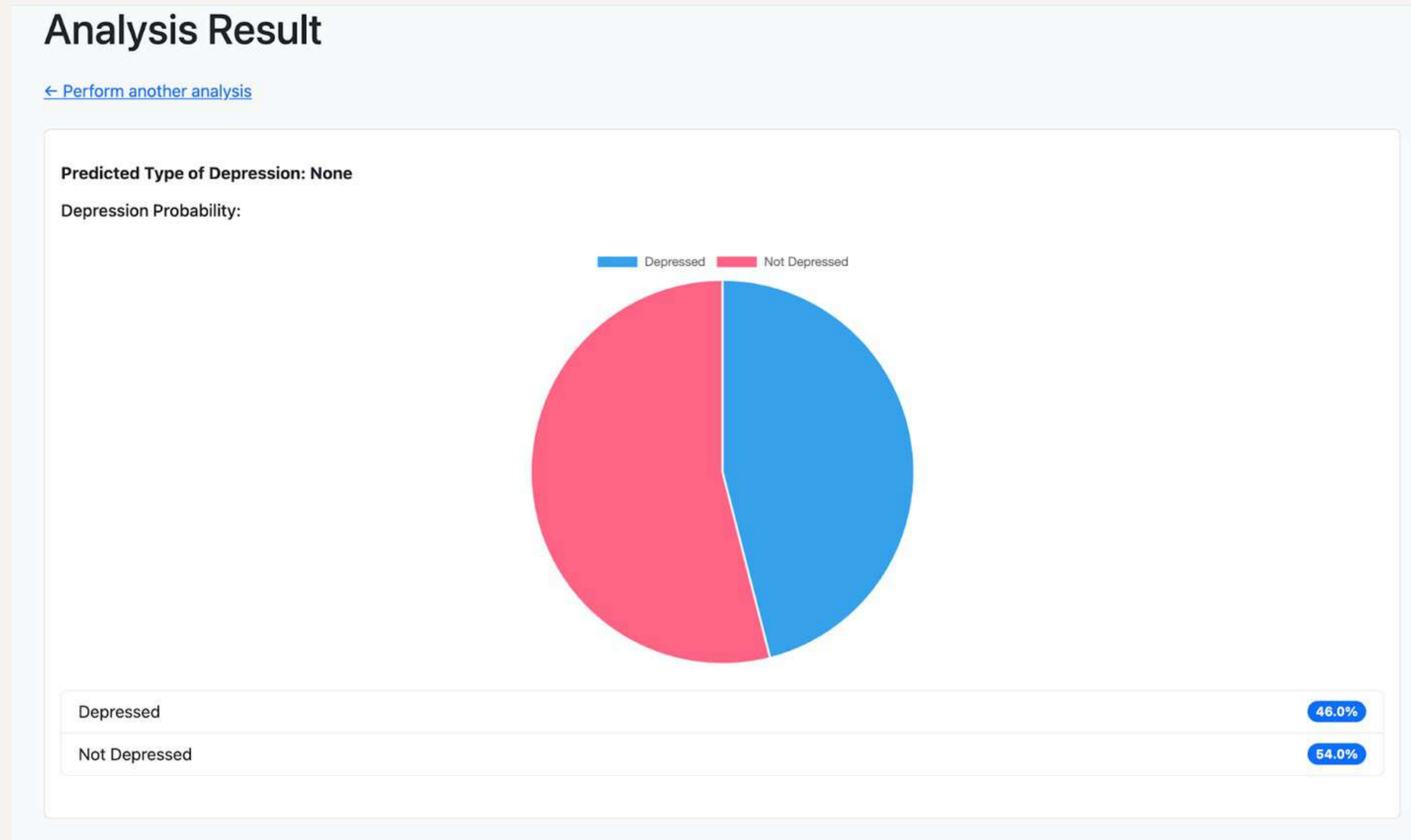
Age: Gender: Age Category:

Enter text for analysis

Or upload a file

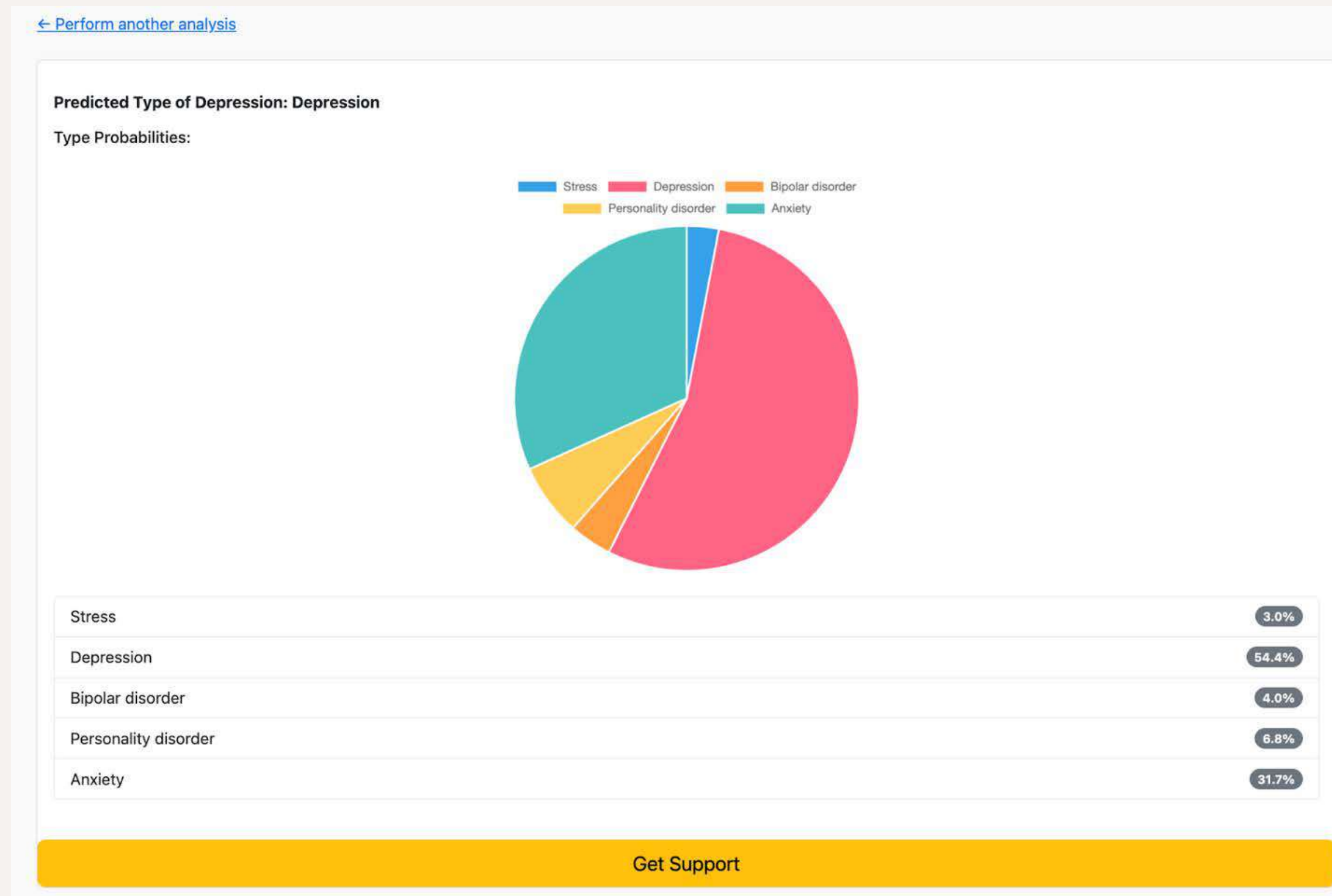
Сторінка аналізу

РОЗРОБЛЕННЯ ВЕБ-ЗАСТОСУНОКУ



Сторінка з підсумками аналізу ментального стану користувача

РОЗРОБЛЕННЯ ВЕБ-ЗАСТОСУНОКУ



Сторінка з аналізом можливих психологічних розладів користувача

РОЗРОБЛЕННЯ ВЕБ-ЗАСТОСУНОКУ

If you need help to cope with Depression

Information about depression and available support.

Helpful resources:

[Free online therapy sessions](#)

[Online support groups](#)

[Articles about depression](#)

[Back to Analysis](#)

Сторінка підтримки

РОЗРОБЛЕННЯ ВЕБ-ЗАСТОСУНОКУ

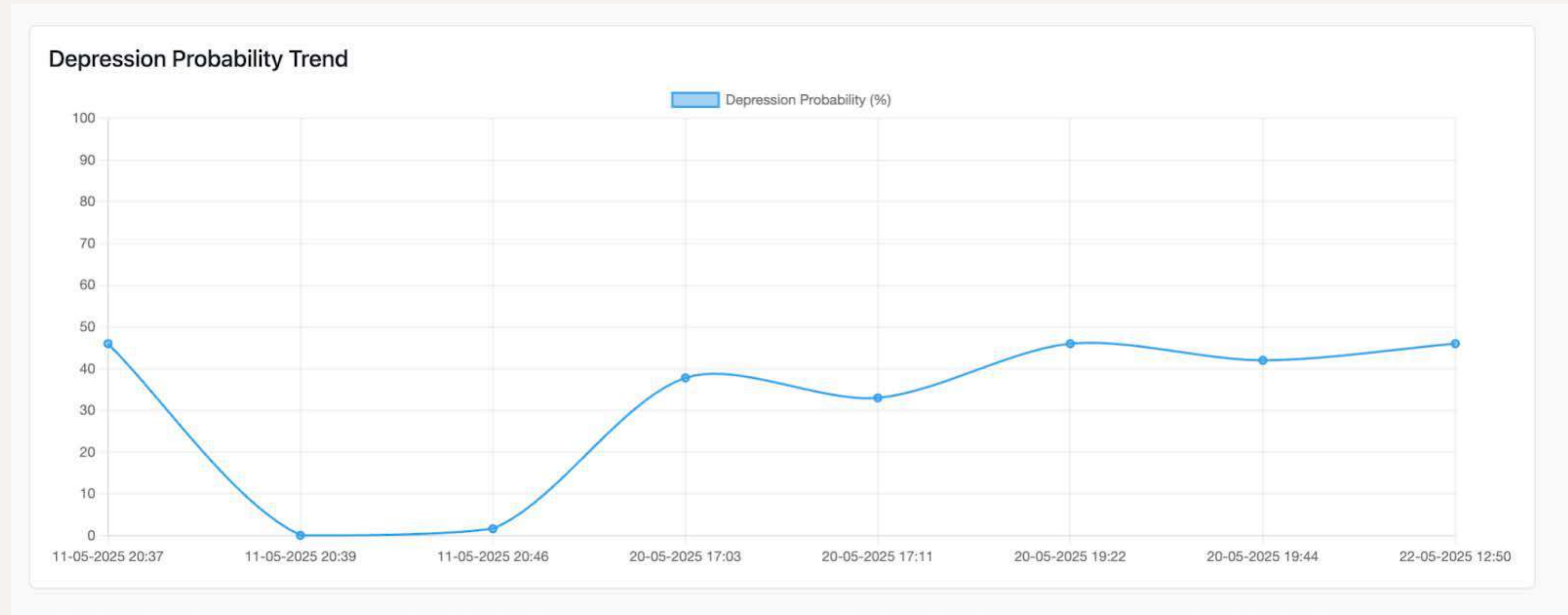
Analysis History

New Analysis

Date	Time	Probability	Result	Actions
11-05-2025	20:37	Depressed: 46.0% Not Depressed: 54.0%	Not Depressed	Delete
11-05-2025	20:39	Depressed: 0.0% Not Depressed: 100.0%	Not Depressed	Delete
11-05-2025	20:46	Depressed: 1.6% Not Depressed: 98.4%	Not Depressed	Delete
20-05-2025	17:03	Depressed: 37.8% Not Depressed: 62.2%	Not Depressed	Delete
20-05-2025	17:11	Depressed: 33.0% Not Depressed: 67.0%	Not Depressed	Delete
20-05-2025	19:22	Depressed: 46.0% Not Depressed: 54.0%	Not Depressed	Delete
20-05-2025	19:44	Depressed: 42.0% Not Depressed: 58.0%	Not Depressed	Delete
22-05-2025	12:50	Depressed: 46.0% Not Depressed: 54.0%	Not Depressed	Delete

Сторінка історії аналізів користувача

РОЗРОБЛЕННЯ ВЕБ-ЗАСТОСУНОКУ



Графік для відстеження динаміки депресії за запитами користувача

Висновки

- Розроблено автоматизовану систему для аналізу текстових повідомлень із використанням ML і NLP для виявлення симптомів депресії та психічних розладів.
- Проведено порівняльний аналіз чотирьох методів векторизації (Bag-of-Words, N-грами, TF-IDF, Word2Vec) і трьох алгоритмів класифікації (Naive Bayes, SVM, Random Forest).
- Найкращу точність і стабільність досягнуто при поєднанні Random Forest із TF-IDF (accuracy $\approx 78\%$, F1 $\approx 78\%$).
- Створено вебзастосунок для швидкої автоматизованої оцінки ментального стану користувача за текстом, що не замінює професійну діагностику.

Дякую за увагу!