Prediction of depression score based on AI art prompts

Klaus Kellerwessel, 2024

Supplementary material

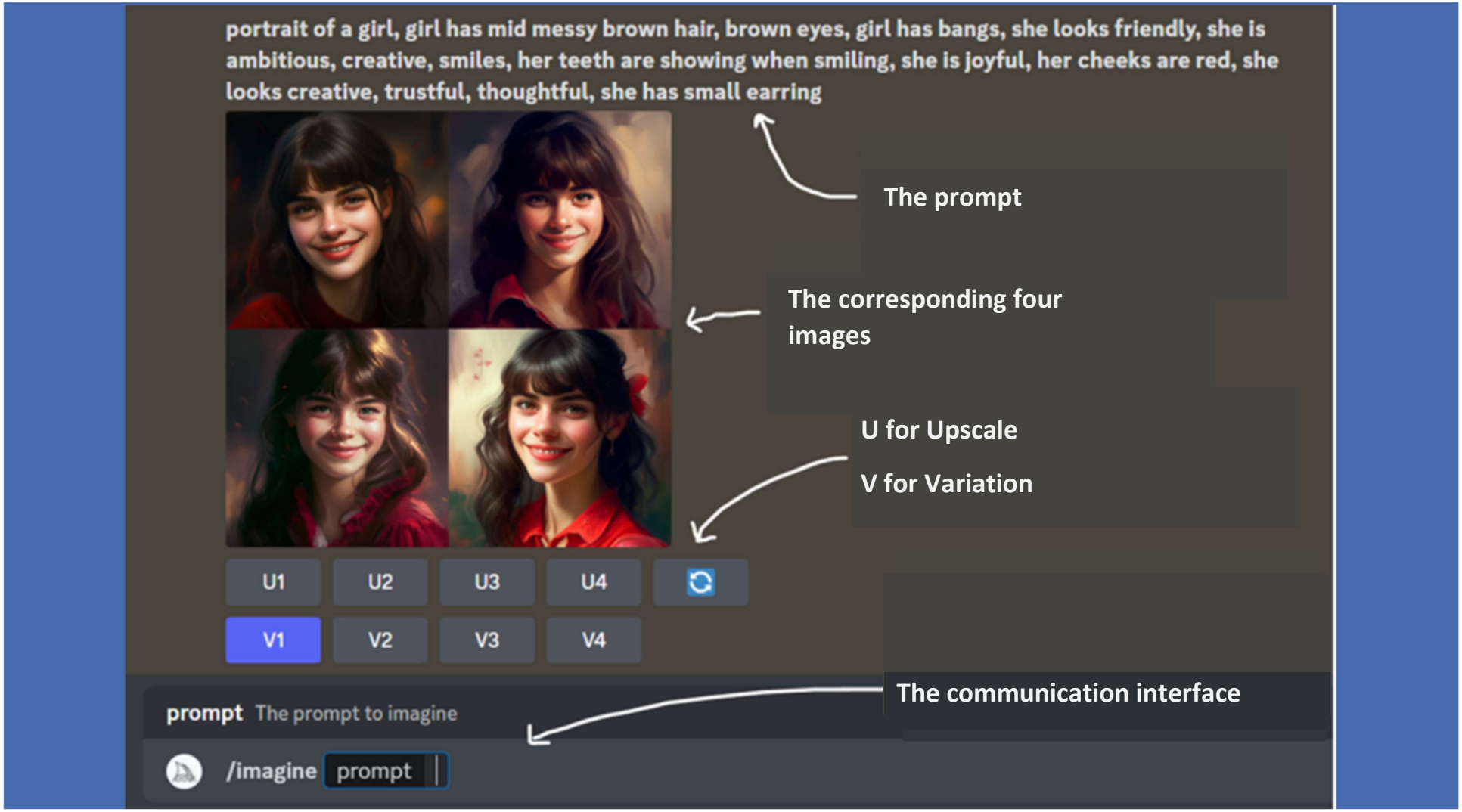
The image creation process

- **Rosenberg Self-Esteem Scale**: Horváth, Zs., Urbán, R., Kökönyei, Gy., Demetrovics, Zs. (2022). *Kérdőíves módszerek a klinikai és egészségpszichológiai kutatásban és gyakorlatban I.* Budapest: Medicina könyvkiadó.
- **Beck Depression Inventory**: Rózsa, S., Szádóczky, E., & Füredi, J. (2001). A Beck depresszió kérdőív rövidített változatának jellemzői hazai mintán. *Psychiatria Hungarica*.
- **Identity Status/EOM-EIS II Questionnaire**: Jámbori, Sz., Kőrössy, J. (2019). A szándékos önszabályozás jelentősége serdülő és fiatal felnőttkorban a társas támogatás, az identitásállapotok és a reziliencia tükrében. *Alkalmazott Pszichológia 19*(3): 33-52.
- **Self-Concept Clarity Scale:** Hargitai, R., Rózsa, S., Hupuczi, E., Birkás, B., Hartung, I., Hartungné Somlai, E., ... & Kállai, J. (2021). Énkép egyértelműség mérése és korrelátumai. *Magyar Pszichológiai Szemle*, *75*(4), 557-580.
- **Almost Perfect Scale (Perfectionism):** Horváth, Zs., Urbán, R., Kökönyei, Gy., Demetrovics, Zs. (2022). *Kérdőíves módszerek a klinikai és egészségpszichológiai kutatásban és gyakorlatban I.* Budapest: Medicina könyvkiadó.
- **Ten-Item Personality Inventory (Big5 – Shortened Version):** Chiorri, C., Bracco, F., Piccinno, T., Modafferi, C., & Battini, V. (2015). Psychometric properties of a revised version of the Ten Item Personality Inventory. *European Journal of Psychological Assessment*.
- **Body Attitudes Questionnaire:** Horváth, Zs., Urbán, R., Kökönyei, Gy., Demetrovics, Zs. (2022). *Kérdőíves módszerek a klinikai és egészségpszichológiai kutatásban és gyakorlatban I.* Budapest: Medicina könyvkiadó.

1. Supplementary Material: Elements of the 25-minute questionnaire package completed in advance by the participants.

I would like to ask you to try to create images of yourself that express who you are both inside and outside, and how you feel about yourself. These images do not have to be realistic, but they can be if you wish. You can start from several basic ideas, the images can be embedded in scenes or situations, and they can deviate from reality as much as you like. During the image creation process, you can specify artistic styles, play with the format, composition, lighting, and colors. The goal is to create as many images as possible that you feel capture something of your personality.

2. Supplementary Material: The standardized instructions read before image creation. I considered it important to mention in the instructions how broad the horizon of possibilities with stable diffusion software is, so that participants would not hold back simply because they were unaware that they could, for example, play with style and setting (Kellerwessel & Ujhelyi, 2024)."



3. Supplementary Material: The user interface of the Midjourney application, showing four images from one of the participants along with a prompt sequence. (Kellerwessel & Ujhelyi, 2024)

4. Supplementary Material: Based on the provided prompt sequence ("girl in dark blue coat standing at the Big Ben, with other people around, listening to music"), the software initially generates four images. The participant can select their preferred images and request further variations (using the V command for variations) or a higher resolution, more detailed version of any selected image (using the U command for Upscale) based on the image numbers. In this instance, the participant requested further variations of the second image from the initial set and then requested an upscaled version of the first image from the subsequent variations (Kellerwessel & Ujhelyi, 2024)

How are you feeling now?

What was it like to complete the task?

What goals did you set for yourself?

Did you have any strategy?

How do you feel about the images you created?

Which ones do you feel are the most expressive of yourself?

Which ones are the least?

Which image would your best friend consider to be the most expressive of you?

How does it feel to look through the images now?

What would you change if you could start over?

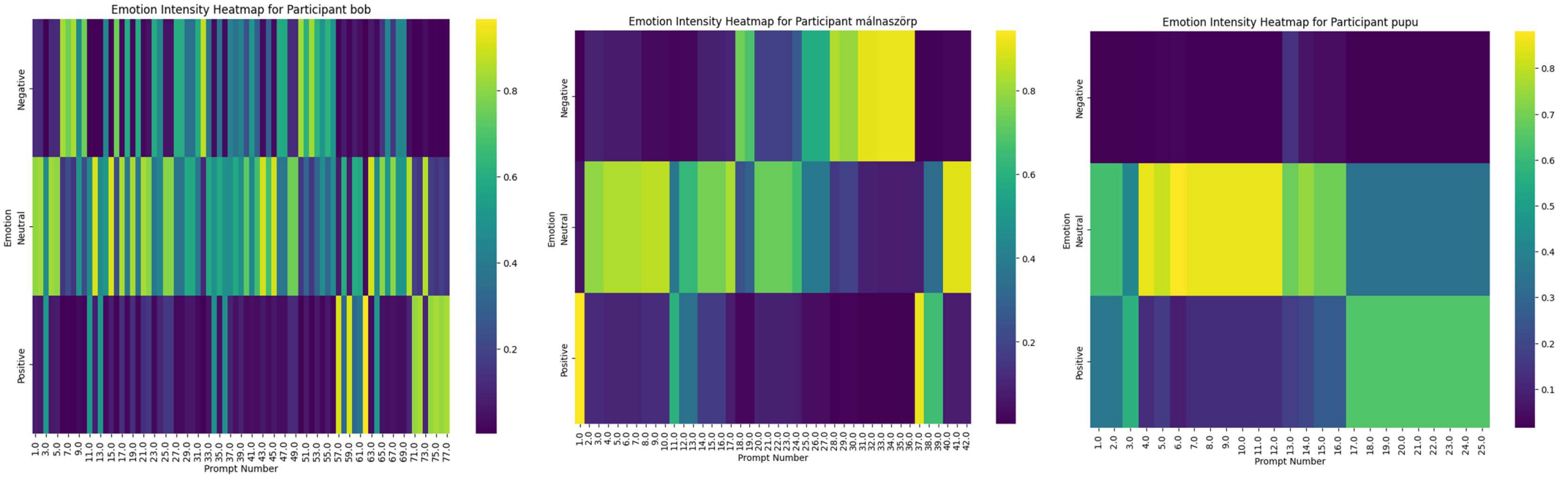
5. Supplementary Material: Guiding questions for the semi-structured interview (Kellerwessel & Ujhelyi, 2024)

Prompts with the strongest positive sentiment	Pseudonym	Image Number	Positive sentiment
Dance class with good music and colourful people, having fun	Apple Pie	7	0.984516
a happy family with big smile and happiness. The mother and father are in love	Oyster Mushroom	30	0.984236
an amazing friend group having fun	Monstera	23	0.984032

Prompts with the strongest neutral sentiment	Pseudonym	Image Number	Neutral sentiment
the roma caravan tries to make deals with the suitmaker next to the campfire.	Ahasverus	33	0.947187
a boy, sitting in the corner of the room, while everyone is talking to eachother in the room	Brainbig	10	0.939233
university students studying	Chimney Cake	79	0.933287

Prompts with the strongest negative sentiment	Pseudonym	Image Number	Negative sentiment
i am not enough, never enough, i do not have the right to live	Mekk Master	85	0.936157
I have no motivation and I am too tired to do anything. My head hurts all the time	Window	6	0.932891
i dont want to live in this world, i dont want to give birth to a child who is gonna die anyway because of the global warming.	Chucken	37	0.930059

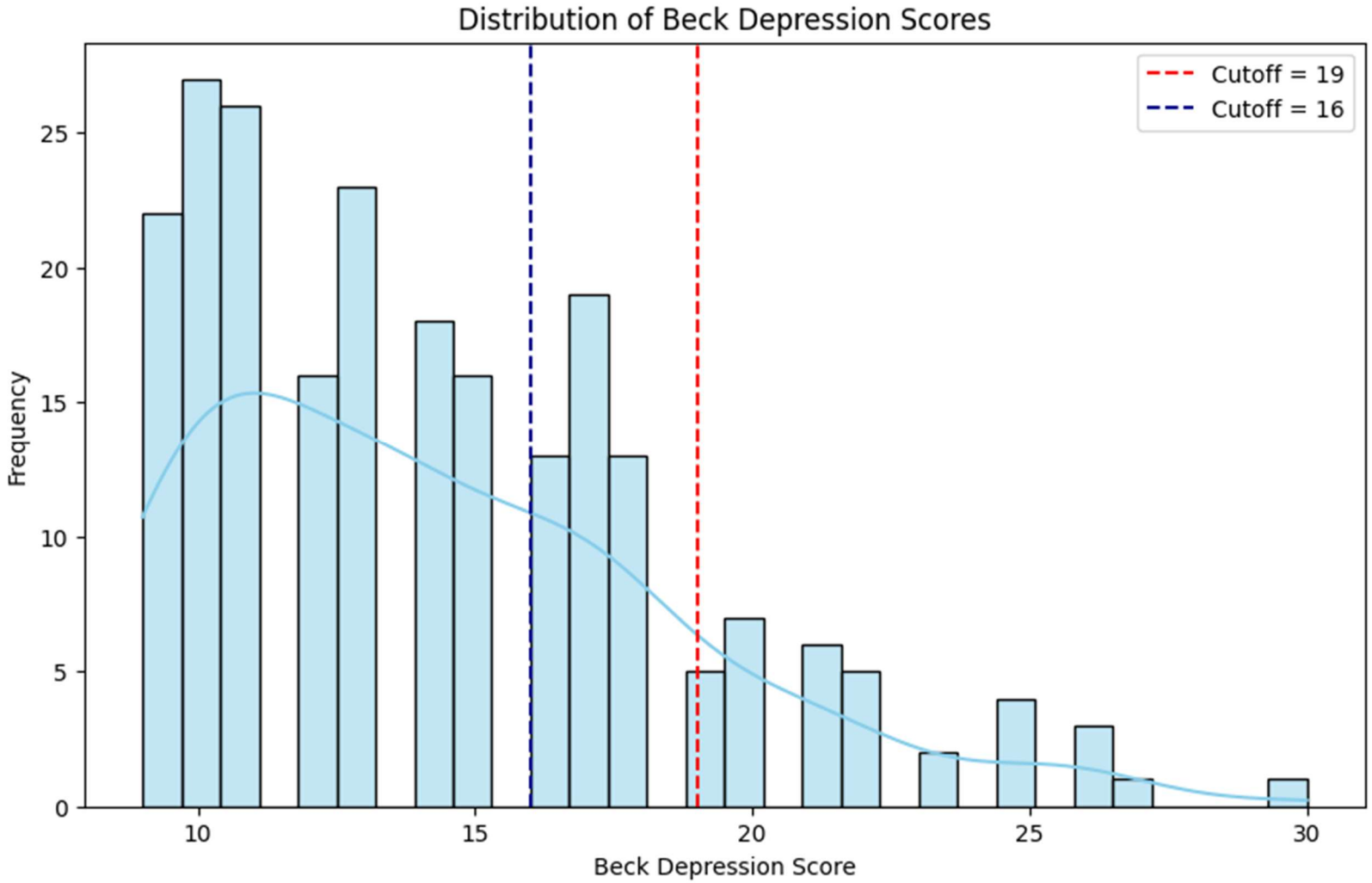
6. Supplementary Material*:* The model assigns three values ranging from 0 to 1 to each prompt based on the three examined sentiments (positive, neutral, negative). To illustrate its functioning, the prompts that received the highest scores in each dimension from the current sample are shown here.



7. Supplementary Material: Sentiment changes during the image creation process for three participants (Bob, Raspberry Syrup – málnaszörp in Hungarian, Pupu). The X-axis represents the sequence of completed images (thus, a temporal variable), while the Y-axis displays three bands corresponding to the three different sentiments (negative, neutral, positive). The higher the emotional sentiment of a given prompt, the more yellow the corresponding cell becomes—if low, it appears purple, and if it is around the median value, it takes on a greenish hue. Such emotion maps can help uncover the temporal patterns of emotions experienced by the participant during the creative process.

- pandas: Reback, J., McKinney, W. et al., (2024). *pandas* (Version 1.5.0) [Software]. Available at https://pandas.pydata.org
- numpy: Harris, C. R., Millman, K. J., van der Walt, S. J., et al. (2024). *NumPy* (Version 1.26.0) [Software]. Available at https://numpy.org
- transformers (Hugging Face): Wolf, T., Debut, L., Sanh, V., et al. (2024). *Transformers: State-of-the-art Natural Language Processing* (Version 5.3.0) [Software]. Available at https://huggingface.co/transformers
- datasets (Hugging Face): Lhoest, Q., et al. (2024). *Datasets: A library for easily accessing NLP datasets* (Version 2.4.0) [Software]. Available at https://huggingface.co/docs/datasets
- tensorflow: Abadi, M., Agarwal, A., Barham, P., et al. (2024). *TensorFlow: Large-scale machine learning on heterogeneous systems* (Version 2.12.0) [Software]. Available at https://tensorflow.org
- torch: Collobert, R., Kavukcuoglu, K., Farabet, C., (2024). *Torch* (Version 1.15.0) [Software]. Available at http://torch.ch
- scipy.stats: Part of the SciPy package: Virtanen, P., Gommers, R., Oliphant, T. E., et al. (2024). *SciPy* (Version 1.10.0) [Software]. Available at https://scipy.org
- seaborn: Waskom, M. L. (2024). *Seaborn: Statistical Data Visualization* (Version 0.13.0) [Software]. Available at https://seaborn.pydata.org
- matplotlib: Hunter, J. D., Dale, D., Firing, E., et al. (2024). *Matplotlib* (Version 3.7.0) [Software]. Available at https://matplotlib.org

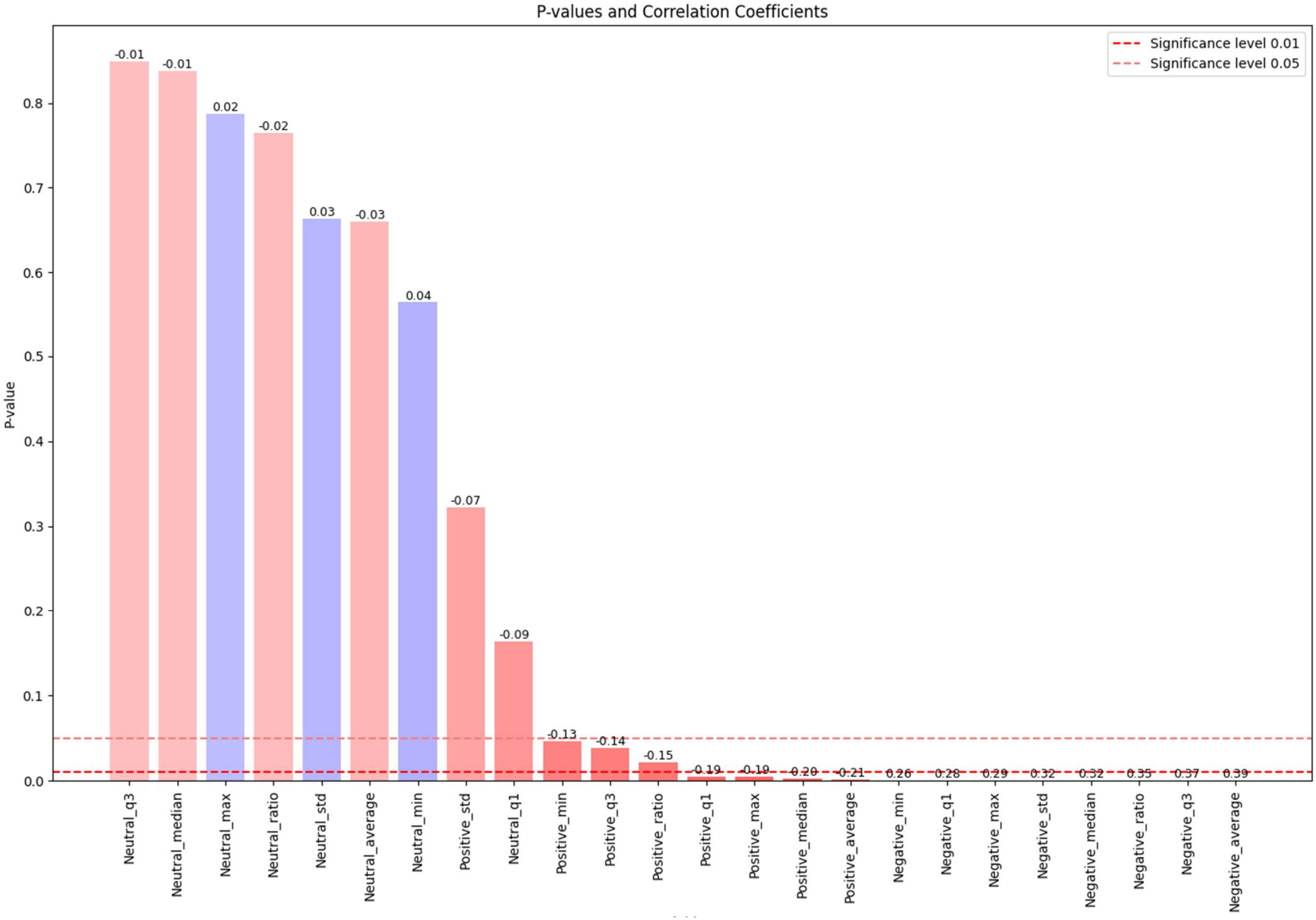
8. Supplementary Material: Python libraries used for data processing, analysis, and creating visualizations.



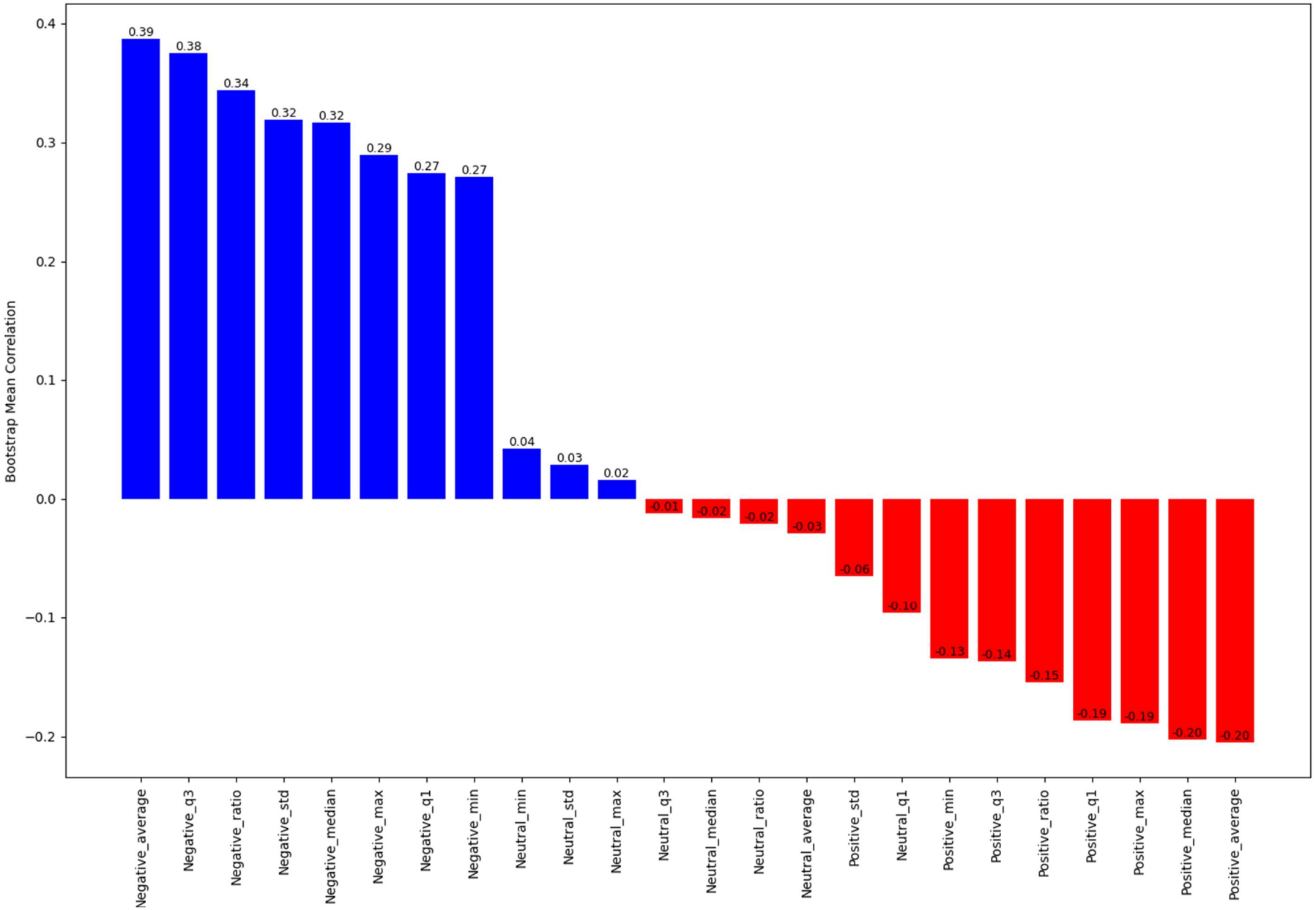
9. Supplementary Material: Distribution of the Hungarian-adapted short version of the Beck Depression Inventory (BDI) scores in the non-clinical sample. The dashed lines represent the standard cutoff of 16 (red) and the stricter cutoff of 19 (blue), illustrating how applying the higher threshold reduces the number of participants classified as high-risk, potentially decreasing false positives but simultaneously increasing class imbalance.

Independent Variable	Correlation Coefficient	p-value (correlation)	Mean bootstrap corr	lower bound	upper bound
Negative_average	**0,388**	0,0000	0,388	0,241	0,525
Negative_std	**0,317**	0,0000	0,312	0,167	0,453
Negative_min	**0,264**	0,0001	0,269	0,170	0,391
Negative_q1	**0,276**	0,0000	0,273	0,126	0,407
Negative_median	**0,322**	0,0000	0,320	0,163	0,463
Negative_q3	**0,372**	0,0000	0,369	0,210	0,510
Negative_max	**0,286**	0,0000	0,286	0,158	0,405
Neutral_average	-0,029	0,6603	-0,027	-0,144	0,087
Neutral_std	0,029	0,6631	0,029	-0,080	0,134
Neutral_min	0,038	0,5642	0,039	-0,073	0,155
Neutral_q1	-0,093	0,1634	-0,092	-0,206	0,026
Neutral_median	-0,014	0,8379	-0,016	-0,149	0,106
Neutral_q3	-0,013	0,8490	-0,012	-0,147	0,115
Neutral_max	0,018	0,7869	0,015	-0,088	0,110
Positive_average	**-0,207**	0,0018	-0,204	-0,312	-0,088
Positive_std	-0,066	0,3218	-0,067	-0,182	0,054
Positive_min	-0,132	0,0467	-0,133	-0,210	-0,063
Positive_q1	**-0,185**	0,0051	-0,187	-0,266	-0,101
Positive_median	**-0,200**	0,0025	-0,196	-0,303	-0,090
Positive_q3	-0,138	0,0379	-0,140	-0,267	-0,006
Positive_max	**-0,189**	0,0043	-0,191	-0,303	-0,063
Negative_ratio	**0,348**	0,0000	0,340	0,187	0,497
Neutral_ratio	-0,020	0,7649	-0,020	-0,143	0,101
Positive_ratio	-0,152	0,0219	-0,152	-0,257	-0,052

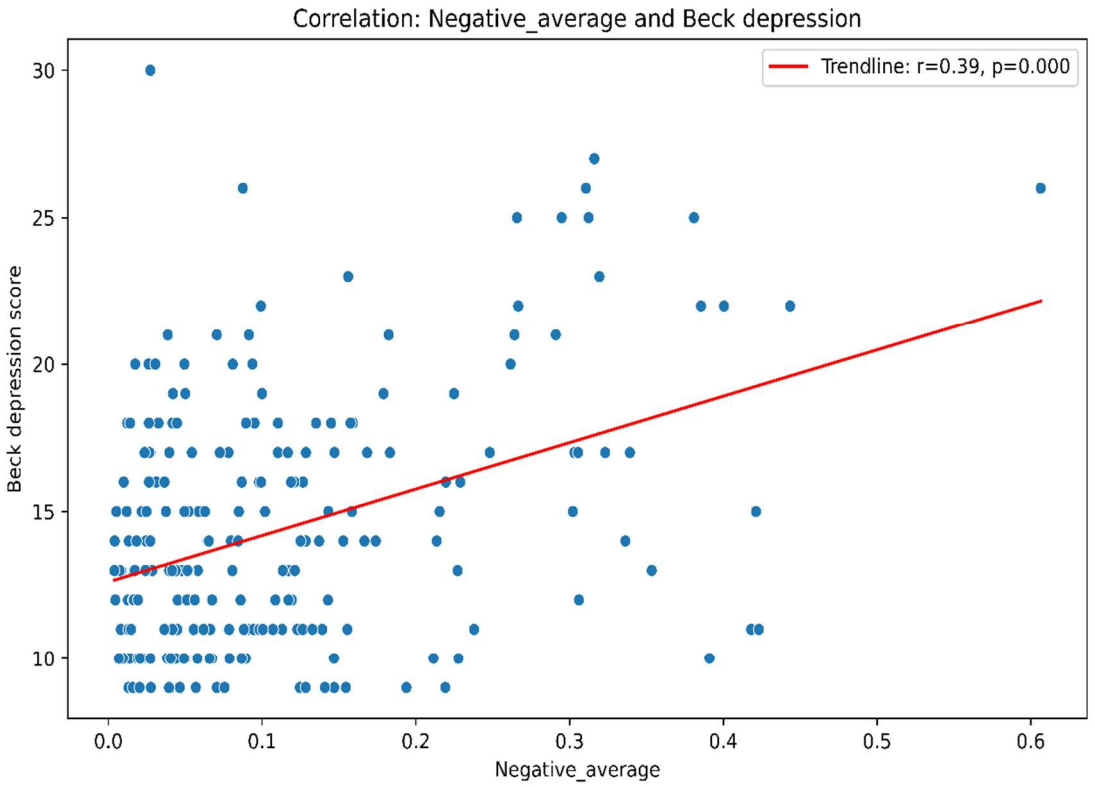
10. Supplementary Material: Results of the correlation analyses presented in table format. P-values below 0.01 are highlighted in gray.

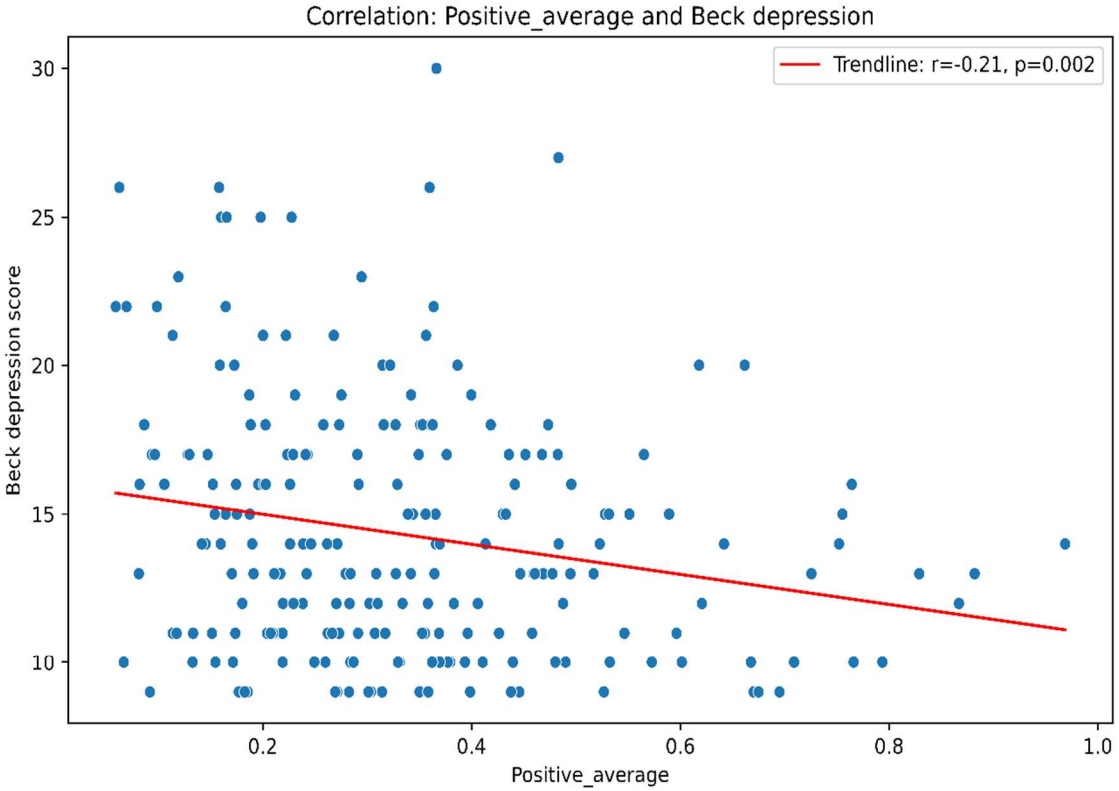


11. Supplementary Material: P-values from the correlation analyses. The x-axis displays the variables under investigation, while the height of the corresponding columns indicates the p-value. The lighter dashed line represents the 0.05 threshold, and the stronger red line marks the 0.01 alpha value used in this thesis. The color of the variable columns corresponds to the direction of the correlation (red for negative, blue for positive), and the label above each column indicates the strength of the correlation.



12. Supplementary Material: Average effect sizes of the bootstrap correlations. I conducted a 15,000-sample bootstrap correlation for each variable to enhance the robustness of the associations: the height of each column represents the average effect size for the respective variables, while the color and direction of the columns indicate whether the correlation was positive or negative.





13. Supplementary Material: Correlation of Negative and Positive Sentiment Averages with Beck Depression Scores. These two scatterplots were selected to visually highlight the core finding of the sentiment analysis: higher negative sentiment is associated with elevated depression scores, whereas higher positive sentiment correlates with lower scores.

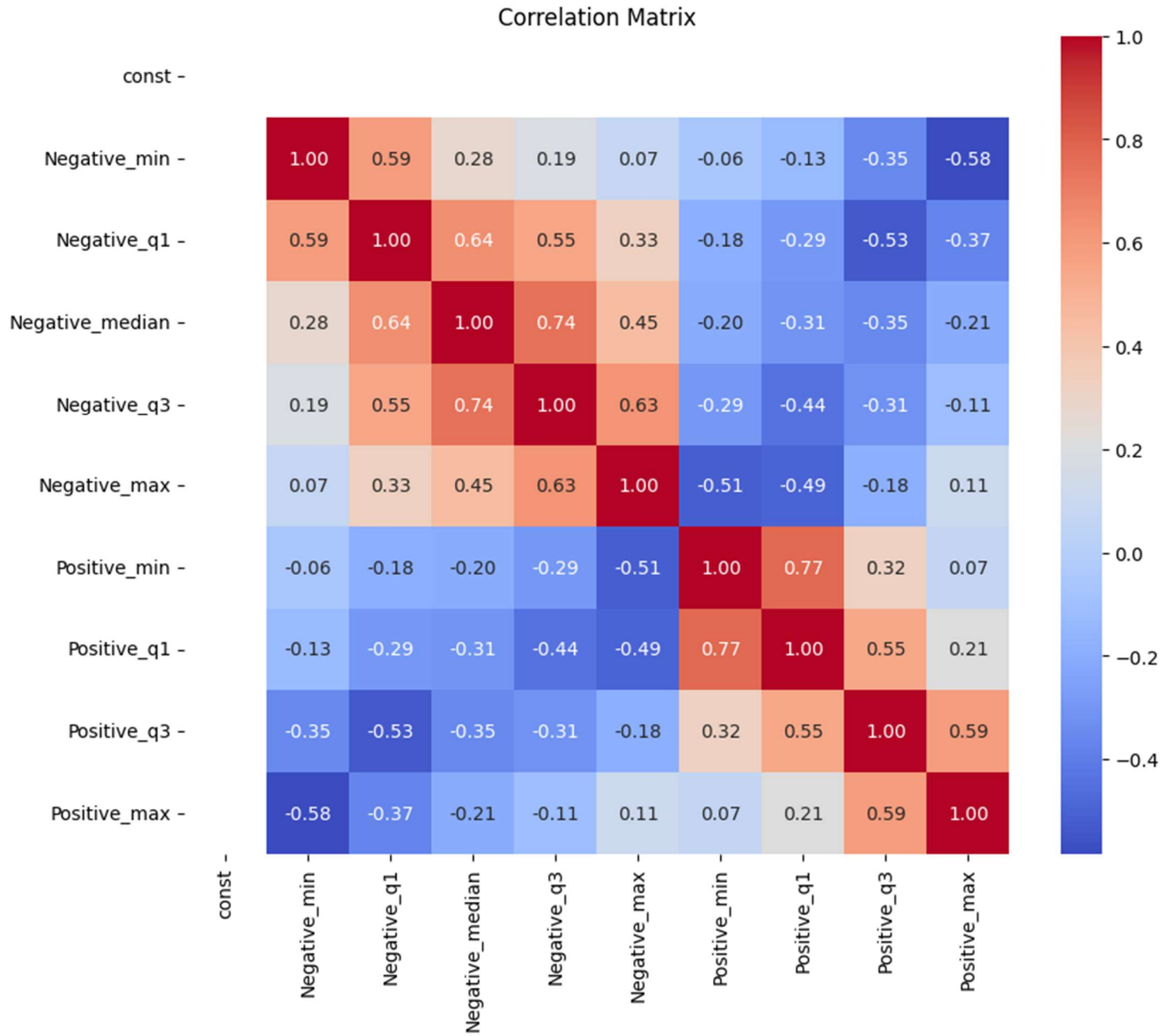
Potential predictor variables

feature	VIF
const	33.181
Negative_average	117.171
Negative_std	60.100
Negative_min	**2.296**
Negative_q1	**5.873**
Negative_median	**9.823**
Negative_q3	**16.151**
Negative_max	**17.547**
Negative_ratio	12.781
Positive_average	101.189
Positive_min	**3.785**
Positive_q1	**11.515**
Positive_median	15.520
Positive_q3	**17.796**
Positive_max	**4.895**

Predictor variables selected for the analyses

feature	VIF
const	24.270
Negative_min	2.200
Negative_q1	3.038
Negative_median	2.727
Negative_q3	3.202
Negative_max	2.234
Positive_min	2.838
Positive_q1	3.690
Positive_q3	2.646
Positive_max	2.337

14. Supplementary Material: These tables illustrate the results of the Variance Inflation Factor (VIF) analysis conducted to diagnose multicollinearity among the initially significant or trend-level predictor variables (left panel) and the final subset of predictors retained for the linear models (right panel). As shown, many of the initial variables exceeded the commonly accepted threshold of VIF > 5, indicating substantial collinearity. To address this issue, I employed an iterative variable selection process: variables with the highest VIF values were removed step by step until all remaining predictors fell below the VIF cutoff.



15. Supplementary Material: Correlation Matrix of the Predictor Variables Selected. The heatmap reveals intriguing relationships among the sentiment features selected for regression and classification analyses. Despite some moderately strong associations, none of the variables exceed the

commonly problematic threshold of |r| = 0.8. This indicates that, following VIF-based selection, the chosen sentiment features remain sufficiently distinct and are suitable for further analyses.

While internal consistency was expected among negative and positive sentiment measures, several unexpected findings emerge: Negative_min and Negative_max show virtually no correlation ($r = 0.07$, $p=0.280$), suggesting that individuals who occasionally use highly negative language (max) do not necessarily maintain a consistently negative emotional tone throughout their prompts (min) – as is the case with Positive_min and Positive_max ($r = 0.07$, $p=0.316$). On the other hand, Negative_min and Positive_max exhibit a moderate negative correlation ($r = -0.58$, $p<0.0001$), indicating that participants who have higher minimum negativity scores don't express extreme high levels of positivity. Similar cross-sentiment patterns emerge between other distribution measures as well: Negative_q3 negatively correlates with Positive_q1 ($r= -0,44$, $p<0.0001$), and the relationships appear symmetrical between negative and positive sentiment distributions (Positive_min and Negative_max, $r = -0.51$, $p<0.0001$; Positive_q3 and Negative_q1, $r = -0.53$, $p<0.0001$).

Collectively, these findings highlight that sentiment distributions convey nuanced emotional patterns that simple averages or extreme values alone cannot fully capture. Such relationships emphasize the importance of analyzing the full distribution of sentiment features, rather than relying solely on single metrics, to accurately reflect the complex emotional dynamics underlying the creative process.

Linear regression (OLS)

Beck depression

R^2	0.198	Durbin-Watson	1.791	
R^2 (adjusted)	**0.164**	Omnibus	19.456	
F-stat	5.935	Jarque-Bera	22.340	
p-value (F-stat)	< 0.001	AIC	1278	
Log-Likelihood	-628.80	BIC	1312	

Variable	**Coeff**	**Std err**	**T-value**	**P-value**	**95% conf lower**
Const	13.259	1.291	10.266	0.000	10.713
Negative_min	66.674	36.267	1.838	0.067	-4.806
Negative_q1	-5.9157	17.086	-0.346	0.729	-39.591
Negative_median	2.8067	4.025	0.697	0.486	-5.127
Negative_q3	5.0299	2.342	2.148	0.033	0.414
Negative_max	1.9628	1.259	1.559	0.121	-0.519
Positive_min	0.6982	3.398	0.205	0.837	-5.999
Positive_q1	-0.7899	2.905	-0.272	0.786	-6.516
Positive_q3	1.9744	1.616	1.222	0.223	-1.211
Positive_max	-2.6578	1.694	-1.569	0.118	-5.996

16. Supplementary Material: A statistically significant model ($F(9, 217) = 5.935$, $p < .001$) was fitted to the data, with an adjusted R^2 of 0.164. Although the Durbin–Watson test (1.791) suggested independent residuals, the Omnibus (19.456, $p < .001$) and Jarque–Bera (22.340, $p < .001$) tests pointed to deviations from normality, likely reflecting non-normal distributions of one or more predictors. However, the Breusch–Pagan test (16.868, $p = 0.051$) showed no significant heteroscedasticity, supporting the homoscedasticity assumption of the OLS model.

GLM

Beck depression

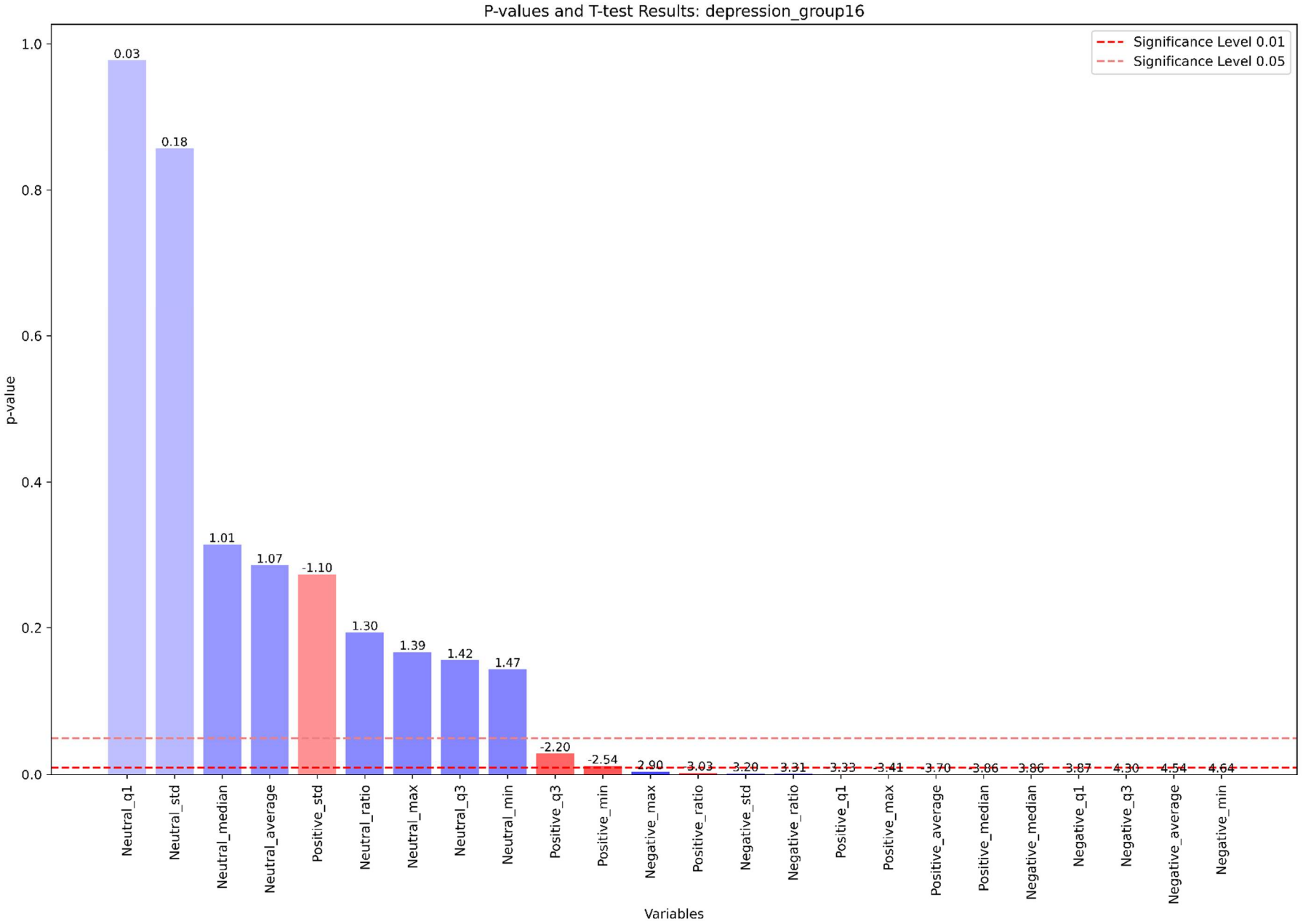
	Number of observations	230	Log-Likelihood	-628.80
	Model family	Gaussian	Deviance	3385.3
	Link function	Identity	Pearson chi2	3390
	Scale	15.600	**Pszeudo-R^2 (CS)**	**0.2105**

	Coeff	SD	Z-value	p-value	95% conf interval
Constant	14,335	0,256	55,927	0,000	13,832
Negative_min	0,715	0,337	2,120	0,034	0,054
Negative_q1	-0,158	0,452	-0,350	0,727	-1,045
Negative_median	0,302	0,574	0,526	0,599	-0,822
Negative_q3	1,008	0,621	1,622	0,105	-0,210
Negative_max	0,611	0,386	1,583	0,113	-0,145
Positive_min	0,091	0,264	0,344	0,731	-0,426
Positive_q1	-0,137	0,409	-0,335	0,738	-0,939
Positive_q3	0,521	0,467	1,114	0,265	-0,395
Positive_max	-0,629	0,358	-1,756	0,079	-1,331

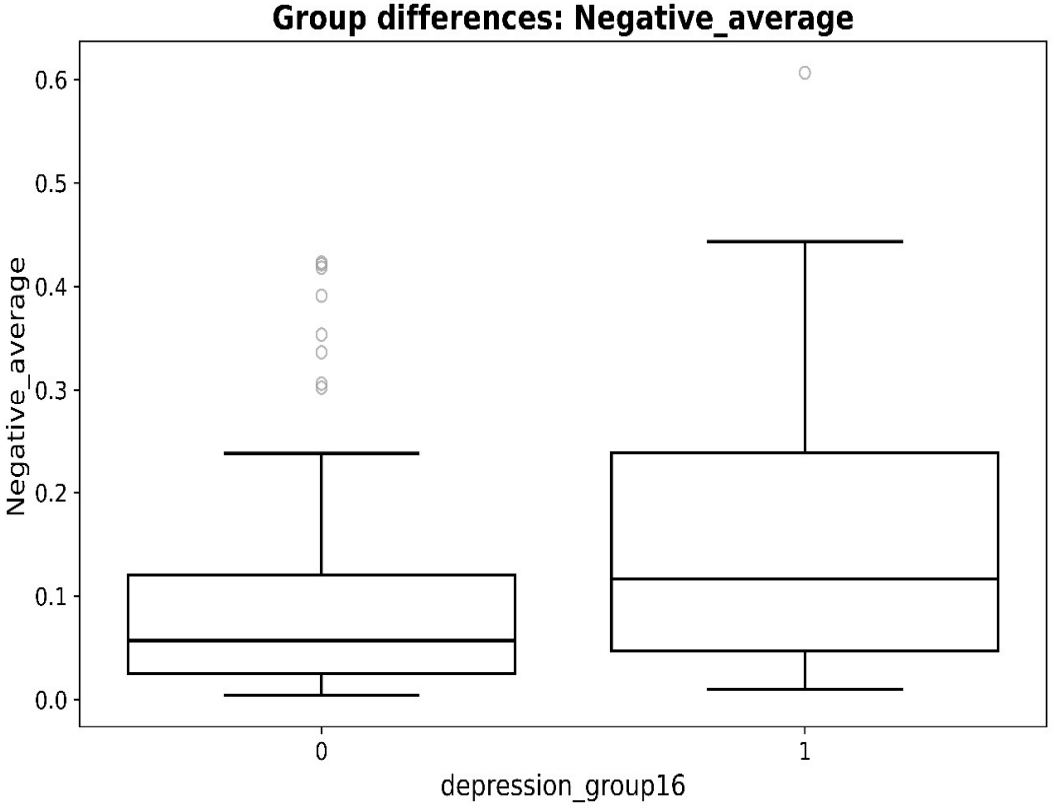
17. Supplementary Material: Results of the generalized linear model (GLM). Similar to the ordinary least squares (OLS) regression, this model was also statistically significant, with a pseudo-R^2 value of 0.21, indicating that the included predictors explained 21% of the variance in depression scores.

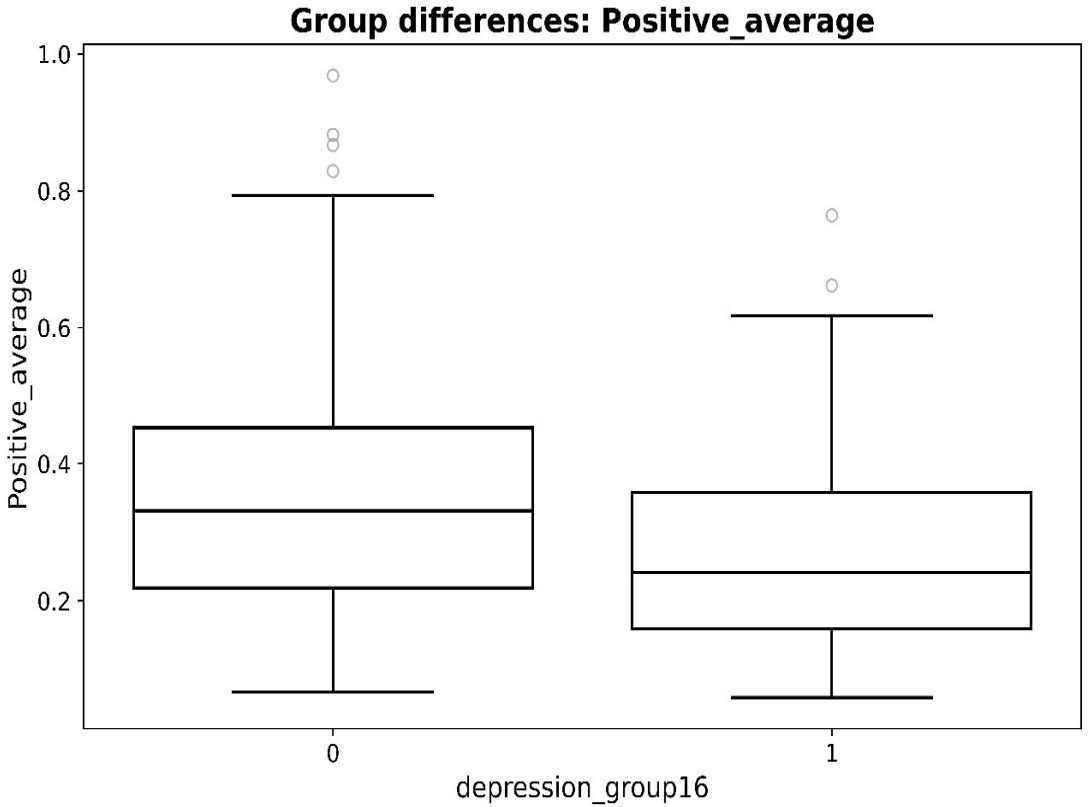
Independent_Variable	t_statistics_16	p-value_16	t_statistics_19	p-value_19
Negative_average	**4,543**	0,000	**5,683**	0,000
Negative_std	**3,196**	0,002	**4,109**	0,000
Negative_min	**4,636**	0,000	2,478	0,014
Negative_q1	**3,875**	0,000	**3,606**	0,000
Negative_median	**3,863**	0,000	**5,376**	0,000
Negative_q3	**4,296**	0,000	**5,842**	0,000
Negative_max	**2,899**	0,004	**3,774**	0,000
Neutral_average	1,069	0,286	-0,597	0,551
Neutral_std	0,180	0,858	1,348	0,179
Neutral_min	1,467	0,144	-1,031	0,304
Neutral_q1	0,028	0,978	-1,387	0,167
Neutral_median	1,008	0,315	-0,296	0,767
Neutral_q3	1,424	0,156	-0,362	0,718
Neutral_max	1,387	0,167	0,701	0,484
Positive_average	**-3,698**	0,000	**-2,704**	0,007
Positive_std	-1,097	0,274	-0,528	0,598
Positive_min	-2,545	0,012	-1,876	0,062
Positive_q1	**-3,332**	0,001	-2,537	0,012
Positive_median	**-3,855**	0,000	-2,424	0,016
Positive_q3	-2,200	0,029	-1,758	0,080
Positive_max	**-3,406**	0,001	-1,746	0,082
Negative_ratio	**3,309**	0,001	**4,921**	0,000
Neutral_ratio	1,304	0,194	-0,208	0,835
Positive_ratio	-3,030	0,003	-2,134	0,034

18. Supplementary Material: T-test results comparing the two groups defined by each depression criterion (≥16 and ≥19). The table displays the t-statistics and p-values for each independent variable, with statistically significant group differences highlighted in gray. At first glance, it may seem surprising that more statistically significant differences emerged using the lower cutoff compared to the higher cutoff. However, when applying the stricter criterion, less than half as many participants were classified as depressed, which likely reduced the power of the t-test- This could lead to a higher chance of Type II errors, where potentially meaningful effects may only appear as trends rather than reaching statistical significance.

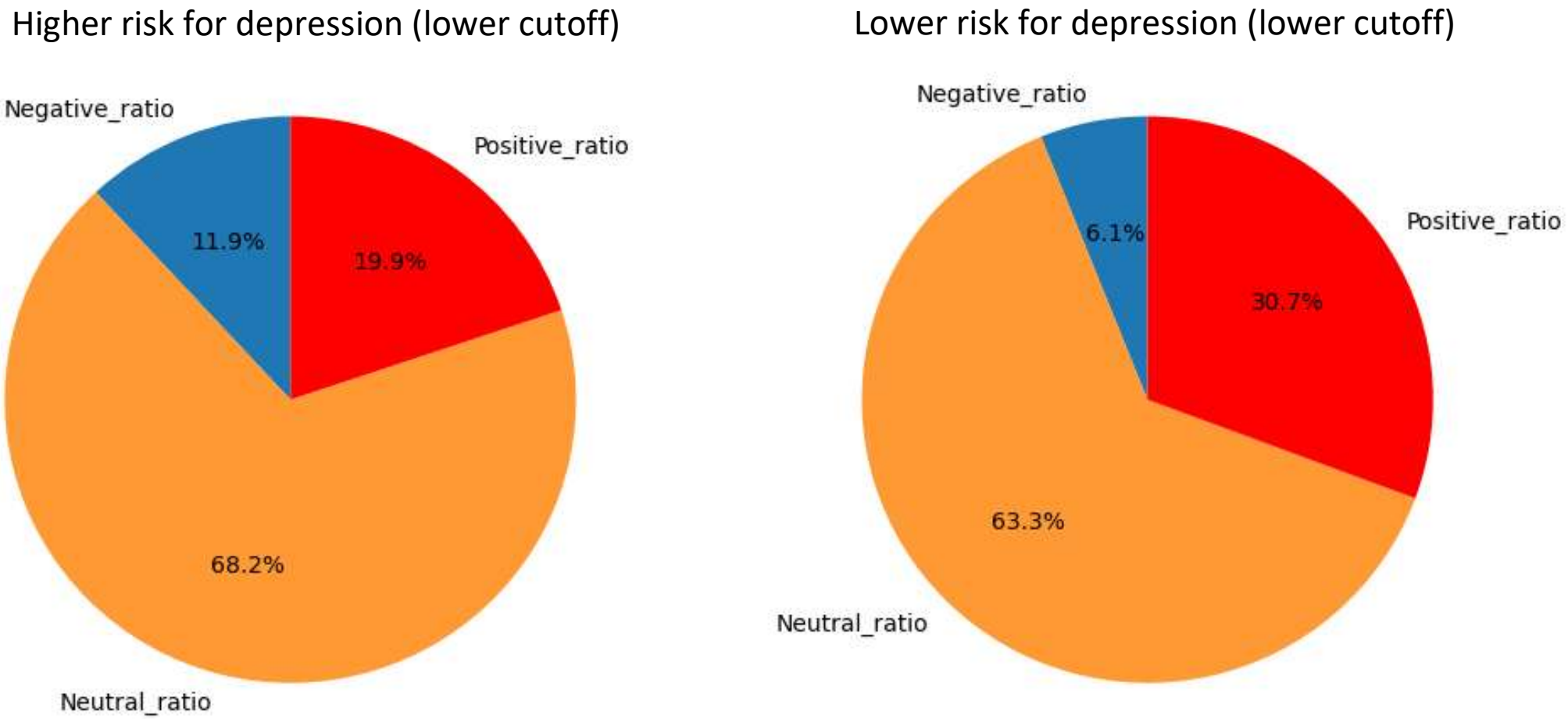


19. Supplementary Material: p-values from independent samples t-tests (lower cutoff). The x-axis represents the variables examined, with the height of the corresponding columns indicating the p-value. The lighter dashed line marks the 0.05 level, and the stronger red line marks the 0.01 p-value level, designated as the alpha threshold in this paper. The color of the variable columns corresponds to the direction of the group difference (red - negative, blue - positive), and the label above the column indicates the t-value of the difference.

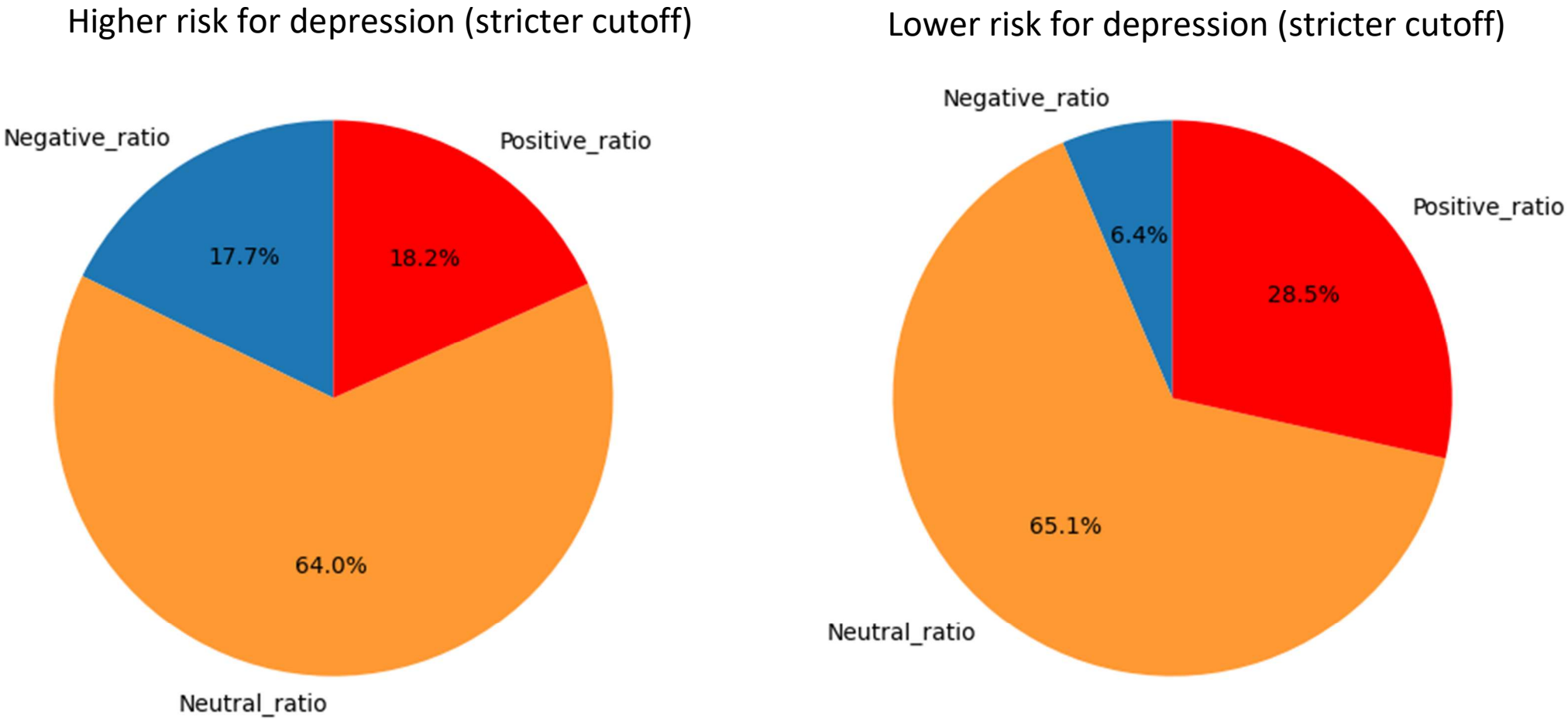




20. Supplementary Material: Intergroup differences in negative and positive mean (lower cutoff).



21. Supplementary Material**:** The different distributions of the dominant emotional tone of the images in the groups with high and low risk for depression (lower cutoff). Although the order of dominant emotional tones is the same for both groups, the average proportion of positive and negative prompts is closer in the high-risk group. However, the high variance within groups indicates significant individual differences.

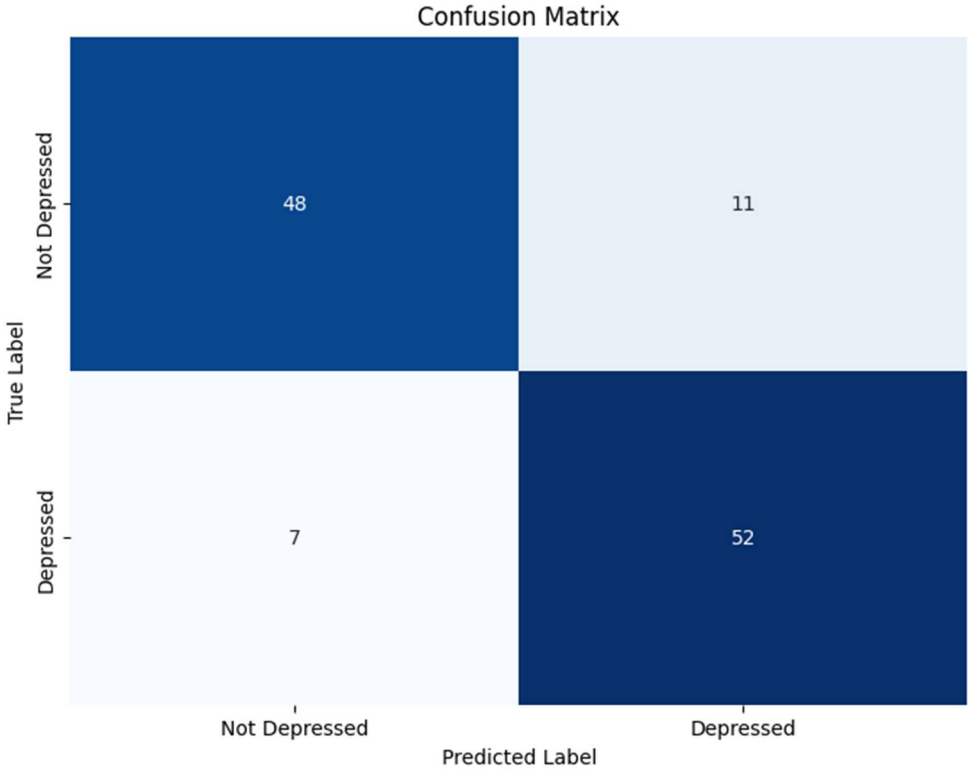


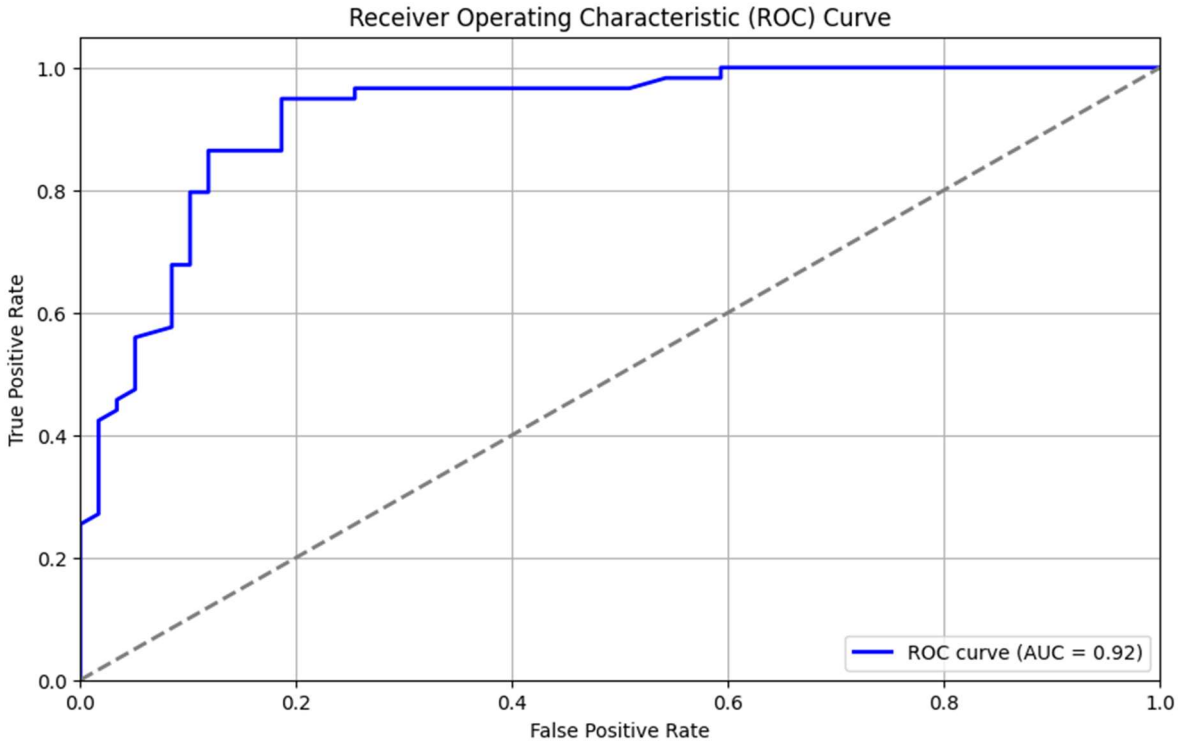
22. Supplementary Material: Distribution of negative, neutral, and positive prompts among participants classified as higher risk for depression (left) versus lower risk (right) under the stricter cutoff (≥19). The high-risk group displays a substantially higher proportion of negative prompts (17.7% vs. 6.4%) and a lower proportion of positive prompts (18.2% vs. 28.5%), while neutral prompts remain the largest share in both groups.

ML-classifier metrics (lower cutoff)

Model	AUC mean	AUC std	Accuracy mean	Accuracy std	Precision mean	Precision std	Recall mean	Recall std	F1 mean	F1 std
Logistic Regression	**0,703**	0,097	**0,646**	0,088	**0,644**	0,087	**0,663**	0,099	**0,652**	0,086
Decision Tree	**0,666**	0,076	**0,665**	0,076	**0,653**	0,078	**0,722**	0,125	**0,680**	0,083
KNN	**0,719**	0,104	**0,663**	0,092	**0,632**	0,083	**0,796**	0,122	**0,701**	0,086
Random Forest	**0,826**	0,103	**0,769**	0,099	**0,758**	0,097	**0,803**	0,103	**0,777**	0,091
SVM	**0,675**	0,087	**0,583**	0,063	**0,586**	0,063	**0,590**	0,128	**0,581**	0,078
XGBoost	**0,786**	0,109	**0,722**	0,100	**0,707**	0,084	**0,762**	0,151	**0,727**	0,112
AdaBoost	**0,749**	0,090	**0,682**	0,061	**0,693**	0,080	**0,675**	0,088	**0,679**	0,061

23. Supplementary Material**:** Performance metrics of other machine-learning classifiers (lower cutoff). The **AUC score** indicates how well the model can distinguish between the two groups compared to random classification, while the **accuracy** indicates how many of the classifications were correct (high-risk as high-risk, low-risk as low-risk). Beyond this, **precision** shows how many of the samples identified as high-risk by the model were indeed high-risk, while **sensitivity** shows how many of the actual high-risk samples were correctly identified by the model. The **F1 score**, which is the harmonic mean of precision and sensitivity, represents the general measure of the model's performance.
    In this study, 10-fold cross-validation was used to evaluate the models, where the dataset was divided into ten equal parts. Each model was trained on nine subsets and tested on the remaining one, and this process was repeated for all ten folds. This approach reduced the risk of bias from any specific data split, improving the reliability and generalizability of performance estimates. The standard deviation (SD) reflects the variability in model performance across cross-validation folds, indicating the consistency and robustness of the results.
    Several different ML classification models were tested, and overall, tree-based models outperformed linear and kernel-based approaches. The **Random Forest** classifier achieved the **highest performance** with an average accuracy of 76.9% (SD = 9.9) and an AUC of 0.83 (SD = 0.10), effectively identifying 80.3% (SD = 10.3) of high-risk participants. In contrast, the **logistic regression** model exhibited moderate accuracy (64.6%) and a lower AUC (0.70), suggesting that depression risk classification requires non-linear decision boundaries. SVM performed the worst, with an accuracy of only 58.3% and an AUC of 0.67, indicating that it struggled to effectively distinguish between the two groups. These results highlight the advantage of ensemble methods and decision tree-based approaches in capturing the complex relationships between prompt sentiment distributions and depression risk.





24. **Supplementary Material:** Confusion matrix (left) and ROC curve (right) from a single Random Forest run under the stricter cutoff (≥19 points). Compared to the less stringent criterion (≥16) of Figure 2 (Page 8 of the main text), this model exhibits substantially higher classification metrics, suggesting robust performance, as reflected in the elevated AUC. However, these results should be interpreted with caution, as they may not fully generalize beyond this specific dataset.

ML-classifier metrics (stricter cutoff)

Model	AUC mean	AUC std	Accuracy mean	Accuracy std	Precision mean	Precision std	Recall mean	Recall std	F1 mean	F1 std
Logistic Regression	**0,703**	0,063	**0,646**	0,064	**0,705**	0,092	**0,525**	0,128	**0,590**	0,093
Decision Tree	**0,816**	0,039	**0,816**	0,038	**0,792**	0,054	**0,867**	0,053	**0,826**	0,032
KNN	**0,859**	0,045	**0,781**	0,061	**0,717**	0,064	**0,943**	0,055	**0,813**	0,048
Random Forest	**0,952**	0,031	**0,883**	0,051	**0,861**	0,059	**0,918**	0,062	**0,887**	0,048
SVM	**0,799**	0,068	**0,684**	0,064	**0,730**	0,092	**0,597**	0,110	**0,650**	0,083
XGBoost	**0,934**	0,032	**0,875**	0,058	**0,855**	0,073	**0,914**	0,071	**0,881**	0,054
AdaBoost	**0,859**	0,042	**0,776**	0,047	**0,784**	0,064	**0,781**	0,118	**0,774**	0,055

25. Supplementary Material**:** As with the lower cutoff analyses, the models were evaluated using 10-fold cross-validation, ensuring more robust and generalizable performance estimates. However, under the stricter cutoff (≥19), the class imbalance is higher, necessitating more intensive oversampling – while this procedure helps mitigate bias toward the majority class, it can also introduce artificial patterns, and thus the results should be interpreted with caution.

    Despite these caveats, tree-based methods again outperformed linear and kernel-based approaches, underscoring the advantage of ensemble models. The Random Forest classifier achieved notably strong performance, with an average accuracy of 88.3% (SD = 5.1) and an AUC of 0.95 (SD = 0.03), indicating near-excellent discrimination between high-risk and low-risk participants. By contrast, logistic regression's AUC (0.70) and accuracy (65.4%) were markedly lower, suggesting that depression risk classification requires non-linear decision boundaries. These enhanced metrics compared to the lower cutoff reflect the more homogeneous nature of the high-risk group under the stricter threshold, but also highlight the challenges of working with a smaller, more imbalanced dataset.