Predicting evaluation of the day: the role of morning expectation and mood during the day

Grotto Tommaso

1. Introduction

Understanding how individuals evaluate their day is a fundamental question in the study of human behavior and well-being. This paper investigates the factors that shape the evening evaluation of a day, focusing on the interplay between morning expectations and mood status throughout the day. While prior research has extensively explored mood prediction and the influence of contextual factors on emotions, the evaluation of daily experiences, such as the subjective assessment of how the day went, remains an underexplored area. This study aims to fill this gap by examining the relative contributions of morning expectations and daily mood in predicting evening evaluations.

The advent of mobile sensing technologies has transformed the ability to collect and analyze behavioral data, enabling a deeper understanding of mood dynamics. Previous studies have demonstrated the effectiveness of mobile sensing in predicting mood states by integrating data on daily activities, social interactions, and environmental contexts. For instance, Meegahapola et al. (2023) highlighted the importance of personalized and context-specific models in mood inference, while LiKamWa et al. (2013) demonstrated the potential of app usage patterns for mood prediction. However, the evaluation of daily experiences, as opposed to momentary mood states, presents unique challenges that require a different methodological approach.

This study leverages data from the WeNet project, which collected longitudinal data from university students via the ILog app. The app provided a comprehensive dataset, including self-reported mood, daily activities, social contexts, and locations, alongside pre-study assessments of personality traits, human values and demographic informations on the subjects. By analyzing this data, we aim to answer the following research questions: how do morning expectations and mood during the day predict the evening evaluation of the day? Which of these factors is more influential, and how do their effects vary between workdays and holidays?

To address these questions, we employed a combination of regression and machine learning techniques, including Linear Regression, Random Forest Regressor, and Gradient Boosting Regressor. Each method was applied to different groupings of variables, encompassing personality traits, demographic characteristics, temporal patterns, mood metrics, and all variables combined. This approach allowed us to systematically evaluate the relative importance of various predictors and assess their predictive power under different contexts.

By integrating insights from prior research and leveraging advanced analytical techniques, this study contributes to a deeper understanding of the factors that influence daily evaluations, providing a foundation for future research in mood dynamics and daily experience evaluations.

2. Literature Review

The study of mood prediction and evaluation has gained significant attention in recent years, particularly with the advent of mobile sensing technologies. These technologies enable the unobtrusive collection of data on human behavior, emotions, and interactions, providing new opportunities for understanding the determinants of mood and well-being.

Several studies have explored the potential of mobile sensing to infer mood states. For example, Meegahapola et al. (2023) examined mood inference across diverse geographical regions, highlighting the challenges of generalization in mood prediction models. Their findings showed that country-specific models, especially when partially personalized, outperformed generic multi-country models. Similarly, LiKamWa et al. (2013) introduced MoodScope, a smartphone-based system that uses communication history and app usage to predict users' mood with increasing accuracy over time. These studies underscore the importance of tailoring mood prediction models to individual and contextual factors to enhance their reliability.

The role of daily activities and context in mood prediction has also been a focus of research. Assi et al. (2023) emphasized the need to move beyond simple activity recognition to detect complex daily activities that better reflect people's lives. Their findings demonstrated the effectiveness of multimodal sensors and machine learning in capturing nuanced behaviors, which are crucial for understanding mood dynamics. Similarly, Rachuri et al. (2010) developed EmotionSense, a platform that integrates emotion sensing with activity and interaction tracking, revealing the interplay between social context and emotional states.

Large-scale and longitudinal studies have provided valuable insights into mood dynamics over time. Servia-Rodríguez et al. (2017) conducted one of the largest in-the-wild studies of mood using smartphone data, showing strong correlations between routines and psychological variables. Their findings demonstrated that mobile sensing could predict mood with high accuracy, particularly during weekends. Wang et al. (2014) contributed to this research field by examining the impact of workload on students' mood and well-being, revealing significant patterns of stress and mood fluctuations throughout an academic term.

Mobile sensing has also been used to identify behavioral markers associated with mental health. Sano and Picard (2013) demonstrated that features such as screen usage, mobility, and activity levels were significantly correlated with stress. Similarly, Canzian and Musolesi (2015) showed that mobility patterns could predict depressive mood changes, highlighting the potential of passive sensing for mental health monitoring.

These studies collectively emphasize the importance of integrating diverse data sources, including personality traits, daily activities, and contextual variables, for mood prediction. While much of the existing work focuses on mood inference, the evaluation of daily experiences, such as the evening evaluation of the day, remains relatively underexplored. This study builds on the existing literature by investigating how morning expectations and mood during the day contribute to evening evaluations, leveraging both traditional regression techniques and advanced machine learning models.

3. Research Question and Data Description

The research question of this study is: how do morning expectations and mood during the day predict the evening evaluation of the day? Which of these factors is more influential, and how do their effects vary between workdays and holidays?

The data comes from the WeNet project, which used an app called ILog to collect daily data from 241 Unitn students for a month, starting the 14 of November 2020. The app collects the smartphone information and asks the user every half hour what they are doing, where they are, with whom, and their mood.

In this study, the data that is being used is the one from the first two weeks of the research because past this period the push notifications arrive every hour and not every 30 minutes. The subjects also filled out a pre-study questionnaire, where their demographic characteristics were collected and their personality traits and human values scores were assessed.

The ILog app also asks users how they expect the day to go (expectday) at 8 am and how the day was (howwasday) at 10 pm. We considered the subjects who responded to both the morning and evening questions at least 10 times out of the 16 total days. After cleaning the data the subjects remaining were 85 and there were 1162 records in total.

We used the standardized version of the ‘how was day’, ‘expect day’, and ‘mood’ for the data visualization and we used the numeric version for the regression and machine learning models.

The dependent variable is the evening day evaluation, a numeric variable ranging from 1 to 5. The independent variables encompass multiple dimensions, including personality traits, human values, demographic information, temporal patterns and mood metrics. These variables are:

Personality traits (BFI):
extraversion: tendency toward sociability, energy, and positive emotions; agreeableness: propensity for cooperation, trust, and altruism; conscientiousness: tendency for self-discipline, organization, and goal-directed behavior; neuroticism: susceptibility to negative emotions like anxiety, depression, and emotional instability; openness: inclination toward creativity, curiosity, and openness to new experiences.

Human values (BVS):
conformity, tradition, benov, univers, self, stim, hedon, achieve, power, security, open, selfenh, selftran, conserv, excitements, suprapersonal, interactive, promotion, existence, normative, linguistic, logicmath, spatial, bodykines, musical, interpersonal, intrapersonal, environmental, spiritual.

Demographic characteristics:
gender, department, age, residence in relation to the university.

Temporal variable:
day of the week

Mood metrics
mean, max and min mood of the day, of the morning (8:00 - 11:30), of the afternoon (12:30 - 18:30), and of the evening (19:00 - 22:00); expect day, mean mood^2: A non-linear transformation of the mean mood to capture potential quadratic effects.

4. Methodology

To analyze the determinants of daily evaluations (howwasday), we employed a combination of regression and machine learning techniques. Specifically, we utilized Linear Regression, Random Forest Regressor, and Gradient Boosting Regressor to identify the most influential variables and assess predictive accuracy. Each method was applied with varying sets of variables to explore their relative importance and predictive power.

To systematically assess the impact of different factors, we created four variable groupings: personality traits, demographic characteristics and temporal variables, mood and expectation variables, and all variables combined.

The first three models of linear regression, trained on the first three variables groupings, were built using OLS regression to examine direct relationships between predictors and the dependent variable. Three further models of linear regression, trained on the all variables set, were created using different regularizations: Ridge regression (L2 regularization), to address multicollinearity by penalizing large coefficients; Lasso regression (L1 regularization), to both address multicollinearity and perform feature selection by shrinking irrelevant coefficients to zero; Principal Component Analysis (PCA), which reduced dimensionality by transforming the predictor space into uncorrelated components while retaining 95% of the variance, followed by regression on the principal components.

The Random Forest regressor is an ensemble method that was employed to capture non-linear relationships and interactions between predictors. By averaging predictions from multiple decision trees, Random Forest provide robust predictions and feature importance scores, highlighting the most influential variables.

The Gradient Boosting regressor is an advanced ensemble method that was used to iteratively improve predictions by focusing on errors from previous iterations. Gradient Boosting is particularly effective for capturing complex, non-linear patterns in the data.

For each model, we evaluated performance using standard regression metrics: Mean Squared Error (MSE), which measures the average squared difference between observed and predicted values; Mean Absolute Error (MAE), which captures the average absolute differences between predictions and actual values; and R-squared (R^2), which indicates the proportion of variance in the dependent variable explained by the model.

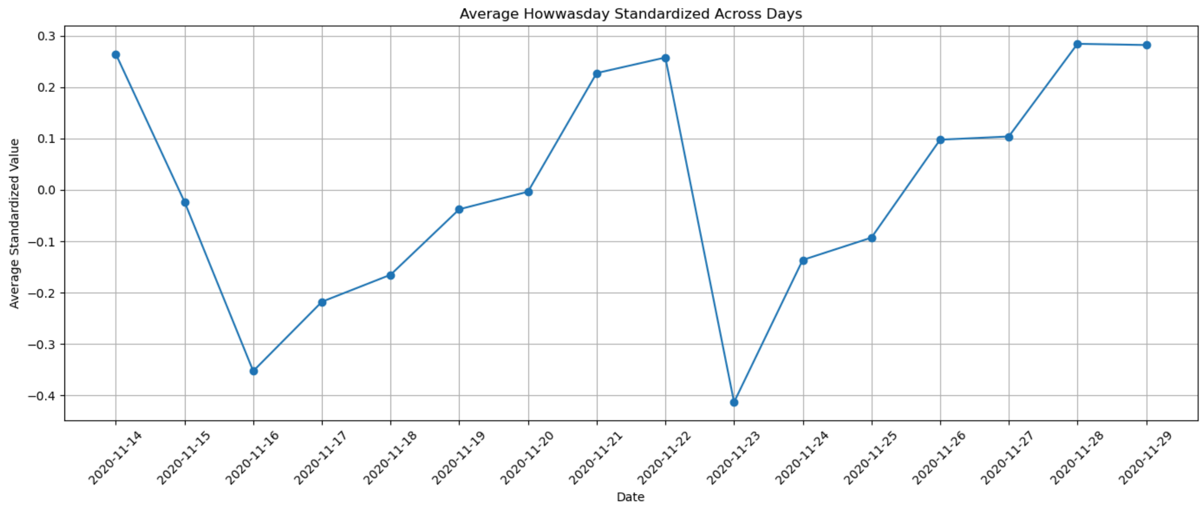
Models were trained on a set comprising 70% of the observations and tested on a set comprising 30% of the observations to assess generalizability.

Employing a diverse set of methods and variable groupings allowed us to systematically evaluate the determinants of daily evaluations while addressing challenges such as multicollinearity and non-linearity in the data. We also created these same models but trained and tested on only the workdays and separately on only the holidays, to better understand the day evaluation and compare them to the models trained on the full data.

5. Results

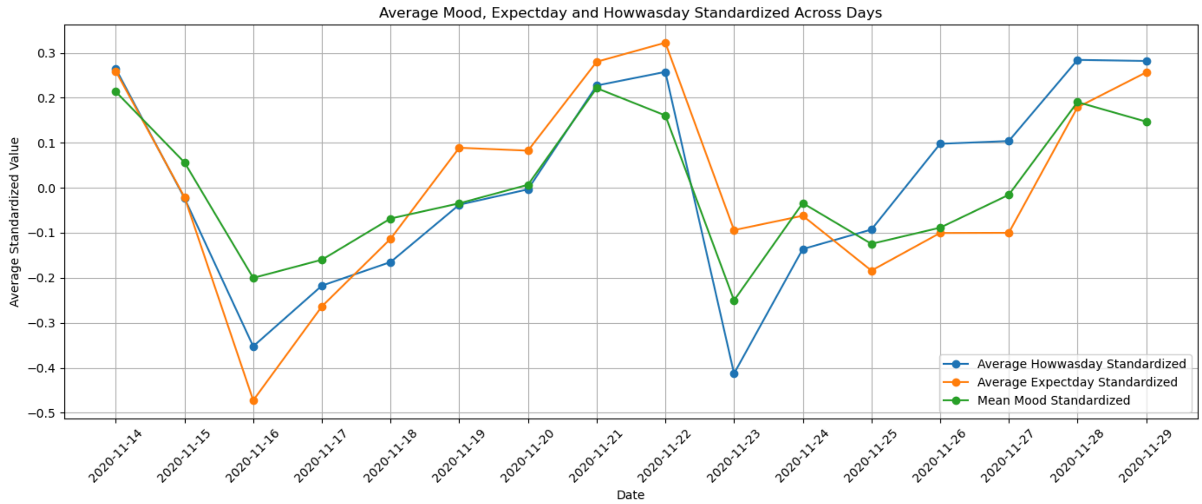
We begin our analysis by displaying the result. First, we show the distribution of the main variables.

Graph 1: Average standardized how was day trend across days



Looking at the distribution of the evening evaluation of the day we can see that the highest scores are registered on the weekend days and the lowest are on Monday. The trend shows that the how was day score constantly increases from Monday to Sunday.

Graph 2: Average standardized how was day, expect day and mood across days



We can see that the average mood and expect day follow very closely the day evaluation trend, which is increasing from Monday, the lowest point, to Sunday, the highest point.

Table 1: Linear regression model with personality traits features

Feat.	Coeff.	S.E.	P-value
cost.	4.3334	0.378	0.000
Extraversion	-0.0027	0.002	0.153
Agreeableness	0.0043	0.002	0.037
Conscientiousness	-0.0040	0.002	0.025
Neuroticism	-0.0086	0.002	0.000
Openness	-0.0037	0.003	0.153
Pconformity	1.946e+05	9.44e+04	0.040
Ptradition	1.946e+05	9.44e+04	0.040
Pbenov	-2.144e+05	8.19e+04	0.009
Punivers	-2.144e+05	8.19e+04	0.009
Pself	1.254e+05	9.78e+04	0.200
Pstim	1.254e+05	9.78e+04	0.200

Phedon	-1.371e+05	8.29e+04	0.098
Pachieve	-1.371e+05	8.29e+04	0.098
Ppower	-1.371e+05	8.29e+04	0.098
Psecurity	1.946e+05	9.44e+04	0.040
Popen	-2.508e+05	1.96e+05	0.200
Pselfenh	4.114e+05	2.49e+05	0.098
Pselftran	4.288e+05	1.64e+05	0.009
Pconserv	-5.837e+05	2.83e+05	0.040
Pexcitements	-0.0254	0.045	0.576
Psuprapersonal	0.0453	0.050	0.361
Pinteractive	-0.0645	0.040	0.110
Ppromotion	-0.0166	0.045	0.712
Pexistence	-0.0296	0.047	0.532
Pnormative	0.0239	0.030	0.426
Linguistic	0.0043	0.002	0.055
Logicmath	-0.0028	0.002	0.173
Spatial	0.0041	0.003	0.122
Bodykines	0.0013	0.002	0.458
Musical	0.0022	0.001	0.113
Interpersonal	6.957e-05	0.003	0.983
Intrapersonal	-0.0114	0.003	0.000
Environmental	0.0112	0.002	0.000
Spiritual	0.0042	0.002	0.087

R^2	0.159
N	813

This table presents the results of a linear regression model predicting how was day using personality traits and human values.

The model explains 15.9% of the variance in the dependent variable. While this indicates some predictive power, the relatively low value suggests that other factors not included in this model might significantly influence daily evaluations.

The extraversion coefficient is negative but not statistically significant, suggesting no strong evidence that extraversion directly impacts daily evaluations. Agreeableness has a small positive and significant effect, indicating that higher agreeableness is associated with slightly better daily evaluations. Conscientiousness has a small but significant negative effect, suggesting that higher conscientiousness may slightly reduce positive evaluations. This could reflect stress or perfectionism in conscientious individuals. Neuroticism has a significant and stronger negative effect, indicating that higher neuroticism is strongly associated with worse daily evaluations. This aligns with neuroticism's link to negative emotionality. The openness coefficient is not significant, suggesting no clear impact of openness on daily evaluations.

Pconformity, Ptradition, Psecurity positively influence daily evaluations, suggesting that individuals who prioritize conformity, tradition, and security may experience better daily outcomes. Pbenov, Punivers negatively impact daily evaluations, potentially reflecting tension or unmet expectations in individuals who prioritize benevolence and universalism. Pselftran has a significant positive effect, indicating that self-transcendence is linked to better daily evaluations. Pconserv has a significant negative effect, suggesting that individuals with a strong focus on conservation values may experience worse daily evaluations.

Intrapersonal has a significant negative effect, indicating that individuals with strong intrapersonal skills might evaluate their day more critically, possibly due to higher self-awareness. Environmental has a significant positive effect, suggesting that individuals with strong environmental awareness tend to evaluate their day more positively. Linguistic is marginally significant, indicating a potential positive effect of linguistic skills on daily evaluations.

The strongest predictors include Neuroticism (negative), Environmental (positive), Intrapersonal (negative), and certain human values like Pselftran (positive) and Pconserv (negative). Many variables, such as Extraversion, Openness, and Musical, show no significant relationship with daily evaluations, indicating that these traits may not directly influence the dependent variable in this context.

The model highlights the importance of personality traits and human values in shaping daily evaluations. While the findings align with theoretical expectations (e.g., the negative impact of neuroticism and the positive influence of environmental awareness), the relatively low explanatory power suggests the need for a broader set of predictors to better understand the determinants of daily evaluations.

Table 2: Linear regression model with demographic characteristics and temporal features

Feat.	Coeff.	S.E.	P-value
cost.	4.2378	0.186	0.000
Female (reference)			
Male	0.2747	0.068	0.000
Business and Economics (reference)			
Engineering and Applied Sciences	-0.1123	0.122	0.356
Humanities	0.2418	0.150	0.108
Law	0.2050	0.126	0.105
Medicine and veterinary medicine	-0.1492	0.226	0.509
Natural Sciences	-0.0394	0.123	0.749
Social Sciences	0.1366	0.128	0.286
17 - 18 years old (reference)			
19 years old	-0.1745	0.140	0.213
20 years old	-0.2449	0.150	0.104
21 years old	-0.0525	0.149	0.724
22 years old	0.2121	0.166	0.203
23 years old	-0.0556	0.163	0.734
24 years old	-0.2230	0.166	0.179
25 - 26 years old	0.0386	0.166	0.817

27+ years old	0.0008	0.181	0.997
residence in Trento (reference)			
residence in another city close to university	0.1909	0.092	0.038
residence in another city away from university	0.1051	0.076	0.169
residence in Rovereto	0.0965	0.126	0.444
Monday (reference)			
Tuesday	0.1643	0.109	0.131
Wednesday	0.1853	0.107	0.084
Thursday	0.2440	0.108	0.024
Friday	0.2423	0.108	0.025
Saturday	0.4803	0.097	0.000
Sunday	0.3706	0.100	0.000
R^2	0.094		
N	813		

This table presents the results of a linear regression model predicting how was day using demographic characteristics and contextual variables. Below is an interpretation of the results.

The model explains 9.4% of the variance in how was day, indicating limited explanatory power. This suggests that while demographic characteristics and contextual factors contribute to daily evaluations, other variables likely play a more significant role.

Male participants have significantly higher daily evaluations compared tofemale subjects. The positive coefficient suggests a notable gender difference in how days are evaluated.

None of the departments and age groups show statistically significant differences compared to their reference group. Residence close to Trento has a significant positive effect indicating that students living near Trento tend to evaluate their days more positively compared to those living in Trento. Residence in Rovereto and residence in another city away from the university show non-significant effects. Significant positive effects are observed for Thursday, Friday, Saturday, and Sunday, these results suggest that participants evaluate their days more positively on weekends and toward the end of the week, compared to Monday. Wednesday and Tuesday show non-significant effects.

Male gender, residence close to Trento, and certain days of the week (Thursday, Friday, Saturday, Sunday) are significant predictors of daily evaluations. The strong effects of weekend days (Saturday and Sunday) highlight the role of leisure and reduced academic/work pressures in improving daily evaluations. Academic department and age group show no significant impact on daily evaluations, suggesting that these demographic factors may not directly influence how participants perceive their days.

This model highlights the influence of gender, residence, and day of the week on daily evaluations, with weekends and proximity to Trento contributing positively. However, the low explanatory power suggests that demographic characteristics alone are insufficient for fully understanding the determinants of daily evaluations.

Table 3: Linear regression model with expectation and mood features

Feat.	Coeff.	S.E.	P-value
cost.	2.1731	0.121	0.000
Expectday	0.4693	0.030	0.000
Mean mood	0.5239	0.051	0.004
Mean mood^2	-0.1451	0.034	0.000
Max mood	-0.0450	0.023	0.051
Min mood	0.0475	0.018	0.009
R^2	0.464		
N	813		

This table presents the results of a linear regression model predicting how was day using expectation and mood-related features. Below is an interpretation of the findings.

The model explains 46.4% of the variance of how was day, indicating a very good explanatory power. The high R-squared suggests that mood metrics are a fundamental aspect of the day evaluations.

Expectday has a significant positive coefficient indicates that higher morning expectations of how the day will go are strongly associated with better end of the day evaluations. This result highlights the importance of initial optimism in shaping daily perceptions.

Mean Mood has a significant positive coefficient, suggesting that higher average mood throughout the day is associated with better end of the day evaluations. This aligns with the expectation that sustained positive emotions contribute to favorable daily assessments.

Mean Mood^2 has a significant negative coefficient, suggesting a non-linear relationship between mean mood and daily evaluations. Specifically, while higher mean mood improves evaluations, the effect diminishes at very high levels of mood, potentially reflecting diminishing returns of positive affect.

Max Mood has a marginally significant negative coefficient, suggesting that the highest mood experienced during the day may slightly reduce daily evaluations. This could indicate that extreme mood highs are less influential than sustained positive moods.

Min Mood has a significant positive coefficient, indicating that higher minimum mood levels are associated with better daily evaluations. This suggests that avoiding extreme negative moods is crucial for favorable day assessments.

Morning expectations significantly influence end of the day evaluations, supporting the idea that initial optimism or pessimism can set the tone for the entire day. Mean Mood is a strong predictor, reinforcing the importance of overall emotional experiences throughout the day. The quadratic term (Mean Mood^2) reveals that the relationship between mood and daily evaluations is not strictly linear, suggesting that overly high moods may not always enhance day assessments. Min Mood has a positive effect, emphasizing the importance of maintaining a baseline level of emotional stability to avoid negative day evaluations. Max Mood shows a marginally negative effect, which might indicate that short-lived emotional highs are less impactful than the overall mood trend.

This model highlights the critical role of expectations and mood in shaping daily evaluations. Morning optimism (Expectday) and overall mood status (Mean Mood and Min Mood) are significant predictors, while the non-linear effect of mood suggests diminishing returns of extreme positivity.

Table 4: Performance on the test set for linear regression models

	MSE	MAE	R^2
Personality traits model	0.56	0.58	0.15
Demographic and temporal variables model	0.61	0.59	0.07
Expectation and mood model	0.37	0.45	0.44
Ridge model	0.33	0.43	0.50
Lasso model	0.34	0.44	0.48
PCA model	0.34	0.44	0.48

This table presents the performance metrics of various linear regression models evaluated on the test set. Below is an interpretation of the results.

The personality traits model explains only 15% of the variance in the dependent variable. While personality traits provide some predictive power, their ability to explain day evaluations is limited, suggesting that other factors play a more significant role.

The demographic characteristics and temporal variables model has the lowest R^2, indicating that demographic and temporal variables alone are insufficient for explaining daily evaluations. The relatively high MSE and MAE further suggest that this model lacks predictive accuracy.

The expectation and mood model significantly outperforms the previous two, explaining 44% of the variance in the dependent variable. The lower MSE and MAE indicate better predictive accuracy, highlighting the importance of expectations and mood in shaping daily evaluations.

Ridge regression, with its ability to handle multicollinearity through L2 regularization, achieves the best performance, explaining 50% of the variance. The lowest MSE and MAE indicate the highest predictive accuracy among all models.

Lasso regression performs slightly worse than Ridge but still achieves strong predictive accuracy. Its feature selection capability (via L1 regularization) likely helped reduce noise, resulting in a model that balances simplicity and performance.

PCA performs similarly to Lasso, suggesting that dimensionality reduction effectively captures the most important predictors while reducing redundancy.

The expectation and mood model's strong performance underscores the importance of mood metrics and expectations as key predictors of daily evaluations. Ridge, Lasso, and PCA models outperform simpler models, highlighting the value of addressing multicollinearity and reducing overfitting when using a large number of predictors. Personality traits and demographic models' poor performance suggests that personality traits and demographic factors alone are insufficient for predicting daily evaluations.

The results highlight the importance of expectation and mood metrics in predicting daily evaluations, with regularized methods (Ridge and Lasso) and dimensionality reduction (PCA) providing the best performance. Simpler models based on personality traits or demographic factors explain only a small portion of the variance, emphasizing the need for more comprehensive and nuanced predictors.

Table 5: Performance on the test set for Random Forest models

	MSE	MAE	R^2
Personality traits model	0.47	0.53	0.29
Demographic and temporal variables model	0.64	0.60	0.02
Expectation and mood model	0.40	0.45	0.38
All variables model	0.37	0.42	0.44

This table presents the performance metrics of Random Forest models evaluated on the test set, using different sets of features. Below is an interpretation of the results.

The personality traits model explains 29% of the variance in the dependent variable. Personality traits contribute moderately to predicting daily evaluations, indicating their relevance but also suggesting the need for additional features to improve accuracy.

The demographic and temporal variables model has very low R^2, suggesting that demographic and temporal variables alone are poor predictors of daily evaluations. The high MSE and MAE further confirm the model's weak predictive performance, indicating that these variables do not capture sufficient variance in the dependent variable.

The expectation and mood model performs significantly better than the first two, explaining 38% of the variance. The lower MSE and MAE highlight the importance of expectations and mood metrics in predicting daily evaluations. The all variables model achieves the best performance, explaining 44% of the variance. The inclusion of all variables allows the Random Forest model to leverage complex interactions and non-linear relationships, resulting in the highest predictive accuracy.

The all variables model outperforms the others, emphasizing the value of combining diverse predictors, such as personality traits, demographic characteristics, and mood metrics. The ability of Random Forest to handle high-dimensional data and capture non-linear interactions likely contributes to this superior performance. The expectation and mood model demonstrates strong performance, reinforcing the critical role of these variables in shaping daily evaluations. The demographic and temporal variables model performs poorly, with an

R^2 of only 0.02. This suggests that these features, while informative in some contexts, are insufficient for predicting daily evaluations on their own. The personality traits model performs better than the demographic model but worse than the mood-based models. This indicates that while personality traits are relevant, they are not as impactful as mood-related factors.

The results demonstrate the superior performance of models that include expectation and mood metrics, with the all variables model achieving the highest predictive accuracy. Demographic and temporal variables alone provide minimal explanatory power, while personality traits contribute moderately. These findings highlight the importance of integrating diverse predictors and leveraging advanced machine learning methods like Random Forest to capture the complexity of daily evaluations.

Table 6: Features importance for the all variables Random Forest model

Feat.	Importance	Feat.	Importance
Expectday	0.248204	Pexcitements	0.005138
Mean mood	0.236968	Ptradition	0.005135
Min mood	0.040962	Saturday	0.005112
Mean mood afternoon	0.031857	Sunday	0.005110
Min mood afternoon	0.029730	Friday	0.004996
Mean mood^2	0.027389	Interpersonal	0.004859
Max mood afternoon	0.027075	Psecurity	0.004743
Max mood	0.016879	Pconformity	0.004522
Mean mood evening	0.011806	Phedon	0.004308
Pexistence	0.011756	Ppower	0.004205
mean_mood_morning	0.011376	Wednesday	0.003890
Pnormative	0.011082	Pstim	0.003582
Neuroticism	0.009948	Monday	0.003411
max_mood_evening	0.009854	Environmental	0.003115

Bodykines	0.009330	Tuesday	0.002768
Musical	0.009259	Male	0.002115
Spatial	0.009182	Thursday	0.002110
Linguistic	0.008875	cohort_27+	0.001951
max_mood_morning	0.008802	cohort_20	0.001672
Pself	0.008660	cohort_21	0.001606
Pselfenh	0.008518	dep Natural Sciences	0.001458
Pconserv	0.008359	dep Business/ economics	0.001093
min_mood_evening	0.008341	cohort_22	0.001093
min_mood_morning	0.008163	close to university	0.001087
Agreeableness	0.008147	cohort_19	0.001006
Psuprapersonal	0.008050	Rovereto	0.000959
Extraversion	0.008035	Trento	0.000947
Spiritual	0.007888	cohort_24	0.000837
Popen	0.007034	dep Social Science	0.000828
Pbenov	0.006972	far away from university	0.000751
Pachieve	0.006933	cohort_23	0.000741
Logicmath	0.006721	Female	0.000737
Ppromotion	0.006688	cohort_17-18	0.000596
Intrapersonal	0.005878	dep Law	0.000590
Openness	0.005585	dep Medicine	0.000527
Punivers	0.005459	dep Engineering	0.000381
Pselftran	0.005332	dep Humanities	0.000256

Pinteractive	0.005321	cohort_25-26	0.000177
Conscientiousness	0.005170		

This table lists the features contribution to the predictions of the all variables Random Forest model, ranked by their importance scores.

Expectday (0.248) and Mean Mood (0.237) are the two most influential features, significantly outweighing other predictors. This reinforces the idea that morning expectations and overall mood are the primary determinants of how individuals evaluate their day. Other mood-related metrics, such as Min Mood (0.041), Mean Mood Afternoon (0.032), and Min Mood Afternoon (0.030), also rank highly, suggesting that both overall emotional states and specific periods of the day (e.g., the afternoon) strongly influence evaluations. The inclusion of Mean Mood² (0.027) highlights the importance of capturing non-linear effects, where extreme moods (both positive and negative) may have diminishing impacts on evaluations.

While mood and expectation dominate, personality traits such as Neuroticism (0.010), Agreeableness (0.008), and Extraversion (0.008) contribute moderately. These traits likely provide additional context by capturing individual differences in emotional sensitivity and stability. Human values such as Pexistence (0.012), Pnormative (0.011), and Pself (0.009) appear among the mid-ranking features, indicating their relevance but secondary importance compared to mood and expectation variables.

Temporal variables like Saturday (0.005), Sunday (0.005), and Friday (0.005) show modest importance, suggesting that the day of the week plays a limited but notable role in shaping daily evaluations. This may reflect differences in routines or expectations associated with weekends versus weekdays.

Contextual variables, such as location and social context, rank lower, with features like close to university (0.001) and Trento (0.001) having minimal impact. This suggests that physical and social environments are less critical than intrinsic and emotional factors in determining daily evaluations.

Demographic variables such as gender (Male: 0.002, Female: 0.001), cohort, and academic department contribute minimally to the model. Their low importance indicates that demographic characteristics are less predictive of daily evaluations compared to mood, expectations, and personality traits.

Metrics tied to specific times of the day, such as Mean Mood Morning (0.011), Max Mood Afternoon (0.027), and Min Mood Evening (0.008), suggest that certain periods may disproportionately influence end-of-day evaluations. This could reflect the recency effect, where later experiences weigh more heavily on overall impressions.

The overwhelming importance of Expectday and Mean Mood highlights their central role in shaping daily evaluations, underscoring the value of emotional and anticipatory states in determining subjective well-being. The inclusion of quadratic terms like Mean Mood² indicates that extreme moods have a less pronounced effect, emphasizing the importance of balanced emotional states.

While demographic, temporal, and contextual variables contribute to some extent, their predictive power is overshadowed by mood and personality-related features. Although secondary, personality traits provide valuable insights into individual differences.

The results align with prior results emphasizing the primacy of expectations and mood in subjective evaluations, while also highlighting the complementary role of personality traits.

Table 7: Performance on the test set for Gradient Boosting models

	MSE	MAE	R^2
Personality traits model	0.46	0.53	0.30
Demographic and temporal variables model	0.57	0.57	0.14
Expectation and mood model	0.37	0.43	0.43
All variables model	0.33	0.41	0.50

This table presents the performance of Gradient Boosting models across different sets of features in predicting daily evaluations. Below is an interpretation of the results:

The personality traits model explains 30% of the variance, performing moderately well. Personality traits provide valuable information but lack sufficient predictive power on their own, suggesting the need for additional variables to improve accuracy.

The demographic and temporal variables model explains only 14% of the variance, indicating limited predictive power. Demographic and temporal features are relatively weak predictors of daily evaluations.

The expectation and mood model performs significantly better, explaining 43% of the variance. The low MSE and MAE underscore the importance of mood metrics and morning expectations in predicting daily evaluations.

The all variables model explains 50% of the variance, with the lowest error metrics. Combining all available features allows Gradient Boosting to leverage complex interactions and non-linear relationships, achieving the highest predictive accuracy.

The all variables model achieves the highest R^2 (0.50) and lowest error metrics, showcasing the power of Gradient Boosting to handle diverse and high-dimensional datasets effectively.

The expectation and mood model performs nearly as well as the full model, confirming the strong influence of mood patterns and morning expectations on daily evaluations.

The personality traits model performs moderately well but cannot match the predictive power of mood-related features. The demographic and temporal variables model has the weakest performance, indicating these variables provide limited value in isolation.

The results confirm the strength of Gradient Boosting in predicting daily evaluations. The all variables model achieves the best performance, while the expectation and mood model highlights the central role of emotional and temporal predictors. Personality traits and demographic features, while valuable, are insufficient on their own. These findings emphasize the need for comprehensive models that integrate diverse predictors to achieve accurate and robust predictions.

Table 8: Features importance for the all variables Gradient Boosting model

Feat.	Importance	Feat.	Importance
Expectday	0.287063	Saturday	0.002305
Mean mood	0.273077	Pconserv	0.002140
Mean mood^2	0.066500	Male	0.001910
Min mood	0.035746	Ppower	0.001768
Max mood afternoon	0.023676	Pbenov	0.001603
Pachieve	0.020592	cohort_27+	0.001561
Neuroticism	0.020047	cohort_20	0.001371
Min mood afternoon	0.016834	dep Social Sciences	0.001328
Pself	0.016191	Tuesday	0.001277
Mean mood afternoon	0.015995	min_mood_morning	0.001262
Interpersonal	0.013056	cohort_22	0.001261
Bodykines	0.012742	Openness	0.000859
max_mood_morning	0.011790	residence_Rovereto	0.000713
Pselfenh	0.011752	Conscientiousness	0.000688
Musical	0.010918	residence_Trento	0.000634
max_mood_evening	0.009793	dep Humanities	0.000604
Spatial	0.008985	Psecurity	0.000513

Agreeableness	0.008617	cohort_19	0.000507
Pstim	0.008159	Sunday	0.000447
Pnormative	0.007430	Friday	0.000365
Pexcitements	0.007334	Psuprapersonal	0.000216
Spiritual	0.007256	Environmental	0.000112
max_mood	0.007211	cohort_21	0.000000
Ppromotion	0.006994	dep Business/ economics	0.000000
Pexistence	0.006691	dep Engineering	0.000000
Intrapersonal	0.006087	dep Law	0.000000
Pselftran	0.005817	dep Medicine	0.000000
Logicmath	0.005807	dep Natural Sciences	0.000000
Ptradition	0.005496	cohort_17-18	0.000000
Monday	0.005203	Female	0.000000
Popen	0.004892	Phedon	0.000000
mean_mood_morning	0.004522	cohort_24	0.000000
Extraversion	0.004157	Wednesday	0.000000
mean_mood_evening	0.004041	cohort_25-26	0.000000
Pinteractive	0.003712	Thursday	0.000000
Punivers	0.003539	close to university	0.000000
Pconformity	0.003164	far from university	0.000000
min_mood_evening	0.003070	cohort_23	0.000000
Linguistic	0.002601		

This table presents the features contribution to the predictions of the all variables Gradient Boosting model, ranked by their importance scores.

Expectday (0.287) and Mean Mood (0.273) are the two most influential features, significantly surpassing all others. This reaffirms the critical role of morning expectations and overall mood in shaping how individuals evaluate their day. Mean Mood^2 (0.067) ranks third, emphasizing the importance of capturing non-linear effects of mood, where extreme values may have diminishing or different impacts on daily evaluations. Other mood-related metrics, such as Min Mood (0.036), Max Mood Afternoon (0.024), and Mean Mood Afternoon (0.016), highlight the importance of specific periods and emotional lows in influencing end-of-day evaluations.

Among personality traits, Neuroticism (0.020) and Agreeableness (0.009) show moderate importance, likely reflecting their influence on emotional stability and social tendencies. Human values such as Pachieve (0.021), Pself (0.016), and Pselfenh (0.012) are among the mid-ranking features, suggesting their relevance in capturing individual priorities and self-perception. The relative importance of Pnormative (0.007) and Pexistence (0.007) further indicates that values tied to societal norms and existential concerns contribute meaningfully, albeit less so than mood and expectation variables.

Temporal features such as Monday (0.005) and Saturday (0.002) have low importance, indicating that the day of the week has minimal predictive power for daily evaluations. Contextual factors, such as location and social context, are also less influential, with features like residence in Rovereto (0.001) and residence in Trento (0.001) ranking near the bottom. This suggests that environmental factors are secondary to intrinsic emotional states and expectations.

Mood metrics tied to specific times of the day, such as Max Mood Morning (0.012) and Mean Mood Evening (0.004), indicate that different periods contribute uniquely to end of the day evaluations. This aligns with the idea that emotional highs and lows during distinct times may disproportionately affect overall perceptions.

Demographic features, including gender (Male: 0.002, Female: 0.000), cohort, and academic department, show minimal to no importance. Their lack of predictive power suggests that demographic characteristics are far less relevant than mood and expectation variables in determining daily evaluations.

The overwhelming importance of Expectday and Mean Mood underscores their central role in shaping daily evaluations, suggesting that interventions should prioritize these factors. The inclusion of Mean Mood^2 highlights the need to account for non-linear dynamics, where extreme moods may not have a straightforward impact on evaluations. While secondary, personality traits and human values provide meaningful insights into individual differences, helping to tailor interventions and predictive models. The limited importance of day-of-week and location variables suggests that daily evaluations are driven more by internal states than external circumstances.

The results align with prior results emphasizing the primacy of expectations and mood status.

6. Results for workdays

We now display the result of the models trained on the workdays data and compare them with the results of the models trained and tested on the complete data.

Table 9: Performance on the test set for linear regression models on workdays and comparison with the models trained on the complete data

	MSE	MAE	R^2
Personality traits model	0.46 (-0.10)	0.52 (-0.06)	0.21 (+0.06)
Demographic and temporal variables model	0.56 (-0.05)	0.56 (-0.03)	0.03 (-0.04)
Expectation and mood model	0.32 (-0.05)	0.41 (-0.04)	0.44 (+0.00)
Ridge model	0.26 (-0.07)	0.39 (-0.04)	0.55 (+0.05)
Lasso model	0.27 (-0.07)	0.38 (-0.06)	0.54 (+0.06)
PCA model	0.33 (-0.01)	0.43 (-0.01)	0.43 (-0.05)

This table presents the performance of linear regression models specifically on workdays, compared to the general models evaluated across all days. The comparison provides insights into how the models perform in a more structured temporal context, where daily routines are likely more consistent.

The personality traits model exhibits a notable improvement (-0.10) in Mean Squared Error (MSE) on workdays, suggesting better predictive accuracy in this context. The Mean Absolute Error (MAE) also decreases (-0.06), indicating that the average prediction error is smaller on workdays. The R-squared value increases (+0.06), suggesting that personality traits explain a larger proportion of the variance in mood evaluations on workdays compared to the general model. This may be due to the more predictable structure of workdays, where personality traits have a stronger influence on mood dynamics.

The demographic and temporal variables model's MSE decreases (-0.05), reflecting slightly better accuracy on workdays. The MAE shows a modest improvement (-0.03), but the error remains relatively high compared to other models. A decrease in R-squared (-0.04) suggests that demographic and temporal variables are less effective in explaining mood evaluations on workdays. This may indicate that these variables are more relevant in capturing variability across a broader range of contexts.

The expectation and mood model's MSE and MAE metrics improve (-0.05 and -0.04, respectively), highlighting enhanced predictive accuracy and reduced error on workdays. The R-squared value remains unchanged (+0.00), indicating that the proportion of explained variance is consistent across general and workday-specific models. This suggests that mood and expectation variables are robust predictors regardless of the day type.

The Ridge model demonstrates significant improvements in both MSE (-0.07) and MAE (-0.04), showcasing its ability to handle workday-specific patterns effectively. The R-squared value increases (+0.05), reinforcing the model's ability to capture more variance in mood evaluations on workdays. This highlights the strength of regularization techniques in refining predictions under structured conditions.

The Lasso model shows the largest reduction in MAE (-0.06) among all models, emphasizing its effectiveness in minimizing prediction errors on workdays. A notable increase (+0.06) in R-squared underscores the model's ability to identify and leverage key predictors for mood evaluations in this context. Lasso's feature selection capability likely enhances its performance on workdays.

PCA model's MSE and MAE improvements are modest (-0.01 for both metrics), suggesting that PCA's dimensionality reduction is less impactful in the workday context. A slight decrease (-0.05) in R-squared indicates that PCA may lose some explanatory power when applied specifically to workdays. This could reflect a limitation in capturing the nuances of structured daily routines.

The expectation and mood model continues to perform robustly across contexts, maintaining high R^2 and low error metrics, underscoring the importance of these variables in predicting mood evaluations.

Regularization techniques, particularly Lasso regression, show strong performance improvements on workdays, likely due to their ability to prioritize relevant features under structured conditions.

The personality traits model benefits significantly from the predictable nature of workdays, with notable gains in both accuracy and explanatory power.

The demographic and temporal variables model shows limited improvement, suggesting that these variables may be less critical in explaining mood evaluations on workdays compared to other predictors.

These findings highlight the value of tailored models for specific temporal contexts, such as workdays, and reinforce the importance of incorporating diverse variable sets to optimize predictive accuracy.

Table 10: Performance on the test set for Random Forest models on workdays and comparison with the models trained on the complete data

	MSE	MAE	R^2
Personality traits model	0.46 (-0.01)	0.50 (-0.03)	0.20 (-0.09)
Demographic and temporal variables model	0.75 (+0.11)	0.61 (+0.01)	-0.30 (-0.32)
Expectation and mood model	0.34 (-0.06)	0.40 (-0.05)	0.42 (+0.04)
All variables model	0.31 (-0.06)	0.38 (-0.04)	0.47 (+0.03)

This table evaluates the performance of Random Forest models on workdays compared to their performance across all days. The results highlight variations in predictive accuracy and the explanatory power of different variable groupings when applied to a more structured context.

Personality traits model's MSE decreases slightly (-0.01), indicating a marginal improvement in predictive accuracy on workdays. The MAE shows a moderate reduction (-0.03), suggesting that predictions on workdays are slightly more precise. A significant decrease (-0.09) in R-squared suggests that personality traits explain less variance in mood evaluations on workdays compared to the general model. This might indicate that personality traits play a less prominent role in influencing mood within the structured routines of workdays.

The demographic and temporal variables model shows a notable increase (+0.11) in MSE reflecting a reduced predictive accuracy on workdays. The MAE increases slightly (+0.01), further indicating less precise predictions. A dramatic decrease (-0.32) results in a negative R-squared value, suggesting that the model performs worse than a simple mean-based prediction. This indicates that demographic and temporal variables alone may not adequately capture the dynamics of mood on workdays, potentially due to their limited variability in this context.

The expectation and mood model's MSE decreases (-0.06), reflecting improved accuracy in predicting mood evaluations on workdays. The MAE also shows a meaningful reduction (-0.05), highlighting better precision in predictions. An increase (+0.04) in R-squared demonstrates that expectation and mood variables are even more effective at explaining variance in mood evaluations on workdays. This aligns with the idea that mood and expectations are closely tied to structured daily routines.

All Variables Model shows a decrease (-0.06) in MSE indicates improved predictive accuracy on workdays. The MAE also improves (-0.04), suggesting more precise predictions. A slight increase (+0.03) in R-squared shows that incorporating all variables remains highly effective for explaining mood evaluations, even in the structured context of workdays. The comprehensive nature of this model allows it to capture the multifaceted determinants of mood more effectively than individual variable groupings.

The expectation and mood model consistently demonstrates strong performance, with notable improvements in both accuracy and explanatory power on workdays. This underscores the importance of mood dynamics and expectations in structured daily routines.

The demographic and temporal variables model performs poorly on workdays, with increased error metrics and a significant drop in R-squared. This suggests that these variables are less relevant in predicting mood when daily routines are more uniform.

The all variables model continues to perform well, with improvements in all metrics. Its ability to incorporate diverse predictors likely contributes to its robust performance across different contexts.

The personality traits model shows marginal gains in accuracy but a substantial drop in R-squared, suggesting that personality traits may play a less central role in influencing mood on workdays compared to other variable groupings.

These results emphasize the importance of context-specific modeling, particularly the enhanced relevance of mood and expectation variables on workdays. While the all variables model remains the most comprehensive and robust, the underperformance of demographic and temporal variables highlights the need for more nuanced approaches to modeling mood dynamics in structured daily contexts.

Table 11: Performance on the test set for Gradient Boosting models on workdays and comparison with the models trained on the complete data

	MSE	MAE	R^2
Personality traits model	0.46 (-0.01)	0.51 (-0.02)	0.20 (-0.10)
Demographic and temporal variables model	0.59 (+0.02)	0.55 (-0.02)	-0.02 (-0.16)
Expectation and mood model	0.34 (-0.03)	0.43 (+0.00)	0.42 (+0.01)
All variables model	0.31 (-0.02)	0.40 (-0.01)	0.47 (-0.03)

This table evaluates the Gradient Boosting models' performance on workdays, highlighting variations in accuracy, error, and explanatory power compared to their general performance across all days.

The personality traits model shows a small reduction (-0.01) indicates slightly improved predictive accuracy on workdays. A moderate reduction (-0.02) suggests more precise predictions. A significant decrease (-0.10) in R-squared reflects a reduced ability of personality traits to explain mood variance on workdays. This suggests that personality traits may play a diminished role in structured contexts, such as workdays, compared to more variable contexts like weekends or holidays.

The demographic and temporal variables model shows a slight increase (+0.02) in MSE indicates reduced accuracy in predictions for workdays. A small reduction (-0.02) in MAE suggests marginally better precision despite the higher error in squared terms. A substantial drop (-0.16) in R-squared, resulting in a near-zero value, suggests that demographic and temporal variables are less effective in explaining mood changes on workdays. This aligns with the idea that such variables might not capture the nuances of mood dynamics in structured daily routines.

The expectation and mood model shows a moderate reduction (-0.03) reflects improved predictive accuracy on workdays. No change (+0.00) in MAE indicates consistent precision in predictions. A slight increase (+0.01) demonstrates that expectation and mood variables maintain their strong explanatory power on workdays. This highlights the importance of these variables in structured daily contexts, where expectations and mood fluctuations are likely more predictable.

All variables model shows a minor reduction (-0.02) suggests improved predictive accuracy on workdays. A small reduction (-0.01) indicates slightly more precise predictions. A slight decrease (-0.03) in R-squared suggests a marginally reduced ability to explain mood variance compared to the general model. Despite this, the all variables model remains the best-performing model, underscoring its robustness and comprehensive approach to capturing mood determinants.

The expectation and mood model consistently performs well, with improvements in accuracy and explanatory power on workdays. This reinforces the significance of these variables in structured contexts.

The demographic and temporal variables model exhibits increased error and a significant drop in R-squared, suggesting limited relevance of these variables in predicting mood on workdays.

Despite a slight reduction in R-squared, the all variables model remains the most comprehensive and effective, capturing the multifaceted determinants of mood changes on workdays.

The personality traits model demonstrates marginal gains in accuracy but a notable drop in R-squared, indicating a diminished role of personality traits in structured contexts like workdays.

These results emphasize the importance of using context-specific variables to predict mood changes. The expectation and mood model continues to perform strongly, while the all variables model remains the most robust and generalizable. The limited performance of demographic and temporal variables suggests the need for additional factors to better capture mood dynamics on workdays.

7. Results for holidays

We now display the result of the models trained on the holidays data and compare them with the results of the models trained and tested on the complete data.

Table 12: Performance on the test set for linear regression models on holidays and comparison with the models trained on the complete data

	MSE	MAE	R^2
Personality traits model	0.56 (+0.00)	0.58 (+0.00)	0.15 (+0.00)
Demographic and temporal variables model	0.62 (+0.01)	0.60 (+0.01)	0.06 (-0.01)
Expectation and mood model	0.37 (+0.00)	0.45 (+0.00)	0.44 (+0.00)
Ridge model	0.33 (+0.00)	0.43 (+0.00)	0.50 (+0.00)
Lasso model	0.34 (+0.00)	0.44 (+0.00)	0.48 (+0.00)
PCA model	0.34 (+0.00)	0.44 (+0.00)	0.48 (+0.00)

This table evaluates the linear regression models' performance on holidays, comparing their predictive accuracy, error, and explanatory power with their general performance across all days.

The personality traits model's MSE, MAE, and R^2 show no changes compared to the general model suggest that personality traits influence mood evaluations consistently across holidays and other contexts. The model's moderate performance indicates that personality traits alone provide limited explanatory power for mood changes on holidays, likely due to the greater variability in activities and contexts on such days.

The demographic and temporal variables model shows a slight increases in MSE and MAE (+0.01) suggest marginally reduced accuracy and precision on holidays. A small decrease in R^2 (-0.01) in explanatory power indicates that demographic and temporal variables are slightly less effective at explaining mood changes on holidays. This may be due to reduced structure in daily routines, which weakens the predictive power of temporal patterns and demographic characteristics.

The expectation and mood model's MSE, MAE, and R^2 show no changes compared to the general model highlight the consistent importance of expectation and mood variables in explaining mood changes, regardless of the day type. The strong performance of this model underscores the relevance of these variables on holidays, where mood dynamics are likely shaped by personal expectations and emotional states.

Ridge Model's MSE, MAE, and R^2 show no changes compared to the general model suggest that the Ridge model performs robustly across different contexts. Its ability to handle multicollinearity makes it well-suited for capturing the complex interplay of variables influencing mood changes on holidays.

Lasso Model's MSE, MAE, and R^2 show no changes compared to the general model indicate consistent performance on holidays. The Lasso model's feature selection capability likely retains the most relevant predictors for mood changes, ensuring stable performance across contexts.

PCA Model's MSE, MAE, and R^2 show no changes compared to the general model suggest that the dimensionality reduction approach of PCA captures the key factors influencing mood changes on holidays. Its performance remains comparable to that of the Lasso model.

Most models exhibit no changes in performance metrics, indicating that their predictive power is not significantly affected by the unique dynamics of holidays.

The expectation and mood model continues to demonstrate strong performance, reinforcing the importance of these variables in understanding mood changes, even on holidays when activities and routines are more flexible.

The slight reduction in explanatory power for the demographic and temporal variables model highlights its limited relevance on holidays, where routines and temporal structures may be less rigid.

Ridge and Lasso models maintain stable performance across contexts, demonstrating their robustness in handling multicollinearity and selecting relevant predictors.

The results indicate that mood dynamics on holidays are best explained by expectation and mood variables, while demographic and temporal variables play a less significant role. The stability of Ridge, Lasso, and PCA models underscores their suitability for analyzing mood changes across diverse contexts. These findings highlight the need for context-specific variables to better capture the nuances of mood dynamics on holidays.

Table 13: Performance on the test set for Random Forest models on holidays and comparison with the models trained on the complete data

	MSE	MAE	R^2
Personality traits model	0.47 (+0.00)	0.53 (+0.00)	0.29 (+0.00)
Demographic and temporal variables model	0.57 (-0.07)	0.57 (-0.03)	0.14 (+0.12)
Expectation and mood model	0.40 (+0.00)	0.45 (+0.00)	0.38 (+0.00)
All variables model	0.37 (+0.00)	0.42 (+0.00)	0.44 (+0.00)

This table evaluates the performance of Random Forest models in predicting mood evaluations on holidays, with comparisons to their general performance across all days.

The personality traits model's MSE, MAE, and R^2 show no changes compared to the general model indicate that personality traits consistently influence mood dynamics, regardless of the day type. The model's moderate R^2 (0.29) suggests that personality traits capture some aspects of mood variation but are not the primary determinants on holidays, where external factors like activities or social contexts might play a larger role.

The demographic and temporal variables model show a notable decrease in MSE(-0.07) suggests improved error performance on holidays. A smaller decrease in MAE (-0.03) indicates slightly more precise predictions. A significant increase in R^2(+0.12) highlights that demographic and temporal variables are more predictive on holidays compared to other days. This improvement may reflect the greater variability in routines and temporal patterns on holidays, making these variables more relevant for explaining mood changes.

The expectation and mood model's MSE, MAE, and R^2 show no changes compared to the general model indicate that expectation and mood metrics are robust predictors across both holidays and general contexts. This model's strong performance ($R^2 = 0.38$) reinforces the importance of these variables in explaining mood dynamics, even when external structures and activities differ, as on holidays.

All Variables Model's MSE, MAE, and R^2 show no changes compared to the general model suggest that the inclusion of all variable types ensures stable performance, regardless of the day type. The model's strong R^2 (0.44) highlights the benefit of integrating diverse predictors to capture the multifaceted determinants of mood changes.

The personality traits, expectation and mood, and all variables models exhibit no changes in performance metrics, indicating their robustness in capturing mood dynamics across diverse contexts, including holidays.

The improved performance of the demographic and temporal variables model on holidays suggests that these factors become more relevant in less structured daily contexts. This highlights the importance of contextual variables in explaining mood changes.

The strong performance of the expectation and mood model underscores the centrality of these variables in understanding mood dynamics, even when external activities and routines vary significantly.

The stable performance of the all variables model demonstrates the Random Forest's capability to capture complex interactions between diverse predictors, providing robust predictions across contexts.

The results emphasize the importance of context-specific factors, such as demographic and temporal variables, in explaining mood changes on holidays. However, expectation and mood variables remain the most consistent and influential predictors. Random Forest models effectively integrate diverse predictors, offering stable and reliable performance in understanding mood dynamics on holidays.

Table 14: Performance on the test set for Gradient Boosting models on holidays and comparison with the models trained on the complete data

	MSE	MAE	R^2
Personality traits model	0.46 (+0.00)	0.53 (+0.00)	0.30 (+0.00)
Demographic and temporal variables model	0.57 (+0.00)	0.58 (+0.01)	0.13 (-0.01)
Expectation and mood model	0.37 (+0.00)	0.43 (+0.00)	0.43 (+0.00)
All variables model	0.33 (+0.00)	0.41 (+0.00)	0.50 (+0.00)

This table evaluates the Gradient Boosting models' performance in predicting mood evaluations on holidays, comparing results to their general performance across all days.

The personality traits model's MSE, MAE, and R^2 show no changes compared to the general model indicate that personality traits are equally predictive on holidays and across all days. The model's moderate R^2 (0.30) reflects its ability to capture some aspects of mood variability, but personality traits alone may not fully explain mood changes, especially on holidays when external influences like social interactions or activities might play a larger role.

The demographic and temporal variables model show no change in MSE suggesting that the error in predictions remains consistent across holidays and the general dataset. A slight increase in MAE (+0.01) indicates marginally less precise predictions on holidays. A small decrease in R^2 (-0.01) suggests that demographic and temporal variables have slightly less explanatory power on holidays compared to other days. This may reflect reduced variability in temporal patterns or demographic influences on mood during less structured holiday contexts.

The expectation and mood model's MSE, MAE, and R^2 show no changes compared to the general model highlight the robustness of expectation and mood variables in predicting mood changes across both structured (workdays) and unstructured (holidays) contexts. The model's strong R^2 (0.43) underscores the centrality of these variables in explaining mood dynamics, regardless of the day type.

All Variables Model's MSE, MAE, and R^2 show no changes compared to the general model suggest that combining all variable types ensures stable performance across diverse contexts. The high R^2 (0.50) indicates that integrating diverse predictors effectively captures the multifaceted determinants of mood changes, even in the more variable context of holidays.

The personality traits, expectation and mood, and all variables models show no changes in performance metrics, indicating their robustness in capturing mood dynamics regardless of the day type.

The small decrease in R^2 for the demographic and temporal variables model suggests that these factors may be less relevant on holidays when daily routines and temporal patterns are less structured.

The strong performance of the expectation and mood model reaffirms the central role of these variables in understanding mood changes, even when external routines and activities vary significantly.

The stable performance of the all variables model highlights Gradient Boosting's capability to capture intricate, non-linear relationships among diverse predictors, providing reliable predictions across contexts.

The results highlight the robustness of Gradient Boosting models, particularly when incorporating expectation and mood variables, in explaining mood changes on holidays. While personality traits and all variables models maintain consistent performance, demographic and temporal variables are slightly less predictive, emphasizing the importance of context-specific factors in understanding mood dynamics during holidays.

8. Conclusions

This study investigated the factors influencing daily evaluations, focusing on the roles of morning expectations and mood dynamics, and analyzed how these relationships vary between workdays and holidays. By employing linear regression, Random Forest, and Gradient Boosting models across different variable groupings, the analysis revealed key insights into the determinants of daily evaluations and the predictive power of various methodologies.

Across the full dataset, morning expectations (Expectday) and mood-related metrics, particularly the mean mood, emerged as the most significant predictors of end of the day evaluations. These variables consistently demonstrated strong predictive power across all models, underscoring their central role in shaping daily perceptions. The inclusion of quadratic terms (e.g., Mean $Mood^2$) revealed non-linear relationships, suggesting that higher mood status has a less pronounced increase on the impact on day evaluation.

The expectday variable resulted slightly more relevant the the mood metrics for predicting the end of the day evaluation.

Advanced machine learning models, particularly Gradient Boosting, outperformed non-regularized linear regression and Random Forest models in predictive accuracy, with the all variables model achieving the highest R^2 (0.50). These methods effectively captured complex interactions among predictors, highlighting the importance of leveraging advanced techniques for mood evaluation research.

Ridge, Lasso, and PCA models also displayed a very good performance in predicting the day evaluations, with an accuracy on the level of the best Gradient Boosting model. This highlights the importance of regularization and dimensionality reduction methods in addressing multicollinearity and reducing overfitting for building more precise models.

The comparative analysis of workdays and holidays revealed nuanced differences in model performance.

Linear regression models generally performed better on workdays, particularly the personality traits model, which saw an improvement in R^2 (+0.06). This suggests that personality traits may play a more prominent role in structured, routine-driven contexts. Ridge and Lasso models also demonstrated higher predictive accuracy on workdays, likely due to their ability to emphasize key predictors and handle multicollinearity.

Random Forest and Gradient Boosting models showed slight improvements in R^2 for the expectation and mood model (+0.04 and +0.01, respectively) and the all variables model (+0.03). However, demographic and temporal variables showed a sharp decline in R^2 (-0.32 and -0.16, respectively), indicating their diminished relevance in structured contexts.

Model performance on holidays remained consistent with the general dataset, with no significant improvements or declines. The expectation and mood model maintained its strong R^2 across all methods, reaffirming the importance of these variables in predicting daily evaluations regardless of context. Advanced machine learning models, particularly Gradient Boosting, achieved the highest R^2 (0.50) with the all variables model, demonstrating their robustness even in less structured, unplanned contexts.

Morning expectations and mood metrics consistently emerged as the most critical predictors across all models and contexts. These findings highlight the importance of fostering positive expectations and managing mood fluctuations throughout the day to enhance daily evaluations.

The varying performance of models between workdays and holidays emphasizes the need for tailored approaches in mood prediction. Personality traits appear more relevant on workdays, while mood and expectations remain dominant predictors across both contexts.

Random Forest and Gradient Boosting models demonstrated stable and superior performance across all contexts, particularly when incorporating diverse predictors. These methods are recommended for future research aiming to capture complex, non-linear relationships in mood dynamics.

While less accurate than machine learning models, linear regression methods offer greater interpretability, making them valuable for understanding the relationships between predictors and outcomes. The inclusion of regularization techniques, such as Ridge and Lasso, enhances their utility by addressing multicollinearity and emphasizing key predictors.

Future research directions that could be taken are to investigate interaction effects between mood, expectations, and personality traits to deepen understanding of their combined influence, explore potential biases in data collection, particularly the temporal dominance of afternoon mood metrics in shaping daily evaluations, and validate the findings across diverse populations and contexts to assess their generalizability.

By systematically evaluating the determinants of daily evaluations across different contexts, this study highlights the pivotal roles of mood dynamics and morning expectations.

9. Bibliography

Assi, K., Meegahapola, L., Droz, W., Kun, P., De Götzen, A., Bidoglia, M., ... & Gatica-Perez, D. (2023, April). Complex daily activities, country-level diversity, and smartphone sensing: A study in denmark, italy, mongolia, paraguay, and uk. In Proceedings of the 2023 CHI conference on human factors in computing systems (pp. 1-23).

Canzian, L., & Musolesi, M. (2015, September). Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing (pp. 1293-1304).

LiKamWa, R., Liu, Y., Lane, N. D., & Zhong, L. (2013, June). Moodscope: Building a mood sensor from smartphone usage patterns. In Proceeding of the 11th annual international conference on Mobile systems, applications, and services (pp. 389-402).

Meegahapola, L., Droz, W., Kun, P., De Götzen, A., Nutakki, C., Diwakar, S., ... & Gatica-Perez, D. (2023). Generalization and personalization of mobile sensing-based mood inference models: an analysis of college students in eight countries. Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies, 6(4), 1-32.

Rachuri, K. K., Musolesi, M., Mascolo, C., Rentfrow, P. J., Longworth, C., & Aucinas, A. (2010, September). EmotionSense: a mobile phones based adaptive platform for experimental social psychology research. In Proceedings of the 12th ACM international conference on Ubiquitous computing (pp. 281-290).

Sano, A., & Picard, R. W. (2013, September). Stress recognition using wearable sensors and mobile phones. In 2013 Humaine association conference on affective computing and intelligent interaction (pp. 671-676). IEEE.

Servia-Rodríguez, S., Rachuri, K. K., Mascolo, C., Rentfrow, P. J., Lathia, N., & Sandstrom, G. M. (2017, April). Mobile sensing at the service of mental well-being: a large-scale longitudinal study. In Proceedings of the 26th international conference on world wide web (pp. 103-112).

Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., ... & Campbell, A. T. (2014, September). StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing (pp. 3-14).

The code used for this analysis is available on:
https://github.com/TommasoGrotto2/SDL