IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Ingénierie des Systèmes d'Information
Vol. 29, No. 6, December, 2024, pp. 2345-2356
Journal homepage: http://iieta.org/journals/isi

Supervised Learning for Emotional Prediction and Feature Importance Analysis Using SHAP on Social Media User Data

Erna Hikmawati[1*], Nur Alamsyah[2]

[1] School of Applied Science, Telkom University, Bandung 40257, Indonesia
[2] Information Systems Study Program, Faculty of Technology & Informatics, University of Informatics and Business Indonesia, Bandung 40285, Indonesia

Corresponding Author Email: ernahikmawati@telkomuniversity.ac.id

https://doi.org/10.18280/isi.290622

Received: 24 June 2024
Revised: 8 November 2024
Accepted: 21 November 2024
Available online: 25 December 2024

Keywords:
emotional prediction, supervised learning, random forest, SHAP analysis, social media data

ABSTRACT

This study aimed to predict emotional states using supervised learning models and analyze the importance of features in social media user data. We implemented the Random Forest algorithm to predict happiness, neutrality, and sadness based on various social media activity metrics, including daily usage time, posts per day, and interactions such as likes and comments received. Data preprocessing involved handling missing values, coding categorical features using One-Hot Encoding, and scaling numerical features with StandardScaler. We assessed the model's performance utilizing Mean Squared Error (MSE) and R-squared (R^2) measures. The results showed that the model had a high prediction accuracy, with R^2 values of 0.997 for happiness, 0.863 for neutrality, and 0.851 for sadness. SHapley Additive exPlanations (SHAP) were used to perform a thorough feature importance analysis, which revealed that daily usage time and user interaction significantly influenced emotional states. These findings underscore the efficacy of combining supervised learning with SHAP for interpretable and accurate emotional predictions, providing valuable insights for the development of tools and strategies to monitor and enhance emotional health in the digital era.

1. INTRODUCTION

Social media's explosive growth has changed how people engage, communicate, and exchange information [1]. Social media sites like Facebook, Instagram, and Twitter have become a vital part of daily life because they allow people to express their ideas, feelings, and experiences [2]. However, these digital engagements also impact emotional well-being, both positively and negatively. Understanding these impacts is critical to developing tools and strategies to improve emotional health in the digital age [3].

Prior studies have examined multiple facets of social media's impact on emotional conditions. For example, Bohnert and Gracia [4] found that increased Facebook use was associated with decreased subjective well-being. Similarly, Brailovskaia and Margraf [5] highlighted that passive use of Facebook can lead to decreased mood. On the other hand, positive interactions and social support on this platform have been associated with increased emotional well-being [6].

In a machine learning context, supervised learning models such as Random Forest have shown a great potential in predicting emotional states from social media data [7]. Random Forest is preferred for its robustness and readability compared to other models [8]. But one of the biggest obstacles to applying machine learning models is realizing how different features affect the outcome of the predictions [9-11]. In order to improve the model's comprehensibility, SHAP was developed as a solution that offers a consistent assessment of the relevance of characteristics [12]. Some other literature also supports the importance of this analysis. For example, research conducted by Roy et al. [13] showed how sentiment analysis using machine learning can predict the stress level of social media users. Meanwhile, research by Peng et al. [14] used deep learning models to identify users' emotions based on texts posted on social media, showing that features such as posting intensity and content type can influence the prediction results. Furthermore, the research by Oliveira et al. [15] investigated the use of ensemble models to social media users' happiness prediction, highlighting the importance of strong models in the processing of diverse and complicated data.

Study by Uban et al. [16] emphasized the importance of temporal analysis in predicting emotional states, by showing that time patterns in social media activity can be an important indicator of mood changes. In addition, Chancellor and De Choudhury [17] demonstrated that users' mental health, including anxiety and depression, may be predicted using linguistic information gleaned from social media posts. The research underscores the significance of various data kinds and analytical methods in comprehending the impact of social media on mental health.

This research intends to close this gap by combining SHAP with supervised learning approaches to examine the significance of variables in social media user data and forecast emotional states. The main contributions of this research are

as follows: First, we apply Random Forest algorithm to predict happiness, neutrality, and sadness based on social media activity metrics. Second, we use SHAP to conduct a thorough feature importance analysis, providing insights into the factors that most significantly influence emotional states. Third, we detail a comprehensive data pre-processing pipeline, including missing value handling, categorical feature coding, and numerical feature scaling. Fourth, we employed the MSE and R^2 metrics to rigorously evaluate the model's performance, and the findings indicated exceptional predictive accuracy. Finally, our findings provide valuable insights for developing tools and strategies to monitor and improve emotional health in the digital age.

2. LITERATURE REVIEW

2.1 Evolution of emotion prediction methods

Recent years have seen an increasing interest among researchers in the impact of social media on emotional states. The study by Wirtz et al. [18] found that an increase in Facebook use was related to a decrease in subjective well-being. Users who spent more time on Facebook reported feeling worse over time. The study by Karsay et al. [19] showed that passive use of Facebook, such as browsing posts without interacting, can lead to a decrease in mood. In contrast, Lin and Kishore [20] emphasized the positive impact of social support and meaningful interactions on emotional well-being, suggesting that active and supportive engagement can improve emotional health.

Recent studies have utilized diverse machine learning and deep learning models to forecast emotional states. The Random Forest model has been widely employed to forecast emotional states based on social media data [21]. Random Forest as a powerful and reliable model because it is able to overcome overfitting and provides good interpretation through the importance of features [22]. Braig et al. [23] showed how sentiment analysis using machine learning can predict the stress level of social media users, using various linguistic and behavioral features. Deep learning model to identify users' emotions based on text posted on social media [24]. They found that features such as posting frequency, content type, and social interactions can affect emotion prediction results. The study by Mukta et al. [25] explored the use of ensemble models to predict the happiness of social media users, emphasizing the importance of robust models in the analysis of complex and varied data. Study by Meena et al. [26] used convolutional neural networks for sentiment analysis on Twitter, showing that deep learning-based approaches can produce very accurate predictions.

2.2 Feature importance analysis methods

Interpreting model predictions is crucial for understanding the underlying factors that influence emotional states. SHAP offers a comprehensive metric for assessing feature significance, allowing researchers to comprehend each feature's contribution to model predictions [27]. SHAP has been utilized widely to understand the contribution of each feature to model predictions, making it an essential tool in analyzing the impact of social media metrics on emotional states. In the context of sentiment analysis, VADER has been used in many studies to measure sentiment in social media texts with high accuracy [28]. Feature importance methods such as SHAP enable researchers to identify which social media activities, such as posting frequency and content interactions, are the most influential in predicting user emotions.

2.3 Social media user emotion analysis

The relationship between social media interactions and user emotions is multifaceted. Study by Jordan et al. [29] emphasize the importance of temporal analysis in predicting emotional conditions, showing that time patterns in social media activity can be an important indicator of mood change. The study by Gao et al. [30] explores the influence of different types of social media interactions on emotional well-being. They found that the type and quality of interactions, not just the quantity, had a significantly impact on users' emotional state. Xu et al. [31] examined how group dynamics in social media can affect individuals' moods, suggesting that social context and social networks play an important role in users' emotional experiences.

The study aimed to fill the gap by combining supervised learning techniques with SHAP to predict emotional states and analyze the importance of features in social media user data. We apply Random Forest algorithm to predict happiness, neutrality, and sadness based on social media activity metrics, and use SHAP to perform a thorough feature importance analysis. Our findings are expected to provide valuable insights for developing tools and strategies to monitor and improve emotional health in the digital age. Our study builds on this body of work by combining supervised learning techniques, specifically Random Forest, with SHAP analysis to offer a comprehensive understanding of how social media activities influence emotional states. This research aims to fill existing gaps by applying the Random Forest algorithm to predict happiness, neutrality, and sadness, and utilizing SHAP for feature importance analysis. Our findings are expected to contribute valuable insights for developing strategies to monitor and improve emotional health in the digital era.

3. MATERIAL AND METHOD

We outline the data sources, pre-processing steps, machine learning algorithms applied, as well as the evaluation methods used to measure model performance. In addition, we also describe the use of SHAP for model interpretability analysis. Our proposed method is presented in Figure 1.

3.1 Data collected

The dataset used in the study consists of three main files collected from the Kaggle portal: train.csv, val.csv, and test.csv. The dataset comprises a total of 10,000 samples, split into 7,000 samples for training, 1,500 samples for validation, and 1,500 samples for testing. Each sample represents a user's social media activity, captured through several features, such as Daily Usage Time, Comments Received Per Day, Messages Sent Per Day, and platform-specific interactions. The data was collected over a period of six months, ensuring a variety of social media usage patterns. The emotional categories predicted in this study include happiness, neutrality, and sadness. The distribution of these categories is as follows: 40% happiness, 35% neutrality, and 25% sadness, indicating a

reasonable representation of different emotional states. A description of these dataset features is provided in Table 1, outlining the variables used for emotion prediction and their relevance in capturing user behavior and emotional well-being.

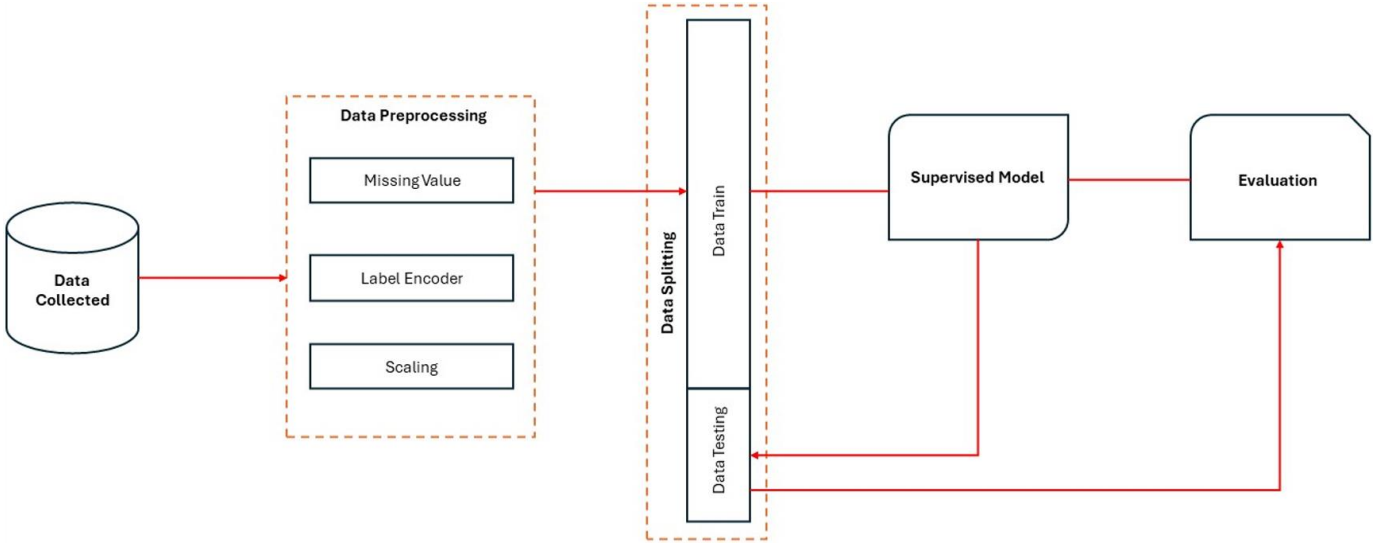

Figure 1. Proposed method

Table 1. Description of dataset

Feature Name	Description
User_ID	Unique ID of the user
Daily_Usage_Time	Daily social media usage time in minutes
Posts_Per_Day	Number of posts per day
Likes_Received_Per_Day	Number of likes received per day
Comments_Received_Per_Day	Number of comments received per day
Messages_Sent_Per_Day	Number of messages sent per day
Platform_Facebook	Indicator if the user uses Facebook
Platform_Twitter	Indicator if the user uses Twitter
Platform_Instagram	Indicator if the user uses Instagram
Platform_Whatsapp	Indicator if the user uses Whatsapp
Dominant_Emotion_Happiness	Indicator of the user's dominant emotion of happiness
Dominant_Emotion_Neutral	Indicator of the user's dominant emotion of neutrality
Dominant_Emotion_Sadness	Indicator of the user's dominant emotion of sadness
Dominant_Emotion_Anger	Indicator of the user's dominant emotion of anger
Dominant_Emotion_Anxiety	Indicator of the user's dominant emotion of anxiety
Dominant_Emotion_Boredom	Indicator of the user's dominant emotion of boredom

Train Dataset (train.csv), Validation Dataset (val.csv), and Test Dataset (test.csv). The majority of the data employed in the machine learning process is contained in the Train Dataset, which is utilized to train the predictive model. The Validation Dataset is employed to verify the model during training, preventing overfitting and ensuring its efficacy on unseen data. The model's ultimate performance is assessed utilizing the Test Dataset. This data offers an unbiased assessment of the model's performance and is not utilized in the training or validation phases.

3.2 Data preprocessing

Data preparation is an essential phase in data analysis and machine learning that guarantees data quality and uniformity before it is applied to the model [32]. The pd.read_csv function with error handling is used to load the dataset from a CSV file in the first step, addressing troublesome rows. Once the dataset is loaded, the next step is to identify categorical and numerical columns. For numerical columns, we employed three methods for handling missing values: mean imputation, K-Nearest Neighbors (KNN) Imputation, and Iterative Imputation. KNN Imputation estimates missing values based on the similarity to other data points, while Iterative Imputation models each missing value as a function of the other features in an iterative process. While for categorical columns, the missing value is filled with the mode of the column. Next, the categorical features are converted into numerical representations using the One-Hot Encoding technique with the sklearn.preprocessing.OneHotEncoder library. Next, to make sure that every feature has the same scale, the numerical features are scaled using StandardScaler from the sklearn library. The preprocessed data is then saved back into a CSV file for use in the next stage of model training.

3.3 Data splitting

The data splitting process is very important in machine learning to ensure that the built model can be well evaluated and has good generalization to data that has never been seen before [33]. The dataset used in the study is divided into three primary categories: testing, validation, and training data. The

training data is used to train the prediction model, which helps the model identify patterns in the data. Validation data is used throughout the training process to verify the model's performance and avoid overfitting, which occurs when the model fits the training data too well and performs poorly on new data. We can tweak the model's parameters to enhance overall performance by keeping an eye on how the model performs using the validation data. After training is finished, test data is utilized to evaluate the model's overall performance. This data offers an unbiased assessment of the model's capacity to predict entirely fresh data and is not used in the training or validation phases of the process. We can make sure that the developed model has strong generalization capabilities and is dependable in real-world scenarios by segmenting the dataset into training, validation, and testing data.

3.4 Supervised model

In the study, the model used to predict the emotional states of social media users is the Random Forest Regressor. With many separate decision trees cooperating as a group, Random Forest is a potent and adaptable machine learning system. Every tree in the forest makes a prediction, and the total result is calculated by summing together each prediction. This model is selected due to its robustness while handling a variety of connected and different input variables, as well as its capacity to handle vast and complex datasets. To address the key parameters and tuning process of the Random Forest model in this study, we provide an explanation of each primary parameter and the methodology used for optimal selection. Random Forest is an ensemble method that constructs multiple decision trees and aggregates their outputs, enhancing predictive accuracy and robustness. The performance of Random Forest depends significantly on several key parameters: (1) Number of Trees (n_estimators), which defines the number of decision trees in the forest. Generally, a higher number of trees can reduce variance, enhancing accuracy, but it also increases computational cost. (2) Maximum Depth of Trees (max_depth), which limits how deep each tree grows. Setting an appropriate depth helps balance complexity and overfitting, with deeper trees often learning more complex patterns, though at risk of overfitting smaller datasets. (3) Minimum Samples per Leaf (min_samples_leaf), which determines the minimum number of samples at a leaf node, aiding in noise reduction by preventing overly specific splits. (4) Minimum Samples per Split (min_samples_split), controlling the number of samples needed to split an internal node, which helps prevent highly specific rules and further reduces overfitting. (5) Maximum Features (max_features), limiting the number of features considered at each split. This introduces additional randomness, reducing variance and improving generalization.

To select the optimal values for these parameters, we employed a Randomized Search Cross-Validation approach, balancing computational efficiency with solution quality. Randomized search, preferred over grid search, enabled us to explore a broader parameter range. The ranges defined for tuning included values for n_estimators (e.g., 50 to 500), max_depth (e.g., 10 to None), min_samples_split (e.g., 2 to 15), min_samples_leaf (e.g., 1 to 6), and max_features (e.g., ‘auto’, ‘sqrt’, ‘log2’). Accuracy served as the primary evaluation metric during tuning, aligning with the study’s objectives, and cross-validation ensured robust performance across different folds, preventing overfitting. The best-performing parameters were selected based on the highest cross-validated accuracy score, yielding an optimal configuration used in the final Random Forest model. This tuning approach delivered a balanced model that minimizes overfitting while ensuring strong predictive accuracy and robustness on unseen data.

To strengthen the model selection, we have also included a comparison with another mainstream algorithm, XGBoost. XGBoost is a gradient boosting algorithm known for its efficiency and high performance on structured data. It builds trees sequentially, where each new tree attempts to correct the errors of the previous ones, resulting in a more refined prediction model.

The first step in building a model is extracting training and testing subsets from the pre-processed dataset. Through the use of the training dataset, the model learns to recognize patterns and correlations in the data. To optimize the model's performance during training, hyperparameters influencing the number of trees in the forest and the trees' maximum depth are changed. The equation for prediction in Random Forest can be expressed as Eq. (1).

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} \hat{y}_i \tag{1}$$

where, $\hat{y}$ is the final predicted value, N is the total number of trees in the forest and $\hat{y}_i$ is the prediction of the ith tree. Once the model is trained, the performance of the model is evaluated using the test dataset. This evaluation is done by calculating performance metrics such as MSE and R^2. The equation for MSE should be as Eq. (2).

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{2}$$

where, n is the total number of samples, $y_i$is the actual value of the i sample and $\hat{y}_i$ is the predicted value of the i sample. R^2 measures the proportion of variance in the data that is explained by the model, and is given by Eq. (3).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3}$$

where, y_i is the actual value of the ith sample, $\hat{y}_i$ is the predicted value of the ith sample and $\bar{y}$ is the average of the actual values.

Models with high $R^2$and low MSE are regarded as performing well. We employed SHAP to offer a more thorough interpretation and comprehend the contribution of each feature to the model prediction. Researchers can comprehend the impact of each feature on the prediction outcomes by using SHAP, a technique that offers a standardized and uniform measure of feature relevance. With this approach, the Random Forest model is not only able to predict emotional states with high accuracy but also provides valuable insights into the factors that influence the emotional states of social media users.

3.5 Evaluation

The evaluation of the model's performance is a crucial step

in assessing its accuracy and reliability in predicting the emotional states of social media users [34]. To achieve this, we use several key metrics: MSE and R^2. The average squared difference between the expected and actual numbers is measured by the Mean Squared Error or MSE. Better model performance is indicated by a smaller MSE, which shows that the predictions are more accurate. R^2 shows the percentage of the dependent variable's volatility that can be predicted based on the independent variables. R^2 value closer to 1 signifies that a large proportion of the variance in the dependent variable has been explained by the model, indicating a high level of predictive accuracy.

In addition to these metrics, visualizations such as scatter plots of actual versus predicted values and bar charts displaying MSE and R^2 values for each target emotion are employed to provide a more intuitive understanding of the model's performance. Scatter plots help identify patterns of overfitting or underfitting, while bar charts summarize the evaluation metrics in a clear and concise manner. Overall, the combination of these quantitative metrics and visual tools provides a comprehensive evaluation of the model's effectiveness in predicting emotional states, ensuring that the model is both accurate and interpretable.

4. RESULTS AND DISCUSSION

MSE and R^2 are included in the assessment for the anticipated emotional states of neutrality, sadness, and happiness. Additionally, visualizations such as scatter plots of actual versus predicted values are provided to illustrate the model's performance. To gain deeper insights into the feature contributions, SHAP summary plots and SHAP feature importance plots are also presented for each dominant emotion. The significance of several variables in predicting the emotional states of social media users is examined, along with the consequences of the model's predictions, based on an analysis and discussion of the data.

4.1 Results comparison for missing value imputation methods

In this section, we compare the performance of three missing value imputation techniques: mean/mode imputation, KNN Imputation, and Iterative Imputation. The results for each method are presented for three target variables: Dominant_Emotion_Happiness, Dominant_Emotion_Neutral, and Dominant_Emotion_Sadness.

Across all three methods, the RMSE (as represented by the MSE values) and R^2 scores for each target variable remain consistent, it can be seen in Figure 2 and Figure 3. This indicates that all three imputation techniques produce similar predictive accuracy for this particular dataset, with only negligible differences observed in the model's performance.

This similarity suggests that, in this case, the choice of imputation method does not significantly impact the model's performance. However, advanced methods like KNN and Iterative Imputation might still be preferable in datasets with more complex missing data patterns, as they leverage relationships within the data.

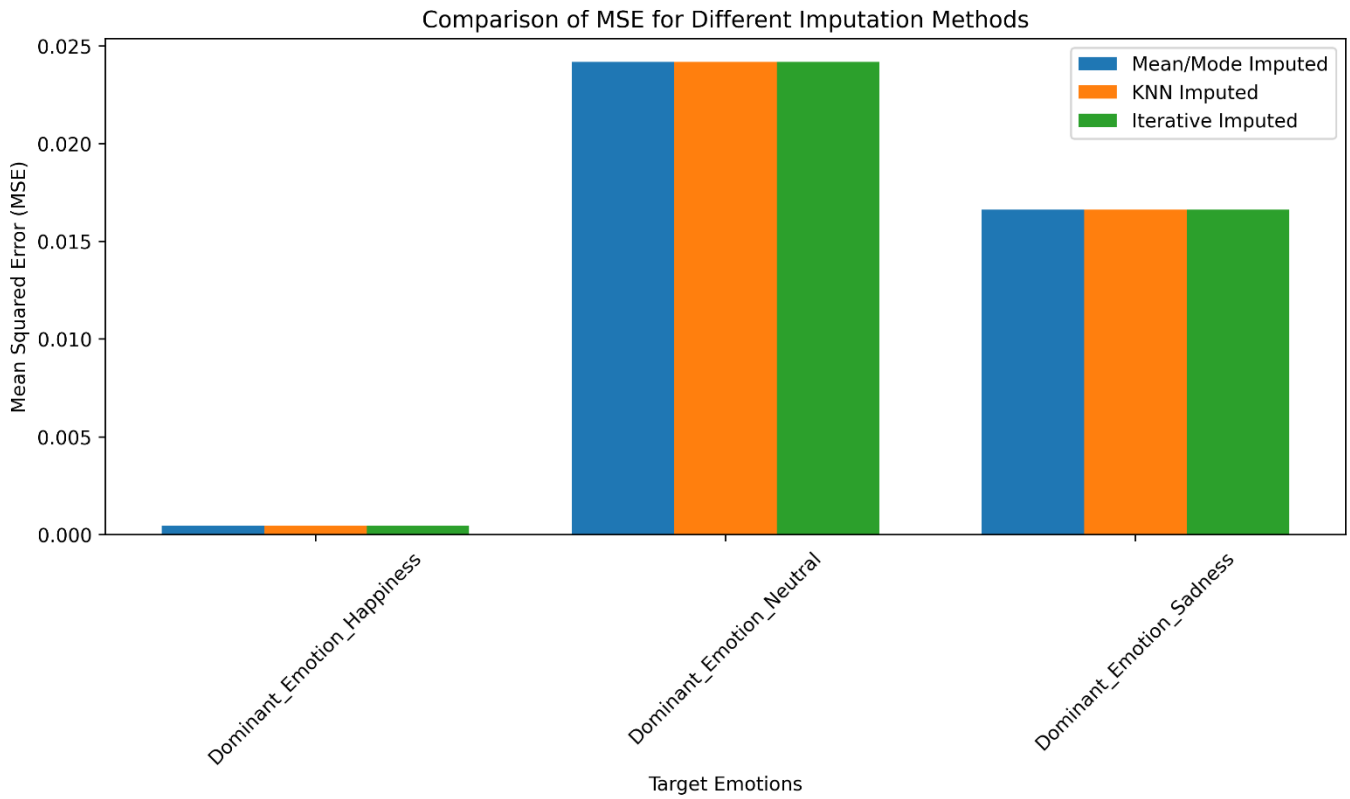

Figure 2. Comparison MSE for different imputation methods

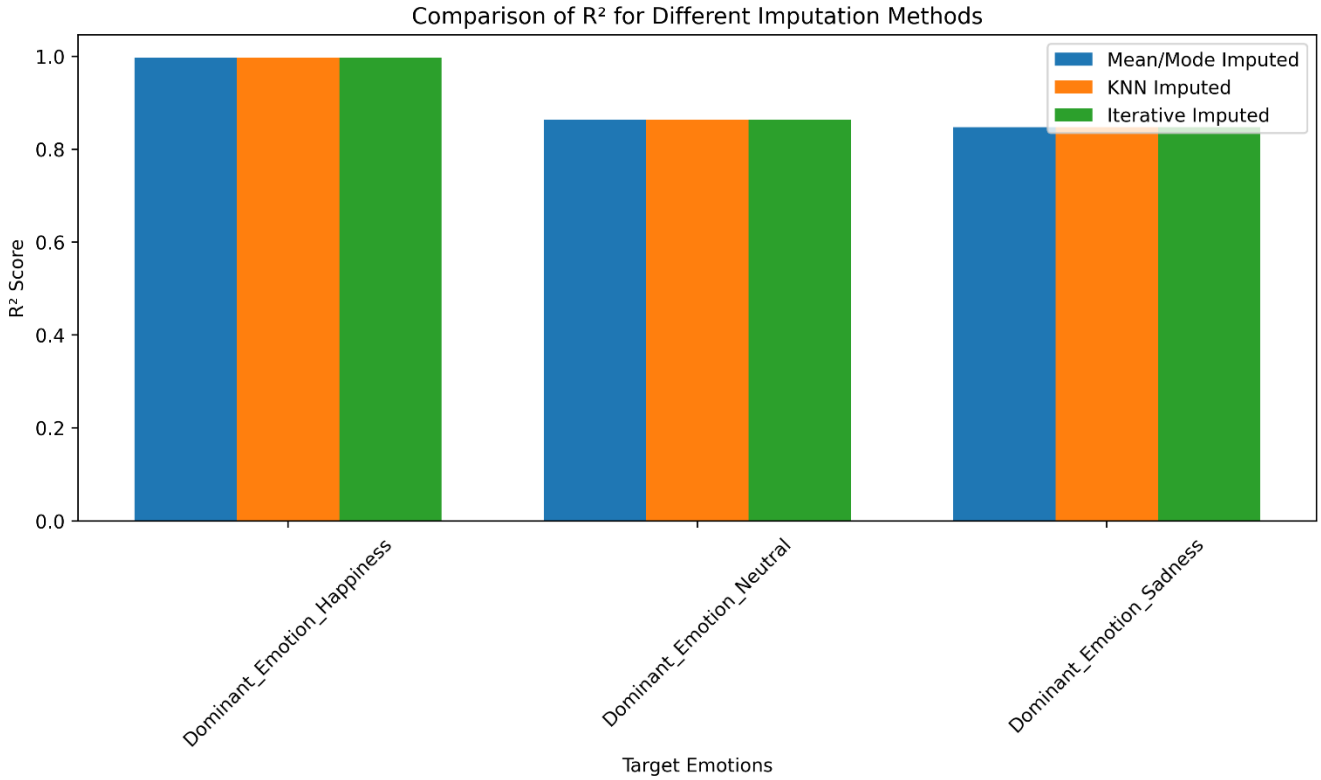

Figure 3. Comparison R^2 for different imputation methods

4.2 Results for random forest regressor

The analysis of the model's performance, as shown in the bar charts, shows how well the Random Forest Regressor predicts the emotions of neutrality, happy, and sorrow. The MSE evaluation of the model, as presented in Figure 4, quantifies the prediction errors by measuring the average squared difference between predicted and actual values. Lower MSE values indicate higher accuracy.

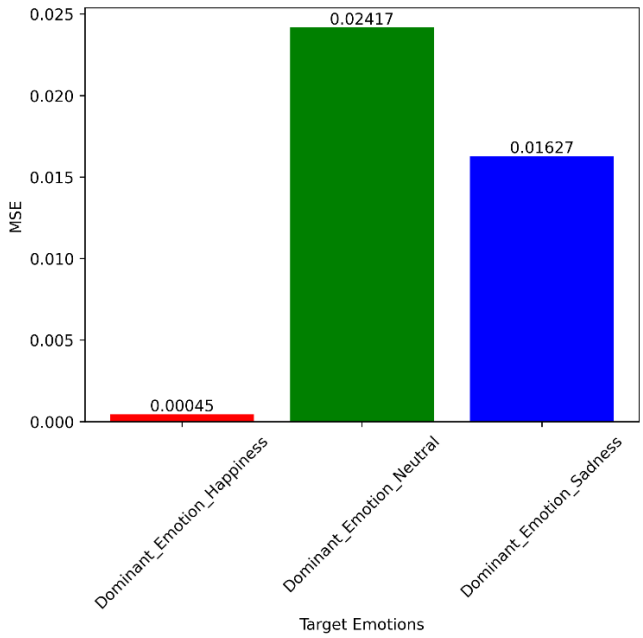

Figure 4. The MSE evaluation

For happiness, the model achieves an exceptionally low MSE of 0.00045, reflecting very high accuracy. In contrast, the neutrality emotion has an MSE of 0.02417, indicating relatively larger prediction errors. The MSE for sadness is 0.01627, reflecting moderate accuracy.

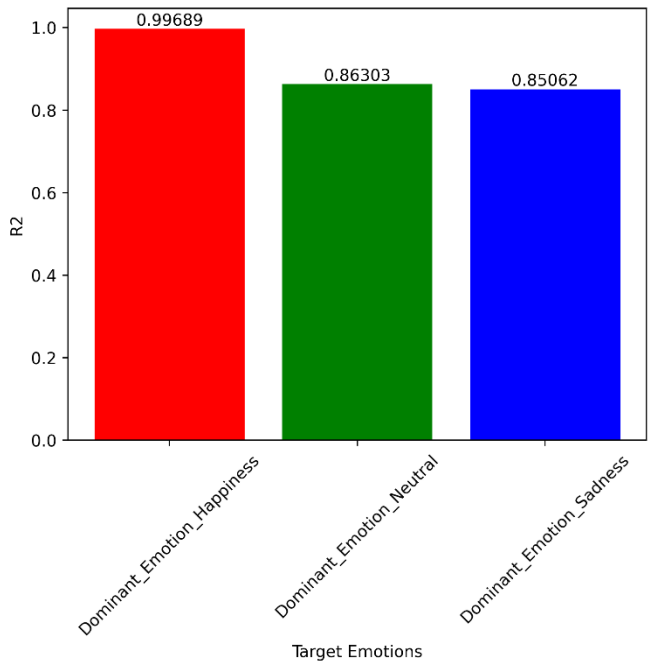

Figure 5. The R^2 evaluation

MSE values reveal the accuracy of the predictions by measuring the average squared differences between the predicted and actual values. A lower MSE indicates more accurate predictions. For the emotion of happiness, the model achieves an exceptionally low MSE of 0.00045, indicating

very high accuracy. In contrast, the MSE for neutrality is higher at 0.02417, suggesting that predictions for this emotion are less accurate compared to happiness. The sadness emotion has an MSE of 0.01627, reflecting moderate accuracy that is better than neutrality but not as high as happiness.

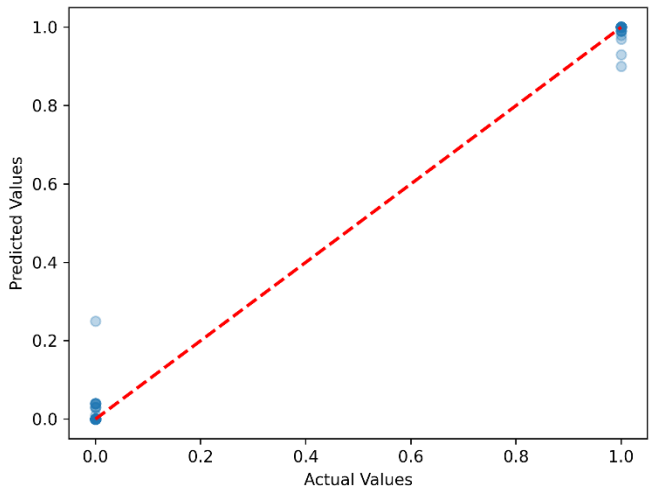

Figure 6. Actual vs Predicted values for dominant emotion happiness

The percentage of variance in the actual values that can be predicted from the independent variables is shown by R^2 values, which provide additional insight into the model's performance. The R^2 of our model is presented in Figure 5. An R^2 value closer to 1 signifies a better model fit. The model exhibits an R^2 value of 0.99689 for happiness, indicating that nearly all the variance in the happiness data is explained by the model. For neutrality, the R^2 value is 0.86303, showing strong predictive power, although not as high as for happiness. The R^2 value for sadness is 0.85062, slightly lower than neutrality but still representing a good level of variance explained by the model.

We also evaluate the results of the prediction model by using a scatter plot to compare the actual and predicted values of each dominant emotion, namely happiness, neutrality, and sadness. This scatter plot helps visualize the performance of the Random Forest Regressor in predicting these emotions based on social media usage data. The first scatter plot showing actual versus predicted values for the dominant emotion happiness is presented in Figure 6.

R^2 value is 0.99689, while the mean squared error (MSE) is 0.00045. The ideal situation, in which the projected values precisely match the actual values, is represented by the red dashed line. The points closely follow the diagonal line, indicating that the model has made highly accurate predictions for the emotion of happiness, with very low prediction errors.

The second scatter plot depicts the actual versus predicted values for the dominant emotion of neutrality is presented in Figure 7. The MSE is 0.02417, and the R^2 value is 0.86303. The points are more dispersed compared to the happiness plot, indicating larger prediction errors.

However, the overall trend still follows the diagonal line, suggesting that the model has a good predictive power for neutrality, though not as high as for happiness. The third scatter plot shows the actual versus predicted values for the dominant emotion of sadness is presented in Figure 8.

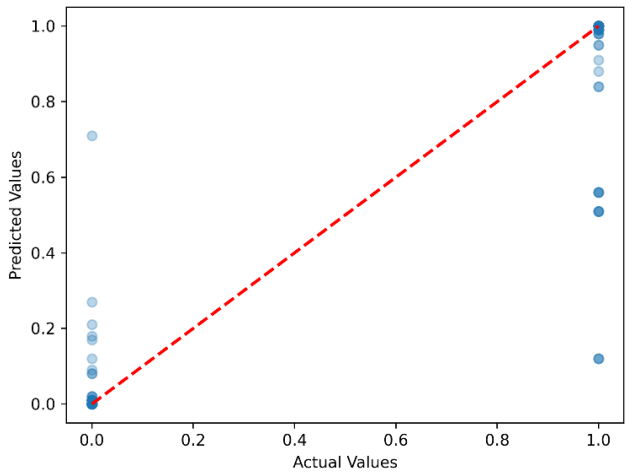

Figure 7. Actual vs Predicted values for dominant emotion neutral

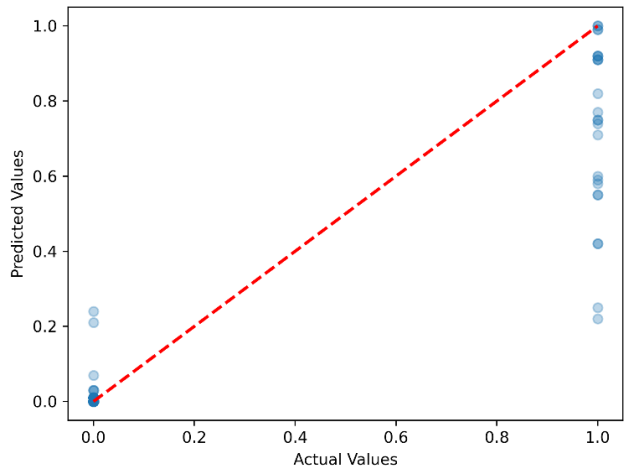

Figure 8. Actual vs Predicted values for dominant emotion sadness

The MSE is 0.01627, and the R^2 value is 0.85062. Similar to the neutrality plot, the points are more scattered around the diagonal line, indicating moderate prediction errors. The model performs reasonably well in predicting sadness, but the accuracy is lower compared to happiness. Overall, the scatter plots indicate that the model performs best for predicting happiness, followed by neutrality and sadness. The close alignment of points along the diagonal line for happiness shows that the model has high accuracy for this emotion, while the increased dispersion for neutrality and sadness reflects relatively higher prediction errors.

To gain a deeper understanding of feature contributions and interactions, we analyze the SHAP Feature Importance and SHAP Summary Plots for each dominant emotion. Figure 9 displays the SHAP Feature Importance for happiness, showing the mean absolute SHAP values for each feature. The most influential feature is Daily Usage Time (minutes), followed by Comments Received Per Day and Messages Sent Per Day. This suggests that time spent on social media and social interactions through comments and messages are key indicators of happiness. Other influential features include Dominant Emotion Anxiety, Dominant Emotion Anger, and Posts Per Day.

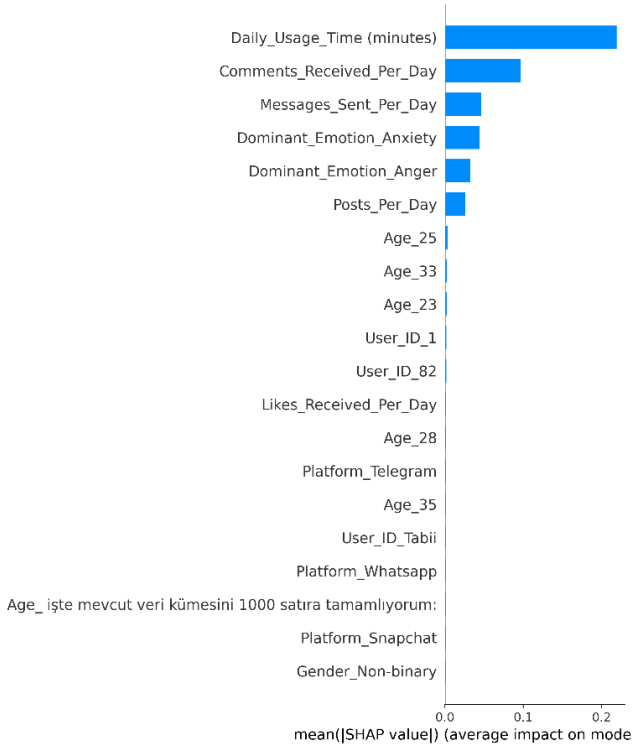

Figure 9. SHAP feature importance for dominant emotion happiness

The features have been arranged in order of significance. The most influential feature is the Daily Usage Time (minutes), followed by Comments Received Per Day and Messages Sent Per Day. This suggests that social media usage and the relationships that occur through messages and comments are important indicators of pleasure. Other notable features include Dominant Emotion Anxiety, Dominant Emotion Anger, and the number of Posts Per Day.

In the SHAP Summary Plot for happiness (Figure 10), each dot represents a Shapley value for a feature in a specific instance. The color gradient from red to blue indicates high to low feature values. A broader spread of SHAP values for Daily Usage Time suggests a strong influence on happiness prediction. Higher daily usage is associated with positive SHAP values, implying that increased time on social media positively correlates with happiness. Moreover, we observe interactions between Comments Received Per Day and Messages Sent Per Day, where higher values of both features contribute positively to happiness, highlighting the role of social engagement.

Every dot in the dataset represents a single instance of a feature's Shapley value. The color of the dots indicates the feature value, with red representing high values and blue representing low values. The horizontal spread of the dots shows the range of SHAP values for that feature. For example, Daily Usage Time (minutes) has a wide spread, indicating a strong influence on the prediction. High values of Daily Usage Time are associated with higher SHAP values, suggesting that more time spent on social media is positively correlated with happiness. Similarly, high values of Comments Received Per Day and Messages Sent Per Day also show a positive impact on the predicted happiness.

For neutrality, Figure 11 presents SHAP Feature Importance, where Comments Received Per Day has the highest influence, followed by Dominant Emotion Boredom and Dominant Emotion Anxiety. This indicates that neutral emotions are significantly influenced by social feedback and boredom. Figure 12 shows the SHAP Summary Plot for neutrality, revealing interactions between Comments Received Per Day and other emotional indicators like boredom and anxiety. High comment counts are associated with higher SHAP values, suggesting that neutral emotions correlate with frequent social interactions. Furthermore, higher values for boredom and anxiety are linked to neutrality, indicating that certain emotional states (like boredom and anxiety) are predictive of a neutral stance on social media.

The features are ranked based on their importance. The most influential feature is Comments Received Per Day, followed by Dominant Emotion Boredom and Dominant Emotion Anxiety. This indicates that the number of comments received and the levels of boredom and anxiety significantly impact the prediction of the neutral emotion. Other notable features include Daily Usage Time (minutes), Dominant Emotion Anger, and the number of Posts Per Day.

Next, we evaluated Summarry Plot for Dominant Emotion Neutral. The visualization presented in Figure 10. For each dot represents a Shapley value for a feature and a single instance in the dataset. The color of the dots indicates the feature value, with red representing high values and blue representing low values. The horizontal spread of the dots shows the range of SHAP values for that feature. For example, Comments Received Per Day has a wide spread, indicating a strong influence on the prediction. High values of Comments Received Per Day are associated with higher SHAP values, suggesting that receiving more comments is positively correlated with neutral emotion. Similarly, high values of Dominant Emotion Boredom and Dominant Emotion Anxiety also show a significant impact on the predicted neutrality.

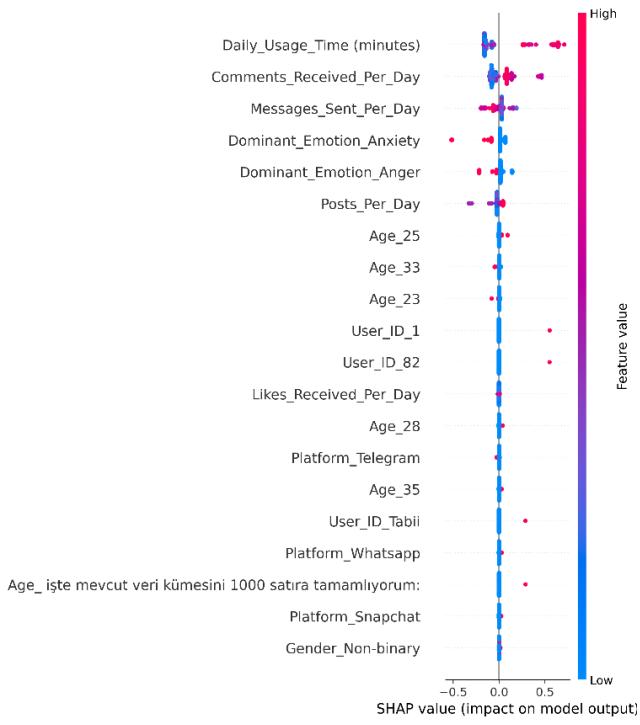

Figure 10. SHAP summary plot for dominant emotion happiness

For sadness, the SHAP Feature Importance in Figure 13 ranks Comments Received Per Day as the most impactful feature, followed by Daily Usage Time and Snapchat Platform. This suggests that social media interaction and

platform-specific use are significant predictors of sadness. Figure 14, the SHAP Summary Plot for sadness, illustrates the interaction between Comments Received Per Day and Daily Usage Time. Higher values of these features tend to increase the SHAP value for sadness, signifying a positive correlation with predicted sadness. Notably, instances with higher Snapchat Platform usage display elevated sadness predictions, highlighting platform-specific behaviors influencing emotional outcomes.

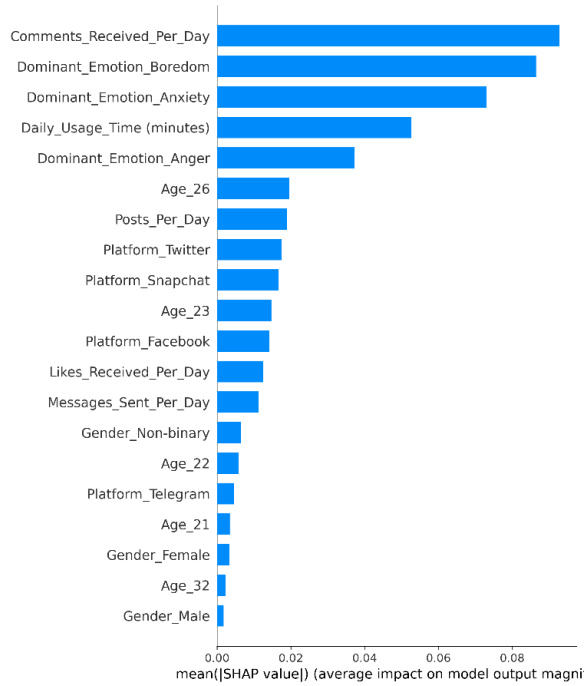

Figure 11. SHAP feature importance for dominant emotion neutral

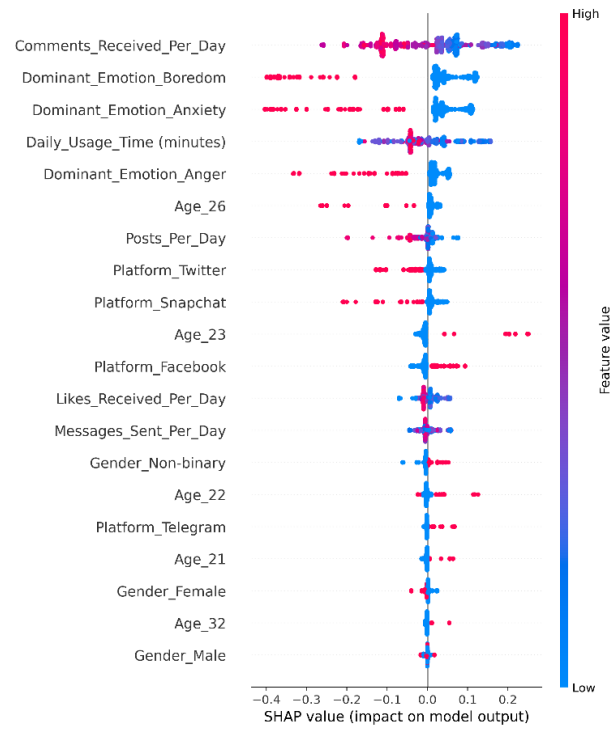

Figure 12. SHAP summary plot for dominant emotion neutral

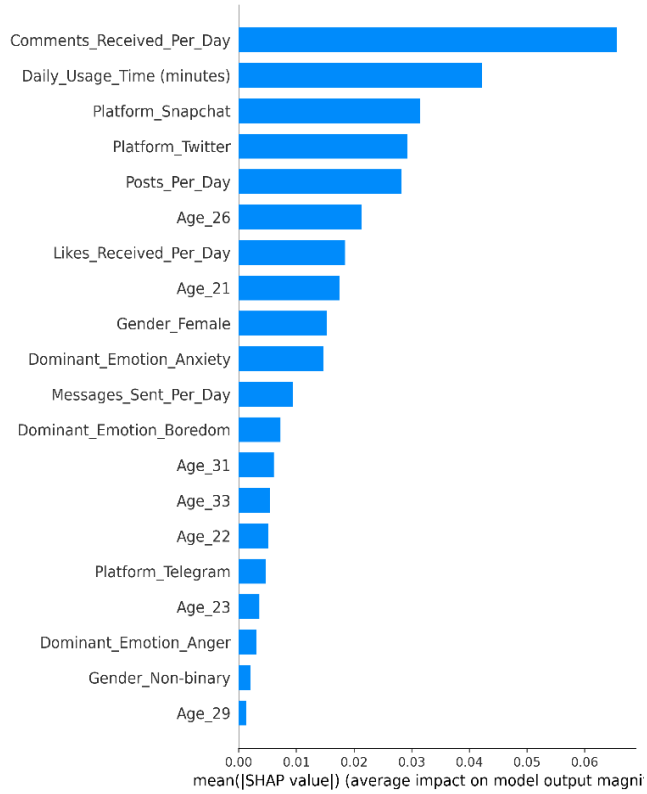

Figure 13. SHAP summary plot for dominant emotion sadness

The quantity of comments received daily (Comments Received Per Day) was the most important factor, followed by the amount of time spent using Snapchat each day (Daily Usage Time in Minutes) and the amount of time spent on the platform (Snapchat Platform). This shows that the number of comments received, time spent on social media, and Snapchat usage are significant predictors of sadness emotions. Other important features included the use of the Twitter platform (Twitter Platform), the number of posts per day (Posts Per Day), and age 26.

The following SHAP Feature Importance for dominant emotion sadness is presented in Figure 14.

Every dot in the dataset represents the Shapley value for a particular feature occurrence. The value of the characteristic is represented by the color of the dot, where red indicates high values and blue indicates low values. The horizontal spread of the dots indicates the range of SHAP values for the feature. For instance, there is a significant variation in the quantity of comments received each day, indicating a significant impact on forecast. Higher SHAP values are connected with higher numbers of comments received daily, suggesting a positive relationship between the quantity of comments received and the feeling of melancholy. Similarly, high values of daily usage time and Snapchat platform usage also showed a significant impact on grief prediction. Overall, the SHAP analysis reveals that happiness is most strongly associated with active social engagement (comments and messages), while neutrality and sadness have complex interactions with features like boredom, anxiety, and platform-specific usage. These findings emphasize the importance of social engagement and platform choice in predicting emotions, with nuanced interactions that enhance our understanding of social media behaviors and emotional states.

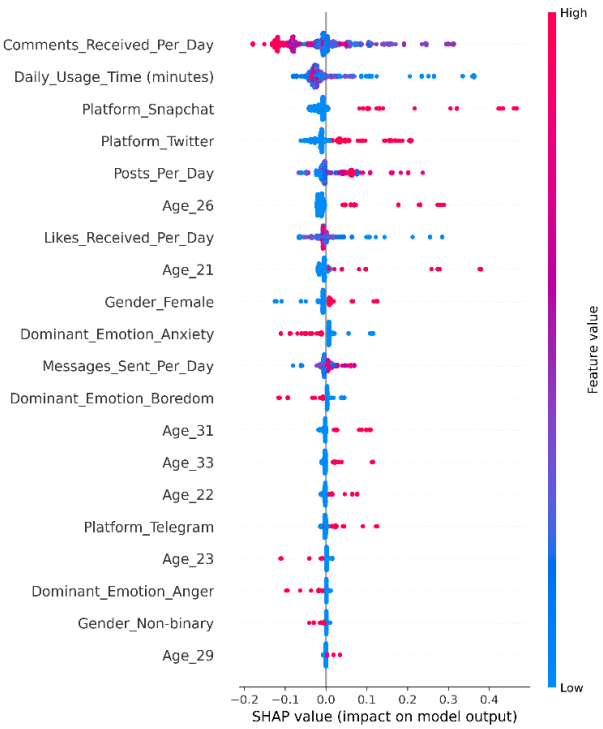

Figure 14. SHAP summary plot for dominant emotion sadness

4.3 Results for XGBoost

XGBoost achieved an overall accuracy of 99.05%, demonstrating strong predictive performance. The classification report provides detailed metrics for each class, including precision, recall, and F1-score, which offer insights into how well the model performs across different target classes.

For each class (0 through 5), the precision, recall, and F1-scores are close to 1.0, indicating near-perfect performance in identifying each category correctly. This suggests that XGBoost is effective in distinguishing between different emotional states with minimal error. The macro and weighted averages for precision, recall, and F1-score are all 0.99, which further emphasizes the model's reliability across all classes.

In comparison to the initial model (Random Forest Regressor), XGBoost provides a high level of accuracy and robust performance across the board. The choice of Random Forest as the primary model remains justified due to its interpretability and stability, but XGBoost demonstrates comparable accuracy, making it a strong alternative. The results suggest that both models are capable of handling the complexities of the dataset effectively, with XGBoost offering a slight edge in precision and recall across specific classes.

4.4 Discussion

The results obtained from the model predictions and the SHAP analysis provide significant insights into the factors influencing emotional states as predicted from social media usage. The evaluation measures, which include R^2 and MSE, show how well the model performs in relation to the three main prevailing emotions: sorrow, neutrality, and happiness. The analysis revealed that the Daily Usage Time and Comments Received Per Day are consistently among the top influential features for predicting emotional states. Specifically, for happiness, significant predictors include the amount of time spent on social media and the volume of interactions via messages and comments. This aligns with existing literature suggesting that social interactions on digital platforms can enhance positive emotional experiences.

To address prediction error cases, we analyzed instances where the model's predictions deviated significantly from the actual values. For happiness, error cases often occurred when users displayed inconsistent engagement patterns, such as fluctuating levels of social interaction, which confused the model. For neutrality, errors were common when users exhibited mixed emotional signals, such as high engagement alongside signs of underlying anxiety or boredom, indicating the challenge of capturing emotional balance. For sadness, mispredictions were frequently associated with atypical platform usage patterns, where the emotional impact of comments varied depending on the context, such as the nature or tone of the comments. These error analyses highlight the need for more context-aware features and further refinement of the model to account for these complexities.

The SHAP summary plots provided a detailed view of how individual feature values impact the model's predictions. High values of comments received and daily usage time, for instance, were shown to have a substantial positive correlation with the predicted emotional states. The ability to read the model in great detail is essential for comprehending how the model makes decisions and for creating tactics that will favorably affect user emotions. In terms of practical application, the research findings offer valuable insights for social media platform design and user emotion intervention strategies. For example, platforms can be designed to promote healthier engagement by encouraging positive interactions, such as supportive comments and meaningful messaging. Additionally, user emotion interventions could include personalized recommendations for content or reminders to take breaks based on detected emotional states, helping to manage well-being and mitigate negative emotions. These applications can enhance user experience while promoting emotional health on digital platforms.

5. CONCLUSION

This study demonstrates the efficacy of supervised learning models, particularly Random Forest, in predicting emotional states based on social media usage patterns. High accuracy was demonstrated by the model evaluation using metrics like MSE and R^2, particularly for the happiness prediction. The SHAP analysis provided valuable insights into feature importance, highlighting that Daily Usage Time, Comments Received Per Day, and platform-specific interactions significantly influence emotional states.

The findings underline the critical role of social interactions and engagement on social media platforms in shaping users' emotions. Positive interactions, as indicated by comments and active usage, are strongly correlated with happiness, whereas negative emotions like sadness are influenced by the nature and frequency of social validations and specific platform usage. This implies that social media interactions—both in terms of volume and quality—are crucial indicators of emotional health.

These insights have practical implications for developing interventions aimed at enhancing digital well-being. By understanding which features most impact emotional states,

platforms can design features and algorithms that promote positive interactions and mitigate negative experiences. Additionally, users can be educated on how their social media habits may affect their emotional health, encouraging more mindful and balanced usage.

However, this study has several limitations. One major limitation is the representativeness of the data, as the dataset is derived from specific social media platforms, which may not generalize well to other platforms or user demographics. Additionally, the features used in the model may not comprehensively capture all relevant aspects of social media usage that influence emotions, such as the content or sentiment of posts and interactions. The model's adaptability is also a concern, as it has not been validated across different platforms to test its generalization capability. Future research should address these limitations by incorporating more diverse datasets, enhancing feature completeness, and performing cross-platform validation experiments. Furthermore, longitudinal studies are needed to explore the long-term effects of digital interactions on emotional health and refine models to better understand the dynamic nature of social media's impact on emotions.

REFERENCES

[1] Appel, G., Grewal, L., Hadi, R., Stephen, A.T. (2020). The future of social media in marketing. Journal of the Academy of Marketing Science, 48(1): 79-95. https://doi.org/10.1007/s11747-019-00695-1

[2] Sykora, M., Elayan, S., Hodgkinson, I.R., Jackson, T.W., West, A. (2022). The power of emotions: Leveraging user generated content for customer experience management. Journal of Business Research, 144: 997-1006. https://doi.org/10.1016/j.jbusres.2022.02.048

[3] Steinert, S., Dennis, M.J. (2022). Emotions and digital well-being: On social media's emotional affordances. Philosophy & Technology, 35(2): 36. https://doi.org/10.1007/s13347-022-00530-6

[4] Bohnert, M., Gracia, P. (2021). Emerging digital generations? Impacts of child digital use on mental and socioemotional well-being across two cohorts in Ireland, 2007–2018. Child Indicators Research, 14: 629-659. https://doi.org/10.1007/s12187-020-09767-z

[5] Brailovskaia, J., Margraf, J. (2022). The relationship between active and passive Facebook use, Facebook flow, depression symptoms and Facebook Addiction: A three-month investigation. Journal of Affective Disorders Reports, 10: 100374. https://doi.org/10.1016/j.jadr.2022.100374

[6] Canale, N., Marino, C., Lenzi, M., Vieno, A., Griffiths, M.D., Gaboardi, M., Giraldo, M., Cervone, C., Massimo, S. (2022). How communication technology fosters individual and social wellbeing during the COVID-19 pandemic: Preliminary support for a digital interaction model. Journal of Happiness Studies, 23(2): 727-745. https://doi.org/10.1007/s10902-021-00421-1

[7] Babu, N.V., Kanaga, E.G.M. (2022). Sentiment analysis in social media data for depression detection using artificial intelligence: A review. SN Computer Science, 3(1): 74. https://doi.org/10.1007/s42979-021-00958-1

[8] Mollas, I., Bassiliades, N., Tsoumakas, G. (2022). Conclusive local interpretation rules for random forests. Data Mining and Knowledge Discovery, 36(4): 1521-1574. https://doi.org/10.1007/s10618-022-00839-y

[9] Chekroud, A.M., Bondar, J., Delgadillo, J., Doherty, G., Wasil, A., Fakkema, M., Cohen, Z., Belgrave, D., DeRubeis, R., Iniesta, R., Dwyer, D., Choi, K. (2021). The promise of machine learning in predicting treatment outcomes in psychiatry. World Psychiatry, 20(2): 154-170. https://doi.org/10.1002/wps.20882

[10] Hikmawati, E., Nugroho, H., Surendro, K. (2024). Improve the quality of recommender systems based on collaborative filtering with missing data imputation. In Proceedings of the 2024 13th International Conference on Software and Computer Applications, Bali Island Indonesia, pp. 75-80. https://doi.org/10.1145/3651781.3651793

[11] Hikmawati, E., Maulidevi, N.U., Surendro, K. (2023). Improved classification accuracy by feature selection using adaptive support method. In Proceedings of the 2023 12th International Conference on Software and Computer Applications, Kuantan, Malaysia, pp. 171-176. https://doi.org/10.1145/3587828.3587854

[12] Sahlaoui, H., Nayyar, A., Agoujil, S., Jaber, M.M. (2021). Predicting and interpreting student performance using ensemble models and shapley additive explanations. IEEE Access, 9: 152688-152703. https://doi.org/10.1109/ACCESS.2021.3124270

[13] Roy, A., Nikolitch, K., McGinn, R., Jinah, S., Klement, W., Kaminsky, Z.A. (2020). A machine learning approach predicts future risk to suicidal ideation from social media data. NPJ digital medicine, 3(1): 78. https://doi.org/10.1038/s41746-020-0287-6

[14] Peng, S., Cao, L., Zhou, Y., Ouyang, Z., Yang, A., Li, X.G., Jia, W.J., Yu, S. (2022). A survey on deep learning for textual emotion analysis in social networks. Digital Communications and Networks, 8(5): 745-762. https://doi.org/10.1016/j.dcan.2021.10.003

[15] Oliveira, F.B., Haque, A., Mougouei, D., Evans, S., Sichman, J.S., Singh, M.P. (2022). Investigating the emotional response to COVID-19 news on Twitter: A topic modeling and emotion classification approach. IEEE Access, 10: 16883-16897. https://doi.org/10.1109/ACCESS.2022.3150329

[16] Uban, A.S., Chulvi, B., Rosso, P. (2021). An emotion and cognitive based analysis of mental health disorders from social media data. Future Generation Computer Systems, 124: 480-494. https://doi.org/10.1016/j.future.2021.05.032.

[17] Chancellor, S., De Choudhury, M. (2020). Methods in predictive techniques for mental health status on social media: A critical review. NPJ Digital Medicine, 3(1): 43. https://doi.org/10.1038/s41746-020-0233-7

[18] Wirtz, D., Tucker, A., Briggs, C., Schoemann, A.M. (2021). How and why social media affect subjective well-being: Multi-site use and social comparison as predictors of change across time. Journal of Happiness Studies, 22: 1673-1691. https://doi.org/10.1007/s10902-020-00291-z

[19] Karsay, K., Matthes, J., Schmuck, D., Ecklebe, S. (2023). Messaging, Posting, and Browsing: A mobile experience sampling study investigating youth's social media use, affective well-being, and loneliness. Social Science Computer Review, 41(4): 1493-1513. https://doi.org/10.1177/08944393211058308

[20] Lin, X., Kishore, R. (2021). Social media-enabled healthcare: A conceptual model of social media

affordances, online social support, and health behaviors and outcomes. Technological Forecasting and Social Change, 166: 120574. https://doi.org/10.1016/j.techfore.2021.120574
[21] Geetha, G., Saranya, G., Chakrapani, K., Ponsam, J.G., Safa, M., Karpagaselvi, S. (2020). Early detection of depression from social media data using machine learning algorithms. In 2020 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS), Chennai, India, pp. 1-6. https://doi.org/10.1109/ICPECTS49113.2020.9336974
[22] Otchere, D.A. (2024). Fundamental error in tree-based machine learning model selection for reservoir characterisation. Energy Geoscience, 5(2): 100229. https://doi.org/10.1016/j.engeos.2023.100229
[23] Braig, N., Benz, A., Voth, S., Breitenbach, J., Buettner, R. (2023). Machine learning techniques for sentiment analysis of COVID-19-related twitter data. IEEE Access, 11: 14778-14803. https://doi.org/10.1109/ACCESS.2023.3242234
[24] Wang, J., Liu, Y.L. (2023). Deep learning-based social media mining for user experience analysis: A case study of smart home products. Technology in Society, 73: 102220. https://doi.org/10.1016/j.techsoc.2023.102220
[25] Mukta, M.S.H., Ahmad, J., Zaman, A., Islam, S. (2024). Attention and meta-heuristic based general self-efficacy prediction model from multimodal social media dataset. IEEE Access, 12: 36853-36873. https://doi.org/10.1109/ACCESS.2024.3373558
[26] Meena, G., Mohbey, K.K., Indian, A., Khan, M.Z., Kumar, S. (2024). Identifying emotions from facial expressions using a deep convolutional neural network-based approach. Multimedia Tools and Applications, 83(6): 15711-15732. https://doi.org/10.1007/s11042-023-16174-3
[27] Prendin, F., Pavan, J., Cappon, G., Del Favero, S., Sparacino, G., Facchinetti, A. (2023). The importance of interpreting machine learning models for blood glucose prediction in diabetes: An analysis using SHAP. Scientific Reports, 13(1): 16865. https://doi.org/10.1038/s41598-023-44155-x
[28] Qi, Y., Shabrina, Z. (2023). Sentiment analysis using Twitter data: A comparative application of lexicon-and machine-learning-based approach. Social Network Analysis and Mining, 13(1): 31. https://doi.org/10.1007/s13278-023-01030-x
[29] Jordan, D.G., Slavish, D.C., Dietch, J., Messman, B., Ruggero, C., Kelly, K., Taylor, D.J. (2023). Investigating sleep, stress, and mood dynamics via temporal network analysis. Sleep Medicine, 103: 1-11. https://doi.org/10.1016/j.sleep.2023.01.007
[30] Gao, W., Wei, J., Li, Y., Wang, D., Fang, L. (2023). Motivations for social network site use and users' well-being: Mediation of perceived social support, positive self-presentation and honest self-presentation. Aslib Journal of Information Management, 75(1): 171-191. https://doi.org/10.1108/AJIM-08-2021-0224
[31] Xu, X., Liu, J., Liu, J.H. (2024). The effect of social media environments on online emotional disclosure: Tie strength, network size and self-reference. Online Information Review, 48(2): 390-408. https://doi.org/10.1108/OIR-04-2022-0245
[32] Alamsyah, N., Budiman, B., Yoga, T.P., Alamsyah, R.Y.R. (2024). A stacking ensemble model with SMOTE for improved imbalanced classification on credit data. TELKOMNIKA (Telecommunication Computing Electronics and Control), 22(3): 657-664. http://doi.org/10.12928/telkomnika.v22i3.25921
[33] Alamsyah, N., Kurniati, A.P. (2023). A novel airfare dataset to predict travel agent profits based on dynamic pricing. In 2023 11th International Conference on Information and Communication Technology (ICoICT), Melaka, Malaysia, pp. 575-581. https://doi.org/10.1109/ICoICT58202.2023.10262694
[34] Putrada, A.G., Alamsyah, N., Pane, S.F., Fauzan, M.N. (2023). Feature Importance on Text Analysis for a Novel Indonesian Movie Recommender System. In 2023 11th International Conference on Information and Communication Technology (ICoICT), Melaka, Malaysia, pp. 34-39. https://doi.org/10.1109/ICoICT58202.2023.10262504