Analyzing the Impact of Lifestyle Patterns on Daily Screen Usage

Viktorija Projkova, Angela Nastovska, Hristijan Gjoreski
Faculty of Electrical Engineering and Information Technologies,
University of Ss. Cyril and Methodius in Skopje, Macedonia
projkova.viktorija@gmail.com, angelanastovska23@gmail.com, hristijang@feit.ukim.edu.mk

***Abstract*—Understanding and predicting technology usage time is becoming increasingly relevant in modern society, especially due to its impact on health and lifestyle. In this paper, we analyze the effectiveness of two machine learning approaches, classification and regression, for estimating screen time (TUE) based on lifestyle-related factors. The dataset contains health and behavior data from individuals in Mexico, Peru, and Colombia, with features such as diet, physical activity, and hydration habits. In the classification task, screen time was treated as an ordinal categorical variable (low, medium, high usage), while in the regression task, it was numerically encoded to allow continuous prediction. A comparative analysis was conducted between the best performing models, XGBoost classification model and the Random Forest regression model, with mapped outputs. The classification approach achieved over 77% accuracy and stable F1 and recall scores, while regression models, although numerically close, lacked interpretability and performed weaker in classification-equivalent tasks. Overall, the classification model proved to be more effective for predicting digital behavior patterns, offering clearer insights and decision-making advantages in domains such as education, digital health, and well-being.**

Keywords — machine learning, technology usage, lifestyle habits, classification and regression

I. Introduction

Excessive screen time is a growing concern in modern society, with the proliferation of digital devices contributing to increased sedentary behavior and potential hazards to physical health, mental health, and overall well-being [1]. Excessive use of electronic devices such as smartphones, computers, and televisions has been associated with sedentary behavior, poor sleep quality, and mental fatigue. Long hours in front of screens can reduce physical activity, contributing to an unbalanced lifestyle and potentially leading to various health problems. Monitoring and understanding screen time behavior is therefore essential, especially in modern societies where digital consumption is constantly increasing.

The amount of time individuals spend using electronic devices can reflect their lifestyle patterns and habits, such as their level of physical activity, dietary choices, hydration, and use of transportation. Numerous studies have highlighted a strong correlation between daily screen time and various lifestyle habits. Individuals who spend extended periods in front of screens are more likely to engage in sedentary behavior, skip regular physical activity, and adopt irregular dietary routines. High screen time has also been associated with insufficient hydration, poor sleep quality, and increased consumption of fast food or sugary snacks. These lifestyle patterns do not only influence physical well-being, but also contribute to mental fatigue, decreased motivation, and overall lower quality of life. As such, screen time can be seen not only as an outcome, but also as a reflection of broader behavioral trends [2].

This paper aims to explore whether these behavioral factors can be used to predict screen time using machine learning techniques. By applying both classification and regression models, we investigate which approach provides more accurate and interpretable results for screen time prediction. The ultimate goal is to identify meaningful patterns in digital behavior and support future efforts in promoting healthier and more balanced lifestyles.

II. Related Work

Most existing studies in public health and behavioral science focus on physical health outcomes such as obesity or chronic diseases, while the behavioral aspect of screen time is less explored, especially using machine learning techniques. Our work aims to fill this gap by analyzing “Time of using technology” as a modern lifestyle indicator.

The dataset and general methodology used in this study are inspired by the work of Palechor and Manotas (2019) [3], even though their research follows a fundamentally different direction compared to ours, which is understandable, since their most accented, target variable is “NObeyesdad” – Level of obesity. Their study mainly focuses on describing the dataset and balancing the obesity classes using synthetic oversampling (SMOTE), but it does not go further into building predictive models or analyzing which specific features contribute to obesity.

Our work started from a similar idea – examining how lifestyle factors affect people – but it takes a different direction in both focus and goal. Instead of analyzing obesity, we focus on daily screen time (TUE) as a behavior that reflects the modern world. Unlike the research paper by Palechor et al. [3], our study applies both classification and regression models to predict screen time. In addition, we performed data preprocessing, exploratory data analysis (EDA), feature engineering, feature selection, and hyper parameter optimization to identify the best performing models in each approach. These steps are crucial for understanding the dataset and achieving more accurate and reliable results.

To explore which modeling strategy performs best for predicting screen time, we trained and compared a variety of machine learning models. For the classification task, we used Random Forest, XGBoost, and Support Vector Machine (SVM). For the regression task, we applied Linear Regression, Random Forest Regressor, K – Nearest Neighbors Regressor (KNN) and XGBoost Regressor. This diverse selection allowed us to assess both simple interpretable models and more complex ensemble methods, as well as to examine how different model types handle categorical vs. continuous target representations.

In contrast, even though we shift the research focus toward a different outcome - screen time - we aim to explore not just the prediction of this variable, but also how it is related to other lifestyle factors such as physical activity, nutrition, hydration, and transport habits.

III. DATA

The dataset used in this study originates from the UCI Machine Learning Repository and it contains lifestyle and health – related data from 2111 individuals originating from Mexico, Peru and Colombia. It includes 17 features (categorical and numeric) that describe participant's demographic characteristics, eating habits, physical activity and other behavioral factors. The information in the dataset were collected through a web – based survey, and approximately 77% of the data is synthetically generated using SMOTE technique to balance class labels in the original classification task.

The categorical features in the dataset include: *Gender* – indicates the biological sex of the participant (Male/Female); *family history with overweight* – whether there is a known family history of overweight; *FAVC (Frequent consumption of high-calorie food)* – Yes/No indicator; *CAEC (Consumption of food between meals)* – frequency-based: No, Sometimes, Frequently, Always; *SMOKE* – smoking status (Yes/No); *SCC* ***(****Calories consumption monitoring****)*** – Yes/No*;* *CALC (Alcohol consumption)* – frequency: No, Sometimes, Frequently; *MTRANS (Transportation used)* – categories such as Walking, Public Transport, Motorbike, Bike, Automobile.

The numerical features in the dataset include: *Age* - representing the participant's age in years; *Height* - measured in meters; *Weight* - measured in kilograms; *FCVC (Frequency of vegetable consumption)* - rated on a scale from 1 to 3; *NCP (Number of main meals per day)* - typically ranging between 1 and 4; *CH2O* - indicating daily water intake in liters; *FAF (Physical activity frequency)* - which reflects the weekly frequency of physical activity, also expressed as a scaled value; and *TUE (Time Using Electronic devices)*, representing the number of hours per day spent in front of screens.

All categorical features were encoded based on their type and the nature of their values. Features with two possible values, such as *FAVC*, *SMOKE*, *SCC*, and *family history with overweight*, were encoded using *binary encoding* (0 for No, 1 for Yes). Features with ordinal relationships, such as *CAEC* and *NObeyesdad*, were encoded using *ordinal encoding*, preserving the inherent order in their categories. Multiclass categorical features without a natural order, like *Gender* and *MTRANS*, were transformed using *one-hot encoding* to create separate binary indicators for each category.

As for the numerical features, including *Age****,*** *Height****,*** *Weight****,*** *FCVC****,*** *NCP****,*** *CH2O****,*** *FAF****,*** and *TUE*, we applied *z-score standardization*. This transformation scaled the features to have a mean of 0 and a standard deviation of 1, improving model performance and ensuring comparability across variables.

As we said before, our target variable is TUE (Time of using electronic devices), which refers to the number of hours per day an individual spends in front of screens (television, computer, smartphone etc.). This variable is originally categorical (0-2 hours, 3-5 hours and more than 5 hours). To gain a deeper understanding of screen time behavior, we approach the problem using two perspectives: Classification modeling, where we retrain the original categorical labels (0,1,2) and train models to predict the correct class of screen time range. Regression modeling, where TUE is numerically encoded (0 -> 1.0h, 1 -> 4.0h, 2 - > 5.5h), allowing us to predict approximate screen time as a continuous variable. In order to enable a valid and direct comparison between the classification and regression models, it is essential to operate with the same format of the target variable. Therefore, as part of the comparison process, an inverse mapping is applied – transforming the predicted regression values back into their original categorical classes (0,1,2). This step ensures consistency in the output format of both models, allowing for the application of the same evaluation metrics and a fair comparison within a unified results table.

In addition to the original features, we performed feature engineering by deriving new variables such as BMI (Body Mass Index) and BMR (Basal Metabolic Rate) that could improve the model's ability to detect patterns related to screen time. These engineered features were also standardized using z-score normalization to maintain consistency with the rest of the numerical data.

Furthermore, we evaluated that the target variable Tue_Target is not evenly distributed across it's classes, specifically class 1 (representing moderate screen time) is significantly overrepresented. This class imbalance could lead to biased model performance, therefore, it is necessary to apply balancing techniques (Fig. 1)

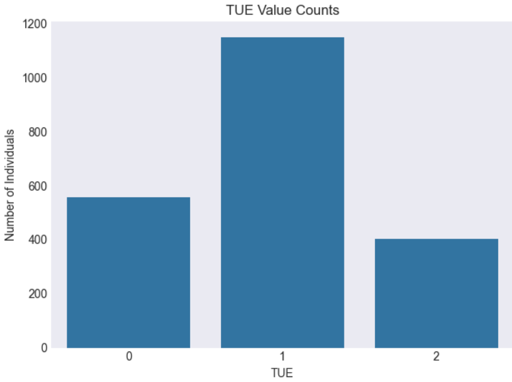

Figure 1. Distribution of TUE Classes before balancing

To address the imbalance in the TUE_Target variable, we applied a two-step resampling strategy on the training data: Oversampling with SMOTE: Classes 0 and 2, which were underrepresented in the original distribution, were synthetically oversampled to 600 samples each using the SMOTE algorithm. This ensures that these minority classes have sufficient representation in the training process. Undersampling the majority class: Class 1, which initially had the highest number of instances, was randomly undersampled to 600 samples to prevent the model from being biased toward this class. As a result, the training dataset became perfectly balanced, with 600 samples per class, creating a fair foundation for building robust and unbiased classification models (Fig.2).

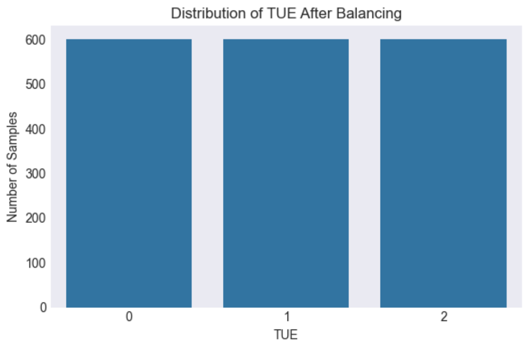

Figure 2. Distribution of TUE Classes after balancing

After completing the resampling procedure, the newly balanced dataset was used exclusively for the classification task, ensuring equal representation of all three screen time categories. For the regression task, however, the original unbalanced dataset was retained. This decision was made in order to preserve the natural variation and distribution of the continuous TUE values, which are necessary for meaningful regression modeling.

IV. METHODS

In this study, we developed two types of predictive models: classification models to predict the screen time category (low, medium, high), and regression models to estimate the approximate number of hours a person spends using electronic devices. For the *classification task*, we trained: *Random Forest Classifier* [4] – an ensemble method based on bagging, combining multiple decision trees; *Support Vector Machine (SVM)* [5] – a margin-based classifier effective for high-dimensional spaces; *XGBoost Classifier* [6] – a gradient boosting algorithm that iteratively improves the model by focusing on misclassified samples. For the *regression task*, we used: *Linear Regression* [7] – a baseline linear model to identify direct proportional relationships; *Random Forest Regressor* [4] – capable of modeling complex non-linear interactions; *K-Nearest Neighbors Regressor (KNN)* [8] – a non-parametric method relying on similarity between data points; *XGBoost Regressor* [6] – an advanced boosting technique with built-in regularization and high performance for structured data. To evaluate the models, the dataset was split into training, validation, and test sets using stratified sampling for classification, in order to preserve class proportions. This allowed us to tune hyper parameters using the validation set while fairly assessing model performance on the held-out test set. All models were trained on the preprocessed dataset. Model performance was evaluated using the following metrics: For classification: *Accuracy***,** *Precision***,** *Recall*, and *F1-score*; For regression: *Mean Absolute Error (MAE)***,** *Root Mean Squared Error (RMSE)***,** and *R² score***.**

V. EXPERIMENTAL SETUP

To train and evaluate the models, we used different data partitioning strategies depending on the task. For the classification models, we applied an 80/20 train-test split, where 80% of the data was used for training and 20% for testing. An additional 20% of the training data was set aside as a validation set to tune hyper parameters and avoid overfitting. The final model performance was evaluated on the held-out test set, which remained completely unseen during training and validation. For classification, we used Accuracy**,** Precision**,** Recall, and F1-score to evaluate model performance. Given the presence of class imbalance, special attention was given to F1-score and Recall, which better reflect model effectiveness across all screen time categories.

For the regression models, we evaluated performance using the following metrics:

$$RMSE = \sqrt{\frac{1}{q}\sum_{1}^{q}(TUE_{estimated} - TUE_{true})^2} \tag{1}$$

$$MAE = \frac{1}{q}\sum_{1}^{q}|TUE_{estimated} - TUE_{true}| \tag{2}$$

$$R^2 = 1 - \frac{\sum(TUEtrue - TUEestimated)^2}{\sum(TUEtrue - \overline{TUEtrue})^2} \tag{3}$$

Greater emphasis was placed on R^2, as it provides an interpretable indication of how well the model explains the variance in screen time behavior. A value of R^2=1 indicates perfect prediction, while a value close to 0 indicates poor predictive performance.

VI. EXPERIMENTAL RESULTS FOR CLASSIFICATION

In order to build effective classification models for predicting technology usage (TUE) categories, we began by conducting feature selection to reduce dimensionality, remove irrelevant inputs, and improve overall model performance. Initially, we examined the correlation matrix to gain insights into the relationships between features and to identify variables that were either weakly correlated with the target or highly collinear with each other. This step was crucial for reducing noise and ensuring that the model would not be overwhelmed by redundant or uninformative features (Fig. 3).

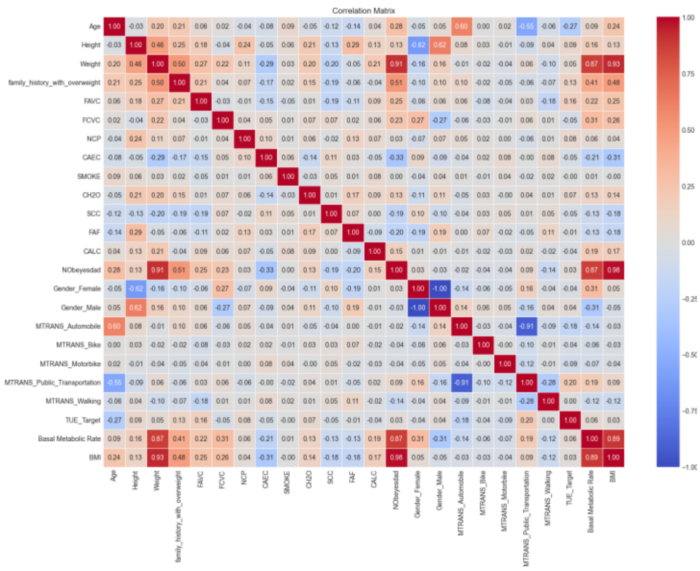

Figure 3. Correlation Matrix on all features

Following the correlation analysis, we implemented an iterative feature elimination process. Specifically, we adopted a manual, model-driven approach where we trained classification models repeatedly while removing one feature at a time. After each iteration, we observed the effect on the model's accuracy. Features that consistently had little to no impact—or even a negative impact—on performance were marked for exclusion. This method enabled us to make data-driven decisions rather than relying solely on automatic selection techniques. As a result, we excluded the first seven features, which demonstrated low predictive value and had minimal influence on the classification outcome (Table 1).

Table 1. Accuracy impact after individual feature removal (feature selection analysis)

	Removed Feature	Accuracy
0	FAF	0.789100
1	MTRANS_Public_Transportation	0.789100
2	FAVC	0.786730
3	MTRANS_Automobile	0.786730
4	CAEC	0.781991
5	Basal Metabolic Rate	0.779621
6	SCC	0.779621

With the refined feature set, we proceeded to train and evaluate several classification algorithms. The models tested included *SVM*, *Random Forest*, and *XGBoost*. For each model, we assessed the classification performance using multiple evaluation metrics: *Accuracy*, which indicates the proportion of correct predictions, *Precision and Recall*, which are especially important for imbalanced class distributions, *F1-score*, which balances precision and recall, *Confusion Matrix*, to visualize the distribution of correct and incorrect classifications across the three TUE categories. After evaluating all models, *XGBoost emerged as the most successful*, delivering the highest performance across all metrics (Fig. 4).

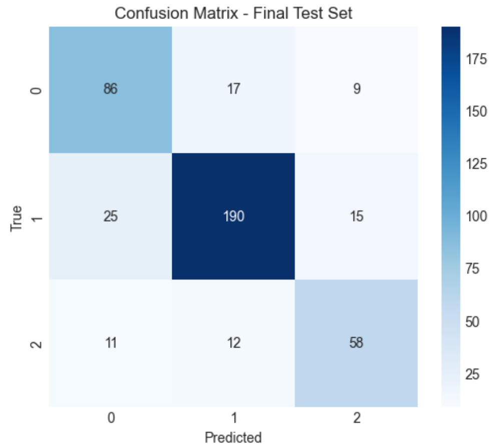

Figure 4. Performance of the XGBoost Classifier: Confusion Matrix on the Test Set

Its ability to handle complex relationships and automatically manage feature interactions gave it a distinct advantage. To further enhance its performance, we applied hyper parameter tuning using grid search. This process led to optimized values for parameters such as learning rate, maximum tree depth, and number of estimators. The tuned XGBoost model showed consistent improvements in both training and validation accuracy, confirming its robustness and suitability for the classification task (Table 2).

Table 2. Validation vs Test Performance of the Final Classifier

	Accuracy	F1 Score	Precision	Recall
Validation	0.815166	0.814652	0.814935	0.815166
Test	0.789598	0.791073	0.793822	0.789598

Overall, the feature selection process played a critical role in enhancing the effectiveness of the classifiers. By focusing on the most informative features and excluding those with limited contribution, we were able to build more efficient and interpretable models. The final XGBoost classifier not only achieved the best results but also provided reliable predictions that could support practical applications in analyzing and understanding technology usage patterns.

VII. Experimental Results for Regression

For the regression task, the TUE variable was treated as a continuous target, instead of the three discrete categories used in the classification setting. Although the original dataset contained discrete screen time intervals, we remapped the classes into continuous numerical values (1.0h, 4.0h, 5.5h) to better capture the variation in technology usage across individuals. We used the imbalanced distribution of TUE_Target for regression models. The performance of each model was assessed using four key metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the R^2 Score. These metrics were calculated on the validation set (Fig. 5).

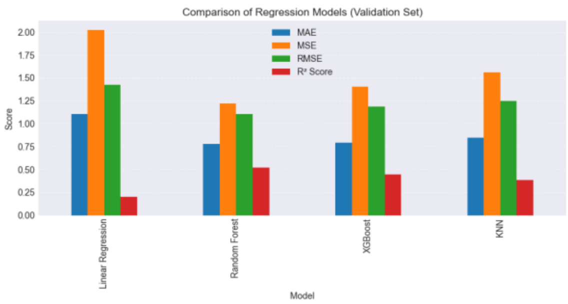

Figure 5. Comparison of Regression Models on the validation set using four evaluation metrics

From this comparison, Random Forest Regressor clearly outperformed the others, achieving the lowest RMSE and MAE, along with the highest R^2 score, indicating that it was the most effective at capturing the underlying structure in the data. To further improve the model, we applied *feature selection*. Using a similar methodology as in the classification task, we analyzed the effect of iteratively removing features while monitoring the model's performance. Features that had little or no impact were excluded (Table 3).

Table 3. R^2 score after individual feature removal during regression feature selection

	Removed Feature	R^2 Score
0	SMOKE	0.52011
1	Gender_Male	0.51852
2	FAVC	0.51749
3	Gender_Female	0.51724
4	family_history_with_overweight	0.51534
5	FAF	0.51421
6	SCC	0.51340

After selecting the most relevant features through iterative elimination, we retrained all regression models on the optimized feature set to re-evaluate their performance. This step ensured that each model had a fair chance to benefit from the reduced input complexity (Table 4).

Table 4. Performance of Regression Models after Feature Selection

	Model	MAE	MSE	RMSE	R^2 Score
0	Linear Regression	1.1342	2.1071	1.4516	0.1673
1	Random Forest	0.7554	1.1772	1.0850	0.5348
2	XGBoost	0.7539	1.3179	1.1480	0.4792
3	KNN	0.8495	1.5763	1.2555	0.3771

As shown, Random Forest continues to outperform the other models, achieving the lowest MAE (0.7554) and RMSE (1.0850), as well as the highest R^2 Score (0.5348). These results confirm its robustness and suitability for regression in this context.

To further improve the performance of the Random Forest Regressor, we conducted a two-step hyper parameter optimization process. Initially, we applied Randomized Search to explore a broader range of parameter combinations in a computationally efficient manner. This step allowed us to identify promising regions in the parameter space without the high cost of exhaustive search. Once the most relevant parameter ranges were identified, we proceeded with Grid Search for fine-tuning.

While the validation performance slightly fluctuated due to parameter adjustments, the final model tested on unseen data achieved the *lowest error values (MAE, MSE, RMSE)* and a *marginally higher R^2 Score*, indicating better generalization and robustness. These results confirm that the two-step tuning approach—combining randomized and grid search—was effective in refining the model for optimal predictive accuracy (Fig. 6).

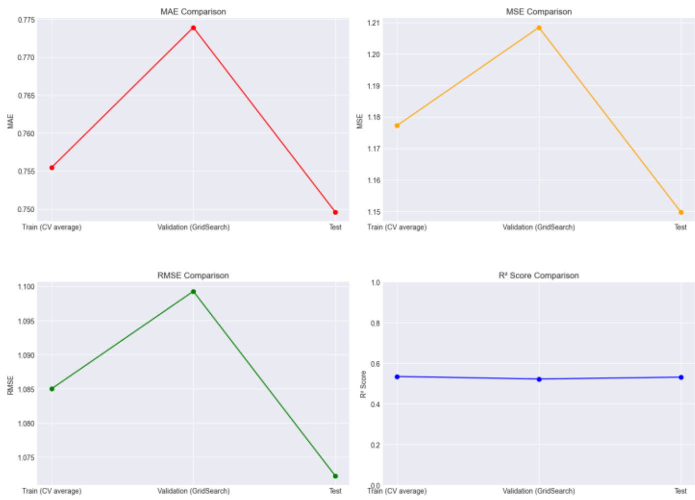

Figure 6. Comparison of evaluation metrics before and after hyper parameter optimization for the Random Forest Regressor

Overall, the regression analysis demonstrated that with proper feature selection and hyperparameter tuning, the Random Forest model is capable of accurately estimating screen time based on lifestyle-related inputs.

VIII. COMPARISON

In order to enable a valid and direct comparison between the classification and regression models, it is essential to operate with the same format of the target variable. Specifically, the regression model initially predicts continuous values (1.0h, 4.0h, 5.5h), which were previously mapped from the original classification categories 0, 1, and 2. However, such continuous outputs cannot be directly compared to the discrete predictions generated by the classification model. Therefore, as part of the comparison process, an inverse mapping is applied—transforming the predicted regression values back into their original categorical classes (0, 1, 2). This step ensures consistency in the output format of both models, allowing for the application of the same evaluation metrics and a fair comparison within a unified results table. This approach guarantees an objective and consistent analysis of model performance in terms of accuracy, F1 score, precision, and recall.

We performed a direct comparison between the best-performing classification model (XGBoost) and the top regression model (Random Forest), whose continuous predictions were mapped back into categorical labels. This approach enabled us to apply consistent evaluation metrics across both models, ensuring a fair assessment of their predictive capabilities (Fig. 7).

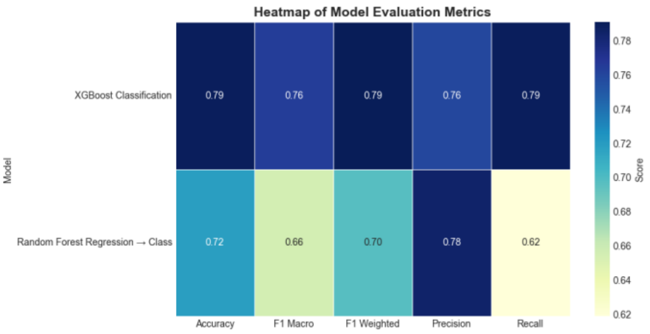

Figure 7. Heatmap of Model Evaluation metrics

As illustrated, the classification model consistently outperformed the regression-based method across all metrics. The differences are particularly notable in recall and macro-averaged F1 score, highlighting the advantage of modeling the target variable directly as a categorical label rather than through indirect regression mapping (Fig.8)

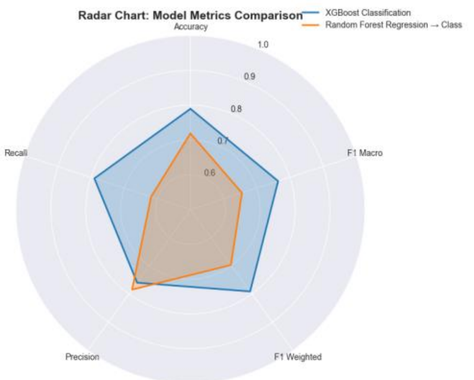

Figure 8. Radar Chart: Model metrics comparison

The final comparison between the classification-based model (XGBoost) and the regression-based model (Random Forest with mapped classes) demonstrates that the classification approach is significantly more effective in predicting screen time categories. The classification model achieves a better balance across key performance metrics such as accuracy, recall, and F1 score, with improvements exceeding 10% in some cases. Although the regression model shows slightly higher precision for certain classes, its overall ability to correctly classify all categories is considerably weaker. Therefore, it can be concluded that the classification model offers a more complete and efficient solution for the task of categorizing daily electronic device usage.

IX. CONCLUSION

This study investigated the prediction of screen time (TUE) using lifestyle-related features, applying both classification and regression approaches. The classification model, based on XGBoost, achieved the best overall results and proved more interpretable and actionable. Its ability to directly classify users into meaningful categories (low, medium, high usage) makes it suitable for practical applications in education, healthcare, and digital well-being. The regression model, using Random Forest, offered precise numerical predictions and explained 53% of the variance ($R^2 = 0.5314$). However, despite optimization, its interpretability and practical value were limited.

In the future, this analysis could be extended by incorporating additional behavioral or psychological variables, enabling even richer models. The developed methodology could also support tools for personalized recommendations, digital health monitoring, or awareness-raising platforms that promote healthier screen time habits.

REFERENCES

[1] K. A. Devi and S. K. Singh, "The hazards of excessive screen time: Impacts on physical health, mental health, and overall well - being", J. Educ. Health Promot, vol. 12, p.413, 2023

[2] D. R. S. Irawati, "Screen time and its health consequences in children and adolescents", ResearchGate, 2023

[3] F. M. Palechor, A. H. Manotas, "Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from Colombia, Peru and Mexico," *Data in Brief*, vol. 25, p. 104344, 2019.

[4] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[5] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support Vector Machines for Classification," *IEEE Intelligent Systems and their Applications*, vol. 13, no. 4, pp. 18–28, 1998.

[6] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794. arXiv:1603.02754.

[7] J. P. Keener, "Linear Regression," *Applied Linear Statistical Models*, University of Colorado Boulder, ch. 12

[8] K. Chaudhuri, "Nearest Neighbors I: Regression and Classification," Simons Institute for the Theory of Computing, University of California, San Diego, Tutorial, 2018.