PREDICTING SOCIAL MEDIA ADDICTION USING MACHINE LEARNING

- Introduction to Machine Learning -

JBP041-B-6

May 29, 2025

By

Coen Hiddink - *2156559*

Angelo Filiol de Raimond - *2154269*

Vasilis Loridas - *2154802*

Contents

1 Introduction

1.1 Introduction and Motivation

Social media platforms have become an integral part of daily life. According to [1], 51% of teenagers spend more than 4.8 hours per day online, and nearly one in three users exhibit behaviors indicative of addiction without being aware of it [2]. Traditional self-assessment tools often fail to detect these at-risk behaviors early, due to cognitive biases and social desirability effects. Our goal is to develop a machine learning model that reliably predicts the level of social media addiction using objective usage and interaction data.

1.2 Problem Statement

- **Context:** Intensive use of social platforms and potential mental health impacts.
- **Limitation of existing approaches:** Self-reported questionnaires are subject to underestimation and lack early warning capabilities.
- **Operational challenge:** Build a quantitative, automated, and scalable tool to identify at-risk users and alert them before an addiction becomes entrenched.

1.3 Research Questions

- Which usage metrics (session duration, return frequency, notifications clicked, posting vs. browsing rate, etc.) best discriminate addictive from moderate behavior?
- How can we transform raw logs into structured features suitable for machine-learning algorithms?
- What minimum performance thresholds (precision, recall, F1-score) must our model meet to be useful in a prevention context?
- How can we integrate this model into a privacy-preserving, scalable pipeline for public deployment?

1.4 Methodology: CRISP-DM

We follow the CRISP-DM framework:

1. **Business Understanding:** Define objectives (addiction prediction, user awareness) and constraints (scoring latency, user adoption).
2. **Data Understanding:** Collect irregular logs from a platform and perform exploratory analysis to identify patterns.
3. **Data Preparation:** Removal of anomalous sessions and handling of missing values; scaling of numeric features and encoding of categorical variables into numeric representations.
4. **Modelling:** Experiment with regression, random forests, boosting, using cross-validation to compare performance.
5. **Evaluation:** Target a minimum precision of 85% for classifying at-risk users; measure precision, recall, and F1-score and analyze false positives/negatives.
6. **Deployment (planned):** Design a lightweight API for real-time scoring and consider integration scenarios such as browser extensions, mobile applications, and alert dashboards.

1.5 Results and Conclusions

The results show that ensemble tree-based models, particularly Random Forest, achieved the best predictive performance in estimating self-reported social media addiction, with a mean R^2 of 0.44. Linear models like Ridge and Lasso identified key predictors such as Watch Time and Frequency but explained less variance overall. Polynomial regression underperformed due to overfitting, and stacking models failed to outperform the best individual models.

1.6 Perspectives

These results pave the way for a large-scale public deployment aimed at raising awareness — especially among heavy users — of the risks of social media addiction. In the long term, integrating personalized support modules and adaptive feedback could further enhance the effectiveness of preventive interventions.

2 Related Work

2.1 Latent Profile Analysis and Classification Heterogeneity

In [3], a latent profile analysis (LPA) was applied to two independent samples (UK: $n = 573$, US: $n = 474$) using the six items of the Bergen

Social Media Addiction Scale (BSMAS). Rather than imposing an arbitrary cut-off on a total score, the authors identified three empirically derived risk classes (low, moderate, high) based on co-occurrence patterns of salience, tolerance, mood modification, withdrawal, conflict, and relapse. They then benchmarked four common classification schemes—strict monothetic, monothetic, strict polythetic, and polythetic—against the LPA "gold standard" by computing sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). Prevalence estimates varied dramatically: from 1% (strict monothetic) to 15% (monothetic) in the UK sample, and from 0% to 11% in the US sample. The *polythetic* scheme provided the best balance—sensitivity around 73–74%, high specificity, and PPV of 88–94%—making it the most robust empirical classifier of social media addiction.

2.2 Logistic Regression During the COVID-19 Pandemic

While LPA addresses classification validity, other researchers have focused on practical predictors. Rashid et al. (2022), in [4], employed a logistic regression model on usage data collected during lockdown in Bangladesh ($n \approx 1200$). Predictor variables included objective measures (total daily screen time, session count, peak connection hours) and subjective self-reports (emotional distress, sleep quality, loneliness on Likert scales). After feature selection and L2 regularization to mitigate overfitting, the model achieved 94% overall accuracy, with recall of 91% and specificity of 89% in distinguishing "addicted" from "non-addicted" users. This work demonstrates that a simple, interpretable linear model can perform remarkably well under extreme usage variability, while underscoring the necessity of testing generalizability beyond the pandemic context.

2.3 Random Forests with Behavioral and Psychological Features

In [5], researchers trained a random forest classifier on a multimodal dataset ($n \approx 2000$) combining fine-grained usage logs (session duration, average inter-session interval, scroll depth) with validated psychological scales (emotional dependency, anxiety, depression). The model reached approximately 90% accuracy and an F1-score of 0.87. Feature-importance analysis revealed that total time spent online was the top predictor, followed by the active/passive time ratio, emotional dependency score, and anxiety level. The ensemble method's ability to capture non-linear interactions between behavioral and mental-health dimensions proved crucial for robust prediction.

2.4 Synthesis and Implications for Our Project

Across these studies, three key insights emerge:

1. **Definition variability:** Prevalence estimates depend heavily on the choice of classification scheme, making empirical benchmarking (e.g., via LPA) essential.
2. **Power of simple models:** Regression, when properly regularized, can serve as a strong, interpretable baseline, especially in contexts with pronounced usage shifts.
3. **Value of ensemble methods:** Random forests and other ensemble approaches, combined with rich feature sets that merge usage metrics and psychological indicators, deliver superior generalization and capture complex, non-linear effects.

Accordingly, our solution will try to:

1. Adopt the polythetic scheme.
2. Compare simple regression models and random forests using cross-validation to monitor and prevent overfitting.
3. Validate model performance (target $\geq 85\%$ accuracy) on culturally and temporally diverse subsamples to ensure real-world robustness.

3 Proposed Methodology and ML Techniques

3.1 End-to-End Data Science Pipeline

Before implementing our machine learning models, we followed a structured data science workflow to ensure data quality, interpretability, and relevance of the predictors. This pipeline includes the following stages:

1. **Data Inspection:** Checking the structure of the dataset, identifying column types, missing values, and inconsistencies.

2. **Exploratory Data Analysis (EDA):** Descriptive statistics, correlation matrices, and histograms will be used to uncover distributions, potential outliers, and potential linear/non-linear relationships between features.

3. **Data Cleaning:** Addressing missing values (e.g. filling or dropping), dropping irrelevant fields, and filtering sessions with unrealistic values.

4. **Feature Engineering:** One-hot encoding categorical variables, and optimising every variable for modelling.

5. **Feature Selection:** Selecting impactful features for the models, retaining statistically significant predictors to reduce dimensionality.

6. **Modelling:** Training and testing linear & polynomial regression models, as well as tree-based models. Each model is trained individually, and will be used to try to find an optimal (stacked) model.

7. **Finetuning Hyperparameters:** Optimising the hyperparameters of each individual model for best performance.

8. **Evaluation:** Comparing evaluation metrics like Mean Squared Error and R^2 using Cross Validation to assess generalisation performance.

This structured pipeline ensures the reproducibility and interpretability of our results, and bridges raw interaction logs with meaningful addiction risk prediction.

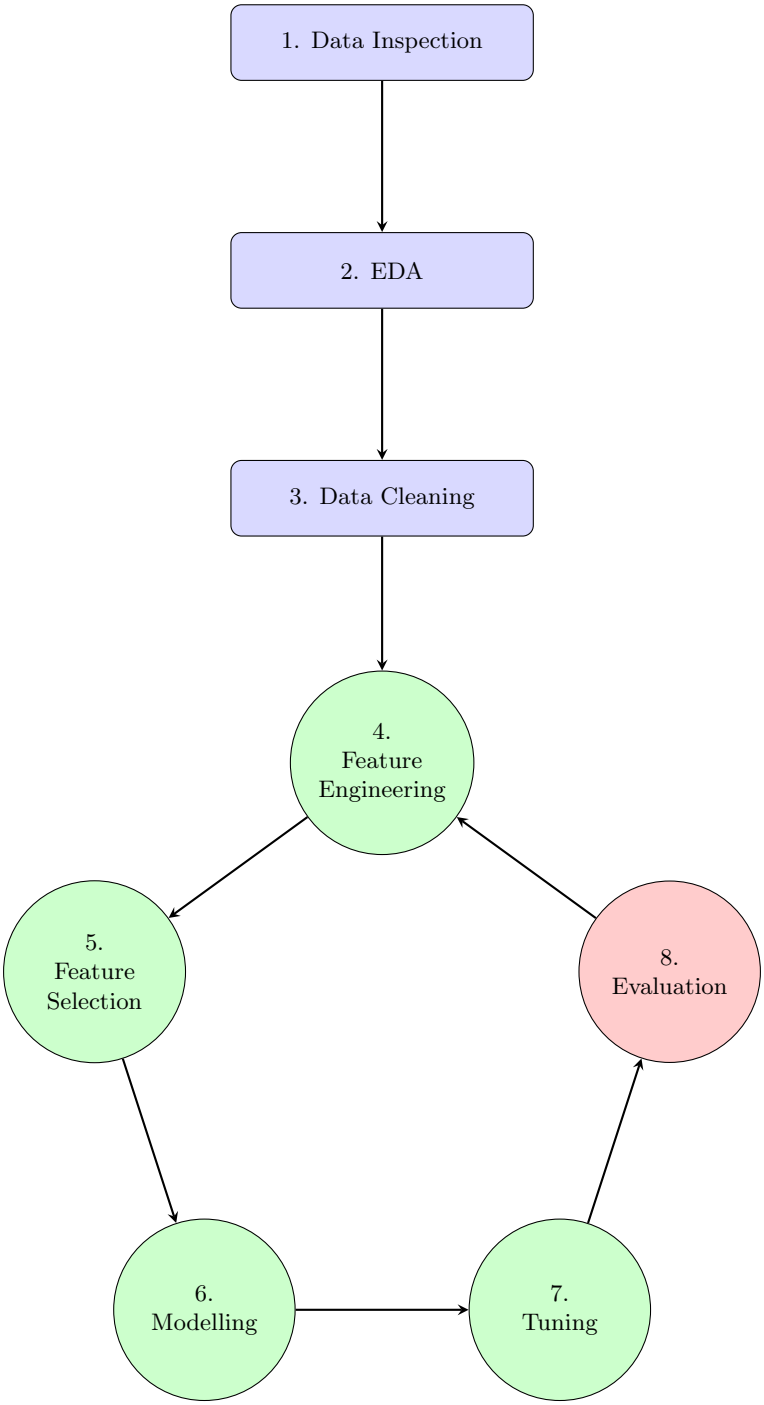

Figure 1: Machine learning pipeline: linear data preparation (steps 1–3) followed by an iterative modelling loop (steps 4–8).

Feature	Type	Range	Description
UserID	Num.	1–1000	*Unique identifier*
Age	Num.	18–64	*Age of user*
Gender	Cat.	3 cats.	*Male, Female, Other*
Location	Cat.	10 cats.	*US, Germany, India, Phillipines, Pakistan, Mexico, Vietnam, Brazil, Indonesia, Japan*
Income ($)	Num.	20138–99676	*Annual income of user*
Debt	Bool.	–	*If user has debt*
Owns Property	Bool.	–	*If user owns property*
Profession	Cat.	9 cats.	*Students, Artist, Teacher, Cashier, Labor/Worker, Waiting staff, Engineer, Manager, Driver*
Demographics	Cat.	2 cats.	*Rural, Urban*
Platform	Cat.	4 cats.	*Instagram, Facebook, Tiktok, Youtube*
Total Spent (min)	Num.	0–297	*Total time spent on social media by user*
Number of Sessions	Num.	0–37	*Number of sessions this user has used social media*
VideoID	Num.	1000–4999	*Unique identifier for video*
Video Category	Cat.	9 cats.	*Pranks, Vlogs, Gaming, Jokes/memes, Entertainment, ASMR, Lifehacks, Trends, Comedy*
Video Length (min)	Num.	0–45	*Length of the video*
Engagement score	Num.	0–9952	*Score on how engaged the user was with the video*
Importance	Num.	1–10	*Score on how important the video was experienced by the user*
Time Spent On Video	Num.	0–45	*Time spent on the video*
Videos Watched	Num.	0–45	*Total number of videos watched*
Scroll Rate (%)	Num.	0–100	*Scrolling rate of the user*
Frequency	Cat.	4 cats.	*When user mostly used social media: Morning, Afternoon, Evening, Night*
ProductivityLoss	Num.	1–10	*Score on productity lost by user afterwards*
Satisfaction	Num.	1–10	*Satisfaction score of the user*
Watch Reason	Cat.	4 cats.	*Procrastination, Habit, Entertainment, Boredom*
DeviceType	Cat.	2 cats.	*Smartphone, Computer, Tablet*
OS	Cat.	4 cats.	*iOS, Android, Windows, MacOS*
Watch Time	Cat.	-	*Time of day*
Self Control	Num.	1–10	*Score on self control*
Addiction Level	Num.	0–10	*Score on addiction level*
CurrentActivity	Cat.	4 cats.	*Commuting, At home, At school, At work*
ConnectionType	Cat.	3 cats.	*Mobile data, WiFi*

Table 1: Raw data overview

4 Experimental Evaluation

4.1 Dataset Description

The dataset used in this study is titled *Time-Wasters on Social Media* and originates from a synthetic source designed to simulate user behaviour and demographics in relation to social media usage. The dataset was retrieved from `www.kaggle.com`. It comprises a total of 1,000 user entries, each representing an individual profile with associated demographic, financial, and behavioral information. The dataset includes a mix of numerical, categorical, and binary variables. In Table 1, the structure of the dataset is depicted more elaborately.

4.2 Experimental Settings

Our goal in this research was to assess how user attributes, behavioural metrics, and content engagement relate to the `Addiction_Level` variable, and to test the hypothesis that certain content categories and user demographics significantly predict social media addiction.

4.2.1 Dependent and Independent Variables

The dependent variable is `Addiction_Level`, a numerical measure of an individual's addiction to social media platforms. The independent variables include a combination of demographic features (e.g. `Age`, `Gender`, `Location`), behavioural

patterns (e.g. `Watch_Time`, `Scroll_Rate`), content-related variables (e.g., `Video_Category`, `Time_Spent_On_Video`), and technical aspects (e.g. `Device_Type`, `Connection_Type`).

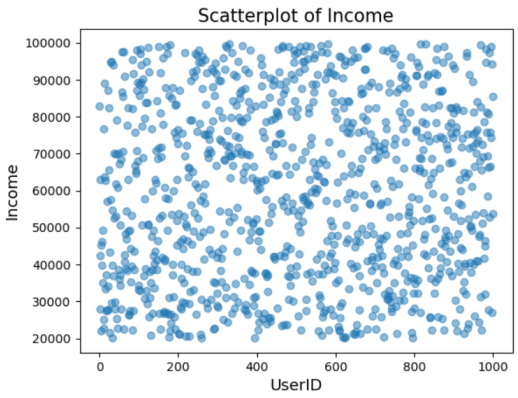

Figure 2: Example scatterplot for inspecting outliers

4.2.2 Data Preparation

First, exploratory data analysis was conducted to inspect distributions and outliers. For each variable, scatterplots were inspected; no records were excluded.

For each categorical variable, it was checked how many values occurred and what they were to get a better first understanding of the data. All features were standardised in the same format, and some inconsistencies were taken care of (e.g. `Barzil` → `Brazil`). Missing or duplicate values were not present. Moreover, two pseudo-numerical variables (`Watch_Time` and `Frequency`), which indicated the time of the day in a categorical format, were converted to be utilised as numerical variables. For `Watch_time` (e.g. 10:00 AM), the hour of the day (0-24) was used as an integer and `Frequency` (e.g. Morning) was enumerated according their order.

Next, a Pearson correlation matrix was computed and inspected. It was noted that `Self_Control`, `Satisfaction` and `Productivity_Loss` were (negatively) perfectly correlated with `Addiction_Level`. This probably arises from the way the data was synthesized, as this is not realistic for a dataset. In addition, as these two variables are self-evaluated, it does not make sense to include them in future modelling. Hence, they were excluded. In Figure 3, the relevant correlations are depicted; the other correlations were nearly zero. Lastly, categorical variables were one-hot encoded, resulting in a final number of 59 numerical predictor variables to be included in modelling.

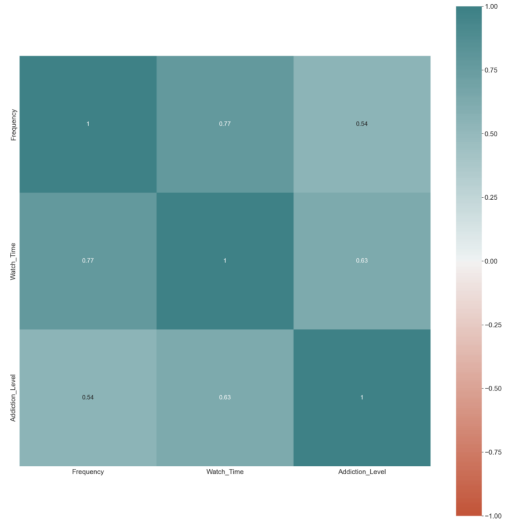

Figure 3: Relevant features in correlation matrix

4.2.3 Modelling

The dataset was split into a training set (80%) and test set (20%), using a fixed random seed to ensure reproducibility. 20% is a valid test size for a dataset of 1,000. This split provides a realistic scenario of evaluating generalisation to unseen users. Before fitting, the predictor variables were scaled.
We trained and compared several regression models:

- Linear models: Backwards selection (OLS), Lasso, Ridge
- Non-linear models: Polynomial regression
- Tree-based models: Decision Tree, Random Forest, XGBoost, GradientBoosting
- Ensemble methods: Stacking Regressor

Model performance was evaluated using three metrics:

- R^2 score: proportion of variance in the target explained by the model
- Mean Squared Error (MSE): average squared error between predicted and actual values
- Cross Validation: Assessing a model's ability to generalise to unseen data and to reduce the risk of overfitting.

4.3 Results

In this section, results of all explored models will be discussed. The result of each model shown is the result after parameter tuning and optimizing. At the end of this section, all models will be compared and discussed.

To start off, a baseline linear regression model was initialised. Based on this model, it was noted that the most variance in prediction values was in the lowest values. This might indicate that it would be the most challenging to predict if a user evaluates itself as not addicted.

4.3.1 Linear Regression Models

Predictor	**p-value**
Location Pakistan	0.04
Income	0.00
Frequency	0.00
Watch time	0.00
Initial R^2: *0.50*	

Table 2: Results BFS

To improve the baseline model, several regularisation techniques were introduced. Firstly, through backwards feature selection using OLS technique four features were selected to be used in the regression. The p-value threshold was set at 0.05. In Table 2, the results are shown. Next, Lasso and Ridge regularizations were applied. Due to the high number of predictors, one might expect that Lasso would be the best approach. However, as seen in the correlation matrix, the selected feature in BFS are also the ones highly correlated, suggesting Ridge might be a better approach. In Table 3 and 4, the results are shown.

Predictor	**Coefficient**
Income	-0.12
Current Activity Commuting	-0.12
Location Phillippines	-0.12
Video Category Gaming	0.13
Location US	-0.13
Location Vietnam	-0.14
Video Category Trends	0.15
Location Pakistan	-0.20
Frequency	0.23
Watch time	1.06
Initial R^2: *0.47*	

Table 3: Results Ridge

Model	**MSE**
Location Pakistan	-0.02
Video Category Jokes/Memes	-0.02
Income	-0.02
Frequency	0.19
Watch time	0.98
Initial R^2: *0.48*	

Table 4: Results Lasso

These linear models already suggest a relatively large predictive role of `Watch_Time`. In all models it is the strongest predictor. At first glance, their R^2 values are moderate, so the models are not discarded for implementation in future stacking models.

4.3.2 Polynomial Regression

In order to investigate potential non-linearity within the dataset, polynomial regression was applied using a poly-degree transformation of all predictor variables, using a linear regression. An R^2 value of –2.11 was observed, indicating that the model performed worse than a naïve mean-based prediction. In addition, the mean squared error (MSE) increased sharply to 15.16. As further polynomial degrees were tested, performance declined even further, suggesting that the underlying relationships in the data are most likely not quadratic. However, this does not mean that the relationships in the data are completely linear. Hence, we continue with the tree based models.

4.3.3 Tree Based Models

In order to capture more complex, non-linear relationships in the data, tree-based machine learning models were also evaluated. Unlike linear regression approaches, tree-based models operate through recursive binary partitioning and are inherently capable of modelling interactions and non-linear patterns without requiring explicit feature engineering or transformation. For this reason, they were considered a promising alternative to both linear and polynomial regressions.

4.3.4 Comparison of models

To systematically evaluate model performance, each model was assessed over 5 K-fold splits using key regression metrics: mean R^2, mean squared error (MSE) and standard deviation (STD) of each metric. The results are summarised in Table 5.

Model	Mean R^2	STD. R^2	Mean MSE	STD. MSE
BFS Linear Regression	0.40	0.09	2.41	0.41
Ridge Regression	0.36	0.09	2.58	0.40
Lasso Regression	0.39	0.08	2.53	0.40
Decision Tree	0.42	0.08	2.36	0.39
Random Forest	0.44	0.08	2.23	0.37
XGBoost	0.43	0.07	2.38	0.35
GradientBoosting	0.42	0.08	2.38	0.36

Table 5: Comparison of the models

Also, for each model, the predicted values are plotted against the actual values using a Two-Dimensional Kernel Density Estimate plot. With this visualisation, the performance of the models can be interpreted better. The plots are depicted in Figures 4 to 10 and all follow the same colour range.

Upon inspection of the 2D KDE plots, it can be observed that the overall predictive performance across models is broadly similar in structure. This is supported by the results in Table 5, which show that the difference in the mean coefficient of determination (R^2) between the best-performing model (Random Forest, $R^2 = 0.44$) and the lowest-performing model (Ridge Regression, $R^2 = 0.36$) is only 8%. Thus, the proportion of variance in `Addiction_Level` explained by the models remains relatively consistent across approaches.

However, what is particularly noteworthy from the visual plots is the alignment of the density mass with the diagonal perfect prediction line. For the tree-based models, the tangent of the predicted values appears to be more closely aligned with the diagonal, indicating more accurate predictions across the full range of actual values. In contrast, the linear regression models, including Ridge and Lasso, display a noticeably flatter orientation in the upper range. This suggests a tendency to underpredict higher levels of addiction, likely due to the linear models' limited capacity to capture non-linear or interaction effects present in the data.

Moreover, for all models, the 2D KDE plots indicate a clear higher performance around the central values – in this case 3-4 as the highest entry is 7, even though the theoretical range was 0-10 – and seems to have most variance around the extremes. This could arise from the fact that users who entered 0 as their addiction level – indicating no addiction – can exhibit a wide range of social media behaviour which does not follow a clear trend, whereas only the 'addicted' users follow a more distinct pattern in their behaviour.

Among all the models tested, the Random Forest Regressor achieved the highest mean R^2 score of 0.44, indicating that it explained the greatest proportion of variance in addiction scores across validation folds. It also recorded the lowest mean MSE (2.23), demonstrating superior predictive accuracy. Furthermore, its relatively low standard deviations ($STD.R^2 = 0.08; STD.MSE = 0.37$) indicate consistent performance across folds.

In contrast, the regularised linear models — Ridge and Lasso regression — exhibited slightly lower explanatory power (mean $R^2 = 0.36$ and 0.39, respectively) and higher mean errors ($MSE = 2.53 - 2.58$). The linear models appear to be too simple to capture the complex behavioural patterns underlying social media addiction.

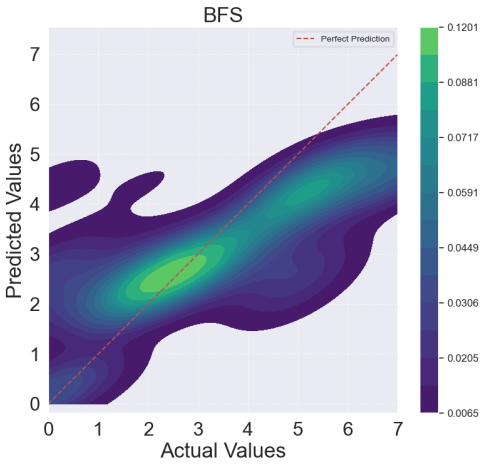

Figure 4: 2D KDE plot of BFS Model

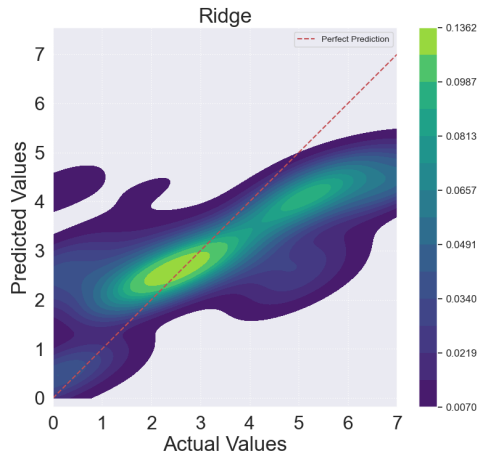

Figure 5: 2D KDE plot of Ridge Model

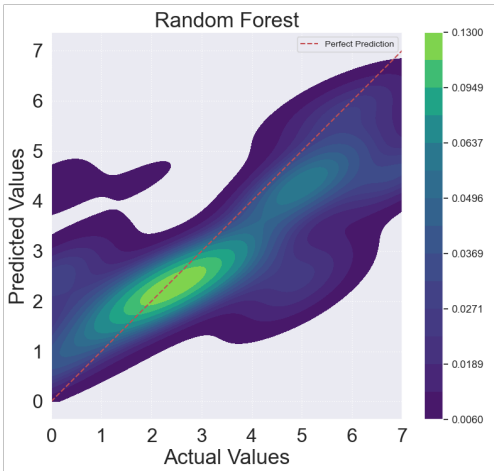

Figure 8: 2D KDE plot of Random Forest Model

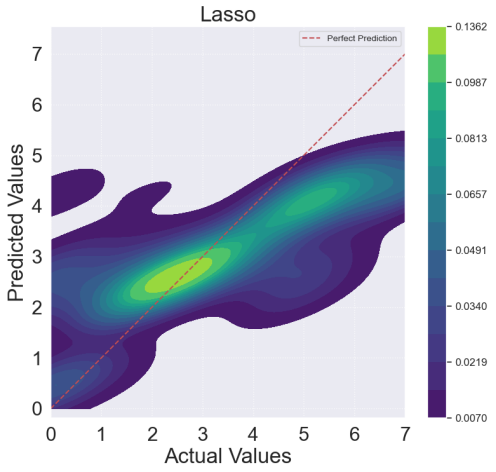

Figure 6: 2D KDE plot of Lasso Model

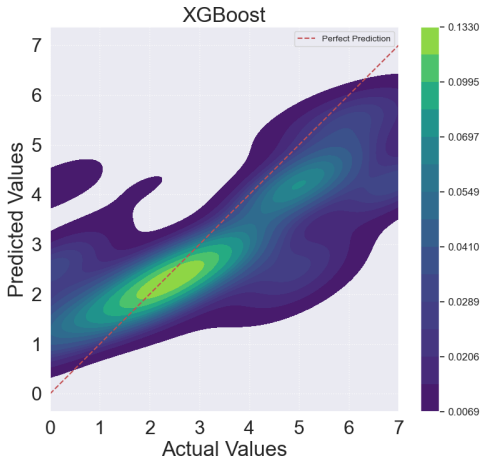

Figure 9: 2D KDE plot of XGBoost Model

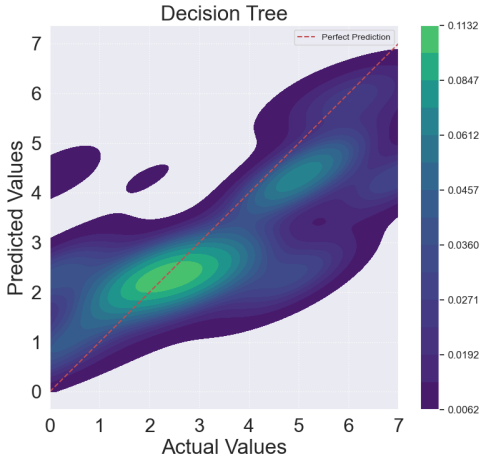

Figure 7: 2D KDE plot of Decision Tree Model

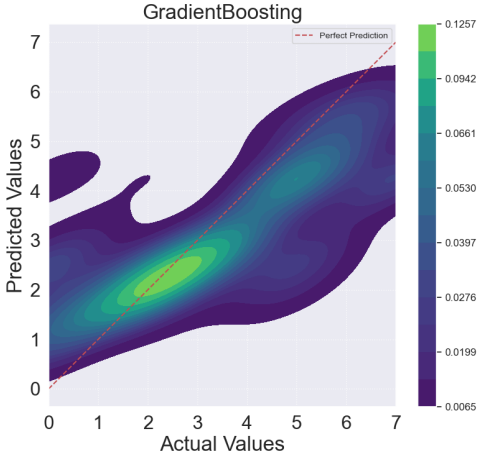

Figure 10: 2D KDE plot of Gradient Boosting Model

Base-learners	Final Estimator	Mean R^2	STD. R^2
Ridge + Linear	GBR	0.40	0.08
DT + GBR + XGB	GBR	0.39	0.06
DT + GBR + Lasso	GBR	0.39	0.06
RF + Lasso	GBR	0.39	0.05
Lasso + Ridge	Lasso	0.39	0.09

Table 6: Best performing combinations of stacking models

In summary, ensemble tree-based models, particularly Random Forest, consistently outperformed both linear models and single decision trees in terms of accuracy and stability. The better performance can be read from Table 5, but can particularly also be noticed from the density plots. These findings support the earlier visual and metric-based observations and confirm the suitability of ensemble methods for predicting social media addiction.

4.3.5 Stacking model

In the final step to improve the performance of the models and increase the success of explaining the variance in addiction, a stacking ensemble model approach was employed. All combinations of base-learners and final estimators were used, and the results concatenated in a dataframe for comparison. Through testing all combinations (2 or 3 base-models + final estimator) the optimal combination can be found.

Despite this optimisation approach, the stacking models did not outperform the best individual models. As shown in Table 6, the top-performing stacked configuration – Ridge + Linear regression as base models with Gradient Boosting as final estimator – only achieved an R^2 of 0.40. This is only marginally better than other stacked combinations (e.g., DT + GBR + XGB or RF + Lasso), but still fails to exceed the R^2 of the top standalone models, indicating no improvement from stacking in this case. This would suggest that the base models may be too similar and do not provide sufficiently diverse or complementary predictions, in that their combinations capture different variations in the data.

4.3.6 Concluding model performance

The results from the modelling phase reveal a small but consistent effect: tree-based ensemble models outperform both linear and polynomial regression approaches in predicting social media addiction from this dataset. Starting from a basic linear regression using backward feature selection (Table 2), the results showed modest explanatory power with an R^2 of 0.50 before cross-validation, primarily driven by features like Watch Time, Frequency, Income, and Location (Pakistan). Regularised models such as Ridge and Lasso regression (Tables 3 and 4) were introduced to address multicollinearity and overfitting risks. While these models retained slightly more features than BFS and delivered marginally improved generalisation compared to the baseline, their R^2 only reached 0.36–0.39. The corresponding KDE plots (Figures 5 and 6) further indicated their limitations in capturing the full range of addiction scores, particularly in regions of high self-reported addiction.

Polynomial regression, intended to capture potential non-linearities, instead resulted in a negative R^2 (–2.11), indicating a serious overfit and confirming its unsuitability for this task. Hence, it was discarded.

Advancement to the tree-based models – Decision Tree, Random Forest, XGBoost, and Gradient Boosting – showed a clear advantage. These models handle complex feature interactions and non-linearities with lower risk of overfitting like the polynomial regression. Random Forest achieved the highest cross-validated mean R^2 (0.44) and the lowest MSE (2.23), as shown in Table 5. Visual support from the 2D KDE plot of the Random Forest model (Figure 8) confirmed this, showing a better density alignment along the diagonal, indicative of accurate prediction across a wide range of values. Also, in general, the tree-based models seemed to have a better alignment with the diagonal, indicating a slightly better performance around the extremes.

Attempts to enhance accuracy through model stacking, as detailed in Table 6, did not lead to performance gains beyond the standalone Random Forest model. Even the best-performing stacked model (Ridge + Linear as base learners with Gradient Boosting as final estimator) reached only $R^2 = 0.40$.

In conclusion, ensemble tree-based approaches—especially Random Forest—stand out in their ability to model the complex patterns in the data, making them the most suitable candidates for potential deployment or further refinement.

4.4 Discussion

A substantial limitation of this research lies in the fact that the entire dataset was synthetically generated. It remains unknown whether any deliberate patterns, dependencies, or structures were purposely embedded into the data — or if the data was generated in a manner that reflects realistic behaviour at all. This uncertainty casts doubt on the interpretability and generalisability of the modelling results. One obvious pattern observed in the Pearson correlation matrix was the perfect (negative) correlation between `Addiction_Level`, `Productivity_Loss`, `Satisfaction`, and `Self_Control`. This cluster may represent the only clearly embedded pattern within the dataset. However, since these variables were excluded from the model, it is possible that the remaining predictors contained insufficient patterns to support a high-performing model.

Despite this, for the purpose of analysis and evaluation, the synthetic nature of the data will be disregarded. From this point onward, it is assumed that the dataset reflects a plausible record of social media usage and user-reported experience, enabling a discussion as if the data were genuinely observed.

Another limitation concerns the subjectivity of the target variable. The `Addiction_Level` variable appears to be based on self-evaluation, rather than on objective behavioural criteria or even third-party assessments. As such, it is sensitive to bias, misjudgement, and personal perception. Individuals may underreport or overstate their addiction due to personal traits such as denial, lack of insight, or even personality-driven exaggeration. Without objective data capturing individual differences in self-perception or honesty, this label noise cannot be corrected or even explained. This likely contributes to the model's unexplained variance and moderate R^2 scores observed across all models. In practice, models trained on subjective targets are inherently limited in predictive power, and careful interpretation of results is required.

An additional point of ambiguity encountered during the project was the nature of the prediction task itself. In this research, `Addiction_Level` was treated as a continuous regression target. However, it might have yielded better results — both conceptually and practically — to convert the task to a classification problem. Addiction scores could be binned into categories such as low, moderate, and high, and addressed using classification algorithms. Such a reframing might improve model interpretability and allowed for more robust evaluation metrics (e.g., accuracy, precision, recall). However, it might also have generalised too much as the score is converted from eleven possible scores into three.

Finally, while an R^2 of approximately 0.40 may be interpreted as weak to moderate, it is important to recognise that there are likely many meaningful predictors of social media addiction that were not captured in this dataset. For example, a user's sensitivity to addiction might explain parts of the variance. Furthermore, total screen time, application usage logs, or psychological profiles such as impulsivity or anxiety traits may offer stronger explanatory power. Contextual information, such as life satisfaction, peer influence, or even social support networks, may also influence addiction tendencies and were entirely absent from the dataset. In short, the model's moderate performance may not reflect a fundamental limitation of machine learning for this task, but rather a data limitation, where key explanatory variables are simply missing. In this light, the R^2 of 0.4 can still be considered a medium to good performance. To confirm this, future research would benefit from richer, real-world datasets incorporating both behavioural and psychosocial predictors of addictive tendencies.

5 Conclusion

In this project, a machine learning approach was employed to predict social media addiction using a variety of models applied to the dataset. The structured pipeline (Figure 1), based on the CRISP-DM framework, was followed and ensured that each stage – from data exploration and cleaning to feature engineering, modelling, and evaluation – was carried out systematically.

A diverse set of modelling strategies was implemented, including linear models, polynomial regression, and advanced ensemble techniques. While multiple models achieved moderate levels of performance, the tree-based ensemble methods

– especially Random Forest – consistently demonstrated the highest predictive accuracy and lowest error rates. These findings underline the value of flexible, non-linear approaches in modelling complex human behaviours such as social media usage.

It was observed that even with a rich feature space covering demographics, behavioural patterns, and engagement metrics, the majority of these features did not seem to predict the variance in the addiction level. The increase in R^2 from the BFS model, with only four features, to the Random Forest model was very limited (4%). In general, the models were only able to explain a portion of the variance in self-reported addiction scores.

Despite these constraints, the project succeeded in setting up the operational steps of a machine learning framework capable of identifying meaningful patterns in social media usage data and handling inconsistencies in the process. Additionally, this work provided valuable experience in handling real-world machine learning challenges, such as managing high-dimensional data, addressing potential bias, tuning hyperparameters, and selecting appropriate evaluation metrics.

As discussed in section 4.4, the performance of the model can still be considered moderate to good, as other predictors of addiction behaviour may not have been included in this dataset. With improvements in data quality and feature richness, similar approaches may ultimately yield a better-performing model aimed at early detection and intervention for digital addiction.

6 Future Work

A possible limitation of the current modelling approach lies in the ambiguous definition of the target variable, Addiction Level. As mentioned before, it might be beneficial if Addiction Level were instead to be considered a categorical or subjective label (e.g., "low", "moderate", "high"). The use of regression may be misaligned with the problem's nature. Therefore, for future studies, a classification framework – particularly ordinal or multiclass classification – may provide a more suitable and interpretable solution.

As mentioned in the discussion, one of the recommendations for future work would definitely be to include more psychological or behavioural predictors in the model. This would at least show the impact of those features, hence providing a better understanding of the explanation in variance of addiction by the predictors in this dataset.

Lastly, a key recommendation for future research is to move beyond subjective self-reporting and instead develop a computational definition of addiction grounded in observed user behaviour. Therefore, future research should prioritise redefining a computation of the target variable using behavioural data, such as:

- Time spent on social media platforms
- Frequency and timing of usage
- Engagement patterns (e.g., posting, liking, scrolling behaviour)

This shift would allow for the development of a more quantitative and reproducible target, potentially enabling regression models to more accurately capture addiction-related behaviours. This would first raise the question: 'What defines addiction?'. Answering this question might lead to some more stable quantitative indicators which can 'score' a user on their addiction level. Subsequently, with this more reliable addiction level score, a better-performing model based on social media usage data can be trained.

References

[1] Gallup. "U.s. teens spend average of 4.8 hours per day on social media". Accessed: 2025-05-01. (2023), [Online]. Available: https://news.gallup.com/poll/512576/teens-spend-average-hours-social-media-per-day.aspx.

[2] Odoxa, *Cyberdépendance : 14,5 millions de Français présentent une pratique à risque*, https://www.odoxa.fr/sondage/cyberdependance-145-millions-de-francais-presentent-une-pratique-a-risque/, Sondage réalisé pour GAE Conseil, publié le 19 octobre 2022, Oct. 2022. (visited on 05/01/2025).

[3] C. Cheng, O. V. Ebrahimi, and J. W. Luk, "Heterogeneity of prevalence of social media addiction across multiple classification schemes: Latent profile analysis", *Journal of Medical Internet Research*, vol. 24, no. 1, e27000, 2022. DOI: 10.2196/27000. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8787656/.

[4] M. Rashid, M. S. Alam, M. Mohiuddin, *et al.*, "A machine learning approach to predict social media addiction during covid-19 pandemic", *IEEE Access*, vol. 10, pp. 54 494–54 505, 2022. DOI: 10.1109/ACCESS.2022.3176086. [Online]. Available: https://ieeexplore.ieee.org/document/9793193.

[5] T. Ehsan and J. Basit, "Machine learning for detecting social media addiction patterns: Analyzing user behavior and mental health data", *International Journal of Innovations in Science & Technology*, vol. 6, no. 4, 2024. [Online]. Available: https://journal.50sea.com/index.php/IJIST/article/view/1070.