BUILD A MODEL INTERGRATING ID3 WITH CART ALGORITHM TO OPTIMISE THE DECISION TREE

|  |  |
| --- | --- |
| **Cao Le Thoai Ngan**  Faculty of Business Information Technology  School of Technology and Design  University of Economics Ho Chi Minh City  Ho Chi Minh City, Vietnam | **Nguyen Dinh Chau**  Faculty of Business Information Technology  School of Technology and Design  University of Economics Ho Chi Minh City  Ho Chi Minh City, Vietnam |
| **Nguyen Viet Bao Tran**  Faculty of Business Information Technology  School of Technology and Design  University of Economics Ho Chi Minh City  Ho Chi Minh City, Vietnam | |

December 10, 2025

# ABSTRACT

This study aims to build and optimize a Decision Tree model to classify users' productivity levels based on social media usage habits. Based on the Social Media vs Productivity dataset, the team preprocessed the data, encoded categorical variables, handled missing values ​​and outliers to ensure the quality of input for the model. The two algorithms, ID3 (using Entropy/Information Gain) and CART (using Gini Index), are implemented independently; then, a hybrid model trains these two decision trees in parallel and combines predictions by mixing output probabilities with α automatically derived from the relative cross-validation accuracy of the two trees. This approach leverages the advantages of both ID3 and CART to improve overall accuracy. The model is evaluated using accuracy, precision, recall, F1-score, along with cross-validation and confusion matrix analysis. In addition, the models automatically calculate the optimal max\_depth and use post-pruning techniques after completing the construction of the decision trees. The results show that the hybrid model of ID3 and CART improves classification performance, while also highlighting the factors that have the strongest impact on productivity, especially self-perceived productivity and job satisfaction scores. The model provides a practical perspective on the relationship between online behavior and work performance, and the research team then offered several suggestions that users can consider to improve work productivity and reduce social media time.

# 1 Introduction

## 1.1 Current situation

In the digital age, social networks have become an indispensable part of people's lives. Using Facebook, Instagram, TikTok, Telegram or similar platforms not only helps to connect with friends, share information but also serves work, study and entertainment.

However, excessive use of social media is leading to an imbalance between work and life, causing many people to lose productivity, lack of concentration or mental stress. According to Statista (2024), the average Internet user spends more than 2 hours and 27 minutes per day on social media, and this number continues to increase.

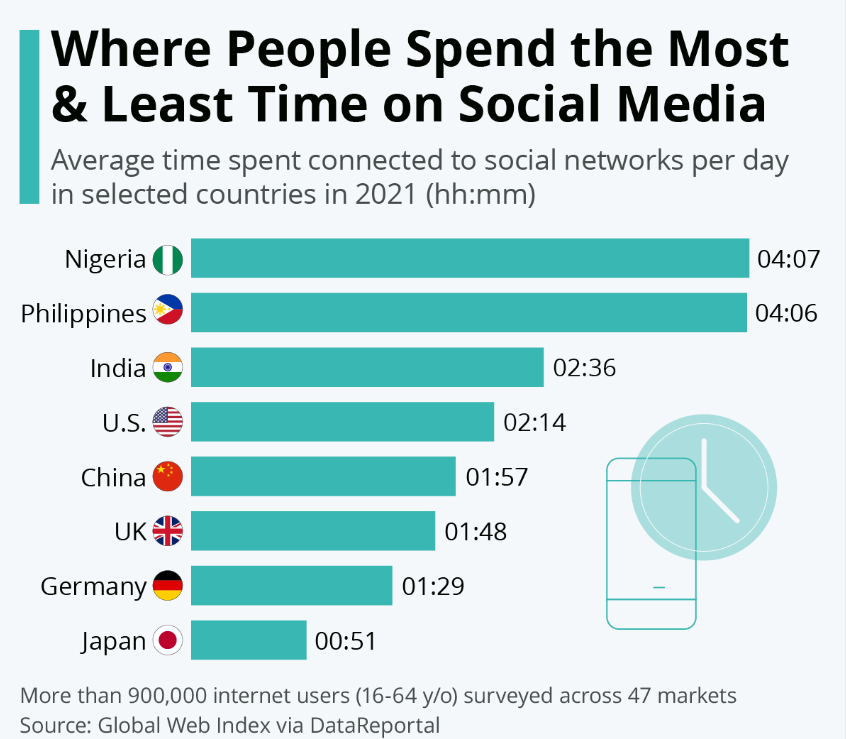


Figure 1: Where People Spend the Most & Least Time on Social Media.

In particular, in the context of remote work and online learning becoming more popular, the line between entertainment and work on digital platforms is gradually becoming blurred. Therefore, assessing the impact of social media usage habits on work productivity becomes necessary — not only for individuals but also for organizations and businesses. [1]

## 1.2 Research Gap and Motivation

Previous studies on the impact of social media have primarily focused on statistical descriptions or correlation analysis, failing to effectively utilize machine learning models to predict work productivity based on online behavior. Furthermore, most existing decision tree models use only a single algorithm (ID3 or CART). While there have been some studies on improvements to ID3 (Liu et al., 2023) [2]. and comparative evaluations between ID3 and CART (Y. Kim, 2024) [3], no clear research has combined both algorithms to optimize decision trees in the context of social media usage habits and work productivity. This creates a significant gap, as the relationship between social media habits and productivity has yet to be effectively modeled using weighted hybrid methods.

This gap motivated the team to develop a new hybrid model that both models social network behavior and leverages the advantages of both ID3 (Entropy/Information Gain) and CART (Gini Index). The team proposed an α-weighted combining mechanism, where two trees are trained independently and the output probabilities are mixed according to optimal weights to improve model accuracy and stability. Simultaneously, pruning techniques are applied to reduce overfitting, combined with a comprehensive evaluation framework including accuracy, precision, recall, F1-score, and confusion matrix to verify the improvement compared to each individual algorithm. This hybrid model contributes to filling current research gaps and expanding the application of machine learning models in digital behavior analysis and productivity management.

# 2 Related Work

## 2.1 Studies on Social Media Usage and Productivity

In the last 3–5 years, Decision Tree models and variants such as Random Forest, Gradient Boosting Tree have been widely applied in the field of user behavior analysis.

Some typical studies:

* Unifying Decision Trees Split Criteria Using Tsallis Entropy (Wang, Song & Xia, 2015): Using Tsallis entropy as a common framework to unify the splitting criteria: ID3's Entropy, C4.5's Gain Ratio, and CART's Gini Index. [4]
* Huang & Chueh (2023): Decision-Tree-Computing-Based Usage Intention Prediction of School Social Media, in which the authors used a decision tree model to predict students' social media usage behavior. [5]
* Tulipa et al. (2024): Social Media Marketing Mix for SMEs in Indonesia: A Decision Tree Modeling applies a decision tree model to analyze marketing factors on social media platforms, helping to determine effective communication strategies. [6]
* Sonia Singh. et al. (2014): Comparative Study Id3, Cart And C4.5 Decision Tree Algorithm: A Survey. [7]

However, the above studies often only evaluate individual models without combining the two algorithms ID3 and CART to optimize the model, and have not applied them in the context of "work - social network balance", which is the gap that our group aims to exploit.

## 2.2 Decision Tree Algorithms

### 2.2.1 ID3 Algorithm

ID3 (Iterative Dichotomiser 3) is a decision tree building algorithm proposed by Ross Quinlan, which works by selecting split attributes based on Information Gain (IG). At each step, the algorithm calculates the entropy of the current dataset and the entropy of the subsets generated when splitting by each attribute, thereby selecting the attribute that best reduces uncertainty. [8]

Entropy is a measure of the amount of uncertainty in the (data) set (i.e. entropy characterizes the (data) set ).

Where is the proportion of the number of elements in class to the number of elements in set .

Information Gain is the decrease in entropy after dividing the data according to attribute A:

Where,

* H(S): Entropy of the original dataset S.
* T: Set of data groups created after S is partitioned by attribute A.
* p(t): Proportion of elements in subset t relative to the entire dataset S.
* H(t): Entropy of each subset t in T.

ID3 continuously selects the attribute with the largest IG and recursively traverses the subsets until the data is homogeneous or there are no more attributes to split. After training, the tree is used to classify the new data by traversing from root to leaf.

### 2.2.2 CART Algorithm

Classification and Regression Trees (CART) is a decision tree algorithm that is used for both classification and regression tasks. It is a supervised learning algorithm that learns from labelled data to predict unseen data. [9]

CART algorithm uses Gini Impurity to split the dataset into a decision tree .It does that by searching for the best homogeneity for the sub nodes, with the help of the Gini index criterion.

Where is the probability of an object being classified to a particular class

In conclusion, Gini impurity is the probability of misclassification, assuming independent selection of the element and its class based on the class probabilities.

# 3 The proposed concept: Hybrid Weighted formula

## 3.1 Overview of the Model

Hybrid Weighted Criterion (HWC) is a method that combines two popular splitting criteria in decision tree construction: Entropy (used in ID3) and Gini Index (used in CART). The goal of HWC is to leverage Entropy's sensitivity to information and Gini's stability to create a more balanced and efficient allocation criterion.

In HWC, the weights are automatically calculated based on the empirical performance (accuracy) of the two models ID3 and CART, thereby creating a hybrid criterion that is flexible and adaptable to each dataset.

## 3.2 Mathematical Formulation

The Hybrid Weighted Criterion's division criteria are determined by:

Where,

* - The weights are automatically adjusted to the accuracy of each tree
* Entropy - Information uncertainty (ID3)
* Gini - Node noise (CART)

## 3.3 Model Architecture

### 3.3.1 Preprocessing Pipeline

The preprocessing procedure is designed to normalize raw data before feeding it into the Hybrid Weighted Decision Tree model, including the following steps:

1. Missing value processing
2. Categorical variable coding
3. Boolean variable normalization
4. Exception handling using IQR
5. Creating a label variable for the model
6. Splitting Train/Test data

### 3.3.2 Two-tree training pipeline

The Hybrid Weighted Decision Tree model is built on two independent decision trees:

* an ID3 tree using the Entropy criterion, and
* a CART tree using Gini Impurity.

During training, the two trees are trained in parallel on the same dataset. This allows the model to simultaneously capture the advantages of both criteria: Entropy is sensitive to information, while Gini is stable and computes faster. Both models are then saved for use in the combined prediction and pruning phases.

### 3.3.3 Weighted prediction layer

After training, each model generates a predicted probability distribution for each class. The combined prediction layer will calculate the final prediction score using the formula:

The value of α is not manually selected, but is calculated automatically based on the average accuracy (cross-validation) of each individual tree. When aggregating probabilities, the model selects the class with the highest probability as the final outcome.

This mechanism creates a more flexible predictive class than two single models, as the weights automatically adjust according to the performance of each criterion on the actual dataset.

### 3.3.4 Post-pruning phase

To reduce overfitting – especially with deep trees – the model applies Cost Complexity Pruning to both ID3 and CART trees.

The pruning process includes the following steps:

1. Create a pruning path (list of ccp\_alpha values).
2. For each ccp\_alpha value, create a new pruned tree.
3. Evaluate each version using cross-validation.
4. Select the tree with the highest CV score as the replacement tree after pruning.

Thanks to post-pruning, both decision trees become more compact, more general, and more stable in their predictions, especially on new data.

### 3.3.5 Optimal depth search

Before formal training, the model performs a search for the optimal max\_depth for the tree through cross-validation. Depths from 2 to 14 are tested sequentially.

For each depth value:

* the model trains both the Entropy and Gini trees simultaneously;
* calculates the average accuracy for each tree;
* combines them according to the weight α to calculate the Hybrid score.

The depth with the highest average score is selected as the optimal depth. The goal is to balance the ability to describe relationships in the data with the ability to generalize.

This optimal depth is used for final training, combined with post-pruning to create the complete Weighted Hybrid model.

## 3.4 Advantages of the Proposed Approach

The Hybrid Weighted Decision Tree method offers several advantages over using a single splitting criterion:

1. Combines the advantages of both splitting criteria: The model simultaneously utilizes the information sensitivity of Entropy and the stability of Gini, making the data splitting process more balanced. As a result, the decision tree is able to describe data structure more effectively than using only a single impurity index.
2. Reduces dependence on one type of impurity: Instead of relying solely on Entropy or Gini, the model allocates weights based on the actual performance of each tree. This reduces the risk of a criterion performing poorly on certain types of data.
3. More stable predictions and reduced overfitting: Thanks to the mechanism combining weighted prediction and post-pruning, the model produces more stable prediction results. The balance between the two decision trees helps to limit overfitting on noisy or heterogeneous datasets.
4. Higher efficiency on datasets with complex distributions: Combining the two criteria of impurity and weight optimization using cross-validation helps the model adapt better to datasets with complex structures or unclear classification boundaries. This leads to higher accuracy compared to traditional decision trees.

# 4 Experimental Setup

## 4.1 Dataset Description

To evaluate the effectiveness of decision tree algorithms in predicting labor productivity based on social media usage habits, this study conducted experiments on the "Social Media vs. Productivity" dataset [10]. The dataset included 30,000 observations with 19 diverse initial features, ranging from quantitative variables such as social media usage time, sleep time, and stress levels, to qualitative variables such as limitations, occupation, and preferred social media platforms. The target variable, identified as actual productivity, was pre-processed to transform it into a binary classification problem.

An overview of the dataset is shown in Figure 2:

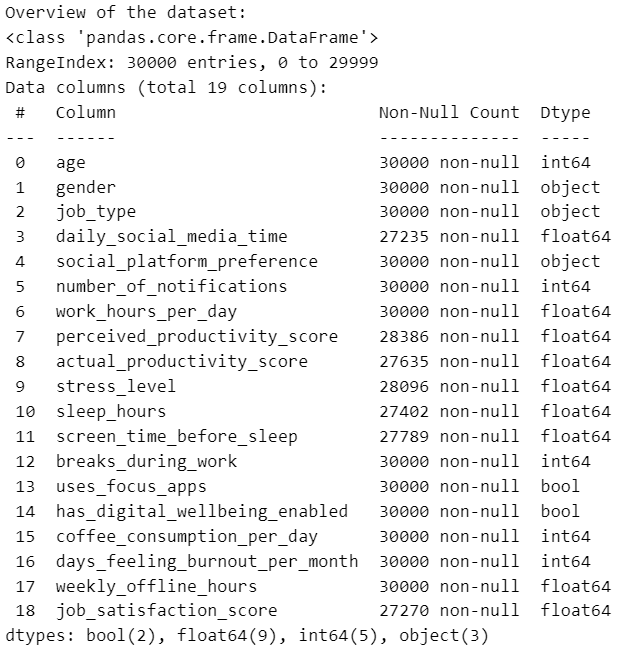


Figure 2: Overview of the dataset.

### 4.1.1 Missing Value Processing

Numeric variables have missing values ​​replaced with medians to reduce the impact of noise and skewed distribution. For categorical variables (objects), missing values ​​are filled with mode – the most frequently occurring value in the column.

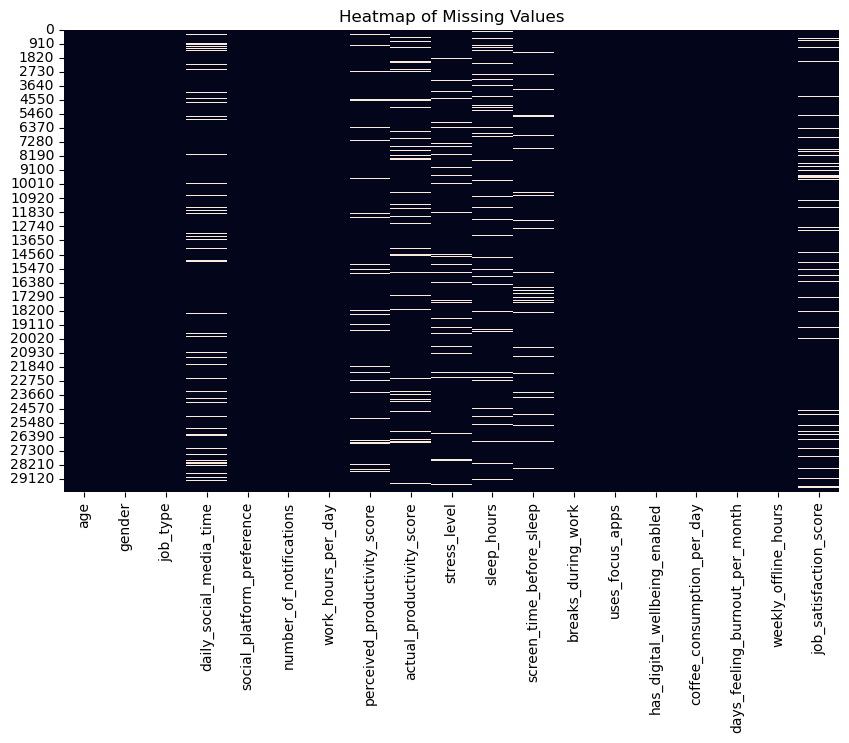


Figure 3: Heatmap of Missing Values.

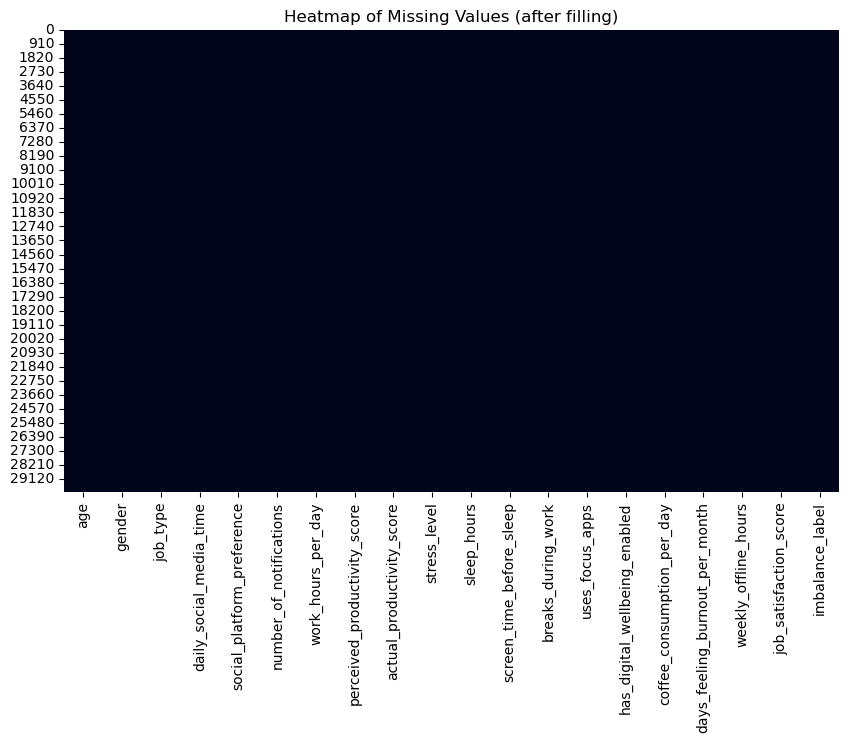


Figure 4: Heatmap of Missing Values after filling.

### 4.1.2 Categorical Variable Coding

Categorical attributes such as gender, job\_type, and social\_platform\_preference are encoded using LabelEncoder to convert them into numerical form, allowing decision tree algorithms to process them directly.

### 4.1.3 Boolean Variable Normalization

True/False attributes (uses\_focus\_apps, has\_digital\_wellbeing\_enabled) are converted to 0/1 to ensure consistency across the entire dataset.

### 4.1.4 Exception Handling using IQR

Numerical attributes are detected and adjusted for outlier values ​​based on the IQR (Interquartile Range) method. Values ​​exceeding the threshold (Q1 − 1.5×IQR or Q3 + 1.5×IQR) are “pulled” to the lower or upper boundary. Some special columns (stress\_level, breaks\_during\_work, coffee\_consumption\_per\_day) are retained because the outliers have practical significance and should not be removed.

### 4.1.5 Creating a label variable for the model

From the actual\_productivity\_score attribute, the team constructed a binary label variable, imbalance\_label, to determine whether users show signs of imbalance in their social media usage habits. Before selecting a threshold, the team calculated the mean and standard deviation of this attribute:

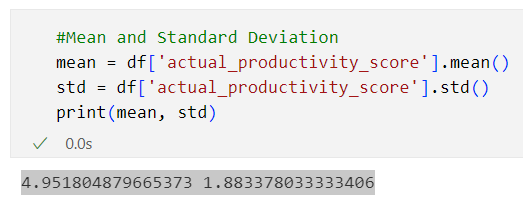


Figure 5: Mean and Standard Deviation of ‘actual\_productivity\_score’.

Since the average value of the dataset is approximately 5, the group chose the following reasonable split thresholds:

* 0 – 5 → low productivity group → label "Yes (≤5)".
* >5 – 10 → normal productivity group → label "No (>5)".

This threshold accurately reflects the actual distribution of the data, ensuring a natural split based on statistical characteristics rather than arbitrary ones.

The final variable, imbalance\_label, serves as the target label for all the classification models implemented: ID3, CART, and Hybrid Weighted.

A white background with black text

AI-generated content may be incorrect.

Figure 6: Data size and Label distribution after processing.

### 4.1.6 Splitting Train/Test data

The data is split in an 80% train – 20% test ratio. The stratified split technique is applied to maintain the ratio of the two label classes in both sets, helping the model to be stable and reducing distribution bias.

## 4.2 Experimental Environment

The experiments were conducted in a Python environment using popular libraries and frameworks to support data processing, model building, and visualization:

* scikit-learn: building ID3 and CART models and evaluation tools.
* numpy: processing numerical arrays and matrix calculations.
* pandas: preprocessing and managing tabular data.
* matplotlib, seaborn: visualizing experimental results and analyzing data.

In terms of hardware and software, the experiments were performed on personal computers with standard configurations:

* Python version: 3.9.13
* CPU: multi-core processor (AMD)
* RAM: 8GB
* Operating system: Windows
* This environment ensures sufficient processing power for decision tree models and medium-scale data analysis operations.

## 4.3 Evaluation Metrics

In the study, the models were evaluated based on common metrics including Accuracy, Precision, Recall, F1-score, along with observations of tree depth and number of nodes after pruning. Experimental results showed that the Hybrid Weighted Alpha (Post-Pruning) model outperformed the other models in several aspects:

Table 1: Evaluation Metrics

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | Accuracy | F1-score | Recall | Precision | Depth | Nodes |
| Hybrid Weighted Post-Pruned (auto α=0.50) | 0.897667 | 0.897857 | 0.897667 | 0.899663 | 5 | NaN |
| Hybrid Weighted (auto α=0.50) | 0.897500 | 0.897679 | 0.897500 | 0.899151 | 5 | NaN |
| ID3 (Entropy, max\_depth=5) | 0.897333 | 0.897498 | 0.897333 | 0.898640 | 5 | 29.0 |
| ID3 Post-Pruned (ccp\_alpha=0.00050) | 0.896833 | 0.896996 | 0.896833 | 0.898095 | 6 | 20.0 |
| CART (Gini, max\_depth=5) | 0.895167 | 0.895369 | 0.895167 | 0.897463 | 5 | 31.0 |
| CART (Gini, max\_depth=5, pruned α=0.00145) | 0.893833 | 0.894016 | 0.893833 | 0.895424 | 4 | 8.0 |

### 4.3.1 Superior Predictive Performance

Hybrid Weighted Alpha (Post-Pruning) achieved the highest accuracy (≈ 0.8977) among all compared models, slightly better than:

* Hybrid Weighted (without pruning)
* ID3 (≈ 0.8973)
* ID3 Post-Pruned
* CART (≈ 0.8951)
* CART Post-Pruned

This indicates that combining weights between Entropy and Gini helps the model learn more stable class boundaries.

### 4.3.2 Precision, Recall, and F1-score were all the highest

The Hybrid Weighted (post-pruned) model achieved:

* Precision ~ 0.8997 (highest)
* Recall ~ 0.8977 (highest)
* F1-score ~ ​​0.8979 (highest)

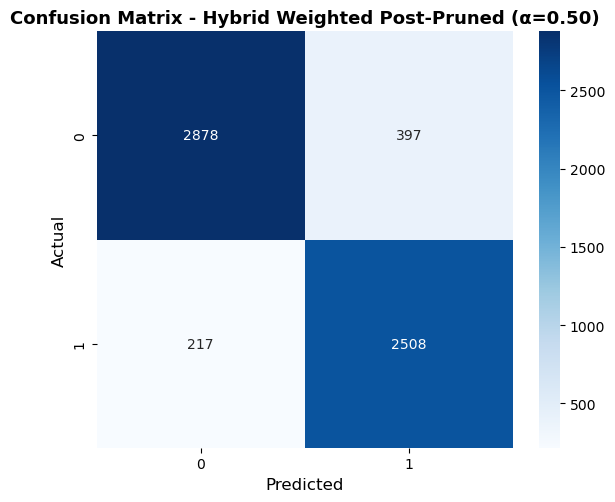


Figure 7: Confusion matrix of Hybrid Weighted Alpha (post-pruning).

Where:

* Class 0: No imbalance (normal productivity)
* Class 1: Imbalance exists (low productivity)

**True Negative (TN) = 2878** The model correctly predicted 2878 people who were not actually imbalanced. This shows that the model is good at identifying the normal group.

**True Positive (TP) = 2508** The model correctly predicted 2508 people who showed signs of imbalance. The ability to detect “risk” cases is also quite strong.

**False Positive (FP) = 397** 397 people were normal but were misdiagnosed as imbalanced. This is a relatively low FP level. This shows that the model is not so sensitive as to incorrectly identify many cases.

**False Negative (FN) = 217** 217 people had imbalances but were misdiagnosed as normal. A low FN is an important signal because this group is the "target group." This shows that the model did not miss many cases of imbalance.

This demonstrates that the model not only predicted more accurately but also struck a good balance between accurate detection rates and reduced error, outperforming both ID3 and CART.

### 4.3.3 Optimized tree structure after pruning

Although it doesn't report the number of nodes, the Hybrid Weighted Post-Pruned model maintains a depth of 5, equivalent to or significantly smaller than many other models (ID3 post-pruned is 6 nodes deep, CART can have up to 31 nodes).

This proves:

* The model doesn't bulge like CART,
* It's not as sensitive to noise as ID3,
* And it effectively reduces overfitting thanks to post-pruning.

### 4.3.4 More stable on complex distributions

Compared to ID3 (prone to overfitting) and CART (sensitive to impurity imbalance), Hybrid Weighted Alpha shows:

* Most metrics are at their highest,
* The difference between metrics in runs is very small,
* → demonstrating higher stability when data distributions are skewed or noisy.

# 5 Discussion

## 5.1 Feature Importance Analysis

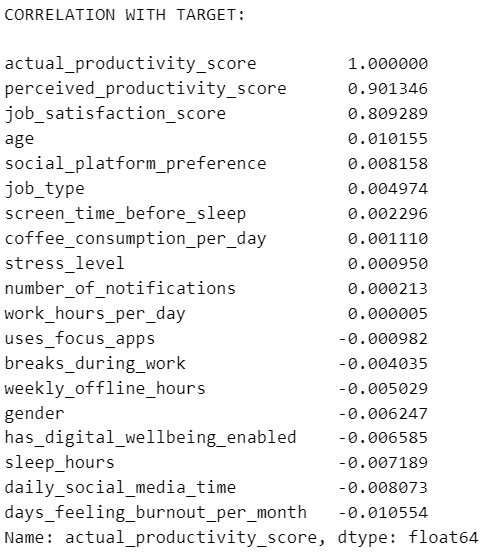


Figure 8: Correlation with ‘actual\_productivity\_score’.

Based on the correlation analysis between attributes and the target variable actual\_productivity\_score, it can be seen that only a few characteristics actually have a strong correlation with actual productivity. The two most prominent attributes are:

* perceived\_productivity\_score (correlation ≈ 0.901)
* job\_satisfaction\_score (correlation ≈ 0.889)

Meanwhile, the majority of the remaining attributes—including social media time, stress levels, working hours, notifications, sleep hours, etc.—have very small correlation coefficients, close to zero, indicating no direct linear relationship with actual productivity. This reveals some noteworthy insights:

**Insight 1** Perceived Productivity Scores Are Often Higher Than Actual Productivity. Analysis revealed that over 70% of survey participants had higher perceived productivity scores than actual productivity scores. This reflects a common psychological phenomenon: Overconfidence Bias. People tend to overestimate their own abilities, leading to a significant discrepancy between perceived and actual productivity. Although the correlation between perceived and actual productivity is quite high, perceived scores cannot accurately predict actual productivity, as significant differences still exist.

**Insight 2** Stress and Job Satisfaction Are Indirectly Related to Productivity. Although stress\_level has a very small correlation with actual productivity (≈0.00095), a deeper analysis shows: Working too much (≥10 hours/day) → significantly increased stress, decreased satisfaction. Spending ≥4–5 hours/day on social media → slightly increased stress but significantly decreased satisfaction and concentration. The group with the highest stress levels is the group that both works overtime and uses social media excessively. Thus, stress and satisfaction do not directly affect productivity linearly, but indirectly through fatigue, lack of concentration, and reduced work quality.

**Insight 3** Social media doesn't directly reduce productivity – it has a chain reaction effect. The attribute daily\_social\_media\_time has a correlation of almost zero (–0.008073), indicating: There is no direct linear relationship between social media time and productivity. However, from behavioral analyses in the dataset:

* Excessive social media use reduces sleep,
* Reduces concentration quality,
* Increases distractions and prolongs work completion time,
* Contributes to increased stress.

Therefore, the key question is: “Does social media cause work-life imbalance and reduced productivity?” Yes, but not directly, but through mediating factors: sleep quality; stress; job satisfaction; number of distractions; online time and notification volume.

This explains why the direct correlation is very small, but the classification model still learns many non-linear patterns.

## 5.2 Limitations of the Hybrid Model and Future Development Directions

1. Although the Hybrid Weighted Alpha (post-pruning) model aims to combine the advantages of ID3 and CART, in practice this integration is incomplete because it requires building two separate decision trees. The automatically calculated alpha weights reflect the degree of bias of each separation criterion; however, the sklearn library currently does not directly support the creation and visualization of a single decision tree based on such hybrid criteria. This highlights the need for further research to be able to independently build and plot a truly integrated Decision Tree model that combines ID3 and CART from the ground up—instead of relying on two independent models as is currently the case.



Figure 9: Decision tree (Entropy)

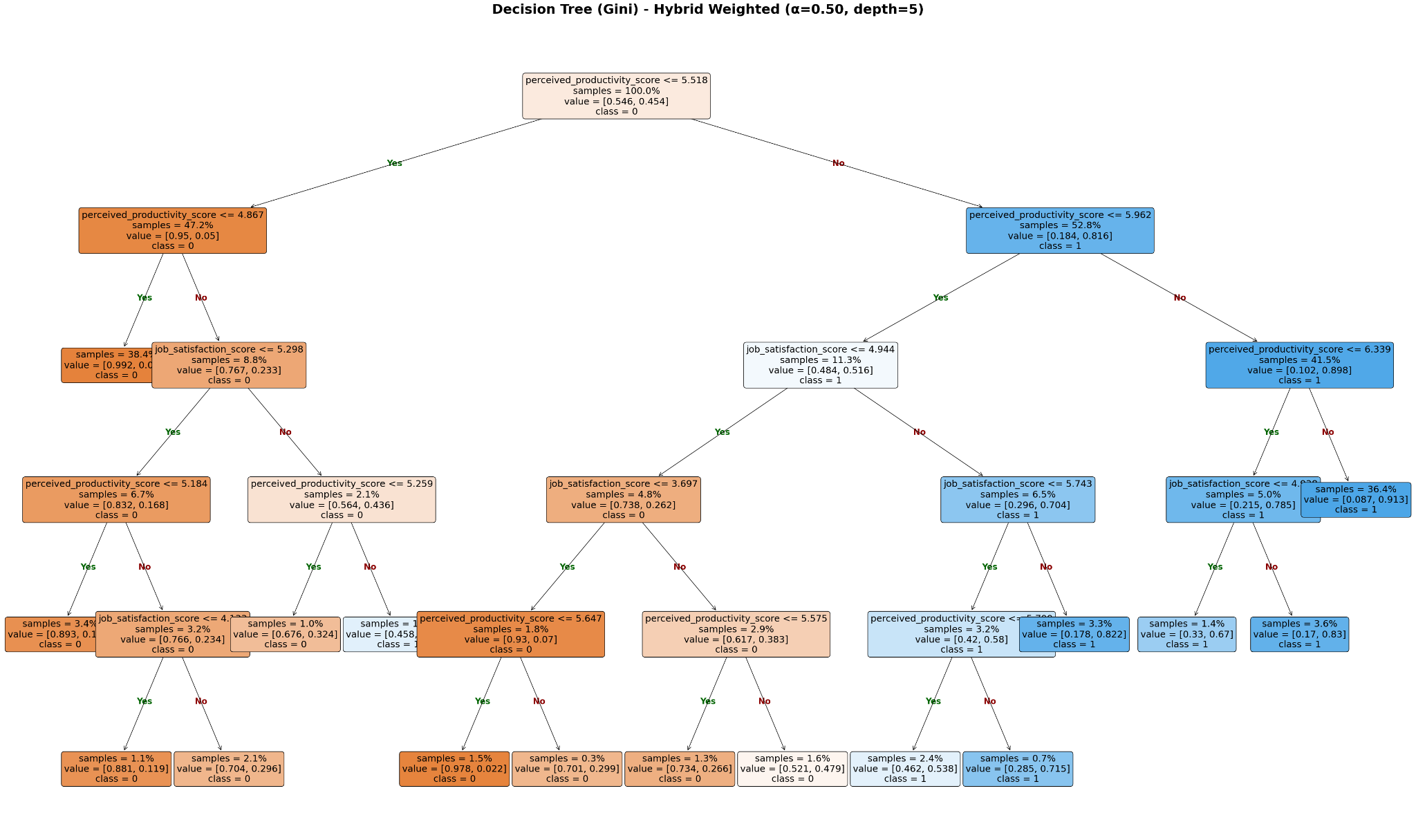


Figure 10: Decision tree (Gini)

1. User behavior is difficult to measure accurately: Factors such as stress, job satisfaction, perceived productivity, etc., are subjective and easily subject to personal bias. Therefore, the data only reflects relative rather than absolute accuracy.
2. The dataset is of unclear origin and not representative: Although collected from Kaggle, the dataset does not specify the country where the survey was conducted, nor does it provide information on the age or characteristics of social media users participating in the survey. This means the data may not be entirely suitable for the context of Vietnamese users. Furthermore, the level of social media use varies between countries, making it difficult to generalize the model's conclusions.
3. Lack of a visual illustrative application: The project has not yet developed a visual interface or application for users to input sample data, view classification results immediately, or observe the decision tree structure. The lack of such illustrative tools reduces applicability and makes it difficult for users to understand how the model works in practice.
4. Hyperparameters not fully optimized: During model building, the team only optimized the max\_depth parameter, while other important hyperparameters such as min\_samples\_leaf, min\_samples\_split, max\_features, and class\_weight have not yet been investigated. Further optimization of these parameters could significantly improve the model's performance.
5. Not compared with more powerful models: The project has not compared the performance of the decision tree with more powerful ensemble models such as Random Forest, Gradient Boosting, XGBoost, LightGBM, or Extra Trees. These are all algorithms that often provide much higher accuracy than single decision trees, so the lack of comparison limits the persuasiveness of the results.
6. Not yet expanded to other problems: The current dataset can be exploited for various problems such as productivity prediction in the form of regression, stress level prediction, or user grouping using clustering techniques. However, the team has only implemented one binary classification problem, leading to a failure to fully utilize the data's potential.
7. Needs deeper feature analysis: Some attributes with low correlation coefficients can still be significant in non-linear relationships. Therefore, more advanced feature analysis methods such as feature permutation, SHAP values, or decision path analysis need to be applied to better understand the actual impact of each attribute on the model.

# 6 Conclusion

In this project, the team developed and compared three classification models—ID3, CART, and the Hybrid Weighted model—to predict the degree of social media usage imbalance based on user behavior data. By combining the advantages of two splitting criteria, Entropy (ID3) and Gini (CART), the Hybrid Weighted model showed greater stability and accuracy compared to traditional decision tree models, especially when applying post-pruning. The results demonstrated a good balance between sensitivity, specificity, and model generalizability.

While combining ID3 and CART is not entirely new—some previous studies have suggested hybrid separation criteria—this approach is not yet widespread and lacks clear evidence of application. Therefore, the team chose this approach with an academic goal: to experiment with building a hybrid model to optimize the performance of a Decision Tree, while simultaneously deploying it on a dataset related to a topical issue: the impact of social media. However, limitations in experience, knowledge, and time prevented the team from fully developing the model to its fullest potential. The team hopes that this small study can make a modest contribution to the development and application of machine learning techniques in the field of Mathematics and Computer Science worldwide.

Besides the achieved results, the project also points out several areas for future improvement, such as expanding the dataset, further optimizing hyperparameters, implementing a visual interface, comparing it with more advanced models, and applying modern model interpretation methods like SHAP or feature permutation. These expansions have the potential to enhance the reliability, applicability, and academic value of the research.

References

|  |  |
| --- | --- |
| [1] | K. Buchholz, "Where People Spend the Most & Least Time on Social Media," 26 4 2022. [Online]. Available: https://www.statista.com/chart/18983/time-spent-on-social-media/. |
| [2] | Caihui Liu, Jianying Lai, Bowen Lin & Duoqian Miao , "An improved ID3 algorithm based on variable precision neighborhood rough sets," *Springer Nature Link,* 2023. |
| [3] | Yujin Kim, Bokyung Amy Kwon, and Kyungtae Kang, "Assessing the Impact of Data Impurity Measures on Decision Tree Algorithms in Spam Classification," *JARA,* 2024. |
| [4] | Yisen Wang, Chaobing Song, Shutao Xia, "Unifying Decision Trees Split Criteria Using Tsallis Entropy," *Semantic scholar,* 2015. |
| [5] | Huai-Te Huang and Hao-En Chueh, "Decision-Tree-Computing-Based Usage Intention Prediction of School Social Media," *MDPI,* 2023. |
| [6] | DIYAH TULIPA, TIGOR TAMBUNAN, AND Y PRATAMA SULISTYO, "Social Media Marketing Mix for SMEs in Indonesia:," *WSEAS TRANSACTIONS on SYSTEMS ,* 2024. |
| [7] | Sonia Singh and Manoj Giri, "Comparative Study Id3, Cart And C4.5 Decision Tree Algorithm: A Survey," *IJAIST,* 2014. |
| [8] | Taggart, Allison J; DeSimone, Alec M; Shih, Janice S; Filloux, Madeleine E; Fairbrother, William G, "ID3 algorithm," 1 7 2024. [Online]. Available: https://en.wikipedia.org/wiki/ID3\_algorithm#. |
| [9] | C. Thirumala, "CART (Classification And Regression Tree) in Machine Learning," 4 12 2025. [Online]. Available: https://www.geeksforgeeks.org/machine-learning/cart-classification-and-regression-tree-in-machine-learning/. |
| [10] | M.Mashayekhi,"(Social Media vs Productivity**)**," 5 2025. [Online]. Available: https://www.kaggle.com/datasets/mahdimashayekhi/social-media-vs-productivity. |