Proceedings of the 2nd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/109/2024.17893

Investigating the Impact of Age and Education on Delay Discounting: A Predictive Random Forest Model for Personalized Marketing and Mental Health Risk Detection

Huazhang Lu

University of Electronic Science and Technology of China, Chengdu, 611731, China

2022110901023@std.uestc.edu.cn

Abstract. This study uses a random forest model to predict delay discounting rates, aiming to support mental health interventions and personalized marketing strategies. Delay discounting refers to the tendency to prefer smaller immediate rewards over larger delayed rewards, a behavior closely linked to impulsivity and relevant in both psychological and commercial contexts. Data from Garofalo et al., including 357 healthy Italian adults categorized by age and education level, were used. Two-way ANOVA showed that education significantly influences delay discounting rates, while age does not. The model was utilized to anticipate delay discounting behaviors influenced by various factors. The random forest algorithm reached an impressive accuracy rate of 92%, showing particular proficiency in forecasting both high and low delay discounting classifications. Additionally, the model successfully pinpointed individuals whose actual discounting rates diverged significantly from the predictions, offering crucial insights for proactive mental health interventions. Based on the predictions, personalized strategies were recommended for individuals with different delay discounting rates, demonstrating the value of this predictive model in the fields of mental health and commercial marketing.

Keywords: Delay discounting, Random forest, Predictive model, Personalized marketing, Mental health.

1. Introduction

When presented with the choice between receiving \$80 immediately or \$100 in one week, how would people choose? Delay discounting refers to the tendency to prefer smaller immediate rewards over larger delayed ones when balancing time and reward size. The rate of delay discounting quantifies this preference, where a greater rate signifies a heightened tendency to favor immediate gratification over delayed rewards.[1] This conduct is intricately associated with impulsivity and holds significant importance in psychological research, especially concerning mental health studies. [2] Over the years, investigations into delay discounting have evolved from their psychological origins to encompass applications within the realm of commercial marketing..[3]

In existing delay discounting research, while numerous studies have explored the relationship between delay discounting and psychological or behavioral factors, there is still a lack of studies using machine learning to effectively predict discounting rates. This study aims to address this gap by employing a random forest model to predict individuals' delay discounting rates based on their age and

DOI: 10.54254/2755-2721/109/2024.17893

education levels. Delay discounting rates were divided into five stages based on the dataset distribution: low (below 10%), lower (10%-30%), moderate (30%-70%), higher (70%-90%), and high (above 90%). The model can predict and classify discounting levels based on age and education.

The research data comes from the Italian sample by Garofalo et al. [4], including 357 healthy adults. The study analyzed the distribution of delay discounting rates across different education and age groups. Two-way ANOVA results showed that education significantly influenced delay discounting, while age was not significant. Based on this, the random forest model developed in this study accurately predicts delay discounting categories using age and education information, enabling personalized marketing strategies for individuals with different discounting rates. Furthermore, the model identifies individuals whose actual discounting rates deviate significantly from predictions, offering psychological health alerts and supporting early interventions for potential mental health issues.

2. Literature review

The study by Garofalo et al. provides a crucial data foundation for this paper [4]. The research examined delay discounting (DD) behavior in 357 healthy Italian adults, offering comprehensive demographic data including gender, age, education, and DD tendencies. Garofalo et al. classified participants into three age groups (21-39, 40-59, 60+) and three educational levels (3-8, 9-13, 14-23 years). This paper utilizes these classifications to analyze DD behavior across demographic segments.

In their data analysis, Garofalo et al. found that education had a significant impact on delay discounting rates, with individuals with higher education levels showing a stronger preference for delayed rewards and thus exhibiting lower discounting rates. In contrast, age was found to have no significant effect.

This paper utilizes the Area Under the Curve (AUC) method to measure individual delay discounting rates. AUC calculates the area under the delay discounting curve, reflecting an individual's preference for rewards at different time points. The formula is as follows: $\mathrm{AUC} = \sum_{i=1}^{n-1}(x_{i+1} - x_i) \times \frac{y_i + y_{i+1}}{2}$ .where x represents the delay time, and y represents the corresponding discount rate. Reduced AUC values suggest a greater inclination towards immediate rewards, which correlates with elevated rates of delay discounting. [5].

High delay discounting rates indicate greater vulnerability to impulsive decision-making, often associated with addiction, gambling, substance abuse, and ADHD[6]. In addiction studies, this measure serves as a key predictor of impulsivity, aiding psychologists in identifying those drawn to immediate rewards and facilitating intervention. Conversely, lower rates may reflect better self-regulation, enabling planning for future gains[7]. However, excessively low rates could correlate with psychological concerns like OCD, anxiety, and perfectionism.

The examination of delay discounting rates has significant implications in business, especially in marketing. Those with high delay discounting are more influenced by flash sales and limited-time offers, as they prioritize immediate rewards. Businesses can boost purchasing behavior by creating urgency through time-sensitive promotions[8]. In contrast, individuals with low delay discounting prefer long-term rewards, like membership programs and delayed promotions, and they tend to prioritize long-term planning over short-term incentives[9]. However, offering substantial long-term value can foster strong consumer loyalty in this group. For those with moderate delay discounting, promotional strategies that effectively combine short-term incentives with long-term benefits yield the best results.

This paper's model builds upon the aforementioned studies.

3. Methodology

3.1. Dataset and Two-Way ANOVA

The dataset used in this study is sourced from Garofalo et al. , comprising 357 healthy Italian adults [4]. The sample characteristics show that there are 91 participants in the 21-39 age group, 140 in the 40-59 age group, and 126 in the 60 and above age group. In terms of education, 102 participants had 3-8 years of education, 128 had 9-13 years, and 127 had 14-23 years of education. The average age of the sample is 52.18 years (SD = 16.71), and the average years of education is 12.69 years (SD = 4.36). These statistics provide a basic demographic profile of the sample population.

During the data preprocessing phase, individuals were grouped according to their original categories based on age and education levels, resulting in nine groups. The top and bottom 5% of AUC values in each group were removed to minimize the impact of outliers. The mean AUC for each group was then calculated as a representative measure of delay discounting behavior for that group. The processed results are presented in Table 1.

To examine the effects of age and education level on delay discounting, a two-way analysis of variance (ANOVA) was conducted. Two-way ANOVA is a statistical method commonly used to study the effects of multiple factors on a dependent variable.[10] In this study, the main effects of age and education on delay discounting rates (AUC) were analyzed to assess their independent influences.

The two-way ANOVA was conducted using the following model:

$$Y = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \tag{1}$$

where Y is the dependent variable (AUC value), μ represents the overall mean effect, α_i is the effect of the i-th age group, β_j is the effect of the j-th education level group, and ε_{ijk} is the residual error. This model allows for testing whether age, education, and their interaction have significant effects on delay discounting rates.

Table 1. Mean AUC Values for Different Age and Education Level Groups

Age.lvl \ Edu.lvl (AUC)	1	2	3
1	0.344549583	0.361485696	0.44655766
2	0.331341166	0.442463011	0.47229565
3	0.359591875	0.375203232	0.439958167

3.2. Development and application of the random forest model

In this study, we utilize a random forest model to predict delay discounting rates and provide personalized recommendations and mental health alerts for individuals with different discounting rates. Random forest is an ensemble learning algorithm that improves classification accuracy and robustness by constructing multiple decision trees and combining their predictions. In contrast to an individual decision tree, a random forest significantly mitigates overfitting while preserving robust classification efficacy in high-dimensional datasets.[11]

Given that the random forest algorithm excels at handling multi-dimensional feature sets and effectively captures the interactions between various factors influencing delay discounting rates, it is regarded as one of the best choices for completing this task[8]. Additionally, the random forest has the ability to assess feature importance, helping to identify which variables are most critical in predicting delay discounting rates[7].In the implementation, the dataset was first preprocessed. To ensure that the data could be recognized and processed by the model, categorical variables were converted into numerical ones using One-Hot encoding. Next, the AUC values were divided into percentiles, classifying the sample delay discounting rates into five levels: low, lower, moderate, higher, and high. For data splitting, a stratified sampling strategy was used to ensure that the training and test sets maintained the same class proportions. This method divided the data into 80% training set and 20% test set. Stratified sampling helps ensure that each category is proportionally represented in both the training and test sets, preventing sample imbalance from affecting model performance. The model was developed utilizing Python's 'sklearn.ensemble.RandomForestClassifier', a widely adopted library for executing the random forest algorithm. It generates classification predictions employing an ensemble

DOI: 10.54254/2755-2721/109/2024.17893

voting system across numerous decision trees. During the training phase, the model was fit to the training set, where the decision trees partitioned the features, and the results of the trees were aggregated to produce the final prediction. The ensemble learning mechanism of random forest effectively handles multi-dimensional features and reduces overfitting. In the evaluation phase, the model's accuracy and classification performance were assessed on the test set.

Based on the classification results, personalized push notifications were proposed for individuals with different delay discounting rates. For individuals classified with a "high delay discounting rate," the recommendation is to push short-term promotional offers, such as limited-time discounts and flash sales, as these individuals are more inclined to prefer immediate rewards. In contrast, individuals exhibiting a "low delay discounting rate" are better candidates for long-term incentives or high-value products, since they are inclined to prioritize long-term rewards. Conversely, for those with a "moderate delay discounting rate," a promotional strategy that harmonizes both short-term and long-term offers is advisable to align with their consumption preferences.

By applying the random forest model, this study not only accurately predicts individual delay discounting rates but also provides scientific support for personalized commercial notifications and mental health interventions.

4. Results

4.1. Results of ANOVA analysis

According to the results of the ANOVA, education level had a significant effect on delay discounting rates. The sum of squares for education was 0.017496, with 2 degrees of freedom, an F-value of 11.1142, and a p-value of 0.0233, which is below the significance level of 0.05. This indicates that education significantly influences delay discounting rates, with individuals who have higher levels of education tending to delay gratification and exhibit lower discounting rates. In contrast, the effect of age on delay discounting was not significant. The sum of squares for age was 0.001592, with 2 degrees of freedom, an F-value of 1.0111, and a p-value of 0.4412, which is above the significance level of 0.05. This suggests that there is no significant difference in delay discounting behavior across age groups. The residual sum of squares was 0.003148, with 4 degrees of freedom, indicating a small portion of unexplained variance in the model.

Based on these statistical analyses, we conclude that education level has a significant influence on delay discounting behavior, while the effect of age is not significant. These findings are consistent with the conclusions of Garofalo et al., further supporting the notion that education is a key factor influencing delay discounting behavior.

4.2. Results of the random forest model prediction

In the random forest model's prediction of delay discounting rate categories, the test set contained a total of 72 samples across five delay discounting categories: low delay discounting (7 samples), lower delay discounting (15 samples), moderate delay discounting (29 samples), higher delay discounting (14 samples), and high delay discounting (7 samples). The overall classification accuracy of the model was 92%, with particularly strong performance in the high and low delay discounting categories. The classification report showed that the precision, recall, and F1-score for both the high and low delay discounting categories were 1.00. In contrast, the model’s performance fluctuated slightly for the higher and lower delay discounting categories. The recall for the higher delay discounting category was 0.86, while both the precision and recall for the lower delay discounting category were 0.87, resulting in an F1-score of 0.87. For the moderate delay discounting category, the model achieved a precision of 0.88, a recall of 0.97, and an F1-score of 0.92, indicating a few misclassifications. Detailed results are shown in Table 2.

The model not only performs classification but also has the capability to identify outlier individuals. By analyzing samples in the test set where the actual delay discounting rates significantly deviated from the predicted values, the model identified one key outlier (Sample No. 139). This sample represents a

DOI: 10.54254/2755-2721/109/2024.17893

49-year-old female with 17 years of education, whose actual delay discounting rate was classified as "low," but the model misclassified her as "lower." This misclassification suggests that the individual may have tendencies toward excessive control or compulsive behavior, recommending further psychological evaluation to preempt potential mental health issues. By identifying such outliers, the model not only excels in classification tasks but also provides a foundation for personalized mental health alerts.

The model provided customized marketing strategies for accurately classified samples. For individuals with low delay discounting rates, it recommended long-term discounts and premium offerings. Those with high rates were suggested time-sensitive promotions, while a balanced strategy combining short-term incentives and long-term value was advised for moderate rates. It also generated personalized push notifications for 66 correctly classified samples, demonstrating its practicality in real-world applications.

To ensure the robustness of the model, five-fold cross-validation[12] was used for further evaluation. The cross-validation results showed accuracy rates of 0.875, 0.861, 0.859, 0.845, and 0.944, with an average accuracy of 88% and a standard deviation of 0.03. These results indicate that the model performed consistently across different data splits, with minimal fluctuation in accuracy, demonstrating strong generalization capability and robustness.

Table 2. Classification Report for Random Forest model

	Precision	Recall	F1-score	Support
high_delay_discounting	1.00	1.00	1.00	7
higher_delay_discounting	1.00	0.86	0.92	14
low_delay_discounting	1.00	0.86	0.92	7
lower_delay_discounting	0.87	0.87	0.87	15
medium_delay_discounting	0.88	0.97	0.92	29
accuracy			0.92	72
macro avg	0.95	0.91	0.93	72
weighted avg	0.92	0.92	0.92	72

5. Conclusion

The core of the study is to investigate the impact of education level and age on delay discounting rates based on the sample, and to build a random forest model using these two features for classification prediction of discounting rates. Through two-way ANOVA, the results showed that education level had a significant effect on delay discounting rates, while age did not have a significant impact. In the classification task, the random forest model demonstrated high accuracy. The model not only categorized delay discounting behavior but also pinpointed individuals whose actual delay discounting rates markedly diverged from the anticipated values, establishing a foundation for early intervention in mental health.

The model also provided personalized commercial recommendations based on delay discounting rates. For those with low rates, it suggested long-term discount plans, while high-rate individuals were recommended limited-time offers. Moderate rate individuals received a mix of short-term incentives and long-term value strategies. This personalized marketing approach underscores the model's potential in mental health applications and commercial utility. However, the study has limitations. The sample was confined to specific regions in Italy, potentially limiting result generalizability due to cultural differences. Future research should involve diverse populations to assess the model's applicability. Moreover, only age and education were considered, neglecting influences from economic status, family background, and personality traits[1][4][13]. Future studies should incorporate additional variables to enhance the model's predictive accuracy.

References

[1] Odum, A. L. (2011). Delay discounting: trait variable?. Behavioural processes, 87(1), 1-9.

[2] Rung, J. M., & Madden, G. J. (2018). Experimental reductions of delay discounting and impulsive choice: A systematic review and meta-analysis. Journal of experimental psychology: general, 147(9), 1349.
[3] Oliveira, L. L., & Green, L. (2012). Discounting and impulsivity: Overview and relevance to consumer choice. Handbook of developments in consumer behaviour, 285-322.
[4] Garofalo, S., Degni, L. A., Sellitto, M., Braghittoni, D., Starita, F., Giovagnoli, S., & Benassi, M. (2022). Unifying evidence on delay discounting: Open task, analysis tutorial, and normative data from an Italian sample. International Journal of Environmental Research and Public Health, 19(4), 2049.
[5] Beck, R. C., & Triplett, M. F. (2009). Test–retest reliability of a group-administered paper–pencil measure of delay discounting. Experimental and clinical psychopharmacology, 17(5), 345.
[6] Amlung, M., Marsden, E., Holshausen, K., Morris, V., Patel, H., Vedelago, L., ... & McCabe, R. E. (2019). Delay discounting as a transdiagnostic process in psychiatric disorders: A meta-analysis. JAMA psychiatry, 76(11), 1176-1186.
[7] Shamosh, N. A., DeYoung, C. G., Green, A. E., Reis, D. L., Johnson, M. R., Conway, A. R., ... & Gray, J. R. (2008). Individual differences in delay discounting: relation to intelligence, working memory, and anterior prefrontal cortex. Psychological science, 19(9), 904-911.
[8] Kim, S., Yoon, S., Baek, T. H., Kim, Y., & Choi, Y. K. (2021). Temporal and social scarcities: effects on ad evaluations. International Journal of Advertising, 40(7), 1115-1134.
[9] Lewis, M. (2004). The influence of loyalty programs and short-term promotions on customer retention. Journal of marketing research, 41(3), 281-292.
[10] Sawyer, S. F. (2009). Analysis of variance: the fundamental concepts. Journal of Manual & Manipulative Therapy, 17(2), 27E-38E.
[11] Ali, J., Khan, R., Ahmad, N., & Maqsood, I. (2012). Random forests and decision trees. International Journal of Computer Science Issues (IJCSI), 9(5), 272.
[12] Valente, G., Castellanos, A. L., Hausfeld, L., De Martino, F., & Formisano, E. (2021). Cross-validation and permutations in MVPA: Validity of permutation strategies and power of cross-validation schemes. Neuroimage, 238, 118145.
[13] MacKillop, J. (2013). Integrating behavioral economics and behavioral genetics: delayed reward discounting as an endophenotype for addictive disorders. Journal of the experimental analysis of behavior, 99(1), 14-31.