Jurnal Teknologi Pendidikan, April 2025, 27 (1), 277-290
DOI: http://dx.doi.org/10.21009/JTP2001.6
p-ISSN: 1411-2744 e-ISSN: 2620-3081
Accredited by Directorate General of Strengthening for Research and Development
Available online at http://journal.unj.ac.id/unj/index.php/jtp

Student's Digital Intentions Prediction Using CatBoost

Zulwisli[1], Ambiyar[2], Muhammad Anwar[3], Andhika Herayono[4]
[1,3]Department of Teknik Elektronika, Universitas Negeri Padang, Padang, Indonesia
[2]Department of Teknik Mesin, Universitas Negeri Padang, Padang, Indonesia
[4]Doctoral Program, Universitas Negeri Padang, Padang, Indonesia

Received: : March 9, 2025
Revised: : April 7, 2025
Accepted: : April 28, 2025

Abstract
Digital Education and Entrepreneurial (DE) represents a new paradigm in Education and Entrepreneurial that leverages digital technology to create, manage, and expand businesses. By integrating advanced digital tools and platforms, DE plays a crucial role in reshaping traditional business models, driving innovation faster, and enabling enterprises to reach broader markets. This transformative approach benefits individual entrepreneurs and contributes to broader economic development. One of DE's most significant impacts is its ability to foster economic growth. By embracing digital Education and Entrepreneurial, businesses can create new jobs, increase competitiveness, and adapt more effectively to the demands of the digital age. These factors collectively ensure that economies are better positioned to thrive in a technology-driven world. A recent study developed a predictive model using the CatBoost algorithm to understand better the factors influencing digital Education and Entrepreneurial. This advanced machine learning method was applied to data collected from thousands of college students, encompassing various demographic, psychological, and business-related variables. The results demonstrated the model's high accuracy in predicting intentions toward digital Education and Entrepreneurial, offering a reliable framework for analysis and application. The study identified three key factors influencing students' intentions to pursue digital Education and Entrepreneurial. These are digital skills, which reflect their ability to navigate and utilize digital tools effectively; self-efficacy, their confidence in their entrepreneurial capabilities; and Education and Entrepreneurial education, which equips them with the knowledge and skills needed to innovate and create businesses. These findings provide valuable insights for educational institutions and policymakers. By emphasizing digital skills training, fostering self-efficacy, and enhancing Education and Entrepreneurial education programs, they can better prepare students to succeed in the digital economy. Such targeted initiatives empower individuals and contribute to the sustainable growth of digital Education and Entrepreneurial, reinforcing its role as a driver of innovation and economic progress.

Keywords: Digital Education and Entrepreneurial; CatBoost Algoritihm; Entrepreneurial Intention; Machine Learning; Education and Entrepreneurial Education

(*) Corresponding Author: zulwisli@ft.unp.ac.id

How to Cite: Zulwisli, Ambiyar, Muhammad Anwar, & Andhika Herayono. (2025). Student's Digital Intentions Prediction Using CatBoost. *JTP - Jurnal Teknologi Pendidikan*, *27*(1), 277–290. https://doi.org/10.21009/jtp.v27i1.54035

INTRODUCTION

Digital Education and Entrepreneurial (DE), influenced by Industrial Revolution 4.0, is a concept of Education and Entrepreneurial that uses digitalisation to disrupt traditional business models and practices. This gives business actors new opportunities to innovate, provide services, and interact with customers in a rapidly changing market (Baig et al, 2022). By digitally altering ecosystems and business models, DE also plays a significant role in promoting innovation and economic progress. Even while DE is becoming increasingly significant, there are still not enough college graduates turning into digital entrepreneurs, according to recent research. According to research, only 5% of graduates in Malaysia selected DE as their professional path. (A.R Mohd Nor et al, 2023). According to Italian research, 4.5% of graduates became DEs, and just 1.3% of students launched a firm after graduation Ferrante et al, 2023). Despite the growing significance of digital Education and Entrepreneurial, a different survey revealed that very few college graduates are still pursuing careers as digital entrepreneurs (Nguyen et al, 2024).

There are several reasons why there aren't more college graduates worldwide becoming digital entrepreneurs. According to a survey, two major challenges are a lack of long-term finance and capital limits (Handayati et al, 2021). According to other research, one more thing impeding entrepreneurial endeavours is the inability to obtain ED counsel (Rashid, 2019). Another obstacle preventing graduates from becoming digital entrepreneurs is their lack of the specialised abilities needed by ED, such as inventive cognition and digital competence (Siddoo et al, 2019). Graduates who want to become digital entrepreneurs face additional obstacles, such as the high risk of founding a startup and the potential for failure due to poor business decisions or a lack of market demand (Darmanto et al, 2022). According to additional studies, one major barrier is college graduates' poor intention to become digital entrepreneurs (Biclesanu et al, 2023). Numerous studies have demonstrated that a major barrier preventing many college graduates from entering the field of education design is their poor ambition to become digital entrepreneurs. (Alzougool et al, 2024), (Muafi et al, 2021).

Because of the many obstacles and variables influencing the growth of digital entrepreneurial intention (DEI), it is necessary to predict DEI. Numerous research works have found several indicators of DEI, such as creativity, self-efficacy in Education and Entrepreneurial, education and knowledge in digital Education and Entrepreneurial, and awareness of entrepreneurial potential. (Alzougool et al, 2024), (Laguía et al, 2019). Subjective norms, psychological resilience variables, and the calibre of higher education services have all been linked to good effects on DEI, according to other research (Song et al, 2024). Meanwhile, other research shows that digital entrepreneurial competency elements, innovative cognition, social media literacy, and role models contribute greatly to DEI (Miniesy et al, 2022),(Mbhele and Beharry-Ramrai, 2024).

Prediction of DEI is very important. The prediction results can be used to design effective Education and Entrepreneurial education programs that can improve information technology knowledge for students and can be used to provide direct training to become digital entrepreneurs. In addition, DEI prediction helps bridge the gap between entrepreneurial intentions and behaviour in the real world, as suggested by the theory of planned behaviour (Zulwisli et al, 2024). Understanding DEI will guide educators and policymakers in building digital Education and Entrepreneurial. DEI is very important to predict because it indicates the possibility of individuals pursuing digital businesses (P.N.D. Nguyen et al, 2024). This indication can be used to treat students who have high intentions. Psychological factors such as subjective norms, resilience, and quality of higher education services also positively affect DEI among students (Wardana et al, 2023). The factors influencing DEI are important knowledge that needs to be strengthened so that DEI becomes strong, too.

The impact of personality factors and entrepreneurial mentality on entrepreneurial intentions was investigated using SEM and route analysis (Al-Ghazali et al, 2022). The association between education, knowledge, awareness, and intentions in digital Education and Entrepreneurial has been examined using partial least squares SEM (PLS-SEM) (Wibowo et al, 2023). Numerous studies utilising various techniques have been carried out to predict digital entrepreneurial inclinations. Artificial intelligence (AI) is one of them; it examines ChatGPT adoption via technostress and entrepreneurial efficacy (Bui & Duong et al, 2024). DEI prediction has also been done using machine learning models that employ Kernel Extreme Learning Machine (M. Zekić-Sušac and A. Has, 2015). The impact of knowledge and inspiration gained from digital Education and Entrepreneurial education on entrepreneurial intents is examined using a quantitative method utilising structural equation modelling (SEM) (Wibowo et al, 2023). Students' intentions are predicted using artificial intelligence techniques like Random Forest and SVM, optimised with chaotic search methods (J. Tu et al, 2019). Meanwhile, integrating PCA and artificial neural networks improves predictions based on demographics, social norms, and self-efficacy (M. Zekić-Sušac and A. Has, 2015).

However, traditional statistical methods used to predict digital entrepreneurial intentions have weaknesses when dealing with non-linear, multivariate, and complex data. So, it is less accurate in predicting the dynamic digital business context. This method also tends to be static and inflexible in dealing with rapid changes in predictor variables (Bakator et al, 2023). Certain traits are present in the machine learning model that has been employed to forecast the intentions of digital enterprises. Random Forest tends to overfit when working with data that is hard to interpret and has many irrelevant features (Zhao et al, 2023). Support Vector Machine (SVM) is also utilised for DEI prediction. However, it performs poorly against noise and missing data and is less successful with large datasets (J. Tu et al, 2019). Neural networks require training large and high-quality datasets because overfitting is a concern (Gonzalez-Diaz et al, 2022) (Xie et al, 2022). Extreme Learning Machine's (ELM) susceptibility to outliers and capacity to regulate generalisation (Songet al, 2020)(Y. Park and J. C. Ho, 2021).

By integrating many models to provide more robust and accurate predictions, some Machine Learning Ensemble models, like XGBoost and AdaBoost, can be utilised to solve the overfitting issue in Random Forest (I. Salehin and D. K. Kang, 2023). According to additional research, Deep Neural Networks (DNN) with Dropout techniques can reduce overfitting and increase training efficiency, thereby addressing the drawbacks of conventional neural networks (I. Salehin and D. K. Kang, 2023). In the meantime, another study shows that SVM is more effective at handling big and complicated datasets when Support Vector Regression (SVR) with an optimised kernel is used. (M. Xie, L. Xie, and P. Zhu, 2021). However, CatBoost, a tree-based boosting algorithm, has an advantage over other models because it can handle categorical data with little preprocessing and because it makes use of the Ordered Boosting technique, which lessens overfitting and predictive bias and increases accuracy and efficiency, particularly when handling complex and imbalanced data (Prokhorenkova et al, 2021), (J. T. Hancock and T. M. Khoshgoftaar, 2020),(Ibrahim et al, 2020).

Students' intentions to engage in digital Education and Entrepreneurial are predicted by the dataset utilised in this study. Age, gender, education in Education and Entrepreneurial, digital skills, self-efficacy, social influence, risk tolerance, exposure to business, and access to capital are some of its key characteristics. Particularly the intention to engage in digital Education and Entrepreneurial, some of these variables—particularly those that are categorical—may not be distributed evenly. According to the explanation of CatBoost's prior capabilities, Catboost is highly appropriate for predicting DEI, given the type of data and the possibility of imbalance in the dataset utilised in this study. Therefore, this study proposes a prediction of DEI using CatBoost. Thus, this research aims to develop an accurate prediction model to predict students' Digital Education and Entrepreneurial Intention (DEI) using CatBoost.

METHODS

The target population for this study was roughly 52,200 students, and the data were gathered from students enrolled at Universitas Negeri Padang. Using a stratified random sampling technique, 4800 respondents made up the final sample size, ensuring a representative sample. Important demographic, psychological, and business-related characteristics are included in the gathered dataset. These target variables include age, gender, business exposure, Education and Entrepreneurial education, digital skills, self-efficacy, social influence, risk tolerance, and intentions for digital Education and Entrepreneurial. New qualities have also been included, including the following: Year of Enrollment, Program of Study, Parent's Occupation, Parent's Education, Digital Literacy, Creativity, Business Attitude, Subjective Norm, and Perceived Behavioral Control. The Region of Origin is Kabupaten. Combined, these attributes offer extensive information for examining elements impacting digital Education and Entrepreneurial.

A questionnaire was created to gather information on digital entrepreneurial goals' business, psychological, and demographic elements to measure the factors described in this study. This tool measures digital abilities, self-efficacy, social influence, risk tolerance, and business attitudes using a 5-point Likert scale. The questionnaire uses a multiple-choice style for demographic information like age, gender, and place of origin. Furthermore, a scale gauges respondents' degree of engagement or mastery with new factors like digital literacy, creativity, and perceived behavioural control, which is used to evaluate exposure to business and Education and Entrepreneurial courses and new variables like these. This tool thoroughly assesses the factors influencing Universitas Negeri Padang students' desire to pursue Education and Entrepreneurial.

CatBoost is a machine learning technique that handles categorical data using gradient boosting instead of one-hot encoding or additional preparation. This method reduces the possibility of overfitting by using an ordered boosting strategy on categorical data sequentially. CatBoost is more efficient than other boosting algorithms like XGBoost and AdaBoost because it can handle categorical characteristics directly and decrease computational complexity. CatBoost is a powerful technique in modern data analysis since the ordered target statistics formula yields more accurate predictions on complex and unbalanced datasets.

Ordered target statistics is the primary formula in CatBoost and is utilised to handle categorical data. The average target statistics are updated to prevent data leaking at each iteration to calculate this formula. This equation can be understood as:

$$\frac{\sum_{j=1}^{p-1} \left[x_{\sigma_{j,k}} = x_{\sigma_{p,k}}\right] Y_{\sigma_j} + a.P}{\sum_{j=1}^{p-1} \left[x_{\sigma_{j,k}} = x_{\sigma_{p,k}}\right] + a} \quad (1)$$

The data order index is represented by pIn this formula, the target label is represented by Y, the categorical feature value by x, A smoothing parameter that helps avoid overfitting is represented by a · P. With CatBoost, ordered boosting enables the model to reduce prediction bias by updating the target statistics in each iteration based on prior data. This method allows CatBoost to handle multi-category data with greater stability and produce more accurate predictions on unbalanced datasets. In machine learning, several performance metrics are generally used to evaluate models. Here are some of the most commonly used metrics and an explanation of each metric:

a. Accuracy

Accuracy is the ratio of the number of correct predictions to the total predictions made by the model. The formula is shown in Equation 2.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

Where: - TP = True Positive (true positive prediction)
- TN = True Negative (true negative prediction)
- FP = False Positive (false positive prediction)
- FN = False Negative (false negative prediction)

b. Precision

Precision measures how many positive predictions are positive. The formula is shown in Equation 3.

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

c. Recall (Sensitivity or True Positive Rate)

The number of positive cases the model successfully locates is measured by recall. Equation 4 displays the formula.

$$Recall = \frac{TP}{TP + FN} \qquad (4)$$

d. F1-Score

The harmonic mean of precision and recall, or F1-Score, strikes a balance between the two, particularly in cases where the data is unbalanced. Eq. 5 displays the formula.

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \qquad (5)$$

e. Area Below the ROC Curve (AUC-ROC)

AUC-ROC measures the model's ability to discriminate between positive and negative classes. Plotting a True Positive Rate (Recall) against a False Positive Rate (FPR) is known as ROC. Eq. 6 displays the formula.

$$FPR = \frac{FP}{FP + TN} \qquad (6)$$

RESULTS & DISCUSSION

Results

1. Overview of the Dataset

The dataset used in this study, which includes several significant characteristics that affect Digital Education and Entrepreneurial Intention (DEI), is broadly described in the following table. These characteristics include self-efficacy, digital skills, entrepreneurial education, and other business and psychological components. The categories in this table indicate the percentage of respondents who fit each characteristic value. This data offers a solid basis for future research on the forecast and evolution of DEI since it captures various characteristics that influence students' intentions to pursue careers in digital Education and Entrepreneurial.

Table 1. General Description of Datasets

Feature/Target	Category 1 (% value)	Category 2 (% value)	Category 3 (% value)
Education and Entrepreneurial Education	Yes (70%)	No (30%)	
Digital Skills	High (40%)	Medium (40%)	Low (20%)
Self-efficacy	High (50%)	Medium (30%)	Low (20%)
Social Influence	High (40%)	Medium (30%)	Low (30%)
Risk Tolerance	High (35%)	Medium (40%)	Low (25%)
Business Exposure	Yes (60%)	No (40%)	
Access to Capital	Yes (55%)	No (45%)	
Digital Literacy	High (45%)	Medium (35%)	Low (20%)
Creativity	High (50%)	Medium (30%)	Low (20%)
Business Attitude	Positive (60%)	Neutral (30%)	Negative (10%)

Subjective Norm	Strong (45%)	Moderate (35%)	Weak (20%)
Perceived Behavioral Control	High (50%)	Medium (30%)	Low (20%)
Digital Education and Entrepreneurial Intentions	Yes (60%)	No (40%)	

The dataset utilised in this study includes a wide range of qualitative and numerical attributes linked to Digital Education and Entrepreneurial Intention (DEI). Examples of categorical variables that, when processed using CatBoost, do not require additional preprocessing, such as one-hot encoding, are features like Digital Skills (with High, Medium, and Low levels), Education and Entrepreneurial Education (with Yes and No categories), Self-efficacy, Social Influence, and Business Attitude. In addition, there is an imbalance in the number of categories and a diverse distribution of other attributes among the respondents, such as inventiveness and financial availability. The imbalance in these variables makes the dataset ideal for processing because CatBoost is a method developed to handle category data directly and alleviate the overfitting issue commonly occurring in regular boosting models.

Using an ordered boosting strategy, catBoost processes category data sequentially to prevent target leakage. This means that to maintain accuracy in predictions, the algorithm learns from past examples without consulting future data in each iteration. This method works particularly well with unbalanced datasets, such as the one containing the Digital Education and Entrepreneurial Intentions variable, where notable variations exist between the Yes (60%) and No (40%) categories. Another benefit of CatBoost is that it can handle categorical characteristics without converting them to numeric form, which expedites training and enhances the model's ability to recognise patterns in the data.

2. Performance of Prediction Models

The following confusion matrix displays the model's performance in predicting positive and negative classes over the whole dataset. Based on the model's actual and anticipated values, this matrix displays the number of positive, negative, accurate, and wrong forecasts for each class.

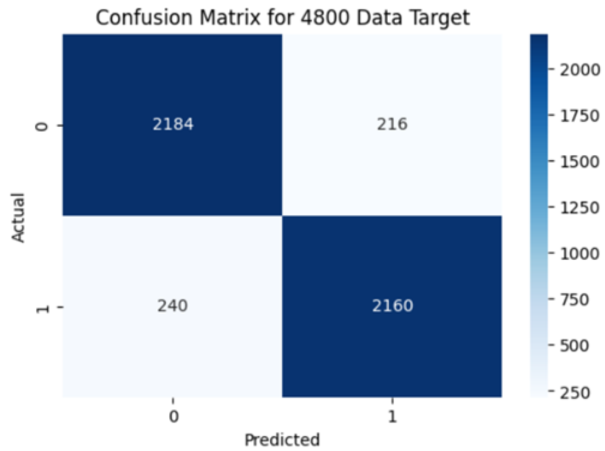

Figure 1. Confusion Matrix

The confusion matrix results show that the digital Education and Entrepreneurial intention (DEI) prediction model distinguishes between samples with intention and those without. This model shows high accuracy, with 2184 correct predictions for the negative class (true negatives) and 2160 correct predictions for the positive class (true positives). However, there are still some prediction errors. Namely, 216 samples were incorrectly classified as having intention (false positives), and 240 samples that had intention were classified as not (false negatives). False negative errors suggest that there are people with entrepreneurial potential who the model does not identify, even though the model is generally accurate. This can impede the delivery of essential support or intervention.

These outcomes align with earlier studies by (Prokhorenkova et al, 2021), who discovered that while CatBoost considerably lessens overfitting, minimising false negatives in imbalanced datasets is still challenging. This is also supported by (J. T. Hancock and T. M. Khoshgoftaar, 2020), who noted that increasing accuracy in classification often has to be balanced with decreasing error rates in minority classes. According to [36], additional methods like threshold optimisation or ensemble models might be required to lower the possibility of false negatives in large and complicated datasets with high overall accuracy.

The model's performance is summarised here using several widely used evaluation indicators. The table below displays metrics that indicate how well the model predicts the tested dataset. These metrics include accuracy, precision, recall, F1-score, and ROC-AUC.

Table 2. Model Performance Metrics

Metric	**Value**
Accuracy	0.92
Precision	0.91
Recall	0.90
F1-Score	0.91
ROC-AUC	0.93

Every indication points to a well-performing model within the predicted theoretical bounds. Theoretically, a model is considered good if its accuracy is 0.80 or higher; therefore, a high accuracy indicates that it can produce general predictions. A high precision suggests that the model rarely makes an error in predicting the positive class, according to the theory emphasizing precision's importance in reducing false positives. In important applications, a high recall value indicates that the model meets the recall criteria by being able to capture the most positive samples. In theory, the model's effectiveness in accurately detecting positives is reflected in the balanced F1 score that balances precision and recall. Furthermore, a high ROC-AUC indicates the model's exceptional ability to distinguish between positive and negative classifications. The following graph presents the ROC curve, which illustrates the model's ability to separate positive and negative classes based on various thresholds. The area under the curve (AUC) is shaded to indicate how well the model can differentiate between the two classes, with an AUC value of 0.93 indicating excellent model performance.

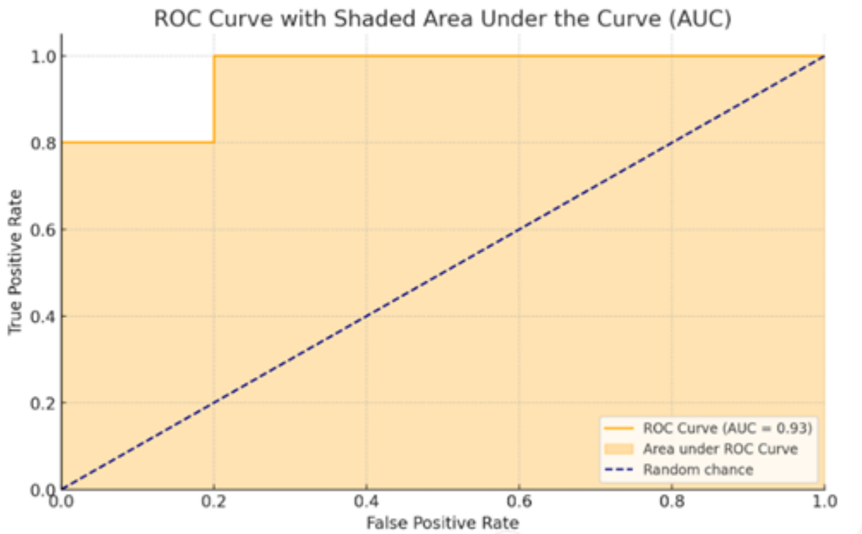

Figure 2. ROC-AUC curve

The ROC curve graph above shows the model's ability to separate positive and negative classes based on various thresholds. The ROC line shows the relationship between the True Positive Rate (TPR) and False Positive Rate (FPR), with the area under the curve (AUC) reaching 0.93. This AUC value indicates that the model has a very good ability to distinguish positive and negative classes. The diagonal line on the graph represents random predictions, and the ROC curve model above this line indicates better performance than random predictions.

The bar chart below displays the top ten CatBoost model features with the most influence. These traits are arranged according to feature relevance, which sheds light on how each feature affects the likelihood that a person will pursue digital Education and Entrepreneurial.

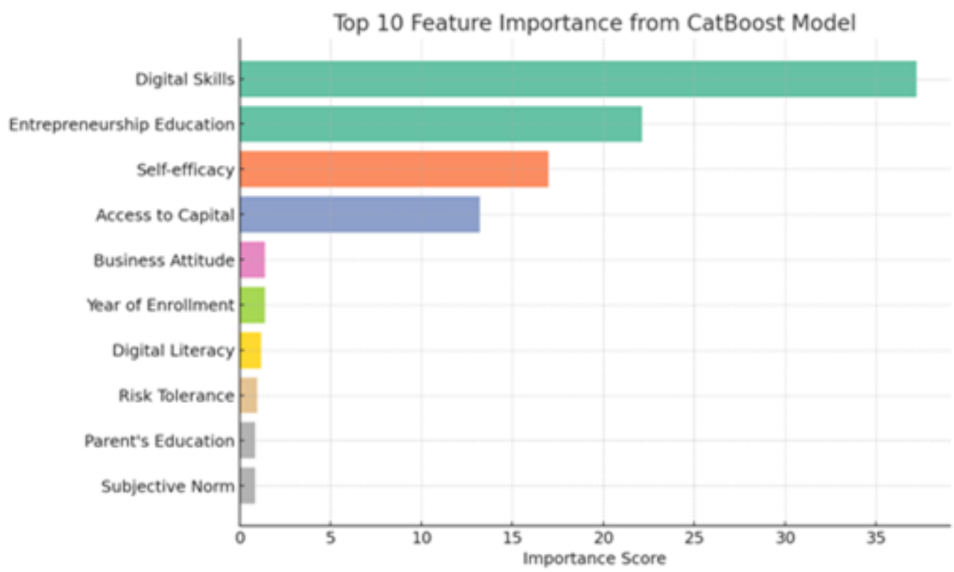

Figure 3. Importance Feature

According to the graph above, digital skills have the biggest impact on predicting the intention to engage in digital Education and Entrepreneurial. This is followed by self-efficacy and Education and Entrepreneurial education. Even if they have smaller effects, characteristics like business attitude and access to capital also make substantial contributions. Even if they have less of an impact on the model, other factors, including risk tolerance, digital literacy, and year of enrollment, are nevertheless important in predicting the outcomes. This demonstrates that digital skills and Education and Entrepreneurial education are the primary determinants of students' intention to engage in digital Education and Entrepreneurial.

Discussion

The outcomes demonstrate that, without extra pre-processing like one-hot encoding, the CatBoost model can handle datasets containing qualitative and numerical properties, particularly categorical data. This model effectively processes digital skills, entrepreneurial education, self-efficacy, social impact, and business-related attitudes. Furthermore, the model may correct imbalances in the variables' distribution, including financial availability and creativity, which are frequently problems for other prediction models. It has been demonstrated that CatBoost lowers the chance of overfitting, resulting in more precise and reliable predictions on complicated and unbalanced data.

This result is consistent with several earlier research projects that have demonstrated the efficiency of boosting models in managing unbalanced and categorical data. According to (Prokhorenkova et al, 2021), using the ordered boosting technique, CatBoost greatly reduces overfitting by processing categorical data. CatBoost is also very good at handling complicated datasets without requiring further pre-processing, which improves prediction accuracy, as (J. T. Hancock and T. M. Khoshgoftaar, 2020) also confirmed. According to [36], CatBoost is perfect for various categorisation applications since it can manage unbalanced data, particularly in business data and digital Education and Entrepreneurial.

According to the evaluation results, the prediction model performs exceptionally well within the anticipated theoretical bounds. According to theory, a model is deemed good if its accuracy is 0.80 or higher. The model's accuracy in this study was high, indicating its capacity to generate generally accurate predictions. According to the theory highlighting the significance of precision in lowering false positives, high precision means that the model rarely makes mistakes in predicting the positive class. The model can theoretically satisfy the recall requirements if it has a high recall in significant applications, indicating that it can account for most positive samples. The balanced F1 score shows how well the model correctly identifies the positive class by balancing recall and precision. Furthermore, this model's exceptional ability to distinguish between positive and negative classes is confirmed by its high ROC-AUC (0.93). This demonstrates how the model can produce reliable forecasts even when utilising different thresholds. This finding is corroborated by the (Prokhorenkova et al, 2021) study, which emphasised that a high ROC-AUC indicates the model's reliability in classification tasks. An AUC above 0.90 indicates excellent model performance, especially in complex and imbalanced classification tasks, as noted by (J. T. Hancock and T. M. Khoshgoftaar, 2020). Table II summarises the performance metrics of the CatBoost model in predicting Digital Education and Entrepreneurial Intentions (DEI), showing excellent results in various evaluation aspects. With an accuracy of 0.92, the model successfully made correct predictions on 92% of the total data, reflecting CatBoost's ability to handle complex and heterogeneous datasets with various types of features, both numeric and categorical. This high accuracy value suggests that CatBoost can minimise prediction performance and prevent overfitting using its Ordered Boosting technique (Prokhorenkova et al, 2021).

Based on the CatBoost model research, the Top 10 Features in Figure 3 offer significant insights into the critical aspects driving Digital Entrepreneurial Intention (DEI). The most important factor, digital skills, suggests that students' technical aptitude plays a crucial role in determining whether or not they are prepared to enter the field of digital Education and Entrepreneurial. This aligns with research highlighting the importance of digital skills in today's workforce (Dorogushet al, 2020). Furthermore, according to Education and Entrepreneurial education, which comes in second, formal Education and Entrepreneurial training significantly influences students' willingness to launch a digital firm (J. T. Hancock and T. M. Khoshgoftaar, 2020).

Additional variables like capital availability and business attitudes imply that financial assistance and favourable business attitudes are significant determinants of students' entrepreneurial decision-making. This is consistent with studies that emphasise the role of personal and economic factors in Education and Entrepreneurial success (A. Garcez, M. Franco, and R. Silva, 2023). The application of CatBoost is highly appropriate in this situation since it allows for a more thorough investigation of the relationships between features and can handle categorical data and non-linear interactions without the need for laborious preprocessing. The effectiveness of ordered boosting in CatBoost has been proven to reduce the risk of overfitting, especially when working with complex and diverse datasets (Prokhorenkova et al, 2021). Consequently, this model provides precise forecasts and a more profound comprehension of the factors influencing students' aspirations to participate in digital Education and Entrepreneurial (Peng et al, 2024).

CONCLUSION

This study's prediction model, which uses the CatBoost algorithm, has demonstrated high performance in predicting DEI. Additionally, this study discovered several significant factors that affect DEI, including business attitude, self-efficacy, digital skills, Education and Entrepreneurial education, and access to capital. These characteristics are the primary predictors in the CatBoost model, which was developed to influence students' intentions to participate in digital Education and Entrepreneurial.

Future studies should investigate additional aspects like innovation and technological adaptability. Accuracy and efficiency could also be compared using predictive methods like ensemble learning or neural networks. Furthermore, demographic and cultural variables are analysed to comprehend the variations in digital entrepreneurial aspirations among environments. Longitudinal studies and entrepreneurial interventions are also crucial to show how these aspirations change over time and how successful the support is.

This study has some limitations, including not examining demographic and cultural aspects, not comparing with other models, and not explaining non-dominant traits. Furthermore, CatBoost's efficacy and efficiency in predicting Digital Entrepreneurial Intention (DEI) cannot be determined by comparing it with other models like XGBoost or AdaBoost.

With CatBoost's ability to handle unbalanced and categorical data without further preprocessing, this study offers a clearer understanding of the factors impacting students' Digital Education and Entrepreneurial Intention (DEI). Its benefit is that ordered boosting lowers the chance of overfitting, which increases accuracy and efficiency. The findings of this research can be used to create more successful digital Education and Entrepreneurial education initiatives, assisting organisations and decision-makers in identifying and assisting students with a high degree of entrepreneurial potential.

The findings of this study suggest that educational establishments create courses on digital Education and Entrepreneurial with an emphasis on self-efficacy, digital skills, and Education and Entrepreneurial education. To get more thorough findings, it is also advised to apply additional prediction models, including XGBoost and AdaBoost. More research is required to understand better how cultural and demographic factors affect ambitions for digital Education and Entrepreneurial.

REFERENCES

Al-Ghazali, B. M., Shah, S. H. A., & Sohail, M. S. (2022). The role of five big personality traits and entrepreneurial mindset on entrepreneurial intentions among university students in Saudi Arabia. *Frontiers in Psychology, 13*, Article 964875. https://doi.org/10.3389/fpsyg.2022.964875

Alzougool, B. (2024). Predictors of digital entrepreneurial intention in Kuwait. *Information Sciences, 27*. https://doi.org/10.28945/5339

Baig, U., Hussain, B. M., Meidute-Kavaliauskiene, I., & Davidavicius, S. (2022). Digital education and entrepreneurial: Future research directions and opportunities for new business model. *Sustainability*, *14*(9), 5004. https://doi.org/10.3390/su14095004

Bakator, M., Cockalo, D., Kavalić, M., Terek Stojanović, E., & Gluvakov, V. (2023). An application of statistical methods in data mining techniques to predict ICT implementation of enterprises. *Applied Sciences (Switzerland)*, *13*(6), 4055. https://doi.org/10.3390/app13064055

Biclesanu, I., Savastano, M., Chinie, C., & Anagnoste, S. (2023). The role of business students' entrepreneurial intention and technology preparedness in the digital age. *Administrative Sciences, 13*(8), 177. https://doi.org/10.3390/admsci13080177

Bui, H. N., & Duong, C. D. (2024). ChatGPT adoption in education and entrepreneurial and digital entrepreneurial intention: A moderated mediation model of technostress and digital entrepreneurial self-efficacy. *Equilibrium. Quarterly Journal of Economics and Economic Policy, 19*(2), 391–428. https://doi.org/10.24136/eq.3074

Darmanto, S., Darmawan, D., Ekopriyono, A., & Dhani, A. U. (2022). Development of digital entrepreneurial intention model in uncertain era. *Uncertain Supply Chain Management, 10*(3), 1091–1102. https://doi.org/10.5267/j.uscm.2022.7.050

Dorogush, A. V., Ershov, V., & Gulin, A. G. (n.d.). CatBoost: Gradient boosting with categorical features support. [Online]. Available: https://github.com/Microsoft/LightGBM

Ferrante, F., Federici, D., & Parisi, V. (2019). The entrepreneurial engagement of Italian university students: Some insights from a population-based survey. *Studies in Higher Education, 44*(11), 1813–1836. https://doi.org/10.1080/03075079.2018.1458223

Garcez, A., Franco, M., & Silva, R. (2023). The influence of the pillars of digital academic education and entrepreneurial on university students' entrepreneurial intention. *European Journal of Innovation Management, ahead-of-print*(ahead-of-print). https://doi.org/10.1108/EJIM-01-2023-0051

Gonzalez-Diaz, R., Gutiérrez-Naranjo, M. A., & Paluzo-Hidalgo, E. (2022). Topology-based representative datasets to reduce neural network training resources. *Neural Computing and Applications, 34*(17), 14397–14413. https://doi.org/10.1007/s00521-022-07252-y

Handayati, P., Wibowo, A., Narmaditya, B. S., Kusumojanto, D. D., Setiawan, A. B., & Tung, D. T. (2021). The university students enterprises development: Lesson from Indonesia. *Cogent Education, 8*(1), 1973286. https://doi.org/10.1080/2331186X.2021.1973286

Hancock, J. T., & Khoshgoftaar, T. M. (2020). CatBoost for big data: An interdisciplinary review. *Journal of Big Data, 7*(1), 94. https://doi.org/10.1186/s40537-020-00369-8

Hu, M., Peng, H., Zhang, X., Wang, L., & Ren, J. (2024). Building gender-specific sexually transmitted infection risk prediction models using CatBoost algorithm and NHANES data. *BMC Medical Informatics and Decision Making, 24*(1), 24. https://doi.org/10.1186/s12911-024-02426-1

Ibrahim, A. A., Ridwan, R. L., Muhammed, M. M., Abdulaziz, R. O., & Saheed, G. A. (2020). Comparison of the CatBoost classifier with other machine learning methods. [Online]. Available: www.ijacsa.thesai.org

Laguía, A., Moriano, J. A., & Gorgievski, M. J. (2019). A psychosocial study of self-perceived creativity and entrepreneurial intentions in a sample of university students. *Thinking Skills and Creativity, 31*, 44–57. https://doi.org/10.1016/j.tsc.2018.11.004

Mbhele, N., & Beharry-Ramraj, A. (2024). Digital education and entrepreneurial in the fourth industrial revolution and beyond. https://doi.org/10.5772/intechopen.1005449

Miniesy, R., Shahin, M., & Fakhreldin, H. (2022). The determinants of digital education and entrepreneurial by informal micro and small enterprises (MSEs) in Egypt. *World Journal of Education and Entrepreneurial, Management and Sustainable Development, 18*(4), 425–445. https://doi.org/10.47556/J.WJEMSD.18.4.2022.1

Mohd Nor, A. R., Anual, N., Ahmad Alemi, N. H., & Ab Jalil, S. A. (2023). The effect of compromising factors on career choice tendency: A case of Malaysia community college graduates. *International Journal of Academic Research in Business and Social Sciences, 13*(9). https://doi.org/10.6007/ijarbss/v13-i9/12889

Muafi, M., Syafri, W., Prabowo, H., & Nur, S. A. (2021). Digital education and entrepreneurial in Indonesia: A human capital perspective. *Journal of Asian Finance, Economics and Business, 8*(3), 351–359. https://doi.org/10.13106/jafeb.2021.vol8.no3.0351

Nguyen, P. N. D., & Nguyen, H. H. (2024). Unveiling the link between digital education and entrepreneurial education and intention among university students in an emerging economy. *Technological Forecasting and Social Change, 203*, 123330. https://doi.org/10.1016/j.techfore.2024.123330

Park, Y., & Ho, J. C. (2021). Tackling overfitting in boosting for noisy healthcare data. *IEEE Transactions on Knowledge and Data Engineering, 33*(7), 2995–3006. https://doi.org/10.1109/TKDE.2019.2959988

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (n.d.). CatBoost: Unbiased boosting with categorical features. [Online]. Available: https://github.com/catboost/catboost

Rashid, L. (2019). Education and entrepreneurial education and sustainable development goals: A literature review and a closer look at fragile states and technology-enabled approaches. *Sustainability, 11*(19), 5343. https://doi.org/10.3390/su11195343

Salehin, I., & Kang, D. K. (2023). A review on dropout regularization approaches for deep neural networks within the scholarly domain. *Electronics, 12*(14), 3106. https://doi.org/10.3390/electronics12143106

Siddoo, V., Sawattawee, J., Janchai, W., & Thinnukool, O. (2019). An exploratory study of digital workforce competency in Thailand. *Heliyon, 5*(5), e01723. https://doi.org/10.1016/j.heliyon.2019.e01723

Song, C., Pan, L., Liu, Q., Jiang, Z., & Jia, J. (2020). Extreme learning machine under minimum information divergence criterion. *IEEE Access, 8*, 122026–122035. https://doi.org/10.1109/ACCESS.2020.3007522

Song, P., et al. (2024). Analyzing psychological resilience in college students: A decision tree model. *Heliyon, 10*(11), e32583. https://doi.org/10.1016/j.heliyon.2024.e32583

Tu, J., Lin, A., Chen, H., Li, Y., & Li, C. (2019). Predict the entrepreneurial intention of fresh graduate students based on an adaptive support vector machine framework. *Mathematical Problems in Engineering, 2019*, 2039872. https://doi.org/10.1155/2019/2039872

Wardana, L. W., et al. (2023). Do digital literacy and business sustainability matter for creative economy? The role of entrepreneurial attitude. *Heliyon, 9*(1), e12763. https://doi.org/10.1016/j.heliyon.2022.e12763

Wibowo, A., Narmaditya, B. S., Saptono, A., Effendi, M. S., Mukhtar, S., & Mohd Shafiai, M. H. (2023). Does digital education and entrepreneurial education matter for students' digital entrepreneurial intentions? The mediating role of entrepreneurial alertness. *Cogent Education, 10*(1), 2221164. https://doi.org/10.1080/2331186X.2023.2221164

Xie, M., Xie, L., & Zhu, P. (2021). An efficient feature weighting method for support vector regression. *Mathematical Problems in Engineering, 2021*, 6675218. https://doi.org/10.1155/2021/6675218

Xie, X., Xie, M., Moshayedi, A. J., & Noori Skandari, M. H. (2022). A hybrid improved neural networks algorithm based on L2 and dropout regularization. *Mathematical Problems in Engineering, 2022*, 8220453. https://doi.org/10.1155/2022/8220453

Zhao, X., Wu, Y., Lee, L., & Cui, W. (n.d.). iForest: Interpreting random forests via visual analytics.

Zulwisli, Ambiyar, Anwar, M., & Herayono, A. (2024). The effect of theory of planned behavior (TPB) and creativity-based industry perception on digital education and entrepreneurial: An innovativeness as mediator. *Paper Asia, 40*(3), 96–105. https://doi.org/10.59953/paperasia.v40i3b.100