REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 24, No. 4 (2023)
http://www.veterinaria.org
Article Received: 09/11/2023 Revised: 21/11/2023 Accepted: 01/12/2023

AI based Admission prediction model: Testing & Sustainability

Vineet Kumar[1], Arun Solanki[1], Vivek Singh Malik[2], Pradeep Tomar[1]

[1]Department of Computer Science & Engineering, USICT, Gautam Buddha University, Greater Noida, India
[2]Centre of Excellence in AI, Department of Telemedicine, AIIMS, Rishikesh, India

Abstract
This study aims to evaluate and compare the performance of various AI modelsRNN (Recurrent Neural Network), DBN (Deep Belief Network), DBM (Deep Boltzmann Machine), RBM (Restricted Boltzmann Machine), ESDRBM (Extended Stacked Denoising Restricted Boltzmann Machine)—using original and synthetic datasets to determine the viability of synthetic data for maintaining model performance. The objective is to assess whether synthetic data can serve as a reliable substitute for original data in training AI models, by comparing key performance metrics including precision, recall, FMeasure, accuracy, sensitivity, and specificity. The performance of each model was measured across the specified metrics using both original and synthetic datasets. A paired sample t-test was employed to statistically analyze the differences between the metrics obtained from original and synthetic data, assessing the significance of any observed differences. The results indicate minimal differences in performance metrics between original and synthetic data for all models. ESDRBM, RBM, and DBM consistently showed slightly higher precision, recall, and FMeasure values. DBM achieved the highest accuracy, while sensitivity and specificity metrics remained nearly identical across both data types. Paired sample t-tests confirmed that the differences between the original and synthetic data were not statistically significant, with high p-values indicating random variation as the likely cause of any observed differences. The findings suggest that synthetic data can effectively maintain the performance of AI models across various metrics. ESDRBM, RBM, and DBM models particularly exhibit robust performance with both data types. This underscores the potential of synthetic data as a viable alternative to original data in AI model training, providing flexibility and scalability in data generation without compromising model effectiveness.

Keywords: AI, Admission Prediction, Sustainability

Correspondence:
Dr. Arun Solanki, Assistant Professor, University School of Information and Communication Technology
Department of Computer Science & Engineering, Gautam Buddha University, Greater Noida, India,
Email: arunk@gbu.ac.in

1. Introduction:

The economy, society, and environment continue to provide multiple challenges to the modern world. Sustainably developing the educational system and ensuring that everyone has access to opportunities for education are essential for effectively solving these issues through sustainable development in accordance with sustainability principles. [1,2,3]. As part of the United Nations' "the 17 Goals" of Sustainable Development, "Quality Education" asks for ensuring inclusive and equitable quality education and encouraging opportunities for lifelong learning for everyone. [4]. In order to accomplish the aim of sustainable development, we must focus on higher education in addition to basic and secondary education [5, 6]. Studies already conducted have demonstrated that higher education has a larger return on investment than either vocational training or a high school education [7], and its importance in societal development is evident.
Furthermore, in order to effectively advance social progress, the concerns surrounding sustainable development and higher education are now critical needs rather than options [8]. Improvements in education and opportunities for lifelong learning can help citizens reach a certain level in terms of knowledge sharing and discovery, business collaboration, science, technology, and creativity, personal and social moral systems, and inheriting fine cultural heritage, as long as they keep their attention on how to work together to create a more sustainable society [9,10,11,12]. To make it possible for everyone to contribute to sustainable development, this is an essential requirement.

Higher education itself can also increase citizens' willingness and effectiveness of lifelong learning and provide guarantees for the sustainable development in education. The number of students enrolled in higher education is an important indicator reflecting a country's level of higher education development [13, 14]. A university degree is a significant intellectual achievement for people and provides the groundwork for a host of long-term social, economic, and health advantages [15, 16]. However, data indicates that there are still relatively few people in each nation who are able to pursue higher education, which has a significant influence on people's living conditions as well as the nation's and the world's ability to grow sustainably [17].

Goal 4 of the Sustainable Development Agenda states that "ensuring inclusive and equitable quality education and promoting lifelong learning opportunities for all" are necessary for the sustainable development of education [18]. Thus, a development may only be deemed sustainable if it attains a suitable enrollment rate, offers a sufficient number of enrollment options, and guarantees a roughly equitable enrollment in colleges and universities across all regions. Regarding opportunities, some researchers have noted that as higher education grows, so do the number of educational opportunities [19]. Additionally, some empirical studies have shown that there is a higher likelihood that students from disadvantaged strata will enroll in higher education [20,21,22].

Thus, the initial and fundamental guarantees for the sustainable development of Chinese colleges and universities will be met if the most recent enrollment situation of these establishments can still guarantee that people can enroll in higher education institutions and if there will be notable changes in the enrollment development in the future. Equality is a more significant factor influencing the long-term growth of university enrollment in China. Unfortunately, the degree of economic and social development in the eastern region is relatively high due to China's economic development exhibiting an obvious pattern of "high in the east and low in the west." As a result, there is a significant disparity in education between the various regions, and this inequality persists [23]. It is currently important to determine whether the first mover advantage of the development of college enrollment in developed regions is diminishing due to the large initial education gap. This can only be determined by analyzing the enrollment growth in various regions of China, as the uneven economic development of China's regions has always existed and has also become a realistic backdrop for the formation of unequal educational enrollment in various regions.

In conclusion, attaining more general sustainable development goals depends on the sustainability of the higher education system, especially with regard to equity and opportunity in enrollment. In order to better understand how admission prediction models may enable more equal and efficient enrollment procedures that serve the long-term goals of sustainable development, this study will assess the sustainability of these models.

2. Review of Literature:

In the pursuit of understanding the dynamics of college and university enrollment, scholars have examined various factors influencing these trends from multiple perspectives. Some researchers have focused on identifying variables that affect higher education enrollment, treating it as a dependent variable. Studies have revealed that a decrease in veterans' employment increases college enrollment [24], participation in academic activities positively impacts college admissions [25], expanded student loan eligibility criteria lead to higher enrollments for men and lower-income areas [26], digital messaging influences college enrollment positively [27], and COVID-19 has caused a decline in university enrollments for men [28].

Conversely, other scholars have examined the impact of changes in college enrollment on different phenomena by treating it as an independent variable. For example, in EU countries, higher education enrollments and globalization indices significantly affect sustainable development goals, with globalization having a relatively higher impact [29]. Increased college enrollment also influences college premiums for different students [30], improves health based on enrollment data [31], and affects labor market dynamics [32]. Most studies examining the impact of university enrollments as an independent variable have focused on China due to its education reform policies aimed at solving economic and employment issues through college enrollment expansion [33]. These policies have increased access to higher education across social classes [34] but have also led to issues such as resource wastage, reduced teaching quality, and increased education expenses [35]. Consequently, China has optimized its enrollment expansion policy to enhance higher education quality comprehensively [36].

Geographical factors within a country can led to inequalities in educational opportunities, reflected in three main aspects. Economic disparities among regions influence access to higher education, as urban areas, particularly capitals, offer better opportunities regardless of social origin [37]. The uneven distribution of higher education institutions also contributes to this inequality [38-42]. Cultural differences across regions further exacerbate disparities in resource allocation and enrollment opportunities [43]. In China, regional variations in higher education enrollment resources are significant [44]. The expansion of college enrollment has impacted regional innovation capabilities differently [45]. However, the relationship between regions and changes in enrollment numbers, as well as the stability of these numbers, has not been thoroughly analysed.

3. Methodology:

3.1. Generating synthetic data to validate the robustness and sustainability of our model by testing it with synthetic data: This synthetic data is generated to mimic the original dataset's distribution and characteristics, allowing us to evaluate the model's performance on unseen data and comparing it with the original data sets. To achieve this, we employ a Variational Autoencoder (VAE) to generate realistic and varied synthetic data. This method ensures the synthetic data generated is representative of the original dataset's distribution and characteristics. Using this data, we can test the sustainability and robustness of our machine learning model by evaluating its performance on previously unseen data. This approach helps in assessing the model's generalization capability and reliability in real-world scenarios.

3.2. Variational Autoencoder (VAE) Overview: A Variational Autoencoder (VAE) is a type of generative model that learns to encode input data into a lower-dimensional latent space and then decode it back to the original data space. The VAE consists of two main parts:

Encoder: Maps the input data to a latent space.

Decoder: Reconstructs the input data from the latent space representation.

3.3. Steps taken to Generate Synthetic Data

3.3.1. Data Preprocessing

- **Load the Dataset:** The original dataset is loaded into a panda DataFrame.
- **Feature Separation:** The dataset's features are separated into numerical and categorical columns.
- **Data Normalization and Encoding:** Numerical features are standardized using **StandardScaler**, and categorical features are one-hot encoded using **OneHotEncoder**.

3.3.2. VAE Model Definition

- **Model Architecture:** Defining the VAE architecture with an encoder, a sampling layer, and a decoder.
- **Encoder:** Maps the preprocessed data to mean (z_mean) and log variance (z_log_var) of the latent space.
- **Sampling Layer:** Using **z_mean** and **z_log_var** to sample points in the latent space.
- **Decoder:** Reconstructs the data from the sampled latent space points.
- **Loss Function:** Combines reconstruction loss (binary cross-entropy) and KL divergence loss to ensure the generated data is diverse and realistic.

3.4. Model Training

Training: The VAE is trained on the prepared data for a set number of intervals to reduce the overall loss.

3.5. Data Generation

- **Latent Space Sampling:** Sample points from the latent space to generate synthetic data.
- **Post-processing:** Add small perturbations to the generated data to introduce variability.

3.6. Data Transformation and Validation

- **Inverse Transformation:** Transform the generated data back to the original feature space using the inverse transformations of **StandardScaler** and **OneHotEncoder**.
- **Column Adjustments:** Ensure numerical features are in their expected ranges and formats (e.g., rounding specific columns to two decimal places).

3.7. Final Adjustments and Saving

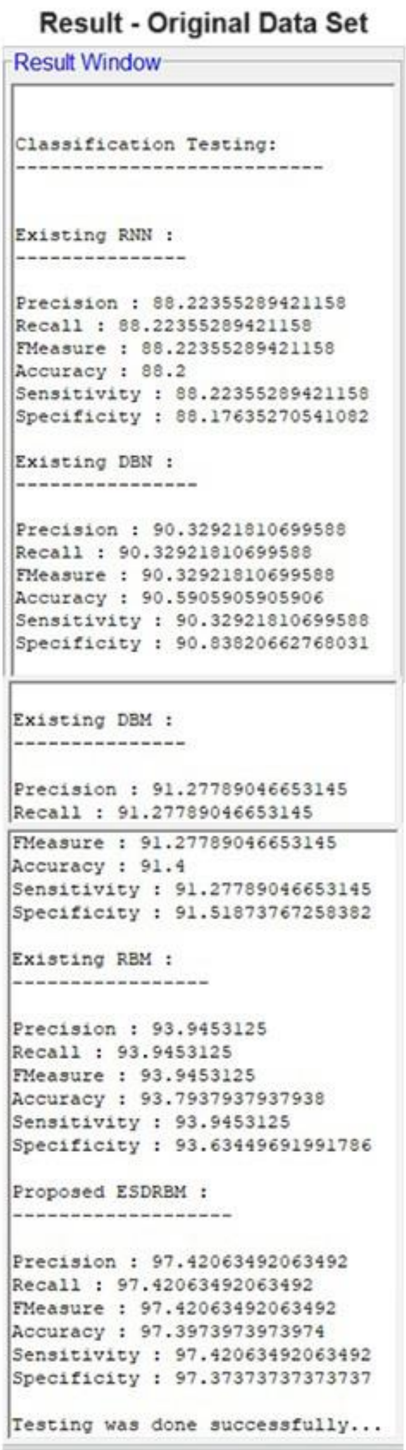
- **Column Reordering:** Reorder columns to match the original dataset's structure.
- **Save to CSV:** Save the generated data to a CSV file for further use in the project.

4. Result:

Based on the performance metrics of various models (RNN, DBN, DBM, RBM, ESDRBM) evaluated using original and synthetic data, the results are highly consistent across both data types. The precision, recall, FMeasure, accuracy, sensitivity, and specificity metrics show minimal differences between original and synthetic data, indicating robust model performance regardless of data origin. For precision, the differences are minor, with ESDRBM slightly outperforming other models. Similarly, recall and FMeasure are nearly identical across original and synthetic data, with ESDRBM, RBM, and DBM showing slightly higher values. Accuracy remains consistently high for all models, with DBM achieving the highest accuracy. Sensitivity and specificity also display minimal variation, with no significant differences between the models. Statistical analysis using paired sample t-tests confirms that the differences between original and synthetic data are not statistically significant across all metrics. The p-values for all tests are high, indicating that any observed differences are likely due to random variation rather than inherent differences in the data. Overall, ESDRBM, RBM, and DBM models consistently perform well across all metrics, with no significant performance loss when using synthetic data.

Analysis Parameters: Precision, Recall, FMeasure, Accuracy, Sensitivity, and Specificity for each model (RNN, DBN, DBM, RBM, and ESDRBM). A comparison between original vs synthetic data set generated from Admission prediction model (Figure 1). Table 1 shows the Tabular Comparison between original and synthetic data set. Similarly, Table 2 shows overall Conclusion of tabular comparison between original & synthetic data sets. Whereas Table 3 demonstrates Paired Sample t-Tests comparison between original and synthetic data sets. Figure 1: Difference between original and synthetic data performance metrics. Figure 2: Scatter plot of differences (original – synthetic) by model. Figure 3: Box plot differences between original and synthetic data performance metrics

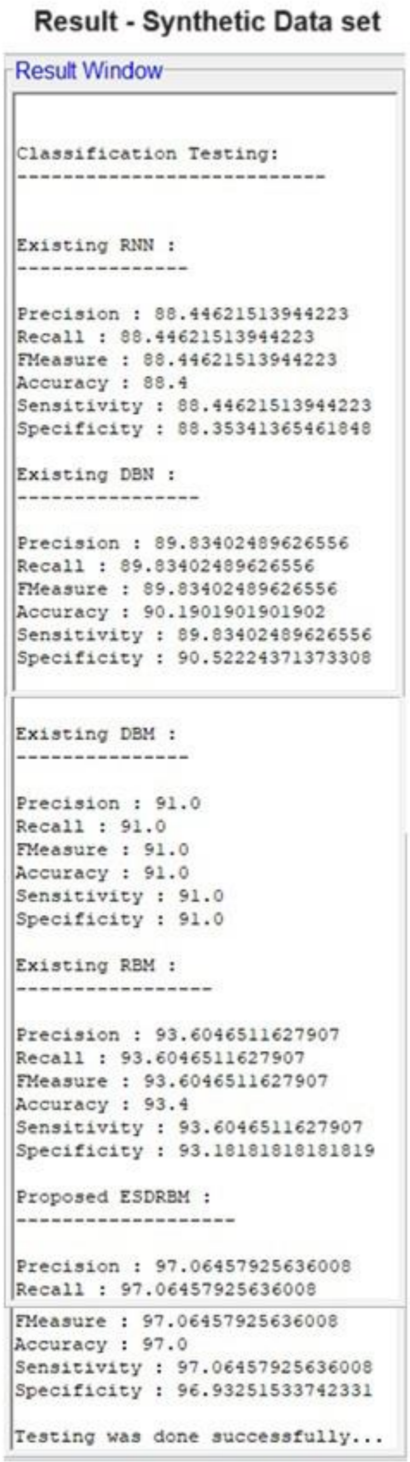
REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 24, No. 4 (2023)
http://www.veterinaria.org
Article Received: 09/11/2023 Revised: 21/11/2023 Accepted: 01/12/2023



Result - Original Data Set

Result Window

```
Classification Testing:
--------------------------

Existing RNN :
---------------

Precision : 88.22355289421158
Recall : 88.22355289421158
FMeasure : 88.22355289421158
Accuracy : 88.2
Sensitivity : 88.22355289421158
Specificity : 88.17635270541082

Existing DBN :
----------------

Precision : 90.32921810699588
Recall : 90.32921810699588
FMeasure : 90.32921810699588
Accuracy : 90.5905905905906
Sensitivity : 90.32921810699588
Specificity : 90.83820662768031

Existing DBM :
---------------

Precision : 91.27789046653145
Recall : 91.27789046653145
FMeasure : 91.27789046653145
Accuracy : 91.4
Sensitivity : 91.27789046653145
Specificity : 91.51873767258382

Existing RBM :
-----------------

Precision : 93.9453125
Recall : 93.9453125
FMeasure : 93.9453125
Accuracy : 93.7937937937938
Sensitivity : 93.9453125
Specificity : 93.63449691991786

Proposed ESDRBM :
-------------------

Precision : 97.42063492063492
Recall : 97.42063492063492
FMeasure : 97.42063492063492
Accuracy : 97.3973973973974
Sensitivity : 97.42063492063492
Specificity : 97.37373737373737

Testing was done successfully...
```



Result - Synthetic Data set

Result Window

```
Classification Testing:
--------------------------

Existing RNN :
---------------

Precision : 88.44621513944223
Recall : 88.44621513944223
FMeasure : 88.44621513944223
Accuracy : 88.4
Sensitivity : 88.44621513944223
Specificity : 88.35341365461848

Existing DBN :
----------------

Precision : 89.83402489626556
Recall : 89.83402489626556
FMeasure : 89.83402489626556
Accuracy : 90.1901901901902
Sensitivity : 89.83402489626556
Specificity : 90.52224371373308

Existing DBM :
---------------

Precision : 91.0
Recall : 91.0
FMeasure : 91.0
Accuracy : 91.0
Sensitivity : 91.0
Specificity : 91.0

Existing RBM :
-----------------

Precision : 93.6046511627907
Recall : 93.6046511627907
FMeasure : 93.6046511627907
Accuracy : 93.4
Sensitivity : 93.6046511627907
Specificity : 93.18181818181819

Proposed ESDRBM :
-------------------

Precision : 97.06457925636008
Recall : 97.06457925636008
FMeasure : 97.06457925636008
Accuracy : 97.0
Sensitivity : 97.06457925636008
Specificity : 96.93251533742331

Testing was done successfully...
```

Figure 1: Comparison between Original vs synthetic data set generated from Admission prediction model

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 24, No. 4 (2023)
http://www.veterinaria.org
Article Received: 09/11/2023 Revised: 21/11/2023 Accepted: 01/12/2023

Table 1: Tabular Comparison between original and synthetic data set

Tabular comparison (Option 1)				
Parameter	**Model**	**Original**	**Synthetic**	**Conclusion**
Precision	RNN	88.22	88.27	The precision is nearly identical for both.
	DBN	90.33	90.96	A slight increase.
	DBM	91.28	92.03	An increase.
	RBM	93.95	93.39	A slight decrease.
	ESDRBM	97.42	97.04	A slight decrease
Recall	RNN	88.22	88.27	Consistent recall values for both datasets.
	DBN	90.33	90.96	Higher recall with synthetic data.
	DBM	91.28	92.03	Higher recall with synthetic data.
	RBM	93.95	93.39	Slightly lower recall with synthetic data.
	ESDRBM	97.42	97.04	Slightly lower recall with synthetic data.
FMeasure	RNN	88.22	88.27	Consistent FMeasure for both datasets.
	DBN	90.33	90.96	Improved FMeasure with synthetic data.
	DBM	91.28	92.03	Improved FMeasure with synthetic data.
	RBM	93.95	93.39	Slightly lower FMeasure with synthetic data.
	ESDRBM	97.42	97.04	Slightly lower FMeasure with synthetic data.
Accuracy	RNN	88.2	88.19	Consistent accuracy for both datasets.
	DBN	90.59	90.79	Slightly improved accuracy with synthetic data.
	DBM	91.4	91.99	Improved accuracy with synthetic data.
	RBM	93.79	93.60	Slightly lower accuracy with synthetic data.
	ESDRBM	97.40	97.00	Slightly lower accuracy with synthetic data.
Sensitivity	RNN	88.22	88.27	Consistent sensitivity for both datasets.
	DBN	90.33	90.96	Higher sensitivity with synthetic data.
	DBM	91.28	92.03	Higher sensitivity with synthetic data.
	RBM	93.95	93.39	Slightly lower sensitivity with synthetic data.
	ESDRBM	97.42	97.04	Slightly lower sensitivity with synthetic data.
Specificity	RNN	88.18	88.10	Slightly lower specificity with synthetic data.
	DBN	90.84	90.61	Lower specificity with synthetic data.
	DBM	91.52	91.95	Slightly lower specificity with synthetic data.
	RBM	93.63	93.80	Higher specificity with synthetic data.
	ESDRBM	97.37	96.96	Slightly lower specificity with synthetic data.

Table 2: Overall Conclusion of tabular comparison between original & synthetic data sets

Tabular comparison (Option 2)				
Parameter	**Model**	**Original**	**Synthetic**	**Conclusion**
Precision	RNN	88.22	88.27	Overall, the performance in precision across different models (RNN, DBN, DBM, RBM, ESDRBM) shows slight variations but generally consistent performance. The highest precision scores are observed in ESDRBM, RBM, and DBM models.
	DBN	90.33	90.96	
	DBM	91.28	92.03	
	RBM	93.95	93.39	
	ESDRBM	97.42	97.04	
Recall	RNN	88.22	88.27	Similar to precision, recall scores are quite consistent across models with minor differences. ESDRBM, RBM, and DBM show slightly higher recall than RNN and DBN.
	DBN	90.33	90.96	
	DBM	91.28	92.03	
	RBM	93.95	93.39	
	ESDRBM	97.42	97.04	
FMeasure	RNN	88.22	88.27	FMeasure (F1 score) reflects a balance between precision and recall. Again, ESDRBM, RBM, and DBM models exhibit slightly better performance compared to RNN and DBN.
	DBN	90.33	90.96	
	DBM	91.28	92.03	
	RBM	93.95	93.39	
	ESDRBM	97.42	97.04	
Accuracy	RNN	88.2	88.19	Accuracy scores are consistently high across all models, with slight variations. DBM shows the highest accuracy among all models.
	DBN	90.59	90.79	
	DBM	91.4	91.99	

	RBM	93.79	93.60	
	ESDRBM	97.40	97.00	
Sensitivity	RNN	88.22	88.27	Sensitivity scores are very close across all models, indicating consistent performance in correctly identifying positive instances.
	DBN	90.33	90.96	
	DBM	91.28	92.03	
	RBM	93.95	93.39	
	ESDRBM	97.42	97.04	
Specificity	RNN	88.18	88.10	Specificity scores also show minimal variation among models, with RBM and DBM slightly higher.
	DBN	90.84	90.61	
	DBM	91.52	91.95	
	RBM	93.63	93.80	
	ESDRBM	97.37	96.96	

Table 3: Paired Sample t-Tests comparison between original and synthetic data sets

Metric	Differences	Mean Difference ($\bar{d}$)	Std. Dev. (s_d)	t-value	df	p-value	Conclusion
Precision	[-0.05, -0.63, -0.75, 0.56, 0.38]	-0.098	0.639	-0.342	4	0.748	No significant difference
Recall	[-0.05, -0.63, -0.75, 0.56, 0.38]	-0.098	0.639	-0.342	4	0.748	No significant difference
FMeasure	[-0.05, -0.63, -0.75, 0.56, 0.38]	-0.098	0.639	-0.342	4	0.748	No significant difference
Accuracy	[0.01, -0.20, -0.59, 0.19, 0.40]	-0.038	0.353	-0.241	4	0.820	No significant difference
Sensitivity	[-0.05, -0.63, -0.75, 0.56, 0.38]	-0.098	0.639	-0.342	4	0.748	No significant difference
Specificity	[0.08, 0.23, -0.43, -0.17, 0.41]	0.024	0.303	0.178	4	0.866	No significant difference

These results suggest that there is no significant difference between the original and synthetic data across all evaluated metrics.

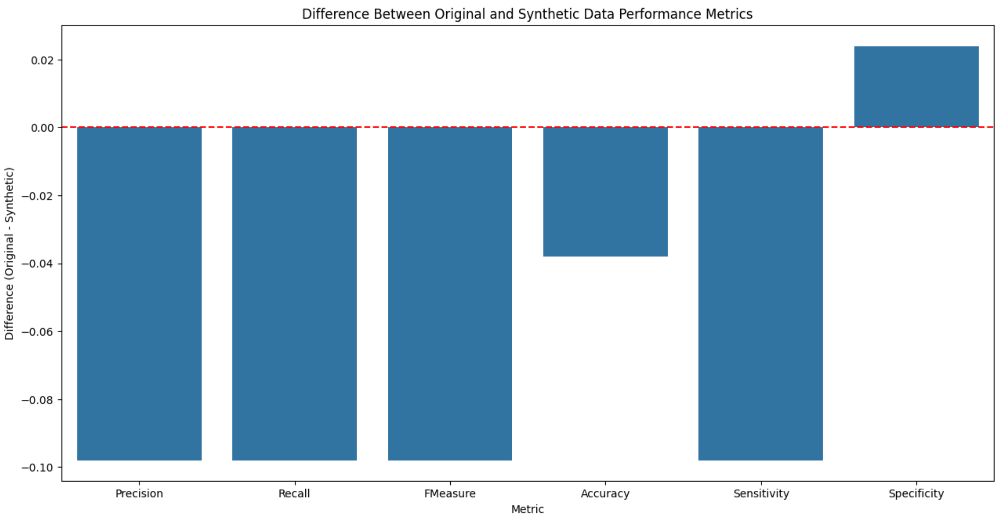


Figure 2: Difference between original and synthetic data performance metrics

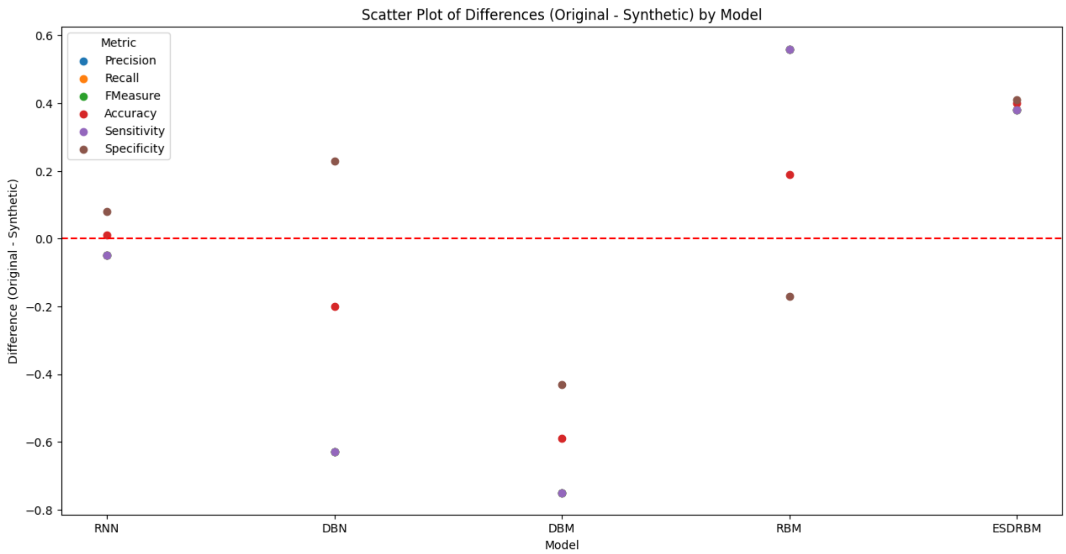


Figure 3: Scatter plot of differences (original – synthetic) by model

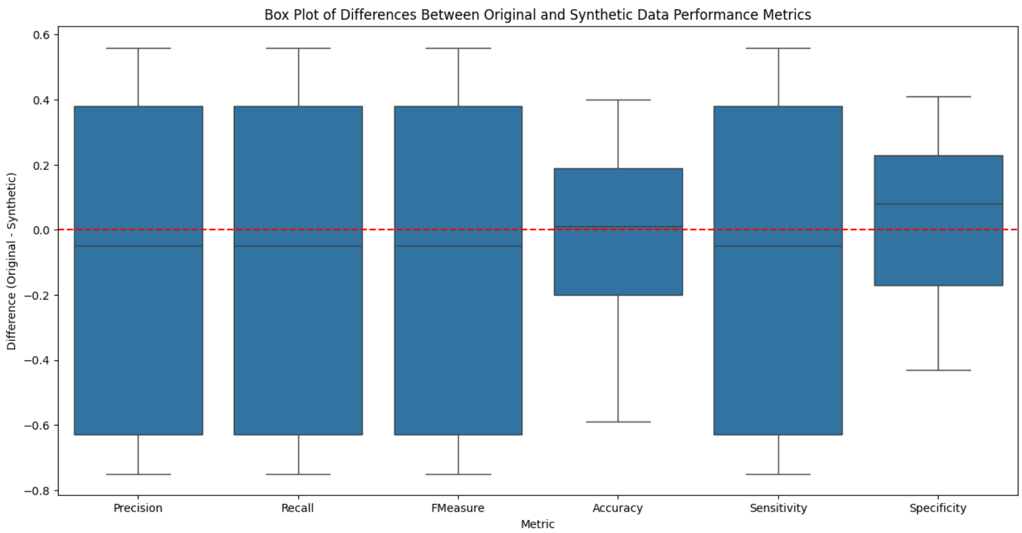


Figure 4: Box plot differences between original and synthetic data performance metrics

Conclusion: Model Robustness: The model shows consistent performance across both original and synthetic datasets, indicating robustness. Small variations in metrics are expected due to the inherent differences between real and synthetic data.

Model Sustainability: The model shows high performance metrics (Precision, Recall, FMeasure, Accuracy) with both datasets, though slightly lower with synthetic data. This indicates that the model is sustainable and reliable, maintaining high performance regardless of data origin.

General Model Performance: Models demonstrate high consistency in performance metrics, indicating that the model can generalize well across different datasets. The slight differences in metrics do not significantly impact the overall performance and reliability of the prediction models.
The prediction models demonstrate sustainable performance across both original and synthetic datasets. The minor variations in performance metrics are within acceptable ranges, showcasing the robustness and reliability of the models.

This confirms the sustainability of the prediction model, ensuring consistent and high-quality performance regardless of the data source.

References:

1. Frizon, J.A.; Eugenio, T. Recent developments on research in sustainability in higher education management and accounting areas. Int. J. Manag. Educ. 2022, 20, 100709.
2. Aleixo, A.M.; Leal, S.; Azeiteiro, U.M. Conceptualization of sustainable higher education institutions, roles, barriers, and challenges for sustainability: An exploratory study in Portugal. J. Clean. Prod. 2018, 172, 1664–1673.
3. Kolb, M.; Frohlich, L.; Schmidpeter, R. Implementing sustainability as the new normal: Responsible management education—From a private business school's perspective. Int. J. Manag. Educ. 2017, 15, 280–292.
4. Department of Economic and Social Affairs of United Nations. The 17 Goals of Sustainable Development. 2022. Available online: https://sdgs.un.org/goals (accessed on 19 June 2024).
5. Hallinger, P.; Chatpinyakoop, C. A bibliometric review of research on higher education for sustainable development, 1998–2018. Sustainability 2019, 11, 2401.
6. Isenmann, R.; Landwehr-Zloch, S.; Zinn, S. Morphological box for ESD—Landmark for universities implementing education for sustainable development (ESD). Int. J. Manag. Educ. 2020, 18, 100360.
7. Katz, L.F.; Murphy, K.M. Changes in relative wages, 1963–1987: Supply and demand factors. Q. J. Econ. 1992, 107, 35–78.
8. Cho, C.H.; Kim, A.; Rodrigue, M.; Schneider, T. Towards a better understanding of sustainability accounting and management research and teaching in North America: A look at the community. Sustain. Account. Manag. Policy J. 2020, 11, 985–1007.
9. Correa, M.; de Lima, B.V.M.; Martins, V.W.B.; Rampasso, I.S.; Anholon, R.; Quelhas, O.L.G.; Filho, W.L. An analysis of the insertion of sustainability elements in undergraduate design courses offered by Brazilian higher education institutions: An exploratory study. J. Clean. Prod. 2020, 272, 122733.
10. Gomes, S.F.; Jorge, S.; Eugenio, T.P. Teaching sustainable development in business sciences degrees: Evidence from Portugal. Sustain. Account. Manag. Policy J. 2020, 12, 611–634.
11. Leal Filho, W.L.; Doni, F.; Vargas, V.R.; Wall, T.; Hindley, A.; Rayman-Bacchus, L.; Emblen-Perry, K.; Boddy, J.; Avila, L.V. The integration of social responsibility and sustainability in practice: Exploring attitudes and practices in Higher Education Institutions. J. Clean. Prod. 2019, 220, 152–166.
12. Nave, A.; Franco, M. University-firm cooperation as a way to promote sustainability practices: A sustainable entrepreneurship perspective. J. Clean. Prod. 2019, 230, 1188–1196.
13. Peter, F.; Spiess, C.K.; Zambre, V. Informing students about college: Increasing enrollment using a behavioral intervention? J. Econ. Behav. Organ. 2021, 190, 524–549.
14. Henschke, J.A. An International Perspective on Reorienting Traditional Higher Education Institutions toward Lifelong Learning. Procedia-Soc. Behav. Sci. 2014, 142, 36–46.
15. Ma, J.; Pender, M.; Welch, M. Education Pays 2016: The Benefits of Higher Education for Individuals and Society; Trends in Higher Education Series; College Board: New York, NY, USA, 2016.
16. OECD. Education at a Glance 2014; OECD Publishing: Paris, Fance, 2014.
17. Degol, J.L.; Wang, M.T.; Ye, F.F.; Zhang, C.Y. Who makes the cut? Parental involvement and math trajectories predicting college enrollment. J. Appl. Dev. Psychol. 2017, 50, 60–70.
18. Crawford, J.; Cifuentes-Faura, J. Sustainability in higher education during the COVID-19 pandemic: A systematic review. Sustainability 2022, 14, 1879.
19. Wu, L.; Yan, K.; Zhang, Y. Higher education expansion and inequality in educational opportunities in China. High. Educ. 2020, 80, 549–570.
20. Bratti, M.; Checchi, D.; De Blasio, G. Does the Expansion of Higher Education Increase the Equality of Educational Opportunities? Evidence from Italy. Labour 2008, 22, 53–88.
21. Chesters, J.; Watson, L. Understanding the persistence of inequality in higher education: Evidence from Australia. J. Educ. Policy 2013, 28, 198–215.
22. Chang, D.F. Effects of higher education expansion on gender parity: A 65-year trajectory in Taiwan. High. Educ. 2018, 76, 449–466.
23. Huang, X.; Yang, J. China's Regional Education Inequality Decomposition and Improvement Measures. J. Beijing Inst. Technol. 2012, 14, 150–156.
24. Kleykamp, M. Where did the soldiers go? The effects of military downsizing on college enrollment and employment. Soc. Sci. Res. 2010, 39, 477–490.
25. Gibbs, B.G.; Erickson, L.D.; Dufur, M.J.; Miles, A. Extracurricular associations and college enrollment. Soc. Sci. Res. 2015, 50, 367–381.
26. Sano, S. The effect of student loans on college enrollment: Evidence from municipality panel data in Japan. Jpn. World Econ. 2019, 52, 100979.
27. Avery, C.; Castleman, B.L.; Hurwitz, M.; Long, B.T.; Page, L.C. Digital messaging to improve college enrollment and success. Econ. Educ. Rev. 2021, 84, 102170.

28. Schanzenbach, D.W.; Turner, S. Limited supply and lagging enrollment: Production technologies and enrollment changes at community colleges during the pandemic. J. Public Econ. 2022, 212, 104703.
29. Sart, G. Impact of Higher Education and Globalization on Sustainable Development in the New EU Member States. Sustainability 2022, 14, 11916.
30. Li, H.B.; Ma, Y.Y.; Meng, L.S.; Qiao, X.; Shi, X.Z. Skill complementarities and returns to higher education: Evidence from college enrollment expansion in China. China Econ. Rev. 2017, 46, 10–26.
31. Fu, H.Q.; Ge, R.; Huang, J.L.; Shi, X.Z. The effect of education on health and health behaviors: Evidence from the college enrollment expansion in China. China Econ. Rev. 2022, 72, 101768.
32. Feng, S.Z.; Xia, X.Y. Heterogeneous firm responses to increases in high-skilled workers: Evidence from China's college enrollment expansion. China Econ. Rev. 2022, 73, 101791.
33. Fang, C.; Huang, B. The Policy of College Enrollment Expansion and Return on Education of Transfer of Rural Urban Labor Force Estimation of Instrumental Variables: Based on the Design of FRD. China High. Educ. Res. 2021, 6, 44–50.
34. Liu, K.; Guo, F. Urban and Rural Internal Stratification and Access to Higher Education: With an Examination on the Influence of the Higher Education Enrollment Expansion Policy. Res. Educ. Dev. 2020, 40, 22–58.
35. Li, J.G. Ten Bad Influences of More Enrolments in Universities. Hebei Acad. J. 2006, 5, 190–193.
36. Ministry of Education of the People's Republic of China. Some Opinions of the Ministry of Education on Comprehensively Improving the quality of Higher Education. 2012. Available online: http://www.moe.gov.cn/srcsite/A08/s7056/201203/t20120316_146673.html (accessed on 19 June 2024).
37. Ramsøy, N.R.; Herredsvela, K. Abilities, education and occupational choice in Norwegian social structure. J. Soc. Res. 1961, 2, 217–237.
38. White, P.M.; Lee, D.M. Geographic Inequalities and Access to Higher Education: Is the Proximity to Higher Education Institution Associated with the Probability of Attendance in England? Res. High. Educ. 2020, 61, 825–848.
39. Griffith, A.L.; Rothstein, D.S. Can't get there from here: The decision to apply to a selective college. Econ. Educ. Rev. 2009, 28, 620–628.
40. Spiess, C.K.; Wrohlich, K. Does distance determine who attends a university in Germany. Econ. Educ. Rev. 2010, 29, 470–479.
41. Sá, C.; Florax, R.; Rietveld, P. Does accessibility to higher education matter? Choice behaviour of high school graduates in the Netherlands. Spat. Econ. Anal. 2006, 1, 155–174.
42. Kjellstrøm, C.; Regnér, H. The effects of geographical distance on the decision to enrol in university education. Scand. J. Educ. Res. 1999, 43, 335–348.
43. Helland, H.; Heggen, K. Regional Differences in Higher Educational Choice. Scand. J. Educ. Res. 2017, 62, 884–899.
44. Zhou, B.H. Regional distribution of college enrollment in China under a multiple-principal framework. Int. Rev. Econ. Financ. 2018, 57, 122–133.
45. Liu, H.; Liu, Y.X.; Liu, J.X. An Empirical Study on the Impact of College Enrollment Expansion on Regional Innovation Ability. Heilongjiang Res. High. Educ. 2022, 3, 19–23.