2025 IEEE International Conference on Artificial Intelligence and Mechatronics Systems (AIMS) | 979-8-3315-7805-3/25/$31.00 ©2025 IEEE | DOI: 10.1109/AIMS66189.2025.11229547

Multi-Dimensional Quality Assessment of Synthetic Data across ERP Modules

Kurnia Cahya Febryanto
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
6025241065@student.its.ac.id

Shintami Chusnul Hidayati
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
shintami@its.ac.id

Riyanarto Sarno
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
riyanarto@its.ac.id

***Abstract*—The rapid adoption of Enterprise Resource Planning (ERP) systems has highlighted critical data availability and quality challenges, particularly for predictive modeling and analytics. This research presents a comprehensive framework for synthetic data generation in ERP environments using advanced generative models, including GANs, CGANs, VAEs, Beta-VAEs, and Normalizing Flows. The framework encompasses data processing across Sales, Purchase, and Human Resource modules, implementing specialized loss functions for maintaining business rules and inter-module relationships. Experimental results demonstrate exceptional performance of the VAE architecture in Purchase and Human Resource modules with accuracy scores of 99.3 percent and 95.2 percent, respectively. In comparison, CGANs achieve 95.1 percent accuracy in Sales module synthesis.**

***Index Terms*—Enterprise Resource Planning, Synthetic Data Generation, Generative Models, Quality Assessment**

I. Introduction

Enterprise Resource Planning (ERP) systems are fundamental to modern business operations, yet predictive model development faces significant challenges due to data scarcity in historical adoption records and Critical Success Factor ratings [1]. This limitation compromises implementation effectiveness across sectors. Additionally, enterprise data privacy concerns and regulatory compliance requirements create barriers to accessing quality training data for predictive modeling.

Autoencoders have demonstrated effectiveness in capturing complex data relationships [2], showing significant potential for ERP data synthesis. Research in multicolor space adaptation [3] and structured body shape modeling [4] highlights the importance of preserving multi-dimensional attributes. Furthermore, studies on user profiles influencing social media popularity [5] emphasize the value of contextual integration for predictive performance. This research utilizes these insights to develop a framework addressing ERP dataset complexity through structured modeling, feature adaptation, and metadata integration.

Conventional data augmentation fails to capture complex ERP system dependencies and temporal patterns. Despite Generative AI advancements showing promise for synthetic data creation [6], [7], current implementations lack frameworks addressing enterprise-specific constraints and inter-module relationships. This research gap is particularly evident in ERP environments, where maintaining logical consistency across interconnected modules remains challenging [1].

This research addresses ERP data limitations through a novel multimodel comparative approach grounded in financial forecasting, performance evaluation systems, and process mining advancements. The study delivers three key contributions: a methodology for synthetic data generation across interrelated ERP modules, multi-dimensional quality assessment metrics preserving business rules, and evidence-based guidelines for model selection based on module characteristics. These contributions enhance predictive modeling capabilities while maintaining data privacy and regulatory compliance in enterprise environments.

II. Related Works

Generative models address ERP data scarcity effectively. VAEs excel in representation learning, while GANs maintain statistical integrity. However, current frameworks inadequately address enterprise-specific constraints and inter-module dependencies critical for maintaining consistent relationships across ERP components [8].

Quality assessment is crucial for validating synthetic ERP data [9]. Each module presents unique requirements: HR demands the preservation of organizational hierarchies, Sales requires accurate transaction patterns, and Purchasing depends on consistent inventory constraints [10]. While recent methods combine statistical fidelity with domain-specific indicators [11], ensuring module-level integrity and cross-module consistency remains challenging [12]. Standardized metrics are needed to assess synthetic data while protecting business processes [13].

III. Proposed Method

The methodology integrates data processing, model implementation, and evaluation (Fig. 1).

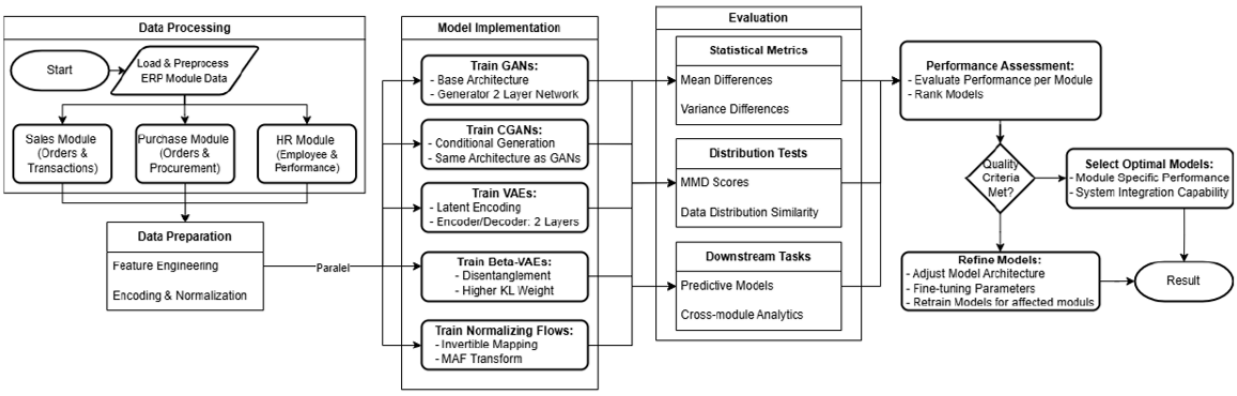

Fig. 1. Illustration of the proposed approach.

A. Data Processing

Maharana *et al.* [14] establish that data pre-processing constitutes up to 80% of classification processes, fundamentally enhancing model performance. The implemented pipeline utilizes SAP Datasphere sample content across ERP modules, representing actual enterprise data from real-world business operations. Processing operations include non-essential column removal, datetime standardization, and referential integrity preservation via table merging, as detailed in Table I. Stratified sampling maintains distribution balance with approximately 5,000 samples per module while preserving critical business patterns across transaction metrics. Median imputation addresses numerical fields, while interquartile range-based outlier removal achieves 85% data retention with maintained statistical significance.

B. Data Preparation

In accordance with established ERP data preparation frameworks [9], we carry out data transformation and cleaning through the following steps.

1) Feature Engineering: Module-specific transformations derive critical business metrics across HR (salary ratios, departmental statistics), Purchase (financial margins, delivery indicators), and Sales (revenue growth factors) modules, preserving complex relationships and business logic while maintaining data validity.

2) Data Transformation: The transformation approach employs variable-appropriate encoding techniques: one-hot encoding for low-cardinality variables (status codes, departments), label encoding for high-cardinality fields (products, customers), and binary encoding for boolean flags. Numerical features undergo MinMaxScaler normalization to [0,1] range, while missing values are addressed through median imputation (numeric fields) and mode imputation (categorical variables). Datetime features are decomposed into component parts, creating a normalized feature space that preserves statistical relationships while facilitating cross-module model training.

TABLE I
ERP MODULE TABLES AND COLUMNS STRUCTURE.

Sales Module	
Business Partners	PartnerId, PartnerRole, CompanyName, EmailAddress, PhoneNumber, WebAddress, AddressId, LegalForm, Currency
Sales Orders	SalesOrderId, PartnerId, SalesOrg, FiscalYearVariant, FiscalYearPeriod, GrossAmount, NetAmount, TaxAmount, LifecycleStatus, BillingStatus, DeliveryStatus
Sales Order Items	SalesOrderId, SalesOrderItem, ProductId, Currency, GrossAmount, NetAmount, TaxAmount, Quantity, QuantityUnit, DeliveryDate, ItemAtpStatus
Purchase Module	
Customer	CustomerId, CustomerTypeId, Country, Language
Financial Transactions	TransactionId, AccountId, AccountTypeId, CustomerId, ProductId, ProductCategoryId, ProfitCenterId, Version, Date, Value
Products	ProductId, TypeCode, ProductCategoryId, Language, SupplierPartnerId, TaxTariffCode, QuantityUnit, WeightMeasure, WeightUnit, Currency, Price, Width, Depth, Height, DimensionUnit
HR Module	
Employee Personal	EmployeeId, FirstName, LastName, FullName, Gender, MaritalStatus, Ethnicity, YearOfBirth, LocationId
Employee Position	EmployeeId, JobId, NewToPosition, PayGradeId, FullTime, CompanyCode, BusinessArea
Employee Performance	EmployeeId, Performance, ImpactOfLoss, FutureLeader
Departments	DepartmentId, Language
Department Texts	DepartmentId, Language, ShortDescription, MediumDescription, LongDescription
Divisions	DivisionId, Language
Division Texts	DivisionId, Language, ShortDescription, MediumDescription, LongDescription
Jobs	JobId, JobClassificationId, Language
Job Classification	JobClassificationId, Language
Job Texts	JobId, Language, ShortDescription, MediumDescription, LongDescription

C. Generative Models Implementation

1) Generative Adversarial Network (GAN): Building on prior research [15], [16], the GAN architecture synthesizes ERP data using generator G projecting noise

$p_z \sim \mathcal{N}(0,1)$ and discriminator D. The objective function (1) combines adversarial loss with feature matching $\mathcal{L}_{feature}$ (2) for module correlations and business constraints $\mathcal{L}_{business}$ (3) for domain compliance.

$$\begin{aligned}\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{erp}}[\log D(x)]+ \\ \mathbb{E}_{z\sim p_z}[\log(1-D(G(z)))]+ \\ \lambda_f \mathcal{L}_{feature} + \lambda_b \mathcal{L}_{business}\end{aligned} \tag{1}$$

$$\mathcal{L}_{feature} = \sum_{m=1}^{M}\sum_{i=1}^{N_m} \|f_i^m(x) - f_i^m(G(z))\|_2^2 \tag{2}$$

$$\mathcal{L}_{business} = \sum_{r=1}^{R} \omega_r \|\phi_r(G(z)) - y_r\|_2^2 \tag{3}$$

2) Conditional GAN (CGAN): CGAN enhances synthesis by leveraging conditional adversarial training [17], [18]. The objective function (4) models the real distribution p_{data}, noise vectors z, and conditioning variables c. Additionally, it integrates $\mathcal{L}_{module}$ (5) to capture characteristic patterns via weights ω_m and $\mathcal{L}_{constraint}$ (6) to enforce domain rules through functions ϕ_r across modules.

$$\begin{aligned}\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}}[\log D(x|c)]+ \\ \mathbb{E}_{z\sim p_z}[\log(1-D(G(z|c)))]+ \\ \lambda_m \mathcal{L}_{module} + \lambda_c \mathcal{L}_{constraint}\end{aligned} \tag{4}$$

$$\mathcal{L}_{module} = \sum_{m\in M} \omega_m \|f_m(x) - f_m(G(z|c))\|_2^2 \tag{5}$$

$$\mathcal{L}_{constraint} = \sum_{r=1}^{R} \alpha_r \|\phi_r(G(z|c)) - y_r\|_2^2 \tag{6}$$

3) Variational Autoencoder (VAE): Drawing from insurance data synthesis [12], generative approaches [16], and linked data frameworks [10], the VAE architecture implements module-specific ERP synthesis through evidence lower bound optimization (7). This balances reconstruction accuracy and latent regularization via encoder $q_\phi(z|x)$ and decoder $p_\theta(x|z)$ networks. The loss function (8) combines reconstruction fidelity with KL divergence regularization, where λ_{KL} controls module-specific regularization strength.

$$\begin{aligned}\mathcal{L}_{VAE}(\theta,\phi;x) = \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)]- \\ \lambda_{KL} D_{KL}(q_\phi(z|x)||p(z))\end{aligned} \tag{7}$$

$$\mathcal{L}_{total} = \sum_{m=1}^{M} \omega_m (\mathcal{L}_{VAE}^m + \alpha_m \|f_m(x) - f_m(p_\theta(z))\|_2^2) \tag{8}$$

4) Beta-VAE: Beta-VAE enhances ERP synthesis by optimizing reconstruction quality and latent factor disentanglement (9). Parameter β regulates encoder $q_\phi(z|x_m)$ and decoder $p_\theta(x_m|z)$ networks. Module-specific patterns incorporate weighted loss functions (10) with penalties α_m, while disentanglement adapts via annealing (11) based on training progress t, duration T, and β ranges.

$$\begin{aligned}\mathcal{L}_\beta(\theta,\phi;x_m) = \mathbb{E}_{q_\phi(z|x_m)}[\log p_\theta(x_m|z)]- \\ \beta D_{KL}(q_\phi(z|x_m)||p(z))\end{aligned} \tag{9}$$

$$\mathcal{L}_{module} = \sum_{m\in M} \omega_m (\mathcal{L}_\beta^m + \alpha_m \|\hat{x}_m - x_m\|_2^2) \tag{10}$$

$$\beta(t) = \min(\beta_{max}, \beta_{min} + (\beta_{max} - \beta_{min})\frac{t}{T}) \tag{11}$$

5) Normalizing Flows: Based on foundations from [19] and [20], the Normalizing Flows framework implements exact likelihood computation for ERP synthesis through the log-density formula (12). The invertible transformation f maps between latent variable z and module data $\mathbf{x}_m$ with Jacobian J_f, while incorporating module-specific loss terms $\mathcal{L}_{module}$ (13) that enforce data characteristics through weights ω_m, penalties α_m, and feature extractors $h_m(\cdot)$ tailored for each module.

$$\begin{aligned}\log p_{\mathbf{x}_m}(\mathbf{x}_m) = \log p_z(f(\mathbf{x}_m)) + \log|\det J_f(\mathbf{x}_m)|+ \\ \sum_{m\in M} \mathcal{L}_{module}\end{aligned} \tag{12}$$

$$\mathcal{L}_{module} = \sum_{m\in M} \omega_m (\alpha_m \|h_m(\mathbf{x}_m) - h_m(f^{-1}(z))\|_2^2) \tag{13}$$

D. Evaluation Framework

1) Statistical Metrics: Distributional similarity between real and synthetic data is assessed through mean and variance differences, as defined in (14) and (15). Variables μ_r and μ_s represent feature means, σ_r^2 and σ_s^2 denote variances, f indicates specific features, and N represents the total numeric features evaluated.

$$\text{Mean Diff}_f = \frac{1}{N}\sum_{i=1}^{N} |\mu_{r,f} - \mu_{s,f}| \tag{14}$$

$$\text{Var Diff}_f = \frac{1}{N}\sum_{i=1}^{N} |\sigma_{r,f}^2 - \sigma_{s,f}^2| \tag{15}$$

2) Distribution Tests: Distribution similarity is evaluated using Maximum Mean Discrepancy (MMD) (16). MMD utilizes a Gaussian kernel function $k(x_i, x_j)$ with bandwidth σ determined by median heuristic. Variables x_i and y_j represent samples from real and synthetic

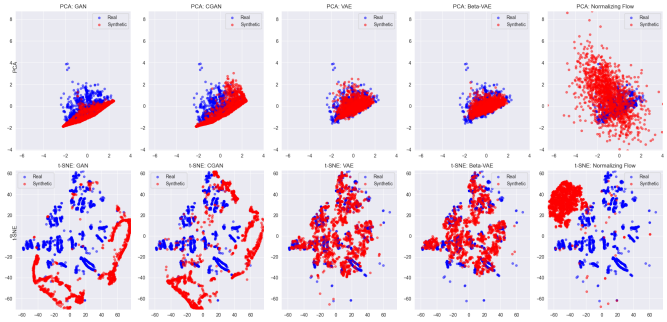

Fig. 2. HR: real (blue), synthetic (red); PCA (top), t-SNE (bottom).

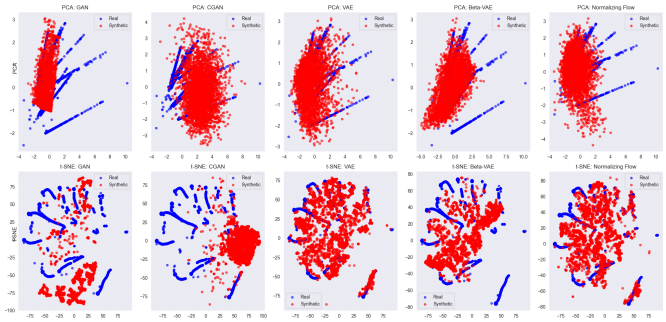

Fig. 3. Purchase: real (blue), synthetic (red); PCA (top), t-SNE (bottom).

distributions with sizes m and n.

$$\text{MMD}(X,Y) = \sqrt{\frac{1}{m^2}\sum_{i,j=1}^{m} k(x_i,x_j) + \frac{1}{n^2}\sum_{i,j=1}^{n} k(y_i,y_j) - \frac{2}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} k(x_i,y_j)} \tag{16}$$

3) Downstream Tasks: Synthetic data utility is evaluated through module-specific machine learning tasks using F1-score (17) and accuracy. Variables TP_t, TN_t, FP_t, and FN_t represent true positives, true negatives, false positives, and false negatives for task t, with precision-recall balance factor $\beta = 1$.

$$\text{F1}_t = \frac{(1+\beta^2)\cdot TP_t}{(1+\beta^2)\cdot TP_t + \beta^2 \cdot FN_t + FP_t} \tag{17}$$

IV. Results and Discussion

A. Visual Distribution Analysis

Model distribution matching is visualized using PCA and t-SNE across HR, Purchase, and Sales modules (Fig. 2, 3, 4). PCA captures global variance while t-SNE reveals local neighborhood preservation. The overlap between real and synthetic data points provides qualitative assessment of generation quality across architectural approaches.

B. Statistical Metric Analysis

Table II and Fig. 5 show VAE excels in HR and Purchase modules with lowest MMD scores (HR: 0.0072, Purchase: 0.1178), mean differences, and optimal variance preservation. Conversely, CGAN demonstrates superior performance in Sales module (MMD: 0.0987, variance difference: 0.0118). Results confirm VAE's

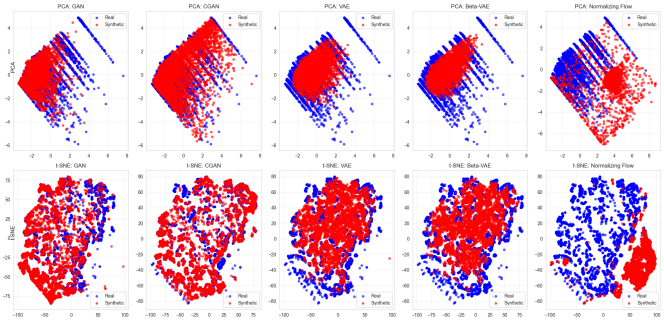

Fig. 4. Sales: real (blue), synthetic (red); PCA (top), t-SNE (bottom).

TABLE II

Comparison of statistical quality metrics, including Maximum Mean Discrepancy (MMD), Mean Difference (Mean Diff), and Variance Difference (Var Diff).

Model	Metric	Module		
		Sales	HR	Purchase
GAN	MMD	0.2859	0.0819	0.3455
	Mean Diff	0.1093	0.0764	0.0075
	Var Diff	0.0241	0.0570	0.0062
CGAN	MMD	**0.0987**	0.0604	0.4703
	Mean Diff	0.0202	0.1453	0.1800
	Var Diff	**0.0118**	0.0539	0.0279
VAE	MMD	0.2117	**0.0072**	**0.1178**
	Mean Diff	**0.0040**	**0.0049**	**0.0025**
	Var Diff	0.0272	**0.0099**	**0.0020**
Beta-VAE	MMD	0.2211	0.0078	0.2634
	Mean Diff	0.0092	0.0046	0.0032
	Var Diff	0.0282	0.0132	0.0033
Normalizing Flow	MMD	0.6573	0.0174	0.1725
	Mean Diff	0.3142	0.0429	0.0021
	Var Diff	0.0187	0.0155	0.0027

effectiveness for HR/Purchase synthesis while CGAN specializes in Sales data generation.

C. Feature Preservation Analysis

Table III demonstrates distinct feature preservation patterns across modules. In HR, Beta-VAE excels in pattern preservation (KL: 54.53) while VAE performs best in age distribution (30.81). For Purchase data, VAE achieves lowest KL values in value (31.17), price (58.53), and weight (68.72). Sales module highlights CGAN's exceptional performance with minimal KL divergence in gross amount (3.97), net amount (3.89), and quantity (78.82). VAE and Beta-VAE effectively preserve numerical attributes in financial data, while Normalizing Flow maintains consistent moderate scores, balancing preservation and distribution matching.

D. Downstream Task Performance

Table IV demonstrates varied model capabilities across nine critical tasks. Beta-VAE excels in HR module (F1: 0.873) for performer identification and management tracking, while CGAN dominates Sales tasks (F1: 0.960) in order prediction and delivery assessment. For Purchase module, Normalizing Flow achieves exceptional consistency (F1: 0.999) in transaction detection and pricing prediction, with VAE showing strong performance in high-value transactions (0.999) and bulk

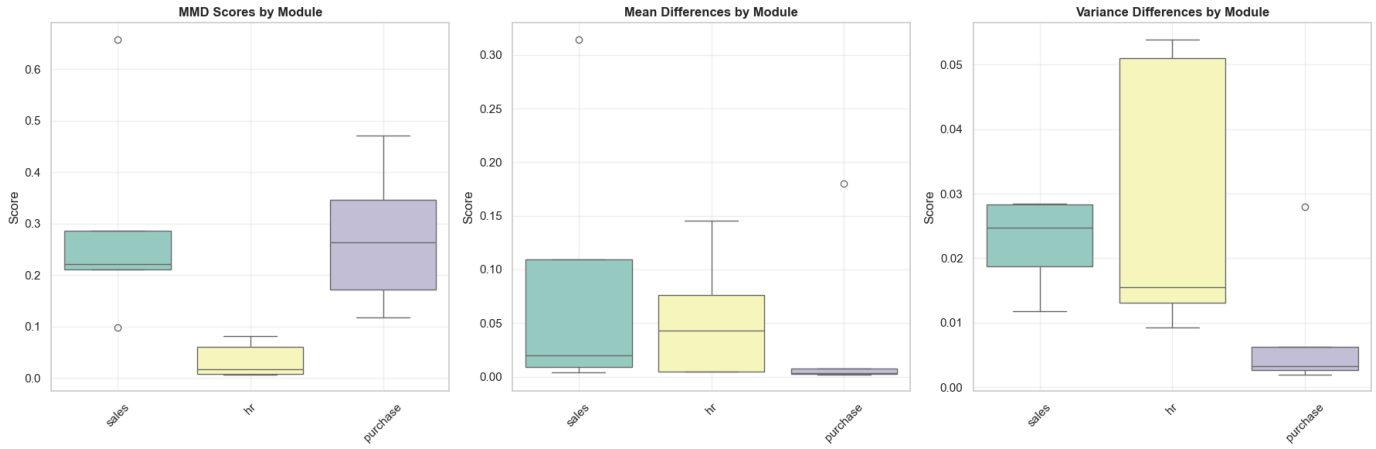

Fig. 5. Statistical quality metrics distribution across ERP modules showing MMD Scores (left), Mean Differences (middle), and Variance Differences (right).

TABLE III
KL DIVERGENCE VALUES FOR FEATURE PRESERVATION ANALYSIS ACROSS MODULES (LOWER VALUES INDICATE BETTER RETENTION).

Model	HR Module			Purchase Module			Sales Module		
	Age	Salary	Perf	Value	Price	Weight	Gross	Net	Qty
GAN	62.40	287.18	123.25	150.54	310.76	464.86	20.99	19.09	81.32
CGAN	67.20	80.46	120.04	53.90	109.32	223.41	3.97	3.89	78.82
VAE	30.81	14.15	61.96	31.17	58.53	68.72	52.88	160.89	302.79
Beta-VAE	92.62	23.37	54.53	121.73	61.47	70.22	201.22	87.97	300.19
N-Flow	28.51	12.91	128.84	58.34	67.14	68.96	127.33	143.72	123.38

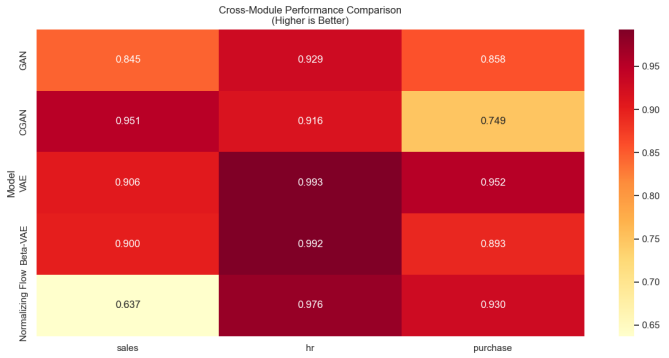

Fig. 6. Cross-module performance heatmap distribution.

purchases (0.895). GAN exhibits moderate overall performance with effectiveness in specific metrics across modules.

E. Comparative Module Analysis

Cross-module evaluation results in Table V and Fig. 6 demonstrate distinctive model performance patterns across ERP components. The VAE exhibits consistent excellence, achieving top rankings in HR (0.993, 1st) and Purchase (0.952, 1st) while maintaining strong Sales performance (0.906, 2nd). The CGAN shows specialized proficiency in Sales (0.951, 1st) despite lower Purchase performance (0.749, 5th). At the same time, Beta-VAE maintains balanced effectiveness across modules, particularly in HR (0.992, 2nd). Normalizing Flow presents varying capabilities from strong Purchase results (0.930, 2nd) to reduced Sales efficiency (0.637, 5th), highlighting the inherent trade-offs between architectural choices and module-specific synthesis requirements.

V. CONCLUSION

This study evaluates synthetic ERP data quality based on statistical fidelity, feature preservation, and downstream task performance across HR, Sales, and Purchase modules. VAE shows strong statistical fidelity for HR and Purchase data, while CGAN captures Sales distributions. Beta-VAE best preserves HR features, CGAN retains key financial attributes in Sales, and Normalizing Flow provides balanced feature preservation. For downstream tasks, CGAN leads in Sales, Beta-VAE in HR, and Normalizing Flow in Purchase classifications, with VAE effective in financial predictions. Future work will explore hybrid models and domain constraints to enhance fidelity while addressing privacy, security, and cross-vendor generalizability to extend the framework to Oracle and Microsoft Dynamics platforms with different data structures.

REFERENCES

[1] K. C. Hong, A. S. B. Shibghatullah, T. C. Ling, and S. M. Sarsam, "Generative AI-powered predictive analytics model: Leveraging synthetic datasets to determine ERP adoption success through critical success factors," *"Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 5, 2024.

[2] Z.-M. Liu, Y.-Y. Chen, S. Hidayati, S.-C. Chien, F.-C. Chang, and K.-L. Hua, "3D model retrieval based on deep autoencoder neural networks," in *Proc. Int. Conf. Signals and Systems*, 2017, pp. 290–296.

[3] S. C. Hidayati, M. Valda Rizky Nur Firdaus, R. W. Nur Dianto, and Sarwosri, "Unleashing attributes-content adaptation with multi-color spaces for food photo aesthetic assessment," in *Proc. APSIPA Annual Summit and Conf.*, 2024, pp. 1–6.

[4] S. C. Hidayati and Y. Anistyasari, "Body shape calculator: Understanding the type of body shapes from anthropometric measurements," in *Proc. ACM Int. Conf. Multimedia Retrieval*, 2021, pp. 461–465.

[5] S. C. Hidayati, M. R. Fiqih Thalib, and A. Munif, "The influence of user profile and post metadata on the popularity of image-based social media: A data perspective," in *Proc. Int. Conf. Artificial Intelligence in Information and Communication*, 2024, pp. 806–811.

TABLE IV
F1-SCORES FOR DOWNSTREAM TASKS ACROSS HR, SALES, AND PURCHASE MODULES.

Model	HR Module Tasks			Sales Module Tasks			Purchase Module Tasks		
	High Perf	**High Imp**	**Mgmt**	**High Val**	**Lg Qty**	**Fast Del**	**High Val**	**Bulk**	**Prem Price**
Real	1.000	0.847	0.769	1.000	1.000	1.000	1.000	1.000	0.999
Beta-VAE	1.000	0.843	0.776	0.987	1.000	0.323	1.000	0.911	0.909
VAE	1.000	0.847	0.759	0.985	1.000	0.419	0.999	0.895	0.909
GAN	1.000	0.849	0.766	0.787	1.000	0.481	1.000	0.775	0.670
CGAN	1.000	0.702	0.766	0.995	1.000	0.887	0.932	0.778	0.614
Norm Flow	0.748	0.849	0.766	0.485	0.890	0.315	1.000	0.999	0.998

TABLE V
CROSS-MODULE PERFORMANCE AND MODEL RANKINGS.

Model	**Module**		
	Sales	**HR**	**Purchase**
GAN	0.845 (4^{th})	0.929 (5^{th})	0.858 (4^{th})
CGAN	**0.951** (1^{st})	0.916 (4^{th})	0.749 (5^{th})
VAE	0.906 (2^{nd})	**0.993** (1^{st})	**0.952** (1^{st})
Beta-VAE	0.900 (3^{rd})	0.992 (2^{nd})	0.893 (3^{rd})
Normalizing Flow	0.637 (5^{th})	0.976 (3^{rd})	0.930 (2^{nd})

[6] M. Goyal and Q. H. Mahmoud, "A systematic review of synthetic data generation techniques using generative AI," *Electronics*, vol. 13, no. 17, p. 3509, 2024.

[7] S. Liu, J. Chen, Y. Feng, Z. Xie, T. Pan, and J. Xie, "Generative artificial intelligence and data augmentation for prognostic and health management: Taxonomy, progress, and prospects," *Expert Systems with Applications*, vol. 255, p. 124511, 2024.

[8] S. Assefa, D. Dervovic, M. Mahfouz, T. Balch, P. Reddy, and M. Veloso, "Generating synthetic data in finance: opportunities, challenges and pitfalls," in *Proc. Conf. Neural Information Processing Systems, Works. AI in Financial Services*, Vancouver, Canada, 2019, pp. 1–10.

[9] A. A. A. Fernandes, M. Koehler, N. Konstantinou, P. Pankin, N. W. Paton, and R. Sakellariou, "Data preparation: A technological perspective and review," *SN Computer Science*, vol. 4, p. 425, 2023.

[10] R. Dos Santos and J. Aguilar, "A synthetic data generation system based on the variational-autoencoder technique and the linked data paradigm," *Progress in Artificial Intelligence*, vol. 13, pp. 149–163, 2024.

[11] L. B. Iantovics and C. Enăchescu, "Method for data quality assessment of synthetic industrial data," *Sensors*, vol. 22, no. 4, p. 1608, 2022.

[12] C. Jamotton and D. Hainaut, "Variational AutoEncoder for synthetic insurance data," *Intell. Syst. Appl.*, vol. 24, p. 200455, 2024.

[13] G. Park and M. Song, "Predicting performances in business processes using deep neural networks," *Decis. Support Syst.*, vol. 129, p. 113191, 2020.

[14] K. Maharana, S. Mondal, and B. Nemade, "A review: Data preprocessing and data augmentation techniques," *Global Transitions Proceedings*, vol. 3, pp. 91–99, 2022.

[15] D. Carvajal-Patiño and R. Ramos-Pollán, "Synthetic data generation with deep generative models to enhance predictive tasks in trading strategies," *Res. Int. Bus. Finance*, vol. 62, p. 101747, 2022.

[16] A. M. Venugopal, T. S. Tran, and M. Endres, "Synthetic data generation: A comparative study," in *Proc. Int. Database Engineering & Applications Symposium*, ser. IDEAS'22. ACM, 2022, pp. 94–102.

[17] A. H. Aziira, N. A. Setiawan, and I. Soesanti, "Generation of synthetic continuous numerical data using generative adversarial networks," *Journal of Physics: Conference Series*, vol. 1577, no. 1, p. 012027, 2020.

[18] A. Ruiz-Gándara and L. Gonzalez-Abril, "Generative adversarial networks in business and social science," *Applied Sciences*, vol. 14, no. 17, p. 7438, 2024.

[19] I. Kobyzev, S. J. Prince, and M. A. Brubaker, "Normalizing flows: An introduction and review of current methods," *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 43, no. 11, pp. 3964–3979, 2021.

[20] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan, "Normalizing flows for probabilistic modeling and inference," *J. Machine Learning Research*, vol. 22, pp. 1–64, 2021.