

Data Description

2024 年度プロジェクト : DB2024-18①

Data Number (will be filled by the Commons Center)	DPPSCdbp_2024-18①en
Title	Synthetic Journal Entry Data (Medical Domain / Clinics) Simply Generated Using a Variational Autoencoder
Creator	Sota Mashiko, Ryoki Motai, Ryota Shin, Yuji Kawamata, Yukihiro Okada
Contact information	Yukihiro Okada okayu@sk.tsukuba.ac.jp
Summary (Subject and/or Methodology for generation)	This dataset consists of synthetic journal entry data generated from a Variational Autoencoder (VAE), trained on real journal entries provided by Tax Accountant Corporation Nihon Keiei, with prior consent obtained from individual client entities. The synthetic entries are generated by inputting randomly sampled vectors into a trained Variational Autoencoder (VAE). Each data record contains three features: debit account name, credit account name, and amount.
Update History (Versions, File names, Dates)	First edition • Synthetic_JE_Basic • 2025/06/05
Data formats	csv
Data sizes	753KB
Things to be aware of when using the data (e.g., required metadata)	
Related reports and/or papers	Motai, R., Mashiko, S., Kawamata, Y., Shin, R., & Okada, Y. (2025). Generating Synthetic Journal-Entry Data Using Variational Autoencoder. <i>Intelligent Systems in Accounting, Finance and Management</i> , 32(1), e70005. https://doi.org/10.1002/isaf.70005
Other remarks	

Data Description

2024 年度プロジェクト : DB2024-18②

Data Number (will be filled by the Commons Center)	DPPSCdbp_2024-18②en
Title	Synthetic Journal Entry Data (Medical Domain / Clinics) with Temporal Realism Generated Using a Variational Autoencoder
Creator	Sota Mashiko, Ryoki Motai, Ryota Shin, Yuji Kawamata, Yukihiro Okada
Contact information	Yukihiro Okada okayu@sk.tsukuba.ac.jp
Summary (Subject and/or Methodology for generation)	This dataset consists of synthetic journal entry data generated from a Variational Autoencoder (VAE), trained on real journal entries provided by Tax Accountant Corporation Nihon Keiei, with prior consent obtained from individual client entities. The synthetic entries are generated using a trained Variational Autoencoder (VAE) in a way that preserves the original chronological order of dates. Each data record contains four features: date, debit account name, credit account name, and amount.
Update History (Versions, File names, Dates)	First edition • Synthetic_JE_DateAdded • 2025/06/05
Data formats	csv
Data sizes	956KB
Things to be aware of when using the data (e.g., required metadata)	
Related reports and/or papers	Motai, R., Mashiko, S., Kawamata, Y., Shin, R., & Okada, Y. (2025). Generating Synthetic Journal-Entry Data Using Variational Autoencoder. <i>Intelligent Systems in Accounting, Finance and Management</i> , 32(1), e70005. https://doi.org/10.1002/isaf.70005
Other remarks	Entries dated 13th month indicate that they are year-end adjustment entries.