

Minería de procesos e inteligencia artificial para el diseño, sostenibilidad y mejora de procesos colaborativos.

Martín Rubio

¹Instituto de Computación, Facultad de Ingeniería,
Universidad de la República, Uruguay

mrubio@fing.edu.uy

Abstract. *In recent decades, process mining has played a key role in improving organizational processes. Today's growing complexity, with collaborative processes involving multiple actors, demands new perspectives for design and improvement. At the same time, sustainability and artificial intelligence (AI) have gained relevance. This doctoral work aims to expand the scope of process mining and AI by integrating techniques, practices, and tools for the design, sustainability and improvement of collaborative processes.*

Resumen. *En las últimas décadas, la minería de procesos ha sido clave para mejorar procesos en las organizaciones. Hoy, la complejidad creciente, con procesos colaborativos que involucran múltiples actores, exige nuevas perspectivas de diseño y mejora. Al mismo tiempo, la sostenibilidad y la inteligencia artificial (IA) han cobrado protagonismo. Este trabajo doctoral busca ampliar el alcance de la minería de procesos y la IA, integrando técnicas, prácticas y herramientas para el diseño, sostenibilidad y mejora de procesos colaborativos.*

1. Introducción

La minería de procesos es una disciplina que emerge de la necesidad de analizar datos desde una perspectiva de procesos [van der Aalst 2016]. Para esto, se integran enfoques que provienen de la ciencia de datos y ciencia de procesos con el objetivo de analizar los datos generados a partir de la ejecución de procesos de negocio. Un proceso de negocio es un conjunto de actividades o tareas que son llevadas a cabo por las organizaciones en un entorno organizacional y técnico, para cumplir con sus objetivos [Weske 2019]. Por ejemplo, un proceso bancario en un banco o un proceso de salud en un hospital.

El principal insumo para la minería de procesos son los logs de eventos (event log) que se generan a partir de la ejecución de los distintos procesos. El enfoque tradicional abarca una perspectiva intra-organizacional, en donde existe una única organización o actor principal asociado a todo el proceso en estudio. Las realidades cada vez más complejas donde se deben integrar ecosistemas y tecnologías diversos, definen procesos colaborativos [van der Aalst 2011], con dependencias e interacciones entre distintos participantes en un único proceso general. Esta naturaleza distribuida, inmersa en diferentes contextos técnicos, complejiza la recolección y preparación de datos de los procesos.

Adicionalmente, surge el interés en disminuir el impacto de las actividades humanas en el medioambiente, donde los procesos están naturalmente involucrados. Para dimensionar y mejorar la sostenibilidad de procesos es clave poder medir su ejecución

con elementos de sostenibilidad, por lo que interesa definir para la minería de procesos una nueva dimensión de sostenibilidad a la hora de analizar los procesos. La sostenibilidad puede definirse como la habilidad de emplear cualquier recurso en el presente sin comprometer el uso de ese recurso para satisfacer las necesidades propias de las generaciones futuras, además asegurando el balance entre el crecimiento económico, bienestar social y medioambiental [Brundtland 1987].

En los últimos años la inteligencia artificial (IA) ha irrumpido en todas las áreas, incluyendo la minería de procesos, propiciando la integración de nuevos enfoques de diseño, automatización y análisis de datos, como aprendizaje automático (AA) para predicción de eventos futuros en los procesos con base en eventos existentes, o creación de nuevo contenido con modelos de lenguaje de gran tamaño (LLMs) como ChatGPT entrenados con una gran cantidad de datos, para asistir en diversas tareas, por ejemplo, el modelado de procesos o el análisis de trazas de ejecución de procesos.

En este contexto, es que esta tesis doctoral busca extender el alcance actual de la minería de procesos con inteligencia artificial, integrando nuevos artefactos como técnicas, prácticas, metodologías y herramientas para el diseño, la sostenibilidad y mejora de procesos colaborativos. El resto del artículo se organiza de la siguiente forma: en la sección 2 se plantean las preguntas y objetivos de investigación, en la sección 3 se presenta el estado del arte, luego en la sección 4 se muestra la metodología de investigación, la propuesta de solución y el avance actual se presenta en la sección 5 y finalmente en la sección 6 el trabajo pendiente.

2. Preguntas y objetivos de la investigación

El objetivo general es definir nuevas técnicas, prácticas, metodologías y herramientas de aplicación de minería de procesos e inteligencia artificial para el diseño, sostenibilidad y mejora de procesos colaborativos, a los efectos de reducir el esfuerzo de modelado, mejorar la ejecución y sostenibilidad de procesos colaborativos y reducir el gap existente entre modelos y su ejecución real. Las preguntas específicas son:

- **RQ1.** ¿Qué técnicas, prácticas, metodologías y herramientas basadas en minería de procesos son requeridas para el diseño, sostenibilidad y mejora de procesos colaborativos?
- **RQ2.** ¿Cuál es el alcance actual de las propuestas existentes y que limitaciones presentan?
- **RQ3.** ¿Cómo puede integrarse la inteligencia artificial para mejorar la aplicación de minería de procesos para el diseño, sostenibilidad y mejora de procesos colaborativos?

3. Estado del arte

La aplicación de minería de procesos (descubrimiento, conformidad, extensión, predicción) sobre los datos de ejecución de procesos de negocio [van der Aalst 2011] se han enfocado principalmente en procesos de tipo orquestación que se realizan en una única organización (intra-organizacionales), y más recientemente, a procesos colaborativos (inter-organizacionales) [Hernandez et al. 2021, Liu et al. 2023, Peña et al. 2024, Corradini et al. 2024], principalmente descubrimiento de procesos. Se ha definido una nueva propuesta de log de eventos orientado a objetos, Object Oriented Event Data

(OCED) [Fahland et al. 2024] como el sucesor al estándar tradicional XES (IEEE 1849-2016) [Group et al. 2016] para la representación de los logs de procesos, particularmente con una visión orientada a objetos. El log orientado a objetos permite ampliar las posibilidades de análisis de los procesos, en particular de procesos colaborativos.

Por otro lado, ha cobrado importancia de la orientación hacia procesos más sostenibles y la relevancia de enfoques como Green BPM. En [Couckuyt and Looy 2020] la revisión muestra que el foco principal ha sido medir y controlar las emisiones o la energía en los procesos de negocio, en la dimensión medioambiental. La evaluación de las TICs en [Hernández González et al. 2019] para gestionar recursos energéticos de forma eficiente controlando el impacto de las actividades que se realizan, y [Calero et al. 2021] la influencia del software de soporte en el consumo de energía, ej. plataformas BPMS, algoritmos de inteligencia artificial [Schwartz et al. 2020] y descubrimiento de procesos [Delgado et al. 2023]. La noción de “Ciencia de Datos Verde” (GDS) se introduce en [Van Der Aalst 2016] por la posible “contaminación” que puede causar la ciencia de datos.

Los avances en la inteligencia artificial han propiciado la integración de nuevos enfoques de automatización y análisis de datos, como aprendizaje automático (AA) para predicción de eventos futuros [Di Francescomarino and Ghidini 2022, Ceravolo et al. 2024], o creación de nuevo contenido con modelos de lenguaje de gran tamaño (LLMs) como ChatGPT para asistir en diversas tareas. En particular, se han utilizado LLMs para generación de descripciones textuales desde datos de procesos y logs de eventos, modelado de procesos desde su descripción textual, o uso de descripciones textuales de modelos y logs de eventos para minería de procesos [Grohs et al. 2024, Kourani et al. 2024, Berti et al. 2024b], así como evaluación de requisitos para uso de LLMs [Berti et al. 2024a] y herramientas asociadas.

El foco de este trabajo de tesis es en procesos colaborativos, en el cual se podrán utilizar de base propuestas y herramientas existentes para procesos tradicionales adaptándolas y/o extendiéndolas al contexto más complejo de ese tipo de procesos.

4. Metodología de investigación

Para guiar la investigación se siguen los lineamientos de Design Science para Sistemas de Información e Ingeniería de Software [Wieringa 2014, Hevner et al. 2004, Venable 2006]. En Design-Science para trabajos de investigación se atacan problemas no resueltos aún, utilizando enfoques únicos e innovadores, o problemas resueltos pero con enfoques más efectivos o eficientes, contribuyendo en ambos casos, al conocimiento en fundamentos (teorías, frameworks, construcciones, modelos, métodos, instancias) y metodologías (técnicas de análisis de datos, formalismos, medidas, criterios de validación). Se trabajará principalmente con el primer enfoque, en la creación de artefactos como framework general y metodología de trabajo, modelos (procesos, AA, LLMs e ingeniería del prompt), algoritmos de: minería de procesos, registro y manejo de datos, medidas de sostenibilidad, entre otros. A continuación se describen las principales etapas:

1. En primer lugar se define el problema identificando hipótesis y preguntas de investigación que guiarán el trabajo a realizar, que se presentaron en la Sección 2.
2. Realizar el estado del arte al inicio de la investigación sobre las temáticas seleccionadas siguiendo los lineamientos de revisiones sistemáticas de la literatura y map-

- ping studies planteados en [Kitchenham 2004, Kitchenham and Charters 2007].
3. Con base en el estado del arte realizado, se identifican propuestas que pueden servir de base para su aplicación en sostenibilidad de procesos colaborativos, de minería de procesos, aprendizaje automático e inteligencia artificial generativa.
 4. Para la validación de resultados se realizarán estudios de caso y/o experimentos controlados [Yin 2014, Wohlin et al. 2012]. Por ejemplo, experimentos controlados para validar la adecuación de las propuestas generadas con usuarios expertos (ej. medidas y análisis de sostenibilidad), estudios de caso en el contexto de procesos colaborativos de e-Government (AGESIC)¹.
 5. Comunicación de resultados. Serán validados en la comunidad científica con publicaciones en congresos y revistas de primer nivel regional e internacional.

5. Solución propuesta y avances

La solución propuesta busca materializar una visión integral para la mejora de procesos colaborativos, considerando la perspectiva de sostenibilidad mediante la minería de procesos y la inteligencia artificial como se muestra en la figura 1.

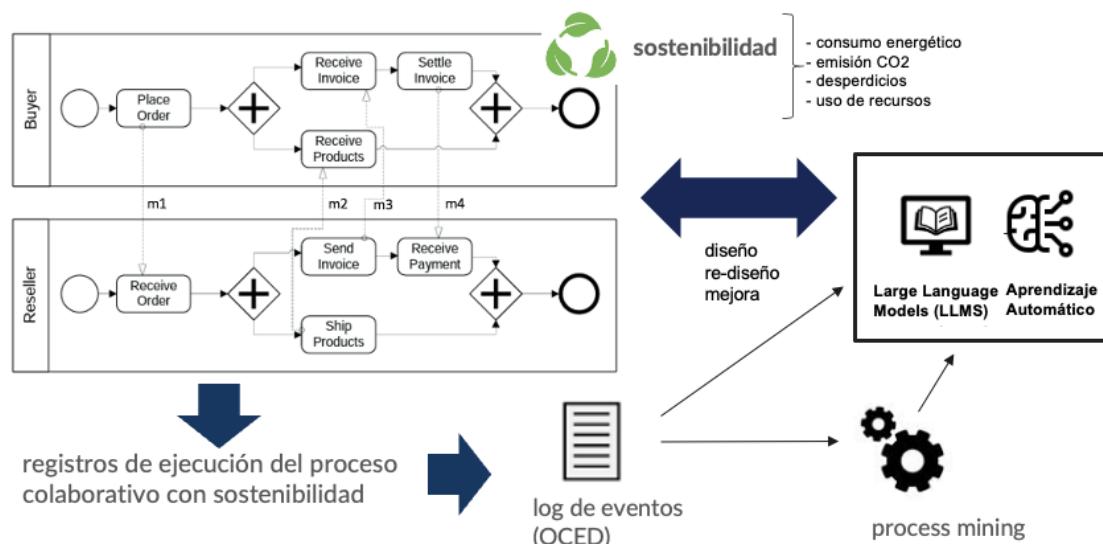


Figura 1. Visión general de la solución propuesta

La propuesta incorpora elementos de sostenibilidad en el registro de los logs de procesos, con base en los descritos en [Hernández González et al. 2019, Mancebo et al. 2017] y métricas asociadas, utilizando logs de eventos OCED. Tanto los logs de eventos como los modelos de procesos colaborativos, serán insumo para minería de procesos potenciada con inteligencia artificial (LLMs, aprendizaje automático) para apoyar el diseño, re-diseño y análisis de trazas de ejecución y modelos de procesos, para su sostenibilidad y mejora asociada. Por ejemplo, en la integración con LLMs se considerarán elementos como prompt engineering específicos [Grohs et al. 2024, Kourani et al. 2024, Berti et al. 2024a], generación aumentada por recuperación (RAG en inglés) para contexto específico, para procesos colaborativos. De esta manera, se espera mejorar los resultados obtenidos con minería de procesos tradicional.

¹<https://www.gub.uy/agencia-gobierno-electronico-sociedad-informacion-conocimiento/>

La propuesta incluirá la definición de técnicas, prácticas, metodologías y herramientas de soporte que guíen y promuevan el uso del enfoque definido como parte de las tareas de análisis y mejora de procesos en las organizaciones.

5.1. Revisión sistemática de la literatura

Actualmente se encuentra en curso una revisión sistemática para recolectar la información existente acerca de la minería de procesos aplicada con un enfoque de sostenibilidad. Para llevar a cabo esta revisión se siguen los lineamientos propuestos por Kitchenham [Kitchenham 2004][Kitchenham and Charters 2007]. Para la elaboración de la revisión se está utilizando en la plataforma Parsifal [Parsifal 2025] que brinda soporte a los lineamientos antes mencionados, lo cual ayuda a la ejecución, documentación y análisis de los resultados obtenidos en la revisión.

Partiendo de las preguntas de investigación definidas se genera la cadena de búsqueda que luego fue utilizada en cada una de las fuentes. Las fuentes se seleccionaron con base en el cubrimiento de publicaciones internacionales tanto de conferencias como revistas de primer nivel en el área, incluyendo Springer, IEEE Xplore, Scopus, Science Direct. Además, de los resultados obtenidos en las fuentes antes mencionadas también se consideró importante incluir artículos de conferencias específicas del área [Genero Bocco et al. 2023], particularmente los artículos presentados en el workshop Process Mining 4 Sustainability en 2024. [István Koren 2024]

A modo de resumen de los resultados preliminares se obtuvo un total de 434 artículos, de los que solamente 11 fueron considerados artículos relevantes. También se realizará snowballing [Wohlin 2014] partiendo de los estudios primarios para extender los resultados. Actualmente, se están procesando los 11 artículos relevantes que fueron clasificados con el objetivo de obtener los estudios primarios y de esta manera continuar con la realización de la revisión sistemática. Los estudios primarios serán analizados desde seis perspectivas, donde para cada una se cuenta con una pregunta que será calificada con base en la escala de Likert. Las perspectivas definidas son: técnicas de minería de procesos aplicada para sostenibilidad, medidas de sostenibilidad, tipo de log de eventos utilizado (XES/OCED), técnicas de inteligencia artificial utilizadas (si alguna, aprendizaje automático, IA generativa), datos de sostenibilidad definidos, tipo de procesos utilizados (orquestación, colaborativos).

5.2. Generación de logs OCED

Obtener logs de procesos, suele tener una complejidad considerable. Contar con el conjunto de datos de procesos de calidad es un problema recurrente tanto para el análisis de procesos como para tareas de investigación. Generar logs en forma sintética es una forma de atacar este inconveniente y de obtener datos nuevos y confiables para la experimentación utilizando la propuesta de integración de elementos de sostenibilidad.

Dentro de las técnicas de inteligencia artificial para entrenar redes profundas se encuentran las Redes Generativas Antagónicas (GANS) que han demostrado muy buenos resultados a la hora de generar datos sintéticos [Goodfellow et al. 2014]. Particularmente, las Conditional Tabular GANs (CTGANs)[Xu et al. 2019] son redes que se especializan en generar datos sintéticos a partir de datos tabulares. En ese sentido, se está trabajando en el entrenamiento de distintos modelos generativos utilizando CTGANs para la generación de logs de eventos de procesos sintéticos basados en OCED. Los primeros resultados

muestran que es posible generar log sintéticos de buena calidad utilizando estas tecnologías, lo cual será un insumo relevante para la experimentación con distintos datasets.

6. Trabajo pendiente

Un elemento clave es la identificación de elementos de sostenibilidad y la definición de prácticas basadas en integración de datos a los procesos. Particularmente, en mi tesis de maestría [Rubio et al. 2021] [Rubio et al. 2023] trabajé en la integración de datos organizacionales a los logs de procesos, que se trasladará a la nueva propuesta con OCED, utilizando el metamodelo y la implementación de [Calegari and Delgado 2023].

En este año, se trabajará en el análisis de las herramientas que brinda la inteligencia artificial, particularmente el aprendizaje automático y la inteligencia artificial generativa, para el diseño y modelado de procesos considerando su sostenibilidad, mejora de modelos existentes, análisis de sostenibilidad con base en trazas de ejecución reales, entre varias otras posibilidades. La integración de estos elementos a la minería de procesos es parte clave de la propuesta que se generará en este trabajo doctoral.

Se integrarán y definirán técnicas, prácticas, metodologías y herramientas de soporte adecuadas, incluyendo prompt engineering específicos, uso de generación aumentada por recuperación para contexto específico, diseño y re-diseño de procesos sostenibles, considerando y extendiendo/adaptando propuestas existentes, y definiendo nuevas para alcanzar los objetivos planteados.

A partir de los trabajos en curso y las definiciones propuestas, se prevé realizar experimentos y validar la propuesta con casos reales de procesos colaborativos del Estado uruguayo, utilizando datos de AGESIC.

Referencias

- Berti, A., Kourani, H., Häfke, H., Li, C.-Y., and Schuster, D. (2024a). Evaluating large language models in process mining: Capabilities, benchmarks, and evaluation strategies. In *Enterprise, Business-Process and Inf.Sys. Modeling*, pages 13–21. Springer.
- Berti, A., Schuster, D., and van der Aalst, W. M. P. (2024b). Abstractions, scenarios, and prompt definitions for process mining with llms: A case study. In *Business Process Management Workshops*, pages 427–439. Springer.
- Brundtland, G. (1987). Report of the world commission on environment and development: our common future. In *United Nations*.
- Calegari, D. and Delgado, A. (2023). A model-driven engineering perspective for the object-centric event data (oced) metamodel. In *Int. Conf. on Business Process Management*, pages 508–520. Springer.
- Calero, C., Moraga, M. Á., and Piattini, M. (2021). *Software Sustainability*, pages 1–15. Springer.
- Ceravolo, P., Comuzzi, M., De Weerdt, J., Di Francescomarino, C., and Maggi, F. M. (2024). Predictive process monitoring: concepts, challenges, and future research directions. *Process Science*, 1(1):2.
- Corradini, F., Pettinari, S., Re, B., Rossi, L., and Tiezzi, F. (2024). A technique for discovering bpmn collaboration diagrams. *Software and Systems Modeling*, 23:1323–1343.

- Couckuyt, D. and Looy, A. V. (2020). A systematic review of green business process management. *Business Process Management Journal*, 26(2):421–446.
- Delgado, A., García, F., Moraga, M. Á., Calegari, D., Gordillo, A., and Peña, L. (2023). Adding the sustainability dimension in process mining discovery algorithms evaluation. In *Business Process Management Forum*, pages 163–177. Springer.
- Di Francescomarino, C. and Ghidini, C. (2022). *Predictive Process Monitoring*, volume 448 of *Lecture Notes in Business Information Processing*, pages 320–346. Springer.
- Fahland, D., Montali, M., Lebherz, J., van der Aalst, W. M. P., van Asseldonk, M., Blank, P., Bosmans, L., Brenscheidt, M., Ciccio, C. D., Delgado, A., Calegari, D., Peeperkorn, J., Verbeek, E., Vugs, L., and Wynn, M. T. (2024). Towards a simple and extensible standard for object-centric event data (OCED) - core model, design space, and lessons learned. *CoRR*, abs/2410.14495.
- Genero Bocco, M., Piattini Velthuis, M. G., Cruz Lemus, J. A., and Díaz García, (2023). *Métodos de investigación en informática*. Ra-Ma.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, 27.
- Grohs, M., Abb, L., Elsayed, N., and Rehse, J.-R. (2024). Large language models can accomplish business process management tasks. In *Business Process Management Workshops*, pages 453–465. Springer.
- Group, X. W. et al. (2016). Ieee standard for extensible event stream (xes) for achieving interoperability in event logs and event streams. *IEEE Std*, 1849:1–50.
- Hernandez, J. D., Tello, E., Marin, H. M., Ramirez, U. M., and Mata, J. A. (2021). Merging event logs for inter-organizational process mining. In *New Perspectives on Enterprise Decision-Making Applying AI Techniques*, pages 3–26. Springer.
- Hernández González, A., Calero, C., Pérez Parra, D., and Mancebo, J. (2019). Approaching green BPM characterisation. *Journal of Software: Evolution and Process*, 31(2).
- Hevner, A. R., March, S. T., Park, J., and Ram, S. (2004). Design science in information systems research. *MIS Q.*, 28(1):75–105.
- István Koren, Janina Bauer, N. G. B. P. (2024). 1st international workshop on process mining for sustainability (pm4s). In *Process Mining Workshops: ICPM 2024 International Workshops, Denmark, Revised Selected Papers*.
- Kitchenham, B. (2004). Procedures for performing systematic reviews. Technical Report TR/SE-0401, Keele University and Empirical Soft. Eng. Joint Report, Keele, UK.
- Kitchenham, B. A. and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical Report EBSE 2007-001, Keele University and Durham University Joint Report, Keele, UK.
- Kourani, H., Berti, A., Hennrich, J., Kratsch, W., Weidlich, R., Li, C.-Y., Arslan, A., Schuster, D., and van der Aalst, W. M. P. (2024). Leveraging large language models for enhanced process model comprehension. *CoRR*, abs/2408.08892.

- Liu, C., Li, H., Zhang, S., Cheng, L., and Zeng, Q. (2023). Cross-department collaborative healthcare process model discovery from event logs. *IEEE Transactions on Automation Science and Engineering*, 20(3):2115–2125.
- Mancebo, J., García, F., Pedreira, O., and Moraga, M. Á. (2017). Bpms-game: Tool for business process gamification. Business Process Management Forum, pages 127–140. Springer.
- Parsifal (2025). Parsifal systematic review tool. Accessed: 2025-03-29.
- Peña, L., Andrade, D., Delgado, A., and Calegari, D. (2024). An approach for discovering inter-organizational collaborative business processes in BPMN 2.0. In *Process Mining Workshops*, pages 487–498. Springer Nature Switzerland.
- Rubio, M., Delgado, A., and Tansini, L. (2021). Process mining and data mining integration frameworks for evidence-based business intelligence: a systematic review. In *24th Iberoamerican Conf. on Software Engineering (CIbSE)*, pages 28–41. Curran Assoc.
- Rubio, M., Delgado, A., and Tansini, L. (2023). An integrated approach to process and organizational data mining. In *Latin American Computer Conf.*, pages 1–10. IEEE.
- Schwartz, R., Dodge, J., Smith, N. A., and Etzioni, O. (2020). Green ai. *Commun. ACM*, 63(12):54–63.
- Van Der Aalst, W. (2016). Green data science - using big data in an “environmentally friendly” manner. In *Proceedings of the 18th International Conference on Enterprise Information Systems - Volume 1: ICEIS*, pages 9–21. SciTePress.
- van der Aalst, W. (2016). *Process Mining: Data Science in Action*, 2nd.ed. Springer.
- van der Aalst, W. M. P. (2011). Intra- and inter-organizational process mining: Discovering processes within and between organizations. In *The Practice of Enterprise Modeling*, pages 1–11, Cham. Springer.
- Venable, J. (2006). A framework for design science research activities. In *Emerging Trends and Challenges in Information Technology Management*, pages 184–187. Idea group publishing.
- Weske, M. (2019). *Business Process Management - Concepts, Languages, Architectures, 3rd Ed.* Springer.
- Wieringa, R. J. (2014). *Design Science Methodology for Inf. Systems and Software Engineering*. Springer.
- Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th international conference on evaluation and assessment in software engineering*, pages 1–10.
- Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., and Regnell, B. (2012). *Experimentation in Software Engineering*. Springer.
- Xu, L., Skouliaridou, M., Cuesta-Infante, A., and Veeramachaneni, K. (2019). Modeling tabular data using conditional gan. *Advances in neural information processing systems*, 32.
- Yin, R. K. (2014). *Case Study Research: Design and Methods*, 5th Ed. SAGE Publications, Inc.