

# Temas

# Diversidade e Interdisciplinaridade

- Objetivo: Comparar a diversidade das universidades brasileiras e criar um ranking de diversidade
- Zhang, L., & Leydesdorff, L. (2021). The scientometric measurement of interdisciplinarity and diversity in the research portfolios of Chinese universities. *Journal of data and information science*, 6(4), 13-35.
- Dados para trabalhar:
  - e-Mec
  - Scopus
- Ferramentas:
  - ML, Redes, e outros

# Mobilidade (internacional de pesquisadores)

- Objetivo: Estudar a mobilidade de pesquisadores por meio das análises das bolsas de pesquisa de pós-doutorado CAPES e das FAP's
- Murray, D., Yoon, J., Kojaku, S., Costas, R., Jung, W. S., Milojević, S., & Ahn, Y. Y. (2023). Unsupervised embedding of trajectories captures the latent structure of scientific migration. *Proceedings of the National Academy of Sciences*, 120(52), e2305414120.
- Dados
  - Sucupira
  - Fapesp, Fapsc, Fapemg, Faprij, etc.

# Mapeamento da inovação em IA (academia – patente)

- Objetivo: Estudar a produção científica que gera inovação em IA no Brasil.
- Gong, H., Zou, H., Liang, X., Meng, S., Cai, P., Xu, X., & Qu, J. (2025). A Global Dataset Mapping the AI Innovation from Academic Research to Industrial Patents. *Scientific Data*, 12(1), 1261.
- Dados:
  - INPI,
  - LENS,
  - WIPO

# Uso de Redes Complexas para Análise de mobilidade de pesquisadores

- Objetivo: Criar uma rede complexa para analisar a mobilidade dos pesquisadores no Brasil.
- Wang, Y., Luo, H., & Shi, Y. (2019). Complex network analysis for international talent mobility based on bibliometrics. *International Journal of Innovation Science*, 11(3), 419-435.
- Dados:
  - Scopus
  - OpenAlex
  - SciELO

# Relação entre Patentes e Pesquisa científica

- Objetivo: Mensurar a produção científica brasileira e a sua relação com a produção ou não de patentes.
- Xu, S., Zhai, D., Wang, F., An, X., Pang, H., & Sun, Y. (2019). A novel method for topic linkages between scientific publications and patents. *Journal of the Association for Information Science and Technology*, 70(9), 1026-1042.
- Dados:
  - Lens,
  - PATSTAT
  - Scopus

# Transferência de tecnologias usando co-citação de patentes

- Objetivo: Estudar o modelo de citações entre patentes Brasileiras, para descobrir (dado um exemplo) quais tecnologias foram beneficiadas.
- Wang, X., Zhao, Y., Liu, R., & Zhang, J. (2013). Knowledge-transfer analysis based on co-citation clustering. *Scientometrics*, 97(3), 859-869.
- Dados
  - Lens,
  - Patstat

# Identificação de tecnologias emergentes a partir de dados de patentes usando ML

- Objetivo: Identificar nos dados de patentes brasileiras ou na escolha de alguma tecnologia, uma emergente.
- Choi, Y., Park, S., & Lee, S. (2021). Identifying emerging technologies to envision a future innovation ecosystem: A machine learning approach to patent data. *Scientometrics*, 126(7), 5431-5476.
- Dados
  - WIPO
  - LENS
  - PATSTAT



# Previsão de desempenho acadêmico usando ML

- Objetivo: Prever o desempenho acadêmico de alunos ou instituições brasileiras utilizando algoritmos de aprendizado de máquina
- Kuzilek, Jakub, Zdenek Zdrahal, and Viktor Fuglik. "Student success prediction using student exam behaviour." *Future Generation Computer Systems* 125 (2021): 661-671.
- Chen, Y., Zhai, L. A comparative study on student performance prediction using machine learning. *Educ Inf Technol* 28, 12039–12057 (2023). <https://doi.org/10.1007/s10639-023-11672-1>
- Fernandes, Eduardo, et al. "Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil." *Journal of business research* 94 (2019): 335-343.
- Fonseca, Stella Oggioni da, and Anderson Amendoeira Namen. "Data mining on INEP databases: An initial analysis aiming to improve brazilian educational system." *Educação em Revista* 32 (2016): 133-157.
- Dados
  - microdados do ENEM e do Censo Escolar disponibilizados pelo INEP

# Previsão da evasão de alunos usando ML

- Objetivo: Prever alunos com risco de evasão ou identificar os determinantes de evasão em instituições brasileiras utilizando algoritmos de aprendizado de máquina
- Ramos, Jorge Luis Cavalcanti, et al. "Um estudo comparativo de classificadores na previsão da evasão de alunos em EAD." *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*. Vol. 29. No. 1. 2018.
- de Jesus, Jeferson Andrade, and Renê Pereira de Gusmão. "Investigação da evasão estudantil por meio da mineração de dados e aprendizagem de máquina: Um mapeamento sistemático." *Revista Brasileira de Informática na Educação* 32 (2024): 807-841.
- Silva, A. R. da, Pereira, F. C., Mendonça, R. de S., Almeida, R. G. M. de, Barbosa, E. H. dos S., & Delgado, K. V. (2025). Uma Análise Quantitativa dos Determinantes da Evasão no Ensino Superior EaD. *EaD Em Foco*, 15(1), e2322.  
<https://doi.org/10.18264/eadf.v15i1.2322>
- Araújo, Ana Cléssia Pereira Lima de, Francisca Zilania Mariano, and Celina Santos de Oliveira. "Determinantes acadêmicos da retenção no Ensino Superior." *Ensaio: Avaliação e Políticas Públicas em Educação* 29.113 (2021): 1045-1066.
- Paula, Gustavo Bruno de, and Felícia Picanço. "Desigualdades após o acesso: origem social e evasão do sistema de ensino superior." *Educação & Sociedade* 45 (2024): e281915.
- Neves, Patrícia Takaki. "FICARE: um framework inteligente de coleta e análise de dados para mitigar a reprovação e a evasão na educação superior a distância." Tese UFSC. (2024).
- Dados
  - Censo da Educação Superior
  - Microdados do ENEM

# Avaliação da coesão textual em redações no contexto do ENEM

- Objetivo: Estimar notas relacionadas a coesão de redações no contexto do ENEM
- Rosa, Bruno Alexandre, et al. "Combinação de Modelos de Aprendizado de Máquina utilizando Teoria de Resposta ao Item para Avaliação de Coesão Textual em Redações no contexto do ENEM." Simpósio Brasileiro de Informática na Educação (SBIE). SBC, 2024.
- Marinho, J., Anchiêta, R., and Moura, R. (2022a). Essay-br: a brazilian corpus to automatic essay scoring task. Journal of Information and Data Management, 13(1).
- Oliveira, H., Mello, R. F., Miranda, P., Alexandre, B., Cordeiro, T., Bittencourt, I. I., and Isotani, S. (2023b). Classificação ou regressão? avaliando coesão textual em redações no contexto do ENEM. In Anais do XXXIV Simpósio Brasileiro de Informática na Educação, pages 1226–1237. SBC
- Dados
  - Essay-BR estendido, que compreende 6.563 redações no estilo do Exame Nacional do Ensino Médio (ENEM)

# Detecção de discrepâncias em dados financeiros de instituições de educação brasileiras

- Objetivo: Detecção de instituições que apresentam valores financeiros de despesas discrepantes (outliers). Para isso, são realizados agrupamentos visando à criação de grupos coesos e são aplicados métodos não supervisionados de detecção de outliers.
- Freitas, Nathan C., et al. "Detecção de Outliers em Finanças de Instituições de Ensino Superior Brasileiras Utilizando Aprendizado de Máquina Não Supervisionado." *Simpósio Brasileiro de Informática na Educação (SBIE)*. SBC, 2023.
- Dados
  - Censo da Educação Superior

# Análise do perfil dos estudantes e instituições de ensino superior e escolas

- Objetivo: analisar a infraestrutura das escolas, os perfis de estudantes e os perfis das Instituições de Ensino Superior.
- Carvalho, Jonathan, Lisandra Cruz, and Roberta Gouveia. "Descoberta de conhecimento com aprendizado de máquina supervisionado em dados abertos dos censos da educação básica e superior." *Anais dos workshops do congresso brasileiro de informática na educação*. Vol. 6. No. 1. 2017.
- Dados
  - Censo da Educação Superior e Censo de Educação Básica

# Predição do tempo de permanência na graduação em instituições brasileiras

- Objetivo: descobrir conhecimentos acerca do tempo de permanência de discentes em cursos de graduação de instituições de ensino superior públicas brasileiras.
- Rodrigues, Ebony M., and Roberta MM Gouveia. "Técnicas de machine learning para predição do tempo de permanência na graduação no âmbito do ensino superior público brasileiro." *Congresso sobre Tecnologias na Educação (Ctrl+ e)*. SBC, 2021.
- Dados
  - Exame Nacional de Desempenho dos Estudantes (ENADE) e do Censo da Educação Superior

# Análises de padrões na educação escolar em instituições brasileiras

- Objetivo: capturar relações complexas entre os dados, facilitando a extração de informações relevantes, bem como a geração de insights educacionais significativos utilizando Redes Neurais de Grafos (GNNs) como uma abordagem inovadora para enfrentar os desafios associados à alta dimensionalidade, à falta de padronização e às dificuldades de interpretação dos dados do Censo Escolar.
- Rodrigues, Bruno Cesar Silva. "Análise de padrões na educação via redes neurais e aprendizado de máquina." TCC UNESP. (2024).
- Dados
  - Censo Escolar do INEP

# Análises de padrões na educação superior em instituições brasileiras

- Objetivo: analisar as características de Instituições de Ensino Superior (IES) públicas brasileiras e à descoberta de conhecimento sobre os contextos de formação de discentes de cursos de graduação das IES observadas.
- Rodrigues, Ebony Márques, et al. "APRENDIZADO DE MAQUINA PARA AGRUPAMENTO E ASSOCIACAO DE DADOS DO ENSINO SUPERIOR PUBLICO BRASILEIRO." *Revista de Sistemas e Computação-RSC* 13.1 (2023).
- Dados
  - Censo da Educação Superior



# Orientação vocacional utilizando ferramentas de IA com dados do POSCOMP

- Objetivo: propor uma orientação vocacional inteligente baseado em clusterização a partir da análise exploratória dos resultados do POSCOMP.
- de Carvalho Costa, Jean Carlos, Lucas Mesquita Rodrigues Ferreira, and Reginaldo Cordeiro dos Santos Filho. "Orientação Vocacional Inteligente Baseado no Aprendizado de Máquina dos Microdados do POSCOMP." *iSys-Brazilian Journal of Information Systems* 17.1 (2024): 19-1.
- Dados
  - Microdados do POSCOMP

# Avaliação da capacidade das LLM's para resolver questões do POSCOMP

- Objetivo: avaliar a habilidade do ChatGPT e BARD para resolver questões do POSCOMP
- Saldanha, Mateus Santos, and Luciano Antonio Digiampietri. "ChatGPT and Bard Performance on the POSCOMP Exam." *Proceedings of the 20th Brazilian Symposium on Information Systems*. 2024.
- Dados
  - Microdados do POSCOMP