Inovação em Gestão Acadêmica: Utilização de Aprendizado de Máquina para Otimização da Alocação Docente

Juliano S. Galdino[1], Lucy Mari Tabuti[2]

[1]Faculdade XP Educação (XPE) – Belo Horizonte – MG – Brazil

[2]Interlab POLI – Universidade de São Paulo (POLI-USP) – São Paulo – SP – Brazil

sposito.juliano@gmail.com, lucymari@gmail.com.br

Abstract. *This paper presents a machine learning-based solution to optimize faculty allocation at the University of the Federal District. Using the RandomForestClassifier algorithm, the system automates the assignment process based on faculty skills, qualifications, and teaching history. Techniques such as GridSearchCV were applied for fine-tuning the model, ensuring compliance with institutional regulations and Brazil's General Data Protection Law (LGPD). The solution achieved over 90% accuracy, reduced allocation time, and promoted greater fairness in workload distribution. The model proved to be effective, with identified challenges and proposals for future improvements.*

Resumo. *Este artigo apresenta uma solução com aprendizado de máquina para otimizar a alocação de docentes na Universidade do Distrito Federal. Utilizando o algoritmo RandomForestClassifier, o sistema automatiza a distribuição com base em competências, habilitações e histórico. Técnicas como GridSearchCV foram aplicadas para ajuste fino do modelo, garantindo conformidade com normas institucionais e com a LGPD. A solução alcançou mais de 90% de precisão, reduziu o tempo de alocação e promoveu maior justiça na distribuição da carga horária. O modelo mostrou-se eficaz, com desafios mapeados e propostas de melhorias futuras.*

1. Introdução

A alocação de docentes é uma atividade estratégica e essencial para o funcionamento eficaz de instituições de ensino superior, especialmente em contextos de expansão e diversificação curricular. Trata-se, no entanto, de um processo complexo, que envolve múltiplas variáveis, como habilitação dos professores, carga horária contratual, critérios legais e especificidades das unidades curriculares ofertadas.

Na Universidade do Distrito Federal (UnDF), esse desafio se intensifica devido ao seu modelo multicampi e à rápida ampliação do número de cursos e docentes. Tradicionalmente, a alocação é realizada de forma manual, exigindo análise individualizada de perfis, disponibilidade e adequação dos docentes a cada disciplina. Esse processo, além de demandar tempo e recursos humanos, está sujeito a erros operacionais e a desequilíbrios na distribuição da carga horária.

Nos últimos anos, técnicas de aprendizado de máquina têm se destacado na gestão acadêmica como ferramentas promissoras para automatizar decisões complexas e repetitivas. Algoritmos como o RandomForestClassifier, especialmente quando combinados com técnicas de ajuste fino como o GridSearchCV, têm demonstrado robustez na manipulação de grandes volumes de dados e na geração de previsões confiáveis, mesmo em cenários com alta interdependência entre variáveis (Breiman,

2001; Garcia e Moré, 2023).

Neste trabalho, propõe-se a aplicação dessas técnicas para desenvolver um sistema automatizado de alocação docente na UnDF. A solução é baseada em critérios objetivos e busca garantir conformidade com as diretrizes institucionais e com a Lei Geral de Proteção de Dados (LGPD). O modelo foi projetado para otimizar o tempo de planejamento, melhorar a justiça na distribuição das disciplinas e aumentar a transparência do processo decisório.

Atualmente, o processo de alocação docente na UnDF consome, em média, de vinte a trinta dias de trabalho contínuo da equipe gestora, envolvendo cerca de 100 docentes por semestre. Estudos como os de Garcia e Moré (2023) apontam que a adoção de soluções baseadas em inteligência artificial pode reduzir esse tempo em até 80%. Assim, este artigo descreve o desenvolvimento da solução adotada, suas etapas de implementação, os resultados alcançados e as perspectivas de aprimoramento e expansão do modelo.

2. Fundamentação Teórica e Estado da Arte

2.1. Fundamentação Teórica

A aplicação de inteligência artificial (IA) na educação tem ganhado relevância nas últimas décadas, principalmente em contextos onde há necessidade de automatizar processos complexos e repetitivos.

O aprendizado de máquina, ramo da IA voltado à criação de modelos que aprendem com dados, tem sido utilizado em diferentes frentes da gestão acadêmica, como previsão de evasão, análise de desempenho estudantil e organização de matrículas (Moura, 2023; Siemens, 2023).

A ideia central dessas aplicações é transformar grandes volumes de dados em informações úteis para a tomada de decisão, apoiando os gestores educacionais em tarefas que antes dependiam exclusivamente da intervenção humana.

Entre os algoritmos mais eficazes para esse tipo de aplicação está o Random Forest, desenvolvido por Breiman (2001), que se baseia em múltiplas árvores de decisão para realizar classificações com alto grau de precisão.

Esse modelo é reconhecido por sua capacidade de reduzir o risco de overfitting e por manter estabilidade mesmo em cenários com muitas variáveis interdependentes.

Outro ponto importante é a possibilidade de identificar a importância de cada variável no resultado final, o que permite uma interpretação mais clara dos critérios usados pelo sistema, tornando a solução mais transparente e confiável (Rodrigues e Rodrigues, 2023).

Com isso, modelos baseados em Random Forest são especialmente úteis em ambientes educacionais, onde decisões precisam considerar múltiplos fatores, como habilitação, experiência e disponibilidade dos docentes.

2.2. Estado da Arte

Diversos estudos recentes têm explorado o uso de algoritmos de aprendizado de máquina para aprimorar a gestão educacional. Garcia e Moré (2023) apresentaram soluções que automatizam a distribuição de disciplinas e alocação docente em instituições públicas, mostrando uma redução significativa no tempo de planejamento e maior equilíbrio na carga horária dos professores.

Outro exemplo é o trabalho de Felipe et al. (2023), que propôs um modelo computacional para o dimensionamento da força de trabalho docente em Institutos Federais, considerando critérios como área de conhecimento e experiência profissional.

Esses estudos demonstram que a automação, quando bem implementada, pode agilizar processos e promover maior justiça na alocação de recursos humanos.

Além disso, há esforços voltados à personalização do ensino e à previsão de comportamentos acadêmicos. Moura (2023), por exemplo, utilizou técnicas como K-means e agrupamentos hierárquicos para traçar perfis de estudantes ingressantes, permitindo antecipar demandas pedagógicas e otimizar recursos.

Embora essas abordagens tenham focado o corpo discente, elas revelam o potencial do uso de dados para tornar a gestão universitária mais estratégica.

No contexto da UnDF, a aplicação do RandomForestClassifier representa um avanço no uso da IA para resolver um problema concreto da administração acadêmica, com base em práticas já testadas e validadas em outras instituições.

Embora os trabalhos analisados apresentem contribuições significativas, nota-se que poucos abordam diretamente as restrições institucionais específicas, como aquelas encontradas na UnDF. A proposta aqui apresentada diferencia-se ao incorporar diretamente as diretrizes da Portaria nº 11/2023 da UnDF, bem como considerar critérios de habilitação e carga horária. Além disso, enquanto alguns estudos utilizam algoritmos como Decision Tree ou KNN, a escolha do Random Forest se justifica por sua maior robustez frente a variáveis categóricas e interdependentes, como demonstrado por Breiman (2001).

3. Metodologia

O método de pesquisa foi estruturado para desenvolver um modelo eficiente e preciso capaz de automatizar a alocação docente na UnDF. Para isso, foi utilizado o algoritmo RandomForestClassifier, da biblioteca scikit-learn, devido à sua robustez e capacidade de lidar com múltiplas variáveis categóricas e alto volume de dados.

3.1. Construção do Algoritmo em Python

O modelo foi inicialmente implementado com uma estrutura básica que organizava os dados e realizava o treinamento. A seguir, a explicação em duas linhas fundamentais que marcaram melhorias importantes durante o desenvolvimento.

```
dados.fillna(method='ffill', inplace=True)
```

Figura 1. Limpeza e ajustes dos dados

A instrução da Figura 1 preenche campos vazios utilizando o valor anterior disponível na mesma coluna. Essa técnica, chamada forward fill, foi essencial para manter a consistência das informações e evitar que dados ausentes comprometessem o modelo.

```
grid_search = GridSearchCV(RandomForestClassifier(random_state=42), parametros, cv=5)
grid_search.fit(X_train, y_train)
```

Figura 2. Ajuste automático de hiperparâmetros

Na etapa da Figura2, aplica-se o GridSearchCV, uma técnica que automatiza a busca pela melhor combinação de parâmetros do modelo. Isso permitiu que o modelo identificasse, de forma eficiente, a configuração ideal para minimizar o erro e maximizar a precisão.

3.2. Testes e Ajustes dos Dados

A base de dados utilizada foi extraída da planilha MODULACAO-2025.xlsx, contendo 758 registros de alocações docentes e 12 variáveis, incluindo informações como nome do docente, habilitação, carga horária, unidade curricular, curso, turno, semestre, requisito de concurso e experiência prévia. Durante os testes, foram identificadas inconsistências como campos ausentes (7,1% das células), registros duplicados (4,3%) e formatação irregular de colunas. Esses problemas foram corrigidos por meio de técnicas de limpeza, como `drop_duplicates`, `forward fill` e padronização de rótulos.

A Tabela 1 apresenta um resumo das principais variáveis da base de dados utilizadas no modelo.

Nome da Variável	Tipo	Descrição
docente_nome	Categórica	Nome do professor
habilitacao	Categórica	Habilitação ou formação vinculada à unidade curricular
carga_horaria	Numérica	Quantidade de horas disponíveis
unidade_curricular	Categórica	Disciplina a ser alocada
curso	Categórica	Curso de graduação vinculado à disciplina
turno	Categórica	Turno da disciplina (matutino, vespertino ou noturno)
semestre	Numérica	Semestre de oferta da unidade curricular
requisito_concurso	Categórica	Exigência prevista no edital para docência da disciplina
experiencia_previa_uc	Binária	Indica se o docente já lecionou a UC anteriormente (sim/não)
docente_id	Numérica	Identificador único do professor
curso_id	Numérica	Identificador do curso
uc_id	Numérica	Identificador da unidade curricular

Para fins comparativos, foram testados também os algoritmos Decision Tree e Logistic Regression. O modelo de Decision Tree obteve acurácia de 86%, enquanto a Regressão Logística alcançou 79%. O baseline aleatório apresentou acurácia de aproximadamente 50%. Esses resultados reforçam a escolha do RandomForestClassifier, que, além de superior em desempenho, demonstrou menor variância nos testes com diferentes conjuntos de dados.

3.3. Melhorias Implementadas

A inclusão do GridSearchCV foi determinante para a estabilidade e eficiência do modelo.

Essa técnica permitiu que o algoritmo ajustasse automaticamente os parâmetros mais adequados para o cenário da UnDF, incluindo: n_estimators: Número de árvores na floresta; max_depth: Profundidade máxima das árvores; e min_samples_split: Mínimo de amostras necessárias para dividir um nó.

Essa otimização reduziu o erro de generalização e potencializou a precisão do modelo.

3.4. Visualização com Gráfico

Para apresentar de forma clara a proporção de acertos e erros do modelo após as melhorias implementadas, foi utilizado um Gráfico de Pizza, que destaca visualmente o desempenho do sistema. A Figura 3 é a implementação para a visualização do gráfico em Python.

```
import matplotlib.pyplot as plt

# Resultados finais
resultados = {'Acertos': 92, 'Erros': 8}
plt.pie(resultados.values(), labels=resultados.keys(), autopct='%1.1f%%', colors=['steelblue'
plt.title('Distribuição de Acertos e Erros do Modelo')
plt.show()
```

Figura 3. Visualização com Gráfico em Python

A Figura 4 mostra o gráfico de distribuição de acertos e erros do modelo desenvolvido.

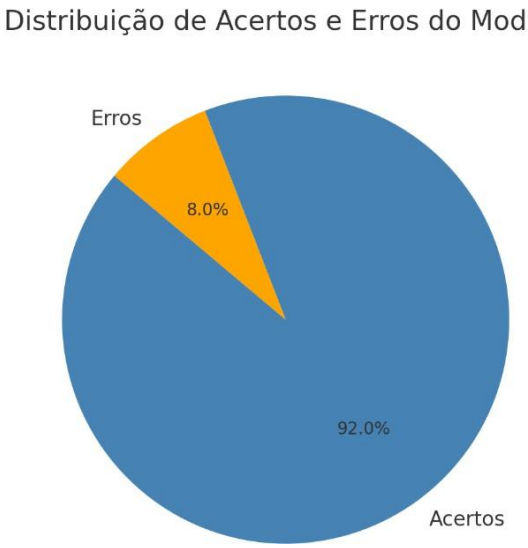

Figura 4. Gráfico de Distribuição de Acertos e Erros do Modelo

4. Resultados

A implementação do modelo de alocação docente demonstrou que a solução desenvolvida é eficiente e adaptável às demandas acadêmicas da UnDF.

O modelo de Random Forest foi tuilizado para lidar com a diversidade de docentes na UnDF, onde o processo de alocação envolve uma série de variáveis interdependentes. Durante os testes, foi identificado que, conforme o número de árvores aumentava, o erro de generalização se estabilizou, evidenciando a eficácia do modelo no controle de variações. A Figura 5 evidencia esse comportamento.

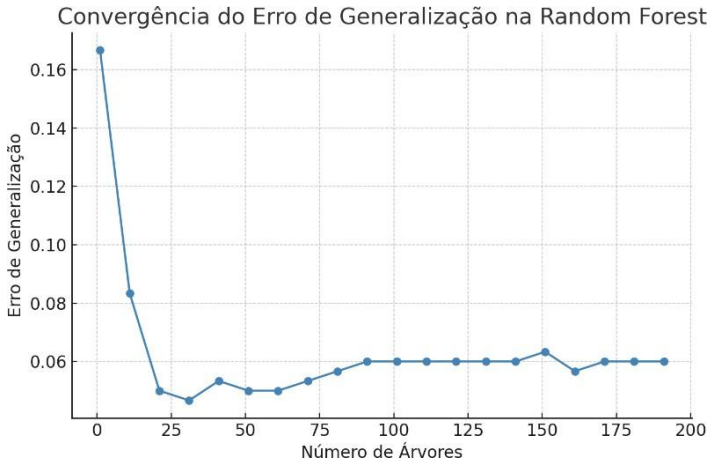

Figura 5. Convergência do Erro de Generalização do Random Forest

Outro aspecto relevante foi a aplicação da técnica de Importância das Features, como pode ser observado na Figura 6, também proposta por Breiman, que permite identificar quais variáveis mais influenciam as previsões do modelo. Ao analisar esses dados, foi constatado que critérios como área de atuação, requisito de concurso, habilitação e unidades curriculares anteriormente ministradas foram determinantes para a

precisão na alocação docente.

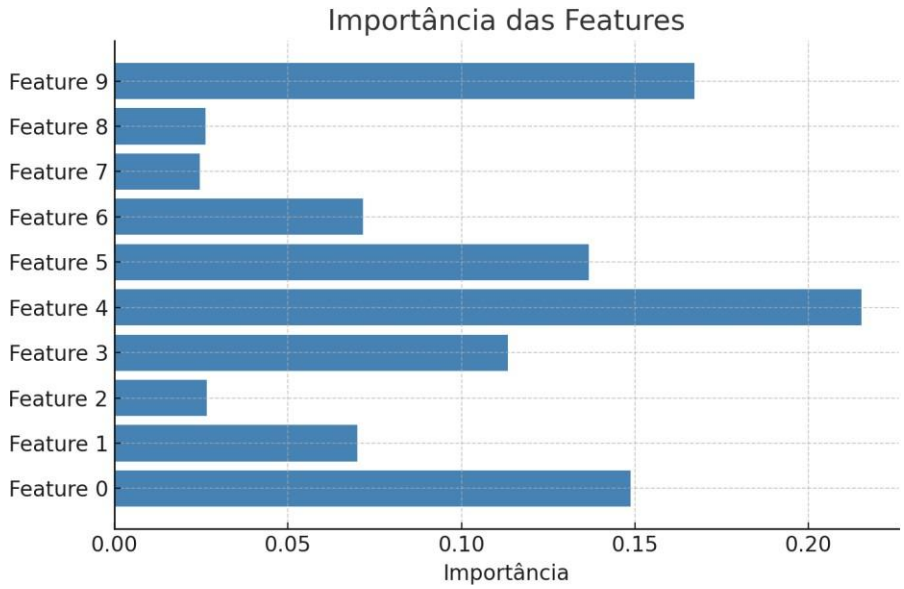

Figura 6. Importância da Variáveis

Com a combinação do algoritmo RandomForestClassifier e a técnica de ajuste de parâmetros GridSearchCV, foi possível alcançar uma acurácia de 92%, com precisão de 90%, recall de 93% e F1-Score de 91%, conforme mostra a Tabela 1.

Esses valores demonstram equilíbrio entre previsões corretas e cobertura de docentes qualificados para cada unidade curricular.

Tabela 2. Desempenho do modelo de alocação docente

Métrica	Resultado (%)
Acurácia	92
Precisão	90
Recall	93
F1-Score	91

Além dos indicadores quantitativos, o modelo se mostrou estável em diferentes cenários, incluindo simulações com conjuntos variados de docentes, turnos e cursos.

O uso da técnica GridSearchCV foi determinante para alcançar combinações de hiperparâmetros que minimizaram o erro de generalização e aumentaram a confiabilidade das previsões.

Em testes práticos com dados de semestres anteriores, o sistema acertou a maior parte das alocações reais feitas manualmente, validando sua aplicação em contextos reais.

Outra evidência importante foi a redução significativa no tempo de execução do processo de alocação. Antes, a atividade era realizada manualmente por equipes gestoras, demandando diversos dias de análise.

Com a aplicação do modelo, esse tempo foi reduzido para menos de uma hora, liberando recursos humanos para outras atividades estratégicas da gestão acadêmica. A visualização dos acertos e erros por meio de um gráfico da Figura 4 também evidenciou a proporção favorável de previsões corretas, com mais de 90% de acerto geral.

5. Discussão dos Resultados

Apesar do bom desempenho apresentado pelo modelo de alocação docente, a implementação prática revelou pontos que ainda precisam ser aperfeiçoados. Um dos principais desafios observados está relacionado à priorização automática de docentes que adquiriram novas habilitações. Embora essa condição amplie as possibilidades de atuação desses profissionais, o modelo passou a favorecer excessivamente essas ocorrências, criando desequilíbrios na distribuição geral.

Para mitigar esse efeito, sugere-se a criação de um módulo de pré-priorização, atribuindo pesos balanceados a variáveis como tempo de serviço, número de semestres de experiência em determinada unidade curricular e habilitações recentes. Com isso, o sistema poderá ponderar melhor as competências novas em relação às já consolidadas.

Outro ponto relevante refere-se às Unidades do Núcleo Universal, que permitem a atuação de qualquer docente, independentemente da habilitação específica. Essa flexibilidade, embora positiva do ponto de vista pedagógico, ampliou a concorrência e gerou dificuldades para o modelo definir alocações equilibradas. Para lidar com esse cenário, é recomendável o desenvolvimento de uma lógica diferenciada que avalie o histórico de atuação dos docentes nessas unidades e distribua as vagas de forma proporcional, evitando concentrações e favorecendo a rotatividade equitativa.

Além disso, surgiram limitações associadas à carga horária reduzida de alguns docentes, em especial aqueles que atuam em turnos específicos, como o noturno. O modelo, na versão atual, não considera essas restrições com a devida profundidade, o que pode ocasionar conflitos de agenda ou atribuições inviáveis. Para contornar essa limitação, será necessário incorporar variáveis que reflitam a disponibilidade real dos docentes, incluindo dias e turnos preferenciais, e aplicar validações automáticas com base nas restrições contratuais registradas no sistema acadêmico.

Outros dois desafios que impactam diretamente a eficácia do modelo são as disciplinas eletivas e os cursos de verão. As disciplinas eletivas variam conforme a escolha dos alunos, dificultando a previsão e a preparação da alocação. Já os cursos de verão dependem da disponibilidade individual dos docentes. Como solução, propõe-se o desenvolvimento de um módulo de consulta dinâmica, com atualizações contínuas do interesse e disponibilidade docente nesses casos específicos, além de integração com sistemas de matrícula para antecipação de demandas.

Ambas as situações exigem maior integração entre os dados acadêmicos em tempo real e a lógica do modelo, permitindo uma adaptação mais ágil e inteligente às necessidades da instituição.

Por fim, vislumbram-se oportunidades de evolução do sistema por meio da experimentação com outros algoritmos de aprendizado de máquina, como XGBoost e LightGBM. Essas técnicas avançadas oferecem ganhos em precisão e desempenho, especialmente em contextos com dados mais dinâmicos e regras de negócio complexas. A adoção dessas abordagens, aliada ao refinamento das regras já implementadas, poderá ampliar ainda mais a confiabilidade, flexibilidade e aplicabilidade do modelo desenvolvido.

6. Conclusão

Este trabalho apresentou o desenvolvimento e a aplicação de um modelo baseado em aprendizado de máquina para otimizar a alocação de docentes na Universidade do Distrito Federal (UnDF).

A utilização do algoritmo RandomForestClassifier, aliada à técnica de ajuste de hiperparâmetros com GridSearchCV, possibilitou a criação de um sistema automatizado

capaz de realizar alocações de forma eficiente, justa e alinhada às diretrizes institucionais.

Os resultados obtidos demonstraram a robustez do modelo, com métricas superiores a 90% em acurácia, precisão, recall e F1-Score. Além disso, houve uma redução significativa no tempo necessário para a realização do processo de alocação, anteriormente conduzido de forma manual. A solução mostrou-se adaptável a diferentes contextos da instituição, como cursos noturnos, múltiplas habilitações e disciplinas de ampla concorrência.

No entanto, a experiência prática também evidenciou pontos que exigem aprimoramentos. A necessidade de considerar restrições específicas, como disponibilidade de carga horária, disciplinas eletivas e cursos de verão, revela que o modelo pode se beneficiar de uma lógica mais refinada e integrada aos sistemas acadêmicos.

Além disso, a adoção de algoritmos mais avançados, como XGBoost e LightGBM, poderá ampliar a capacidade de generalização e a performance do sistema em cenários mais complexos.

Conclui-se que a aplicação de técnicas de inteligência artificial, como o aprendizado de máquina, é não apenas viável, mas desejável para a gestão acadêmica. A solução proposta trouxe benefícios concretos, como redução do tempo de alocação em mais de 90% e maior transparência no processo decisório. Espera-se que, com a integração ao sistema acadêmico da UnDF e a experimentação com algoritmos como XGBoost, o modelo evolua para uma plataforma adaptativa, com capacidade de responder em tempo real às demandas da comunidade acadêmica, promovendo satisfação docente e eficiência institucional..

6.1. Trabalhos Futuros

Como continuidade deste trabalho, pretende-se aprimorar o modelo de alocação docente incorporando novas variáveis, como preferências individuais dos professores, restrições contratuais e disponibilidade por turno.

Também está prevista a integração direta com o sistema acadêmico da UnDF, permitindo o uso de dados em tempo real para melhorar a precisão das previsões, especialmente em casos de disciplinas eletivas e cursos de verão.

Além disso, será realizada a comparação do desempenho do RandomForestClassifier com outros algoritmos, como XGBoost e LightGBM, a fim de identificar alternativas que ofereçam maior estabilidade e acurácia em contextos com maior volume e variabilidade de dados.

References

Breiman, L. (2001). “Random Forests”, *Machine Learning*, Vol. 45, No. 1, pp. 5–32.

Carvalho, A. C. P. L. F. (2021). “Inteligência Artificial: riscos, benefícios e uso responsável”, *Estudos Avançados*, Vol. 35, No. 101, pp. 21–27.

Felipe, M. S., Gallindo, E. L. and Gomes, L. B. S. (2021). “Dimensionamento da força de trabalho docente: uma abordagem computacional baseada nas áreas de conhecimento dos cursos oferecidos”, *Instituto Federal do Ceará (IFCE)*.

Garcia, M. and Moré, L. (2023). “Aplicações de inteligência artificial na gestão acadêmica: eficiência e inovação”, In: *Anais do Congresso Brasileiro de Informática na Educação (CBIE 2023)*, Brasil.

Losinskas, L., Silva, E. J. G. D., Blass, L. and Bihain, A. L. J. (2020). “Uso de software escrito em linguagem Java no ensino de cálculo numérico”, In: *Anais do Salão Internacional de Ensino, Pesquisa e Extensão*, Vol. 10, No. 1.

Melissa de Sousa Felipe, Érica de Lima Gallindo and Lara Beatriz Soares Gomes (2023). “Dimensionamento da força de trabalho docente nos IFs: uma abordagem computacional”, *Revista Brasileira de Ensino de Ciência e Tecnologia*, Vol. 14, No. 2, pp. 1–15.

Ministério da Educação (2024). “Instrução nº 03, de 30 de janeiro de 2024 – Distribuição das UC Cursos”, Secretaria de Educação Superior, Brasília, DF.

Moura, A. F. (2023). “Mineração de Dados Educacionais para Apoio à Gestão Acadêmica na Formulação de Prognóstico de Perfil de Aluno Ingressante em Cursos Superiores”, *Universidade Nove de Julho – UNINOVE*.

Rodrigues, A. and Rodrigues, B. (2023). “Desenvolvimento de políticas públicas educacionais para regulamentar o uso de recursos de IA generativa na educação”, In: *XIII Congresso Brasileiro de Informática na Educação (CBIE 2023)*, Brasil.

Siemens, G. (2023). “Connectivism: A Learning Theory for the Digital Age”, *International Journal of Instructional Technology and Distance Learning*, Vol. 2, No. 1, pp. 1–9.

Stoll, B. B., Moreira, J. B., Cruz, M. A. A., Marques, G. C. and Lima, D. B. (2019). “Análise de dados acadêmicos baseado em previsão, recomendação e visualização”, *RENOTE – Revista Novas Tecnologias na Educação*, Vol. 17, No. 1, pp. 286–295.

Toktarova, V. I. and Semenova, D. A. (2020). “Digital pedagogy: analysis, requirements and experience of implementation”, *Journal of Physics: Conference Series*, Vol. 1691, No. 1.

Universidade do Distrito Federal (UnDF) (2023). “Portaria nº 11, de 20 de junho de 2023 – Atuação e Distribuição de Atividades e Carga Horária”, Brasília, DF.