



11ª Jornada Científica e Tecnológica do IFSULDEMINAS & 8º Simpósio de Pós-Graduação

PREDIÇÃO DE EVASÃO ESCOLAR DE GRADUANDOS UTILIZANDO REDES NEURAIS ARTIFICIAIS

Amanda B.A. da SILVA¹; Aline M. D. Valle²; Gisselida do P. SIQUEIRA³

RESUMO

Ingressar em uma universidade pública no Brasil é a concretização de um sonho para muitos estudantes, a trajetória até a aprovação envolve muito estudo, disciplina, concorrência e investimento financeiro. Após a aprovação, o discente depara-se com outros fatores desafiadores que deverão ser enfrentados até a formatura, a fim de concluir a graduação com êxito. Diante desse desafio, muitos alunos sentem-se desmotivados para concluir o curso e acabam abandonando os estudos, gerando um número alarmante no índice de evasão no ensino superior brasileiro. As instituições de ensino armazenam diversos dados sobre a vida acadêmica desses alunos. Indicadores importantes relacionados à evasão podem ser obtidos a partir destes dados, gerando informações para que ações efetivas para a retenção de alunos sejam feitas. Assim, este artigo apresenta um estudo utilizando dados acadêmicos de alunos de Ciência da Computação do IFSULDEMINAS Campus Muzambinho. Os resultados demonstram que utilizando as primeiras notas semestrais dos discentes é possível identificar com precisão de 88% a situação final do aluno no curso.

Palavras-chave: Mineração de Dados; KDD; Inteligência Artificial; Ciência da Computação.

1. INTRODUÇÃO

A disseminação do uso de tecnologias da informação proporcionou inúmeros benefícios para a sociedade, facilitando a vida das pessoas e organizações por meio da internet e demais ferramentas. Cada vez mais as empresas e outras entidades vêm utilizando desses recursos, que auxiliam no controle de seus produtos, funcionários e clientes, melhorando assim sua produtividade e eficiência.

Dessa maneira, o número de dados gerados aumentam, assim como a concorrência entre as entidades. Surge então o interesse de analisar esses dados a fim de buscar avanços dentro das próprias organizações e melhorar a competitividade no mercado (GOLDSCHMIDT; PASSOS, 2005).

Esses fatores influenciaram na elaboração de uma técnica de Descoberta de Conhecimento em Base de Dados (*Knowledge Discovery in Databases - KDD*), que consiste na busca, análise e interpretação de padrões retirados de uma base de dados.

¹ Aluna, IFSULDEMINAS - Campus Muzambinho. E-mail: 12121002076@muz.ifsuldeminas.edu.br

² Orientadora, IFSULDEMINAS - Campus Muzambinho. E-mail: aline.valle@muz.ifsuldeminas.edu.br

³ Coorientadora, IFSULDEMINAS - Campus Muzambinho. E-mail: gisselida.siqueira@muz.ifsuldeminas.edu.br

Uma das etapas principais dentro do KDD é a Mineração de Dados (*Data Mining*), que tem sua origem em áreas como Estatística, Inteligência Artificial e Banco de Dados. O objetivo da técnica é explorar os dados já coletados, aplicando algoritmos, em busca de padrões e anomalias, a fim de obter ou prever resultados. As informações obtidas auxiliam na tomada de decisões e planejamento.

Ainda que a *Data Mining* seja frequentemente mais aplicada no contexto de negócios e finanças, a técnica também está sendo amplamente utilizada para resolver problemas na área educacional. Anualmente, o Ministério da Educação (MEC) e o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) realizam um levantamento estatístico a respeito da oferta de cursos de graduação e sequenciais de formação específica, perfil de alunos e docentes, e realidade das Instituições de Educação Superior. De acordo com esses dados, o número de ofertas de vagas no ensino superior tem aumentado gradativamente a cada ano, porém, o número de vagas ocupadas não é proporcional, e o índice de evasão é preocupante.

Para encontrar respostas por meio desses dados e compreender as causas de evasão são utilizadas técnicas de Mineração de Dados Educacionais (*Educational Data Mining – EDM*), que tem como principal objetivo apoiar determinadas práticas pedagógicas e gerar melhorias para os sistemas de ensino. Segundo Costa (2013), “a Mineração de Dados Educacionais procura desenvolver ou adaptar métodos e algoritmos de mineração existentes, de tal modo que se prestem a compreender melhor os dados em contextos educacionais”. Assim, por meio de técnicas específicas aplicadas em base de dados, é possível compreender de forma mais eficaz e adequada os alunos, como sua forma de aprendizagem, além de outros fatores importantes para melhoria dos sistemas de ensino.

O Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais, Câmpus Muzambinho, mantém dados de seus alunos do curso de Ciência da Computação desde 2009. Esses dados são referentes ao histórico do discente na instituição, e ao “Questionário Socioeconômico e Cultural”, que é aplicado no momento da inscrição para o vestibular, com o objetivo de obter informações a respeito do candidato. Com a mineração desses dados é possível definir o perfil de estudantes propensos a não conclusão do curso, bem como descobrir quais são os fatores que influenciam a evasão dos mesmos na instituição, contribuindo para que ações efetivas sejam feitas, visando a permanência dos alunos e diminuindo assim, o alto índice de evasão do curso.

Assim, o objetivo deste trabalho é prever o risco de evasão escolar dos alunos do curso de Bacharelado em Ciência da Computação do IFSULDEMINAS, Câmpus Muzambinho, por meio de

de técnicas de Mineração de Dados Educacionais, utilizando o algoritmo de redes neurais artificiais do tipo Multilayer Perceptron.

2. MATERIAL E MÉTODOS

A princípio foi realizada uma Revisão Sistemática de Literatura (RSL) de trabalhos correlatos, com o intuito de estabelecer quais dados dos discentes seriam utilizados, assim como as técnicas de Mineração de Dados, linguagem ou ferramenta para aplicação dos algoritmos.

Após a leitura e tabulação dos resultados dos 23 trabalhos⁴ estudados, foi definido que seriam utilizados dados do histórico acadêmico dos discentes e dados socioeconômicos. Foi escolhido o algoritmo de Redes Neurais Artificiais do tipo *Multilayer Perceptron (MLP)* para aplicação e a linguagem *Python* juntamente com suas bibliotecas de *Machine Learning*, *Sklearn* e *Pandas*, foi escolhida para codificação.

Os dados utilizados no trabalho foram obtidos do sistema acadêmico. Foi realizado o pré-processamento dos dados, no qual os mesmos foram preparados para serem submetidos ao algoritmo.

3. RESULTADOS E DISCUSSÕES

Inicialmente, a base de dados era composta por duas tabelas, a primeira referente a dados socioeconômicos dos alunos, com 403 registros e 19 atributos. A segunda tabela continham dados do histórico acadêmico dos discentes, composta por 11249 registros referente à todas disciplinas cursadas por todos alunos que já passaram pelo curso de Ciência da Computação, bem como suas características como Matriz Curricular, período e ano de curso, percentual de faltas, nota e situação final da disciplina.

O pré-processamento iniciou-se com a junção das tabelas de dados socioeconômicos e histórico acadêmico, no qual foram considerados apenas dados de disciplinas referentes ao primeiro período do curso, como Algoritmos e Estrutura de Dados I (AED1), Física, Cálculo Diferencial e Integral I e Lógica Matemática. Na fase de limpeza dos dados, foram removidos 103 registros referentes a alunos regulares em 2019, que ainda estão em curso, e 27 registros incompletos, que não continham todas as informações necessárias para análise.

A base de dados foi submetida ao algoritmo de cálculo de correlação, a fim de detectar a relação estatística entre as variáveis. Instâncias com grau inferior a 0.3 (negativo ou positivo) foram removidas da base, pois demonstram correlação fraca. Para obter um bom desempenho dos

⁴A tabela com todos os artigos estudados está disponível online em: <https://bit.ly/2xv3pOD>

algoritmos, variáveis categóricas, como auxílio, resultados das disciplinas e status final do discente, foram transformadas em categorias numéricas.

Ao final, a base foi composta por 273 registros e 11 variáveis, sendo elas: percentual de faltas, nota e situação final da disciplina Algoritmos e Estruturas de Dados 1, percentual de faltas, nota e situação final da disciplina de Cálculo Diferencial e Integral 1, percentual de faltas, nota e situação final da disciplina de Física, percentual de faltas, nota e situação final da disciplina Lógica Matemática, ano de nascimento do discente, auxílio estudantil e status final do aluno (Desistente ou Formado).

Na implementação do algoritmo de redes neurais MLP, os atributos da base foram divididos em classe e elementos previsores, sendo a classe constituída pelo status final do aluno e os previsores compostos pelo restante dos atributos. Foram utilizados 75% dos dados disponíveis para treinamento e o restante para testes.

O algoritmo foi executado por volta de 235 vezes e obteve uma acurácia de 88%.

4. CONCLUSÕES

Com a conclusão deste trabalho, fica evidente que o desempenho nas disciplinas iniciais do curso de Ciência da Computação, assim como fatores socioeconômicos como o auxílio estudantil estão diretamente relacionados a conclusão ou desistência do curso, visto que o algoritmo obteve uma alta taxa de precisão.

Espera-se que com esse trabalho, alunos com risco de abandonarem o sistema de ensino sejam identificados precocemente, para que a instituição possa traçar estratégias para combater a situação.

Para trabalhos futuros, outros algoritmos de aprendizagem de máquina serão testados e comparados, assim como outras variáveis socioeconômicas disponíveis.

REFERÊNCIAS

- COSTA, E. e. a. **Mineração de dados educacionais: conceitos, técnicas, ferramentas e aplicações**. Jornada de Atualização em Informática na Educação, v. 1, n. 1, p.1–29, 2013.
- GOLDSCHMIDT, R.; PASSOS, E. **Data mining: um guia prático**. [S.l.]: Gulf Professional Publishing, 2005.