



DADOS DO TRABALHO

CÓDIGO DE IDENTIFICAÇÃO: 1208

TÍTULO

Inteligência artificial no combate a evasão na educação em Minas Gerais

ÓRGÃO/ENTIDADE EXECUTOR(A)

Secretaria de Estado de Educação

CATEGORIA

Ideias Inovadoras Implementáveis

TEMÁTICA GERAL

Inovação em Gestão Pública

ÁREA TEMÁTICA ESPECÍFICA

Tecnologia da Informação e Comunicação

RESUMO

A Inteligência Artificial (IA) já ocupa espaços importantes no dia a dia de milhares de pessoas. Na educação, essa tecnologia tem o poder de transformar a realidade pedagógica de professores e alunos ajudando a combater a evasão escolar, problema crônico da educação brasileira, que, segundo o Censo Escolar 2018, atingiu 82 mil estudantes da rede estadual de ensino de Minas Gerais, dos quais cerca de 75% deles são do ensino médio.

O objetivo principal desta iniciativa proposta é identificar quais os alunos matriculados na rede estadual de Minas Gerais têm maior risco de abandonar os estudos e como esse risco evolui ao longo do período letivo utilizando técnicas de IA.



Com isso, a partir da análise de diversos dados - como desempenho acadêmico e frequência escolar - será possível auxiliar professores, coordenadores e diretores no desenho de intervenções pedagógicas personalizadas para cada estudante para controlar a evasão escolar e reverter esse quadro no estado de Minas Gerais.

PALAVRAS-CHAVE

Inteligência Artificial Evasão Escolar Educação

CANVAS DE ESTRUTURAÇÃO DE PROJETOS INOVADORES

INSUMOS	O QUE? POR QUE?	FATORES IMPORTANTES
RECURSO NECESSÁRIOS Softwares são de código aberto 3 servidores efetivos capacitados da SEE MG Não são necessários recursos financeiros	DESAFIO/PROBLEMA Evasão escolar na rede estadual de Minas Gerais	PÚBLICO-ALVO Alunos matriculados do 6º Ano ao 3º Ano do Ensino Médio Rede estadual de Minas Gerais
PARCERIAS Fundação João Pinheiro - dados socioeconômicos Polícia Militar de Minas Gerais - registros de ocorrência SEDESE - indivíduos assistidos por programas sociais	IDEIA/INICIATIVA Identificação de alunos em risco de evasão escolar Identificação das principais causas de evasão escolar Auxílio na formulação de políticas públicas personalizadas Toma ensino adaptativo ao estudante em sua jornada	RISCOS E INCERTEZAS Volume grande de dados Qualidade dos dados
ANÁLISE FINANCEIRA Não são necessários investimentos financeiros	VALOR GERADO Desenho de serviços públicos personalizados Políticas públicas focadas no indivíduo	ESTRATÉGIA DE APRIMORAMENTO E MULTIPLICAÇÃO Uso de dados longitudinais Identificação de abandono e de reprovação escolar

DETALHAMENTO DA IDEIA/INICIATIVA

O QUE? POR QUE?

DESAFIO/PROBLEMA

Evasão escolar é o ato de deixar de frequentar as aulas, ou seja, abandonar o ensino em decorrência de qualquer motivo. Esse problema social é, infelizmente, comum no Brasil, e apresenta altos índices. Quase quatro (36,5%) em cada dez brasileiros de 19 anos não concluíram o ensino médio em 2018, idade considerada ideal para esta etapa de ensino. Entre eles, 62% não frequentam mais a escola e 55% pararam de estudar ainda no ensino fundamental. Em



Minas Gerais, segundo o Censo Escolar 2018, 82 mil estudantes da rede estadual de ensino abandonaram os estudos durante o ano - cerca de 75% deles são do ensino médio.

As causas para a evasão escolar são as mais diversas. Segundo Soares, Fernandes, Nóbrega e Nicoletta (2015) dentre as características individuais que estes citam como propícios fatores para o abandono destacam-se: o baixo desempenho educacional; comportamento e atitudes do aluno (tais como o pouco envolvimento acadêmico no aspecto das atividades escolares e das atividades sociais, o absentefismo e as expectativas educacionais); assim como características demográficas - cor/raça, gênero, idade.

Dentre as características institucionais, os mesmos autores apontam três aspectos: a estrutura (se tradicional ou monoparental, por exemplo) e mudanças nessa estrutura familiar (como a separação dos pais) ao longo do processo escolar; a renda e outros recursos familiares; e o capital social (como altas expectativas educacionais, acompanhamento do progresso escolar dos filhos e a participação na vida escolar dos filhos).

As variáveis são muitas como se pode observar, entretanto, utilizando a tecnologia e análise de dados é possível, de forma sistemática, agregar diversas informações e de forma preliminar e previsível saber se um determinado aluno da rede estadual de ensino de Minas Gerais possui a probabilidade alta de abandonar a escola e o Estado, por meio da Secretaria de Estado de Educação pode, anteriormente a sua evasão, realizar ações de acompanhamento e prevenção, impedindo ao máximo a evasão.

IDEIA/INICIATIVA

O objetivo principal desta iniciativa proposta é identificar quais os alunos matriculados na rede estadual de MG têm maior risco de abandonar os estudos e como esse risco evolui ao longo do período letivo utilizando técnicas de Inteligência Artificial (IA).

Para alcançar este objetivo pretende-se utilizar das informações disponíveis no Sistema Mineiro de Administração Escolar, no Diário Escolar Digital (DED) e no Censo Escolar, além de outros dados externos Secretaria de Educação a fim de endossar o projeto. O primeiro gerencia a vida escolar do estudante e dispõe de informações socioeconômicas a nível individual como sexo, raça, idade, residência, mobilidade interescolar e acadêmicas como as notas dos alunos de modo que, juntamente com os dados de frequência diária disponíveis no DED, torna possível criar um perfil do estudante e gerar importantes parâmetros sobre seu comportamento. De forma complementar, serão utilizados para contextualizar as escolas estaduais o Indicador de Nível Socioeconômico desenvolvido pelo INEP.

A partir daí, por meio das técnicas de classificação supervisionada binária a Inteligência Artificial estabelece um comparativo entre as informações de cada aluno e permite identificar quais estudantes estão mais propensos a abandonar a escola. As técnicas também permitem mensurar quais as principais variáveis que contribuem para o risco de abandono escolar para cada aluno.

Esse conjunto de informações tornará possível à SEE identificar o perfil de seus estudantes auxiliando na formulação de políticas públicas personalizadas e no monitoramento das ações durante todo período escolar. Para cada perfil de estudante poderá ser recomendado uma ação ou um conjunto de ações, como, por exemplo, acompanhamento familiar, reforço escolar ou indicação para Educação em Tempo Integral. Além disso, o monitoramento dessas ações servirá como feedback para melhoria dos algoritmos desenvolvido.

De forma mais ampla, esse ensino adaptativo acompanha o estudante em sua jornada de aprendizagem, onde a escola poderá propor conteúdos e avaliar o seu aproveitamento no ritmo de cada aluno. Além disso, a IA consegue repassar esses insights aos professores, que conhecem melhor a turma e cada aluno individualmente. Assim, eles podem atuar de maneira ainda mais eficaz. Também é possível potencializar este próximo passo por meio de uma política pública especializada, que pode trazer novas ideias e maior eficiência na execução dos planos de ação junto aos estudantes



ESTUDOS PRELIMINARES

A principal referência e inspiração para esta iniciativa é o trabalho que vem sendo desenvolvido pela Universidade de Chicago por meio do programa Data Science for Social Good, o qual procura auxiliar organizações governamentais e não governamentais sem interesse lucrativo em projetos de inteligência artificial.

Nos anos recentes esse programa auxiliou os sistemas de educação de condados americanos - Chicago (2014, 2015, 2016) - na identificação dos estudantes com risco de não se formarem no Ensino Médio em tempo adequado, melhorando os sistemas de identificação já existentes e elevando significativamente o poder preditivo.

Considera-se que a Secretaria de Estado de Educação pode-se valer dessas experiências e dos algoritmos que foram desenvolvidos e que estão disponíveis em repositório de código aberto DSSG (2015) para adaptar a proposta para sua realidade e conseguir desenvolver metodologia própria para identificação de alunos em risco de evasão escolar.

GRAU DE NOVIDADE

Não foram identificados demais órgãos ou entidades no Governo do Estado de Minas Gerais que realizam monitoramento de fenômenos sociais a partir do processamento de dados e aplicação de algoritmos e técnicas de Inteligência Artificial

Pode-se considerar que o grau de novidade e inovação é alto. O tema Inteligência Artificial é relativamente novo e o seu uso em políticas públicas é ainda bem inicial e atual. Em se tratando no seu uso em um grau de 3612 escolas e 1,4 milhão de estudantes, distribuídos em 700 mil nos anos finais do Ensino Fundamental e 700 mil no Ensino Médio, seria um case inédito de nível mundial devido à proporção e impacto que pode-se gerar com o seu uso.

VALOR GERADO

O sistema aqui proposto tem como propósito inicial identificar para cada aluno matriculado na rede estadual de ensino de Minas Gerais qual é o risco de abandono escolar a ele associado e como esse risco evolui ao longo do período letivo de acordo com suas medidas de desempenho escolar como nota e frequência.

Desta forma, aliar a capacidade do motor cognitivo da IA a processos internos estruturados da educação viabiliza na identificação dos fatores que levam o aluno à desistência. É essa análise tecnológica preditiva capaz de cruzar os dados individuais de cada perfil, o desempenho acadêmico, e faltas do aluno e, com isso, gerar importantes parâmetros sobre o seu comportamento e assim, permite detectar quais estudantes estão mais propensos a abandonar a sua inscrição e, tendo em mãos os principais indicadores desses motivos, a instituição de ensino tem a possibilidade de criar estratégias para evitar a evasão.

Com os algoritmos desenvolvidos será possível identificar quais são os fatores que mais contribuem para o risco de abandono associado a cada aluno e identificar o perfil do estudante em questão, possibilitando uma ação imediata e o acompanhamento diferenciado durante todo o período escolar. Assim, ela pode oferecer um feedback imediato a estes atores e de forma pessoal e cuidadosa auxiliar o aluno a solucionar essas lacunas que estão propiciando-o a abandonar a escola.

De forma mais ampla, esse ensino adaptativo acompanha o estudante em sua jornada de aprendizagem, onde a Escola estadual poderá propor conteúdos e avaliar o seu aproveitamento no ritmo de cada aluno. Além disso, a IA consegue repassar esses insights aos professores, que conhecem melhor a turma e cada aluno individualmente. Assim, eles



podem atuar de maneira ainda mais eficaz. Também é possível potencializar este próximo passo por meio de uma política pública especializada, que pode trazer novas ideias e maior eficiência na execução dos planos de ação junto aos estudantes.

FATORES IMPORTANTES

PÚBLICO ALVO

A identificação do abandono escolar é realizada anualmente para todas as redes de ensino do país pelo INEP por meio dos dados do Censo Escolar que faz parte do processo de cálculo das taxas de rendimento - aprovação, reprovação e abandono.

Por questões metodológicas e de disponibilidade de dados, o INEP calcula essas taxas apenas para o Ensino Regular, o rendimento não é calculado para a EJA e a Educação Profissional.

Os alunos que abandonaram os estudos já identificados pelo INEP no Censo Escolar servirão como base para a definição dos algoritmos de classificação, os quais serão utilizados para identificar os alunos que têm risco de abandonar.

Por esses motivos a iniciativa proposta tem como foco apenas o Ensino Regular ofertado na rede estadual de Minas Gerais. Além disso, dado que as taxas de abandono se tornam mais expressivas a partir do 6º Ano do Ensino Fundamental, o público-alvo da iniciativa são todos os estudantes matriculados entre o 6º Ano e o 3º Ano do Ensino Médio.

RISCOS E INCERTEZAS

Em se tratando da atuação da Secretaria de Educação em uma enorme gama de alunos e escolas, por consequência, o volume de dados que o sistema utilizará será proporcionalmente extenso. Com isso, o tratamento dessa informação é de extrema importância e carece de um cauteloso trabalho para que a qualidade dos dados seja assegurada. Corroborando para esse trabalho o comprometimento da SEE-MG no lançamento tempestivo das notas e frequências dos estudantes SEE-MG (2019).

Outro ponto que se destaca é que como a iniciativa se propõe a trabalhar com previsões, ou seja, realidades estatísticas e não com realidades literais, resultados errôneos podem ser gerados com a atuação do sistema de Inteligência Artificial.

Embora todos os riscos da IA sejam reais, o padrão de referência adequado não é a perfeição, mas sim a busca da melhor resposta disponível. A vantagem dos sistemas baseados em máquinas é que eles podem ser melhorados sempre, buscando atrelar mais e melhores dados ao longo do tempo, incorporando os erros que são identificados, visando uma resposta cada vez mais consistente e acurada em sua previsão de abandono escolar dos alunos da rede pública estadual.

É importante ressaltar que os algoritmos de classificação desenvolvidos utilizando-se os dados da coorte de estudantes do ano de 2018 serão validados com uma coorte diferente, a do ano de 2019. O objetivo é aumentar a robustez dos algoritmos ao aplicá-los em dados nunca antes vistos.

Como foi mencionado, 82 mil alunos abandonaram a escola em 2018. Para auxiliar na tomada de decisão e definição de políticas públicas, os estudantes identificados com risco de abandono escolar poderão ser agrupados em grupos de



risco priorizando-se aqueles que se encontram em situação mais vulnerável. Isso será possível pois os algoritmos de classificação utilizados calculam a probabilidade associada ao abandono.

ESTRATÉGIA DE APRIMORAMENTO E MULTIPLICAÇÃO

A iniciativa proposta irá utilizar os dados de caráter socioeconômico dos estudantes e também dados de desempenho acadêmico e frequência escolar ao longo do período letivo. Esse conjunto de dados contempla dados que são estáticos - inerentes ao estudante - e dados que são dinâmicos e que refletem o comportamento do aluno em relação aos estudos e desempenho educacional.

Como forma de aprimoramento da iniciativa apresentada podem ser pensadas estratégias para utilizar dados históricos dos estudantes. Realizando um acompanhamento longitudinal dos estudantes será possível construir trajetórias acadêmicas dos estudantes e identificar quais os principais gargalos que dificultam uma trajetória de sucesso e em que momento da vida acadêmica do estudante eles mais ocorrem. Ao se utilizar de dados históricos melhoram-se os algoritmos de identificação de abandono escolar e ganha-se mais tempo para se realizar as intervenções necessárias dado que a identificação pode ser antecipada Lakkaraju et. al (2015) e Aguiar et. al (2015).

O trabalho realizado pelo INEP na construção de indicadores de fluxo escolar utilizando dos dados do Censo Escolar INEP (2017) pode servir de inspiração para que a Secretaria de Estado de Educação desenvolva trabalho semelhante, já que se dispõe de informações históricas dos estudantes da rede estadual de ensino de Minas Gerais. Pelo grau de maturidade da SEE MG no que tange à capacidade de inteligência de dados optou-se nesta iniciativa pela proposta de utilização dos dados dos estudantes apenas para 1 ano letivo pela complexidade técnica elevada associada a construção de uma base de acompanhamento longitudinal. Ressalta-se que essa opção constitui um primeiro passo importante. Outro ponto de melhoria consiste na transição do uso de técnicas de classificação supervisionada binária para o que é chamado na literatura de machine learning como multi class classification. A melhoria mencionada propõe que se identifique para além do abandono também o aluno que tem risco elevado de reprovação. Dessa forma, também poderiam ser pensadas estratégias para evitar a reprovação estudantil e garantir que os alunos aprendam mais e na idade certa.

O monitoramento do desempenho acadêmico de cada estudante e de sua frequência das aulas será realizado constantemente pelo Simade e pelo Diário Escolar Digital de modo que será possível acompanhar cada aluno, sua evolução ao longo do período letivo e as respostas às atividades de intervenção pedagógicas propostas.

INSUMOS

RECURSOS NECESSÁRIOS

RECURSOS HUMANOS

A Secretaria de Estado de Educação dispõe atualmente de uma equipe de 3 servidores efetivos capacitados para realizar essa iniciativa.



INFRA-ESTRUTURA

A Secretaria de Estado de Educação dispõe das máquinas e equipamentos necessários para a construção dos modelos de machine learning. Os softwares em Python são de dados abertos e possuem licenças gratuitas para o uso. Não há necessidade de contratação ou aquisições. E toda a infraestrutura utilizada são os próprios computadores já disponíveis na secretaria.

ANÁLISE FINANCEIRA

CUSTOS DE IMPLANTAÇÃO/MANUTENÇÃO

Não há nenhum custo de implantação ou manutenção. Todos os softwares são gratuitos para sua utilização e todos os computadores e mão de obra já estão disponíveis pela Secretaria de Educação.

RECURSOS ORÇAMENTÁRIOS E FINANCEIROS

Para desenvolver o desenvolvimento de algoritmo de machine learning para identificação de alunos com provável risco de abandono escolar não é necessário nenhum tipo de aporte financeiro, todas as licenças de software utilizadas são gratuitas, assim não há o que se falar ou especificar recursos internos ou externos.

RECURSOS ORÇAMENTÁRIOS E FINANCEIROS

Para desenvolver o desenvolvimento de algoritmo de machine learning para identificação de alunos com provável risco de abandono escolar não é necessário nenhum tipo de aporte financeiro, todas as licenças de software utilizadas são gratuitas, assim não há o que se falar ou especificar recursos internos ou externos.

PARCERIAS

A proposta de identificação dos alunos com risco de abandono escolar aqui elencada se faz uso de dados e informações disponíveis no Simade, DED e Censo Escolar, os quais a SEE-MG tem acesso direto.

Para além desses dados, identifica-se como oportuno o estabelecimento de parcerias com órgãos do Poder Executivo Estadual para utilização de informações que podem aprimorar a proposta e auxiliar na identificação da evasão escolar. São estes os informativos utilizados de outras áreas extra Secretaria de Educação:

- Registros de ocorrência de violência nas escolas e em suas proximidades extraídos pela Polícia militar;
- Dados socioeconômicos para contextualizar cada realidade territorial fornecidos pela Fundação João Pinheiro;
- Dados para identificação de alunos assistidos por programas sociais, como por exemplo o programa Bolsa Família,



extraídos pela SEDESE.

ESTRATÉGIA DE IMPLANTAÇÃO

A primeira etapa desta iniciativa consiste no tratamento e preparação da base de dados que reunirá informações do Simade, do DED e do Censo Escolar, das quais a SEE-MG já dispõe.

Em seguida, será o momento de desenvolver os algoritmos de machine learning analisando o resultado das predições e avaliando as principais características que levam ao abandono escolar na rede estadual, é nessa etapa que também serão definidos critérios para priorização dos estudantes e identificação de diferentes grupos de risco.

No início do próximo ano, a SEE-MG terá identificados os alunos que requerem maior grau de atenção e poderá direcionar políticas pedagógicas personalizadas para auxiliar esses estudantes e assim direcionar políticas para contenção no que diz respeito ao abandono de alunos nas escolas da rede pública estadual.

CRONOGRAMA

DESCRIÇÃO	INICIO	TÉRMINO	STATUS	ENTREGA
Coleta dos dados - Simade, DED, Censo Escolar	01/07/2019	26/07/2019	Concluído	Relatórios de dados a serem usados
Preparação da base de dados	29/07/2019	29/09/2019	Em andamento	Base de dados pronta para utilização
Análise exploratória dos dados	30/09/2019	01/11/2019	A iniciar	Variáveis-chave da evasão escolar identificadas
Desenvolvimento dos algoritmos de machine learning	04/11/2019	27/12/2019	A iniciar	Algoritmos de machine learning
Validação dos algoritmos com a coorte dos estudantes de 2019	06/01/2020	31/01/2020	A iniciar	Algoritmos validados e melhorados
Identificação dos estudantes matriculados em 2020 com risco	03/02/2020	28/02/2020	A iniciar	Lista de estudantes agrupados por grupo de risco