

Inteligência Artificial na Educação 4.0: Auxiliando o Professor no Ensino Individualizado dos Alunos

Autores:

Masanori Iha

Raphael Coqui Camargo

Rodrigo Tassin Nappi

William Honorato dos Santos

Orientador:

Prof. Me. Gabriel Lara Baptista

Resumo

Tradicionalmente a evolução tecnológica esteve ligada a indústria e, como consequência da modernização e em resposta às necessidades emergentes, proporcionou o desenvolvimento e a implementação de novas metodologias de ensino. Este fato fica mais evidente ao considerar o papel de mediador assumido pelo professor na Educação 4.0 e as ferramentas utilizadas com foco na interação e na produção de conteúdo por parte dos alunos. Neste contexto, esta pesquisa tem como objetivo utilizar a Inteligência Artificial – especialmente o algoritmo supervisionado *K-Nearest-Neighbors* (KNN) – para auxiliar o professor a identificar as dificuldades dos alunos de forma individualizada, a partir de uma avaliação composta por questões objetivas e vinculadas a temas específico. Assim, buscou-se aferir a acurácia alcançada pelo KNN na classificação de diferentes padrões de respostas e, conseqüentemente, na indicação do melhor material de estudo possível considerando as necessidades individuais de aprendizagem de cada um dos discentes. Os resultados indicam que o KNN aparenta ser uma alternativa possível, sugerindo um bom desempenho quando aplicado a uma base de dados inicialmente pequena, fazendo com que ele possa ser utilizado de maneira rápida e simples.

Palavras-chave: Educação 4.0; Inteligência Artificial; *Machine Learning*, KNN.

Abstract

Technological evolution has traditionally been linked to industry and, because of modernization and in response to emerging needs, it has provided the development and implementation of new teaching methodologies. This fact is more evident if we consider the role of mediator assumed by the teacher in Education 4.0 and the tools used with a focus on interaction and content production by students. In this context, this research aims to use Artificial Intelligence - especially the supervised K-Nearest-Neighbors (KNN) algorithm - to assist the teacher in identifying the students' difficulties individually, based on an assessment composed of objective and linked to specific topics. Thus, we sought to assess the accuracy achieved by the KNN in the classification of different response patterns and, consequently, in the indication of the best possible study material considering the individual learning needs of each of the students. The results indicate that KNN appears to be a possible alternative, suggesting a good performance when applied to an initially small database, making it able to be used quickly and simply.

Keywords: Education 4.0; Artificial Intelligence; Machine Learning; KNN.

1 Introdução

Segundo Führr (2018b), historicamente a educação foi influenciada pela indústria. Este fato se torna ainda mais evidente nos períodos de transição entre as fases da industrialização marcados pelas três Revoluções Industriais. Essa influência ocorre devido à necessidade de formar mão-de-obra capaz de se adaptar aos novos métodos de produção, de realizar as novas atividades e de corresponder às expectativas do mercado de trabalho resultantes da modernização dos meios de produção. Hoje, com a chegada da Indústria 4.0 e as oportunidades criadas por ela, os formatos utilizados no ensino, assim como seus métodos e suas metodologias, voltam a ser questionados, dando origem a Educação 4.0 FÜHR (2018b).

Neste contexto a tecnologia assume um papel importante sendo a principal responsável no suporte a alunos e professores. A Internet, além de ser utilizada como plataforma para distribuição de conteúdo – principal característica da Educação 3.0 – passa a coletar dados e a transformá-los em informação por meio de algoritmos em tempo real, identificando o perfil do agente e possibilitando que o conteúdo se adapte a ele, considerando não apenas o seu potencial, mas também as suas preferências pessoais e os formatos que mais lhe agradam BARCIA et al. (1999).

Esse avanço foi possível graças ao aumento da capacidade de processamento computacional proporcionado pela computação em nuvem, o que nos permite aplicar técnicas de Inteligência Artificial cada vez mais complexas na análise dos dados gerados na interação entre homem e máquina, originando uma resposta cada vez mais assertiva e rápida, atendendo as necessidades de controle e automação de tarefas provenientes da indústria, assim como a de identificar o perfil do agente e suas preferências pessoais, possibilitando a classificação e a entrega de conteúdos relevantes a ele no momento correto e no formato adequado (FÜHR, 2018a).

Desta maneira, este artigo tem como objetivo sugerir uma proposta de aplicação da Inteligência Artificial – em especial o Aprendizado de Máquina Supervisionado – que seja capaz de auxiliar o professor a identificar as dificuldades dos alunos de forma individualizada e assertiva, sugerindo a ele o melhor conteúdo possível para cada discente.

2 Referencial Teórico

A Educação 4.0 está em fase inicial, não se sabe ao certo como será a transição entre o modelo que conhecemos hoje e as novas possibilidades promovidas por ela. Porém, conforme citado por Führr (2018b), algumas perguntas podem ser feitas para decifrar esse caminho, tais como: Como a Educação 4.0 pode garantir a aprendizagem do aluno enquanto desenvolve as competências mínimas e necessárias para o século XXI? Quais os impactos das mudanças oriundas do advento do Mundo 4.0 para o docente?

2.1 As Fases da Educação

Segundo Führr (2018b), a Educação 4.0 está diretamente ligada a quarta Revolução Industrial e deve interagir com as novas competências exigidas pela indústria. Em um contexto temporal de evolução, a forma de ensino passou por três grandes mudanças - Educação 1.0, 2.0 e 3.0 – em que é possível observar que cada uma delas foi acompanhada de mudanças significativas na maneira pela qual se ensina e se aprende.

Segundo GOMES et al. (2013), durante a Educação 1.0 a tecnologia que o docente dispunha para auxiliá-lo na disseminação de seus conhecimentos eram escassas e concentravam-se nas mídias de massa. Os conteúdos produzidos em larga escala pelas editoras

eram impressos e não editáveis, e as avaliações eram feitas exclusivamente por meio de exames e questionários.

Influenciado pela Segunda Revolução Industrial, deu-se início a um novo modelo de escola baseado nas características da produção industrial. Essa abordagem tinha como objetivo principal treinar o aluno para o mercado de trabalho e a sociedade. Conforme definido por Almeida e Simoes (2019), a Educação 2.0 tem foco no desenvolvimento de novos projetos em grupo e na utilização de novas tecnologias.

A Educação 3.0 surgiu com a era da informação e da globalização, fazendo com que o ensino fosse adaptado às tecnologias emergentes e as novas possibilidades de interação, dentro e fora do espaço físico das instituições. Conforme Barcia et al. (1999), a tecnologia nos computadores apresenta inúmeros recursos, entre eles possibilita várias formas de relações, enriquecendo as experiências dos alunos, contribuindo em seu desenvolvimento e propiciando a construção do conhecimento através de explorações autônomas e independentes por parte do sujeito.

Assim como aconteceu na Educação 2.0, a atividade industrial tem grande influência sobre a Educação 4.0, ditando as regras e definindo as necessidades do mercado de trabalho e da sociedade, garantindo assim que os seus objetivos sejam atendidos. Neste contexto e considerando o fato de o conhecimento não estar centralizado como antigamente, mas sim distribuído pela Internet e praticamente onipresente, a figura do professor se assemelha cada vez mais a de um orientador, auxiliando os alunos a filtrar, organizar e transformar toda a informação disponível em conhecimento para ajudar a sociedade (FÜHR, 2018a).

Segundo De Jesus (2019), avaliando o contexto apresentado não é possível imaginar as escolas isoladas em grandes bolhas, mas também é preciso considerar que a tecnologia não atende a totalidade, sejam os alunos ou professores, de forma homogênea. Observa-se também que os educandos aprendem em variados ambientes, sejam eles complexos ou não, resolvem problemas de forma autônoma e adquirem novas habilidades técnicas rapidamente, compartilhando os riscos e objetivos de forma flexível (FÜHR, 2018a).

Para Balsan (2019), outro grande pilar da Educação 4.0 é o método de avaliação. Hoje, esta avaliação é feita por meio de trabalhos e provas que verificam se os alunos conhecem os conteúdos vistos em sala de aula, porém não é raro de se perceber que em numerosos casos, os alunos não compreendem os conceitos apresentados e apenas memorizam dados e fórmulas. Esse método tradicional tem como objetivo verificar acertos e erros, ignorando o conteúdo aprendido de fato.

Balsan (2019) também defende a ideia de que as escolas já possuem *softwares* para gestão, que computam notas, faltas e materiais de apoio, facilitando a aplicação de provas para avaliar o conhecimento do aluno e reduzindo a atividade do professor a corrigir e registrar a nota obtida neste sistema, que apenas aceita essa informação com o intuito de concluir ou reprovar a disciplina, sem nenhum critério que possa auxiliar o aluno na evolução do conhecimento.

Desta maneira, é possível observar que o professor esteve presente em inúmeros papéis ao longo desse ciclo evolutivo. Na Educação 1.0 ele foi tido como uma autoridade incontestada; na Educação 2.0 ele surgiu como fonte do conhecimento, e não apenas mais como uma figura autoritária; e na Educação 3.0 começou a exercer o papel de mentor, tutor e facilitador na busca pelo aprendizado e conhecimento (GOMES et al., 2013).

Além disso, até a Educação 2.0 os alunos tinham um papel passivo, no qual observavam e aprendiam exclusivamente com o professor. Também durante esse período, a autonomia do estudante era restrita ao estudo de casa e limitada ao material escolar ou a biblioteca. Durante a Educação 3.0 e na Educação 4.0, os alunos passam a ter um papel reflexivo, participando e contribuindo para o aprendizado do coletivo, além do acesso online à informação (GOMES et al., 2013).

Hoje, a Educação 4.0 utiliza a tecnologia para transformar a experiência de aprendizagem, incentivando que os alunos se esforcem nos estudos guiados pelo interesse, auxiliando o professor na tarefa de monitoria deste processo com dados mais detalhados e possibilitando uma melhoria na experiência de ensino e aprendizagem (ANDRADE, 2018).

2.2 Inteligência Artificial - IA

Segundo Russell e Norvig (2013), são inúmeras as definições para Inteligência Artificial, sendo esse o reflexo das particularidades de cada campo de conhecimento. A Tabela 1 mostra oito destas definições agrupadas em duas dimensões: as relacionadas ao pensamento, processos e raciocínio; e as relativas à comportamento.

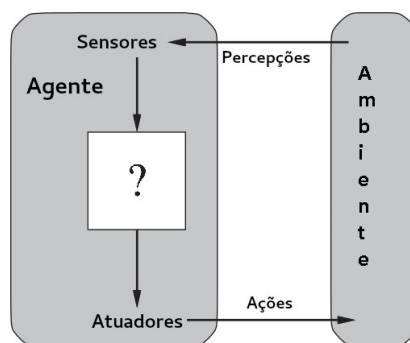
Tabela 1 - Algumas definições de Inteligência Artificial.

Pensando como um humano	Pensando racionalmente
<p>“O novo e interessante esforço para fazer os computadores pensarem (...) máquinas com mentes, no sentido total e literal.” Haugeland (1985 <i>apud</i> Russell e Norvig, 2013)</p> <p>“[Automatização de] atividades que associamos ao pensamento humano, atividades como a tomada de decisões, a resolução de problemas, o aprendizado...” Bellman (1978 <i>apud</i> Russell e Norvig, 2013)</p>	<p>“O estudo das faculdades mentais pelo uso de modelos computacionais.” Charniac e McDermott (1985 <i>apud</i> Russell e Norvig, 2013)</p> <p>“O estudo das computações que tornam possível perceber, raciocinar e agir.” Winston (1992 <i>apud</i> Russell e Norvig, 2013)</p>
Agindo como seres humanos	Agindo racionalmente
<p>“A arte de criar máquinas que executam funções que exigem inteligência quando executadas por pessoas.” Kurzweil (1990 <i>apud</i> Russell e Norvig, 2013)</p> <p>“O estudo de como os computadores podem fazer tarefas que hoje são mais bem desempenhadas por pessoas.” Rich e Knight (1991 <i>apud</i> Russell e Norvig, 2013)</p>	<p>“Inteligência computacional é o estudo do projeto de agentes inteligentes.” Poole et al. (1998 <i>apud</i> Russell e Norvig, 2013)</p> <p>“AI... está relacionada a um desempenho inteligente de artefatos.” Nilson (1998 <i>apud</i> Russell e Norvig, 2013)</p>

Fonte: Adaptado de Russell e Norvig (2013).

Para Russell e Norvig (2013), essas definições são seguidas até hoje no estudo da IA sendo cada uma dessas abordagens realizada por pessoas e métodos diferentes. Ainda conforme o autor, um agente é tudo aquilo pode fazer algum tipo de interação com o ambiente, percebendo alterações através de sensores e agindo sobre esse ambiente com atuadores, como mostra a Figura 1.

Figura 1 - Esquema de um agente e o ambiente.



Fonte: Russell e Norvig (2013).

O humano pode ser considerado um tipo de agente que possui visão, audição e outros órgãos que funcionam como sensores, além de boca, mãos e pernas que se comportam como atuadores. Já um robô é um agente que se utiliza de câmeras e outros componentes como sensores e motores para interagir com o ambiente.

2.3 Aprendizado de Máquina (*Machine Learning*)

Aprendizado de Máquina (AM) é uma das várias áreas pertencentes à Inteligência Artificial, tendo como um de seus objetivos desenvolver técnicas computacionais que sejam capazes de adquirir conhecimento de forma autônoma. Um programa de computador que consegue tomar decisões baseando-se em experiências anteriores é tido como um sistema de aprendizado. Apesar dos diversos sistemas de aprendizado de máquina possuírem aspectos particulares e comuns, é possível classificá-los quanto à linguagem de descrição, modo, paradigma e forma utilizados pelos algoritmos (MONARD; BARANAUSKAS, 2003).

Conforme Souto et al. (2003), as técnicas utilizadas em *Machine Learning* podem ser divididas, de uma forma geral, em duas dimensões: aprendizado supervisionado e não supervisionado. Para ser considerado aprendizado supervisionado, o indutor, no momento da aprendizagem, recebe um conjunto de dados conhecidos no qual cada informação é fundada por atributos de entrada e saída (rótulos). Já no aprendizado não supervisionado, apenas os atributos de entrada estão disponíveis ao indutor.

Para Russell e Norvig (2013), ambas as partes que integram um agente podem ser melhoradas através dos dados utilizando técnicas de aprendizagem. Essas melhorias e as técnicas usadas para construí-las dependem de quatro fatores principais: qual componente tem que ser melhorado, conhecimento anterior que o agente possui, que tipo de representação é utilizada para os dados e para os componentes e qual o *feedback* utilizado para aprendizagem.

2.4 *K-Nearest Neighbors* – KNN

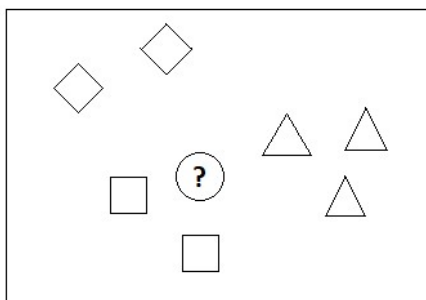
Um aspecto importante relacionado à pesquisa acadêmica é a criação de métodos capazes de prever o comportamento dos estudantes, de modo a possibilitar a intervenção de docentes ou demais envolvidos, aspirando a resgatar o estudante antes que ele seja reprovado (MACFADYN; DAWSON, 2010).

Um sistema de recomendação pode ser definido como um programa com o objetivo de inferir as preferências e necessidades do usuário e indicar os itens mais adequados de acordo com o seu perfil. Para realizar essas recomendações podem ser utilizados os dados do usuário, os itens consumidos por ele ou a relação entre ambos. (LU et al., 2015)

O KNN é um dos algoritmos supervisionados mais utilizados no processo de classificação e recomendação de conteúdo. Também conhecido como algoritmo dos K's vizinhos mais próximos, ele pertence à família dos IBL (*Instance-based Learning*) (COVER & HART, 1967), o que significa que sua implementação foi baseada em um paradigma que utiliza dados armazenados ao invés de um conjunto de regras induzidas e aprendidas para a classificação de novos conjuntos de dados. A classificação de um novo conjunto é baseada na quantidade de dados similares K mais próximos, de acordo com a métrica de distância Euclidiana (AHA; KIBLER; ALBERT, 1991).

Abaixo, a Figura 2 mostra uma representação de algumas classes formadas pelos conjuntos de treinamento e um novo dado desconhecido que deve ser classificado.

Figura 2 - Representação de grupos conhecidos em relação a um dado desconhecido.

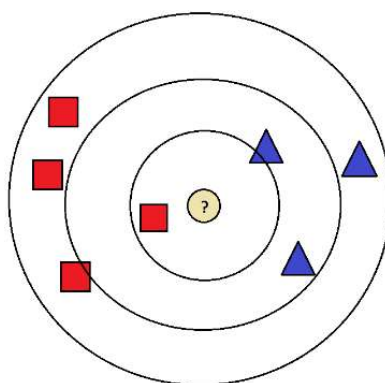


Fonte: Elaborada pelo autor.

Utilizando a notação $VP(k, x_q)$ que indica os k vizinhos mais próximos de x_q , que para classificar, primeiro é encontrado os números de vizinhos de x_q a partir $VP(k, x_q)$ e então é verificada a quantidade de votos, sendo assim possível estimar a qual classe esse novo ponto pertence. Vale observar que, para evitar empates na quantidade de votos, k deve ser sempre um número ímpar (RUSSELL; NORVIG, 2013).

Passos et al. (2015) diz que um ponto importante que deve ser considerado é o valor que será atribuído a k . Se o valor de k for muito pequeno, o resultado do algoritmo pode estar sujeito a *outliers* (discrepâncias), e por outro lado se for um valor muito grande, pode ocorrer a inclusão de muitos pontos inerentes à classes distintas. A Figura 3 mostra um exemplo de como o valor de k pode influenciar na classificação de um dado desconhecido. Se $k = 1$, o novo dado inserido vai pertencer aos quadrados, se $k = 3$ então pertence aos triângulos e por fim se $k = 7$ o dado volta a fazer parte da classe dos quadrados.

Figura 3 - Influência o valor de k .



Fonte: Elaborada pelo autor.

O KNN faz parte do conjunto de algoritmos que utilizam o aprendizado baseado em casos (*Case Based Learning*) empregando o princípio de continuidade – no qual casos semelhantes devem estar dentro de uma mesma classe – comum a todos os algoritmos que se utilizam do conceito de semelhança direta para determinar a classificação de um dado. Assim, o KNN faz a busca a partir da aproximação dentre os pontos a serem analisados, aplicando uma função que calcula a distância entre cada par destes pontos. Existem várias formas possíveis de definir a função utilizada para calcular esta distância, um exemplo é a *Distância Euclidiana* que é definida por $\sqrt{\sum (a_i - b_i)^2}$, além da *Distância de Manhattan* dada por $\sum |a_i - b_i|$ (PASSOS et al., 2015).

2.5 K-Fold e Leave One Out – LOO

Conforme Pedregosa et al. (2020), colocar um modelo para aprender os parâmetros de uma função de previsão e realizar os testes com os mesmos dados é um erro metodológico, pois um modelo que apenas repete os rótulos das amostras que aprendeu teria uma pontuação perfeita, mas não conseguiria uma boa acurácia ao classificar novos dados não sendo capaz de fazer uma boa generalização. Para evitar esse problema, é uma prática comum nos experimentos de aprendizado de máquina supervisionados reservar uma parte dos dados para validação e utilizar o restante no treinamento do modelo, garantindo resultados mais precisos (DUCHESNE; RÉMILLARD, 2005).

Tendo k como sendo a quantidade de partes em que a base de dados será fracionada, ou seja, considerando uma base com 100 amostras e k igual a cinco, o conjunto de dados será dividido em cinco partes iguais. Após a divisão, um dos fragmentos será utilizado para a validação do algoritmo e as demais para o treino (SCHREIBER et al., 2017), conforme mostra a Figura 4.

Figura 4 - Divisão e iterações na base de dados com *K-Fold*.

	Parte 1	Parte 2	Parte 3	Parte 4	Parte 5
Iteração 1	Teste	Treino	Treino	Treino	Treino
Iteração 2	Treino	Teste	Treino	Treino	Treino
Iteração 3	Treino	Treino	Teste	Treino	Treino
Iteração 4	Treino	Treino	Treino	Teste	Treino
Iteração 5	Treino	Treino	Treino	Treino	Teste

Fonte: Elaborada pelo autor.

Além do *k-fold*, outro método utilizado para validação de modelos, criados ou não a partir do KNN, é o *Leave One Out* (LOO). Este método pode ser definido como um caso particular de aplicação do *K-fold*, tendo k igual ao total de amostras - 1 e utilizando o fragmento restante para validar o modelo, conforme exemplifica a Figura 5.

Figura 5 – Aplicação do *Leave One Out* em uma base contendo 100 amostras.

	Parte 1	Parte 2	Parte 3	...	Parte 100
Iteração 1	Teste	Treino	Treino	...	Treino
Iteração 2	Treino	Teste	Treino	...	Treino
Iteração 3	Treino	Treino	Teste	...	Treino
...					
Iteração 100	Treino	Treino	Treino	...	Teste

Fonte: Elaborada pelo autor.

Assim, pode-se observar que a complexidade computacional tanto do *k-fold* quanto do LOO é alta, visto que todos os segmentos da base de dados são comparados entre si, resultando na acurácia média obtida entre todas as iterações (SCHREIBER et al., 2017).

3 Metodologia

Tendo como foco testar as hipóteses de aplicação da Inteligência Artificial no auxílio ao professor, facilitando a classificação e a indicação de conteúdo relevante aos discentes por meio de uma visão individualizada destes alunos, espera-se que este estudo revele a acurácia do *K-Nearest-Neighbors* (KNN) – algoritmo supervisionado de classificação – ao relacionar o conhecimento do docente e a sua visão referente às dificuldades de cada aluno ao resultado apontado pelo algoritmo.

Para isso, utilizou-se o *Notebook Jupyter* em conjunto com o *Python*, criando um ambiente propício à instalação de bibliotecas relacionadas ao tratamento e a classificação de dados, como a *scikit-learn* – uma biblioteca especializada em aprendizagem de máquina – utilizada para aplicação do KNN.

Após a definição das ferramentas a serem utilizadas, foram seguidos dois processos distintos na tentativa de atingir os objetivos deste estudo, no qual um se mostrou eficaz e o outro não, conforme descrito a seguir.

3.1 Experiência realizada com uma base de dados criada randomicamente

Considerando os dados que representam as respostas dos alunos como variáveis independentes e os materiais de estudo sugeridos pelo sistema e pelo professor como dependentes, realizou-se uma experiência cujo método adotado foi isolar as variáveis independentes, controlando-as para garantir a sua integridade, analisar as dependentes e comparar os resultados teóricos obtidos ao comportamento esperado em um cenário real.

Nesta experiência, as respostas dos alunos foram representadas por meio de um algoritmo que, a partir de um determinado questionário, responde às perguntas de forma aleatória. Para sugerir os conteúdos com base na combinação dos padrões de respostas, simulando a atuação do professor, foi atribuído um fator relacionado a aderência de um determinado material de estudo a cada uma das alternativas, seguindo um critério lógico. Este fator foi utilizado pelo algoritmo de tal forma que a soma obtida em cada um dos temas abordados pelo questionário pôde sugerir o conteúdo a ser indicado, tendo como referência o maior valor dentre as somas, conforme pode ser visualizado na Tabela 2.

Tabela 2 - Indicação do material conforme a quantidade de erros por tema abordado no questionário.

	Aluno	Questões que formam o questionário												Conteúdo indicado
		Tema 1				Tema 2				Tema 3				
		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	
Acertos/Erros	A1	0	0	0	0	0	0	1	1	1	0	1	0	Referente ao tema 1
		Qtd. erros: 4				Qtd. erros: 2				Qtd. erros: 2				
	A2	1	1	1	1	0	0	1	1	1	1	1	1	Referente ao tema 2
		Qtd. erros: 0				Qtd. erros: 2				Qtd. erros: 0				

Fonte: Elaborada pelo autor.

A partir da aplicação do algoritmo descrito, criou-se uma base de dados coerente contendo as representações das respostas dos alunos e as sugestões indicadas pelo professor. Neste momento, foi extraída uma amostra desta base de conhecimento com o intuito de submetê-la ao KNN para que os resultados obtidos pudessem ser relacionados às indicações auferidas anteriormente, dando origem à acurácia.

Após a análise dos resultados obtidos, verificou-se que a base de dados criada aleatoriamente não foi capaz de representar o comportamento de pessoas reais, não gerando combinações de respostas relevantes ou que puderam ser classificadas pelo KNN, impactando diretamente na acurácia observada e aumentando a probabilidade de erros relacionados à interpretação dos resultados. Observou-se também que este comportamento ocorre devido à quantidade de combinações de respostas possível, maior que 16 milhões, o que torna improvável que uma base de dados gerada de forma aleatória possa representar o comportamento de um grupo limitado de pessoas.

Considerando os pontos de atenção descritos e o impacto causado nos resultados observados, fez-se necessária a submissão de um questionário composto de perguntas objetivas a um público real, dando origem a uma base de dados autêntica e possibilitando uma análise mais relevante.

3.2 Experiência realizada com uma base de dados real

Nesta segunda experiência, foi criado um modelo baseado em perguntas objetivas e materiais de estudo, no qual os alunos responderam os questionários disponibilizados pelo professor e, por meio da análise da combinação destas respostas, que o algoritmo pudesse classificar o melhor conteúdo possível para cada um destes discentes, tendo como base as dificuldades individuais estimadas.

Para isso, foi criado um questionário composto por doze perguntas objetivas, sendo cada uma delas com quatro alternativas e apenas uma opção correta. Estas perguntas foram distribuídas de forma proporcional aos temas abordados no questionário que, neste caso, foram três: pronomes pessoais, pronomes de tratamento e pronomes relativos. Assim, definiu-se três grupos formados por quatro perguntas e relacionados a cada um dos temas definidos para a avaliação.

Este questionário foi disponibilizado ao público geral por meio de um formulário on-line com o objetivo de reunir o maior número possível de respostas, para que estes dados fossem

utilizados como base para a aplicação do KNN e para atingir os objetivos gerais e específicos definidos por esta pesquisa.

Tendo em mãos as respostas coletadas, iniciou-se a avaliação dos resultados considerando o número de erros por tema. Assim, foi possível definir o melhor material de estudo para cada conjunto de alternativas selecionadas, contribuindo para que cada aluno tivesse uma indicação de material de estudo personalizada, com foco em suprir as suas necessidades particulares de aprendizagem.

Após a consolidação das respostas coletadas, a avaliação destes dados e a indicação do material de estudo, iniciou-se a normalização da base de dados, dos quais as alternativas selecionadas pelos participantes e o conteúdo indicado para consulta passaram a ser representados por um identificador numérico, possibilitando a aplicação do KNN. Conforme pode ser observado na Tabela 3.

Tabela 3 - Combinação de alternativas selecionadas por cada um dos alunos e o material de estudo indicado.

	Aluno	Questões que formam o questionário												Conteúdo indicado
		Pronomes Pessoais				Pronomes Tratamento				Pronomes Relativos				
		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	
Alternativas selecionadas	Aluno 1	52	53	60	62	65	70	73	79	82	86	89	93	5
	Aluno 2	50	53	60	62	68	70	73	77	83	87	91	94	5
	Aluno 3	51	53	60	64	68	71	76	80	84	88	90	95	6

Fonte: Elaborada pelo autor.

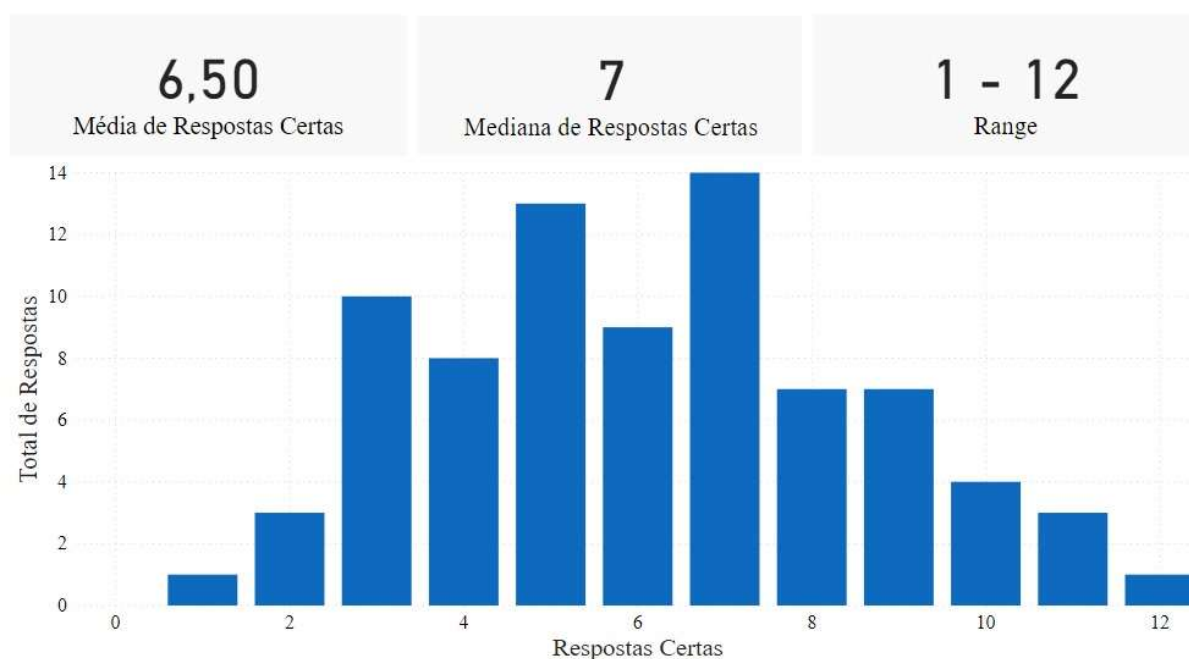
Ainda antes da aplicação do KNN, foi definido que os participantes que acertaram todas as respostas seriam retirados da base de dados, visto que nestes casos a indicação de um material de estudo não se faz necessária e que poderiam influenciar os resultados de forma negativa.

Após a obtenção das respostas, a consolidação da base, a normalização e a limpeza dos dados, foi possível verificar a relevância do KNN na indicação de materiais de estudo com base no perfil estimado de cada um dos participantes por meio da análise da acurácia obtida. Também se fez necessário validar a aplicação do modelo proposto, observando a acurácia alcançada por ele através de dois métodos distintos, foram eles: *K-Fold* e *Leave One Out* (LOO).

4 Resultados

A aplicação do questionário por meio de um formulário online, composto por perguntas objetivas e relacionadas a temas específicos, deu origem a uma base de respostas coerente, contendo um total de 80 tentativas distintas e distribuídas conforme mostrado na Figura 6. Aqui já se observa o fato de que as atuais ferramentas existentes para elaborar questionários, relacioná-los a assuntos específicos e disponibilizá-los aos alunos através da internet é uma realidade. No caso específico desta pesquisa, foi utilizada a plataforma Google para a execução desse processo.

Figura 6 - Distribuição da quantidade de acertos em função do total de pessoas.



Fonte: Elaborado pelo autor.

Com os resultados armazenados, criou-se uma base de dados organizada e normalizada, possibilitando a classificação e a indicação de materiais de estudo por meio da análise destas informações utilizando técnicas de Inteligência Artificial e algoritmos de *Machine Learning*, como o KNN. O *dataset* ajustado pode ser acessado em <https://github.com/fttec10/tcc-knn/blob/master/notebook/data/base-dados-normalizada.xlsx>.

Analisando a base normalizada, entendeu-se que a relação observada entre as alternativas selecionadas pelos participantes da pesquisa, os temas abordados pelas questões e os materiais de estudo sugeridos após a avaliação de suas respostas, mostrou-se relevante na identificação das dificuldades encontradas por cada indivíduo.

Esta base normalizada se apresenta ainda mais coerente quando combinada com o KNN, servindo como um conjunto de dados rotulados válido para aplicação deste algoritmo de classificação, sendo utilizado nas fases de treinamento e validação do modelo obtido, conforme mostra a Figura 7, podendo ser acessado em <https://github.com/fttec10/tcc-knn/blob/master/notebook/otimizacao-modelo.ipynb>.

Figura 7 - Divisão da base de dados entre treino e teste.

```
# Carregando Dataset
df = pd.read_csv('data/input-knn.csv',engine='python', sep=',')
df.head()

# Separando Alternativas selecionadas(X) e conteúdos sugeridos(y)
X = df.drop('Content',axis=1).values
y = df['Content'].values

# Normalizando os id's das alternativas
scaler = StandardScaler()
X = scaler.fit_transform(X)

# Dividindo a base em Treino(80%) e Teste(20%)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1, stratify=y)

# Definindo os parâmetros que serão utilizados pelo KNN
knn = KNeighborsClassifier(n_neighbors = 7)
knn.fit(X_train, y_train)

# Algoritmo final com os hyper-parâmetros otimizados
#KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
#                      metric_params=None, n_jobs=None, n_neighbors=7, p=2,
#                      weights='uniform')
```

Fonte: Elaborado pelo autor.

A utilização do método LOO em conjunto com a variação de K entre um e dez, nos permitiu determinar que sete é o melhor valor possível para K, além da acurácia obtida pelo modelo de 64,56%, conforme pode ser observado na Figura 8 e acessado em <https://github.com/fttec10/tcc-knn/blob/master/notebook/otimizacao-modelo.ipynb>.

Figura 8 - Aplicando o *Leave-One-Out* (LOO) para definir o melhor valor de K e a acurácia média.

```
neighbors = range(1,10,2)
train_accuracy_mean, test_accuracy_mean = [],[]
test_accuracy_max, best_k = 0,0

for i in neighbors:
    train_accuracy, test_accuracy = np.empty(len(X)), np.empty(len(X))

    loo = LeaveOneOut()
    for train, test in loo.split(X):
        X_train, X_test, y_train, y_test = X[train], X[test], y[train], y[test]

        knn = KNeighborsClassifier(n_neighbors=i)
        knn.fit(X_train,y_train)

        #Calculando acurácia para treinamento e teste
        train_accuracy[test], test_accuracy[test] = knn.score(X_train, y_train), knn.score(X_test, y_test)

    train_accuracy_mean.append(train_accuracy.mean())
    test_accuracy_mean.append(test_accuracy.mean())

    if (np.amax(test_accuracy_mean) > test_accuracy_max):
        best_k = i
        test_accuracy_max = test_accuracy.mean()

train_accuracy_mean, test_accuracy_mean = np.array(train_accuracy_mean), np.array(test_accuracy_mean)

print(f"Melhor K: {best_k} - Acurácia: {round(test_accuracy_max * 100, 2)}")
# output = Melhor K: 7 - Acurácia: 64.56
```

Fonte: Elaborado pelo autor.

Ao aumentar a quantidade de tentativas utilizadas pelo KNN de forma gradual, observou-se a acurácia alcançada pelo modelo desenhado em relação a predição de materiais de estudo para novas tentativas, distintas ou não das existentes, em relação ao volume de dados rotulados utilizados no processo, conforme mostra a Tabela 4 abaixo.

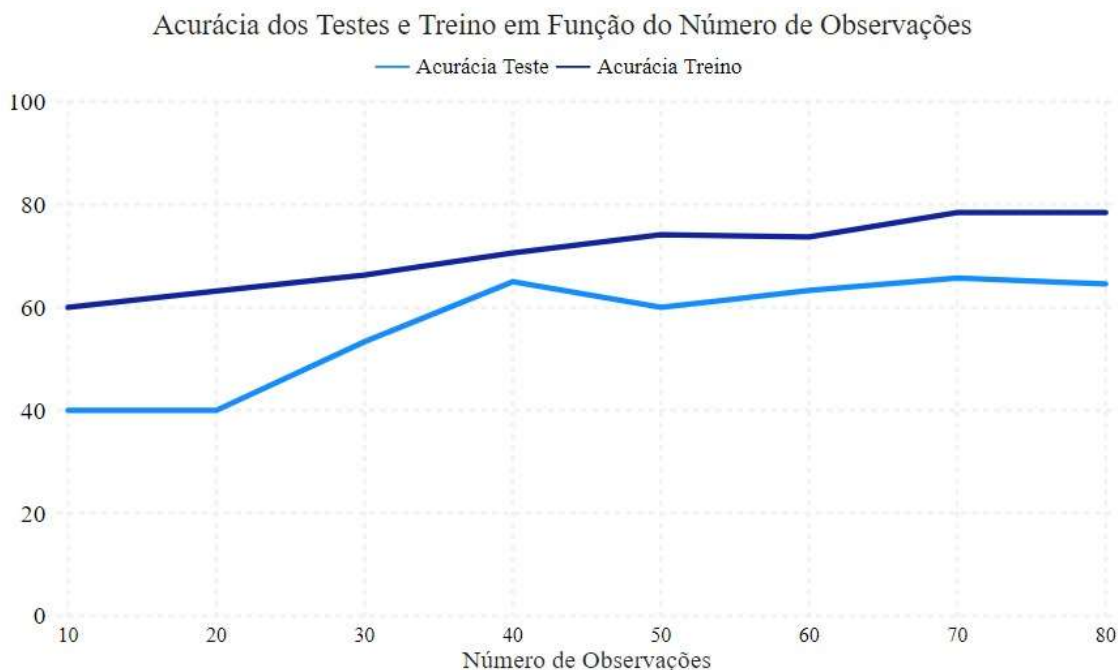
Tabela 4 - Relação entre a variação da quantidade de dados e a acurácia do KNN.

Base de dados rotulados (qtd. tentativas)	Acurácia Treino (%)	Acurácia Teste (%)
10	60,00	40,00
20	63,16	40,00
30	66,32	53,33
40	70,58	65,00
50	74,12	60,00
60	73,70	63,33
70	78,43	65,71
80	78,45	64,56

Fonte: Elaborada pelo autor.

Sendo assim, é possível observar que a utilização do KNN como algoritmo de classificação foi capaz de atingir uma média de 63,72% de acurácia com uma base de dados rotulados variando entre 40 e 80 respostas, o que sugere que não há necessidade de um grande conjunto de dados rotulados para que sua aplicação se torne relevante, podendo ser utilizado pelo professor em turmas relativamente pequenas, exigindo uma quantidade mínima de 40 respostas validadas e alcançando uma acurácia de até 65,71% na sugestão de materiais de estudo, como mostra a Figura 9 e pode ser visto em <https://github.com/fttec10/tcc-knn/blob/master/notebook/aplicacao-modelo.ipynb>.

Figura 9 - Acurácia do KNN em relação a quantidade de dados rotulados na base.



Fonte: Elaborado pelo autor.

Sob a luz dos resultados apresentados, é conveniente frisar que os objetivos propostos no início desta pesquisa foram atingidos. Ou seja, a proposta de utilização da tecnologia, da Inteligência Artificial, da análise de dados e dos sistemas de informação, de forma a auxiliar o professor a identificar as dificuldades de seus alunos de forma individualizada, simples e assertiva se mostrou relevante, originando uma interessante perspectiva para estudos futuros.

5 Considerações Finais

A partir da pesquisa realizada e dos resultados apresentados, verificou-se que o objetivo de utilizar o KNN para identificar as dificuldades individuais dos alunos, e indicar materiais de estudo ao professor, sugere ser uma possibilidade viável. Porém, é importante ressaltar que o tratamento prévio dos dados utilizados na aplicação do KNN tem influência direta nos resultados obtidos pelo modelo. Ressalta-se, neste caso, a organização e a filtragem destas informações para que o algoritmo possa ser utilizado de forma mais eficaz.

Ainda sobre a base de dados, observou-se que a utilização de informações que não são capazes de generalizar o comportamento real de um grupo pode comprometer a utilização da Inteligência Artificial, conforme mostra o primeiro experimento realizado.

Verificou-se também que a quantidade de respostas obtidas, além da aplicação de um único questionário e a utilização de apenas um algoritmo de classificação, possivelmente simplificaram os resultados e, por isso, são consideradas limitações desta pesquisa. Outro ponto de atenção é o fato do KNN ser capaz de sugerir apenas um material de estudo por tentativa, o que compromete a sua aplicação em casos em que o estudante necessite de uma quantidade maior de materiais por avaliação.

Além das limitações citadas, a necessidade de uma base de dados rotulada – comum aos algoritmos supervisionados – também pode se tornar uma barreira para utilização desta tecnologia, visto que exigirá a classificação e indicação de material de estudo por parte do professor nas primeiras tentativas.

Para estudos futuros, indica-se a validação desta proposta em parceria com instituições de ensino, disponibilizando avaliações relacionadas a temas diversos e dentro de um contexto escolar, observando o comportamento do KNN na classificação e na generalização de perfis em cenários complexos, os quais os dados de escolaridade, preferências e características pessoais, e a necessidade de adaptação do meio utilizado serão considerados.

Também se faz relevante implementar a proposta de classificação de perfis e indicação de materiais de estudo em um sistema no qual o professor possa realizar o cadastro de questionários, os alunos tenham acesso a estas avaliações e ainda consigam respondê-las de forma simples. Desta maneira, o professor poderia indicar materiais de estudo, que serviria como base para o funcionamento do KNN na identificação das dificuldades destes discentes e em futuras sugestões de conteúdo.

6 Referências

AHA, D.W.; KIBLER, D.; ALBERT, M.K. **Instance-based learning algorithms. Machine Learning**. 6, 1991, p.37-66.

ALMEIDA, F.; SIMOES, J. **The Role of Serious Games, Gamification and Industry 4.0 Tools in the Education 4.0 Paradigm**. CONTEMPORARY EDUCATIONAL TECHNOLOGY. 2019.

ANDRADE, K. **O desafio da Educação 4.0 nas escolas**. 2018. Disponível em: <https://canaltech.com.br/mercado/o-desafio-da-educacao-40-nas-escolas-109734/>. Acesso em: 19 nov. 2019.

BALSAN, L. **Método de avaliação utilizando Educação 4.0**. Olhares & Trilhas. Uberlândia, 2019.

BARCIA, R. M. et al. **A transformação do ensino através do uso da tecnologia da educação**. In: XIX Congresso Nacional da Sociedade Brasileira de Computação. Rio de Janeiro, PUC. Anais, 1999.

DE JESUS, J. S. **Educação 4.0: Uma proposta de aprendizagem para o futuro**. Bahia, Brasil, 2019.

DUCHESNE, Pierre; RÉMILLARD, Bruno. **Statistical Modeling and Analysis for Complex Data Problems**. Berlim, Alemanha: Springer, 2005.

FÜHR, R. C. **A Tecnopedagogia na esteira da Educação 4.0: Aprender a aprender na cultura digital**. V CONEDU, 2018, Olinda - PE. V Congresso Nacional de Educação - V CONEDU, 2018a.

FÜHR, R. C. **Educação 4.0 e seus impactos no século XXI**. In: V Congresso Nacional de Educação. V CONEDU, 2018, Olinda - PE. V Congresso Nacional de Educação - V CONEDU, 2018b.

GOMES, M. J. et al. **Aprender a Qualquer hora e em qualquer lugar, learning anytime anywhere**. Braga, Portugal, 2013.

LU, J. et al. **Recommender system application developments: A survey**. Decision Support Systems. v. 74, p. 12–32, (2015). ISSN 01679236.

MACFADYN, L.P.; DAWSON, S. **Mining LMS Data to Develop an "Early Warning System" for Educators: A Proof of Concept**. Computers & Education, 2010. no. 54, p.588-599.

MONARD, M. C.; BARANAUSKAS, J. A. **Conceitos Sobre Aprendizado de Máquina. Sistemas Inteligentes Fundamentos e Aplicações**. 1 ed. Barueri-SP: Manole Ltda, 2003.

PASSOS, U. R. C. et al. **Um Estudo Comparativo entre Técnicas de Inteligência Computacional para o Reconhecimento Ótico de Caracteres Manuscritos**. 2015. 12 f. TCC (Graduação) - Curso de Análise e Desenvolvimento de Sistemas, Ucam – Universidade Cândido Mendes, Porto de Galinhas, 2015.

PEDREGOSA, F. et al. **Scikit-learn: Machine learning in Python. Journal of machine learning research**. Disponível em: < https://scikit-learn.org/stable/modules/cross_validation.html >. Acessado em junho 2020.

RUSSELL, S.J.; NORVIG, P. **Inteligência Artificial: Tradução da Terceira Edição**. Rio de Janeiro: Elsevier, 2013.

SCHREIBER, J. N. C. et al. **Técnicas de Validação de Dados para Sistemas Inteligentes: Uma Abordagem do Software SDBAYES**. Universidade de Santa Cruz do Sul, Mar del Plata, Argentina, 2017.

SOUTO, M. C. P. et al. **Técnicas de aprendizado de máquina para problemas de biologia molecular**. Universidade de São Paulo. São Carlos, 2003.