

## **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 se considerarmos 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, essa pesquisa tem como objetivo propor uma forma de aplicação em que, a tecnologia, a inteligência artificial – especialmente do aprendizado de máquina supervisionado – a análise de dados e os sistemas de informação, possam auxiliar o professor a identificar as dificuldades dos alunos de forma individualizada, simples e assertiva. A partir de uma avaliação composta por questões objetivas e vinculadas a um tema específico, buscou-se aferir a acurácia alcançada pelo *K-Nearest-Neighbors* (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 considerada inicialmente pequena, podendo ser utilizado de forma rápida.

Introduzir KNN no trecho amarelo e só mencionar KNN depois

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

### **Abstract**

Technological evolution has traditionally been linked to industry and, as a result 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 propose a form of application in which, technology, artificial intelligence - especially supervised machine learning - data analysis and information systems, can help the teacher to identify students' difficulties in an individualized, simple and assertive way. From an evaluation composed of objective questions and linked to a specific topic, we sought to assess the accuracy achieved by the K-Nearest-Neighbors (KNN) in the classification of different

response patterns and, consequently, in the indication of the best study material possible considering the individual learning needs of each of the students. Thus, it was concluded that KNN appears to be a possible alternative, suggesting a good performance when applied to an initially small labeled database, which can be used quickly.

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

## 1 Introdução

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 é necessária por conta da 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.

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, e considerando não apenas o seu potencial, mas também as suas preferências pessoais e os formatos que mais lhe agradam.

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.

Sendo assim, este artigo tem como objetivo sugerir uma proposta de ferramenta que, por meio da inteligência artificial – em especial o Aprendizado de Máquina supervisionado – auxilie 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 Führ (2018b), algumas perguntas podem ser feitas para decifrar esse caminho, são elas: 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?

Segundo Führ (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 - onde é possível observar que cada uma delas foi acompanhada de mudanças significativas na maneira pelo 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 voltado para a importância de desenvolver novos projetos em grupo e utilizar novas tecnologias.

A educação 3.0 surge com a era da informação e da globalização, fazendo com que o ensino fosse adaptado às tecnologias emergentes e a novas possibilidades de ensino dentro e fora dos locais físicos 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.

## 2.1 Educação 4.0

Assim como na Educação 2.0, a atividade industrial tem grande influência a respeito da 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 se tornando 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 4.0 (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, adquirem rapidamente complexas habilidades técnicas e compartilham com os outros 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 que a ideia de que as escolas já possuem *softwares* para gestão, que computam notas, faltas e materiais de apoio onde os alunos conseguem realizar consultas. Assim, a aplicação da prova para avaliar o conhecimento do aluno, reduz a atividade do professor a corrigir e registrar a nota obtida neste sistema, que aceita essa informação apenas com o intuito de concluir ou reprovar a disciplina, sem nenhum critério que possa auxiliar o aluno na evolução do conhecimento.

### 2.1.1 Papel do Professor

O professor está presente em inúmeros papéis ao longo desse ciclo evolutivo. Na Educação 1.0 ele é tido como uma autoridade incontestada, na Educação 2.0 ele surge como fonte do conhecimento, e não apenas mais como uma figura autoritária, e na educação 3.0 começa a exercer o papel de mentor, tutor e facilitador na busca pelo aprendizado e conhecimento (GOMES et al., 2013).

A quarta Revolução Industrial permite a tecnologia transformar a experiência de aprendizagem, permitindo que os alunos se esforcem nos estudos porque se guiam pelo interesse, auxiliando o professor na tarefa de monitoria do processo ensino-aprendizagem, com dados mais detalhados, possibilitando uma melhoria nessa experiência de ensino e aprendizagem (ANDRADE, 2018).

### 2.1.2 Papel do Aluno

Até a Educação 2.0, os alunos, no geral, têm um papel passivo, onde observam e aprendem com o professor, exclusivamente. Também durante esse período, a autonomia do estudante é 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).

## 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. São listadas oito delas e agrupadas em duas dimensões as relacionadas ao pensamento, processos e raciocínio e as relativas à comportamento. A Tabela 1 mostra as definições divididas entre o desempenho humano e de racionalidade.

Pensando como um humano	Pensando racionalmente
<p>“O novo e interessante esforço para fazer os computadores pensarem (...) <i>máquinas com mentes</i>, no sentido total e literal.” (Haugeland, 1985)</p> <p>“[Automatização de] atividades que associamos ao pensamento humano, atividades como a tomada de decisões, a resolução de problemas, o aprendizado...” (Bellman, 1978)</p>	<p>“O estudo das faculdades mentais pelo uso de modelos computacionais.” (Charniak e McDermott, 1985)</p> <p>“O estudo das computações que tornam possível perceber, raciocinar e agir.” (Winston, 1992)</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)</p> <p>“O estudo de como os computadores podem fazer tarefas que hoje são melhor desempenhadas pelas pessoas.” (Rich and Knight, 1991)</p>	<p>“Inteligência Computacional é o estudo do projeto de agentes inteligentes.” (Poole <i>et al.</i>, 1998)</p> <p>“AI... está relacionada a um desempenho inteligente de artefatos.” (Nilsson, 1998)</p>

Tabela 1 - Algumas definições de inteligência artificial. Fonte: Russell e Norvig (2013).

Para Russell e Norvig (2013), historicamente, estratégias citadas anteriormente são seguidas até hoje para o estudo da IA, onde cada uma dessas abordagens são realizadas por pessoas e métodos diferentes.

Segundo Russell e Norvig (2013), o 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. A Figura 1 mostra esse esquema de forma simples. Um exemplo de um agente humano que tem visão, audição e entre outros órgãos que funcionam como sensores, e tem boca, mãos, pernas, como outras partes do corpo, que servem como atuadores. Já um robô é um agente que se utiliza de câmeras e detectores de diversos tipos que funcionando como sensores e motores que servem como atuadores.

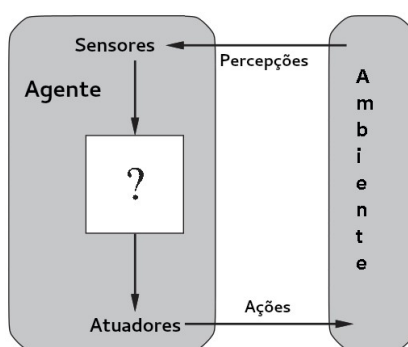


Figura 1 - Esquema de um agente e o ambiente. Fonte: Russell e Norvig (2013).

### 2.3 Aprendizado de Máquina (*Machine Learning*)

Aprendizado de Máquina (AM) é uma das várias áreas pertencentes à Inteligência Artificial, sendo um dos seus objetivos desenvolver técnicas computacionais que sejam capazes de aprender e adquirir conhecimento de forma automática. Um programa de computador que consegue tomar decisões baseando-se nas experiências anteriores ao solucionar problemas de forma bem-sucedida é dito 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 utilizado para o aprendizado utilizado 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, onde cada informação é fundada por atributos de entrada e saída (rótulos) corretos. Já o aprendizado não supervisionado, para cada informação, apenas os atributos de entrada estão disponíveis ao indutor, essa técnica é utilizada quando o objetivo for descobrir padrões ou tendências que auxiliem o entendimento desses dados.

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-los 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 NEIGHBOURS - KNN

Um dos aspectos importantes, relacionados às pesquisas, é desenvolver métodos capazes de prever o comportamento dos estudantes, de modo a possibilitar a intervenção de docentes, ou demais envolvidos, aspirando 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 dados do usuário, dos itens ou na relação entre ambos. (LU et al., 2015)

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

A Figura 2 mostra uma representação de algumas classes formadas pelos conjuntos de exemplos de treinamento e um novo dado desconhecido entre os grupos que deve ser classificado.

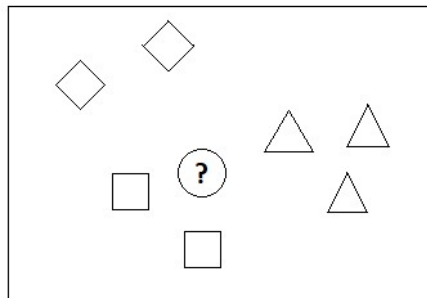


Figura 2 - Representação dos agrupamentos e um novo dado desconhecido.  
Fonte: Elaborada pelo autor.

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

Passos et al. (2015) diz que um ponto importante que tem que 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 muitos pontos inerentes à classes distintas. A Figura 3 mostra um exemplo de como o valor de  $k$  influencia na decisão de qual classe pertence o 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.

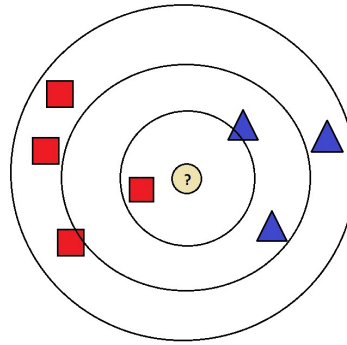


Figura 3 - Influência o valor de  $k$ . Fonte: Elaborada pelo autor.

O algoritmo KNN de classificação está inserido no campo dos algoritmos que realizam o aprendizado baseado em casos (*Case Based Learning*). Portanto, este algoritmo parte do princípio em que casos semelhantes deverão estar dentro de uma mesma classe, sendo esse princípio de continuidade comum a todos os algoritmos que utilizam o conceito de semelhança diretamente para determinar a classe. O KNN faz a busca a partir da semelhança dentre os pontos a serem analisados, utilizando uma função que calcula a distância, que é aplicada a pares de casos. Existem várias formas possíveis de definir funções de distância, um exemplo é a distância euclidiana, que é definida por  $\sqrt{\sum (a_i - b_i)^2}$ . Assim como a distância de *Manhattan* dada por  $\sum |a_i - b_i|$  onde, nos dois casos  $a_i$  e  $b_i$  equivalem os valores de  $i$  nos casos **a** e **b** (PASSOS et al., 2015).

## 2.5 K-fold

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, já que um modelo que apenas repetisse os rótulos das amostras que acabou de aprender teria uma pontuação perfeita, mas não conseguiria uma boa assertividade ao prever novos dados, já que o modelo não vai fazer uma boa generalização, essa situação é chamada de sobre ajuste. Para evitar esse problema, é uma prática comum, realizar um experimento de aprendizado de máquina supervisionado é deixar uma parte dos dados disponíveis para testes.

*k-fold* é um método computacional que utiliza as amostras da base de dados para o teste e também para o treinamento, atingindo resultados precisos (DUCHESNE; RÉMILLARD, 2005).

Onde o  $k$  indica a quantidade de partes que o conjunto de dados será dividido, para exemplificar, pode-se considerar uma base com cem amostras e  $k$  igual a dez, a base é dividida em dez partes, onde cada parte terá dez amostras. Então, uma das partes será utilizada para validação e as demais para treino (SCHREIBER et al., 2017). A complexidade computacional do  $k$ -fold é alta porque no primeiro passo ele vai utilizar a primeira parte para validação e as demais para treino, no segundo passo ele vai utilizar a parte dois para validação e as demais para treino, incluindo a primeira parte e assim por diante até que todas as amostras sejam uma das partes utilizadas no teste. O resultado é a média de desempenho do  $k$ 's testes (SCHREIBER et al., 2017).

**A Erro! Fonte de referência não encontrada.** abaixo mostra um outro exemplo de uma base de dados dividida em 5 partes e as iterações entre as partes, onde em cada iteração uma parte é usada como teste e as demais como treino e cada iteração é selecionada uma outra parte para teste ainda não utilizada.

	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

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

Fonte: Elaborada pelo autor.

## 2.6 Leave one out - LOO

Este método é um caso particular do *K-fold*, sendo o treinamento aplicado com  $n-1$  dados e o teste com apenas uma das amostras, portanto este método define o número de partes igual ao de amostras na base como exemplificado na **Erro! Fonte de referência não encontrada.**, sendo então, que uma base com cem amostras possuirá cem partes para testes, ou seja, como a mesma técnica do *K-fold*, porém como dito acima, o valor de  $k$  é igual à quantidade de dados (SCHREIBER et al., 2017).

	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

Figura 5 - Leave One Out em uma base com 100 partes.

Fonte: Elaborada pelo autor.

## 3 Metodologia

Tendo como foco testar as hipóteses de aplicação da inteligência artificial como ferramenta de 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 as dificuldades de cada aluno ao resultado apontado pelo algoritmo.

Para isso, foram seguidos dois processos distintos, 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 onde o 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, respondeu 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.

	Aluno	Questões quem 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				

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

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 as indicações auferidas anteriormente, dando origem a acurácia.

Após a análise dos resultados obtidos, verificou-se que a base de dados criada randomicamente 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 a interpretação dos resultados. Observou-se também que este comportamento ocorre devido à quantidade de combinações de respostas possível, maior que dezesseis 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, 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, onde 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.

	Aluno	Questões quem 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

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

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 execução desse processo.

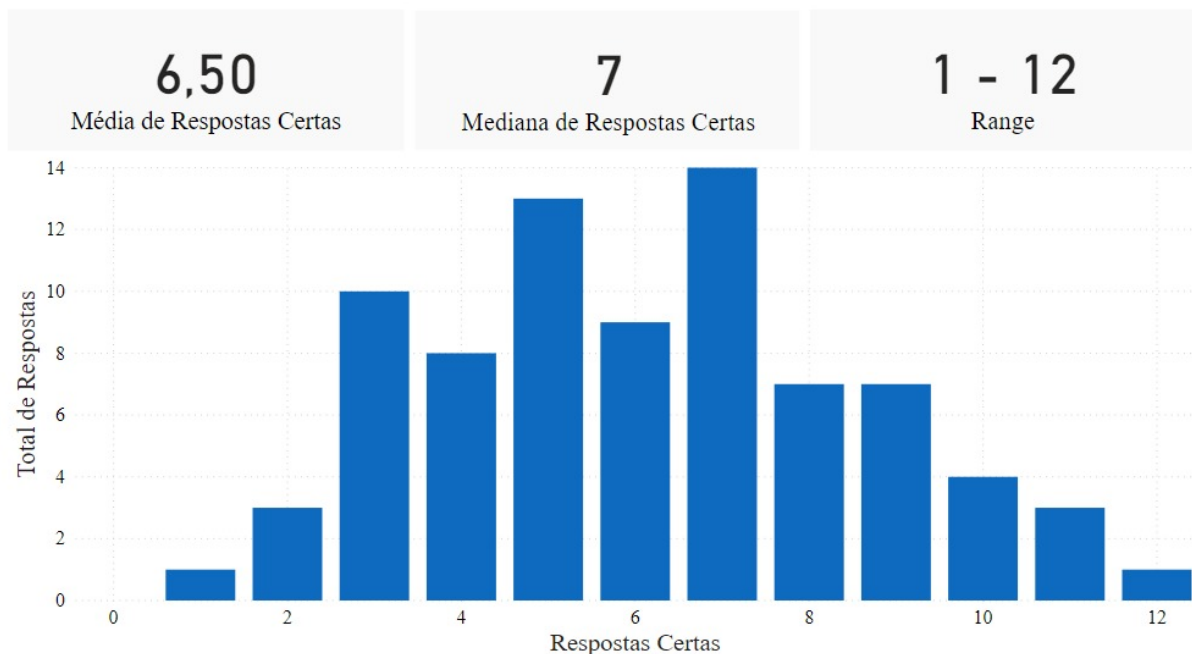


Figura 6 - Distribuição da quantidade de acertos em função do total de pessoas.

Fonte: Elaborado pelo autor.

Com os resultados armazenados, foi possível criar uma base de dados organizada e normalizada, que possibilitou 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, conforme pode ser acessada em

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 quando utilizada para identificar as dificuldades encontradas por cada indivíduo, fato este considerado como essencial para sugerir que o perfil destas pessoas pôde ser identificado e que o material de estudo foi recomendado considerando suas necessidades de forma individual.

Os resultados se apresentam ainda mais coerentes quando combinados com o KNN, servindo como uma base de dados rotulados válida para aplicação deste algoritmo de classificação, e que foi utilizada para as fases de treinamento e validação do modelo obtido, possibilitando-nos observar a acurácia alcançada pelo modelo desenhado em relação a

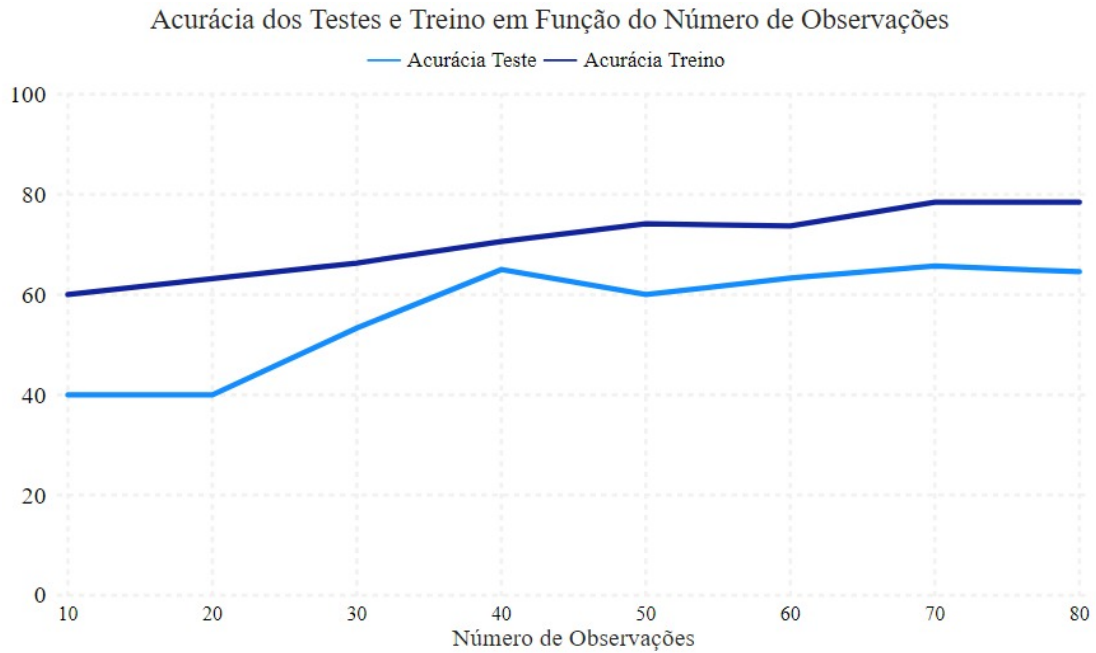
predição de materiais de estudo para novas tentativas, distintas ou não das existentes, além do impacto gerado pelo volume de dados rotulados utilizados neste processo, conforme mostra a Tabela 4.

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

*Tabela 4 – Tabela relacionando a variação da base com a acurácia do KNN.*

*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 5.



*Figura 7 – 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 de o KNN ser capaz de sugerir apenas um material de estudo por tentativa, o que compromete a sua aplicação em casos onde 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, onde os dados de escolaridade, preferências pessoais e a necessidade de adaptação do meio utilizado serão considerados, além dos casos onde os alunos apresentem algum tipo de deficiência física.

Também se faz relevante a implementação desta proposta de classificação de perfis e indicação de materiais de estudo em um sistema onde o professor possa realizar o cadastro de questionários, os alunos tenham acesso a estas avaliações e consigam respondê-las de forma simples. Assim, o professor poderá indicar materiais de estudo, que servirão 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](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.