Universidade do Estado do Rio de Janeiro
Centro de Tecnologia e Ciências
Instituto de Matemática e Estatística

Ighor Opiliar Mendes Rimes

Segurança Digital e Inteligência Artificial: Novos Caminhos para o Ensino de Matemática

Rio de Janeiro
2024

Ighor Opiliar Mendes Rimes

Segurança Digital e Inteligência Artificial: Novos Caminhos para o Ensino de Matemática

Tese apresentada, como requisito parcial para obtenção do título de Doutor, ao Programa de Pós-Graduação em Ciências Computacionais, da Universidade do Estado do Rio de Janeiro.

Orientadora: Prof[a]. Dra. Cristiane Oliveira de Faria
Orientador: Prof. Dr. Carlos Antonio de Moura

Rio de Janeiro
2024

CATALOGAÇÃO NA FONTE

UERJ/REDE SIRIUS/BIBLIOTECA CTC/A

R575 Rimes, Ighor Opiliar Mendes.
Segurança digital e inteligência artificial: novos caminhos para o ensino de matemática/ Ighor Opiliar Mendes Rimes – 2024.
112 f.: il.

Orientadora: Cristiane Oliveira de Faria
Coorientador: Carlos Antonio de Moura
Tese (Doutorado em Ciências Computacionais) - Universidade do Estado do Rio de Janeiro, Instituto de Matemática e Estatística.

1. Matemática - Estudo e ensino (Ensino superior) - Teses. 2. Modelagem matemática - Teses. 3. Segurança de dados - Teses. I. Faria, Cristiane Oliveira de. II. Moura, Carlos Antonio de. III. Título.

CDU 51:37

Patricia Bello Meijinhos CRB7/5217 - Bibliotecária responsável pela elaboração da ficha catalográfica

_____ _____

Assinatura Data

Ighor Opiliar Mendes Rimes

Segurança Digital e Inteligência Artificial: Novos Caminhos para o Ensino de Matemática

Tese apresentada, como requisito parcial para obtenção do título de Doutor, ao Programa de Pós-Graduação em Ciências Computacionais, da Universidade do Estado do Rio de Janeiro.

Aprovada em 19 de julho de 2024.
Banca Examinadora:

Profª. Dra. Cristiane Oliveira de Faria (Orientadora)
Instituto de Matemática e Estatística – UERJ

Prof. Dr. Carlos Antonio de Moura (Orientador)
Instituto de Matemática e Estatística – UERJ

Prof. Dr. Augusto Cesar de Castro Barbosa
Instituto de Matemática e Estatística – UERJ

Prof. Dr. Jayme Luiz Szwarcfiter
COPPE – PESC – Universidade Federal do Rio de Janeiro

Prof. Dr. José Abdalla Helayel Neto
Centro Brasileiro de Pesquisas Físicas

Prof. Dr. Luís Alfredo Vidal de Carvalho
Universidade Federal do ABC

Profª. Dra. Maria Clícia Stelling de Castro
Instituto de Matemática e Estatística – UERJ

Rio de Janeiro
2024

AGRADECIMENTOS

Ao concluir esta jornada de doutorado, gostaria de expressar minha sincera gratidão aos meus orientadores, professora Cristiane, professor Moura e professor Augusto. Este trabalho não teria sido possível sem a orientação, o apoio e o incentivo contínuos de cada um de vocês.

À professora Dr[a]. Cristiane, meu profundo agradecimento por usa orientação estratégica e apoio. Sua perspectiva e capacidade de ver o quadro geral foram essenciais para moldar a direção e o alcance desta tese. Agradeço por me desafiar a buscar novas abordagens para os problemas.

Ao professor Dr. Moura, sou grato pelo vasto conhecimento e pela forma como sempre esteve disponível para discutir ideias e oferecer conselhos. Sua experiência e contribuições enriqueceram significativamente este trabalho. Agradeço também pelo apoio, principalmente emocional, durante este período.

Ao professor Dr. Augusto, agradeço imensamente por sua dedicação, paciência e por sempre acreditar no meu potencial. Suas orientações e sugestões valiosas foram fundamentais para o desenvolvimento desta tese e seu bom humor durante às reuniões foram importantes em muitas ocasiões. Agradeço também, por mesmo não sendo creditado de forma oficial como orientador desta tese, possuir tanto crédito neste trabalho quanto os outros orientadores.

A cada um de vocês, minha gratidão, por seu apoio incondicional e por terem acreditado em mim em cada etapa deste caminho. Esta conquista é tanto de vocês quanto minha, e serei eternamente grato por tudo o que fizeram por mim.

Gostaria de expressar minha profunda gratidão também aos membros da banca examinadora, Professores Dr. Jayme, Dr. Helayel, Dr. Luís Alfredo e Dr[a]. Maria Clícia, por terem aceitado o convite para avaliar e discutir esta tese.

A todos vocês, meu mais profundo reconhecimento pelo tempo e esforço dedicados à leitura e avaliação desta tese. Suas contribuições são fundamentais para o desenvolvimento deste trabalho e para meu crescimento acadêmico e profissional. Sinto-me honrado e privilegiado por ter tido a oportunidade de contar com a expertise e o apoio de cada um de vocês.

Por fim, gostaria de agradecer à minha mãe, Joselene, mais uma vez. Mais uma vez, pois desde meu trabalho de especialização repito sempre o seu nome nesta seção. Obrigado mãe por todo suporte, financeiro, emocional e qualquer outro que me ofereceu durante toda a caminhada, não só acadêmica, a da vida. Mesmo a senhora não sabendo nada sobre títulos, me apoiou totalmente.

É impossível vencer alguém que nunca desiste.

Babe Ruth

RESUMO

RIMES, Ighor Opiliar Mendes. *Segurança Digital e Inteligência Artificial: Novos Caminhos para o Ensino de Matemática.* 2024. 112 f. Tese (Doutorado em Ciências Computacionais) – Instituto de Matemática e Estatística, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2024.

A homologação da Base Nacional Comum Curricular (BNCC), em 2017, introduziu disciplinas que conectam a teoria com dispositivos essenciais usados no dia a dia dos alunos, destacando-se a Educação Financeira, Ambientes de Aprendizagem e Tecnologias Digitais. Nesta tese, apresentamos propostas de conexões de conceitos matemáticos estudados no Ensino Fundamental imersos em novas tecnologias computacionais, como Segurança Digital e Inteligência Artificial. Dentro do escopo de Segurança Digital, trabalhamos o conceito de moeda digital através do *Bitcoin*, que já está bastante difundido. A partir desta motivação é possível explorar conceitos de modelagem matemática no ensino de funções. Para que a metodologia de modelagem matemática seja utilizada, inicialmente, aborda-se uma revisão sistemática sobre a tecnologia *Blockchain*, fundamental para diversas ferramentas tecnológicas atuais. Em seguida, são discutidos conceitos de criptografia e funções matemáticas (linear, quadrática, geométrica, exponencial e logarítmica) aplicadas ao ajustes de curvas do preço do Bitcoin. Utilizamos planilhas eletrônicas para análise de dados e o coeficiente de Pearson para avaliar o melhor ajuste, na tentativa de previsão do preço da moeda. Aproveitando o fato de que cada vez mais estamos envolvidos com aplicativos que utilizam Inteligência Artificial, em seguida, são vistos conceitos básicos necessários para entender os métodos não supervisionados de Aprendizagem de Máquina aliados à escolha de decisões que devem ser tomadas utilizando um linguajar matemático que pode ser aplicado em até turmas de ensino fundamental. A tese culmina em uma proposta de sequência didática que pode ser utilizada por docentes, adaptada para cursos de graduação em Matemática, incentivando a aplicação prática da modelagem matemática, e estimulando a compreensão dos alunos sobre criptoeconomia, técnica de análise de dados e Inteligência Artificial.

Palavras-chave: modelagem matemática; moedas digitais; aprendizado de máquina; ensino de matemática.

ABSTRACT

RIMES, Ighor Opiliar Mendes. *Digital Security and Artificial Intelligence: Some New Paths in Mathematics Teaching.* 2024. 112 f. Tese (Doutorado em Ciências Computacionais) – Instituto de Matemática e Estatística, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2024.

The approval of the National Common Curricular Base (BNCC) in 2017 introduced subjects that connect theory with essential devices used in students' daily lives, highlighting Financial Education, Learning Environments, and Digital Technologies. In this thesis, we present proposals for connections among mathematical concepts studied in elementary school and immersed in new computational technologies, such as digital security and artificial intelligence. Within the scope of Digital Security, we work on the concept of digital currency through Bitcoin, which is already quite widespread. Based on this motivation, it is possible to explore mathematical modeling concepts in teaching functions. For the mathematical modeling methodology to be used, a systematic review of Blockchain technology, fundamental to several current technological tools, is addressed. Next, cryptography concepts and mathematical functions (linear, quadratic, geometric, exponential, and logarithmic) applied to Bitcoin price curve adjustments are discussed. We use electronic spreadsheets for data analysis and the Pearson coefficient to evaluate the best fit to predict the currency's price. Taking advantage of the fact that we are increasingly involved with applications that use Artificial Intelligence, we then look at basic concepts necessary to understand unsupervised Machine Learning methods combined with the choice of decisions that must be made using a mathematical language that can be applied in even elementary school classes. The thesis culminates in a proposal for a didactic sequence that teachers can use and adapt for undergraduate mathematics courses, encouraging the practical application of mathematical modeling and stimulating students' understanding of cryptoeconomics, data analysis techniques, and artificial intelligence.

Keywords: mathematical modeling; digital coins; machine learning; mathematics teaching.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

BNCC	Base Nacional Comum Curricular
Coremec	Comitê de Regulação e Fiscalização dos Mercados Financeiros, de Capitais, de Seguros, de Previdência e Capitalização
ENEF	Estratégia Nacional de Educação Financeira
UERJ–FFP	Universidade do Estado do Rio de Janeiro – Faculdade de Formação de Professores
IA	Inteligência Artificial
SCLSWE	Aprendizagem Colaborativa e Avaliação do Trabalho Estudantil
BCT	Tecnologias *blockchain*
DES	Data Encryption Standard
IDEA	International Data Encryption Algorithm
AES	Advanced Encryption Standard
RC 4	Rivest Cipher
LSB	Least Significant Bit
MDC	Máximo Divisor Comum
RSA	Rivest, Shamir and Adleman
SVD	Decomposição em Valores Singulares
MMQ	Método dos Mínimos Quadrados
B3	Brasil, Bolsa e Balcão
CPF	Cadastro de Pessoas Físicas
AM	Aprendizado de Máquina
AMC	Aprendizado de Máquina Clássico
VP	Verdadeiros Positivos
VN	Verdadeiros Negativos
FP	Falsos Positivos
FN	Falsos Negativos

SUMÁRIO

INTRODUÇÃO

Em 20 de dezembro de 2017 houve a homologação da Base Nacional Comum Curricular (BNCC) pela Portaria 1570 (DIÁRIO..., 2017), e com ela o governo brasileiro escolheu seguir uma tendência global: a inserção de disciplinas em seu currículo escolar que aproximem mais a prática dos conceitos já abordados de forma teórica no Ensino Médio com situações reais do dia a dia dos alunos. Um dos exemplos é a inserção da disciplina de Educação Financeira, criada pelo Comitê de Regulação e Fiscalização dos Mercados Financeiros, de Capitais, de Seguros, de Previdência e Capitalização (Coremec), por meio da proposta intitulada Estratégia Nacional de Educação Financeira (ENEF).

Pensar em Educação Financeira é compreender os meios para que se tenham economias mais sustentáveis, evitar gastos desnecessários, poupar dinheiro para algum objetivo e, atualmente com a facilidade dos aplicativos de corretoras, também pensar em investimentos. Nos últimos anos, dentre os investimentos mais comentados temos as criptomoedas e, em especial, o Bitcoin (SPOLADOR, 2017).

Antes do entendimento dessa moeda, existe a necessidade de compreendermos o que é de fato uma criptomoeda. Estas são assim denominadas porque sua viabilização ocorre a partir de métodos criptográficos, ou seja, de um conjunto de técnicas que permitem proteger dados transmitidos e armazenados, a partir da transformação de informações legíveis em códigos ininteligíveis (CARVALHO et al., 2017). Estes códigos são viabilizados por criptografia. Já a criptomoeda utiliza conceitos de informação imutável e para que seja criada esta se utiliza de *Blockchain* que permite a imutabilidade. Atualmente, existem diversos setores que se utilizam de criptografia além das criptomoedas, como as mensagens de *Whatsapp* e as senhas de banco.

Já o Bitcoin é uma moeda digital *peer-to-peer* (par a par, ou simplesmente, de ponto a ponto), descentralizada, de código aberto. O que significa que não depende de uma autoridade central, a transação é feita diretamente entre os interessados (ULRICH, 2017). Ainda de acordo com Ulrich, um dos fatos que tornam o Bitcoin único, é ele possuir todas as melhores características do melhor dinheiro, sendo escasso, divisível, portátil, mas vai, inclusive, na direção do ideal monetário, por ser ao mesmo tempo "sem peso e sem espaço" – é incorpóreo. Isso impossibilita a transferência de propriedade a despeito da geografia, a um custo virtualmente nulo e sem depender de um intermediário, contornando, dessa forma, todo o sistema bancário, completamente subvertido pela intervenção governamental. Isso torna a moeda especialmente interessante para o investidor que pensa em retorno financeiro rápido, devido às flutuações do preço dessa moeda e também pode ser um elemento para que os alunos comecem a compreender o poder da criptoeconomia. Aproveitando este tema, poderia utilizá-lo também em sala de aula como uma proposta de educação financeira mais atual do que aquelas atualmente trabalhadas as quais por

vezes não atraem os alunos do Ensino Médio, por utilizar situações fora de seus contextos sociais ou que não lhes interessam até o momento da aquisição, como imóveis e moradia.

Outro conteúdo, na atualidade, muito discutido é como deve ser apresentada ao ciclo básico o conceito de Funções. Sierpinska (1992) afirma que, pelo menos desde o início do século XX, o conceito de função foi considerado como um dos fundamentais da Matemática. Importante não somente para os alunos que desejam seguir a área de Ciências Exatas, pois, o conteúdo se torna extremamente necessário para modelar qualquer tipo de fenômeno. No entanto, o estudo deste tópico no Ensino Médio brasileiro segue ainda uma ordenação tradicional. Os temas são geralmente tratados de forma independente e sem conexão alguma entre eles (BARRETO, 2007). Este modelo vai em sentido oposto às ideias matemáticas, que são resultado de um processo que procura explicar e compreender fatos e fenômenos observados na realidade (D'AMBROSIO, 1999). O desenvolvimento dessas ideias e sua organização intelectual dão-se a partir de elaborações de suas representações da realidade. Tais representações constituem o que se costuma chamar de "modelos matemáticos", cuja obtenção, aplicação e avaliação compõem a modelagem matemática. Em geral, são apresentados cinco argumentos para a inclusão de Modelagem no currículo (BASSANEZI, 1994): motivação, facilitação da aprendizagem, preparação para utilizar a matemática em diferentes áreas, desenvolvimento de habilidades gerais de exploração e compreensão do papel sócio-cultural da matemática. Além disso, com a modelagem nesse conteúdo, pode-se trabalhar a interdisciplinaridade em qualquer outra área do conhecimento; os conceitos fundamentais da estatística como frequências, médias, e ainda utilizar ferramentas como planilhas eletrônicas, importantes para o aluno que vive cercado por um mundo que exige dele maior domínio sobre recursos tecnológicos.

Motivado pelos modelos matemáticos encontrados, uma outra dúvida surge: como os computadores utilizam modelagens para tomar decisões? Segundo Carvalho (2005) o conceito de inteligência é explorado basicamente desde que o ser humano se reconhece como ser pensante, então é isto que ainda nos faz humanos? É de conhecimento da sociedade que os processos realizados pela computação envolvem conceitos matemáticos, mas que conceitos são esses? Afinal, quais são os caminhos que a máquina utiliza para desenvolver seus métodos? De acordo com (SAYAD, 2023) a inteligência artificial não é subjetiva, nem neutra. Os sistemas são desenvolvidos por seres humanos, portanto grande parte das implicações éticas podem ser mitigadas na base de dados à qual estão expostos.

Apoiado nesses argumentos, aliados a essa nova forma de se pensar em moeda, este trabalho desenvolve um estudo sobre a criptografia e criptoeconomia, passando por uma modelagem matemática com funções sobre o preço da moeda Bitcoin, que pode ser o incentivo inicial para se compreender esses temas de forma mais didática e contextualizada no universo em que o aluno está inserido. Com isso, este trabalho pretende desenvolver em seu primeiro capítulo uma revisão sistemática sobre tudo que vem sendo estudado sobre *Blockchain*. No segundo, o contexto histórico e atual da criptografia com exemplos

práticos e didáticos de suas aplicações. O Capítulo três aborda um maior conhecimento sobre como funciona a ideia das criptomoedas e como elas podem modificar o modelo econômico que já conhecemos. Apresenta também a modelagem matemática focada na área educacional e criam-se ajustes de curvas com diferentes funções sobre o preço da moeda Bitcoin. No quarto capítulo, utilizam-se conceitos fundamentais de Aprendizado de Máquina para se compreender melhor a modelagem baseada em funções.

O advento da Inteligência Artificial é um marco evolutivo da humanidade e é impossível negligenciá-lo na formação educacional brasileira (FREITAS, 2020). Por isso, esta tese também propõe uma sequência didática para cada capítulo, a partir do Capítulo 2, a qual teve como ementa os cursos de graduação de licenciatura em matemática da UERJ–FFP. Porém, como muitas das disciplinas são vistas em outros cursos, o mesmo pode ser adaptado com facilidade para algumas disciplinas específicas em cursos de exatas.

Apesar de ser apresentada esta tese na sequência citada acima, vale ressaltar que esta não foi a ordem orgânica na qual o trabalho foi desenvolvido. Iniciamos, como motivação, com a proposta de descobrir qual seria o preço da moeda *Bitcoin* no futuro para um público de nível médio e início de graduação; e com o aprofundamento do estudo sobre o tema, vimos a necessidade de compreender o que seria o *blockchain* (Capítulo 3). Quando nos debruçamos sobre esse conteúdo, nos deparamos com as diversas funcionalidades que esta tecnologia já cria, muito além de um modelo econômico, e sentimos a necessidade de produzir uma revisão sistemática para compreendermos melhor o poder de tal tecnologia (Capítulo 1). Ainda no estudo do *blockchain*, notamos que seu desenvolvimento é todo apoiado em criptografia e sentimos assim a necessidade de revisitar e explorar conceitos matemáticos, relacionando-os com as ciências computacionais (Capítulo 2). Ao fim dessa pesquisa, nos invadiu a dúvida de que se muitas decisões são tomadas por inteligência artificial, quais são os caminhos que tornam viável essa tal inteligência existir (Capítulo 4). Com tudo isso, é possível ler esta tese também na ordem em que foi concebida, sem nenhuma perda de compreensão do assunto. Inclusive, sugerimos que seja feita desta forma, pois assim é possível nos acompanhar por uma trilha didática até mais interessante, visto que, ao fim de cada capítulo, também é despertada a mesma curiosidade que nos motivou a escrever o próximo, e assim seguimos um caminho que condiz com o método científico, negligenciado em alguns momentos durante o aprendizado.

1 REVISÃO SISTEMÁTICA

Iniciamos com a revisão sistemática de trabalhos sobre o tema de Blockchain na área de educação. A revisão se baseia no trabalho de (ALAMMARY et al., 2019) e nas orientações fornecidas por (OKOLI; KIRA, 2010), enumeradas abaixo.

1. Identificação do propósito e questionamento das pesquisas. Ter esse processo bem claro auxilia em encontrar artigos de maior relevância com maior facilidade.

2. Redação de protocolo detalhado para a revisão. Um protocolo é um plano que detalha as etapas e procedimentos específicos a serem seguidos na revisão.

3. Busca de artigos relevantes. Atualmente, os recursos eletrônicos são a fonte predominante de procura literária.

4. Triagem de artigos para inclusão. Nesta etapa, os revisores decidem quais artigos devem ser considerados para a revisão e quais devem ser eliminados. Eles também precisam declarar por quais razões práticas deve-se excluir cada artigo.

5. Avaliação da qualidade dos artigos. Nesta etapa, os revisores precisam determinar quais artigos são de qualidade suficiente para serem incluídos na revisão sistemática. Esta etapa serve a dois propósitos. Primeiro, em revisões sistemáticas onde há um padrão mínimo de qualidade para inclusão, a qualidade é empregada para excluir artigos que não atendem ao padrão dos revisores. Em segundo lugar, em toda revisão sistemática, é preciso haver algum tipo de avaliação de qualidade, pois a qualidade da revisão depende disso.

6. Extração de dados de artigos. Após identificar todos os artigos que serão incluídos na revisão, os revisores precisam extrair sistematicamente os dados apropriados de cada artigo. Esses dados devem servir como matéria-prima para a etapa de síntese. O tipo de dado a ser extraído é determinado com base nas questões de pesquisa estabelecidas durante o estágio inicial da revisão.

7. Análise dos dados extraídos. Também conhecida como síntese de dados, esta etapa envolve agregar, organizar, comparar e discutir os fatos extraídos dos artigos. O procedimento envolvido nesta etapa depende de serem os artigos incluídos qualitativos, quantitativos ou mistos. Estudos qualitativos, quantitativos e mistos podem ser analisados qualitativamente, enquanto apenas estudos quantitativos podem ser analisados quantitativamente.

8. Redação da revisão sistemática. Nesta etapa, os principais padrões de pesquisa devem ser seguidos. A revisão deve ser relatada com detalhes suficientes para que seu resultado possa ser reproduzido independentemente.

As seções a seguir descrevem como estes oito passos foram implementados neste trabalho.

1.1 **Definição das perguntas para a pesquisa**

Com base no objetivo deste trabalho, as seguintes perguntas foram formuladas.

1. Quais produtos foram desenvolvidos com a tecnologia *blockchain* para fins educacionais?
2. Quais benefícios essa tecnologia pode trazer para a educação?
3. Quais são os desafios que surgem com a adoção dessa tecnologia na educação?

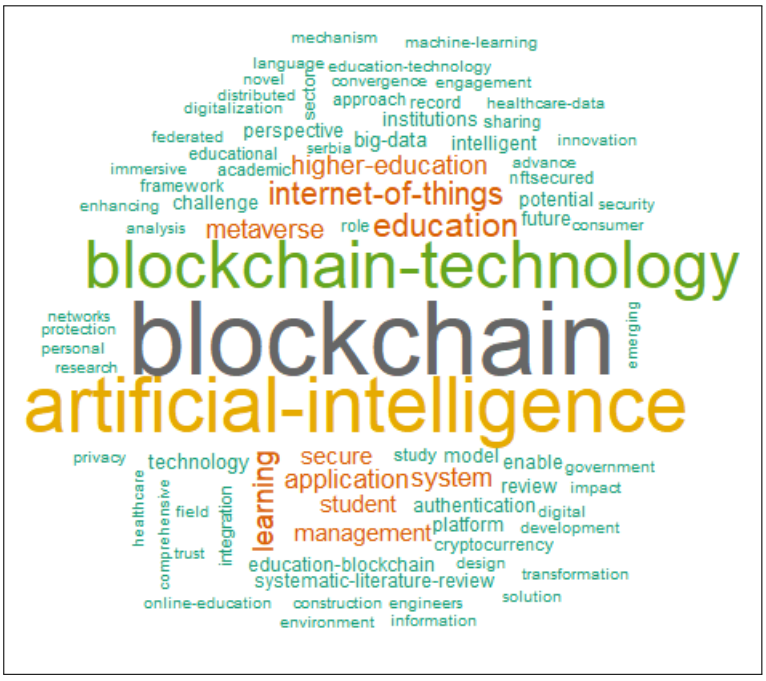
1.2 **Busca por artigos relevantes**

Para reunir artigos relevantes, foi decidido utilizar o Google Acadêmico e se basear nas recomendações obtidas. Entre essas recomendações, algumas bases de dados científicas encontradas foram: IEEE Xplore, Springer, MDPI, entre jornais e revistas de determinados países. É importante salientar aqui que só foram incluídos artigos e livros publicados a partir do ano de 2022 e que essa escolha se deve ao fato de que o tema envolve a área de tecnologia, portanto, se infere serem os resultados sujeitos a uma atualização fortemente dinâmica e quanto mais atual os trabalhos, melhor é a compreensão de avanços dos estudos na área. A busca ocorreu no dia sete de abril de 2023. A escolha de um dia específico se deve principalmente ao critério de artigos mais relevantes encontrados. Caso houvesse uma busca mais duradoura, a ordem de relevância poderia ter alterações.

Os termos utilizados na busca foram inicialmente: “Blockchain + education”, algo que retornou aproximadamente 133.000 trabalhos; o termo “Artificial Intelligence + education” possui cerca de 3.510.000 artigos. Por fim, decidimos unir as três palavras-chave “Blockchain + Artificial Intelligence + education”, acreditando ter mais precisão na busca de trabalhos que possuem o mesmo escopo de nosso interesse e assim encontramos o total de 72.300, cerca de 15.400 só a partir de 2022. Foi obtida também, com o auxílio de um algoritmo de mineração de textos, que cria nuvem de palavras se utilizando de métodos Fuzzy, o RStudio (CRUZ; LANZILLOTTI, 2020) e, com a colaboração da aluna de dou-

torado Carla Passos, uma nuvem de palavras mais buscadas sobre o tema na Internet (Figura 1).

Figura 1 - Nuvem de palavras na área de Blockchain.



Fonte: O autor, 2023.

1.3 **Inclusão e exclusão de artigos**

Após o término da busca, foi realizada a triagem dos títulos e resumos dos artigos usando critérios de inclusão e exclusão pré-definidos. Um artigo foi excluído se: (1) o tema não estivesse relacionado à educação e tecnologia Blockchain, (2) o texto completo não estivesse disponível *on-line* de forma gratuita, (3) o texto não estivesse entre os duzentos mais recomendados pela plataforma e (4) o artigo não tivesse sido publicado por algum recurso acadêmico.

Por fim, o texto integral de cada artigo foi lido para garantir que continha todos os dados necessários para esta revisão sistemática.

Nenhuma avaliação de qualidade foi conduzida. A qualidade desta revisão foi assegurada pelo uso de artigos revisados por editoras respeitadas, e assim incluímos somente artigos supostamente de boa qualidade.

1.4 **Extração de dados**

Um formulário de extração de dados foi usado para recolher informações dos estudos conduzidos. O formulário foi projetado especificamente para esta revisão e continha os itens mostrados na tabela 1.

Depois de extrair os dados dos artigos, foi realizada a análise dos dados. Esses dados foram agrupados em cinco temas principais que foram considerados importantes para a pesquisa.

Tabela 1 - Itens da extração de dados.

Item	Descrição
Título	Título do artigo
Autor(es)	O nome dos autores
Data	Ano de publicação
Países	Países dos autores
Objetivo do artigo	O objetivo do artigo conforme declarado pelos autores
Implementação dos artigos	Resumo de como o artigo foi implementado (caso haja)

Fonte: O autor, 2023.

Depois de extrair os dados dos artigos, foi realizada a análise dos dados. Os dados extraídos foram analisados usando dois temas principais pré-determinados, que emergiram das questões de pesquisa – objetivo e implementação do artigo. Para cada um desses principais temas, diversos subtemas emergiram da análise dos dados.

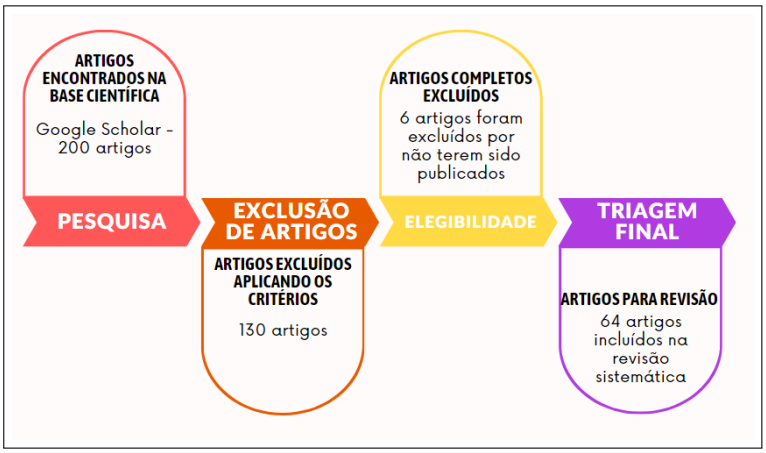
1.5 **Resultados**

Um total de 200 artigos foram recuperados das bases de dados científicas. A triagem inicial desses artigos, que foi baseado em título e resumo, resultou na exclusão de 130 artigos. A grande maioria desses artigos foi excluída devido à sua abrangência, ou seja, os artigos não apresentavam aplicação de tecnologia *blockchain* na educação. Outros foram excluídos porque não havia o texto de leitura completo; seis dos artigos foram excluídos por não terem sido publicados em nenhum meio acadêmico (figura 2).

A figura 3 mostra a distribuição geográfica dos artigos incluídos. China, Estados Unidos, Índia e Indonésia foram os que mais contribuíram no levantamento, mas há contribuição de todos os continentes.

A distribuição dos artigos por fontes de publicação é mostrado na figura 4. A grande maioria desses artigos (72,8%) foi publicada em jornais/revistas, 18,5% foram publicados em conferências e o restante (8,7%) se distribuiu entre livros, workshops e colóquios.

Figura 2 - Fluxograma do processso de revisão sistemática.

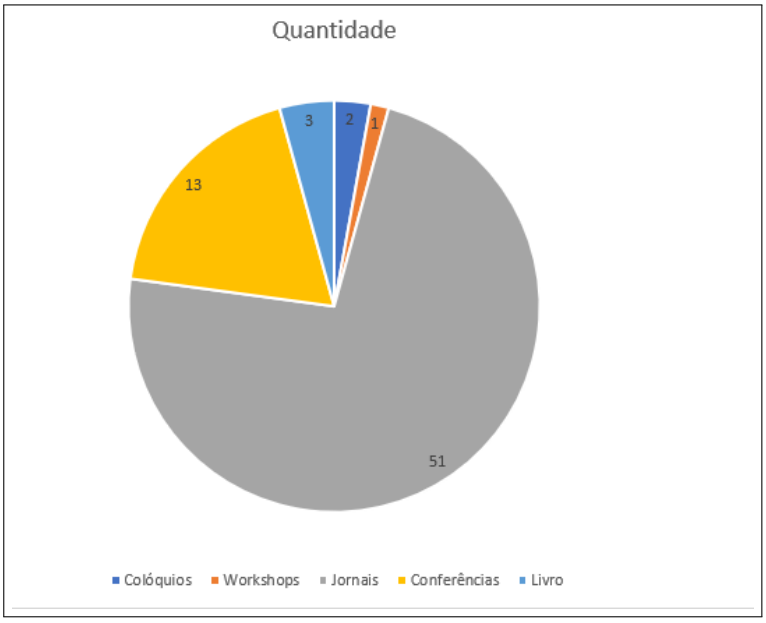


Fonte: O autor, 2023.

Figura 3 - Distribuição geográfica dos artigos incluídos.

Fonte: O autor, 2023.

Figura 4 - Distribuição de artigos por fontes de publicação.



Fonte: O autor, 2023.

1.6 **Aplicações de Blockchain**

Como mostrado na tabela 2, as aplicações foram classificadas em quatro categorias: validação de informações acadêmicas com *blockchain*, melhora a compreensão atual das aplicações de *blockchain*, utilização da inteligência artificial e *blockchain* na área de ensino, *blockchain* como ferramenta para classificar a qualidade de ensino e *blockchain* para a melhora da segurança.

Tabela 2 - Aplicações identificadas nos artigos revisados.

Categoria das aplicações	Artigos
Validação de informações acadêmicas ou projetos com Blockchain	(ANWAR et al., 2022); (ZHAO; DI; HE, 2022); (MAINETTI et al., 2022); (SHAIKH et al., 2022); (TALPUR; TALPUR; HASEEB, 2022); (MEIROBIE et al., 2022); (MASADEH, 2022); (MUHATI; RAWAT; SADLER, 2022); (SHAHZAD; ASEERI; SHAH, 2022); (LEVITSKAYA; POKROVSKAIA; RODIONOVA, 2022); (GIUSEPPE, 2022); (ZHANG; GOYAL, 2022); (KIM; JANG, 2022); (YANG; WANG, 2022); (YING, 2022); (KUMAR et al., 2022); (BJELOBABA et al., 2022); (ZHENG, 2022); (WANG et al., 2022); (MUSTI; KANT; KHANNA, 2022)

Melhora da compreensão atual das aplicações Blockchain	(KULETO et al., 2022), (SAVELYEVA; PARK, 2022); (CHEN, 2022a); (MOORE; FELO, 2022); (NEMORIN et al., 2023); (BHUTORIA, 2022); (LI et al., 2023); (LIU; FAN; QI, 2022); (THOMASON, 2022); (KHAN et al., 2022); (CHAKA, 2023); (NOKITI et al., 2022); (OCHEJA et al., 2022); (SUNNY et al., 2022); (E., 2022); (FENG; XU; WEYMOUTH, 2022); (ZHAO et al., 2023); (HANNAN, 2022); (SUPRIATI et al., 2022)
Utilização da IA e Blockchain na área de ensino	(SOUZA, 2020); (FERDIG et al., 2022); (BUCEA-MANEA-ŢONIş et al., 2022); (CHOI; CHOI; PARK, 2022); (VESELOV et al., 2022); (THUAN et al., 2022); (CHINYAMUNJIKO; C.; BHIBHI, 2022); (MOZUMDER et al., 2023); (MOURTZIS; ANGELOPOULOS; PANOPOULOS, 2022); (HWANG; CHIEN, 2022)
Blockchain para a classificação da qualidade de ensino	(GARG et al., 2021); (WANG; SUN; BIE, 2022); (CHIVU et al., 2022); (NUROVIC; POTURAK, 2021); (PANAGIOTIDIS, 2022); (LI, 2022a); (MORA-ZAMBRANO et al., 2022); (MIN; BIN, 2022); (KHOLISHOTULAILA; LAILA; ANGGA, 2022); (CHEN, 2022b); (KOMATH; O., 2022); (LI, 2022b); (WANG, 2022); (AL-MASKARI; T.; S., 2022); (KOSASI et al., 2022)

Fonte: O Autor, 2023.

Alguns desafios foram levantados pelos artigos consultados. Um deles foi a discussão sobre diferentes tipos de segurança e privacidade que podem ser experimentadas utilizando a tecnologia *blockchain*, como ataques maliciosos e vazamento de dados. Com isso, algumas instituições ainda estão relutantes em compartilhar seus dados utilizando a tecnologia *blockchain*. Há também uma discussão sobre quanto custa adotar essa tecnologia na área de educação.

De acordo com alguns artigos, a tecnologia poderia permitir que os alunos agissem como seus próprios registradores de conquistas educacionais e que poderia minar ou modificar o papel central das instituições educacionais como agentes de certificação. Isso é visto, principalmente, nos trabalhos que criam cursos *on-line* autossustentáveis, que são capazes de melhorar sua capacidade didática pela interação com o usuário.

A primeira categoria na tabela 2 foca em algumas aplicações relacionadas ao gerenciamento de certificados. Esta categoria diz respeito ao tratamento de todas as formas de credenciais acadêmicas, transcrições, certificados dos alunos, ou quaisquer outras formas de registros. No campo da educação, muitos aplicativos utilizam *blockchain* para emissão

de certificados. Um exemplo de artigo nessa categoria é o intitulado "A model for secure inter-institutional communication based on artificial intelligence and blockchain" (TALPUR; TALPUR; HASEEB, 2022), que propõe um modelo e arquitetura para validação de informações digitais entre universidades.

A segunda categoria concentrou-se na compreensão atual das aplicações do *blockchain* de forma geral. Foi dada maior atenção à construção de tal tecnologia e quais processos e mudanças essa tecnologia poderia realizar em áreas não relacionadas obrigatoriamente à educação, mas possuindo algum ponto de interseção com o tema. Um exemplo de artigo nessa categoria é o intitulado "The potential of blockchain technology in higher education as perceived by students in Serbia, Romania and Portugal" (KULETO et al., 2022). Este artigo investiga uma alternativa em que as instituições de ensino superior incluem uma rede *blockchain* para fornecer o melhor sistema educacional sustentável.

Já a terceira categoria aborda as aplicações envolvidas entre a utilização da inteligência artificial e o *blockchain*. Dentro dessa área existem diversos assuntos abordados, como classificação de testes, desenvolvimento de professores sobre o conhecimento de inteligência artificial, entre outros. Podemos citar, para elucidar melhor, o artigo "Blockchain-Centered Educational Program Embodies and Advances 2030 Sustainable Development Goals" (CHOI; CHOI; PARK, 2022), que desenvolve um plano de programa educacional centrado em *blockchain* utilizando técnicas de gamificação com inteligência artificial.

A quarta categoria busca principalmente artigos que propõem soluções inteligentes utilizando *blockchain* para melhorar a qualidade dos cursos principalmente à distância. A principal temática dessa categoria é buscar que os cursos geridos nesse modelo *on-line* sejam melhorados de forma sustentável. Isso é visto, por exemplo, no artigo "Blockchain in Education – The Case of Language Learning" (PANAGIOTIDIS, 2022), que analisa características e vantagens das aplicações em *blockchain* no setor educacional aplicadas à aprendizagem de línguas.

É importante salientar que, de todos os artigos coletados para esta revisão, apenas 12,5% abordam de alguma forma o Ensino Básico (tabela 3). Todo o restante se concentra em abordar a educação, ou seus processos educacionais, visando diretamente o público de graduação ou acima. Dentre eles, é importante destacar (CHEN, 2022b), que visa dimensionar os impactos que a utilização de *blockchain* e inteligência artificial associada à utilização de óculos de realidade virtual podem ter no ensino de Química. O laboratório intitulado Smart Cloud Lab foi utilizado para um experimento semestral e utilizou as notas dos alunos bem como resultados de questionários e entrevistas antes e depois da implementação desse modelo de ensino e o estudo mostrou bons resultados. Outro artigo que merece destaque é (WANG, 2022), que analisa os efeitos positivos de utilizar IA e *blockchain* na qualidade de ensino da educação da China, a partir de aspectos de *design*, construção de ambiente educacional inteligente, construções de recursos educacionais de alta qualidade, entre outros. O trabalho utiliza a análise de regressão para testar benefícios

dessa política. O restante dos artigos voltados a esse público, se concentram em realizar análises reflexivas sobre os prós e contras de se utilizar a tecnologia na geração de ensino atual e discute maneiras de implementação.

Tabela 3 - Artigos classificados por nível de escolaridade.

Nível de escolaridade	Artigos
Graduação e Pós-Graduação	(ANWAR et al., 2022); (MAINETTI et al., 2022); (SHAIKH et al., 2022); (TALPUR; TALPUR; HASEEB, 2022); (MEIROBIE et al., 2022); (MASADEH, 2022); (MUHATI; RAWAT; SADLER, 2022); (SHAHZAD; ASEERI; SHAH, 2022); (LEVITSKAYA; POKROVSKAIA; RODIONOVA, 2022); (GIUSEPPE, 2022); (ZHANG; GOYAL, 2022); (KIM; JANG, 2022); (YANG; WANG, 2022); (YING, 2022); (KUMAR et al., 2022); (BJELOBABA et al., 2022); (ZHENG, 2022); (MUSTI; KANT; KHANNA, 2022) (KULETO et al., 2022), (SAVELYEVA; PARK, 2022); (MOORE; FELO, 2022); (NEMORIN et al., 2023); (BHUTORIA, 2022); (LIU; FAN; QI, 2022); (THOMASON, 2022); (KHAN et al., 2022); (CHAKA, 2023); (OCHEJA et al., 2022); (SUNNY et al., 2022); (E., 2022); (FENG; XU; WEYMOUTH, 2022); (ZHAO et al., 2023); (HANNAN, 2022); (SOUZA, 2020); ; (BUCEA-MANEA-ŢONIş et al., 2022); ; (VESELOV et al., 2022); (THUAN et al., 2022); (CHINYAMUNJIKO; C.; BHIBHI, 2022); (MOZUMDER et al., 2023); (MOURTZIS; ANGELOPOULOS; PANOPOULOS, 2022); (GARG et al., 2021); (WANG; SUN; BIE, 2022); (CHIVU et al., 2022); (NUROVIC; POTURAK, 2021); (PANAGIOTIDIS, 2022); (LI, 2022a); (MIN; BIN, 2022); (KHOLISHOTULAILA; LAILA; ANGGA, 2022); (CHEN, 2022b); (KOMATH; O., 2022); (LI, 2022b); (WANG, 2022); (AL-MASKARI; T.; S., 2022); (KOSASI et al., 2022)
Ensino básico	(FERDIG et al., 2022); (CHOI; CHOI; PARK, 2022); (LI et al., 2023); (MORA-ZAMBRANO et al., 2022); (CHEN, 2022a); (NOKITI et al., 2022); (WANG et al., 2022); (ZHAO; DI; HE, 2022); (SUPRIATI et al., 2022); (HWANG; CHIEN, 2022)

Fonte: O Autor, 2023.

Figura 5 - Distribuição de artigos por fontes de publicação no continente asiático.

Fonte: O autor, 2023.

A tabela (4) apresenta os produtos criados pelos artigos selecionados para esta revisão sistemática. É importante salientar que foram considerados para esta tabela somente os artigos que de fato produziram algum tipo de plataforma. Sendo assim, trabalhos com propostas de criação de arquiteturas ou análises qualitativas ou quantitativas, não foram consideradas nesta seleção.

Tabela 4 - Itens da extração de dados.

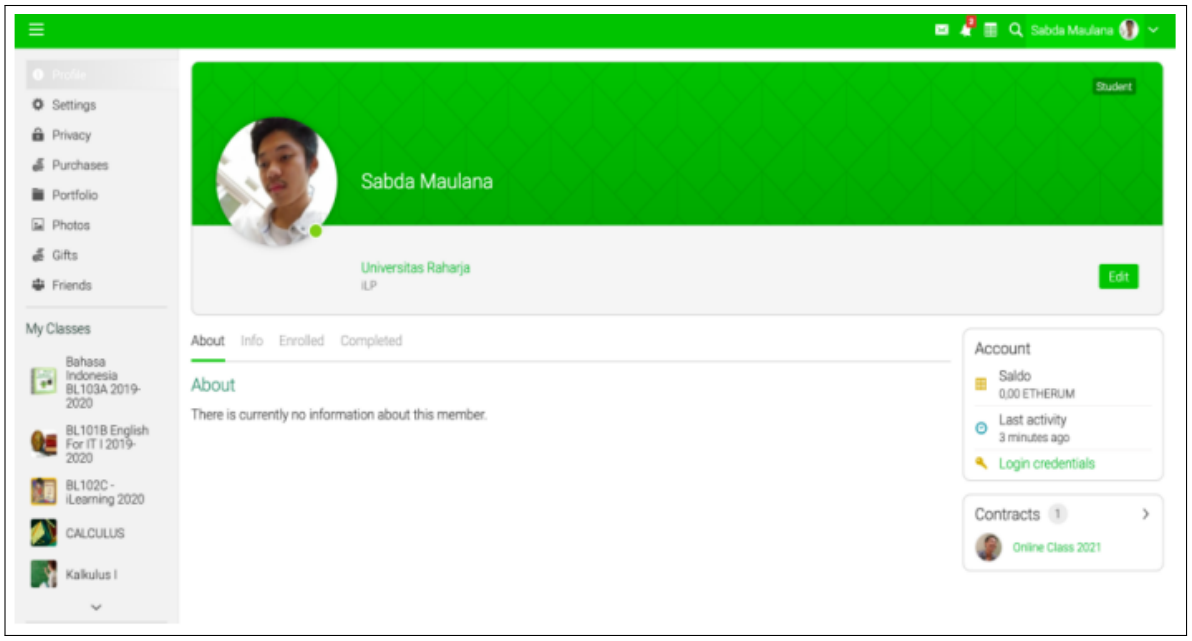
Autores	Produtos gerados
(ANWAR et al., 2022)	Framework para validação de atividades acadêmicas.
(GARG et al., 2021)	Tecido Hyperledger para classificar conteúdo educacional *on-line*.
(WANG; SUN; BIE, 2022)	Sistema para compartilhamento de dados educacionais.
(SHAIKH et al., 2022)	Tecido Hyperledger para avaliação de credenciais educacionais.
(BJELOBABA et al., 2022)	Modelo colaborativo para avaliação de trabalhos dos alunos.

Fonte: O autor, 2023.

No trabalho (ANWAR et al., 2022), o autor apresenta um modelo de confiança para fornecer um local seguro com o objetivo de apoiar a aprendizagem colaborativa . A proposta é que todos os cursos e certificados educacionais possam estar credenciados e serem possíveis de serem acessados em um mesmo local com a segurança utilizando o *blockchain* de iLearning centrado no aluno (SCi-B). O uso dessa tecnologia credencia e

melhora a qualidade da educação. Esta metodologia aproveita os benefícios da própria inovação e aborda fundamentalmente o extraordinário desenvolvimento na educação, uma vez que leva em conta a forte confirmação da aquisição de competências por alunos e, além disso, garante que eles estejam preparados de acordo com as circunstâncias, empregos reais e necessidades atuais da indústria; permite também que a empresa prepare estimativas relacionadas ao ciclo de desenvolvimento do aluno utilizando de forma sensata, programática e descentralizada. A aplicação ainda está em seus estágios iniciais, porém já há produção de uma interface para experimentos iniciais (figura 6). Essa interface se baseia em contratos inteligentes adaptados de *Ethereum* [1].

Figura 6 - Interface baseada em contratos inteligentes.



Fonte: (ANWAR et al., 2022).

No artigo (GARG et al., 2021) propõem-se sistemas Hyperledger, ou seja, uma iniciativa da Linux para desenvolver ecossistemas de códigos abertos baseados em tecnologia *blockchain*, descentralizados de revisão *on-line* para validar a confiabilidade da classificação e permitir a integração baseada em consórcio de especialistas no assunto. Este sistema garante uma revisão segura e transparente, onde ninguém pode fornecer avaliações falsas. Simultaneamente, as empresas de educação *on-line* obtêm uma visão clara das melhorias necessárias para subir na classificação e torná-las populares. A figura 7 mostra o *layout* do sistema desenvolvido.

No desenvolvimento do trabalho de (WANG; SUN; BIE, 2022) é construído um

[1] Ethereum é uma plataforma que permite a programação de aplicativos descentralizados, contratos inteligentes e transações da criptomoeda Ether e vários tokens. Alguns consideram uma evolução no conceito da tecnologia *blockchain*.

Figura 7 - Interface para ranking de cursos *on-line*.

Fonte: (GARG et al., 2021).

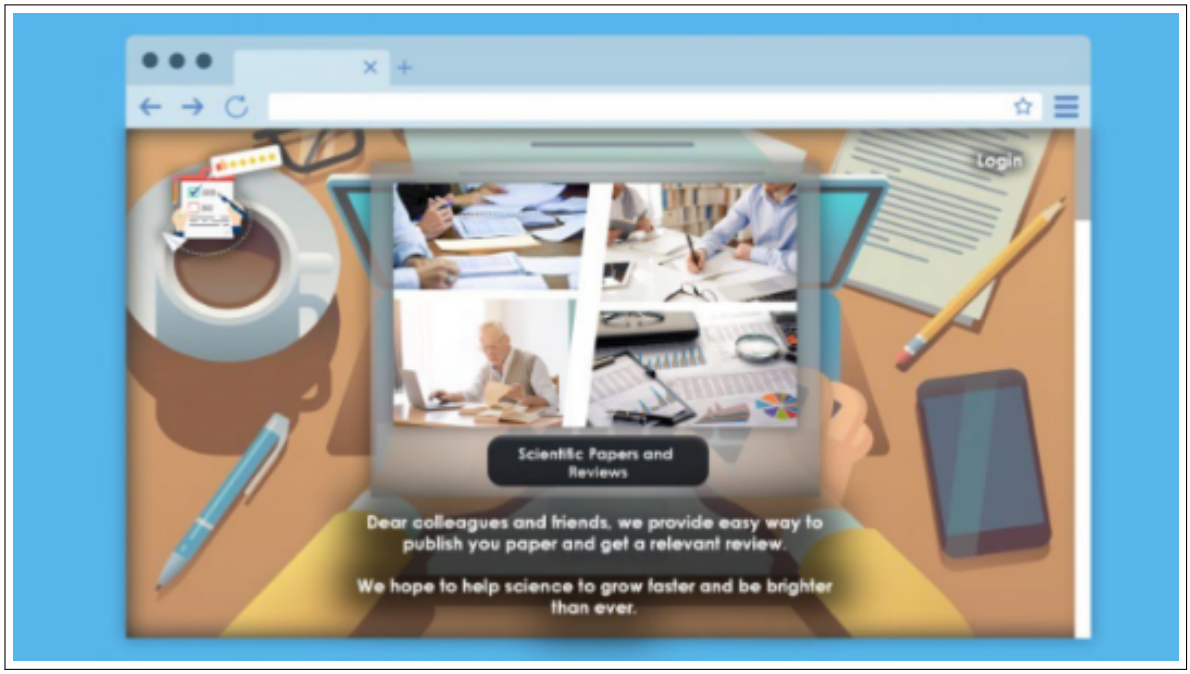
modelo de gerenciamento de dados educacionais *on-line* baseado em *blockchain*, que resolve os problemas de recursos educacionais *on-line*, autenticação e armazenamento em cadeia em *blockchain*. Com base no código de verificação temporário e na chave pública de terceiros, um mecanismo de compartilhamento de dados educacionais *on-line*, baseado em contrato inteligente é proposto, e a eficácia do mecanismo é verificada pela análise de segurança.

O trabalho de (SHAIKH et al., 2022) discute o sigilo e a privacidade das credenciais do certificado e os desafios relacionados à verificação segura no banco de dados centralizado. Por esse motivo, propuseram um processo de credenciamento seguro e de verificação de credencial de certificado habilitado para *blockchain hyperledger*. Este modelo proposto cria uma plataforma analítica para investigar credenciais de certificados, enquanto processa a avaliação (pré-verificação) usando uma rede neural artificial para classificação de registros antes de enviá-los para posterior atestação. No entanto, as partes interessadas participam no processo global de investigação e obtêm detalhes (apenas leitura) de todas as transações educativas através da rede do consórcio.

Por fim, no artigo de (BJELOBABA et al., 2022) é proposto um modelo de Aprendizagem Colaborativa e Avaliação do Trabalho Estudantil (CLSWE) baseado em tecnologias *blockchain* (BCTs), abrangendo conceitos selecionados do processo de revisão por pares de pesquisas científicas. Os BCTs são usados com o fim de desenvolver uma plataforma segura para armazenar e trocar dados sobre projetos e avaliações dos alunos. O modelo CLSWE oferece a possibilidade de melhorar a cooperação entre instituições de ensino superior e empresas que procuram as "competências empregatícias" de estudantes proativos. Foi construída uma plataforma com banco de dados criado na linguagem

MySQL para o modelo de testes, como é visto na figura 8.

Figura 8 - Interface do aplicativo para avaliação do trabalho dos alunos.



Fonte: (BJELOBABA et al., 2022).

O que aprendemos neste capítulo?

Vimos no Capítulo 1 artigos publicados em 2022 que desenvolveram estudos na área de *blockchain*. A escolha de artigos a partir deste ano se deve ao fato de estarmos tratando sobre um tema na área de tecnologia, que se modifica com uma velocidade muito maior que em outras áreas. Abordamos os critérios para nossa revisão sistemática, notamos uma maior concentração de publicações no continente asiático, principalmente China, e estabelecemos os artigos em quatro áreas. Além disso, pudemos mostrar algumas implicações e produtos produzidos por alguns destes trabalhos. Com isso em mente, podemos compreender melhor sobre o surgimento da tecnologia que o *blockchain* se insere.

2 PROBLEMA 1: CRIPTOGRAFIA

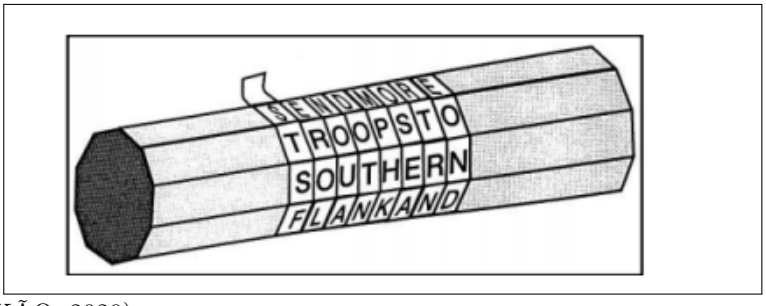
No Capítulo 1 foi apresentada uma revisão sistemática sobre Blockchain. Compreender, essa tecnologia é apenas um ramo pequeno de uma área da Matemática que vem sendo de grande utilidade desde que foi criada, a Criptografia. A Criptografia é denominada a arte de escrever em códigos, de forma a permitir que somente o destinatário a decifre e compreenda (TAMAROZZI, 2001). Essa área tem um papel importante nos dias atuais, pois é utilizada nos recursos humanos (auditoria eletrônica, lacre de arquivos de pessoal e pagamentos), em compras e vendas (autenticação de ordens eletrônicas de pagamento), nos processos jurídicos, na automação de escritórios, nos navegadores de *Internet*, nos aplicativos de mensagens, entre outras atividades da vida moderna (GROENWALD; OLGIN, 2011). Além dessas situações modernas, esse ramo foi muito importante historicamente em diversas áreas. É o que abordamos na próxima seção.

2.1 **A criptografia na história**

Esta seção se baseia no livro "O Livro dos Códigos" do autor Simon Singh (SINGH, 2001) e na dissertação de mestrado de Jéssica Paixão (PAIXÃO, 2020). Um dos primeiros relatos de criptografia foi feita por Heródoto, um grande historiador que narra o embate entre Pérsia e Grécia no século V antes da era cristã. A criptografia pode ser dividida em dois ramos: transposição e substituição.

Na criptografia de transposição as letras são permutadas como num anagrama. Se não houver um sistema de comunicação pré-definido entre emissor e receptor essa modalidade pode ser desvantajosa caso o tamanho da palavra seja pequeno, pois há poucas possibilidades de rearranjo. Já se a mensagem for longa, não há tempo hábil suficiente para que esta seja decifrada sem nenhuma orientação. Um exemplo desse tipo de criptografia é a cítala espartana (figura 9).

Figura 9 - Exemplo de cítala espartana.



Fonte: (PAIXÃO, 2020)

No caso da criptografia de substituição, cada letra do texto comum é substituída por uma letra diferente. A primeira vez que esta técnica aparece são em cartas do imperador Júlio César no século I antes da era cristã. Segundo este documento, o governante escreveu uma carta trocando as letras do alfabeto romano por letras gregas. Outra mudança utilizada por ele consistia na substituição de cada letra do alfabeto por outra que tivesse três posições adiante. Este método ficou conhecido como Cifra de César (figura 10).

Uma outra forma de rearranjar o alfabeto cifrado seria a escolha de uma palavra-chave ou frase-chave. Desta forma, iniciaríamos o alfabeto cifrado com as letras da palavra/frase-chave, excluindo-se repetições e, depois, seguiríamos a ordem do alfabeto com as letras que restassem. Por exemplo, suponha que a palavra-chave escolhida fosse COMANDANTE VALDEZ. Primeiro, eliminaríamos as letras repetidas obtendo COMANDTEVLZ. Em seguida, acrescentaríamos o restante das letras do alfabeto na ordem em que aparecem. O resultado é apresentado na figura 10 (PAIXÃO, 2020).

Figura 10 - Exemplo de Cifra de César.

Alfabeto normal	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Alfabeto cifrado	C	O	M	A	N	D	T	E	V	L	Z	B	F	G	H	I	J	K	P	Q	R	S	U	W	X	Y

Fonte: (PAIXÃO, 2020)

No século IX os árabes mostraram que era possível quebrar as cifras. Eles são considerados os inventores da ciência que permite decifrar uma cifra sem a chave, a criptoanálise. Foi com o estudo de várias temáticas, incluindo matemática, estatística e linguística que os árabes, através do entendimento da frequência relativa com que aparecem as letras, decifraram as cifras de substituição. Esse estudo mostra que as características do idioma formam padrões, como letras que mais aparecem, que facilita construir essas relações para decifrar a mensagem. Se o texto for curto, essas relações podem ser mais difíceis de ser encontradas.

A criptografia se desenvolveu séculos depois no Ocidente. O primeiro livro que menciona a criptografia aparece apenas no século XIII, na Europa. No século seguinte, o uso dessa ferramenta estava mais difundido, pois cientistas e alquimistas a usavam a fim de manter suas descobertas em sigilo. Dois séculos depois foi utilizada para sigilo político. Uma batalha entre criptógrafos (que criavam os códigos) e os criptoanalistas (que decifravam esses códigos) estava instaurada. Quem ainda não dominava a arte de decifrar se mantinha cegamente nas cifras de substituição, enquanto os que possuíam mais domínio, já sabiam que não era completamente seguro este tipo de cifra.

Com a análise de frequência e sua base nos padrões da língua materna, uma forma simples de dificultar a quebra do código era acrescentar símbolos que não significavam nada, os nulos. Em conjunto com os nulos, usar a escrita de forma incorreta, sem perda

de compreensão, também era uma possibilidade. Assim, CASA poderia ser escrita como CAZA para dificultar o encontro da padronização das letras.

O uso da análise de frequência relativa das letras tornou a criptografia de substituição fraca, tanto que, em 1586, Maria, a rainha da Escócia, foi executada. Sua correspondência com conspiradores era considerada tão indecifrável por eles que cometeram o erro de colocar todas as suas informações nas cartas.

Ao fim do século XVI, é proposto por Leon Battista Alberti o uso de dois ou mais alfabetos cifrados (figura 11).

Figura 11 - Exemplo de Cifra de Alberti.

Alfabeto original	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
Alfabeto cifrado 1	F	Z	B	V	K	I	X	A	Y	M	E	P	L	S	D	H	J	O	R	G	N	Q	C	U	T	W
Alfabeto cifrado 2	G	O	X	B	F	W	T	H	Q	I	L	A	P	Z	J	D	E	S	V	Y	C	R	K	U	H	N

Fonte: (SINGH, 2001)

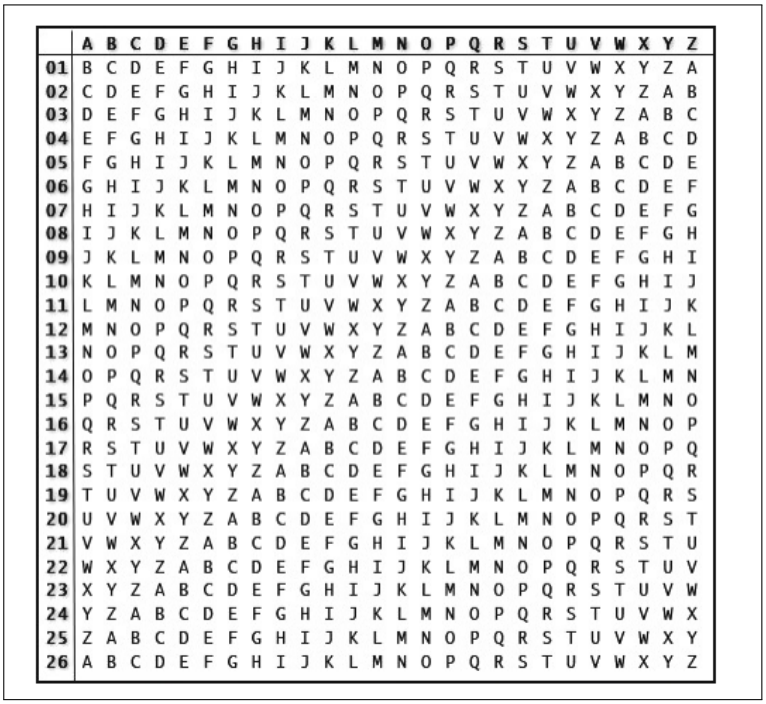
Um exemplo para entender este método é como seria feita a cifra da palavra MOEDA. Em primeiro lugar se mudaria a letra M pela letra L do alfabeto 1, depois O seria modificada por J do alfabeto 2, a terceira letra E seria modificada por K novamente pelo alfabeto 1, a letra D pela B do alfabeto 2 e, por fim, a letra A pela letra F no alfabeto 1, e assim a palavra MOEDA seria escrita como LJKBF. Aqui, percebe-se que a vantagem é que uma mesma letra pode ter duas opções de codificação. O contrário também ocorre, assim a letra do alfabeto original possui duas representações no alfabeto cifrado.

Após anos, Blaise de Vigenère aperfeiçoa o método de Alberti e cria o Quadrado de Vigenère (figura 12). Essa cifra consiste em utilizar vinte e seis alfabetos e para utilizá-la é preciso o uso de uma palavra-chave acordada entre emissor e receptor.

O quadrado de Vigenère oferecia a vantagem de ser imune à análise por frequência, que era a mais utilizada na época e também possuía inúmeras chaves, visto que era um acordo estabelecido entre o receptor e emissor, e isso tornava inviável a decodificação por tentativa. Essa cifra era tão complexa para a época que acabou sendo utilizada apenas dois séculos depois de sua criação. Assim, a próxima busca dos criptógrafos foi encontrar uma cifra que possuísse uma dificuldade intermediária, porém, entre os séculos XVII e XIX todas as técnicas propostas se mostraram não tão eficazes, o que desencadeou o uso da técnica de Vigenère com o principal objetivo de reforçar a proteção de mensagens enviadas por telegrama.

No século XIX é construído o telégrafo e o responsável pela primeira linha foi Samuel Morse, que também criaria o código com seu nome. O aparelho tinha alcance de 60 quilômetros e o código era constituído de pontos e traços que representavam as letras do alfabeto e o envio desse código era em forma de bips audíveis (figura 13). Esse código não é uma forma de criptografia, tendo assim que ser enviado a um especialista para

Figura 12 - O Quadrado de Vigenère.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
01	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
02	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
03	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
04	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
05	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
06	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
07	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
08	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
09	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
10	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
11	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
12	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
13	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
14	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
15	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
16	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
17	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
18	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
19	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
20	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
21	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
22	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
23	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
24	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
25	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
26	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z

Fonte: (SINGH, 2001)

criptografá-lo e depois enviar a mensagem. Utilizando o quadrado de Vigenère para cifrar as mensagens, nem o responsável pelo envio sabia de que tratava o conteúdo da mesma. Assim, não havia o risco de que esses trabalhadores pudessem ser subornados para passar informações. Em 1863, Friedrich Wilhelm Kasiski publicou a técnica de como decifrar o quadrado de Vigenère. A partir daí, procura-se uma nova codificação mais segura para o telégrafo. No final do século XIX, Guglielmo Marconi, atraído por circuitos elétricos, cria o rádio, que possuía uma grande vantagem quanto ao telégrafo: não eram necessário fios.

O rádio logo chamou a atenção dos militares, principalmente com a chegada da Primeira Guerra Mundial. Porém, existia uma grande fragilidade nele: uma vez que fosse descoberta a frequência das ondas emanadas, qualquer receptor poderia ter a informação da mensagem. Por isso, era importante ser a mensagem criptografada. As cifras durante esse período foram modificadas, mas sempre após algum tempo as mesmas eram decifradas pelas tropas inimigas.

2.2 **A Criptografia na Segunda Guerra Mundial**

No século XX, Scherbius, um inventor alemão, dedicou-se a trabalhar com a tecnologia da época para desenvolver uma máquina que substituísse o uso de papel e lápis, já que, por exemplo, na 1ª Guerra Mundial, só os franceses chegaram a interceptar 1000 mensagens criptografadas. Assim, é criada a máquina *Enigma*, que era quase uma versão

Figura 13 - Código Morse internacional.

Símbolo	Código	Símbolo	Código
A	·–	W	·––
B	–···	X	–··–
C	–·–·	Y	–·––
D	–··	Z	––··
E	·	1	·————
F	··–·	2	··———
G	––·	3	···––
H	····	4	····–
I	··	5	·····
J	·–––	6	–····
K	–·–	7	––···
L	·–··	8	–––··
M	––	9	––––·
N	–·	0	–––––
O	–––	ponto final	·–·–·–
P	·––·	vírgula	––··––
Q	––·–	ponto de interrogação	··––··
R	·–·	dois pontos	–––···
S	···	ponto e vírgula	–·–·–·
T	–	hífen	–····–
U	··–	barra	–··–·
V	···–	aspas	·–··–·

Fonte: (SINGH, 2001)

elétrica do disco de cifras criado por Leon Albertini no século XV (figura 14). Basicamente, consistia em dois discos de cobre de tamanhos diferentes com alfabeto escrito neles. Os dois discos eram fixados, um no outro, através de um eixo, com o disco menor ficando em cima do maior, de modo que fosse possível a rotação. Com essa máquina era possível mecanizar a Cifra de César e a Cifra de Vigenère. A inovação da *Enigma* era possuir um teclado para que se digitasse o texto normal, uma unidade misturadora e um teclado cujas letras podiam ser iluminadas por suas várias lâmpadas que indicavam as letras do texto cifrado.

Figura 14 - Disco de cifras.

Fonte: (SINGH, 2001)

O disco misturador era responsável por traduzir as letras do texto original para o codificado. A ideia consistia em um giro de um vinte e seis avos do misturador a cada letra digitada. Aqui se apresentava uma possibilidade de falha, pois após 26 voltas o disco voltava a sua posição inicial; assim, se alguém tivesse acesso a uma das máquinas, poderia digitar indefinidamente um mesmo botão 26 vezes e o padrão se repetiria, o que torna a criptografia insegura. Contra esse revés era possível adicionar mais um misturador. O método de cifragem seria o mesmo no primeiro misturador, enquanto o segundo disco permaneceria imóvel, e só giraria quando o primeiro misturador tivesse completado sua volta.

A *Enigma* (figura 15) na verdade contava com três misturadores e um refletidor, o que permitia desfazer o processo de cifragem usando a própria máquina. Os três misturadores podiam ter seus lugares alterados e havia um painel de tomadas de forma que era

possível trocar algumas letras antes que elas entrassem no misturador, o que dificultava o uso de análise de frequências.

Figura 15 - Máquina *Enigma*.

Fonte: (GALILEU, 2018)

Utilizando os três misturadores e as vinte e seis letras do alfabeto se tem 26 x 26 x 26 = 17576 possibilidades de posições iniciais, e é justamente esta posição que determina a forma como a mensagem é cifrada. Ao se pensar em todos os fatores citados, se teria um número acima de 10 quatrilhões de chaves possíveis.

As chaves da *Enigma* estavam indicadas em um livro de códigos e uma cópia era dada a todos da comunicação. Cada página indicava as configurações de um dia e serviam tanto para o emissor, quanto para o receptor da mensagem, uma vez que para desfazer a cifragem, bastava que o operador fosse digitando a mensagem cifrada e o texto decifrado seria iluminado no painel, tudo isso graças ao uso do refletor.

Scherbius após a patente de sua máquina tenta vendê-la e não obtém sucesso pelo alto custo. Só depois de 5 anos, graças a um livro inglês que revelou que os britânicos conseguiram decifrar as mensagens dos alemães na Primeira Guerra Mundial é que existe um grande investimento por parte destes militares. Nos vinte anos seguintes, mais de 30 mil máquinas foram compradas pelos militares e sua decifragem fez com que os alemães dessem como certa a vitória de seu país.

Após os investimentos iniciais de Hans-Thilo Schmidt, um dos que sofriam com o pós-guerra, decidiu entregar à França dois documentos que forneciam o modo de usar a *Enigma*. Essa informação foi também compartilhada com a Polônia. Embora tivessem a possibilidade de criar uma réplica da máquina, os franceses continuavam com o grande problema de não saber a disposição dos componentes, ou seja, a máquina sem a chave era inútil. Já a Polônia, com medo da invasão iminente, se debruçou nos documentos. Decidiram contratar matemáticos experientes em alemão, ao invés de peritos em

linguística.

O matemático Marian Rejewski foi capaz de decifrar a máquina após uma série de análises e uma cópia da *Enigma*. Porém, em 1938, a Alemanha passou a contar com cinco misturadores e para que fossem decifráveis as novas modificações seriam necessários altos investimentos que o governo polonês não era capaz de custear. Com medo da invasão alemã, decidem dividir todas suas descobertas com os franceses e britânicos.

A maior lição que os britânicos tiraram dos poloneses foi a contribuição que os matemáticos traziam como decifradores de códigos e os mantiveram. Os novos recrutas já dominavam a técnica em um ano e seguiam a rotina: à meia-noite, os operadores alemães mudavam para uma nova chave diária e qualquer avanço feito pelos britânicos no dia anterior era perdido; assim, todo trabalho era reiniciado, o código descoberto e as mensagens acumuladas do dia eram decifradas. Esse processo ganhou alguns atalhos, mas não houve nenhum avanço substancial até a chegada de Alan Turing, que atacou impiedosamente esse objetivo e venceu a batalha.

2.2.1 Alan Turing

Nascido em 1911, Alan Turing foi admitido ao King's College, Cambridge, no ano de 1931; nessa época o campo matemático discutia a existência das questões indecidíveis. Inspirado por esse assunto, ele escreveu seu artigo "Sobre os números computáveis", e nele imaginava uma série de máquinas que efetuariam diversas operações matemáticas – as *máquinas de Turing*. A chamada *máquina universal* tinha como objetivo responder a qualquer questão que pudesse ser feita utilizando lógica. Apesar de nunca ter conseguido construí-la, foi a inspiração que deu aos matemáticos para a criação do computador moderno.

Turing analisou mensagens antigas e notou que havia uma estrutura rígida a ser seguida. Assim, notou alguns padrões no texto cifrado e se dedicou, primeiramente, a resolver o problema dos misturadores utilizando três máquinas *Enigma* conectadas por fios. Em 1940, o primeiro protótipo de decodificação estava pronto, mas a máquina demorava uma semana para encontrar a chave.

A segunda máquina com ajustes ficaria pronta quatro meses depois, mas nesse período os alemães modificaram sua decodificação. Alguns meses depois, cria-se uma teceira máquina que adequa-se muito bem e se esperava encontrar a chave em uma hora. Porém, os alemães utilizavam diversos sistemas de comunicação, cada um com um livro de códigos diferentes e algumas máquinas com oito misturadores. A solução para facilitar as decifrações era produzir *colas*, ou seja, encontrar padrões nos códigos gerados. Para isso, o governo britânico lançava minas sobre um local escolhido e obrigava os navios alemães a enviarem avisos a outros navios. Dentre as informações estava a localização, que já era

conhecida pelos ingleses e assim era obtida a *cola*.

Não se pode afirmar ao certo que as descobertas de Turing e a Escola de Cifras e Códigos do Governo foram um fator decisivo para a vitória, mas os especialistas afirmam que auxiliaram a encurtar o período de guerra. Em 1974, após se certificarem de que a *Enigma* caíra em desuso, que o segredo terminara, o serviço de informações deu o aval para a publicação do livro "The Ultra Secret", que narrava as atividades conduzidas no período da guerra.

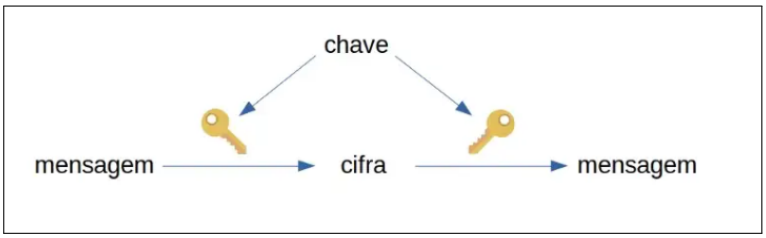
2.3 **As chaves atuais da Criptografia**

Ao longo da história foi visto que o acesso às chaves derrubava a segurança das mensagens criptografadas. Com o desenvolvimento da tecnologia, houve uma modificação substancial nos sistemas criptográficos. Na computação as "chaves" são baseadas em um conjunto de *bits* em algum algoritmo que é capaz de codificar e decodificar informações ou dados. O sistema ainda funciona com o receptor e o emissor possuindo chaves compatíveis.

2.3.1 Criptografia Simétrica

As chaves simétricas são as mais simples, são aquelas em cujo emprego tanto o emissor quanto o receptor recorrem à mesma chave. Em outras palavras, uma mesma chave, a que é usada para criptografar uma mensagem, obrigatoriamente precisa ser utilizada para descriptografar essa mensagem, como foi visto na "Cifra de Júlio César". Portando essa chave deve ser conhecida por emissor e pelo receptor.

Figura 16 - Criptografia de chave simétrica.



Fonte: (SOUZA, 2020)

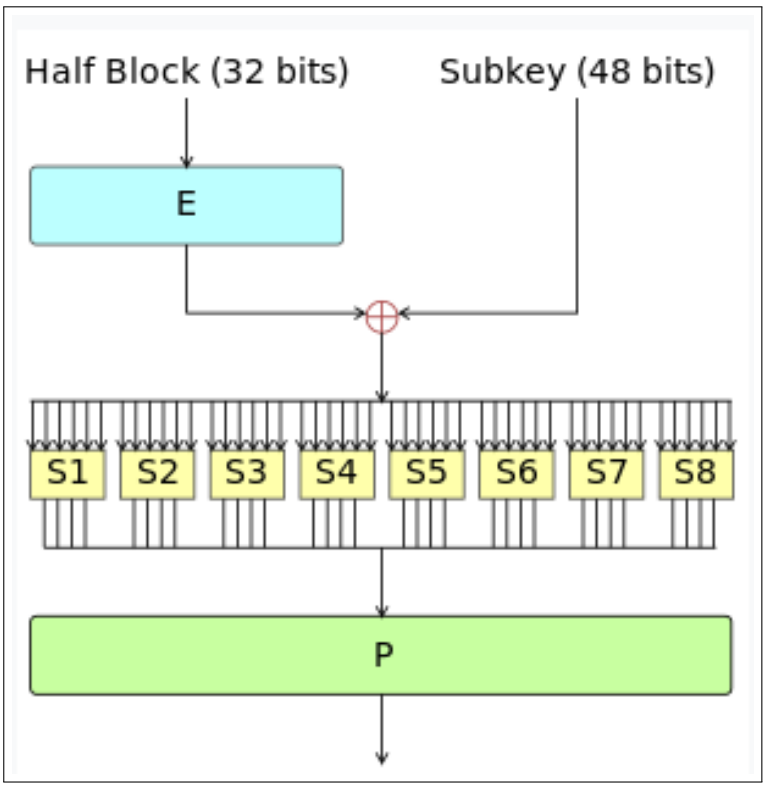
Quando o emissor criptografa (ou cifra) uma mensagem, é utilizado um algoritmo que transforma o conteúdo em texto cifrado, e para o receptor descriptografar (ou decifrar) a mensagem, utiliza-se o mesmo algoritmo para converter o texto na mensagem original (figura 16). Porém, se um interceptor souber o algoritmo de encriptação ele pode decifrar a mensagem facilmente. Entra então a chave simétrica (ou privada). A ideia é que quando

o emissor utilizar o algoritmo de encriptação, coloque também uma chave de segurança; por sua vez o receptor utiliza o mesmo algoritmo e a mesma chave para decifrar o texto. Assim, se o interceptor não possuir esta chave não conseguirá decifrar a mensagem.

No que se refere a esse tipo de algoritmo de chaves temos (VIANA et al., 2022):

- **DES (Data Encyption Standard):** criado em 1977, utiliza chaves de 56 bits mais 8 bits de paridade, tendo assim 72 quatrilhões de combinações que podem decifrar a informação; um número baixo para um computador potente. Foi o algoritmo mais disseminado no mundo até o AES. Possui variações como o 3DES (com 168 bits) e DESX (com 120 bits), que aumentaram a quantidade de bits para maior segurança, porém, ainda não de forma tão significativa e por isso não são mais eficazes atualmente (figura 17).

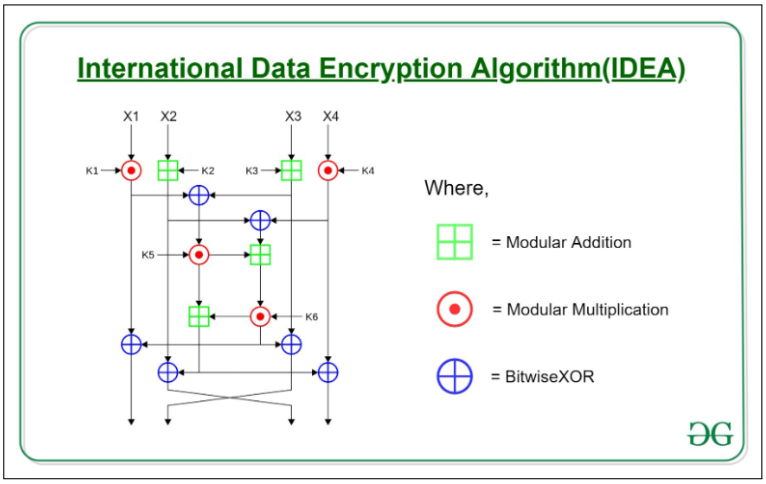
Figura 17 - Criptografia DES.



Fonte: (WIKIPEDIA, 2001)

- **IDEA (International Data Encryption Algorithm):** criado em 1991, esse já tem base em chaves de 128 bits, possui estrutura semelhante ao DES, mas opera blocos de informação de 64 bits e sua implantação é mais fácil, em comparação ao anterior (figura 18). Ele utiliza princípios de confusão, que impedem o realinhamento das informações. É usado principalmente no mercado financeiro e para programas de criptografia para e-mail pessoal.

Figura 18 - Criptografia IDEA.



Legenda: Todas as operações realizadas em seus subblocos de 16 bits.
Fonte: (GEEKS, 2023)

- **AES (Advanced Encryption Standard):** um dos algoritmos mais populares e seguros desde 2006, com alto nível de eficiência e confidencialidade, padrão adotado pelo governo dos Estados Unidos. Tem um bloco de tamanho fixo com 128 bits, e uma chave com tamanho de 128, 192 ou 256 bits. É conhecido como um algoritmo quase imune a todos os tipos de ataques exceto os de força bruta (tentativa e erro) (figura 19). Trata-se de uma variação da cifra de bloco Rijndael, sendo abordado com mais detalhes na seção 2.3.1.1.

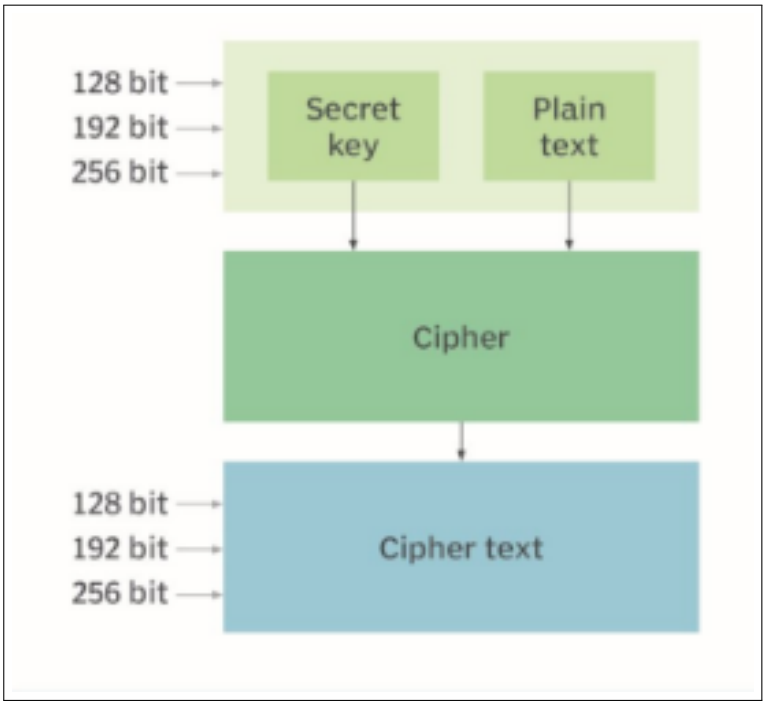
- **RC 4 (Ron's Code ou Rivest Cipher):** criado pela empresa RSA Data Security em 1987, varia de 8 a 1024 bits e é muito utilizado por provedores de e-mails devido a sua simplicidade e velocidade de operação (figura 20).

- **Blowfish:** criado por Bruce Schneir em 1993, varia entre 32 a 488 bits, popular em e-commerce, com confiabilidade e velocidade em lidar com métodos de pagamento, além de ser *open source*, ou seja, não patenteado. Segmenta as informações em blocos de 64 bits.

- **Twofish:** variação do Blowfish por blocos de 128 bits e chaves de 256 bits, também é de uso livre, sem restrição para qualquer usuário.

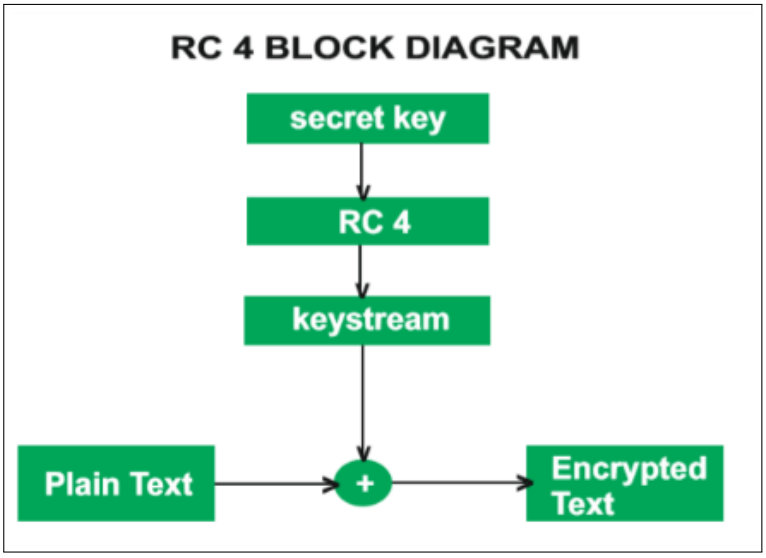
A vantagem desse tipo de criptografia está na facilidade de implementação e velocidade de processamento, porém, o problema desse tipo, geralmente, é que a chave deve ser compartilhada entre origem e destino e armazenada em local seguro, mas corre o risco de ser interceptada durante o compartilhamento. Além disso, essa modalidade não garante os princípios de autenticidade e não repúdio.

Figura 19 - Criptografia AES.



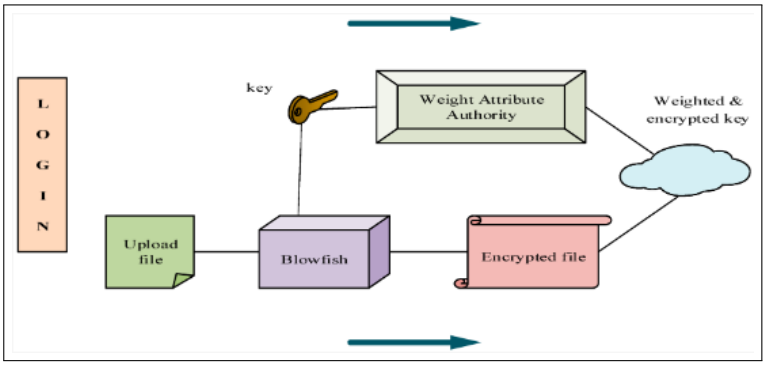
Legenda: A criptografia AES se utiliza de um subconjunto da criptografia de Rjindael.
Fonte: (AWATI; BERNSTEIN; COBB, 2024)

Figura 20 - Criptografia RC4.



Legenda: A criptografia RC4 codifica o texto original utilizando dois algoritmos (RC4 e keystream) em uma chave secreta disponibilizada pelo usuário.
Fonte: (GEEKS, 2021)

Figura 21 - Criptografia Blowfish.



Fonte: (GHOSH; KARAR, 2018)

2.3.1.1 Aplicação atual 1: Esteganografia

É possível criar também uma assinatura com Esteganografia.

A Esteganografia é a ciência que estuda a ocultação de mensagens. Não devemos confundir a esteganografia com a criptografia, que oculta o significado da mensagem, e não a mensagem em si. Na esteganografia digital é possível esconder imagens, áudios ou textos. Em sua forma mais comum, seleciona-se determinado pixel de uma imagem e troca-se o *bit* menos significante por um bit de informação (CAMPOS; TIZZIOTTI, 2015). Para melhor eficácia na inserção de bits, foi descoberto o método chamado de inserção por matriz, que faz com que o receptor não precise saber onde o emissor fez as alterações na imagem. Basicamente, este método utiliza conceitos de álgebra linear e pode ou não utilizar matriz em sua construção.

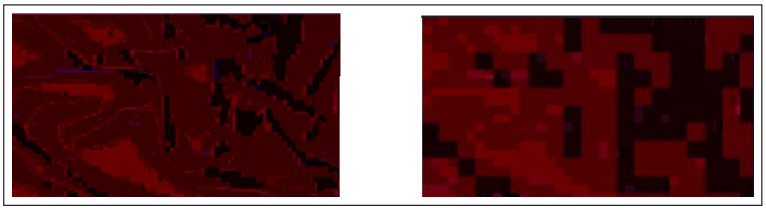
Para compreendermos melhor, utilizamos o trabalho escrito por Davi Fajardo e Ighor Rimes na Décima Sexta Semana do IME em 2023.

O artigo visava, primariamente, o artifício de esconder mensagens de texto, o que também se aplica a assinaturas digitais e em arquivos de imagem. O método de esconder a mensagem altera os pixels da imagem de maneira imperceptível à vista humana, contudo, observável pelo "olho" do software. Essa técnica é conhecida como LSB (*Least Significant Bit* ou Bit Menos Significativo). O algoritmo primeiramente comprime a mensagem a ser escondida e, logo em seguida, através de números pseudo-aleatórios, seleciona os pixels a serem modificados (figura 22).

Note que a matriz à esquerda seria a matriz formada por todos os pixels da imagem original e a matriz à direita seria aquela com modificações em alguns pixels para esconder a mensagem desejada. Na prática, o que acontece é uma coloração levemente diferente da original sendo nela armazenada a informação a ocultar. Neste caso, para que seja percebida a modificação, foi realizado um aumento gigantesco em uma pequena área da imagem e assim é possível localizar a diferença.

O aplicativo utilizado foi o (STEGHIDE, 2003), que é um programa de esteganogra-

Figura 22 - Imagens de aplicação de Esteganografia.



Legenda: A imagem à esquerda é formada pelos pixels originais e à direita a mesma imagem com as alterações.
Fonte: O autor, 2024.

fia capaz de ocultar dados em vários tipos de arquivos de imagem e áudio. As frequências de amostra de cor, respectivamente, não são alteradas, tornando a incorporação resistente a testes estatísticos de primeira ordem. Suas características são:

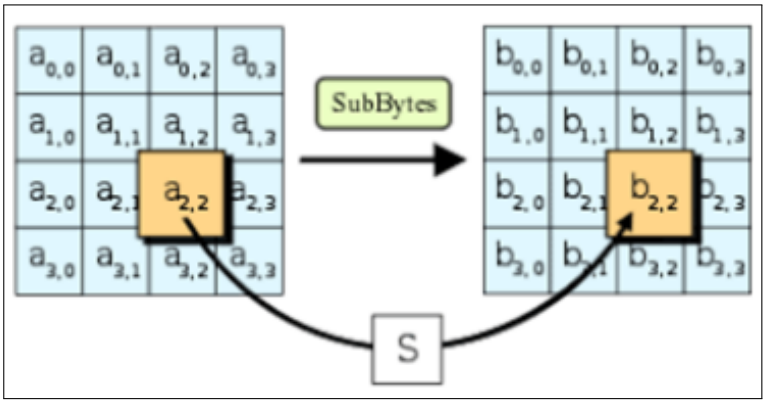
- compactação de dados incorporados;
- criptografia de dados incorporados;
- incorporação de verificação para integridade dos dados extraídos;
- suporte para arquivos JPEG, BMP, WAV e AU.

O Steghide é de Licença Pública Geral, o que permite a modificação e distribuição do programa, desde que essas modificações sejam disponibilizadas livremente.

O modo para tal é a encriptação de Rijndael (DAEMEN; RIJMEN, 1999), que utiliza uma chave simétrica e que no caso de imagens que irão conter textos, os valores aleatórios da chave e do texto são convertidos para binário como em um plano cartesiano e substituídos pelas suas coordenadas em outra matriz de base hexadecimal (figura 23).

Essa técnica é entitulada como *SubBytes*. É importante salientar que existem quatro processos de transformações em encriptação de Rijndael: SubBytes, ShiftRows, MixCollumns e AddRoundKey. Para melhor compreensão, é visto cada um desses passos utilizando um exemplo. Em primeiro lugar, para a encriptação formamos uma matriz de bits da região que é substituída em nossa imagem. Esta matriz é convertida em base hexadecimal. A base hexadecimal possui um processo de formação bem similar à base binária. É importante salientar que, nesta base, a notação é representada pelos algarismos decimais, porém, a partir do número 10 há uma troca de simbologia, onde 10 = A, 11 = B, 12 = C, 13 = D, 14 = E, 15 = F. Por exemplo, é convertido o número 438 em hexadecimal (figura 24). O processo é dividir o número desejado por 16 e, com o resto encontrado em cada divisão, escrever esse número de trás para frente. Uma pergunta que pode ser refletida quando estudamos este processo é qual a causa de modificarmos o número decimal para base hexadecimal. O motivo é que quando o número está em base

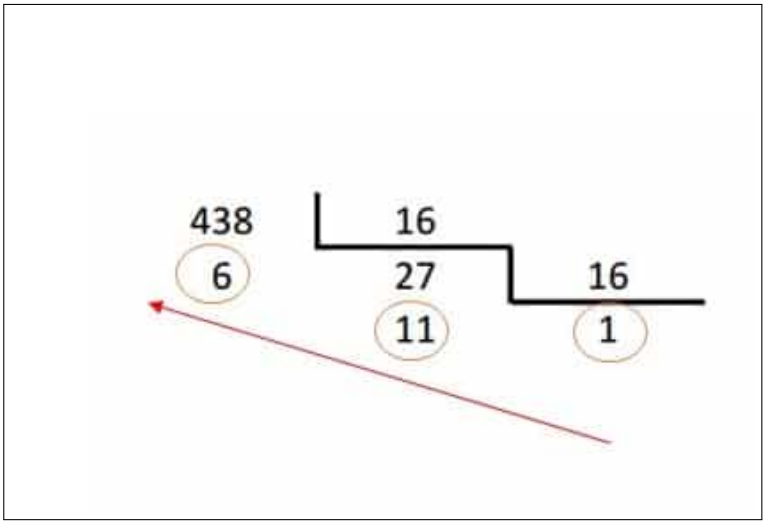
Figura 23 - Matriz de substituição em Esteganografia.



Legenda: A matriz à esquerda é formada pelos pixels originais da imagem e à direita a matriz com as alterações.
Fonte: (DAEMEN; RIJMEN, 1999)

hexadecimal, transformá-lo em base binária é muito mais simples e rápido. É visto essa transformação a seguir.

Figura 24 - Conversão de base decimal para hexadecimal.



Legenda: Utilizamos o algoritomo de Euclides para modificar a base do número.
Fonte: (SOUZA, 2016)

Assim, o número $438 = 1B6$. É muito comum na frente o número $1B6$ também ser escrito como $Ox1B6$ ou $\#1B6$. Também é possível utilizarmos a nomenclatura do operador *mod*. A aritmética modular trata de conceitos de divisibilidade e congruência que são trabalhados com conjunto dos números inteiros.

Seja m um número natural. Diremos que dois números inteiros a e b são congruentes módulo m se os restos de sua divisão euclidiana por m são iguais. Quando os inteiros a e

b são congruentes módulo m, escreve-se

$$a \equiv b \ (mod \ m).$$

Aqui tomaremos sempre $m > 1$, pois o resto da divisão de qualquer inteiro por 1 é zero e isso não nos é interessante.

Segue do algoritmo da divisão que todo inteiro a é congruente módulo m a um inteiro b, tal que $0 \leq b < m$.

De fato, considerando $a \in \mathbb{Z}$ e a divisão euclidiana de a por m e quociente q, tem-se:

$$a = qm + b, \quad \text{onde} \quad 0 \leq b < m.$$

Como $0 \leq b < m$, o resto da divisão de b por m é exatamente b. Logo $a \equiv b(mod \ m)$.

Esta definição, conhecida também como Algoritmo de Euclides, é vista em turmas de 6º ano do ensino fundamental II, principalmente, envolvendo os cálculos de "prova real" e Máximo Divisor Comum (MDC).

Vamos compreender como seria a utilização de *mod* no exemplo do 438. Nosso cálculo seria

$$438 \equiv 6 \ (mod \ 16)$$

$$27 \equiv 11 (mod \ 16)$$

$$1 \equiv 1 (mod \ 16)$$

Repare que o processo é idêntico ao mostrado anteriormente. Aqui utilizamos os restos 6, 11 e 1 identificados por b em nossa aritmética modular obtendo $438 = 1B6$.

Esse processo, que demanda mais atenção, pode parecer desnecessário já que utilizamos o Algoritmo de Euclides nele também. Porém, muitos algoritmos de conversão utilizam exatamente o princípio da aritmética modular, então é importante compreendermos como essa conversão é efetuada com o computador. É importante salientar também que a mudança para qualquer base segue a mesma lógica. Isso é essencial ser compreendido, pois a base binária, que é a base para todos os sistemas de computação, também se utiliza deste método. Outro ponto que nos chama atenção é que, como observado anteriormente, o conceito de *mod* é visto de outra forma nos anos iniciais de nossa formação, mas que sua formalização é encontrada em cursos iniciais de Álgebra, na maior parte das carreiras de graduação em Ciências Exatas. Assim, o método de Rijndael e modificações de base podem ser apresentados como práticas a serem incorporadas nesse curso como mais um motivador do estudo desse conteúdo.

Compreendido como transformar um número de base decimal para um de base hexadecimal, necessitamos compreender como transformar um número binário em hexadecimal, pois o método de Rjindael já se utiliza de bits nessa base para iniciar seu

procedimento de encriptação. Vamos transformar o número binário 00110010 como um exemplo. Para se transformar o número binário para hexadecimal é necessário inicialmente separar, da direita para a esquerda, o número em grupos de quatro dígitos. A quantidade de dígitos aumenta de acordo com o número que se deseja transformar. Com o auxílio da tabela na figura 25 podemos fazer a modificação e obtemos $0011 = 3$ e $0010 = 2$.

Figura 25 - Tabela de conversão hexadecimal para binários.

Hexadecimal	Binário		Hexadecimal	Binário
0	0000		8	1000
1	0001		9	1001
2	0010		A	1010
3	0011		B	1011
4	0100		C	1100
5	0101		D	1101
6	0110		E	1110
7	0111		F	1111

Fonte: (SOUZA, 2016)

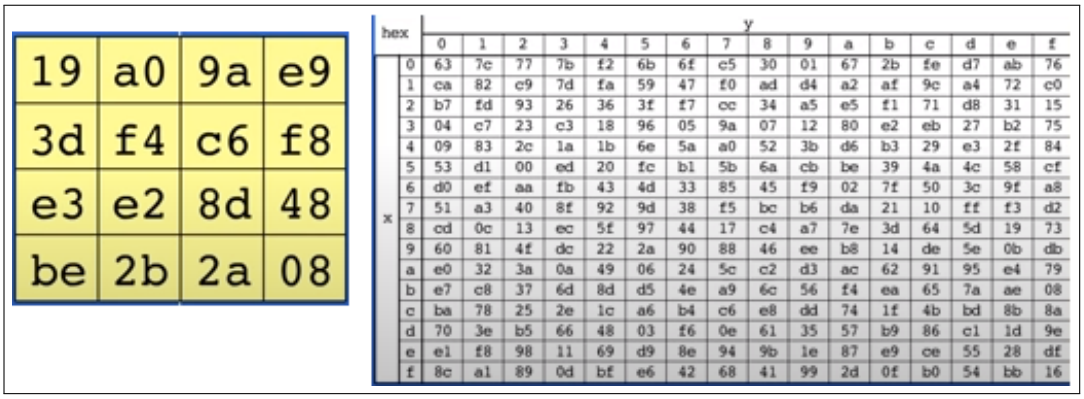
Finalmente, podemos iniciar nosso algoritmo de esteganografia. O exemplo a seguir foi retirado do canal do Youtube entitulado “AppliedGo” criado por Enrique Zabala (ZABALA, 2017). Suponha uma matriz de bytes hexadecimais. Essa matriz será encriptada e para iniciar o processo é utilizado uma matriz suporte que trocautilizando o *blockchain* de iLearning centrado no aluno (SCi-B) seus elementos com nossa matriz inicial (figura 26).

Nosso primeiro processo, o SubBytes, basicamente utiliza a matriz à direita como uma matriz de troca de elementos. Por exemplo, nosso elemento $a_{11} = 19$. Com isso, será buscado qual elemento está na linha 1 e coluna 9 dessa matriz suporte modificando o 19 por $d4$. O mesmo será feito a todos os elementos de nossa matriz original. A matriz resultado do método de SubBytes pode ser vista na figura 27.

O segundo processo é o *ShiftRows* onde um byte é modificado para o fim da linha na posição da segunda linha da matriz, dois bytes na terceira e assim sucessivamente (figura 28).

O terceiro passo intitulado *MixColumns* seleciona cada coluna da matriz que estamos transformando e multiplica por uma matriz que é obtida utilizando campos de Galois, conceito esse oriundo de anéis e corpos de Álgebra (figura 29) e pode ser melhor explorado em (DAEMEN; RIJMEN, 1999). Aqui vale ressaltar que, mesmo se a matriz

Figura 26 - Tabela de troca de elementos da matriz a ser encriptada.



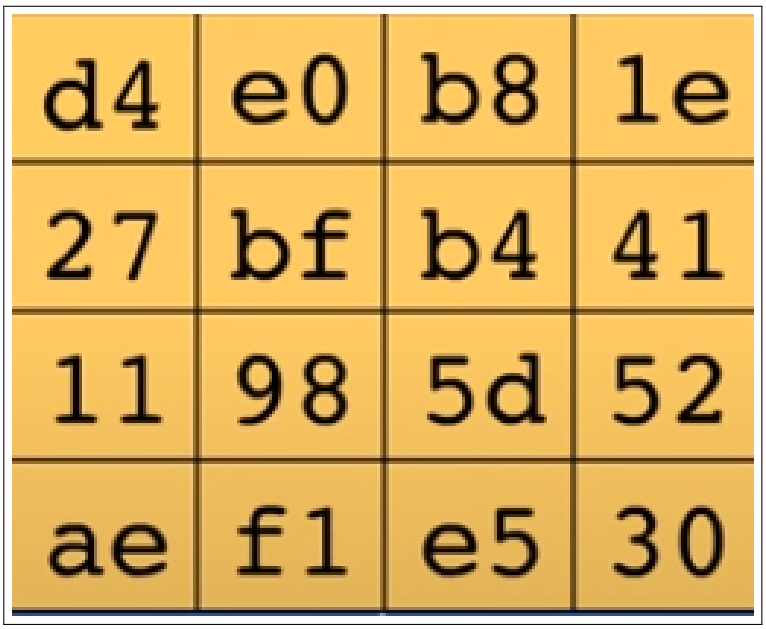
19	a0	9a	e9
3d	f4	c6	f8
e3	e2	8d	48
be	2b	2a	08

hex	y															
	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
0	63	7c	77	7b	f2	6b	6f	c5	30	01	67	2b	fe	d7	ab	76
1	ca	82	c9	7d	fa	59	47	f0	ad	d4	a2	af	9c	a4	72	c0
2	b7	fd	93	26	36	3f	f7	cc	34	a5	e5	f1	71	d8	31	15
3	04	c7	23	c3	18	96	05	9a	07	12	80	e2	eb	27	b2	75
4	09	83	2c	1a	1b	6e	5a	a0	52	3b	d6	b3	29	e3	2f	84
5	53	d1	00	ed	20	fc	b1	5b	6a	cb	be	39	4a	4c	58	cf
6	d0	ef	aa	fb	43	4d	33	85	45	f9	02	7f	50	3c	9f	a8
7	51	a3	40	8f	92	9d	38	f5	bc	b6	da	21	10	ff	f3	d2
8	cd	0c	13	ec	5f	97	44	17	c4	a7	7e	3d	64	5d	19	73
9	60	81	4f	dc	22	2a	90	88	46	ee	b8	14	de	5e	0b	db
a	e0	32	3a	0a	49	06	24	5c	c2	d3	ac	62	91	95	e4	79
b	e7	c8	37	6d	8d	d5	4e	a9	6c	56	f4	ea	65	7a	ae	08
c	ba	78	25	2e	1c	a6	b4	c6	e8	dd	74	1f	4b	bd	8b	8a
d	70	3e	b5	66	48	03	f6	0e	61	35	57	b9	86	c1	1d	9e
e	e1	f8	98	11	69	d9	8e	94	9b	1e	87	e9	ce	55	28	df
f	8c	a1	89	0d	bf	e6	42	68	41	99	2d	0f	b0	54	bb	16

x

Fonte: (ZABALA, 2017)

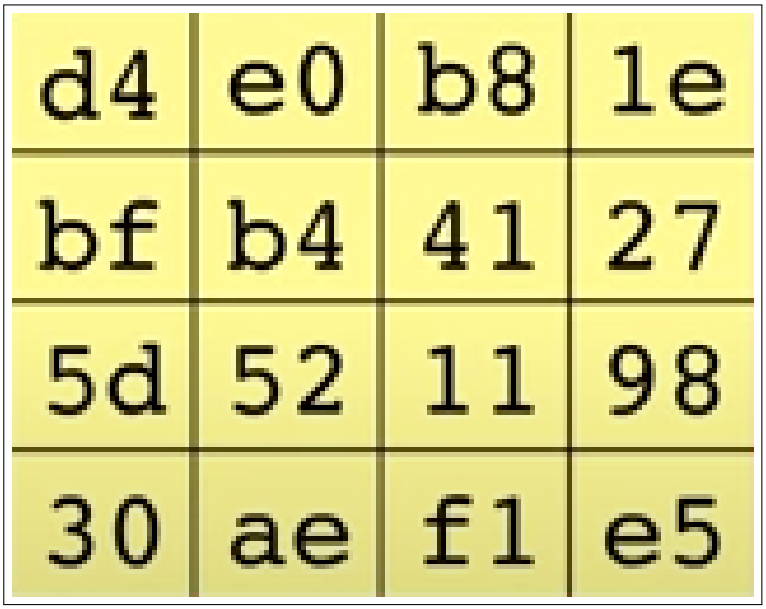
Figura 27 - Matriz original depois de passar pelo processo de SubBytes.



d4	e0	b8	1e
27	bf	b4	41
11	98	5d	52
ae	f1	e5	30

Fonte: (ZABALA, 2017)

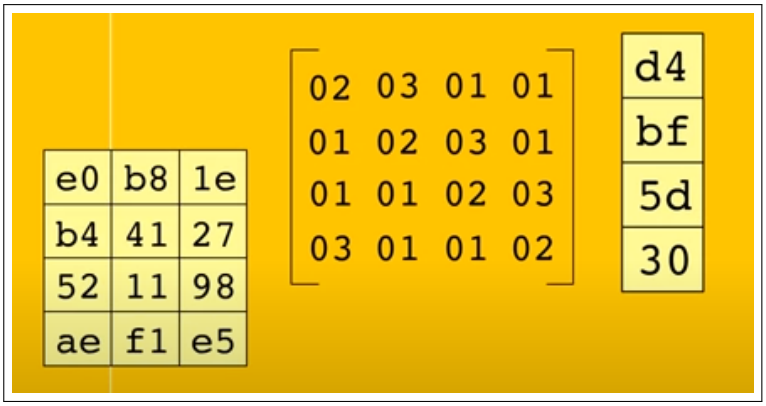
Figura 28 - Matriz de bytes depois de passar pelo processo de ShitfRows.



Fonte: (ZABALA, 2017)

a multiplicar vier de teoria matemática aprofundada, a utilização de multiplicação de matrizes é lecionada já na 2ª série do ensino médio regular em algumas escolas.

Figura 29 - Matriz de bytes exigida no processo de MixCollums.

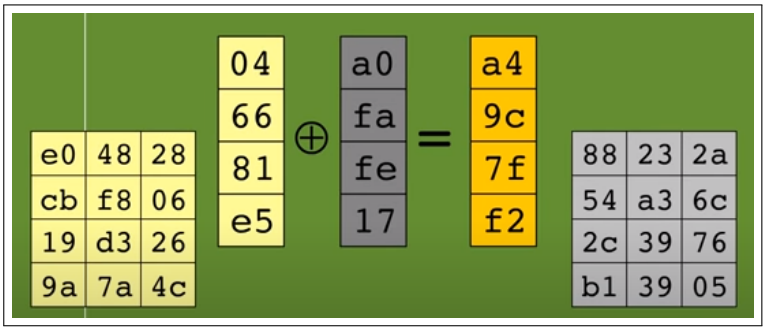


Fonte: (ZABALA, 2017)

Por último, o processo *AddRoundKey* que produz, utilizando a soma de colunas entre matrizes, uma nova coluna (figura 30). Esse processo é aplicado nove vezes e uma décima sem a implementação do *MixColumns*. O resultado prático pode ser visto na figura 22. A demonstração será efetuada a partir de imagens e com a ajuda do software.

Em primeiro lugar, instalamos o aplicativo (figura 31). Na figura 32 são vistos os comandos e o que realizam junto à sigla ao lado ao qual representa o que o comando executa. Tendo-se noção de qual comando será utilizado, o usuário pode reunir as informações e executar o código (figura 33). Nesta figura, é selecionado o *software*, depois é

Figura 30 - Matriz de bytes exigida no processo de AddRoundKey.



Fonte: (ZABALA, 2017)

passada a informação de que o arquivo “oculos.png” serve de cobertura e depois que o arquivo de texto “mina.odt” é embutido na imagem. Após todo o processo, para assegurar que a mensagem foi escondida, é realizada a análise da foto do próprio autor utilizando o Steghide (figura 34). É possível criar uma senha também para que a decodificação se torne mais dificultada (figura 35). Com o código “steghide extract - sf [nome-do-arquivo]” é possível realizar a extração dos dados. Assim, a imagem pode ser veiculada em redes sociais sem qualquer levantamento de suspeitas e preocupação com o armazenamento de senhas em caso de perda do dispositivo, pois a pessoa pode baixar a imagem por qualquer outro aparelho eletrônico.

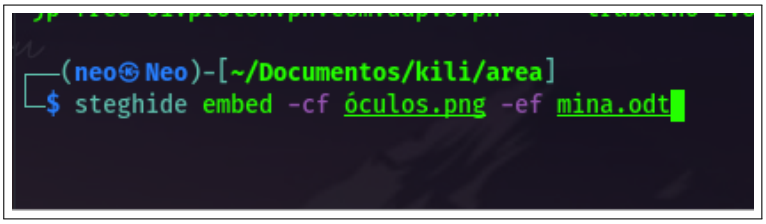
Figura 31 - Instalação do aplicativo Steghide.

Fonte: O autor, 2023.

Figura 32 - Comandos do aplicativo Steghide.

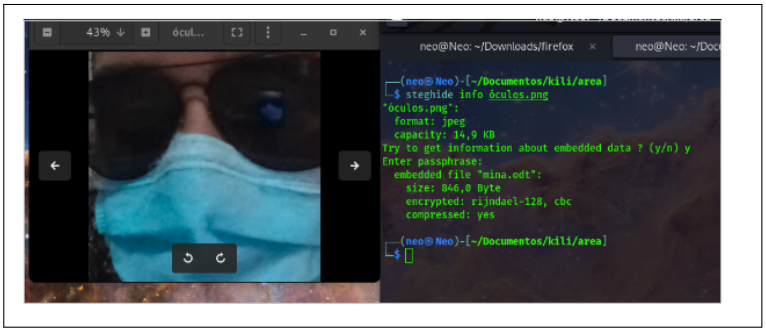
Fonte: O autor, 2023.

Figura 33 - Comandos para implementação no Steghide.



Fonte: O autor, 2023.

Figura 34 - Exemplo de esteganografia no Steghide.



Legenda: Foto do aluno utilizada para esconder suas senhas no aplicativo Steghide.
Fonte: O autor, 2023.

Figura 35 - Extração das informações no Steghide.

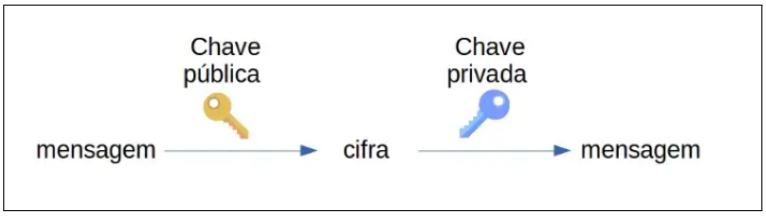
Legenda: Utilização de senha para extração das informações no aplicativo Steghide.
Fonte: O autor, 2023.

2.3.2 Criptografia Assimétrica

O sistema de Criptografia Assimétrica é um processo que utiliza uma chave pública para cifrar as mensagens e uma diferente para decifrar, chamada de chave privada.

Assim, um interceptor pode ter acesso à chave pública, mas não consegue decifrar a informação. Quando o emissor quiser receber dados criptografados deve criar a "chave pública" a qual disponibiliza para outros enviarem seus dados criptografados, mas somente ele, que criou a chave inicial, possui a possibilidade de desembaralhar o que recebeu. Com isso, soluciona-se um problema, pois agora esse método fornece garantia de confidencialidade, autenticidade e não repúdio enquanto a chave privada estiver segura. Assim, qualquer um pode enviar uma mensagem criptografada sem o risco de quebrar a confidencialidade, por uma interceptação, é mais segura e complexa, com isso também é mais lenta que a simétrica.

Figura 36 - Criptografia de chave assimétrica.



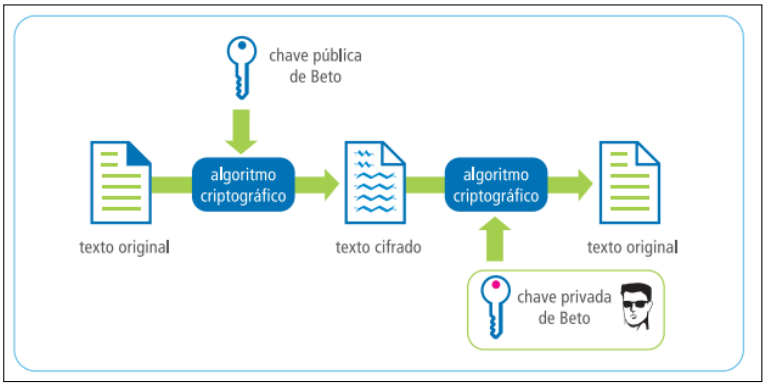
Fonte: (SOUZA, 2020)

Alguns algoritmos que utilizam chave assimétrica (VIANA et al., 2022):

- **RSA (Rivest, Shamir and Adleman):** criado em 1977, nos laboratórios do MIT, sendo esse o algoritmo mais utilizado das chaves assimétricas, através dos números primos. O processo ocorre com a multiplicação de dois números primos grandes e sua decodificação seria descobrir os dois números iniciais tentando fatorar o resultado de sua multiplicação. Algo que pode tornar essa tarefa praticamente inviável (figura 37). Os dois números primos que o geraram são a chave privada e o terceiro a chave pública.

- **ElGamal:** criado pelo egípcio Taher ElGamal em 1984, opera com a manipulação de grandes quantidades de números, também de forma cumulativa. Com o intuito de se tornar mais seguro, este faz uso do problema matemático conhecido por logaritmo discreto. Na matemática são grupos análogos a logaritmos naturais, sendo frequentemente utilizados em assinaturas digitais.

Figura 37 - Criptografia RSA.



Fonte: (LADEIRA; RAUGUST, 2017).

2.3.2.1 Aplicação atual 2: Assinatura Digital

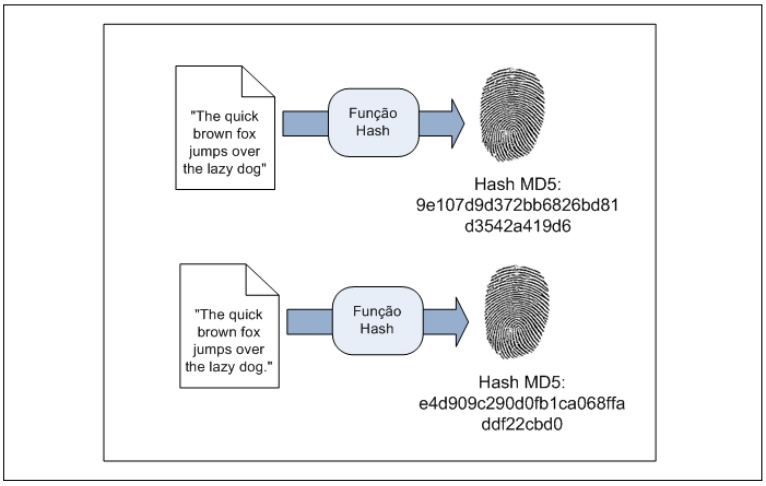
Outro mecanismo de segurança utilizado para garantir integridade de dados ou documentos é a "assinatura digital". Esta se utiliza de criptografia assimétrica. Assim, como uma assinatura de próprio punho, a digital também só pode ser reproduzida de forma autêntica por uma pessoa, entretanto todos têm acesso para verificar. Assim, a assinatura digital se baseia no processo inverso da criptografia assimétrica, pois enquanto a criptografia assimétrica utiliza duas chaves, uma pública para cifrar e outra privada para decifrar, a assinatura digital utiliza a pública para decifrar e a privada para cifrar.

Para realizar a validação dessa assinatura e saber se a informação é verdadeiramente do remetente que esperava receber ou se foi adulterada, a assinatura digital é apoiada pela função *hash*. Esta função, unidirecional quando é aplicada, gera para cada entrada uma saída única e exclusiva. Assim, uma pequena modificação em um arquivo, como a troca do valor de um bit, altera completamente o resultado da aplicação desta função (figura 38). Com isso, a função *hash* também é associada como impressão digital de um documento e por isso muito utilizado em assinaturas.

Outro exemplo que podemos citar sobre assinatura digital é a autenticação de imagens. Esse tipo de assinatura é reconhecida como marca d'água digital e seu uso pode dificultar a falsificação e distribuição ilegal de conteúdos digitais. Uma boa marca d'água deve ser invisível aos olhos humanos e interferir o mínimo possível no conteúdo original.

A autenticação de imagens utilizando este método pode ser dividida em três etapas: inserção, extração e verificação. Na inserção da marca d'água realizamos o processo chamado *Decomposição em Valores Singulares*, SVD como abreviação de seu nome em inglês. Essa técnica utiliza matrizes quadradas ortogonais, posto, conceitos de autovalor e autovetor entre outros processos e algoritmos (FURTADO; FARIA; SASAKI, 2020). Vamos recordar algumas definições e conceitos matemáticos a seguir, onde todos são baseados no livro "Álgebra Linear" (BOLDRINI et al., 1980).

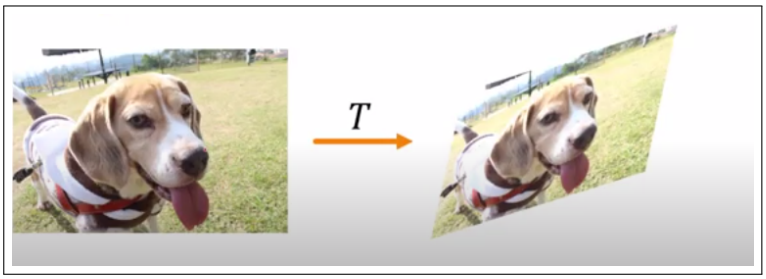
Figura 38 - Função *Hash*.



Fonte: (VIANA et al., 2022).

O exemplo retirado do canal no Youtube "Matemateca – Ester" (VELASQUEZ, 2021) supõe que estamos trabalhando com o processamento de imagens e desejamos redimensionar a imagem original como a da imagem à sua direita (figura 39). Para modificar essa imagem é necessário utilizar uma transformação linear: multiplicar a matriz de pixels da imagem por uma matriz, o que vai distorcer a imagem dada (figura à direita). Para compreendermos melhor este processo, devemos perceber que a foto original é formada por diversos vetores.

Figura 39 - Foto rotacionada de um cachorro.



Legenda: A figura mostra a imagem de um cachorro original e à sua direita a mesma foto após submissão a uma transformação linear.

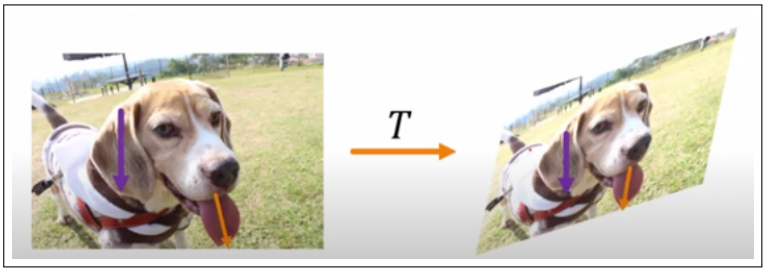
Fonte: (VELASQUEZ, 2021).

Vamos centralizar nossa atenção para a orelha e língua do animal (figura 40).

Olhar para a diferença das imagens talvez nos leve a crer que houve grande modificação, porém ao tomar os vetores como referência, não se detectam tantas alterações. Assim, o vetor da orelha se mantém na mesma direção modificando apenas seu comprimento, enquanto o da língua sofre uma modificação. Vamos entender matematicamente o que houve e para isso precisamos definir vetores no plano cartesiano.

Dados dois pontos P e Q do plano cartesiano, podemos considerar o segmento

Figura 40 - Foto do cachorro com vetores.



Legenda: A figura mostra a imagem do cachorro original e rotacionada com dois vetores.
Fonte: (VELASQUEZ, 2021).

de reta orientado $\overrightarrow{PQ}$, com ponto inicial P e ponto final Q. Considerando o segmento orientado com ponto inicial na origem, o mesmo pode ser denominado como um vetor no plano.

Usando esta correspondência entre vetores e pontos do plano, costumamos representar um vetor $v = \overrightarrow{OP}$ pelas coordenadas de seu ponto final, $P(a, b)$, ou $v = \begin{bmatrix} a \\ b \end{bmatrix}$ e também $v(a, b)$, ou simplestmente, $\overrightarrow{v}$.

Note que embora como conjunto de pontos os segmentos $\overrightarrow{PQ}$ e $\overrightarrow{QP}$ sejam iguais, como segmentos orientados eles são distintos. Com isso, os denominamos de segmentos opostos ou vetores em direção oposta.

Retornando ao nosso exemplo, se chamarmos o vetor que está na orelha do cachorro de $\overrightarrow{v}$ e multiplicarmos por uma matriz A de qualquer ordem $(A.\overrightarrow{v})$, que fará a distorção da imagem, geraremos vetores todos com a mesma direção com comprimentos diferentes, também conhecidos como vetores colineares.

Para que dois vetores sejam colineares, ou proporcionais, temos que:

- Multiplicar um vetor $\overrightarrow{v}$ por um número escalar $k > 0$ é considerar um novo vetor $\overrightarrow{w} = k\overrightarrow{v}$, que possui a mesma direção de $\overrightarrow{v}$ e tem comprimento k vezes o comprimento de $\overrightarrow{v}$.

- Multiplicar um vetor $\overrightarrow{v}$ por um número escalar $k < 0$ é considerar um novo vetor $\overrightarrow{w} = kv$, que possui mesma direção e sentido oposto de $\overrightarrow{v}$ e tem comprimento k vezes o comprimento de $\overrightarrow{v}$.

- Multiplicar um vetor $\overrightarrow{v}$ por $k = 0$ é considerar um novo vetor $\overrightarrow{w} = k\overrightarrow{v}$, como vetor nulo.

Assim, é possível observar que a transformação feita de um vetor para o outro foi multiplicar o vetor $\overrightarrow{v}$ por um $k > 0$. Note que o módulo do vetor resultado foi diminuído em relação ao vetor original, o que ocorre porque $k > 0$, porém $0 < k < 1$.

Observando agora o vetor da língua, que entitularemos como $\overrightarrow{u}$ e multiplicando pela matriz que realiza a distorção, obtemos um vetor completamente diferente em direção e comprimento. Note que utilizamos a ideia de *transformação linear* sem suas definições matemáticas, pois neste exemplo, apenas seu conceito simplificado era necessário.

Com as duas transformações realizadas, é possível definirmos autovalor e autovetor.

Dada uma matriz quadrada A de ordem n, com números reais, dizemos que um número $\lambda \in \mathbb{R}$ é um autovalor de A quando existe um vetor não nulo $\overrightarrow{v}$ tal que A $\overrightarrow{v} = \lambda \overrightarrow{v}$

Neste caso, $\overrightarrow{v}$ é dito um autovetor de A associado a λ.

Isso significa que nosso vetor $\overrightarrow{v}$, orelha do cachorro, passou pela transformação linear e seu vetor resultado foi um múltiplo do vetor original, o que nos mostra que o mesmo possui um autovetor, $\overrightarrow{v}$, pois multiplicar esse vetor pela matriz distorção A, conduz a um vetor múltiplo do original e este vetor é λ vezes o $\overrightarrow{v}$, sendo λ então o autovalor. É importante ressaltar que o exemplo citado poderia ser facilmente utilizado em aulas no currículo de Álgebra Linear, mas também se encaixa bem como mais uma aplicabilidade do conceito de vetores, que é visto somente nas disciplinas de Física no ensino médio na área de Cinemática.

Em posse dos conceitos gerais de autovalor, autovetor e transformações, tendo consciência de que eles são necessários para o algoritmo de SVD, é possível abordarmos agora a marca d'água, mais uma aplicação atual de assinatura digital. Para recordar, a decomposição SVD pode ser utilizada para autenticação de imagens em formato de marca d'água.

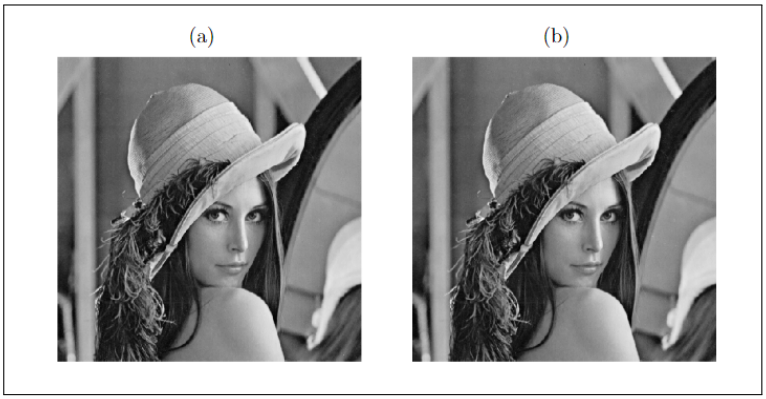
O processo de inserção da marca d'água é realizado tomando-se a decomposição SVD da imagem hospedeira e da imagem utilizada como marca d'água. Em seguida, modificam-se os valores singulares da imagem hospedeira e se reconstrói a imagem autenticada. Somente os bits menos representativos são modificados.

Os processos de verificação e extração podem ser realizados aplicando operações inversas, isto é toma-se a decomposição SVD da imagem autenticada para a imagem hospedeira. Em seguida, calculam-se os valores singulares da marca d'água a ser extraída da imagem autenticada e se reconstrói a marca d'água. A figura 41 mostra um exemplo da autenticação por marca d'água.

Já a figura 42 mostra a marca d'água que foi inserida na figura 41, qual o resultado de ter ela sido extraída da imagem autenticada e como ela fica sendo extraída da imagem sem autenticação.

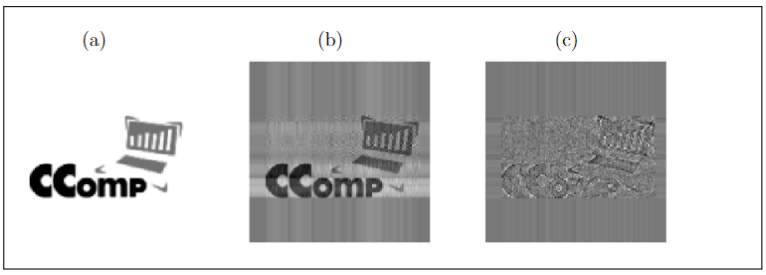
É visto então que em um mundo altamente conectado, a criptografia é indispensável para garantir a nossa segurança em todos os meios que facilitam nosso cotidiano. Alguns exemplos que podemos citar são: quando se utiliza o aplicativo de banco, arquivos na nuvem, *e-mails* e até uma simples mensagem SMS. Ela está presente também nas redes sem fio para que permita a conexão somente a quem inserir a senha correta. As redes sociais também a utilizam na autenticação de seus usuários, como a criptografia de

Figura 41 - Foto da imagem original (a) e (b) imagem autenticada com a marca d'água.



Fonte: (FURTADO; FARIA; SASAKI, 2020).

Figura 42 - Marca–d'água original (a), (b) extraída de imagem autenticada e (c) extraída de imagem não autenticada.



Fonte: (FURTADO; FARIA; SASAKI, 2020).

ponta a ponta utilizada pelo *Whatsapp* que seria para que nem a empresa pudesse ler as mensagens de seus usuários. A criptografia de ponta a ponta é do tipo assimétrica, porém existe um ponto de atenção quando falamos do *backup* em nuvem. Essa criptografia protege as informações em casos de interceptações durante a troca de mensagens e quando são armazenadas na nuvem em dispositivos Android, por meio do *Google Drive* para os *backups*, é utilizada a criptografia AES de 256 bits (GOOGLE, 2024). Já para os usuários Apple, que se utilizam do ICloud, os dados são criptografados em sistema de ponta a ponta, chamado de Proteção de Dados Avançada (APPLE, 2024). Vale ressaltar que essa criptografia é a mesma utilizada por vários aplicativos na atualidade. Comprovando a força da criptografia, as chamadas criptomoedas só podem ter seu modelo descentralizado graças à segurança que esse sistema confere a quem as utiliza.

É notável que os sistemas de computadores e invasores venha evoluindo tanto em hardware quanto em software e, com isso, tem sido necessário o desenvolvimento contínuo da cibersegurança. É indiscutível que as tecnologias quânticas são as mudanças mais fortes no paradigma do modelo computacional. Ao passo que a estrutura principal da computação clássica é o bit, na computação quântica a estrutura se chama qubit (bit quântico). Enquanto um bit clássico sempre armazena apenas um valor, 0 ou 1, o qubit corresponde a um conjunto de bits, podendo assumir 0, 1 ou uma superposição linear dos dois estados, ou seja, ser 0 e 1. Em termos práticos, enquanto a informação total armazenada pelos bits é igual à soma direta deles ($1 + 1 + 1 + 1 ... = n$), a informação armazenada pelos qubits cresce exponencialmente (2 x 2 x 2.... $= 2^n$). Essa diferença já nos traz diversos desafios para a área de criptografia, mas mais em geral ainda, modifica todo o sistema de algoritmos ao qual estamos adaptados. Abrindo assim uma gama gigantesca de futuros problemas e propostas em todas as áreas de ciências computacionais.

2.4 **Sequência Didática deste capítulo**

A proposta desta sequência tem como objetivo principal a compreensão dos métodos criptográficos e a matemática por trás desses algoritmos.

1. Para alcançar esse objetivo, inicialmente seria feita uma introdução sobre as aplicações de criptografia, como seu uso em aplicativos de celular, senhas de banco, entre outros. Porém, para melhor entender o avanço tecnológico que esta área proporcionou, é necessário contextualizarmos este conteúdo para o aluno de forma a motivá-lo, apresentando o problema com o qual a sociedade se deparou inicialmente e, brevemente, o que a sociedade estava vivendo nesse período histórico. Essa abordagem é interessante para que o aluno compreenda que a ciência, por inúmeras vezes, é criada por diversos motivos, reviravoltas, problemas que envolvem seu criador, e que o cientista, ao contrário do mito que se criou, não faz descobertas trancando-se

em seu escritório ou laboratório, completamente afastado da realidade que o cerca. Em meu planejamento, esse tópico seria desenvolvido em 2 tempos de aula.

2. Após essa contextualização, é possível explorarmos os métodos criptográficos. Assim, é possível abordar o tema de criptografia assimétrica. Citaria seus principais algoritmos e comentaria sobre suas vantagens e desvantagens. Dentro deste mesmo encontro, abordaríamos uma de suas aplicações práticas, a esteganografia. Mostramos o desenvolvimento do trabalho já realizado utilizando foto e o aplicativo Steghide, e temos a proposta para ser replicado em um laboratório de informática. Essas atividades podem ser desenvolvidas em 4 tempos de aula.

3. No encontro seguinte, explorar a matemática que está sendo utilizada no algoritmo. Com isso, explicamos os passos que a técnica empregada pelo Steghide usa para fazer a substituição dos bits na imagem da foto proposta pelos alunos. Além disso, durante esse processo de compreensão do algoritmo, revisitaríamos alguns temas importantes de matemática básica, como transformação entre números hexadecimais e binários, algoritmo de Euclides, aritmética modular; conceitos que, em muitos cursos, são lecionados apenas teoricamente, ou sem observarmos sua conexão entre outras disciplinas do curso. Este tema é planejado para ser apresentado durante cerca de 8 tempos de aula.

4. No próximo encontro, é discutido sobre criptografia assimétrica. Assim, como na anterior, citamos seus principais algoritmos, vantagens e desvantagens. O exemplo prático a ser abordado é o de assinatura digital. Utilizado para criar marca d'água em imagens, a matemática vista para implementar esse método passa por conceitos da álgebra linear. Em nosso caso, autovalor e autovetor. Com a imagem de uma foto sendo rotacionada, podemos mostrar aos alunos quais são as aplicações práticas desses vetores e também perceber neste momento a importância de sabermos esse conteúdo e definirmos tais conceitos. Este tópico teria seu planejamento para 8 tempos de aula.

5. Por fim, há uma leve introdução sobre a diferença de um *bit* e o *qubit*, e como este último pode revolucionar toda a estrutura da computação à qual já estamos adaptados. Acredito que este tema é apenas um despertar para reflexões futuras que envolvem a criptografia. Para isso, um tempo de aula é o suficiente.

Plano de aula

Tema: Introdução à Criptografia.

Público-alvo: alunos de graduação nas disciplinas de Álgebra II e Álgebra Linear I.

Objetivos: Introduzir os conceitos fundamentais da criptografia, abrangendo sua história, tipos de chave, criptografia simétrica e assimétrica, criptografia de Rijndael, esteganografia, autovalores e autovetores, e conversão entre sistemas numéricos binários e hexadecimais.

Duração do curso: 15 tempos de aula.

Conteúdos:

- Introdução à criptografia antiga (Egito, Grécia e Roma);
- Criptografia na idade média (cifras de substituição e transposição);
- A criptografia moderna (máquina Enigma, criptografia de chave pública e privada);

Metodologia: aula expositiva, com uso de tecnologia de apoio, como projeção de *slides* sobre a matéria; apostilas e livro.

Avaliação:

- Projeto integrando os conceitos aprendidos nos diferentes módulos;
- Questões discursivas referente aos temas.

O que aprendemos neste capítulo?

No Capítulo 2 passamos pelos principais pontos na história do desenvolvimento da técnica de Criptografia. Em seguida, é frisado em como essa tecnologia está na atualidade abordando alguns algoritmos utilizados. Dentro de uma das técnicas, a criptografia simétrica, ilustramos em um exemplo prático como pode ser utilizada dentro de Esteganografia. Com outra técnica, a criptografia assimétrica, mostramos como ela é utilizada em assinaturas digitais e em nosso exemplo específico, como marca d'água. Todo esse entendimento é importante para que vejamos que este campo da Matemática e Ciências Computacionais possuem muitas aplicações cotidianas, sendo uma delas o Bitcoin, que será abordado no próximo capítulo.

3 APLICAÇÃO ATUAL 3: AS CRIPTOMOEDAS E SEU MODELO ECONÔMICO

É inegável o papel modificador da Internet em toda sociedade atual. A mesma modificou a forma de se comunicar, de se trabalhar, de se divertir e até de se pensar sobre o dinheiro, este último, podendo ser facilmente transferido sem nenhuma barreira e controle aparente. Cada vez se torna mais raro uma pessoa ir à sua agência bancária para retirar dinheiro ou pagar suas contas. A velocidade com que a sociedade necessita produzir demonstra incabível a possibilidade de aguardar horas numa fila de atendimento.

Tornando-se o tempo um bem escasso e a Internet modificando a forma de se relacionar com os compromissos financeiros, gerou-se a necessidade do desenvolvimento da moeda digital. A moeda digital permite ao usuário final efetuar transações de pagamento, de transferência de moeda, pagamento de contas, tudo isso ainda sendo gerido por um banco da escolha do usuário. As criptomoedas são totalmente digitais (SILVA, 2017). Não existe a possibilidade de ir a um banco para sacá-la, ela não possui cédulas, mas pode ser utilizada como dinheiro em poucas transações. Muitos confundem ainda moeda digital com criptomoedas, sendo que esta última possui como sua principal característica a independência da economia de qualquer estado, ou seja, não há um banco regulador das transações feitas por essa modalidade, sendo necessário apenas o acesso à internet (SILVA, 2017). Em contrapartida, a moeda digital ainda é gerida por um banco. Vale ressaltar que não há qualquer tipo de regulamentação a respeito do uso de Criptomoedas, o que é visto de forma amedrontadora por uns – já que não existe inicialmente uma forma de rastrear as transações –, enquanto preocupa outros o futuro da economia por conta de impostos ou processos políticos em determinados países, que poderiam trazer flutuações ao valor de tal moeda. Existiam, no ano de 2021, entre 10000 a 15000 criptomoedas (MALAR, 2021) e foi decidido neste trabalho explorar apenas aquela denominada Bitcoin (NAKAMOTO, 2008), por ser a primeira criada e ser a mais famosa.

3.1 **Bitcoin**

O protocolo *Bitcoin* foi originalmente anunciado em um artigo publicado em novembro de 2008 por Satoshi Nakamoto (2008); surgia assim uma nova forma de se pensar em dinheiro, que não dependia mais de um governo ou banco para produção de sua moeda. Com essa proposta, a pergunta que se fazia era como as transações e posse desse dinheiro seriam validadas. É necessário entender que, mesmo muitos sendo aversos a ir ao banco, são eles que garantem alguns direitos, como extorno de uma compra, restituição do dinheiro caso haja roubo, entre outros. Como haveria garantia das transações, ou quem

possui essa moeda, já que a mesma é totalmente digital?

No sistema tradicional, sabe-se que quem tem o dinheiro é quem possui a cédula ou a conta bancária, sendo sinalizado pelo banco qual o valor. Quanto ao Bitcoin, não são armazenadas de forma centralizada ou local e, portanto, não existe um dono. Eles existem como registros em um livro contábil distribuído chamado *Blockchain*, cujas cópias são compartilhadas por uma rede voluntária de computadores conectados. Assim, possuir uma *Bitcoin*, simplesmente significa ter a capacidade de transferir o controle para outra pessoa, criando um registro da transferência na *blockchain* (RIKWALDER, 2014).

Para compreendermos melhor o que é o *blockchain*, é utilizado aqui algumas analogias e definições retiradas do documentário "The Giant Beast that is the Global Economy", disponível no catálogo da Amazon Prime Video. Imagine que transações financeiras tradicionais são como o casamento. Dois indivíduos se unem e trocam algo de valor, como um anel. E essas transações necessitam de uma licença de um Estado e um oficial ordenado para verificá-las, "um padre", por exemplo. Isso as torna centralizadas. O banco é o padre que tradicionalmente verifica a transação. Por alguma razão, os jovens de hoje não se unem apenas por meio do casamento tradicional. Alguns ficam satisfeitos em se unirem na frente de testemunhas de confiança que certificam, mesmo que involuntariamente, a concepção da união tirando fotos, marcando os noivos e publicando-as *on-line*, acessíveis a todos na internet. Compreendido isto, podemos ver os noivos como duas pessoas que querem trocar *bitcoins* e os convidados tirando as fotos como as pessoas que garantem a segurança e confiabilidade da operação entre eles. Esta é a ideia central do *blockchain*, utilizar pessoas ao invés de um banco regulador para as operações com a moeda. Com esse conceito um pouco mais esclarecido, é interessante se entender também como pode-se comprar *bitcoins*. Isso pode ser feito trocando criptomoedas, mas antes que isso aconteça as moedas devem ser criadas digitalmente. Não há nenhuma instituição física para cunhá-las ou uma reserva federal que controle o suprimento, como o papel-moeda. Ao invés disso, o Bitcoin encontrou um modo de gerar digitalmente um número finito de moedas com o poder da computação. Esse processo se chama *mineração*. Com isso, mineradoras de criptomoedas surgiram por todo o mundo. Operam constantemente computadores poderosos em uma corrida para gerar e acumular criptomoedas. Retornando à analogia do casamento, os mineradores são os convidados que verificam as transações tirando fotos, marcando os noivos e postando *on-line*. O Bitcoin é criado como incentivo para que os mineradores construam e mantenham o *blockchain*. O *Bitcoin* é entregue a um minerador de sorte. Isso acontece de forma constante, então outros mineradores também o receberão. Quanto mais trabalham para manter o *blockchain*, mais chances têm de serem recompensados com a moeda.

Com esse sistema inovador, descentralização de um governo que não influencia no valor da moeda e difícil capacidade de rastreamento, o *Bitcoin* se mostra à sociedade como a evolução dos sistemas bancários e isso é refletido no valor da moeda. Segundo o

Tabela 5 - Valor do Bitcoin durante os anos.

Período	Preço do Bitcoin
Fevereiro de 2011	1
8 de julho de 2011	31
Dezembro de 2011	2
Dezembro de 2012	13
Abril de 2013	266
Maio de 2013	130
Junho de 2013	100
Novembro de 2013	350 – 1250
Dezembro de 2013	600 – 1000
Janeiro de 2014	750 – 1000
Fevereiro de 2014	550 – 750
Março de 2014	450 – 700
Abril de 2014	340 – 530
Maio de 2014	440 – 630
Março de 2015	200 – 300
Novembro de 2015	395 – 504
Fevereiro de 2017	1222
Março de 2017	1270
Agosto de 2017	4400
1 de setembro de 2017	5000
12 de setembro de 2017	2900
13 de outubro de 2017	5600
21 de outubro de 2017	6180
8 de dezembro de 2017	18300
17 de Dezembro de 2017	19900
2 de fevereiro de 2018	6375
20 de Fevereiro de 2018	11785
Julho de 2018	9900 – 5995
19 de novembro de 2018	5000 – 4900
Janeiro de 2019	3700 – 3400
Abril de 2019	5000
Julho de 2019	12958
Janeiro de 2020	6844 – 9000
Fevereiro de 2020	10400 – 8543
Março de 2020	3800 – 6400
Abril de 2020	9100 – 6000
8 Junho de 2020	9141
1 Agosto de 2020	11727
Novembro de 2020	18000
Dezembro de 2020	28000

Legenda: Os valores estão sendo calculados em dólar.
Fonte: (BIT2ME, 2020).

site Bit2me (BIT2ME, 2020), em 2009, quando houve a primeira transação, um *Bitcoin* custava 0,00076392 dólar, ou seja, bem menos que 1 centavo de dólar. Em julho de 2010 o custo era de 0,08 dólar, um aumento de 10472%. Na tabela 5 temos o acompanhamento de alguns valores marcantes da moeda.

Com a compreensão inicial do que são criptomoedas e, em especial, o Bitcoin, além de notar que com o passar do tempo, a mesma só vem se valorizando, parece ser interessante a possibilidade de prever qual seria seu valor em um espaço curto de tempo para saber se ainda vale a pena investir nesse ativo, ou se o melhor momento para o investimento já ocorreu. Para prever esses valores, pode-se utilizar o conceito de função. Todavia, ao entrar em contato com essas definições, os alunos muitas vezes apresentam dificuldade na compreensão de seu significado. Essa dicotomia, de não serem as funções inseridas em um contexto real e atual, sinaliza a necessidade de se levar em consideração recursos semióticos que aliam seu uso prático e teórico. No contexto de sala de aula, unir recursos linguísticos e extralinguísticos valendo-se da oralidade, da escrita e até mesmo de ferramentas da tecnologia digital, como imagens e gráficos são considerados recursos semióticos (ARZARELLO, 2006). Essa vertente fortifica a utilização da área de Modelagem Matemática, já que esta perspectiva pode ser percebida como elemento integrador entre a realidade e o conteúdo matemático a ser ensinado.

3.2 **Modelagem Matemática**

Segundo Bassanezi (2012), a ciência é desenvolvida pelo ser humano que procura compreender melhor os fenômenos que o cercam. As ciências da natureza evoluíram com o auxílio da Matemática, mas outras ciências, entre as quais Biologia, Psicologia, Economia, empregavam apenas a linguagem para seu desenvolvimento. Tal abordagem conduziu à imprecisão e falta de respaldo para se comprovar determinados argumentos e interpretações de alguns fenômenos. Com o fluir do tempo, a Matemática passou a ser inserida utilizando apenas alguns suportes estatísticos. Devido ao crescimento científico, propostas de modelos matemáticos são vistas em muitas áreas nas quais antes não eram aplicadas. Assim, aliaram de forma equilibrada a abstração e a formalização, não perdendo de vista as características das respectivas áreas do conhecimento. Esse processo se convencionou chamar de Matemática Aplicada e se iniciou, declaradamente, no século XX, ganhando força após a segunda guerra mundial. Quando se procura refletir sobre uma parte da realidade, na tentativa de explicar, de entender, ou de agir sobre ela, formaliza-se por meio de um modelo. O método científico é formulado baseado em experimentos que devem ser bem descritos, de tal forma a possibilitar serem reproduzidos, com a exigente comparação empírica, trazendo assim a ajuda da matemática ou da lógica para inferir mais afirmações sobre o estudo. Com o passar do tempo, muitas modificações alteraram os métodos

existentes, surgiram outros novos, mas são alcançados os objetivos quando se cumprem ou se propõem a cumprir as seguintes etapas (MARCONI; LAKATOS, 2003):

- seleção do problema;
- definição precisa do problema;
- procura de conhecimentos ou instrumentos relevantes ao problema;
- tentativa de solução do problema com auxílio dos métodos identificados;
- geração de novas ideias (hipóteses, teorias ou técnicas) ou produção de novos dados empíricos que prometam resolver o problema;
- obtenção de uma solução exata ou aproximada;
- investigação das consequências a partir da solução obtida. Em se tratando de uma teoria, é a busca de prognósticos que induzam ser feitos com seu auxílio. Em se tratando de novos dados, é o exame das consequências que possam ter para as teorias relevantes;
- prova (comprovação) da solução. Se o resultado é satisfatório, a pesquisa é dada como concluída, até novo aviso. Do contrário, passa-se para a etapa seguinte;
- correção das hipóteses, teorias, procedimentos ou dados empregados na obtenção da solução insatisfatória. Esse é, naturalmente, o começo de um novo ciclo de investigação.

Essas etapas podem ser vistas também como na figura 43.

Assim, a modelagem é o processo de criação de modelos, onde estão definidas as estratégias de ação do indivíduo sobre a realidade. Já a Modelagem Matemática é a validação utilizando modelos matemáticos.

3.2.1 Modelagem Matemática no ensino

D'Ambrosio (1991) argumenta que quando o professor aplica a modelagem como estratégia pedagógica na sala de aula, explorando as aplicações matemáticas no dia a dia, a construção de modelos e o relacionamento entre a matemática utilizada na modelagem e o conteúdo programático, o professor oferece ao aluno a oportunidade de conviver com conteúdos vivos, práticos, úteis e com bastante significado.

A modelagem aplicada ao ensino pode ser um caminho para despertar maior interesse, ampliar o conhecimento do aluno e auxiliar na estruturação de sua maneira de

Figura 43 - Método Científico.

Legenda: As etapas do método mostram que é um processo cíclico.
Fonte: (MENEZES, 2022).

pensar e agir (BASSANEZI, 2002). D'Ambrosio (1991) defende ainda que, em relação às escolas, o maior desafio dos matemáticos e educadores matemáticos é "fazer uma Matemática integrada no pensamento e no mundo moderno" e aponta a Modelagem Matemática como um caminho para contribuir no enfrentamento deste desafio.

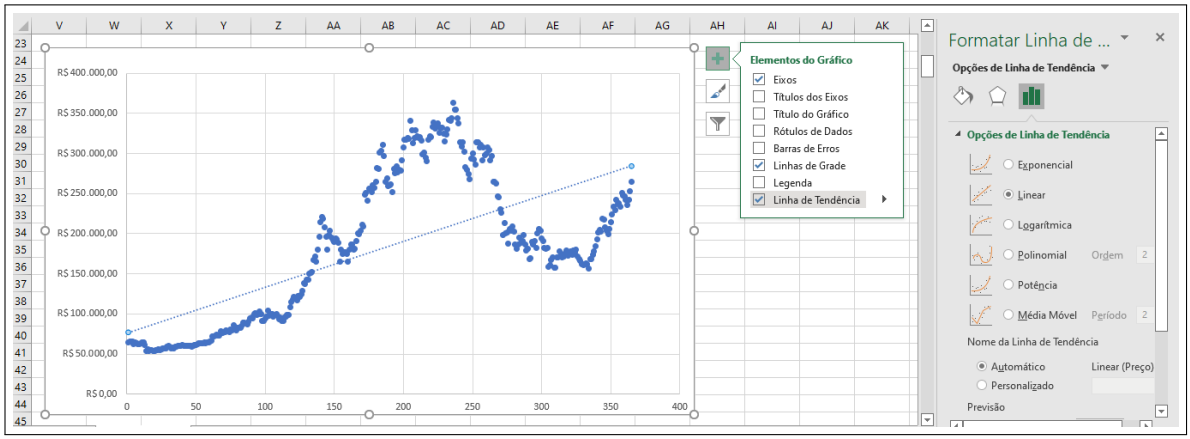
3.2.2 Ajustes de curvas

Nesta seção são definidos os modelos de ajuste linear, exponencial, polinomial de grau 2 e logarítmico, e apresentados os resultados obtidos ao ajustar os valores do preço do *Bitcoin* utilizando esses modelos. Essa escolha é feita por serem funções estudadas e aprofundadas na 1ª série do Ensino Médio. As definições e suas devidas demonstrações são baseadas no livro escrito por Rodney Carlos Bassanezi (2002). A obtenção dos modelos é feita via a ferramenta interna "*Linha de Tendências*" do aplicativo Excel (veja a figura 44).

Ajustar uma curva significa determinar os coeficientes de uma função associada a essa curva, de modo que, no intervalo de valores considerados, esta função e os dados coletados experimentalmente sejam "próximos". Note que só podemos garantir a proximidade entre a curva ajustada e os pontos dados no intervalo onde tais pontos foram observados. Realizar previsões de valores futuros utilizando estes modelos, que é o principal objetivo de uma modelagem, sempre possui embutido um desvio assim, nem sempre essa previsão se confirma em um resultado satisfatório. De qualquer forma, os modelos conduzidos

nesse intervalo limitado são fundamentais para o processo.

Figura 44 - Ferramenta Linha de Tendência do aplicativo Excel.



Legenda: Observe no quadro à direita as opções de funções que o aplicativo permite utilizar.

Fonte: O autor, 2022.

Um dos métodos mais usados para estimação dos parâmetros de uma função é o Método dos Mínimos Quadrados – MMQ (BASSANEZI, 2002, p.57) – com o qual podemos ajustar as curvas desejadas.

Considere um conjunto de n dados observados $\{(\overline{x}_i, \overline{y}_i)\}, i = 1, 2, 3, \cdots, n$ e uma função $y(x) = f(x, a_1, a_2, \ldots, a_j)$, onde a_j $(j = 1, 2, \ldots, j)$ são parâmetros. O Método dos Mínimos Quadrados (MMQ) consiste em determinar estes parâmetros de modo que se minimize o valor de

$$S = \sum_{i=1}^{n} (y_i - \overline{y}_i)^2 = \sum_{i=1}^{n} [f(\overline{x}_i, a_1, a_2, \ldots, a_j) - \overline{y}_i]^2, \tag{1}$$

isto é, deve-se minimizar a soma dos quadrados dos desvios entre os $\overline{y}_i$ observados e os valores ajustados $y_i = f(\overline{x}_i, a_1, a_2, \ldots, a_j)$.

Assim, devemos encontrar os valores dos parâmetros coerentes com os dados observados nos ajustes que são definidos no próximo capítulo.

Com o método dos mínimos quadrados, pode-se modelar a curva que melhor se adapta ao problema. A proposta mais simples é a de ajuste linear, que é procurar a reta que melhor se aproxima dos dados.

3.2.2.1 Ajuste linear

Um ajuste é linear se for da forma

$$f(x) = ax + b\,, \quad \text{onde} \quad a, b \in \mathbb{R}, \tag{2}$$

ou seja, é a equação de uma reta, e devemos encontrar os valores dos parâmetros a e b que tornam mínimo o valor da soma dos quadrados dos desvios S,

$$S = \sum_{i=1}^{n} (b + a\overline{x_i} - \overline{y_i})^2\,, \tag{3}$$

onde $\overline{x}_i$ e $\overline{y}_i$ são as médias dos valores observados. No caso proposto neste trabalho, $\overline{x}_i$ seria a média dos dias e $\overline{y}_i$ a média dos valores correspondentes da moeda nesses dias selecionados.

Tais valores devem satisfazer, necessariamente, às condições de minimalidade de S:

$$\begin{cases} \dfrac{\partial S}{\partial b} = 0 \;\Leftrightarrow\; \sum\limits_{i=1}^{n} 2(b + a\overline{x}_i - \overline{y}_i) = 0 \\ \\ \dfrac{\partial S}{\partial a} = 0 \;\Leftrightarrow\; \sum\limits_{i=1}^{n} 2\overline{x}_i(b + a\overline{x}_i - \overline{y}_i) = 0 \end{cases}, \tag{4}$$

ou seja,

$$\begin{cases} a = \dfrac{n\sum \overline{x}_i\overline{y}_i - \sum \overline{x}_i \sum \overline{y}_i}{n\sum \overline{x}_i^2 - (\sum \overline{x}_i)^2} = \dfrac{\sum \overline{x}_i\overline{y}_i - n\overline{xy}}{\sum \overline{x}_i^2 - n\overline{x}^2} \\ \\ b = \dfrac{\sum \overline{x}_i^2 \sum \overline{y}_i - \sum \overline{x}_i\overline{y}_i}{n\sum \overline{x}_i^2 - (\sum \overline{x}_i)^2} = 0 \Leftrightarrow b = \dfrac{\sum \overline{y}_i}{n} - a\dfrac{\sum \overline{x}_i}{n} = \overline{y} - a\overline{x} \end{cases}. \tag{5}$$

Quando se realiza um ajuste linear para relacionar duas variáveis, não sabemos inicialmente se o modelo encontrado é, de fato, o melhor. Um dos critérios mais utilizados para verificar a existência e o grau de relação entre as variáveis é o coeficiente de correlação.

A correlação linear mede a relação existente entre as variáveis x e y, através do posicionamento dos pontos $(x_i,\, y_i)$ dados, em torno de uma reta ajustada $y = ax + b$.

O coeficiente de correlação de Pearson R é um instrumento de medida de associação, ou força, do grau de relação entre duas variáveis. Existem algumas formas de se escrever este cálculo. Utilizaremos a que se desenvolve utilizando alguns conceitos vistos no ensino médio e cursos introdutórios de Estatística. Para isso, é necessário definir Desvio Padrão Populacional. As equações à seguir são adaptadas do livro "Conexões: ma-

temática e suas tecnologias" (MODERNA, 2020). Para compreendermos esta definição relembramos a definição de média aritmética.

Média aritmética é o quociente entre a soma dos valores observados e o número de observações. Indicamos a média aritmética por $\overline{x}$,

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}, \tag{6}$$

em que $x_1, x_2, \ldots, x_n$ são valores que a variável assume e n é a quantidade de valores no conjunto de dados.

Com a média bem definida, podemos definir desvio padrão.

O desvio padrão é a raiz quadrada da média aritmética dos quadrados dos desvios. Indicamos o desvio padrão por DP,

$$DP = \frac{\sqrt{(x_1 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2}}{n} = \frac{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2}}{n}, \tag{7}$$

em que $(x_1 - \overline{x})$ é chamado de desvio.

Podendo também definir Covariância. A covariância, ou *Cov*, pode ser calculada como os desvios das duas variáveis que desejam ser estudadas se possuem correlação.

$$Cov_{x,y} = \frac{(x_i - \overline{x}) \times (y_i - \overline{y})}{n} \tag{8}$$

Note que o DP nada mais é do que a soma desses desvios, no caso seus valores absolutos, sendo calculada a raiz quadrada. Ou seja, os desvios são como os erros de cálculo se comparados à média e o desvio padrão nos apresenta o quanto os valores estão distantes da média aritmética dessa distribuição. Assim, quanto mais próximo do valor zero, menor o "erro" da média.

Já a Variância é definida como a média aritmética dos quadrados dos desvios. Indicamos a variância por Var,

$$Var = \frac{(x_1 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2}{n} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n}. \tag{9}$$

Note que o DP é exatamente a raiz quadrada da Variância. Utilizamos o DP apenas para retornar o resultado obtido na mesma escala da variável. Por exemplo, em um problema envolvendo massa, é como se encontrássemos a variância em quilogramas

ao quadrado e o DP apenas retorna o resultado em quilogramas.

Tendo retomado tais definições, pode-se escrever a correlação linear de Pearson como:

$$R = \frac{Cov_{x,y}}{DP_x \times DP_y}, \tag{10}$$

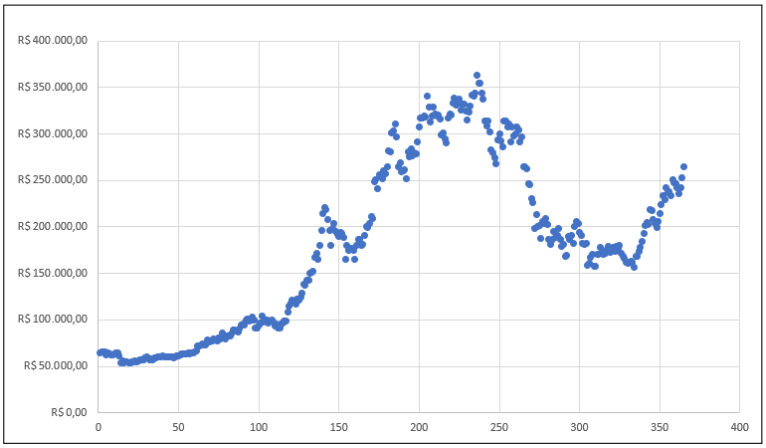
onde DP_x é o desvio-padrão da variável x e DP_y é o desvio-padrão da variável y.

O intervalo de variação de R é $[-1, 1]$, isto é, $-1 \leq R \leq 1$.

A correlação será tanto mais forte quanto mais próximo R estiver de 1 ou de -1. Ou seja, caso a correlação se aproxime de 1 significa que as grandezas estão tendendo a ser diretamente proporcionais (quando uma aumenta, a outra tende a aumentar) e caso o valor cada vez se aproxime mais de -1, as grandezas são inversamente proporcionais (caso uma aumente a outra tende a diminuir ou vice-versa). Se $R = 0$, então não existe nenhuma correlação entre as variáveis ajustadas. O sinal de R é o mesmo sinal do coeficiente angular da reta modelada, assim, quando encontramos valores para R próximos a 1 estamos encontrando o coeficiente angular da reta modelada mais próximo de $f(x) = x$ e quando próximo de -1 a reta modelada se aproxima de $f(x) = -x$, por isso a conclusão de serem diretamente ou inversamente proporcionais.

Neste trabalho, para o ajuste linear, a variável independente x representa os dias contados a partir de vinte e dois de agosto de dois mil e vinte, sendo este o dia um, o dia vinte e três de agosto de dois mil e vinte como dia dois e assim sucessivamente. Já a variável y representa os valores do *Bitcoin* nesses respectivos dias a zero hora medida na moeda Real (INFOMONEY, 2021). A figura 45 mostra esse levantamento.

Figura 45 - Valores durante um ano da moeda *Bitcoin*.

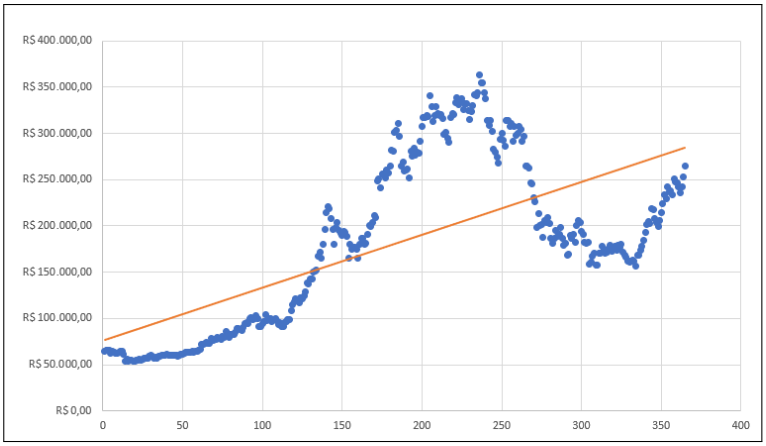


Legenda: Valores apresentados do dia 22/08/2020 ao dia 22/08/2021.
Fonte: O autor, 2022.

Pode-se visualizar a reta na figura 46. É possível observar que em certos intervalos,

o modelo se distancia bastante do valor observado. Para saber se esta reta se encaixa bem ao problema proposto, é possível calcular o coeficiente de correlação de Pearson. Este coeficiente é de grande ajuda para se determinar se as grandezas são proporcionais. Porém, existe um outro coeficiente que procura explicar a variável y em função da variável x. Este se chama coeficiente de determinação.

Figura 46 - Valores durante um ano da moeda Bitcoin com ajuste linear.



Legenda: O ajuste linear é representado pela reta de cor alaranjada no gráfico.
Fonte: O autor, 2022.

O coeficiente de determinação tem sua definição mais geral como R^2, onde R é o coeficiente de correlação de Pearson.

Num contexto de um modelo de regressão linear simples, em que a variável explanatória (ou preditora) é x e a variável resposta (ou a prever) é y, o coeficiente de determinação R^2 dá a porcentagem de variabilidade dos y's (variável a prever), que fica explicada em função da variabilidade dos x's. Assim, um valor de $R^2 \approx 1$ significa que, em princípio, a nuvem de pontos apresentada no diagrama de dispersão está próxima da reta de regressão, considerada para o modelo de regressão. Quando $R^2 \approx 0$ já não se vislumbra uma estrutura linear (MONTGOMERY; RUNGER, 2013).

Embora esta medida, como dito anteriormente, seja normalmente utilizada como uma indicação da adequação do modelo de regressão ao conjunto de pontos inicialmente dado, ela deve ser usada com precaução, pois nem sempre um valor de R^2 grande (próximo de 1) é sinal de que o correspondente modelo ajusta bem os dados. Do mesmo modo, um valor baixo de R^2, pode ser provocado por um *outlier*, enquanto a maior parte dos dados se ajustam razoavelmente bem a uma reta. Uma visualização prévia dos dados num diagrama de dispersão é fundamental.

Em tese, quanto maior o R^2, melhor o modelo de ajuste se adapta à amostra. No caso do ajuste feito, encontra-se a função

$f(x) = 569,21x + 76500\,.$

Esta função possui $R^2 = 0,4396$. Esse valor demonstra que o modelo está distante de se encaixar aos dados observados nesse intervalo.

3.2.2.2 Ajuste linear de crescimento exponencial

Outro modelo que pode ser utilizado para modelar o problema de forma smplificada é o da função exponencial

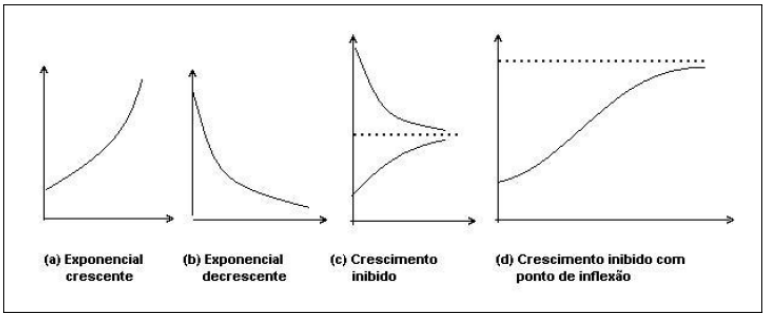
$$f(x) = be^{ax}, \quad b > 0 \quad e \quad a \in \mathbb{R}. \tag{11}$$

Se for considerada a mudança de variável $z = lny$, teremos a equação 11 na forma de uma reta:

$$z = ln\, y = ax + ln\, b, \tag{12}$$

onde $\alpha = a$ e $\beta = ln\, b$. Se $a > 0$, a exponencial será crescente e se $a < 0$, decrescente.

Figura 47 - Modelos de funções exponenciais.



Fonte: (BASSANEZI, 2012)

As curvas dadas na figura 47 servem para modelar fenômenos em que as taxas de crescimento (ou decaimento) das variáveis de estado positivas são funções das próprias variáveis. Se as taxas de variação são constantes, temos as curvas do tipo (a) se a taxa é positiva ou do tipo (b), se é negativa. Se a taxa de crescimento (positiva) é decrescente como função da variável de estado, temos os modelos dados pelas curvas do tipo (c) ou (d). Se for considerado o modelo exponencial para o ajuste de dados, seu cálculo é facilitado se acrescentar dados auxiliares na tabela com mudança de variável $z_i = ln\, y_i$, juntamente

com os componentes da expressão (5). Assim, encontramos a reta

$$f(x) = 0,0042386237x + 11,17611668\,.$$

Como $\beta = ln\ b$ e $\alpha = a$, então a curva exponencial ajustada é

$$f(x) = be^{ax} = 71404,53606 \cdot e^{0,0042386237} \quad \text{, para} \quad x \geq 0\,.$$

Uma observação interessante também sobre este modelo é que como $a^x = e^{xln\ a}$, temos que

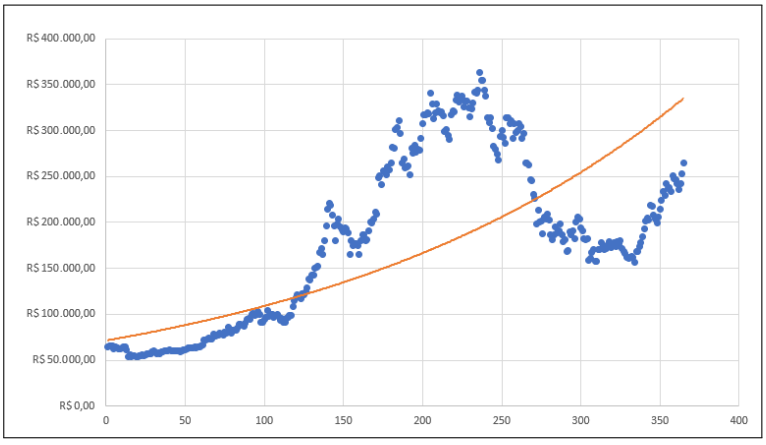
$$e^{0,0042386237} = e^{xln(1,004247619)} = 1,004247619^x$$

Assim, o ajuste exponencial pode ser escrito na forma

$$f(x) = 71404,53606 \cdot 1,004247619^x = 71404,53606 \cdot (1 + 0,004247619)^x\,.$$

A expressão $(1 + 0,004247619)$ indica que para cada unidade de tempo (dia) há um acréscimo de $0,004247619$, ou seja, a sua taxa média diária de crescimento ou juro, no período, é de $0,424\%$ ao dia. Esse ponto é importante por permitir utilizar a modelagem exponencial para conectar os conteúdos de juros e Matemática Financeira no currículo do aluno do Ensino Médio, ou pelo menos retomar alguns pontos que são de grande importância na comprensão maior da Educação Financeira. Na figura 48 observa-se a função exponencial ajustada.

Figura 48 - Valores durante um ano da moeda Bitcoin com ajuste exponencial.



Fonte: O autor, 2022.

Esta função possui $R^2 = 0,3019$. Esse valor demonstra que o modelo está muito distante de se encaixar aos dados observados nesse intervalo. Se comparada ao ajuste

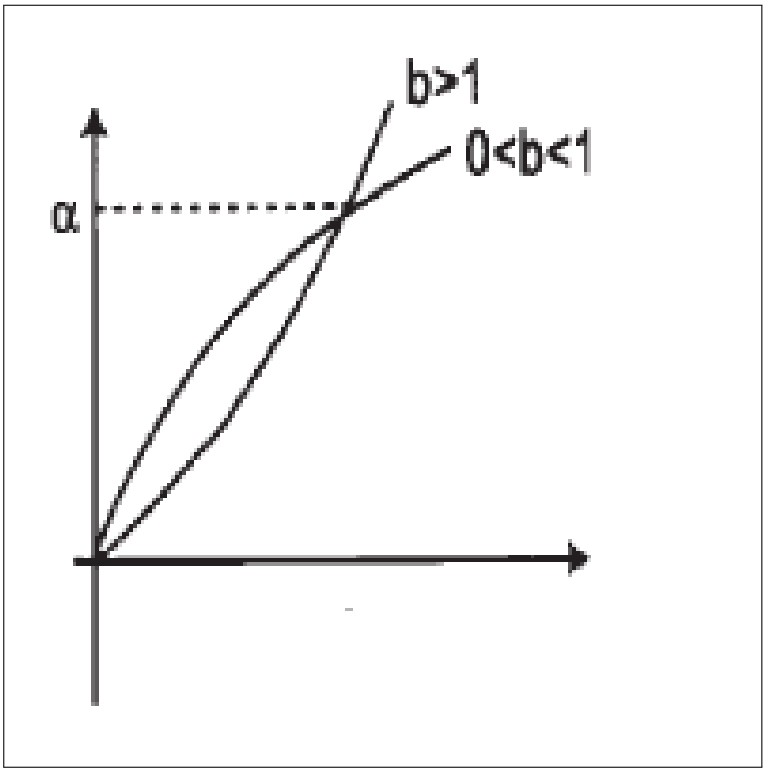
linear de $R^2 = 0,4396$ pode-se inferir que o objetivo está ficando mais distante. É interessante observar que no início do intervalo analisado, o modelo se aproxima razoavelmente bem dos dados observados, porém após o 120^o dia os valores ajustados estão muito longe dos valores esperados. Como o MMQ utiliza a menor soma dos desvios em todo o intervalo, e este é muito discrepante, o modelo encontrado tende a procurar valores mais próximos da mediana.

3.2.2.3 Ajuste linear de modelo geométrico

Os modelos geométricos são dados por funções potências

$$f(x) = ax^b, \qquad a > 0 \quad \text{e} \quad b > 0\,. \tag{13}$$

Figura 49 - Função potência.



Fonte: (BASSANEZI, 2002)

A configuração da curva é do tipo da figura 49 e o ajuste dos parâmetros pode ser efetuado, via ajuste linear, com as seguintes mudanças de variável

$$Y = ln\, y \qquad\qquad X = ln\, x\,. \tag{14}$$

De fato, a função potência (13) pode ser escrita como

$$ln\ y = ln\ a + b\ ln\ x\,.$$

Portanto, nas novas variáveis Y e X, temos

$$Y = \alpha + bX, \qquad \text{onde} \quad \alpha = ln\ a \quad \Leftrightarrow \quad a = e^{\alpha}\,.$$

O ajuste linear encontrado é dado por

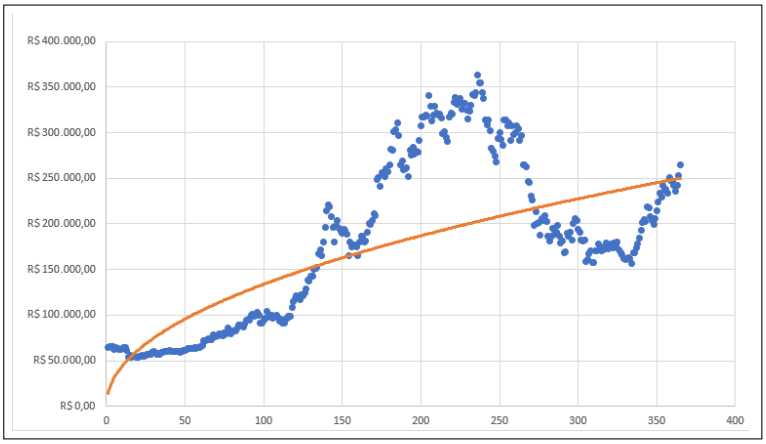
$$Y = 0,4832X + 9,579\,.$$

Assim, temos que $b = 0,4832$ e o parâmetro $\alpha = 9,579$. Recordando que $\alpha = ln\ a$ temos que $a = e^{9,579} = 14457,95$.

A função potência ajustada (figura 50) é dada então por

$$f(x) = 14457,95x^{0,4832}\,.$$

Figura 50 - Valores durante um ano da moeda Bitcoin com ajuste geométrico.



Legenda: O ajuste geométrico é representado pela curva alaranjada no gráfico.
Fonte: O autor, 2022..

Nesse caso, a função possui $R^2 = 0,5133$. Esse valor indica que o modelo ainda está distante de representar bem os valores reais do problema, porém, ainda é melhor que os ajustes linear e exponencial.

3.2.2.4 Ajuste quadrático

Os modelos quadráticos são parábolas

$$f(x) = a + bx + cx^2, \quad \text{onde} \quad a, b, c \quad \in \mathbb{R}. \tag{15}$$

Sua característica principal é possuir pontos críticos (máximos ou mínimos locais) para a variável independente y em um intervalo limitado de variação x.

A determinação dos parâmetros a, b e c também é feita mediante a aplicação do método dos mínimos quadrados, minimizando a expressão

$$f(a, b, c) = \sum_{i=1}^{n} (y_i - y)^2 = \sum_{i=1}^{n} [y_i - (a + bx_i + c{x_i}^2)]^2 \, .$$

As condições necessárias para o mínimo de f são:

$$\frac{\partial f}{\partial a} = 0, \quad \frac{\partial f}{\partial b} = 0, \quad \frac{\partial f}{\partial c} = 0 \, .$$

Estas equações fornecem o sistema de ajustamento para o cálculo de a, b, c, isto é,

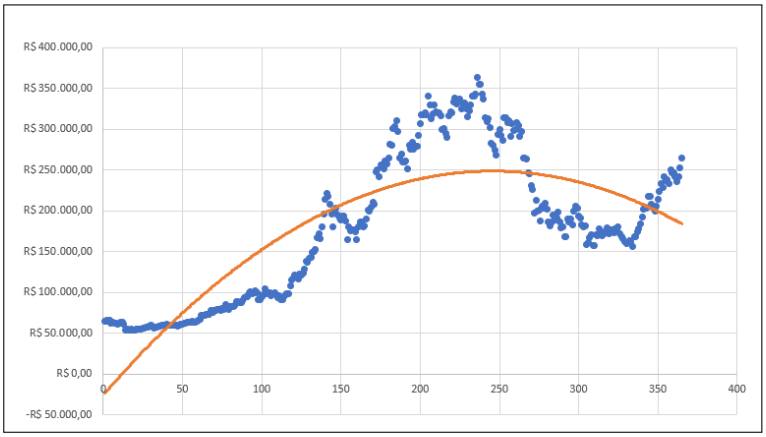
$$\begin{cases} \sum y_i = na + b \sum x_i + c \sum {x_i}^2 \\ \\ \sum x_i y_i = a \sum x_i + b \sum {x_i}^2 + c \sum {x_i}^3 \\ \\ \sum {x_i}^2 y_i = a \sum {x_i}^2 + b \sum {x_i}^3 + c \sum {x_i}^4 \end{cases} . \tag{16}$$

Assim como no ajuste linear, a equação polinomial de grau 2 pode ser facilmente calculada utilizando o aplicativo Excel com o comando linha de tendência. A função quadrática ajustada (figura 51) é dada por

$$f(x) = -4,543x^2 + 2231,9x - 25203 \, .$$

Nesse caso, a função possui $\mathrm{R}^2 = 0,6883$. Esse valor sinaliza que o modelo polinomial é melhor para modelar os valores reais do problema, porém o quadrático ainda está distante do desejado. Vale ressaltar que, mesmo tendo o melhor R^2 dentre os ajustes trabalhados, o início do ajuste quadrático inicia com valores negativos, o que não faz sentido algum para o problema proposto. Isso levanta uma pergunta interessante: será

Figura 51 - Valores durante um ano da moeda Bitcoin com ajuste quadrático.



Legenda: O ajuste quadrático é representado pela curva alaranjada no gráfico.
Fonte: O autor, 2022.

que apenas analisando o coeficiente de determinação pode-se afirmar que aquele ajuste é o ideal para todo o intervalo pesquisado? Este modelo responde que não, já que os modelos linear, exponencial e geométrico possuem R^2 menores, mas, por exemplo, no intervalo dos 50 primeiros dias apresentam valores mais satisfatórios.

O próximo passo lógico após este modelo seria calcular os ajustes de funções polinomiais de graus maiores. Porém, esta aplicação visa o ensino de modelagem de funções estudadas nas escolas no 1º ano do Ensino Médio regular e, portanto, esses modelos de graus maiores não estariam sendo abordados nesse momento do ciclo estudantil e, por isso, não são feitos esses ajustes no trabalho. Porém, ainda pode ser ajustado o modelo logarítmico, muito visto neste ciclo e de grandes aplicações para as mais diversas áreas das ciências da natureza e exatas.

3.2.2.5 Ajuste linear no modelo logarítmico

Seja a função de regressão uma função logarítmica da forma

$$y = a \cdot ln(x) + b, \tag{17}$$

onde a e b são constantes reais. Linearizando (17), sendo $X = ln(x)$, $\alpha = a$ e $\beta = b$, passa-se a ter o problema de estimar os parâmetros da reta

$$f(x) = \alpha X + \beta\,.$$

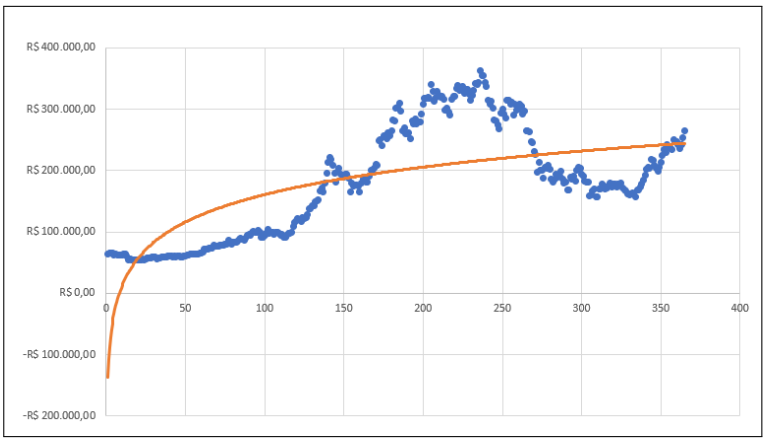
Com essa mudança de variáveis e utilizando os mesmos cálculos feitos para estimar a e b

feitos no ajuste linear, encontra-se o ajuste logarítmico

$$f(x) = 64631ln(x) - 136704\,.$$

Esse ajuste possui $R^2 = 0,4789$, valor abaixo de outros modelos já estimados neste trabalho (figura 52). Assim como no modelo quadrático, seus primeiros dias possuem valor negativo, o que não faz sentido com o problema proposto. Em certos dias os valores propostos até são próximos aos valores encontrados na situação real, porém como visto pelo R^2, em geral eles ainda se mantêm mais afastados que o modelo quadrático.

Figura 52 - Valores durante um ano da moeda Bitcoin com ajuste logarítmico.



Legenda: O ajuste logarítmico é representado pela curva alaranjada no gráfico.
Fonte: O autor, 2022.

3.2.2.6 Preço da moeda no futuro

Um dos maiores propósitos ao encontrar uma função que ajusta dados observados em um determinado problema é conseguir, a partir desses estudos, predizer o futuro. Assim, com as funções de ajustes calculadas, é possível calcular valores futuros, que estão fora do intervalo dos dados modelados, entre agosto de 2020 a agosto de 2021, e saber se a diferença de preço especulado é muito distante do valor que de fato ocorreu naquele dia. Como a proposta era usar como critério de medida o coeficiente de determinação R^2, dentre todos os ajustes calculados na seção anterior, o modelo quadrático é o escolhido. Para que o modelo possa ter maior chance de encontrar bons resultados, é recomendável escolher os dias mais próximos logo após o fim da base de dados. Assim, são utilizados nesta análise os dias do período de 23/8/2021 até 28/8/2021 (INFOMONEY, 2021) e os resultados são apresentados na tabela 6.

Em média, temos que o valor encontrado pela modelagem é 31% menor do que

Tabela 6 - Ajustes encontrados.

Ajuste	Modelo	R^2
Linear	$f(x) = 569,21x + 76500$	$R^2 = 0,4396$
Exponencial	$f(x) = 71404,53606 \cdot e^{0,0042386237}$	$R^2 = 0,3019$
Geométrico	$f(x) = 14457,95x^{0,4832}$	$R^2 = 0,5133$
Quadrático	$f(x) = -4,543x^2 + 2231,9x - 25203$	$R^2 = 0,6883$
Logarítmico	$f(x) = 64631ln(x) - 136704$	$R^2 = 0,4789$

Fonte: O autor, 2022.

aquele apresentado efetivamente no dia (tabela 7).

Tabela 7 - Comparação entre os preços do *Bitcoin*.

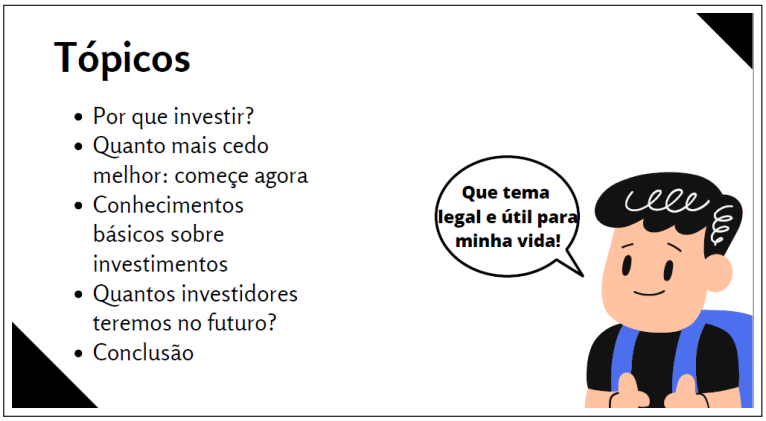
Dias	Preço do dia	Ajuste quadrático
22/08/2021	R$ 263.189,02	R$ 183.110,29
23/08/2021	R$ 264.992,56	R$ 182.012,17
24/08/2021	R$ 266.479,27	R$ 180.904,97
25/08/2021	R$ 251.534,78	R$ 179.788,68
26/08/2021	R$ 255.448,94	R$ 178.663,30

Fonte: O autor, 2022.

Se em relação à previsão do preço futuro da moeda, não encontramos uma função satisfatória, o mesmo não pode ser dito quanto à atividade desenvolvida por alunos da 2ª série do Ensino Médio.

A proposta geral era introduzir os motivos de se investir dinheiro em produtos vendidos pela atual bolsa de valores do Brasil, a Brasil, Bolsa e Balcão (B3), que aparece na figura 53. E como visto também nesta imagem, o último objetivo de pesquisa dos alunos era tentar prever quantos investidores na bolsa de valores teremos no futuro.

Figura 53 - Sumário do trabalho apresentado pelos alunos na escola.



Fonte: O autor, 2022.

Utilizando as funcionalidades da ferramenta e seguindo a proposta deste capítulo, os alunos aproveitaram que a informação da quantidade de cadastro de poessoas físicas (CPF) era disponibilizado pela B3 e encontraram o resultado apresentado na figura 54.

Figura 54 - Modelo quadrático encontrado pelos alunos utilizando o Excel.

Fonte: O autor, 2022.

Ao fim da apresentação, o grupo ainda exibiu matérias de jornal que corroboravam que o aumento da quantidade de pessoas investindo na B3 seria próximo ao valor que o grupo especulou no trabalho, demonstrando assim grande confiança em suas estimativas e, consequentemente, na matemática que a envolve.

Vale ressaltar que essa atividade desenvolvida na escola possui como tema a escolha livre dos alunos, ou seja, o tema investimentos foi uma decisão exclusiva desse grupo. Foram propostos diversos outros temas relacionados a problemas biológicos, financeiros, ambientais, entre outros, e como são mais simplificados, foi possível encontrar previsões mais satisfatórias que o preço da moeda *Bitcoin*, porém, utilizamos este problema como motivador para os alunos. Com essa diversidade de temas, percebemos como a modelagem matemática e o conteúdo de função tem grande usabilidade em diversas áreas do conhecimento, além das exatas.

3.3 **Sequência Didática deste capítulo**

A proposta desta sequência tem como objetivo principal a compreensão dos ajustes de curvas utilizando as funções estudadas no ensino médio.

1. Inicialmente necessitamos compreender qual seria a ideia do conceito de funções. Sua usabilidade é bem vasta nos conteúdos de ciências, mas podemos chamar a

atenção para um tema muito falado nesses últimos anos, o *Bitcoin*. Explicar a ideia dessa moeda, sua proposta de mundança estrutural na economia e algumas possibilidades, é a motivação nesse início. Esse tópico teria seu planejamento para 2 tempos de aula.

2. Em nosso segundo encontro, é introduzido o conceito revolucionário por trás da moeda, o *Blockchain*. Entender que toda a mudança proposta é possível apenas pela criação de Nakamoto e aproveitamos para citar e exemplificar outros projetos e produtos que foram criados em consequência dessa tecnologia. Esse tema pode ser desenvolvido em 2 tempos.

3. Após toda essa contextualização e justificativa, podemos iniciar o processo de exemplificação englobando como foram colhidos os dados, organizados e selecionados. Expor também o processo de modelagem matemática e sua importância. Calcular os ajustes para procurar a melhor função que se encaixa ao tema proposto recordando as funções vistas no ensino básico sobre outra ótica; que parâmetros estatísticos foram utilizados para auxiliar no julgamento da melhor função; como esses processos são desenvolvidos matematicamente e, por fim, mostrar os valores obtidos expondo uma conclusão sobre o modelo encontrado. Esse tópico se subdivide em apresentar as funções e métodos estatísticos com embasamento nas definições vistas pelos alunos antes da graduação, mas como em geral, os alunos chegam com embasamento matemático muito diverso, o professor tem de esmiuçar o conteúdo, o que provavelmente vai acarretar uma demanda grande de tempo. Com isso, uns 20 tempos de aula é uma estimativa razoável, levando em consideração que não há aprofundamento, como propriedades e características particulares, além do esboço do gráfico, de cada função individualmente por não haver necessidade para o cálculo da previsão futura do modelo.

É importante que tanto professor, como aluno, tenha entendimento que escolher o tema, decidir qual amostra pode ser utilizada e procurar fontes confiáveis, fazem parte do protagonismo do aluno que deve ser incentivado pelo professor. Além disso, também encontramos conceitos básicos de pesquisa estatística que, apesar de não ser nosso foco neste capítulo, convergem com o método científico utilizado dentro de modelagem matemática citado.

Plano de aula

Tema: Modelagem matemática Aplicada ao Bitcoin.

Público-alvo: alunos de graduação nas disciplinas de Pré-cálculo e Estatística.

Duração do curso: 24 tempos de aula;

Objetivos: Capacitar os alunos a utilizarem técnicas de modelagem matemática para analisar dados e prever tendências, aplicando esses conhecimentos ao estudo do preço do Bitcoin.

Conteúdos:

- Introdução ao *Bitcoin* e *Blockchain*;
- Fundamentos de Modelagem Matemática;
- Análise de dados com Excel;
- Coeficiente de correlação de Pearson.

Metodologia: aula expositiva, com uso de tecnologia de apoio, como projeção de *slides* sobre a matéria e demonstrações práticas no Excel; discussões em grupo e resolução de exercícios; apresentação e debates; apostilas e livro.

Avaliação:

- Projeto de Modelagem matemática no qual os alunos devem escolher um tema de interesse e desenvolver um modelo matemático utilizando as ferramentas e técnicas aprendidas no curso;

O que aprendemos neste capítulo?

Neste capítulo é explicado o que são as criptomoedas e sua proposta para uma nova versão do modelo econômico mundial. Aproveitando-nos deste tema muito discutido em toda a sociedade e pouco compreendido, o utilizamos como um motivador para trabalhar com modelagem matemática no ensino. Assim, com o auxílio do Excel, que pode ser utilizado de forma *on-line* e gratuita, ensinamos aos alunos alguns passos iniciais de como funciona o processo de modelagem em matemática, alguns conceitos estatísticos e também de funções, que são estudadas em toda a 1ª série do Ensino Médio, mas não possuem diálogo com temas mais atuais para este público. Com isso, mostramos uma das ideias mais fortes do conceito de funções – o de tentar prever o futuro.

4 INTELIGÊNCIA ARTIFICIAL

Antes de falar sobre Inteligência Artificial, devemos tentar compreender como historicamente a sociedade conceituou o que de fato é inteligência. Os parágrafos abaixo são um resumo do Capítulo 5 do livro “Data Mining – A Mineração de Dados no Marketing, Medicina, Economia, Engenharia e Administração” (CARVALHO, 2005).

Do vasto cosmos, infinitamente grande, até as partículas elementares, infinitamente pequenas, talvez a questão mais intrigante que o ser pensante enfrenta em seu mundo físico esteja dentro de si mesmo e seja a mente humana. Na busca por entender o fenômeno mental, o ser humano desenvolveu uma ampla variedade de teorias, geralmente baseadas em analogias com fenômenos físicos observáveis e compreendidos ao longo da História. Na antiga Índia, cerca de 500 anos antes da era Cristã, a mente era descrita como algo a ser dominado pela vontade do "eu real"(a alma).

No Egito antigo, a teoria da mente se baseava na ideia de que ela era uma entidade invisível e imortal, que seria julgada após a morte do corpo pelos atos cometidos durante a vida. Nos primórdios do Judaísmo, cerca de 1500 anos antes da era Cristã, acreditava-se que o coração era o centro da razão, enquanto os rins eram responsáveis pelos sentimentos. A mente seria imortal e habitava um local específico após a morte do corpo físico, podendo ser consultada ou trazida temporariamente para se expressar.

Na Grécia antiga, cada filósofo tinha sua própria teoria sobre a mente, com pequenas variações em relação aos pensadores de sua escola. Empédocles (450 a.C.) afirmava que a mente, situada no coração, era uma duplicata material do indivíduo que sobreviveria após sua morte. Platão confirmava a imortalidade da alma, situando o cérebro como a sede da razão, enquanto as emoções nobres residiam no peito e os instintos, nas vísceras. Aristóteles, por sua vez, postulava que a alma não se restringia à mente e localizava a mente humana no cérebro. O século XVIII foi marcado por diversos ataques à teoria de Descartes, o que fortaleceu o materialismo, ao argumentar que uma mente imaterial não poderia controlar um cérebro material sem violar o Princípio da Conservação da Energia. No século XIX, o anatomista Franz Joseph Gall foi o primeiro a identificar regiões específicas do cérebro responsáveis por atividades mentais. Mais de um século depois, a doutrina das células no cérebro foi substituída pela dos neurônios. Após a Segunda Guerra Mundial, os matemáticos McCulloch e Pitts foram os pioneiros em modelar matematicamente um neurônio e simular de forma primitiva funções mentais. Nos anos 60, foram desenvolvidas teorias sobre o processamento de informações análogas ao funcionamento dos neurônios no cérebro, estabelecendo as bases do paradigma conhecido atualmente como Redes Neurais Artificiais. Foi nesse período que as descobertas sobre inteligência humana e artificial começaram a se entrelaçar. Talvez, para o bem da humanidade, a mente humana continue a ter sua sede em um local desconhecido, e as mais complexas

metáforas computacionais do século XXI sejam apenas equivalentes a um lago (a mente), cuja superfície agitada (a atividade sensorial) impede a visão de seu fundo, como sugeriam os hindus.

O início da comercialização de computadores com capacidade para armazenamento de informações se inicia na década de 1950. Com isso, começaram a aparecer os primeiros computadores em ambientes universitários (VALENTE, 1999). Somente em 1970 é que chega às escolas brasileiras essa tecnologia. Com o passar dos anos e com o impulsionamento da Internet, o computador se tornou indispensável de algum modo para auxiliar nas tarefas de professores, alunos e funcionários da escola.

A Inteligência Artificial (IA) é um tópico imenso atualmente. Comunicar-se com seu *smartphone* faz com que ele aprenda mais sobre seu modo de falar e também seus interesses. Este é um exemplo de IA e sem perceber você faz uso dessa tecnologia o tempo todo, como em uma recomendação de filme, por exemplo. Essa técnica se baseia na hipótese de que o pensamento mecanizado é possível.

O nascimento da IA, conforme a conhecemos hoje, ocorreu com a publicação de "Computing Machinery and Intelligence", de Alan Turing, em 1950. Nesse artigo, Turing explorou a ideia de como determinar se máquinas podem pensar. Para isso, foi determinado um jogo envolvendo três indivíduos. O jogador A é um computador e o jogador B, um ser humano. Cada um deve convencer o jogador C, um jogador humano que não pode ver nem o jogador A e nem o B, de que é humano. Se o jogador C não puder determinar de forma coerente quem é humano e quem não é, o computador vence. Esse jogo ficou mundialmente conhecido como "Teste de Turing".

Dentro de IA existem sub-áreas que podem ser melhor compreendidas com o auxílio de conceitos matemáticos. Uma delas é o Aprendizado de Máquina (Machine Learning). Antes de escrevermos sobre este tema, vale ressaltar que existem quatro tipos de aprendizagem de máquina: o supervisionado, o não-supervisionado, o semi-supervisionado e o por reforço.

No supervisionado, o sistema aprende com base nos exemplos que recebeu. No não-supervisionado o sistema aprende sem receber exemplos do que deve ou não ser considerado certo, cabendo ao próprio sistema essa decisão baseado em características dos dados que são recebidos. Já no semi-supervisionado há uma mistura entre os dois tipos de treinamento citados anteriormente e na aprendizagem por reforço, o sistema utiliza a lógica de tentativa e erro, ou seja, o aprendizado é feito por experiência, assim o sistema aprende com seus próprios erros (ZENDESK, 2024). Na seção a seguir vemos um exemplo de Aprendizado de Máquina supervisionado .

4.1 **Aprendizado de Máquinas**

O Aprendizado de Máquinas (AM) conta com algoritmos para analisar gigantescos conjuntos de dados. Assim, ele pode realizar análise preditiva bem mais rápido que qualquer ser humano (MUELLER; MASSARON, 2019). Nosso processo atual com IA é fazer análise e o aprendizado de máquinas é uma das ferramentas para que isso ocorra da forma mais ágil e eficaz. Ela permite que a IA execute as seguintes tarefas:

- adaptar-se a novas circunstâncias não previstas pelo desenvolvedor original;
- detectar padrões em todos os tipos de fontes de dados;
- criar novos comportamentos com base em padrões reconhecidos;
- tomar decisões com base no sucesso ou na falha desses comportamentos.

Compreender o funcionamento desse aprendizado é de grande valia para nós, já que ainda são os seres humanos que devem tomar as decisões sugeridas pelo computador.

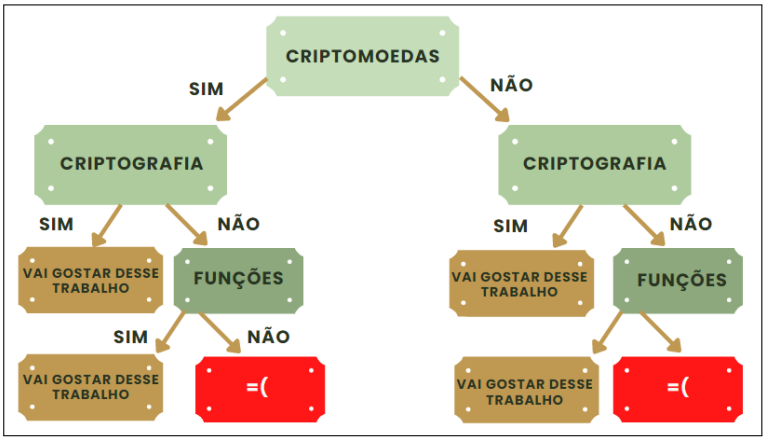
Nesse contexto, "aprendizagem" é um processo computacional iterativo em que os parâmetros de um modelo matemático são ajustados a fim de minimizar uma medida de erro. Assim, quanto menor o erro cometido pelo modelo na execução da tarefa, melhor será seu desempenho.

Para se obter uma melhor compreensão do processo computacional, é necessário maior conhecimento de matemática e, assim, parece indispensável proporcionar oportunidades para a construção pessoal desse conhecimento, o que nos remete à educação escolar. O assunto AM não é tratado nas escolas brasileiras, mas poderia ser facilmente trabalhado utilizando Modelagem Matemática. Para isso, é necessário conhecer as direções principais no AM. Atualmente, são elas: Aprendizado de Máquinas Clássico (AMC), Aprendizagem por Reforço, Métodos *Ensemble*, Redes Neurais e *Deep Learning* (GOMES, 2019). Foram vistos no capítulo de Ajuste de Curvas alguns métodos de regressão e estes se encaixam na direção de Aprendizado de Máquinas clássico. Porém, mesmo o melhor modelo ajustado ainda está bem distante do desejado para se encontrar o valor futuro do preço da moeda. Isso se deve ao fato de que os dados que se desejam modelar são muito aleatórios e, consequentemente, de difícil ajuste em uma regressão linear.

Antes de prosseguir, vamos compreender um pouco melhor o que seria o AMC, baseados na série de vídeos do canal do Youtube "StatQuest with Josh Starmer", criada por Josh Starmer e publicada nos anos entre 2019 e 2022 (STARMER, 2019), observe a figura 55.

Este é um exemplo simples, mas ilustra árvores de decisões, um dos métodos utilizados em AMC. O propósito dessa árvore, em particular, é predizer se alguém vai ou não

Figura 55 - Árvore de Decisões.



Fonte: O autor, 2022.

gostar desse trabalho. Alternativamente, pode-se dizer que esta árvore de decisões classifica uma pessoa como alguém que gosta, ou não deste trabalho. Em geral, Aprendizado de Máquinas é sobre realizar previsões e classificações.

Um questionamento importante, tanto para o estudo de funções e de AM, é se com a perda de alguns dados, ainda seria possível encontrar um bom ajuste para o modelo. Uma das formas de desenvolver esse aprendizado é incorporar inicialmente menos dados à inteligência que você deseja desenvolver. Vamos compreender isto melhor com outro exemplo.

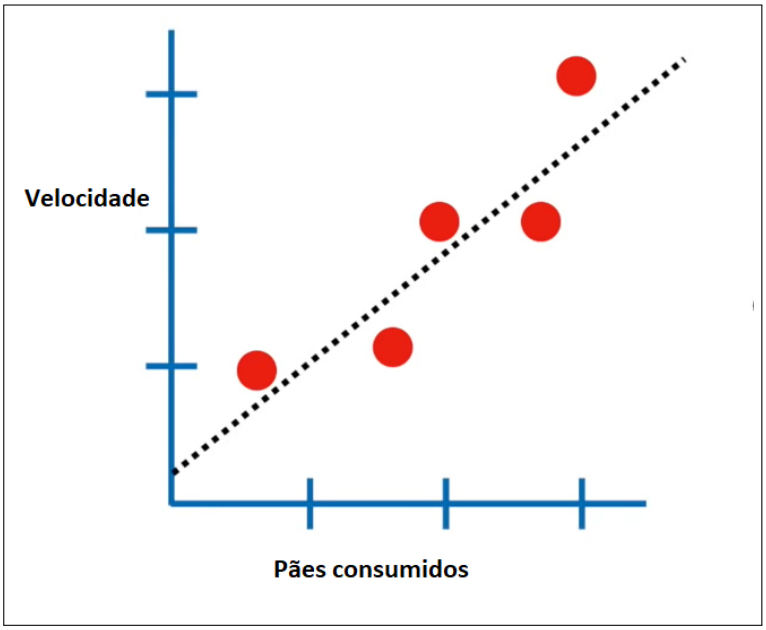
Imagine que medimos o quão rápido alguém consegue correr 100 metros e quantos pães ele come. Antes de mais nada, na linguagem de AM os dados originais são chamados de dados de treinamento, representados pelos pontos vermelhos, e então pode-se criar uma linha preta para mostrar a tendência, mas também pode-se usar a linha preta para fazer previsões (figura 56).

Por exemplo, se alguém disser que consumiu uma certa quantidade de pães, então é possível prever o quão rápido essa pessoa corre (figura 57). A linha preta é um tipo de aprendizado de máquinas, porque pode-se utilizá-la para tais previsões.

Agora que compreendemos inicialmente como são feitas as previsões e classificações, é possível compreendermos algumas ideias principais do AM. Como dito anteriormente, os pontos vermelhos são chamados de dados de treinamento e a linha preta é ajustada aos dados de treino. Porém, existem linhas, como o rabisco verde (figura 58), que se ajusta melhor aos dados de treino que a linha preta. Mas lembre-se que o objetivo do AM é prever. Então, é necessário uma forma de saber se o rabisco verde é melhor ou pior que a linha preta nesse sentido.

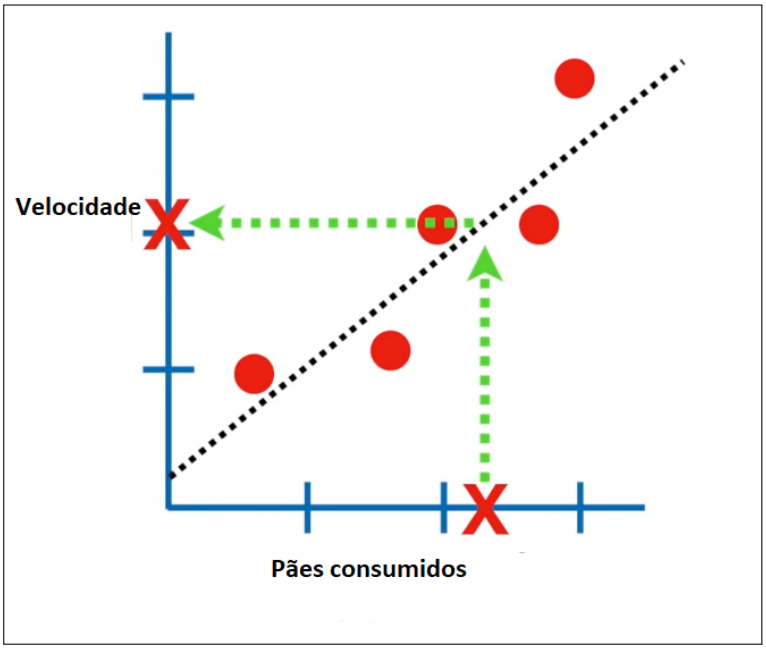
Para isso, encontra-se uma nova pessoa e mede-se o quão rápido ela corre e quanto de pão ingere. E depois encontram-se mais pessoas, todas representadas como pontos azuis na figura 59. Todas estas são chamados em AM como dados de teste. Utilizam-se

Figura 56 - Gráfico Pães consumidos x velocidade.



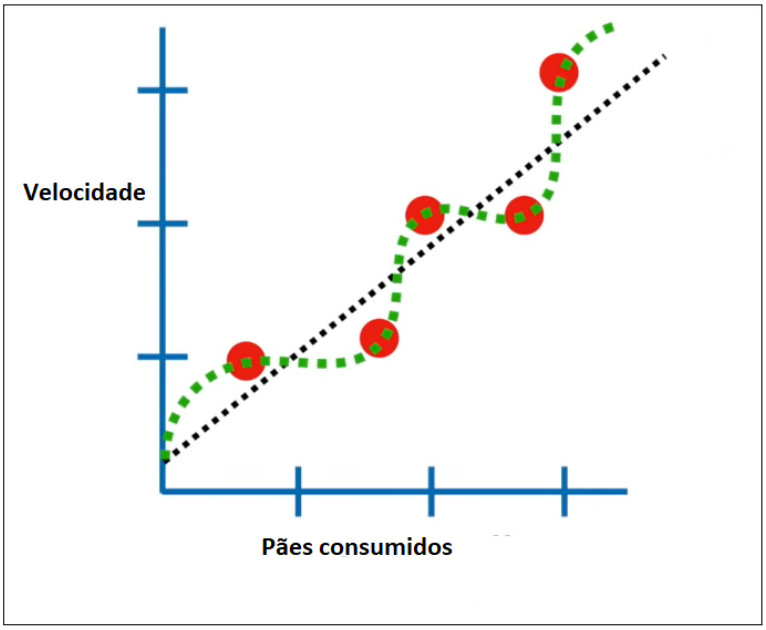
Legenda: O gráfico mostra a relação entre pães consumidos e a velocidade de um indivíduo.
Fonte: O autor, 2022.

Figura 57 - Gráfico com a previsão da velocidade.



Legenda: O gráfico mostra qual seria a velocidade do indivíduo em relação aos pães consumidos.
Fonte: O autor, 2022.

Figura 58 - Gráfico linha de tendência e linha ajustada.



Legenda: O gráfico mostra o rabisco verde que passa por todos os pontos e a linha preta de tendência.

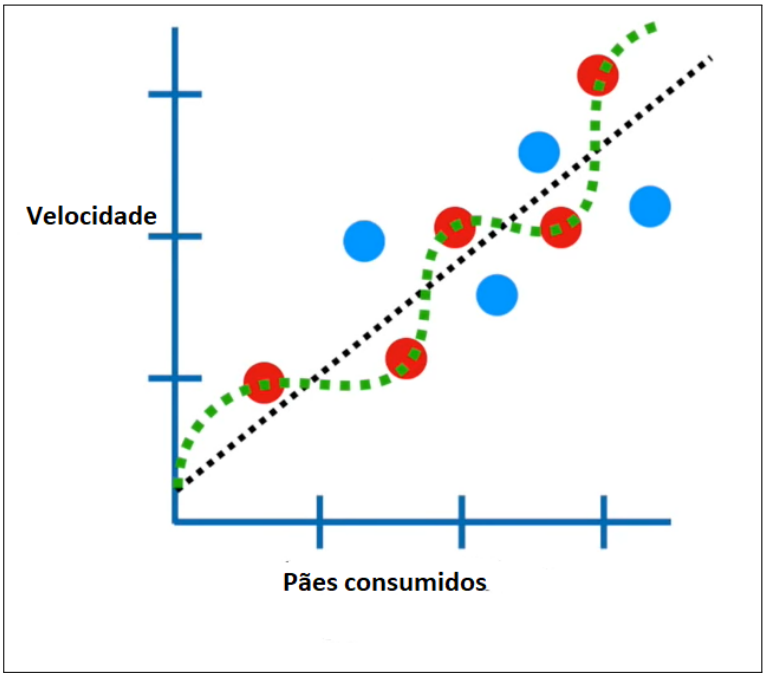
Fonte: O autor, 2022.

esses dados para comparar com as previsões feitas pela linha preta e aquelas feitas pelo rabisco verde.

Iniciamos observando qual o desempenho da velocidade de cada pessoa em relação à linha preta nos dados de teste. Vejamos um caso. A pessoa da figura 60 consome muito pão e corre muito rápido, porém a linha preta prevê que uma pessoa que consumiu essa quantidade de pão deveria correr um pouco mais lentamente (figura 61).

Medindo então a distância entre a velocidade real e a prevista em todos os dados de teste com a linha preta, teremos um valor que é representado por uma escala preta no canto inferior direito da figura 62. Da mesma forma, calculam-se as velocidades reais e previstas só que do rabisco verde. Lembrando que o rabisco verde é excelente ajustando os dados de teste, mas quando estamos falando de AM o interesse é saber como prevê o rabisco verde em relação aos novos dados. Calculando então as distâncias das velocidades reais e previstas, pode ser notado que a soma das distâncias é maior para o rabisco verde do que para a linha preta. Em outras palavras, mesmo sendo melhor o ajuste do rabisco verde aos dados de treinamento do que o que corresponde à linha preta, a última fez um trabalho melhor prevendo as velocidades com os dados de teste. Logo, se tiver que escolher entre usar uma ou outra, seria melhor escolher a linha preta. Denominamos de *underfitting* quando o modelo não consegue aprender as relações mais importantes entre as classes, ou seja, fica abaixo do esperado. Este caso não se apresenta em nosso modelo. Quando o modelo faz uma super relação entre os dados, chama-se de *overfitting*,

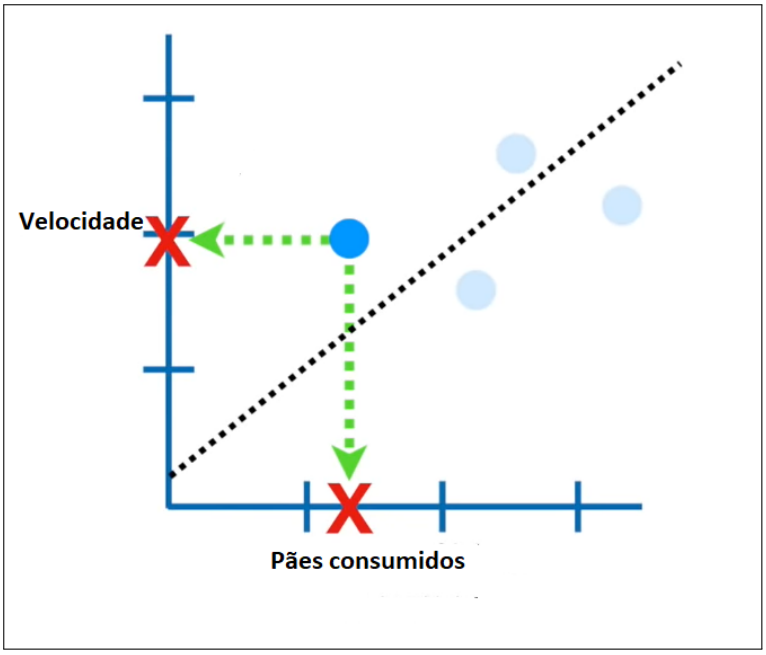
Figura 59 - Gráfico com os dados de teste.



Legenda: O gráfico mostra o rabisco verde, a linha preta de tendência, os dados de treino e os dados de teste.

Fonte: O autor, 2022.

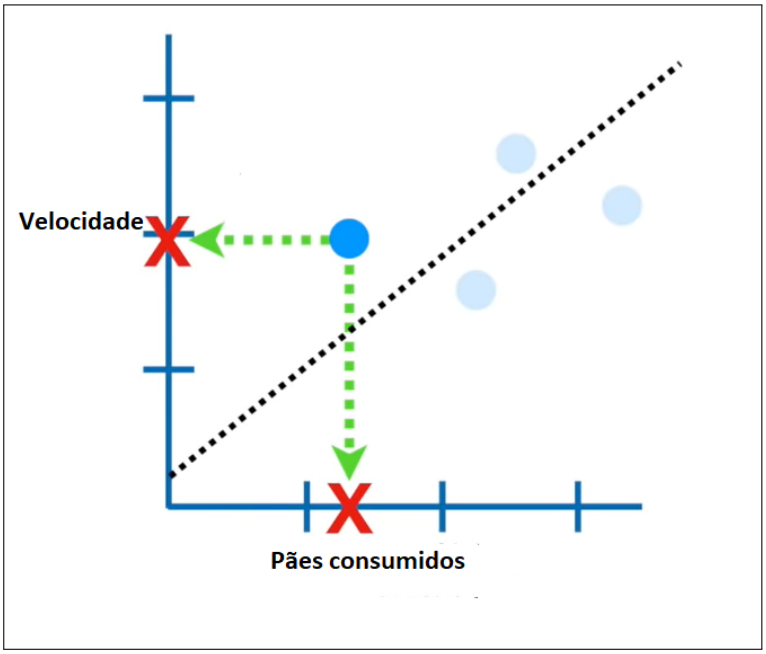
Figura 60 - Gráfico com a verificação dos dados de teste.



Legenda: O gráfico mostra qual a quantidade de pães ingeridos e a velocidade de um dos dados de teste.

Fonte: O autor, 2022.

Figura 61 - Gráfico com a distância dos dados de teste.



Legenda: O gráfico mostra a distância entre o valor previsto e real das velocidades de um indivíduo.

Fonte: O autor, 2022.

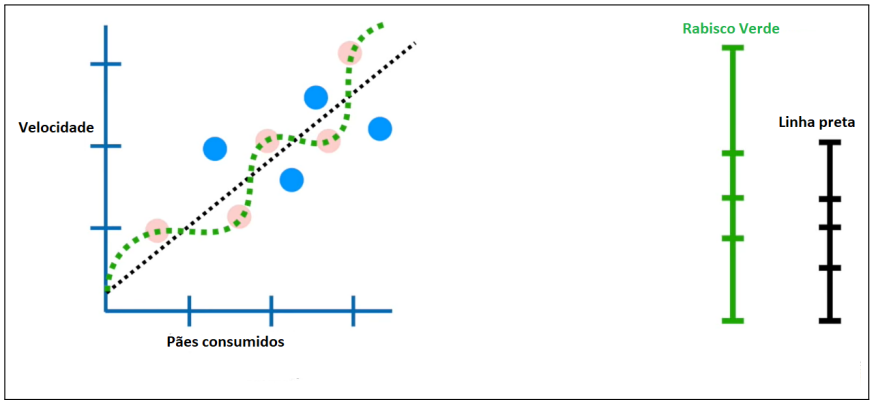
que em nosso exemplo seria o modelo de rabisco verde. É dito neste modelo que não houve aprendizado e sim que o mesmo memorizou os resultados apresentados para seu treinamento.

Fazendo um paralelo, essas distâncias encontradas nada mais são que o Método dos Mínimos Quadrados, sendo apresentado de forma mais simplificada para os alunos do ensino médio. Este exemplo ensina duas ideias principais de AM. Primeiro, nós usamos os dados de teste para validar os métodos de AM. Depois nós não nos deixamos enganar pelo quão bem o método de AM se ajustou aos dados de treino. É importante observar que se ajustar bem aos dados de treino, mas fazer previsões ruins, é chamado de balanço viés-variância. Antes de prosseguir, é importante compreender um pouco melhor esse conceito.

4.1.1 Viés e variância

Na seção anterior, comentamos sobre um dos fundamentos de AM: viés e variância. Para compreendermos melhor esse conceito, suponha que medimos o peso e a altura de um grupo de ratos e esboçamos os dados em um gráfico (figura 63). Repare que os ratos magros têm a tendência a serem baixos e os mais pesados tendem a ser mais altos. Mas após um certo nível de peso, eles não crescem mais, apenas se tornam mais obesos. Com

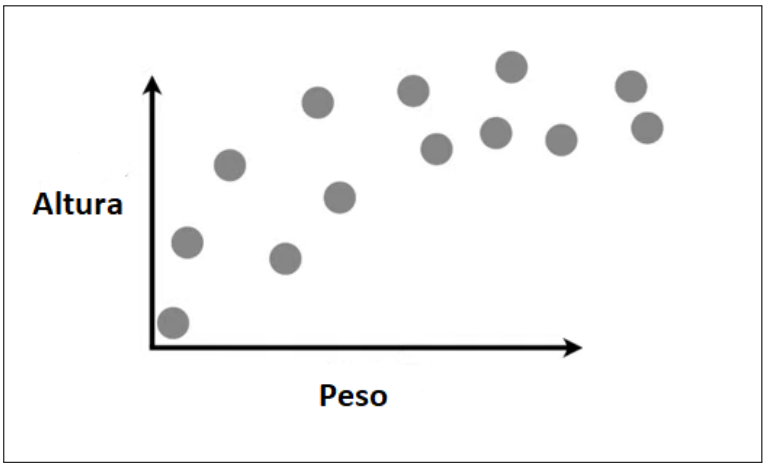
Figura 62 - Gráfico de comparação entre as distâncias.



Legenda: O gráfico mostra a distância encontrada pelo ajuste preto e o rabisco verde.
Fonte: O autor, 2022.

isso, gostaríamos de prever a altura de um rato.

Figura 63 - Gráfico peso x altura de ratos.

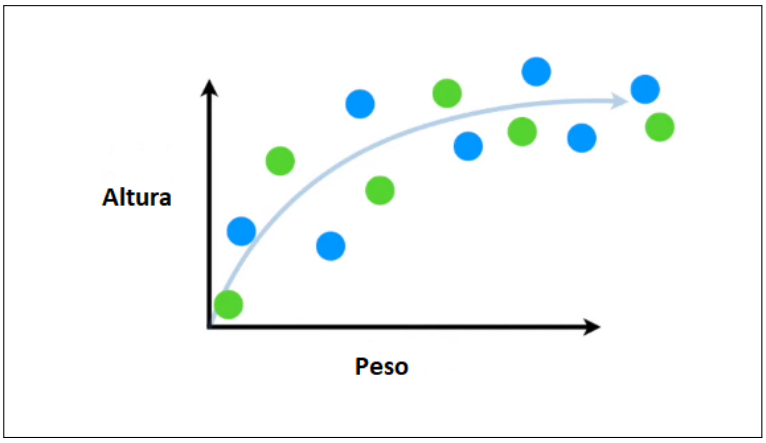


Legenda: O gráfico mostra os dados de peso e altura de um grupo fictício de ratos.
Fonte: O autor, 2022.

Em uma situação ideal, teríamos uma equação matemática para descrever a relação entre peso e altura. Mas neste caso não a conhecemos. Ainda assim, vamos deixar a relação "verdadeira" descrita na figura 64 como a curva em azul claro, para se ter uma referência. Utilizamos dois métodos de AM para realizar uma aproximação dessa relação.

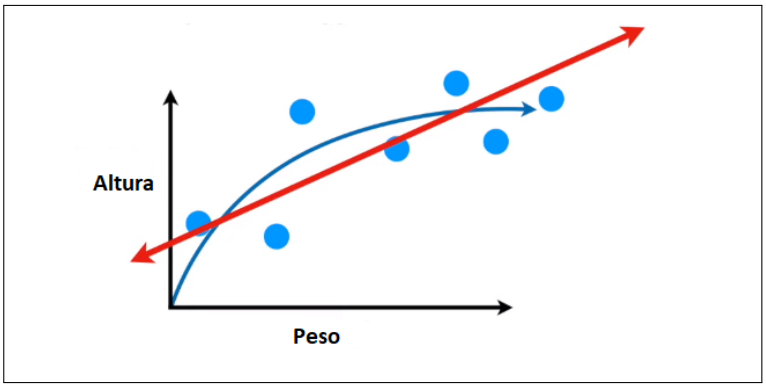
A primeira ação a executar é dividir os dados em dois conjuntos, um para treinamento do algoritmo e outro para o teste. Os pontos azuis são o conjunto de treino e os verdes, o grupo de teste (figura 64). Vamos pensar inicialmente no conjunto de treinamento. Utilizando o algoritmo baseado em Regressão Linear, também conhecido como Método dos Mínimos Quadrados, citado no Capítulo 3, para o AM, o algoritmo faz o ajuste de uma linha reta ao conjunto de treinamento (figura 65) .

Figura 64 - Gráfico dados de treino e teste.



Legenda: O gráfico mostra os dados de treino e teste e a relação “verdadeira” entre eles.
Fonte: O autor, 2022.

Figura 65 - Gráfico com o ajuste linear para altura e peso.

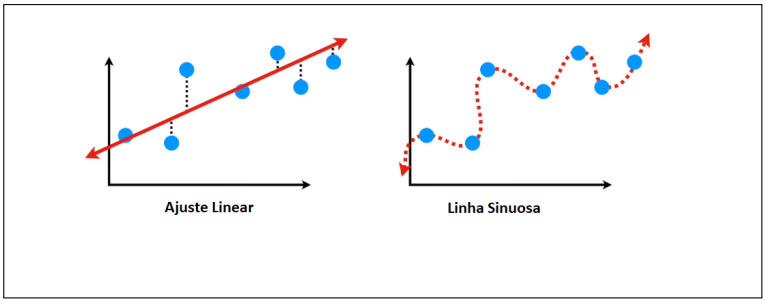


Legenda: O gráfico mostra os dados de treino e teste, a relação “verdadeira” entre eles e o ajuste linear.
Fonte: O autor, 2022.

Note que essa linha não possui a flexibilidade necessária para replicar fielmente o conjunto escolhido. Não importa o quanto tente ajustar a linha, ela nunca se curva. Portanto, a linha reta nunca representa bem a relação entre peso e altura, não importando o quão bem a ajustemos ao conjunto de treinamento. A incapacidade de um algoritmo de aprendizagem de máquina em conseguir representar a relação fiel é chamada de Viés. Como a linha reta não pode ser curvada até se encaixar na relação "verdadeira", dizemos que ela possui uma quantidade relativamente alta de viés. Apesar de ser um conceito simples, o cálculo de viés só é abordado em cursos de graduação.

Outro método de AM poderia ajustar uma linha sinuosa ao conjunto de treinamento. A linha sinuosa poderia ser bastante flexível, e seguiria o conjunto de treinamento ao longo do arco da relação "verdadeira". Com esse feito, podemos dizer que ela possui pouco viés. Podemos comparar o quão bem a linha reta e a linha sinuosa se ajustam aos dados de treinamento calculando a soma dos quadrados (figura 66). Ou seja, medimos todas as distâncias das linhas de ajuste aos dados, calculamos todos os seus quadrados e os somamos.

Figura 66 - Gráfico com o ajuste linear e linha sinuosa (Grupo de treinamento).



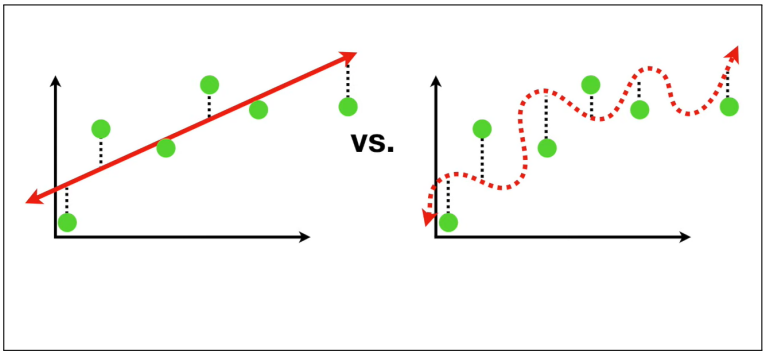
Legenda: O gráfico mostra os dados com o ajuste linear e outro ajuste não linear para o grupo de treinamento.

Fonte: O autor, 2022.

Nota-se que a linha sinuosa se ajusta tão bem aos dados que a distância entre elas e os dados são todas zero. Na disputa de ver qual ajuste é melhor para o conjunto de treinamento o segundo modelo vence. Mas, lembre-se que até agora só calculamos o MMQ sobre o grupo de treinamento. Realizando o mesmo cálculo sobre o conjunto de teste, podemos notar que a linha reta se ajusta melhor (figura 67).

Apesar de realizar a linha sinuosa um ótimo trabalho ao tentar se ajustar ao grupo de treinamento, ela não obtém o mesmo sucesso ao tentar se ajustar ao conjunto de teste. Em termos de AM, a diferença de ajustes entre esses dois conjuntos de dados é chamada de Variância. A linha sinuosa possui um baixo viés, pois é bastante flexível e consegue se ajustar bem à curva da relação "verdadeira" entre peso e altura. Mas a linha sinuosa possui alta variabilidade, pois ela resulta em uma alta diferença de soma de quadrados para conjunto de dados diferentes. Assim, é difícil afirmar o quanto a linha sinuosa iria dar

Figura 67 - Gráfico com o ajuste linear e linha sinuosa.



Legenda: O gráfico mostra os dados com o ajuste linear e outro ajuste não linear para o grupo de teste.

Fonte: O autor, 2022.

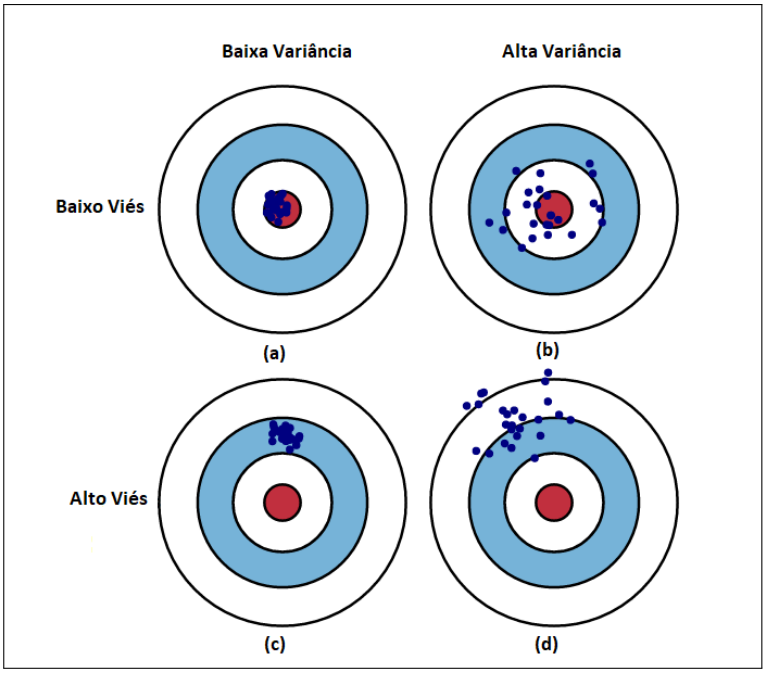
certo com outros conjuntos de dados. Esse conceito, por mais que esteja sendo abordado dentro de AM, também é visto na educação básica e foi revisitado em nosso capítulo de Modelagem matemática, na seção 3.2.2.1 ajuste linear.

Assim, a linha sinuosa possui pouco viés, pois é bastante flexível e consegue se ajustar bem à curva de relação "verdadeira". Mas essa mesma linha possui alta variabilidade pois resulta em uma alta soma de quadrados para conjunto de dados diferentes. Em outras palavras, é difícil dizer se esse modelo iria funcionar bem com outro conjunto de dados. Por outro lado, a linha reta possui um viés relativamente alto, pois ela não é capaz de representar bem a curva entre a relação de peso e altura. Porém, este modelo possui uma variância relativamente baixa, porque a soma dos quadrados resulta em valores similares para conjunto de dados diferentes. Com isso, podemos dizer que a linha reta pode nos dar boas previsões, mas não predições ótimas, mesmo assim ainda são consistentes. Uma outra forma de compreendermos melhor a diferença entre Viés e Variância pode ser utilizando a figura 68. Nela, observamos o algoritmo ter baixa variância, significa que os dados se concentram todos em regiões próximas, enquanto ter uma alta variância significa que os dados ficam mais dispersos. Já o viés pode ser visualizado como a chance de estarem os dados concentrados na região que se deseja. Ter um baixo viés significa então que o algoritmo está próximo aos resultados "ideais". Já possuir um alto viés pode ser entendido que o algoritmo não consegue encontrar resultados muito satisfatórios.

Ainda observando a figura 68 é possível compreendermos o que a combinação dos dois conceitos acarretam aos dados do algoritmo. Em (a) observamos que um baixo valor para viés e variância resulta em um algoritmo que fez um treinamento perfeito, ou seja, alcança os resultados ideais e todos se concentram bem próximos a estes resultados. No caso (b), o baixo viés faz com que os dados fiquem próximos ao seu alvo porém, com a variabilidade alta, os dados se dispersam sensivelmente deste alvo. Já em (c) os dados possuem alto viés, o que causa nos dados encontrados valores distantes do alvo, mas baixa

variabilidade, o que faz com que estes resultados estejam todos na vizinhança. Por último (d), temos o pior caso, onde tanto viés quanto variância são altos, acarretando em modelos com resultados distantes dos ideais e bem espalhados.

Figura 68 - Gráfico Dilema Viés x Variância.



Legenda: O gráfico mostra um exemplo didático para compreender a diferença entre viés e variância.

Fonte: (KOMESU, 2022)

Esse problema é também conhecido como Dilema Viés-Variância. De um lado, aumentar a complexidade de um modelo normalmente aumentará sua variância e reduzirá seu viés. Por outro lado, reduzir a complexidade de um modelo aumenta seu viés e reduz sua variância (tabela 8).

Tabela 8 - Dilema Viés-Variância.

$\uparrow$ Complexidade $\Longrightarrow$	$\uparrow$ variância e $\downarrow$ viés
$\downarrow$ Complexidade $\Longrightarrow$	$\downarrow$ variância e $\uparrow$ viés

Fonte: O autor, 2024.

Para finalizar, nossa linha sinuosa se encaixa em um modelo *overfitting* (visto na seção de Aprendizagem de Máquina) e com esse exemplo podemos ilustrar o que a AM busca encontrar: um modelo com baixo viés, que consiga representar adequadamente a relação "verdadeira" e tenha baixa variabilidade, para produzir previsões consistentes independente do conjunto de dados.

Compreendendo melhor esses conceitos de AM, uma pergunta razoável a ser feita é se com o dilema visto acima, sempre perder-se algum tipo de informação, e qual método

de AM podemos utilizar para nos auxiliar a tomar esta decisão. É isto que vemos na próxima seção.

4.1.2 Validação Cruzada

Existem diversos métodos de AM que podem ser utilizados, como Regressão Logística, K-vizinhos mais próximos, Máquina de Vetores de Suporte, entre outros. Como decidir qual deles utilizar? A Validação Cruzada nos permite comparar diferentes métodos de AM e ter uma ideia de como eles se sairiam na prática (KOHAVI, 1995). Imagine todos os dados coletados sobre o preço da moeda Bitcoin. Com esses dados precisamos:

1. **Estimar os parâmetros para os métodos de aprendizagem de máquina**. Isso significa que para utilizar regressão logística temos que usar alguns dados para estimar o formato desta curva. Em AM estimar parâmetros é chamado de "treinar o modelo".

2. **Avaliar quão bem os métodos de AM funcionam**. Isto é, torna-se necessário descobrir se a curva fará um bom trabalho categorizando novos dados. Em AM, avaliar um método é chamado de "testar o modelo".

Assim, em jargão de AM, são necessários os dados para treinar e testar os métodos de AM. Uma abordagem considerada ruim seria utilizar todos os dados para treinar e testar, pois, precisamos saber como o modelo funciona em dados nos quais não foram treinados. Uma ideia um pouco melhor seria utilizar 75% dos dados inicias como treinamento e os outros 25% como teste. Assim, podemos comparar os métodos observando como cada um categoriza os dados de testes. Mas, como saber se o ideal seria utilizar a parte final para teste e não a inicial? Ou um pouco de cada parte? Ao invés de nos preocuparmos com isso, a validação cruzada testa todas as possibilidades. Por exemplo, começaria utilizando os 3 primeiros blocos de 25% para treinar o modelo e, então, o último para testar. Daí registram-se os resultados obtidos pelo modelo com os dados de teste. Depois utiliza-se essa combinação de blocos para treinar o modelo. Repete-se este método variando blocos ou dados dentro dos blocos. No final, todos os blocos de dados foram utilizados para testes e podemos comparar métodos observando quão bem o desempenharam. Neste caso, foi utilizada a divisão em quatro blocos, também chamada de quatro dobras, porém o número de blocos é arbitrário. Em casos extremos, poderíamos tratar cada paciente ou amostra como um bloco, técnica chamada de Validação Cruzada "Leave One Out". É muito comum dividir os dados em dez dobras (STARMER, 2019).

Utilizando a validação cruzada, é possível determinarmos qual método funciona melhor para os nossos dados. Então utilizando o conjunto de treinamento, treinamos

todos os métodos nos quais estávamos interessados e, em seguida, testamos cada um dos métodos com o conjunto de testes. Precisamos identificar como se saiu cada método, quando comparado ao conjunto de testes. Uma maneira de obter tal resultado é criando uma Matriz de Confusão para cada método. Para ficar mais claro, suponha que queremos prever se, em uma dada amostra observada, existe observação de grávida. Com isso, teremos uma matriz de ordem 2 como mostra a figura 69.

Figura 69 - Matriz de confusão para teste de gravidez.

Legenda: A tabela mostra a matriz de confusão para os casos de teste de gravidez.
Fonte: O autor, 2024.

As linhas em uma matriz de confusão correspondem à previsão do algoritmo de AM e as colunas correspondem ao que é tido como verdade (referência). Como existem apenas duas categorias para se escolher “estar grávida” ou “não estar grávida”, então o canto esquerdo superior contém os verdadeiros positivos (VP). Estas são pacientes que estão grávidas e que foram corretamente identificadas pelo algoritmo. Os verdadeiros negativos (VN) estão no canto inferior direito. Estas são pacientes que não estão grávidas, e foram corretamente identificadas pelo algoritmo. O canto inferior esquerdo contém os falsos negativos. Falsos negativos ocorrem quando uma paciente está grávida, mas segundo o algoritmo ela não estava. Por fim, o canto superior direito contém os falsos positivos (FP). Falsos positivos são pacientes que não estão grávidas, mas o algoritmo as rotulou como estando. Em resumo, a matriz informa o que o algoritmo aprendeu e o que não aprendeu, de forma correta. Vale salientar que, em termos de casos reais, o teste falha cerca de $1,5\%$ na identificação da gravidez em que as mulheres realmente estejam grávidas (FN). Deste modo, o teste tem confiabilidade de $98,5\%$ no que diz respeito aos falsos negativos. Por outro lado, $0,5\%$ das mulheres que não estão grávidas têm a mudança de cor dos reagentes (FP). Então, a exatidão, no que diz respeito aos falsos positivos é de $99,5\%$. Assim, a porcentagem entre $98,5\%$ e $99,5\%$ é 99%. Com isso, observa-se que os erros em relação aos falsos negativos são o triplo dos erros relativos aos falsos positivos (RIMES; FURST, 2017).

Quando a Matriz de Confusão entregar resultados muito parecidos nos VP e VN, dificultando a escolha de qual método de AM é mais adequado para esses dados, teremos que introduzir métricas mais sofisticadas como Sensibilidade e Especificidade (FERREIRA; PATINO, 2017). Neste caso, a Sensibilidade determina qual a porcentagem de pacientes com gravidez foram corretamente identificados. A sensibilidade é calculada como:

Sensibilidade é a probabilidade de termos verdadeiros positivos em um experimento, ou:

$$Sensibilidade = \frac{Verdadeiros\ Positivos}{Verdadeiro\ Positivos + Falsos\ Negativos} \tag{18}$$

A Especificidade determina qual a porcentagem de pacientes sem estarem grávidas que foram corretamente identificados.

Especificidade é a probabilidade de termos verdadeiros negativos em um experimento, ou:

$$Especificidade = \frac{Verdadeiros\ Negativos}{Verdadeiro\ Negativos + Falsos\ Positivos}. \tag{19}$$

Para compreendermos melhor, vamos supor valores para a tabela de confusão (figura 69). Utilizando um dos métodos de AM suponhamos o seguinte valor dado na (figura 70):

Figura 70 - Matriz de confusão para teste de gravidez com quantidade de pacientes.

	ESTAR GRÁVIDA	NÃO ESTAR GRÁVIDA
ESTAR GRÁVIDA	139	20
NÃO ESTAR GRÁVIDA	32	112

Fonte: O autor, 2024

Vamos iniciar calculando a sensibilidade. Para os verdadeiros positivos utilizamos 139 e para os falsos negativos, 32. Assim, temos:

$$Sensibilidade = \frac{139}{171} = 0.81.$$

A Sensibilidade nos conta que 81% das mulheres na gravidez foram corretamente identificadas pelo modelo de aprendizagem escolhido. Agora, vamos calcular a Especificidade. Para os verdadeiros negativos, vamos utilizar o valor de 112 pacientes, já para os falsos positivos, 20. Com isso,

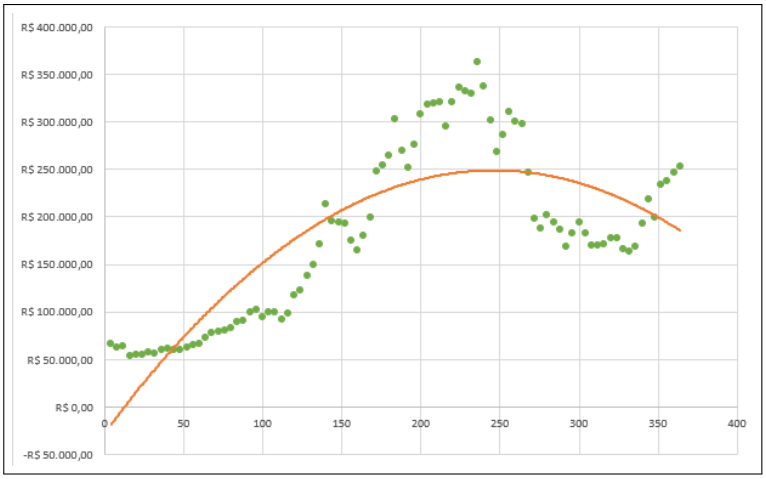
$$Especificidade = \frac{112}{132} = 0.85.$$

Ou seja, 85% das pacientes foram corretamente identificadas pelo modelo selecio-

nado. Com esses dois conceitos definidos, decidimos qual modelo de aprendizagem utilizar dependendo do que consideramos mais relevante para nossa pesquisa. Por exemplo, caso seja mais importante identificar corretamente as pacientes que de fato estão grávidas, devemos escolher o método que possui maior Sensibilidade. Caso seja identificar as que não estão grávidas, o ideal seria utilizar os modelos que possuem maior Especificidade. Para situações que envolvem uma Matriz de Confusão maior que de ordem 2, teríamos que calcular a Sensibilidade e Especificidade de todas as categorias. De forma geral, teríamos que calcular $(n-1)$ Sensibilidade e Especificidade, sendo n a ordem da matriz.

Baseado nos processos de aprendizado de máquina, foi decidido retirar os dias múltiplos de quatro dos dados anuais apresentados no capítulo anterior, já que esses dias representam a porcentagem aceita como teste, sendo formadas quatro dobras, utilizando os restantes dos dias como os dados de teste e calcular novamente os ajustes (figura 71). As funções e seus respectivos R^2 se encontram na tabela 9.

Figura 71 - Gráfico com a retirada dos dias múltiplos de quatro e seu ajuste quadrático.



Legenda: O gráfico mostra os dados de treinamento para o ajuste do preço da moeda *Bitcoin*.

Fonte: O autor, 2024.

Tabela 9 - Ajuste de curvas com os dados reduzidos.

Ajuste	Função	R^2
Linear	$f(x) = 565,9x + 77079$	$0,4358$
Exponencial	$f(x) = 71809 \cdot e^{0,0042x}$	$0,3001$
Geométrico	$f(x) = 12162x^{0,5172}$	$0,5084$
Quadrático	$f(x) = -4,5706x^2 + 2247,9x - 27203$	$0,6868$
Logarítmico	$f(x) = 69190ln(x) - 160042$	$0,4979$

Fonte: O autor, 2022.

Nota-se que mesmo utilizando os dados de treinamento o melhor ajuste continua

sendo o quadrático. É importante ressaltar que retirando-se os 25% dos dados ainda assim a diminuição do coeficiente de determinação (R^2) é mínima, nos levando a acreditar que não há perda substancial de informação e que provavelmente um modelo linear não tem a capacidade de ajustar melhor esse exemplo de dados (Figura 71) e que o objetivo da AM, pouco viés e baixa variância, citado na seção anterior, não deve ser alcançado.

4.2 **Sequência Didática deste capítulo**

1. Em um primeiro encontro, o foco da nossa proposta de ensino seria em compreender como a máquina aprende. Para tal, vamos iniciar refletindo sobre o que é inteligência de fato. Podemos relembrar também, do capítulo de criptografia, o que era inteligência para Alan Turing, ressaltando como os assuntos possuem relação e motivando aos alunos as razões desse tema ser tão importante atualmente em nossa sociedade. Para esta etapa do planejamento, 2 tempos de aula seriam o suficiente.

2. Em um segundo encontro, abordamos algumas definições sobre AM para iniciar o processo de formalização do aprendizado. É importante salientar que todo esse processo de ensino é seguido de exemplos de contexto simples e que se relacionam diretamente com nosso capítulo de funções. Como para o público-alvo esses conceitos já foram vistos em um curso de pré-cálculo, a retomada desse tema se torna mais confortável e engloba mais um significado de fácil compreensão. Para este tema ser trabalhado com exemplos para melhor assimilação, 4 tempos de aula devem ser suficientes.

3. O tópico anterior desperta curiosidade para nosso próximo conteúdo, viés e variância. Utilizamos um exemplo motivador no plano cartesiano e com ele é necessário definir novamente os conceitos de média, variância, desvio padrão, e apresentar o dilema viés-variância. Aqui, pode haver indagações sobre os motivos de não serem abordados moda e mediana, já que são temas vistos no ensino médio. Ocorre que nosso objetivo é sugerir uma modificação no processo de ensino já estabelecido onde definimos os conceitos para só após termos consciência de quais deles utilizamos. A proposta desta tese é formalizarmos conteúdos que necessitam aparecer em um problema motivador proposto. Tornando assim a sequência mais orgânica e próxima do método científico que conhecemos. Neste bloco, de 6 a 8 tempos são o necessário.

4. Por último, é introduzido o conceito de validação cruzada e com um exemplo prático sobre teste de gravidez, e aproveitamos para explicar sobre os comuns erros de testes e conceituar sensibilidade e especificidade, definições muito utilizadas em validação cruzada e diversas áreas da saúde, contextualizando assim, também, esse tópico e

motivando sua importância não somente em AM. Esse tema pode ser desenvolvido entre 4 a 6 tempos de aula.

Plano de curso

Tema: Introdução à Inteligência Artificial.

Público-alvo: alunos de graduação nas disciplinas de Estatística.

Duração do curso: 24 tempos de aula;

Objetivos: Os alunos deverão ser capazes de compreender e aplicar técnicas básicas de Aprendizado de Máquinas, interpretar os resultados de modelos e avaliar o desempenho utilizando métricas apropriadas.

Conteúdos:

- Introdução à Inteligência Artificial e Aprendizado de Máquinas;
- Viés e Variância;
- Estatísticas e medidas de desempenho;
- Avaliação de modelos.

Metodologia: aula expositiva, com uso de tecnologia de apoio, como projeção de *slides* sobre a matéria; apostilas e livro.

Avaliação: serão avaliados com questões discursivas referente ao tema.

O que aprendemos neste capítulo?

No Capítulo 4 aborda-se uma dificuldade que foi proposta no capítulo anterior: encontrar o melhor modelo matemático para um problema. Neste Capítulo 4 procurar este modelo parte de alguns conceitos matemáticos, mas por dificuldade em encontrar resultados mais satisfatórios, a decisão final acaba sendo feita por nós humanos, pelo nosso conhecimento e julgamento. Mas, como uma máquina possui tais atributos? Para compreender melhor, retornamos à ideia de inteligência e exploramos alguns conceitos de como a máquina aprende. Quais artifícios matemáticos são utilizados por elas e como

realizam os julgamentos que muitos de nós consideram como ser capaz apenas com a inteligência humana e que não pode ser ensinada a um ser inanimado... Ou será que pode?

CONCLUSÃO

Este trabalho procurou desenvolver uma experimentação didática de um tema muito atual, quando iniciamos a contrução da tese, envolvendo tecnologia: o *Blockchain*. Iniciamos construindo uma revisão sistemática para compreender melhor onde esta tecnologia é utilizada atualmente. Encontrar aplicações na área de validação de informações acadêmicas, metodologias em ensino e classificação da qualidade de cursos nos faz perceber muitas outras formas de se pensar nessa tecnologia e seu poder para o futuro em diversas áreas.

Encontrar motivos para compreender o futuro nos fez indagar sobre o passado. A tecnologia atual sempre vem de algum lugar que precisa ser explorado, pois, o ser humano, está sempre inserido em um contexto e suas ferramentas, em geral, são resultados de suas necessidades. Com isso, revisitamos a história da criptografia e toda sua evolução até seus modelos atuais, que são utilizados em diversas atividades banais no cotidiano. De senhas de *e-mails*, redes sociais até substituição de imagens, a criptografia continua evoluindo a ponto de talvez ser o *blockchain* a ferramenta mais forte construída por ela nos últimos anos.

Procurar exemplos que possam estar mais inseridos na realidade dos alunos é primordial para que haja de fato aprendizado (D'AMBROSIO, 1991). Com esse tema central, buscamos alinhar esse tema, muito ligado a *Bitcoin* e moedas digitais, a temas do conteúdo de Matemática que possuem, geralmente, o mesmo exemplo em livros didáticos e que acabam não motivando alguns alunos. Com isso, encontramos áreas do conhecimento que possuem ligações já bem conhecidas, mas às vezes pouco exploradas matematicamente, e outras que muitas vezes sequer são citadas por nós professores por falta de conhecimento e que podem, como um exemplo introdutório até, ser um motivador para dar início à compreensão de um conceito, ou mesmo de trabalhos que podem ser desenvolvidos pelo aluno, motivando assim a utilização e familiarização do método científico.

Notar que um modelo matemático que pode não funcionar de forma satisfatória para um problema real é um ponto de partida para se aprofundar em modelos mais complexos no futuro, porém, outra problemática ao fim do capítulo despertou um interesse imediato a ser explorado, como a máquina toma decisões com base nos dados disponibilizamos. Exploramos alguns conceitos de IA iniciais para responder essa dúvida e encontramos diversos conceitos matemáticos que podem ser utilizados de forma integrada à IA, e termos outra ótica desses mesmos assuntos em uma área que é, inevitavelmente, um marco para desenvolvimento da sociedade nas próximas gerações.

Outro ponto importante de salientar é que, durante as pesquisas, encontramos raras informações sobre como os algoritmos computacionais funcionavam matematicamente. O pouco que descobrimos era sempre de fontes estrangeiras, demonstrando assim que são

temas que podem ser melhor explorados em trabalhos futuros. As perguntas “Como a máquina aprende” e “Como realiza seus processos matemático- computacionais” ainda possuem muito material a ser explorado e este é um recorte que fez sentido para desenvolver nosso tema central, mas que com a escolha de outro, podemos abordar diferentes conteúdos matemáticos, tanto elementares, quanto de ensino superior, só dependendo de qual algoritmo computacional estaremos utilizando.

Um objetivo que este trabalho também propôs a explorar é de sugerir uma sequência didática para seus capítulos. Nessa vertente, foi idealizada uma sequência que não se propõe a iniciar pelo conteúdo mais básico, pelo historicamente descoberto primeiro e nem pela ementa tradicional dos cursos, e sim pela sua motivação inicial. Iniciamos a escrita desejando compreender melhor sobre a moeda *Bitcoin* e criptoeconomia. Porém, para alcançarmos essa meta, precisamos passar por conceitos como *Blockchain*, que nos fez debruçar no tema de Criptografia, o qual para ser compreendido tinha a necessidade de outro tema e assim sucessivamente, criando, assim, uma trilha natural entre os temas e suas aplicações. Com isso, assuntos vistos em cursos de graduação aparecem de forma menos engessada e são abordados como uma etapa natural a ser explorada para que se tenha avanço em nossa pergunta inicial. Essa sequência didática nos incentiva a estudar tópicos para solucionar questionamentos que se apresentam e pode ser um caminho para que o aluno seja mais motivado a aprender determinados assuntos.

Por fim, por mais que a proposta desta tese seja de propor ideias e sugestões de sequências didáticas, dentro do eixo central *Blockchain*, muitos caminhos ainda ficam abertos para melhor estruturar essa sequência. Como explicar os modelos e recordar muitas informações de diversos níveis matemáticos foi o foco principal, há uma possibilidade de artigo somente focado nessa estruturação didática para que possa ser seguida em algum curso da área de Matemática, além de diversos outros artigos que pensem na sequência didática de outros temas matemático-computacional que estejam em alta, como as apostas esportivas *on-line*, por exemplo. Isso desperta inúmeras possibilidades e trabalhos futuros que pretendo atacar nos próximos anos.

REFERÊNCIAS

AL-MASKARI, A.; T., Riyami; S., Ghnimi. Factors affecting students' preparedness for the fourth industrial revolution in higher education institutions. *Journal of Applied Research in Higher Education*, 2022. ISSN 2050-7003.

ALAMMARY, A. et al. Blockchain-based applications in education: A systematic review. *Applied Sciences*, v. 9, 2019. ISSN 2400;. Disponível em: <www.mdpi.com/journal/applsci>.

ANWAR, S. A. et al. iLearning model approach in creating Blockchain based higher education trust. s.n., [S.l., 2022. Disponível em: <https://ijair.id/index.php/ijair/article/view/258>. Acesso em: 7 abr. 2023.

APPLE. Segurança e criptografia de dados do icloud. s.n., [S.l., 2024. Disponível em: <https://support.apple.com/pt-br/102651#:~:text=Os%20dados%20do%20iCloud%20s%C3%A3o,nos%20data%20centers%20da%20Apple.> Acesso em: 14 jun. 2024.

ARZARELLO, Ferdinando. Semiosis as a multimodal process. *Relime*, p. 267–299, 2006.

AWATI, R.; BERNSTEIN, C.; COBB, M. Padrão de criptografia avançado (AES). 2024. Disponível em: <https://www.techtarget.com/searchsecurity/definition/Advanced-Encryption-Standard>. Acesso em: 13 abr. 2024.

BARRETO, Marina Menna. *Matemática e Educação Sexual*: modelagem do fenômeno da absorção de anticoncepcionais orais diários. 2007. 216 f. Dissertação (Mestrado Profissionalizante em Ensino de Matemática) — Instituto de Matemática, Universidade Federal do Rio Grande do Sul, 2007.

BASSANEZI, Rodney. *Modelagem Matemática*. Blumenau: Dynamis, 1994. 55 p.

_____. Ensino-aprendizagem com modelagem matemática. Editora Contexto, p. 389, 2002.

_____. *Temas e Modelos*. 1ª. ed. [S.l.: s.n.], 2012. 227 p.

BHUTORIA, Aditi. Personalized education and artificial intelligence in the United States, China, and India: A systematic review using a Human-In-the-Loop model. *Computers and Education: Artificial Intelligence*, v. 3, p. 100068, 2022. ISSN 2666-920X. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2666920X22000236>.

BIT2ME. Preço histórico do Bitcoin. Bit2me, 2020. Disponível em: <https://academy.bit2me.com/pt/pre%C3%A7o-hist%C3%B3rico-de-bitcoin>. Acesso em: 23 dez. 2021.

BJELOBABA, Goran et al. Blockchain technologies and digitalization in function of student work evaluation. *Sustainability*, v. 14, n. 9, 2022. ISSN 2071-1050. Disponível em: <https://www.mdpi.com/2071-1050/14/9/5333>.

BOLDRINI, J. L. et al. *Álgebra Linear*. 3ª. ed. [S.l.]: Harper Row do Brasil, 1980. 407 p.

BUCEA-MANEA-ŢONIş, Rocsana et al. Artificial intelligence potential in higher education institutions enhanced learning environment in Romania and Serbia. *Sustainability*, v. 14, n. 10, 2022. ISSN 2071-1050. Disponível em: <https://www.mdpi.com/2071-1050/14/10/5842>.

CAMPOS, Alexandre Henrique; TIZZIOTTI, Guilherme Chaud. Esteganografia e o método de inserção por matriz. *Revista Eletrônica Matemática e Estatística em Foco*, Minas Gerais, Volume 3, n. 1, maio 2015.

CARVALHO, Carlos Eduardo et al. Bitcoin, criptomoedas, blockchain: Desafios analíticos, reação dos bancos, implicações regulatórias. *Fórum Liberdade Econômica*, São Paulo, 2017.

CARVALHO, Luiz Alfredo Vidal de. *Datamining*: - a mineração de dados no marketing, medicina, economia, engenharia e administração. Rio de Janeiro: Editora Ciência Moderna Ltda., 2005. 225 p.

CHAKA, Chaka. Fourth industrial revolution -— a review of applications, prospects, and challenges for artificial intelligence, robotics and blockchain in higher education. *Chaka Research and Practice in Technology Enhanced Learning*, 2023. Disponível em: <https://rptel.apsce.net/index.php/RPTEL/article/view/2023-18002/324>.

CHEN, Xiaorong. Explore the application and challenges of Blockchain technology in the field of higher education in China. In: *Proceedings of the 2022 8th International Conference on Humanities and Social Science Research (ICHSSR 2022)*. Atlantis Press, 2022. p. 2612–2616. ISBN 978-94-6239-580-0. ISSN 2352-5398. Disponível em: <https://doi.org/10.2991/assehr.k.220504.474>.

CHEN, Yan. The impact of artificial intelligence and blockchain technology on the development of modern educational technology. *Mobile Information Systems*, v. 2022, p. 12, 2022. Disponível em: <https://downloads.hindawi.com/journals/misy/2022/3231698.pdf>.

CHINYAMUNJIKO, N.; C., Simon; BHIBHI, P. Rethink thinking Zimbabwean tertiary education in the fourth industrial revolution: The case of a state university. *International Journal of Research Publications*, v. 103, p. 653–674, 2022. Disponível em: <https://ijrp.org/paper-detail/3351>.

CHIVU, Raluca-Giorgiana (Popa) et al. The role of blockchain technologies in the sustainable development of student's learning process. *Sustainability*, v. 14, n. 3, 2022. ISSN 2071-1050. Disponível em: <https://www.mdpi.com/2071-1050/14/3/1406>.

CHOI, E.; CHOI, Y.; PARK, N. Blockchain-centered educational program embodies and advances 2030 sustainable development goals. s.n., [S.l., 2022. Disponível em: <https://www.mdpi.com/2071-1050/14/7/3761>. Acesso em: 7 abr. 2023.

CRUZ, C. C. Passos; LANZILLOTTI, R. S. Mineração de texto na identificação de palavras-chave no contexto do COVID-19 na modelagem Fuzzy. *Cadernos do IME - Série Matemática*, v. 15, 2020. Disponível em: <https://www.e-publicacoes.uerj.br/cadmat/article/view/54666/36443>.

DAEMEN, Joan; RIJMEN, Vincent. AES proposal: Rijndael. Bélgica, 1999. Disponível em: <https://www.cs.miami.edu/home/burt/learning/Csc688.012/rijndael/rijndael_doc_V2.pdf>. Acesso em: 16 mar. 2024.

D'AMBROSIO, Ubiratan. Matemática, ensino e educação: uma proposta global. Revista da Sociedade Brasileira de Educação Matemática, Rio Claro, p. 16, 1991.

_____. Dos fatos reais à modelagem: uma proposta de conhecimento matemático. s.n., [S.l., p. 11, 1999. Disponível em: <http://vello.sites.uol.com.br/modelos.htm>. Acesso em: 9 jan. 2022.

DIÁRIO OFICIAL DA UNIÃO. Brasília: Imprensa Nacional, dez. 2017. 146 p.

E., Enang C. Emerging technologies in teaching and learning of business education programmes in the new normal in tertiary institution in Nigeria. *Nigerian Journal of Business Education*, v. 9, n. 2, 2022. ISSN 2756-5912. Disponível em: <http://www.nigjbed.com.ng/index.php/nigjbed/article/view/589>.

FENG, Yebo; XU, Jiahua; WEYMOUTH, Lauren. University blockchain research initiative (UBRI): Boosting blockchain education and research. *IEEE Potentials*, v. 41, n. 6, p. 19–25, 2022.

FERDIG, Richard E. et al. Examining blockchain protocols, cryptocurrency, NFTs, and other web 3.0 affordances in teacher education. *Journal of Technology and Teacher Education*, Society for Information Technology Teacher Education, Waynesville, NC USA, v. 30, n. 1, p. 5–19, January 2022. ISSN 1059-7069. Disponível em: <https://www.learntechlib.org/p/221200>.

FERREIRA, Juliana Carvalho; PATINO, Cecilia Maria. Entendendo os testes diagnósticos. parte 1. Jornal Brasileiro de Pneumologia, [S.l., p. 2, 2017. Disponível em: <https://www.scielo.br/j/jbpneu/a/rHy8WhCg5cWVWypdf4phDXj/?lang=pt&format=pdf>. Acesso em: 9 ago. 2024.

FREITAS, Mara Luiza Gonçalves. Presença da Inteligência Artificial os projetos pedagógicos de cursos de administração: uma análise. s.n., [S.l., 2020. Disponível em: <https://acrobat.adobe.com/id/urn:aaid:sc:VA6C2:4f5cc71b-e826-42de-bbdb-53992511e753>. Acesso em: 15 jun. 2024.

FURTADO, Danilo P.; FARIA, Cristiane O. De; SASAKI, Diana. Aplicando a decomposição em valores singulares no processamento de dados. *Cadernos Do IME - Série Matemática*, Rio de Janeiro, n. 15, dez. 2020.

GALILEU, Revista. *Os segredos da máquina nazista Enigma são "quebrados" em exames de raio X*. São Paulo: Globo, nov. 2018. Disponível em: <https://revistagalileu.globo.com/Ciencia/noticia/2018/11/segredos-da-maquina-nazista-enigma-sao-quebrados-em-exame-de-raios-x.html>. Acesso em: 16 fev. 2022.

GARG, A. et al. Blockchain-based online education content ranking. *Education and Information Technologies volume*, 2021. Disponível em: <https://link.springer.com/article/10.1007/s10639-021-10797-5>.

GEEKS, Geeks for. O que é criptografia RC4? 2021. Disponível em: <https://www.geeksforgeeks.org/what-is-rc4-encryption/>. Acesso em: 13 abr. 2024.

_____. Algoritmo internacional simplificado de criptografia de dados (IDEA). 2023. Disponível em: <https://www.geeksforgeeks.org/simplified-international-data-encryption-algorithm-idea/>. Acesso em: 13 abr. 2024.

GHOSH, Smarajit; KARAR, Vinod. Blowfish hybridized weighted attribute-based encryption for secure and efficient data collaboration in cloud computing. *Applied Sciences*, v. 8, n. 7, 2018. ISSN 2076-3417. Disponível em: <https://www.mdpi.com/2076-3417/8/7/1119>.

GIUSEPPE, Terranova. The new geography of asylum: digital identity, artificial intelligence and blockchain. *AIMS Geosciences*, 2022. Disponível em: <http://www.aimspress.com/aimspress-data/aimsgeo/2022/3/PDF/geosci-08-03-022.pdf>.

GOMES, Pedro C. T. Machine Learning para todos, de forma simples e com exemplos! Data Geeks, 2019. Disponível em: <https://www.datageeks.com.br/machine-learning/>. Acesso em: 13 abr. 2022.

GOOGLE. Primeiros passos com arquivos criptografados no drive, no documentos, no planilhas e no apresentações. s.n., [S.l., 2024. Disponível em: <https://support.google.com/docs/answer/10519333?hl=pt-BR&co=GENIE.Platform%3DAndroid>. Acesso em: 14 jun. 2024.

GROENWALD, Claudia Lisete Oliveira; OLGIN, Clarissa de Assis. Criptografia e o currículo de matemática no ensino médio. Revista de Educação Matemática, São Paulo, 2011.

HANNAN, Shaikh Abdul. Application and scope of blockchain in technical research and higher education. *NeuroQuantology*, v. 20, p. 6185–6191, 12 2022.

HWANG, Gwo-Jen; CHIEN, Shu-Yun. Definition, roles, and potential research issues of the metaverse in education: An artificial intelligence perspective. *Computers and Education: Artificial Intelligence*, v. 3, p. 100082, 2022. ISSN 2666-920X. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2666920X22000376>.

INFOMONEY. Cotações da moeda Bitcoin no gráfico. Infomoney, 2021. Disponível em: <https://www.infomoney.com.br/cotacoes/cripto/ativo/bitcoin-btc/grafico/>. Acesso em: 22 ago. 2021.

KHAN, H. U. et al. Transforming the capabilities of artificial intelligence in GCC financial sector: A systematic literature review. *Wireless Communications and Mobile Computing*, v. 2022, 2022. Disponível em: <https://www.hindawi.com/journals/wcmc/2022/8725767/>.

KHOLISHOTULAILA, Siti; LAILA, Kholishotu; ANGGA, Ayu Lestari. Benefits provided by blockchain technology in the field of education. *Blockchain Frontier Technology*, 2022.

KIM, S.; JANG, E. The intelligent blockchain for the protection of smart automobile hacking. *Journal of Multimedia Information System*, v. 9, n. 1, p. 33–42, 2022. Disponível em: <https://koreascience.kr/article/JAKO202211757489199.page>.

KOHAVI, Ron. A study of cross-validation and bootstrap for accuracy estimation and model selection. International joint Conference on artificial intelligence, [S.l., p. 1137–1145, 1995.

KOMATH, C. M.A; O., Saylir. How will blockchain transform the ed-tech industry? integration of blockchain technology in learning management systems (LMSS) in Turkey. *Socrates Journal of Interdisciplinary Social Studies*, v. 22, 2022. Disponível em: <https://socratesjournal.org/index.php/pub/article/view/167>.

KOMESU, Daniel. *Trade off entre viés e variância (links)*. 2022. Disponível em: <https://dkko.me/posts/20220130-trade-off-vies-variancia/index.html>. Acesso em: 28 mar. 2024.

KOSASI, Sandy et al. Blockchain technology - emerging research themes opportunities in higher education. In: *2022 International Conference on Science and Technology (ICOSTECH)*. [S.l.: s.n.], 2022. p. 1–8.

KULETO, V. et al. The potential of Blockchain technology in higher education as perceived by students in Serbia, Romania, and Portugal. 2022. Disponível em: <https://medium.com/prognosys/criptografia-sim%C3%A9trica-6b4271ff697c>. Acesso em: 07 abr. 2023.

KUMAR, Nayana N. et al. Decentralized storage of educational assets using NFTs and Blockchain technology. In: *2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT)*. [S.l.: s.n.], 2022. p. 260–266.

LADEIRA, Ricardo de la R.; RAUGUST, Anderson S. Uma análise da complexidade do algoritmo RSA implementado com o teste probabilístico de Miller-Rabin. *Revista de Empreendorismo, Inovação e Tecnologia*, v. 4, n. 1, 2017. ISSN 2359-3539. Disponível em: <https://seer.atitus.edu.br/index.php/revistasi/article/view/1639/1296>.

LEVITSKAYA, Aleksandra N.; POKROVSKAIA, Nadezhda N.; RODIONOVA, Elena A. Blockchain platforms as a tool for resolving contradictions in the labor market and improving the interaction of the applicant and employer. In: *2022 Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus)*. [S.l.: s.n.], 2022. p. 1698–1703.

LI, C. The education cloud platform for digital resources with Blockchain under intelligent learning environment. *Creative Education*, v. 13, 2022. ISSN 599-608. Disponível em: <https://www.scirp.org/journal/paperinformation.aspx?paperid=115367>.

LI, Jie. Adaptive learning model of english vocabulary based on Blockchain and deep learning. *Mobile Information Systems*, v. 2022, 2022. Disponível em: <https://www.hindawi.com/journals/misy/2022/4554190/>.

LI, Joey et al. Methods and applications for artificial intelligence, big data, internet of things, and Blockchain in smart energy management. *Energy and AI*, v. 11, p. 100208, 2023. ISSN 2666-5468. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2666546822000544>.

LIU, Feng; FAN, Hao-Yang; QI, Jia-Yin. Blockchain technology, cryptocurrency: Entropy-based perspective. *Entropy*, v. 24, n. 4, 2022. ISSN 1099-4300. Disponível em: <https://www.mdpi.com/1099-4300/24/4/557>.

MAINETTI, Luca et al. Digital brick: Enhancing the student experience using Blockchain, Open Badges and Recommendations. *Education Sciences*, v. 12, n. 8, 2022. ISSN 2227-7102. Disponível em: <https://www.mdpi.com/2227-7102/12/8/567>.

MALAR, João Pedro. Criação de criptomoedas se tornou mais simples e pode ocorrer em horas; entenda. CNN Brasil, 2021. Disponível em: <https://www.cnnbrasil.com.br/business/criacao-de-criptomoedas-se-tornou-mais-simples-e-pode-ocorrer-em-horas-entenda/#:~:text=Os%20levantamentos%20variam%2C%20mas%20indicam,crescente%20de%20criar%20uma%20criptomoeda.> Acesso em: 3 fev. 2022.

MARCONI, Marina de Andrade; LAKATOS, Eva Maria. *Fundamentos de metodologia científica*. 5ª. ed. Blumenau: Atlas, 2003. 310 p.

MASADEH, Rawan. Study of NFT-secured Blockchain Technologies for High Security Metaverse Communication. *Symposium of Student Scholars*, v. 26, 2022. Disponível em: <https://digitalcommons.kennesaw.edu/cgi/viewcontent.cgi?article=1986&context=undergradsymposiumksu>.

MEIROBIE, I. et al. Framework authentication e-document using Blockchain technology on the government system. *Artificial Intelligence Research*, v. 6, 2022. ISSN 2579-7298. Disponível em: <http://ijair.id/index.php/ijair/article/view/294>.

MENEZES, Pedro. O que é ciência? 2022. Disponível em: <https://www.significados.com.br/ciencia/>. Acesso em: 31 jan. 2022.

MIN, L.; BIN, G. Online teaching research in universities based on blockchain. *Education and Information Technologies*, v. 27, 2022. Disponível em: <hhttps://link.springer.com/article/10.1007/s10639-022-10889-w#citeas>.

MODERNA, Editora. Conexões: matemática e suas tecnologias. Editora Moderna, São Paulo, 2020. Disponível em: <https://pnld.moderna.com.br/wp-content/uploads/2021/05/FP_0193P21202_3_MP_PDF_CARAC.pdf>. Acesso em: 10 fev. 2024.

MONTGOMERY, D. C.; RUNGER, G. C. *Applied statistics and probability for engineers*. 6ª. ed. [S.l.]: John Wiley Sons, 2013. 811 p.

MOORE, Walter B.; FELO, Andrew. The evolution of accounting technology education: Analytics to STEM. *Journal of Education for Business*, Routledge, v. 97, n. 2, p. 105–111, 2022. Disponível em: <https://doi.org/10.1080/08832323.2021.1895045>.

MORA-ZAMBRANO, Eugenio Rafael et al. Revisión sistemática de la implementación de blockchain en el sector educativo. *Ciencia Huasteca Boletín Científico de la Escuela Superior de Huejutla*, v. 10, n. 19, p. 28–34, ene. 2022. Disponível em: <https://repository.uaeh.edu.mx/revistas/index.php/huejutla/article/view/8327>.

MOURTZIS, D.; ANGELOPOULOS, J.; PANOPOULOS, N. A teaching factory paradigm for personalized perception of education based on extended reality (XR). In:

Proceedings of the 12th Conference on Learning Factories (CLF 2022). [S.l.: s.n.], 2022. p. 09.

MOZUMDER, Md Ariful Islam et al. Technological roadmap of the future trend of metaverse based on IoT, blockchain, and AI techniques in metaverse education. In: *2023 25th International Conference on Advanced Communication Technology (ICACT)*. [S.l.: s.n.], 2023. p. 1414–1423.

MUELLER, John P.; MASSARON, Luca. *Aprendizado de Máquinas para Leigos*. 1ª. ed. [S.l.]: Alta Books, 2019. 432 p.

MUHATI, Eric; RAWAT, Danda B.; SADLER, Brian M. A new cyber-alliance of artificial intelligence, internet of things, blockchain, and edge computing. *IEEE Internet of Things Magazine*, v. 5, n. 1, p. 104–107, 2022.

MUSTI, Adam Shettima; KANT, Shri; KHANNA, Tejaswi. Degchain: Development of blockchain framework for generation and verification of educational certificates. In: *2022 IEEE 7th International conference for Convergence in Technology (I2CT)*. [S.l.: s.n.], 2022. p. 1–7.

NAKAMOTO, Satoshi. A peer-to-peer eletronic cash system. s.n., [S.l., p. 9, 2008. Disponível em: <https://bitcoin.org/bitcoin.pdf>. Acesso em: 7 dez. 2021.

NEMORIN, Selena et al. AI hyped? a horizon scan of discourse on artificial intelligence in education (AIED) and development. *Learning, Media and Technology*, Routledge, v. 48, n. 1, p. 38–51, 2023. Disponível em: <https://doi.org/10.1080/17439884.2022.2095568>.

NOKITI, Abdullah et al. Is blockchain the answer? a qualitative study on how blockchain technology could be used in the education sector to improve the quality of education services and the overall student experience. *Jisuanji Jicheng Zhizao Xitong/Computer Integrated Manufacturing Systems, CIMS*, v. 28, p. 1006–5911, 11 2022.

NUROVIC, E.; POTURAK, M. Rethinking the concept of the education th-rough digitalization of higher education institutions and blockchain. *International Journal of Social Sciences Educational Studies*, v. 8, p. 8, 2021. ISSN 2520-0968. Disponível em: <https://ijsses.tiu.edu.iq/wp-content/uploads/2021/06/Rethinking-the-Concept-of-the-Education-Through-Digitalization-of-/Higher-Education-Institutions-and-Blockchain.pdf>.

OCHEJA, P. et al. Blockchain in education: A systematic review and practical case studies. *IEEE EDUCATION SOCIETY SECTION*, 09 2022.

OKOLI, Chitu; KIRA, Schabram. A guide to conducting a systematic literature review of information systems research. Social Science Research Network, p. 50, 2010. Disponível em: <https://www.scinapse.io/papers/1539987097>. Acesso em: 20 jan. 2023.

PAIXÃO, Jéssica Shayanne da. *Criptografia*: história, atividades e divulgação científica. 2020. 177 p. Dissertação (Mestrado Profissional em Matemática em Rede Nacional (PROFMAT)) — Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2020.

PANAGIOTIDIS, P. Blockchain in education-the case of language learning. s.n., [S.l., 2022. Disponível em: <https://revistia.org/index.php/ejed/article/view/5907>. Acesso em: 7 abr. 2023.

RIKWALDER, Eric. A matemática por trás do Bitcoin. 2014. Disponível em: <<https://pt.bitcoinonair.com/math-behind-bitcoin>>. Acesso em: 20 dez. 2021.

RIMES, Ighor O. M.; FURST, Patrícia. *Probabilidade e Estatística no Ensino Médio – problemas e situações que despertam o interesse do aluno –*. 2017. 37 f. Dissertação (Especialização em Aprendizagem em Matemática) — Instituto de Matemática e Estatística, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2017.

SAVELYEVA, Tamara; PARK, Jae. Blockchain technology for sustainable education. *British Journal of Educational Technology*, v. 53, n. 6, p. 1591–1604, 2022. Disponível em: <https://bera-journals.onlinelibrary.wiley.com/doi/abs/10.1111/bjet.13273>.

SAYAD, Alexandre Le Voci. *Inteligência artificial e pensamento crítico*: : caminhos para a educação midiática. 1ª. ed. São Paulo: Instituto palavra Aberta, 2023. 160 p.

SHAHZAD, Khurram; ASEERI, Ahmad O.; SHAH, Munam Ali. A Blockchain-based authentication solution for 6G communication security in tactile networks. *Electronics*, v. 11, n. 9, 2022. ISSN 2079-9292. Disponível em: <https://www.mdpi.com/2079-9292/11/9/1374>.

SHAIKH, Zaffar Ahmed et al. Blockchain hyperledger with non-linear machine learning: A novel and secure educational accreditation registration and distributed ledger preservation architecture. *Applied Sciences*, v. 12, n. 5, 2022. ISSN 2076-3417. Disponível em: <https://www.mdpi.com/2076-3417/12/5/2534>.

SILVA, Luiz Gustavo Doles. *A regulação do uso de criptomoedas no Brasil*. 2017. 132 f. Dissertação (Mestrado em Direito Político), Universidade Presbiteriana Mackenzie,São Paulo, 2017.

SINGH, Simon. *O livro dos códigos*. 6ª. ed. [S.l.]: Record, 2001. 448 p.

SOUZA, Fábio. *Conversão entre sistemas de numeração*. 2016. Disponível em: <https://embarcados.com.br/conversao-entre-sistemas-de-numeracao/>. Acesso em: 28 mar. 2024.

SOUZA, Fernando. Criptografia simétrica e assimétrica. Medium, 2020. Disponível em: <https://medium.com/prognosys/criptografia-sim%C3%A9trica-6b4271ff697c>. Acesso em: 25 jan. 2023.

SPOLADOR, Rodrigo Mesquita. Precisamos falar de Bitcoin! *Encontro de Iniciação Científica do Centro Universitário Antonio Eufrásio de Toledo de Presidente Prudente*, v. 13, n. 13, 2017.

STARMER, Josh. Statquest with Josh Starmer. s.n., [S.l., 2019. Disponível em: <https://www.youtube.com/channel/UCtYLUTtgS3k1Fg4y5tAhLbw>. Acesso em: 12 ago. 2023.

STEGHIDE. Bem-vindo ao site da Steghide! 2003. Disponível em: <https://steghide.sourceforge.net/>. Acesso em: 14 abr. 2024.

SUNNY, Farhana Akter et al. A systematic review of Blockchain applications. *IEEE Access*, v. 10, p. 59155–59177, 2022.

SUPRIATI, Ruli et al. Utilizing the potential of blockchain technology for leading education 4.0. In: *2022 International Conference on Science and Technology (ICOSTECH)*. [S.l.: s.n.], 2022. p. 01–08.

TALPUR, M. H.; TALPUR, F.; HASEEB, A. A model for secure inter-institutional communication based on Artificial Intelligence and Blockchain. s.n., [S.l., 2022. Disponível em: <https://www.researchgate.net/profile/Asadullah-Kehar/publication/357928593_A_Model_for_Secure_Inter-Institutional_Communication_Based_on_Artificial_Intelligence_AI_and_Blockchain/links/61e7bf889a753545e2df34df/A-Model-for-Secure-Inter-Institutional-Communication-Based-on-Artificial-Intelligence-AI-and-Blo pdf>. Acesso em: 7 abr. 2023.

TAMAROZZI, Antônio Carlos. Codificando e decifrando mensagens. Revista do Professor de Matemática, Rio Claro, 2001.

THOMASON, Jane. Metaverse, token economies, and non-communicable diseases. *Global Health Journal*, v. 6, n. 3, p. 164–167, 2022. ISSN 2414-6447. Special Issue on Prevention and Control of Obesity and Related Non-communicable Diseases. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2414644722000458>.

THUAN, Nguyen Dinh et al. Using blockchain and artificial intelligence to build a job recommendation system for students in information technology. In: *2022 RIVF International Conference on Computing and Communication Technologies (RIVF)*. [S.l.: s.n.], 2022. p. 364–369.

ULRICH, Fernando. *Bitcoin*: A moeda na era digital. 1ª. ed. São Paulo: LVM Editora, 2017. 127 p.

VALENTE, Jorge A. O computador na sociedade do conhecimento. *Coleção Informática para mudança na educação*, UNICAMP, 1999.

VELASQUEZ, Ester. O que são autovalores e autovetores?: Exercícios passo a passo | Álgebra linear. 2021. Disponível em: <https://www.youtube.com/watch?v=3UzV21Ak3uc>. Acesso em: 21 mar. 2024.

VESELOV, Gennady et al. Training of engineers: Approaches to customization of educational programs. In: *2022 IEEE Global Engineering Education Conference (EDUCON)*. [S.l.: s.n.], 2022. p. 590–596.

VIANA, Cleberton Junio et al. *Criptografia e Segurança*. [S.l.]: Revista Científica E-Locução, dez. 2022.

WANG, Yaofei; SUN, Qiurui; BIE, Rongfang. Blockchain-based secure sharing mechanism of online education data. *Procedia Computer Science*, v. 202, p. 283–288, 2022. ISSN 1877-0509. International Conference on Identification, Information and Knowledge in the internet of Things, 2021. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1877050922005713>.

WANG, Zhipeng. The empirical study on developing strategies of AI education in China. In: *Proceedings of the 2nd International Conference on Internet, Education and Information Technology (IEIT 2022)*. Atlantis Press, 2022. p. 743–748. ISBN 978-94-6463-058-9. ISSN 2352-538X. Disponível em: <https://doi.org/10.2991/978-94-6463-058-9_116>.

WANG, Zhifeng et al. Smart contract vulnerability detection for educational blockchain based on graph neural networks. In: *2022 International Conference on Intelligent Education and Intelligent Research (IEIR)*. [S.l.: s.n.], 2022. p. 8–14.

WIKIPEDIA. Data Encryption Standard. 2001. Disponível em: <https://en.wikipedia.org/wiki/Data_Encryption_Standard>. Acesso em: 13 abr. 2024.

YANG, M.; WANG, J. The security of student information management system based upon Blockchain. *Journal of Electrical and Computer Engineering*, v. 2022, p. 1, 2022. Disponível em: <https://www.hindawi.com/journals/jece/2022/9781939/>.

YING, Zhang. Increasing cyber defense in the music education sector using blockchain zero-knowledge proof identification. *Computational Intelligence and Neuroscience*, v. 2022, 2022. Disponível em: <https://www.hindawi.com/journals/cin/2022/9922167/>.

ZABALA, Enrique. AES Rijndael cipher explained as a flash animation. s.n., [S.l., 2017. Disponível em: <https://www.youtube.com/watch?v=gP4PqVGudtg>. Acesso em: 7 mar. 2024.

ZENDESK. Machine learning: tudo sobre a tecnologia de aprendizagem de máquina. s.n., [S.l., 2024. Disponível em: <https://www.zendesk.com.br/blog/machine-learning/>. Acesso em: 9 ago. 2024.

ZHANG, X.; GOYAL, S. B. Security and privacy challenges using IoT-Blockchain technology in a smart city: Critical analysis. *International Journal of Electrical and Electronics Research (IJEER)*, v. 10, n. 2, p. 190–195, 2022. ISSN 2347-470X. Disponível em: <https://ijeer.forexjournal.co.in/papers-pdf/ijeer-100224.pdf>.

ZHAO, Gang; DI, Bingbing; HE, Hui. A novel decentralized cross-domain identity authentication protocol based on blockchain. *Transactions on Emerging Telecommunications Technologies*, v. 33, n. 1, p. e4377, 2022. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/ett.4377>.

ZHAO, Mao et al. Blockchain in online learning: A systematic review and bibliographic visualization. *Sustainability*, v. 15, n. 2, 2023. ISSN 2071-1050. Disponível em: <https://www.mdpi.com/2071-1050/15/2/1470>.

ZHENG, Wen Qi. Design and implementation of an intelligent education platform based on Blockchain technology. In: *CIBDA 2022; 3rd International Conference on Computer Information and Big Data Applications*. [S.l.: s.n.], 2022. p. 1–6.