

## **Análise preditiva de dados: uma abordagem de conceitos e algoritmos**

### **Predictive data analytics: a concepts and algorithms approach**

**Jonatas Cerqueira Dias** 

Fatec Praia Grande  
jonatas.dias2@fatec.sp.gov.br

**Jeferson Cerqueira Dias** 

Fatec Itaquera  
jefersoncdias@hotmail.com

#### **RESUMO**

Com o crescimento da quantidade de dados, atrelado ao barateamento das tecnologias de coleta e gestão dos dados, é notável o aumento do interesse acerca da tecnologia Big Data Analytics, tendo também em vista seus resultados e “cases” de sucesso. Porém, é notado também uma lacuna de conhecimento em leitores interessados no tema, principalmente no que tange a predição. Desta forma, o objetivo deste estudo é abordar o assunto a partir da ótica de “marketing” e conduzir o leitor para o entendimento técnico da obtenção da característica da análise preditiva para a tomada de decisão.

**PALAVRAS-CHAVE:** Big Data Analytics; Machine Learning; Análise Preditiva; Algoritmo; Flor de Iris.

#### **ABSTRACT**

*With the growth of the amount of data, coupled with the cheapening of data collection and management technologies, it is also notable the growth of interest in the Big Data Analytics technology, also in view of its results and success "cases". However, it is also noted that there is a gap in the knowledge of readers interested in the subject, especially regarding prediction. Thus, the objective of this study is to address the subject from the perspective of marketing and lead the reader to the technical understanding of obtaining the characteristic of predictive analytics for decision making.*

**KEY WORDS:** Big Data Analytics; Machine learning; Predictive analysis; Algorithm; Iris flower.

#### **INTRODUÇÃO**

O crescimento exponencial da informação durante os últimos anos, atrelado ao barateamento de tecnologias que permitem a coleta e o armazenamento de grandes volumes de dados, fez com que organizações de diversos segmentos começassem a encontrar dificuldades relacionadas ao gerenciamento de dados, dificultando a obtenção de informações valiosas e informações que permitissem a tomada de decisões (DAVENPORT, 2017; MARQUESONE, 2018). Desta forma, uma abordagem tecnológica conhecida como “Big Data Analytics” busca

auxiliar no gerenciamento e tratamento da informação e, principalmente, extrair conhecimento a partir deste volume de dados, com o objetivo de minimizar riscos e incertezas nas tomadas de decisões, bem como potencializar a vantagem competitiva organizacional, a partir de previsões baseadas em dados (Ibidem).

Esta tecnologia atrai empresas como *Facebook*, *Twitter*, *Master Card*, dentre outras, pois, além de fornecer uma finalidade para a coleta da informação, o “Big Data Analytics” auxilia as organizações a compreenderem o cenário atual de seu negócio a partir dos dados, permitindo-lhes conhecer seus pontos fortes e fracos, diminuir riscos, ou, potencializar seu negócio com estratégias personalizadas e atendimentos ao cliente (COHEN, 2018; SOTO, 2017).

Mesmo com a ascensão do “Big Data Analytics”, a literatura apresenta diversos desafios para o aperfeiçoamento desta tecnologia (PANDOVE; GOEL; RANI, 2018; PHILIP CHEN; ZHANG, 2014). Estes desafios podem ser: éticos (como a privacidade dos dados), computacionais (como a quantidade de dados e sua variabilidade), qualitativos (como o processo de limpeza ou redução dos dados), dentre outros como o processamento de dados em tempo real (CHEN; ZHANG; 2014; SOTO, 2017). Apesar destas dificuldades, seu uso é indispensável para o mercado, pois fornece ganho competitivo e fortalecimento estratégico.

Um outro ponto observado no uso desta nova tecnologia, é a compreensão de sua sistemática de funcionamento, voltada para a obtenção de informação, oferecendo um cenário de tomada de decisão com o menor risco e incerteza. Antes, porém, faz-se necessário compreender duas linhas de estratégia de análise dos dados: a diagnóstica e a prognóstica.

A estratégia de diagnose consiste em analisar e compreender os fatos que já ocorreram, minimizando os riscos na tomada de decisão. Em linhas gerais, segundo Marquesone (2018), esta forma de análise visa responder o que aconteceu (análise descritiva) e o porquê aconteceu (análise diagnóstica).

A outra estratégia, com foco no prognóstico, é uma forma de análise que oferece a habilidade de compreender o passado e visualizar o futuro, a partir da análise dos dados registrados (análise preditiva), possuindo também a capacidade de descrever possíveis eventos (análise prescritiva) (MARQUESONE, 2018). Estes tipos de análise são aplicados em diversos campos de estudo ou do mercado, reduzindo principalmente as incertezas na tomada de decisão (GANDOMI; HAIDER, 2015; MARQUESONE, 2018).

A proposta de estratégia adotada para este estudo é voltada para a análise prognóstica, conduzindo o leitor a partir de uma visão de “marketing” desta tecnologia, para uma visão técnica. Com relação a sistematização do funcionamento desta tecnologia, existe, no grande

público, uma grande dificuldade em compreender e ilustrar a maneira como este prognóstico é realizado. São dúvidas frequentes: Como é possível elaborar e obter resultados de uma análise preditiva? Quais ferramentas auxiliam neste processo? Como de fato é feita uma análise preditiva?

A análise prognóstica e sua sistematização, conforme abordado no tópico anterior, torna-se objeto de estudo uma vez que seus desafios e suas dificuldades são uma realidade atual. O mercado reconhece a vantagem competitiva que esta tecnologia proporciona, tendo em vista o constante crescimento do interesse por parte dos gestores e acadêmicos a partir do ano de 2011, quando o tema passa a ter ascensão (GOOGLE TRENDS, 2022)<sup>1</sup>, de acordo com o gráfico apresentado na Figura 1.



Após apresentar um vislumbre da sistematização da forma preditiva de análise e extração do conhecimento, é essencial compreender os principais conceitos técnicos e abordagens preditivas presentes na literatura. Desta forma, o conteúdo apresentado aborda este assunto com foco na seguinte problemática: a partir de uma visão comercial de análise preditiva para a tomada de decisão, de que forma realista esta predição pode ser efetuada?

<sup>1</sup> *Big Data Analytics* pelo *Google Trends* - A busca considerou as seguintes variáveis: no mundo todo, a partir de 2004 em todas as categorias. Também foram avaliadas categorias como: ciência, comercial e indústria, livros e leitura, bem como pessoas e sociedade; todas demonstraram a mesma ascensão no interesse pelo assunto a partir de 2011, conforme apresentado no gráfico.

Para responder ao problema de pesquisa, inicialmente foram identificadas as hipóteses que vão delinear a busca da solução para o problema proposto, como segue:

1. Os dados utilizados na análise preditiva, podem ser avaliados por algoritmos estudados na área de Machine Learning, tal que estes algoritmos possam ser utilizados com maior ou menor grau de acerto das causalidades.

2. Uma compreensão melhor das limitações, técnicas e ferramentas de uso permitem ao gestor usufruir dos benefícios empregados na extração de valor dos dados.

Com estas hipóteses definidas, buscou-se um objetivo que norteasse a pesquisa, resultando no seguinte: apresentar os conceitos específicos do Big Data Analytics e do aprendizado de máquina, levando o leitor para uma esfera de conhecimento da ciência de dados focada em prever situações ou elementos.

## **1. MÉTODO APLICADO A REALIZAÇÃO DA PESQUISA**

Para alcançar o objetivo proposto, a presente pesquisa, em conformidade com suas características, pode ser classificada, em termos de natureza, como uma pesquisa aplicada, pois se trata da busca da solução de um problema real existente (GARCIA, 2016; GIL, 2008), e se utilizou de uma estratégia com duas abordagens: a) uma que direciona o raciocínio lógico por meio dos resultados para obtenção da conclusão e generalização do conhecimento obtido, denominado de “Método de Abordagem” e b) outra de procedimentos, que esclarece acerca dos “Procedimentos Técnicos” utilizados (GARCIA, 2016; GIL, 2008; MARCONI; LAKATOS, 2003; SESC SÃO PAULO/CEBRAP, 2016; SILVA, 2015).

**Método de Abordagem** - O método utilizado para a formação das conclusões foi de caráter dedutivo, pois, parte-se das observações e de um conhecimento prévio estabelecido pelo repertório bibliográfico existente (GIL, 2008; SILVA, 2015), aliado a uma abordagem experimental, pois, se procurou interpretar e analisar o fenômeno observado a partir de dados disponíveis em base de dados de acesso público, na interpretação da solução proposta para o problema de pesquisa (Ibidem).

**Procedimentos Técnicos** - Em termos de procedimentos técnicos, utilizou-se da pesquisa bibliográfica, pois, o uso de uma abordagem de pesquisa deste tipo, provoca um aprofundamento da investigação das questões relacionadas ao fenômeno em estudo, procurando não captar apenas a aparência do fenômeno, mas também suas essências, explicando sua origem, relações e mudanças, e tentando perceber as consequências (GIL, 2008; SILVA, 2015).

Com a adoção desta linha metodológica, foi necessário identificar quais seriam as fontes primárias e secundárias; decidiu-se pelas fontes primárias: livros, relatórios técnicos, artigos em revistas científicas, anais de congressos. As teses universitárias não foram contempladas nesta oportunidade. Como fontes secundárias foram considerados: os trabalhos de artigos de revisão bibliográfica<sup>2</sup>, enciclopédias e os artigos de divulgação<sup>3</sup>.

Como apoio, esta pesquisa utilizou bases de dados referenciais que listam referências bibliográficas de determinados assuntos, cujo conteúdo abrange a descrição dos dados dos artigos de periódicos. Quatro destas bases foram selecionadas: Google Acadêmico; Web of Science, Scopus e IEEE. Para elucidar o conhecimento abordado no desenvolvimento deste trabalho, de uma forma experimental, escolheu-se a plataforma Weka<sup>4</sup>, permitindo compreender o comportamento de uma análise de dados.

## 1.1. ESTRATÉGIA DE APLICAÇÃO DA ANÁLISE PREDITIVA

Para a demonstração da análise preditiva, este estudo aborda uma estratégia utilizando-se da abstração<sup>5</sup>, aliada à verossimilhança<sup>6</sup>, a partir de duas situações:

1. A primeira, simplificada e consagrada na literatura, como é o caso da classificação da flor de Iris<sup>7</sup> em um dos três grupos existentes (Versicolor, Setosa, Virginica), ilustrados na Figura 2;

---

<sup>2</sup> **Artigos de revisão bibliográfica** se concentra na organização de textos publicados apresentando os principais caminhos da pesquisa acerca de um determinado assunto, resumindo, analisando e discutindo informações já publicadas (GRUPO ANIMA EDUCAÇÃO, 2014; MARTINO, 2018).

<sup>3</sup> **Artigos de divulgação científica** é um tipo de texto expositivo e argumentativo mais elaborado. São produzidos mediante pesquisas, aprofundamentos teóricos e resultados de investigações sobre determinado tema. Possuem a finalidade principal de “popularizar a ciência”, ou seja, difundir o conhecimento científico, transmitindo assim diversas informações de valor (GIACHETI, 2006).

<sup>4</sup> **Weka** é uma coleção de algoritmos que trabalha segundo a filosofia de aprendizagem de máquina, os quais operam para solucionar problemas de mineração de dados. Contém ferramentas para pré-processamento, classificação, regressão e agrupamento. A biblioteca possui código aberto e é amparada pela *GNU General Public License* (UNIVERSIDADE DE WAIKATO, 2019).

<sup>5</sup> **Abstração** é uma operação intelectual que consiste em isolar um aspecto complexo do objeto real a fim de simplificar a sua avaliação, classificação, ou para permitir a comunicação do mesmo. (“Dicio - Dicionário Online de Português”, 2019; “Que Conceito - Dicionário”, 2019; PRIBERAM, 2019).

<sup>6</sup> **Verossimilhança** é a ligação, semelhança entre fatos, ideias, etc. numa obra literária, ainda que os elementos imaginários ou fantásticos sejam determinantes no texto; coerência (HOLANDA, 2010).

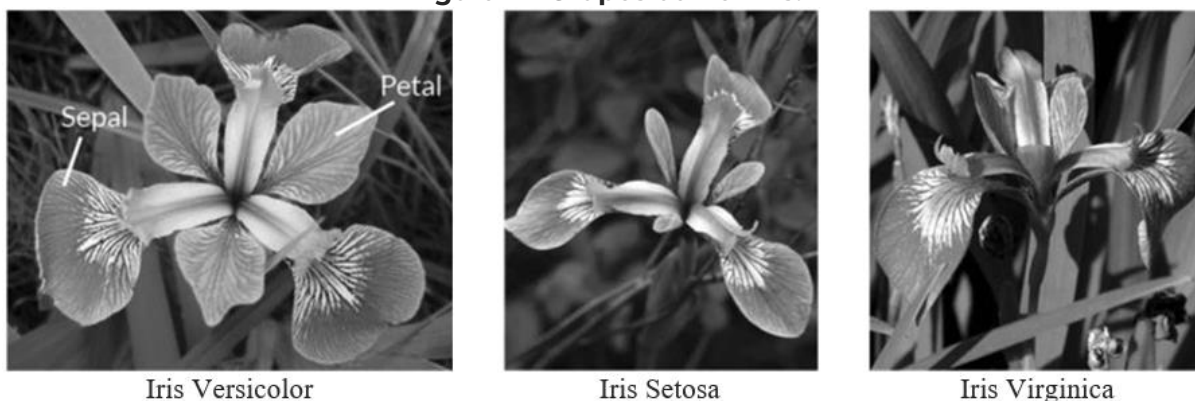
<sup>7</sup> **Iris** é uma base de dados de espécies de flores da família das Iridáceas chamada Iris. Existem três classes registradas neste banco de dados: a Iris-setosa, a Iris-versicolor e a Iris-virginica. Cada uma contém 50 instâncias, totalizando 150 destas. Cada instância possui 4 atributos: o comprimento da sépala, a largura da sépala, o comprimento da pétala e a largura da pétala (UNIVERSIDADE DE WAIKATO, 2019). A pétala e sépala são apontadas na **Figura 2**, primeiro quadro (*petal*, *sepal*).

2. A segunda, próxima da realidade, relacionada a necessidade de identificar se um determinado cliente, na solicitação de um empréstimo, é ou não bom pagador, reduzindo a taxa de empréstimos fraudulentos (GÜLTEKIN; ŞAKAR, 2018).

Observa-se que estas duas situações são similares, pois, não saber a qual dos três grupos pertence uma determinada flor de íris coletada ao acaso, equivale a não saber em qual grupo de pagadores pertence um novo cliente de empréstimo. Esta estratégia deve conduzir o leitor na compreensão da técnica aplicada a primeira situação (1) e transporta esta compreensão para a segunda (2), percebendo o *modus operandi* de uma das técnicas de predição.

A classificação da flor Iris está disponível na plataforma Weka. É conhecido pela literatura e didaticamente utilizado por diversos autores. Com uma massa de registros proposta por Fischer em 1936, concede espaço a aplicação de técnicas preditivas.

**Figura 2 - Grupos da flor Iris.**



Fonte: SAWLA (2018).

## 1.2. CARACTERÍSTICAS E IMPACTO NA PREDIÇÃO

O princípio comum entre estas situações é a identificação de variáveis que caracterizam o objeto de interesse (AMARAL, 2016; MWADULO, 2016; PROVOST; FAWCET, 2016).

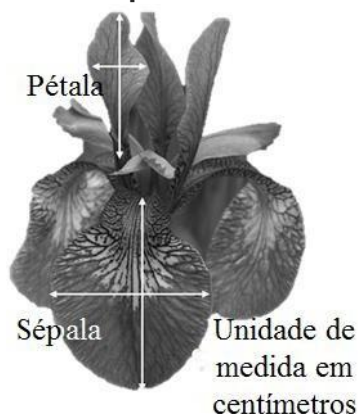
No caso da flor Iris, são utilizados o comprimento e a largura da sépala<sup>8</sup> e da pétala<sup>1</sup> para a sua classificação, pois, estas variáveis determinam as características discrepantes, permitindo distinguir uma flor dentre todas as outras (FAYYAD; GRINSTEIN; WIERSE, 2002; FISCHER, 1936), identificando em qual dos três grupos ela pertence. A Figura 3 ilustra a localização das medidas tomadas a partir da flor de íris, permitindo sua classificação.

<sup>8</sup> **Sépalas** são as partes semelhantes a folhas que envolvem o botão da flor. Elas se dobram e protegem o botão do clima ou de lesões. **Pétalas**, em geral coloridas, têm a função de atrair os insetos para polinizar, trazendo o pólen de outra flor da mesma espécie, depositando-o no estigma (BARRETO, 2017).

Para encontrar características relevantes de elementos em uma massa de dados, é preciso utilizar técnicas de seleção de recursos e de métodos estatísticos que indiquem um grau de relevância entre elas (MWADULO, 2016). Outra maneira, é o entendimento do negócio (“Business Understanding”) e dos dados (“Data Understanding”) e a partir daí definir quais serão os recursos/características utilizados no treino e classificação do modelo (PROVOST; FAWCET, 2016).

Para o caso da classificação do cliente solicitante de empréstimo, as características determinantes podem ser: salário, saldo bancário, e o limite do cartão de crédito. Porém, estas características podem variar de acordo com a necessidade da classificação, influenciadas pela particularidade da operação da organização e pela região geográfica em que está situada. Considera-se que, para organizações com mesmo ramo comercial, as variáveis não são necessariamente as mesmas utilizadas para predição (MWADULO, 2016; PROVOST; FAWCET, 2016).

**Figura 3 - Indicação das características que determinam o grupo ao qual a flor pertence.**



Fonte: Modificado de BARRETO (2017).

## 2. RESULTADO

Para a realização da tarefa de classificação, é necessário iniciar o Weka e abrir o arquivo ARFF contendo o conjunto de dados Iris, configurando a classificação do Weka de acordo com a Tabela 1. Este conjunto possui 150 registros de medidas de flores, já rotuladas com os grupos a que pertencem. Para que ocorra a classificação, parte dos registros terão seu rótulo removido (“não-classificados”), sendo após isto classificados por um algoritmo de árvore de decisão,

especificamente o algoritmo C4.5<sup>9</sup> (identificado como J48 no Weka), este pelo qual será treinado por outra parte dos registros (rotulados). Esta divisão da massa de registros é apresentada na Tabela 1.

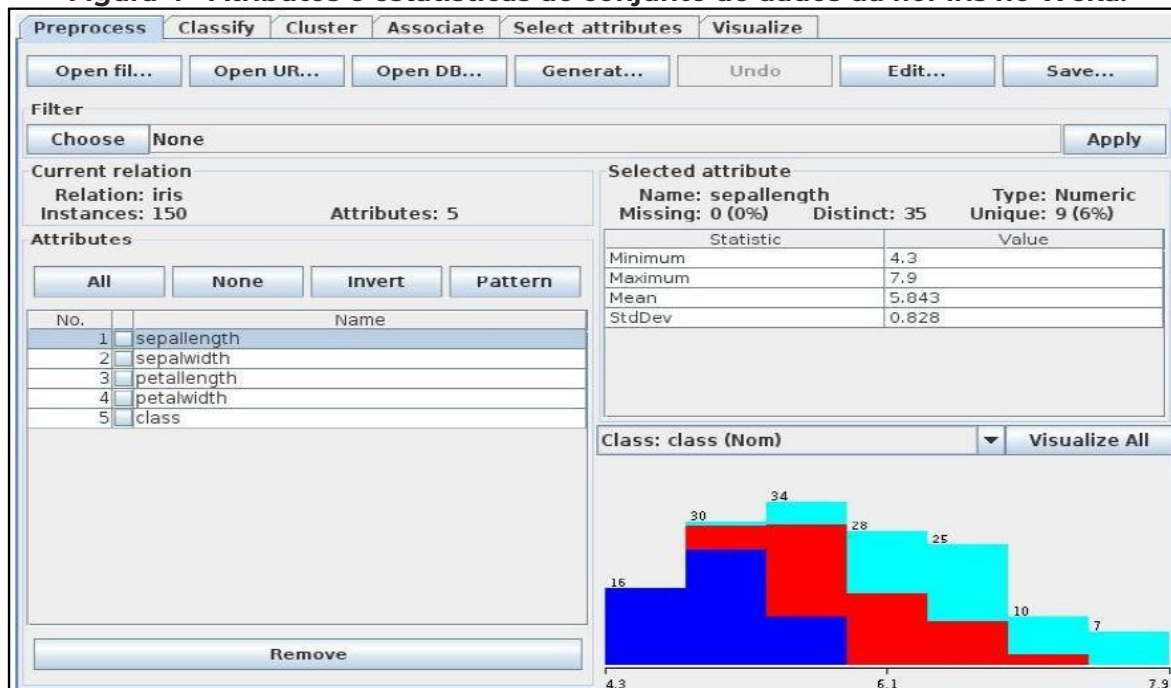
**Tabela 1 - Divisão do conjunto de dados de flores de Iris.**

	Divisão do conjunto de dados	%
Massa total de dados (150 registros)	Massa de dados para treinamento do modelo: 105 registros rotulados(classificados).	70%
	Massa de dados para teste do modelo: 45 registros não- rotulados (não-classificados).	30%

Fonte: Autores (2019).

Esta forma de classificação permite visualizar de maneira clara como o algoritmo classifica uma flor. Na Figura 4, é possível observar o conjunto de dados já aberto com todos os atributos, inclusive a classe. Além disso, é possível visualizar alguns dados estatísticos como o valor máximo e mínimo, média e desvio padrão.

**Figura 4 - Atributos e estatísticas do conjunto de dados da flor íris no Weka.**



Fonte: Autores (2022).

<sup>9</sup> C4.5 é um algoritmo utilizado para criar uma árvore de decisão e foi desenvolvido por Ross Quinlan. O algoritmo J48 surgiu da necessidade de recodificar o algoritmo C4.5, que, originalmente, é escrito na linguagem C, para a linguagem Java. (QUINLAN, 1993).



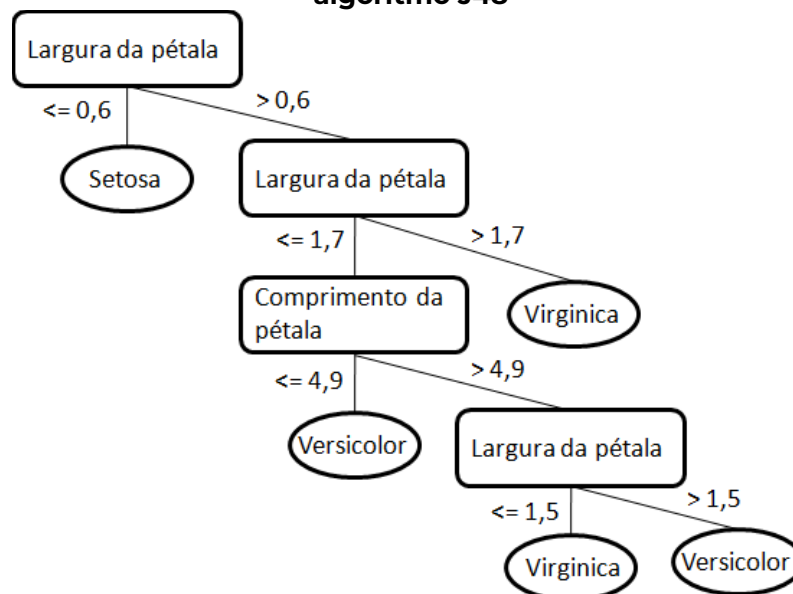
Após o carregamento do conjunto de dados, deve-se executar o algoritmo J48 na aba de classificadores (Classify), com o algoritmo J48 selecionado, utilizando para o treinamento 70% do conjunto de dados e a árvore de decisão utilizando a estrutura de representação do Weka.

Cinco regras através da árvore gerada são definidas. São estas:

- Largura pétala  $\leq 0.6$  cm  $\rightarrow$  Setosa;
- Largura pétala  $> 1.7$  cm  $\rightarrow$  Virginica;
- Largura pétala  $\leq 1.7$  cm e Comprimento Pétala  $\leq 4.9$  cm  $\rightarrow$  Versicolor;
- Comprimento pétala  $> 4.9$  cm e Largura Pétala  $\leq 1.5$  cm  $\rightarrow$  Virginica;
- Comprimento pétala  $> 4.9$  cm e Largura Pétala  $> 1.5$  cm  $\rightarrow$  Versicolor.

Além das regras, pode-se perceber que as medidas relativas as sépalas não foram utilizadas para a determinação da espécie da flor. Isto caracteriza que as medidas das sépalas foram irrelevantes na tarefa de classificação, conforme observado na árvore de decisão apresentada na Figura 5.

**Figura 5 - Gráfico da árvore de decisão gerado a partir do modelo gerado pelo algoritmo J48**



Fonte: Autores (2022)

A partir disto, a seguinte matriz de confusão, apresentada na Tabela 2, foi gerada após a predição.

**Tabela 2 - Matriz de confusão gerada a partir da predição de flores de Iris**

	Setosa (14)	Versicolor (18)	Virginica (13)
Setosa (14)	14	0	0
Versicolor (16)	0	16	0
Virginica (15)	0	2	13

Fonte: Autores (2022)

Observando a tabela anterior, percebe-se, com os 45 elementos testados, apenas duas ocorrências de erros (ao comparar a última linha com a última coluna), representando um percentual de acerto de 95%. Embora tal acurácia seja satisfatória, vale ressaltar que para situações em que isto não ocorra, o algoritmo é submetido a novos treinamentos.

Os dados são reavaliados ou são utilizados outros algoritmos. Ao analisar uma nova flor de íris, cujo grupo seja desconhecido, por exemplo: uma flor com largura da pétala > 1,7 cm, consegue-se identificar, por meio do uso do modelo criado pelo algoritmo (representado pela Figura 5), que ela pertence ao grupo Virginica.

Expondo estes resultados, referentes à flor de Iris, é possível compreender a situação de classificação de um cliente como bom ou mau pagador, antes de efetivar um empréstimo. Lembrando que, assim como a flor de Iris, algumas variáveis precisam ser identificadas como determinantes, para caracterizar o bom e o mau pagador. Estas variáveis compõem uma base de dados, das quais são submetidas a um processo semelhante ao apresentado até o momento, gerando uma árvore de decisão, conforme a Figura 5.

Abstraindo<sup>10</sup>, ao utilizar um conjunto de dados de clientes (e suas características discrepantes) e seus empréstimos (bem e malsucedidos) para a construção de um modelo (árvore de decisão, por exemplo), é possível, ao efetuar um novo empréstimo, classificar o cliente como um bom ou mau pagador, evitando um empréstimo de risco.

<sup>10</sup> No caso de um projeto real, que se utilize de técnicas de aprendizado de máquina, aplica-se uma metodologia de extração de conhecimento (como o CRISP, por exemplo) para definir a forma de construção do modelo, quais dados e características desses dados serão submetidas ao modelo e qual algoritmo será utilizado para este modelo.

### 3. CONSIDERAÇÕES FINAIS

Este trabalho teve como principal objetivo identificar e apresentar os recursos técnicos que auxiliam na tomada de decisão a partir da predição, levando o discurso de uma visão de “marketing” para uma ótica técnica, apresentando uma análise e extração do conhecimento a partir de dados. Para isto, se utilizou de duas situações: uma existente e consagrada na literatura, a partir da qual se tornou base prática e técnica para elucidar uma situação do mundo real. No âmbito técnico, um algoritmo de árvore de decisão (C4.5/J48) foi utilizado para comprovar a predição, desta forma cumprindo com o objetivo proposto inicialmente.

Espera-se que, com este texto, o leitor possa visualizar uma pequena parte do mundo que a ciência dos dados abrange, compreendendo os processos para aplicar uma predição simples, e entender a complexidade de uma análise preditiva envolvendo uma grande massa de dados com Big Data Analytics.

### REFERÊNCIAS

AMARAL, F. **Aprenda Mineração de Dados - Teoria e Prática**. Rio de Janeiro: Alta Books, 2016.

BARRETO, F. C. **Dicionário - A Compreensão de Cada Termo da Biologia**. Rio de Janeiro: Câmara Brasileira do Livro, 2017.

DICIO - Dicionário Online de Português. Disponível em: <<https://www.dicio.com.br>>. Acesso em: 30 jun. 2020.

FAYYAD, U.; GRINSTEIN, G. G.; WIERSE, A. **Information Visualization in Data mining and Knowledge Discovery**. San Diego: MK - Morgan Kaufman Publishers, 2002.

FISCHER, R. A. **The Use of Multiple Measurements in Taxonomic Problems**. 1936: [s.n.]. Disponível em: <[http://www.comp.tmu.ac.jp/morbier/R/Fisher-1936-Ann.\\_Eugen.pdf](http://www.comp.tmu.ac.jp/morbier/R/Fisher-1936-Ann._Eugen.pdf)>. Acesso em: 19 jul. 2020.

GARCIA, E. **Pesquisa Bibliográfica Versus Revisão Bibliográfica - Uma Discussão Necessária**. 2016: [s.n.].

GIACHETI, L. DE J. M. **José Reis: a ciência que fala**. São Paulo: Annablume, 2006. GIL, A. C. **Métodos e Técnicas de Pesquisa Social**. 6. ed. São Paulo: Atlas, 2008.

GOOGLE TRENDS. **Big data Analytics**, 2022. Disponível em: <<https://trends.google.com.br/trends/ explore?date=all&q=Big% 20data% 20analytics>>. Acesso em: 04 jul. 2022.

GRUPO ãNIMA EDUCAÇÃO. **Manual - Revisão Bibliográfica Sistemática Integrativa: pesquisa Baseada em Evidências**. 2014: [s.n.]. Disponível em: <<http://biblioteca.cofen.gov.br/wp-content/uploads/2019/06/manual-revisao-bibliografica-sistemica-integrativa.pdf>>. Acesso em: 19 jul. 2021.

GÜLTEKIN, B.; SAKAR, B. E. **Variable Importance Analysis in Default Prediction using Machine Learning Techniques**. 7th International Conference on Data Science, Technology and Applications -2018: [s.n.]. Disponível em: <<https://www.scitepress.org/papers/2018/68724/68724.pdf>>. Acesso em: 19 jul. 2021.

HOLANDA, A. B. DE. **Dicionário Aurélio da Língua Portuguesa**. Curitiba: Positivo, 2010.

MARCONI, M. DE A.; LAKATOS, E. M. **Fundamentos de metodologia científica**. São Paulo: Atlas, 2003.

MARTINO, L. M. S. **Métodos de Pesquisa em Comunicação - Projetos, Ideias e Práticas**. Petrópolis: Editora Vozes, 2018.

MWADULO, M. W. **A Review on Feature Selection Methods For Classification Tasks**. International Journal of Computer Applications Technology and Research - 2016: [s.n.]. Disponível em: <<https://pdfs.semanticscholar.org/0112/fbdf443158dbaf3226028c6bed68de5fb774.pdf>>. Acesso em: 20 jul. 2021.

PRIBERAM, D. **Dicionário da Língua Portuguesa**. Disponível em: <<https://dicionario.priberam.org>>. Acesso em: 7 mar. 2022.

PROVOST, F.; FAWCET, T. **Data Science para Negócios**. Rio de Janeiro: Alta Books, 2016. Que Conceito - Dicionário. Disponível em: <<https://queconceito.com.br/>>. Acesso em: 30 jun. 2021.

SESC SÃO PAULO/CEBRAP. **Métodos de pesquisa em Ciências Sociais: Bloco Qualitativo**. 2016.

SILVA, A. M. DA. **Metodologia da Pesquisa**. Fortaleza: EdUECE - Editora da Universidade Estadual do Ceará, 2015.