



UNIVERSIDADE FEDERAL DO MARANHÃO

Curso de Ciência da Computação

Phillipe Idivaldo Mendonça Silva

Benchmarking em classificação de dados: Redes neurais e árvores de decisão

São Luís - MA

2021

Phillipe Idivaldo Mendonça Silva

Benchmarking em classificação de dados: Redes neurais e árvores de decisão

Trabalho de conclusão de curso apresentada para Ciência da Computação da Universidade Federal do Maranhão como parte dos requisitos necessários para obtenção do grau de bacharel em Ciência da Computação.

Orientador: Prof. Dr. Ivo José da Cunha Serra

São Luís - MA

2021

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Diretoria Integrada de Bibliotecas/UFMA

Silva, Phillipe Idivaldo Mendonça.

Benchmarking em classificação de dados : Redes neurais e árvores de decisão / Phillipe Idivaldo Mendonça Silva. - 2021.

41 f.

Orientador(a): Ivo José da Cunha Serra.

Monografia (Graduação) - Curso de Ciência da Computação, Universidade Federal do Maranhão, São Luís - MA, 2021.

1. Árvores de decisão. 2. Benchmarking. 3. Classificação de dados. 4. Redes neurais. I. Serra, Ivo José da Cunha. II. Título.

Trabalho aprovado em São Luís - MA, 28 de setembro de 2021

Prof. Dr. Ivo José da Cunha Serra
Orientador
Universidade Federal do Maranhão

**Prof. Msc. Carlos Eduardo Portela
Serra de Castro**
Banca examinadora
Universidade Federal do Maranhão

Profa. Dra. Simara Vieira da Rocha
Banca examinadora
Universidade Federal do Maranhão

Agradecimentos

Bom em primeiro lugar agradecer a minha família, em especial uma grande mulher e batalhadora que é minha mãe, te amo.

À todos os professores de curso por seus ensinamentos, em especial para Anselmo, Simara, Vanessa, Greiciane, Geraldo, vocês são demais.

Aos meus amigos de curso que me possibilitaram chegar aqui, sempre incentivando e ajudando a cada período do curso.

Ao meu orientador e professor Ivo, por sua paciência e diligência para concluir mais esta etapa na minha vida.

Às minhas filhas Xena e Cheetara, por me alegrar e me ouvir todos mesmo sem entender nada, amo vocês!

Ao meu amigo Anderson, Deus na computação e titã na defesa dos anciãos 2. Iremos ganhar o the international ainda.

Ao meu amigo Vitu, por me carregar até onde deu no curso e por aturar meus rages durante os jogos, e sua capacidade de ensinar em três horas o que eu não aprendia em um período.

À todo grupo do PETCOMP, o melhor grupo e mais incrível de todos os tempos e tive o prazer de fazer parte, Lucas, Vitu, Bruno, Hugo, Dudu, Anderson, Alexandre, Artur, Ari, Rod, Pereirão, vou guardar vocês para sempre comigo.

Não sou uma pessoa de muitos amigos, mas posso garantir que nesta trajetória eu conheci excelentes pessoas, Renan, Marvin, Carol, Nélia, Sky, Ems, Amanda, Paulo, vocês são feras.

E por fim a Glacyane, que mesmo não estamos tão próximos como antes eu tenho certeza que não estaria aqui, obrigado por seus conselhos, incentivos e carinho.

E se eu esqueci alguém, gente por favor me perdoem estou ficando velho, amo todos vocês.

“From knowledge come skill.”
(Xin, the Ember Spirit, Defense of the Ancients 2)

Resumo

Com o avanço tecnológico e a dificuldade de lidar com grande volume de informações, surgiu a necessidade de a ciência criar novas possibilidades para ajudar tanto em automatizar, como em ofertar meios de lidar sejam elas estratégias, softwares, dispositivos, etc. Uma das ferramentas utilizadas para auxiliar nesta tarefa é o benchmarking. A princípio um termo não tão novo para o mundo corporativo, mas que conduzem a boas práticas sejam elas comparativas ou não, auxiliando o desempenho e buscando melhores estratégias para o crescimento de um determinado empreendimento. Suas diretrizes se popularizaram e colaboram para todo o âmbito científico, trazendo mais efetividade e confiabilidade para as pesquisas. Em algoritmos de classificação, auxiliam a viabilizar por exemplo qual a melhor estratégia aplicar, levando em consideração análises métricas e desempenho de cada algoritmo. Neste trabalho, foram analisados dois algoritmos classificadores através de um experimento realizado na ferramenta WEKA, utilizando uma abordagem de redes neurais, e uma outra perspectiva por árvores de decisão, afim de mostrar diferentes pontos de vista aplicados em uma mesma tarefa.

Palavras-chaves: Benchmarking, Redes neurais, Árvores de decisão, classificação de dados

Abstract

With technological advances and a difficult to deal with a large volume of information, the need for science to create new possibilities to help both in automating and in offering ways to deal, be they strategies, software, devices, etc., arose. One of the tools used to assist in this task is benchmarking. At first, a term not so new to the corporate world, but that lead to good practices, whether they are comparative or not, helping performance and seeking better strategies for the growth of a given enterprise. Its guidelines have become popularized and collaborate across the scientific field, bringing more effectiveness and reliability to research. In classification algorithms, they help to enable, for example, which is the best application strategy, taking into account metric analysis and performance of each algorithm. In this work, classifier algorithms were traversed through an experiment carried out in the WEKA tool, using a neural network approach, and another perspective by decision trees, in order to show different points of view given in the same task.

Keywords: Benchmarking, Neural networks, Decison trees, Data classification

Lista de ilustrações

Figura 1 – Gerações do benchmarking	16
Figura 2 – Tela inicial do software WEKA	21
Figura 3 – Guia Setup do software WEKA	22
Figura 4 – Guia Run do software WEKA	23
Figura 5 – Guia Analyse do software WEKA	24
Figura 6 – Arquivo iris.arff	26
Figura 7 – Exemplo de rede neural multicamada	27
Figura 8 – Acurácia dos classificadores J48 e MLP	29
Figura 9 – Recall e Precisão dos classificadores J48 e MLP	30
Figura 10 – F-score dos classificadores J48 e MLP	31
Figura 11 – TFP e TFN dos classificadores J48 e MLP	31
Figura 12 – Exemplo de Curvas ROC	32
Figura 13 – Área sob curva ROC dos classificadores J48 e MLP	32
Figura 14 – Árvore de decisão gerada da base de dados IRIS	33
Figura 15 – Árvore de decisão gerada da base de dados PIMA Indians	34
Figura 16 – Rede neural gerada da base de dados IRIS	34
Figura 17 – Rede neural gerada da base de dados IRIS	35

Lista de tabelas

Tabela 1 – Atributos da base de dados iris	25
Tabela 2 – Atributos da base de dados PIMA Indians	25
Tabela 3 – Matriz de confusão	28
Tabela 4 – Resultados gerais extraídos do experimento	35

Lista de abreviaturas e siglas

AUC	Area under the ROC curve
ARFF	Attribute-Relation File Format
CSV	Comma Separated Values
FN	Falsos Negativos
FP	Falsos Positivos
MLP	Multilayer Perceptron
TVP	Taxa de Verdadeiros Positivos
TVN	Taxa de Verdadeiros Negativos
TFP	Taxa de Falsos Positivos
TFN	Taxa de Falsos Negativos
VN	Verdadeiros Negativos
VP	Verdadeiros Positivos
WEKA	Waikato Environment for Knowledge Analysis
WEE	Weka Experiment Environment

Sumário

1	INTRODUÇÃO	12
1.1	Objetivo Geral	12
1.1.1	Objetivos Específicos	12
1.2	Organização do Trabalho	13
2	BENCHMARKING	14
2.1	Definições	14
2.2	Contexto histórico	14
2.3	Evolução do benchmarking	15
2.4	Tipos de benchmarking	17
2.5	Como ocorre o processo de Benchmarking	17
2.6	Questões complementares acerca do benchmarking	18
2.7	Benchmarking no âmbito computacional	18
3	BENCHMARKING COM ALGORITMOS DE REDES NEURAIS E ÁRVORES DECISÃO	20
3.1	Software WEKA	20
3.2	Ambiente WEKA	20
3.3	Conjunto de dados utilizado	24
3.4	Tipo de arquivos WEKA	25
3.5	Classificadores utilizados para o experimento	26
3.5.1	Algoritmo Multilayer Perceptron	26
3.5.2	Algoritmo J48	27
3.6	Resultados	28
3.6.1	Métricas de avaliação	28
3.6.1.1	Acurácia	29
3.6.1.2	Recall	29
3.6.1.3	Precisão	30
3.6.1.4	F-score	30
3.6.1.5	Taxa de Falsos Positivos e Taxa de Falsos Negativos	31
3.6.1.6	Curvas ROC	31
3.6.1.7	Árvore de decisão gerada pelo algoritmo J48	33
3.6.1.8	Rede neural gerada pelo algoritmo MLP	34
4	CONCLUSÃO	37

REFERÊNCIAS 39

1 Introdução

Desde os primórdios, nós seres humanos visamos criar ferramentas que nos auxiliem a lidar com variados tipos de tarefas, criando meios e estratégias para abordar cada tipo problema. O que antes era manual, e difícil de lidar pelo volume de informações, hoje está se tornando automatizado e buscando melhores resultados para reduzir a margem de erros. Com este avanço tecnológico, apresentar soluções mais eficientes e confiáveis requerem testes incessantes para todas as áreas científicas.

Um termo que se popularizou bastante nas últimas décadas, devido a necessidade de buscar meios de melhorar e comparar através de testes e que trouxe influências para todo o âmbito científico, foi o benchmarking. Esta abordagem, no contexto da computação, trouxe aplicabilidade e melhores técnicas para efetuar testes em sistemas, hardwares, softwares, dispositivos e etc. Segundo Berry et al. (1991), a técnica de benchmark é geralmente utilizada para medir a performance de alguns aspectos de sistemas computacionais de forma sistemática. O rigor nas medições visa o controle do ambiente no qual as avaliações são realizadas. Em outras palavras, avaliar por meio de testes, as análises de performance, e a eficiência geral de sistemas computacionais em um contexto específico.

Neste contexto, será apresentado por meio de um experimento, como dois algoritmos de classificação de dados se comportam, utilizando redes neurais e árvores de decisão, analisando suas performances aplicadas em duas bases de dados distintas. A ferramenta escolhida para realizar o experimento foi o WEKA, onde suas aplicações oferecem diversos mecanismos de avaliações em classificação de dados que serão melhor discutidas no Capítulo 3. O propósito central deste trabalho, é trazer os conceitos de benchmarking e sua aplicação no meio computacional, a fim de evidenciar sua aplicabilidade em análise e desempenho de algoritmos de classificação de dados.

1.1 Objetivo Geral

Verificar, por meio de um experimento de benchmarking, o desempenho dos classificadores de redes neurais e árvores de decisão, na tarefa de classificação de dados, analisando-os e comparando-os por meio de seus respectivos desempenhos.

1.1.1 Objetivos Específicos

- Apresentar os conceitos acerca do benchmarking, ressaltando sua aplicação no contexto computacional.

- Apresentar e discutir um experimento de benchmarking, utilizando algoritmos de redes neurais e árvores de decisão.
- Disponibilizar dados para o âmbito científico a partir de um experimento de benchmarking com relação aos algoritmos de redes neurais e árvores de decisão aplicados em duas bases distintas.

1.2 Organização do Trabalho

Este trabalho está organizado em 4 capítulos, sendo o primeiro contendo a introdução, objetivos gerais e específicos. No capítulo dois será abordado sobre os conceitos acerca do benchmarking, suas influencias e aplicações. O capítulo três será efetuado um experimento entre dois algoritmos classificadores de dados, mostrando também como utilizar a ferramenta WEKA para realizar o experimento. Ao final do capítulo três será apresentado os resultados com tabelas e gráficos das métricas utilizadas, geradas pelo experimento através do WEKA. Por fim, o capítulo quatro trazendo as considerações finais a respeito do trabalho.

2 Benchmarking

O presente capítulo tem como objetivo trazer as principais fundamentações acerca de benchmarking, apresentando seus variados conceitos, a fim de direcionar um melhor entendimento e trazer suas aplicações para o âmbito computacional, evidenciando suas vantagens para o uso dessa abordagem em algoritmos de classificação de dados.

2.1 Definições

A palavra benchmarking, muito utilizada no mundo dos negócios, tem sua origem derivada da palavra "benchmark" na língua inglesa e é utilizada como substantivo, que se traduzida literalmente para a língua portuguesa temos por "bancada alvo". No qual refere-se a métodos de comparação entre diferentes empresas para indicar qual o mais eficiente (COLLINS, 2021).

Para [Camp \(1993\)](#) benchmarking é um processo pró ativo e estruturado que conduz a mudanças, ao desempenho e a vantagem competitiva. Esse processo busca novas ideias para métodos, práticas e processos para então adotá-las e praticá-las. É a busca de melhores práticas na indústria que conduzirão ao desempenho superior e ao alcance da eficácia global.

Já para CIMA (1996, p.20), citado em ([MADEIRA, 1999](#)), o conceito de benchmarking pode ser dito como uma instauração de práticas, de recolha de dados, de objetivos e comparadores, com os quais é possível identificar determinados tipos de performance.

O objetivo do benchmarking é, sistematicamente, identificar e comparar o processo e o desempenho de um setor, entre seus competidores, em um ambiente que está constantemente mudando. ([HONG et al., 2012](#))

Observando os conceitos abordados, pode-se inferir uma concordância entre os autores em utilizar a técnica de benchmarking como mecanismo comparativo e também aperfeiçoar uma determinada tarefa ou sistema. Seguindo este conceito, aprimorar-se utilizando métodos comparativos em diversas áreas da ciência é de suma importância para obter-se uma eficácia e classificar de forma mais concisa determinado experimento e trazer maior legitimidade para aplicar, ou qual melhor tomada de decisão seguir.

2.2 Contexto histórico

Por volta da Segunda Grande Guerra, muitas empresas japonesas já trabalhavam com alguns conceitos de comparação para melhoria de seus produtos e uma melhor

inserção no mercado. Porém o termo surgiu décadas depois em 1970 com a empresa XEROX Corporation.

O caso da Xerox Corporation iniciou quando seu presidente, Charles Christ, leu um anúncio no jornal New York Times, em 1979, que um concorrente oferecia copiadoras muito parecidas com as fabricadas na Xerox pelo preço equivalente ao de seu custo de produção. A Xerox tinha desenvolvido a copiadora 914 e teve a sua participação de mercado de 80% reduzida para 20%. O valor das ações da Xerox e sua participação no mercado caíram enormemente e a empresa pediu concordata. Seu novo presidente, David Kearns, antecipava as mudanças afirmando: “nós precisamos mudar significativamente o jeito e o modo de conduzir o nosso negócio”. (ALBERTIN; KOHL; ELIAS, 2016)

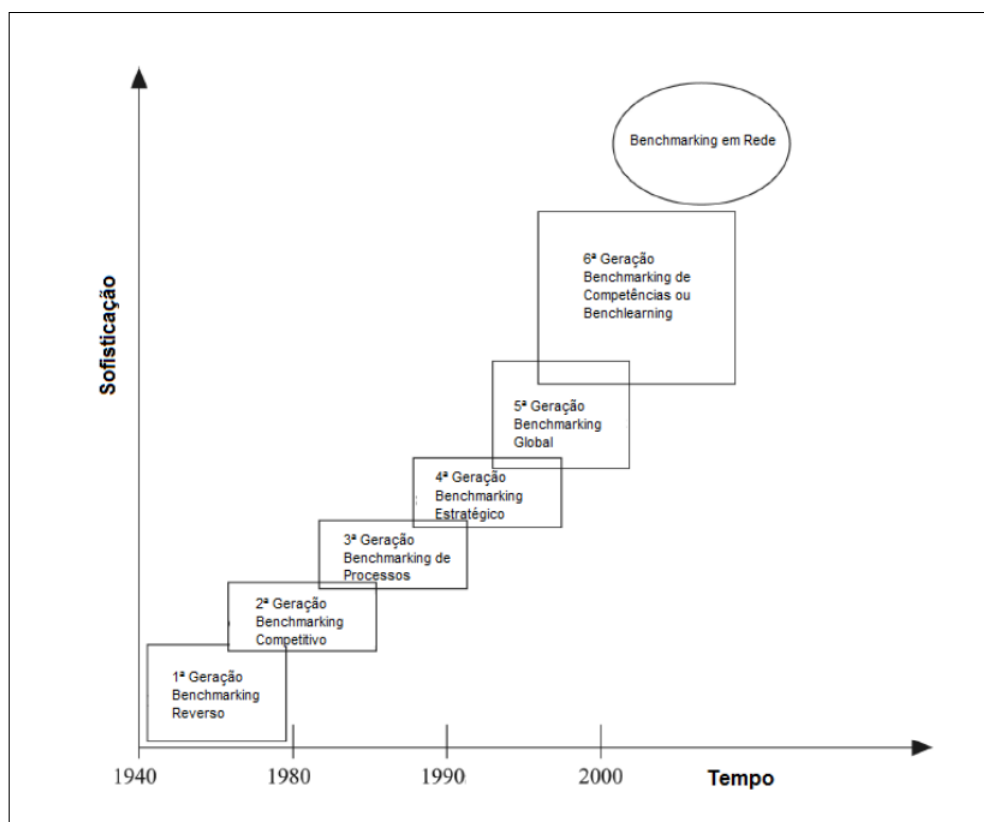
Para lidar com este problema, o presidente da empresa Xerox Corporation enviou uma equipe de funcionários ao Japão, para estudar e entender esta nova técnica de sucesso, para aumentar a qualidade e diminuir os custos de produção. Os resultados alcançados permitiram a recuperação da empresa e o desenvolvimento do programa “Leadership through Quality” (Liderança através da Qualidade), em 1983. Em 1989, a Xerox ganhou o prêmio “Malcolm Baldrige Quality Award”, divulgando mundialmente as práticas de benchmarking. (ALBERTIN; KOHL; ELIAS, 2016)

2.3 Evolução do benchmarking

Segundo Watson (1993), a concepção por benchmarking foi se modificando em gerações evolutivas como o próprio progresso empresarial. Seu desenvolvimento coincide ao padrão exemplar de “arte, transição para ciência” referente a transformação de uma nova doutrina de gerenciamento. Como podemos identificar na figura 1 a evolução de sua sofisticação a cada geração ao longo do tempo.

- **1ª geração** - Segundo Kyrö (2003), a primeira geração de benchmarking, intitulada de engenharia reversa, foi orientada ao produto, de forma que eram comparadas as características dos produtos, funcionalidades e desempenho.
- **2ª geração** - Benchmarking competitivo: Inicialmente elaborado pela empresa Xerox em 1976 a 1986 que percebeu suas despesas na produção estavam equivalentes ao custo de distribuição da concorrência. Modificou-se então o foco para a eficiência dos processos.
- **3ª geração** - Benchmarking de processos: Entre 1982 a 1988, esta fase marcada pelo entendimento de que valia mais a pena aprender com empresas de diferentes ramos, do que se comparar com as rivais de mesmo setor empresarial.

Figura 1 – Gerações do benchmarking



Fonte: Ahmed; Rafiq, (1998) e Kyrö (2003)

- **4ª geração** - Benchmarking estratégico: Caracterizada por expandir e adequar planos táticos vindas externamente de múltiplos associados.
- **5ª geração** - Benchmarking global: Após essa expansão de ideias vindas da geração anterior, utilizar-se de forma mais adequada para que as diferentes culturas entre nações tenham um impacto positivo para o mercado
- **6ª geração** - Kyrö (2003) acrescenta uma sexta geração chamada de benchlearning, traz como ponto central de que o princípio da mudança organizacional se obtém nas mudanças de ações e comportamentos dos indivíduos e das equipes, e tem o intuito de tornar-se uma organização de aprendizagem.
- **7ª geração** - Costa (2008) complementa uma sétima geração chamada “Networking benchmarking”. O autor destaca que esse tipo de benchmarking é denominado na literatura como benchmarking colaborativo. Ao invés de outras gerações, que têm enfoque na competição esse tem o enfoque no aprendizado baseado na colaboração em redes.

2.4 Tipos de benchmarking

Nesta seção são indicados os principais tipos de benchmarking encontrados na literatura. Esta tipologia auxilia a compreensão da abrangência e diversidade de potenciais aplicações de projetos de benchmarking. Os tipos de benchmarking são caracterizados pelas suas formas, vantagens e desvantagens de aplicações (KOHL, 2007). Há uma certa discordância entre os autores em relação aos tipos de benchmarking introduzidos por Spendolini (1992), e discutidos posteriormente por Camp (1993), Madeira (1999), Araújo (2000) e Dias (2008). Apesar das divergências, os principais citados por eles são:

- **Benchmarking interno** - método de comparar internamente, entre unidades operacionais ou funcionais, as práticas de negócio, dentro da mesma indústria. Partem do princípio que na empresa há unidades ou divisões mais eficientes que as outras, justificam esse esforço de aprendizagem;
- **Benchmarking competitivo** - envolve a comparação dos produtos, serviços e processos de trabalho da empresa com os conceitos diretos. Podem participar várias empresas do mesmo ramo, o que permite posicionar a sua eficiência em relação à do mercado.
- **Benchmarking funcional** - consiste em identificar as melhores práticas de qualquer tipo de organização, que tem uma reputação de excelência na área funcional sujeita ao benchmarking (também conhecido como benchmarking operacional ou genérico).
- **Benchmarking estratégico** - é um tipo de benchmarking competitivo cujo principal objetivo é estabelecer uma determinada ação estratégica ou uma mudança organizacional.

2.5 Como ocorre o processo de Benchmarking

De acordo com (WATSON, 1993), o benchmarking é baseado em quatro passos, que se refere ao ciclo de Edwards Deming (planejar – fazer – controlar – agir):

- **Planejamento do projeto:** Resume-se a dois questionamentos principais, qual será o objeto de benchmarking? E quem devemos analisar?
- **Recolha de dados:** Analisar as informações de domínio público, conjuntamente examinar os dados diretamente em contato com as empresas de ascensão no mercado, através de questionários, visitas ou entrevistas.
- **Análise de dados:** Trata-se da análise das informações recolhidas seguindo dois pontos de vista; primeiro avaliando por meio das diferenças de desempenho; segundo identificando os pontos fortes para os bons resultados.

- **Adaptação e melhoria:** Esta é a fase final, que conduz a ação e o empenho para mudar o modo de operar, a fim de melhorar o desempenho de acordo com as conclusões das fases anteriores do ciclo.

Como observado, o processo de benchmarking para ser bem executado, muitos desses quesitos citados devem ser trabalhados, para melhor atender as partes envolvidas durante todo o processo, e assim gerar a todos resultados satisfatórios a longo prazo.

2.6 Questões complementares acerca do benchmarking

Alguns quesitos importantes que vale a pena ressaltar, antes, durante e após a implantação do processo de benchmarking. Segundo Vaziri (1993), citado em (OLIVEIRA et al 2017) enfatiza os questionamentos a serem realizados antes da implementação:

- existe interesse da empresa em torno dos clientes, funcionários e melhoria contínua dos processos?
- é o benchmarking a estratégia de qualidade certa para sua organização?
- o que você deve comparar?
- o que você deve medir?
- quem deve avaliar o desempenho?
- como você pode coletar dados de modo eficiente?
- qual é a melhor maneira de implementar o que você aprendeu com a iniciativa de benchmarking

Além das soluções às perfeitas acima, a coleta de dados por parte da equipe de benchmarking deve seguir um método estruturado. As métricas devem estar bem especificadas em termos de unidades e intervalos de coleta e análise. O procedimento de obtenção dos dados deve ser testado anteriormente [Freytag e Hollensen \(2001\)](#).

2.7 Benchmarking no âmbito computacional

Até então, muito se foi discutido a respeito de processo de benchmarking em empresas de modo geral, porém com o passar do tempo, as aplicações de seus conceitos se tornaram abrangentes em diversas áreas da ciência.

No âmbito da ciência da computação destacamos aqui alguns trabalhos, como podemos exemplificar o trabalho de [Ciferri et al. \(1995\)](#), que tinha por objetivo propor

um conjunto de transações primitivas que representem a carga de trabalho de um benchmark voltado à análise de desempenho de sistemas de informações geográficas, além de caracterizar aspectos relacionados aos dados.

Segundo Ciferri et al. (1995) citado em (QUEIROZ et al., 2013) e (SANTOS, 2018) para efetuar a análise de desempenho em um sistema, existem dois modelos que podem ser seguidos, que são: modelo analítico e modelo de simulação ou experimental. As técnicas que integram o modelo experimental, são: monitoração e benchmarking.

A técnica de monitoração não possui seguimentos padrões. Esta abordagem possui uma técnica de avaliação própria, o que dificulta o uso de diferentes ferramentas, e obter resultados de forma justa. De uma maneira diferente, o benchmarking usa técnicas padronizadas para testar em diferentes modelos, garantindo uma comparação dos resultados de forma precisa.

O trabalho de SANTOS (2018), apresentou uma abordagem de benchmarking em algoritmos de classificação, avaliando desempenhos por métricas de avaliação definidas em um cenário de data mining, ou seja avaliar uma grande base de dados e trazer padrões de comportamento de um contexto específico.

Em particular na área de ciência da computação, a análise de desempenho feita pelo benchmarking é usada em diferentes segmentos. De acordo com Queiroz et al. (2013) para todos esses campos apurados, os principais resultados que se deseja obter são:

- Avaliar a capacidade máxima do sistema;
- Comparar tecnologias diversas;
- Avaliar a viabilidade de um sistema aplicado em um contexto específico;
- Mensurar a relação custo benefício.

Hoje em dia, em ciência da computação, as aplicações de benchmarking se estendem diversas subáreas. Podemos citar como exemplo suas aplicações e análises de desempenho em: sistemas operacionais, computação paralela, mineração de dados, banco de dados, redes de computadores etc. Para Gray (1992) e citado em Queiroz et al. (2013), uma das aplicações mais comumente utilizada, é o benchmark em banco de dados, onde dada a base de dados é definido um grupo de instruções, em seguida, estas instruções são executadas em sistemas de avaliação, para extrair os resultados de desempenho.

Em vista dos argumentos destacados pelos autores, podemos destacar a importância e versatilidade de aplicação do benchmarking em várias áreas do conhecimento, e trazer seus conceitos em diferentes pontos no meio computacional, seja na comparação de sistemas, testes de algoritmos ou softwares, que por consequência nos auxiliam a ampliar o campo de visão para solucionar uma determinada problemática em diferentes contextos.

3 Benchmarking com algoritmos de redes neurais e árvores decisão

Neste presente capítulo será elaborado um experimento, que consiste na aplicação de um processo de benchmarking sobre dois algoritmos de classificação, que serão avaliados por meio de seus respectivos desempenhos: o J48 que faz o uso de árvores de decisão; e o Multilayer Perceptron (MLP), que se baseia em redes neurais multicamadas. Esta avaliação se dará por métricas recomendadas em experimentos em classificação de dados e melhor abordado na seção 3.6.1. Para a execução deste experimento será utilizada a ferramenta WEKA, e particularmente sua aplicação WEKA Experiment Environment(WEE).

3.1 Software WEKA

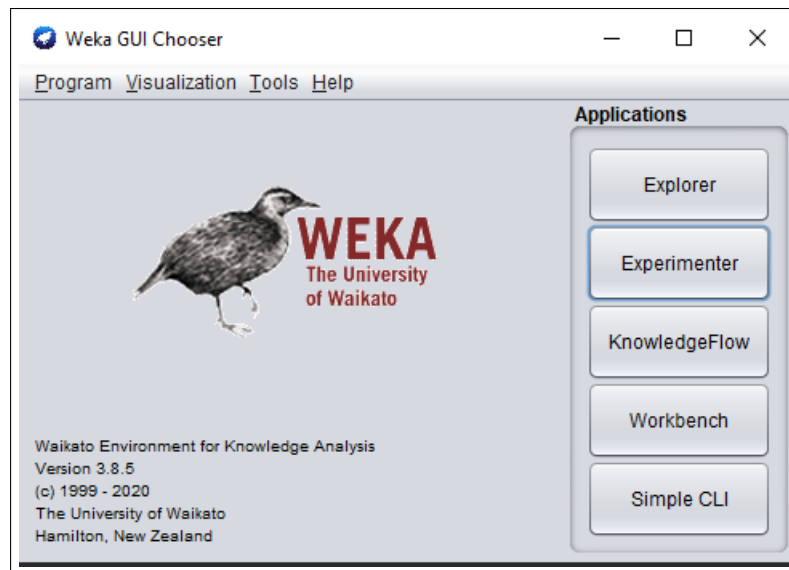
WEKA é um software de código aberto emitido sob a GNU General Public License, desenvolvido pela Universidade de Waikato na Nova Zelândia em 1993 sendo um acervo de algoritmos de aprendizado de máquina para realizar tarefas de mineração de dados. Nele contém diversas ferramentas para preparação de dados, classificação, regressão, agrupamento, mineração de regras de associação e visualização. Com uma interface de fácil entendimento e ambientação com o usuário, além de oferecer diferentes meios de aprendizados e cursos específicos para sua utilização fornecidas em seu próprio site.

Neste trabalho os experimentos realizados foram feitos na versão 3.8.5 do WEKA, seguindo as orientações oferecidas pela documentação do site “Data Mining: Practical Machine Learning Tools and Techniques”, por sua quarta edição (WITTEN; FRANK; MARK,).

3.2 Ambiente WEKA

Como dito anteriormente no Capítulo 2, o processo de benchmarking pode ser uma recolha de dados e comparações feitas para identificar variados tipos de performance. O Weka possui várias funcionalidades específicas, dentre eles podemos citar o Explorer, Experimenter, KnowledgeFlow, Workbench, Simple CLI. Na figura 2 podemos observar estas guias e suas respectivas funções.

Figura 2 – Tela inicial do software WEKA



Fonte: elaborada pelo autor

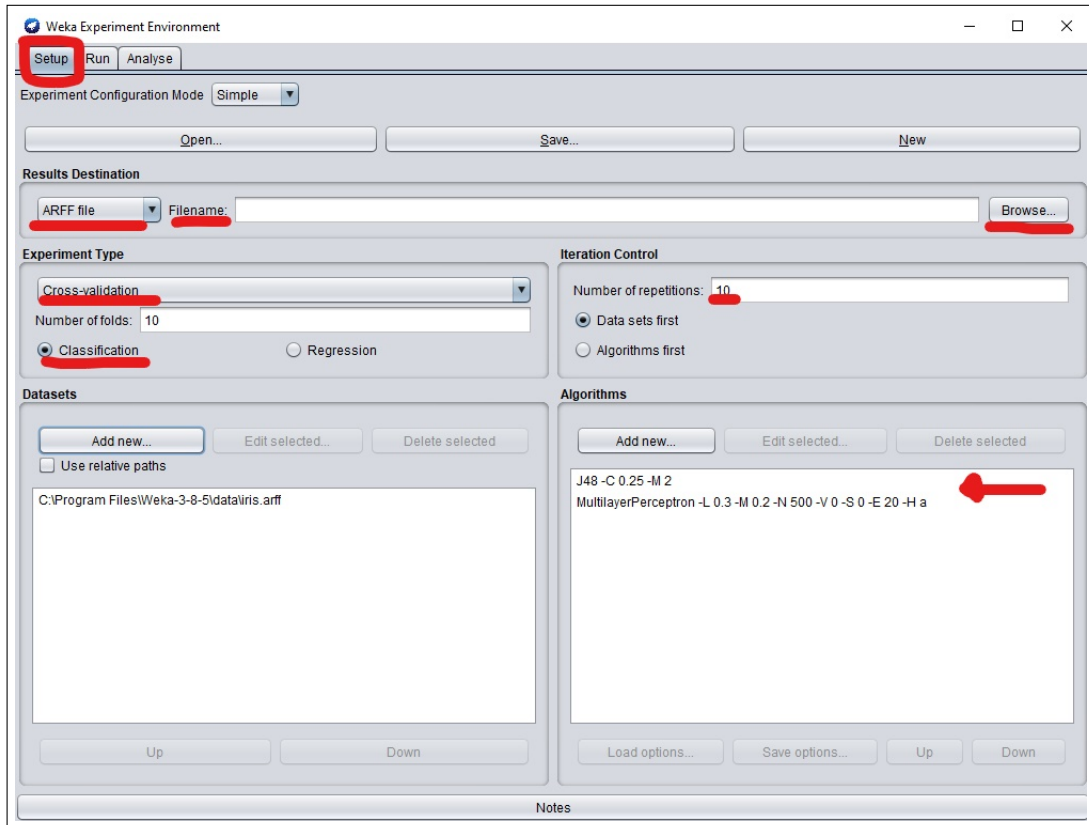
- **Explorer:** Esta aplicação é utilizada para testes individuais do acervo oferecido pelo WEKA, podendo testar bases, aplicar algoritmos, classificar dados e etc.
- **Experimenter:** Já nesta aplicação, seu uso é mais indicado para testar múltiplos algoritmos de forma simultânea, verificar análises e desempenhos para compará-los de forma mais eficiente através do benchmarking oferecido pela ferramenta WEKA.
- **KnowledgeFlow:** Esta funcionalidade permite a criação de fluxos de dados por meio de uma interface. Utilizando filtros, classificadores criando assim uma sequência para processamento.
- **Workbench:** Nessa opção é a junção de todas as aplicações.
- **Simple CLI:** Nesta aplicação permite a criação de scripts para inserir em testes, complementar algoritmos e obter melhores resultados.

Para executar este experimento, será utilizada uma das aplicações WEKA, o Experimenter. Nesta guia são feitas análises e comparações com diversos algoritmos presentes em sua biblioteca. Será empregado em uma base de dados específica concedido pelo software WEKA (citada na seção 3.3). é importante destacar também as três funcionalidades principais desta aplicação: Setup, Run e Analyze.

- **Setup:** Observa-se na figura 3 a guia de Setup, onde informa-se os parâmetros para o experimento. Pode-se carregar um experimento já realizado, assim como criar um novo, selecionando a opção desejada. Quando é criado um novo experimento,

temos a opção de salvar os resultados em três tipos de arquivos: CSV, ARFF, JDBC. Informa-se também a base de dados, e os algoritmos a serem aplicados.

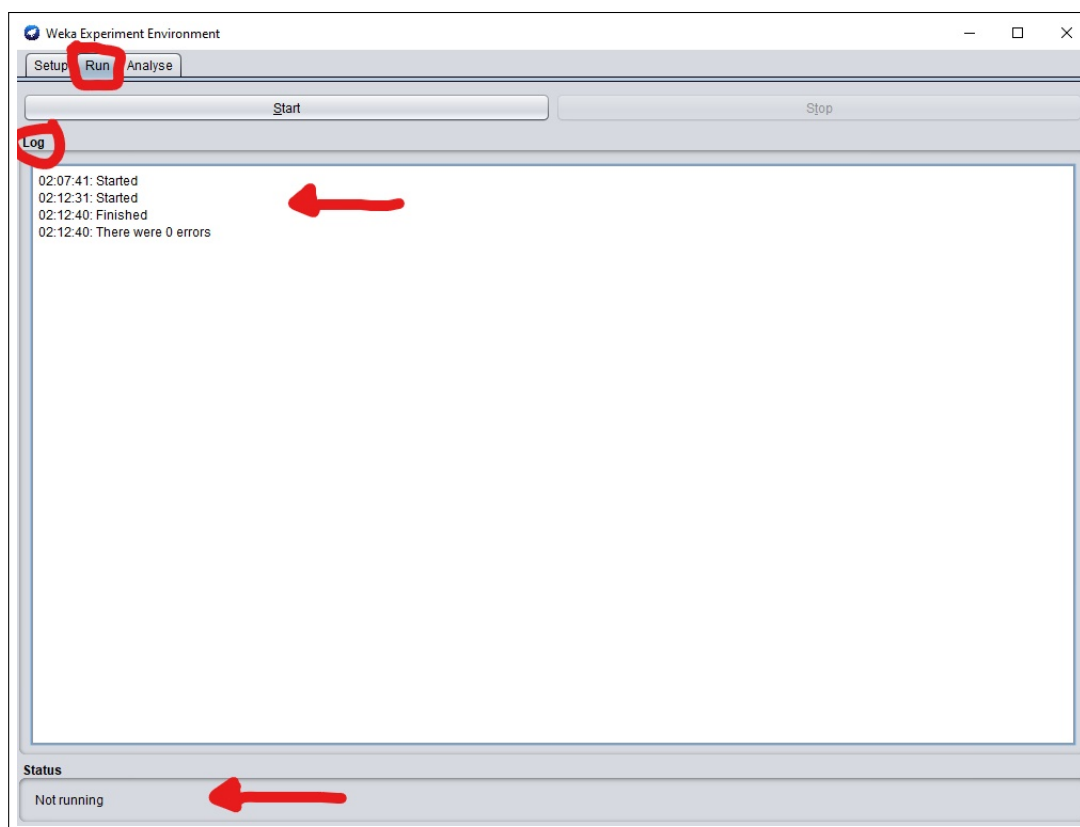
Figura 3 – Guia Setup do software WEKA



Fonte: elaborado pelo autor

- **Run:** Nesta guia podemos visualizar pela figura 4, a execução do experimento. Nela são mostrados os logs de execução, com informações de início e término do experimento, possíveis mensagens de erro, parar e retomar o experimento e o seu progresso de execução.

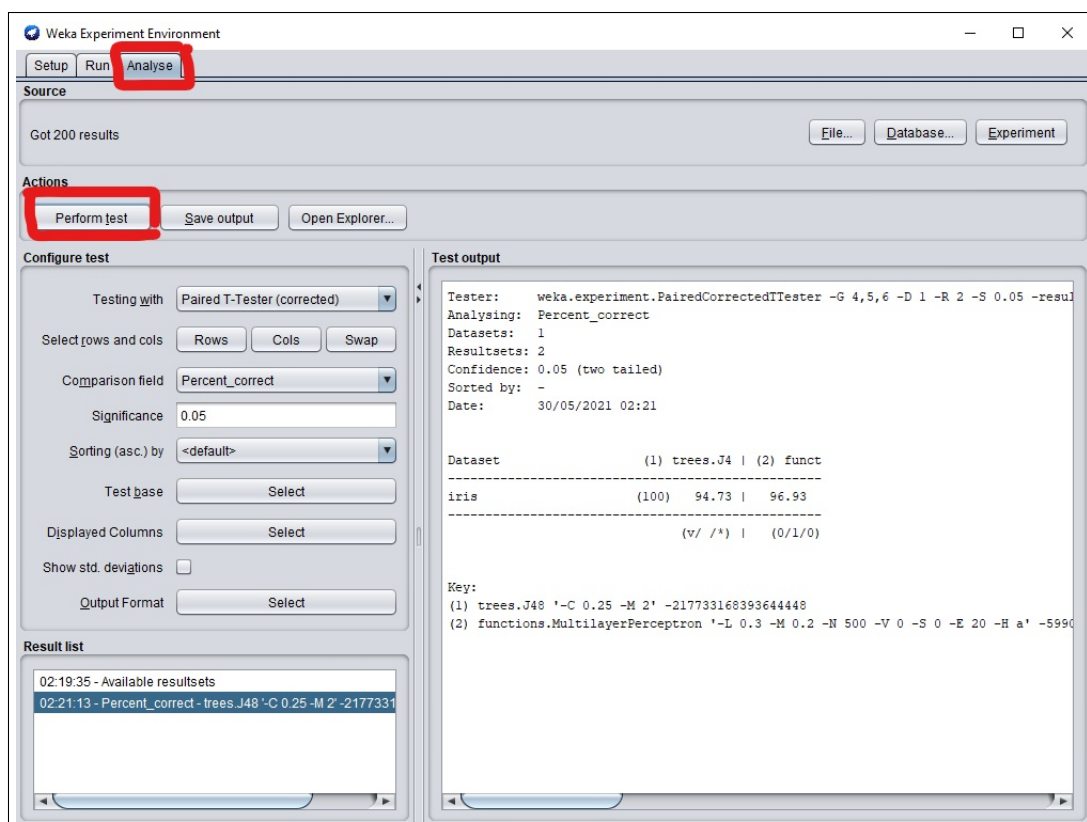
Figura 4 – Guia Run do software WEKA



Fonte: elaborado pelo autor

- **Analyze:** Após passar pela guia Run, execução do experimento, passamos para guia de análise indicada pela figura 5. Nesta guia mostra os resultados de análise dos algoritmos aplicados durante o experimento, onde também podem ser feitos pequenos testes comparativos com os dados obtidos, e assim observar por sua tela de saída como estes dados se comportam, gerando soluções, assim como aprender e manipular melhor os valores obtidos de análise.

Figura 5 – Guia Analyse do software WEKA



Fonte: elaborado pelo autor

3.3 Conjunto de dados utilizado

Para realizar os experimentos foram empregadas duas bases de dados fornecidas pelo próprio software WEKA. A primeira base utilizada foi IRIS (Iris siberica), que descreve um tipo de flor caracterizado por sua diversidade de cores e tipos. O conjunto de dados IRIS contém quatro características (comprimento e largura das sépalas e pétalas) de 150 amostras de três espécies de IRIS (IRIS setosa, IRIS virginica e IRIS versicolor). Com o uso dessas medidas, foram utilizadas para criar um modelo discriminante linear para classificar as espécies. O conjunto de dados é frequentemente usado em mineração de dados, classificação e exemplos de agrupamento e para testar algoritmos. Na tabela 1 são mostrados os atributos, a descrição e o tipo de dado da base IRIS.

A segunda base de dados escolhida foi a PIMA Indians (SIGILLITO, 1990). Essa base pertence originalmente ao Instituto Nacional de Diabetes e Doenças Digestivas e Renais, contém informações de 768 mulheres de uma população próxima a Phoenix, Arizona, EUA. O objetivo do conjunto de dados é prever diagnosticamente se um paciente terá diabetes ou não, com base nos dados clínicos obtidos dos pacientes e incluídos no conjunto de dados. Em particular, todos os pacientes aqui são mulheres com pelo menos 21 anos de herança indígena PIMA. Composta por 768 instancias, oito atributos e uma

Tabela 1 – Atributos da base de dados iris

Atributos	Descrição	Tipo de dados
sepalength	Comprimento das sépalas	Numérico
sepalwidth	Altura das sépalas	Numérico
petallength	Comprimento das pétalas	Numérico
petalwidth	Altura das pétalas	Numérico
class	Atributo nominal de saída	Iris-setosa, Iris-versicolor, Iris-virginica

Fonte: elaborado pelo autor

classe de saída, conforme mostra a tabela 2.

Tabela 2 – Atributos da base de dados PIMA Indians

Atributos	Descrição	Tipo de dados
preg	Quantidade de vezes que a paciente já esteve grávida	Numérico
plas	Concentração plasmática de glicose de duas horas após feito um teste oral de tolerância a glicose.	Numérico
pres	Pressão sanguínea diastólica(mm Hg).	Numérico
skin	Espessura da pele na região do tríceps (mm).	Numérico
insu	Insulina medida com 2 horas (mu U/ml).	Numérico
mass	Índice de massa corpórea(peso/kg , altura/m)	Numérico
pedi	Probabilidade de desenvolver a doença baseado no histórico familiar	Numérico
age	Idade	Numérico
class	Classe (0 - não apresenta diabetes; 1 apresenta diabetes)	Numérico

Fonte: elaborado pelo autor

3.4 Tipo de arquivos WEKA

Weka nos permite operar com diferentes fontes de dados, criando assim uma viabilidade para importar arquivos de diferentes formatos além do padrão disponibilizado pelo software, o ARFF (Attribute-Relation File Format/Arquivo no Formato Atributo Relação). Weka especifica este tipo de arquivo em um exemplo simples, separando em duas seções, cabeçalho e dados. Na parte cabeçalho, onde é informada o nome do conjunto de dados, é também feita a especificação dos atributos, seus tipos de dados, incluindo a classe. Observe o exemplo a seguir ilustrado pela figura 6 em formato ARFF do arquivo iris.

Figura 6 – Arquivo iris.arff

```
64 @RELATION iris
65
66 @ATTRIBUTE sepallength REAL
67 @ATTRIBUTE sepalwidth REAL
68 @ATTRIBUTE petallength REAL
69 @ATTRIBUTE petalwidth REAL
70 @ATTRIBUTE class {Iris-setosa,Iris-versicolor,Iris-virginica}
71
72 @DATA
73 5.1,3.5,1.4,0.2,Iris-setosa
74 4.9,3.0,1.4,0.2,Iris-setosa
75 4.7,3.2,1.3,0.2,Iris-setosa
76 4.6,3.1,1.5,0.2,Iris-setosa
77 5.0,3.6,1.4,0.2,Iris-setosa
78 5.4,3.9,1.7,0.4,Iris-setosa
79 4.6,3.4,1.4,0.3,Iris-setosa
80 5.0,3.4,1.5,0.2,Iris-setosa
81 4.4,2.9,1.4,0.2,Iris-setosa
82 4.9,3.1,1.5,0.1,Iris-setosa
83 5.4,3.7,1.5,0.2,Iris-setosa
84 4.8,3.4,1.6,0.2,Iris-setosa
85 4.8,3.0,1.4,0.1,Iris-setosa
```

Fonte: elaborado pelo autor.

O arquivo começa com a instrução **@relation**, que determina qual o nome de nossa base de dados. As seguintes instruções, **@attribute** que indica os atributos e seus respectivos valores que podem ser do tipo numeric, nominal, string e date. e por último a instrução **@data**, onde informa os dados no arquivo separados por vírgula.

3.5 Classificadores utilizados para o experimento

O Weka disponibiliza em sua biblioteca diversos algoritmos para classificação e efetuar os mais variados testes e análise. Para o desenvolvimento deste trabalho e aplicação nos experimentos, foi escolhido os algoritmos J48 de árvores de decisão e o algoritmo Multilayer Perceptron baseado em redes neurais. Para por em evidência estes diferentes tipos de algoritmos e verificar seus comportamentos e desempenhos para as bases aplicadas.

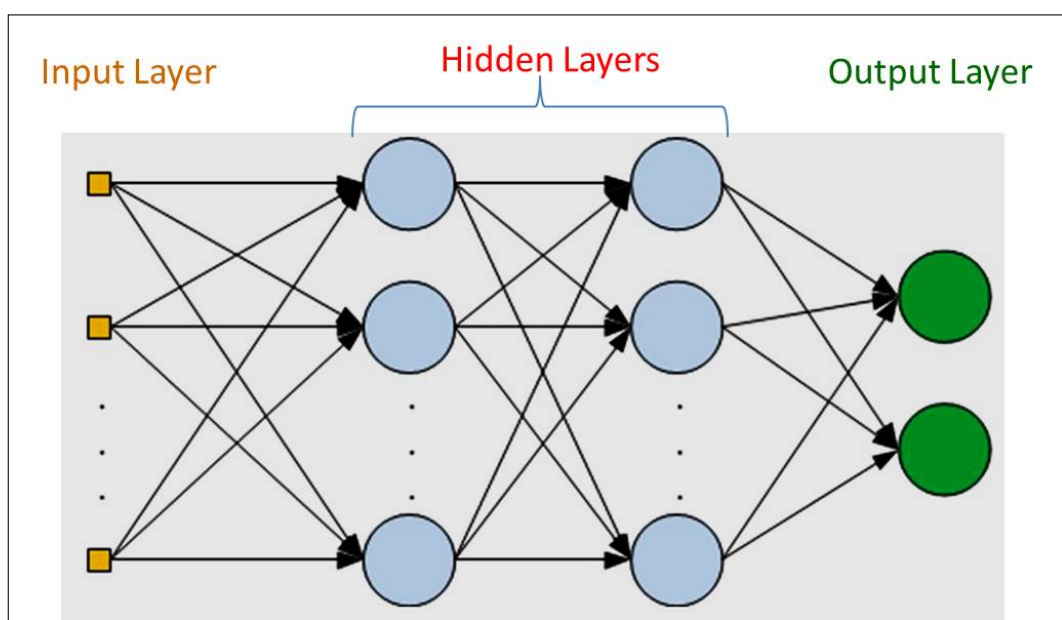
3.5.1 Algoritmo Multilayer Perceptron

Redes neurais Multilayer Perceptron consistem em unidades organizadas em camadas. Cada camada é composta de nós e na rede totalmente conectada considerada aqui, cada nó se conecta a todos os nós nas camadas subsequentes. Cada MLP é composto por um mínimo de três camadas consistindo em uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. A camada de entrada distribui as entradas para as camadas subsequentes. Os nós de entrada têm funções de ativação. Cada nó de unidade oculto e cada nó de saída têm limites associados a eles, além dos pesos. Os nós de unidade ocultos têm funções de ativação não lineares e as saídas têm funções de ativação

linear. Portanto, cada sinal que alimenta o nó em uma camada subsequente tem a entrada original multiplicada por um peso com um limite adicionado e, em seguida, é passado por uma função de ativação que pode ser linear ou não linear (unidades ocultas) (NAZZAL; EL-EMARY; NAJIM, 2008). Podemos observar no exemplo da figura 7 uma rede neural multicamadas genérica:

- Camada de entrada(Input Layer) (X).
- Camada(s) oculta(s)(Hidden Layers) .
- Conjunto de saída(Output Layer) (Y).

Figura 7 – Exemplo de rede neural multicamada



Fonte: Spinelli(2019),retirada de "Reconhecimento de caracteres escritos a mão utilizando Multilayer Perceptron (MLP)

3.5.2 Algoritmo J48

Árvores de decisão constituem um método de classificação baseado na análise dos valores de atributos com finalidade de representar de forma simples e eficiente o conhecimento obtido a partir de um conjunto de amostras de entrada. Normalmente são usadas em práticas de mineração de dados, ou seja, atividades de extração de conhecimentos, previamente desconhecidos, de uma base de dado. Em uma árvore de decisão, os nós internos correspondem aos testes realizados sob os valores dos atributos, nos arcos são encontrados os possíveis valores de saída para determinado teste e as folhas mostram os resultados da classificação das amostras. O algoritmo J48 gera árvores de decisão escolhendo o atributo mais apropriado para cada teste ou situação. As árvores são geradas do topo para base e cada nó analisa a relevância do atributo de forma individual.(BONINI,

2016) Este algoritmo é caracterizado pela evolução de seus antecessores os quais são ID3, C4.5, C5.0, onde os mesmos são implementados na linguagem C, e o J48 tem a mesma lógica só que este é implementado na linguagem JAVA. Uma das vantagens da aplicação deste algoritmo na tomada de decisão é que mesmo se mostra adequado para os procedimentos, envolvendo as variáveis qualitativas contínuas e discretas presentes nas bases de dados, e o mesmo permite a construção de árvores de decisão que classifica e apresenta em suas ramificações os atributos de maior relevância.(TELOKEN et al., 2016)

3.6 Resultados

Para a exposição dos dados estatísticos com as bases utilizadas, foram analisadas pelo WEE 786 instancias do PIMA Indians e 150 instancias da base IRIS, fazendo uma validação cruzada seguindo as orientações padrões do software WEKA. Estes resultados obtidos foram salvos em um arquivo CSV para praticidade na demonstração e produção dos respectivos gráficos de cada métrica de avaliação aplicada no processo de classificação de dados.

3.6.1 Métricas de avaliação

Para o processo de validação e mostrar se um algoritmo é eficiente ou não, é necessário analisar as métricas aplicadas durante o experimento de classificação de dados. Para introduzir melhor esses indicadores, é importante ressaltar os conceitos da matriz de confusão. Nesta matriz cada linha representa uma classe, que também é chamada de classe real e cada coluna uma classe prevista. Dela são extraídos quatro conceitos. Sendo eles: verdadeiros positivos (VP), Falsos Positivos (FP), Verdadeiros Negativos (VN) e Falsos Negativos (FN.).

Tabela 3 – Matriz de confusão

		Classe predita	
		1	0
Classe real	1	Verdadeiro Positivo(VP)	Falso Negativo (FN)
	0	Falso Positivo(FP)	Verdadeiro Negativo (VN)

Fonte: elaborado pelo autor

Com base nessas informações da tabela 3 pode-se definir:

- **Verdadeiro positivo (VP):** Conjunto de dados da classe real buscada foram corretamente previstos.
- **Falso positivo (FP):** conjunto de dados da classe real, a classe que está sendo buscada foi prevista incorretamente.

- **Verdadeiro Negativo (VN):** conjunto de dados da classe real, a classe que não está sendo buscada, foi prevista corretamente.
- **Falso negativo (FN):** conjunto de dados da classe real, a classe que não está sendo prevista, foi prevista incorretamente.

A partir dos conceitos tirados da matriz de confusão, podemos extrair as fórmulas das principais métricas de avaliação utilizadas em análise de desempenho de algoritmos de classificação. Segundo (SOKOLOVA; LAPALME, 2009), as medidas de mais relevância de desempenho a serem avaliadas são: precisão (precision), exatidão (accuracy), taxa de verdadeiros positivos (TVP), taxa de verdadeiros negativos (TVN), taxa de falsos positivos (TFP), taxa de falsos negativos (TFN), curvas ROC e F-score.

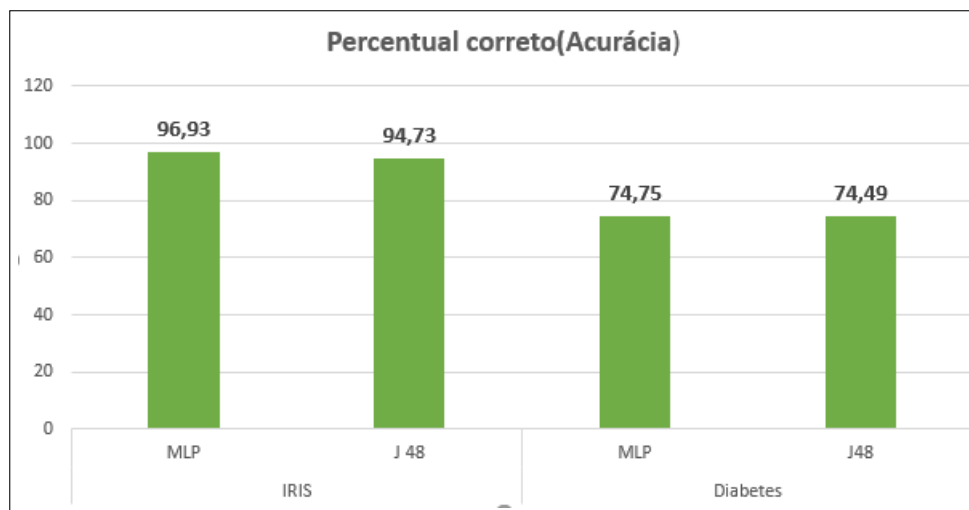
3.6.1.1 Acurácia

Número de acertos dividido pelo número de previsões, indicador importante e muito utilizado como parâmetro inicial em classificadores, e sua fórmula é dada por:

$$\text{Acurácia} = \frac{VP + VN}{VP + FP + VN + FN} = \frac{\text{previsões corretas}}{\text{todas as previsões}}$$

A figura 8 mostra os dados extraídos do experimento.

Figura 8 – Acurácia dos classificadores J48 e MLP



Fonte: elaborado pelo autor.

3.6.1.2 Recall

Outra métrica de avaliação utilizada, e que serve para identificar taxa de verdadeiros positivos e mostrar quão bom o modelo reagiu na previsão de positivos. Sua fórmula é dada por:

$$\text{Recall} = \frac{VP}{VP + FN}$$

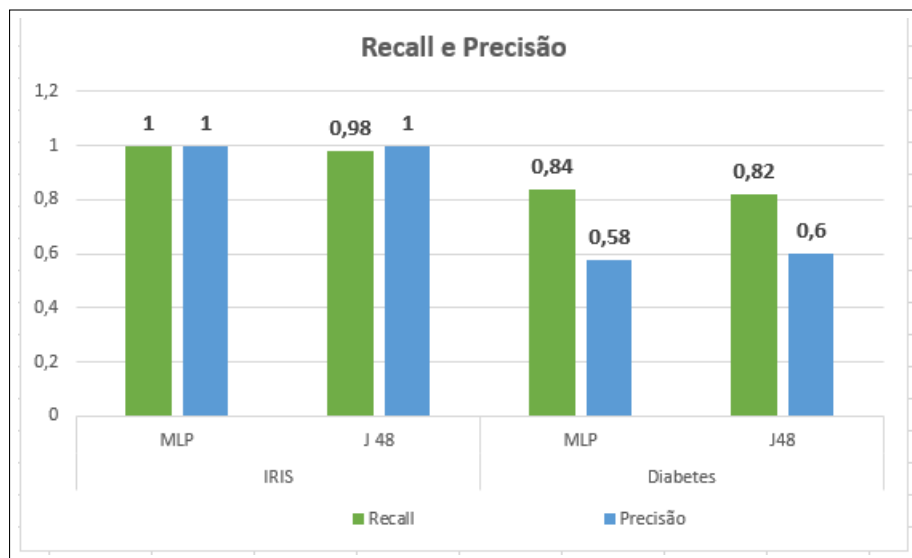
3.6.1.3 Precisão

Uma métrica também bastante comum para analisar taxa de verdadeiros negativos em classificadores e orientar quão bom o algoritmo atuou de maneira geral. Sua fórmula é dada por :

$$\mathbf{Precisão} = \frac{VP}{VP + FP}$$

A figura 9 mostra os dados de recall e precisão retirados do experimento.

Figura 9 – Recall e Precisão dos classificadores J48 e MLP



Fonte: elaborado pelo autor.

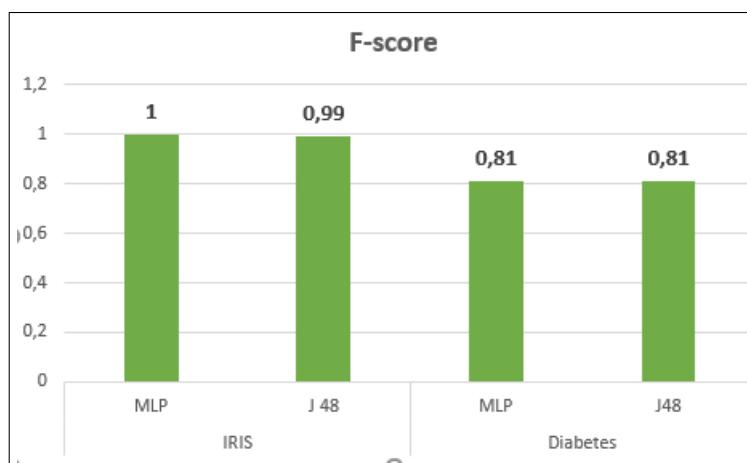
3.6.1.4 F-score

Outra métrica muito utilizada é a f-score. Tem como finalidade gerar um balanço por meio de uma média harmônica ponderada entre o recall e a precisão. E sua fórmula é dada por:

$$\mathbf{F-score} = 2 * \frac{\text{precisão} * \text{recall}}{\text{precisão} + \text{recall}}$$

A figura 10 apresenta os dados de f-score realizados durante o experimento:

Figura 10 – F-score dos classificadores J48 e MLP



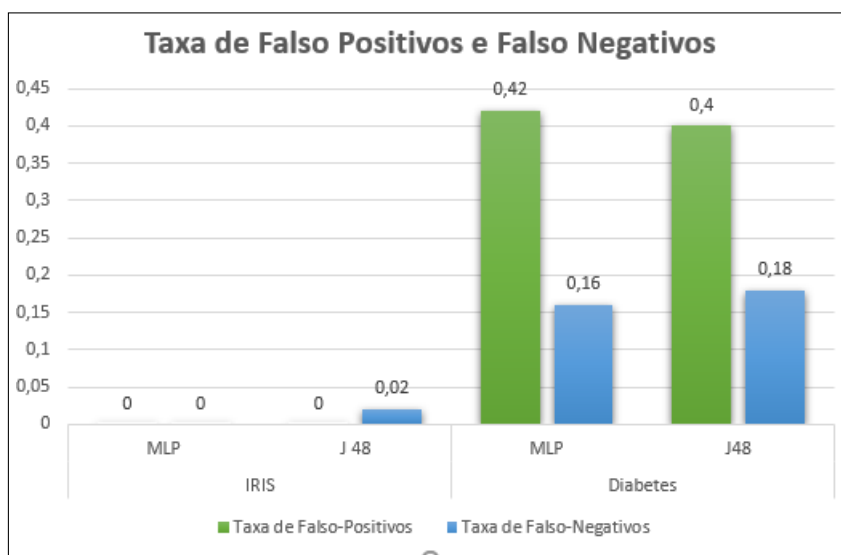
Fonte: elaborado pelo autor.

3.6.1.5 Taxa de Falsos Positivos e Taxa de Falsos Negativos

Por fim, duas métricas restantes utilizadas para a análise de desempenhos dos algoritmos foram as Taxa de Falsos Positivos e Taxa de Falsos Negativos. Estas métricas são definidas pelas fórmulas: $\mathbf{TFP} = \frac{FP}{VN+FP}$ e $\mathbf{TFN} = \frac{FN}{VP+FN}$

A partir destes cálculos, foram aplicadas no experimento e geraram os seguintes dados mostrados na figura 11:

Figura 11 – TFP e TFN dos classificadores J48 e MLP



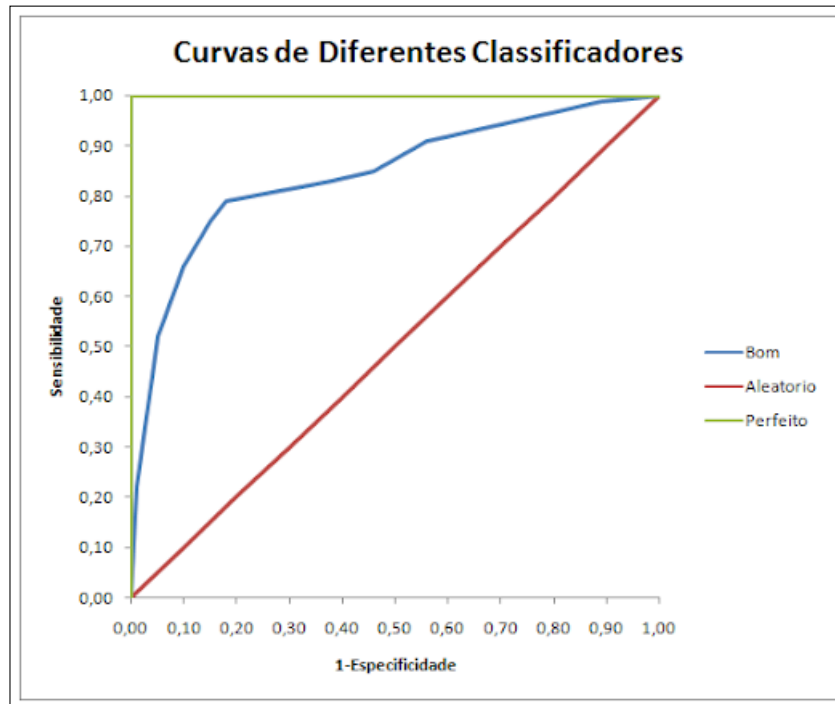
Fonte: elaborado pelo autor.

3.6.1.6 Curvas ROC

As curvas ROC (Receiver Operating Characteristic Curve) servem para avaliar um classificador de maneira gráfica através de seu desempenho, bom ou ruim, com base nas características das taxas de verdadeiros positivos (recall ou sensibilidade) eixo y e taxa

de falsos positivos eixo x do gráfico(Especificidade). É traçada uma diagonal no gráfico. Quanto mais próximo estiver do eixo Y melhor será o desempenho do algoritmo, quando mais próximo do eixo X pior será sua performance,como ilustra a figura 12.

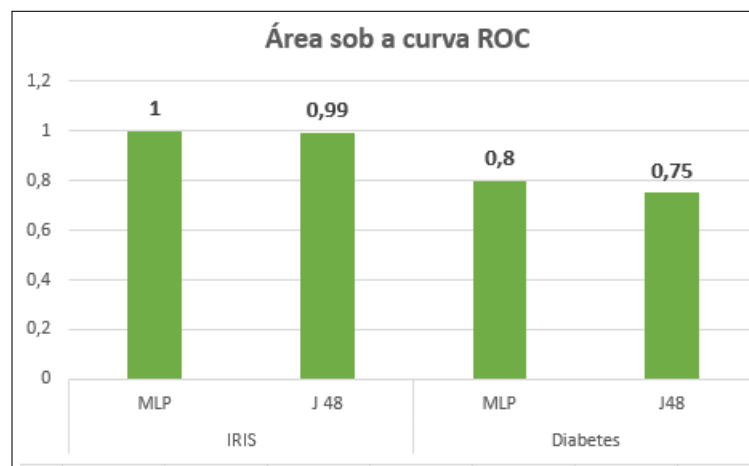
Figura 12 – Exemplo de Curvas ROC



Fonte: (BITTAR, 2021) retirada de "Métricas para avaliação de modelos de machine learning".

Outro fator importante e derivado desse conceito, é a área sob a curva ROC. Ela é capaz de definir a eficiência do algoritmo pelos pontos da curva ROC, quanto mais próximo o valor estiver de 1 melhor será sua performance. A figura 13 mostra os dados da área sob a curva ROC.

Figura 13 – Área sob curva ROC dos classificadores J48 e MLP

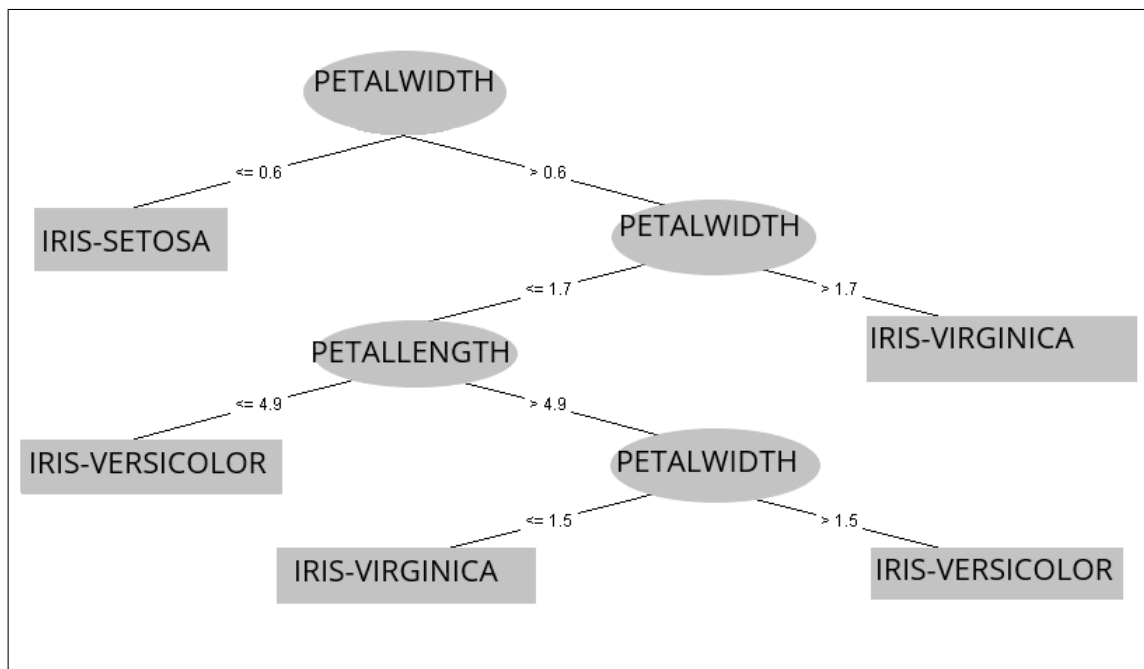


Fonte: elaborado pelo autor.

3.6.1.7 Árvore de decisão gerada pelo algoritmo J48

Na figura 14, podemos observar a árvore gerada pelo algoritmo J48 no conjunto da base IRIS, teve um tamanho de 9 elementos, contando com o número de folhas igual a 2. Identificou aproximadamente 142 instâncias perfeitamente, com uma taxa de sucesso de 94,73% dos dados, e 8 identificados erroneamente, sendo 5,27% dos dados como citado na figura 8 anteriormente.

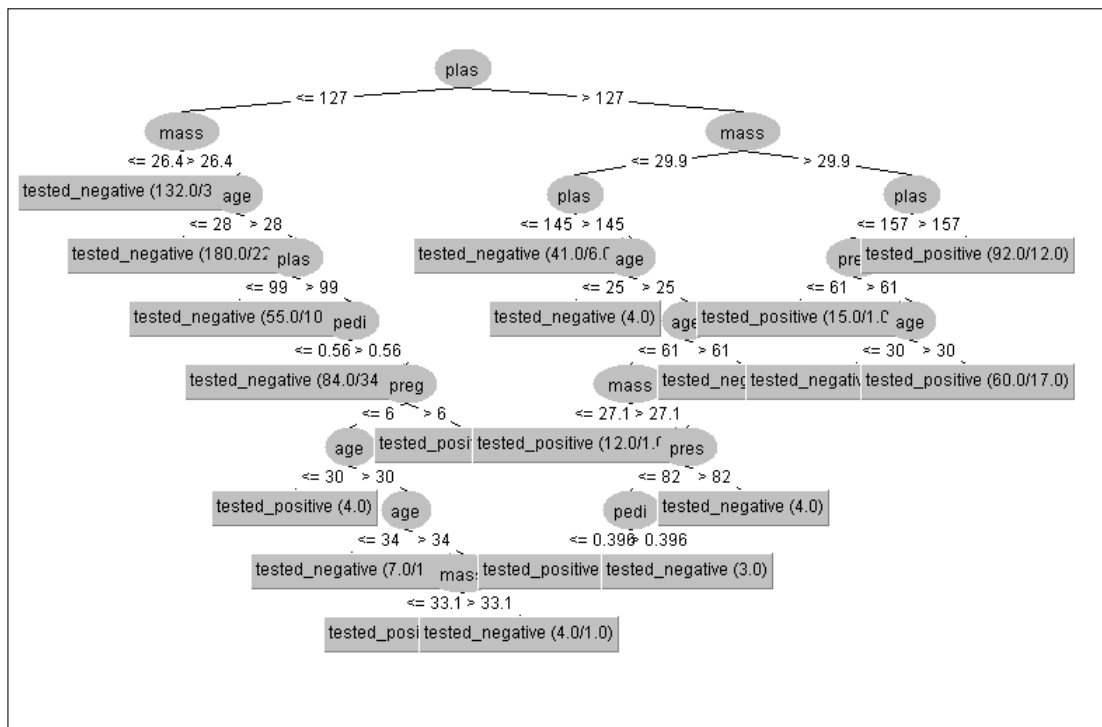
Figura 14 – Árvore de decisão gerada da base de dados IRIS



Fonte: elaborado pelo autor.

Já na figura 15, podemos visualizar a árvore gerada pelo conjunto de dados da base PIMA Indians, que teve um tamanho 39 elementos, com o número de folhas igual a 20. Classificando aproximadamente 572 de forma correta, tendo um êxito de 74,49%, e cerca de 195 equivocadas totalizando 25,51% das intâncias.

Figura 15 – Árvore de decisão gerada da base de dados PIMA Indians

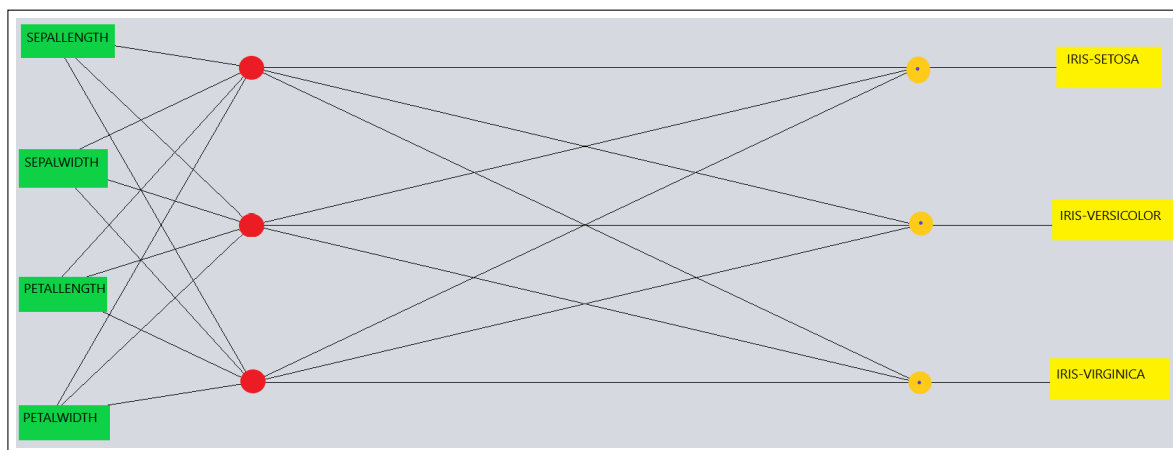


Fonte: elaborado pelo autor.

3.6.1.8 Rede neural gerada pelo algoritmo MLP

Para as redes neurais geradas pelo algoritmo MLP, obteve-se na base IRIS, 1 camada oculta de 3 neurônios, para 4 neurônios na camada de entrada, gerando 3 neurônios na camada de saída, apontado pela figura 16. Cerca de 145 instancias classificadas corretamente, e somente 5 instancias erroneamente identificadas, com uma taxa de sucesso de 96,93% e margem de erro de 3,07%.

Figura 16 – Rede neural gerada da base de dados IRIS

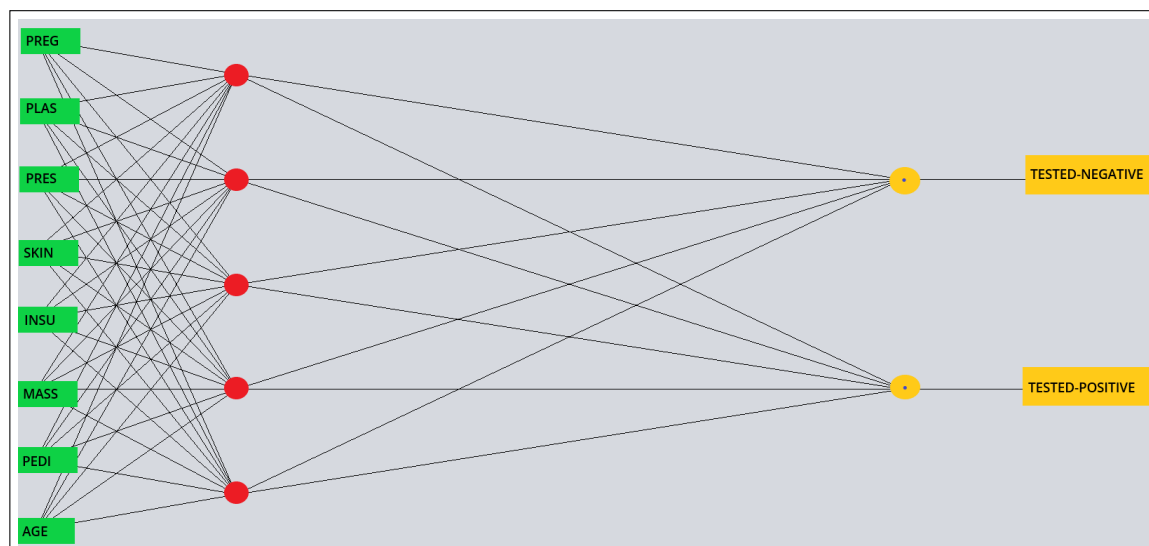


Fonte: elaborado pelo autor.

Já para base PIMA Indians, por ser uma base de dados maior, gerou uma rede neural

mais robusta, porém bem eficiente com 8 neurônios na camada de entrada, formando 1 camada oculta de 5 neurônios, para 2 neurônios de saída indicado pela figura 17. Classificando corretamente em torno de 574 instâncias tendo uma taxa percentual de 74,75%, e por volta de 193 incorretas totalizando 25,25%.

Figura 17 – Rede neural gerada da base de dados IRIS



Fonte: elaborado pelo autor.

Na tabela 4 , apresenta-se os dados gerais obtidos através do experimento aplicado nos conjuntos de dados da base iris e pima diabetes:

Tabela 4 – Resultados gerais extraídos do experimento

Métrica	data set - Iris		data set - Pima__diabetes	
	MLP	J48	MLP	J48
Número corretos	14,54	14,21	57,41	57,21
Número incorretos	0,46	0,79	19,39	19,49
Percentual correto	96,93	94,73	74,75	74,49
Percentual incorreto	3,07	5,27	25,25	25,51
F-score	1,00	0,99	0,81	0,81
Taxa Verdadeiro Positivos	1,00	0,98	0,84	0,82
Taxa Verdadeiro Negativos	1,00	1,00	0,58	0,60
Taxa de Falso Positivos	0	0	0,42	0,40
Taxa de Falso Negativos	0	0,02	0,16	0,18
Área sob curva ROC	1,00	0,99	0,80	0,75

Fonte: elaborado pelo autor

Considerando os valores obtidos pelas métricas de avaliação, após a execução dos experimentos de benchmarking, pode-se observar que o algoritmo MLP devido a sua flexibilidade de aplicação em variados conjuntos de dados, apresentou melhores resultados que o algoritmo J48. Os MLP's por ser um algoritmo que trabalham muito bem em

problemas de classificação quando as entradas são de classes e rótulos, como o conjunto de dados IRIS e o PIMA diabetes. Observa-se ainda, que apesar desta leve vantagem, o algoritmo J48 também teve um bom percentual de classificação dos dados, chegando até a empatar em várias métricas de avaliação como percentual correto , percentual incorreto e F-score, evidenciados pela tabela 4 .

4 Conclusão

Este trabalho proporcionou através de um experimento realizado a avaliação de desempenho entre dois algoritmos seguindo técnicas de benchmarking. Utilizou-se o MLP, que é um dos classificadores mais tradicionais quando se fala em redes neurais multicamadas, e o J48 que é um algoritmo de árvore de decisão. Existem muitos algoritmos baseados em árvores de decisão como, Hoeffding tree, Random Forest, REPTree, entre outros. A ideia de selecionar o J48 deve-se ao fato de ser bem utilizado em pesquisas do meio científico e acadêmico, e também por ser um bom algoritmo para realização de testes comparativos com eficácia na mesma proporção a outros algoritmos encontrados na literatura. Por sua vez, o MLP são um tipo clássico de rede neural, onde se adequam muito bem a problemas de classificação com entrada de dados supervisionados, atuando também de maneira eficaz em conjunto de dados tabulares, previsões em classificação e regressão.

Utilizando as definições e técnicas de benchmarking apresentadas durante o Capítulo 2, pode-se verificar que seu uso serve de base para uma sustentação em critérios avaliativos em diversas áreas do meio científico. Constatou-se que aplicando as técnicas de maneira correta e padronizada possibilitou um melhor direcionamento para as comparações de desempenho e obtenção de resultados precisos do experimento.

As bases de dados utilizadas, PIMA Indians e a base IRIS, dois conjuntos de dados distintos e descritos na seção 3.3. Com estas bases, foram feitos testes de análise de desempenho usando métricas de avaliação recomendadas para algoritmos classificadores, tendo seus dados estatísticos extraídos por meio de uma aplicação da ferramenta WEKA, o WEE. Como resultado do benchmarking feito entre os dois algoritmos classificadores, observou-se uma pequena vantagem de desempenho para o algoritmo MLP, se mostrando mais preciso para as métricas de avaliação como acurácia, que é o percentual de dados classificados corretamente, observou-se também a eficácia na identificação de taxa de verdadeiros negativos e positivos, igualmente chamados de precisão e recall. Com o auxílio da métrica de Curvas ROC, foi possível analisar melhor graficamente o desempenho de cada algoritmo para as bases utilizadas.

Este trabalho deixa sua colaboração, ao apresentar uma análise de desempenhos dos algoritmos J48 de árvores de decisão e MLP de redes neurais. Expondo pontos de vista distintos sobre uma mesma problemática e melhor aplicação, ressaltando que não existe um algoritmo que transcende sobre o outro, então deve-se realizar experimentos a fim de buscar soluções mais apropriadas para exercerem em determinadas situações.

Como trabalhos futuros, podemos sugerir a utilização de outros algoritmos de classificação como Support Vector Machine (SVM), Random Forest, entre outros em

experimentos de benchmarking.

Referências

- ALBERTIN, M. R.; KOHL, H.; ELIAS, S. J. B. Manual do benchmarking. Imprensa Universitária, 2016. Citado na página 15.
- BERRY, M.; CYBENKO, G.; LARSON, J. Scientific benchmark characterizations. *Parallel Computing*, Elsevier, v. 17, n. 10-11, p. 1173–1194, 1991. Citado na página 12.
- BONINI, J. A. Aplicação de algoritmos de árvore de decisão sobre uma base de dados de câncer de mama. *Revista ComInG-Communications and Innovations Gazette*, v. 1, n. 1, p. 57–67, 2016. Citado na página 28.
- CAMP, R. C. A bible for benchmarking, by xerox. *Financial executive*, Financial executives international, v. 9, n. 4, p. 23–28, 1993. Citado 2 vezes nas páginas 14 e 17.
- CIFERRI, R. R. et al. Um benchmark voltado a análise de desempenho de sistemas de informações geográficas. [sn], 1995. Citado 2 vezes nas páginas 18 e 19.
- COSTA, D. B. Diretrizes para a realização de processo de benchmarking colaborativo visando à implementação de melhorias em empresas de construção civil. 2008. Citado na página 16.
- FREYTAG, P. V.; HOLLENSSEN, S. The process of benchmarking, benchlearning and benchaction. *The TQM magazine*, MCB UP Ltd, 2001. Citado na página 18.
- GRAY, J. *Benchmark handbook: for database and transaction processing systems*. [S.l.]: Morgan Kaufmann Publishers Inc., 1992. Citado na página 19.
- HONG, P.; HONG, S. W.; ROH, J. J.; PARK, K. Evolving benchmarking practices: a review for research perspectives. *Benchmarking: An International Journal*, Emerald Group Publishing Limited, 2012. Citado na página 14.
- KOHL, H. *Integriertes Benchmarking für kleine und mittlere Unternehmen: eine Methode zur Integration von Best-Practice-Informationen in das interne Unternehmenscontrolling*. [S.l.]: Fraunhofer-IRB-Verlag, 2007. Citado na página 17.
- KYRÖ, P. Revising the concept and forms of benchmarking. *Benchmarking: An International Journal*, MCB UP Ltd, 2003. Citado 2 vezes nas páginas 15 e 16.
- MADEIRA, P. Benchmarking: a arte de copiar. *Jornal do Técnico de Contas e da Empresa (JTCE)*, Publístudos, p. 364–367, 1999. Citado 2 vezes nas páginas 14 e 17.
- NAZZAL, J. M.; EL-EMARY, I. M.; NAJIM, S. A. Multilayer perceptron neural network (mlps) for analyzing the properties of jordan oil shale 1. Citeseer, 2008. Citado na página 27.
- QUEIROZ, L. T. et al. Um benchmark para avaliação de técnicas de busca no contexto de análise de mutantes sql. Universidade Federal de Goiás, 2013. Citado na página 19.
- SANTOS, G. d. O. Benchmarking em algoritmos de classificação na mineração de dados. Universidade Federal do Maranhão, 2018. Citado na página 19.

- SIGILLITO, V. Pima indians diabetes database. *UCI Machine Learning Repository* [[http://archive.ics.uci.edu/ml/datasets/Pima Indians Diabetes](http://archive.ics.uci.edu/ml/datasets/Pima%20Indians%20Diabetes)]. Phoenix, AZ: National Institute of Diabetes and Digestive and Kidney Diseases, 1990. Citado na página 24.
- SOKOLOVA, M.; LAPALME, G. A systematic analysis of performance measures for classification tasks. *Information processing & management*, Elsevier, v. 45, n. 4, p. 427–437, 2009. Citado na página 29.
- SPENDOLINI, M. J. The benchmarking process. *Compensation & Benefits Review*, Sage Publications Sage CA: Thousand Oaks, CA, v. 24, n. 5, p. 21–29, 1992. Citado na página 17.
- TELOKEN, A. et al. Estudo comparativo entre os algoritmos de mineração de dados random forest e j48 na tomada de decisão. *Simpósio de Pesquisa e Desenvolvimento em Computação*, v. 2, n. 1, 2016. Citado na página 28.
- WATSON, G. H. *Strategic benchmarking: How to rate your company's performance against the world's best*. [S.l.]: John Wiley & Sons Incorporated, 1993. Citado 2 vezes nas páginas 15 e 17.
- WITTEN, I. H.; FRANK, E.; MARK, A. Hall, and christopher j pal. 2017. data mining: Practical machine learning tools and techniques. *Morgan Kaufmann, Cambridge, MA*, v. 10, p. 3086818. Citado na página 20.