UNIVERSIDADE FEDERAL DO AMAZONAS
INSTITUTO DE CIÊNCIAS EXATAS E TECNOLOGIA-ICET

GEOVANNA BEATHRYZ
GUSTAVO SOUZA
IASMIM BRAGA
JEAN BARAÚNA
KAIO SOBRAL
PEDRO JHEIVISON

TRABALHO PRÁTICO 2 – IHC E MACHINE LEARNING

Classificação Supervisionada Aplicada a Problemas de Usabilidade

ITACOATIARA - AM
2025

GEOVANNA BEATHRYZ
GUSTAVO SOUZA
IASMIM BRAGA
JEAN BARAÚNA
KAIO SOBRAL
PEDRO JHEIVISON

TRABALHO PRÁTICO 2 – IHC E MACHINE LEARNING

Classificação Supervisionada Aplicada a Problemas de Usabilidade

Relatório do Trabalho Prático 2 (TP2) apresentado à disciplina de Interação Humano-Computador (IHC), do Instituto de Ciências Exatas e Tecnologia (ICET) da Universidade Federal do Amazonas (UFAM), como requisito parcial para a obtenção de nota.

Prof. Dr. Andrey Rodrigues

ITACOATIARA - AM
2025

Sumário

RELATÓRIO FINAL

1. Contexto e Motivação

O problema de classificação foi formulado no contexto de um aplicativo hipotético chamado **"Agenda Fácil"**, projetado para profissionais autônomos (como manicures, barbeiros e personal trainers) gerenciarem seus agendamentos de clientes. A principal proposta de valor do aplicativo é ser simples e eficiente, minimizando a chance de erros comuns em agendamentos manuais.

A motivação para este estudo é investigar se é possível treinar um modelo de machine learning para classificar automaticamente o nível de usabilidade de uma sessão de uso, com base em dados quantitativos da interação do usuário. Tal modelo poderia, no futuro, ajudar desenvolvedores a identificar pontos de atrito na interface e a validar melhorias de design de forma automatizada, prevendo se uma interação foi positiva ou negativa.

2. Atributos Preditores

Para treinar o modelo, foram selecionados 5 atributos preditores que representam diferentes aspectos de uma tarefa central do aplicativo: "realizar um novo agendamento". Esses atributos foram escolhidos por serem métricas clássicas de usabilidade, como tempo de tarefa e número de erros.

- **tempo_para_agendar (Numérico):** Tempo total, em segundos, que o usuário levou para concluir a tarefa de agendamento. Mede a *eficiência*.
- **passos_ate_concluir (Numérico):** Número total de cliques ou toques necessários para finalizar o agendamento. Mede a *complexidade do fluxo de interação*.
- **usou_lista_clientes (Nominal: {sim, não}):** Indica se o usuário utilizou a funcionalidade de selecionar um cliente já cadastrado, uma ação que representa um caminho mais eficiente.
- **ativou_lembrete (Nominal: {sim, nao}):** Verifica se o usuário ativou a função de lembrete automático, indicando engajamento com funcionalidades chave do sistema.

- **erros_no_fluxo (Numérico):** Quantidade de erros cometidos durante o processo, como clicar em um botão desabilitado ou tentar agendar em um horário já ocupado. Mede a *eficácia*.

3. Classe-Alvo

A classe-alvo, ou seja, o atributo que desejamos prever, foi definida para representar a qualidade geral da usabilidade da sessão.

- **Nome da Classe:** nível usabilidade
- **Valores Possíveis:** {Alta, Media, Baixa}
- **Tipo de Problema:** Multiclasse

Cada valor representa uma avaliação da interação: **Alta** indica uma experiência fluida e eficiente; **Média** indica que a tarefa foi concluída, mas com alguma dificuldade; e **Baixa** indica uma experiência frustrante e ineficiente.

4. Regras usadas para gerar a classe-alvo

Usabilidade Alta: tempo_agendar < 45 segundos E erros_fluxo = 0 Usabilidade Baixa: tempo_agendar > 120 segundos OU erros_fluxo > 2 Usabilidade Média: Casos que não se enquadram nas condições anteriores.

5. Descrição da Base Sintética

A base de dados utilizada neste trabalho, denominada base_sintetica.arff, foi gerada artificialmente para simular o comportamento de usuários no aplicativo "Agenda Fácil". O objetivo foi criar um conjunto de dados controlado para treinar um modelo de Machine Learning capaz de classificar o nível de usabilidade de uma interação com base em métricas de desempenho.

A geração foi guiada por um conjunto de regras explícitas, detalhadas na proposta do projeto. A base de dados contém 200 instâncias e 6 atributos, sendo 5 atributos preditores e 1 atributo-alvo (a classe).

Os atributos são descritos a seguir:

tempo_agendar (Numérico): Representa o tempo total, em segundos, que o usuário levou para concluir um agendamento**.**

passos_concluir (Numérico): Indica o número de cliques ou toques que o usuário realizou para finalizar a tarefa.

usou_lista_clientes (Nominal: {sim, nao}): Informa se o usuário utilizou a lista de clientes pré-cadastrados, um recurso que otimiza o fluxo.
ativou_lembrete (Nominal: {sim, nao}): Verifica se o usuário ativou funcionalidades adicionais, como lembretes para o agendamento.
erros_fluxo (Numérico): Quantifica o número de erros cometidos pelo usuário durante o processo de agendamento**.**
nivel_usabilidade (Nominal: {Alta, Media, Baixa}): Este é o atributo-alvo (classe). Ele classifica a interação do usuário em três níveis de usabilidade, com base nas regras de geração de dados.
Esse conjunto de dados serve como alicerce para os experimentos, permitindo a aplicação de algoritmos de classificação para identificar padrões que se correlacionam com uma experiência de usuário positiva ou negativa.

6. Descrição dos Experimentos no Weka

Nesta etapa, foi utilizada a ferramenta Weka Explorer para a análise exploratória e experimental da base de dados base_sintetica.arff.

6.1 Análise Visual dos Dados

Inicialmente, os dados foram carregados na aba “Visualize”, onde foi gerada uma Matriz de Gráficos de Dispersão (Plot Matrix) para observar a distribuição dos atributos e suas possíveis correlações. As instâncias foram coloridas de acordo com a variável-alvo nivel_usabilidade, conforme a legenda:

- Azul escuro: Alta usabilidade
- Vermelho: Média usabilidade
- Ciano: Baixa usabilidade

A partir dessa análise, foi possível identificar os seguintes padrões:

- **Correlação entre atributos numéricos:** Observou-se uma correlação positiva entre os atributos tempo_agendar e passos_concluir. As instâncias formam uma tendência diagonal, indicando que tarefas que exigem mais passos também demandam mais tempo. A classe Alta (azul escuro) concentra-se em valores baixos de tempo e passos, a classe Média (vermelho) apresenta valores intermediários, e a classe Baixa (ciano) predomina em valores elevados.
- **Separação das classes:** Os atributos numéricos demonstraram uma boa capacidade de distinguir as classes de usabilidade. tempo_agendar e

passos_concluir baixos tendem a indicar usabilidade Alta, enquanto valores altos apontam para usabilidade Baixa. O atributo erros_fluxo também se mostrou um forte preditor: 0 erros está associado à predominância da classe Alta, 1 a 2 erros à Média , e acima de 2 erros à Baixa.

- **Influência dos atributos categóricos:** Verificou-se que a maioria das instâncias da classe Alta ocorre quando usou listaclientes e ativou lembrete possuem valor "sim", sugerindo que o uso dessas funcionalidades está associado a uma melhor percepção de usabilidade.

A exploração inicial indicou padrões visuais bem definidos e uma separação clara entre as classes, o que sugere que os algoritmos de aprendizado de máquina devem ser capazes de aprender regras de classificação com alta precisão.

7. Resultados dos Experimentos de Classificação

Para avaliar a capacidade de predição do nível de usabilidade, foram treinados e testados cinco algoritmos de classificação: ZeroR, OneR, J48, Naive Bayes e IBk. A base de dados foi dividida com a abordagem *Percentage Split*, utilizando 66% dos dados para treino. A seguir, são apresentados os resultados detalhados para cada modelo.

7.1 ZeroR

O ZeroR, um classificador de linha de base, previu sempre a classe majoritária "Media".

- **Acurácia:** 46,9697%

Matriz de Confusão:

```
 a  b  c   <-- classified as
 0 18  0 |  a = Alta
 0 31  0 |  b = Media
 0 17  0 |  c = Baixa
```

7.2 OneR

O OneR (One Rule) gerou uma regra de classificação baseada no atributo que melhor prediz a classe-alvo.

- **Acurácia:** 98,4848%

Matriz de Confusão:

```
 a  b  c   <-- classified as
17  1  0 |  a = Alta
 0 31  0 |  b = Media
 0  0 17 |  c = Baixa
```

7.3 J48 (Árvore de Decisão)

O J48 é uma implementação do algoritmo C4.5 que gera uma árvore de decisão para classificar as instâncias.

- **Acurácia:** 100%

Matriz de Confusão:

```
 a  b  c   <-- classified as
18  0  0 |  a = Alta
 0 31  0 |  b = Media
 0 17  0 |  c = Baixa
```

7.4 Naive Bayes

Este classificador probabilístico utiliza o teorema de Bayes com a suposição de independência entre os atributos.

- **Acurácia:** 100%

Matriz de Confusão:

```
 a  b  c   <-- classified as
18  0  0 |  a = Alta
 0 31  0 |  b = Media
 0 17  0 |  c = Baixa
```

7.5 IBk (k-Nearest Neighbors)

O IBk é um classificador baseado em instâncias. Foi utilizado com a configuração padrão (k=1).

- **Acurácia:** 98,4848%

Matriz de Confusão:

```
 a  b  c   <-- classified as
18  0  0 |  a = Alta
 0 30  1 |  b = Media
 0  0 17 |  c = Baixa
```

7.6 Tabela Comparativa dos Resultados

Algoritmo	Acurácia	Instâncias Corretamente Classificadas
ZeroR	46,97%	31 / 66
OneR	98,48%	65 / 66
IBk (kNN)	98,48%	65 / 66
J48	100%	66 / 66
Naive Bayes	100%	66 / 66

8. Análise do Modelo da Árvore de Decisão (J48)

Uma das principais vantagens do algoritmo J48 é a alta interpretabilidade do seu modelo. A árvore de decisão gerada revela um conjunto de regras hierárquicas e claras que o algoritmo "descobriu" a partir dos dados , fornecendo insights valiosos sobre a relação entre os atributos e o nível de usabilidade.

- O Fator Decisivo: Tempo na Tarefa (tempo_agendar)

  A raiz da árvore, e portanto o atributo mais importante, é o tempo_agendar. O modelo aprendeu um limiar crítico de 120 segundos. Se o tempo para concluir a tarefa excede esse valor, a usabilidade é incondicionalmente classificada como Baixa. Isso demonstra que, no contexto do sistema avaliado, a eficiência é o principal pilar da percepção de usabilidade.

- O Diferencial para a Qualidade: Ausência de Erros (erros_fluxo)

  Para tarefas concluídas em tempo aceitável (<=120 segundos), o fator determinante para a qualidade da experiência passa a ser a ocorrência de erros. O modelo aprendeu que:

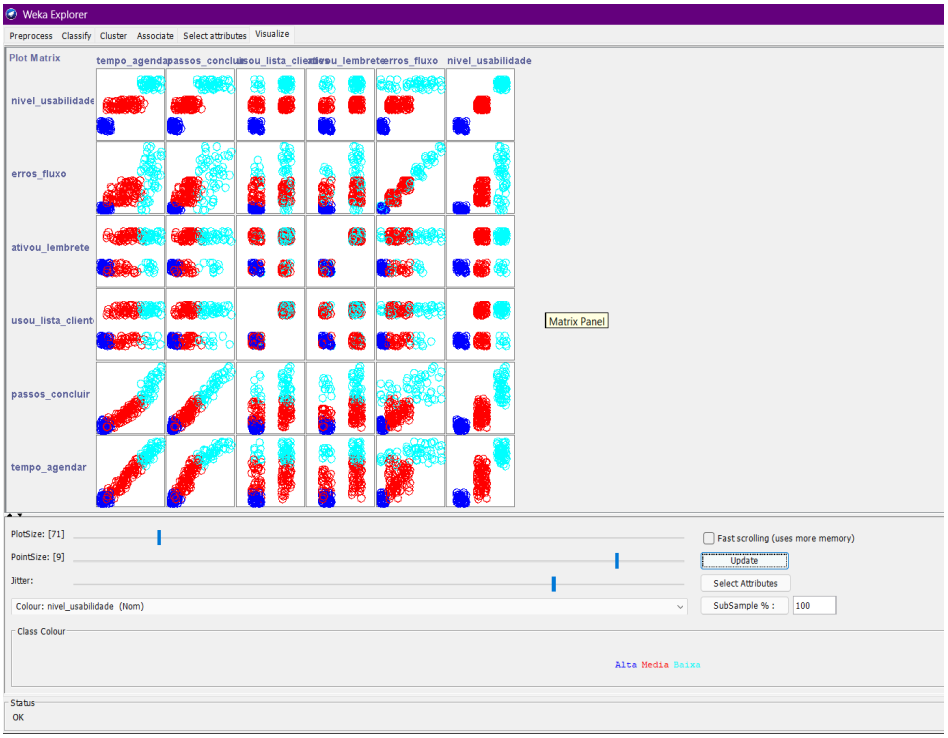
➢ Se o usuário comete pelo menos um erro (>0), a usabilidade é classificada como **Média**.

- ➢ Apenas quando a tarefa é concluída de forma eficiente e sem nenhum erro (<=0), a usabilidade atinge o nível **Alta**.

As regras extraídas da árvore de decisão validam empiricamente conceitos fundamentais de Interação Humano-Computador. O modelo confirmou que uma boa usabilidade é uma função direta da eficiência (baixo tempo para completar a tarefa) e da eficácia (ausência de erros). A clareza do modelo J48 não só resultou em uma acurácia perfeita, mas também forneceu uma explicação lógica e alinhada com a teoria de IHC para o problema de classificação proposto.

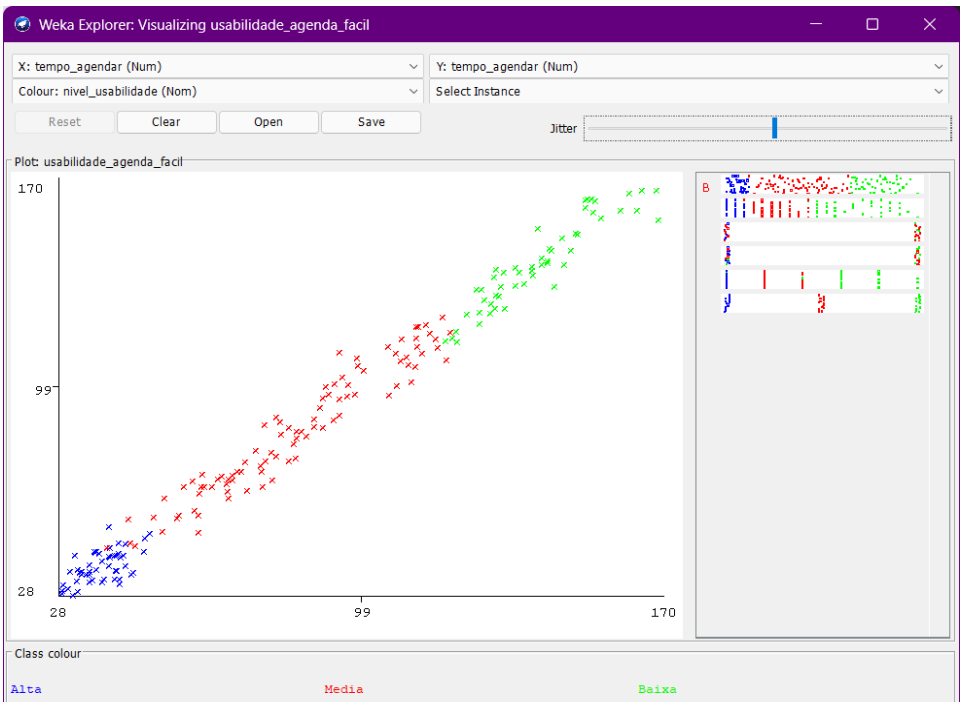
9. Resultados

Figura 1 - Aba “Visualize”



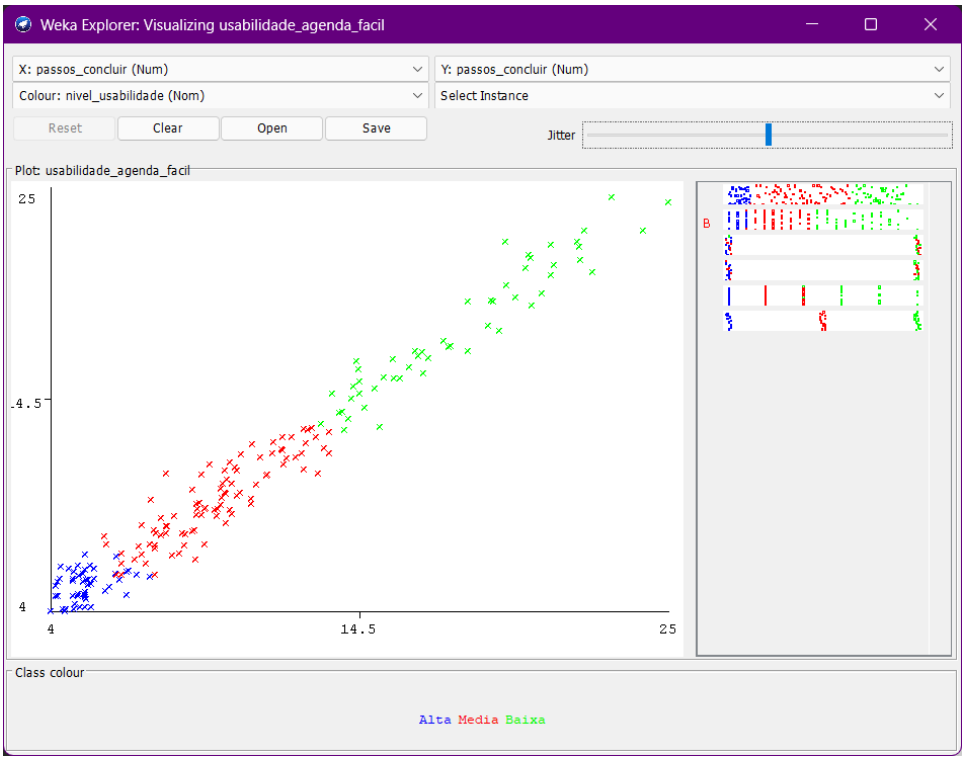
Fonte: Elaborado pelos autores (2025), a partir do software Weka.

Figura 2- Gráfico de Dispersão : tempo_agendar



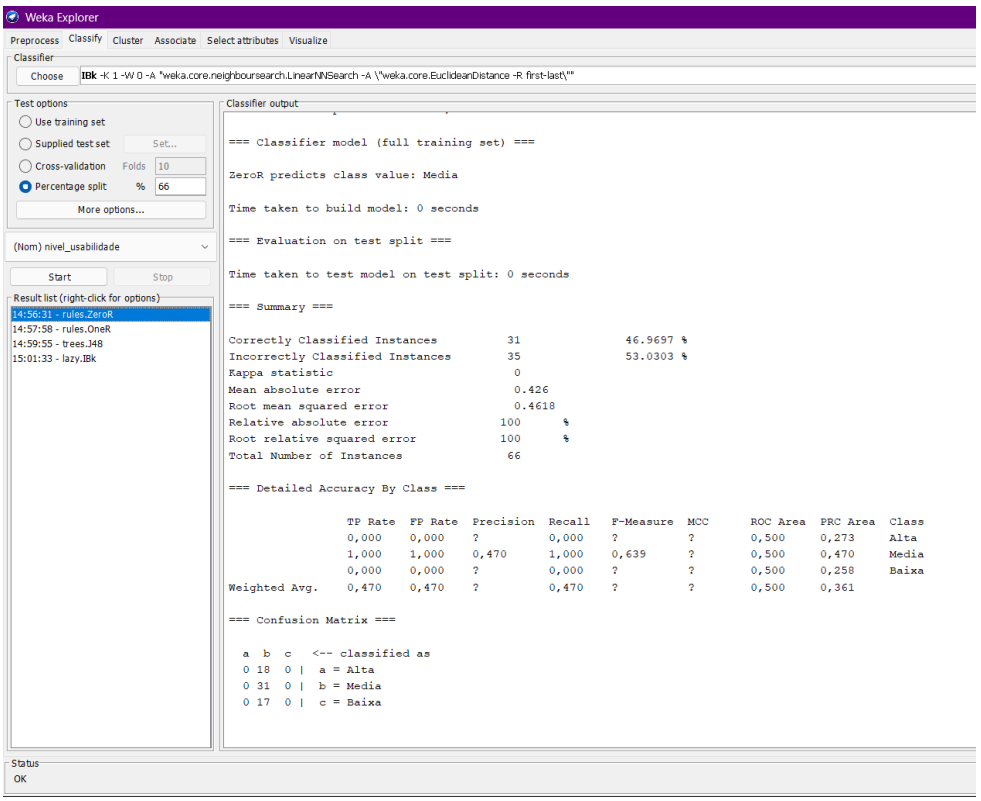
Fonte: Elaborado pelos autores (2025), a partir do software Weka.

Figura 3- Gráfico de Dispersão: passos_concluir



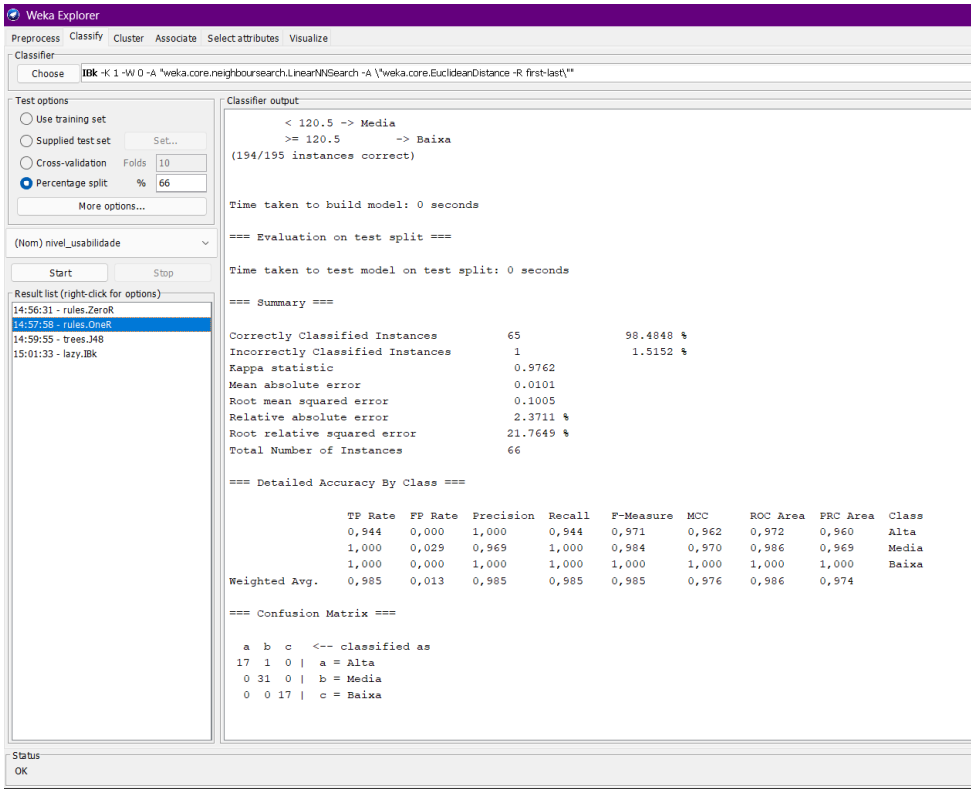
Fonte: Elaborado pelos autores (2025), a partir do software Weka.

Figura 4- ZeroR (Baseline)



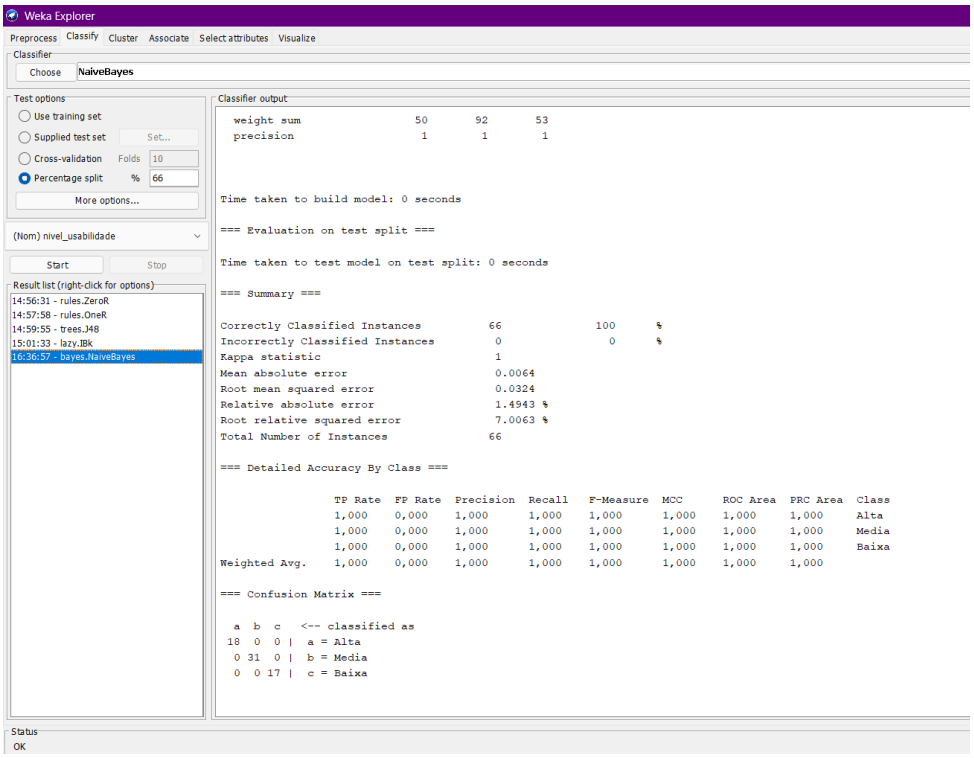
Fonte: Elaborado pelos autores (2025), a partir do software Weka.

Figura 5- OneR (Baseline)



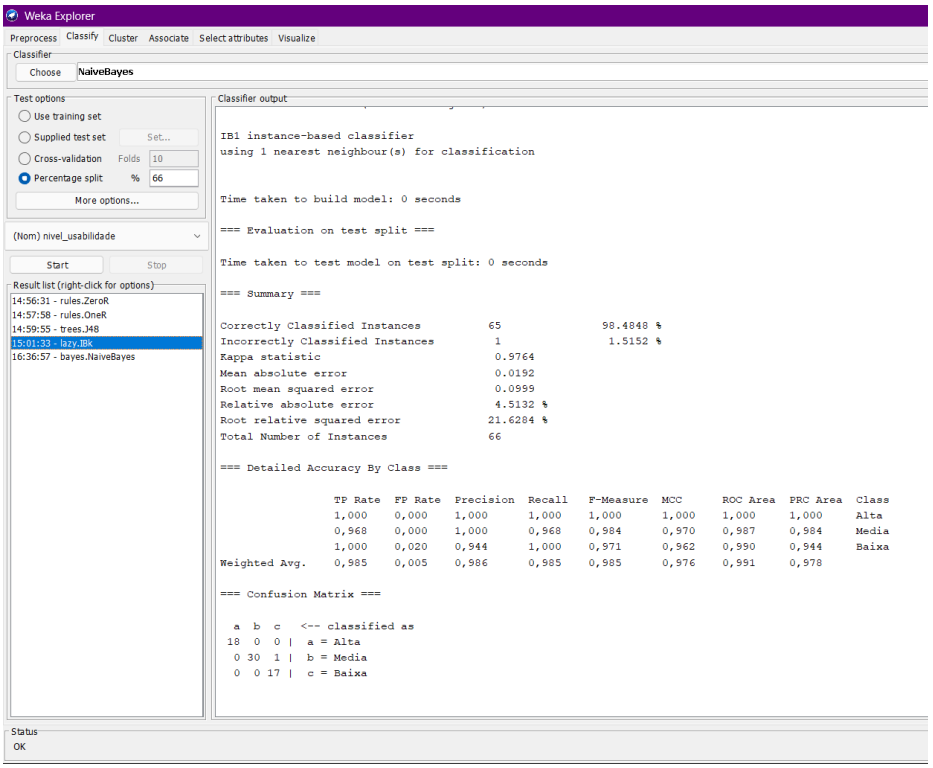
Fonte: Elaborado pelos autores (2025), a partir do software Weka.

Figura 6- Algoritmo NaiveBayes



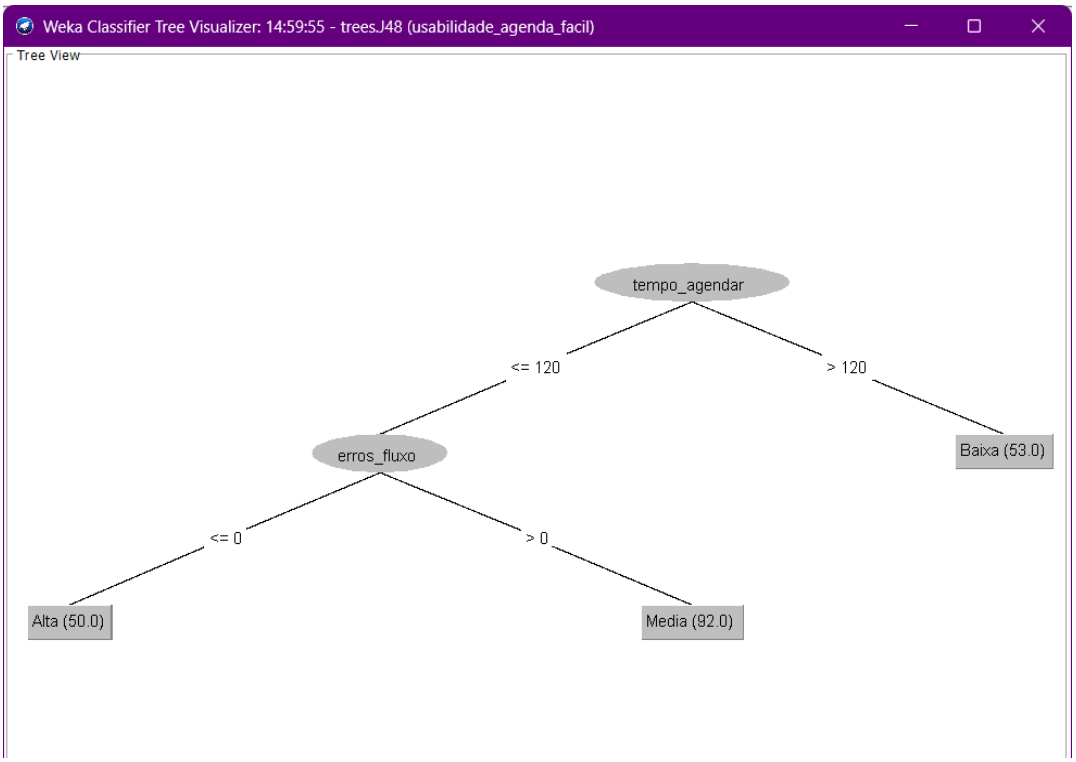
Fonte: Elaborado pelos autores (2025), a partir do software Weka.

Figura 7- Algoritmo IBk (k-NN)



Fonte: Elaborado pelos autores (2025), a partir do software Weka.

Figura 8- J48 (Árvore de Decisão)



Fonte: Elaborado pelos autores (2025), a partir do software Weka.

10. Análise crítica dos resultados em relação ao domínio de IHC

A análise comparativa do desempenho dos modelos no conjunto de teste (66 instâncias) revelou uma alta previsibilidade dos dados, permitindo extrair conclusões significativas sobre a avaliação da experiência do usuário no domínio de IHC aplicado ao sistema "Agenda Fácil".

ZeroR (Baseline): Atingiu apenas 46,97% de acurácia. Como esperado, ele classificou todas as instâncias de teste como a classe majoritária, "Média''. Isso define a linha de base: qualquer modelo útil deve superar esse valor.

OneR (Baseline): Demonstrou um salto significativo, alcançando 98,48% de acurácia. Ele errou apenas 1 das 66 instâncias, indicando que um único atributo (provavelmente tempo_agendar ou erros_fluxo) era um preditor muito forte.

IBk (k-NN): Teve um desempenho idêntico ao OneR, com 98,48% de acurácia. sua matriz de confusão mostra que também cometeu apenas um erro, classificando uma instância "Media" como "Alta".

J48 e **Naive Bayes**: Ambos os modelos atingiram o desempenho máximo de 100% de acurácia. Eles conseguiram classificar corretamente todas as 66 instâncias de teste sem nenhum erro.Os resultados demonstram que a usabilidade, neste contexto, não é um conceito puramente subjetivo. Pelo contrário, ela pode ser quantificada e prevista com um nível de precisão notavelmente alto. O desempenho perfeito dos algoritmos J48 e Naive Bayes indica que os atributos de interação selecionados possuem uma forte correlação com a classificação da experiência do usuário.

A análise do modelo J48 (Árvore de Decisão), que se destaca pela sua interpretabilidade, oferece os insights mais claros:

A Eficiência como Pilar Central: A árvore elegeu o atributo tempo_agendar como o nó raiz, identificando-o como o fator mais decisivo para a classificação da usabilidade. O modelo aprendeu que um tempo de tarefa superior a 120 segundos resulta, incondicionalmente, em uma classificação de usabilidade "Baixa". Isso

sugere que, para o perfil de usuário deste aplicativo, a eficiência e a velocidade na conclusão da tarefa são os principais pilares da experiência percebida.

A Eficácia como Diferencial para a Excelência: Após o critério de eficiência ser atendido (tempo <= 120s), o modelo utiliza a eficácia (erros_fluxo) como o fator de desempate. A árvore demonstra que apenas as interações concluídas sem nenhum erro (erros_fluxo <= 0) alcançam o nível "Alta". A ocorrência de qualquer erro (erros_fluxo > 0), mesmo que o tempo tenha sido satisfatório, rebaixa a classificação para "Média''.Notavelmente, atributos secundários, como usou_lista_clientes ou ativou_lembrete, não foram selecionados pela árvore J48, indicando que seu impacto na classificação é irrelevante quando comparado à velocidade e à ausência de erros.

A Validação das Regras: O que o J48 “Descobriu" vs o que foi Definido. Um dos objetivos centrais do trabalho era verificar se o modelo conseguiria "aprender" os padrões de usabilidade predefinidos. Ao comparar as regras originais com as extraídas da árvore de decisão J48, confirma-se que o modelo não só encontrou os padrões, como também os otimizou Para a classe **Baixa**: A regra original era tempo_agendar > 120 segundos ou erros_fluxo > 2. O J48 aprendeu uma regra mais direta: apenas tempo_agendar >120. Isso revela que o tempo de tarefa foi um indicador tão forte que se sobrepôs à necessidade de analisar os erros para identificar uma experiência ruim.

Para a classe Alta: A regra original era tempo_agendar < 45 segundos eerros_fluxo = 0. O J48 generalizou essa condição para tempo_agendar <= 120 e erros_fluxo <= 0. O modelo expandiu o critério de tempo, entendendo que qualquer tarefa concluída em um tempo razoável e sem erros poderia ser classificada como de alta usabilidade. Em suma, os resultados validam empiricamente que a experiência do usuário no "Agenda Fácil" é uma função direta dos dois principais pilares da usabilidade: eficiência (tempo) e eficácia (erros). Notavelmente, atributos secundários de engajamento, como usou_lista_clientes ou ativou_lembrete, não foram selecionados pela árvore J48, indicando que seu impacto na classificação da usabilidade é irrelevante quando comparado à velocidade e à ausência de erros.Para o domínio de IHC, isso comprova que um modelo de machine learning pode ser uma ferramenta poderosa de diagnóstico. Ele permite à equipe de design

focar em métricas objetivas: para otimizar a experiência do usuário deste sistema, os esforços devem ser direcionados a garantir que o fluxo da tarefa principal possa ser concluído em menos de 120 segundos e de forma tão intuitiva que minimize a ocorrência de erros.

CONCLUSÃO

Este trabalho demonstrou com sucesso a aplicabilidade de técnicas de Machine Learning na avaliação automática de usabilidade, integrando de forma prática os domínios de Interação Humano-Computador (IHC) e Aprendizado de Máquina. A partir de um cenário realista de um aplicativo de agendamentos, foi possível construir, treinar e validar modelos capazes de classificar a qualidade da experiência do usuário com base exclusivamente em métricas objetivas de interação.