

3. Exploração e Classificação no Weka

Nesta etapa, a base de dados sintética, contendo 300 instâncias e 7 atributos relacionados à usabilidade de um site educacional, foi carregada na ferramenta Weka. A primeira fase do experimento consistiu na exploração visual dos dados, conforme detalhado a seguir.

3.1. Análise Visual das Distribuições e Padrões

Utilizando a aba "Visualize" do Weka, foi realizada uma análise dos atributos para identificar padrões visuais e correlações com a classe-alvo, **nivel_usabilidade**. Esta classe é dividida em três categorias: 'baixa' (representada pela cor azul nos gráficos), 'media' (vermelho) e 'alta' (ciano/verde).

Para uma visão geral, a Matriz de Dispersão (*Plot Matrix*) foi gerada, cruzando todos os atributos entre si.

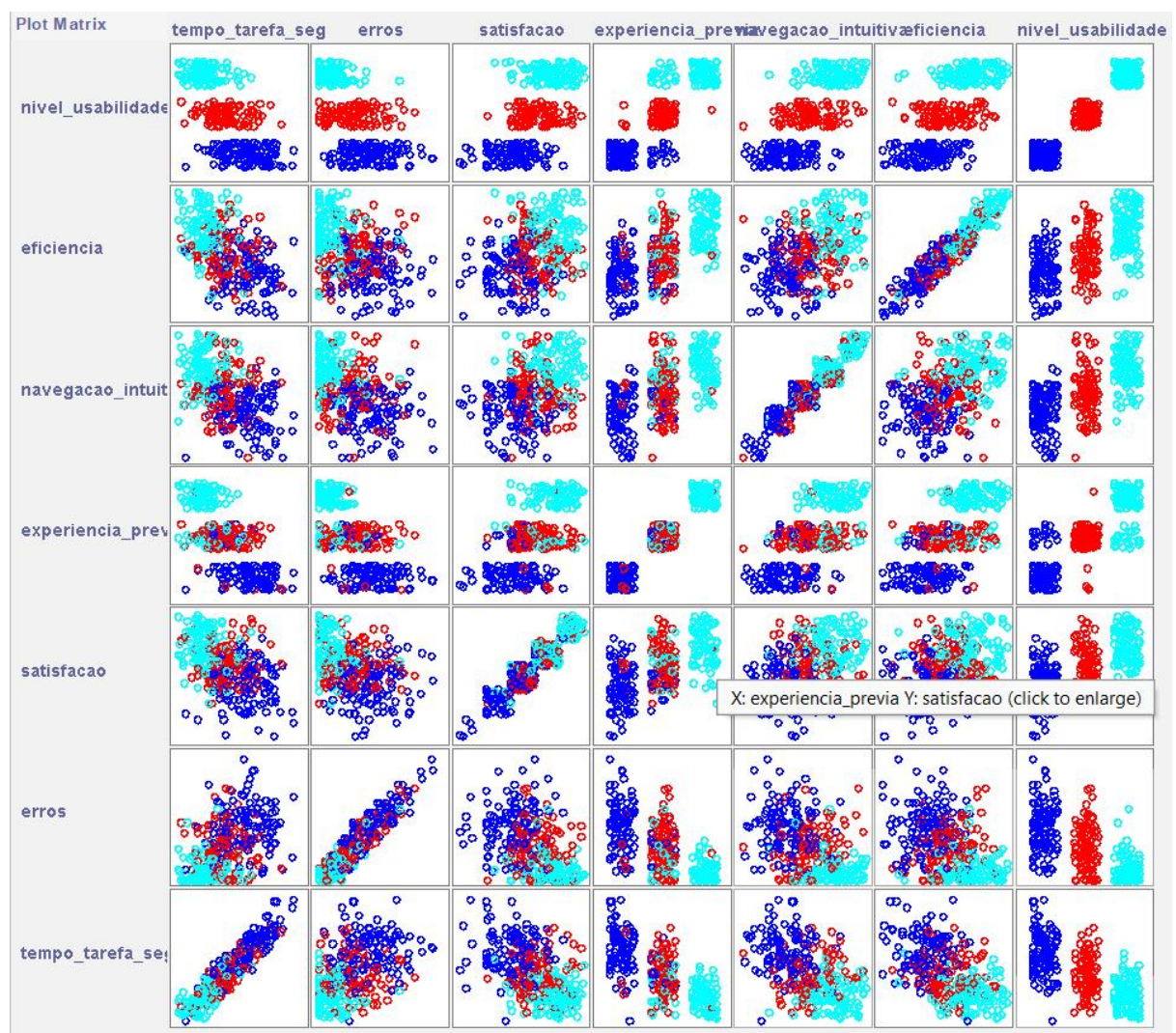


Figura 1: Matriz de Dispersão (*Plot Matrix*) de todos os atributos da base de dados.

A análise da matriz já sugere que a maioria dos atributos possui um alto poder de separação das classes. A seguir, detalhamos a análise para cada atributo preditor.

3.1.1. Atributo **eficiencia**

Este atributo numérico demonstrou ser o preditor mais poderoso, exibindo uma separação de classes praticamente perfeita. O gráfico de dispersão abaixo mostra que os dados estão organizados em três grupos distintos e não sobrepostos, indicando que **eficiencia** será, muito provavelmente, o atributo principal utilizado pelos modelos de classificação, como o atributo raiz em uma árvore de decisão J48.

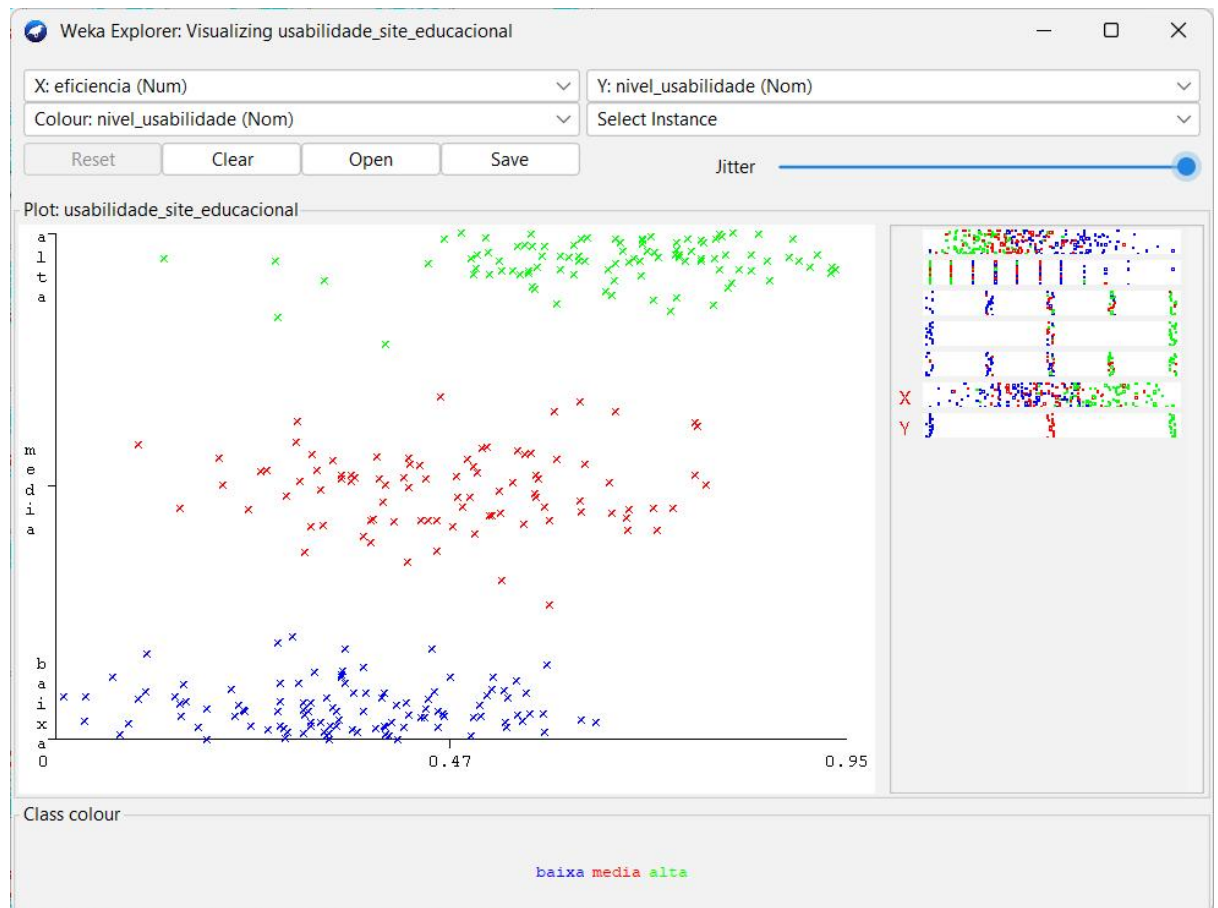


Figura 2: Dispersão do atributo **eficiencia** em relação à classe **nivel_usabilidade**.

3.1.2. Atributos **tempo_tarefa_seg** e **erros**

Ambos os atributos apresentam uma forte correlação negativa com a usabilidade, um padrão clássico em estudos de IHC. A análise visual mostra que instâncias de usabilidade 'alta' estão concentradas em valores baixos de tempo e erro, transitando claramente para as classes 'media' e 'baixa' conforme os valores desses atributos aumentam.



Figura 3: Dispersão do **tempo_tarefa_seg** em relação à classe.



Figura 4: Dispersão do número de **erros** em relação à classe.

3.1.3. Atributos **satisfacao** e **navegacao_intuitiva**

Sendo atributos nominais, **satisfacao** e **navegacao_intuitiva** também se destacam como preditores de elite. Os gráficos mostram um comportamento "em blocos", onde cada categoria do atributo está quase que exclusivamente associada a uma única classe de usabilidade. Este padrão é ideal para algoritmos baseados em regras, pois as associações são diretas e evidentes.



Figura 5: Relação entre **satisfacao** e a classe **nivel_usabilidade**.

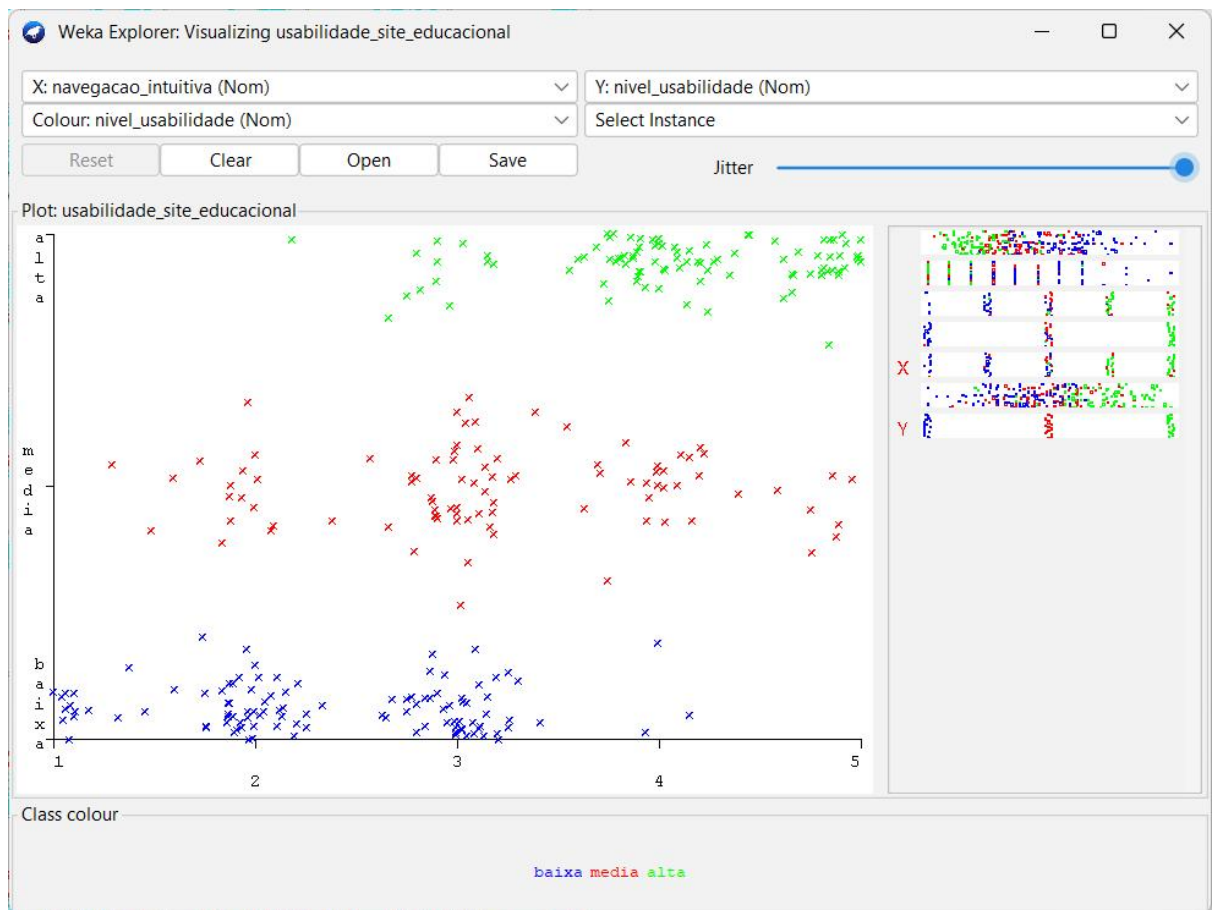


Figura 6: Relação entre *navegacao_intuitiva* e a classe *nivel_usabilidade*.

3.1.4. Atributo *experiencia_previa*

Em contraste com os demais, o atributo *experiencia_previa* se mostrou o preditor visualmente mais fraco. As cores que representam as classes estão misturadas em todas as categorias de experiência, indicando que, isoladamente, este atributo não contribui significativamente para a distinção do nível de usabilidade.



Figura 7: Dispersão do atributo **experiencia_previa** em relação à classe.

3.1.5. Análise da Relação entre Atributos Preditores

Adicionalmente, a Matriz de Dispersão permite observar a correlação entre os próprios atributos. A relação mais relevante para o domínio de IHC é entre **tempo_tarefa_seg** e **erros**.



Figura 8: Relação de correlação positiva entre o número de **erros** e o **tempo_tarefa_seg**.

O gráfico acima demonstra uma clara **correlação positiva**: quanto mais erros um usuário comete, mais tempo ele tende a levar na tarefa. Este padrão é esperado, pois erros demandam tempo para serem corrigidos. Essa alta correlação também sugere uma **redundância informacional** parcial entre os dois atributos, o que pode influenciar como os modelos de aprendizado de máquina os utilizam.

3.2. Conclusão da Análise Visual

A exploração visual dos dados permitiu concluir que os atributos **eficiencia**, **tempo_tarefa_seg**, **erros**, **satisfacao** e **navegacao_intuitiva** são preditores muito fortes e devem ser altamente relevantes para os algoritmos de classificação. A clara separação visual das classes sugere que os modelos terão alta probabilidade de alcançar um excelente desempenho na classificação do nível de usabilidade.

3.3. Treinamento e Avaliação dos Classificadores

Seguindo a análise visual, foram treinados e avaliados cinco algoritmos de classificação distintos, utilizando o método de avaliação *hold-out* com 66% dos dados para treino e 34% para teste, conforme especificado no trabalho. Os algoritmos testados foram: ZeroR e OneR (como *baselines*), e Naive Bayes, J48 (Árvore de Decisão) e IBK (k-NN) como modelos principais.

ZeroR

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      110           36.6667 %
Incorrectly Classified Instances    190           63.3333 %
Kappa statistic                     0
Mean absolute error                  0.4432
Root mean squared error              0.4708
Relative absolute error              100           %
Root relative squared error          100           %
Total Number of Instances           300

=== Detailed Accuracy By Class ===
```

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-----|----------|----------|-------|
| | 1,000 | 1,000 | 0,367 | 1,000 | 0,537 | ? | 0,451 | 0,344 | baixa |
| | 0,000 | 0,000 | ? | 0,000 | ? | ? | 0,449 | 0,286 | media |
| | 0,000 | 0,000 | ? | 0,000 | ? | ? | 0,467 | 0,317 | alta |
| Weighted Avg. | 0,367 | 0,367 | ? | 0,367 | ? | ? | 0,456 | 0,318 | |

```
=== Confusion Matrix ===

  a  b  c  <-- classified as
110  0  0 |  a = baixa
 91  0  0 |  b = media
 99  0  0 |  c = alta
```

OneR

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      252           84      %
Incorrectly Classified Instances    48           16      %
Kappa statistic                     0.7611
Mean absolute error                 0.1067
Root mean squared error            0.3266
Relative absolute error             24.0662 %
Root relative squared error        69.3696 %
Total Number of Instances         300

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0,827   0,021   0,958      0,827   0,888      0,835    0,903     0,856     baixa
                0,945   0,206   0,667      0,945   0,782      0,686    0,870     0,647     media
                0,758   0,005   0,987      0,758   0,857      0,814    0,876     0,828     alta
Weighted Avg.   0,840   0,072   0,879      0,840   0,846      0,783    0,884     0,783

=== Confusion Matrix ===

  a  b  c  <-- classified as
91 19  0 | a = baixa
 4 86  1 | b = media
 0 24 75 | c = alta
```

Naive Bayes

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      262           87.3333 %
Incorrectly Classified Instances    38           12.6667 %
Kappa statistic                     0.8103
Mean absolute error                 0.0924
Root mean squared error            0.2509
Relative absolute error             20.8582 %
Root relative squared error        53.2818 %
Total Number of Instances         300

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0,873   0,032   0,941      0,873   0,906      0,856    0,983     0,976     baixa
                0,912   0,144   0,735      0,912   0,814      0,729    0,944     0,807     media
                0,838   0,010   0,976      0,838   0,902      0,864    0,984     0,977     alta
Weighted Avg.   0,873   0,058   0,890      0,873   0,877      0,820    0,971     0,925

=== Confusion Matrix ===

  a  b  c  <-- classified as
96 14  0 | a = baixa
 6 83  2 | b = media
 0 16 83 | c = alta
```

J48

=== Summary ===

| | | |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances | 274 | 91.3333 % |
| Incorrectly Classified Instances | 26 | 8.6667 % |
| Kappa statistic | 0.8694 | |
| Mean absolute error | 0.0781 | |
| Root mean squared error | 0.2342 | |
| Relative absolute error | 17.6125 % | |
| Root relative squared error | 49.7501 % | |
| Total Number of Instances | 300 | |

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|-------|
| | 0,982 | 0,042 | 0,931 | 0,982 | 0,956 | 0,930 | 0,964 | 0,897 | baixa |
| | 0,846 | 0,057 | 0,865 | 0,846 | 0,856 | 0,794 | 0,882 | 0,830 | media |
| | 0,899 | 0,030 | 0,937 | 0,899 | 0,918 | 0,879 | 0,950 | 0,858 | alta |
| Weighted Avg. | 0,913 | 0,043 | 0,913 | 0,913 | 0,913 | 0,872 | 0,934 | 0,864 | |

=== Confusion Matrix ===

| a | b | c | <-- classified as |
|-----|----|----|-------------------|
| 108 | 2 | 0 | a = baixa |
| 8 | 77 | 6 | b = media |
| 0 | 10 | 89 | c = alta |

IBK

| | | |
|----------------------------------|-----------|-----------|
| Correctly Classified Instances | 265 | 88.3333 % |
| Incorrectly Classified Instances | 35 | 11.6667 % |
| Kappa statistic | 0.8242 | |
| Mean absolute error | 0.0815 | |
| Root mean squared error | 0.2775 | |
| Relative absolute error | 18.392 % | |
| Root relative squared error | 58.9415 % | |
| Total Number of Instances | 300 | |

=== Detailed Accuracy By Class ===

| | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|-------|
| | 0,936 | 0,074 | 0,880 | 0,936 | 0,907 | 0,852 | 0,938 | 0,868 | baixa |
| | 0,802 | 0,081 | 0,811 | 0,802 | 0,807 | 0,723 | 0,870 | 0,714 | media |
| | 0,899 | 0,020 | 0,957 | 0,899 | 0,927 | 0,894 | 0,936 | 0,894 | alta |
| Weighted Avg. | 0,883 | 0,058 | 0,885 | 0,883 | 0,883 | 0,827 | 0,917 | 0,830 | |

=== Confusion Matrix ===

| a | b | c | <-- classified as |
|-----|----|----|-------------------|
| 103 | 7 | 0 | a = baixa |
| 14 | 73 | 4 | b = media |
| 0 | 10 | 89 | c = alta |

A tabela abaixo resume o desempenho geral de cada algoritmo em termos de Acurácia e estatística Kappa, que mede a concordância do modelo em relação ao acaso.

| Algoritmo | Acurácia (%) | Kappa | Interpretação Geral |
|-------------------------|--------------|---------------|--|
| ZeroR | 36,67% | 0,0 | Modelo-base que serve apenas como referência mínima de desempenho. |
| OneR | 84,00% | 0,7611 | Bom desempenho, indicando que regras simples já capturam grande parte dos padrões. |
| Naive Bayes | 87,33% | 0,8103 | Modelo probabilístico eficiente com performance consistente. |
| J48 (Árvore de Decisão) | 91,33% | 0,8694 | Melhor desempenho , com alto poder de generalização e baixa taxa de erro. |
| IBK (KNN) | 88,33% | 0,8242 | Bom desempenho, porém ligeiramente inferior ao J48 e mais sensível a ruídos. |

Tabela 1: Comparação Geral do Desempenho dos Algoritmos

Como indicado na tabela, o algoritmo **J48 (Árvore de Decisão)** apresentou o **melhor desempenho geral**, com 91,33% de instâncias corretamente classificadas.

3.4. Análise das Matrizes de Confusão

Para entender em detalhe onde cada modelo acertou e errou, foram analisadas as matrizes de confusão.

- **ZeroR:** O modelo classificou todas as 300 instâncias como 'baixa', a classe majoritária. Isso explica sua acurácia de 36,67%, que corresponde exatamente à proporção de instâncias da classe 'baixa' no dataset.
- **OneR:** Apresentou bom desempenho, mas confundiu principalmente as classes 'media' e 'alta', que são mais próximas entre si.

- **Regra 2:** Se a **experiencia_previa** é '**alta**', o modelo classifica o **nivel_usabilidade** como '**alta**'. Também faz sentido, pois usuários experientes são mais eficientes.
- **Regra 3:** Para usuários com **experiencia_previa** '**media**', a árvore utiliza outros atributos para refinar a decisão. Fatores como **navegacao_intuitiva**, **satisfacao** e **erros** são usados em sequência para determinar se a usabilidade é 'baixa', 'media' ou 'alta'.

Conclusão da Análise do J48:

As regras extraídas pelo J48 são **altamente coerentes, interpretáveis e refletem os padrões esperados em testes de usabilidade**. O algoritmo conseguiu "descobrir" com sucesso que a combinação de experiência, navegação intuitiva, alta satisfação e poucos erros resulta em um maior nível de usabilidade, validando tanto a estrutura da base de dados sintética quanto a eficácia do modelo para este problema de IHC.