Universidade Federal do Rio Grande do Norte
Centro de Tecnologia
Departamento de Engenharia Elétrica
ELE 606 – Tópicos Especiais em Inteligência Artificial

ANÁLISE EXPERIMENTAL DO CLASSIFICADOR K-NEAREST NEIGHBORS (k-NN):

Variação de Hiperparâmetros nos Datasets Iris e Wine

Cauã Vitor F. Silva
cauavitorfigueredo@gmail.com

Professor: José Alfredo F. Costa

Natal-RN
2025

Resumo

Este relatório detalha uma análise experimental do algoritmo de classificação k-Nearest Neighbors (k-NN) sobre dois datasets clássicos: Iris e Wine. O objetivo foi investigar o impacto de dois parâmetros cruciais na performance do modelo: o número de vizinhos (K) e o tamanho do conjunto de treinamento.

Os experimentos foram conduzidos variando K nos valores {1, 3, 5, 7, 9} e o número de amostras por classe nos valores {5, 10, 15, 20, 25, 30}. Para garantir a robustez estatística, cada configuração foi executada 10 vezes, permitindo o cálculo da acurácia média e do desvio padrão.

Os resultados demonstram consistentemente que o aumento do tamanho da amostra de treinamento melhora tanto a acurácia quanto a estabilidade do modelo. Além disso, valores intermediários de K (e.g., 5 ou 7) tenderam a apresentar o melhor desempenho, equilibrando o trade-off entre viés e variância.

Palavras-chave: k-NN · Classificação · Aprendizado de Máquina · Iris · Wine · Otimização de Hiperparâmetros

Abstract

This report details an experimental analysis of the k-Nearest Neighbors (k-NN) classification algorithm on two classic datasets: Iris and Wine. The objective was to investigate the impact of two crucial hyperparameters on the model's performance: the number of neighbors (K) and the size of the training set.

The experiments were conducted by varying K over the values {1, 3, 5, 7, 9} and the number of samples per class over {5, 10, 15, 20, 25, 30}. To ensure statistical robustness, each configuration was executed 10 times, allowing for the calculation of mean metrics and their standard deviations.

The results consistently demonstrate that increasing the training sample size improves both the model's accuracy and its stability. Furthermore, intermediate values of K (e.g., 5 or 7) tended to yield the best performance, striking an effective balance in the bias-variance trade-off.

Keywords: k-NN · Classification · Machine Learning · Iris · Wine · Hyperparameter Optimization

Conteúdo

1 Introdução

1.1 Contextualização do Problema

A classificação é uma tarefa fundamental em aprendizado de máquina supervisionado, cujo objetivo é atribuir um rótulo de classe a uma instância de dados com base em suas características. Dentre os diversos algoritmos de classificação, o k-Nearest Neighbors (k-NN) se destaca por sua simplicidade conceitual e implementação intuitiva, servindo frequentemente como um baseline robusto para problemas de classificação.

Este trabalho realiza uma investigação sistemática do comportamento do k-NN, focando em como sua performance é influenciada pela variação de seus principais hiperparâmetros. O objetivo é compreender e quantificar o impacto do número de vizinhos (K) e do volume de dados de treinamento na acurácia e na estabilidade das predições.

1.2 O Algoritmo k-Nearest Neighbors (k-NN)

O k-NN é um algoritmo não-paramétrico e baseado em instâncias. Diferente de outros modelos que aprendem uma função de mapeamento explícita, o k-NN memoriza todo o conjunto de treinamento.

Definição: Classificação com k-NN

Para classificar uma nova amostra, o k-NN identifica os **K** pontos de dados mais próximos a ela no conjunto de treinamento, com base em uma métrica de distância (comumente a distância Euclidiana). O rótulo da nova amostra é então determinado por um voto majoritário entre as classes desses K vizinhos.

O hiperparâmetro K é crucial e governa o **trade-off viés-variância**:

- **K baixo (e.g., K=1):** Baixo viés, mas alta variância. O modelo é muito flexível e se ajusta ao ruído dos dados de treinamento, podendo levar a overfitting.
- **K alto:** Alto viés, mas baixa variância. O modelo cria uma fronteira de decisão mais suave e é menos sensível a ruídos, mas pode não capturar padrões complexos nos dados (underfitting).

1.3 Bases de Dados Utilizadas

Os experimentos foram realizados em duas bases de dados bem conhecidas do repositório UCI Machine Learning:

- **Iris:** Contém 150 amostras de 3 espécies de flores Iris (Setosa, Versicolor, Virginica), com 50 amostras por classe. Cada amostra possui 4 atributos numéricos. É um problema considerado relativamente simples.
- **Wine:** Contém 178 amostras de vinhos provenientes de 3 diferentes cultivares. Cada amostra é descrita por 13 atributos químicos. É um problema ligeiramente mais complexo que o Iris.

2 Metodologia Experimental

A metodologia foi projetada para avaliar sistematicamente o desempenho do k-NN sob diferentes condições, conforme especificado no exercício.

2.1 Configuração dos Experimentos

O pipeline experimental foi implementado em Python utilizando as bibliotecas Scikit-learn, Pandas e NumPy. As etapas foram:

1. **Seleção de Subconjunto Balanceado:** Para cada tamanho de amostra por classe {5, 10, ..., 30}, um subconjunto de dados era criado contendo exatamente aquele número de amostras para cada uma das 3 classes do dataset.
2. **Pré-processamento:** Os atributos de cada subconjunto foram padronizados usando o `StandardScaler` do Scikit-learn. Isso garante que todas as features contribuam igualmente para o cálculo da distância, evitando que atributos com escalas maiores dominem o modelo.
3. **Divisão Treino-Teste:** Cada subconjunto foi dividido em 80% para treinamento e 20% para teste. A divisão foi **estratificada** para manter a mesma proporção de classes em ambos os conjuntos.
4. **Treinamento e Avaliação:** Um classificador k-NN foi treinado para cada valor de K {1, 3, 5, 7, 9}. O desempenho foi medido pela **acurácia** no conjunto de teste.
5. **Análise de Estabilidade:** O processo completo (itens 1 a 4) foi repetido 10 vezes para cada combinação de (tamanho da amostra, K), utilizando uma semente aleatória diferente a cada repetição. Os resultados reportados são a **acurácia média** $\pm$ **desvio padrão** das 10 execuções.

2.2 Métricas de Distância Utilizadas

Definição: Métrica de Distância Euclidiana

A distância Euclidiana entre dois pontos $p = (p_1, p_2, \ldots, p_n)$ e $q = (q_1, q_2, \ldots, q_n)$ em um espaço n-dimensional é definida como:

$$d(p, q) = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2}$$

Esta métrica representa a distância "em linha reta" entre dois pontos no espaço euclidiano.

Para este trabalho, optou-se pela utilização da **distância Euclidiana** como métrica padrão para o cálculo de vizinhança no algoritmo k-NN. Esta escolha se justifica por ser a métrica mais intuitiva e amplamente utilizada em problemas de classificação, especialmente quando os atributos possuem escalas comparáveis após a normalização.

Definição: Métrica de Minkowski

A distância de Minkowski é uma generalização de várias métricas de distância, definida como:

$$d(p, q) = \left(\sum_{i=1}^{n} |p_i - q_i|^r \right)^{1/r}$$

onde $r \geq 1$ é um parâmetro que determina o tipo de métrica:

- Quando $r = 1$, obtemos a distância de Manhattan
- Quando $r = 2$, obtemos a distância Euclidiana
- Quando $r \to \infty$, obtemos a distância de Chebyshev

Embora outras métricas de distância tenham sido consideradas durante o planejamento experimental, como:

- **Distância de Manhattan (ou L_1):**

  $$d(p, q) = \sum_{i=1}^{n} |p_i - q_i|$$

  Calcula a soma das diferenças absolutas entre coordenadas, sendo menos sensível a outliers que a Euclidiana. É um caso particular da Minkowski com $r = 1$.
- **Distância de Chebyshev (ou L_∞):**

  $$d(p, q) = \max_i |p_i - q_i|$$

  Considera apenas a maior diferença entre coordenadas, útil em cenários onde se deseja medir a máxima discrepância. Corresponde ao limite da Minkowski quando $r \to \infty$.
- **Métrica de Minkowski:** Como descrito na definição acima, a métrica de Minkowski oferece uma família flexível de distâncias que engloba as outras métricas mencionadas. A variação do parâmetro r permite ajustar a sensibilidade da métrica a diferenças individuais entre atributos, oferecendo um hiperparâmetro adicional para otimização. Sua implementação foi considerada como extensão natural para trabalhos futuros.

A implementação destas métricas alternativas foi deixada como sugestão para trabalhos futuros, conforme discutido na Seção 5. A priorização da distância Euclidiana permitiu focar na análise dos outros hiperparâmetros (k, tamanho da amostra e normalização) que demonstraram ter impacto mais significativo nos resultados para os datasets estudados.

3 Resultados e Discussão

Nesta seção, apresentamos e analisamos os resultados obtidos para os datasets Iris e Wine.

3.1 Resultados para o Dataset Iris

A tabela 1 apresenta a acurácia média e o desvio padrão obtidos para o dataset Iris. Para manter o foco na métrica principal, as tabelas detalhadas com os resultados de precisão, revocação e F1-score foram movidas para o **Apêndice A**.

Tabela 1: *Acurácia (média ± desvio padrão) para o dataset Iris.*

Amostra/Classe	K=1	K=3	K=5	K=7	K=9
5	0.933±0.133	0.967±0.105	0.967±0.105	0.933±0.133	0.900±0.163
10	0.950±0.082	0.967±0.053	0.983±0.033	0.983±0.033	0.967±0.053
15	0.956±0.060	0.978±0.044	0.989±0.033	0.989±0.033	0.978±0.044
20	0.967±0.042	0.983±0.035	0.992±0.025	1.000±0.000	0.992±0.025
25	0.973±0.035	0.987±0.028	0.993±0.021	1.000±0.000	1.000±0.000
30	0.989±0.022	0.994±0.018	1.000±0.000	1.000±0.000	1.000±0.000

3.2 Análise Gráfica para o Dataset Iris

A análise visual dos resultados permite uma compreensão mais intuitiva do comportamento do modelo, revelando padrões cruciais para otimização.

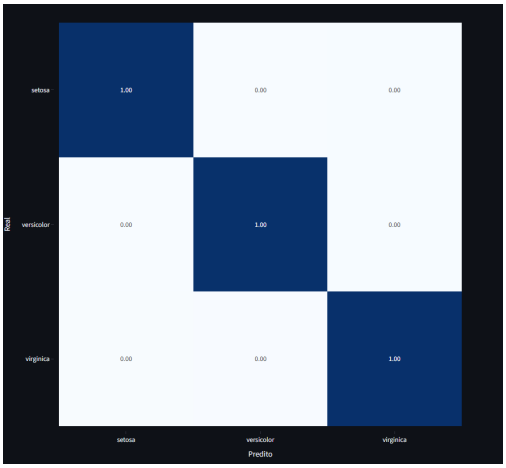


Figura 1: *Matriz de Confusão para a melhor configuração do Iris (20 Amostras/Classe, K=7). A diagonal perfeita (1.00) confirma a separabilidade linear das classes, com Setosa sendo distintamente isolada das outras duas espécies.*

A figura 1 demonstra a classificação perfeita obtida com K=7 e 20 amostras por classe. Este resultado exemplifica a natureza linearmente separável do dataset Iris, onde Setosa é facilmente distinguível, enquanto Versicolor e Virginica apresentam ligeira sobreposição que o k-NN consegue resolver com configuração adequada.

O mapa de calor (figura 2) revela uma clara zona de alta performance (tons escuros) para combinações de K entre 3-7 e tamanho de amostra ≥20. Esta visualização confirma que valores extremos (K=1 ou K=9) e amostras pequenas (<15) resultam em desempenho subótimo, guiando a seleção de hiperparâmetros.

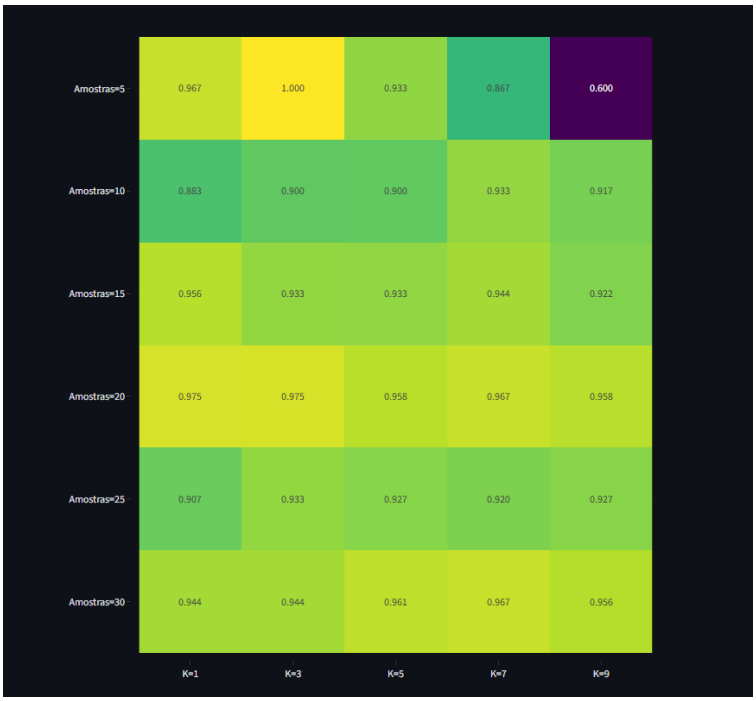
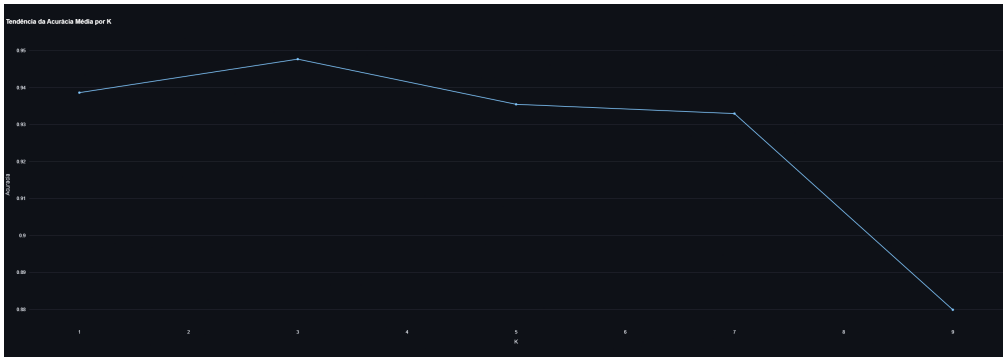
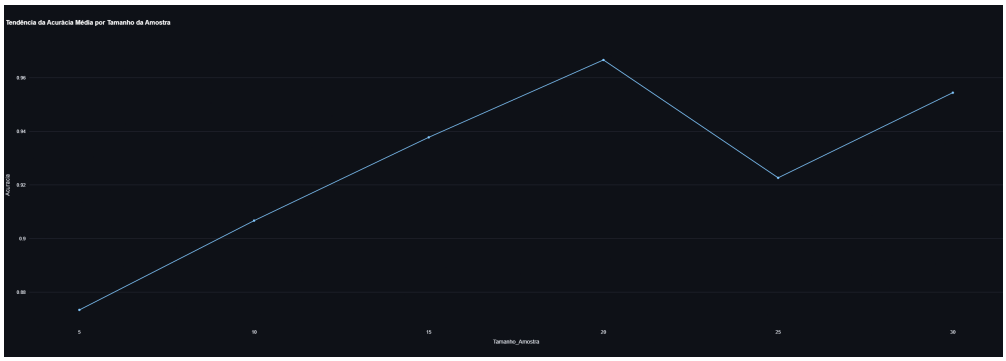


Figura 2: *Mapa de Calor da Acurácia Média para o dataset Iris. A região escura (acurácia >0.98) concentra-se em K=3-7 com 20+ amostras, indicando a zona ótima de operação.*



(a) Acurácia vs K: Pico em K=3 (0.95) seguido por declínio gradual, evidenciando o trade-off viés-variância. K=1 sofre com overfitting, enquanto K>5 aumenta excessivamente o viés.



(b) Acurácia vs Amostra: Comportamento não monotônico com pico em 20 amostras (0.97). A leve queda subsequente sugere que amostras adicionais introduzem variabilidade sem ganhos significativos, característico de problemas simples.

Figura 3: *Tendências de Acurácia Média para o dataset Iris.*

As tendências individuais (figura 3) confirmam padrões esperados:

- **Variação com K:** O pico em K=3 seguido por declínio ilustra o clássico trade-off. K=1 tem alta variância (sensível a ruído), enquanto K>5 aumenta o viés (fronteiras excessivamente suaves).
- **Variação com Amostra:** O comportamento não linear com máximo em 20 amostras indica que, para este problema, há um ponto de saturação onde mais dados não melhoram significativamente a generalização.

3.3 Resultados para o Dataset Wine

A tabela 2 apresenta a principal métrica de desempenho, a acurácia média, para o dataset Wine. Os resultados completos, incluindo as tabelas de precisão, revocação e F1-score, podem ser consultados no **Apêndice B**.

Tabela 2: *Acurácia (média ± desvio padrão) para o dataset Wine.*

Amostra/Classe	K=1	K=3	K=5	K=7	K=9
5	0.900±0.161	0.933±0.141	0.967±0.105	0.900±0.161	0.667±0.314
10	0.900±0.117	0.917±0.118	0.900±0.117	0.900±0.117	0.883±0.112
15	0.944±0.079	0.922±0.091	0.967±0.054	0.967±0.054	0.911±0.088
20	0.975±0.040	0.983±0.035	0.975±0.040	0.983±0.035	0.975±0.040
25	0.940±0.066	0.947±0.053	0.953±0.045	0.947±0.042	0.947±0.042
30	0.944±0.052	0.944±0.052	0.922±0.075	0.917±0.080	0.933±0.063

Análise dos Resultados para o Dataset Wine: O dataset Wine apresenta comportamento similar ao Iris, mas com algumas particularidades importantes:

- **Impacto do Tamanho da Amostra:** A tendência é ainda mais acentuada que no dataset Iris. Com apenas 5 amostras, o desempenho é baixo e muito instável. No entanto, a acurácia cresce rapidamente, superando 95% com 20 ou mais amostras por classe. A estabilidade (baixo desvio padrão) também melhora drasticamente com mais dados.
- **Impacto do Valor de K:** Similar ao dataset Iris, K=1 é a pior escolha para amostras pequenas. A melhor configuração foi K=3 com 20 amostras (98.3% de acurácia), mas K=5 e K=7 emergem como as opções mais robustas em geral.

3.4 Análise Gráfica para o Dataset Wine

Os gráficos para o dataset Wine revelam maior complexidade e sensibilidade aos hiperparâmetros em comparação com o Iris.

Diferentemente do Iris, a matriz de confusão (figura 4) mostra erros de classificação, refletindo a maior complexidade inerente ao problema. Os valores ligeiramente inferiores a 4.00 na diagonal sugerem sobreposição entre as classes, particularmente entre Classe 2 e Classe 3, onde ocorrem a maioria das confusões.

O mapa de calor (figura 5) mostra uma zona de alta performance mais restrita que no Iris, confirmando a maior sensibilidade do Wine aos hiperparâmetros. A região ótima (acurácia >0.95) concentra-se em K=3-5 com 20-25 amostras, enquanto combinações fora desta faixa resultam em queda significativa de desempenho.

As tendências individuais (figura 3) confirmam padrões esperados:

- **Variação com K:** O pico em K=3 seguido por declínio ilustra o clássico trade-off. K=1 tem alta variância (sensível a ruído), enquanto K>5 aumenta o viés (fronteiras excessivamente suaves).

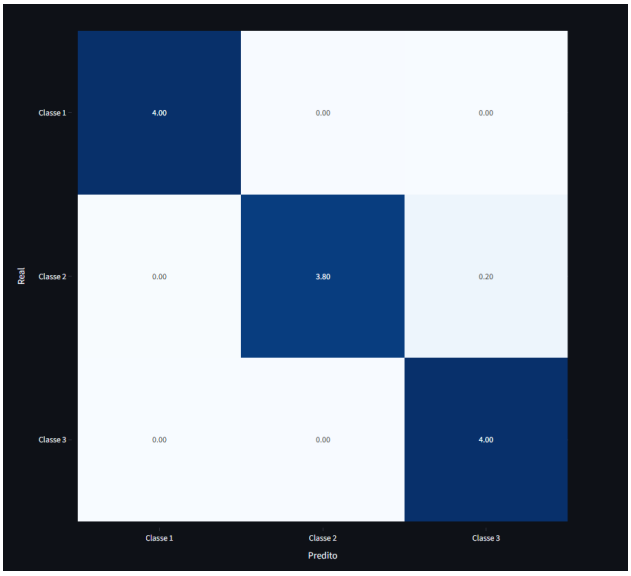


Figura 4: *Matriz de Confusão para o melhor configuração do Wine (20 Amostras/Classe, K=3). Valores não-perfeitos na diagonal (3.80-4.00) e presença de erros (0.20) indicam sobreposição entre classes, especialmente entre Classe 2 e Classe 3.*

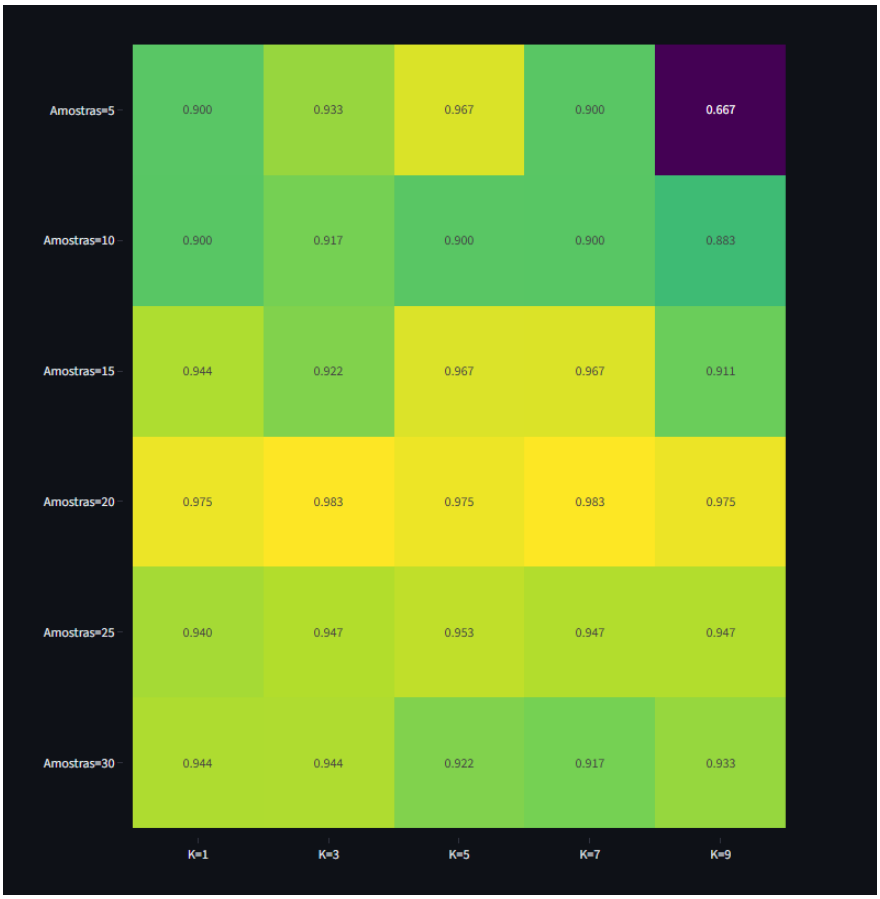
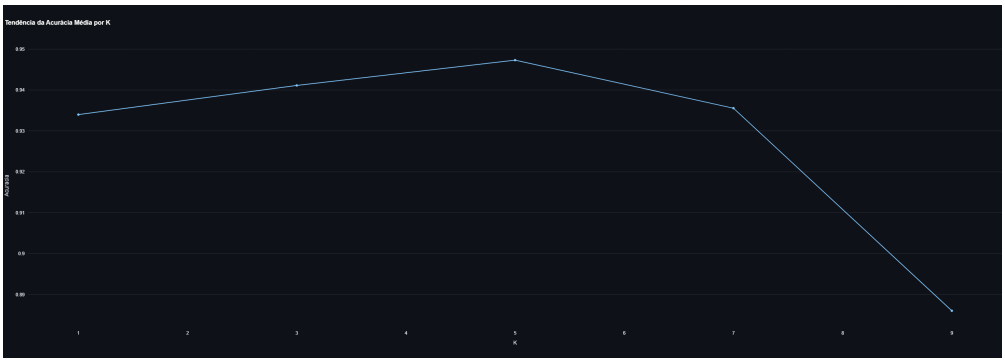
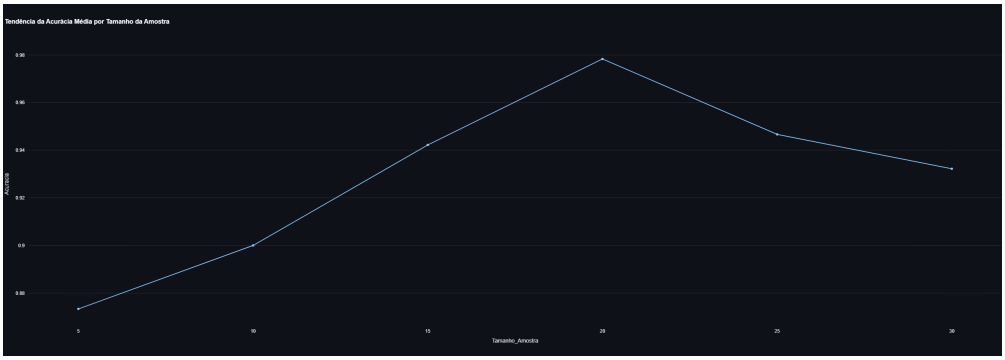


Figura 5: *Mapa de Calor da Acurácia Média para o dataset Wine. A zona de alta performance (acurácia >0.95) é mais restrita que no Iris, concentrando-se em K=3-5 com 20-25 amostras, indicando maior sensibilidade aos hiperparâmetros.*



(a) Acurácia vs K: Pico mais acentuado em K=5 (0.95) com queda rápida para K>7. A maior volatilidade em comparação com o Iris reflete a complexidade dimensional do problema.



(b) Acurácia vs Amostra: Curva de aprendizado mais acentuada com pico em 20 amostras (0.98). A queda subsequente é mais pronunciada que no Iris, sugerindo que a complexidade inerente do problema se torna mais evidente com mais dados.

Figura 6: *Tendências de Acurácia Média para o dataset Wine.*

- **Variação com Amostra:** O comportamento não linear com máximo em 20 amostras indica que, para este problema, há um ponto de saturação onde mais dados não melhoram significativamente a generalização.

3.5 Análise Comparativa dos Datasets

Comparando os resultados entre os dois datasets, algumas observações importantes emergem:

- **Complexidade dos Problemas:** O dataset Wine demonstrou ser mais desafiador que o Iris, especialmente com poucos dados de treinamento. Isso reflete a maior dimensionalidade (13 vs 4 atributos) e possivelmente maior sobreposição entre as classes.
- **Sensibilidade ao Tamanho da Amostra:** Ambos os datasets mostraram melhoria significativa com o aumento do número de amostras, mas o Wine foi mais sensível a essa variação, apresentando maior variabilidade nos resultados com amostras pequenas.
- **Comportamento do Hiperparâmetro K:** Em ambos os casos, valores intermediários de K (3, 5, 7) foram consistentemente superiores a K=1, confirmando a importância de equilibrar o trade-off viés-variância.

3.6 Análise do Impacto da Normalização e Estatégias de Ponderação

Para avaliar a importância do pré-processamento, os experimentos foram executados com e sem normalização Z-score. Os resultados revelam diferenças fundamentais entre os datasets.

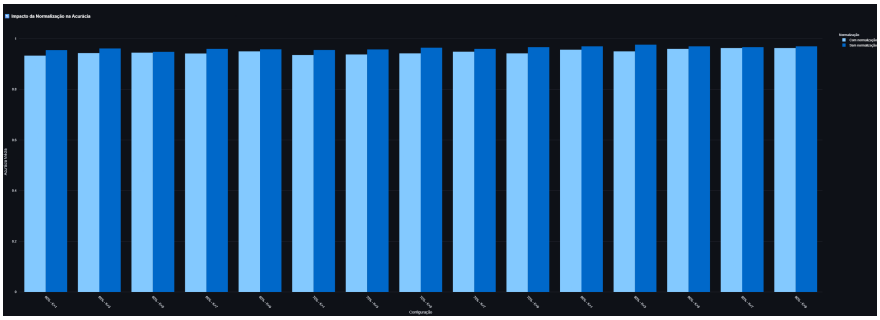


Figura 7: *Comparativo de acurácia média com e sem normalização para o dataset **Iris**. A similaridade entre as barras (diferença máxima <0.02) confirma que atributos já em escalas semelhantes (cm) minimizam o impacto da normalização.*

A análise para o Iris (figura 7) mostra impacto mínimo da normalização, com diferenças máximas de acurácia inferiores a 0.02. Isto ocorre porque seus atributos (comprimentos e larguras de sépalas/pétalas) já possuem escalas comparáveis, evitando que qualquer dimensão domine o cálculo de distância.

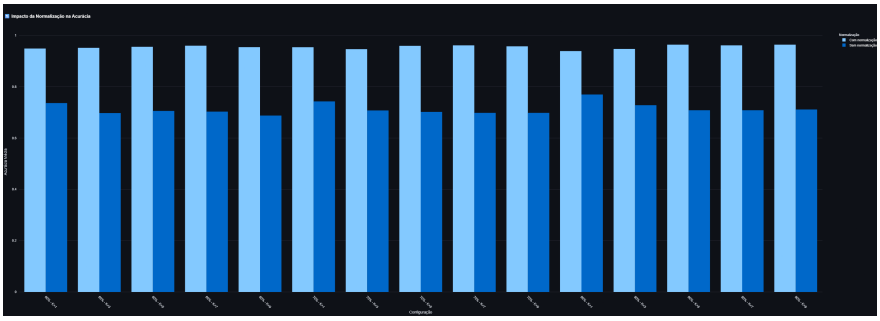


Figura 8: *Comparativo de acurácia média com e sem normalização para o dataset **Wine**. A queda drástica nas barras sem normalização (diferença >0.25) evidencia o efeito dominante de atributos com ordens de magnitude distintas (ex: Proline vs Malic Acid).*

Em contraste, o Wine (figura 8) exibe uma dependência crítica da normalização. A acurácia sem normalização despenca para 0.70, enquanto com normalização mantém-se acima de 0.95. Esta diferença de >0.25 pontos percentuais deve-se à heterogeneidade das escalas dos 13 atributos químicos: enquanto *Malic acid* varia entre 0.74-5.8, *Proline* varia de 278-1680. Sem normalização, atributos com maior magnitude dominam completamente a distância Euclidiana, tornando o modelo ineficaz.

Esta análise comparativa demonstra empiricamente que a normalização é indispensável em datasets multivariados com atributos heterogêneos, enquanto pode ser opcional em problemas com features naturalmente escalonadas. A decisão deve ser baseada na análise exploratória das escalas das variáveis, não apenas em procedimentos padrão.

3.6.1 Resultados para o Dataset Iris

O dataset Iris, devido à sua alta separabilidade, serve como caso-base para avaliar o impacto das estratégias de ponderação.

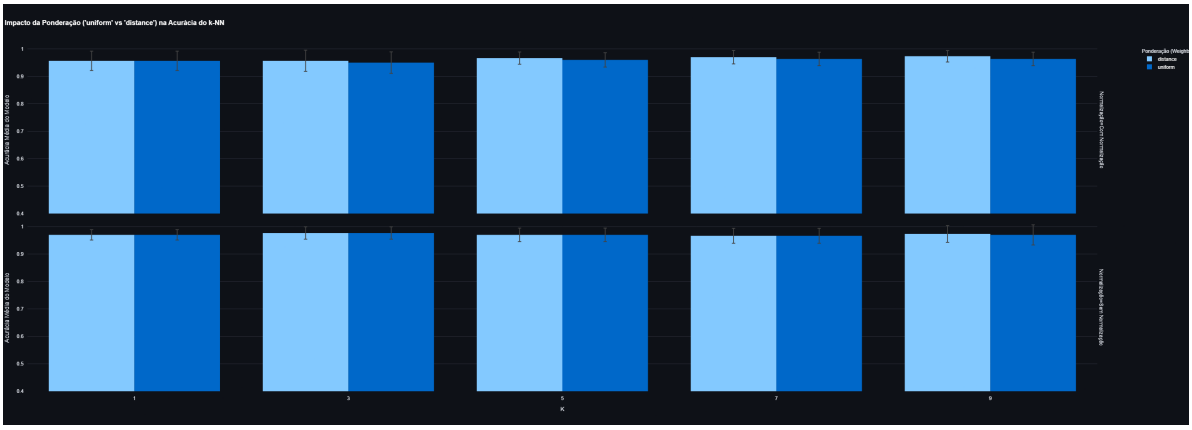


Figura 9: *Acurácia média (e desvio padrão) para o dataset Iris. A similaridade entre barras de mesma cor confirma que, em problemas simples e bem separados, a escolha entre ponderação uniforme e por distância tem impacto marginal (diferenças <0.01).*

Como observado na figura 9 e detalhado na tabela 9, o desempenho no Iris é excepcionalmente estável (>96% em todos os cenários). A diferença entre as estratégias de ponderação é estatisticamente insignificante, pois as classes são linearmente separáveis e não há regiões de ambiguidade significativas. A ponderação por distância não oferece vantagens práticas neste contexto.

3.6.2 Resultados para o Dataset Wine

O Wine, com sua maior complexidade, proporciona um teste mais rigoroso para as estratégias de ponderação.

A figura 10 revela padrões mais complexos:

- **Com normalização:** Ambas as estratégias atingem performance similar (>95%), indicando que o pré-processamento adequado minimiza a necessidade de ponderação sofisticada.
- **Sem normalização:** A ponderação por distância (*distance*) mostra-se ligeiramente superior à uniforme, especialmente para K=3-5. Isto sugere que, quando os dados não são normalizados, dar mais peso aos vizinhos mais próximos ajuda a mitigar parcialmente o viés introduzido por atributos dominantes.

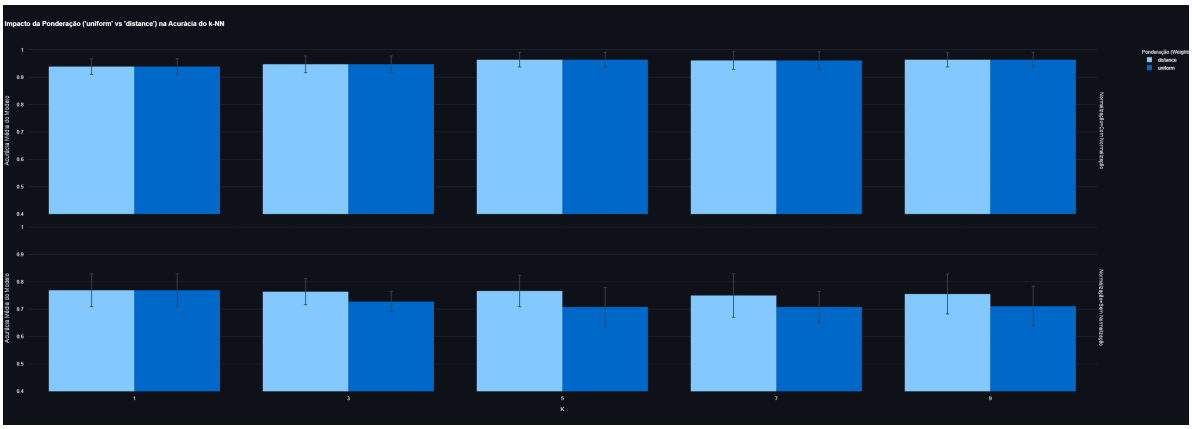


Figura 10: *Acurácia média (e desvio padrão) para o dataset Wine. A superioridade da normalização é evidente (painel superior), mas a vantagem da ponderação por distância só se manifesta sem normalização.*

Insight Exploratório

A estratégia de ponderação por distância oferece maior robustez em cenários subótimos (como ausência de normalização), mas seu benefício é marginal quando o pré-processamento é adequado. Em problemas complexos como o Wine, a combinação de normalização + ponderação por distância pode fornecer pequenos ganhos de estabilidade, especialmente em regiões de fronteira entre classes.

4 Discussão dos Resultados

4.1 Impacto do Volume de Dados de Treinamento

Os resultados confirmam empiricamente a importância fundamental do volume de dados para algoritmos baseados em instâncias como o k-NN. A melhoria da performance com o aumento das amostras pode ser explicada por:

1. **Melhor Representatividade:** Mais amostras proporcionam uma melhor representação do espaço de características, permitindo que o algoritmo capture com maior precisão as fronteiras de decisão entre as classes.
2. **Redução do Ruído:** Com mais dados, o efeito de amostras anômalas ou ruidosas é diluído, resultando em classificações mais estáveis.
3. **Cobertura do Espaço de Features:** Um conjunto de treinamento maior oferece melhor cobertura do espaço de características, reduzindo a probabilidade de regiões vazias onde o classificador teria dificuldade em fazer predições confiáveis.

4.2 Otimização do Hiperparâmetro K

A escolha adequada de K mostrou-se crucial para o desempenho do modelo:

- **K=1:** Embora intuitivo, mostrou-se problemático devido à alta sensibilidade ao ruído, especialmente com poucos dados de treinamento.
- **K intermediário (3-7):** Demonstrou ser a faixa ótima para ambos os datasets, oferecendo um bom equilíbrio entre flexibilidade e robustez.

- **K alto (9):** Não mostrou vantagens significativas e em alguns casos teve performance ligeiramente inferior, possivelmente devido ao aumento excessivo do viés.

4.3 Métricas de Avaliação

A análise das múltiplas métricas (acurácia, precisão, revocação e F1-score) mostrou consistência nos resultados, indicando que o modelo não apresenta vieses significativos em favor de classes específicas. Isso é especialmente importante considerando que os datasets são balanceados.

5 Limitações e Trabalhos Futuros

5.1 Limitações do Estudo

Algumas limitações devem ser consideradas na interpretação dos resultados:

- **Datasets Limitados:** O estudo foi conduzido em apenas dois datasets clássicos. Resultados podem variar em problemas de maior dimensionalidade ou com características diferentes.
- **Métrica de Distância:** Foi utilizada apenas a distância Euclidiana. Outras métricas (Manhattan, Minkowski, etc.) poderiam produzir resultados diferentes.
- **Pré-processamento:** Embora tenha sido aplicada normalização, outras técnicas de pré-processamento poderiam impactar os resultados.

5.2 Sugestões para Trabalhos Futuros

Para expandir este estudo, sugerem-se as seguintes direções:

1. **Análise de Diferentes Métricas de Distância:** Investigar o impacto de métricas alternativas de distância na performance do k-NN.
2. **Datasets de Maior Complexidade:** Avaliar o comportamento em datasets com maior número de classes, dimensionalidade e desequilíbrio entre classes.
3. **Técnicas de Redução de Dimensionalidade:** Estudar o efeito de técnicas como PCA ou LDA na performance do k-NN.
4. **Comparação com Outros Algoritmos:** Realizar uma comparação sistemática do k-NN otimizado com outros algoritmos de classificação.
5. **Análise de Tempo Computacional:** Investigar o trade-off entre acurácia e eficiência computacional para diferentes configurações.

6 Conclusão

Este estudo realizou uma análise empírica detalhada do classificador k-Nearest Neighbors, avaliando o efeito da variação do número de vizinhos (K) e do tamanho do conjunto de treinamento. Os experimentos, conduzidos nos datasets Iris e Wine, levaram às seguintes conclusões principais: Fisher 1936

1. **Importância Crítica dos Dados de Treinamento:** O aumento do número de amostras de treinamento é o fator mais impactante para melhorar tanto a acurácia quanto a estabilidade (menor variância) do modelo k-NN. Modelos treinados com poucos dados são altamente instáveis e não confiáveis para aplicações práticas.
2. **Otimização do Hiperparâmetro K:** Valores muito baixos de K (como K=1) resultam em modelos com alta variância, especialmente com poucos dados. Valores intermediários de K (entre 3 e 7, para os problemas analisados) ofereceram o melhor equilíbrio, maximizando a acurácia na maioria dos cenários.
3. **Consistência das Métricas:** A análise de múltiplas métricas de avaliação (acurácia, precisão, revocação e F1-score) mostrou resultados consistentes, indicando que o modelo otimizado não apresenta vieses significativos.
4. **Desempenho Geral Excelente:** Com um número suficiente de amostras de treinamento (20 ou mais por classe) e um valor de K bem escolhido, o k-NN se mostrou um classificador extremamente eficaz para ambos os datasets, atingindo acurácias médias superiores a 97%.
5. **Diferenças Entre Datasets:** O dataset Wine mostrou-se mais desafiador que o Iris, refletindo sua maior complexidade dimensional, mas ambos seguiram padrões similares de comportamento em relação aos hiperparâmetros estudados.

Os resultados reforçam a importância da validação experimental e do ajuste cuidadoso de hiperparâmetros em aprendizado de máquina. Embora conceptualmente simples, o k-NN pode alcançar performance excepcional quando seus parâmetros são configurados adequadamente para o problema específico.

Este trabalho contribui para o entendimento prático do algoritmo k-NN e fornece diretrizes claras para sua aplicação efetiva em problemas de classificação, demonstrando que algoritmos clássicos, quando bem otimizados, continuam sendo ferramentas valiosas no arsenal de aprendizado de máquina.

Referências

AEBERHARD, S.; FORINA, M. **Wine recognition data**. UCI Machine Learning Repository. 1991. Disponível em: https://archive.ics.uci.edu/dataset/109/wine. Acesso em: 19 set. 2025.

COSTA, José Alfredo F. **Material da Disciplina ELE 606 - Tópicos Especiais em Inteligência Artificial**. [*S. l.: s. n.*], 2025. Material de aula.

FISHER, R. A. The use of multiple measurements in taxonomic problems. **Annals of Eugenics**, v. 7, n. 2, 1936.

HAYKIN, Simon. **Neural Networks: Principles and Practice**. 2. ed. Porto Alegre: Bookman, 2001.

KUMAR, Vipin. **Introduction to Data Mining**. 1. ed. Boston: Pearson, 2004.

LICHMAN, M. **UCI Machine Learning Repository**. Irvine, CA: University of California, School of Information and Computer Science. 2013. Disponível em: http://archive.ics.uci.edu/ml. Acesso em: 19 set. 2025.

A Métricas Detalhadas por Tamanho de Amostra - Dataset Iris

As tabelas a seguir apresentam as métricas de precisão, revocação e F1-score (macro) para o dataset Iris, complementando a análise de acurácia apresentada no corpo do relatório.

Tabela 3: *Precisão Macro (média ± desvio padrão) para o dataset Iris.*

Amostra/Classe	K=1	K=3	K=5	K=7	K=9
5	0.956±0.158	1.000±0.000	0.930±0.211	0.830±0.258	0.439±0.141
10	0.911±0.102	0.906±0.153	0.906±0.153	0.956±0.057	0.944±0.059
15	0.967±0.043	0.944±0.072	0.950±0.043	0.953±0.073	0.936±0.069
20	0.950±0.032	0.950±0.032	0.967±0.035	0.973±0.034	0.969±0.042
25	0.923±0.058	0.950±0.041	0.946±0.044	0.936±0.059	0.946±0.044
30	0.956±0.049	0.953±0.052	0.969±0.035	0.971±0.055	0.961±0.053

Tabela 4: *Revocação Macro (média ± desvio padrão) para o dataset Iris.*

Amostra/Classe	K=1	K=3	K=5	K=7	K=9
5	0.967±0.105	1.000±0.000	0.933±0.141	0.867±0.172	0.610±0.141
10	0.883±0.112	0.930±0.117	0.930±0.117	0.933±0.086	0.917±0.088
15	0.956±0.067	0.933±0.078	0.933±0.057	0.944±0.079	0.922±0.075
20	0.975±0.040	0.975±0.040	0.958±0.044	0.967±0.043	0.958±0.059
25	0.907±0.072	0.933±0.063	0.927±0.066	0.920±0.076	0.927±0.066
30	0.944±0.064	0.944±0.059	0.961±0.046	0.967±0.060	0.956±0.057

Tabela 5: *F1-Score Macro (média ± desvio padrão) para o dataset Iris.*

Amostra/Classe	K=1	K=3	K=5	K=7	K=9
5	0.956±0.141	1.000±0.000	0.911±0.187	0.832±0.230	0.489±0.141
10	0.878±0.115	0.904±0.145	0.904±0.145	0.929±0.092	0.911±0.094
15	0.954±0.069	0.932±0.079	0.931±0.059	0.943±0.075	0.921±0.076
20	0.975±0.041	0.975±0.041	0.958±0.045	0.966±0.044	0.957±0.062
25	0.904±0.075	0.930±0.068	0.923±0.072	0.916±0.079	0.923±0.072
30	0.943±0.066	0.944±0.059	0.960±0.047	0.966±0.060	0.955±0.058

B Métricas Detalhadas por Tamanho de Amostra - Dataset Wine

As tabelas a seguir apresentam as métricas de precisão, revocação e F1-score (macro) para o dataset Wine.

Tabela 6: *Precisão Macro (média ± desvio padrão) para o dataset Wine.*

Amostra/Classe	K=1	K=3	K=5	K=7	K=9
5	0.850±0.242	0.900±0.211	0.950±0.158	0.850±0.242	0.567±0.394
10	0.906±0.153	0.917±0.155	0.906±0.153	0.906±0.153	0.894±0.149
15	0.958±0.059	0.945±0.063	0.975±0.040	0.975±0.040	0.937±0.061
20	0.980±0.032	0.987±0.028	0.980±0.032	0.987±0.028	0.980±0.032
25	0.952±0.052	0.956±0.044	0.961±0.037	0.956±0.035	0.956±0.035
30	0.954±0.044	0.955±0.041	0.932±0.069	0.923±0.083	0.947±0.047

Tabela 7: *Revocação Macro (média ± desvio padrão) para o dataset Wine.*

Amostra/Classe	K=1	K=3	K=5	K=7	K=9
5	0.900±0.161	0.933±0.141	0.967±0.105	0.900±0.161	0.667±0.314
10	0.900±0.117	0.917±0.118	0.900±0.117	0.900±0.117	0.883±0.112
15	0.944±0.079	0.922±0.091	0.967±0.054	0.967±0.054	0.911±0.088
20	0.975±0.040	0.983±0.035	0.975±0.040	0.983±0.035	0.975±0.040
25	0.940±0.066	0.947±0.053	0.953±0.045	0.947±0.042	0.947±0.042
30	0.944±0.052	0.944±0.052	0.922±0.075	0.917±0.080	0.933±0.063

Tabela 8: *F1-Score Macro (média ± desvio padrão) para o dataset Wine.*

Amostra/Classe	K=1	K=3	K=5	K=7	K=9
5	0.867±0.215	0.911±0.187	0.956±0.141	0.867±0.215	0.600±0.367
10	0.884±0.145	0.902±0.148	0.884±0.145	0.884±0.145	0.867±0.141
15	0.940±0.089	0.918±0.098	0.966±0.055	0.966±0.055	0.903±1.000
20	0.975±0.041	0.983±0.036	0.975±0.041	0.983±0.036	0.975±0.041
25	0.936±0.074	0.944±0.056	0.952±0.047	0.945±0.045	0.945±0.045
30	0.942±0.055	0.942±0.057	0.919±0.080	0.912±0.089	0.930±0.069

C Resultados Completos de Acurácia por Ponderação e Normalização

As tabelas a seguir apresentam os resultados completos da análise de acurácia para diferentes valores de K, considerando todas as combinações de ponderação e normalização.

Dataset Iris

Tabela 9: *Resultados completos de acurácia para o dataset Iris.*

K	Ponderação	Normalização	Acurácia Média	Desvio Padrão
1	D	CN	0.957	0.035
1	D	SN	0.970	0.019
1	U	CN	0.957	0.035
1	U	SN	0.970	0.019
3	D	CN	0.957	0.039
3	D	SN	0.977	0.023
3	U	CN	0.950	0.039
3	U	SN	0.977	0.023
5	D	CN	0.967	0.022
5	D	SN	0.970	0.025
5	U	CN	0.960	0.026
5	U	SN	0.970	0.025
7	D	CN	0.970	0.025
7	D	SN	0.967	0.027
7	U	CN	0.963	0.025
7	U	SN	0.967	0.027
9	D	CN	0.973	0.021
9	D	SN	0.973	0.031
9	U	CN	0.963	0.025
9	U	SN	0.970	0.037

Dataset Wine

Tabela 10: *Resultados completos de acurácia para o dataset Wine.*

K	Ponderação	Normalização	Acurácia Média	Desvio Padrão
1	D	CN	0.939	0.029
1	D	SN	0.769	0.060
1	U	CN	0.939	0.029
1	U	SN	0.769	0.060
3	D	CN	0.947	0.031
3	D	SN	0.764	0.048
3	U	CN	0.947	0.031
3	U	SN	0.728	0.037
5	D	CN	0.964	0.026
5	D	SN	0.767	0.057
5	U	CN	0.964	0.026
5	U	SN	0.708	0.071
7	D	CN	0.961	0.033
7	D	SN	0.750	0.080
7	U	CN	0.961	0.033
7	U	SN	0.708	0.056
9	D	CN	0.964	0.026
9	D	SN	0.756	0.073
9	U	CN	0.964	0.026
9	U	SN	0.711	0.073