

Análises utilizando ferramentas de IA sobre estoques de camarões

Pedro Petroski Ruiz

¹Faculdade de Computação e Informática (FCI)
Universidade Presbiteriana Mackenzie – São Paulo, SP – Brasil

²Programa de pós-graduação em Computação Aplicada – Faculdade de Computação e Informática (FCI) – Universidade Presbiteriana Mackenzie – São Paulo, SP – Brasil

10390289@mackenzista.com.br

Resumo. A crescente digitalização do setor pesqueiro e aquícola tem possibilitado novas abordagens de análise para previsão e gestão de estoques. Este trabalho investiga a aplicação de modelos de Inteligência Artificial (IA) para classificar e prever padrões em uma base de dados relacionada ao estoque de diferentes tipos de camarões. Foram utilizados três algoritmos de aprendizado de máquina: Naive Bayes, K-Nearest Neighbors (KNN) e Decision Tree. Os resultados demonstraram que o modelo Naive Bayes apresentou o pior desempenho, com acurácia de apenas 44,35%, enquanto os modelos KNN e Decision Tree alcançaram acurácias de 92,22% e 93,33%, respectivamente. Foram analisadas métricas complementares, como precisão, recall, F1-Score e AUC-ROC, para avaliar a robustez dos modelos. Conclui-se que modelos baseados em vizinhança e árvores de decisão são mais adequados para prever comportamentos na base de dados em questão, podendo ser utilizados como apoio na tomada de decisão gerencial.

Palavras-chave: Inteligência Artificial; Aprendizado de Máquina; Estoque de camarões; Classificação; Análise preditiva.

Abstract. The increasing digitalization of the fishing and aquaculture sector has enabled new analytical approaches for inventory forecasting and management. This study investigates the application of Artificial Intelligence (AI) models to classify and predict patterns in a database related to the stock of different shrimp types. Three machine learning algorithms were used: Naive Bayes, K-Nearest Neighbors (KNN), and Decision Tree. The results showed that the Naive Bayes model presented the lowest performance, with an accuracy of only 44.35%, while the KNN and Decision Tree models achieved accuracies of 92.22% and 93.33%, respectively. Additional metrics such as precision, recall, F1-Score and AUC-ROC were analyzed to assess model robustness. It is concluded that neighborhood-based and decision-tree models are more suitable for predicting behaviors in this database and may support managerial decision-making.

Keywords: Artificial Intelligence; Machine Learning; Shrimp stock; Classification; Predictive analysis.

1. Introdução

A análise e gestão de estoques são atividades essenciais nas cadeias produtivas do setor pesqueiro, especialmente quando envolvem produtos altamente perecíveis, como os camarões. A variabilidade de demanda, as diferenças regionais de captura e a diversidade de espécies tornam o processo de gerenciamento complexo e altamente dependente de informações precisas.

Com o avanço das tecnologias de Inteligência Artificial (IA), tornou-se possível aplicar métodos computacionais para identificar padrões, classificar dados e auxiliar na tomada de decisão. Neste contexto, a utilização de algoritmos de aprendizado de máquina pode contribuir para melhorar a precisão das estimativas e ampliar a visão analítica sobre o comportamento dos estoques.

Este trabalho apresenta a aplicação de três modelos de IA — Naive Bayes, K-Nearest Neighbors (KNN) e Decision Tree — para classificar registros de uma base de dados contendo informações sobre tipo de pessoa, estado, município, local de armazenamento e diferentes quantidades de espécies de camarão. O objetivo é avaliar o desempenho dos modelos e identificar aquele que melhor se adapta aos dados, fornecendo subsídios para processos decisórios futuros.

2. Referencial Teórico

2.1 Inteligência Artificial e Aprendizado de Máquina

A Inteligência Artificial é uma área da computação dedicada ao desenvolvimento de algoritmos capazes de executar tarefas que normalmente requerem inteligência humana. Uma de suas subáreas, o Aprendizado de Máquina (Machine Learning), envolve o desenvolvimento de modelos capazes de aprender padrões a partir de dados.

2.2 Classificação

A classificação é uma das tarefas mais comuns em Machine Learning, consistindo na categorização de dados em classes predefinidas. Ela é amplamente utilizada em áreas como detecção de fraudes, diagnóstico médico e previsões de demanda.

2.3 Algoritmos Utilizados

2.3.1 Naive Bayes

O Naive Bayes é um algoritmo probabilístico baseado no Teorema de Bayes. Sua principal característica é assumir independência entre os atributos, o que pode prejudicar o desempenho quando essa premissa não é verdadeira.

2.3.2 K-Nearest Neighbors (KNN)

O KNN classifica as instâncias com base nos exemplos mais próximos. É amplamente utilizado em bases com padrões não lineares, devido à sua simplicidade e boa capacidade de generalização.

2.3.3 Decision Tree

Árvores de decisão utilizam divisões sucessivas dos atributos para classificar instâncias. São interpretáveis, eficientes e funcionam bem em bases heterogêneas e com variáveis não normalizadas.

3. Metodologia

3.1 Base de Dados

- A base utilizada contém registros de estoque de camarões com os seguintes atributos:
- Data de envio
- Tipo de pessoa
- Estado
- Município
- Local de armazenamento
- Quantidades (Kg) de diversas espécies de camarão:
 - Camarão Rosa
 - Camarão Sete Barbas
 - Camarão Branco
 - Camarão Santana ou Vermelho
 - Camarão Barba-ruça
- Apresentação do produto
- Tipo de embalagem

3.2 Pré-processamento

As etapas incluíram:

- limpeza dos dados
- normalização de variáveis numéricas
- codificação de variáveis categóricas (One-Hot Encoding)
- divisão em conjunto de treino e teste

3.3 Modelagem

Foram treinados três modelos:

- Naive Bayes
- KNN
- Decision Tree

As métricas avaliadas foram:

- Acurácia
- Precisão
- Recall
- F1-Score
- AUC-ROC

4. Resultados e discussão

Os resultados principais obtidos foram:

Modelo Naive Bayes

- **Acurácia:** 44,35%
O baixo desempenho indica que o modelo não conseguiu capturar relacionamentos entre as variáveis. Provavelmente isso ocorre porque:
 1. As variáveis apresentam dependências fortes entre si.
 2. A premissa de independência condicional do Naive Bayes não é atendida.
 3. A base mistura dados numéricos e categóricos com alta heterogeneidade.

Modelo KNN

- **Acurácia:** 92,22%

- **Precisão:** 0.9087
- **Recall:** 0.9222
- **F1-Score:** 0.9146
- **AUC-ROC:** 0.6970

O modelo KNN apresentou excelente desempenho. Sua capacidade de capturar relações locais na base o torna adequado para esse tipo de dado, onde as quantidades de camarões e o tipo de apresentação variam entre regiões e situações de armazenamento.

Modelo Decision Tree

- Acurácia: 93,33%

O Decision Tree foi o melhor modelo entre os três, com acurácia superior. Isso se deve ao fato de que árvores conseguem lidar bem com:

- dados categóricos,
- interações complexas entre atributos,
- escalas diferentes,
- padrões não lineares.

5. Conclusão

Este trabalho demonstrou que modelos de Inteligência Artificial podem ser aplicados com sucesso em análises de estoque de camarões. Entre os três modelos testados, o Decision Tree apresentou o melhor desempenho, seguido de perto pelo KNN, ambos com acurácias acima de 92%. O Naive Bayes mostrou-se inadequado para este tipo de conjunto de dados, devido às dependências entre os atributos.

Conclui-se que o uso de IA pode contribuir para o aprimoramento das operações de logística e armazenamento de produtos pesqueiros, permitindo análises mais precisas, identificação de padrões e suporte à tomada de decisões. Trabalhos futuros podem incluir redes neurais, aumento da base de dados, técnicas de balanceamento e modelos de previsão temporal.

6. Referências bibliográficas

O que é KNN (k-nearest neighbors)?