

Classificação do cavalo Pantaneiro com o uso de algoritmo de machine learning KNN

Igor Pinho Souza

Análise e desenvolvimento de sistemas, IFMS, igor1_souza@hotmail.com

Marcel José Soleira Grassi

Professor do IFMS, marcel.grassi@ifms.edu.br

Sandra Aparecida Santos

Pesquisador da Embrapa Pantanal, sandra.santosl@embrapa.br

Cláudio Pereira Flores

Analista da Embrapa Pantanal, claudio.flores@embrapa.br

Adriana Mello de Araujo

Pesquisador da Embrapa Pantanal, adriana.araujo@embrapa.br

“Machine learning” é um ramo da inteligência artificial que é capaz de analisar uma grande quantidade de dados para encontrar padrões. Com base nesses padrões, consegue fazer determinações ou previsões. Sua principal característica é não precisar ter as rotinas implantadas a mão, o próprio sistema tem a habilidade de aprender com a análise de dados e executar tarefas com uma precisão cada vez maior. Por meio do uso de métodos estatísticos, os algoritmos são treinados para fazer previsões ou classificações ajudando no desenvolvimento de sistemas que são complexos, ou que requerem habilidades ou conhecimentos detalhados específicos. A avaliação e classificação de equinos para registros e outros fins é feita por inspetores capacitados pela Associação Brasileira de Criadores de Cavalos Pantaneiros – ABCCP. Para o registro, os inspetores estabelecem alguns critérios do padrão da raça, um deles se baseia na avaliação fenotípica, ou seja, se refere a características externas que podem ser observadas visualmente e medidas. Atualmente, a ABCCP utiliza para o padrão do julgamento e avaliação de desempenho funcional e das aptidões da raça 15 medidas lineares tomadas por seus inspetores credenciados, que dimensionam a altura, comprimento, largura, espessura e ângulos do cavalo. Além disso, eles incluem performance e taxa de conformação. A conformação corporal do cavalo pode ser estudada e analisada por meio do uso de medidas corporais, que abrangem proporções, dimensões e relações entre as diversas partes. Na presente pesquisa, o foco é dado exclusivamente às medidas lineares e algumas relações corporais. Esse trabalho tem o objetivo de buscar o desenvolvimento de um aplicativo a fim de auxiliar os usuários na classificação do porte dos cavalos Pantaneiros como alta, média e baixa qualidade. Nessa pesquisa, foram utilizadas 296 informações da base de dados da ABCCP, sendo 222 para treino e 74 para validar a taxa de acerto. O algoritmo de classificação utilizou o conjunto mínimo de medidas preliminares, que apontou oito medidas lineares do registro, sendo elas altura na cernelha (AC), altura na garupa (AG), altura do dorso (AD), comprimento do corpo (CC), comprimento da espádua (CE), comprimento do dorso lombo (CDL), largura do peito (LP), largura da anca (LA), sendo acrescentado ao modelo estudado o perímetro torácico (Pt), por ser considerado uma medida de extrema importância para avaliação morfofuncional segundo a ABCCP. Foram consideradas no modelo as proporções lineares, que avaliam o balanço das medidas dos cavalos relacionadas a essas medidas (AD/AG), (LA/AD), (AC/CC). O algoritmo classificador k-vizinhos mais próximos (KNN; K-Nearest Neighbors) é capaz de rotular cada elemento de um conjunto de dados, a partir da distância destes em relação aos seus vizinhos mais próximos em um conjunto de dados semelhante treinado previamente. O modelo com nove medidas lineares e suas relações, utilizando o número de vizinhos para o algoritmo KNN como dois (k=2) apresentou o maior valor de precisão (91%), sendo considerado robusto. Mais estudos vão ser necessários para ajustar o algoritmo de classificação e possibilitar uma melhor taxa de acerto para o usuário.

Palavras-chave: medidas lineares, equinos, padrão racial, automatização.

Apoio/financiamento: Projeto SEG/Embrapa 10.20.02.007.00.02.001 e CNPq