

Resultados de competições e informações genealógicas de equinos da raça crioula: uma abordagem baseada em descoberta do conhecimento

Competition results and genealogical information of creole horses: a knowledge discovery-based approach

Resultados de la competencia e información genealógica de caballos criollos: un enfoque basado en el descubrimiento del conocimiento

Recebido: 09/08/2021 | Revisado: 18/08/2021 | Aceito: 23/08/2021 | Publicado: 24/08/2021

Jessica Della Giustina

ORCID: <https://orcid.org/0000-0002-6333-7730>
Universidade de Caxias do Sul, Brasil
E-mail: jdgiustina@ucs.br

Tiago Piccoli

ORCID: <https://orcid.org/0000-0002-8002-8286>
Universidade de Caxias do Sul, Brasil
E-mail: tapiccoli@gmail.com

Fernanda Pessi de Abreu

ORCID: <https://orcid.org/0000-0002-5006-4833>
Universidade de Caxias do Sul, Brasil
E-mail: fpabreu1@ucs.br

Pedro Lenz Casa

ORCID: <https://orcid.org/0000-0002-4666-9434>
Universidade de Caxias do Sul, Brasil
E-mail: plcasa@ucs.br

Fernando Paixão Lisboa

ORCID: <https://orcid.org/0000-0001-5157-1506>
Universidade de Caxias do Sul, Brasil
E-mail: pxlisboa@gmail.com

Scheila de Avila e Silva

ORCID: <https://orcid.org/0000-0002-3472-3907>
Universidade de Caxias do Sul, Brasil
E-mail: sasilva6@ucs.br

Resumo

O presente artigo tem por objetivo aplicar algoritmos de inteligência artificial na mineração de dados genealógicos de cavalos crioulos premiados dentro da Expointer, a fim de encontrar padrões e características de animais vencedores de competições. O conjunto de dados foi obtido do repositório da Associação Brasileira de Criadores de Cavalos Crioulos. As técnicas e algoritmos aplicados foram as análises de frequências, comparação de médias e regressão linear. No total, foram extraídas as informações de 1866 animais. Os dados foram organizados em dois grupos segundo as provas de Morfologia e Freio de Ouro. A tribo paterna que teve o melhor desempenho entre os campeões de Morfologia e Freio de Ouro foi Aculeo Vastago. Em comparação, as famílias maternas bem-sucedidas demonstraram uma maior variabilidade, mesmo assim foi observado uma dominação de Torhuela, Che Pitanga 565 e BT Fuzarca na prova de Morfologia. As análises de comparação de médias revelaram que existe diferença estatística significativa ($p < 0,001$) entre as médias das notas finais em comparação à posição da prova. A análise do conjunto de dados demonstrou grande variabilidade e poucos padrões. Ainda assim, foi possível identificar as linhagens mais recorrentes entre as melhores colocações nas provas de Morfologia e Freio de Ouro. Os achados deste estudo podem auxiliar no melhoramento genético das raças crioulas, possibilitando um melhor desempenho das linhagens nas competições.

Palavras-chave: Descoberta do conhecimento em banco de dados; Associação brasileira de criadores de cavalos crioulos; Pesquisa genealógica.

Abstract

The current paper aims to perform data mining and artificial intelligence algorithms for genealogical data related to creole horses awarded in the “Expointer” competition, in order to uncover patterns and characteristics of winning animals. The dataset was obtained on a repository from the “Associação Brasileira de Criadores de Cavalos Crioulos”. The techniques and algorithms employed were frequency analysis, means comparison and linear regression. In total,

information from 1866 animals was extracted, which were organized in two groups according to the competition events of “Morfologia” and “Freio de Ouro”. The paternal lineage with better performance out of the champions from both events was Aculeo Vastago. In comparison, the more successful maternal branches demonstrated higher variability; Nevertheless, a dominance was observed for the families Torhuela, Che Pitanga 565 and BT Fuzarca in the “Morfologia” event. The mean comparison analyses revealed a statistically significant difference ($p < 0,001$) between the mean final punctuation in comparison to final position in the event. Overall, the analysis of this dataset demonstrated high variability and few patterns. Still, we were able to identify the most recurrent lineages amongst the better positioned animals in the “Morfologia” and “Freio de Ouro” events. The findings contained in this study can be used as support for genetic enhancement of creole breeds, allowing for a better performance in competitions.

Keywords: Knowledge discovery in databases; Brazilian association of creole horse breeders; Genealogical research.

Resumen

Este artículo tiene como objetivo aplicar algoritmos de inteligencia artificial en la minería de datos genealógicos de caballos criollos premiados en la competencia de Expointer, con el fin de encontrar patrones y características de animales que han ganado concursos. El conjunto de datos se obtuvo del repositorio de la Asociación Brasileña de Criadores de Caballos Criollos. Las técnicas y algoritmos aplicados fueron análisis de frecuencia, comparación de medias y regresión lineal. En total, se extrajo la información de 1866 animales. Los datos se organizaron en dos grupos de acuerdo con las pruebas de Morfología y Golden Brake. La tribu paterna que mejor desempeño tuvo entre los campeones de Morfología y Golden Brake fue Aculeo Vastago. En comparación, las familias maternas exitosas mostraron mayor variabilidad, aun así se observó una dominancia de Torhuela, Che Pitanga 565 y BT Fuzarca en la prueba de Morfología. El análisis de comparación de medias reveló que existe una diferencia estadísticamente significativa ($p < 0,001$) entre las medias de las notas finales en comparación con el puesto de prueba. El análisis del conjunto de datos mostró una gran variabilidad y pocos patrones. Aun así, fue posible identificar las cepas más reincidentes entre las mejor colocadas en las pruebas de Morfología y Golden Brake. Los hallazgos de este estudio pueden ayudar en la mejora genética de razas autóctonas, permitiendo un mejor desempeño de las cepas en competiciones.

Palabras clave: Descubrimiento de conocimiento en la base de datos; Asociación brasileña de criadores de caballos criollos; Investigación genealógica.

1. Introdução

A obtenção de conhecimento a partir de bases de dados requer um processo que abrange etapas desde a coleta e armazenamento de dados até o tratamento e interpretação de resultados. Nesse sentido, a Mineração de Dados compõe uma das etapas da descoberta do conhecimento em banco de dados (KDD, do inglês *knowledge discovery in databases*). Com ela é possível revelar o conhecimento implícito em grandes bases de dados, possibilitando a análise antecipada de eventos e a previsão de tendências e comportamentos futuros (Cardoso & Machado, 2008).

Na saúde animal, a disponibilidade de dados vem crescendo devido a informatização de vários processos. Os avanços nas tecnologias de sequenciamento de genomas, sistemas de vigilância e registros médicos eletrônicos aumentaram a quantidade de dados brutos gerados. Além disso, uma forma de auxiliar a aquisição de dados é a utilização de sensores desenvolvidos para animais. Outras fontes de dados na saúde animal incluem a vigilância governamental sobre doenças animais, registros médicos veterinários eletrônicos, registros de produção agrícola e bancos de dados específicos de espécies (Ouyang et al., 2019).

Como exemplo de aplicação do KDD em saúde animal, cita-se a epidemiologia veterinária. Nesse contexto, essas análises podem ser utilizadas para entender os riscos e minimizar o impacto de problemas adversos à saúde dos animais. Por meio da modelagem epidemiológica é possível combinar diferentes dados ou processos, identificando populações de alto risco. Assim, a identificação de padrões auxilia em ações de vigilância e controle. Como o movimento de animais entre locais é um fator de risco essencial para muitas doenças infecciosas, muitos países implementaram programas obrigatórios de rastreabilidade animal. Por meio desses, as remessas de animais em produção são rastreadas no espaço e no tempo, gerando uma rica fonte de informações para uma resposta rápida a ameaças à saúde (VanderWaal et al., 2017; Neethirajan, 2020).

Lima e colaboradores (2013) fizeram uso da base de dados de um programa de controle de incidência da Anemia Infecciosa Equina na região do Pantanal sul-mato-grossense. O principal objetivo da pesquisa foi modelar os resultados do

acompanhamento da Anemia Infecciosa Equina, traçando o perfil do sucesso ou fracasso do controle da doença. Como resultado, os autores obtiveram um modelo baseado em árvore de decisão com acurácia de 85%, sendo esse efetivo na identificação de fatores e características que culminam na infecção dos animais (Lima et al., 2013).

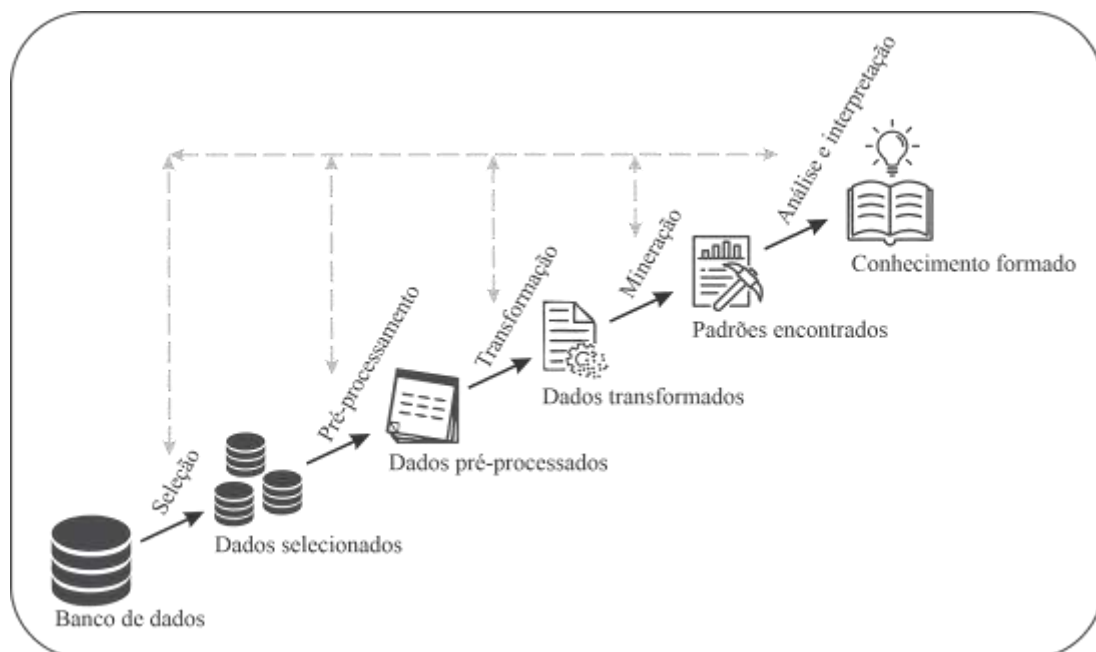
O KDD foi também exemplificado em um estudo realizado pelo programa de melhoramento genético da Universidade Federal de Santa Maria. O trabalho teve como finalidade aprimorar o processo de seleção de poedeiras. Para isso, foram utilizadas técnicas de mineração de dados. A metodologia utilizada estabeleceu quais são as características chave para classificar a produtividade de cada raça de poedeiras de ovos. Assim, tornou-se possível aumentar os ganhos na produção de ovos por intermédio da identificação de características fenotípicas associadas ao ambiente (Ferreira et al., 2013).

Considerando a aplicabilidade do processo de descoberta do conhecimento em dados provenientes de saúde animal, o objetivo do presente trabalho foi aplicar técnicas de mineração de dados em um repositório referente à genealogia de cavalos crioulos premiados na Expointer.

2. Metodologia

Este trabalho baseou-se no processo de KDD, o qual compreende uma série de etapas, entre as quais está a seleção dos dados, pré-processamento, mineração, análise e avaliação (Figura 1) (Fayad; Haussler & Stolorz, 1996; Elmasri & Navathe, 2019). Sendo que presume que a análise da informação seja feita por um especialista de domínio, a fim de gerar conhecimento. O KDD realiza análises através de algoritmos que retornam padrões e tendências nos dados analisados (Silva, 2016; Carvalho et al., 2021).

Figura 1. Etapas do KDD.



Fonte: Adaptado de Fayyad; Piatetsky-Shapiro e Smyth (1996).

A seleção dos dados consistiu na escolha do conjunto de dados. O presente estudo fez uso de dados disponibilizados na base de dados da Associação Brasileira de Criadores de Cavalos Crioulos (ABCCC, 2020). Assim, realizou-se uma pesquisa de mérito dos principais animais da raça crioula premiados nas provas de Morfologia e Freio de Ouro da Expointer, um dos maiores eventos da ABCCC. Vale notar que essas provas são consideradas importantes como subsídio para a seleção e melhoramento (Cucco et al., 2016), visto que podem repercutir diretamente na lucratividade do equino (Gianluppi et al., 2009).

Para a seleção dos dados foram consideradas somente as premiações máximas das provas de seleção da raça (Morfologia). Já nas provas Freio de Ouro foram consideradas as colocações de primeiro até o décimo quarto lugar a partir do ano de 1982. No total, as informações de 1866 animais foram extraídas para o projeto.

O Pré-Processamento envolveu a compreensão de cada variável do conjunto de dados. No Pré-processamento é realizada inicialmente a limpeza e transformação dos elementos, com o objetivo de selecionar e filtrar dados ruidosos, inconsistentes e dados faltantes que podem afetar a qualidade do processo de mineração de dados (Elmasri & Navathe, 2019; Silva, 2016; Goldschmidt; Passos, & Bezerra, 2015). No total, foram considerados 53 atributos para serem trabalhados, dos quais definidos os atributos mais relevantes para a etapa de mineração (Tabela 1).

Tabela 1. Atributos do conjunto de dados e suas classificações.

Atributo	Tipologia
Categoria	Catégorico
Prova	Catégorico
Colocação	Numérico
Ano	Numérico
Sexo	Catégorico
Nome do animal	Catégorico
Genealogia paterna (5 gerações) 2m, 3mm, 4mmm, 5mmmm	Catégorico
Tribo paterna	Catégorico
Linha materna pai	Catégorico
Genealogia materna (5 gerações) 2f, 3ff, 4fff, 5ffff	Catégorico
Família materna	Catégorico
Morfologia	Numérico
Andadura	Numérico
Figura	Numérico
VSP/ESB	Numérico
Mangueira 1 / Mangueira 2	Numérico
Campo 1 / Campo 2	Numérico
Final	Numérico

Fonte: Autores.

Em seguida, foi encaminhada a etapa de Mineração, visto que não foi necessária a etapa de Transformação para o conjunto utilizado. A mineração de dados, se trata de um processo de escavação de conjuntos de dados com grandes volumes, os quais são analisados e deles extraídos informações, padrões, tendências, entre outros (Dani et al., 2020). As ferramentas utilizadas para o processo de mineração de dados foram: Minitab (Minitab 17 Statistical Software, 2010) e Orange Data Mining (Demšar et al., 2013).

As técnicas e algoritmos aplicados foram as análises de frequências, comparação de médias e regressão linear (Demšar et al., 2013). Para os testes inferenciais, todas as premissas foram analisadas a fim de aplicar os testes adequadamente. Em todas as análises, o conjunto de dados foi dividido em dois grupos: (i) um grupo referente à prova de Morfologia, a qual não define pontuação para os animais, apenas uma colocação; e (ii) um grupo relacionado às provas de Freio de Ouro, voltadas para o desempenho dos animais, as quais apresentam pontuações nas provas.

3. Resultados e Discussão

Em relação à análise de frequências, pode-se elencar os animais das famílias materna e paterna mais reincidentes na prova de Morfologia. Os valores apresentados no Quadro 1 foram obtidos pelo cruzamento dos atributos: Família Materna, Categoria e Colocação. A partir dessas variáveis encontrou-se os animais da Família Materna mais reincidentes (troncos maternos dominantes) nas respectivas categorias, considerando as quatro primeiras colocações. A Família Materna denominada Torhuela foi a mais frequente (26 animais campeões), seguida por BT Fuzarca e Che Pitanga 565 (25 animais campeões cada uma).

Quadro 1. Animais da Família Materna mais reincidentes na prova de Morfologia.

Família materna	Cavalo adulto	Cavalo menor	Égua adulta	Égua menor	Gr campeão ato	Melhor exemplar	Potranca maior	Potranca menor	Potranco maior	Potranco menor	Total
Anahi onza	2	1	1		3					2	9
Aragana			1	2	1		1	1	2		8
Aroeira 8 de nazareth		1	2	3	8		1	1			16
Bt argila	1	1			4		1	1	2		10
Bt camanga	4	4		1	3		3		3	1	19
Bt dansarina					8	1			2	1	12
Bt fuzarca	1	2	3	2	12		2		2	1	25
Bt lambisgoia	2		5	1	9	1		1		2	21
Bt sarita		1		2	3			1	1	2	10
Che pitanga 565	1	3	1	3	7		6	1	2	1	25
Rosi 10 de nazareth			2	1	5	1	1	2			12
Santa isabel farrera			1	1	6		1				9
Taleira 486 de nazareth					4			3		1	8
Torhuela	1	1	1	1	10	1	4	3		4	26
Trinidad la invernada	1	1			1		2		4		9
Victoria de santa angelica	1				3		2	1		1	8

Fonte: Autores.

O mesmo procedimento foi realizado para a construção da Quadro 2, porém cruzando Categoria e Colocação com os animais da Tribo Paterna. Essa apresenta os animais da Tribo Paterna mais reincidentes nas respectivas categorias, considerando as quatro primeiras colocações. A Tribo Paterna denominada Aculeo Vastago foi a mais frequente (509 animais campeões), seguida pela tribo Curiche Rigor (131 animais campeões).

Quadro 2. Animais da Tribo Paterna mais reincidentes na prova de Morfologia.

Tribo paterna	Cavalo adulto	Cavalo menor	Égua adulta	Égua menor	Gr campeão ato	Melhor exemplar	Potranca maior	Potranca menor	Potranco maior	Potranco menor	Total
Aculeo vastago	37	55	33	36	130	11	52	50	54	51	509
Casas de polpaico estribillo		3	2	2	7		4	1	3	4	26
Charque chingolo			1						2	2	5
Coiron iii	3	6	4	7	9	1	4	5	9	6	54
Curiche rigor	10	11	8	11	31		14	20	12	14	131
Los tilos cachupin	1	6	4	5	14	2	11	9	6	10	68
Naranjal quebradito	2	3	2	2	10		3	5	2	3	32
No me toques	9	1	2	2	18		2	8	2	7	51
Peralillo tranquilo				1	2		1		1	2	7
Quillacon ii					5					1	6
San justo 400		1	3	2	15		3	7	1	1	33
Santa elsa descarinado	1		1		2			1			5
Santa zulema el huila		1		1	1		1	1		1	6
Totoral colibri		1	1	2			1	1		2	8

Fonte: Autores.

Os Quadros 1 e 2 permitem, adicionalmente, identificar famílias de padrões genéticos precoces ou tardios, de acordo com a reincidência nas categorias da prova. Outro destaque destas tabelas é a reincidência no Grande Campeonato, uma vez que se trata de uma categoria que mostra características superiores em um animal.

O Freio de Ouro consiste em um conjunto de nove provas de desempenho e possui o cálculo de quatro médias, até a média final de cada animal. O Quadro 3 apresenta os animais mais reincidentes da Tribo Paterna, Família Materna e Família Materna do Pai, com as melhores pontuações nas provas do Freio de Ouro (terço superior das notas individuais de cada prova). Em suma, os animais campeões mais reincidentes nas nove provas da Tribo Paterna e Linha Materna do Pai foram Aculeo Vastago e Hacienda Aculeo Sancion, respectivamente. Por outro lado, as Famílias Maternas dos animais campeões nas nove provas apresentaram uma alta variabilidade. Com as informações presentes nessa tabela, é possível aos criadores identificar as genealogias relacionadas às características específicas como, por exemplo: aptidão vaqueira (provas de mangueira e campo), habilidade lateral (provas de envolta sobre pata, figura e bayard).

Quadro 3. Animais da família materna mais reincidentes no Freio de Ouro e suas gerações.

Prova	Tribo paterna	Qtd	Linha materna pai	Qtd	Família materna	Qtd
Morfologia +8	Aculeo vastago	31	Haciendada aculeo sancion	12	Torhuela	2
	Curiche rigor	5	Torhuela	6	Tirana de santa angelica	2
	Los tilos cachupin	5	Mandinga onza 358	4	Mandinga onza 358	2
	Casas de polpaico estribillo	4	Bt dansarina	3	Bt lambisgoia	2
	Coiron iii	2	Bt fuzarca	3	Bt camanga	2
	Clementino aleteo	1	Bt camanga	2	Esmeralda do itapororo 19	1
	Flotador	1	Pozo azul castanuela	1	Pantera de santa angelica	1
	Aculeo coiron iii	1	Taleira 486 de nazareth	1	Rosi 14 de nazareth	1
	No me toques	1	Aculeo indirecta	1	Jardineira 163 merkt do penarol	1
	San justo 400	1	Bt fumaça	1	Che pitanga 591	1
Bayard +15	Aculeo vastago	95	Haciendada aculeo sancion	28	Torhuela	8
	Los tilos cachupin	10	Mandinga onza 358	10	Bt camanga	7
	Curiche rigor	49	Bt fuzarca	10	Miss de santa angelica	5
	Casas de polpaico estribillo	19	La amanecida muchachita	8	Peleadora la invernada	3
	San justo 400	12	Palmira piedrait	8	Mandinga onza 358	3
			Torhuela	8	Huifquenco tranquilla	3
			Bt camanga	8		
Mangueira +11	Aculeo vastago	109	Haciendada aculeo sancion	29	Torhuela	9
	Curiche rigor	53	Bt fuzarca	15	Bt camanga	8
	Casas de polpaico estribillo	30	Torhuela	14	Mandinga onza 358	5
	San justo 400	11	Palmira piedrait	10	Peleadora la invernada	4
	Los tilos cachupin	10	Bt camanga	10	Peleadora la invernada	3
			La amanecida muchachita	8	Manta pampa de san justo	3
		Mandinga onza 358	7	Huifquenco tranquilla	3	
Mangueira 2 +15	Aculeo vastago	79	Haciendada aculeo sancion	24	Torhuela	6
	Curiche rigor	43	Torhuela	15	Bt camanga	6
	Casas de polpaico estribillo	18	Mandinga onza 358	10	Peleadora la invernada	5
	San justo 400	11	Palmira piedrait	8	Miss de santa angelica	4
	Los tilos cachupin	10	Del oeste diferente	8	Mandinga onza 358	4
Campo 1 +12	Aculeo vastago	85	Haciendada aculeo sancion	18	Bt camanga	5
	Curiche rigor	39	Bt fuzarca	11	Torhuela	5
	Casas de polpaico estribillo	23	Mandinga onza 358	11	Bt argila	4
	San justo 400	9	Torhuela	11	Bt fuzarca	4
	Los tilos cachupin	8	Palmira piedrait	7	Reliquia do reponte	3
			Bt camanga	6	Peleadora la invernada	3
Campo 2 +15	Aculeo vastago	111	Haciendada aculeo sancion	34	Bt camanga	7
	Curiche rigor	59	Torhuela	18	Torhuela	5
	Casas de polpaico estribillo	23	Palmira piedrait	14	Mandinga onza 358	5
	San justo 400	13	Mandinga onza 358	14	Che pitanga 565	4
	Los tilos cachupin	12	Bt fuzarca	11	Peleadora la invernada	4
	No me toques	5	Bt camanga	8	Anahi onza	3
Andadura +12	Aculeo vastago	76	Haciendada aculeo sancion	12	Bt camanga	6
	Curiche rigor	30	Torhuela	9	Aroeira 8 de nazareth	4
	Casas de polpaico estribillo	16	Mandinga onza 358	8	Bt argila	3
	San justo 400	9	Bt fuzarca	7	Che napinda 553	3
	Los tilos cachupin	7	Bt camanga	6	Torhuela	3
			Aroeira 8 de nazareth	5	Cabriuva 4 de nazareth	3
		Huifquenco tranquilla	5	Bt bandeada	3	
Figura +12	Aculeo vastago	81	Haciendada aculeo sancion	15	Torhuela	5
	Curiche rigor	42	Bt fuzarca	12	Bt camanga	5
	Casas de polpaico estribillo	17	Palmira piedrait	10	Mandinga onza 358	4
	San justo 400	10	Torhuela	9	Chake la peloncha	3
	Los tilos cachupin	7	Mandinga onza 358	7	Bt insinuante	3
		Huifquenco tranquilla	6			

Vsp/esb +12	Aculeo vastago	61	Haciendada aculeo sancion	14	Peleadora la invernada	5
	Curiche rigor	35	Torhuela	7	Bt camanga	3
	Casas de polpaico estribillo	14	Del oeste diferente	7	Torhuela	3
	San justo 400	11	Mandinga onza 358	6	Huifquenco tranquila	3
	Los tilos cachupin	6	Palmira piedraita	6	Che pitanga 565	3
			La amanecida muchachita	5	Bt insinuante	3
			Bt camanga	4		
		Aroeira 8 de nazareth	3			
Final +19	Aculeo vastago	150	Haciendada aculeo sancion	38	Bt camanga	8
	Curiche rigor	72	Torhuela	22	Torhuela	8
	Casas de polpaico estribillo	26	Mandinga onza 358	15	Peleadora la invernada	7
	San justo 400	18	Palmira piedraita	14	Mandinga onza 358	6
	Los tilos cachupin	15	Bt fuzarca	13	Miss de santa angelica	5
			La amanecida muchachita	9	Bt fuzarca	4
			Del oeste diferente	9	Chake la peloncha	3
		Bt camanga	8			

Fonte: Autores.

Para os Quadros 4 e 5, foram selecionados os animais com maior frequência no Quadro 3, ou seja, os com melhor desempenho nas provas do Freio de Ouro. Assim, por meio dos animais selecionados foram localizadas as respectivas Tribos Paternas e Famílias Maternas.

Quadro 4. Animais da tribo paterna mais reincidentes no Freio de Ouro e suas gerações.

Tribo paterna	2m	Qtd	3mm	Qtd	4mmm	Qtd	5mmmm	Qtd
Aculeo vastago	La invernada hornero	223	Tren tren arrebol	273	Tren tren arrebol	437	Estuco 2	273
	Bt delantero	46	La invernada hornero	418	La invernada hornero	193	Tren tren arrebol	218
	Pirai 1569 do brazao	29	Bt brazao do junco	38	Sendero kalifa	21	La invernada hornero	31
	Bt lucero	29	Bt delantero	29	Ja cartucho	11	Hacienda aculeo veguero	11
	Chicao de santa odessa	28	Crt guapo	28	Estuco 2	11	Sendero kalifa	11
	Nobre tupambae	26	Ja cartucho	21	Bt brazao do junco	7		
Casas de polpaico estribillo	Santa elba comediante	22	Puchaura esquinazo	23	Puchaura esquinazo	19	Casas de polpaico estribo	40
	Mackenna guindo	15	Santa elba comediante	19	La amanecida esperando	17	Puchaura esquinazo	6
	As malke sedutor-te	14	Mackenna guindo	16	Santa elba comediante	6	Com dario pavez guarani	5
	Butia leopardo	7	La amanecida esperando	15	Casas de polpaico estribo	5	Casas de polpaico estribillo	1
	Sbt custodio	3	Lutador do puruna	3	Santa isabel escorpion	1		
La amanecida esperando	3	Butia leopardo	2					
San justo 400	Tanido redoblado	25	Tanido trampolin	25	Tanido trampolin	26	San justo conde	25
	Quero quero de santa edwiges	7	Tanido redoblado	24	San justo conde	4	Tanido trampolin	7
	Rz tanido redoblado ii	5	Mananero abanico	4	Dormido trampero	4	Prendedor guampa	4
	Desafio de santa edwiges	5	Dormido trampero	2	Tanido redoblado	3	Dormido trampero	2
	Rodopio de santa edwiges	4	Ja anarquista	2	Mananero abanico	2		
	Mananero manicero	4	Mananero manicero	2				
Los tilos cachupin	Las hortensias rigolemu	19	Las hortensias rigolemu	23	Curitiro facon	19	El rincon estanquero	26
	Indio da escondida	15	Festeiro do itapororo	15	El rincon estanquero	19	Curiche salteador	19
	Macanudo do itapororo	15	Tinajera buen abrigo	15	Las hortensias rigolemu	18	Los tilos sembrador	19
	Equador de santa edwiges	15	Equador de santa edwiges	11	Tinajera buen abrigo	11	Curitiro facon	11
	Ja impulso	7	Curitiro facon	4	Festeiro do itapororo	4	Las hortensias rigolemu	4
	Oitao de rancho da escondida	4	Indio da escondida	4	Los tilos sembrador	4		
Curiche rigor	Santa elba senuelo	46	Santa elba senuelo	55	Las mercedes taco	87	Las mercedes taco	58
	Muchacho de santa angelica	25	Las mercedes taco	48	Santa elba senuelo	32	Com. Dario pavez comunista	53
	Indio do boeiro	21	Muchacho de santa angelica	30	Mandante de santo angelo	21	El sur gancho amigo	21
	Bt lamborguine	15	Bt farroupilha	15	El aroma de pichidegua campero	18	Santa elba bellaco	18
	Ganadero da harmonia	13	Santa elba bellaco	14	Santa elba bellaco	13	Bellavista ganchito	11
	La invernada pascuero	12	El aroma de pichidegua campero	12	El sur gancho amigo	11	Santa elba puelche	7

Fonte: Autores.

Quadro 5. Animais da família materna mais reincidentes no Freio de Ouro e suas gerações.

Família materna	2f	Qtd	3ff	Qtd	4fff	Qtd	5ffff	Qtd
Torhuela	Fuzarca do itapororo	6	La invernada tortolita iv	16	La invernada jauria	16	La invernada buenas noches	16
	Balisa do itapororo	5	La invernada pama	5	La invernada tortolita iv	6	La invernada jauria	6
	Serpentina do itapororo	5	Chilenero bataraza	4	La invernada la gorda	5	La invernada parana	5
	Tinajera mancha rara	4	Dinastia do itapororo	4	Melincue paula	4	La invernada trilhadora	4
	Tarrafa do puruna	4	La invernada jauria	4	La invernada buenas noches	4	La invernada torhuela ii	3
	Desejada da boa vista	4	La invernada gualtata	3	La invernada pajarera	3	La invernada pepilla	2
	La invernada tortolita iv	4	La invernada anchoa ii	2	La invernada atorhuela	2	La invernada tinaja	2
Bt camanga	Butia fantasia	13	Bt tarca	14	Bt loirosa	14	Taleira 2	13
	Bt nicotina	8	Bt nicotina	9	Taleira 2	13	Taleira 340 de nazareth	13
	Bt familia do junco	4	Bt tirana	3	Bt nicotina	4	Bt nicotina	3
	Bt calhandra do junco	3	Bt loirosa	2	Bt tarca	1	Bt loirosa	1
	Bt loirosa	3	Bt sirigaita	2	Bt tirana	1	Bt tirana	1
	Bt cachucha	2	Bt familia do junco	1	Bt ermana do junco	1		
Miss de santa angelica	Doutora de santa angelica	4	Taguatinga de santa angelica	4	Santa angelica 44	3	Santa angelica 44	7
	Inca de santa angelica	2	Brisa de santa angelica	2			Morena	3
	Orgulhosa de santa angelica	2	Disposta de santa angelica	1				
	Campeira de santa angelica	1						
Peleador a la invernada	Bt labareda	10	Bt eminencia do junco	6	Tijereta	13	Cachimba oriental	13
	As malke mimosa	5	Bt labareda	6	Bt rendeira	6	Bt labareda	6
	Bt relancina	3	Olivia do igiquiqua	2	Dançarina do igiquiqua	2	Tijereta	6
	Bt venus	3	Bt rendeira	1	Bt labareda	1	Bt odessa	2
	Unica do igiquiqua	2						
Mandinga onza 358	Bt raposa	5	Bt nazarena	5	Bt nazarena	5	Castilla	7
	Bt carola do junco	4	Bt carola do junco	4	Bt raposa	4	Bt nazarena	4
	Jotace barbela	2	Bt raposa	4	Castilla	1	Vibora	1
	Bt umbanda	1	Bt odalisca	2	Bt odalisca	1	Bejera	1
	Bt vinagrete	1	Bt brigite	1	Vibora	1		
	Garbosa da 3j	1	Bt soberana	1				
Jotace hortencia	1	Castilla	1					

Fonte: Autores.

Nesses quadros é possível identificar as informações relacionadas à genealogia dos melhores indivíduos transmissores de informações genéticas de cada tribo paterna e, adicionalmente, as tribos paternas com maior destaque na prova de freio de ouro. Trata-se de uma informação que contribui para a melhoria da raça de cavalos crioulos, visto que não é importante apenas pertencer a uma linhagem, mas sim, ser um bom transmissor de características que contribuem para o animal ser vencedor. Apesar da informação genealógica ser altamente valorizada na avaliação de potencial do animal (Gianluppi et al., 2009), dados genéticos de metabolismo, bem como fatores ambientais merecem atenção nessa área (Regatieri & Mota, 2012).

Considerando as variáveis quantitativas relacionadas à prova de Freio de Ouro, as médias das notas finais foram comparadas em relação à posição da prova. Para isso foi aplicado o teste de ANOVA com correção de Brown-Forsythe (devido à violação da premissa da homogeneidade das variâncias) seguido de teste de Bonferroni. A comparação múltipla das médias foi estatisticamente significativa ($p < 0,001$). Os resultados até a décima colocação estão apresentados na Tabela 2.

Tabela 2. Atributos do conjunto de dados, suas respectivas classificações e exemplos de valores.

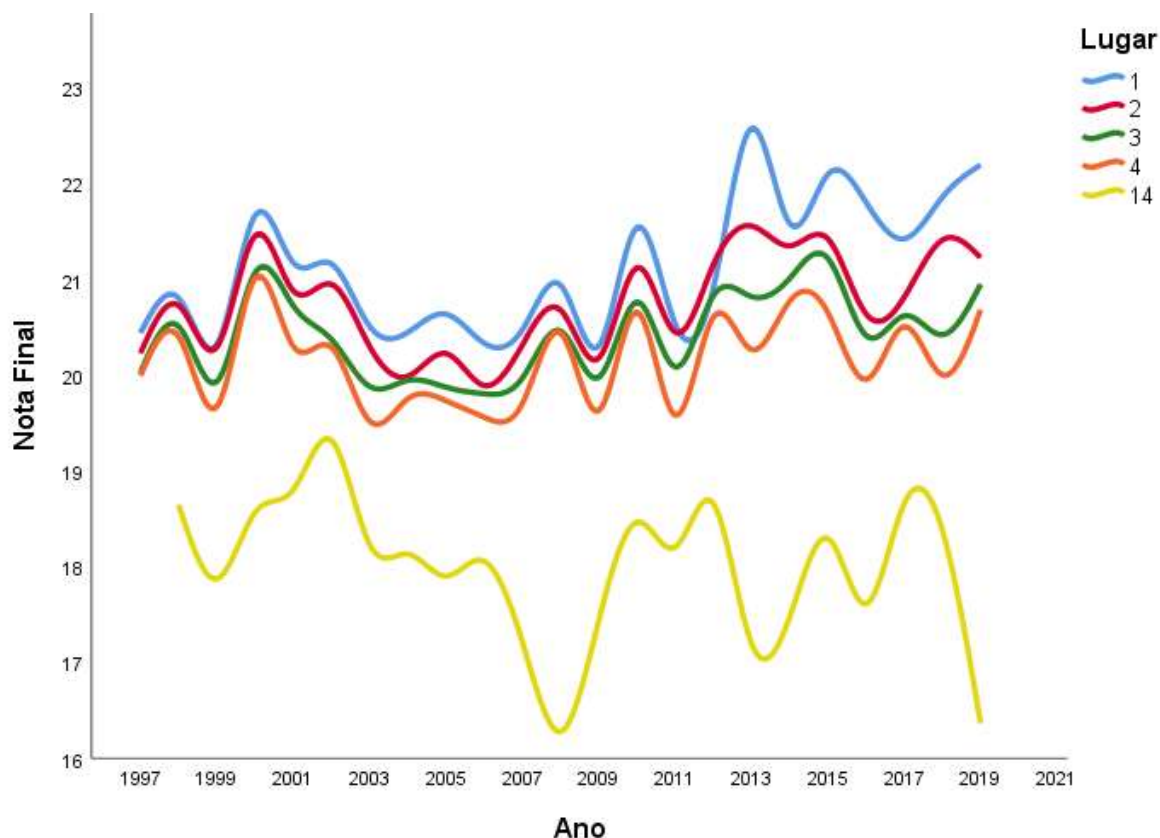
Subconjuntos homogêneos de médias das colocações							
(Médias \pm Desvio Padrão)							
Colocação	N	1	2	3	4	5	6
1	48	21,14 \pm 0,73					
2	48	20,77 \pm 0,4	20,77 \pm 0,4				
3	48		20,41 \pm 0,3	20,41 \pm 0,3			
4	48			20,15 \pm 0,34	20,15 \pm 0,34		
5	48				19,92 \pm 0,4	19,92 \pm 0,4	
6	48				19,78 \pm 0,42	19,78 \pm 0,4	
7	48					19,61 \pm 0,41	19,61 \pm 0,41
8	47					19,45 \pm 0,37	19,45 \pm 0,37
9	47						19,22 \pm 0,62
10	46						19,18 \pm 0,42

Fonte: Autores.

É possível verificar que, apesar das médias apresentarem valores próximos às primeiras posições, são formados grupos com médias maiores que 20 pontos. Apesar de ser um critério importante, considerar apenas a nota da prova isoladamente não contribui para a zootecnia e melhoramento da raça. Como é possível observar na Tabela 2, as médias das primeiras posições possuem pouca variação e o desvio padrão associado também é baixo. Além disso, algumas provas do Freio de Ouro possuem elementos externos que podem diminuir a nota do animal, o que não reflete em suas características genéticas. Um exemplo são as provas de campo, que contam com a presença de bovinos. Nessas provas, o comportamento do bovino pode prejudicar o cavalo, baixando sua nota (Garcia et al., 2020).

Outra análise realizada foi a distribuição das notas ao longo do tempo. Percebe-se que as notas apresentam oscilações ao longo dos anos (Figura 2). Não há nenhuma linha de tendência perceptível das notas finais em função do tempo. Este é outro resultado que mostra a influência dos fatores externos ao cavalo na nota final do Freio de Ouro. Das 8 provas que compõem a nota final, quatro envolvem elementos externos ao cavalo como condições da pista e gado participante.

Figura 2. Notas finais da prova Freio de Ouro.



Fonte: Autores.

Assim, a verificação de tendências não é perceptível. Essa justificativa fundamenta-se, também, com análise de regressão linear, que apresentou R^2 de 0,57 quando se analisou a relação entre a nota final (variável dependente) em função do ano (variável independente).

4. Conclusão

A análise do conjunto de dados mostrou a grande variabilidade existente do ponto de vista da teoria da informação. O conjunto de dados é vasto e com pouca repetição, o que dificulta a aplicação de algoritmos de mineração de dados. A partir dos resultados apresentados, foi possível identificar as linhagens mais recorrentes entre as melhores colocações nas provas de Morfologia e Freio de Ouro. Assim, de acordo com o objetivo do criador, é possível verificar os troncos maternos e paternos que apresentam as características desejadas, bem como a informação sobre os transmissores dessas características. Ainda assim, é importante ressaltar que o bom desempenho em competições depende de um conjunto complexo de fatores. Nesse sentido, existem lacunas a serem preenchidas na integração de informações genealógicas com estudos de caráter morfológico, genético e metabólico.

Referências

ABCCC. ABCCC - Associação Brasileira de Criadores de Cavalos Crioulos. <https://www.cavalocrioulo.org.br/>

Cardoso, O. N. P. & Machado, R. T. M. (2008). Gestão do conhecimento usando data mining: Estudo de caso na Universidade Federal de Lavras. *Revista de Administração Pública*, 42, 495–528.

Faceli, K., Lorena, A. C., Gama, J., Almeida, T. A. & Carvalho, A. C. P. R. F. (2021). *Inteligência Artificial—Uma Abordagem de Aprendizado de Máquina* (2ª ed.). LTC.

- Cucco, D. C., Salles, E. L., Santos, M. R., Ferreira, R., Soriano, V. S., Zampar, A. & Kessler, J. D. (2016). Freio de Ouro como ferramenta de seleção na raça Crioula. *Archivos de Zootecnia*, 65(250), 155–161.
- Dani, G., Sachet, M. & De Avila e Silva, S. (2020) Data Mining in Bioinformática [recurso eletrônico]: contexto computacional e aplicações. Educus.
- Demšar, J., Curk, T., Erjavec, A., Gorup, Č., Hočevar, T., Milutinovič, M., Mozina, M., Poljnar, M., Toplak, M., Staric, A., Stajdohar, M., Umel, L., Zagar, L., Zbonitar, J., Zitnik, M. & Zupan, B. (2013). Orange: data mining toolbox in Python. *the Journal of machine Learning research*, 14(1), 2349-2353.
- Elmasri, R. & Navathe, S. B. (2019). Sistemas de Banco de Dados. Pearson Universidades.
- Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11), 27-34.
- Fayyad, U. M., Haussler, D. & Stolorz, P. E. (1996). *KDD for Science Data Analysis: Issues and Examples*. In *KDD*.
- Ferreira, P. B., Vale, M. M. D., Macedo, A., Boemo, L. S., Rorato, P. R. N. & Beck, T. B. (2013). Classificação de características produtivas fenotípicas de diferentes raças de poedeiras através da mineração de dados. *Ciência Rural*, 43, 164-171.
- Garcia, C. A. S. C., Martins, C. F., Silva, L. L., Brandi, R. A., Feijó, J. O., Barbosa, A. A., Bruhn, F. R. P. & Corrêa, M. N. (2020). Frequência cardíaca, lactacidemia e gasto energético de equinos da raça Crioula em provas credenciadoras ao Freio de Ouro. *Arquivo Brasileiro de Medicina Veterinária e Zootecnia*, 72, 1631–1638.
- Gianluppi, L. D. F., Bortoli, E. C., Schwarz Sobrinho, R., Falcão, T. F. & Silva, T. N. (2009). Agregação de valor em equinos da raça crioula: Um estudo de caso. *Archivos de Zootecnia*, 58(223), 471–474.
- Goldschmidt, R., Passos, E. & Bezerra, E. (2015). *Data Mining*. Elsevier Brasil.
- Lima, H. P., de Abreu, U. G., Oliveira, S. D. M. & Massruhá, S. M. (2013). Mineração de dados aplicada à modelagem da incidência da Anemia Infecciosa Equina (AIE), no Pantanal sul-matogrossense. In: *Embrapa Informática Agropecuária-Artigo em anais de congresso*. Universidade Federal de Mato Grosso.
- Minitab 17 Statistical Software (2010). [Computer software]. Minitab, Inc. (www.minitab.com)
- Neethirajan, S. (2020). The role of sensors, big data and machine learning in modern animal farming. *Sensing and Bio-Sensing Research*, 29, 100367.
- Ouyang, Z., Sargeant, J., Thomas, A., Wycherley, K., Ma, R., Esmailbeigi, R., Versluis, A., Stacey, D., Stone, E., Poljak, Z. & Bernardo, T. M. (2019). A scoping review of “big data”, “informatics”, and “bioinformatics” in the animal health and veterinary medical literature. *Animal Health Research Reviews*, 20(1), 1–18.
- Regatieri, I. C. & Mota, M. D. S. (2012). Melhoramento genético de equinos: Aspectos bioquímicos. *Ars Veterinaria*, 28(4), 227–233.
- Silva, L. A. (2016). *Introdução à Mineração de Dados - Com Aplicações em R GEN LTC*.
- VanderWaal, K., Morrison, R. B., Neuhauser, C., Vilalta, C. & Perez, A. M. (2017). Translating big data into smart data for veterinary epidemiology. *Frontiers in veterinary science*, 4, 110.