Artificial Intelligence in Industry 4.0: The future that comes true: AI

Karabegović, Isak; editor

2024-09-17
https://bastina.anubih.ba/handle/123456789/791

DOI: 10.5644/PI2024.215.18

Predicting Behavioral Patterns in Brazillian Whiteknee Tarantulas - A Multi-Output Machine Learning Analysis of Dietary Effects in Acanthoscurria Geniculata

Amina Radončić*[1]

Abstract: *Machine learning techniques have the potential to act as powerful tools for predicting behavioral patterns in tarantulas. For this particular observation, Brazilian Whiteknee Tarantulas (Acanthoscurria geniculata) are placed in a focal point. The study leverages a multi-output machine learning approach implemented in MATLAB to analyze the impact of dietary compositions—Worms, Crickets, Roaches, Locusts, and Mixed Diet —on tarantula behavior. Based on the collected data from the survey involving 47 owners of the same species, the data on gender, dietary preferences, and activity levels of these tarantulas has been obtained. The findings highlight significant variations in predictive accuracy across different diets, with the mixed diet demonstrating the highest accuracy (0.85714) and the diet consisted only of worms showing robust performance (accuracy: 0.71429, AUC: 0.70833). Despite the limited dataset, this research underscores the potential of machine learning to elucidate the intricate relationship between diet and behavior in tarantulas, offering insights relevant to both ecological studies and captive management practices.*

Keywords: *machine learning, decision tree model, multi-output model, data analysis, tarantula, brazillian whiteknee*

1. Introduction

Every moment in life presents a multitude of choices, each offering numerous possibilities to explore. **Decision making**, therefore, involves the cognitive process of selecting the most suitable course of action among available options. It requires **a thoughtful assessment of the advantages and disadvantages of each alternative while considering all feasible choices**. This intricate process is a fundamental challenge faced by individuals and organizations alike. To address this challenge effectively, **algorithms such as decision trees** provide a structured approach, ensuring comprehensive decision analysis by meticulously weighing essential attributes without overlooking any critical factors [1]. Decision trees serve as indispensable tools in decision making, widely utilized across various domains. These graphical representations map decisions and potential outcomes, encompassing factors like resource allocation, event

*[1]Internatinal Burch University, Natural and Medical Sciences, Sarajevo, Bosnia and Herzegovina
E-mail: amina.radoncic@stu.ibu.edu.ba

probabilities, and utility assessments [1, 2]. Classification trees predict outcomes based on data insights, prevalent in fields like probability and statistics. Regression trees, conversely, determine outcomes from quantitative data, often applied in real estate evaluations [1].
Decision tree forests aggregate diverse models to enhance predictive accuracy, while classification and regression trees combine attributes for logical outcome predictions [1, 2]. Despite varying in accuracy, decision trees, including K Means Clustering, integrate factors for comprehensive decision support across diverse applications: from text classification to medical diagnostics and financial forecasting, and in this case – **behavioral patterns in tarantulas.**

2. The Mysterious Lives of Tarantulas

Current research on tarantula behavioral patterns, particularly regarding the influence of nutrition, is notably sparse in utilizing advanced analytical methods like machine learning and artificial intelligence. Existing studies predominantly focus on ecological observations and basic behavioral assessments.
Taxonomically observed, **tarantulas** belong to the infraorder Mygalomorphae as **the largest ararchinds.** This infraorder comprises approximately 3000 species of spiders distributed globally across 16 families, including trap-door spiders, and sheet-web spiders. These spiders are known for their large size and long lifespan, as well as their diverse silk-producing capabilities [3].
Mygalomorphs are considered an ancient lineage, distinct from the more diverse Araneomorphae, and retain several primitive spider characteristics such as two pairs of book lungs, paraxial chelicerae, simple silk spigots, and comparatively weaker silk [3].

Figure 1. The Brazillian Whiteknee tarantula in captivity

Despite these primitive traits, specific anatomical features support the monophyly of Mygalomorphae, including labial and maxillary cuspules, fewer palpal bulb sclerites, sternal sigillae, segmented posterior lateral spinnerets, reduced anterior lateral spinnerets, and the absence of anterior median spinnerets [3].

The crown jewel of this study is *Acanthoscurria geniculata,* **the Brazillian Whiteknee tarantula**, one of the largest spiders in the world.

These tarantulas are predominantly deep black in body and leg coloration, with occasional lighter shades on their legs and distinctive white bands that enhance their striking appearance. Notably, females of this species are larger and more intensely colored compared to the smaller, less vividly hued males [4]. In terms of size, Brazilian whiteknee tarantulas surpass the average tarantula, with body lengths reaching up to 9 centimeters (3.5 inches). Known for their rapid growth, females typically achieve a mature leg span of 8.5 inches within 3 to 4 years. These tarantulas are popular as pets due to their impressive size, robust nature, and captivating color patterns. However, they exhibit defensive behavior and possess urticating hairs that can cause skin irritation [4].

While the Brazilian whiteknee tarantula can bite if provoked, it typically deploys urticating hairs as a primary defense mechanism. These hairs deter potential threats by causing discomfort and irritation, minimizing the need for biting. Although their venom is generally not medically significant, their large size can potentially cause mechanical harm in the event of a bite [4].

Are their behavioral patterns and temper dictated by the dietary habits? This study aims to fill this research gap and answer this simple, yet complex question, by employing machine learning algorithms to predict tarantula behavior based on dietary factors. By integrating computational approaches with behavioral ecology, this research seeks to contribute novel insights and potentially enhance strategies for managing tarantula care in captivity and conservation efforts in natural habitats.

3. Materials and Methods

This study involved a survey administered anonymously to 47 Brazilian Whiteknee Tarantula owners. Participants provided data on their tarantulas' age categories (3+, 2+, 1+, <1 years), gender (scored with 2 points for females, 1 point for males, and 0 points for unknown gender), activity levels categorized as LOW, NORMAL, or HIGH, and temperament described as DOCILE, NERVOUS, FLICKER (indicating hair flicking behavior), or DEFENSIVE (characterized by defensive postures). Participants also documented their tarantulas' dietary habits, recording consumption of Worms, Crickets, Roaches, Locusts, or a mixed diet using binary scoring (1 for consumption, 0 for non-consumption).

Table 1. Survey questions and criteria

Category	Description	Scoring
Age	to 20	-
Gender	20 – 200	2 points for Female, 1 point for Male, 0 points for Unknown gender
Activity	200 – 500	-
Temper	LOW, NORMAL, HIGH	-
Diet: Worms	Consumption (1) or Non-consumption (0)	Binary
Diet: Crickets	Consumption (1) or Non-consumption (0)	Binary
Diet: Roaches	Consumption (1) or Non-consumption (0)	Binary
Diet: Locusts	Consumption (1) or Non-consumption (0)	Binary
Diet: Mixed	Consumption (1) or Non-consumption (0)	Binary

The obtained results were processed using MATLAB to investigate the correlation between gender, temperament, and dietary preferences of these tarantulas. Initially, categorical variables such as gender (Male, Female, Unknown) and temperament (DOCILE, NERVOUS, DEFENSIVE, FLICKER) were converted into numeric formats suitable for machine learning analysis. The dataset was then divided into training (70%) and testing (30%) sets using a holdout method to ensure robust model validation. For each dietary category (Worms, Crickets, Roaches, Locusts, Mixed), individual decision tree models were trained using the training data. Model performance metrics including accuracy, root mean squared error (RMSE), and area under the ROC curve (AUC) were computed to assess the predictive capability of the models. Visualizations such as confusion matrices, ROC curves, feature importance analyses, and predictive distribution plots were generated to interpret and validate the model's predictions. This approach enabled the identification of nuanced behavioral patterns and dietary preferences among A. geniculata tarantulas, underscoring the effectiveness of decision tree models in analyzing complex behavioral datasets in animal studies.

4. Results and Discussion

4.1. Model Analysis

The accuracy of each decision tree model was evaluated using the test data. Accuracy measures the proportion of correctly predicted outcomes relative to the total number of predictions made. For instance, the model achieved an accuracy of 71.43% for Diet_Worms, indicating that 71.43% of predictions regarding whether a tarantula consumes worms were correct.

RMSE quantifies the average magnitude of the errors between predicted and actual values across all dietary categories. Lower RMSE values indicate better model performance in predicting dietary preferences accurately. For example, the RMSE for Diet_Mixed was 0.37796, suggesting that the model's predictions for tarantulas with mixed diets were generally close to the actual values.

The ROC curve and AUC were used to assess the model's ability to distinguish between positive and negative classes for each dietary category. AUC values close to 1 indicate excellent model performance, while values around 0.5 suggest a model that performs no better than random guessing. The AUC values varied across dietary categories, with Diet_Locust having an AUC of 0.5, indicating that the model's predictive ability for this category was not significantly better than chance. Feature importance analysis was conducted to identify which predictors (gender and temperament) contributed most significantly to the models' predictions. This analysis helps in understanding which variables influence tarantula dietary preferences the most. For instance, the analysis might reveal that female tarantulas and those categorized as DEFENSIVE or NERVOUS tend to prefer specific types of food more than others.

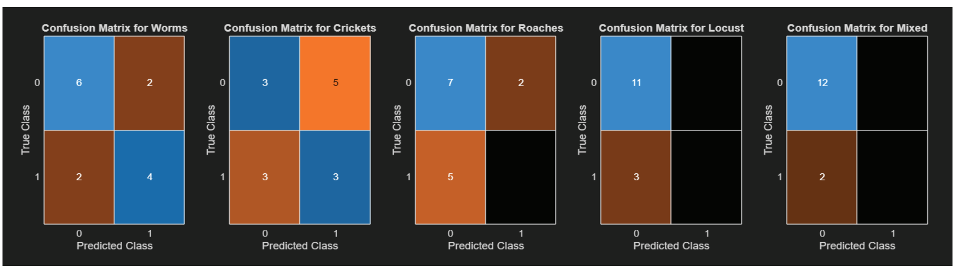

Figure 2. Confusion matrices for each dietary category

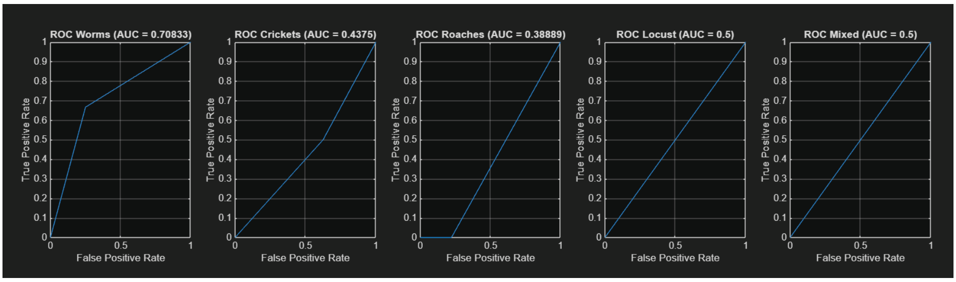

Figure 3. ROC curve for each dietary category, showing false positive and true positive rates

Visualizations (figures above) such as confusion matrices, ROC curves, boxplots, and histograms were generated to provide a comprehensive understanding of model performance and prediction outcomes. Confusion matrices visually represented the distribution of true positive, false positive, true negative, and false negative predictions for each dietary category. ROC curves illustrated the trade-off between sensitivity and specificity, with AUC values quantifying the models' discriminative ability.

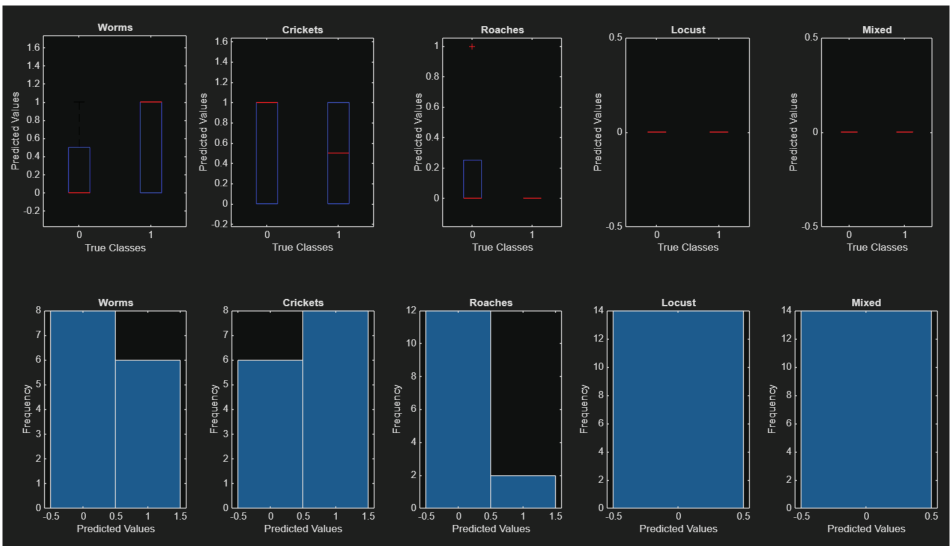

Figure 4. Boxplots (top) and histograms (bottom) portraying the prediction outcomes for dietary categories

5. Conclusion

Through machine learning methodologies and examinations, this study aimed to investigate the interplay between gender, temperament, and dietary preferences in Brazilian Whiteknee tarantulas. Based on the analysis of anonymized survey data from 47 tarantula owners, decision tree models were constructed to predict the dietary habits of these spiders, encompassing preferences for worms, crickets, roaches, locusts, and mixed diets. Despite the limited dataset, he findings revealed nuanced associations between gender, temperament traits, and dietary behaviors, with varying accuracies, root mean squared errors (RMSE), and area under the ROC curve (AUC) values observed across different dietary categories.

Gender and temperament emerged as pivotal predictors influencing dietary preferences, offering insights into tarantula management and conservation practices. Despite the robustness of the predictive models in capturing these relationships, limitations such as the modest sample size and potential biases inherent in survey data warrant cautious interpretation of the results.

Future investigations could expand upon this study by incorporating larger datasets, integrating environmental variables, and extending comparative analyses to include other species within the Theraphosidae family. Continued advancements in machine learning techniques and data collection methodologies hold promise for refining predictive models and deepening our understanding of animal behavior across diverse ecological contexts.

This one-of-a-kind research contributes to advancing knowledge in animal behavior research and underscores the potential of machine learning in informing conservation strategies and management practices for tarantula species. By elucidating the complex relationships between behavioral traits and dietary preferences, this research enhances our ability to implement targeted conservation efforts aimed at safeguarding tarantula populations bred in captivity.

6. References

[1] Navada, A., Ansari, A. N., Patil, S., & onkamble, B. A. (2011). Overview of use of decision tree algorithms in machine learning. 2011 IEEE Control and System Graduate Research Colloquium. doi:10.1109/icsgrc.2011.5991826

[2] M. Somvanshi, P. Chavan, S. Tambadeand S. V. Shinde, "A review of machine learning techniques using decision tree and support vector machine," *2016 International Conference on Computing Communication Control and Automation (ICCUBEA)*, Pune, India, 2016, pp. 1-7, doi: 10.1109/ICCUBEA.2016.7860040.

[3] Pérez-Miles, F., &Perafán, C. (2017). *Behavior and Biology of Mygalomorphae. Behaviour and Ecology of Spiders, 29–54.* doi:10.1007/978-3-319-65717-2_2

[4] Macedo KWR, Costa LJL, de Souza JO, de Vasconcelos IA, de Castro JS, de Santana CJC, Magalhães ACM, Castro MS, Pires OR. Brazilian The raphosidae: a toxicological point of view. J Venom Anim Toxins Incl Trop Dis. 2021 Nov 22;27:e20210004. doi: 10.1590/1678-9199-JVATITD-2021-0004. PMID: 34868282; PMCID: PMC8610171

[5] Malti Bansal, ApoorvaGoyal, Apoorva Choudhary, A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning, Decision Analytics Journal, Volume 3, 2022, 100071, ISSN 2772-6622, https://doi.org/10.1016/j.dajour.2022.100071.

[6] De Ville, B. (2013). Decisiontrees. *Wiley Interdisciplinary Reviews: Computational Statistics*, *5*(6), 448-455.

[7] Ciaburro, G. (2017). *MATLAB for machine learning*. Packt Publishing Ltd.

[8] Suthaharan, S., & Suthaharan, S. (2016). Decision tree learning. *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*, 237-269.

[9] Ferri, C., Flach, P., & Hernández-Orallo, J. (2002, July). Learning decision tree susing the area under the ROC curve. In *Icml* (Vol. 2, pp. 139-146).

[10] Ling, C. X., Huang, J., & Zhang, H. (2003). AUC: a better measure than accuracy in comparing learning algorithms. In *Advances in Artificial Intelligence: 16th Conference of the Canadian Society for Computational Studies of Intelligence, AI 2003, Halifax, Canada, June 11–13, 2003, Proceedings 16* (pp. 329-341). Springer Berlin Heidelberg.

[11] Paluszek, M., & Thomas, S. (2019). *MATLAB machine learning recipes: a problem-solution approach* (pp. 1-347). Berkeley: Apress.

[12] Escoubas, P., &Rash, L. (2004). Tarantulas: eight-legged pharmacists and combinatorial chemists. *Toxicon*, *43*(5), 555-574.

[13] Pechmann, M. (2020). Embryonic development and secondary axis induction in theBrazilian whiteknee tarantula Acanthoscurriageniculata, CL Koch, 1841 (Araneae; Mygalomorphae; Theraphosidae). *Development genesan devolution*, *230*(2), 75-94.

[14] Paluszek, M., Thomas, S., Paluszek, M., & Thomas, S. (2020). MATLAB machine learning toolboxes. *Practical MATLAB deeplearning: A project-based approach*, 25-41.