agriculture

Article

Predictive Modeling of Honey Yield in Rural Apiaries: Insight from Chachapoyas, Amazonas, Peru

Yander M. Briceño-Mendoza [1], José Américo Saucedo-Uriarte [2], Lenin Quiñones Huatangari [2], Jhoyd B. Gaslac-Gomez [2], Hurley A. Quispe-Ccasa [3] and I. S. Cayo-Colca [1,2,*]

[1] Escuela de Posgrado, Programa Doctoral en Ciencias para el Desarrollo Sustentable, Facultad de Ingeniería Zootecnista, Biotecnología, Agronegocios y Ciencia de Datos, Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas, Chachapoyas 01001, Peru; yander.briceno@untrm.edu.pe
[2] Facultad de Ingeniería Zootecnista, Biotecnología, Agronegocios y Ciencia de Datos, Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas, Chachapoyas 01001, Peru; jose.saucedo@untrm.edu.pe (J.A.S.-U.); lenin.quinones@untrm.edu.pe (L.Q.H.); 081021a122@untrm.edu.pe (J.B.G.-G.)
[3] Dirección de Desarrollo Tecnológico Agrario, Instituto Nacional de Innovación Agraria, La Molina, Lima 15024, Peru; hurleyabelqc@gmail.com
* Correspondence: icayo.fizab@untrm.edu.pe; Tel.: +51-979897724

Abstract

Honey production is influenced by multiple factors, including climatic conditions, hive management practices, and harvest scheduling. This study evaluated the predictive capacity of statistical modeling techniques using data mining algorithms (MARS, CHAID, CART, and Exhaustive) and artificial neural network algorithms (Multilayer Perceptron, MLP) to estimate honey yields in apiaries located in northeastern Peru. A structured survey was conducted with sixty-nine beekeepers across nineteen districts in the Chachapoyas province. Variables included beekeeper experience, instruction, hive count, visit frequency, harvest frequency, additional income-generating activities, and geographic location. Descriptive statistics, non-parametric tests, Spearman correlations, and exploratory factor analysis were applied to identify latent structures. A linear mixed-effects model was used to assess the combined influence of predictors on honey production, with district included as a random effect. Results indicated that hive number, beekeeping experience, harvest frequency, and exclusive engagement in apiculture were statistically associated with increased honey yields. The model explained a substantial proportion of variance, supporting the integration of technical and socio-demographic variables in production forecasting. These findings demonstrate the utility of predictive modeling for informing hive management strategies and improving the operational efficiency of small-scale beekeeping systems in Andean regions.

Keywords: bee; beekeeping; hive; correlation

check for updates

Academic Editor: Francesco Marinello

Received: 30 August 2025
Revised: 7 November 2025
Accepted: 12 November 2025
Published: 18 November 2025

Citation: Briceño-Mendoza, Y.M.; Saucedo-Uriarte, J.A.; Quiñones Huatangari, L.; Gaslac-Gomez, J.B.; Quispe-Ccasa, H.A.; Cayo-Colca, I.S. Predictive Modeling of Honey Yield in Rural Apiaries: Insight from Chachapoyas, Amazonas, Peru. *Agriculture* **2025**, *15*, 2377. https://doi.org/10.3390/agriculture15222377

1. Introduction

In South America, Peru was among the first countries to develop beekeeping, particularly with Africanized and stingless bees. However, limited information exists regarding their distribution and the economic impact of apiculture at the national level [1,2]. Despite its high productive potential, beekeeping has not become a major economic activity in Peru. By the mid-20th century, domestic honey production accounted for only about 0.1% of the global market, with approximately 100,000 honeybee colonies primarily dedicated to local consumption. Most apicultural activities were conducted by amateurs (over 70%) and received minimal governmental support, which restricted the sector's development and economic contribution [1,3]. In the Peruvian Amazon, honey is widely used and frequently traded in local markets, mainly as an ingredient in traditional medicine, although it is also consumed as a natural sweetener [4,5]. Honey is a natural substance produced by bees from floral nectar [6,7]. Beyond its organoleptic properties, honey contains bioactive compounds with documented antioxidant, anti-inflammatory, and antimicrobial activities, supporting its classification as a functional food with potential health-promoting effects [8–10]. Additionally, honey production represents an economically viable activity for rural populations, as its sustained demand in domestic and international markets contributes to income generation and supports local socio-economic development [11].

Honey production is influenced by a complex interplay of environmental conditions, apiary location, colony management strategies, and beekeeper expertise. To systematically evaluate these factors, predictive modeling techniques can be employed, enabling the development of robust frameworks for optimizing yield [12]. These approaches employ statistical and computational algorithms—such as data mining methods (CART, CHAID, Exhaustive CHAID, and MARS) and artificial neural network algorithms (Multilayer Perceptron, MLP)—to evaluate variable interactions, generate forecasts [13], and identify key determinants of hive performance [14]. Thus, enhancing honey production requires a comprehensive understanding of the variables affecting colony efficiency, allowing beekeepers to implement data-informed strategies that improve productivity.

2. Materials and Methods

2.1. Study Area

The study was conducted across 19 districts within the Chachapoyas province, Amazonas, Peru, spanning altitudinal gradients from 1800 to 2900 m above sea level (masl). These districts are located within a montane forest ecosystem, typified by temperate climatic conditions (9 °C to 28 °C), moderate precipitation, and high floristic diversity (Figure 1). Apiary distribution across the surveyed districts was as follows: Cheto accounted for the highest proportion with 8 apiaries (11.6%), followed by Chuquibamba, Molinopampa, and Soloco, each with 5 apiaries (7.3%). Chachapoyas, Granada, Huancas, Leimebamba, Magdalena, and San Francisco de Daguas each hosted 4 apiaries (5.8%). Chiliquin, La Jalca, Mariscal Castilla, Montevideo, and Quinjalca registered 3 apiaries each (4.3%), while Levanto, San Isidro de Maino, and Sonche recorded 2 apiaries each (2.9%). Olleros had the lowest representation, with a single apiary (1.5%).

2.2. Data Collection and Predictor Variables

Sixty-nine beekeepers were surveyed from May to June 2022. The studied variables were honey production per harvest per hive per year (HP), time of experience as a beekeeper (TE), number of hives (NH), annual harvest (AH), visit frequency to the beekeeper for technical management of the hives (VF), location (L), instruction (I) and additional activity to beekeeping (AA) (Table 1). For the survey, all beekeepers in the province of Chachapoyas were included without applying any selection criteria or distinctions among them. The administered survey underwent evaluation through expert judgment, ensuring methodological rigor. To assess its internal consistency and reliability, Cronbach's Alpha was applied as a statistical measure ($\alpha = 0.05$) to have a 5% probability of making a Type I error.

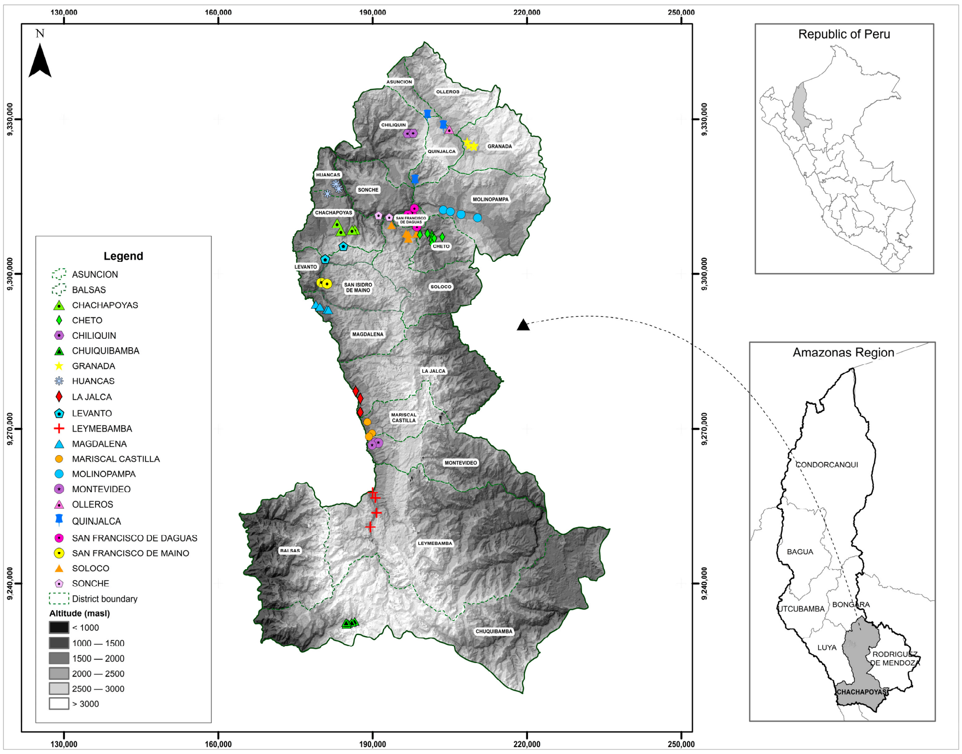

Figure 1. Location map of sampling points in the province of Chachapoyas.

Table 1. Descriptive statistics of variables.

Variables	Mean ± SD
Honey production (kg)	13.1 ± 5.2
Time of experience (year)	7.2 ± 5.9
Number of hives (n)	9.9 ± 14.4
Annual harvest (n)	1.5 ± 0.6
Visit frequency (n)	26.6 ± 16.9
Instruction	**Freq. (%)**
Primary	49 (71.0)
Secondary	12 (17.4)
Superior	8 (11.6)
Activity additional to beekeeping	**Freq. (%)**
Agriculture	32 (46.4)
Cattle raising	8 (11.6)
Construction	8 (11.6)
Agriculture and Cattle raising	6 (8.7)
Craft	3 (4.3)
None	6 (8.7)
Others	6 (8.7)

SD: standard deviation. Freq: frequency, None: only beekeeping, Others: primary or secondary educator, guardian, or trader.

2.3. Statistical Analysis

To evaluate the relationship between categorical variables associated with honey production in apiaries located in Chachapoyas, contingency tables and the Chi-square test ($p < 0.01$) were applied using SPSS v.15.0. Associations were examined among the following variables: harvests per year, visit frequency, additional activities, and beekeepers' instruction. Quantitative variables—years of experience, number of hives, and honey yield—did not follow a normal distribution (Shapiro–Wilk test). Therefore, measures of central tendency were calculated, and their variation across categorical groups was assessed using non-parametric tests: Kruskal–Wallis (for comparisons among more than two groups) and Wilcoxon rank-sum test (for two-group comparisons). Spearman's rank correlation ($p < 0.01$) was used to evaluate associations among quantitative variables.

An exploratory factor analysis was conducted in R v.4.4.3 using the FactoMineR package to identify underlying patterns and reduce the dimensionality of the dataset. The variables included in the analysis were: harvests per year, visit frequency, additional activities, instruction, years of experience, and number of hives. Principal Component Analysis (PCA) was employed for factor extraction, followed by orthogonal Varimax rotation. The adequacy of the factor model was assessed using the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity.

A linear mixed-effects model was fitted in R v.4.4.3 using the lme4 package to assess the joint influence of fixed effects: years of experience, instruction, number of hives, additional activities, harvests per year, and visit frequency. The district was included as a random effect. Honey production was log-transformed and used as the response variable. Multicollinearity among predictors was assessed using the vif() function from the car package. Model performance was evaluated using marginal and conditional R^2 coefficients, calculated with the pR2() function from the MuMin package.

3. Results

3.1. Characterization of Apiaries

Honey production per hive per harvest per year ranged from a minimum of 3 kg to a maximum of 25 kg, with an average yield of 13.1 kg. Table 1 presents the descriptive statistics of the variables analyzed.

3.2. Independence Analysis of Categorical Variables in Beekeeping Practices

The Chi-square test revealed a significant association between harvest frequency per year and VF, with two annual harvests predominating in apiaries visited every 30 days ($p < 0.01$). AA were significantly associated with both instruction and harvest frequency ($p < 0.01$). Apiaries managed by beekeepers with primary were predominantly linked to agricultural activities, whereas those with secondary or higher levels were more frequently associated with other types of additional activities, including handicrafts, services, and others (Table 2).

3.3. Correlation of Quantitative Variables in Beekeeping Practices

Spearman's rank correlation was applied to assess relationships among quantitative variables that did not follow a normal distribution. A significant positive correlation was found between years of experience and number of hives with honey production volume (Table 3).

3.4. Univariate Analysis and Comparison Using the Kruskal–Wallis Test

Mean values of the quantitative variables were calculated and compared across groups using non-parametric tests due to the lack of normal distribution. The Kruskal–Wallis test was applied for comparisons among more than two groups, and the Wilcoxon rank-sum test was used for two-group comparisons. Significant differences ($p < 0.01$) were observed among groups defined by harvest frequency per year, with higher values for years of experience, number of hives, and honey production volume in apiaries achieving three annual harvests. VF and AA were also significantly associated with the number of hives ($p < 0.01$) and honey production ($p < 0.05$). Apiaries visited more often (<15 days) and managed without additional activities—indicating exclusive dedication to beekeeping—exhibited the highest hive counts and honey yields (Table 4).

Table 2. Analysis of categorical variables independence in beekeeping practices [n (%)].

Variables	Total	Instruction			Visit Frequency		Annual Harvests		
		Primary	Secondary	Superior	<15 Days	>30 Days	1	2	3
Visit frequency									
<15 days	30.0 (43.5%)	20.0 (40.8%)	7.0 (58.3%)	3.0 (37.5%)					
>30 days	39.0 (56.5%)	29.0 (59.2%)	5.0 (41.7%)	5.0 (62.5%)					
p-value			0.51						
Annual harvests									
1 harvest/year	35.0 (50.7%)	25.0 (51.0%)	6.0 (50.0%)	4.0 (50.0%)	9.0 (30.0%)	26.0 (66.7%)			
2 harvest/year	31.0 (44.9%)	24.0 (49.0%)	4.0 (33.3%)	3.0 (37.5%)	18.0 (60.0%)	13.0 (33.3%)			
3 harvest/year	3.0 (4.3%)	- -	2.0 (16.7%)	1.0 (12.5%)	3.0 (10.0%)	- -			
p-value			0.09		<0.01				
Activity additional to beekeeping									
None	6.0 (8.7%)	2.0 (4.1%)	3.0 (25.0%)	1.0 (12.5%)	6.0 (20.0%)	- -	- -	4.0 (12.9%)	2.0 (66.7%)
Agriculture	32.0 (46.4%)	31.0 (63.3%)	1.0 (8.3%)	- -	14.0 (46.7%)	18.0 (46.2%)	17.0 (48.6%)	15.0 (48.4%)	- -
Agriculture and livestock	6.0 (8.7%)	6.0 (12.2%)	- -	- -	2.0 (6.7%)	4.0 (10.3%)	2.0 (5.7%)	4.0 (12.9%)	- -
Cattle raising	8.0 (11.6%)	6.0 (12.2%)	1.0 (8.3%)	1.0 (12.5%)	3.0 (10.0%)	5.0 (12.8%)	3.0 (8.6%)	5.0 (16.1%)	- -
Others	17.0 (24.6%)	4.0 (8.2%)	7.0 (58.3%)	6.0 (75.0%)	5.0 (16.7%)	12.0 (30.8%)	13.0 (37.1%)	3.0 (9.7%)	1.0 (33.3%)
p-value			<0.01		0.05			<0.01	

None: only beekeeping, Others: primary or secondary educator, guardian, or trader. Agriculture refers to farms dedicated exclusively to crop production; Livestock refers to those engaged solely in cattle raising; and Agriculture and Livestock refers to farms that combine both activities. The values presented correspond to the frequency, with the percentage of each category shown in parentheses. The associations were analyzed using the χ^2 test at a significance level of $p < 0.01$. N indicates the number of occurrences of beekeepers' agricultural practices, and (%) represents the corresponding percentage.

Table 3. Correlation among quantitative variables of beekeeping practices.

	Experience	Hives	Honey Production
Experience	1	0.659 (**)	0.618 (**)
p-value		<0.01	<0.01
Hives		1	0.933 (**)
p-value			<0.01
Honey production			1

** Significant difference < 0.01.

However, wide interquartile range suggest a degree of uncertainty in these estimates, warranting further analysis using multivariate approaches to better understand the underlying interactions.

3.5. Principal Component Analysis

The variance of factors potentially influencing honey production volume—as the response variable—was assessed using factorial analysis. Six variables (three categorical and three quantitative) were included to evaluate their joint contribution to overall variance. The analysis yielded an explained variance of 65.2% across two principal components, with a dispersed variable distribution (Figure 2). Beekeeping experience, hive count, and AH frequency loaded onto the first component; instruction and presence of additional activities onto the second; and visit frequency onto a third, less dominant component (Table 5).

Table 4. Years of experience, number of hives, and honey production by categorical variables.

Variables	N	Experience (Years)			Hives			Honey Production (kg)		
		Median	P25–P75%		Median	P25–P75%		Median	P25–P75%	
Instruction										
Primary	49.0	6.0	3.0	8.0	4.0	3.0	8.0	45.0	26.0	126.0
Secondary	12.0	5.0	3.0	13.75	5.5	3.0	15.25	56.0	33.0	253.75
Superior	8.0	6.5	4.25	17.75	6.0	3.5	21.25	58.0	30.5	212.5
p-value		0.34			0.54			0.78		
Annual harvests										
1 harvests/year	35.0	5.0	2.0	7.0	3.0	2.0	4.0	30.0	20.0	50.0
2 harvests/year	31.0	7.0	3.0	13.0	8.0	4.0	17.0	100.0	48.0	250.0
3 harvests/year	3.0	14.0	3.0	-	35.0	10.0	-	595.0	2	2171.72
p-value		0.01			<0.01			<0.01		
Visit frequency										
<15 days	30.0	5.0	2.0	12.0	8.0	3.0	12.5	104.0	36.0	251.25
>30 days	39.0	6.0	3.0	8.0	3.0	2.0	5.0	42.0	21.0	64.0
p-value		0.53			<0.01			0.01		
Additional activity to beekeeping										
None	6.0	8.5	3.0	14.75	31.0	10.0	52.75	567.0	99.9	1123.75
Agriculture	32.0	6.0	2.0	8.0	4.0	3.0	7.75	42.50	30.0	124.5
Agriculture and Cattle raising	6.0	7.0	3.0	14.75	4.5	2.0	38.0	61.50	14.5	709.0
Cattle raising	8.0	4.0	3.0	11.5	5.0	3.0	20.75	52.50	31.0	355.5
Others	17.0	5.0	3.5	14.0	3.0	2.5	7.0	36.0	20.5	65.0
p-value		0.66			0.01			0.03		

N (number), Median, and percentiles 25% y 75%. None: only beekeeping, Others: primary or secondary educator, guardian, or trader. Agriculture refers to farms dedicated exclusively to crop production; Livestock refers to those engaged solely in cattle raising; and Agriculture and Livestock refers to farms that combine both activities.

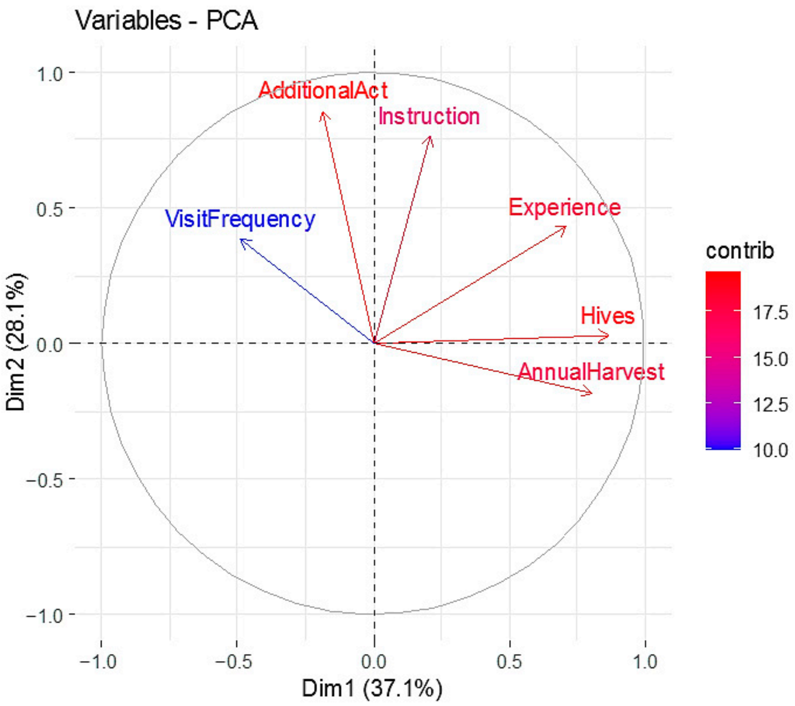

Figure 2. Explained Variance of Recorded Variables Affecting Honey Production.

The Kaiser-Meyer-Olkin measure yielded a value of 0.54, and Bartlett's test of sphericity was statistically significant ($p < 0.001$), supporting the suitability of factorial analysis for characterizing the selected variables.

Table 5. Rotated component matrix [1] of recorded variables related to honey production.

	PC1	PC2	PC3
Experience	0.4768625	0.3316564	0.4421595
Instruction	0.1408419	0.5884626	−0.4600598
Hives	0.5811738	0.0224140	0.2339021
Additional activity to beekeeping	−0.1240016	0.6604619	−0.1631398
Annual harvest	0.5407611	−0.1391135	−0.1334441
Visit frequency	−0.3274200	0.2960591	0.7026435

[1] Orthogonal Varimax rotation.

3.6. Honey Production Volume Estimation Model

Honey production was log-transformed and used as the response variable in a linear mixed-effects model. Fixed effects included beekeeping experience, instruction, number of hives, additional occupational activity, AH frequency, and visit frequency. District was specified as a random effect and contributed to the total variance (Table 6).

Table 6. Random effects contribution to honey production variance.

Variable	Name	Variance	Standard Deviation
Location	Intercept	0.01748	0.1322
Residual		0.33359	0.5776

Table 7 reports the significance levels of the fixed effects. No statistically significant associations were observed between honey production and either instruction or visit frequency. These variables were retained in the model due to their contribution to model fit metrics.

Table 7. Fixed effects of honey production estimation in apiaries in Chachapoyas, Peru.

	Estimate	SE	DF	t-Value	*p*-Value	Sig.
(Intercept)	3.267714	0.426214	56.936322	7.667	2.48×10^{-10}	***
Experience	0.046484	0.018286	59.317116	2.542	0.0137	*
Instruction	0.043957	0.131951	54.413455	0.333	0.7403	
Hives	0.047847	0.007644	61.031473	6.259	4.26×10^{-8}	***
Additional activity to beekeeping	−0.106635	0.052802	58.155984	−2.020	0.0481	*
Annual Harvest	0.404555	0.162787	58.895951	2.485	0.0158	*
Visit frequency	−0.014105	0.011605	58.919428	−1.215	0.2291	

Standard error (SE), degrees of freedom (DF), significance (Sig.), * indicates a significance level of 5% ($p < 0.05$), and *** indicates a significance level of 0.1% ($p < 0.001$).

The estimation equation met the assumption of residual normality, as indicated by the Shapiro–Wilk test ($p = 0.7385$). No evidence of multicollinearity was detected among the predictors, with variance inflation factor (VIF) values ranging from 1.43 to 2.34. The marginal and conditional coefficients of determination were similar ($R^2 = 0.778$ and 0.789, respectively), indicating consistent explanatory power from both fixed and random effects. Predicted values for each apiary were calculated using the coefficients presented in Table 7. Due to the logarithmic transformation applied during model fitting, the exponential function was used to back-transform the estimates. Figure 3 displays the observed and predicted honey production values. The correlation coefficient between observed and predicted values was R = 0.836.

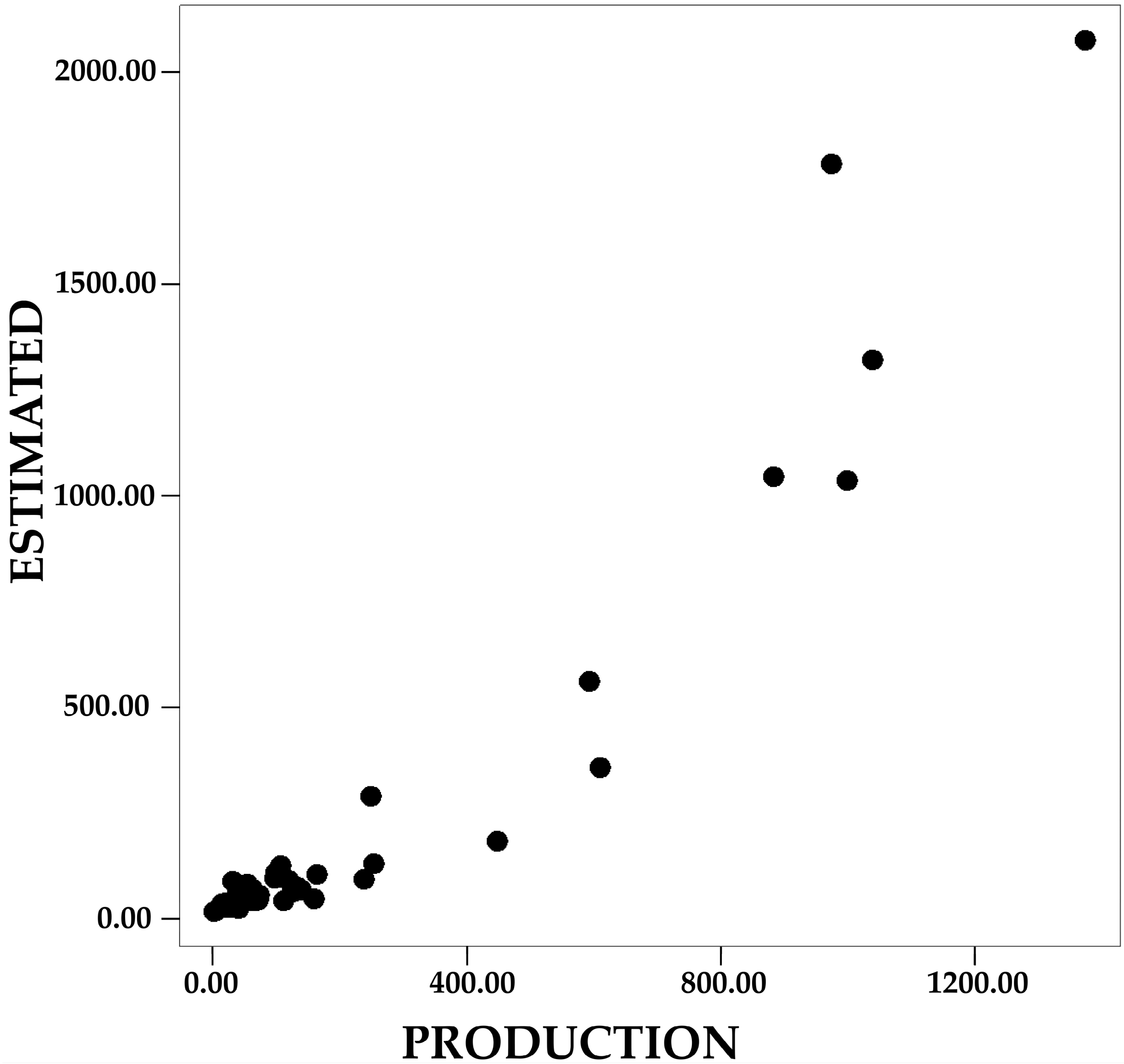

Figure 3. Correlation between observed and predicted honey production values using a linear mixed-effects model.

4. Discussion

4.1. Independence Analysis of Categorical Variables Affecting Honey Production

The independence analysis revealed statistically significant associations between management practices and the sociodemographic characteristics of beekeepers. AH was significantly associated with VF. This relationship indicates that more frequent management—defined as inspections every 15 days or less—is linked to achieving two to three harvests per year. Previous studies have reported that frequent hive inspections play a critical role in the early detection of diseases and in monitoring adverse climatic conditions [15,16]. These practices contribute to maintaining colony health [17,18] and are associated with improved yield outcomes [19,20].

Additional economic activities were also associated with both the beekeeper's level of instruction and the number of annual harvests. Beekeepers with primary tended to combine apiculture with agricultural work, whereas those with secondary or higher diversified into non-agricultural sectors such as crafts or services. These differences may reflect distinct economic sustainability strategies, as well as disparities in access to resources, technical knowledge, market opportunities, and technology adoption [21,22]. In Chachapoyas province, the most prevalent economic activities include agriculture, livestock, and tourism [23]. Beekeepers with lower educational attainment may maintain stronger ties to these traditional sectors, which are embedded in local cultural practices. Diversification into agricultural activities may reduce the time and resources available for apiculture, thereby influencing management frequency and harvest outcomes.

In this context, [24] reported that in southwestern Ethiopia, beekeepers engaged primarily in traditional practices and with limited technical specialization obtained lower yields, whereas those with greater specialization and organizational capacity achieved higher production and economic returns. These findings support the premise that time availability and resource allocation for specialized management are key factors influencing apiary performance.

4.2. Correlation of Quantitative Variables Affecting Honey Production

Spearman correlation analyses identified statistically significant associations between honey production and two key variables: years of beekeeping experience and the number of hives managed. The positive correlation between experience and honey production indicates that beekeepers with longer trajectories tend to achieve higher yields. This finding aligns with previous studies that have documented the role of accumulated experience in improving technical management and decision-making within apicultural systems [25]. Similarly, the number of hives was strongly correlated with honey production, suggesting that apiary scale is a relevant factor in yield outcomes. This association may reflect both the direct effect of colony count [26] and the potential for larger-scale beekeepers to adopt more frequent and technically intensive management practices [27].

Experience was also correlated with hive quantity, indicating that more experienced beekeepers tend to expand their production units over time [28]. These results support the interpretation that honey yield is influenced by the interaction between technical knowledge, management capacity, and production scale. In this context, continuous training and the implementation of scalable production models may contribute to improved yield performance in settings comparable to the Chachapoyas province.

4.3. Univariate Analysis and Mean Comparison Using the Kruskal–Wallis Test

Univariate analysis and mean comparison indicated that beekeeping experience, hive quantity, and honey production were significantly higher in apiaries that achieved three annual harvests. These findings reflect an association between increased management frequency and higher production levels and operational scale [29,30].

The results are consistent with previous studies that link strategic harvest scheduling and beekeeper experience with more effective use of floral resources and increased production output [31,32].

4.4. Principal Component Analysis

Principal component analysis revealed that the first component (Dim1), accounting for 37.1% of the total variance, was strongly influenced by variables directly related to beekeeping management: number of hives, beekeeper experience, and AH frequency. These variables may jointly represent an axis of "productive intensity" or "operational scale". In this regard, a study conducted under Lebanese conditions, using a comprehensive survey and Principal Component Analysis (PCA), revealed that beekeepers' knowledge, cooperative membership, number of hives, and honey storage practices significantly influenced production [33]. Furthermore, geographic region affected management practices, as most beekeepers engaged in year-round migratory beekeeping, highlighting both the challenges and opportunities for improving the sector [33].

The second component (Dim2), which explained 28.1% of the variance, grouped sociocultural variables such as the beekeeper's instruction and the presence of additional economic activities. This indicates that socioeconomic attributes of producers also contribute to explaining differences among beekeeping systems [34]. Furthermore, beekeepers' management philosophy is closely associated with their choice of hive chemicals and overall beekeeping objectives. This suggests that informed decision-making enables beekeepers to adopt management practices suited to their farm size and production philosophy [35].

In contrast, apiary visit frequency was associated with a distinct third component, emerging as a relevant variable that may reflect its transversal role across both technical management and beekeeper profile dimensions [36]. Previous studies, such as that of [37] have demonstrated that management practices and beekeeper characteristics, when analyzed jointly, account for a substantial portion of the variability observed in honey production, thereby supporting the findings of this study. This study demonstrates that management practices and beekeeper characteristics jointly explain a significant proportion of the variability among beekeeping operations. These findings support the notion that variables such as visit frequency can be integrated across both technical and beekeeper profile dimensions to promote sustainable beekeeping [38].

4.5. Honey Production Estimation Model

The applied linear mixed-effects model enabled simultaneous evaluation of multiple predictors of honey yield while accounting for district-level variability as a random intercept. The inclusion of district as a random effect was statistically justified by its significant contribution to the total variance (Table 6), highlighting the influence of geographic factors such as local climate, floral availability, and pest pressure on apiary performance [39].

Regarding fixed effects (Table 7), significant predictors of honey production included beekeeper experience ($p = 0.0137$), number of hives ($p < 0.0001$), AH frequency ($p = 0.018$), and exclusive dedication to beekeeping (absence of additional economic activities, $p = 0.0481$). These results align with patterns identified in univariate and multivariate analyses, confirming that honey yield is closely associated with operational scale, management intensity, and technical specialization [40]. Among these, hive quantity exhibited the strongest positive effect, reinforcing its role as a structurally determinant variable in production systems. Conversely, variables such as beekeeper instruction and apiary visit frequency did not exhibit statistically significant effects within the model. However, their inclusion improved overall model fit and were retained in the final specification. This indicates that while these variables may exert indirect or interaction-based influence, they do not independently account for substantial variance in honey yield [41].

Comparable models have been employed in previous studies, where experience, apiary size, and productive specialization were positively associated with higher yield levels, whereas educational attainment and management frequency were not consistently significant in multivariate frameworks [42]. Recent studies have also incorporated predictive modeling approaches to estimate honey production based on climatic variables, floral resource availability, and management characteristics [43,44]. These studies demonstrate the feasibility of yield forecasting and its potential utility for harvest planning and resource optimization, particularly in rural contexts analogous to Chachapoyas.

The apparent divergence among some statistical results can be attributed to the distinct analytical purposes and scopes of the applied methods. For instance, the Chi-square test revealed significant associations between harvest frequency and apiary visit frequency ($p < 0.01$), as well as between additional economic activities and educational level, indicating that inspection practices and livelihood strategies vary according to managerial and sociodemographic characteristics.

However, when honey production was analyzed using a linear mixed-effects model—which simultaneously incorporates multiple predictors and accounts for district-level variability as a random factor—neither education nor visit frequency exhibited statistically significant direct effects on yield. Although not significant as independent predictors, these variables were retained because they improved the overall model fit according to specific goodness-of-fit criteria.

This contrast reflects the fundamental difference between univariate or bivariate tests and multivariate modeling approaches. While Chi-square analyses detect simple associations, they do not account for confounding or interactive effects. In contrast, multivariate models estimate the adjusted influence of several variables concurrently, offering a more integrated and realistic representation of production dynamics. Consequently, factors that appear significant in isolated analyses may lose direct significance once technical, operational, and spatial sources of variation are controlled. This underscores the complexity of beekeeping systems and suggests the potential influence of latent factors or interactions not explicitly addressed in this study.

Model validation confirmed compliance with key statistical assumptions. The Shapiro–Wilk test applied to residuals ($p = 0.7385$) supported normality, and variance inflation factors (VIF < 10) ruled out significant multicollinearity among predictors. Marginal ($R^2 = 0.778$) and conditional ($R^2 = 0.789$) coefficients of determination indicated high explanatory power for both fixed and random effects, which is essential in mixed models incorporating unit-level heterogeneity (i.e., apiaries). Logarithmic transformation of the dependent variable during modeling, followed by back-transformation via the exponential function (exp()), improved data distribution and stabilized variance, yielding a more accurate fit. This precision was reflected in the strong correlation between predicted and observed honey yields ($R = 0.836$, $p < 0.01$), as illustrated in Figure 3. Despite the overall model performance, several outliers were detected. These may be attributable to external factors not explicitly modeled, including variation in management practices, seasonal floral dynamics, or localized climatic anomalies [45,46].

The weight information was neither collected nor included as a predictor in the honey production estimation model described. The linear mixed-effects model incorporated variables such as beekeeper experience, number of hives, frequency of agricultural applications, and exclusive dedication to beekeeping; however, weight data were not included as part of the fixed or random effects.

A major limitation of this study is the exclusion of the variable related to surrounding vegetation, which may represent a key determinant of beekeeping productivity. Although a positive relationship was identified between the beekeeper's educational background and honey yield, future research should incorporate analyses of the productive potential associated with the floristic composition and availability of nectar and pollen resources within the Amazonian landscape. Areas with greater diversity and abundance of melliferous flora are likely to support higher honey yields, regardless of other factors considered. Therefore, the omission of this environmental variable may constrain the generalizability and accuracy of the explanatory model in identifying the true determinants of beekeeping production.

5. Conclusions

This study identified that factors such as hive quantity, beekeeper experience, annual harvest frequency, and exclusive dedication to beekeeping are significantly associated with higher honey production levels (25 kg of honey per hive per harvest per year) in apiaries located in the province of Chachapoyas, Amazonas. Multivariate statistical analyses confirmed that these management-related variables—particularly hive quantity—explain a substantial proportion of the observed variability. Moreover, the inclusion of district as a random effect underscored the importance of geographic factors in accounting for differences in apiary performance. These findings support the implementation of strategies aimed at strengthening technical capacities, enhancing beekeeper professionalization, and promoting productive scaling to improve the efficiency and sustainability of apiculture in the region. Additionally, territorial specificities should be considered when designing interventions to ensure alignment with local conditions and optimize apiary productivity.

Author Contributions: Conceptualization, methodology, writing—review and editing, supervision, Y.M.B.-M.; data collection, methodology, investigation, J.B.G.-G.; formal analysis, validation, review J.A.S.-U. and H.A.Q.-C.; data curation, statistics, L.Q.H.; conceptualization, writing—review and editing, supervision, I.S.C.-C. All authors have read and agreed to the published version of the manuscript.

Funding: The APC was funded by the Vicerrectorado de Investigación—Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas. This research was funded by the National Council for Science, Technology and Technological Innovation (CONCYTEC) and the National Program for Scientific Research and Advanced Studies (PROCIENCIA) under call E077-2023-01-BM "Scholarship for Doctoral Programs in Interinstitutional Alliances", grant PE501089472-2024, under call E033-2023-01-BM "Interinstituional Alliances for Doctoral Programs", grant PE501084305-2023.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors like to thank the Doctoral Program in Sciences for Sustainable Development of the Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas. Also to the Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica (CONCYTEC) and the Programa Nacional de Investigación Científica y Estudios Avanzados (PROCIENCIA) within the framework of the call E033-2023-01-BM "Interinstitutional Alliances for Doctoral Programs", under grant number (PE501084305-2023).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Pérez-Castro, E.E.; May-Itzá, W.D.J.; Quezada-Euán, J.J.G. Thirty years after: A survey on the distribution and expansion of Africanized honey bees (*Apis mellifera*) in Peru. *J. Apic. Res.* **2002**, *41*, 69–73. [CrossRef]
2. Rasmussen, C.; Rimarachín, L.; Gáslac, E.M.; Gonzalez, V.H. Stingless Bees of Peru: The Use of Plant Resins, Cerumen, and Propolis. In *Stingless Bee Nest Cerumen and Propolis*; Vit, P., Bankova, V., Popova, M., Roubik, D.W., Eds.; Springer: Cham, Switzerland, 2024; Volume 1. [CrossRef]
3. Pérez-Castro, E.E. Evaluación Morfometrica y Molecular del Estado de Ia Africanización en Colonias Manejadas de *Apis mellifera* del Perú. Master's Thesis, Universidad Autónoma de Yucatán, Mérida, Mexico, 2001.
4. Vit, P.; Medina, M.; Eunice Enríquez, M. Quality standards for medicinal uses of Meliponinae honey in Guatemala, Mexico and Venezuela. *Bee World* **2004**, *85*, 2–5. [CrossRef]
5. Rodríguez-Malaver, A.J.; Rasmussen, C.; Gutiérrez, M.G.; Gil, F.; Nieves, B.; Vit, P. Properties of honey from ten species of Peruvian stingless bees. *Nat. Prod. Commun.* **2009**, *4*, 1221–1226. [CrossRef] [PubMed]
6. Hossain, M.L.; Lim, L.Y.; Hammer, K.; Hettiarachchi, D.; Locher, C.A. Review of Commonly Used Methodologies for Assessing the Antibacterial Activity of Honey and Honey Products. *Antibiotics* **2022**, *11*, 975. [CrossRef]
7. Na, S.J.; Kim, Y.K.; Park, J.M. Nectar Characteristics and Honey Production Potential of Five Rapeseed Cultivars and Two Wildflower Species in South Korea. *Plants* **2024**, *13*, 419. [CrossRef]
8. He, L.; Shen, L.; Zhang, J.; Li, R. Comprehensive Investigation of Fluoroquinolone Residues in Apis mellifera and Apis cerana Honey and Potential Risks to Consumers: A Five-Year Study (2014–2018) in Zhejiang Province, China. *Toxics* **2023**, *11*, 744. [CrossRef] [PubMed]
9. Tennakoon, T.M.I.U.K.; Vinodani, L.P.S.; Warnasooriya, W.M.S.N.; Amarasinghe, N.R.; Madushani, J.S. Physical, Chemical and Biological Characteristics of Sri Lankan Bee Honey Varieties. *Asian J. Complement. Altern. Med.* **2023**, *11*, 15–23. [CrossRef]
10. Ye, H.; Yang, J.; Xiao, G.; Zhao, Y.; Li, Z.; Bai, W.; Zeng, X.; Dong, H. A comprehensive overview of emerging techniques and chemometrics for authenticity and traceability of animal-derived food. *Food Chem.* **2023**, *402*, 134216. [CrossRef]
11. Harianja, A.H.; Adalina, Y.; Pasaribu, G.; Winarni, I.; Maharani, R.; Fernandes, A.; Saragih, G.S.; Fauzi, R.; Tampubolon, A.P.; Njurumana, G.N.; et al. Potential of Beekeeping to Support the Livelihood, Economy, Society, and Environment of Indonesia. *Forests* **2023**, *14*, 321. [CrossRef]

12. Robustillo, M.C.; Naranjo, L.; Parra, M.I.; Pérez, C.J. Addressing multidimensional highly correlated data for forecasting in precision beekeeping. *Comput. Electron. Agric.* **2024**, *226*, 109390. [CrossRef]
13. Aksoy, A.; Ertürk, Y.E.; Erdoğan, S.; Eyduran, E.; Tariq, M.M. Estimation of honey production in beekeeping enterprises from eastern part of Turkey through some data mining algorithms. *Pak. J. Zool.* **2018**, *50*, 2199–2207. [CrossRef]
14. Karadas, K.; Kadirhanogullari, I.H. Predicting honey production using data mining and artificial neural network algorithms in apiculture. *Pak. J. Zool.* **2017**, *49*, 1611–1619. [CrossRef]
15. Landaverde, R.; Rodriguez, M.T.; Parrella, J.A. Honey Production and Climate Change: Beekeepers' Perceptions, Farm Adaptation Strategies, and Information Needs. *Insects* **2023**, *14*, 493. [CrossRef]
16. Scott, A.; Hassler, E.; Formato, G.; Rünzel, M.A.S.; Wilkes, J.; Hassan, A.; Cazier, J. Data mining hive inspections: More frequently inspected honey bee colonies have higher over-winter survival rates. *J. Apic. Res.* **2023**, *62*, 983–991. [CrossRef]
17. Kim, H.; Frunze, O.; Maigoro, A.Y.; Lee, M.L.; Lee, J.H.; Kwon, H.W. Comparative Study of the Effect of Pollen Substitute Diets on Honey Bees During Early Spring. *Insects* **2024**, *15*, 101. [CrossRef] [PubMed]
18. Mishukovskaya, G.; Giniyatullin, M.; Shelekhov, D.; Khabirov, A.; Smol-Nikova, E.; Naurazbaeva, A. The Use of Probiotics in Spring Supplementary Feeding of Bee Colonies. *Bulg. J. Agric. Sci.* **2023**, *29*, 131–137.
19. Kasese, L.J.; Mabhuye, E.B.; Ndaki, P. Determinants of Honey Production in Dodoma Region, Central Tanzania A case of Bahi and Kongwa districts. *Discov. Sustain.* **2025**. [CrossRef]
20. Tadesse, B.; Tilahun, Y.; Woyamo, W.; Bayu, M.; Adimasu, Z. Factors influencing organic honey production level and marketing: Evidence from southwest Ethiopia. *Heliyon* **2021**, *7*, e07975. [CrossRef]
21. Bihonegn, A.; Begna, D. Beekeeping Production System, Challenges, and Opportunities in Selected Districts of South Wollo Zone, Amhara, Ethiopia. *Adv. Agric.* **2021**, *2021*, 2278396. [CrossRef]
22. Mulatu, A.; Marisennayya, S.; Bojago, E. Adoption of Modern Hive Beekeeping Technology: The Case of Kacha-Birra Woreda, Kembata Tembaro Zone, Southern Ethiopia. *Adv. Agric.* **2021**, *2021*, 4714020. [CrossRef]
23. Boletín Estadístico Mensual "EL AGRO EN CIFRAS"—2024. 2025. Available online: https://www.gob.pe/institucion/midagri/informes-publicaciones/5380407-boletin-estadistico-mensual-el-agro-en-cifras-2024 (accessed on 10 August 2025).
24. Bayissa, M.; Lauwers, L.; Mitiku, F.; de Graaf, D.C.; Verbeke, W. System Mapping of the Production and Value Chain to Explore Beekeeping Potential in Southwest Ethiopia. *Insects* **2024**, *15*, 106. [CrossRef]
25. Deksisa, K.; Gemechu, A.; Tolera, T. Does cluster beekeeping improve the efficiency of honey production in participant households in southwestern Ethiopia? *Heliyon* **2024**, *10*, e38651. [CrossRef]
26. Tarekegn, K.; Ayele, A. Impact of improved beehives technology adoption on honey production efficiency: Empirical evidence from Southern Ethiopia. *Agric. Food Secur.* **2020**, *9*, 7. [CrossRef]
27. Nganso, B.T.; Soroker, V.; Osabutey, A.F.; Pirk, C.W.W.; Johansson, T.; Elie, N.; Otieno-Ayayo, N.; Ibrahim, M.M.; Ndungu, N.N.; Ayalew, W.; et al. Best practices for colony management: A neglected aspect for improving honey bee colony health and productivity in Africa. *J. Apic. Res.* **2024**, *63*, 438–455. [CrossRef]
28. Vásquez-Ibarra, L.; Iriarte, A.; Rebolledo-Leiva, R.; González-Araya, M.C.; Angulo-Meza, L. Determining the key factors that contribute to the eco-efficiency level of honey production using life cycle approaches. *Int. J. Life Cycle Assess.* **2023**, *28*, 1533–1543. [CrossRef]
29. El Agrebi, N.; Steinhauer, N.; Tosi, S.; Leinartz, L.; de Graaf, D.C.; Saegerman, C. Risk and protective indicators of beekeeping management practices. *Sci. Total Environ.* **2021**, *799*, 149381. [CrossRef]
30. Olate-Olave, V.R.; Verde, M.; Vallejos, L.; Raymonda, L.P.; Cortese, M.C.; Doorn, M. Bee health and productivity in apis mellifera, a consequence of multiple factors. *Vet. Sci.* **2021**, *8*, 76. [CrossRef] [PubMed]
31. Bareke, T.; Roba, K.; Addi, A. Diversity of Bee Floral Resources and Honey Production Calendar in Ethiopia's Southwest Shoa Zone. *Adv. Agric.* **2024**, *2024*, 5428576. [CrossRef]
32. Gounari, S.; Zotos, C.E.; Dafnis, S.D.; Moschidis, G.; Papadopoulos, G.K. On the impact of critical factors to honeydew honey production: The case of Marchalina hellenica and pine honey. *J. Apic. Res.* **2023**, *62*, 383–393. [CrossRef]
33. El Hajj, R.; Nemer, N.; Skaff, W.; Estephan, N. Multivariate approach to analyzing survey data: A case study of beekeeping in Lebanon. *J. Apic. Res.* **2021**, *62*, 459–467. [CrossRef]
34. Andaregie, A.; Astatkie, T. Determinants of beekeeping adoption by smallholder rural households in Northwest Ethiopia. *Cogent Food Agric.* **2021**, *7*, 1954817. [CrossRef]
35. Underwood, R.M.; Traver, B.E.; López-Uribe, M.M. Beekeeping management practices are associated with operation size and beekeepers' philosophy towards in-hive chemicals. *Insects* **2019**, *10*, 10. [CrossRef]
36. Nayak, J.; Mohanty, L.; Saha, G.; Devi, M. Socio-Economic Status and Opportunities of Beekeeping for Farm Women: A Case Study. *Indian J. Entomol.* **2024**, *32*, e24448. [CrossRef]
37. Karakaya, E.; İnci, H.; Topluk, O. Determination of the Factors Affecting the Honey Production Per Coloney in Bingöl Beekeeping Enterprises. *Yuz. Yil Univ. J. Agric. Sci.* **2023**, *33*, 429–440. [CrossRef]

38. Laynes-Magaña, C.; Cetzal-Ix, W.; López-Castilla, H.M.; Tamayo-Cen, I.; Martínez-Puc, J.F.; Noguera-Savelli, E.; Basu, S.K. Honey production, an economic alternative for coastal areas with mangrove ecosystems: A case study in Sabancuy, Campeche, southeastern Mexico. *J. Ethnobiol. Ethnomed.* **2025**, *21*, 21–42. [CrossRef] [PubMed]
39. Mouillard-Lample, L.; Gonella, G.; Decourtye, A.; Henry, M.; Barnaud, C. Competition between wild and honey bees: Floral resources as a common good providing multiple ecosystem services. *Ecosyst. Serv.* **2023**, *62*, 101538. [CrossRef]
40. Kulekçi, M.; Eyduran, E.; Altın, A.Y.; Tariq, M.M. Usefulness of MARS and Bagging MARS Algorithms in Prediction of Honey Production in Beekeeping Enterprises from Elazig Province of Turkey. *Pak. J. Zool.* **2022**, *54*, 1087–1093. [CrossRef]
41. Kim, J.; Lee, J.; Seo, J.; Ghim, Y.C.; Lee, Y.; Na, Y.S. Enhancing disruption prediction through Bayesian neural network in KSTAR. *Plasma Phys. Control. Fusion* **2024**, *66*, 075001. [CrossRef]
42. Guiné, R.P.F.; Oliveira, J.; Coelho, C.; Costa, D.T.; Correia, P.; Correia, H.E.; Dahle, B.; Oddie, M.; Raiments, R.; Karise, R.; et al. Professional Training in Beekeeping: A Cross-Country Survey to Identify Learning Opportunities. *Sustainability* **2023**, *15*, 8953. [CrossRef]
43. Brini, A.; Giovannini, E.; Smaniotto, E. A Machine Learning Approach to Forecasting Honey Production with Tree-Based Methods. *Quant. Methods Soc. Sci.* **2025**, *1*, 1–25. [CrossRef]
44. Gajardo-Rojas, M.; Muñoz, A.A.; Barichivich, J.; Klock-Barría, K.; Gayo, E.M.; Fontúrbel, F.E.; Olea, M.; Lucas, C.M.; Veas, C. Declining honey production and beekeeper adaptation to climate change in Chile. *Prog. Phys. Geogr.* **2022**, *46*, 737–756. [CrossRef]
45. Kileo, G.D.; Masao, C.A.; Andrew, S.M. Honey and beeswax production amidst climate variability in two contrasting agro-ecologies of Tanzania. *Int. J. Biometeorol.* **2025**, *69*, 2369–2381. [CrossRef] [PubMed]
46. Masaka, K. Yearly Fluctuations in Honey Production in Hokkaido, Northern Japan, with Special Reference to Weather Conditions and Masting Behavior. *Int. J. Food Agric. Nat. Resour.* **2023**, *4*, 62–68. [CrossRef]