| Pages: 313-329 | | Vol. 6, No. 2 | | (2023) |

OEconomia | Research Article |

Good Food Score Reflecting Health Status of the Families: A Predicting Approach through Machine Learning

[1]Dr. Hina Ali, [2]Hina Shafiq, [3]Zinaz Aisha, [4]Maria, Iftikhar ,[5]Nazish Iftikhar, [6]Hira Ali

1. Associate Professor, Department of Economics, The Women University Multan. **2,3.** Assistant Professor, Department of Economics, Sardar Bahadur Khan Women's University Quetta, Pakistan.**4,5.** PhD Scholar, Department of Economics, the Women University Multan. **6.** Assistant Professor, Department of English, the Women University Multan, Pakistan.

ABSTRAT

The current scenario investigates the different factors that affect the Household Health Status (HHS) in Southern Punjab Pakistan. Household Good Food Score (GFS) indicate a family's financial access to a variety of foods to meet their nutritional requirements. The primary source of information was 300 instances of households. A particularly coordinated overview for get-together responses was made according to FAO rules. The Weka classification for logistic regression was used to analyze the data. The cross validation estimated coefficients of the GFS, household head income, education level, and covid-19 status all have positive effects on HHS while the attribute household head age and size has a negative effect. The findings of that study indicate the correctly classified instances that out of 300 instances, 298 instances are correctly classified. So, 99% of that is the accuracy of this particular classifier. So, the value of the kappa statistic is 0.99 and it is considered a very good value. The study also suggested key policy recommendations to improve HHS and socioeconomic factors that affect households' health scores.

Keywords:

Household health status, good food score, household socio-economics characteristics, logistic regression, Southern Punjab.

Introduction

A sustainable diet includes economic, social, cultural, and environmental factors in addition to nutrition and health. Sustainable diets are not only environmentally friendly but also cost-effective, socially acceptable, and healthy. Worldwide and regionally, it is anticipated that by 2020, a greater number of people will not be able to get good quality food daily. Nearly three point one billion more people, or 112 million more people, were unable to afford a healthy diet in 2020 than in 2019. With seventy-eight million unable to get good quality food in 2020, Asia was primarily to blame for this rise. Africa came next, with 25 million more people, followed by Europe, North & Latin America, & the Caribbean countries, with more than eight million people (FAO, 2022; Thamrin et al., 2021; WHO, 2022).

Poor nutrition, polluted water, and poor hygiene are the main factors that contribute to both

communicable and non-communicable diseases. Inadequate medical facilities are also deteriorating the health of Pakistan's underprivileged population. Due to its overpopulation, filthy water, poor sanitation, lack of health awareness, and inadequate vaccination coverage, Pakistan is at high risk for epidemics brought on by communicable diseases and accounts for a significant portion of the regional burden of many of these diseases. With 7.6% of its population infected with hepatitis B or C, Pakistan is experiencing a double epidemic; the world's fifth-highest prevalence of tuberculosis and fifth-highest prevalence of malaria (Hafeez et al., 2023; WHO, 2021).
Even though it is undeniable that eating a healthy diet is important, there isn't much information about what foods are available right now. For instance, it is debatable whether an unhealthy diet is brought on by a lack of access. The availability of healthy food options is limited and prices rise in socioeconomically disadvantaged areas. Nutritious food sources frequently come with high prices that low-income families cannot afford. Concerns regarding the instability of food are growing. A region's food security is affected by climate change and variability, land degradation, conflicts caused by humans, economic downturns, or spikes in food prices (economic access). To ensure that everyone has access to food and end hunger, multiple political commitments have been made to the second goal of zero hunger of the United Nations Sustainable Development Goals (UN, 2019). However, achieving the Sustainable Development Goal of ending hunger is challenging. It is necessary to take immediate action if the goal is to be achieved by 2030 (Suiçmez et al., 2023; Westerveld et al., 2021).
The current study employs decision tree classification, which is based on machine learning, to determine whether the various behavioral and cognitive characteristics of distinct eating patterns, as well as their interactions, can be used to distinguish between individuals who engage in recurrent binge eating and those who do not. The decision tree method, which is based on machine learning, has a big advantage because the model finds the best combinations of variables to improve group membership classification. Specialists have demonstrated both direct and nonlinear connections between different healthful elements that influence us (Baitharu & Pani, 2016; Karakan et al., 2022; Linardon et al., 2020).
Positive health is the focus of scientific research into health assets. A factor that enhances health in addition to disease risk factors is a "health asset." Positive Health aims to determine which health assets reduce disease risk and lower healthcare costs while also producing longer, healthier lives. Health determinants include the individual's traits and actions, the physical environment, and the social and economic environment. Human health can be affected by many things. A few examples include housing, financial security, community safety, employment, education, and the environment. These factors are referred to as the broader health determinants (Farooq et al., 2023; Iqbal et al., 2023; Rastogi & Bansal, 2023; Shaikh et al., 2023; Suiçmez et al., 2023).
As the body of evidence regarding the effects of a diet that is not balanced and healthy on health outcomes like obesity, cancer, diabetes, and cardiovascular disease grows, policymakers are interested in interventions that encourage healthy eating patterns. Taste, social similarity, assortment, and moderation are exceptionally significant contemplations when choosing good food. According to numerous studies, customers know about the differentiations between sound

and undesirable food varieties and would lean toward good food varieties at a similar cost. The assembled climate, which is portrayed by financial elements like the accessibility, amount, and availability of refreshing food retailers, is remembered to affect the utilization of good food, as per studies (Birk et al., 2021; Kavitha et al., 2022; Thamrin et al., 2021). Learning more about personalized diets in nutrition and wellness is being done with the help of a disruptive technology called artificial intelligence (AI). In this manner, the ongoing review utilizes AI to recognize different elements that impact families' great wellbeing status in the metropolitan southern Punjab of Pakistan.

Literature Review

This section discusses theoretical clarification and the empirical literature. It discusses the concepts of household characteristics and health status scores.

Shiao et al. (2018) investigated the role of personalized nutrition genes, diet, and the interactive parameters that go along with them as cancer predictors in families with multiethnic colorectal cancer. The principal point of this is to look at five vital qualities in the folate digestion pathway, as well as dietary boundaries and related intelligent boundaries, as indicators of colorectal disease, by estimating the smart dieting file in multiethnic families.

Linardon et al. (2020) analyzed how different eating patterns affected each other to tell recurrent from nonrecurrent binge eating. This study used a decision tree classification analysis based on machine learning on 1,341 participants. the outcomes demonstrate that roughly 70% of members could be precisely named with or without repetitively voraciously consuming food. The gathering probably going to be named having repetitive gorging was the people who had high dichotomous reasoning scores and low instinctive eating scores. This study suggests that the interaction of multiple eating patterns may be able to tell the difference between recurrent binge eating and not having it. Corroborative examinations are expected to test the intelligent speculations created by these exploratory investigations.

Amin et al. (2021) examined many households in the United States that rely on inexpensive, processed foods with low nutritional value because they are unable to obtain healthy food. Due to the multidimensional nature of socioeconomic variables, this study demonstrated that access to healthy food is costly and difficult to quantify. The modified retail food environment index, or the percentage of healthful food retailers in a tract, was predicted using data from census tracts and machine learning for data analysis in this study. The modified retail food environment index is predicted through the use of machine learning. To find the qualities of a "food swamp" with low access and a "food desert" with no entrance. 72% of the sample correctly predicted food deserts and food swamps, according to the final results. Therefore, food deserts and food swamps are distinct concepts that call for distinct policy considerations. The results also demonstrated that population size, the presence of people of color, property values, and pay are all registration plots in the US that contribute to general admission to strengthening food retailers. This model can be used to precisely predict admission to refreshing food retailers.

Menichetti et al. (2021) analyzed the evidence that eating more super-handled food is bad for health, but it is still hard to define handled food. It is true that the food grouping that is based on the ebb and flow handling only includes a small number of items and does not differentiate between different levels of handling. This makes it difficult for customers to make decisions and makes it easier for research into the effects that handled food has on health. A machine learning algorithm was used to estimate the data in this study. Over 73% of the food supply in the United States is ultra-processed, as this algorithm accurately predicts the degree of processing. This study found that nutrient bioavailability decreases and a person's risk of metabolic disease, diabetes, angina, high blood pressure, and natural aging rises when their diet is more tightly controlled. This study's findings suggested that consumers might benefit from having access to information about the level of processing they currently lack. Replacing highly processed foods with less processed ones could improve population health.

Gao et al. (2023) analyzed the relationship between hepatic fibrosis and the quality of the diet. This study utilized cross-sectional connections between three deduced diet quality scores. The dietary approaches score, the alternative healthy eating index, and a modified style diet score, as well as the hepatic fat. For the estimation of the data analysis, this study used a logistic regression model. The finding of this indicates that there is a correlation between a higher-quality diet and better hepatic fat and fibrosis measurements. Our findings suggest that a healthy diet may reduce the risk of obesity, hepatic steatosis, and fibrosis progression.

The summary of this chapter examines the body of health status literature. In metropolitan areas of Southern Punjab, Pakistan, the present study focuses to estimate health status as well as their determinants, which have an impact on household health scores. Therefore, the current study aims to investigate the household's good and poor health scores.

Materials and Methods

This. Section describes the data collection procedures, which include the research design, sampling strategies, sample size, and total data collection. Next present the implemented econometric models and variable descriptions.

Data Collection and Methods

The current study relies on primary data collected from the Lodhran and Vehari district in Southern Punjab. A well-structured questionnaire for collecting responses was developed by FAO (2018) guidelines. To assess the impact of household health status, the good food score of each respondent was gathered for this review by FAO guidelines. In addition, the age of household heads, income, education level, family size, and Covid-19 status of respondents were gathered for socioeconomic characteristics.

The use of multistage sampling in the investigation increased the researchers' chances of obtaining representative sample data by allowing them to collect data from a variety of household groups. Toward the start, Tehsil Lodhran and Vehari were chosen to gather information. For the following round, ten states were selected randomly from metropolitan areas. Thirty respondents were selected at random from each colony during the final stage. As a result, 300 households made up the entire sample size. Through a series of questions, the interviewer gathered information about the health score status.

Description Variables

The variables' descriptions are clarified in this section. The target variables are the health status of the household in addition to other socioeconomic attributes that affect the household health score.

Dependent Variable

Household Health Status (HHS)

A person's health status, which can be rated as excellent, very good, good, fair, or poor, reflects their perception of their health. In our study household health status was taken in dummy form good means yes and otherwise zero (Dong et al., 2023; Gao et al., 2023).

Independent Variables

Good Food score (GFS)

It is a straightforward measure of the overall quality of a diet that gives a score of lower, medium, and higher depending on whether a food recommended in the dietary guidelines is eaten more than once per week (FAO, 2018).

Age of the Household Head

A household is a social unit that consists of several individuals who live independently or in groups. The head of the household is in charge of taking care of the family and earning money to keep the house running. At the hour of information assortment, the age of the family head was kept in the number of years.

Income of the Households

At the time of the interviews, total income is the sum of all incomes earned by all members of the household from all sources over the previous year. In the data, the household head's income is shown in Pakistani rupees. As a result, data on the households' total incomes were gathered.

Education level of the Household

The head of the household was educated over several schooling years. Knowing the household head's education at the time of the interview.

Family Size of Household

The family size of family addresses the number of people that are living respectively in one house. This indicates that households with lower family members have good health scores as compared to more families.

Covid-19 Status of Household

In this study, the covid-19 vaccination status was taken in dummy form if yes then one otherwise zero.

Econometric Model

This study will use the logistic regression method for data analysis because the nature of this data is mostly suitable to use this logistic regression method (Akbar et al., 2023; Dong et al., 2023; Gao et al., 2023). The Weka software (ver. 3.86) has been used for model estimates because it is a collection of data mining-related machine learning algorithms. It contains apparatuses for information arrangement, characterization, relapse, grouping, affiliation rules mining, and representation.

Household health status (HHS) is taken as a predicted (Y) Variable that is dichotomous (binary) in nature. It takes two values 0 and 1. One shows that a health score status of good or zero shows that there is poor health status in the household.

Y =1 If Y > 0

Y =0 If Y < 0.

Many Socioeconomic factors can affect HHS. So, this study used the logistic regression method for data analysis that has been described, the probability of HHS.

$$p = \frac{1}{1+e^{-z}}$$

- p_i : Probability of HHS
- z: taken as a function of independent variables

$$z = B_0 + B_1 X_1 + B_2 X_2 + \ldots + B_k X_k$$

$$1 - p = \frac{1}{1-e^{z}}$$

$1 - p_i$: Probability of having no food choice

The logit model is expressed in terms of logs form is:

$$Ln\left[\frac{P_i}{1-P_i}\right] = B_0 + B_1 X1_i + B_2 X2_i + \ldots\ldots\ldots + Bn_i Xn_i \qquad (1)$$

$Ln\left[\frac{P_i}{1-P_i}\right]$ Shows that the odd ratio of HHS is in favor and also tells the ratio of probability between good health and poor health. ith subscript shows observations that are present in data. Intercept term refers B_0 in this model. $X_1, X_2, X_3, \ldots\ldots\ldots, X_n$ are taken as an explanatory variable, that is GFS, the age of household heads, income, education level, family size, and Covid-19 status.

Empirical Results

The interpretation of the empirical results for various economic models is provided in section 4. Weka's summary of the dataset and the descriptive statistic of variables, socio-demographic information about all univariate attribute distribution households, and zero-r and naive Bayes classifiers in the empirical analysis are presented in next sections.The logistic model's regression algorithm results are discussed.

Weka Summary of Dataset and Descriptive Statistic

The summary of the dataset in the current relation indicates the details about the dataset included relation (name of the data set), instances show the number of rows, and attributes represent the number of columns this attribute shows the summary of descriptive statistics. The descriptive of all the variables used in data estimation is presented in Table

Table 1: Descriptive Statistic

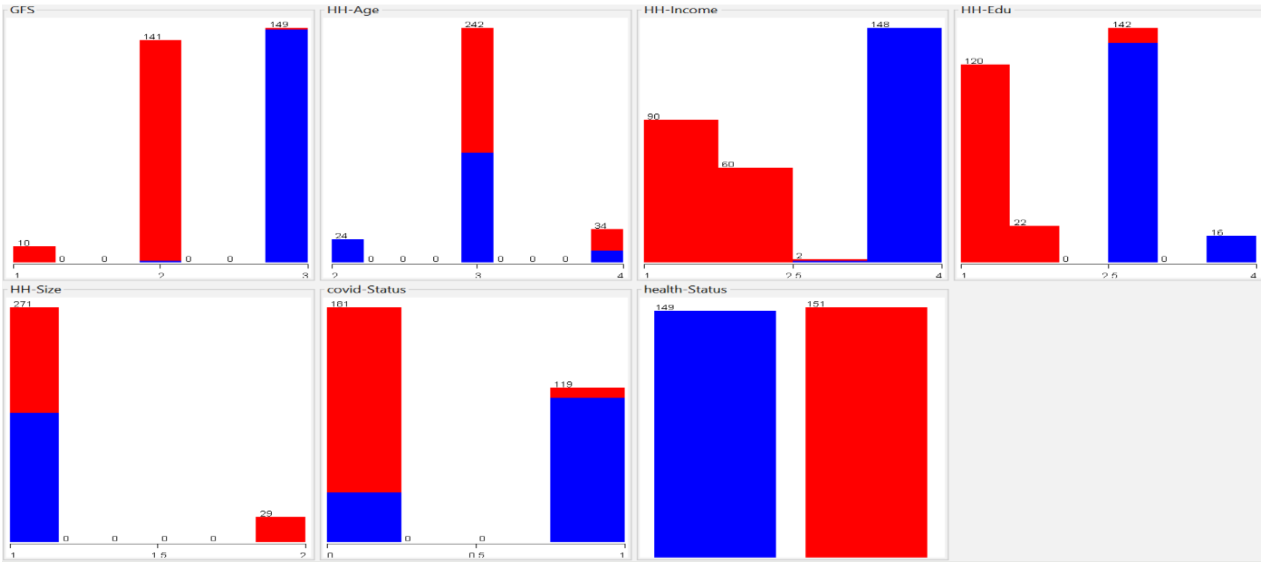
Attributes	NO	Label	Count	Weight
HHS (Y)	1	Good	149	149
	2	Poor	151	151
Variables	**Mean**	**Std-Dev.**	**Min**	**Max**
GFS (X1)	2.463	0.562	1	3
Age (X2)	3.033	0.439	2	4
Income (X3)	2.693	1.344	1	4
Education level (X4)	2.18	1.028	1	4
Family-size (X5)	1.097	0.296	1	2
Covid-19 status (X6)	0.397	0.49	0	1

Table shows the nominal attributes of HHS, we now see that a list of each category and the number of instances for each category are provided. Weightings are also mentioned, this is to give particular instances or attribute values in the dataset more or less weight. It also explains the household demographic factors and mean value of the all variables that are used in the model estimation. The mean value of the attributes X1, 2.463, and X2 is 3.033 while the min value 1 and 2 and max values are 3 and 4 respectively. The mean value of the attributes X3, 2.693, and X4 is 2.18 while the minimum value both are 1 and the maximum values are 4 respectively. The average value of the attributes X5 is 1.097 and X6 0.397 while the min value is one and zero and the max values are 2 and 1 respectively.

Weka All Univariate Attribute Distribution

Figure "A" indicates the all-univariate attribute distribution of the households. The blue color has a positive impact on human health and increases the health status and value of GFS of families. So, it indicates that health status is good and positive impact on households. And a red color indicates a poor and negative impact on households' health status. In the below picture, the good health score is 149 because as per the above dataset family's health status attribute has 149 values that are good and the poor health score is 151. GFS are divided into three groups according to the guidelines of FAO. Group, one indicates a lower GFS that is ten and the value of medium GFS is one forty-one, and three ranges indicate a higher GFS that is one forty-nine and represented in the blue color that has a positive and good impact on household health status.

Figure A: Weka All Univariate Attribute Distribution



Zero-R

Zero-R is the simplest classification technique that ignores all predictors and relies on the target. The Zero Rule algorithm, also known as Zero-R, is a good starting point for both classification and regression problems. The Zero Rule algorithm uses the training dataset's mean to predict a numerical value for a regression predictive modeling problem. The Zero-R classifier only makes predictions for the most common category (class). Zero-R is valuable for setting a gauge execution standard for other order techniques, even though it needs consistency power. Let's demonstrate how the household health status prediction problem affects the Zero Rule algorithm. Table 4.2 indicates that out of 300 instances, we can correctly classify 149. So, other classification algorithms if they perform less than this would perform worse. The rule zero R predicts the poor health score and shows the classification accuracy is 49 percent when one class is assigned to each prediction from two equally represented classes. To demonstrate its expertise, any machine learning algorithm attempting to resolve this issue must achieve an accuracy greater than this value. The dataset is divided into ten parts during the Cross Validation process. The first nine parts are used to train the algorithm, and the tenth part is used to evaluate it. By repeating this procedure, each of the ten parts of the split dataset has a chance to be the held-out test set.

Table 2: Weka Classifiers Rules Zero r

Scheme: Weka. Classifiers. Rules. Zero-R									
Relation: Health-Status **Cases:** 300 **Attributes:** 7 (GFS, HH-Age, HH-Income, HH-Edu, HH-Size, covid-Status & health-Status) **Test mode:** Cross-Validation of Tenfold **Model for a classifier:** (full training set)									
Zero-R predicts class value: poor									
Model construction time: 0 seconds **Cross-validation stratification Summary**									
Correctly-Classified-Instances:149 (49.7%)					**Incorrectly-Classified-Instances:** 151(50.3%)				
Kappa Measurement	**Error Absolute Mean**		**Error² Root mean**		**Absolute & relative error**	**Relative root error²**		**Total # of Cases**	
-0.0121	0.5		0.5		100 %	100 %		300	
Detailed Precision by Class									
	[Rate TP]	**[Rate FP]**	**[Accuracy]**	**[Recall]**	**[Measure F]**	**[MCC]**	**[ROC Area]**	**[PRC Area]**	**[Set]**
	0.094	0.106	0.467	0.094	0.156	-0.020	0.494	0.494	Good
	0.894	0.906	0.500	0.894	0.641	-0.020	0.494	0.500	Poor
weighted Avg.	0.497	0.509	0.483	0.497	0.400	-0.020	0.494	0.497	

[Confusion-Matrix]

		Actual value is categorized as	
		A	**B**
Predicted value	**a = good**	14	135
	b = poor	16	135

Zero-R only correctly predicts the majority class, as shown by the confusion matrix that follows. Zero-R is only useful for selecting a standard execution for other characterization techniques, as previously mentioned.

| Pages: 313-329 | | Vol. 6, No. 2 | | (2023) |

OEconomia | Research Article |

Naive Bayes Classifier

It is a supervised learning algorithm that is used to solve classification problems and is based on the Bayes theorem. It predicts the class with the most vital probability by registering the back probability for each class. As a result, it can solve problems involving multiple classes as well as binary classification. Table indicates that out of 300 instances, we can correctly classify 299. The kappa statistic is 0.993 So, this 99% is better than the baseline classifier.

Table 3: Weka Classification Results for the Naive Bayes Algorithm

[Scheme: Weka. Classifiers. Bayes. Naïve-Bayes] [**Relation:** Health-Status] [**Instances**: 300] **Attributes:** 7 (GFS, HH-Age, HH-Income, HH-Edu, HH-Size, covid-Status & health-Status) **Test mode:** Cross-Validation of Tenfold **Model for a Classifier** (full training set)								
Naive Bayes Classifier								
Attribute	**Class Good (0.5)**				**Class Poor: (0.5)**			
	Mean	**std. dev.**	**weight sum**	**Precision**	**Mean**	**std. dev.**	**weight sum**	**Precision**
GFS	2.99	0.17	149	1	1.94	0.26	151	1
HH-Age	2.92	0.49	149	1	3.15	0.35	151	1
HH-Income	3.99	0.17	149	1	1.41	0.51	151	1
HH-Edu	3.11	0.31	149	1	1.26	0.56	151	1
HH-Size	1	0.17	149	1	1.19	0.39	151	1
covid-Status	0.75	0.44	149	1	0.05	0.22	151	1
Model construction time: 0.01 seconds **Cross-validation Stratified Summary**								
Correctly-Classified-Instances: 299 (99.67%)					**Incorrectly-Classified-Instances**: 1 (0.33%)			
The kappa test	**Error Absolute Mean**	**Error² Root mean**	**Absolute & relative**	**Relative root**	**Total no. of**			

			error	**error^2**	**cases**
0.9933	0.0043	0.0603	0.87 %	12.0668 %	300

Detailed Precision by Class

	[Rate TP]	**[Rate FP]**	**[Accuracy]**	**[Recall]**	**[Measure F]**	**[MCC]**	**[Area ROC]**	**[Area PRC]**	**[set]**
	1.000	0.007	0.993	1.000	0.997	0.99	1.00	1.00	Good
	0.993	0.000	1.000	0.993	0.997	0.99	1.00	1.00	Poor
Mean Avg.	0.997	0.003	0.997	0.997	0.997	0.99	1.00	1.00	

[Confusion-Matrix]

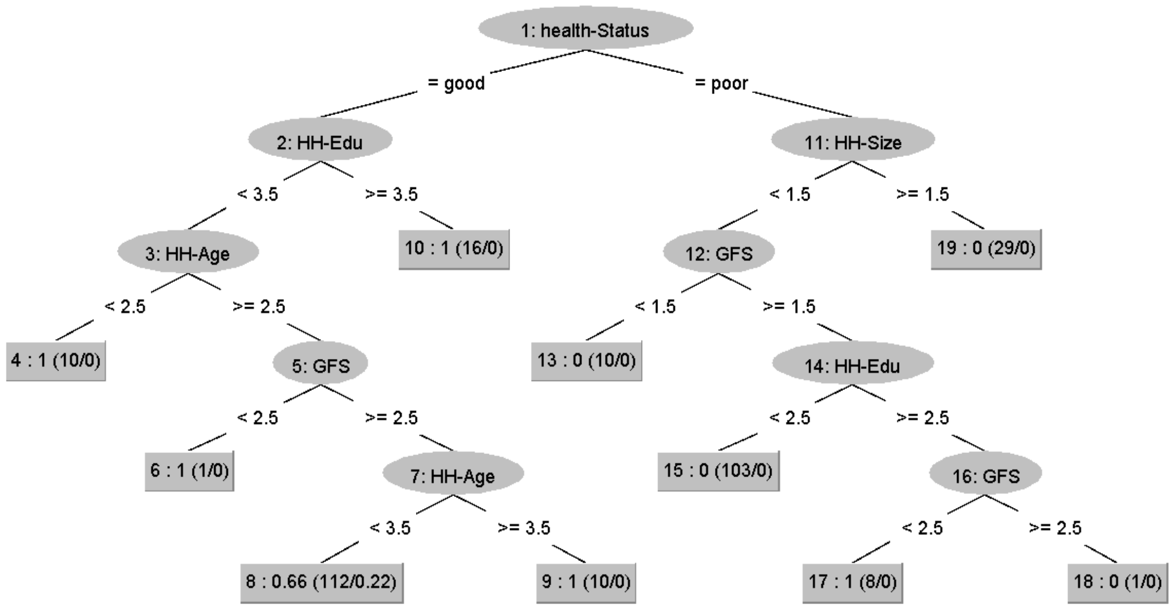
		Actual value is categorized as		
		a	**B**	
Predicted value	**a = good**	149	0	
	b = poor	1	150	

Decision-Tree

In both classification and regression tasks, a non-parametric supervised learning algorithm known as a decision tree is utilized.

It has a root node, internal nodes, leaf nodes, and branches in a hierarchical structure resembling a tree. Figure B depicts the decision tree weka visualization.

Figure (B): An Illustration of a Decision-Tree using Weka



The above figure indicates the health status in the form of a decision tree. The main attribute is "health status". If the health status is good, then the tree further analyzes head education. If head education is higher then the health status is good and the daily food intake score of families is high. tree analyzes that if the household size is lower than health status and GFS is good otherwise poor. Furthermore, the tree analyzes that if the household head status is poor indicates the size of the households is higher shows the lower education level and GFS.

Weka Classification for Logistic Regression Analysis

Family health status is largely determined by the family's financial circumstances. Table 4.4 displays the analysis of the logistic regression algorithm for good or poor households' health status and different characteristics that have a significant impact on households' access to healthy food to predict the HHS in the urban region of Southern Punjab Pakistan.

Table 4: Analysis of the Logistic Regression Algorithm

Scheme:	Weka. Classifiers. Functions. Logistic - R 100.0 -M -1 -num-decimal-places 4
Relation:	Health-Status **Instances:** 300
Attributes:	7 (GFS, HH-Age, HH-Income, HH-Edu, HH-Size, covid-Status & health-Status)
Test mode:	Cross-validation of ten times **Model for a classifier:** (full training set)
Logistic-Regression-with ridge- parameter-of -100.0	

Variable	Coefficients (Class-good)	Odds Ratios (Class-good)
GFS	0.6004	1.8229
HH-Age	-0.1334	0.8751
HH-Income	0.2594	1.2962
HH-Edu	0.2945	1.3425
HH-Size	-0.2982	0.7421
covid-Status	0.4503	1.5688
Intercept	-2.2949	

Model construction time: 0.02 seconds **Summary Cross-validation stratification**

Correctly-Classified-Instances: 298 (99.3%)	**Incorrectly-Classified-Instances:** 2 (0.667%)

kappa test	Error Absolute Mean	Error² Root mean	Absolute & relative error	Relative root error²	Total no. of cases
0.9867	0.2623	0.2706	52.4686 %	54.1137 %	300

Detailed-by-Precision-Class

	[Rate TP]	[Rate FP]	[Accuracy]	[Recall]	[Measure F]	[MCC]	[ROC Area]	[PRC Area]	[Class]
	1.000	0.013	0.987	1.000	0.993	0.987	1.000	1.000	Good
	0.987	0.000	1.000	0.987	0.993	0.987	1.000	1.000	Poor
Mean Avg.	0.993	0.007	0.993	0.993	0.993	0.987	1.000	1.000	

[Confusion-Matrix]

		Actual value is categorized as	
Predicted value		**A**	**B**
	a = good	149	0
	b = poor	2	149

The estimates of the logistic model algorithm's coefficient are shown in Table. In this model dependent variable are household health status and the independent variables are GFS, age of household head (HH), income of HH, education of HH, HH-size and covid-status. The following hypothesis has been used to estimate the model to verify that it is suitable and well-fit. The independent variables do not affect the dependent variable, which is the null hypothesis. Another possibility is that the dependent variable is significantly influenced by independent variables. Odd-ratio values of relative probabilities were computed by exponentiating the coefficients. The regression result demonstrates that the GFS has positive coefficients and that a log-odds increase of 0.6004 with one unit of the variable GFS would be expected to have a positive impact on HHS. The odds ratio (2.739) shows that with an increase in GFS, the probability of a good health score to poor would be expected to increase by a factor of 1.8229 i.e., 274 percent. The relative log odds of a good health score versus a poor health score decrease by 0.1334 with an increase of one unit in HH age. The odds ratio shows that with an increase in HH age, the probability of a good health score would be expected to increase by a factor of 0.8751 i.e. 13 percent. The regression result demonstrates that the HH's income has positive coefficients and that a 0.2945 increase in the log odds of having a positive impact on HHS would be expected for every unit of the variable HH's income. The odds ratio (1.2962) shows that with an increase in HH income, the probability of a good health score to poor would be expected to increase by a factor of 1.3425 i.e., 130 percent. One unit increase in HH education is associated with a 0.2945 increase in the relative log odds of good health score vs. poor score. The odds ratio shows that with an increase in HH education, the probability of preferring a good health score would be expected to increase by a factor of 1.3425 i.e. 29.5 percent.

With a one-unit increase in household size, the ratio of good health score to poor health score decreases by 0.2982. According to the proportion of the chances, an increase in HH size would be expected to increase the likelihood of a good health score by 0.7421 percent, or 30 percent. A 0.4503 increase in the relative log odds of a good health score versus a poor health score is associated with an increase of one unit in the covid-status of families. The odds ratio shows that with an increase in the covid-status of families, the probability of a good health score would be expected to increase by a factor of 1.5688 i.e., 45 percent.

We can see that the correctly classified instances that are out of 300 instances 298 instances are correctly classified. So, 99% of that is the accuracy of this particular classifier. Thus, mistakenly characterized is 2. The kappa test value is 0.99 is a very good value. Kappa compares the classified performance compared to zero or classifies that you have the baseline classifiers as good and poor. So, this 99% is better than the baseline classifier. The root mean square is 0.2706 and this is an absolute error. The values of the true positive rate, false positive rate precision, recall, and F-measure for the good class are 1.000, 0.013, 0.987, 1.000, and 0.993, respectively, according to the detailed accuracy. The detailed accuracy for poor class indicates that the value of Ture positive rate, false positive rate precision, recall, and F-measure are 0.987, 0.000, 1.000, 0.987, and 0.993 respectively. The weighted normal of each class demonstrates that the worth of Ture positive rate, misleading positive rate accuracy, review, and F-measure are 0.993, 0.007, and all 0.993 individually. The confusion matrix shows that the two values a and b a is good and b is poor. The target variable has two values positive shows good health status and negative shows poor health score. The a and b classified the predicted value as the actual values given in the row wise that is the target/outcome value and the column wise indicates the predicted value like classifier output. The model correctly identified 149 good class data points when the true positive (TP) value was set to 149. When the true negative (TN) equaled 149, the model correctly classified 149 data points belonging to the poor class. On the off chance that the model arranged 0 unfortunate class significant pieces of information as having a place with the great class, then the misleading positive (FP) is zero. false negative (FN) equal to 2, meaning the model inaccurately grouped 2 positive class data of interest as having a place with the unfortunate class. So, there are 149 classified as good which is actually good and that is correctly specified.

Conclusion & Policy Recommendations

This study provides a theoretical framework for identifying various socioeconomic factors affecting HHS and GFS by using a logistic regression algorithm for data analysis in Southern Punjab. Samples are taken from 300 metropolitan families in districts Lodhran and Vehari in southern Punjab, Pakistan. This study's main findings indicate that HHS in metropolitan areas is affected by socioeconomic factors like GFS, household head age, income, education level, size, and covid-19 status. This study concluded that when the attribute of GFS higher HHS more as compared to those household heads with a lower level of GFS. This study concluded that when the education level of household heads is higher HHS of households is more as compared to those household heads with a lower level of education. According to our research, the attribute of household head income has positive coefficients, indicating that household heads with higher incomes are more likely to have a positive HHS than those with lower incomes. This study also concluded that the attributes of age and family size have negative coefficients, which indicates in older age HHS is poor. because as they get older, they have fewer options for food than they did when they were younger. Consequently, adolescents' daily food choices are more varied than those of adults or seniors. In contrast to families with larger family sizes and more family heads, families with smaller family sizes have higher HHS. Additionally, the study concluded that the covid-19 status attribute has positive coefficients, indicating that households with higher vaccination rates have higher HHS than households with lower vaccination rates. The study argues some important policy implications that have a significant effect on metropolitan HHS. Pakistan is a developing nation that has questions about HHS. Good HHS has a strong impact on the human body and maintains a reasonable standard for food security. Consuming food that only contains nutrients from a single food item is not beneficial to the human body. Following are strategy suggestions to control populace development, raise the proficiency rate, increment the wellbeing spending plan, control debasement in general wellbeing projects, regionalize medical care benefits, and advance wellbeing training in Pakistan's medical services area in that way we will build the great HHS.

References

Akbar, A., Darma, R., Fahmid, I. M., & Irawan, A. (2023). Determinants of Household Food Security during the COVID-19 Pandemic in Indonesia. *Sustainability*, *15*(5), 4131.

Amin, M. D., Badruddoza, S., & McCluskey, J. J. (2021). Predicting access to healthful food retailers with machine learning. Food Policy, 99, 101985.

Baitharu, T. R., & Pani, S. K. (2016). Analysis of data mining techniques for healthcare decision support system using liver disorder dataset. *Procedia Computer Science*, *85*, 862-870.

Birk, N., Matsuzaki, M., Fung, T. T., Li, Y., Batis, C., Stampfer, M. J., ... & Lake, E. (2021). Exploration of machine learning and statistical techniques in development of a low-cost screening method featuring the global diet quality score for detecting prediabetes in rural India. *The Journal of nutrition*, *151*(Supplement_2), 110S-118S.

Dong, J., Li, X., Fan, R., & Yang, J. (2023). Health-related quality of life among ethnic minority residents in remote Western China: a cross-sectional study. *BMC Public Health*, *23*(1), 1-11.

FAO. (2018). *Dietary assessment: a resource guide to method selection and application in low resource settings*. Food & Agriculture Organization. https://www.fao.org/3/i9940en/I9940EN.pdf

FAO. (2022). *The State of Food Security and Nutrition in the World 2022: Repurposing food and agricultural policies to make healthy diets more affordable*. FAO, IFAD, UNICEF, WFP and WHO.

Farooq, M., & Usman, M. (2023). Pakistan needs an equitable investment in the health system and collaborative efforts. *The Lancet Global Health*, *11*(2), e177-e178.

Gao, V., Long, M. T., Singh, S. R., Kim, Y., Zhang, X., Rogers, G., ... & Ma, J. (2023). Healthy diet is associated with lower risk of hepatic fibrosis. *The Journal of Nutrition*.

Hafeez, A., Dangel, W. J., Ostroff, S. M., Kiani, A. G., Glenn, S. D., Abbas, J., ... & Mokdad, A. H. (2023). The state of health in Pakistan and its provinces and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet Global Health*, *11*(2), e229-e243.

Iqbal, Z., Imran, M., Rahman, G., Miandad, M., Shahid, M., & Murtaza, B. (2023). Spatial distribution, health risk assessment, and public perception of groundwater in Bahawalnagar, Punjab, Pakistan: a multivariate analysis. *Environmental Geochemistry and Health*, *45*(2), 381-391.

Karakan, T., Gundogdu, A., Alagözlü, H., Ekmen, N., Ozgul, S., Tunali, V., ... & Nalbantoglu, O. U. (2022). Artificial intelligence-based personalized diet: A pilot clinical study for irritable bowel syndrome. *Gut Microbes*, *14*(1), 2138672.

Kavitha, C., Mani, V., Srividhya, S. R., Khalaf, O. I., & Tavera Romero, C. A. (2022). Early-stage Alzheimer's disease prediction using machine learning models. *Frontiers in public health*, *10*, 240.

Linardon, J., Messer, M., Helms, E. R., McLean, C., Incerti, L., & Fuller-Tyszkiewicz, M. (2020). +-Interactions between different eating patterns on recurrent binge-eating behavior: A machine learning approach. *International Journal of Eating Disorders*, *53*(4), 533-540.

Menichetti, G., Ravandi, B., Mozaffarian, D., & Barabási, A. L. (2021). Machine learning prediction of food processing. *MedRxiv*, 2021-05.

Rastogi, R., & Bansal, M. (2023). Diabetes prediction model using data mining techniques. *Measurement: Sensors*, *25*, 100605.

Shaikh, B. T., & Ali, N. (2023). Universal health coverage in Pakistan: is the health system geared up to take on the challenge?. *Globalization and Health*, *19*(1), 1-6.

Shiao, S. P. K., Grayson, J., Lie, A., & Yu, C. H. (2018). Personalized nutrition—genes, diet, and related interactive parameters as predictors of cancer in multiethnic colorectal cancer families. *Nutrients*, *10*(6), 795.

Suiçmez, Ç., Yılmaz, C., Kahraman, H. T., Cengiz, E., & Suiçmez, A. (2023). Prediction of Hepatitis C Disease with Different Machine Learning and Data Mining Technique. In *Smart Applications with Advanced Machine Learning and Human-Centred Problem Design* (pp. 375-398). Cham: Springer International Publishing.

Thamrin, S. A., Arsyad, D. S., Kuswanto, H., Lawi, A., & Nasir, S. (2021). Predicting obesity in adults using machine Learning techniques: An analysis of Indonesian Basic Health Research 2018. *Frontiers in nutrition*, *8*, 669155.

Westerveld, J. J., van den Homberg, M. J., Nobre, G. G., van den Berg, D. L., Teklesadik, A. D., & Stuit, S. M. (2021). Forecasting transitions in the state of food security with machine learning using transferable features. *Science of the Total Environment*, *786*, 147366.

WHO. (2021). *Country cooperation strategy for WHO and Pakistan: 2020-2025 / World Health Organization. Regional Office for the Eastern Mediterranean.* https://applications.emro.who.int/docs/9789290226635-eng.pdf

WHO. (2022). *World health statistics 2022: Monitoring health for the SDGs sustainable development goals*. World Health Organization.