International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 6, Nov.-Dec., 2024, pp:409-417 ISSN: 3048-6874
www.ijhssm.org

Decoding the Obesity Puzzle in Peru, Colombia, and Mexico

Rimjhim Singh
University of Warwick

Date of Submission: 07-11-2024 Date of Acceptance: 18-11-2024

Abstract

This study explores the multifaceted factors contributing to obesity in Peru, Colombia, and Mexico, addressing the global epidemic of overweight and obesity. Utilizing data analysis and statistical tools, including regression, correlation, and classifiers, the research investigates potential connections between obesity and eating habits, alcohol consumption, and smoking. Drawing on previous studies for context, the analysis involves creating new variables to facilitate focused inquiries. The study's objective is to provide insights into the relationships between lifestyle, dietary choices, and obesity, offering a comprehensive understanding of this complex health concern. While the findings support associations between certain variables and obesity, caution is emphasized regarding causation, underlining the need for additional research.

I. Introduction

This study aims to explain the factors responsible for the obesity of individuals in Peru, Colombia, and Mexico. Obesity is a global problem now. According to the World Health Organization (WHO), overweight and obesity are characterized by the abnormal or excessive accumulation of body fat, posing health risks. A Body Mass Index (BMI) exceeding 25 is classified as overweight, while a BMI exceeding 30 is categorized as obese. The problem has escalated to an epidemic level, with over 4 million individuals succumbing to the consequences of excess weight in 2017, as reported by the global burden of disease. The rates of overweight and obesity persist in their upward trajectory among both adults and children. Between 1975 and 2016, the prevalence of overweight or obese children and adolescents aged 5–19 years surged more than fourfold, escalating from 4% to 18% on a global scale.

WHO claims that the prevalence of global obesity has surged significantly, nearly tripling since 1975. In 2016, the number of overweight adults, aged 18 and older, exceeded 1.9 billion, with over 650 million classified as obese. In that year, 39% of adults aged 18 and over were identified as overweight, and 13% were categorized as obese. It's noteworthy that a substantial portion of the world's population resides in nations where the impact of overweight and obesity on mortality surpasses that of underweight conditions. In 2020, an alarming 39 million children under the age of 5 were either overweight or obese. Additionally, in 2016, the global count of overweight or obese children and adolescents aged 5-19 surpassed 340 million.

This study delves into answering if there is a relationship between obesity and factors like genetics, eating habits, alcohol consumption or smoking, etc. To explain a health problem like obesity many statistical tools like regression, correlations, and classifiers are used in this study. It is found that relationships between lifestyle, eating habits, and obesity significantly exist.

Background

Banterle and Cavaliere (2008) explore the socio-economic factors influencing obesity in Italy by surveying a consumer sample of 999 individuals. The analysis employs a binary logit model with body mass index (BMI) as the dependent variable, categorized as either seriously overweight and obese (coded as 1) or normal weight (coded as 0). The findings reveal that the likelihood of being seriously overweight and obese rises with age, particularly among individuals aged 65 and above. Gender also plays a significant role, with men exhibiting a higher likelihood of being seriously overweight and obese compared to women.

Smith and Smith (2016) claim that the global rise in overweight and obesity rates can be attributed to an intricate interplay of genetic, environmental, and social factors within a technologically advancing world. The widespread accessibility of affordable, high-calorie foods further exacerbates this phenomenon.

Dutta et. al (2013) aim to evaluate the prevalence of overweight and obesity, as well as their associated factors, among urban adults within Bengal. They found sex differences were significant for obesity ($p < 0.01$) and combined overweight-obesity ($p < 0.01$). Multinomial logistic regression revealed that age and monthly income significantly influenced overweight ($p < 0.05$). Factors such as sex, age, monthly income, marital status, education, and alcohol intake had significant effects on obesity ($p < 0.05$).

Combined overweight-obesity (BMI ≥ 23.00 kg/m²) was significantly influenced by sex, age, monthly income, marital status, and education ($p < 0.05$).

Bollapragada et al. (2017) provide an extensive review on obesity, covering its development, epidemiology, influencing factors, health hazards, management strategies, and potential natural treatments. The authors highlight obesity as a pervasive global health concern affecting diverse populations and emphasize the significance of body mass index (BMI) in its assessment. The escalating prevalence of obesity worldwide is acknowledged, necessitating comprehensive approaches for understanding and addressing this complex issue.

The review explores various factors contributing to obesity, ranging from age, gender, and environmental conditions to psychological influences, genetic predisposition, and lifestyle choices. The intricate interplay of these factors underscores the multifaceted nature of obesity as a disease with diverse determinants. Health hazards associated with obesity are meticulously outlined, encompassing conditions such as hypertension, diabetes mellitus, dyslipidemia, cardiac alterations, respiratory disorders, increased cancer risks, and neurological issues. This comprehensive overview underscores the broad impact of obesity on various physiological systems.

Dixon and Eager (2014), question the commonly held belief that the obesity epidemic and its associated chronic diseases are solely a result of modern lifestyles. It criticizes the term "lifestyle" for being too narrow in its scope, neglecting broader social, economic, and environmental factors, and inadvertently placing blame on individuals. The author advocates for a more comprehensive perspective on lifestyle, considering distal, medial, and proximal determinants, and underscores the importance of analyzing causality across all these levels. The term "anthropogens" is introduced in the article to describe man-made environments, by-products, or lifestyles that could be harmful to human health.

The article systematically reviews anthropogens, concentrating on inducers with a metaflammatory association and providing evidence of their link to various chronic diseases. Obesity is metaphorically portrayed as a "canary in a mineshaft," signaling underlying problems in the broader environment. The implication is that efforts to manage population obesity should prioritize interventions at upstream levels for more effective chronic disease management.

Data

The study utilizes data contributed by Fabio Mendoza Palechor and Alexis de la Hoz Manotas, Universidad de la Costa, CUC, Colombia, https://archive.ics.uci.edu/dataset/544/estimation+of+obesity+levels+based+on+eating+habits+and+physical+condition. The data is about obesity levels among individuals in Mexico, Peru, and Colombia with participants aged between 14 and 61, exhibiting diverse eating habits and physical conditions. Data collection was facilitated through a web platform survey wherein anonymous users responded to each query. The attributes pertaining to eating habits include: Regular consumption of high-calorie food (FAVC), Frequency of vegetable consumption (FCVC), Number of main meals (NCP), Consumption of food between meals (CAEC), Daily water intake (CH20), and Consumption of alcohol (CALC). Attributes linked to physical condition encompass: Monitoring calorie consumption (SCC), Frequency of physical activity (FAF), Time spent using technology devices (TUE), Transportation methods used (MTRANS). Additional variables obtained include: Gender, Age, Height, and Weight. The class variable 'NObeyesdad' encompasses values such as Insufficient Weight, Normal Weight, Overweight Level I, Overweight Level II, Obesity Type I, Obesity Type II, and Obesity Type III. These classifications are determined using Equation (1) and reference information from the World Health Organization (WHO) and Mexican Normativity.

Data Processing

In the data processing stage, two crucial variables were created to facilitate a more focused analysis. Firstly, the class variable "Overweight&Obesity" was formulated, taking on the value of 0 if the participant's body weight status fell under 'Normal_Weight' or 'Insufficient_Weight,' and 1 if categorized as Overweight Level I, Overweight Level II, Obesity Type I, Obesity Type II, or Obesity Type III. This classification aimed to streamline the analysis by grouping participants based on their weight status.

Additionally, a binary variable "Smoking" was introduced, with a value of 0 assigned to participants who reported "no" to smoking and 1 for those who indicated "yes." This binary categorization allowed for a straightforward examination of the association between smoking behavior and other variables of interest.

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 6, Nov.-Dec., 2024, pp:409-417 ISSN: 3048-6874
www.ijhssm.org

Overweight&Obesity = 0 if NObeyesdad = 'Normal_Weight' or 'Insufficient_Weight

= 1 if NObeyesdad = Overweight Level I, Overweight Level II, Obesity Type I,

Obesity Type II, and Obesity Type III

Smoking = 0 if Smoke = "no"

= 1 if Smoke = "yes"

The entire data analysis process was seamlessly executed using the Python language. Python's widespread adoption in data analysis owes itself to its user-friendly syntax, extensive libraries such as NumPy and Pandas, and the strong support of a thriving community. Its cross-platform compatibility encourages collaborative efforts and ensures adaptability to diverse workflows. Python's scalability is particularly advantageous, accommodating projects of various sizes, and its open-source nature fosters continuous innovation within the data analysis realm.

The versatility of Python extends beyond data analysis to encompass other domains, including web development and machine learning. Notably, Python's robust support for big data processing and its provision of interactive development environments, exemplified by tools like Jupyter Notebooks, contribute significantly to its appeal among data professionals. The choice of Python for the data analysis phase underscores its efficacy as a comprehensive and versatile tool in unraveling complex datasets and deriving meaningful insights.

Research questions

1. Is there a relationship between a person smoking and their weight gain?

This research question seeks to explore the potential connection between smoking behavior and changes in an individual's weight. The investigation involves examining whether there is a discernible pattern indicating that individuals who smoke tend to experience weight gain. The study aims to use statistical analyses, such as regression models or correlation coefficients, to quantify and understand the nature and strength of this relationship.

2. Is there a relationship between a person's drinking and their weight gain?

This research question focuses on unraveling the link between alcohol consumption and variations in body weight. The study aims to investigate whether individuals who consume alcohol exhibit a notable trend of weight gain compared to non-drinkers. Through rigorous data analysis, including regression models and correlation assessments, the research intends to quantify the strength and significance of this relationship.

3. Does weight gain have some correlation with the eating habits and lifestyle of an individual?

This broad research question seeks to understand the multifaceted relationship between weight gain and various aspects of an individual's lifestyle and dietary choices. It involves exploring correlations between weight gain and factors such as eating habits, frequency of meals, consumption of high-calorie foods, and engagement in physical activity. By delving into specific lifestyle choices, the study aims to provide a comprehensive understanding of how different aspects of one's daily routine contribute to changes in body weight.

Is there a relationship between smoking and weight gain?

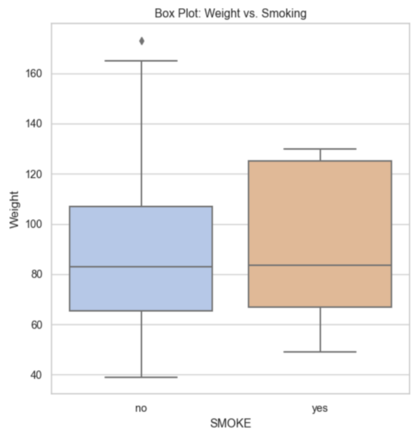


Figure 1

The above graphs show that marginally the median level of weight for smokers is greater than for non-smokers.

Next, a regression analysis is done for the same. This is a dummy variable regression with "weight" as the dependent variable and "smoking" as the independent variable. With a p-value of 0.237, the model is not significant even at a 10 % level of significance.

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 6, Nov.-Dec., 2024, pp:409-417 ISSN: 3048-6874
www.ijhssm.org

Table 1

OLS Regression Results

Dep. Variable:	Weight	R-squared:	0.001
Model:	OLS	Adj. R-squared:	0.000
Method:	Least Squares	F-statistic:	1.399
Date:	Thu, 14 Dec 2023	Prob (F-statistic):	0.237
Time:	11:34:47	Log-Likelihood:	-9887.5
No. Observations:	2111	AIC:	1.978e+04
Df Residuals:	2109	BIC:	1.979e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	86.4877	0.576	150.145	0.000	85.358	87.617
smoking	4.7191	3.990	1.183	0.237	-3.105	12.544

Omnibus:	132.201	Durbin-Watson:	0.291
Prob(Omnibus):	0.000	Jarque-Bera (JB):	65.918
Skew:	0.254	Prob(JB):	4.85e-15
Kurtosis:	2.299	Cond. No.	7.00

Is there a relationship between drinking and weight gain?

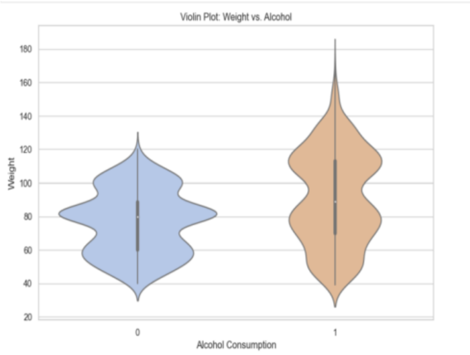


Figure 2

Graphical inspection reveals that the median weight of people consuming alcohol is larger than people not consuming alcohol. Now a dummy regression is run with weight as the dependant variable and smoking as the independent variable.

Table 2

OLS Regression Results

Dep. Variable:	Weight	R-squared:	0.062
Model:	OLS	Adj. R-squared:	0.062
Method:	Least Squares	F-statistic:	140.4
Date:	Thu, 14 Dec 2023	Prob (F-statistic):	2.11e-31
Time:	11:34:52	Log-Likelihood:	-9820.2
No. Observations:	2111	AIC:	1.964e+04
Df Residuals:	2109	BIC:	1.966e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	76.6573	1.003	76.391	0.000	74.689	78.625
alcohol	14.2389	1.202	11.849	0.000	11.882	16.596

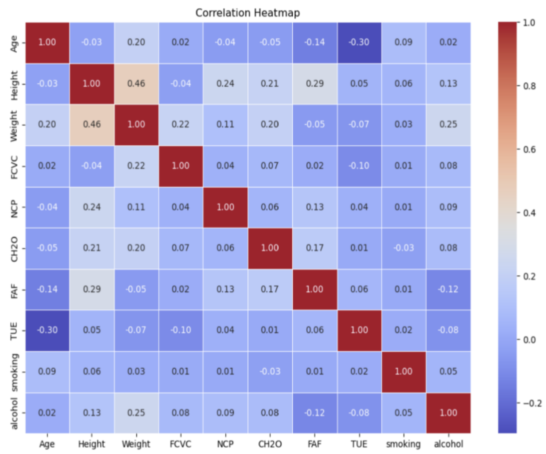
Omnibus:	126.512	Durbin-Watson:	0.361
Prob(Omnibus):	0.000	Jarque-Bera (JB):	48.093
Skew:	0.064	Prob(JB):	3.60e-11
Kurtosis:	2.272	Cond. No.	3.40

The coefficient of determination, is 0.062, indicating that the model explains about 6.2% of the variance in weight. The estimated 'Weight' when the person does not drink is 76.6573 Kg. The estimated change in weight for a one-unit change in 'alcohol' (from not drinking to drinking) is 14.2389 kg. The p-value for 'alcohol' (2.11e-31), is very low (P>|t| < 0.05), suggesting that the effect of drinking on weight is statistically significant. The positive coefficient for 'alcohol' indicates that, on average, individuals who drink have a higher weight compared to those who do not drink, according to this model.

Is there a relationship between obesity and the eating and lifestyle of an individual? Are lifestyle and eating habits a good predictor of obesity?

The correlation heatmap in the figure shows that there is positive correlation between weight of an individual and their age, height, FCVV, number of main meals, Consumption of water, smoking and alcohol consumption.

Figure 3



KNN Classifier

The k-Nearest Neighbors (k-NN) classifier stands out as a versatile and user-friendly supervised machine learning algorithm applicable to both classification and regression tasks. Its fundamental principle revolves around the concept of proximity, assuming that akin data points are situated in close spatial proximity within the feature space. In the realm of classification, the term "neighbors" denotes data points that share proximity in the feature space. When presented with a new data point, the algorithm determines its class by referencing the class labels of its nearest neighbors. The "k" in k-NN designates the number of nearest neighbors taken into account during classification. The class assigned to the new data point is established through a majority or weighted voting mechanism among its k-nearest neighbors. This involves calculating the distance between the new data point and all existing data points in the training set, typically utilizing metrics like Euclidean distance. The

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 6, Nov.-Dec., 2024, pp:409-417 ISSN: 3048-6874
www.ijhssm.org

k-nearest neighbors with the smallest distances are then identified, and for classification, the majority class among these neighbors becomes the predicted class for the new data point. The choice of k significantly impacts the model's performance, with smaller values yielding more adaptable models and larger values producing smoother decision boundaries. Despite its simplicity and effectiveness in capturing intricate decision boundaries, k-NN does grapple with challenges such as sensitivity to outliers and computational demands, particularly for expansive datasets. Researchers commonly opt for k-NN due to its straightforward implementation and applicability across a spectrum of datasets.

Next a KNN Classifier is built to classify the Overweight&Obesity variable. The model achieved an accuracy of approximately 94.33%, reflecting the overall correctness of its predictions on the test set. Precision, which measures the accuracy of positive predictions, is high for both classes. Class 0 has a precision of 0.95, and class 1 has a precision of 0.94

Table 3

Accuracy: 0.9432624113475178
Classification Report:

	precision	recall	f1-score	support
0	0.95	0.84	0.89	118
1	0.94	0.98	0.96	305
accuracy			0.94	423
macro avg	0.95	0.91	0.93	423
weighted avg	0.94	0.94	0.94	423

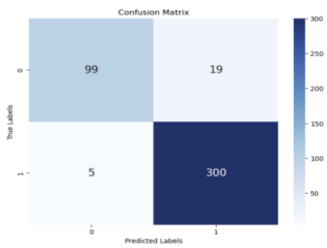


Figure 4

. The model demonstrates good recall, capturing a high percentage of actual positive instances. Class 0 has a recall of 0.84, and class 1 has a recall of 0.98. The F1-Score, which combines precision and recall, is also noteworthy. Class 0 has an F1-score of 0.89, and class 1 has an F1-score of 0.96. The number of instances for each class in the test set is provided. There are 118 instances of class 0 and 305 instances of class 1. Hence, the k-NN classifier demonstrates robust performance in classifying 'Overweight&Obesity,' achieving high accuracy and effectively identifying positive instances.

Figure 5

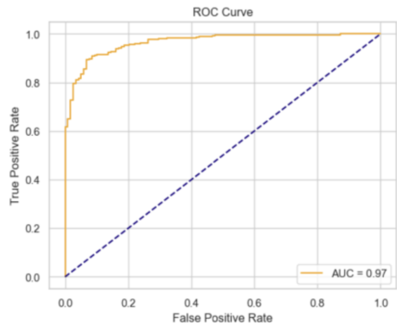


Figure 6

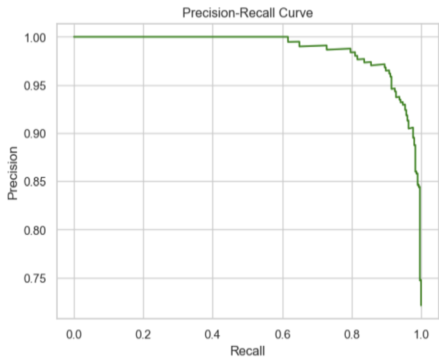


Table 4

Accuracy: 0.9031
Precision: 0.9459
Recall (Sensitivity): 0.9180
F1-Score: 0.9318
True Positive Rate (Sensitivity, Recall): 0.9180
False Positive Rate: 0.1356
True Negative Rate (Specificity): 0.8644
False Negative Rate: 0.0820
Positive Predictive Value (PPV): 0.9459
Negative Predictive Value (NPV): 0.8031

Within the realm of k-Nearest Neighbors (k-NN) classification, diverse performance metrics serve as evaluative tools for gauging the efficacy of the model. These metrics encompass the True Positive Rate (Sensitivity or Recall), False Positive Rate, True Negative Rate (Specificity), False Negative Rate, Positive Predictive Value (Precision), and Negative Predictive Value.

1. True Positive Rate (Sensitivity or Recall): This metric epitomizes the fraction of actual positive instances accurately discerned by the model. Its computation involves dividing the number of true positives by the sum of true positives and false negatives.

2. False Positive Rate: The False Positive Rate characterizes the proportion of actual negative instances erroneously classified as positive. This rate

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 6, Nov.-Dec., 2024, pp:409-417 ISSN: 3048-6874
www.ijhssm.org

is determined by dividing the number of false positives by the sum of false positives and true negatives.

3. True Negative Rate (Specificity): This metric mirrors the fraction of actual negative instances correctly identified by the model. It is derived by dividing the number of true negatives by the sum of true negatives and false positives.

4. False Negative Rate: The False Negative Rate quantifies the proportion of actual positive instances inaccurately categorized as negative. It involves dividing the number of false negatives by the sum of false negatives and true positives.

5. Positive Predictive Value (Precision): Precision gauges the accuracy of positive predictions made by the model. The computation entails dividing the number of true positives by the sum of true positives and false positives.

6. Negative Predictive Value: Negative Predictive Value reflects the accuracy of negative predictions made by the model. It is computed by dividing the number of true negatives by the sum of true negatives and false negatives.

These metrics collectively offer valuable insights into the k-NN classifier's performance by revealing its capacity to accurately identify positive and negative instances and shedding light on potential misclassifications.

True Positive Rate: The model correctly predicts 91.80% of actual positives.

False Positive Rate: The prediction of overweight and obesity is inaccurate in about 13.56% of instances among individuals who are non-overweight and non-obese.

True Negative Rate : Accurate identification of non-overweight and non-obese individuals occurs approximately 86.44% of the time.

False Negative Rate: Identification of overweight and obesity falls short in 8.20% of actual cases.

Positive Predictive Value: Among instances predicted as positive (overweight and obesity), 94.59% are true positives.

Negative Predictive Value: Among instances predicted as negative (non-overweight and non-obese), 80.31% are true negatives.

Naïve Bayes Classifier

Next, a Naïve Bayes classifier is run.

The Naïve Bayes classifier, a widely used probabilistic machine learning algorithm, is built on Bayes' theorem and relies on the assumption of feature independence. Despite its apparent simplification, this algorithm has demonstrated significant effectiveness across diverse classification tasks, notably in applications such as natural language processing and spam filtering. Its core principle involves assessing the probability of a data point belonging to a specific class by considering the conditional probabilities of its features. The "naïve" designation comes from assuming feature independence, streamlining the computation of joint probabilities. Notably, Naïve Bayes excels in handling high-dimensional datasets and exhibits computational efficiency, making it well-suited for real-world scenarios with substantial data volumes. Its applicability extends to both binary and multiclass classification challenges. However, its strict independence assumption may limit performance when features display strong correlations. Despite this constraint, Naïve Bayes retains popularity due to its simplicity, efficiency, and competitive performance in practical situations.

Table 5

Accuracy: 0.9030732860520094
Classification Report:

	precision	recall	f1-score	support
0	0.80	0.86	0.83	118
1	0.95	0.92	0.93	305
accuracy			0.90	423
macro avg	0.87	0.89	0.88	423
weighted avg	0.91	0.90	0.90	423

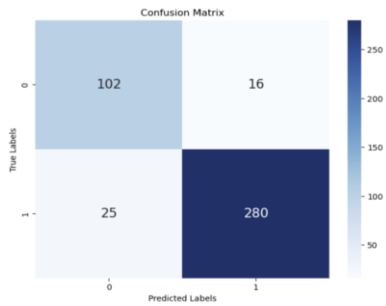
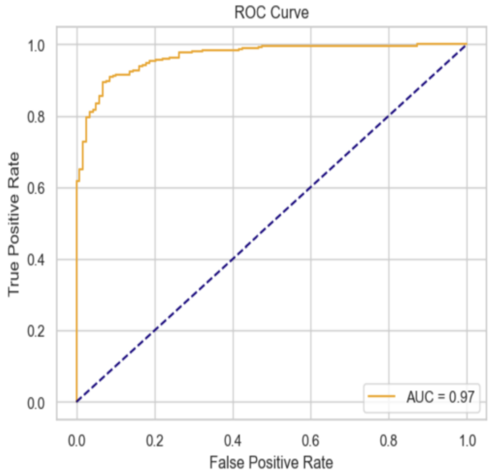
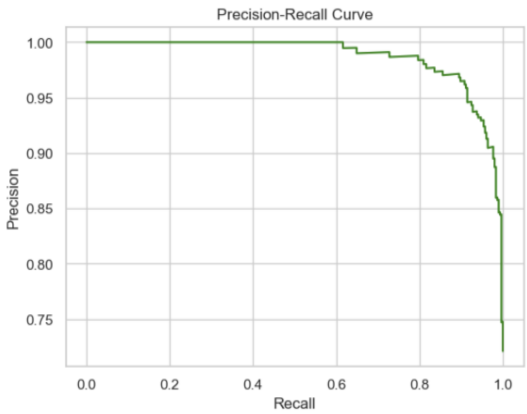


Figure 7

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 6, Nov.-Dec., 2024, pp:409-417 ISSN: 3048-6874
www.ijhssm.org

Figures 8 and 9






```
Accuracy: 0.9031
Precision: 0.9459
Recall (Sensitivity): 0.9180
F1-Score: 0.9318
True Positive Rate (Sensitivity, Recall): 0.9180
False Positive Rate: 0.1356
True Negative Rate (Specificity): 0.8644
False Negative Rate: 0.0820
```

Table 6

The model achieved an accuracy of approximately 90.31%. This indicates that about 90.31% of the predictions on the test set were correct. For class 0 (non-overweight), precision is 0.80. This means that among the instances predicted as non-overweight, 80% were correct. For class 1 (overweight and obese), precision is 0.95. This indicates that among the instances predicted as overweight, 95% were correct. For class 0, recall is 0.86. This means that the model captured 86% of the actual non-overweight instances. For class 1, recall is 0.92. This indicates that the model captured 92% of the actual overweight instances. There are 118 instances of class 0 and 305 instances of class 1.

The true positive rate, reflecting the model's capability to correctly identify positive instances, is at 91.80%. Meanwhile, the false positive rate represents instances where non-overweight and non-obese cases are incorrectly classified as positive, and it stands at 13.56%. The true negative rate, which gauges the accurate identification of non-overweight and non-obese instances, is at 86.44%. The false negative rate, indicating instances where actual positive cases are incorrectly classified as negative, is 8.20%.

Positive Predictive Value (PPV), representing the likelihood of true positives among positive predictions, is at 94.59%. Negative Predictive Value (NPV), signifying the likelihood of true negatives among negative predictions, is at 80.31%. These metrics collectively provide insights into the performance and effectiveness of the prediction system.

While achieving high accuracy with both the models suggests that the lifestyle and eating habits contribute to successful predictions of obesity. There is enough evidence to claim that there is a relationship between the selected variables and the classified variable. However, correlation does not imply causation. The models have a successful predictive model.

II. Concluding Remarks:

This study aimed to investigate the factors contributing to obesity among individuals in Peru, Colombia, and Mexico. The global prevalence of overweight and obesity has reached alarming levels, posing significant health risks. Utilizing data from these three countries, the research explored potential relationships between obesity and various factors, including genetics, eating habits, alcohol consumption, and smoking.

The analysis employed statistical tools such as regression, correlation, and classifiers to uncover patterns and associations within the data. Previous studies were referenced to provide context, with Banterle and Cavaliere (2008) highlighting socio-economic factors in Italy, Smith and Smith (2016) emphasizing the complex interplay of genetic, environmental, and social factors, and Dutta et. al (2013) revealing significant associations in Bengal.

The data processing stage involved the creation of new variables, such as 'Overweight&Obesity' and 'Smoking,' facilitating a more focused analysis.

Research questions were formulated to investigate relationships between smoking, drinking, lifestyle, and eating habits with weight gain and obesity.

Graphical representations and regression analyses were employed to explore these relationships. For instance, the study found a marginal difference in weight between smokers and non-smokers, with regression analysis confirming non-significance. Conversely, alcohol consumption exhibited a significant relationship with weight gain, indicating that individuals who consume alcohol tend to have higher average weights.

Furthermore, a correlation heatmap revealed positive correlations between weight and various factors such as age, height, eating habits, and lifestyle choices. A KNN Classifier demonstrated robust performance in predicting 'Overweight&Obesity,' achieving high accuracy and recall rates.

The Naïve Bayes classifier, while slightly less accurate than the KNN model, still provided meaningful insights into the relationships between the selected variables and obesity.

Hence, the study provides evidence supporting the existence of relationships between lifestyle, eating habits, and obesity. However, it's crucial to acknowledge that correlation does not imply causation, and additional research is needed to establish causal links. The predictive models developed in this study contribute to understanding and addressing the complex issue of obesity, emphasizing the importance of lifestyle choices and eating habits in determining weight outcomes.

Limitations

The study, relying on data from the Universidad de la Costa, CUC, Colombia, has limitations that impact the generalizability of findings. The dataset's specificity may not extend to diverse demographic groups not covered, affecting broader applicability. Data collection via self-reporting introduces potential biases, as individuals may provide inaccurate or socially desirable responses, impacting data reliability.

The cross-sectional nature of the dataset captures a snapshot, limiting the establishment of causal relationships due to the absence of temporal dynamics in lifestyle and health factors. While identifying correlations between variables and weight-related outcomes, the study does not establish causation, emphasizing the need for caution in interpretation. The focus on specific factors without accounting for potential confounders like genetic predispositions, mental health, and socioeconomic status adds complexity.

Geographic limitations include the study's focus on Peru, Colombia, and Mexico, potentially limiting global representativeness. Regional variations in cultural practices, socioeconomic conditions, and healthcare infrastructure may influence outcomes.

The age range of participants (14-61) restricts extrapolation to other age groups, such as children or the elderly, where obesity determinants vary. Lack of longitudinal analysis hinders understanding of how lifestyle changes relate to weight outcomes over time. Despite high predictive model accuracy, caution is required in interpreting results due to the multifaceted nature of obesity. Acknowledging these limitations underscores the need for further research to enhance understanding and address constraints.

Acknowledgment

I would like to express my sincere gratitude to all those who contributed to the completion of this research study on obesity in Peru, Colombia, and Mexico. Undertaking such a multifaceted investigation would not have been possible without the invaluable support and collaboration of various individuals and institutions. I extend my heartfelt appreciation to Fabio Mendoza Palechor and Alexis de la Hoz Manotas from the Universidad de la Costa, CUC, Colombia, for generously providing the dataset essential to this study. Their commitment to advancing research in the field of health and wellness has greatly enriched the quality of this work.

Further Research

Longitudinal studies to track individuals over time can provide insights into the long-term impact of lifestyle and dietary choices on obesity could be conducted. This would allow researchers to establish causal relationships. There could be further exploration into the genetic factors influencing obesity could deepen the understanding of individual susceptibility to weight-related issues. Moreover, Investigating the relationship between mental health factors, such as stress or depression, and obesity could contribute to a more holistic understanding of the issue.

Applications

The study holds broad implications, impacting public health policies, educational initiatives, and healthcare interventions. Policymakers can employ the insights to develop precise obesity interventions, prioritizing lifestyle modifications. Educational campaigns have the potential to enhance awareness regarding the influence of factors such as smoking and alcohol on obesity. Healthcare providers can customize interventions based on identified connections with lifestyle choices. These findings bear relevance to global health initiatives, fostering collaborative efforts to address the intricate challenge of obesity.

References

[1]. Bollapragada, M. K., Shantaram, M. A. N. J. U. L. A., & Kumar, S. U. N. I. L. (2017). Obesity: Development, epidemiology, factors affecting, quantity, health hazards, management and natural treatment-a review. Int J Pharm Pharm Sci, 9(2), 12-26.
[2]. Cavaliere, A., & Banterle, A. (2008). Economic factors affecting obesity: an application in Italy (No. 725-2016-49677).
[3]. Egger, G., & Dixon, J. (2014). Beyond obesity and lifestyle: a review of 21st century chronic disease determinants. BioMed research international, 2014.
[4]. https://www.who.int/health-topics/obesity#tab=tab_1
[5]. Palechor, F. M., & de la Hoz Manotas, A. (2019). Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from Colombia, Peru and Mexico. Data in brief, 25, 104344.
[6]. Sen, J., Mondal, N., & Dutta, S. (2013). Factors affecting overweight and obesity among urban adults: a cross-sectional study. Epidemiology, Biostatistics, and Public Health, 10(1).
[7]. Smith, K. B., & Smith, M. S. (2016). Obesity statistics. Primary care: clinics in office practice, 43(1), 121-135.
[8]. UCI Machine Learning Repository. (n.d.). Estimation of obesity levels based on eating habits and physical condition [Data set]. UCI Machine Learning Repository. https://archive.ics.uci.edu/dataset/544/estimation+of+obesity+levels+based+on+eating+habits+and+physical+condition