Fuzzy Systems and Soft Computing
ISSN : 1819-4362

AN EARLY WARNING SYSTEM FOR ADOLESCENT GIRLS TO IDENTIFY CHRONIC DISEASES USING MACHINE LEARNING CLUSTERING ALGORITHM

Mrs.R.Arulmathi Assistant Professor, Department of Computer Applications, Women's Christian College, Chennai
Dr.R.Lakshmidevi Assistant Professor, Department of Computer Applications, Women's Christian College, Chennai
Ms.D.Sylvia Mary Head of the Department, Department of Computer Applications, Women's Christian College, Chennai

Abstract

Develop an Early Warning System to identify high-risk groups among adolescent girls and facilitate targeted interventions for chronic disease prevention. The EWS utilizes comprehensive dataset comprising demographic information and lifestyle factors such as dietary habits, physical activity, and sleep patterns, collected via surveys. The proposed dataset contains information of 409 students belongs to different age group. In this paper, analyzed this dataset to identify distinct clusters within the population based on shared demographic and lifestyle characteristics. The analysis of the dataset using machine learning clustering algorithms reveals distinct clusters within the adolescent population, each characterized by unique demographic and lifestyle profiles. The EWS successfully identifies high-risk groups for chronic diseases among adolescent girls, enabling healthcare providers to implement targeted interventions aimed at mitigating disease risk and improving long-term health outcomes.

The study employs machine learning clustering algorithms, including k-means clustering, to analyze the dataset and identify distinct clusters within the adolescent population based on shared demographic and lifestyle characteristics. By segmenting the population into homogeneous groups, the EWS facilitates the identification of at-risk individuals and the development of personalized intervention strategies. Introducing an Early Warning System tailored to adolescent girls' health needs using machine learning, advancing proactive chronic disease management.

Keywords: Machine Learning, Chronic disease, Clustering, Adolescent girls, Warning system.

Introduction

Adolescence is a pivotal phase in human development, characterized by significant physical, emotional, and behavioural changes. During this critical period, lifestyle choices have a profound influence on long-term health outcomes. In particular, adolescent girls face a unique set of challenges and opportunities related to their health and well-being. The decisions they make regarding diet, physical activity, sleep, and other lifestyle factors can have a lasting impact on their risk of developing chronic diseases later in life.Chronic diseases, such as type 2 diabetes, obesity, cardiovascular diseases, and PCOD, are of growing concern in modern healthcare. Identifying and addressing risk factors for these diseases in adolescents is crucial for preventive healthcare efforts. Adolescents often represent a population with relatively lower incidence rates of chronic diseases compared to adults, making this phase an ideal window of opportunity for intervention and prevention.This paper introduces an innovative approach to address this challenge—an Early Warning System (EWS) that harnesses the power of machine learning to identify lifestyle clusters among adolescent girls. By examining the shared demographic and lifestyle characteristics within these clusters, this EWS aims to uncover hidden patterns that contribute to chronic disease susceptibility. The ultimate goal is to empower healthcare professionals, parents, and adolescents themselves with the knowledge needed to tailor health interventions and preventive strategies to the specific needs and risks of this demographic.This studyexhibits the EWS's efficacy in discovering meaningful clusters that reveal common trends and associations among adolescent girls. These clusters offer valuable insights into shared risk factors and lifestyle patterns that contribute to chronic disease susceptibility.In a real-world deployment with a cohort of adolescent girls, the EWS successfully identifies and characterizes these clusters. By understanding the unique profiles and shared attributes within each cluster, health interventions can be tailored to address specific lifestyle factors and risks associated with chronic diseases.This paper discusses the practical implications of the proposed approach, emphasizing its potential to enhance public health initiatives by uncovering hidden patterns within demographic and lifestyle data. By focusing on clustering, the EWS empowers healthcare professionals with a deeper understanding of the diverse factors contributing to chronic disease risk among adolescent girls, ultimately facilitating more effective and personalized health interventions.

Literature Review

Machine Learning in Health Research*:* The utilization of machine learning techniques in healthcare has gained prominence in recent years. It offers the potential to process vast amounts of data to extract meaningful patterns, aid in risk prediction, and guide personalized interventions. Various studies have successfully employed machine learning in healthcare for tasks ranging from disease prediction to treatment recommendations [2-3].

Clustering and Identifying Health Patterns*:* Clustering techniques, such as k-means and hierarchical clustering, have been used in health research to uncover hidden structures within large datasets. By grouping individuals with similar demographic and lifestyle profiles, researchers can identify common trends, risk factors, and disease susceptibilities [6-7].

Early Warning Systems for Chronic Diseases*:* Developing early warning systems for chronic diseases is a proactive approach to public health. Such systems can facilitate early detection of risk factors and allow for timely interventions. Previous research has demonstrated the potential of EWS in improving health outcomes, especially in vulnerable populations [8-9].

Adolescent Health and Chronic Disease Risk*:* Adolescent health behaviors have a substantial impact on long-term chronic disease risk. A study in [10] highlights the links between adolescent health and future disease outcomes, emphasizing the importance of early interventions.

Predictive Analytics in Adolescent Health*:* Predictive analytics using machine learning has gained prominence in adolescent health research. In the study [11] demonstrates the use of machine learning models to predict chronic disease risk factors among adolescents, showcasing the potential for early intervention.Personalized Health Interventions*:* Tailoring health interventions to individual needs and risks is a growing trend in healthcare. Research by [12] discusses the benefits of personalized interventions in managing chronic diseases and improving health outcomes.

3.Methodology

3.1Dataset

Adolescent girl Student’s demographic and lifestyle parameters were collected through the questionnaire method. A real time data of 409 adolescent girl’s details were collected. The data was pre-processed by identifying the duplicates and null values in it. The dataset contains 41 attributes related to demographic details such as age, Mother’s occupation and educational qualification, Father’s occupation and educational qualification, income and physical activities like doing exercise per week, scale of walking per day, sleeping hour, experiencing sleeping disturbances, hours spent in socialmedia, stress management and dietary habits such as skipping breakfast, consuming fastfood, eating fruits and vegetables, taking sugary beverages, servings of whole grains per day, having balance diet, visiting doctor for regular checkup, getting digestive issue, any chronic disease in familyhistoryetc.The dataset is shown in Fig 1.

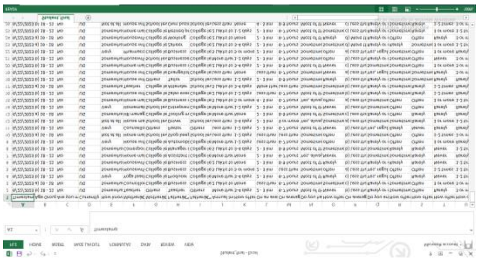


Fig 1: Sample Dataset

3.2AttributeSelection

The dataset attributes relevant to the category of dietary habits, physical activity, stress and sleep patterns were selected to analyse the characteristics of the student. These attributes contain the ordinal categorical values. Since these lifestyle attributes are relevant to analyse the student behaviour, they were selected to convert into numerical values by assigning weights in it.

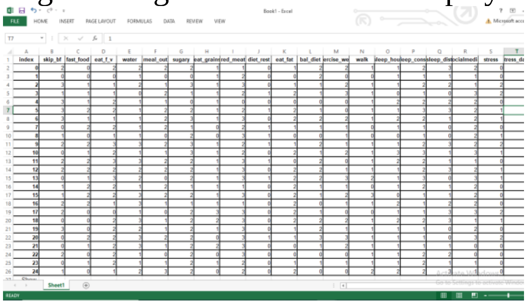
3.3Assigning Weights

Out of 41 attributes, 11 attributes were defined to analyse food habits, 9 attributes were used to extract information of physical and stress activity and regular activity details are collected in 5 attributes. Totally 25 attributes were used to analyse the student lifestyle. The weights were assigned in the scale of 0-3.Example: Under Food habit category, the student were analysed based on skipping breakfast, consuming fast food, frequently consuming meals outside home etc. Assign value based on the given criteria.

Table 1 :Weightage score

Never	**Rarely**	**1-2 times a week**	**3 or more times a week**
3	2	1	0

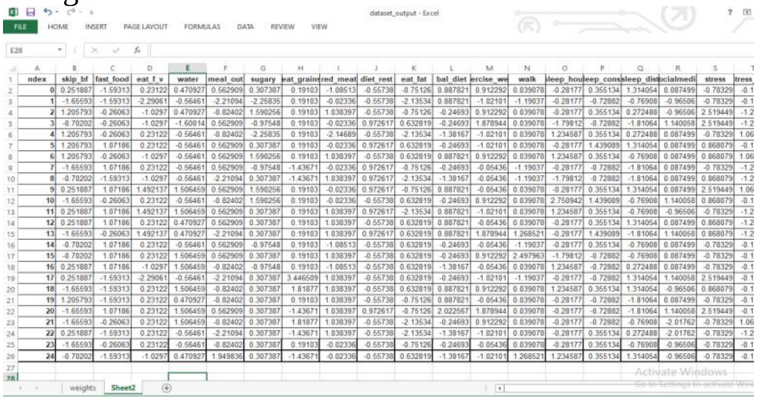
The Ordinal categorical attributes related to diet, physical activity, stress and sleep (20 attributes) were assigned weights as follows and displayed in Fig 2.



Book1 - Excel

T7 · 1

index	skip_bf	fast_food	eat_f_v	water	meal_out	sugary	eat_grain	red_meat	diet_rest	eat_fat	bal_diet	ercise_we	walk	sleep_hou	leep_cons	sleep_dist	ocialmedi	stress	tress_da
0	2	0	2	2	2	2	1	1	0	1	2	2	1	1	2	3	2	0	
1	0	0	0	1	0	0	1	2	0	0	2	0	0	1	1	1	1	0	
2	3	1	1	2	1	3	1	3	0	1	1	2	1	1	2	2	1	2	
3	1	1	1	0	2	1	1	2	1	2	1	3	1	0	1	0	3	2	
4	3	1	2	1	1	0	1	0	0	0	0	0	1	2	2	2	2	0	
5	3	2	2	1	2	2	1	2	1	2	1	0	1	1	3	3	2	1	
6	3	1	1	1	2	3	1	3	0	2	2	2	1	2	2	1	2	1	
7	0	2	2	1	2	1	0	2	1	1	1	1	0	1	1	0	2	0	
8	1	0	1	1	0	2	0	3	1	0	0	1	0	0	1	0	2	1	
9	2	2	3	3	2	3	1	2	1	1	2	1	1	1	2	3	2	2	
10	0	1	2	1	1	3	1	2	0	2	1	2	1	3	3	1	3	1	
11	2	2	3	3	2	2	1	3	1	0	2	0	1	2	2	1	1	0	
12	2	2	2	2	2	2	1	3	0	2	2	1	1	1	2	3	2	1	
13	0	1	3	2	0	2	1	3	1	2	2	3	2	1	3	0	3	1	
14	1	2	2	1	2	1	1	1	0	2	1	1	0	1	2	1	2	0	
15	1	2	2	3	2	2	1	3	0	2	1	2	3	0	1	1	2	0	
16	2	2	1	3	1	1	1	1	0	2	0	1	1	2	1	2	2	0	
17	2	0	0	1	0	2	3	3	0	2	1	0	0	1	1	3	3	2	
18	0	0	2	3	1	2	2	3	0	2	2	2	1	2	2	3	1	1	
19	3	0	2	2	1	2	1	3	0	1	2	1	1	1	1	0	2	0	
20	0	2	2	3	2	2	0	3	1	1	3	3	1	1	1	0	3	2	
21	0	1	2	3	1	2	2	3	0	0	1	2	1	1	1	1	0	0	
22	2	0	2	1	0	2	0	3	0	0	0	0	1	1	2	2	0	0	
23	0	1	2	1	1	2	1	2	0	1	1	1	1	1	2	1	1	0	
24	1	0	1	2	3	2	0	2	0	2	0	0	2	2	2	3	1	0	



dataset_output - Excel

E28

ndex	skip_bf	fast_food	eat_f_v	water	meal_out	sugary	eat_grain	red_meat	diet_rest	eat_fat	bal_diet	ercise_we	walk	leep_hou	eep_cons	sleep_dist	ocialmedi	stress	tress_
0	0.251887	-1.59313	0.23122	0.470927	0.562909	0.307387	0.19103	-1.08513	-0.55738	-0.75126	0.887821	0.912292	0.039078	-0.28177	0.355134	1.314054	0.087499	-0.78329	-0.1
1	-1.65593	-1.59313	-2.29061	-0.56461	-2.21094	-2.25835	0.19103	-0.02336	-0.55738	-2.13534	0.887821	-1.02101	-1.19037	-0.28177	-0.72882	-0.76908	-0.96506	-0.78329	-0.1
2	1.205793	-0.26063	-1.0297	0.470927	-0.82402	1.590256	0.19103	1.038397	-0.55738	-0.75126	-0.24693	0.912292	0.039078	-0.28177	0.355134	0.272488	-0.96506	2.519449	-1.2
3	-0.70202	-0.26063	-1.0297	-1.60014	0.562909	-0.97548	0.19103	-0.02336	0.972617	0.632819	-0.24693	1.878944	0.039078	-1.79812	-0.72882	-1.81064	1.140058	2.519449	-1.2
4	1.205793	-0.26063	0.23122	-0.56461	-0.82402	-2.25835	0.19103	-2.14689	-0.55738	-2.13534	-1.38167	-1.02101	0.039078	1.234587	0.355134	0.272488	0.087499	-0.78329	1.06
5	1.205793	1.07186	0.23122	-0.56461	0.562909	0.307387	0.19103	-0.02336	0.972617	0.632819	-0.24693	-1.02101	0.039078	-0.28177	1.439089	1.314054	0.087499	0.868079	-0.1
6	1.205793	-0.26063	-1.0297	-0.56461	0.562909	1.590256	0.19103	1.038397	-0.55738	0.632819	0.887821	0.912292	0.039078	1.234587	0.355134	-0.76908	0.087499	0.868079	1.06
7	-1.65593	1.07186	0.23122	-0.56461	0.562909	-0.97548	-1.43671	-0.02336	0.972617	-0.75126	-0.24693	-0.05436	-1.19037	-0.28177	-0.72882	-1.81064	0.087499	-0.78329	-1.2
8	-0.70202	-1.59313	-1.0297	-0.56461	-2.21094	0.307387	-1.43671	1.038397	0.972617	-2.13534	-1.38167	-0.05436	-1.19037	-1.79812	-0.72882	-1.81064	0.087499	0.868079	-1.2
9	0.251887	1.07186	1.492137	1.506459	0.562909	1.590256	0.19103	-0.02336	0.972617	-0.75126	0.887821	-0.05436	0.039078	-0.28177	0.355134	1.314054	0.087499	2.519449	1.06
10	-1.65593	-0.26063	0.23122	-0.56461	-0.82402	1.590256	0.19103	-0.02336	-0.55738	0.632819	-0.24693	0.912292	0.039078	2.750942	1.439089	-0.76908	1.140058	0.868079	-0.1
11	0.251887	1.07186	1.492137	1.506459	0.562909	0.307387	0.19103	1.038397	0.972617	-2.13534	0.887821	-1.02101	0.039078	1.234587	0.355134	-0.76908	-0.96506	-0.78329	-1.2
12	0.251887	1.07186	0.23122	0.470927	0.562909	0.307387	0.19103	1.038397	-0.55738	0.632819	0.887821	-0.05436	0.039078	-0.28177	0.355134	1.314054	0.087499	0.868079	-1.2
13	-1.65593	-0.26063	1.492137	0.470927	-2.21094	0.307387	0.19103	1.038397	0.972617	0.632819	0.887821	1.878944	1.268521	-0.28177	1.439089	-1.81064	1.140058	0.868079	-1.2
14	-0.70202	1.07186	0.23122	-0.56461	0.562909	-0.97548	0.19103	-1.08513	-0.55738	0.632819	-0.24693	-0.05436	-1.19037	-0.28177	0.355134	-0.76908	0.087499	-0.78329	-0.1
15	-0.70202	1.07186	0.23122	1.506459	0.562909	0.307387	0.19103	1.038397	-0.55738	0.632819	-0.24693	0.912292	2.497963	-1.79812	-0.72882	-0.76908	0.087499	-0.78329	-0.1
16	0.251887	1.07186	-1.0297	1.506459	-0.82402	-0.97548	0.19103	-1.08513	-0.55738	0.632819	-1.38167	-0.05436	0.039078	1.234587	-0.72882	0.272488	0.087499	-0.78329	-0.1
17	0.251887	-1.59313	-2.29061	-0.56461	-2.21094	0.307387	3.446509	1.038397	-0.55738	0.632819	-0.24693	-1.02101	-1.19037	-0.28177	-0.72882	1.314054	1.140058	2.519449	-0.1
18	-1.65593	-1.59313	0.23122	1.506459	-0.82402	0.307387	1.81877	1.038397	-0.55738	0.632819	0.887821	0.912292	0.039078	1.234587	0.355134	1.314054	-0.96506	0.868079	-0.1
19	1.205793	-1.59313	0.23122	0.470927	-0.82402	0.307387	0.19103	1.038397	-0.55738	-0.75126	0.887821	-0.05436	0.039078	-0.28177	-0.72882	-1.81064	0.087499	-0.78329	-0.1
20	-1.65593	1.07186	0.23122	1.506459	0.562909	0.307387	-1.43671	1.038397	0.972617	-0.75126	2.022567	1.878944	0.039078	-0.28177	-0.72882	-1.81064	1.140058	2.519449	-0.1
21	-1.65593	-0.26063	0.23122	1.506459	-0.82402	0.307387	1.81877	1.038397	-0.55738	-2.13534	-0.24693	0.912292	0.039078	-0.28177	-0.72882	-0.76908	-2.01762	-0.78329	1.06
22	0.251887	-1.59313	0.23122	-0.56461	-2.21094	0.307387	-1.43671	1.038397	-0.55738	-2.13534	-1.38167	-1.02101	0.039078	-0.28177	0.355134	0.272488	-2.01762	-0.78329	-1.2
23	-1.65593	-0.26063	0.23122	-0.56461	-0.82402	0.307387	0.19103	-0.02336	-0.55738	-0.75126	-0.24693	-0.05436	0.039078	-0.28177	0.355134	-0.76908	-0.96506	-0.78329	-0.1
24	-0.70202	-1.59313	-1.0297	0.470927	1.949836	0.307387	-1.43671	-0.02336	-0.55738	0.632819	-1.38167	-1.02101	1.268521	1.234587	0.355134	1.314054	-0.96506	-0.78329	-0.1

Fig 2: Dataset after assigning weights *Fig 3: Dataset after pre-processing*

3.4Scaling the data

Pre-processing technique was used to transform the data to standard range. The standard scalar function brings all the features in the same magnitude and ensures that the each feature the mean is 0 and the variance is 1. It adjusts the feature to make the data suitable for the algorithm and the same is shown in Fig 3.

3.5 Identification Of 'K'

In the proposed dataset, kmeans clustering algorithm is applied for obtaining the cluster. It is the most common algorithm used to identify the cluster. The required number of cluster (k) is calculated by the maximum score of silhouette measure among the values of k from 2 to 9.It is shown in Table 2 and Fig 4.Based on the these score , the required number of cluster is chosen as '2'.

Table 2 :Silhouette Measure

K	2	3	4	5	6	7	8	9
Silhouette score	0.0809	0.0612	0.0549	0.0593	0.0632	0.0582	0.0540	0.0576

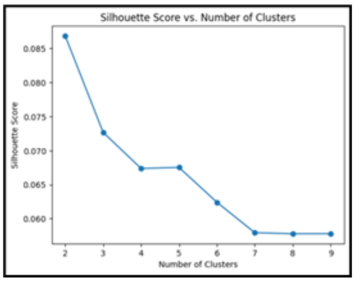


Fig 4: Silhouette Score for selecting no. of cluster

4.Results And Analysis

K-means algorithm is applied in the proposed dataset (409students). The dataset contains the students belongs to different age group, Most of the students (314 students) belongs to 18 – 21 category refer in Fig 5. Students were grouped into 2 clusters. Thenumber of students in each cluster is represented in Table 3.

Table 3 :Students count in cluster

CLUSTER	TOTAL NUMBER OF STUDENTS
0	190
1	219

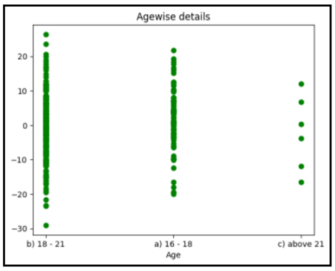
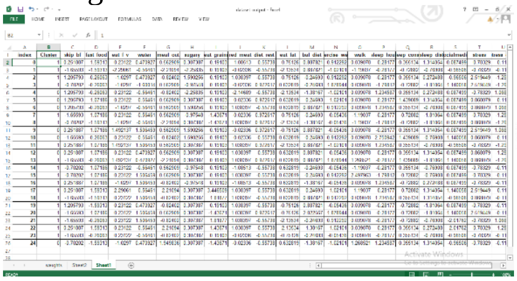


Fig 5: Age category used for the study

The Fig 6.Indicates the dataset with the cluster number for each data.



index	Cluster	skip bf	fast food	eat f v	water	meal out	sugary	eat grain	red meat	diet rest	eat fat	bal diet	ercise we	walk	sleep hou	eep con	sleep dis	cialmedi	stress	trave
0	1	0.251887	1.59313	0.23122	0.470927	0.562909	0.307387	0.19103	1.08613	0.55738	0.75126	0.887821	0.912292	0.039078	0.28177	0.355134	1.314054	0.067459	0.70329	0.11
1	1	-1.66683	-1.59313	-2.29061	-0.55461	-2.21094	-2.25835	0.19103	-0.02336	-0.55738	-2.13534	0.887821	-1.02101	-1.19037	-0.28177	-0.72882	-0.76938	-0.96506	-0.78329	-0.11
2	1	1.205793	0.26063	-1.0297	0.470927	-0.82402	1.590256	0.19103	1.038397	-0.55738	-0.75126	-0.24693	0.912292	0.039078	-0.28177	0.355134	0.272488	-0.96506	2.519449	-1.25
3	1	-0.70202	-0.26063	-1.0297	-1.60314	0.562909	-0.97548	0.19103	-0.02336	0.972517	0.632819	-0.24693	1.078544	0.039078	-1.79812	-0.72882	-1.81064	1.140058	2.519449	-1.25
4	0	1.205793	-0.26063	0.23122	-0.55461	-0.82402	-2.25835	0.19103	-2.14689	-0.55738	-2.13534	-1.38167	-1.02101	0.039078	1.234587	0.355134	0.272488	0.067459	-0.78329	1.368
5	0	1.205793	1.07186	0.23122	0.55461	0.562909	0.307387	0.19103	0.02336	0.972517	0.632819	0.24693	1.02101	0.039078	0.28177	1.439009	1.314054	0.067459	0.868079	0.11
6	0	1.205793	-0.26063	-1.0297	-0.55461	0.562909	1.590256	0.19103	1.038397	-0.55738	0.632819	0.887821	0.912292	0.039078	1.234587	0.355134	-0.76938	0.067459	0.868079	1.368
7	1	1.66593	1.07186	0.23122	0.55461	0.562909	0.97548	1.43671	0.02336	0.972517	0.75126	0.24693	-0.05436	1.19037	0.28177	0.72882	1.81064	0.067459	0.78329	1.25
8	1	-0.70202	-1.59313	-1.0297	-0.55461	-2.21094	0.307387	-1.43671	1.038397	0.972517	-2.13534	-1.38167	-0.05436	-1.19037	-1.79812	-0.72882	-1.81064	0.067459	0.868079	-1.25
9	0	0.251887	1.07186	1.492137	1.536453	0.562909	1.590256	0.19103	-0.02336	0.972517	-0.75126	0.887821	-0.05436	0.039078	-0.28177	0.355134	1.314054	0.067459	2.519449	1.368
10	0	1.66593	0.26063	0.23122	-0.55461	0.82402	1.590256	0.19103	0.02336	0.55738	0.632819	0.24693	0.912292	0.039078	2.753842	1.439009	0.76938	1.140058	0.868079	0.11
11	1	0.251887	1.07186	1.492137	1.536453	0.562909	0.307387	0.19103	1.038397	0.972517	-2.13534	0.887821	-1.02101	0.039078	1.234587	0.355134	-0.76938	-0.96506	-0.78329	-1.25
12	0	0.251887	1.07186	0.23122	0.470927	0.562909	0.307387	0.19103	1.038397	0.55738	0.632819	0.887821	0.05436	0.039078	0.28177	0.355134	1.314054	0.067459	0.868079	1.25
13	1	-1.66683	-0.26063	1.492137	0.470927	-2.21094	0.307387	0.19103	1.038397	0.972517	0.632819	0.887821	1.078544	1.268521	-0.28177	1.439009	-1.81064	1.140058	0.868079	-1.25
14	1	-0.70202	1.07186	0.23122	-0.55461	0.562909	-0.97548	0.19103	-1.08613	-0.55738	0.632819	-0.24693	-0.05436	-1.19037	-0.28177	0.355134	-0.76938	0.067459	-0.78329	-0.11
15	1	-0.70202	1.07186	0.23122	1.536453	0.562909	0.307387	0.19103	1.038397	-0.55738	0.632819	-0.24693	0.912292	2.497963	-1.79812	-0.72882	-0.76938	0.067459	-0.70329	0.11
16	1	0.251887	1.07186	-1.0297	1.536453	-0.82402	-0.97548	0.19103	-1.08613	-0.55738	0.632819	-1.38167	-0.05436	0.039078	1.234587	-0.72882	0.272488	0.067459	-0.78329	-0.11
17	1	0.251887	1.59313	2.29061	0.55461	2.21094	0.307387	3.446509	1.038397	0.55738	0.632819	0.24693	1.02101	1.19037	0.28177	0.72882	1.314054	1.140058	2.519449	0.11
18	1	-1.66683	-1.59313	0.23122	1.536453	-0.82402	0.307387	1.81877	1.038397	-0.55738	0.632819	0.887821	0.912292	0.039078	1.234587	0.355134	1.314054	-0.96506	0.868079	-0.11
19	1	1.205793	-1.59313	0.23122	0.470927	-0.82402	0.307387	0.19103	1.038397	-0.55738	-0.75126	0.887821	-0.05436	0.039078	-0.28177	-0.72882	-1.81064	0.067459	-0.78329	-0.11
20	1	-1.66683	1.07186	0.23122	1.536453	0.562909	0.307387	-1.43671	1.038397	0.972517	-0.75126	2.022567	1.078544	0.039078	-0.28177	-0.72882	-1.81064	1.140058	2.519449	-0.11
21	1	-1.66683	-0.26063	0.23122	1.536453	-0.82402	0.307387	1.81877	1.038397	-0.55738	-2.13534	-0.24693	0.912292	0.039078	-0.28177	-0.72882	-0.76938	-2.01762	-0.78329	1.368
22	1	0.251887	1.59313	0.23122	0.55461	2.21094	0.307387	1.43671	1.038397	0.55738	2.13534	1.38167	1.02101	0.039078	0.28177	0.355134	0.272488	2.01762	0.70329	1.25
23	1	-1.66683	-0.26063	0.23122	-0.55461	-0.82402	0.307387	0.19103	-0.02336	-0.55738	-0.75126	-0.24693	-0.05436	0.039078	-0.28177	0.355134	-0.76938	-0.96506	-0.78329	-0.11
24	0	-0.70202	-1.59313	-1.0297	0.470927	1.949636	0.307387	-1.43671	-0.02336	-0.55738	0.632819	-1.38167	-1.02101	1.268521	1.234587	0.355134	1.314054	-0.96506	-0.78329	-0.11

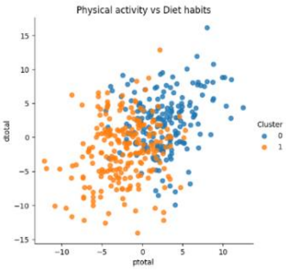
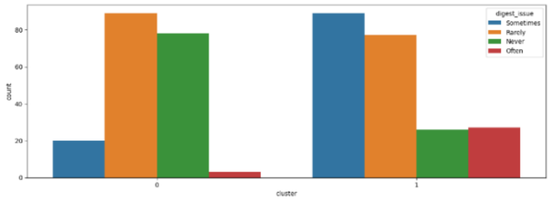


Fig 6: Cluster number for each data pointsFig 7:Scatter graph for physical and dietary habit

For each category like physical activities, dietary habits and lifestyle regular activities, the total was calculated. The given scatter graph Fig 7, represents physical and dietary habits of the student. Students belong to cluster 0 showsthey give more importance to physical activities and dietary habits when compared to cluster 1 student. Fig 8. Shows, Cluster 1 students are suffering with digest issues often when compared to cluster 0 students. More number of students of cluster 0 are never had digestive issues.



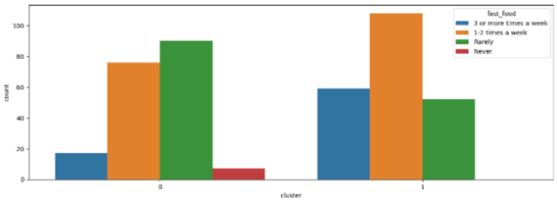


Fig 8:Comparison of Clusters based on digestive issues

Fig 9:Comparison of Clusters based on Consumption of fast food

From the Fig 9, conclude that Cluster 1 students are consuming fast food 3 or more times in a week more than cluster 0 students. One or 2 times consuming fast food by the cluster 0 students is less than the students of cluster 1.

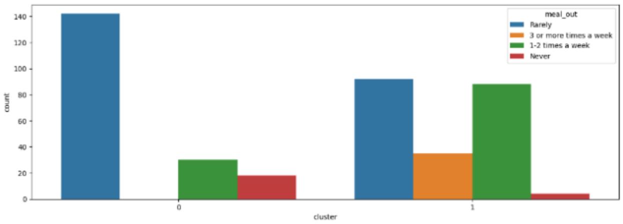


Fig 10:Comparison of Clusters based on consuming meals outside

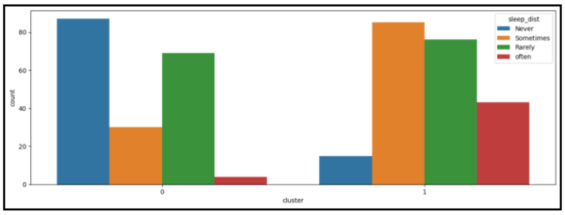
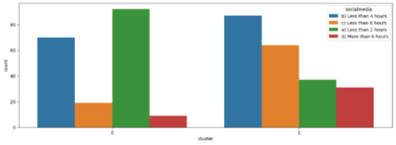


Fig 11:Comparison of clusters based on Sleep disturbances

The students of Cluster 0 consuming Meals outside rarely are high when compared to the students of cluster 1. Less than 40 students of cluster 1 are taking meals outside three or more times in a week which is explored in Fig 10.From the Fig 11. It was found that the number of students (40) belongs to cluster 1 had got often sleep disturbances more than the students in cluster 0. More than 80 students of cluster 0 never had sleep disturbances.



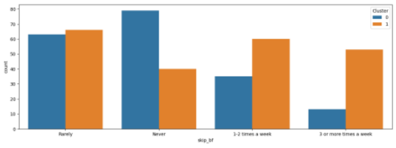


Fig 12:*Comparison of Clusters based on Social media usageFig 13:Comparison of Clusters based on*

skipping breakfast

The number of students of cluster1 are accessing social media (more than 6 hours) is high when compared to cluster 0 students. Many students of cluster 0 are accessing social media less than 2 hours is higher than cluster 1 students. Refer Fig 12.

In cluster 1, More than 60 students are skipping breakfast 1 or 2 times in a week and more than 50 students are doing 3 or more times a week. In cluster 0, more than 75 students are never skipping their breakfast which is exhibited in Fig 13.The students who have taken 1-2 servings of whole grains, don't have chronic disease in their family.

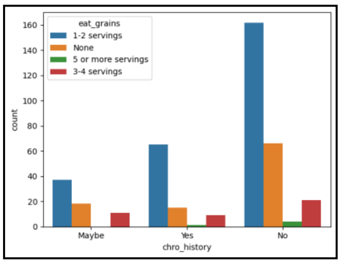


Fig 14:Comparison of Clusters based on eating grains and chronic history in family

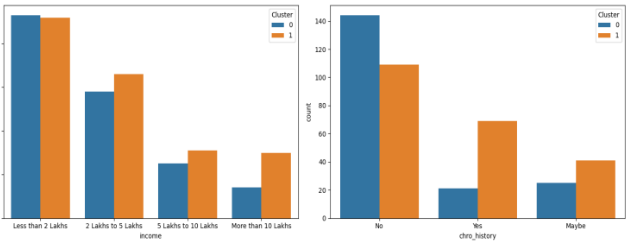
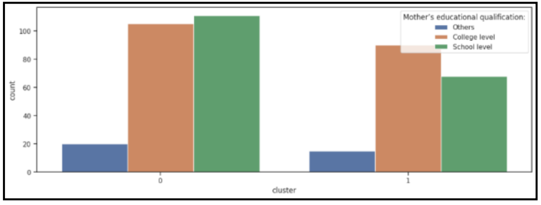


Fig 15: Cluster based on parent's income and family history

The Cluster 1 student's parent income is higher; they are consuming fast food, having their meals outside more number of times in a week. They eat high-fat and high sugar foods more than the students of cluster 0.Shown in Fig 15.



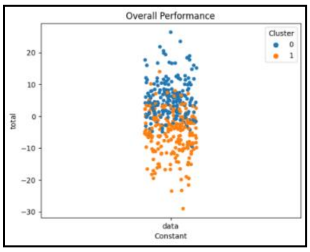


Fig 16:Pattern based on Mothers educational Qualification

Fig 17 Pattern of cluster based on Total

The number of Cluster 0 students mother, have educational qualification more than cluster 1 mother parent, which is shown in Fig 16.The Overall performance of each student is calculated by summarizing the value of dietary habits, physical activities and Demographic values. The values are defined in the Fig 17.

5.Conclusion

The findings in this work, as depicted in Figures 8 to 16, reveal significant variations in lifestyle factors, dietary habits, and health-related behaviours between the two clusters of students, namely Cluster 0 and Cluster 1. These distinctions shed light on the diversity of habits and priorities among the student population, offering valuable insights for future health management and intervention strategies.Cluster 0 students demonstrate a greater emphasis on physical activities and dietary habits, highlighting a proactive approach to maintaining a healthier lifestyle. They are less likely to experience digestive issues, consume fast food in moderation, and rarely eat meals outside the home. This group also exhibits a lower prevalence of sleep disturbances, less time spent on social media, and a strong commitment to not skipping breakfast.

In contrast, Cluster 1 students appear to face distinct challenges. They experience more digestive issues, exhibit higher consumption of fast food and high-fat, high-sugar foods, and tend to eat meals outside more frequently. They are also more likely to have sleep disturbances, spend extended periods on social media, and exhibit a pattern of breakfast skipping.These findings underline the importance of tailored health interventions. It is clear that a one-size-fits-all approach may not be effective in addressing the diverse needs of these two student clusters. The educational qualifications of mothers and family income also play a role in shaping the health behaviours of students, as evidenced in the data.

6.Future Works

Building on these findings, there are several avenues for future research and practical applications. Develop and implement targeted health interventions based on the specific needs of each cluster. These interventions could include educational programs, dietary guidance, and strategies to improve sleep hygiene.Explore the mental health aspects of students in both clusters, as reflected in their sleep patterns and social media usage. Assess how these factors relate to mental well-being and consider implementing mental health support programs.Conduct a comparative study of various clustering algorithms beyond K-means, such as hierarchical clustering, DBSCAN, and spectral clustering. Evaluate their performance in terms of clustering accuracy, interpretability, and computational efficiency.

7.References

[1] Sawyer SM, Afifi RA, Bearinger LH, Blakemore SJ, Dick B, Ezeh AC, et al. Adolescence: A foundation forfuture health. Lancet. 2012 May 5;379(9826):1630-40. DOI: 10.1016/S0140-6736(12)60072-5

[2] Rajkomar A, Oren E, Chen K, Dai AM, Shojania K, Ho J, et al. Scalable and accurate deep learning withelectronic health records. npj Digital Medicine. 2019 Dec;2(1):1-10. DOI: 10.1038/s41746-019-0180-0

[3] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skincancer with deep neural networks. Nature. 2019 Jan 2;542(7639):115-8. DOI: 10.1038/nature21056

[4] Popkin BM, Adair LS, Ng SW. Global nutrition transition and the pandemic of obesity in developingcountries. Nutr Rev. 2012 Jan;70(1):3-21. DOI: 10.1111/j.1753-4887.2011.00456.x

[5] Hu FB, Manson JE, Stampfer MJ, Colditz G, Liu S, Solomon CG, et al. Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. N Engl J Med. 2001 Sep 13;345(11):790-7. DOI: 10.1056/NEJMoa010492

[6] Wang X, Smith C, Montes F, Crow L, Ball M. A systematic review of clustering and classification of lifecourse epidemiology research. Int J Environ Res Public Health. 2017 Oct;14(10):1097. DOI: 10.3390/ijerph14101097

[7] Ng T, Teo T, Yeo WK. Clustering and classification of data patterns for healthcare applications. Int J Environ Res Public Health. 2018 Mar;15(3):417. DOI: 10.3390/ijerph15030417

[8] Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden ofDisease Study 2010. Lancet. 2012 Dec 15;380(9859):2163-96. DOI: 10.1016/S0140-6736(12)61729-2

[9] Gidlow CJ, Ellis NJ, Bostock S, McKenna J. A before and after study of the impact of the '(name removed) Active Work' programme on sedentary behaviour and physical activity. Public Health. 2018 Oct;165:26-34. DOI: 10.1016/j.puhe.2018.07.006

[10] Patton GC, Sawyer SM, Santelli JS, Ross DA, Afifi R, Allen NB, et al. Our future: a Lancet commission onadolescent health and wellbeing. Lancet. 2016 Jun 11;387(10036):2423-78. DOI: 10.1016/S0140-6736(16)00579-1

[11] Charakorn K, Kambhampati C. Predictive analytics and machine learning for adolescents' health risk assessment. ComputBiol Med. 2021 Mar;133:104364. DOI: 10.1016/j.compbiomed.2021.104364

[12] Liu X, Liu J, Liu L, Yan B. Personalized management of chronic diseases through mHealth. J Pers Med.2019 Dec;9(4):47. DOI: 10.3390/jpm9040047

Authors' background

Your Name	Title*	Research Field	Personal website
Mrs.R.Arulmathi	Assistant Professor	Machine Learning	arulmathir@wcc.edu.in
Dr.R.Lakshmidevi	Assistant Professor	Machine Learning	lakshmidevir@wcc.edu.in
Mrs.D.Sylvia Mary	Head of the Department	Machine Learning	sylviamary@wcc.edu.in