International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-4, November 2019

Consumption of Food by College Students using Ml Algorithms

R. RamyaSri, G. Susmitha, K.SaiSurya, V.Bhavani, K Venkata Raju

Abstract—We are well aware of the many problems that our current generations are facing. From all these new enhancements in the real world, it has been quite hard for them to keep up with everything evolving around them. Keeping all this in mind, they work day in and out to make sure that their knowledge of their surroundings up to date. However, we believe that they fail to take care of themselves in the process correctly. No matter how much a particular individual may withstand in terms of workload, stress, or other mental & emotional barriers, our physical body will always be the critical aspect to overcoming them. Most people believe that working out and maintaining physical fitness are the significant aspects to sustain an excellent physical form, but they simply overlook the most crucial element which is their eating habits [1]. Although our body may be physically fit, the nourishment of our body depends on the eating styles that we follow on a day to day basis. Food is what nourishes our body with most of the proteins & minerals that we require; without it, we wouldn't be able to accomplish much. On conducting worldwide research on people's lifestyles, we were able to conclude that over the past 33 years, the obesity rate among human beings has increased by a mere 27.5%. What seems to be the most thoughtful yet intriguing fact is that although many people are overweight as well as obese, they still believe that their eating habits are healthy [2]. Most people are living in the dilemma of the fact that they maintain a healthy lifestyle. We aim to study the views on a healthy lifestyle as per the norms of our current generation. We want to analyze their daily eating habits as well as their thoughts on their lifestyle. So the question that remains is "What exactly is a Healthy Eating Lifestyle?"[3]To know the accuracy of our data set we use the machine learning algorithms like Support Vector Machine (svm), Naïve-Bayes Classifier which are used for both classification and regression [4].

Index Terms: Support vector machine (SVM), Naive Bayes classifier

I. INTRODUCTION

Our project mainly focuses on analyzing the lifestyles in terms of food on our current generation.

For us to accomplish this, we have conducted an extensive survey which consisted mostly of students from our college as well as a few of the colleges in our locality.
By this survey, we came to know that whether they are consuming the nutritional food or not based on their preference given in the study. For this, we have calculated the accuracy of our dataset with the help of machine learning algorithms like support vector machine (SVM) and Naïve-Bayes classifier [5].

1. **Support Vector Machine (SVM):**
It is a supervised learning method used for the classification of both linear and non-linear data. It uses a non-linear mapping to a highly sufficient dimension; data from two classes can always be separated by a hyperplane (a "decision boundary" separating the tuples of one class from another). These are much less prone to overfitting. These support vectors can provide a compactness description of the learned model. SVM can also be used for numeric prediction, regression, and outlier-detection.
SVM is applied in many areas, including handwritten digit recognition, object recognition, speaker identification, and benchmark time-series prediction tests [12].

1.1Linear Support Vector Machine Classifier:
In this, the data points are considered as a p-dimensional vector, and (p-1) dimensional points separate them with a hyper-plane. There may be many hyperplanes separating the linear data, but we have to choose the best hyper-plane, which maximizes the edges (the distance between hyper-plane and the closest data point of either class).
The maximum margin hyper-plane is determined by the data points that are nearest to it, which maximizes the distance between the hyper-plane and data points. The data points which Support our hyper-plane are known as support vectors.

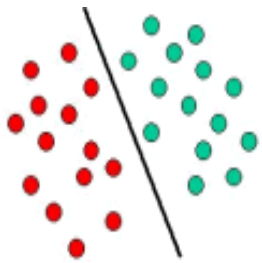

Figure 1.1(a) **Linear Support Vector Machine Classifier**

1.2 Non-Linear Support Vector Machine Classifier:
In this, data points are non-linearly separable in a p-dimensional space. These non-linear hyper-planes are drawn by using kernel tricks. Every kernel holds a non-linear kernel functions.

Manuscript published on November 30, 2019.
* Correspondence Author
R. RamyaSri*, Department of Computer Science and Engineering Koneru Lakshmaiah Education Foundation Vaddeswaram, AP, India
G. Susmitha, Department of Computer Science and Engineering Koneru Lakshmaiah Education Foundation Vaddeswaram, AP, India
K.SaiSurya, Department of Computer Science and Engineering Koneru Lakshmaiah Education Foundation Vaddeswaram, AP, India
V.Bhavani, Asst Professor, Koneru Lakshmaiah Education Foundation, Vaddeswaram, India
K Venkata Raju, Assoc Professor, Koneru Lakshmaiah Education Foundation Vaddeswaram, India

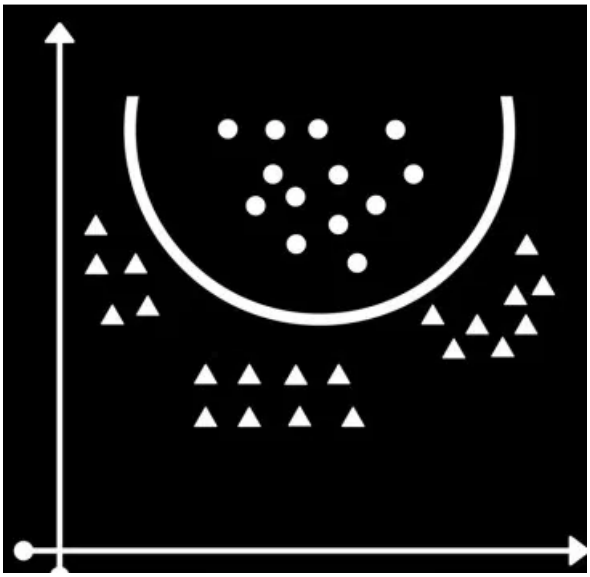

*Figure 1.1(b)***Non-Linear Support Vector Machine Classifier**

2. Naive-Bayes Classifier:

It is a supervised learning method used for the classification of massive datasets. It gives accurate results when we use it for textual data analysis. It mainly works based upon on the Bayes theorem [16].

1. Bayes theorem:

It uses mostly the conditional probability. Conditional probability is the likelihood of an event that will happen, given that an event that has already occurred. By this, we can calculate the possibility of this event using its prior knowledge.

$$P(H \mid E) = \frac{P(E \mid H) * P(H)}{P(E)}$$

Where,

P(H)→probability of hypothesis H being actual. (prior probability).

P(E)→probability of the evidence.

P(E/H)→probability of the evidence given that the hypothesis is true.

P(H/E)→probability of the hypothesis given that evidence is there.

II. PROCEDURE

Our project focused on 4 major stages. The first stage consisted of us gathering the required Data for our Study. The next two consist of us importing the collected data into our program and having it pre-processed. The last stage consists of the implementation of algorithms. Once we have successfully implemented our data with various libraries within '**R**', we can compare the accuracy of the dataset. Let's take a closer look at each step in our process[6].

The proposed system that we are going to use is:

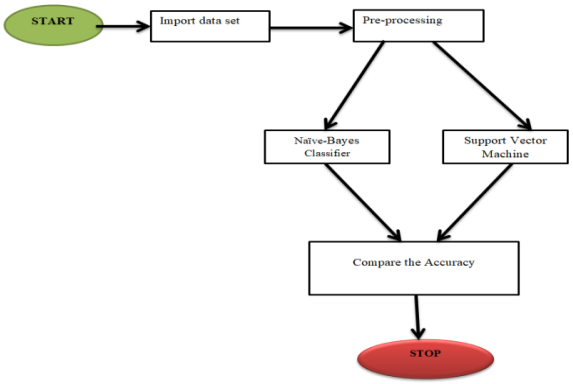

Figure 2.Proposed system of our dataset

1) Collection of Data:

For us to commence our study on the lifestyles of the current generation, we needed data sets to work with. We required data sets that would help us understand 2 significant aspects

i) The Daily Eating Habits of a person.

ii) The person's outlook on whether their eating style is healthy or not.

Keeping these two critical aspects in mind, we created an online survey using google forms for students to fill out. The study mainly focused on a person's daily eating habits focusing on what types of food they consumed. From these habits, we ended our survey with a simple question on their views or opinions on whether their lifestyles are healthy. Once our investigation was completed, we were able to output our data in the form of an Excel sheet. The sheet had all the questions placed along the x-axis, while each entry was considered on the y-axis. Once we have properly saved our excel sheet on our computer, we can now progress to the next stage of implementation and Analysis.

The attributes that we gathered from our survey form are Age, Gender, Diet Plan, time you wake in weekdays, time you wake in weekend, prefer as soon as you wake, Breakfast in a Rush, skip Breakfast, more in Breakfast, Breakfast in Restaurant, Breakfast in Canteen, Breakfast in Street, After your Breakfast, break time, Lunch Time, food in Lunch, type of food in Lunch, skip Lunch, Snacks you eat more, often you have your Snack, Dinner Time, eat more in Dinner, skip your Dinner, after Dinner, kind of food you prefer more and other values for attributes.

.

2) Importing our Metadata:

The most crucial step in any process is gathering all of our pre-requirements. In this stage, we must introduce 2 of the essential requirements into our program, the data along with the required libraries for analysis.

2.1) Importing our Survey Data:

We will be implementing our project with the help of ' R '. For us to implement all our Data-sets that are stored in a single excel sheet into our program, we must undergo the following procedure. The most important part is making sure that our sheet is properly saved in the folder of our wish. From here on out, we will use the setwd() inbuilt within ' R ' so that we can set the path to the folder in which our sheet is saved.

Once we have used the setwd() function in ' R ', the path will be properly set to the designated folder. The next step involves selecting our datasheet so that our program can further access it for future analysis. To do this, we will be using the inbuilt read.csv() function. This function, however, will return our file to have it stored in the form of a vector. To do this, we must first declare a variable and then read our sheet into it.

2.2) Importing Required Libraries:

For us to further access the many tools within 'R**',** we will have first to import the required libraries. One of the essential Libraries in ' **R** ' is 'e1071','klar','caret'.

These are the most important and widely used packages in ' **R** ' as it permits us to analyse our data in many different methods statistically.

It provides us many different functions that allow us to study the structure of our data as well as the variety of data collected. Further services contained in this library permit us to plot the statistical analysis of our data in many different formats. The study of data is entirely flexible with the desires of the user and can provide very accurate values and details.

3) Data preprocessing:

This data has some quality that satisfies the requirements of its use. Many factors consist of data quality, including accuracy, completeness, consistency, timeliness, believability, and interpretability. In real-world, the data is dirty incomplete with lacking attribute values, specific attributes of interest, or containing only the aggregate data, and incomplete data may come from, "Not applicable" data value when collected, Different considerations between the time when the data was collected and when it is analysed. Human/hardware/software problems [6].

3.1 Noisy:

Containing errors or outliers Noisy data (incorrect values) may come from Faulty data collection instruments, Human or computer error at entry of data, Errors in the transmission of data [7].

e.g., Salary= "-10"

3.2 Inconsistent:

Containing some disparities in codes or names, Inconsistent data may come from, Different data sources, Functional dependency violation (e.g., modify some linked data) [8].

Duplicate records also need data cleaning

e.g., Age="52" Birthday="03/07/1998"

e.g., Was rating "1,2,3", now rating "A, B, C".

e.g., discrepancy between duplicate records.

3.3 Data Pre-processing Important:

Quality decisions must be based on quality data, Data warehouse needs consistent integration of quality data, Data extraction, cleaning, and transformation comprises the majority of the work of building a data warehouse [9].

e.g., duplicate or missing data may cause incorrect or even misleading statistics.

3.4 Major Tasks in Data Pre-processing:

One of the major tasks or techniques that we are using in data pre-processing is data cleaning.

3.5 Data cleaning:

It is a process of identifying and changing or removing the data from a data-set, table, or database and which identifies the incomplete, incorrect, inaccurate, or irrelevant parts of the data, and this data is used in replacing, modifying, or deleting the dirty data. It may be executed interactively with data wrangling tools, or as batch processing through scripting. After this, a data set should be consistent with other similar data sets in the system. The inconsistencies detected or removed may have been initially caused by user entry errors, by changing in transmission or storage, or by different data dictionary definitions of similar entities in different stores. It differs from data validation in that validation almost invariably means data is rejected from the system at entry and is executed at the time of entry, rather than on batches of data [10].

The primary process of this is, it may involve removing typographical errors and correcting those values against a list of entities. The correction may be strict (such as rejecting any address that does not have a valid postal code) or fuzzy (such as correcting records that partially match existing, known records). Some of this data cleansing solutions which will clean data by cross-checking with a validated data set. A typical data cleansing practice is data enhancement, where data is made more complete by adding related information. For example, appending addresses with any phone numbers related to that address. Data cleansing may also involve activities like harmonization of data and standardization of data. For example, harmonization of shortcodes (st, rd, etc.) to actual words (street, road, et cetera). Standardization of data is a means of changing a reference data set to a new standard, ex, use of standard codes [11].

3.6 Missing Data:

In this, the data is not always available

E.g., many data records have no recorded value for several attributes, such as customer income in sales data, missing data may need to be inferred [13]

Missing data may be due to:

1. Equipment malfunction
2. Inconsistent data
3. Data not entered due to misunderstanding
4. Certain data may not be considered necessary at the time of entry
5. Not register history or changes of the data

3.7 Handle Missing Data:

Ignore the tuple when the class label is missing (assuming the tasks in classification are not active even when the percentage of missing values per attribute varies [14].

1. Fill the missing value manually: tedious + infeasible?
2. Fill in it automatically with

I.A global constant: e.g., "unknown", a new class?

II.the attribute mean

III.the attribute mean for all samples belonging to the same class: smarter

IV.the most probable value: inference-based such as Bayesian formula or decision tree.

3.8 Noisy Data:

1. Random error or variance in a measured variable
2. Incorrect attribute values may be due to

I. faulty data collection instruments

II. Entry of data problems

III. Transmission of data problems

IV. Inconsistency in naming convention

3. Other data problems which requires data cleaning

I. duplicate records

II. Incomplete data

III. Inconsistent data

3.9 Handle Noisy Data:

- Binning
  - first sort data and partition into (equal-frequency) bins
  - Then one can smooth by bin means, smooth by bin median, smooth by bin boundaries, etc. [15].
- Regression
  - smooth by fitting the data into regression function

4. Implementation of Algorithms:

These algorithms are machine-learning techniques that are used for classification, regression, prediction, and outlier-detection. The machine learning techniques are used for both supervised learning and unsupervised learning [20].

Some of the machines learning techniques are SVM, Decision Tree, Random Forest, Logistic Regression, KNN, K-Classifier, Naïve-Bayes classifier [21].

Now we are considering the two algorithms like SVM and Naïve-Bayes classifier for getting the accuracy of the dataset [22].

4.1 Implementation of SVM-Algorithm:

Using e1071 and caret library of the R language, we implemented a support vector machine algorithm to classify data about the nutritious diet of several users collected from the internet. The data contain attributes like age. Gender, diet plan of different users. The main decision we need to make from the set is whether the person is eating healthy food or not. Here kind of food you eat more is the dependent attribute and remaining all are independent attributes. We divided the entire data into two categories one is for training purposes, and another is for testing purposes. Train and test are in the ratio of 70:30 compositions. Once the training of data is completed, we can predict the outcomes of different data of similar kind to our model through the model we build with the help of a support vector machine in a supervised learning mechanism. This model predicts almost 75 percent accurately [17].

4.2 Implementation of Naïve-Bayes classifier-Algorithm:

Using klar library of the R language, we implemented the Naive Bayes algorithm to classify data about the nutritious diet of several users collected from the internet. The data contain attributes like age. Gender, diet plan of different users. The main decision we need to make from the set is whether the person is eating healthy food or not. Here kind of food you eat more is the dependent attribute and remaining all are independent attributes. We divided the entire data into two categories one is for training purposes, and another is for testing purposes. Train and test are in the ratio of 70:30 compositions. Once the training of data is completed, we can predict the outcomes of different data of similar kind to our model through the model we build with the help of naive Bayes classifier in a supervised learning mechanism. This model predicts almost 73 percent accurately [18].

4.3 Comparison of Algorithms:

After the implementation of SVM and Naïve-Bayes classifier for our dataset. Compare the accuracy's of both algorithm and select the best algorithm [19].

III. RESULTS

The outputs for the SVM algorithm are:

```
> head(data)
  gender   Age      Institute Diet.Plan time..you.wake.in.week.days time.you.wake.in.a.week.end
1 Female 20-25 Vignan Nirula        Yes          7:00 A.M-8:00 A.M          8:00 A.M-9:00 A.M
2 Female 15-20  Kl University       No          6:00 A.M-7:00 A.M         9:00 A.M-10:00 A.M
3 Female 15-20  Kluniversity        No          5:00 A.M-6:00 A.M          7:00 A.M-8:00 A.M
4 Female 25-30          India       No          5:00 A.M-6:00 A.M          5:00 A.M-6:00 A.M
5 Female 15-20            Klu       No          6:00 A.M-7:00 A.M          6:00 A.M-7:00 A.M
6   Male 15-20  Kl university       No          7:00 A.M-8:00 A.M         9:00 A.M-10:00 A.M
  prefer.as.soon.as.u.wake If.other Breakfast.you.prefer.when.you.are.in.a.Rush If.other.1 skip.your.Brea
kfast
1              Warm water                                            Juice                     Some
times
2                  Coffee                                     Instant Items                     Some
times
3              Warm water                                     Instant Items
Never
4              Warm water                                              Milk
Never
5                    Milk                                              Milk                     Some
times
6                     Tea                                              None                     More
Often
  you.prefer.more.in.Breakfast If.other.2 often..you.have.Breakfast.in.a.Restaurant
1                South Indian                                          Monthly
2                South Indian                                          Monthly
3                South Indian                                          Monthly
4                South Indian                                 Weekly once or twice
5                South Indian                                          Monthly
6                South Indian                                 Weekly once or twice
  often..you.have..Breakfast.in.a.Canteen often.you.have..Breakfast.in.a.Street prefer.after.your.Breakfa
st
1                              Monthly                  Weekly once or twice                      No
ne
2                              Monthly                               Monthly                      No
ne
3                              Monthly                               Monthly                     Cof
fe
4                                Never                                 Never                      No
ne
5                              Monthly                               Monthly                      No
ne
6                   Weekly once or twice                 Weekly once or twice                      No
ne
  like.to.have.in.your.break.time     If.other.3 time.you.have.Lunch kind.of.food.you.prefer.more.in.Lunc
h
1                   Bakery Items                  1:00 P.M-2:00 P.M                          Home Foo
d
2                    Fried Items                  1:00 P.M-2:00 P.M             Canteen Food/Restaurant
s
3                    Packet Food                  1:00 P.M-2:00 P.M                          Home Foo
d
```

```
3                    Vegetarian                         Never             Bakery Items, Tea

4                    Vegetarian                         Never                        Coffee

5                    Vegetarian                    Some times                         Other

6            Both Veg and Non-veg                  Some times                  Instant Snack

  often..you.have.your.Snack time.do.you.have.your.Dinner you.prefer.more.in.Dinner If.other.6
1         For every 3 hours            7:00 P.M-800 P.M                   Tiffin
2         For every 3 hours            7:00 P.M-800 P.M                   Tiffin
3         For every 3 hours            7:00 P.M-800 P.M                     Roti
4                                    9:00 P.M-10:00 P.M                   Tiffin
5         For every 3 hours           8:00 P.M-9:00 P.M                     Rice
6         For every 3 hours           8:00 P.M-9:00 P.M                     Roti
  often..you.skip.your.Dinner you.like.to.have.after.Dinner If.other.7 kind.of.food.do.you.prefer.more
1            Twice in a week                         None                              Healthy Food
2            Twice in a week                    Ice creams                                 Junk Food
3                      Never                         None                              Healthy Food
4                      Never                         None                              Healthy Food
5             Once in a week                         None                              Healthy Food
6                      Never                        Juice                              Healthy Food
> x<-data[,-34]
> y<-data[34]
> model_svm <- svm( kind.of.food.do.you.prefer.more ~ ., data = data)
> summary(model_svm)

Call:
svm(formula = kind.of.food.do.you.prefer.more ~ ., data = data)


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  radial
       cost:  1

Number of Support Vectors:  368

 ( 209 159 )


Number of Classes:  2

Levels:
 Healthy Food Junk Food
```

```
Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food
       561        562        563        564        565        566        567        568
Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food
       569        570        571        572        573        574        575        576
Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food
       577        578        579        580        581        582        583        584
Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food
       585        586        587        588        589        590        591        592
Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food
       593        594        595        596        597        598        599        600
Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food
       601        602        603        604        605        606        607        608
Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food
       609        610        611        612        613        614        615        616
Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food
       617        618        619        620        621        622        623        624
Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food
       625        626        627        628        629        630        631        632
Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food
       633
Healthy Food
Levels: Healthy Food Junk Food
> confusionMatrix(pred,y$kind.of.food.do.you.prefer.more)
Confusion Matrix and Statistics

          Reference
Prediction     Healthy Food Junk Food
  Healthy Food          474       159
  Junk Food               0         0

               Accuracy : 0.7488
                 95% CI : (0.7131, 0.7822)
    No Information Rate : 0.7488
    P-Value [Acc > NIR] : 0.5213

                  Kappa : 0

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 1.0000
            Specificity : 0.0000
         Pos Pred Value : 0.7488
         Neg Pred Value :    NaN
             Prevalence : 0.7488
         Detection Rate : 0.7488
   Detection Prevalence : 1.0000
      Balanced Accuracy : 0.5000

       'Positive' Class : Healthy Food

>
```

The outputs for Naïve-Bayes Classifier are:

```
        55         57         64         65         66         71         72         73
Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food   Junk Food Healthy Food
        75         78         84         85         87         93         94        112
  Junk Food Healthy Food   Junk Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food
       114        115        118        119        123        128        135        139
Healthy Food   Junk Food   Junk Food Healthy Food   Junk Food Healthy Food   Junk Food Healthy Food
       141        142        145        156        160        166        170        173
Healthy Food   Junk Food Healthy Food   Junk Food Healthy Food Healthy Food Healthy Food   Junk Food
       177        183        187        189        193        194        202        217
  Junk Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food
       218        222        226        229        231        233        238        239
Healthy Food Healthy Food   Junk Food   Junk Food Healthy Food Healthy Food   Junk Food   Junk Food
       242        244        245        247        248        249        252        253
Healthy Food Healthy Food   Junk Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food
       254        255        256        257        269        271        276        278
Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food
       279        284        286        287        291        294        297        299
Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food   Junk Food Healthy Food Healthy Food
       300        305        312        320        321        324        325        328
Healthy Food   Junk Food Healthy Food Healthy Food Healthy Food   Junk Food   Junk Food   Junk Food
       329        332        334        335        338        339        348        349
  Junk Food Healthy Food   Junk Food Healthy Food Healthy Food Healthy Food   Junk Food Healthy Food
       353        354        359        361        363        366        367        370
Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food   Junk Food Healthy Food
       374        377        378        391        392        393        394        395
  Junk Food   Junk Food Healthy Food Healthy Food   Junk Food Healthy Food Healthy Food Healthy Food
       402        407        412        413        414        417        419        427
  Junk Food Healthy Food   Junk Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food
       428        429        434        435        438        443        459        460
  Junk Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food
       462        466        467        469        472        479        483        492
  Junk Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food   Junk Food
       498        505        511        512        513        515        518        519
Healthy Food Healthy Food   Junk Food Healthy Food Healthy Food Healthy Food   Junk Food Healthy Food
       523        524        527        530        533        538        540        541
Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food
       542        544        546        548        549        553        558        563
Healthy Food Healthy Food Healthy Food   Junk Food Healthy Food Healthy Food Healthy Food Healthy Food
       567        570        576        581        582        584        585        589
Healthy Food Healthy Food   Junk Food Healthy Food Healthy Food Healthy Food Healthy Food Healthy Food
       591        596        600        606        608        613        615        617
Healthy Food Healthy Food Healthy Food   Junk Food   Junk Food Healthy Food Healthy Food Healthy Food
       619        625        626        628        633
Healthy Food Healthy Food Healthy Food   Junk Food Healthy Food
Levels: Healthy Food Junk Food

$posterior
   Healthy Food    Junk Food
1   0.726985592 2.730144e-01
3   0.996613671 3.386329e-03
7   0.455364158 5.446358e-01
```

```
135   0.178144209  8.218558e-01
139   0.687460752  3.125392e-01
141   0.632393628  3.676064e-01
142   0.002524400  9.974756e-01
145   0.992809770  7.190230e-03
156   0.207219119  7.927809e-01
160   0.734999333  2.650007e-01
166   0.848680901  1.513191e-01
170   0.848970733  1.510293e-01
173   0.141620472  8.583795e-01
177   0.498608383  5.013916e-01
183   0.962840799  3.715920e-02
187   0.999754754  2.452458e-04
189   0.999584258  4.157417e-04
193   0.931635477  6.836452e-02
194   0.955810153  4.418985e-02
202   0.895964647  1.040354e-01
217   0.995955727  4.044273e-03
218   0.824646830  1.753532e-01
222   0.565921470  4.340785e-01
226   0.161642204  8.383578e-01
229   0.095144882  9.048551e-01
231   0.790179174  2.098208e-01
233   0.999772247  2.277533e-04
238   0.468355660  5.316443e-01
239   0.163721742  8.362783e-01
242   0.804239421  1.957606e-01
244   0.855648140  1.443519e-01
245   0.134593930  8.654061e-01
247   0.979291441  2.070856e-02
248   0.999960335  3.966519e-05
249   0.980594488  1.940551e-02
252   0.972182264  2.781774e-02
253   0.575875885  4.241241e-01
254   0.935989331  6.401067e-02
255   0.755545171  2.444548e-01
256   0.935783781  6.421622e-02
257   0.970361741  2.963826e-02
269   0.854741979  1.452580e-01
271   0.999643011  3.569889e-04
276   0.638367754  3.616322e-01
278   0.902272829  9.772717e-02
279   0.816041132  1.839589e-01
284   0.817578497  1.824215e-01
286   0.605023452  3.949765e-01
287   0.999132464  8.675365e-04
291   0.570456938  4.295431e-01
294   0.097249950  9.027500e-01
297   0.708056731  2.919433e-01
299   0.692807376  3.071926e-01
```

```
Prediction     Healthy Food Junk Food
  Healthy Food          120        29
  Junk Food              22        18

               Accuracy : 0.7302
                 95% CI : (0.6609, 0.792)
    No Information Rate : 0.7513
    P-Value [Acc > NIR] : 0.7773

                  Kappa : 0.24

 Mcnemar's Test P-Value : 0.4008

            Sensitivity : 0.8451
            Specificity : 0.3830
         Pos Pred Value : 0.8054
         Neg Pred Value : 0.4500
             Prevalence : 0.7513
         Detection Rate : 0.6349
   Detection Prevalence : 0.7884
      Balanced Accuracy : 0.6140
```

IV TABLE1:Accuracy table

Machine learning algorithms	Accuracy
Support vector machine(SVM)	74.88
Naïve-Bayes classifier	73.02

IV. CONCLUSION

We were able to successfully conclude that our modern generation lives in a state of dilemma. From our analysis, we were able to learn that the vast majority of the people that participated in our survey prefer a healthy lifestyle. However, most people according to their eating habits are actually on an unhealthy path yet quite unaware of it. Although they are leading an unhealthy lifestyle when asked how they overview their day to day habits, they confidently answered that they lead a healthy lifestyle. We believe that it should be taken into our hands to further educate our modern generation on the importance of a healthy lifestyle along with what it truly means. They should be able to identify what's healthy and what's not. However, no matter how we look at the analysis of our data, we were able to conclude from our Data-set that eating healthy and maintaining a healthy lifestyle is preferred by most people

From this, we are also able to find that the accuracy that we obtained from SVM and naïve-Bayes classifier, SVM gives more accuracy when compared to naïve-Bayes classifier.

REFERENCES

1. Abraham S, Noriega Brooke R, Shin JY. College students eating habits and knowledge of nutritional requirements. J Nutr Hum Health. 2018;2(1):13-17
2. J. Aravind, J. Dhalia Sweetlin. Nutrient Facts Analysis using Supervised Learning Approaches. 2017 Conference on Information and Communication Technology (CICT'17)
3. NatnichaSuthumchai ; SirinThongsukh ; PacharamaiYusuksataporn ; Songsri Tangsripairoj; FoodForCare: An Android Application for Self-Care with Healthy Food2016 Fifth ICT International Student Project Conference (ICT-ISPC)
4. Paul D Hatzigiannakoglou. Junk-Food Destroyer: Helping adolescents with Down syndrome to understand healthy eating through serious game; 2015 7th International Conference on Games and Virtual Worlds for Serious Applications (VS-Games).
5. Anila, M. & Pradeepini, G. 2017, "Study of prediction algorithms for selecting appropriate classifier in machine learning", Journal of Advanced Research in Dynamical and Control Systems, vol. 9, no. Special Issue 18, pp. 257-268.
6. Bommadevara, H.S.A., Sowmya, Y. & Pradeepini, G. 2019, "Heart disease prediction using machine learning algorithms", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 5, pp. 270-272.
7. Krishna Mohan, G., Yoshitha, N., Lavanya, M.L.N. & Krishna Priya, A. 2018, "Assessment and analysis of software reliability using machine learning techniques", International Journal of Engineering and Technology(UAE), vol. 7, no. 2.32 Special Issue 32, pp. 201-205.
8. Lakhmi Prasanna, P., Rajeswara Rao, D., Meghana, Y., Maithri, K. & Dhinesh, T. 2018, "Analysis of supervised classification techniques", International Journal of Engineering and Technology(UAE), vol. 7, no. 1.1, pp. 283-285.
9. Lakshmi Prasanna, P. & Rajeswara Rao, D. 2017, "Literature survey on text classification: A review", Journal of Advanced Research in Dynamical and Control Systems, vol. 9, no. Special Issue 12, pp. 2270-2280.
10. Leela Sandhya Rani, Y., Sucharita, V. & Satyanarayana, K.V.V. 2018, "Extensive analysis on generation and consensus mechanisms of clustering ensemble: A survey", International Journal of Electrical and Computer Engineering, vol. 8, no. 4, pp. 2351-2357.
11. Mounica Lahari, M., Sravanthi, N., Sai Kalyan, K., Ravinder, N. & Amudhavel, J. 2017, "Nearest neighbour keyword search using R* algorithm", Journal of Advanced Research in Dynamical and Control Systems, vol. 9, no. Special Issue 12, pp. 1587-1595.
12. Murthy, K.V.S.S.R. & Satyanarayana, K.V.V. 2018, "Intrusion detection mechanism with machine learning process A case study with FMIFSSVM, FLCFSSVM, misuses SVM, anomaly SVM, and Bayesian methods", International Journal of Engineering and Technology(UAE), vol. 7, pp. 277-283.
13. Narasinga Rao, M.R., Sajana, T., Bhavana, N., Sai Ram, M. & Nikhil Krishna, C. 2018, "Prediction of chronic kidney disease using machine learning technique", Journal of Advanced Research in Dynamical and Control Systems, vol. 10, pp. 328-332.
14. Pathak, S.S. & Rajeswara Rao, D. 2018, "Reservoir computing for healthcare analytics", International Journal of Engineering and Technology(UAE), vol. 7, no. 2.32 Special Issue 32, pp. 240-244.
15. Razia, S. & Narasingarao, M.R. 2017, "A neurocomputing framework for thyroid disease diagnosis using machine learning techniques", Journal of Theoretical and Applied Information Technology, vol. 95, no. 9, pp. 1996-2005.
16. Sajana, T. & Narasingarao, M.R. 2018, "Classification of imbalanced malaria disease using naïve Bayesian algorithm", International Journal of Engineering and Technology (UAE), vol. 7, pp. 786-790.
17. Moore PW, Burkhart KK, Jackson D.Drugs highly associated with infusion reactions reported using two different data-mining methodologies.J Blood Disorders Transf.2014; 5-195.
18. Kahraman, C., Cebeci, U. and D. Ruan, 2003, "Multi-attribute comparison of catering
19. Liu, L., Bhattacharyya, S., Sclove, S.L., Chen, R. and W.J. Lattyak, 2001, "Data mining on
20. Wiles NJ, Northstone K, Emmett P, Lewis G, Junk food diet and
21. Childhood behavioral problems: results from the ALSPAC cohort,
22. 2009, 63(1), 491–498.
23. Rang HP, Dale MM, Ritter JM, Moore PK. Pharmacology. 5th Ed.
24. Delhi: Churchill Livingstone; 2006. P.394-400
25. Medical devices: early warning of problems is hampered by severe underreporting. US General Accounting Office. GAO/PEMD 87-1; 1987.

AUTHORS PROFILE

R.RamyaSri Pursuing her B.Tech. from KL University, Guntur, Andhra Pradesh. Her current research interest include Food Analysis in Data Mining. She Published Paper in ICCT'19 IEEE Conference conducted by Manipal University ,Jaipur.

G.Susmitha Pursuing her B.Tech. from KL University, Guntur, Andhra Pradesh. Her current research interest include Food Analysis in Data Mining

K .SaiSurya Pursuing his B.Tech. from KL University, Guntur, Andhra Pradesh. His current research interest include Food Analysis in Data Mining

V.Bhavani is an Associate Professor in the Department of Computer Science, K L University. Her current research interest includes Sentiment Analysis and Opinion Mining, Network Security.

K Venkata Raju Pursuing his Ph.D. from Acharya Nagarjuna University, Guntur, Andhra Pradesh. He is an Associate Professor in the Department of Computer Science, K L University. His current research interest include Sentiment Analysis and Opinion Mining, Network Security.