Ilkogretim Online - Elementary Education Online, 2021; Vol 20 (Issue 6): pp. 3381-3389
http://ilkogretim-online.org
doi: 10.17051/ilkonline.2021.06.319

Implications Of Associated Diseases And Its Factors In Predicting The Obesity Disorder

A.Ramya Research Scholar Department of Computer Science, Vels Institute of Science Technology and Advanced Studies Chennai, India., ramyaimayaa7@gmail.com.

Dr. K.Rohini Associate Professor, Department of Computer Science, Vels Institute of Science Technology and Advanced Studies Chennai, India., rrohini16@gmail.com.

Abstract:

Obesity is caused by energy imbalance. A body mass index is a measurement obtained by dividing a person's weight by the square of the persons height. Obesity is further evaluated in terms of fat distribution around the waist. Obesity is a chronic medical condition that increases your risk of diseases and health problems particularly certain types of cancer, Cardiovascular diseases, Type 2 diabetes, Osteoarthritis, Infertility, Irregular periods, Nonalcoholic fatty liver, Obstructive sleep apnea. Datamining is extracting essential information. Machine learning algorithms applied like Naïve Bayes, K-Neighbors, Extreme Gradient Boost, and Decision Tree classifiers.

Keywords: ML – machine learning, **BMI –** Body mass index, Datamining, Obesity, Naïve Bayes, K-Neighbors, Extreme Gradient Boost, and Decision Tree classifiers.

I. INTRODUCTION:

Obesity affects your quality of life and lead to psychological problems such as depression, low self-esteem, shame, guilt and social isolation. Obesity is more common in women than in men. Obesity is leading preventable cause of death worldwide. Obesity is caused by eating more calories than you burn through exercise and normal daily activity. As a result, your body stores these excess calories as fat. Data mining is the process of sorting through large data sets to identify patterns and relationships that can help solve business problems through data analysis. Data mining techniques and tools enable enterprises to predict future trends and make more-informed business decisions. Data mining is a process used by companies to turn raw data into useful information. By using software to look for patterns in large batches of data, businesses can learn more about their customers to develop more effective marketing strategies, increase sales and decrease costs. Data Mining tools are software programs that help in framing and executing data mining techniques to create data models and test them as well. It is usually a framework like R studio or Tableau with a suite of programs to help build and test a data model.

II. LITERATURE REVIEW:

The paper shows that the age breakdown of video game players is 18-35 years. From a study done, 17% of the male respondents and 11% of the female respondents were below 18 years of age. This is a matter of concern because the mind and the social schema is being formed and built in the young age. Parents want their children to have an imaginative and creative mind. They don't want stereotypical thoughts to cloud the child'smind.Young girls get easily targeted through such vile content. Their perception regarding themselves and how they need to look is being constructed. She starts to believe that she is supposed to be submissive, subordinate and have an hourglass-like figure in order to look beautiful. [1]. The paper discusses about understanding of the disease COVIDE-19 the Researchers are using ML, AI (Artificial Intelligence) and natural language processing. We are using big data analytics to track the spread of this coronavirus and Comparison of Tools and algorithms in Big data Analytics using machine learning Algorithm [2]. This paper discussed about The basis of ensemble learning, weak and strong learners, bagging, boosting and stacking [3]. This paper discusses about 13 diagnosis rules were created and major symptom information was verified. On the basis of this study, other decision tree algorithms will be applied to develop additional model and perform comparison analysis for producing an ideal model from now on [4]. In this paper discussed Diabetes is the most common dangerous disease that can lead to additional problems such as heart attack, stroke, blindness, nerve damage, kidney failure, disease of the blood vessels and sexual disempowerment. The proposed method using K-Means and random forest provides better accuracy for predicting type 2 diabetes. To enhance the proposed method the fuzzy system and deep learning method could also be used. [5]. In this paper discovered The results analysis for the various algorithms used under the various platforms, shows that the performance of the ensemble-based models is the best whatever the used platform but the model analysis and the best algorithm are very sensitive to it. [6]. The paper discussed about to perform complex tasks such as reducing large data features. The framework described in this article is adaptive to new attacks especially because it gets updated using live data captured from the internet, which gives it an edge advantage over other types of IDS that are trained either on data not intended for IoT devices or do not use data that gives better results for new attacks [7]. In this paper discussed about process would be more hectic classification and we cannot predict any accurate results. So we proceed with the social media Data mining Process [8]. The paper proposed to categorize the challenges in this area from a new and more comprehensive perspective. [9]. In this paper discussed about Apache Hadoop, Spark, and Flink. Also, major points of each approach and solution is presented. A conclusion in the end summarizes the points discussed in this paper. [10].

III. DATA SHAPE AND RESULTS ABOUT DATA:

```
     Age  Weight  Length  Sex        BMI  DM  HTN  Current Smoker  EX-Smoker  \
0     53      90     175    1  29.387755   0    1               1          0
1     67      70     157    0  28.398718   0    1               0          0
2     54      54     164    1  20.077335   0    0               1          0
3     66      67     158    0  26.838648   0    1               0          0
4     50      87     153    0  37.165193   0    1               0          0
..   ...     ...     ...  ...        ...  ..  ...             ...        ...
298   58      84     168    1  29.761905   0    0               0          0
299   55      64     152    0  27.700831   0    0               0          0
300   48      77     160    0  30.078125   0    1               0          0
301   57      90     159    0  35.599858   1    0               0          0
302   56      85     170    0  29.411765   0    1               1          0

     FH  ...  Dyspnea  FBS  LDL   HDL  BUN    HB    WBC  Lymph  PLT  Cath
0     0  ...        0   90  155  30.0    8  15.6   5700     39  261     1
1     0  ...        0   80  121  36.0   30  13.9   7700     38  165     1
2     0  ...        0   85   70  45.0   17  13.5   7400     38  230     1
3     0  ...        1   78   55  27.0   30  12.1  13000     18  742     0
4     0  ...        1  104  110  50.0   16  13.2   9200     55  274     0
..   ..  ...      ...  ...  ...   ...  ...   ...    ...    ...  ...   ...
298   0  ...        0   92  115  44.0   13  12.3   8500     34  251     1
299   0  ...        1   86   40  23.0   23  12.4  11400     16  377     0
300   1  ...        0   83  112  42.0   13  12.8   9000     35  279     0
301   0  ...        1   96  130  49.0   14  10.1   3800     48  208     0
302   0  ...        0   78  124  34.0   16  14.7   6000     32  302     1

[303 rows x 33 columns]
```

Fig. 1 Data Shape

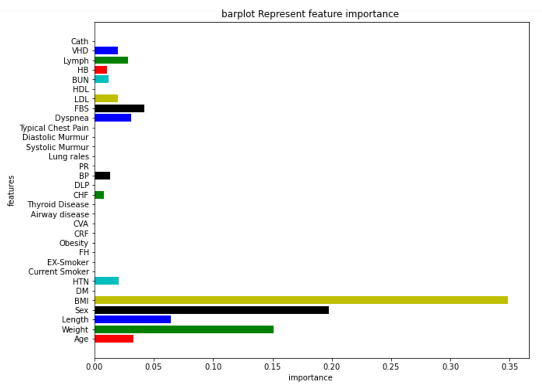

Fig2 Bar plot of Obesity Prediction

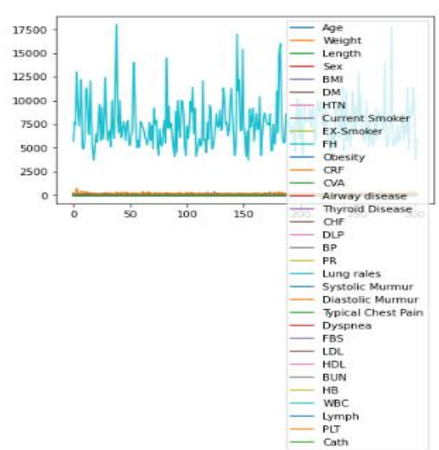

Fig 3 Line chart

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 303 entries, 0 to 302
Data columns (total 33 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   Age                  303 non-null    int64
 1   Weight               303 non-null    int64
 2   Length               303 non-null    int64
 3   Sex                  303 non-null    int64
 4   BMI                  303 non-null    float64
 5   DM                   303 non-null    int64
 6   HTN                  303 non-null    int64
 7   Current Smoker       303 non-null    int64
 8   EX-Smoker            303 non-null    int64
 9   FH                   303 non-null    int64
 10  Obesity              303 non-null    int64
 11  CRF                  303 non-null    int64
 12  CVA                  303 non-null    int64
 13  Airway disease       303 non-null    int64
 14  Thyroid Disease      303 non-null    int64
 15  CHF                  303 non-null    int64
 16  DLP                  303 non-null    int64
 17  BP                   303 non-null    int64
 18  PR                   303 non-null    int64
 19  Lung rales           303 non-null    int64
 20  Systolic Murmur      303 non-null    int64
 21  Diastolic Murmur     303 non-null    int64
 22  Typical Chest Pain   303 non-null    int64
 23  Dyspnea              303 non-null    int64
 24  FBS                  303 non-null    int64
 25  LDL                  303 non-null    int64
 26  HDL                  303 non-null    float64
 27  BUN                  303 non-null    int64
 28  HB                   303 non-null    float64
 29  WBC                  303 non-null    int64
 30  Lymph                303 non-null    int64
 31  PLT                  303 non-null    int64
 32  Cath                 303 non-null    int64
dtypes: float64(3), int64(30)
memory usage: 80.5 KB
```

Fig 4 Data types

Predicted feature impact of Thyroid Disease, Cardio Disease, Airway Disease and Obesity. Fig .5.1 and Fig .5.2 represents Prediction of Obesity. Fig .6.1 and Fig .6.2 represents Prediction of Cardio Disease. Fig 7.1 and Fig .7.2 represents Prediction of Thyroid Disease. Fig 8.1 and Fig .8.2 represents Prediction of Airway Disease.

	Feature	Importance
0	Age	0.033128
1	Weight	0.151174
2	Length	0.064619
3	Sex	0.197489
4	BMI	0.348579
5	DM	0.000000
6	HTN	0.020602
7	Current Smoker	0.000000
8	EX-Smoker	0.000000
9	FH	0.000000
10	Obesity	0.000000
11	CRF	0.000000
12	CVA	0.000000
13	Airway disease	0.000000
14	Thyroid Disease	0.000000
15	CHF	0.007956

Fig .5.1 Impact of features (Predicting Obesity)

	Feature	Importance
16	DLP	0.000000
17	BP	0.013420
18	PR	0.000000
19	Lung rales	0.000000
20	Systolic Murmur	0.000000
21	Diastolic Murmur	0.000000
22	Typical Chest Pain	0.000000
23	Dyspnea	0.030896
24	FBS	0.041927
25	LDL	0.020062
26	HDL	0.000000
27	BUN	0.012211
28	HB	0.010329
29	Lymph	0.028108
30	VHD	0.019499
31	Cath	0.000000

Fig 5.2 Impact of features (Predicting Obesity)

	Feature	Importance
0	Age	0.081593
1	Weight	0.026656
2	Length	0.026951
3	Sex	0.000000
4	BMI	0.032186
5	DM	0.025240
6	HTN	0.044803
7	Current Smoker	0.000000
8	EX-Smoker	0.000000
9	FH	0.040562
10	Obesity	0.017872
11	CRF	0.000000
12	CVA	0.000000
13	Airway disease	0.000000
14	Thyroid Disease	0.000000
15	CHF	0.000000
16	DLP	0.049184

Fig.6.1 Impact of features (Predicting Cardio Disease)

	Feature	Importance
17	BP	0.052824
18	PR	0.034396
19	Lung rales	0.000000
20	Systolic Murmur	0.031674
21	Diastolic Murmur	0.000000
22	Typical Chest Pain	0.224363
23	Dyspnea	0.013063
24	FBS	0.049042
25	LDL	0.035659
26	HDL	0.030671
27	BUN	0.033891
28	HB	0.039142
29	Lymph	0.030752
30	VHD	0.044705
31	Cath	0.034772

Fig 6.2Impact of features (Predicting c)

	Feature	Importance
0	Age	0.000000
1	Weight	0.185874
2	Length	0.000000
3	Sex	0.000000
4	BMI	0.814126
5	DM	0.000000
6	HTN	0.000000
7	Current Smoker	0.000000
8	EX-Smoker	0.000000
9	FH	0.000000
10	Obesity	0.000000
11	CRF	0.000000
12	CVA	0.000000
13	Airway disease	0.000000
14	Thyroid Disease	0.000000
15	CHF	0.000000
16	DLP	0.000000

Fig 7.1 Impact of features (Predicting Thyroid Disease)

c	Feature	Importance
17	BP	0.000000
18	PR	0.000000
19	Lung rales	0.000000
20	Systolic Murmur	0.000000
21	Diastolic Murmur	0.000000
22	Typical Chest Pain	0.000000
23	Dyspnea	0.000000
24	FBS	0.000000
25	LDL	0.000000
26	HDL	0.000000
27	BUN	0.000000
28	HB	0.000000
29	Lymph	0.000000
30	VHD	0.000000
31	Cath	0.000000

Fig 7.2. Impact of features (Predicting Thyroid Disease)

	Feature	Importance
0	Age	0.0
1	Weight	0.0
2	Length	0.0
3	Sex	0.0
4	BMI	0.0
5	DM	0.0
6	HTN	0.0
7	Current Smoker	0.0
8	EX-Smoker	0.0
9	FH	0.0
10	Obesity	0.0
11	CRF	0.0
12	CVA	0.0
13	Airway disease	0.0
14	Thyroid Disease	0.0
15	CHF	0.0
16	DLP	0.0

Fig 8.1. Impact of features (Predicting Airway Disease)

	Feature	Importance
17	BP	0.0
18	PR	0.0
19	Lung rales	0.0
20	Systolic Murmur	0.0
21	Diastolic Murmur	0.0
22	Typical Chest Pain	0.0
23	Dyspnea	1.0
24	FBS	0.0
25	LDL	0.0
26	HDL	0.0
27	BUN	0.0
28	HB	0.0
29	Lymph	0.0
30	VHD	0.0
31	Cath	0.0

Fig 8.2. Impact of features (Predicting Airway Disease)

IV. CONCLUSION

In this Paper Machine learning algorithms applied like Naïve Bayes, K-Neighbors, Extreme Gradient Boost, and Decision Tree classifiers. Predicted feature impact of Thyroid Disease, Airway Disease, Airway Disease and Obesity.

REFERENCES:

1. R. Ansari, B. A. Jaffar, S. Riaz, M. J. Kaur and A. Mushtaq, "Datamining to Alert the Formation of Women Objectification Stereotypes in Video Games," 2019 Amity International Conference on Artificial Intelligence (AICAI), 2019, pp. 521-526, doi: 10.1109/AICAI.2019.8701256.
2. D. R. Krithika and K. Rohini, "Comparative Intrepretation Of Machine Learning Algorithms In Predicting The Cardiovascular Death Rate For Covid-19 Data," 2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), 2021, pp. 394-400, doi: 10.1109/ICCIKE51210.2021.9410777.
3. D. R. Krithika and K. Rohini, "Ensemble Based Prediction of Cardiovascular Disease Using Bigdata analytics," 2021 International Conference on Computing Sciences (ICCS), 2021, pp. 42-46, doi: 10.1109/ICCS54944.2021.00017.
4. J. -K. Choi, C. I. Yoo, K. -A. Kim, Y. Won and J. -J. Kim, "Study on Datamining Techinique for Foot Disease Prediction," 2014 International Conference on IT Convergence and Security (ICITCS), 2014, pp. 1-4, doi: 10.1109/ICITCS.2014.7021816.
5. M. S. Geetha Devasena, R. Kingsy Grace and G. Gopu, "PDD: Predictive Diabetes Diagnosis using Datamining Algorithms," 2020 International Conference on Computer Communication and Informatics (ICCCI), 2020, pp. 1-4, doi: 10.1109/ICCCI48352.2020.9104108.
6. D. Bari, M. Ameksa and A. Ouagabi, "A comparison of datamining tools for geo-spatial estimation of visibility from AROME-Morocco model outputs in regression framework," 2020 IEEE International conference of Moroccan Geomatics (Morgeo), 2020, pp. 1-7, doi: 10.1109/Morgeo49228.2020.9121909.
7. A. E. Ghazi and A. Moulay Rachid, "Machine learning and datamining methods for hybrid IoT intrusion detection," 2020 5th International Conference on Cloud Computing and Artificial Intelligence: Technologies and Applications (CloudTech), 2020, pp. 1-6, doi: 10.1109/CloudTech49835.2020.9365895.
8. S. Umamaheswari and K. Harikumar, "Analyzing Product Usage Based on Twitter Users Based on Datamining Process," 2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM), 2020, pp. 426-430, doi: 10.1109/ICCAKM46823.2020.9051488.
9. N. Shafiei and M. R. Keyvanpour, "Challenges Classification in Search-Based Refactoring," 2020 6th International Conference on Web Research (ICWR), 2020, pp. 106-112, doi: 10.1109/ICWR49608.2020.9122271.
10. M. J. Shayegan Fard and P. A. Namin, "Review of Apriori based Frequent Itemset Mining Solutions on Big Data," 2020 6th International Conference on Web Research (ICWR), 2020, pp. 157-164, doi: 10.1109/ICWR49608.2020.9122295.