Decision Support for Recommendation system at Different Levels of Obesity Classification

A.Ramya,

Research Scholar,
Department of Computer Science,
Vels Institute of Science Technology and Advanced Studies
Chennai, India,

Dr.K.Rohini

Associate Professor,
Department of Computer Science,
Vels Institute of Science Technology and Advanced studies,
Chennai, India,

ramyaimayaa7@gmail.com.rrohini16@gmail.com.

Abstract:

In recent days we are facing various health problems and getting affected to diseases due to environmental conditions and food habits. Earlier stage disease prediction helps the medical people to take quick actions to rectify the complications of getting into worse condition. Datamining is one of the technologies that provides decision support in medical prognosis. Datamining processes the data from patient datasets and yield new knowledgeable information out of it. This research work deals with Body Mass Index, which can be used to identify the intensity of obesity at its various levels. Many supervised machine learning algorithms such as Naïve Bayes, K-Neighbors, Extreme Gradient Boost, and Decision Tree classifiers. Effective implementation of the Algorithms with the Obesity dataset is done in Python. About 6000 patient data are applied to these algorithms and comparative charts of their accuracies are obtained. The proposed Modified Decision Tree Algorithm gives better Accuracy Result compared with the existing ones.

Keywords: ML – machine learning, **BMI –** Body mass index, Datamining, Obesity.

DOI Number:10.14704/nq.2022.20.8.NQ44428 **NeuroQuantology 2022; 20(8): 3974-3984**

I.INTRODUCTION:

Data presented in multiple formats is data acquisition. These raw data are stored and preprocessed to finally get the target data. Datamining uncovers hidden patterns and relationships in data. Processing of datamining involves using software algorithms. Datamining is used in various fields like marketing. Risk management, Fraud detection, Cybersecurity, Medical diagnosis, Mathematics, Cybernetics, and Genetics. In Datamining different types of Machine learning algorithms like classification, regression, and clustering are applied. Data mining is sometimes used interchangeably in data analytics. Various Machine Learning algorithms are applied to enhance the task of datamining and decision support in medical diagnosis.

Supervised type of machine learning is used to classify and predict medical datasets with good accuracy. Here, in this work, the Obesity dataset is used to classify the patients, based on Body

Mass Index, and provide a recommendation result using python.

Obesity is excess fat; it is increased mortality. Body mass index is the accepted standard measure of obesity. The Body Mass Index is represented by kilograms per meter squared. Obesity is associated with a 5-fold higher mortality rate than people who are not obese. There was a study from Wuhan that showed that 88.2% of people who died with covid-19 had a Body Mass Index that was greater than 25, and in a French study, the risk of mechanical ventilation was 7 times higher in obese individuals. So we know that obesity itself seems to be associated with worse outcomes. Obesity is associated with lower-level inflammation and one of the consequences of covid-19 infection is you get. People with obesity have reduced cardiorespiratory fitness and diminished respiratory compliance.

II. LITERATURE REVIEW:

The paper proposed about disease prediction using ANN and KNN algorithms [1]. The paper discusses about the Covid 19 world data to predicting futured cases using bigdata analytics [2]. This paper discussed about Neural network, KNN, Decision trees, naïve bayes algorithms used to predicted patient risk level [3]. This paper discusses about clustering algorithm used to predicted diabetes, heart disease, asthma initial stages [4]. In this paper discussed medical decision support using CBR cycle [5]. In this paper discovered Heart disease prediction based on CNN, KNN and Naïve Bayes Algorithms [6]. The paper discussed about Naïve Bayesian classification method used to predicted heart disease [7]. In this paper discussed about similar type of multiple disease prediction [8]. The paper proposed to heart disease prediction in SVM, ANN, Naïve bayes Hybrid Naïve bayes algorithms [9]. In this paper discussed about hybrid genetic algorithm and k means used to predicted early stage heart disease [10]. In this paper analyzed disease gene prediction [11]. The paper proposed about comparison of grey model, ARIMA and neural network model [12]. This paper discusses modified algorithm BoostAaRoota [13]. The paper discussed about Logistic, random forest and K-nearest Neibhour Algorithms [14]. In this paper proposed about using machine learning algorithm to prediction and treatment recommendation system [15]. The paper discussed about Machine learning algorithm using to predict cardiovascular disease [16]. In this paper analyzed Novel method used to determined disease-pathway [17].

III. DATA SHAPE AND ABOUT DATA:

	age	gender	height	weight	systolic	diastolic	cholesterol	gluc	\
0	37	0	145	74	110	70	1	2	
1	80	1	157	62	130	60	1	1	
2	78	1	162	59	120	70	1	2	
3	52	0	146	55	140	82	1	1	
4	60	1	162	64	110	70	3	1	
...	...	...	...	...	...	...	...	...	
5995	61	1	163	60	110	80	3	2	
5996	35	0	160	53	94	57	3	2	
5997	40	0	151	67	140	95	3	1	
5998	68	1	160	60	120	70	1	1	
5999	47	1	160	106	130	86	1	1	

	smoke	alco	bmi	Obesity	BMILEVEL
0	0	0	35.20	1	4
1	0	0	25.15	0	2
2	0	0	22.48	0	1
3	0	0	25.80	0	2
4	0	0	24.39	0	1
...	...	...	...	...	...
5995	0	0	22.58	0	1
5996	0	0	20.70	0	1
5997	0	0	29.38	0	2
5998	0	0	23.44	0	1
5999	0	0	41.41	1	5

[6000 rows x 13 columns]

Fig 1 Shape of Data

*e*ISSN1303-5150

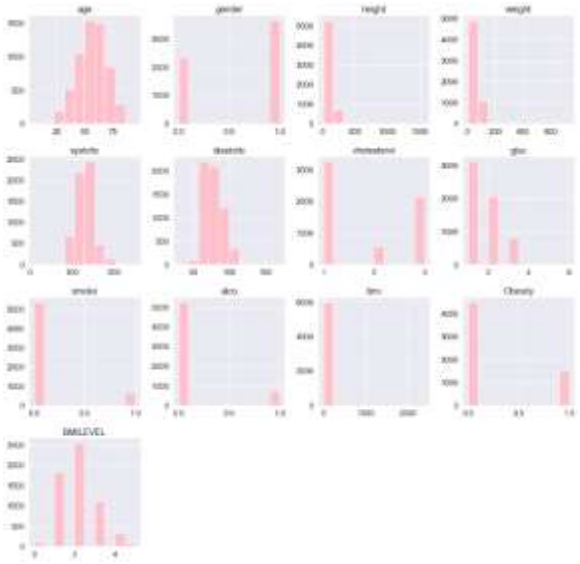

Fig 3 Data Histogram

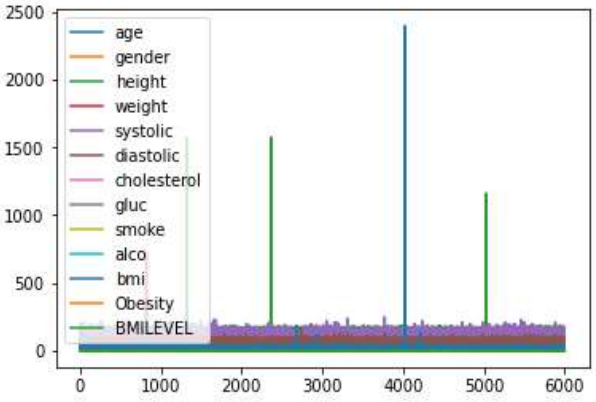

Fig 4 Line Chart of data

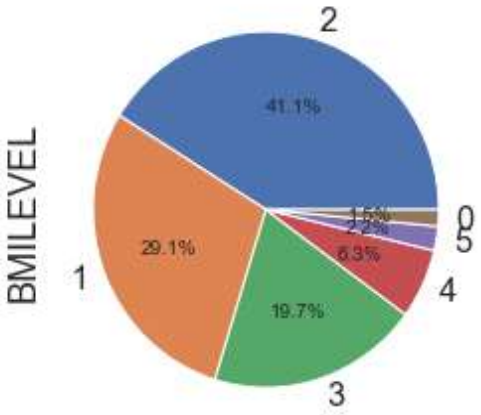

Fig 5 BMI Level Pie chart

```
55-64              0.341167
45-54              0.309000
35-44              0.152333
65-74              0.111167
25-34              0.039667
75-84              0.034667
19-24              0.007500
85+                0.002667
Less than 18       0.001833
Name: Age, dtype: float64
```

Fig 6 Age values normalized

III. ALGORITHM:

The paper proposed algorithm is an enhanced modified decision tree classifier which is used to classify the obesity data set of 6000 data.Thisdataset is applied to few machine learning algorithms are Naïve Bayes, K-Neighbors, XGB, Decision tree classifier and compared with the proposed algorithm. Decision tree algorithm is supervised category and classification algorithm. 6000 Patients data applied in these algorithms. Modified Decision Tree Algorithm gives better Accuracy Result.

```
confussion matrix
[[ 15   6   0   0   0   0]
 [  4 319  22   0   0   0]
 [  0  28 483   0   0   0]
 [  0   0   0 233   3   0]
 [  0   0   0   2  68   0]
 [  0   0   0   0  12   5]]

Train Accuracy of  Naive Bayes model: 94.48

Test Accuracy of Naive Bayes model: 93.58333333333333

              precision    recall  f1-score   support

           0       0.79      0.71      0.75        21
           1       0.90      0.92      0.91       345
           2       0.96      0.95      0.95       511
           3       0.99      0.99      0.99       236
           4       0.82      0.97      0.89        70
           5       1.00      0.29      0.45        17

    accuracy                           0.94      1200
   macro avg       0.91      0.81      0.82      1200
weighted avg       0.94      0.94      0.93      1200
```

Fig 7 Accuracy of Naïve Bayes Algorithm

```
confussion matrix
[[  0  21   0   0   0   0]
 [  0 343   2   0   0   0]
 [  0   0 510   1   0   0]
 [  0   0   6 230   0   0]
 [  0   0   2   0  68   0]
 [  0   0   0   0   0  17]]


Train Accuracy of  Extreme Gradient Boost: 97.56

Test Accuracy of Extreme Gradient Boost: 97.33333333333334
```

	precision	recall	f1-score	support
0	0.00	0.00	0.00	21
1	0.94	0.99	0.97	345
2	0.98	1.00	0.99	511
3	1.00	0.97	0.99	236
4	1.00	0.97	0.99	70
5	1.00	1.00	1.00	17
accuracy			0.97	1200
macro avg	0.82	0.82	0.82	1200
weighted avg	0.96	0.97	0.96	1200

Fig 8 Accuracy of XGB

```
confussion matrix
[[  3  18   0   0   0   0]
 [  0 275  70   0   0   0]
 [  0  74 437   0   0   0]
 [  0   0   1 233   2   0]
 [  0   0   0  56  13   1]
 [  0   0   0   2  11   4]]


Train Accuracy of  K-NeighborsClassifier: 86.33

Test Accuracy of K-NeighborsClassifier: 80.41666666666667
```

	precision	recall	f1-score	support
0	1.00	0.14	0.25	21
1	0.75	0.80	0.77	345
2	0.86	0.86	0.86	511
3	0.80	0.99	0.88	236
4	0.50	0.19	0.27	70
5	0.80	0.24	0.36	17
accuracy			0.80	1200
macro avg	0.79	0.53	0.57	1200
weighted avg	0.80	0.80	0.79	1200

Fig 9 Accuracy of K-Neighbors

*e*ISSN1303-5150

```
confussion matrix
[[ 21   0   0   0   0   0]
 [  0 343   2   0   0   0]
 [  0   0 511   0   0   0]
 [  0   0   0 236   0   0]
 [  0   0   0   0  70   0]
 [  0   0   0   0  17   0]]


Train Accuracy of  DecisionTreeClassifier: 98.46

Test Accuracy of DecisionTreeClassifier: 98.41666666666666

              precision    recall  f1-score   support

           0       1.00      1.00      1.00        21
           1       1.00      0.99      1.00       345
           2       1.00      1.00      1.00       511
           3       1.00      1.00      1.00       236
           4       0.80      1.00      0.89        70
           5       0.00      0.00      0.00        17

    accuracy                           0.98      1200
   macro avg       0.80      0.83      0.81      1200
weighted avg       0.97      0.98      0.98      1200
```

Fig 10 Accuracy of Decision Tree Classifier

```
confussion matrix
[[ 21   0   0   0   0   0]
 [  0 343   2   0   0   0]
 [  0   0 511   0   0   0]
 [  0   0   0 236   0   0]
 [  0   0   0   0  70   0]
 [  0   0   0   0   0  17]]


Train Accuracy of  ModifiedDecisionTreeClassifier: 100.0

Test Accuracy of ModifiedDecisionTreeClassifier: 99.83333333333333

              precision    recall  f1-score   support

           0       1.00      1.00      1.00        21
           1       1.00      0.99      1.00       345
           2       1.00      1.00      1.00       511
           3       0.73      1.00      0.84       236
           4       0.00      0.00      0.00        70
           5       0.00      0.00      0.00        17

    accuracy                           0.93      1200
   macro avg       0.62      0.67      0.64      1200
weighted avg       0.87      0.93      0.90      1200
```

Fig 11 Modified Decision Tree Classifier

	Model	Train Accuracy	Test Accuracy
0	Naive Bayes	94.48	93.583333
1	K-Nearest Neighbour	86.33	80.416667
2	XGB	97.56	97.333333
3	DecisionTreeClassifier	98.46	98.416667
4	ModifiedDecisionTreeClassifier	100.00	99.833333

Fig 12 Algorithm Results

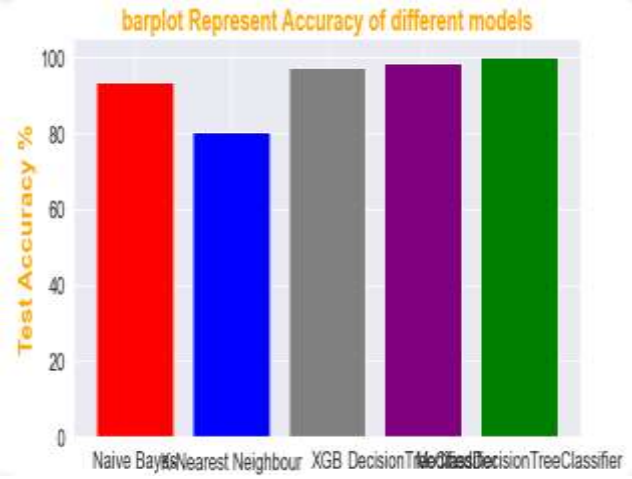

Fig 13 Algorithm Different models in Bar plot

IV. RECOMMENATION SYSTEM:

Based on obesity data recommended chances of diseases using Recommendation system. Body Mass Index levels below 18.5 are insufficient weight. 18.5 to 24.9 is Normal Weight. 25 to 29.9 is Overweight. 30 to 34.9 is Obesity Level 1. 35 to 39.9 is obesity level 2, and above 40 is obesity level. These obesity levels are categorized by values 0,1,2,3,4,5.The below table shows the number of patients at differentBody Mass Index levels based upon 6000 patient data. For this recommendation system delving each level of body mass index gave the result of chances to get list of diseases.

BMI Value	No of Patients	BMI Levels
0	90	Insufficient
1	1746	Normal
2	2468	Over
3	1184	Obesity level1
4	380	Obesity level2
5	132	Obesity level3

Table 1. Number of patients in BMI Levels

	Insufficient weight	Chances of diseases
0	0	Crohn's disease
1	0	Cancer
2	0	Arthritis
3	0	Anemia
4	0	Heart Attack
5	0	Infertility
6	0	Psychiatric disorder
7	0	cognitive behavioral therapy
8	0	postnatal depression

Fig 14 Insufficient Weight and Chances of disease

	normal weight	Chances of diseases
0	1	Less chance to get disease

Fig 15 Normal Weight and Chances of disease

	Over weight	Chances of diseases
0	2	waist circumference
1	2	Abdominal cavity
2	2	Sleep Apnea
3	2	Non-Alcoholic fatty liver
4	2	Symptomatic osteo arthritis
5	2	coronary artery disease
6	2	Peripheral atherosclerosis disease
7	2	Pseudo Hypo para thyroid
8	2	Genetic induced disorder
9	2	Low HDL cholesterol

Fig 16 Overweight and Chances of disease

	Obesity level 1	Chances of diseases
0	3	Metabolic Abnormalities
1	3	Cardiovascular disease
2	3	Hypertension
3	3	Infertility
4	3	Type2 diabetes
5	3	Non-Alcoholic fatty liver
6	3	Irregular periods

Fig 17 Obesity level1 and Chances of disease

	Obesity level 2	Chances of diseases
0	4	Lymphopenia
1	4	Osteo arteritis
2	4	Hypertension
3	4	High LDL Cholesterol
4	4	Low HDL Cholesterol
5	4	Diabetes
6	4	Cancer
7	4	Digestive Problem

Fig 18 Obesity level2 and Chances of disease

	Obesity level 3	Chances of diseases
0	5	Polycystic ovary Syndrome
1	5	Prader Willi syndrome
2	5	Cushing Syndrome
3	5	Stroke
4	5	Gall Bladder Disease
5	5	Heart disease

Fig 19 Obesity level3 and Chances of disease

IV. CONCLUSION

The paper proposed modified decision tree classifier and 6000 Data is applied to these algorithms Naïve Bayes, K-Neighbors, Extreme Gradient Boost, Decision tree classifier modified decision tree classifier. Proposed Algorithm got better Accuracy. Based on Body Mass Index Levels are target value. The paper presents a disease recommendation system based on Body Mass Index. In future work Based upon Body Mass Index predicting Diseases.

REFERENCES:

1. D. Dahiwade, G. Patle and E. Meshram, "Designing Disease Prediction Model Using Machine Learning Approach," *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, 2019, pp. 1211-1215, doi: 10.1109/ICCMC.2019.8819782.
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. J. Thomas and R. T. Princy, "Human heart disease prediction system using data mining techniques," *2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT)*, 2016, pp. 1-5, doi: 10.1109/ICCPCT.2016.7530265.
4. 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.
5. F. Saadi, B. Atmani and F. Henni, "Integration of datamining techniques into the CBR cycle to predict the result of immunotherapy treatment," *2019 International Conference on Computer and Information Sciences (ICCIS)*, 2019, pp. 1-5, doi: 10.1109/ICCISci.2019.8716415.
6. S. Ambekar and R. Phalnikar, "Disease Risk Prediction by Using Convolutional Neural Network," *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, 2018, pp. 1-5, doi: 10.1109/ICCUBEA.2018.8697423.
7. A. N. Repaka, S. D. Ravikanti and R. G. Franklin, "Design And Implementing Heart Disease Prediction Using Naives Bayesian," *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, 2019, pp. 292-297, doi: 10.1109/ICOEI.2019.8862604.
8. J. Gao, L. Tian, J. Wang, Y. Chen, B. Song and X. Hu, "Similar Disease Prediction With Heterogeneous Disease Information Networks," in *IEEE Transactions on NanoBioscience*, vol. 19, no. 3, pp. 571-578, July 2020, doi: 10.1109/TNB.2020.2994983.
9. M. J. A. Junaid and R. Kumar, "Data Science And Its Application In Heart Disease Prediction," *2020 International Conference on Intelligent Engineering and Management (ICIEM)*, 2020, pp. 396-400, doi: 10.1109/ICIEM48762.2020.9160056.
10. M. T. Islam, S. R. Rafa and M. G. Kibria, "Early Prediction of Heart Disease Using PCA and Hybrid Genetic Algorithm with k-Means," *2020 23rd International Conference on Computer and Information Technology (ICCIT)*, 2020, pp. 1-6, doi: 10.1109/ICCIT51783.2020.9392655.
11. K. Yang *et al*., "HerGePred: Heterogeneous Network Embedding Representation for Disease Gene Prediction," in *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 4, pp. 1805-1815, July 2019, doi: 10.1109/JBHI.2018.2870728.
12. Y. Zhao and Y. Su, "Comparison of Three Prediction Models for the Incidence of Epidemic Diseases," *2020 International Conference on Communications, Information System and Computer Engineering (CISCE)*, 2020, pp. 131-136, doi: 10.1109/CISCE50729.2020.00033.
13. A. P and V. Kalyani David, "Feature selection using ModifiedBoostARoota and prediction of heart diseases using Gradient Boosting algorithms," *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, 2021, pp. 19-23, doi: 10.1109/ICCCIS51004.2021.9397154.
14. N. Rubaiyat, A. I. Apsara, A. Al Farabe and I. Ishtiak, "Classification and prediction of

*e*ISSN1303-5150

Orthopedic disease based on lumber and pelvic state of patients," *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, 2019, pp. 1-4, doi: 10.1109/ICECCT.2019.8869540.

15. R. B. Mathew, S. Varghese, S. E. Joy and S. S. Alex, "Chatbot for Disease Prediction and Treatment Recommendation using Machine Learning," *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, 2019, pp. 851-856, doi: 10.1109/ICOEI.2019.8862707.
16. Krithika, D. R., and K. Rohini. "Bigdata Analysis Using Machine Learning Algorithm in Predicting the cardiovascular disease." Intelligent Computing and Innovation on Data Science. Springer, Singapore, 2021. 177-187.
17. A. Ghulam, X. Lei, M. Guo and C. Bian, "Disease-Pathway Association Prediction Based on Random Walks With Restart and PageRank," in *IEEE Access*, vol. 8, pp. 72021-72038, 2020, doi: 10.1109/ACCESS.2020.2987071.