Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

Smart City Citizen Diabetes Predictions using Machine Learning

Dr. Nisha Sharma*
D.A.V. College, Pehowa, Indianishajan6@gmail.com

Dr. Prakash Kumar Udupi
Middle East College, Muscat, Omanprakash@mec.edu.com

Article Info
Page Number: 1557-1563
Publication Issue:
Vol 71 No. 3s2 (2022)

Abstract
Health and wellness is one of the most essential need for creating smart city wellbeing environment and promoting quality of life of smart city citizens. Growing urbanisation, aging population and rapid industrialization demands strategic implementation of healthcare measures for the smart city populations[1]. Deployment of heath parameter observatory measures is one of the most common initiatives, that can be easily incorporated within the smart city healthcare system. This research explores diabetics as a basis and evaluation of its associated parameter such as blood sugar, hemoglobin, body mass index for understanding the relationship between health data, poor health parameters and quality of life of smart city citizens [2]. Information and communication technology can play a significant role in this efforts. Predictive analytics using machine learning is one of such derivatives of information and communication technology. Hence, machine learning can facilitates prediction of associated healthcare parameters, there by enables decision making on health and wellness of smart city citizen easier [3]. This research also identifies, evaluates and explore various machine learning techniques for predicting health related issues and deceases. The research paper further highlight future scopes research with reference to decisions on healthcare and well being of smart city citizens.

Article History
Article Received*: 28 April 2022*
Revised*: 15 May 2022*
Accepted*: 20 June 2022*
Publication*: 21 July 2022*

Keywords: - Smart city, information and communication technology, smart city health care system, machine learning, decision making.

I. Introduction

Smart city enables extensive usage of information and communication technologies, seamless integration of data transmissions between devices and shares information across government, private and public departments. This leads to overall efficiency of operations of the public administration in smart cities. Smart city infrastructures are also integrated with technologies such as IOT (internet of Things), smart networks, smart sensors, connectivity, software and user interfaces [4]. Wearable devices can monitor smart city citizens various health related parameters and continuous transmission of these health data between devices, hospitals, administration departments facilitates continuous monitoring of these health parameters[5]. Smart city health and wellbeing is one of the key factors needs to be considered while drawing smart city strategies and during the time of smart city developments. Growing populations, rapid urbanisations of smart city citizens and industrialization made the smart city populations to expose themselves to rapid changes in the

surrounding living environments and urban life styles. Exposures to other factors such as changes in food habits, improper dietary patters, stressful life styles, less physical activates enables smart city citizens to develop complex issues related to stress, obesity, over weight and diabetes. This in-turn leads to sever health problems related to heart deceases, strokes and few categories of cancers. Diabetes also drives the citizen to poorer quality of life and reduced mental health. Hence addressing the diabetic issues as a part of health and well being can help to improve the quality of life of smart city populations. Technology enabled services available in smart cities ensures timely availability of accurate health data and these data can be further synthesized, filtered and explored for various features to understand health issues relatedto smart city citizens.

Machine learning is a branch of artificial intelligence, facilitates prediction of results based on learning happened by the computer using input data and results[6]. Computer uses various algorithms to learn the data and results, which are provided to computers as an input. Using algorithms, computers learns the patterns available in the data, understand the relationship between these data and results. The learning done by the computers further applied on the new dataset to predict the results or outcomes of new dataset. Machine learning can be applied to health dataset to predict various health parameters of the smart city citizens [7]. Hence, prediction of health parameters helps decision makers to take relevant decisions related to health and well beings of smart citypopulations.

II. Proposed machine learning framework

Basic health parameters such as body mass index, glucose, insulin, blood pressure etc. in the form of data are normally generated through measurements. Capturing of these basic heath related parameters are normally done in the hospitals through physical measurements, electronic devices or wearable devices. Captured data are passed through various machine learning algorithms. Three most common machine learning techniquesused are classification, clustering and regression.

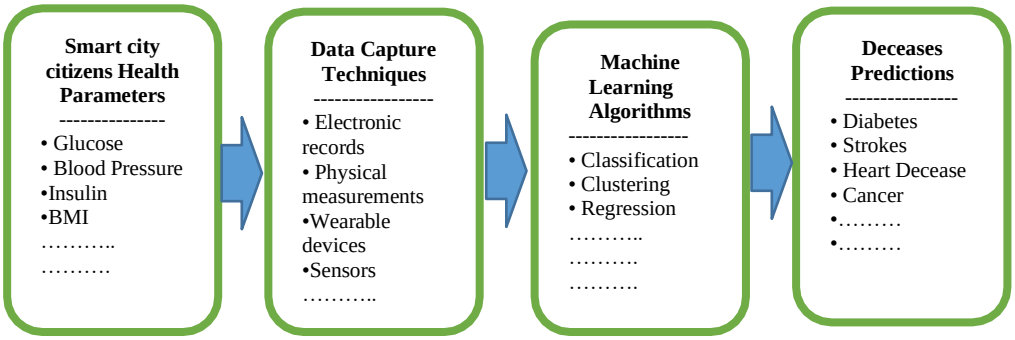

Figure 1. Proposed machine learning framework

Regression is a supervised machine learning techniques, that facilitate the establishment of relationship between variables and demonstrates how one variable effects the other variables [8]. Regression techniques in machine learning provides continuous output as predictions for various values of predictor inputs. The outcome of regression helps to predict whether diabetes is present among the smart city citizens or not. Hence, decision making on improving health related issues and

relevant measurescan be initiated to benefits the smart city populations.

Classification is one of the machine learning techniques facilitates class predictions of for the given input data. Labelled data are the input to the classification techniques [9]. Hence classification also falls under supervised machine learning methods. The outcome of the classification machine learning techniques helps to predict various classes under diabetics. Classification of smart city citizens into various groups based on health parameters can help to make decisions on improving the health and well being of smart city citizens.

Clustering is another machine learning techniques uses unsupervised methods. Clustering uses unlabeled data as inputs and facilitates predictions of clusters by identifying relevant patterns among the input data[10]. Clustering predicts the formations of clusters or groups based on similar patters observed in the health dataset.

III. Data simulation and result analysis

Machine learning techniques uses data and results as inputs to the computers. Machine learns the input data along with results using various algorithms. For the simulation purposes and testing the frameworks, data were assumed and random data is generated by following the diabetes parameters and the machine learning rules. Table 1., demonstrates the diabetic dataset, which are further classified as diabetics, prediabetes and nondiabetics based on four parameters such as random blood sugar, blood sugar after fasting, BMI (Body Mass Index) and hemoglobin. Diabetic column data is categorical data. Hence, diabetics column data also made available as numeric data in the category column, which is used during the time of regression analysis.

Sl No	Random_BS	Fast_BS	BMI	Hemoglobin	Category	Diabetics
1	130	121	26	6.1	2	Prediabetes
2	82	92	23	5.3	1	Nondiabetic
3	110	85	24	5.2	1	Nondiabetic
4	102	97	21	4.4	1	Nondiabetic
5	125	124	26	5.9	2	Prediabetes
6	185	225	44	7.5	3	Diabetes
…..	…..	…..	…..	…..	…..	…..
…..	…..	…..	…..	…..	…..	…..
…..	…..	…..	…..	…..	…..	…..
996	126	121	25	6	2	Prediabetes
997	205	135	29	7.8	3	Diabetes
998	66	73	24	4.9	1	Nondiabetic
999	75	79	19	4	1	Nondiabetic
1000	152	164	32	7.6	3	Diabetes

Table 1. Random health parameters data set

Regression techniques comprises of various algorithms including Boosting algorithm, K-Nearest Neigbor algorithm and Linear Regression algorithm. For the purpose of sample analysis, Boosting regression algorithm is applied and following results captured.

Boosting Regression							
Trees	Shrinkage	Loss function	n(Train)	n(Validation)	n(Test)	Validation MSE	Test MSE
64	0.100	Gaussian	640	160	200	5.290e-5	1.152e-4

Table 2. Results of application of Boosting regression algorithm

From table 2., we observe that, 640 rows of data were used by the computer for training purposes, 160 rows of data used for validating the learning and 200 rows of data used for testing the learning. The model showcased very low MSE, which means excellent accuracy for predictions using regression techniques can be achieved.

Evaluation Metrics

	Value
MSE	0
RMSE	0
MAE	0.003
MAPE	0.68%
R^2	1

Relative Influence Plot

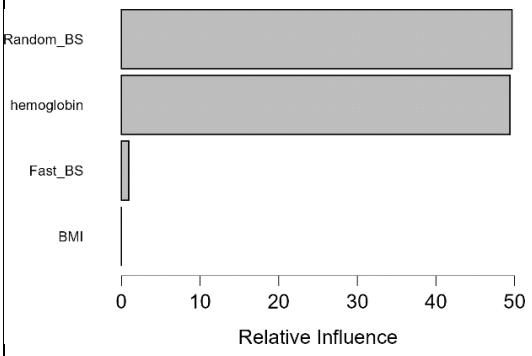

Table 3. Boosting regression algorithm evaluation metrics

From table 3., we observe that, Mean Square Error and Root Mean Square Error is minimum or equal to zero. Minimum MSE or RMSE means, excellent predictions can be achieved. Further R^2 is equal to 1 means, 100% linear relationship possible with the predictor variables. Relative influence plot further demonstrates that random blood sugar and hemoglobin data contributes maximum for predictions of the outcomes.

Classification techniques can be implemented using many algorithms including Linear Discriminant Analysis, Random Forest and K- Nearest Neighbor algorithm. As a part of sample analysis, Boosting classification algorithm is applied and following results captured.

Boosting Classification						
Trees	Shrinkage	n(Train)	n(Validation)	n(Test)	Validation Accuracy	Test Accuracy
100	0.100	640	160	200	1.000	1.000

Table 4. Results of application of Boosting classification algorithm

From table 4., we observe that, 640 rows of data were used by the computer for training purposes, 160 rows of data used for validating the learning and 200 rows of data used for testing the learning.

The model showcased test accuracy as 1, which means 100% accuracy can be achieved for predicting the outcomes using Boosting classification algorithm, for the given input data.

Confusion Matrix				
		Predicted		
		Diabetes	Non- diabetic	Prediabetes
Observed	Diabetes	70	0	0
	Nondiabetic	0	93	0
	Prediabetes	0	0	37

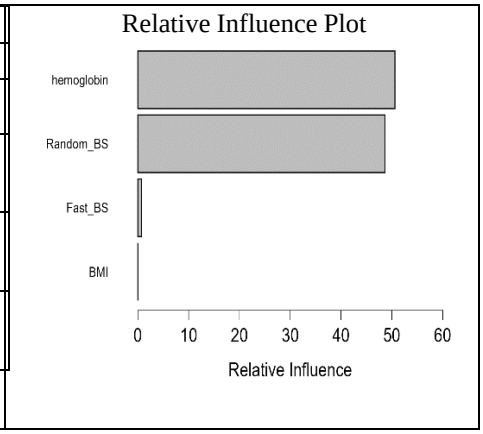

Table 5. Boosting classification algorithm evaluation matrix

Confusion matrix provided in the table 5., demonstrates observed versus predicted outcomes related to diabetics, nondiabetics and prediabetes. All the 200 tested cases were predicted without any errors. Hence, confusion matrix justifies the test accuracy as 100%. Further relative influence chart demonstrates hemoglobin and random blood sugar data are the influencing factors for determining the diabetes among smart city citizens.

Clustering techniques also consists of various algorithms including Hierarchical algorithm, K-Means clustering algorithm and Random Forest algorithm. For the purpose of sample analysis, K-Means clustering algorithm is applied and following results captured.

K-Means Clustering					
Clusters	N	R^2	AIC	BIC	Silhouette
3	1000	0.799	829.190	888.090	0.470

Table 6. K-Means clustering algorithm evaluation metrics

From table 6., we observe that, there are 1000 rows of data used for analysis. In clustering, more silhouette score demonstrates better clusters. If the silhouette score is 1 means best clustering can be formed. Above table demonstrates formation of 3 cluster with 0.470 as silhouette score. Hence good clustering with less overlapping can be achieved.

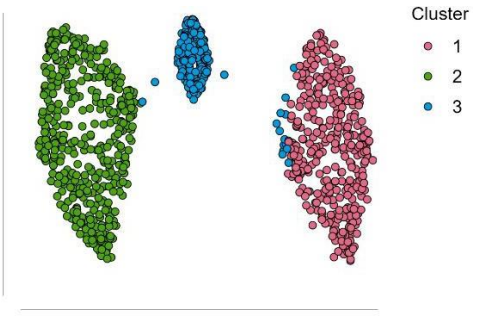

Figure 2. Plotting of clusters

Figure 2., demonstrates the illustration of clustering algorithm results, which showcases that, 3 clustering formation of data can provide better results. It can be noted that, the original data under test also consists of 3 groups categorized as diabetics, prediabetes and nondiabetics.

IV. Conclusion

Smart city facilitates holistic approach towards integration of technologies, computers, devices, sensors along with high speed connectivity and latest data transmission platforms. Capturing of various health related data dynamically and during mobility ensures seamless capturing and transmission of health data availability for tracking and monitoring purposes. Diabetes data is one of the basis, where smart city citizen's health and well being can be monitored and relevant strategy can be applied to overcome various health related issues. Using extended data or along with additional relevant health data, other deceases can also be predicted and strategies can be formulated to provide better quality of life for the smart city citizens. Machine learning facilitate prediction of outcomes in terms of deceases for the given set of health data. Hence applying machine learning techniques on the health dataset enables different types of predictions, classifications, clustering, forecasting of deceases and decision making on improving health and wellness of smart city populations. Machine learning outcome also can further drive the artificial intelligence enabled actions such as automated healthcare recommendations through recommenders to improve the health and wellness of smart city citizens.

References

[1] Huang K, Luo W, Zhang W, Li J. Characteristics and Problems of Smart City Development in China. *Smart Cities*. 2021; 4(4):1403-1419. https://doi.org/10.3390/smartcities4040074

[2] Kalra S, Kataria S, Pandey AK, Girdhar R, Das A, Kardwal N. Sugar smart, heart smart: The way smart cities should be. J Soc Health Diabetes 2016;4:51-4.

[3] Hurbean L, Danaiata D, Militaru F, Dodea AM, Negovan A. Open Data Based Machine Learning Applications in Smart Cities: A Systematic Literature Review. Electronics 2021, 10, 2997. https://doi.org/10.3390/electronics10232997

[4] Eiman AN, Hind AN, Nader M, Jameela AJ. Applications of big data to smart cities. J Internet Serv Appl, 2015, 6, 25. https://doi.org/10.1186/s13174-015-0041-5

[5] Aleksandr O, Viktoriia S, Lucie K, Justyna S, Salwa Saafi, Pavel P, Laura F, Darwin QG, Nadezhda C, Olga C, Asad A, Asma C, Ekaterina S, Waleed BQ, Raúl CM, Sylvia H, Joaquín TS, Sven C, Giuseppe R, Giuseppe A, Radim B, Jiri H, Elena SL.A Survey on Wearable Technology: History, State-of-the-Art and

[6] Sarker IH, Machine Learning: Algorithms, Real-World Applications and Research Directions, SN COMPUT. 2021, SCI. 2, 160. https://doi.org/10.1007/s42979-021-00592-x

[7] Malekloo A, Ozer E, AlHamaydeh M, Girolami M. Machine learning and structural health monitoring overview with emerging technology and high-dimensional data source highlights. Structural Health Monitoring. August 2021. doi:10.1177/14759217211036880

[8] Shipe ME, Deppen SA, Farjah F, Grogan EL. Developing prediction models for clinical use using logistic regression: an overview. J Thorac Dis. 2019 Mar;11(Suppl 4):S574-S584. doi: 10.21037/jtd.2019.01.25. PMID: 31032076; PMCID: PMC6465431.

[9] Van Engelen JE, Hoos HH. A survey on semi-supervised learning. Mach Learn, 2020,

109, 373–440. https://doi.org/10.1007/s10994-019-05855-6

[10] Alashwal H, El Halaby M, Crouse JJ, Abdalla A and Moustafa AA. The Application of Unsupervised Clustering Methods to Alzheimer's Disease. Front. Comput. Neurosci. 2019. 13:31. doi: 10.3389/fncom.2019.00031