International Journal of Engineering in Computer Science



E-ISSN: 2663-3590 P-ISSN: 2663-3582 IJECS 2020; 2(2): 22-25 Received: 10-04-2020 Accepted: 14-05-2020

Tejeshwini Dharoji GATE College, Tirupati, Andhra Pradesh, India

Novel machine learning techniques for detection of diabetes

Tejeshwini Dharoji

DOI: https://doi.org/10.33545/26633582.2020.v2.i2a.36

Abstract

Diabetes mellitus is a typical infection of human body brought about by a gathering of metabolic issue where the sugar levels over a drawn-out period is high. It influences various organs of the human body which in this way hurt an enormous number of the body's framework, specifically the blood veins and nerves. Early expectation in such illness can be controlled and spare human life. AI methods give productive outcome to remove information by developing anticipating models from demonstrative clinical datasets gathered from the diabetic patients. Extricating information from such information can be helpful to anticipate diabetic patients. In this work, we utilize four famous AI calculations, to be specific Support Vector Machine (SVM), Naive Bayes (NB), K-Nearest Neighbor (KNN) and C4.5 Decision Tree (DT), Random forest (RF), Logistic regression (LR) on grown-up populace information to anticipate diabetic mellitus. Logistic regression (LR), Support Vector Machine (SVM), Naive Bayes (GaussianNB) shows highest results.

Keywords: Diabetes mellitus, specific Support Vector Machine (SVM), Naive Bayes (NB), K-Nearest Neighbor (KNN) and C4.5, Decision Tree (DT), Random forest (RF), Logistic regression (LR), Accuracy, Prediction

Introduction

Diabetes mellitus, otherwise called diabetic, is an illness that influences the hormone insulin, bringing about unusual digestion of starches and improve levels of sugar in the blood. This high glucose influences different organs of the human body which thus confounds many reasons for the body, specifically the blood veins and nerves. The reasons for diabetic aren't yet totally found, numerous analysts accepted that both innate components and natural elements are included in that. Regardless, diabetic can used to be generally regular in adults and that is the reason it called grown-up beginning diabetes. It is as of now believed that diabetes mellitus is particularly engaged with the maturing procedure. As demonstrated by Canadian Diabetes Association (CDA), some place in the time of 2010 to 2020, the amount of distinctive individual make sense of to have diabetic in Canada is depended upon to heighten from 2.5 million to around 3.7 million ^[6]. The current overall circumstance isn't not quite the same as this. As demonstrated by the International Diabetes Federation, number of individuals having diabetes mellitus accomplished 382 million out of 2013 ^[7] that bring 6.6% of the world's all out adult populace. As indicated by the world medicinal services clinical information it has been normal that diabetic infection will be increment from 376 billion to 490 billion inside the year 2030 [8]. In addition, diabetic is a possibly autonomous contributing danger factor to small scale vascular traps. Diabetic patients are most likely progressively weak against a raised danger of miniaturized scale vascular harm, along these lines long haul inconvenience influences of cardio-vascular sickness is the main source of death. This small-scale vascular damage and hurried cardio vascular illness in the end brief to retinopathy, nephropathy and neuropathy^[9].

Related Study

Diabetes mellitus, otherwise called diabetic, is an ailment that influences the hormone insulin, bringing about anomalous digestion of starches and recoup levels of sugar in the blood. This high glucose influences different organs of the human body which thusly confounds many reasons for the body, specifically the blood veins and nerves. The reasons for diabetic isn't yet totally found, diabetic can used to be generally regular in adults and that is the reason it called grown-up beginning diabetes. It is at present believed that diabetes, the

Correspondence Tejeshwini Dharoji GATE College, Tirupati, Andhra Pradesh, India current overall circumstance isn't not quite the same as this. As showed by the International Diabetes Federation to expectation to sugar patients' disappointments ^[13]. Diabetic patients are most likely increasingly weak against a lifted danger of small-scale vascular harm, thusly long-haul intricacy influences of cardio-vascular malady is the main source of death. The current frameworks have unfitted to take care of the diabetic patients' concern. In the existing system they will prefer manual methods, so it will take time and high complication for processing. For example, ALjumah *et al.* ^[10] built up a prescient investigation model utilizing bolster vector machine calculation.

Proposed System

The proposed framework to aggregation of diabetes dataset with the applicable traits of the patients preprocesses the numeric worth characteristics, to apply the diverse AI arrangement strategies and relating prescient investigation using such information. To dissected the early expectation of diabetes by considering different hazard factors identified with this infection utilizing AI strategies. Extricating information from genuine human services dataset can be helpful to anticipate diabetic patients. To anticipate diabetes mellitus adequately, we have done our examinations utilizing four famous AI calculations, in particular Support Vector Machine (SVM), Naive Bayes (NB), K-Nearest Neighbor (KNN) and choice tree, Logistic regression, Random forest on grown-up populace information to foresee diabetes mellitus. To help social insurance to take early avoidance and settle on better clinical choices to control diabetes and in this manner spare human life.

SVM

Support Vector Machine or SVM is an administered and direct Machine Learning calculation most normally utilized for tackling grouping issues and is likewise alluded to as Support Vector Classification. There is additionally a subset of SVM called SVR which stands for Support Vector Regression which utilizes similar standards to take care of relapse issues. SVM likewise supports the part strategy additionally called the kernel SVM which permits us to handle non-linearity.

How might a machine distinguish an apple or an orange? Of course, it depends on the attributes that we furnish the machine with. It tends to be size, shape, weight and so forth. The more highlights we consider the simpler it is to recognize and recognize both.

For now, we will simply concentrate on the weight and size (diameter) of apples and oranges. Presently how might a machine utilizing SVM, arrange another natural product as either apple or orange simply dependent on the information on the size and loads of about 20 apples and oranges that were watched and named? The beneath picture delineates how.

The goal of SVM is to draw a line that best isolates the two classes of information focuses".

SVM creates a line that can neatly isolate the two classes. How spotless, you may inquire. There are numerous potential methods of drawing a line that isolates the two classes, notwithstanding, in SVM, it is dictated by the margins and the support vectors.



Fig 1: Support vector machine hyper plane



Fig 2: Over view of the overall process

Algorithm

Here in machine learning it uses step wise processing technique to make operations on the uploaded data set:

Step 2: Data cleaning. Here we are applying some other operations like finding the null values, filling the null values, removing the null values or duplicate values which make redundancy in the data set.

Step 3: If the data set contains any categorical values means we need convert those categorical values to numerical values for easy execution purpose.

Step 4: Slicing and identifying the dependent and independent variables, partitioning the whole data set into two parts for training and testing purpose.

Step 5: Applying the required algorithm (decision tree) on dataset for further performance.

Step 6: Calculating some mathematical issues to know that how our algorithm is performing on given data set.

Step 7: Predicting the result.

Advantages

- Less time for processing.
- Less maintenance is required.
- Less cost.
- More efficient.

Step 1: Importing the data set and displaying the data set.

- Easley identify the diabetic patients
- Treatment support and suggestions medicines to patients.
- Accuracy of results is high.

Results and Discussions

By using the all machine learning upgraded calculations/methods effectively we will get the expectation. In this venture we utilized Logistic regression, Decision tree, Random forest, Naive Bayes, SVM, K-nearest neighbor calculations. Every calculation will assume an alternate job, in light of their exhibition level we will get the precise anticipated yields. Logistic regression, Naïve bayes, Support vector machine gives the great outcomes true to form from the preprocessed information.

View Data									
	Pregnar	ncies Gluco	se BloodPre	essure Skin Thick	kness Insul	InBMI DiabetesPedigr	eeFunction Age Outcome		
0	6	148	72	35	O	33.60.627	50 1		
1	1	85	66	29	0	26.60.351	31 0		
2	8	183	64	0	0	23.30.672	32 1		
3	1	89	66	23	94	28.10.167	21 0		
4	0	137	40	35	168	43.1 2.288	33 1		
5	5	116	74	0	0	25.60.201	30 0		
6	3	78	50	32	88	31.00.248	26 1		
7	10	115	0	0	0	35.30.134	29 0		
8	2	197	70	45	543	30.50.158	53 1		
9	8	125	96	0	0	0.0 0.232	54 1		
10	4	110	92	0	O	37.60.191	30 0		
11	10	168	74	0	0	38.00.537	34 1		
12	10	139	80	0	0	27.11.441	57 0		
13		189	60	23	846	30,1 0.398	59 1		
14	5	166	72	19	175	25.80.587	51 1		

Fig 3: Displaying input data set in data frame.



Fig 4: Model selection for performance purpose.

The selected Model is based on LogisticRegression whose	accuracy score is 76.62337662337663
Select model	
Select avogtar *	
staria.	
🔞 😥 🧱 🔕 🖻	- 20 (2

Fig 5: Applying Logistic regression algorithm.

	Select model	
	Select on option +	
	Surgers?	
	1 million (1997)	

Fig 6: Applying Naïve bayes algorithm.

The	selected Mode	el is based or 7	SupportVector 6.62337662337	Machine wh	ose accurac	y score is
			Select model			
			S.Arrey .			
0 8 📜	🖬 🌍 🕅					- Ref)

Fig 7: Applying the Support Vector Machine algorithm.

After comparison of the all the applied calculations we are getting most noteworthy outcomes for Logistic regression, Naïve bayes, and Support vector machine accuracy as 76.62% and furthermore here we are looking at exactness, accuracy, and review.

Conclusion

In this work, we have dissected the early forecast of diabetes by considering different hazard factors identified with this infection utilizing machine learning strategies. Removing information from genuine human services dataset can be helpful to anticipate diabetic patients. To detect diabetes mellitus adequately, we have done our tests utilizing four famous machine learning calculations, to be specific Support Vector Machine (SVM), Naive Bayes (NB), Logistic regression and C4.5 choice tree, on grown-up populace information to anticipate diabetes mellitus.

References

- Platt John C. "12 fast training of support vector machines using sequential minimal optimization." Advances in kernel methods, 1999, 185-208.
- 2. John George H, Pat Langley. "Estimating continuous distributions in Bayesian classifiers." Proceedings of the Eleventh conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc, 1995.
- Aha David W, Dennis Kibler, Marc K Albert. "Instance-based learning algorithms." Machine learning. 1991; 6(1):37-66.
- 4. Ross Quinlan. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo, CA, 1993.
- 5. Witten IH *et al.* Weka: Practical machine learning tools and techniques with Java implementations, 1999.
- Morteza M, Franklyn P, Bharat S, Linying D, Karim K, Aziz G. Evaluating the Performance of the Framingham Diabetes Risk Scoring Model in Canadian Electronic Medical Records. Canadian journal of diabetes. 2015; 39:152-156.

- V AK, RC. Classification of Diabetes Disease Using Support Vector Machine. International Journal of Engineering Research and Applications. 2013; 3:1797-1801.
- Carlo BG, Valeria M, Jesús DC. The impact of diabetes mellitus on healthcare costs in Italy. Expert review of pharmacoeconomics & outcomes research. 2011; 11:709-19.
- Nahla B Andrew, *et al.* Intelligible support vector machines for diagnosis of diabetes mellitus. Information Technology in Biomedicine, IEEE Transactions. 2010; 14:1114-20.
- Abdullah A Aljumah *et al.* Application of data mining: Diabetes health care in young and old patients, Journal of King Saud University - Computer and Information Sciences. 2013; 25(2):127-136
- 11. Kavakiotis Ioannis, Olga Tsave, Athanasios Salifoglou, Nicos Maglaveras, Ioannis Vlahavas, Ioanna Chouvarda. Machine learning and data mining methods in diabetes research. Computational and structural biotechnology journal, 2017.
- 12. Zheng Tao *et al.* A machine learning-based framework to identify type 2 diabetes through electronic health records. International journal of medical informatics. 2017; 97:120-127.
- Rani A Swarupa, S Jyothi. "Performance analysis of classification algorithms under different datasets." In Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on, 2016, 1584-1589.
- 14. Kandhasamy J Pradeep, Balamurali S. "Performance analysis of classifier models to predict diabetes mellitus." Procedia Computer Science. 2015; 47:45-51.
- Huang Y, McCullagh P, Black N, Harper R. Feature selection and classification model construction on type 2 diabetic patients'data, Artificial Intelligence in Medicine. 2015; 41(3):251-262.
- 16. Meng XH, Huang YX, Rao DP, Zhang Q, Liu Q. Comparison of three data mining models for predicting diabetes or prediabetes by risk factors. The Kaohsiung journal of medical sciences. 2013; 29(2):93-99.