Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2024 SCPE. Volume 25, Issues 4, pp. 3200–3213, DOI 10.12694/scpe.v25i4.2884

PREDICTION OF DIABETES MELLITUS USING ARTIFICIAL INTELLIGENCE TECHNIQUES

G L SUMALATA*, JOSHITHA C†, AND KOLLATI MEENAKSH‡

Abstract. Diabetes Mellitus (DM) is a global health challenge, demanding proficient predictive models for early identification and intervention. This study adopts a comprehensive strategy for diabetes prediction with Machine learning algorithms, utilizing PIMA Indian diabetes dataset which encompasses clinical, demographic and lifestyle data. Employing techniques like Recursive Feature Elimination (RFE) and correlation analysis, the feature selection process identifies influential predictors, including glucose levels, Body Mass Index (BMI), Blood Pressure and diabetic history of family. A distinctive facet of this study involves integrating IBM Auto AI, automating the machine learning pipeline for tasks like feature engineering, hyperparameter tuning and model selection. Through comparative analysis, the research evaluates the efficiency and performance enhancements achieved through automation in contrast to manually-tailored models. Evaluation metrics encompass accuracy, precision, recall, and F1 score. Cross-validation, particularly k-fold cross-validation, ensures model generalization to diverse subsets of the dataset. The research outcomes offer valuable insights into the optimal amalgamation of AI techniques for diabetes prediction, underscoring the significance of interpretability, performance, and automation in healthcare analytics. The proposed Methodology is evaluated with different classifiers with Auto AI and without Auto AI techniques. Using IBM Auto AI,Gradient boosting algorithm performed well with 84. 4 % accuracy and Logistic Regression showed good accuracy of 84. 4% among conventional machine learning techniques without Auto AI using Pima Indian Diabetes Dataset.

Key words: Diabetes Mellitus, IBM, Auto Artificial Intelligence, Extreme Gradient Boosting, Gradient Boosting Classifier

1. Introduction. Diabetes Mellitus is a pervasive metabolic disorder which serves as a harbinger for intricate complications that causes multiple physiological changes in human body. The uncontrolled glycemic imbalances due to Diabetes Mellitus result vast changes in cerebrovascular system. The other effects of the Diabetes mellitus are ocular afflictions leading to visual impairment and also causes nephropathies which affect the renal function [20]. This disease is receiving global attention due to its long term health complications which affects the subjects suffering with the disease. According to World Health Organization (WHO) and International Diabetes Federation(IDF) estimates 387 million people are suffering from the disease. It is forecasted that by the year 2034 the growth will be escalated up to 592 million patients [10]. There are two types of diabetes- type 1 and type 2 [2]. Type-1 Diabetes is caused if pancreatic cells that produce insulin is attacked by the immune system because of which secretion of insulin will be less. This type of diabetes mostly seen in children. WHO stated that number of children having Type-1 diabetes is very high . In patients having Type-1 diabetes, the risk of cardiovascular diseases is more . Not only heart strokes but also various organs like eyes, kidneys etc. get affected [12]. Further, the disease shows gradual deterioration of the circulatory system damaging the retina which may cause diabetic retinopathy causing visual impairement [19]. Type 2 diabetes is caused by overweight or obesity which is due to lack of physical activity and causes resistance to insulin. This further causes one of the dangerous risk factors for cardiovascular diseases. Type 2 diabetes is due to lifestyle, physical activity, diet and also hereditary. Most of the diabetic people i. e 90 % are of Type-2 . The underlying cause of type 1 or type 2 diabetes is irregular blood glucose levels leading to pancreatic dysfunction which finally causes deficiency in insulin production in type 1 and apart from this cellular resistance to insulin in type 2 . The third type of diabetes is gestational diabetes mellitus which occurs due to glucose tolerance during pregnancy[5] . While the exact cause remains elusive, the genetic factors and lifestyle are believed to significantly contribute

*Dept. of ECE Koneru Lakshmaiah Education Foundation, Hyderabad, Telangana, 500075, India (sumaprasadgl@gmail. com),
†Dept. of ECE Koneru Lakshmaiah Education Foundation Hyderabad, Telangana, 500075, India
‡Department of Electronics and Communication Engineering, GRIET Hyderabad, Telangana, 500075, India

to diabetes. Although incurable, the condition can be effectively managed through treatment and medication. Early detection and treatment play a vital role in preventing complications and reducing the risk of severe health issues. As technology is improving in almost all fields [8, 9, 11] , Even in the health care AI many improvements and researches are done. The realm of bioinformatics, numerous researchers have endeavored to address diabetes, creating systems and tools for prediction. Predictive Analysis is a methodology which integrates diverse machine learning algorithms, data mining techniques, and statistical methods and shows great potential in the context of Diabetes Mellitus. By analysing both current and historical data, this approach aims to extract valuable insights and forecast future events related to diabetes. When applied to healthcare data specific to diabetes, predictive analysis becomes a powerful tool for making informed decisions and generating predictions regarding the disease. Employing machine learning techniques for predictive analytics in diabetes care is directed towards achieving precise disease diagnosis, enhancing patient care strategies, optimizing resource allocation, and ultimately improving clinical outcomes in the management of Diabetes Mellitus. Various machine learning algorithms such as Decision Trees, Support Vector Machine (SVM), and Linear Regression, have been commonly employed . Additionally, the use of Artificial Neural Network (ANN) is explored for the same and, more recently, Deep Learning (DL) as an enhancement to ANN, also has shown promising results. The variation in accuracy rates obtained from these methods made the researchers to explore more and more novel classifiers or combinations of existing classifiers to improve accuracy. Many studies in diabetes prediction have utilized the publicly available Pima Indian Dataset from the UCI repository. Some surveys in the field have focused on specific machine learning and deep learning techniques for predicting diabetes [3].

This research paper distinguishes itself by discussing both conventional Machine Learning techniques and implementation using IBM Auto AI models for diabetes prediction, comparison between both the methods. The paper systematically discusses the step by step process of implementation of Auto AI model. The results obtained are subjected to comparative analysis with other research studies employing the same dataset. The paper is organized into subsequent sections. In Section II titled Related works, literature review and a taxonomy of machine learning algorithms related to diabetes prediction are presented. Section III outlines IBM Auto AI services. Section IV delves into the methodology of proposed model for diabetes prediction. Section V discusses the summary of Progress map. The Simulation results and discussions are detailed in Section VI. The Conclusion and Future scope are outlined in Section VII followed by References.

2. Related works. Health care AI intruding in the medical field to reduce the burden of physicians in the decision making. Early detection and diagnosis of Diabetes Mellitus are crucial for timely intervention, improved treatment outcomes, and preventing complications. AI techniques helped the physicians to predict the life challenging diseases such as cancers, tumors using the machine learning techniques[7, 6] . The examination of existing research reveals that predictions are done for diabetes detection through a range of techniques and methods. Some of them are data mining techniques, machine learning algorithms, or combinations of them. As the complexity is increasing many researchers are exploring deep learning algorithms. Different research works using Pima Indian Diabetes dataset have been reviewed thoroughly. First the Research began with introduction of neural networks leading to build adaptive models for diabetes prediction. ADAP [15], a neural network is used for the prediction of Pima Indian population dataset and obtained specificity and sensitivity as 0. 76. Next approaches are done using traditional statistical methods and clinical risk factors. Early models often used simpler algorithms, such as logistic regression. Some of the literatures concluded that the direct and distribution free feature of neural networks is used for prediction but observed that reliability is not upto the mark. Some researchers stated that general regression neural networks (GRNN) are showing good performace in comparison with multilayer perceptron (MLP) and radial basis function (RBF). Next phase of the research is exploring feature selection methods. Broader adoption of machine learning techniques gained importance in this area.

Kemal Palat et al. [13] employed a two stage Generalized Discriminant Analysis and least square support vector machine for the Diabetes Mellitus detection and obtained an accuracy of 82. 05%. Support Vector Machine cannot give probability of chances of accuracy. The use of Automated Machine Learning tools has streamlined the machine learning pipeline and these tools automate the process of feature engineering, hyperparameter tuning, and model selection, reducing the need for extensive manual intervention. Yashi et al. [16] stated advantages of Auto AI Classification is done using Microsoft Azure, which makes the classification

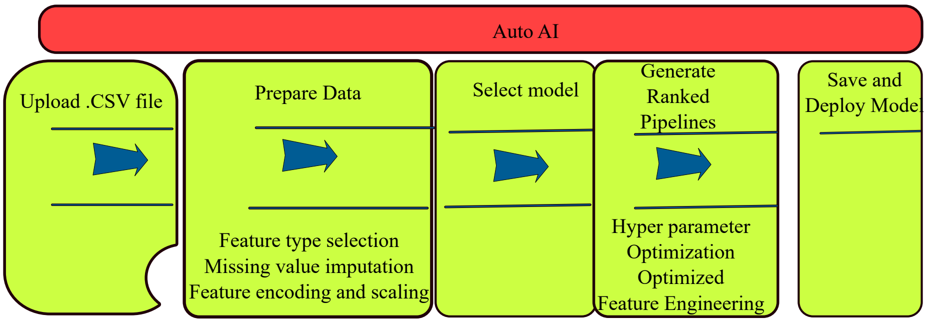

Fig. 3.1: IBM Auto AI procedure in general

easy by reducing the time. The references [4, 14, 18, 1] implemented using the same dataset and obtained results using different classification algorithms. With the surge in computational power, Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks were applied for their ability to capture intricate patterns in data. Ensemble techniques and hybrid models considering the combinations of multiple algorithms are becoming popular. These models often integrated the strengths of different approaches, improving prediction accuracy. Many researchers worked on techniques like bagging, boosting, and stacking for prediction of diabetes. Interpretability and explainablity is the ongoing trend in Explainable AI models especially in healthcare. Tasin et al. [17] Proposed semi supervised model with extreme gradient boosting algorithm with Domain adaptation and by using frameworks like LIME and SHAP ,an explainable AI approach was used for model predictions.

Recent advancements focus the importance of robust validation methods like k-fold cross-validation and model performance can be assessed by considering various Evaluation metrics such as accuracy,precision etc. The main objectives the proposed work is:

- The proposed model employs IBM Auto AI services for the intelligent automation of the entire AI lifecycle for developing a predictive model for the diabetic Mellitus detection.
- Unlike conventional machine learning algorithms, the IBM Auto AI steps of using data sets for training, finding the optimized model for the given data, feature selection, hyperparameter optimization and ranking the pipelines based on evaluation metrics accuracy and precision. The best performing pipelines are brought to train the new data and prediction outcomes are considered based on the model training. This relieves the doctors from the rote of manual processing.
- A comparative study is made with the conventional machine learning algorithms to prove that the auto AI based methods have shown superior performance compared to the conventional machine learning algorithms.

3. IBM Auto AI Services. Proposed work uses IBM Auto AI services. Auto AI reduces experimentation time. So many researches are going on with Auto AI dealing the problems like classification and other applications in almost every field. Auto AI does automation of complete life cycle of AI i. e from data preparation to hyper parameter optimization. It includes development of model, feature engineering etc. with one click deployment. The general architecture of IBM Auto AI Services and its application to the disease Diabetes Mellitus is depicted in Fig. 3.1 and Fig. 3.2. The following services are used in IBM Auto AI:

1. IBM Watson Studio
2. IBM Watson Machine Learning
3. Node-RED

3.1. IBM Watson Studio. IBM Watson Studio is a vital tool to work with Artificial Intelligence and take decisions in optimized manner on IBM Cloud Pak. It integrates many open source frameworks, tools, high level languages like python to program, analyse the outputs, compare various algorithms with different parameters and for identification of best one among them.

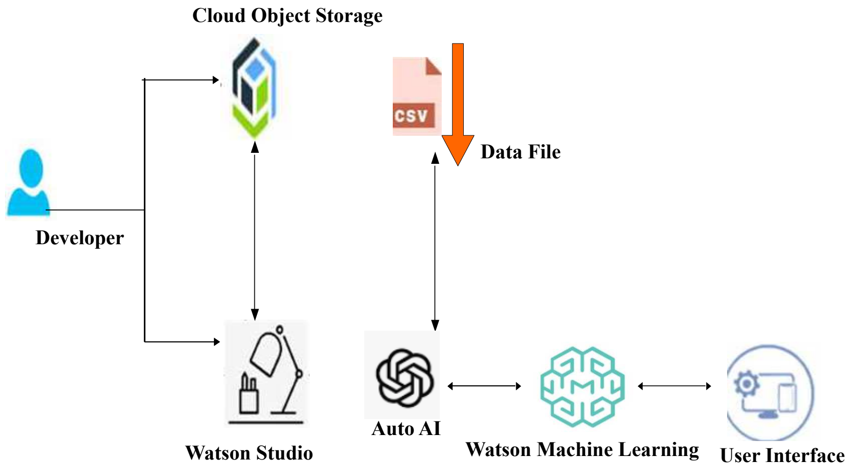

Fig. 3.2: IBM Auto AI for Diabetes Prediction

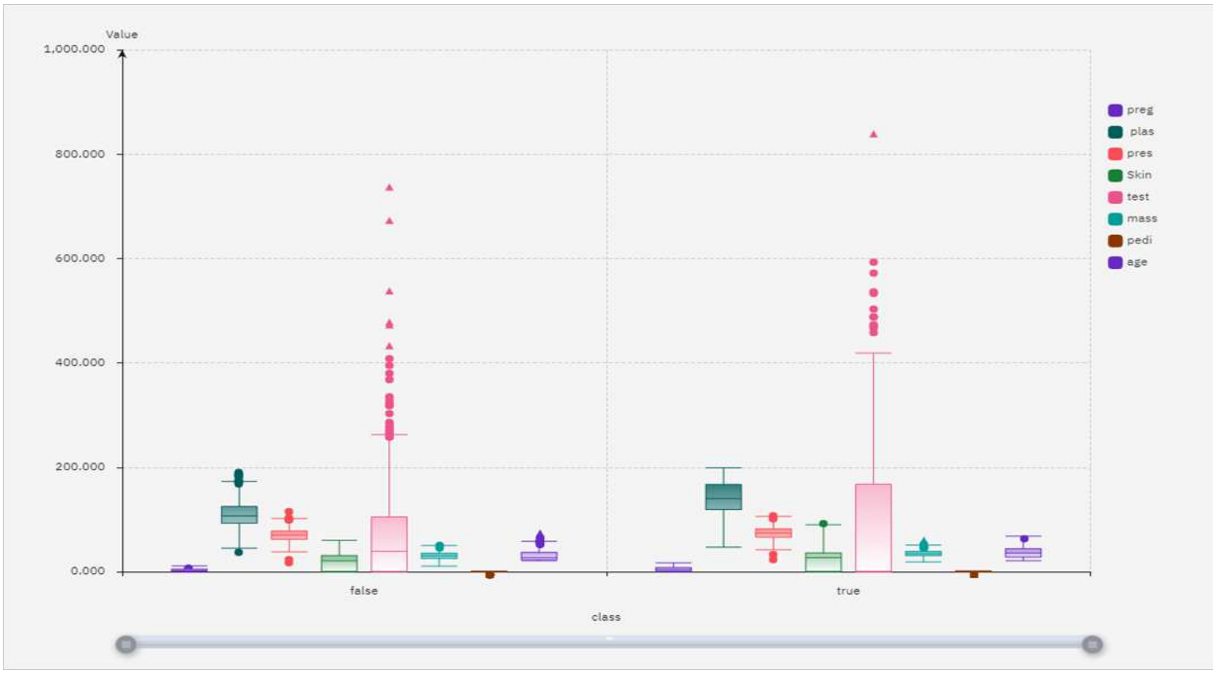

Fig. 3.3: Box Plot Representation of data

3.2. IBM Watson Machine Learning . IBM Watson Machine Learning provides provision for building models, neural networks and analyzing them is possible. It is possible to load the data set and training with our own data to build an application. It is used to test the model and successfully deploy it. All these tasks are possible by tools and wide range of services provided by IBM Watson studio to make the process fully automated which saves the precious time of the user so that prototyping can be done at a faster rate.

3.3. Node-RED. Node-RED is an environment to represent the messages in visual form through a programming technique of following the flow. It is a "Flow based Programming" tool. The program is broken into data and processes. Processes are one which operate on data and these processes are connected together by

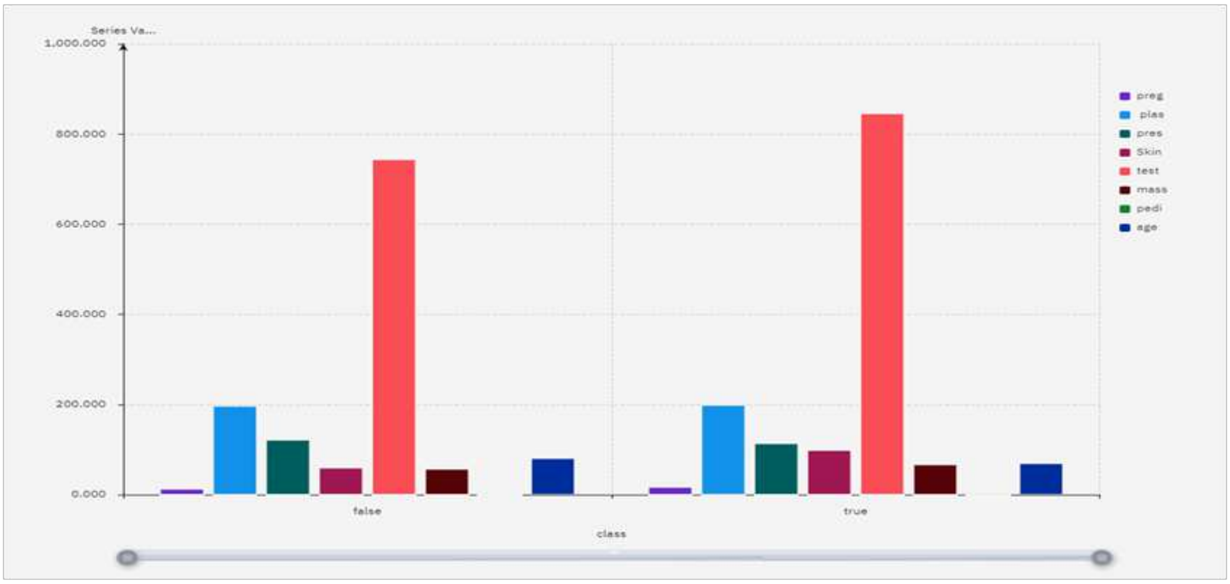

Fig. 3.4: Scatter Matrix Representation of data

network. All this is done to achieve an objective finally. This tool is used to connect together various hardware, APIs etc. Many online services too can be wired up using this tool. It provides editing via browser and with one click deployment can be done. Creation of functions of java script can be done within editor. Reuse of some useful templates also is possible using some functions which are built in. To do the description of metadata Node-RED uses "JavaScript Object Notation i. e JSON". These flows which are created with JSON can be easily shared with others. It can be run anywhere from local computer to cloud. It can run even in hardware devices for building IOT applications like Raspberry Pi too. Due to all these vital features Node-RED is become popular.

3.4. IBM Cloud Object Storage. IBM Cloud Object Storage facilitates low cost solution for storage of data on edge. Performance of it can be increased by increasing the capacity of storage with many more interesting features. It can be used for storage of AI, image repositories etc in an effective manner. Video or analytics of many customers can be stored on cloud. The very important feature is the protection of data with built in authentication tools in it. From any location the stored data can be easily accessed, which is a very important feature i. e simplicity. Time can be saved for searching the data as meta data is created. So easily the stored data can be searched

4. Methodology. The proposed model uses Auto AI Architecture. The General approach of Auto AI using step by step procedure can be depicted in Fig. 4.1. The foremost step starts with loading the data set in . csv file format and then reading the data file, analyzing it by plotting various types of appropriate plots and running the Auto AI experiment by data preparation which undergoes 3 main steps called Feature type selection, Missing value imputation, feature encoding and scaling. Next step is the model selection which can be done by selection of best estimator for the data. After that Generation of Rank model pipelines after hyper parameter Optimization (HPO) and Optimizing Feature Engineering. Finally the best model is saved and deployed. In this proposed work for diabetes Mellitus Prediction, . csv file of Pima Indian dataset is loaded and all the steps in block diagram shown in Fig. 4. 2 specific to Diabetes Mellitus Prediction are executed for the optimized model building.

4.1. DataSet details and Step by Step Approach of Implementation . The proposed system uses 8 vital parameters that can predict the chances of acquiring the diabetes which uses machine learning techniques to learn from hundreds of patient data, fit the best performing model consistent with the training examples

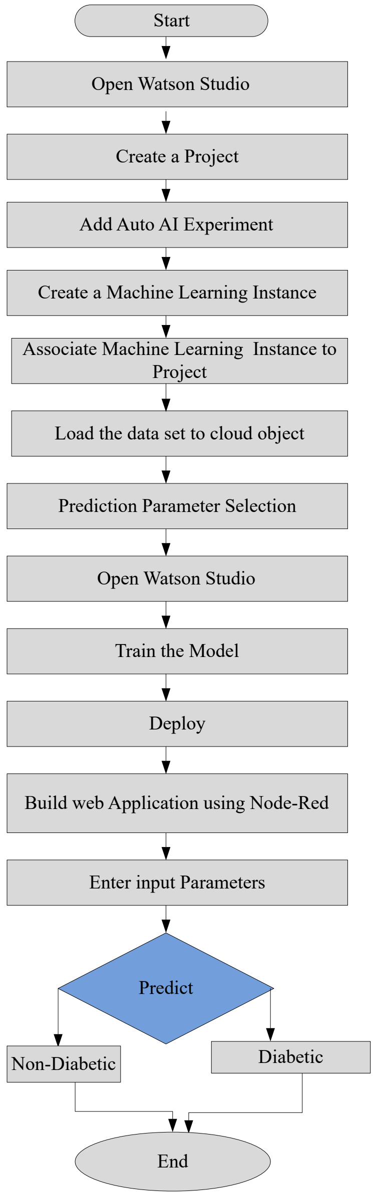

Fig. 4.1: Flowchart for Diabetes Prediction

given and it is tested or validated with unseen data also. It is found to perform with about good accuracy in its predictions. So, an application built on this will be highly reliable with good prediction accuracy. The dataset contains 8 input parameters which are of different data types. Diabetic Pedigree Function and BMI are double and the remaining parameters are in integer format.

4.2. Dataset description is given below:. The data is taken from Kaggle . The link to the dataset is $https : //www.kaggle.com/datasets/uciml/pima - indians - diabetes - database/data$. The ratio used for experiment is 80:20. Majority of the data which is 80% is used for training and 20% for testing.

1. No. of Pregnancies - Number of times pregnant
2. Random Glucose - Plasma glucose concentration a 2 hours in an oral glucose tolerance test

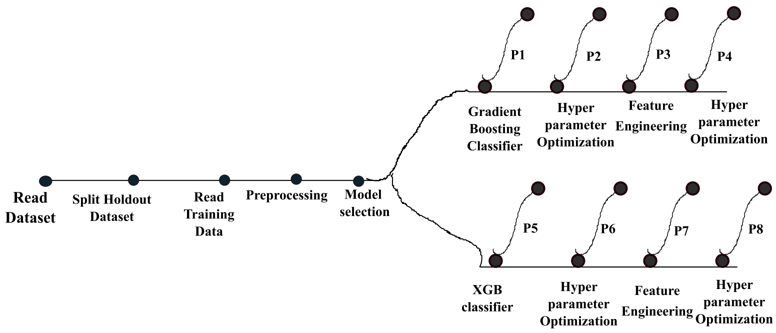

Fig. 5.1: Progress Map

3. Blood Pressure - Diastolic blood pressure (mm Hg)
4. Skin Thickness - Triceps skin fold thickness (mm)
5. Insulin - 2-Hour serum insulin (mu U/ml)
6. Body Mass Index - weight (kg) / [height (m)]2
7. Diabetes Pedigree Function - A function which scores the likelihood of diabetes based on family history. It provided some data on diabetes mellitus history in relatives and the genetic relationship of those relatives to the patient
8. Age - Age (in years)
9. Outcome - The outcome label 1 for Yes (for chances of acquiring diabetes and 0 for No (for no chances of acquiring diabetes)

4.3. Data Visualization. The data can be visualized and analyzed in a better way by checking the distribution between the parameters, finding the correlation between them by using different plots. Fig. 3.3 and Fig. 3.4 shows Box Plot Representation and Scatter Matrix Representation of data respectively.

Steps followed to build the model The step by step procedure is shown in Fig. 4.1 as a flowchart for diabetes prediction using IBM Auto AI.

The main steps involved are:

1. Creation of a project in Watson Studio
2. Diabetes Prediction
3. Adding Auto AI experiment
4. Creation of a Machine Learning instance
5. Association of ML instance to the project
6. Loading the dataset i. e PIMA Indian data set to cloud object storage
7. Selection of the target variable (prediction parameter) in the dataset
8. Training the model
9. Deployment of the model
10. Building web application using Node-Red

5. Summary of Progress Map. Progress map helps to see the steps for creating the pipelines of the model. Each step of the progress map are shown in Fig. 5.1

The process begins with reading the data set, next splitting holdout data. After reading training data , preprocessing in which the features are to be detected and categorization takes place. Then Model Selection takes place. Here different algorithms are applied and which ever matches well for this particular data set will be selected. Here PIMA Indian dataset is applied out of different algorithms, two best algorithms are matched for this data and selected . Comparison of pipelines can be done. The chart in Fig. 5.2 shows 8 pipelines P1,P2,.... P8 with respect to cross validation and holdout scores.

Ranking is given to pipelines based on performance metrics. Comparison between two best classifiers i. e Gradient Boosting Classifier and XGB classifier chosen from data set is done. Different types of analysis can

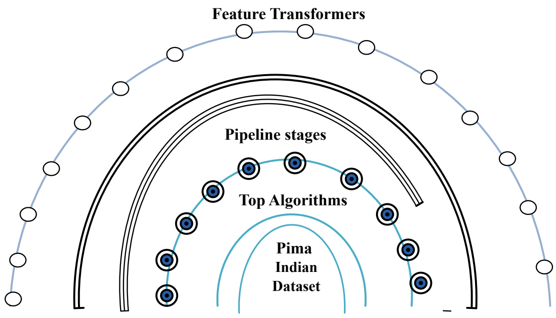

Fig. 5.2: Pipeline Comparision

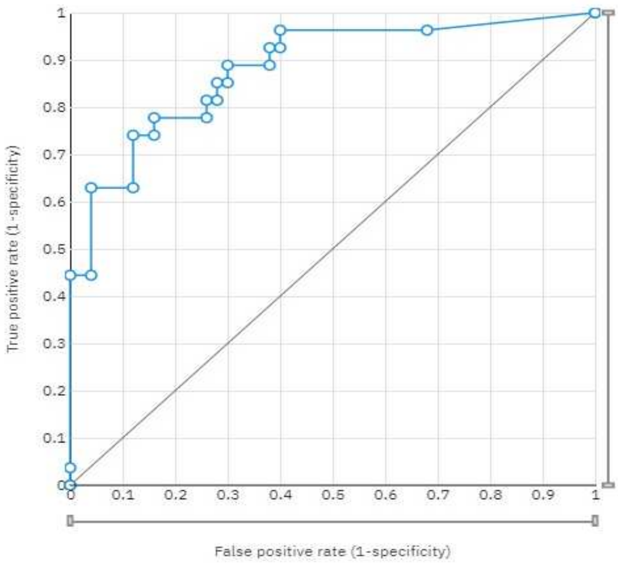

Fig. 5.3: ROC of Gradient Boosting Classifier

be done using relationship maps, Metric charts etc. Relationship map helps to know about the pipelines and the sharing between them, the unique properties of them. Pipelines which are associated to a particular node in the leader board and analysis between the relationship can be known. While creating the relationship map the number of pipelines are decided by algorithms selected. Once the experiment is run then Auto AI generates pipelines using different estimators or enhancements like hyper parameter optimization i. e HPO (used for building number of models and to do comparison between them by automatically exploring parameters) and feature engineering i. e FE (depending on problem which parameters are to be included for accurate prediction by considering raw data). Once the run has completed relationship map can be viewed. The pipeline which has highest ranking is best classifier i. e Gradient Boosting Classifier in this diabetes mellitus prediction model. So it can be saved. Next step is Promotion to deployment space and it can be done using API reference. Next Testing the prediction model can be done after giving input parameters and depending on the parameters output will be shown as 1 or 0 as output diabetic patient or not respectively. Finally user Interface can be built with cloud foundry applications. Using Node-Red building nodes and deployment can be done. It also connects APIs and online services. It is used for building user interface for users to enter data so that the status of diabetic can be predicted.

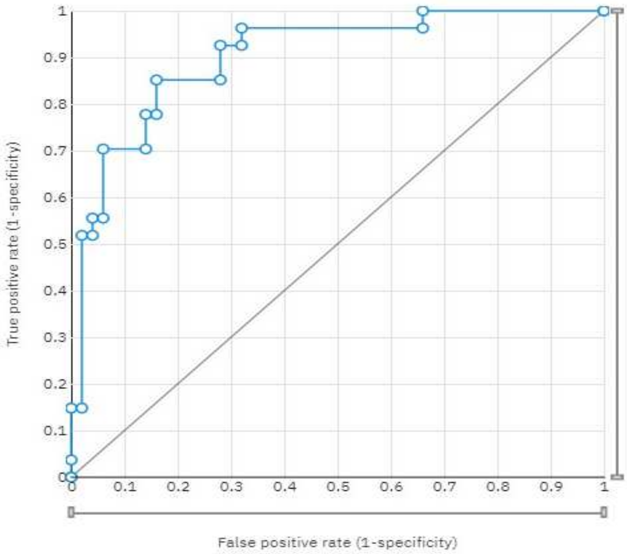

Fig. 5.4: ROC of of XGB Classifier

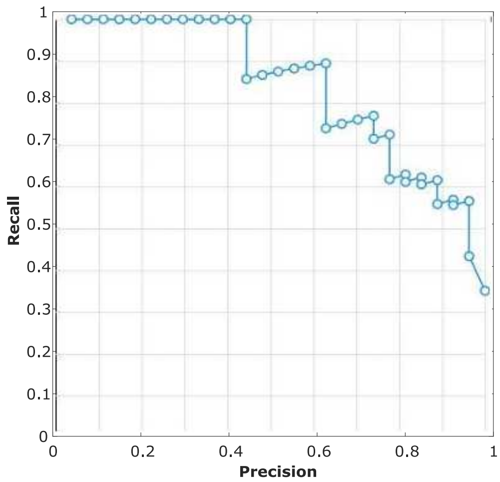

Fig. 5.5: Precision Recall curve of Gradient Boosting Classifier

6. Simulation Results and Discussions.

6.1. Test beds used in the work. The Pima Indian Diabetes dataset itself serves as a foundational test bed. This dataset includes various health-related features, making it suitable for both conventional machine learning and IBM AutoAI approaches. The experiment is performed using Auto AI or without using Auto AI, the performance can be evaluated using few performance metrics like Accuracy, ROCAUC, Precision, Recall and F1 measure. They are shown in Equ. 6. 1,6. 2,6. 3,6. 4,6. 5 and 6. 6.

Accuracy:

$$Accuracy = (TP + TN)/(TP + FP + TN + FN) * 100 \tag{6.1}$$

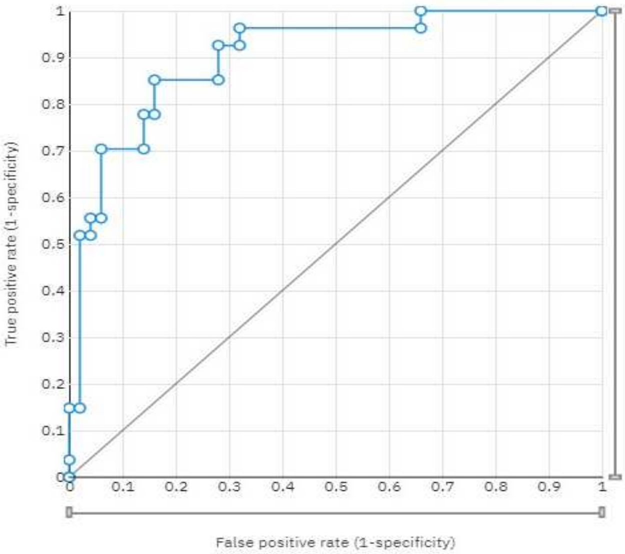

Fig. 5.6: Precision Recall curve of XGB Classifier

Precision:

$$Precision = TP/(TP + FP) * 100 \tag{6.2}$$

Recall or Sensitivity:

$$RecallorSensitivity = TP/(TP + FN) * 100 \tag{6.3}$$

F1 Score:

$$F1 = 2 * (precision * recall)/(precision + recall) * 100 \tag{6.4}$$

Specificity:

$$Specificity = TN/(TN + FP) * 100 \tag{6.5}$$

AUC:

$$AUC = \int_0^1 TruePositiveRate(FPR)d(FalsePositiveRate) \tag{6.6}$$

where TP,TN,FP,FN in Equ. 6. 1,6. 2,6. 3,6. 4,6. 5 and 6. 6 are True positive, True Negative, False Positive and False Negative respectively. In this equation, the True Positive Rate (TPR) is also known as Sensitivity or Recall, and the False Positive Rate (FPR) is the complement of Specificity.

This experiment for prediction of diabetes Mellitus is performed using 2 methods. First method is experimenting by utilizing IBM Auto AI tools and resources for diabetes prediction and same experiment is repeated without Auto AI using conventional Machine Learning Process . Model building using different machine learning algorithms is done manually . Here 12 such algorithms are used for evaluating the model using same PIMA Indian dataset and different performance metrics are obtained such as Accuracy,Precision,Recall,F1 score, ROC etc. and results are compared with IBM Auto AI results. Comparison of AI with Auto AI is done in this work. The advantage of using Auto AI is faster execution and good performance. Table. 6.1 depicts results obtained by Auto AI process. In Table 1 GB referes to Gradient Boosting Classifier and XGB refers to XG Boost Classifier. Comparision of performance metrics without Auto AI are tabulated in Table. 6.2 . Similarly for Table. 6.2 the abbreviations are listed in detail.

Table 6.1: Comparision of performance Metrics using IBM Auto AI best models

Model	GB (Auto AI)	XGB (Auto AI)
Accuracy	84.4	80.5
ROC AUC	90.3	82.1
Precision	85.7	75
Recall	66.7	66.7
F1	75	70.6

Table 6.2: Comparison of performance Metrics without Auto AI

Model	L R	R F	SVM	KNN	GB	GNB	XGBt	XT	LGB	AB	BNB	DT
Accuracy	84.415	80.51	80.51	79.22	77.92	77.92	77.92	75.32	74.02	74.02	72.72	66.23
ROC AUC	80.69	78.69	77.752	77.71	74.84	75.79	75.79	71.94	72.85	70.96	69.04	64.14
Precision	81.81	70.37	72	67.85	68	66.66	66.66	64	60	61.53	60	50
Recall	69.23	73.07	69.23	73.07	65.38	69.23	69.23	61.53	69.23	61.53	57.69	57.69
F1	75	71.69	70.58	70.37	66.66	67.92	67.92	62.74	64.28	61.53	58.82	53.57

1. LR: Logistic Regression
2. RF: Random Forest
3. SVM: Support Vector Machine
4. KNN: k-Nearest Neighbors
5. GB: Gradient Boosting
6. GNB: Gaussian Naive Bayes
7. XGB: XGBoost
8. XT: Extra Trees
9. LGB: LightGBM
10. AB: AdaBoost
11. BNB: Bernoulli Naive Bayes
12. DT: Decision Tree

Using Method 1, Gradient Boosting Classifier is selected by the Auto AI experiment as the best performing model after fine tuning all the hyper-parameters. Fig. 5.3 and Fig. 5.4 depicts ROC curve of Gradient Boosting and XGB classifiers respectively. Fig. 5.5and Fig. 5.6 depicts Precision Recall Curve of Gradient Boosting and XGB classifiers respectively.

The Area Under the Curve (AUC) is also satisfactory which depicts the TPR (sensitivity) and FPR (specificity). The models having higher AUC are said to perform better.

In method2 ,the experiment is repeated without Auto AI i. e conventional machine learning algorithms are optimized using a new approach called as grid search and consolidated Accuracy and ROC AUC results are tabulated in Table. 6.3

Optimized Logistic Regression is good at Accuracy i. e 84. 4 percentage and ROCAUC is 88. 98%. When the Experiment is performed with IBM Auto AI, Gradient Booster Classifier is best among others which achieved hold out accuracy of 84. 41%. The cross validation score of 77. 1%. The Area under ROC,Precision,Recall,F1 score, Average Precision and log loss of Gradient Boosting Classifier which is top Performing Classifiers using IBM Auto AI are 90. 3%, 85. 7%, 66. 7%, 75%, 83. 6%, 38. 8% respectively. The next best classifier is XGB classifier having holdout accuracy as 80. 5% and cross validation score of 76. 1%. The Area under ROC, Precision, Recall, F1 score, Average Precision and log loss of XGB classifier are 88. 6, 75, 66. 7, 70. 6, 83. 3 and 39. 0 respectively. If we compare Gradient Boosting Classifier in AI and Auto AI technique 6. 42% increase can be observed in accuracy in Auto AI technique. Similarly in XG Boost classifier 2. % increase can be observed in Accuracy. However using AI techniques Logistic Regression (LR) model is performing well in accuracy i. e 84. 4%. A Comparison with Existing Works The Comparison is done with existing works both

Table 6.3: Optimized results after using grid Search

CLASSIFIER	ACCURACY	ROC – AUC
LR	84.4	88.98
KNN	75.32	82.35
SVC	84.4	89.14
GAUSSIAN NB	78	87.48
BERNOULI NB	73	79.41
DECISION TREE	66	64.14
RANDOM FOREST	80.51	85.89
EXTRA TREES	75.32	82.5
ADABOOST	79.22	87.4
GRADIENT BOOST	77	81.82
LIGHT GBM	74	83
XG BOOST	77.92	84.31
GRADIENT BOOSTING (AUTO AI)	84.4	90.3
XGB CLASSIFIER (AUTO AI)	80.5	76.1

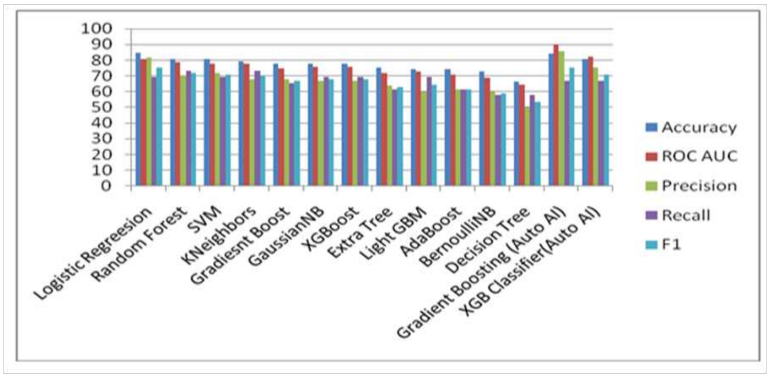

Fig. 6.1: Graphical Representation of AI and Auto AI Algorithms

AI and Auto AI in literature and results are consolidated in Table6.4 . The graphical representation of AI and auto AI algorithms are shown in Fig. 6.1.

7. Conclusions and Future Scope. The paper presents diabetes mellitus prediction using different approaches of Artificial Intelligence. The endeavor to do this prediction employing both conventional methods and cutting-edge technologies like IBM AutoAI, holds significant promise for giving early alerts and lifestyle changes for a better and healthy life. The deployment of advanced algorithms of Auto AI will enhance accuracy in predicting diabetes. The comparative analysis between Auto AI and conventional methods provides us with valuable inputs on the strengths and limitations of each approach but also explores the potential of automation in streamlining the predictive modeling process. Using Auto AI technique for Pima Indian diabetes data set,Gradient boosting classifier showed better performance with 84. 4% accuracy and conventional Machine learning techniques gave Logistic Regression,the better performance with 84. 4% accuracy. The findings of this research will contribute to the growing body of knowledge in healthcare analytics and machine learning applications. The insights gained are crucial for healthcare professionals and doctors in adopting proactive strategies for diabetes management. The emphasis on interpretability, validation across diverse datasets, and usability in realworld healthcare settings underscores the practicality and applicability of the developed models. The future scope of this research extends towards continual refinement and adaptation of predictive models. Further investigations could explore the integration of additional clinical parameters, genetic data, and lifestyle factors to enhance the predictive power of the models. Exploring the potential of explainable AI (XAI) techniques

Table 6.4: Comparison of proposed work with Existing works

Authors /papers	Algorithm	Accuracy	Precision	Recall	F1score	ROC-AUC
Victor chang et al.[4]	J48 decision tree	75.65	70.86	60.36	79.26	80.84
	NB	77.83	81.25	86.09	83.6	84.1
	RF	73.91	80.79	79.74	80.26	81.77
shafi salliah et al. [14]	NB	74.28	75.7	76.1	75.8	81.7
	SVM	63.1	42.2	64.9	51.1	54.4
	DT	71.81	73.3	73.6	73.4	74.9
Ahmed Nazin et al.[1]	NB	78.95	72	68	70	87.98
	DT	76.32	68	68	68	74.15
	RF	80.25	79	62	64	86.96
	SVM	80.26	83	58	70	87.76
	LR	77.63	79	58	68	85.49
	GB	78.95	72	70	70	83.83
	KNN	75	70	60	63	78.65
Neha Prernax et al.[18]	LR	74.4	85.6	77.5	81.3	76.5
	KNN	70.8	83.2	74.8	78.7	81.5
	SVM	74.4	85.6	77.5	81.3	77.1
	NB	68.9	84	82	83	76
	DT	69.7	74.4	78.1	76.2	84.2
	RF	75	84	78.9	81.3	1
Yashi srivastava et al.[16]	LR Auto AI	77.8	75.6	45.9	57.1	77.8
Proposed Model without Auto AI	LR	84.4	81.8	69.2	75	80.6
	RF	80.5	70.3	73	71.6	78.6
	SVM	80.5	72	69.2	70.5	77.7
	K Neighbors	79.2	67.8	73	70.3	77.7
	Gradient Boost	77.9	68	65.3	66.6	74.8
	Gaussian NB	77.9	66.6	69.2	67.9	75.7
	XG Boost	77.9	66.6	69.2	67.9	75.7
	Extra Tree	75.3	64	61.5	62.7	71.9
	Light GBM	74	60	69.2	64.2	72.8
	Ada Boost	74	61.53	61.5	61.5	70.9
	BernoulliNB	72.7	60	57.6	58.8	69
	Decision Tree	66.2	50	57.6	53.5	64.1
Proposed Model using IBM Auto AI	Gradient Boosting (Auto AI)	84.4	85.7	66.7	75	90.3
	XGB Classifier (Auto AI)	80.5	88.6	66.7	70.6	88.6

could contribute to improve the interpretability of machine learning predictions, fostering better collaboration between algorithms and healthcare professionals. Additionally, the generalization and scalability of the models across diverse populations are also to be focussed. The customization of models should be done by focussing on specific demographic groups, considering variations in healthcare practices, genetic predispositions, and also environmental factors. The collaboration between data scientists, healthcare experts, and technology developers will play a key role in ensuring that predictive models not only meet rigorous standards but also align with the dynamic landscape of healthcare advancements. As Health care analytics is the ever evolving field , the integration of machine learning for diabetes prediction is very important. The continuous developments in innovation and practical implementation will pave the way for diabetes management and prevention in future.

REFERENCES

[1] N. Ahmed, R. Ahammed, M. M. Islam, M. A. Uddin, A. Akhter, M. A. Talukder, and B. K. Paul, *Machine learning based diabetes prediction and development of smart web application*, International Journal of Cognitive Computing in Engineering, 2 (2021), pp. 229–241.

[2] M. Borgman and E. McErlean, *What is the metabolic syndrome?: Prediabetes and cardiovascular risk*, Journal of Cardiovascular Nursing, 21 (2006), pp. 285–290.

[3] J. Chaki, S. T. Ganesh, S. Cidham, and S. A. Theertan, *Machine learning and artificial intelligence based diabetes mellitus detection and self-management: A systematic review*, Journal of King Saud University-Computer and Information Sciences, 34 (2022), pp. 3204–3225.

[4] V. Chang, J. Bailey, Q. A. Xu, and Z. Sun, *Pima indians diabetes mellitus classification based on machine learning (ml) algorithms*, Neural Computing and Applications, 35 (2023), pp. 16157–16173.

[5] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, *Machine learning and data mining methods in diabetes research*, Computational and structural biotechnology journal, 15 (2017), pp. 104–116.

[6] K. Meenakshi, A. Adepu, V. V. T. Nagandla, and S. Agarwal, *A machine learning based melanoma skin cancer using hybrid texture features*, in 2023 3rd International Conference on Intelligent Technologies (CONIT), IEEE, 2023, pp. 1–5.

[7] K. Meenakshi and D. Kishore, *Parkinson's disease classification using machine learning techniques*, in 2023 International Conference on Computer Communication and Informatics (ICCCI), 2023, pp. 1–5.

[8] G. NagaJyothi and S. SriDevi, *Distributed arithmetic architectures for fir filters-a comparative review*, in 2017 International conference on wireless communications, signal processing and networking (WiSPNET), IEEE, 2017, pp. 2684–2690.

[9] G. NagaJyothi and S. Sridevi, *High speed and low area decision feed-back equalizer with novel memory less distributed arithmetic filter*, Multimedia Tools and Applications, 78 (2019), pp. 32679–32693.

[10] N. Nai-Arun and R. Moungmai, *Comparison of classifiers for the risk of diabetes prediction*, Procedia Computer Science, 69 (2015), pp. 132–142.

[11] R. Natarajan, G. H. Lokesh, F. Flammini, A. Premkumar, V. K. Venkatesan, and S. K. Gupta, *A novel framework on security and energy enhancement based on internet of medical things for healthcare* 5.0, Infrastructures, 8 (2023), p. 22.

[12] D. M. Nathan, *Long-term complications of diabetes mellitus*, New England journal of medicine, 328 (1993), pp. 1676–1685.

[13] K. Polat, S. Güneş, and A. Arslan, *A cascade learning system for classification of diabetes disease: Generalized discriminant analysis and least square support vector machine*, Expert systems with applications, 34 (2008), pp. 482–487.

[14] S. Shafi and G. A. Ansari, *Early prediction of diabetes disease & classification of algorithms using machine learning approach*, in Proceedings of the International Conference on Smart Data Intelligence (ICSMDI 2021), 2021.

[15] J. W. Smith, J. E. Everhart, W. Dickson, W. C. Knowler, and R. S. Johannes, *Using the adap learning algorithm to forecast the onset of diabetes mellitus*, in Proceedings of the annual symposium on computer application in medical care, American Medical Informatics Association, 1988, p. 261.

[16] Y. Srivastava, P. Khanna, and S. Kumar, *Estimation of gestational diabetes mellitus using azure ai services*, in 2019 Amity International Conference on Artificial Intelligence (AICAI), IEEE, 2019, pp. 321–326.

[17] I. Tasin, T. U. Nabil, S. Islam, and R. Khan, *Diabetes prediction using machine learning and explainable ai techniques*, Healthcare Technology Letters, 10 (2023), pp. 1–10.

[18] N. P. Tigga and S. Garg, *Prediction of type 2 diabetes using machine learning classification methods*, Procedia Computer Science, 167 (2020), pp. 706–716.

[19] W. L. Yun, U. R. Acharya, Y. V. Venkatesh, C. Chee, L. C. Min, and E. Y. K. Ng, *Identification of different stages of diabetic retinopathy using retinal optical images*, Information sciences, 178 (2008), pp. 106–121.

[20] K. Zarkogianni, E. Litsa, K. Mitsis, P.-Y. Wu, C. D. Kaddi, C.-W. Cheng, M. D. Wang, and K. S. Nikita, *A review of emerging technologies for the management of diabetes mellitus*, IEEE Transactions on Biomedical Engineering, 62 (2015), pp. 2735–2749.

Edited by: Dhilip Kumar V
Special issue on: Unleashing the Power of Edge AI for Scalable Image and Video Processing
Received: Nov 20, 2023
Accepted: Mar 13, 2024