**Chapter 1 Introduction**

**1.1 Introduction**

Diabetes is an emerging global health concern. According to the International Diabetes Federation, the comparative prevalence of diabetes is likely to increase to 7.3% by 2025 (Liu, 2007). The symptoms of diabetes can be referred to as three Ps. Polyuria (i.e., increased need to urinate), Polydipsia (i.e., thirst) and, Polyphagia (i.e., uncontrolled blood sugar). A diabetic patient may have one or more symptoms altogether.

Early prediction of diabetes may help in better management of the disease and may lead the patients to undergo a suitable treatment and save lives. Machine learning (ML) is an emerging scientific field of data science dealing with the ways in which machines learn from experience and has helped to make an early diagnosis of diabetes by learning the symptoms or features of diabetic patients (Fuchsberger, 2016). In the prior research study, early detection of diabetes has been done with high prediction accuracy. Many researchers have proposed several predictive diabetic models. But the models they have proposed are black box type and not interpretable. A robust interpretable model representing a good correlation with cardinal features of diabetes can provide more confidence to medical practitioners and patients in the model prognostics.

Since the number of wrong predictions is very small, many researchers overlook the effect of the wrong prediction on humans. The wrong prediction of a diabetic patient as a non-diabetic patient may increase his/her health risk. At the same time prediction of a non-diabetic patient as diabetic may cause him/her to undergo unnecessary treatment which also increases health risks and waste of money. Therefore, it is very crucial to focus on the explainability of an ML model to build trust in its decision (Alghamdi, 2017).

Explainable artificial intelligence (XAI) is a set of processes and methods that allows human users to comprehend and trust the results and output created by machine learning algorithms. Explainable AI is used to describe an AI model and its expected impact. Various XAI tools were used to increase the explainability of our neural network model. SHAP is a unified approach of feature importance with desirable properties, and it provides both explanations for the structure of the model and for a specific prediction. SHAP values were used to provide a direct measure of the influence of patient variables on the actual predictions under the interaction with other variables. LIME is used for explaining model outputs and locally approximating the selected model with an interpretable one. The interpretable models are trained on small samples of the original observation. Thus, they provide an excellent local approximation. ELI5 does this by showing weights for each feature depicting how influential it might have been in contributing to the final prediction decision.

Therefore, the main contribution made in this thesis is designing an explainable and interpretable diabetes prediction model that will predict diabetes in patients and at the same time explain the reason behind the decision of the model.

**1.2 Objective**

* To develop an explainable AI model that predicts diabetes

**1.3 Justification of Study**

Many diabetic patients are suffering due to late diagnosis where the treatment only starts after the severity of diabetes reaches to a chronic level. But, if the prognosis could be done in an early predictable manner, then the preventive measures could have easily been taken. As we know, prevention is better than cure. So, the main motivation of our study is to build a system which could make an early prediction of diabetes so that, preventive measures can be taken and a healthy way of life can be maintained.

**1.4 Scope of study**

We can implement this study in our modern medical sectors. By implementing this model, early predictions can be made to detect diabetes. Hence, patients can take precautionary measures as well as early treatment before it becomes chronic or uncontrollable.

Concepts of this study can also be applied to arthritis, cancer, and other prolonged chronic diseases where early prediction can provide a cure or, reverse the condition.

**Chapter 2 Literature Review**

**2.1 Background study**

Early prognosis of diabetes would help in patient's treatment. Previous studies didn’t provide explanations for their predictive models. In this model, expandability and interpretability are also much emphasized. Two diabetes datasets from the UCI repository, which were collected from regular patients, are used. A wide variety of popular classification algorithms are used to build a diabetes prediction model. The preferred Machine learning techniques are incredibly useful for a diabetic diagnosis. After that, various explainable tools to interpret machine learning algorithms are used to examine and explain the unusual data behavior. Shapely plots explain the extent of attribute impact in diabetes prediction in a positive or negative direction. Various tools were used to interpret the model outcome, like Local Interpretable Model Agnostic Explanations (LIME), Shapley Additive Explanations (SHAP), tree-based model feature importance, partial dependence plots. LIME plots explain contrasting risk scores for a couple of diabetes patients with contributing features. Logistic regression performs decently, which suggests a linear separation between the classes in this dataset. For the cases of more complex and non-linear datasets, linear-based algorithms may not be ideal for classification tasks. Partial dependence plots and decision tree explain the directional and quantitative impact of top contributions attributes.

**2.2 Related Works**

In ‘Diabetes Prognosis Using White Box Machine Learnin Frame Work for Interpretability of Results’, the authors provided an insight into the feasibility and importance of explainable artificial intelligence solutions for the healthcare sector. A case-study on diabetes in Pima Indian females aids this research motive. The study has maintained good explainability of the predictions and high accuracy by the machine learning models used. This study used a white-box machine learning framework, local interpretable model-agnostic explanations, to prove the cause. The framework successfully interpreted case-by-case predictions of some machine learning models. (Khan, Meehan, Kevin, 2021)

In ‘Diabetes Disease Prediction Using Data Mining’ the authors concentrated on reacnhing medical conclusion by learning design through the gathered data of diabetes and to create smart therapeutic choice emotionally supportive network to help the physicians. The primary target of this examination was to assemble intelligent diabetes disease prediction system that gives analysis of diabetes malady utilizing diabetes patient's database. In this system, the authors applied algorithms like Bayesian and KNN (K-Nearest Neighbor) to apply on diabetes patient's database and analyze them by taking various attributes of diabetes for prediction of diabetes disease. (Shetty, Rit, Shaikh, and Patil, 2017)

The authors of ‘Diabetes Prediction Using Different Machine Learning Algorithms’ focused on predicting the occurrence of diabetes by applying different classification methods (e.g., random forest and K-NN). This paper utilized pima Indian diabetes dataset. The authors used black box type model and didn’t explain the reason behind their model’s decision. (Reddy, Krishnaveni, Nikitha, and Vijaykanth, 2021)

A disease index grading method based on logistic regression algorithm was proposed for diabetes prediction in ‘Prediction of Score of Diabetes Progression Index Based on Logistic Regression Algorithm’. According to the score of disease progression index, the results were divided into two categories, namely, the index score was greater than or equal to 150 and less than 150, and the problem of which interval the target value will belong to, was well applicable to the logistic regression model. Compared with using linear regression algorithm to predict the impact of a feature on the progression of diabetes, logistic regression algorithm was used. This model lacks explainability and interpretability. (Lei, 2020)

Three models based on neural network for the classification and prediction of diabetes was presented. These models include a feedforward network, a pattern network, and a cascade forward architecture. The performance of the three models were compared in terms of accuracy, sensitivity, and specificity. All models were implemented and tested in MATLAB. (Diab, Husain, and Jarndal, 2020)

Study focused on the performance of a machine-learning (ML) algorithm to identify the presence of diabetes on the PIMA Indian diabetes dataset (PIDD) which referenced from the University of California, Irvine (UCI) ML repository. This research paper worked on the prediction technique for diabetes classification with outliers and missing values in data with class imbalance. Using an adaptive synthetic sampling method (ADASYN) and reduced the impact of class imbalance on the performance of the prediction model. Then, this algorithm improved the generalization using a feature selection technique and multilayer perceptron classifiers to make predictions and evaluations. (Yadav, Maravi, Agrawal, and Mishra, 2021)

A comparison among the results experimentally obtained, and using three machine learning algorithms in the prediction of diabetes was presented. The three considered algorithms are support vector machine, Naive Bayes, and random forest. The aim of the paper was to analyze the performance of the algorithms considering different metrics in order to compare different techniques to obtain better accuracy. (Costea, Moisi, and Popescu, 2021)

Predictding diabetes onset: an ensemble supervised learning approach was presented and five widely used classifiers were used for the ensembles and a meta-classifier was used to aggregate their outputs. The results were presented and compared with similar studies that used the same dataset within the literature. It was shown that by using the proposed method, diabetes onset prediction can be done with higher accuracy. (Nnamoko, Hussain, and England, 2018)

‘Diabetes Prediction Using Machine Learning Techniques’, where the aim was to predict diabetes via three different supervised machine learning methods including: SVM, Logistic regression, KNN. This project proposed an effective technique for earlier detection of diabetes. The authors didn’t use ML techniques which could improve the model’s performance. Also, the model lacks explainability. (Joshi, and Chawan, 2018)

In ‘Implementation of Diabetes Incidence Prediction Using a Multilayer Perceptron Neural Network’, the authors have presented a diabetes incidence prediction system using a multilayer perceptron neural network that allows users to enter simple data and increases the convenience and accessibility of the prediction capabilities. The proposed system was evaluated on Pima Indians Diabetes and showed a good performance in predicting diabetes incidences. (Song, and Lee, 2021)

**Chapter 3 Methodology**

**3.1 Methodology**

Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy. The ultimate goal of machine learning is to design algorithms that automatically help a system gather data and use that data to learn more.

The basic machine learning steps that we have followed are-

* Collecting Data: The first step of ML is to find and collect reliable data. The quality of the data will impact the accuracy of the outcome. That’s why data needs to be collected from a reliable source.
* Preparing the Data: This requires cleaning the data or removing unwanted data, missing values and duplicate values, visualizing the data, splitting the cleaned data into two sets- a training set and a testing set.
* Choosing a Model: A machine learning model determines the output after running a machine learning algorithm on the collected data.
* Training the Model: In training, the prepared data is passed to the machine learning model to find patterns and make predictions.
* Evaluating the Model: After training the model, model evaluation is done to check the performance. This is done by testing the performance of the model on the test data.
* Parameter Tuning: Parameters are the variables in the model. Parameter tuning is done to check if the accuracy can be improved in any way by tuning parameters present in the model.
* Making Predictions: In the end, the model is used to make required predictions.

The strategy that we have followed in completing our study is shown in the following figure-

Trained Model

Data

Collection

Preprocess

Test Data

Train Data

Built Model

Feed

Prediction

Eplain with XAI Tools

Input

Fig 3.1: Methodology of Machine Learning Model

**3.2 Description of Methodology**

**3.2.1 Data Preprocessing**

Data preprocessing is one of the important processes. Mostly healthcare related data contains missing value and other impurities that can cause effectiveness of data. To improve the quality and effectiveness, data preprocessing is done. To use Machine Learning Techniques on the dataset effectively this process is essential for accurate result and successful prediction. For our diabetes dataset, we have performed pre processing in two steps.

**3.2.2 Missing Values Removal**

We have removed all the instances that have zero (0) as worth. Having zero as worth is not possible. Therefore, this instance is eliminated. Through eliminating irrelevant features/instances we make feature subset and this process is called features subset selection, which reduces diamentonality of data and help to work faster.

**3.2.3 Splitting of Data**

After cleaning the data, data is normalized in training and testing the model. When data is spitted then we train algorithm on the training data set and keep test data set aside. This training process will produce the training model based on logic and algorithms and values of the feature in training data. Basically, aim of normalization is to bring all the attributes under same scale.

**3.2.4 Apply Machine Learning Algorithms**

When data has been ready, we apply Machine Learning techniques. We use different classification and ensemble techniques, to predict diabetes. The methods have been applied on our diabetes datasets. Main objective to apply Machine Learning Techniques is to analyze the performance of these methods and find accuracy of them, and also be able to figure out the responsible/ important feature which play a major role in prediction. The Techniques are follows-

**3.2.5 Explain with AI Tools**

Explainable artificial intelligence (XAI) is a set of processes and methods that allows human users to comprehend and trust the results and output created by machine learning algorithms. Explainable AI is used to describe an AI model and its expected impact. XAI tools and frameworks help us understand and interpret predictions made by our machine learning models. XAI can debug and improve model performance, and help others to understand the models' behavior. The three XAI tools that we have used are SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-agnostic Explanations) and ELI5 (Explain Like I’m 5) which are later explained in chapter 5.

**Chapter 4 Requirement Analysis & Model Description**

**4.1 Dataset**

We have utilized two different datasets to implement the thesis.

**4.1.1** **Dataset I** - **Pima Indian Dataset**

This dataset is collected from UCI repository which is named as Pima Indian Diabetes Dataset. The dataset has many attributes of 768 patients. This dataset has data of 768 patients among which all are female.

|  |  |
| --- | --- |
| **Features** | **Description** |
| Pregnancy | Number of Times pregnant |
| Glucose | 2 hours Plasma glucose concentration, in an oral glucose tolerance test |
| Blood Pressure | Diastolic blood pressure (mm Hg) |
| Skin Thickness | Triceps skin fold thickness (mm) |
| Insulin | 2-Hour serum insulin (mu U/ml) |
| BMI (Body Mass Index) | Body mass index (weight in kg/height in m2) |
| Diabetes Pedigree Function | diabetes mellitus history in relatives |
| Age | Age (years) |

Table 4.1: List of features in Dataset I

**4.1.2 Dataset II** - **Diabetes Dataset of Sylhet**

This dataset is prepared by documenting the patient’s data from Sylhet Diabetes Hospital, Bangladesh. This dataset is also collected from the UCI repository. This dataset has patient’s critical attributes which are helpful for diabetes prediction. This diabetes dataset has 15 features which are Sex, Polyuria, Polydipsia, Sudden Weight Loss, Weakness, Polyphagia, Genital Thrush, Visual Blurring, Itching, Irritability, Delayed Healing, Partial Paresis, Muscle Stiffness and Alopacia. This dataset contains the data of 520 patients among which 328 are male patients and 192 are female patients.

|  |  |
| --- | --- |
| **Features** | **Description** |
| Sex | Male or, female |
| Ployuria | Excessive urination |
| Polydipsia | Excessive thirst |
| Sudden Weight Loss | Loss in body weight |
| Weakness | Feeling fatige or, lethargy |
| Polyphagia | Excessive hunger |
| Genital Thrush | Yeast infection in genitalia |
| Visual Blurring | Diffuculty in focusing eyesight |
| Itching | Irritating sensation on the skin |
| Irritability | Feeling restless and insomnia |
| Delayed Healing | Abnormal delay in wound recovery |
| Partial Paresis | Weakening of a group of muscle |
| Muscle Stiffness | Difficulty in moving due to stiff muscles |
| Alopecia | Abnormal hair loss |
| Obesity | Excessive amount of body fat |

Table 4.2: List of features of Dataset II

**4.2 Language & Tool Requirement**

**4.2.1 Python**

Python is a high-level, interpreted, interactive and object-oriented scripting language. Its design philosophy emphasizes [code readability](https://en.wikipedia.org/wiki/Code_readability) with the use of [significant indentation](https://en.wikipedia.org/wiki/Off-side_rule). It is often described as a "batteries included" language due to its comprehensive [standard library](https://en.wikipedia.org/wiki/Standard_library). Various Modules and packages are supported by Python which fosters program modularity and code reuse. The Pythonb interpreter and its substantial standard library are free to download and distribute in source or binarky form for all major platforms.

**4.2.2 Jupyter Notebook**

The Jupyter Notebook is a free, open-source web application that allows data scientists to create and share documents that include live code, equations, computational output, visualizations and other multitude of features as well as explanatory text. The Jupyter Notebook App is a server-client application that allows editing and running [notebook documents](https://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/what_is_jupyter.html#notebook-document) via a web browser. The Jupyter Notebook App can be executed on a local desktop requiring no internet access (as described in this document) or can be installed on a remote server and accessed through the internet.

**4.3 Machine Learning Algorithms**

**4.3.1 Logistic Regression**

Logistic regression is a powerful supervised ML algorithm used for binary classification problems. It is a process of modeling the probability of a discrete outcome given an input variable. The most common logistic regression models a binary outcome; something that can take two values such as true/false, yes/no, and so on. Logistic regression fits an S shaped logistic function. The curve goes from 0 to 1. That means the curve shows the probability whether our prediction will be true or, false.

**4.3.2 Random Forest**

Random forests or random decision forests is an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. Random forest is so called because each tree in the forest is built by randomly selecting a sample of data. It can be used for both regression and classification. It is based on ensemble learning which is a method of merging several classifiers to solve a difficult problem and increase the model’s performance. Random forest uses a number of decision trees on different subsets of a given dataset and takes the average to increase the dataset’s projected accuracy. Instead of relying on a single decision tree, the random forest collects forecast from each tree and predicts the final output based on the tree’s majority votes.

**4.3.3 Decision Tree**

The decision tree algorithm is a method for solving both regreesion and classification. The primary goal of constructing a decision tree is to create a training model or, training set. A decision tree works as a decision support tool that uses a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements.

**4.3.4 XgBoost Classifier**

[XGBoost](https://xgboost.ai/), which stands for Extreme Gradient Boosting, is a scalable, distributed [gradient-boosted](https://en.wikipedia.org/wiki/Gradient_boosting) decision tree (GBDT) machine learning library. It provides parallel tree boosting and is the leading machine learning library for regression, classification, and ranking problems. XGBoost classifier is a ML algorithm that is applied for structured and tabular data. XGBoost is an implementation of gradient boosted decision trees designed for speed and performance.

**4.3.5 Support Vector Machine**

Support-vector machines are supervised learning models with associated learning algorithms that analyze data for classification and regression analysis. SVM works by mapping data to a high-dimensional feature space so that data points can be categorized, even when the data are not otherwise linearly separable. Each object we want to classify is represented as a point in an n-dimensional space and the coordinates of this point are usually called features. SVM performs classification test by drawing a hyperplane that is a line in 2D or, a plane in 3D, in such a way that all points of one category are in one side of the hyperplane and all points of the other category are on the other side. And while there could be multiple such hyperplanes, SVM tries to fine the one that best separates the two categories in the sense that it maximizes in distance to points in either category. This distance is called the margin and the points that fall exactly on the margin are called the supporting vectors. To find this hyperplane, SVM requires a training set or, a set of points that are already labeled with the correct category.

**4.4 Explainable AI Tools**

Explainable artificial intelligence (XAI) is a set of processes and methods that allows human users to comprehend and trust the results and output created by machine learning algorithms. Explainable AI is used to describe an AI model and its expected impact. XAI tools and frameworks help us understand and interpret predictions made by our machine learning models. XAI can debug and improve model performance, and help others to understand the models' behavior. The three XAI tools that we have used are SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-agnostic Explanations) and ELI5 (Explain Like I’m 5) which are later explained in chapter 5.

**4.5 Description of the Proposed Model**

We have collected our datasets from University of Canada, Irvine ML Repository. Then we have preprocessed our dataset by removing all the null values and replacing them with mean values. After data has been cleaned and preprocessed, we have splitted our data into train data (80%) and test data (20%). We have trained the ML models using the train dataset. For this purpose, we have used Logistic Regression, Random Forest, XgBoost Classifier, Decision Tree and Support Vector Machine. Then we have used our test dataset to evaluate our trained model. The evaluation metrices based on which we have compared our ML models are precision, recall, F-1 score and accuracy.

We have used three explainabe AI tools to explain the predictive result of our ML model. Each of the tools explaing in their own uniqe way, enhancing the interpretability and explainability of our model.

**Chapter 5 Result and Analysis**

**5.1 Confusion Matrix**

**True Positive (TP):** True positive measures the extent to which the model correctly predicts the positive class. That is, the model predicts that the instance is positive, and the instance is actually positive. True positives are relevant when we want to know how many positives our model correctly predicts.

**False Positive (FP):** False positives occur when the model predicts that an instance belongs to a class that it actually does not. False positives can be problematic because they can lead to incorrect decision-making.

**True Negative (TN):** True negatives are the outcomes that the model correctly predicts as negative. For example, if the model is predicting whether or not a person has a disease, a true negative would be when the model predicts that the person does not have the disease and they actually don’t have the disease.

**False Negative (FN):** A false negative occurs when a model predicts an instance as negative when it is actually positive. False negatives can be very costly, especially in the field of medicine. For example, if a cancer screening test predicts that a patient does not have cancer when they actually do, this could lead to the disease progressing without treatment.

**5.1.1 Confusion Matrix for Dataset I**

|  |  |
| --- | --- |
| Logistic Regression | |
| TP  88 | FN  9 |
| FP  23  Fig5.1: Confusion Matrix for LR | TN  34 |

|  |  |
| --- | --- |
| Random Forest | |
| TP  87 | FN  10 |
| FP  20 | TN  37 |

|  |  |
| --- | --- |
| XgBoost Classifier | |
| TP  81 | FN  16 |
| FP  24 | TN  33 |

Fig5.3: Confusion Matrix for XgBoost Classifier

Fig5.2: Confusion Matrix for RF

|  |  |
| --- | --- |
| Decision Tree | |
| TP  80 | FN  17 |
| FP  23 | TN  34 |

|  |  |
| --- | --- |
| Support Vector Machine | |
| TP  91 | FN  6 |
| FP  29 | TN  28 |

Fig5.5: Confusion Matrix for SVM

Fig5.4: Confusion Matrix for DT

**5.1.2 Confusion Matrix for Dataset II**

|  |  |
| --- | --- |
| Logistic Regression | |
| TP  55 | FN  10 |
| FP  6 | TN  85 |

|  |  |
| --- | --- |
| Random Forest | |
| TP  61 | FN  4 |
| FP  3 | TN  88 |

|  |  |
| --- | --- |
| XgBoost Classifier | |
| TP  60 | FN  5 |
| FP  3 | TN  88 |

Fig5.8: Confusion Matrix for XgBoost Classifier

Fig5.7: Confusion Matrix for RF

Fig5.6: Confusion Matrix for LR

|  |  |
| --- | --- |
| Decision Tree | |
| TP  62 | FN  3 |
| FP  6 | TN  85 |

|  |  |
| --- | --- |
| Support Vector Machine | |
| TP  44 | FN  12 |
| FP  9 | TN  91 |

Fig5.10: Confusion Matrix for SVM

Fig5.9: Confusion Matrix for DT

**5.2 Evaluation Metrices**

**Precision:** Precision score measures the proportion of positively predicted labels that are actually correct. Precision is also known as the positive predictive value. Precision Score = TP / (FP + TP).

**Recall:** Recall score represents the model’s ability to correctly predict the positives out of actual positives. This is unlike precision which measures how many predictions made by models are actually positive out of all positive predictions made. Recall Score = TP / (FN + TP).

**Accuracy:** Accuracy is a machine learning model performance metric that is defined as the ratio of true positives and true negatives to all positive and negative observations. In other words, accuracy tells us how often we can expect our machine learning model will correctly predict an outcome out of the total number of times it made predictions.

Accuracy Score = (TP + TN)/ (TP + FN + TN + FP).

**F1 Score:** F1 score represents the model score as a function of precision and recall score. F-score is a machine learning model performance metric that gives equal weight to both the Precision and Recall for measuring its performance in terms of accuracy, making it an alternative to Accuracy metrics.

F1 Score = 2\* Precision Score \* Recall Score/ (Precision Score + Recall Score).

We have achieved the following result using the machine learning algorithms on our datasets:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Evaluation Metric** | **Logistic Regression** | **Random Forest** | **Decision Tree** | **XgBoost Classifier** | **SVM** |
| **Precision** | 0.79 | 0.81 | 0.78 | 0.77 | 0.76 |
| **Recall** | 0.91 | 0.90 | 0.82 | 0.84 | 0.94 |
| **F-1 Score** | 0.85 | 0.85 | 0.80 | 0.80 | 0.84 |
| **Accuracy** | 0.79 | 0.81 | 0.74 | 0.74 | 0.77 |

Table 5.11: Evaluation Metrics for Dataset I

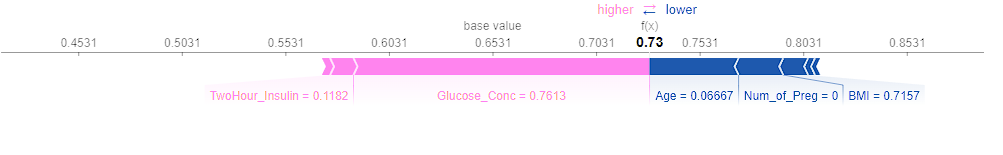
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Evaluation Metric** | **Logistic Regression** | **Random Forest** | **Decision Tree** | **XgBoost Classifier** | **SVM** |
| **Precision** | 0.90 | 0.95 | 0.91 | 0.95 | 0.58 |
| **Recall** | 0.85 | 0.94 | 0.95 | 0.92 | 1.00 |
| **F-1 Score** | 0.87 | 0.95 | 0.93 | 0.94 | 0.74 |
| **Accuracy** | 0.90 | 0.96 | 0.94 | 0.95 | 0.58 |

Table 5.12: Evaluation Metrics for Dataset II

**5.2 Interpretation with XAI Tools:**

**5.2.1 SHAP**

SHAP (SHapley Additive exPlanations) is a game theoretic approach to explain the output of any machine learning model. SHAP values are used to provide a direct measure of the influence of patient variables on the actual predictions under the interaction with other variables

Dataset I:

Dataset II:



Fig 5.3: SHAP plot- patient with high-risk score

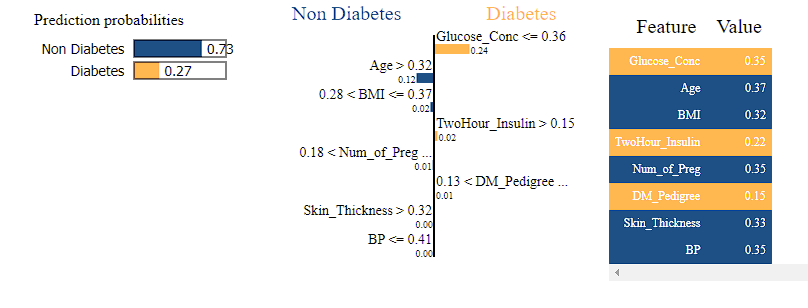
From the above figures we can see which features are mostly responsible for the prediction. For dataset1 the base line value is 0.73 which means the instance of data we have chosen for the plot, it has 73% possibility of being diabetes positive. The features that pushed probability towards the baseline or, the features that are mostly responsible for the prediction are glucose concentration and two-hour insulin level. On the other hand, the features that are responsible for pushing the prediction away from the baseline are age, number of pregnency and BMI.

And dataset 2, the baseline value of the prediction for chosen instance is 0.81. Features responsible for the prediction are polyuria, obesity and age. On the other hand, polydipsia, alopecia and gender are responsible for a negative prediction in the selected instance.

**5.2.2 LIME**

LIME can give explanations for any given supervised learning model. It explains how much each individual features contribute to the final decision. LIME gives explanations that are locally faithful within the surroundings or vicinity of the observation/sample being explained.

Dataset I:

Dataset II:

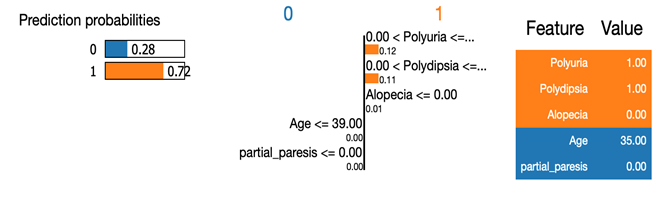


Fig 5.4: LIME plot- Explaining feature contribution to patients with high and low risk probability

The above figures show the risk probability of diabetes for the patient. The figure for the chosen sample of dataset 1 showing that this patient has 73% chance of being diagnosed as non-diabetic. The features that contributed to the prediction are age, BMI, number of pregnency, skin thickness and BP. And the features which could be of concern are two-hour insulin level and DM pedigree (a synthesis of the diabetes mellitus history in relatives and the genetic relationship of those relatives to the subject).

**5.2.3 ELI5**

ELI5 is a python package that is used to inspect ML classifiers and explain their predictions. ELI5 shows weights for each feature depicting how influential it might have been in contributing to the final prediction decision.

Dataset I: Dataset II:

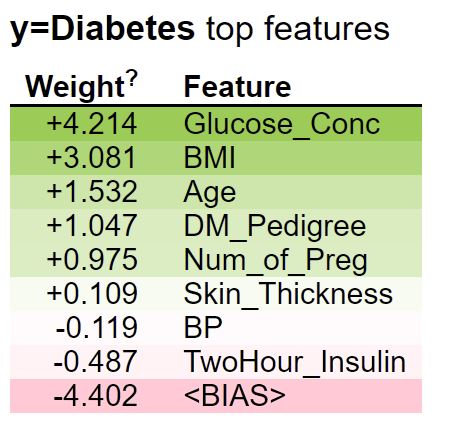
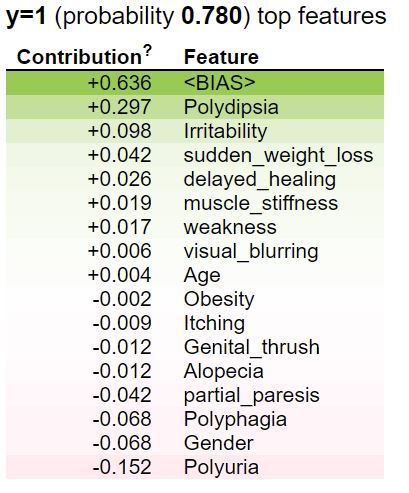


Fig 5.5: ELI5 – Feature Importance Plot

Above figures shows the the probability of the selected sample and showing the weights of the features which are responsible for the prediction.

**Chapter 6 Conclusion**

**6.1 Conslusion**

The main aim of our work was to design and implement diabetes prediction using Machine learning methods as well as explanation of those methods and it has been achieved successfully. The proposed approach uses various classification and ensemble learning method in which SVM, Knn, Random Forest, Decision Tree, Logistic Regression and Gradient Boosting classifiers are used.

**6.2 Limitations**

In case of medical prediction, the higher the accuracy, the better the reliability. Machine learning methods always give some false positive and false negative values in their prediction. More trusted and reliable models could be built if we could explain the reason behind these false values.

**6.2 Future Work**

Attributes such as physical activity, ancestral health condition, body weight, and energy levels, can be considered in the future to diagnose diabetes. Our model can be applied to early predict other prolonged chronic diseases such as cancer, heart diseases, arthritis etc. A mobile application can be developed for predicting and monitoring diabetes for new and old patients using our model.

**References:**

Alghamdi, M., Al-Mallah, M., Keteyian, S., Brawner, C., Ehrman, J. and Sakr, S., (2017), ‘Predicting diabetes mellitus using SMOTE and ensemble machine learning approach: The Henry Ford ExercIse Testing (FIT) Project’, PloS one, 12(7), p.e0179805.

Costea, N.E., Moisi, E.V. and Popescu, D.E., (2021, June), ‘Comparison of Machine Learning Algorithms for Prediction of Diabetes’. 2021 16th International Conference on Engineering of Modern Electric Systems (EMES) (pp. 1-4). IEEE.

Diab, M.S., Husain, S. and Jarndal, A., (2020, November), ‘Diabetes Classification and Prediction using Artificial Neural Networks’. 2020 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI) (pp. 1-5). IEEE.

Fuchsberger, C., Flannick, J., Teslovich, T.M., Mahajan, A., Agarwala, V., Gaulton, K.J., Ma, C., Fontanillas, P., Moutsianas, L., McCarthy, D.J. and Rivas, M.A., (2016), ‘The genetic architecture of type 2 diabetes. Nature’, 536(7614), pp.41-47.

Khan, P.F. and Meehan, K., (2021, January), ‘Diabetes prognosis using white-box machine learning framework for interpretability of results’, 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC) (pp. 1501-1506). IEEE.

Lei, L., (2020, July), ‘Prediction of Score of Diabetes Progression Index Based on Logistic Regression Algorithm’. 2020 International Conference on Virtual Reality and Intelligent Systems (ICVRIS) (pp. 954-956). IEEE.

Liu, Y., Michael, M.D., Kash, S., Bensch, W.R., Monia, B.P., Murray, S.F., Otto, K.A., Syed, S.K., Bhanot, S., Sloop, K.W. and Sullivan, J.M., (2007), ‘Deficiency of adiponectin receptor 2 reduces diet-induced insulin resistance but promotes type 2 diabetes. Endocrinology’, 148(2), pp.683-692.

Nnamoko, N., Hussain, A. and England, D., (2018, July), ‘Predicting diabetes onset: an ensemble supervised learning approach’, 2018 IEEE Congress on Evolutionary Computation (CEC) (pp. 1-7). IEEE.

Reddy, S.K., Krishnaveni, T., Nikitha, G. and Vijaykanth, E., (2021, September), Diabetes ‘Prediction Using Different Machine Learning Algorithms’, 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA) (pp. 1261-1265). IEEE.

Shetty, D., Rit, K., Shaikh, S. and Patil, N., (2017, March), ‘Diabetes disease prediction using data mining’, 2017 international conference on innovations in information, embedded and communication systems (ICIIECS) (pp. 1-5). IEEE.

Song, H. and Lee, S., (2021, December), ‘Implementation of Diabetes Incidence Prediction Using a Multilayer Perceptron Neural Network’, 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 3089-3091). IEEE.

Yadav, S., Maravi, Y.P., Agrawal, J. and Mishra, N., (2021, June). ‘A Neural Network based Diabetes Prediction on Imbalanced Data’, 2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT) (pp. 515-521). IEEE.