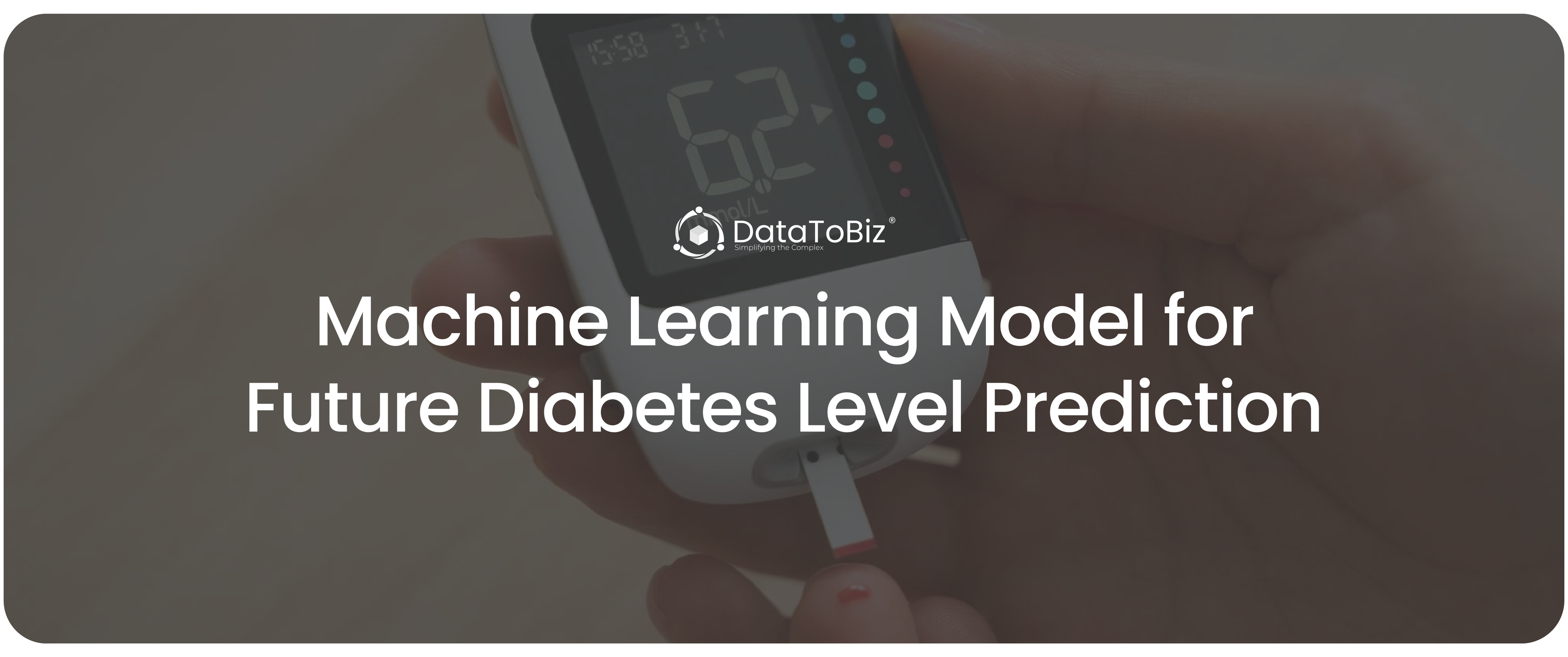


Machine Learning Model for Future Diabetes Level Prediction

Healthcare & Life Sciences

- Artificial Intelligence (AI)
- Data Warehousing
- Predictive Analytics

India

- IT and Technology Support

Agile Based Iterative Delivery

ABOUT CLIENT

- To determine the scope of the Hbalc level for diabetes, the Machine Learning model was applied with predictive analysis. It was performed with the aim to create future awareness and also reduce diabetes levels among individuals.

PROBLEM STATEMENT

ML Model Building Process:

- Problem understanding
- Statistical Analysis of data through Exploratory Data Analysis (EDA)
- Data cleaning/wrangling
- Model using ML Algorithms
- Model evaluation/deployment

Exploratory Data Analysis:

- An overview of all predictor variables whether it is numerical, categorical, or dependent target variables.
- Distributing data as per the target variables.
- Finding and describing the shape, head, and information.
- Analyzing the numerical and categorical features.
- Listing and filling the features of missing values.
- Transforming the log.
- Building a relation of all features to the target variables.
- Listing the numerical features and their coefficient correlation to the target.

SOLUTION

Overall Product Structure and Architecture

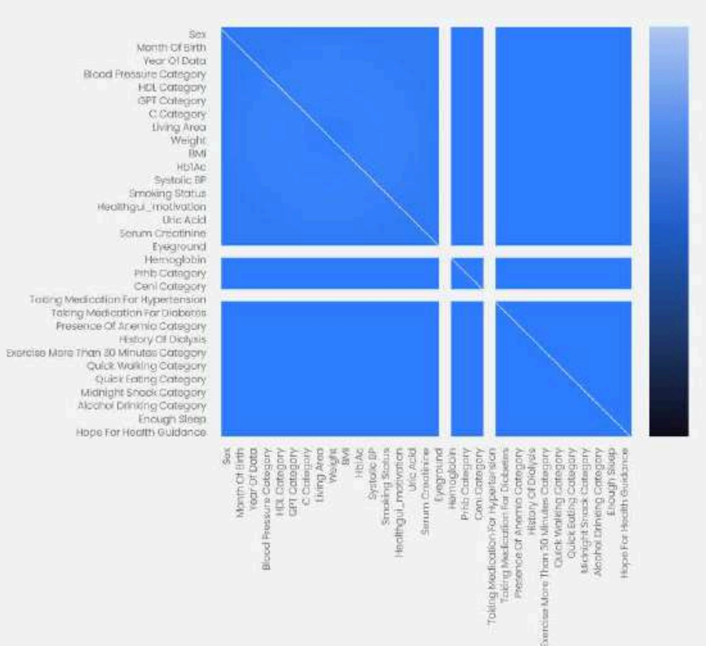
Above is the project's working flow diagram, which involves multiple steps to be followed, from data acquisition to deploying the files to the android device.

Modeling
Applying the machine learning model to generate a pipeline for the loan amount prediction. The algorithms used:

- Linear Regression
- KNeighbors Regressor
- Decision Tree Regressor
- SVR or Support Vector Regressor
- Random Forest Regressor
- XGB Regressor

Evaluation
The metrics used for evaluating predictions on regression machine learning are:

- Mean Absolute Error (MAE) – the sum of the absolute difference between predictions and actual values.
- Mean Squared Error (MSE) – the mean absolute error that provides a gross idea of the magnitude of the error
- R Squared or Root Mean Square Error – provides an indication of the goodness of a fitting set of predictions to their actual values.



Correlation Between The Features Applied To The Prediction Of The HbIAc

The diagram depicts the correlation between the input features like blood pressure, age, sex, weight, height etc are related to other features and between themselves.

TECHNICAL ARCHITECTURE

Sample Dataset Feature Inputs
We considered the human behavior attributes like smoking status, BMI, height, weight, age, hemoglobin, etc that played a vital role in the prediction of the Hb1Ac:

Dataset Feature Inputs Of The Sample Dataset

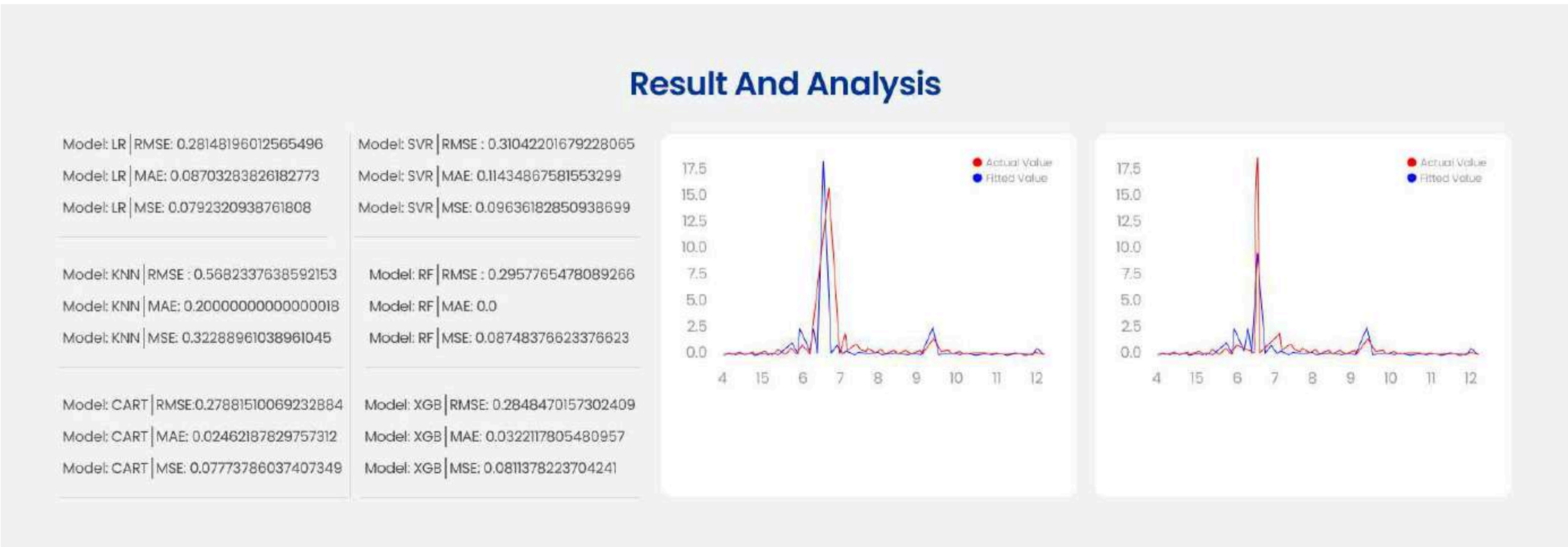
Considering the human behavior like smoking status, BMI, height, weight, age, Hemoglobin ete is playing vital role in the prediction of the HbIAc

Age	Living Area	Hight	Weight	Waist Circumstances (CM)	BMI	Triglyceride	HDL	LDL	Systolic BP	Diastolic BP	Smoking status	Healthgui active	Healthgui Motivation
68	2.0	147.5	48.0	77	22	101	64	133	130	88	0	0	0
69	2.0	147.5	48.5	75	21	111	58	120	130	80	0	0	0
70	2.0	147.5	51.3	82	22	85	61	133	150	80	0	0	0
73	2.0	148.2	50.7	83	23	205	57	146	148	79	0	0	0
74	2.0	147.0	48.0	82	23	164	56	134	127	74	0	0	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-
72	7.0	156.0	58	89	23	82	126	126	150	62	1	0	0
73	7.0	155.6	60	84	25	78	115	115	110	58	0	0	0
67	7.0	155.5	53	71	21	62	97	97	133	79	0	0	0
72	7.0	158.1	65	95	26	129	96	96	140	70	0	0	0
73	7.0	158.0	67	92	26	101	93	93	164	74	0	0	0

Deep Learning Approach
Leveraging the state-of-the-art Deep Learning algorithm MiME (Multi-level Medical Embedding) we implemented the EHR data records. MiME is a novel EHR embedding algorithm that takes into account the connections between diagnoses and corresponding treatments and it has already shown promising results if data is provided appropriately. The embedding included different levels of information, like:

- Visit Level
- Diagnostic Level
- Procedure level
- Medication level

Results and Analysis



BUSINESS IMPACT