COMP90049 Introduction to Machine Learning
Semester 2 2025
Assignment 2: Group project

Groupname: Phoenix, Group ID: G06
Xinyang Sun – 1783329 – xinyang.sun.3@student.unimelb.edu.au
Nina Lock – 1786109 – ninalock@student.unimelb.edu.au
Vadym Musiienko – 1785479 – vmusiienko@student.unimelb.edu.au

Abstract

This study aims to investigate the use of different machine learning models to predict the diabetes status, which can be either non-diabetic, prediabetic or diabetic, of individuals based on health and lifestyle indicators from the 2015 Behavioral Risk Factor Surveillance System (BRFSS) dataset. Here, four classification models were developed and compared with each other: K-Nearest Neighbors (KNN), Logistic Regression (LR), Decision Tree (DT) and a Multilayer Perceptron (MLP) Neural Network. The performances for this classification task are compared between the classical models and the Neural Network using different evaluation metrics, specifically F1-score.

Results show, that despite using oversampling and undersampling techniques, the prediabetes class remained the most challenging to predict across all models, reaching only very poor F1 scores most likely due to the dataset's imbalance. Feature importance analysis revealed that the features "HighBP" and "GenHlth" were listed as the most influential factors when predicting the diabetic status.

1 Introduction

According to the World Health Organization (WHO), diabetes is a chronic disease characterized by high blood glucose levels which over time can lead to serious damage. There were about 830 million people worldwide suffering from diabetes in 2022, however, more than half of that lived without proper treatment.
In general, there are two types of diabetes. Type 1 diabetes is defined by defective insulin production which therefore requires regular insulin administration every day. Type 2 diabetes prevents the body from effectively using insulin and is often linked to unhealthy lifestyles or genetics. It is therefore often preventable with help of early detection. Both types come with significantly higher risks of getting health problems such as heart attacks and kidney failure. (World Health Organization, 2024)
The Behavioral Risk Factor Surveillance System (BRFSS) is an annual health-related telephone survey conducted by the Centers for Disease Control and Prevention (CDC) in the United States. The 'Diabetes Health Indicators Dataset' which is used in this research is derived from the 2015 BRFSS. More specifically, the dataset contains three files, with this research focusing on the file 'diabetes_012_health_indicators_BRFSS2015.csv' which is a clean dataset consisting of 253,680 survey responses and 21 feature variables. Moreover, the target is divided into no diabetes (0), prediabetes (1) and diabetes (2).
The features represent a wide range of demographic (e.g age, sex, income), behavioral (e.g. physical activity, alcohol consumption), and health status indicators (e.g BMI, general health). (Teboul, 2021)

By utilising the dataset, this research addresses the following research questions:

- How accurately can individuals be classified as diabetic, prediabetic or non-diabetic using the selected BRFSS health survey data?
- Which health and behavioral indicators are the most predictive of diabetes?

This research aims to compare the predictive performance of classical machine learning models and a neural network. It also looks at which factors are most important when predicting the occurrence of diabetes.

2 Literature review

Machine learning has been a useful technique in healthcare to support disease prediction and risk assessment. In the context of diabetes prediction, several studies have demonstrated that machine learning algorithms such as K-Nearest Neighbors (KNN), Gaussian Naive Bayes, Random Forest, Logistic Regression, and Decision Trees can classify individuals as diabetic or non-diabetic (Mujumdar and Vaidehi, 2019) (Rani, 2020). Moreover, Lakshmi et al. (2023) compared different classical approaches for the BRFSS dataset in 2023. In contrast to those classical approaches, some other studies have focused on the implementation of neural networks that cover the prediction of diabetes such as El_Jerjawi and Abu-Naser (2018). Furthermore, Prasetyo and Izdihar (2024) investigated the performance of a Multi Layer Perceptron on the BRFSS dataset as well.

These previous studies highlight the importance of reliable and well-performing models for diabetes prediction. Building on this, this research aims to compare the performance of both classical machine learning algorithms and a neural network approach. Furthermore, by using the BRFSS dataset, this study seeks to classify individuals into three categories: diabetes, prediabetes, and no diabetes.

3 Methods

3.1 Data Preprocessing and Data Analysis

Before applying machine learning models, the dataset will undergo a series of preprocessing and exploratory data analysis (EDA) steps to ensure data quality and model suitability.
Although the dataset used in this research is a cleaned version of the BRFSS survey, the first preprocessing step includes verifying the absence of missing values and confirming appropriate data types. Upon inspection, no missing values were found across the 21 features or the target variable. Moreover, all features and the target variable are of the float64 data type, so no additional type conversions are necessary.
To better understand the characteristics of the dataset, descriptive statistics were computed and exploratory data analysis was performed for all 21 features.
The dataset contains 21 features, 14 of which are binary health indicators such as “HighBP”, “HighChol” and “Smoker”, all coded 0 or 1. It also includes numerical features like “BMI” (continuous) and “MentHlth” and “PhysHlth” (discrete counts of unhealthy days in the past 30 days). Respondents reported an average of 3.18 days of poor mental health and 4.24 days of poor physical health, both right-skewed. Ordinal categorical features include “GenHlth”, “Education”, “Income” and “Age”, with “Education”, “Age” and “Income” showing slight left skew, while “GenHlth” is right-skewed.

For more details about the features, please refer to the dataset documentation available on Kaggle. (Teboul, 2021)

As mentioned previously, the target is divided into “No Diabetes” (0), “Prediabetes” (1) and “Diabetes” (2) with a strong imbalance (more than 80%) towards the “No Diabetes” class. This could lead to bias towards the majority class in the decision making process and therefore to poor detection of the minority class. Because of this, the Synthetic Minority Oversampling Technique (SMOTE) is applied on the training data after splitting it using stratification. This approach ensures that models are exposed to a more balanced representation of each class during training, reducing the likelihood of bias against minority classes. However, due to high computational costs, SMOTE technique cannot be applied to the Multilayer Perceptron Neural Network. Instead, an undersampling technique is used for this MLP classifier to address the class imbalance while also maintaining reasonable computational costs.
In the preprocessing pipeline, numerical features are always scaled using “StandardScaler” to normalize their distributions, while ordinal features are scaled for some scale-sensitive models using "MinMaxScaler". Binary features are left unchanged without scaling them.

3.2 Feature Selection

To assess potential redundancy among the features, a correlation matrix was computed across all 21 features using Pearson’s correlation coefficient. The maximum observed correlation was 0.52 between the features “GenHlth” and “PhysHlth”. This relationship seems to be intuitive, as individuals who report poorer overall health are more likely to also report more days of physical illness. However, a correlation of 0.52 seems to be moderate, such that both features will be considered for the following feature selection.

To identify which features contribute most strongly to distinguishing between “No Diabetes”, “Prediabetes”, and “Diabetes”, we calculated

the mutual information of all features using the "mutual_info_classif"-function from scikit-learn. This method estimates the dependency between each feature and the target variable.
The analysis highlights "GenHlth" and "HighBP" as the most informative predictors of diabetes status. At the lower end of the ranking, we identified four features having almost zero mutual information. Those features are "Stroke", "MentHlth", "HvyAlcoholConsump", and "NoDocbcCost". This is why we decided to cut off these four irrelevant features in our following models.

3.3 Models

To evaluate predictive performance and interpretability, four models were selected: K-Nearest Neighbor, Logistic Regression and Decision Tree as classical machine learning models, and a Neural Network to represent a more complex approach. This combination allows for meaningful comparisons between interpretable and black-box approaches. Grid search hyperparameter tuning was performed for all models using F1-macro score as scoring strategy and cross-validation splitting strategy of cv = 3.
In the following the chosen models as well as its tuned hyperparameters will be mentioned briefly.

3.3.1 K-Nearest Neighbor (KNN)

K-Nearest Neighbors (KNN) is a supervised machine learning algorithm which can be used for classification. It works by identifying the "k" closest neighbors to a new data point in the training set and predicting the class based on the majority class among those neighbors. Due to its simplicity and intuitivity it makes sense to use KNN as a first model to classify between diabetes, prediabetes and no diabetes. KNN models generally have high variance and low bias.
The hyperparameters that were tuned for this KNN classifier can be found in table 1.
The best F1-macro score of approximately 0.413 was achieved by using the hyperparameters n_neighbors = 7, weights = 'distance' and p = 1.

Hyperparameter	Tested Values
`n_neighbors`	3, 7, 9, 11
`weights`	'uniform', 'distance'
`p (metric)`	1, 2

Table 1: Tested Hyperparameters during Tuning for KNN Classifier

3.3.2 Logistic Regression

Logistic Regression (LR) is a supervised machine learning algorithm. It models the probability that a given instance belongs to a particular class by fitting a logistic (sigmoid) function to the data. The model estimates coefficients for each feature, indicating their influence on the predicted outcome. Thus, LR was chosen because of its interpretability of coefficients.
LR models typically exhibit low variance and high bias, which means they offer a strong baseline for comparison with more flexible but less interpretable algorithms such as Neural Networks.
The hyperparameters that were tuned for this Logistic Regression can be found in table 2.
The best F1-macro score of approximately 0.4228 was achieved by using the hyperparameters C = 0.01 and class_weight = 'None'.

Hyperparameter	Tested Values
`C`	0.01, 0.03, 0.1, 0.3
`class_weight`	None, 'balanced'

Table 2: Tested Hyperparameters during Tuning for Logistic Regression

3.3.3 Decision Trees

A Decision Tree is a supervised machine learning model that splits data into groups based on different features that best separate the classes. By doing this repeatedly, it can learn complex, non-linear patterns. However, DT is prone to high variance and low bias, where it can easily overfit the training data if allowed to grow too deep.
The hyperparameters that were tuned for this Decision Tree can be found in table 3.
The best F1-macro score of approximately 0.4432 was achieved by using the hyperparameters criterion = 'gini', max_depth = 12, min_samples_split = 5, min_samples_leaf = 1 and class_weight = 'None'.

Hyperparameter	Tested Values
`criterion`	'gini', 'entropy'
`max_depth`	11, 12, 13, 14
`min_samples_split`	2, 5, 10
`min_samples_leaf`	1, 3, 5, 10
`class_weight`	None, 'balanced'

Table 3: Tested Hyperparameters during Tuning for Decision Tree

3.3.4 Neural Network (MLP)

A Multi-Layer Perceptron (MLP) is a type of neural network that can learn complex, non-linear relationships through multiple layers of perceptrons. Each neuron (or perceptron) computes a weighted sum of its inputs followed by an activation function, allowing the network to model intricate feature interactions that simpler models may miss. The model learns by calculating the error and updating the weights through backpropagation. This process is repeated for many iterations, called epochs, until the loss converges or the learning rate becomes sufficiently small. However, MLPs can be computationally expensive to train and are highly uninterpretable, making it difficult to understand why the model produced a particular prediction. The hyperparameters that were tuned for this Multilayer Perceptron Neural Network can be found in table 4.

The best F1-macro score of approximately 0.42 was achieved by using the hyperparameters alpha = 0.1, hidden_layer_sizes = (100,).

Hyperparameter	Tested Values
`hidden_layer_sizes`	(50,), (50, 50), (100,)
`alpha`	0.0001, 0.001, 0.01, 0.1

Table 4: Tested Hyperparameters during Tuning for Multilayer Perceptron Neural Network

3.3.5 Evaluation Metrics

The following metrics will be used for performance evaluation in order to compare the different models with each other.

- Overall accuracy, as a general indicator of performance.
- Macro- and weighted-averaged precision, recall, and F1-score, to ensure balanced evaluation across all three classes.
- Confusion matrices, to identify common misclassification patterns.

3.4 Results

3.4.1 K-Nearest Neighbor (KNN)

Looking at the results of the K-Nearest Neighbor model (see table 5), one observes that the model achieved an overall accuracy of 0.71.

Nevertheless, the model performs very differently for the different classes which is indicated by a fairly low F1-macro score of 0.41. One can clearly observe that the model is predicting class 0 (No Diabetes) the best with a F1-score of 0.83 whereas for class 1 (Prediabetes) and class 2 (Diabetes) the model reaches scores lower F1-values of 0.04 and 0.38. The model especially struggles to detect the prediabetes cases which can be observed in the confusion matrix as well (see fig. 1).

Only 9% of the prediabetes cases are detected as those, meaning that in 91% of the prediabetes cases the model either predicts class 0 or class 2. The confusion matrix 1 shows very clearly that it performs best in predicting no diabetes. Only 24% of class 0 are predicted as class 1 or 2.

Class	Prec.	Rec.	F1	Supp.
0 (No Diabetes)	0.91	0.76	0.83	64111
1 (Prediabetes)	0.03	0.09	0.04	1389
2 (Diabetes)	0.30	0.51	0.38	10604
Macro Avg	0.41	0.45	0.41	76104
Weighted Avg	0.81	0.71	0.75	76104
Accuracy		0.71		

Table 5: KNN Classification Report on Test Set

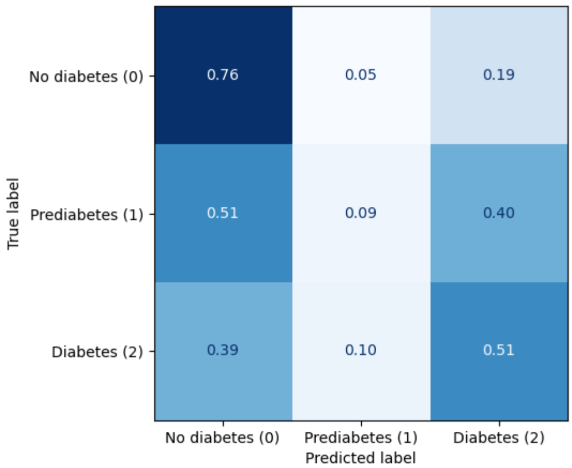

Figure 1: KNN Confusion Matrix

Feature importance analysis could not be directly performed for the KNN model, since it is an instance-based learning algorithm that relies on distance calculations rather than explicitly learning for example feature weights or feature-target interactions.

3.4.2 Logistic Regression

Overall, the LR model achieved moderate predictive performance, with an accuracy of 0.64 and a macro F1-score of 0.42 (see table 6). As shown in figure 2, the confusion matrix shows that the model performed well for the no diabetes class, correctly classifying most non-diabetic individuals (precision 0.95, recall 0.66). In contrast, the performances for the prediabetes and diabetes classes were lower, with prediabetes showing the weakest results (with precision 0.03, recall 0.30, F1-score 0.05). The macro recall of 0.51 indicates that while the model captured a fair portion of positive cases overall, many prediabetic individuals were misclassified as non-diabetic or diabetic. The LR model demonstrates solid performance in predicting non-diabetic cases, while offering fair but limited detection of diabetic and prediabetic individuals.

Class	Precision	Recall	F1-score	Support
0 (No Diabetes)	0.95	0.66	0.78	64111
1 (Prediabetes)	0.03	0.30	0.05	1389
2 (Diabetes)	0.35	0.58	0.44	10604
Macro Avg	0.44	0.51	0.42	76104
Weighted Avg	0.85	0.64	0.72	76104
Accuracy		0.64		

Table 6: Logistic Regression Classification Report on Test Set

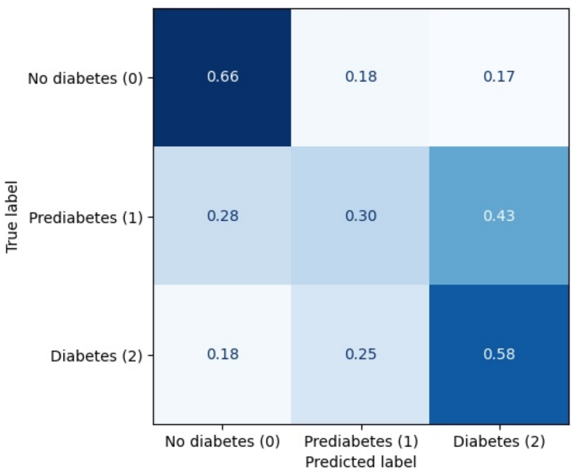

Figure 2: Logistic Regression Confusion Matrix

To identify the most important predictors of diabetes in LR, we extracted the model coefficients as each coefficient represents the weight of a feature in influencing the decision boundary for one of the outcome classes. Because our model was trained for multiple classes, we obtained a coefficient matrix with one row per class and one column per feature. To summarize the importance of features across all classes, we calculated the average of the absolute coefficient values for each feature. CholCheck was found to be the most influential feature, followed by HighChol, HighBP, GenHlth, and BMI, highlighting the strong impact of cardiovascular and general health factors. Other notable predictors include AnyHealthcare, HeartDiseaseorAttack, and Sex. In general, the results suggest that lifestyle and cardiovascular indicators play a key role in determining risk of diabetes.

3.4.3 Decision Tree

The DT model has achieved a predicting performance with an accuracy of 0.77 and a macro F1-score of 0.44. Similarly, the model performed strongly for the no diabetes class, with a precision of 0.91 and recall of 0.82 (see table 7), indicating reliable classification for non-diabetic individuals. However, its performance is very low for the prediabetes class (precision 0.03, recall 0.04, F1-score 0.03), showing difficulty in distinguishing prediabetic cases. Overall, while the DT model exhibited good accuracy and recall for the majority class, it still faced challenges in predicting the minority classes, particularly prediabetes which can be observed in its confusion matrix as well 3.

Class	Precision	Recall	F1-score	Support
0 (No Diabetes)	0.91	0.82	0.86	64111
1 (Prediabetes)	0.03	0.04	0.03	1389
2 (Diabetes)	0.36	0.56	0.43	10604
Macro Avg	0.43	0.47	0.44	76104
Weighted Avg	0.82	0.77	0.79	76104
Accuracy		0.77		

Table 7: Decision Tree Classification Report on Test Set

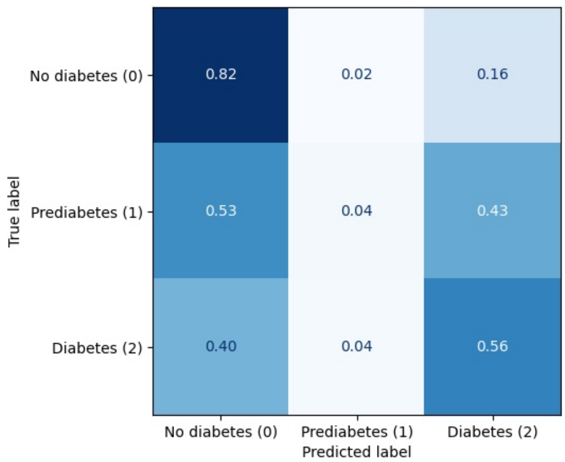

Figure 3: Decision Tree Confusion Matrix

Feature importance analysis for the Decision Tree model was conducted based on the impurity reduction contributed by each feature across all splits. The results show that HighBP (high blood pressure) and GenHlth were among the most influential predictors. Together, these two features contributed the most to the predictions, showing a strong connection between cardiovascular and overall health conditions and diabetes risk. The next key contributors were BMI, HighChol, and Age. Lesser but still relevant predictors included Smoker, PhysHlth, and Income, suggesting that lifestyle and socioeconomic factors also played secondary roles in classification.

3.4.4 Neural Network (MLP)

The MLP (Multilayer Perceptron) model achieved an overall accuracy of 0.72 and a macro F1-score of 0.42. The model demonstrated strong performance for the no diabetes class, with a precision of 0.95 and a recall of 0.72 (see Table 8), indicating that it effectively identified non-diabetic individuals, though with some false negatives. In contrast, the model failed to correctly classify any prediabetic cases (precision 0.00, recall 0.00, F1-score 0.00), showing a complete inability to detect this minority class. For the diabetes class, the model achieved moderate predictive performance, with a recall of 0.79 but a relatively low precision of 0.31, suggesting that it correctly identified most diabetic instances but also misclassified several non-diabetic cases as diabetic.

Overall, while the MLP model showed reasonable discriminative ability for the majority (no diabetes) and diabetic classes, it struggled with imbalanced data, particularly in recognizing prediabetic individuals. This limitation is further illustrated in its confusion matrix (Figure 4), where misclassifications between diabetes and no diabetes dominate the off-diagonal regions.

Class	Precision	Recall	F1-score	Support
0 (No Diabetes)	0.95	0.72	0.82	64111
1 (Prediabetes)	0.00	0.00	0.00	1389
2 (Diabetes)	0.31	0.79	0.44	10604
Macro Avg	0.42	0.50	0.42	76104
Weighted Avg	0.84	0.72	0.75	76104
Accuracy		0.72		

Table 8: Neural Network (MLP) Classification Report on Test Set

Feature importance analysis could not be directly performed for the MLP neural network, due to its nature of not inherently providing feature weights like tree-based models. Therefore, it is not straightforward to determine exactly how or why the neural network made its predictions.

3.5 Discussion

This study wanted to evaluate the performances of different machine learning approaches, specifically

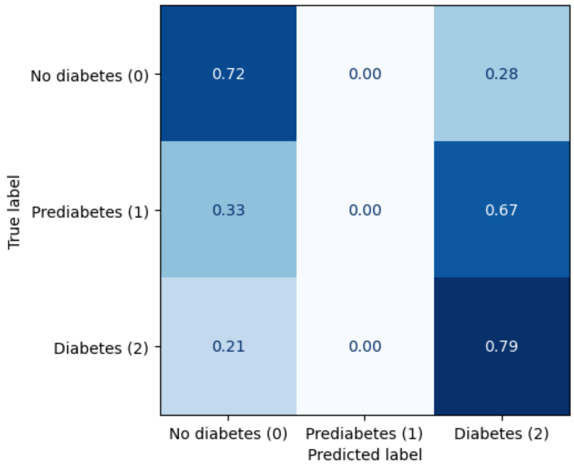

Figure 4: Neural Network (MLP) Confusion Matrix

of K-Nearest Neighbor, Logistic Regression, Decision Tree and Multilayer Perceptron Neural Network in classifying individuals as either non-diabetic, prediabetic or diabetic based on several health indicators that were sampled from the 2015 Behavioral Risk Factor Surveillance System (BRFSS) dataset found on Kaggle. (Teboul, 2021)

Each of these models come with different limitations. For KNN, these include difficulties in selecting the best distance function as well as an appropriate value of K (number of neighbors). Moreover, KNN gets very expensive with larger datasets like our dataset BRFSS.
For Logistic Regression, despite providing some interpretability regarding feature importance, it can only learn linear data-feature relationships. Because of this limitation, we also implemented non-linear models such as Decision Tree and Multilayer Perceptron Neural Network.
Typical limitations of Decision Trees include overfitting, loss of information for continuous variables such as BMI and complex calculations in the case of many classes. In our case we only dealt with one continuous variable (BMI) and three different target classes such that those limitations might not be that significant in this research.
The main limitations of neural networks are their data requirements (they require a large amount of data), it's hard to interpret how input features are transformed into outputs (not interpretable), and the high computational cost.

We obtained the following results for all these models which are presented in bar plots below.

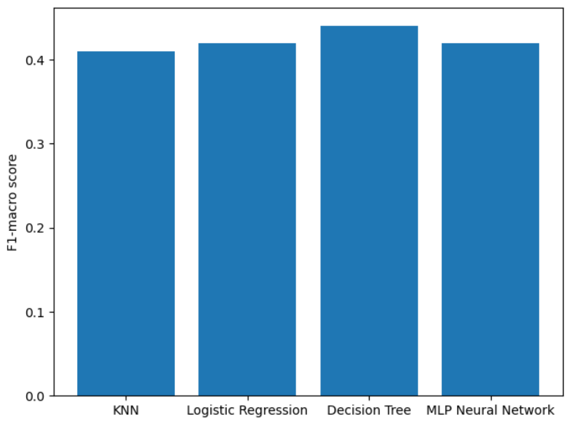

Figure 5: F1-macro Scores across Models

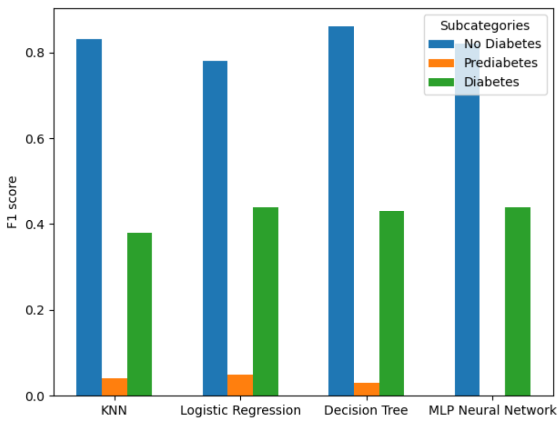

Figure 6: F1 Scores across Classes for all Models

In figure 5, one observes that all models have a similar F1-macro score of around 0.4. When having a closer look at the results, one sees that the Decision Tree model performs slightly better than the other models. Nevertheless, a F1-macro score of around 0.4 shows a in general poor performance. This low F1-macro score is especially due to models' low classifying power of the prediabetic class as shown in figure 6, which presents the F1-score across all three classes for all models. One observes, that for all models the prediabetic cases have a close-to-zero F1 score. This is most likely due to the dataset's imbalance as described earlier. Even though we used either SMOTE oversampling or undersampling technique, the mitigation effect was limited possibly due to similarity in feature values across the three different classes. The confusion matrices in the previous results section show that the prediabetic individuals were frequently misclassified as either non-diabetic or diabetic across all models. This suggests that the boundary between prediabetes and the other two classes is blurred, which reflects real-world diagnostic challenges in distinguishing prediabetes symptoms from non-diabetic and diabetic states.

To overcome this data imbalance we tried to combine undersampling followed by SMOTE technique specifically for the MLP classifier. The undersampling step first reduced the number of majority class (no diabetes) samples, preventing the model from having too many non-diabetic cases. Then, SMOTE synthetically generated new minority samples which produced a balanced dataset. Moreover, undersampling keeps the computational costs of the MLP classifier reasonable. Using undersampling and SMOTE techniques in the MLP model leads to the results shown in table 9 as well as in the confusion matrix displayed in figure 7.

Class	**Prec.**	**Rec.**	**F1**	**Supp.**
0 (No Diabetes)	0.95	0.70	0.81	64111
1 (Prediabetes)	0.03	0.01	0.02	1389
2 (Diabetes)	0.30	0.80	0.44	10604
Macro Avg	0.43	0.50	0.42	76104
Weighted Avg	0.84	0.70	0.74	76104
Accuracy		0.70		

Table 9: Neural Network (MLP) Classification Report on Test Set using Undersampling and SMOTE

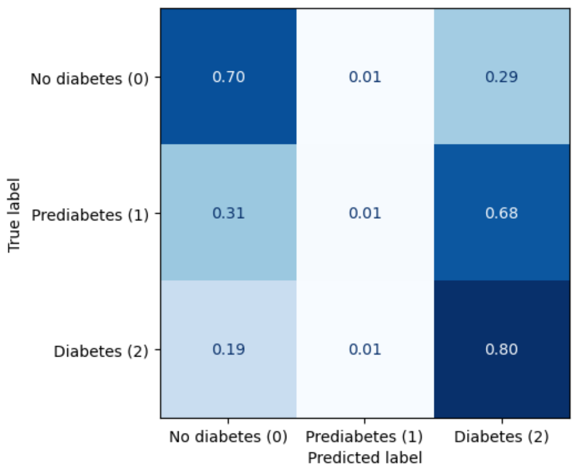

Figure 7: Neural Network (MLP) Confusion Matrix using Undersampling and SMOTE

These results show that despite using undersampling and SMOTE technique, no major improvement was achieved. In general, the overall accuracy decreased from 0.72 when using only undersampling (see table 8) to 0.70 when additionally applying SMOTE technique (see table 9). However, the F1 score for the prediabetic class increased

slightly from 0.00 to 0.02. Nevertheless, the overall F1-macro score of 0.42 still indicates a very poor performance. This again demonstrates that features might have similar values across classes, which makes it hard to distinguish properly between classes.
The second research question is regarding feature selection. The BRFSS dataset consists of 21 features. As mentioned previously (see section 3.2), we performed univariate feature selection to estimate the mutual information between all features and the target variable via "mutual_info_classif"-function. This function measures the dependency between two variables. Four out of those 21 features had a mutual information of almost zero (zero up to second decimal place), which is why we did not further consider those. These include the features "Stroke", "MentHlth", "HyvAlcoholConsump" and "NoDocbcCost". The leading features include "GenHlth", "HighBP" and "AnyHealthcare" with a mutual information above 0.05.
Moreover, for the Logistic Regression and the Decision Tree models we further performed feature importance analysis after building the model to find out which features were most important. For Logistic Regression we extracted the coefficients of the model and summarized how important each feature is across all target classes. The top five most important features for the trained Logistic Regression model can be found in figure 8.

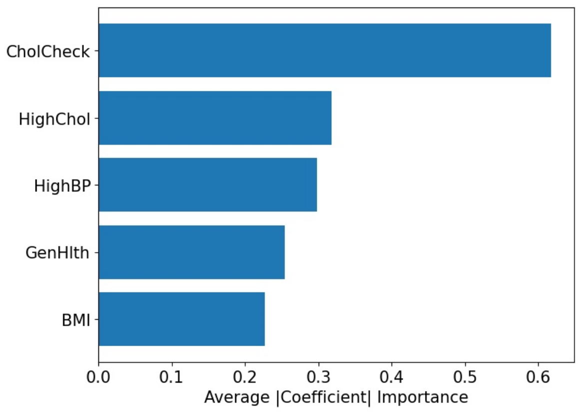

Figure 8: Importance of Features in Logistic Regression Model

For the Decision Tree model, the feature importance is measured as the total reduction in impurity that the feature contributes to across the entire tree. Here, the top five most important features can be found in Figure 9.

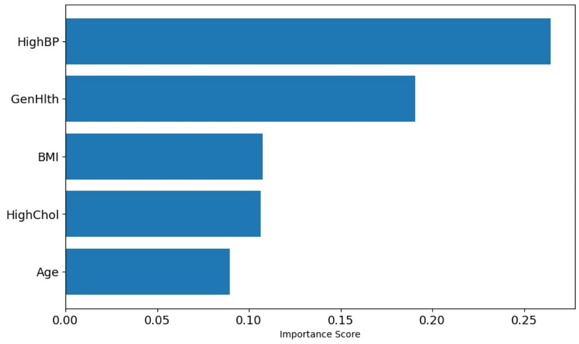

Figure 9: Importance of Features in Decision Tree Model

One can observe that four out of the five most important features are shared across the Logistic Regression and Decision Tree models, though in a different order. These include "HighChol", "HighBP", "GenHlth", and "BMI". Moreover, the features "GenHlth" and "HighBP" appear not only in the lists of the most important features for both Logistic Regression and Decision Tree models but also among the features with the highest mutual information. This suggests that especially these two features contribute significantly to the performance of the different models.

3.6 Conclusion

In this research study, we investigated the ability of three classical machine learning models and one Neural Network model to accurately classify individuals into non-diabetic, prediabetic, and diabetic classes using the BRFSS health survey data. Moreover, we also had a look into feature selection and the importance of those features for the different models. We found that among all four models, the Decision Tree reached the highest accuracy of 0.77 as well as the highest F1-macro score of 0.44. Nevertheless, even after using over- and/or undersampling techniques, all models struggled with the dataset's imbalance, which explains the poor F1-macro scores across all models. To improve the predictive ability of especially the prediabetic class, further over- and undersampling techniques could be considered.
Regarding the feature importance, we noticed that the features "GenHlth" and "HighBP" seem to be the most relevant when predicting diabetes because those were among the leading important features not only when considering mutual information but also when measuring feature importance for the trained Logistic Regression and Decision Tree models.

References

N. El_Jerjawi and S. Abu-Naser. 2018. Diabetes prediction using artificial neural network. *Journal of Advanced Science*, 124:1–10.

H. N. Lakshmi, A. S. Reddy, and K. Naidu. 2023. Analysis of diabetic prediction using machine learning algorithms on brfss dataset. *2023 7th International Conference on Trends in Electronics and Informatics (ICOEI)*, pages 1024–1028.

A. Mujumdar and V. Vaidehi. 2019. Diabetes prediction using machine learning algorithms. *Procedia Computer Science*, 165:292–299. 2nd International Conference on Recent Trends in Advanced Computing ICRTAC -DISRUP - TIV INNOVATION , 2019 November 11-12, 2019.

S. Y. Prasetyo and Z. Izdihar. 2024. Multi-layer perceptron approach for diabetes risk prediction using brfss data. pages 303–308.

KM Rani. 2020. Diabetes prediction using machine learning. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, pages 294–305.

A. Teboul. 2021. Diabetes health indicators dataset (based on cdc's 2015 survey). Data set. Kaggle. Accessed: 2025-10-06.

World Health Organization. 2024. Diabetes. Accessed: 2025-10-06.