Type 2 Diabetes

Leveraging BRFSS Survey Data to Predict Diabetes With Known Risk Factors

Course: Data Mining Visualization
Instructor: Dr. Sherry Ni

By: Khushbuben Patel, Saiyida Zainab Jabeen, Dipti Paldhikar, and Kalyn Simmons

Table of contents

01

Introduction

02

Data

03

Data Exploration

04

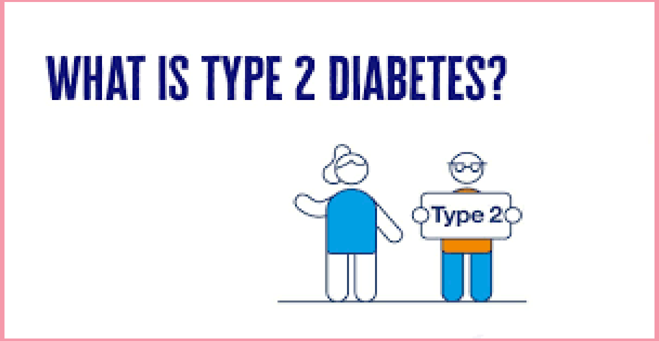
Raw Dataset

05

Processed Dataset

06

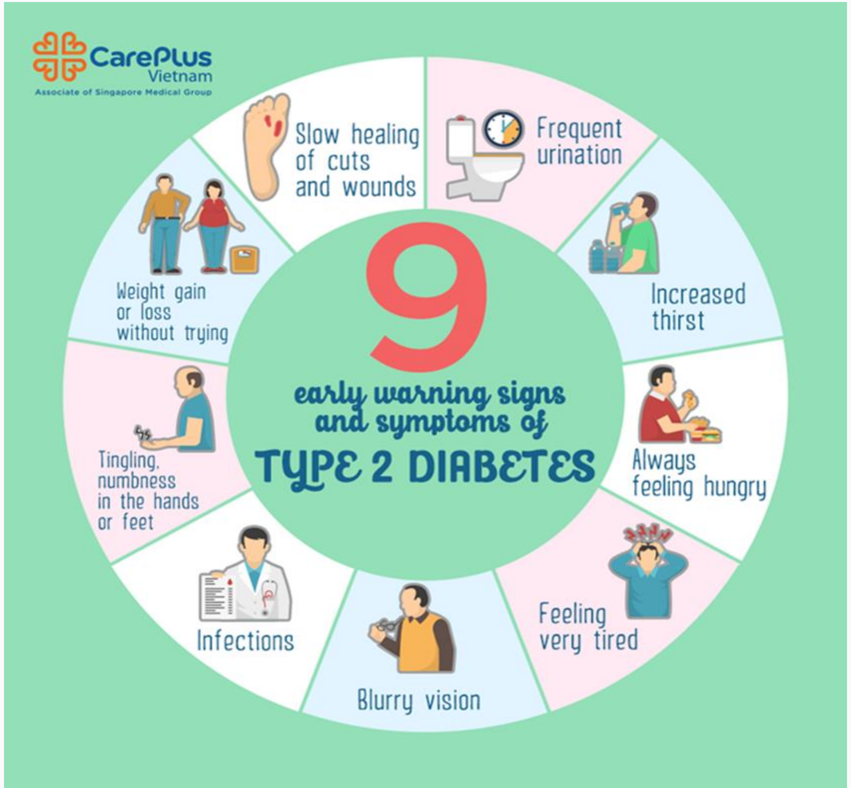
Problem Statement



Introduction

- Diabetes is a chronic disease that affects millions of people in the US, leading to serious complications that can reduce life expectancy.
- We, as healthcare professionals, are interested in analyzing the risk factors of diabetes to help prevent its onset.
- There are two types of diabetes: Type I and Type II, with Type II being the most common form.
- Our aim is to identify the risk variables in the provided dataset that are most effective in predicting risk for developing Type 2 Diabetes.

Symptoms of the disease


CarePlus Vietnam
Associate of Singapore Medical Group
9
early warning signs and symptoms of
TYPE 2 DIABETES
Slow healing of cuts and wounds
Frequent urination
Increased thirst
Always feeling hungry
Feeling very tired
Blurry vision
Infections
Tingling, numbness in the hands or feet
Weight gain or loss without trying

Behavioral Risk Factor Surveillance System

● **United States of America**

The Behavioral Risk Factor Surveillance System (BRFSS) is the nation’s premier system of health-related telephone surveys that collect state data about U.S. residents regarding their health-related risk behaviors, chronic health conditions, and use of preventive services. Established in 1984 with 15 states, BRFSS now collects data in all 50 states as well as the District of Columbia and three U.S. territories. BRFSS completes more than 400,000 adult interviews each year, making it the largest continuously conducted health survey system in the world

Description of Data

441,455 individuals and has 330 features

Source: Kaggle using dataset for the year 2015

253,680 survey responses with 22 feature variables

Source: We will be using diabetes_012_health_indicators_BRFSS2015.csv dataset

Target Variable: **(Diabetes_012) with 3 classes**

The data represented in the dataset is records from the **telephonic survey which was collected annually by the CDC.**

Source: Kaggle

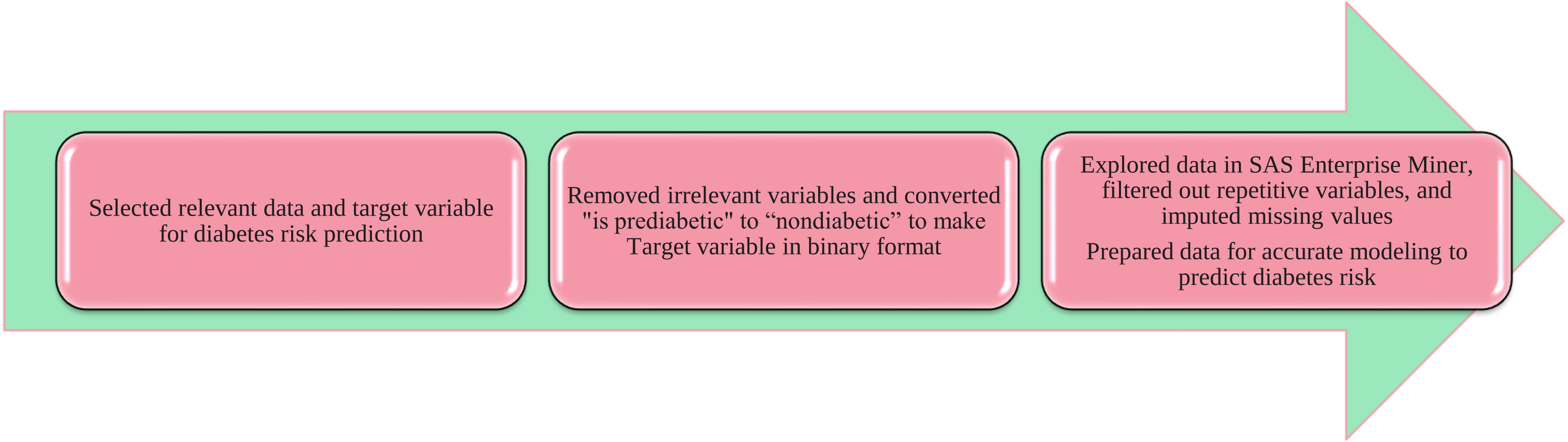
Diabetes_012 Dataset

- Diabetes_012 is the target variable in the dataset.
- It has 3 classes: 0 is for no diabetes or only during pregnancy, 1 is for prediabetes, and 2 is for diabetes.

Diabetes_012	Frequency	Percent
Has no diabetes or only during pregnancy	8218	83.94%
Has diabetes	1606	16.06%
Is prediabetic	176	1.76%

Data Exploration Steps

It is the first step of data analysis which is used to explore and visualize data to uncover insights from the start to identify patterns to dig into more.



Selected relevant data and target variable for diabetes risk prediction

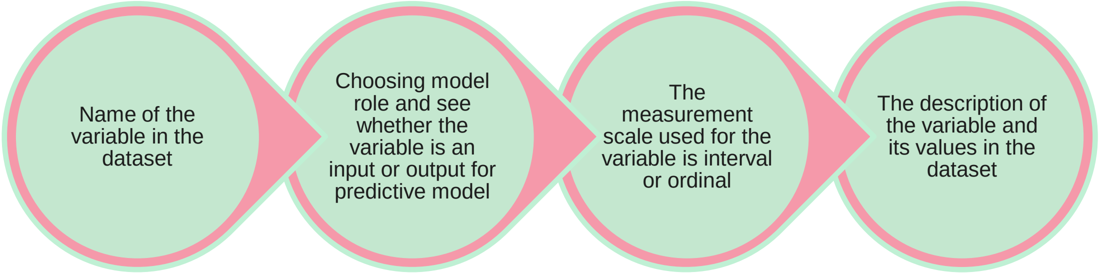
Removed irrelevant variables and converted "is prediabetic" to “nondiabetic” to make Target variable in binary format

Explored data in SAS Enterprise Miner, filtered out repetitive variables, and imputed missing values

Prepared data for accurate modeling to predict diabetes risk

Data Preprocessing Steps

It is a component of data preparation which describes any type of processing performed on raw data to prepare it for another data processing procedure



Name of the variable in the dataset

Choosing model role and see whether the variable is an input or output for predictive model

The measurement scale used for the variable is interval or ordinal

The description of the variable and its values in the dataset

Problem , Goal and Constraints

- The BRFSS survey from 2015 provided a dataset of 21 feature variables to predict the risk of developing diabetes.

- The goal is to develop a prediction model that can accurately forecast an individual's risk of developing diabetes.

- However, class imbalance in the dataset, with the majority of replies coming from class 0, presents a major obstacle.

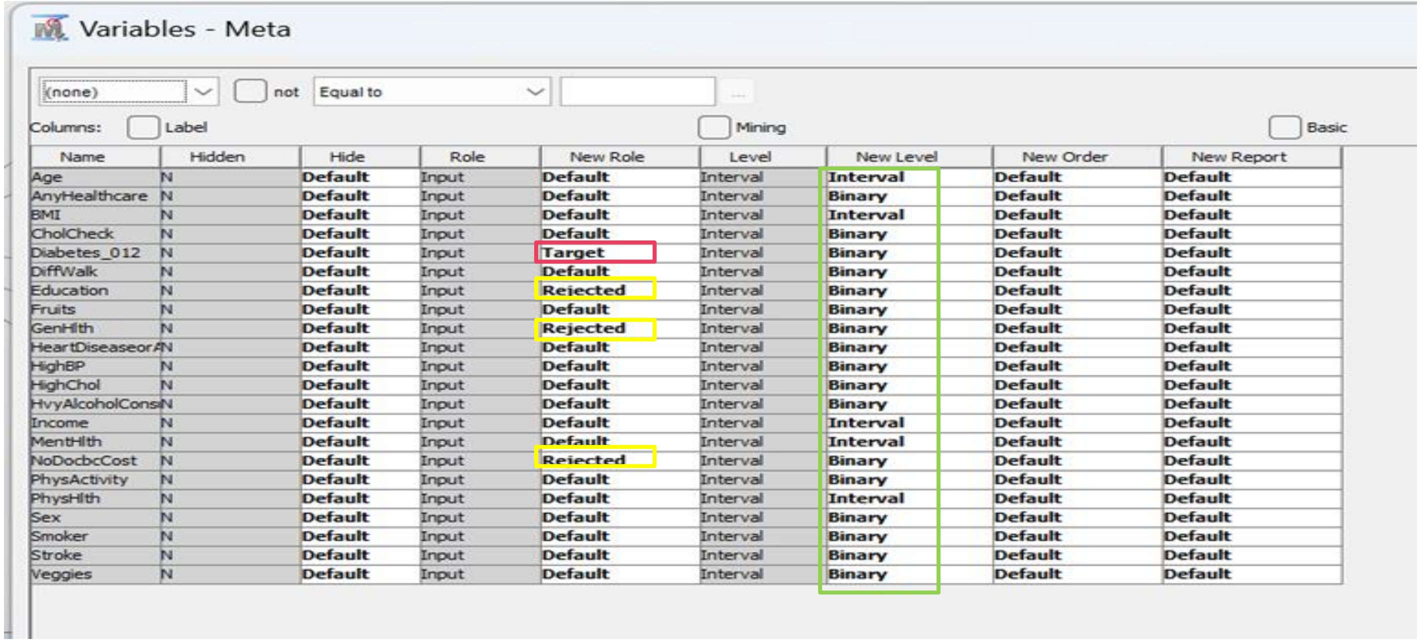
Plan for Data Mining

- The most significant diabetes risk variables were selected using variables graph study and selection approaches.

- SAS Enterprise Miner, a data mining and predictive analytics tool, was used to build the model.

- The plan is broken down into four stages: **data exploration and preprocessing, variable selection, model selection and evaluation.**

Data Preprocessing Result



Variables - Meta

(none) □ not Equal to

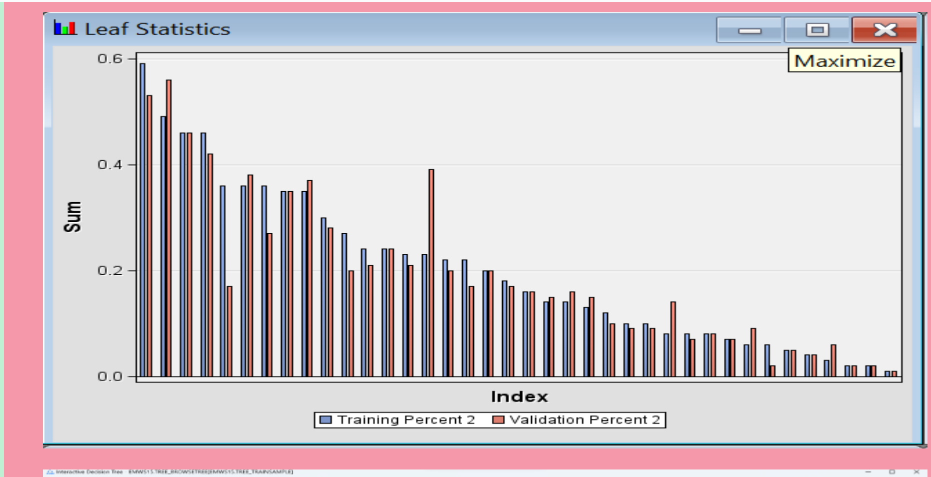
Columns: □ Label □ Mining □ Basic

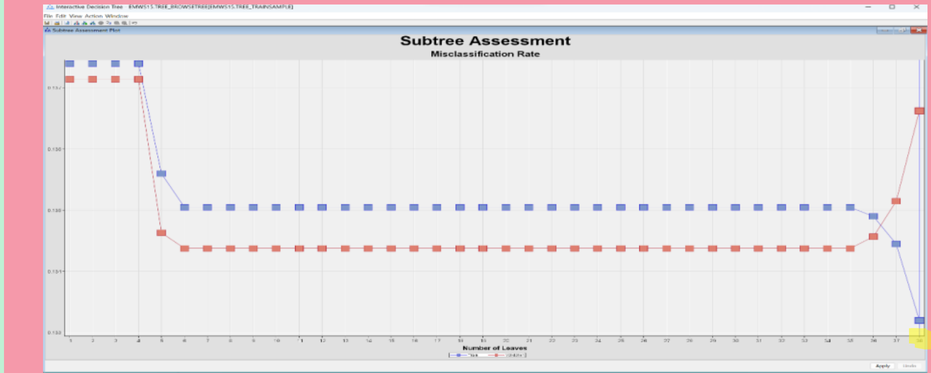
Name	Hidden	Hide	Role	New Role	Level	New Level	New Order	New Report
Age	N	Default	Input	Default	Interval	Interval	Default	Default
AnyHealthcare	N	Default	Input	Default	Interval	Binary	Default	Default
BMI	N	Default	Input	Default	Interval	Interval	Default	Default
CholCheck	N	Default	Input	Default	Interval	Binary	Default	Default
Diabetes_012	N	Default	Input	Target	Interval	Binary	Default	Default
DiffWalk	N	Default	Input	Default	Interval	Binary	Default	Default
Education	N	Default	Input	Rejected	Interval	Binary	Default	Default
Fruits	N	Default	Input	Default	Interval	Binary	Default	Default
GenHlth	N	Default	Input	Rejected	Interval	Binary	Default	Default
HeartDiseaseorA	N	Default	Input	Default	Interval	Binary	Default	Default
HighBP	N	Default	Input	Default	Interval	Binary	Default	Default
HighChol	N	Default	Input	Default	Interval	Binary	Default	Default
HvyAlcoholCons	N	Default	Input	Default	Interval	Binary	Default	Default
Income	N	Default	Input	Default	Interval	Interval	Default	Default
MentHlth	N	Default	Input	Default	Interval	Interval	Default	Default
NoDocbcCost	N	Default	Input	Rejected	Interval	Binary	Default	Default
PhysActivity	N	Default	Input	Default	Interval	Binary	Default	Default
PhysHlth	N	Default	Input	Default	Interval	Interval	Default	Default
Sex	N	Default	Input	Default	Interval	Binary	Default	Default
Smoker	N	Default	Input	Default	Interval	Binary	Default	Default
Stroke	N	Default	Input	Default	Interval	Binary	Default	Default
Veggies	N	Default	Input	Default	Interval	Binary	Default	Default

Model Analysis

Model 1- Maximal Decision Tree

- The maximal tree had 38 leaves, with High BP as the root of the splitting tree.
- Subtree assessment model shows the tree was optimal until 7 leaves, after which it becomes overfitting.
- Misclassification rate for validation was 0.135855.
- However, due to the high level of overfitting, we decided to create another tree model to reduce the noise.
- Further analysis and modeling was conducted to develop an accurate and reliable model for predicting the outcome variable.

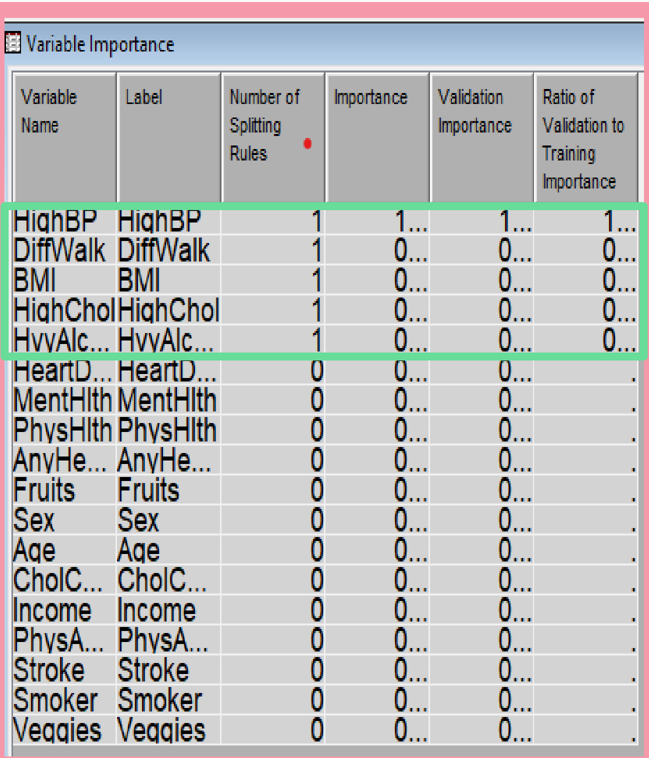



Interactive Decision Tree - EMWS15.TREE_BROWSETREE[EMWS15.TREE_TRAINSAMPLE]
File Edit View Action Window
Subtree Assessment Plot
Subtree Assessment
Misclassification Rate
Number of Leaves
Apply

Model Analysis

Model 2- Default Decision Tree

- Important Variables: High BP, DiffWalk, BMI, High Chol, HvyAlcohol
- Tree: 6 leaves, High BP as root
- Misclassification validation rate: 0.134375
- Purest leaf: HighBP=0/missing, Diffwalk=0/missing
- High BP is the main contributing factor in developing Type 2 Diabetes
- Skipped Average Square Error Tree Model due to categorical target variable.

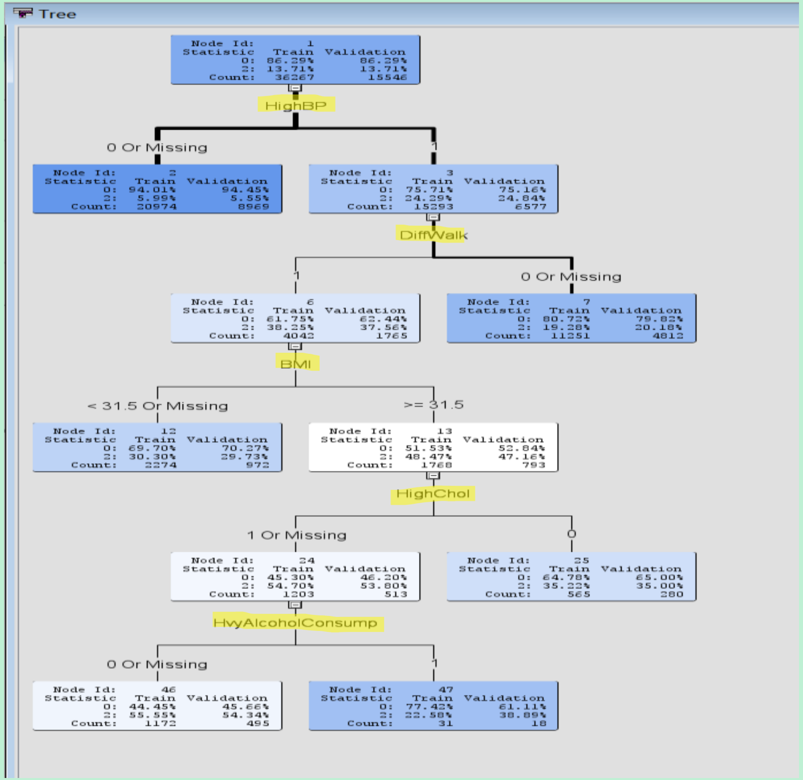


Variable Importance

Variable Name	Label	Number of Splitting Rules	Importance	Validation Importance	Ratio of Validation to Training Importance
HighBP	HighBP	1	1...	1...	1...
DiffWalk	DiffWalk	1	0...	0...	0...
BMI	BMI	1	0...	0...	0...
HighChol	HighChol	1	0...	0...	0...
HvyAlc...	HvyAlc...	1	0...	0...	0...
HeartD...	HeartD...	0	0...	0...	.
MentHlth	MentHlth	0	0...	0...	.
PhysHlth	PhysHlth	0	0...	0...	.
AnyHe...	AnyHe...	0	0...	0...	.
Fruits	Fruits	0	0...	0...	.
Sex	Sex	0	0...	0...	.
Age	Age	0	0...	0...	.
CholC...	CholC...	0	0...	0...	.
Income	Income	0	0...	0...	.
PhysA...	PhysA...	0	0...	0...	.
Stroke	Stroke	0	0...	0...	.
Smoker	Smoker	0	0...	0...	.
Veggies	Veggies	0	0...	0...	.

Model Analysis

Model 2- Default Decision Tree



```
Node Id:        1
Statistic   Train  Validation
       0:  86.29%      86.29%
       2:  13.71%      13.71%
   Count:  36267       15546
```

```
Node Id:        2
Statistic   Train  Validation
       0:  94.01%      94.45%
       2:   5.99%       5.55%
   Count:  20974        8969
```

```
Node Id:        3
Statistic   Train  Validation
       0:  75.71%      75.16%
       2:  24.29%      24.84%
   Count:  15293        6577
```

```
Node Id:        6
Statistic   Train  Validation
       0:  61.75%      62.44%
       2:  38.25%      37.56%
   Count:   4042        1765
```

```
Node Id:        7
Statistic   Train  Validation
       0:  80.72%      79.82%
       2:  19.28%      20.18%
   Count:  11251        4812
```

```
Node Id:       12
Statistic   Train  Validation
       0:  69.70%      70.27%
       2:  30.30%      29.73%
   Count:   2274         972
```

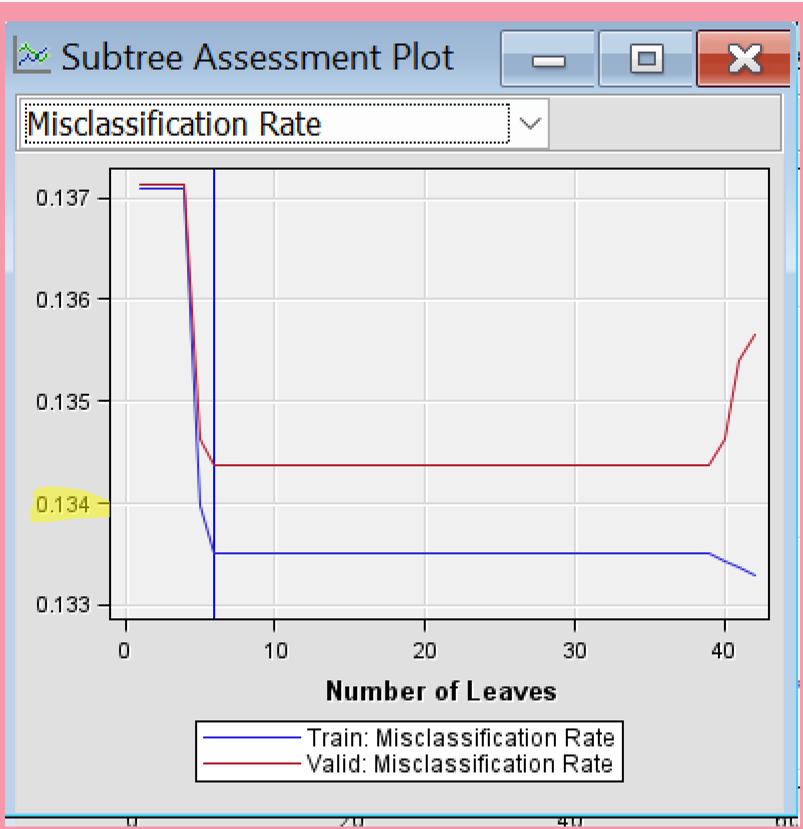
```
Node Id:       13
Statistic   Train  Validation
       0:  51.53%      52.84%
       2:  48.47%      47.16%
   Count:   1768         793
```

```
Node Id:       24
Statistic   Train  Validation
       0:  45.30%      46.20%
       2:  54.70%      53.80%
   Count:   1203         513
```

```
Node Id:       25
Statistic   Train  Validation
       0:  64.78%      65.00%
       2:  35.22%      35.00%
   Count:    565         280
```

```
Node Id:       46
Statistic   Train  Validation
       0:  44.45%      45.66%
       2:  55.55%      54.34%
   Count:   1172         495
```

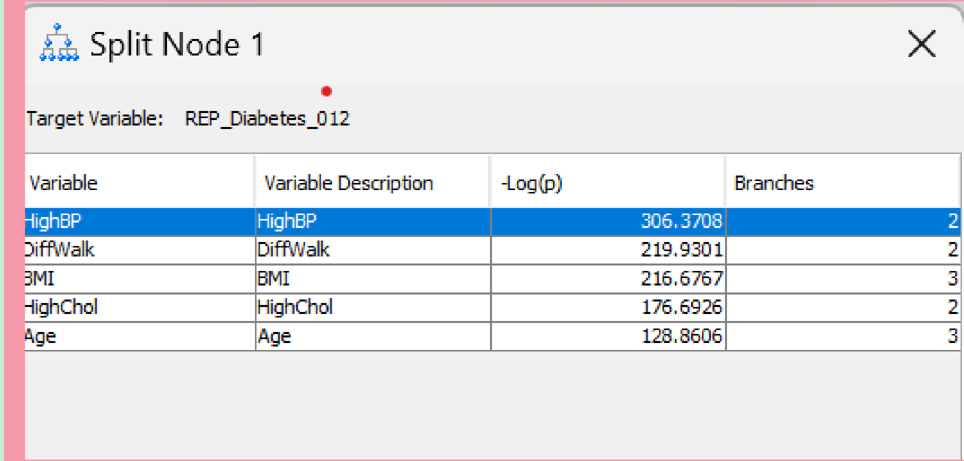
```
Node Id:       47
Statistic   Train  Validation
       0:  77.42%      61.11%
       2:  22.58%      38.89%
   Count:     31          18
```



Model Analysis

Model 3- Max 3 Decision Tree

- High BP is best for first split
- Competing variables: DiffWalk, BMI, HighChol, Age
- 7 leaves in subtree assessment plot
- Validation misclassification rate: 0.133861
- Purest leaf splitting rule:
  High BP=0 or missing High BP=1
- High BP is major factor in predicting type 2 Diabetes

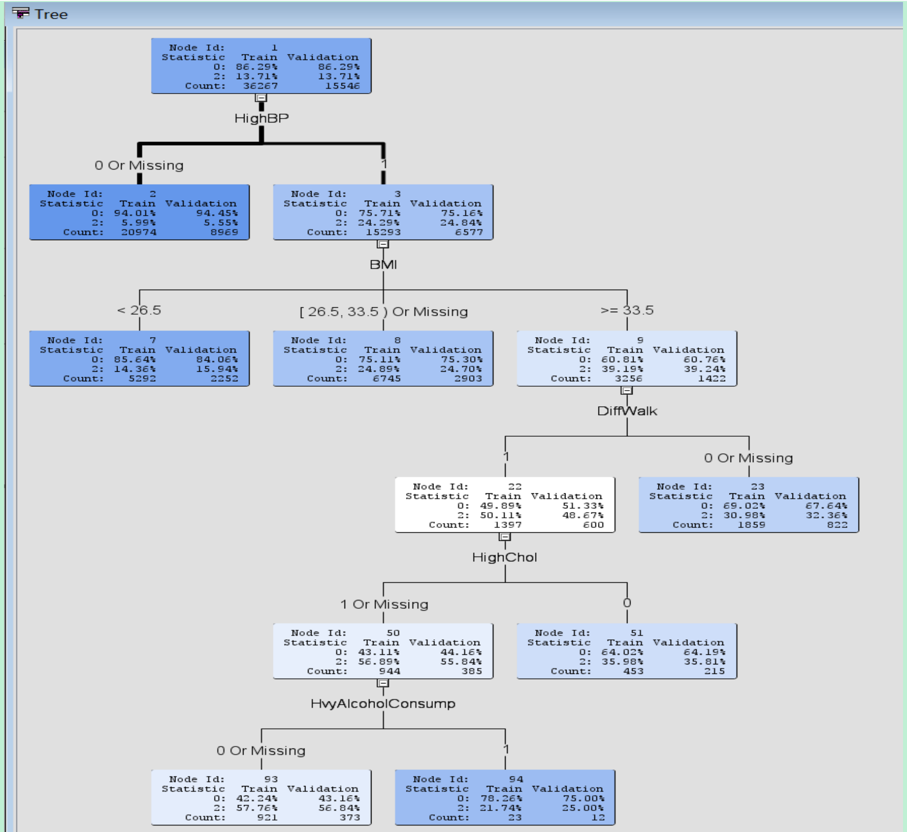


Target Variable: REP_Diabetes_012

Variable	Variable Description	-Log(p)	Branches
HighBP	HighBP	306.3708	2
DiffWalk	DiffWalk	219.9301	2
BMI	BMI	216.6767	3
HighChol	HighChol	176.6926	2
Age	Age	128.8606	3

Model Analysis

Model 3- Max 3 Decision Tree



```
  Node Id:        1
Statistic   Train  Validation
       0:  86.29%      86.29%
       2:  13.71%      13.71%
   Count:   36267       15546
```

```
  Node Id:        2
Statistic   Train  Validation
       0:  94.01%      94.45%
       2:   5.99%       5.55%
   Count:   20974        8969
```

```
  Node Id:        3
Statistic   Train  Validation
       0:  75.71%      75.16%
       2:  24.29%      24.84%
   Count:   15293        6577
```

```
  Node Id:        7
Statistic   Train  Validation
       0:  85.64%      84.06%
       2:  14.36%      15.94%
   Count:    5292        2252
```

```
  Node Id:        8
Statistic   Train  Validation
       0:  75.11%      75.30%
       2:  24.89%      24.70%
   Count:    6745        2903
```

```
  Node Id:        9
Statistic   Train  Validation
       0:  60.81%      60.76%
       2:  39.19%      39.24%
   Count:    3256        1422
```

```
  Node Id:       22
Statistic   Train  Validation
       0:  49.89%      51.33%
       2:  50.11%      48.67%
   Count:    1397         600
```

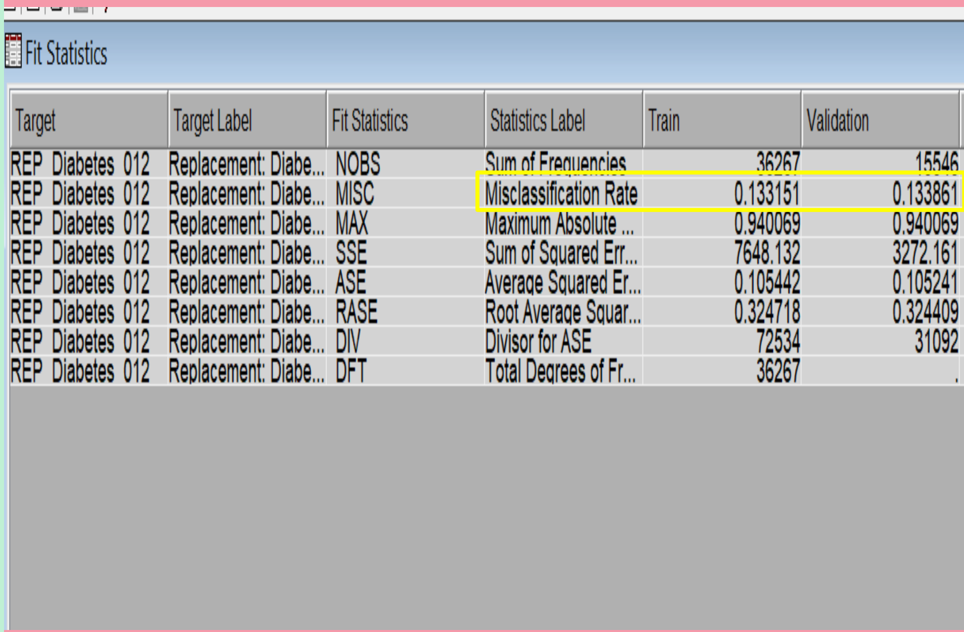
```
  Node Id:       23
Statistic   Train  Validation
       0:  69.02%      67.64%
       2:  30.98%      32.36%
   Count:    1859         822
```

```
  Node Id:       50
Statistic   Train  Validation
       0:  43.11%      44.16%
       2:  56.89%      55.84%
   Count:     944         385
```

```
  Node Id:       51
Statistic   Train  Validation
       0:  64.02%      64.19%
       2:  35.98%      35.81%
   Count:     453         215
```

```
  Node Id:       93
Statistic   Train  Validation
       0:  42.24%      43.16%
       2:  57.76%      56.84%
   Count:     921         373
```

```
  Node Id:       94
Statistic   Train  Validation
       0:  78.26%      75.00%
       2:  21.74%      25.00%
   Count:      23          12
```

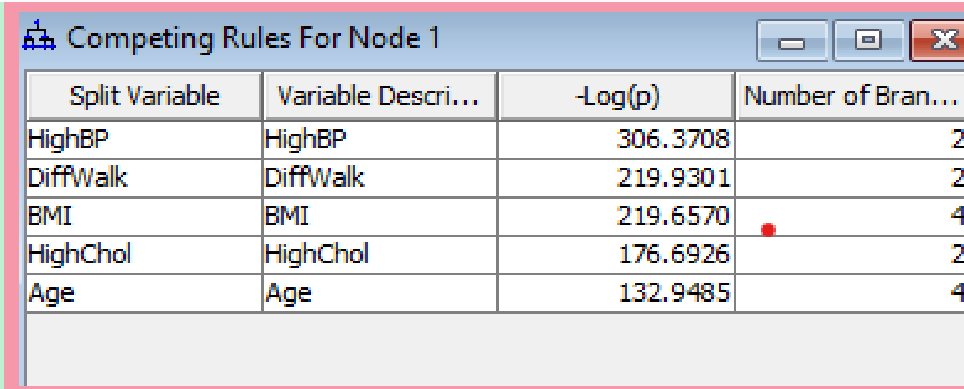


Target	Target Label	Fit Statistics	Statistics Label	Train	Validation
REP_Diabetes_012	Replacement: Diabe...	NOBS	Sum of Frequencies	36267	15546
REP_Diabetes_012	Replacement: Diabe...	MISC	Misclassification Rate	0.133151	0.133861
REP_Diabetes_012	Replacement: Diabe...	MAX	Maximum Absolute ...	0.940069	0.940069
REP_Diabetes_012	Replacement: Diabe...	SSE	Sum of Squared Err...	7648.132	3272.161
REP_Diabetes_012	Replacement: Diabe...	ASE	Average Squared Er...	0.105442	0.105241
REP_Diabetes_012	Replacement: Diabe...	RASE	Root Average Squar...	0.324718	0.324409
REP_Diabetes_012	Replacement: Diabe...	DIV	Divisor for ASE	72534	31092
REP_Diabetes_012	Replacement: Diabe...	DFT	Total Degrees of Fr...	36267	.

Model Analysis

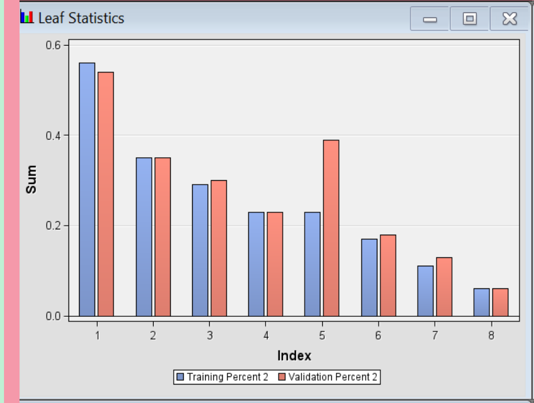
Model 4-Max 4 Decision Tree

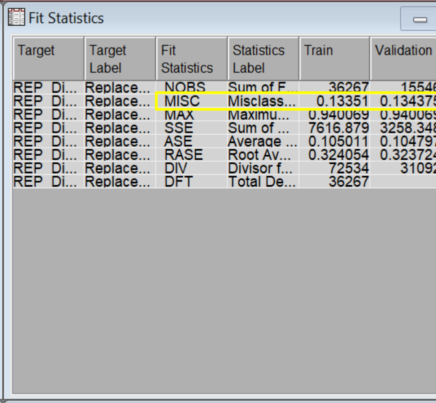
- High BP, DiffWalk, BMI, HighChol, and Age were the selected variables. The tree had 8 leaves with High BP as the root.
- Purest leaf splitting rule: High BP=0 or missing High BP=1 BMI< 23.5.
- Validation misclassification rate: 0.134375.
- Next tree model was not pursued due to higher misclassification rate compared to Max 3 tree.



Competing Rules For Node 1

Split Variable	Variable Descri...	-Log(p)	Number of Bran...
HighBP	HighBP	306.3708	2
DiffWalk	DiffWalk	219.9301	2
BMI	BMI	219.6570	4
HighChol	HighChol	176.6926	2
Age	Age	132.9485	4





Fit Statistics

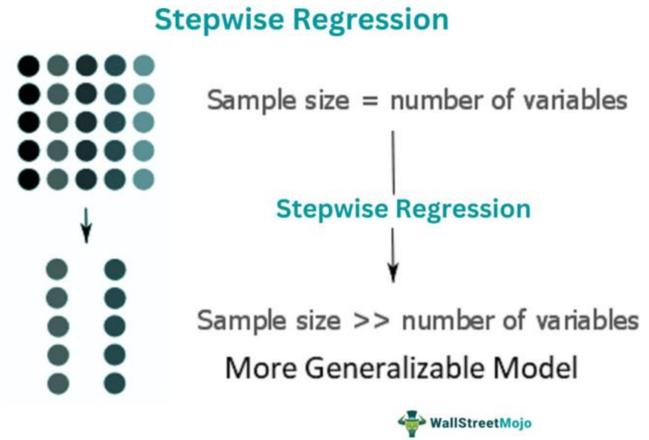
Target	Target Label	Fit Statistics	Statistics Label	Train	Validation
REP_Di...	Replace...	NOBS	Sum of F...	36267	15546
REP_Di...	Replace...	MISC	Misclass...	0.13351	0.134375
REP_Di...	Replace...	MAX	Maximu...	0.940069	0.940069
REP_Di...	Replace...	SSE	Sum of ...	7616.879	3258.348
REP_Di...	Replace...	ASE	Average ...	0.105011	0.104797
REP_Di...	Replace...	RASE	Root Av...	0.324054	0.323724
REP_Di...	Replace...	DIV	Divisor f...	72534	31092
REP_Di...	Replace...	DFT	Total De...	36267	

Model Analysis: Alternative Models

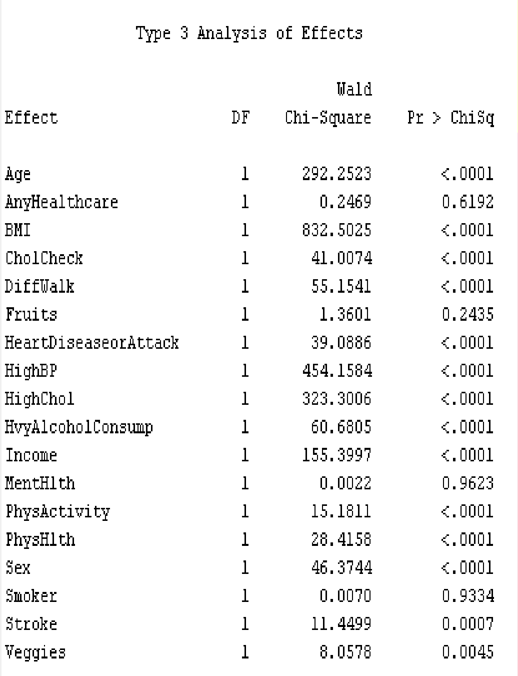
.

- Two logistic regression models were chosen:
(a) Default Regression
(b) Stepwise Regression

- Impute node was added to remove missing values and redundant variables.



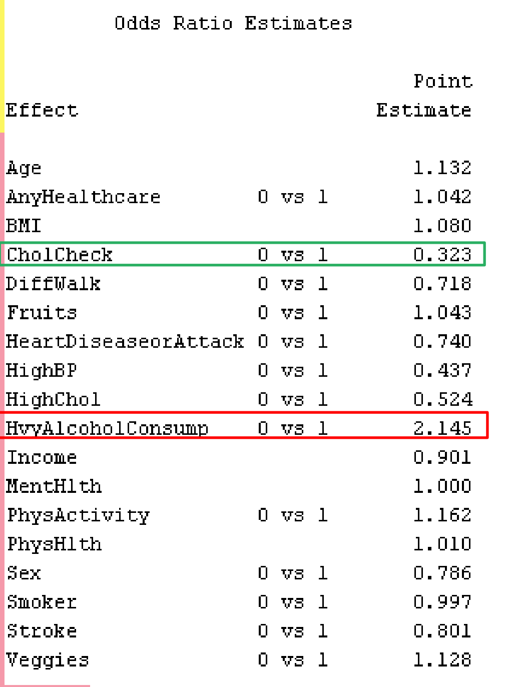
Model Analysis
Model 5- Default Regression Model



```
                 Type 3 Analysis of Effects

                                       Wald
Effect                   DF     Chi-Square    Pr > ChiSq

Age                       1       292.2523        <.0001
AnyHealthcare             1         0.2469        0.6192
BMI                       1       832.5025        <.0001
CholCheck                 1        41.0074        <.0001
DiffWalk                  1        55.1541        <.0001
Fruits                    1         1.3601        0.2435
HeartDiseaseorAttack      1        39.0886        <.0001
HighBP                    1       454.1584        <.0001
HighChol                  1       323.3006        <.0001
HvyAlcoholConsump         1        60.6805        <.0001
Income                    1       155.3997        <.0001
MentHlth                  1         0.0022        0.9623
PhysActivity              1        15.1811        <.0001
PhysHlth                  1        28.4158        <.0001
Sex                       1        46.3744        <.0001
Smoker                    1         0.0070        0.9334
Stroke                    1        11.4499        0.0007
Veggies                   1         8.0578        0.0045
```



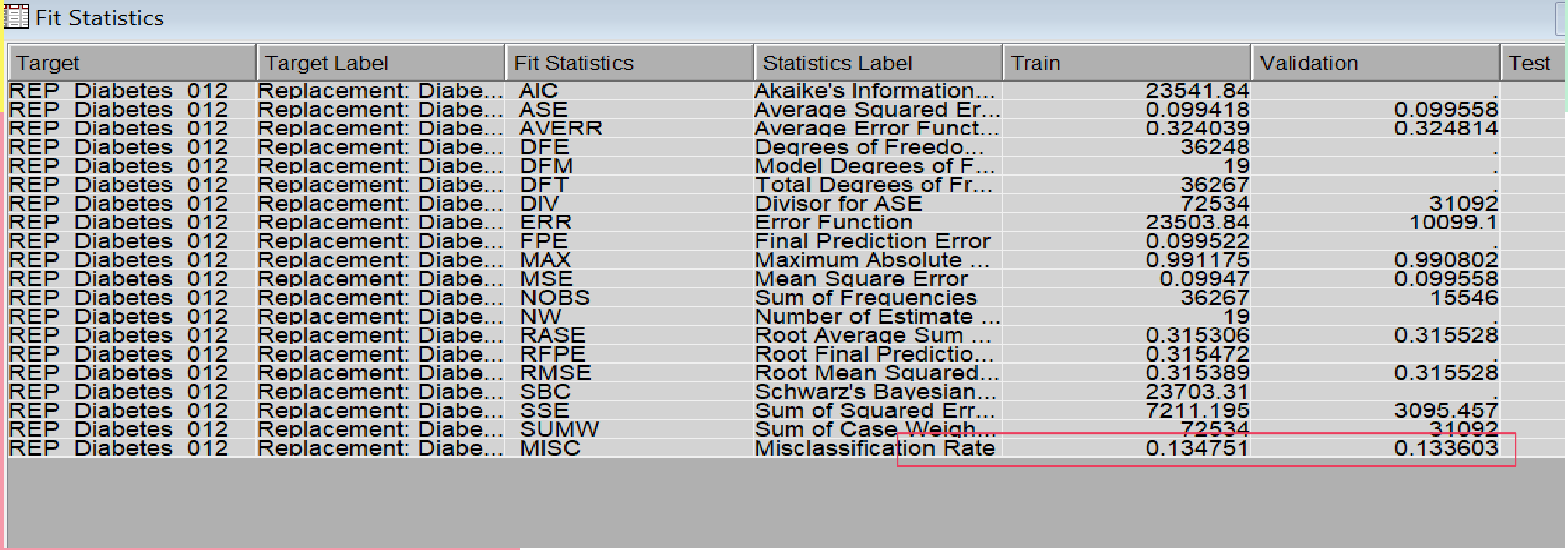
```
            Odds Ratio Estimates

                                           Point
Effect                                  Estimate

Age                                        1.132
AnyHealthcare          0 vs 1              1.042
BMI                                        1.080
CholCheck              0 vs 1              0.323
DiffWalk               0 vs 1              0.718
Fruits                 0 vs 1              1.043
HeartDiseaseorAttack   0 vs 1              0.740
HighBP                 0 vs 1              0.437
HighChol               0 vs 1              0.524
HvyAlcoholConsump      0 vs 1              2.145
Income                                     0.901
MentHlth                                   1.000
PhysActivity           0 vs 1              1.162
PhysHlth                                   1.010
Sex                    0 vs 1              0.786
Smoker                 0 vs 1              0.997
Stroke                 0 vs 1              0.801
Veggies                0 vs 1              1.128
```

- Type 3 Analysis of Effects shows Age, BMI, DiffWalk, HeartDiseaseAttack, HighBP, HighChol, HvyAlcoholConsump, Income, PhysActivity, PhysHlth, and sex are most significant (p=< 0.0001).
- Odds Ratio
shows HvyAlcoholConcsump, PhysActivity, and Age have the most significant effect on the target variable.
- Misclassification rate on validation set was 0.133606, and ASE was 0.099558.

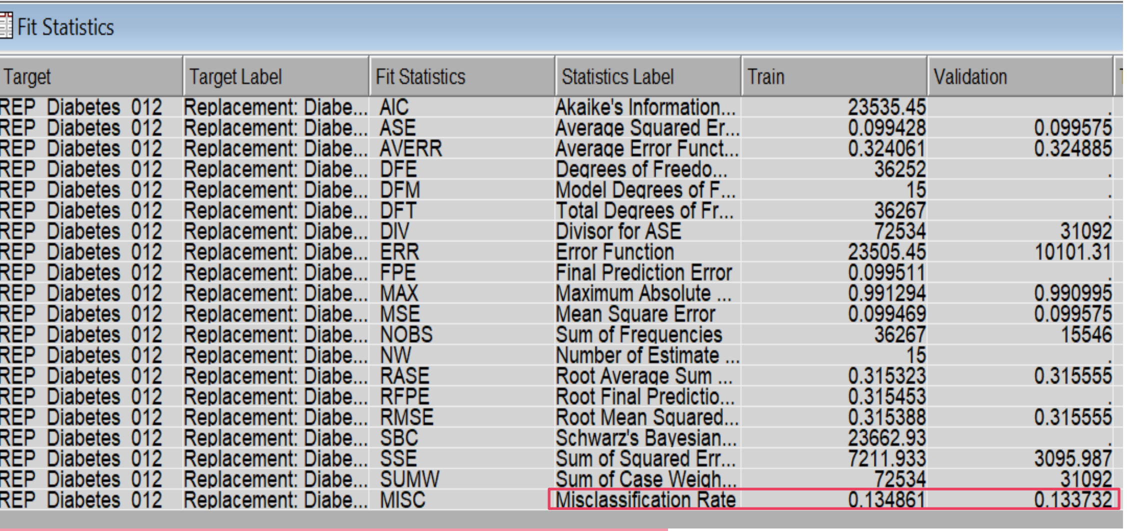
Model 6- Default Regression Model Results



Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
REP_Diabetes_012	Replacement: Diabe...	AIC	Akaike's Information...	23541.84	.	
REP_Diabetes_012	Replacement: Diabe...	ASE	Average Squared Er...	0.099418	0.099558	
REP_Diabetes_012	Replacement: Diabe...	AVERR	Average Error Funct...	0.324039	0.324814	
REP_Diabetes_012	Replacement: Diabe...	DFE	Degrees of Freedo...	36248	.	
REP_Diabetes_012	Replacement: Diabe...	DFM	Model Degrees of F...	19	.	
REP_Diabetes_012	Replacement: Diabe...	DFT	Total Degrees of Fr...	36267	.	
REP_Diabetes_012	Replacement: Diabe...	DIV	Divisor for ASE	72534	31092	
REP_Diabetes_012	Replacement: Diabe...	ERR	Error Function	23503.84	10099.1	
REP_Diabetes_012	Replacement: Diabe...	FPE	Final Prediction Error	0.099522	.	
REP_Diabetes_012	Replacement: Diabe...	MAX	Maximum Absolute ...	0.991175	0.990802	
REP_Diabetes_012	Replacement: Diabe...	MSE	Mean Square Error	0.09947	0.099558	
REP_Diabetes_012	Replacement: Diabe...	NOBS	Sum of Frequencies	36267	15546	
REP_Diabetes_012	Replacement: Diabe...	NW	Number of Estimate ...	19	.	
REP_Diabetes_012	Replacement: Diabe...	RASE	Root Average Sum ...	0.315306	0.315528	
REP_Diabetes_012	Replacement: Diabe...	RFPE	Root Final Predictio...	0.315472	.	
REP_Diabetes_012	Replacement: Diabe...	RMSE	Root Mean Squared...	0.315389	0.315528	
REP_Diabetes_012	Replacement: Diabe...	SBC	Schwarz's Bayesian...	23703.31	.	
REP_Diabetes_012	Replacement: Diabe...	SSE	Sum of Squared Err...	7211.195	3095.457	
REP_Diabetes_012	Replacement: Diabe...	SUMW	Sum of Case Weigh...	72534	31092	
REP_Diabetes_012	Replacement: Diabe...	MISC	Misclassification Rate	0.134751	0.133603	

Model Analysis
Model 6- Stepwise Regression Model

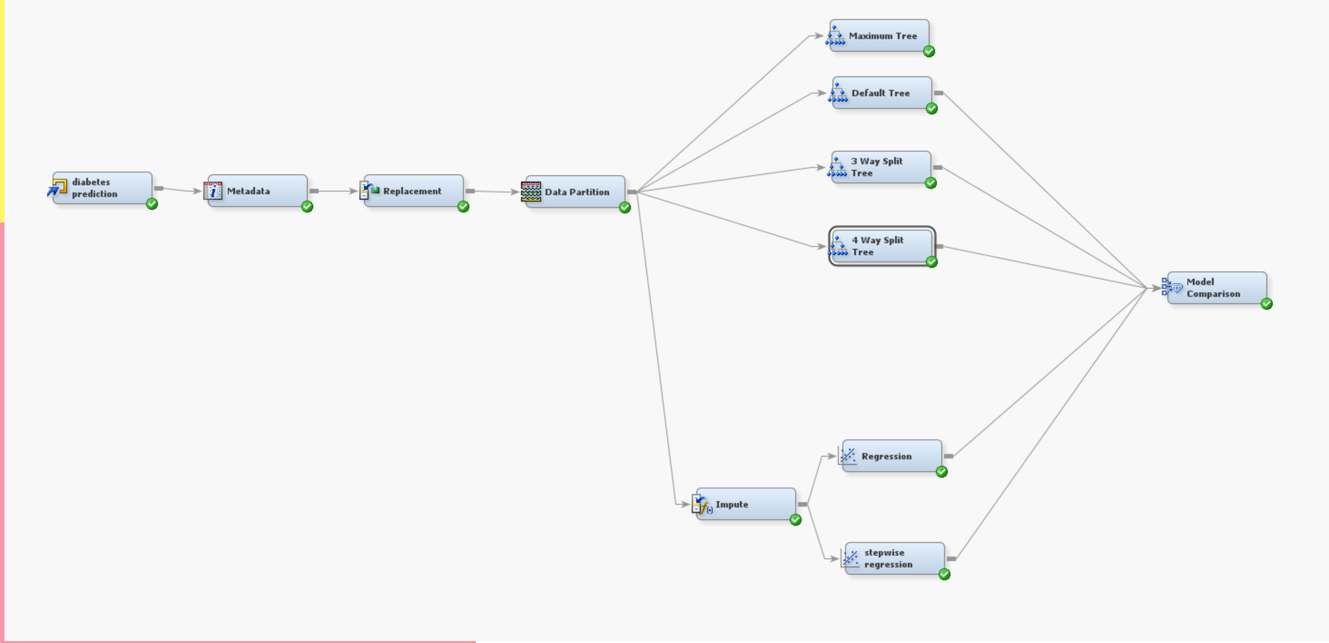


Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation
REP Diabetes 012	Replacement: Diabe...	AIC	Akaike's Information...	23535.45	.
REP Diabetes 012	Replacement: Diabe...	ASE	Average Squared Er...	0.099428	0.099575
REP Diabetes 012	Replacement: Diabe...	AVERR	Average Error Funct...	0.324061	0.324885
REP Diabetes 012	Replacement: Diabe...	DFE	Degrees of Freedo...	36252	.
REP Diabetes 012	Replacement: Diabe...	DFM	Model Degrees of F...	15	.
REP Diabetes 012	Replacement: Diabe...	DFT	Total Degrees of Fr...	36267	.
REP Diabetes 012	Replacement: Diabe...	DIV	Divisor for ASE	72534	31092
REP Diabetes 012	Replacement: Diabe...	ERR	Error Function	23505.45	10101.31
REP Diabetes 012	Replacement: Diabe...	FPE	Final Prediction Error	0.099511	.
REP Diabetes 012	Replacement: Diabe...	MAX	Maximum Absolute ...	0.991294	0.990995
REP Diabetes 012	Replacement: Diabe...	MSE	Mean Square Error	0.099469	0.099575
REP Diabetes 012	Replacement: Diabe...	NOBS	Sum of Frequencies	36267	15546
REP Diabetes 012	Replacement: Diabe...	NW	Number of Estimate ...	15	.
REP Diabetes 012	Replacement: Diabe...	RASE	Root Average Sum ...	0.315323	0.315555
REP Diabetes 012	Replacement: Diabe...	RFPE	Root Final Predictio...	0.315453	.
REP Diabetes 012	Replacement: Diabe...	RMSE	Root Mean Squared...	0.315388	0.315555
REP Diabetes 012	Replacement: Diabe...	SBC	Schwarz's Bayesian...	23662.93	.
REP Diabetes 012	Replacement: Diabe...	SSE	Sum of Squared Err...	7211.933	3095.987
REP Diabetes 012	Replacement: Diabe...	SUMW	Sum of Case Weigh...	72534	31092
REP Diabetes 012	Replacement: Diabe...	MISC	Misclassification Rate	0.134861	0.133732

- Variables selected by the algorithm: Age, BMI, CholCheck, DiffWalk, HeartDiseaseAttack, HighBP, HighChol, HvyAlcoholConsump, Income, PhysActivity, PhysHlth, Sex, and Veggies
- Misclassification validation rate: 0.133732
- ASE value: 0.099575

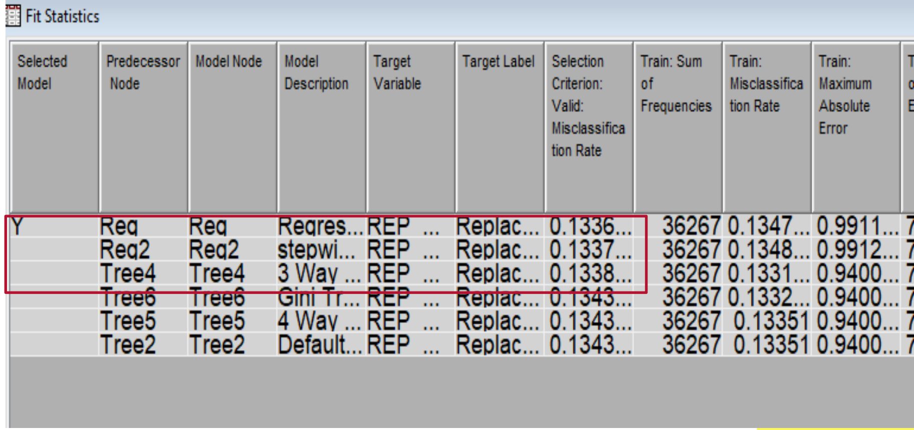
Comparison of Experimental results



Model Comparison Node was added, and all the models were assessed.

Comparison of Experimental results by Fit Statistics

- Fit statistics measure how well a model fits a dataset by comparing the predicted outcome to the actual outcome.
- The lower the misclassification rate, the better the model performs.
- The regression model in the study showed slightly lower misclassification rates and Average squared errors compared to the tree models.
- The Default Regression Model had the lowest validation misclassification rate of 0.1336, indicating that it is the best model for predicting an individual's risk of type 2 diabetes.
- The Stepwise Regression was the second-best choice for the model.



Fit Statistics

Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Valid: Misclassification Rate	Train: Sum of Frequencies	Train: Misclassification Rate	Train: Maximum Absolute Error	T o E
Y	Reg	Reg	Regres...	REP ...	Replac...	0.1336...	36267	0.1347...	0.9911...	7
	Reg2	Reg2	stepwi...	REP ...	Replac...	0.1337...	36267	0.1348...	0.9912...	7
	Tree4	Tree4	3 Way ...	REP ...	Replac...	0.1338...	36267	0.1331...	0.9400...	7
	Tree6	Tree6	Gini Tr...	REP ...	Replac...	0.1343...	36267	0.1332...	0.9400...	7
	Tree5	Tree5	4 Way ...	REP ...	Replac...	0.1343...	36267	0.13351	0.9400...	7
	Tree2	Tree2	Default...	REP ...	Replac...	0.1343...	36267	0.13351	0.9400...	7

Conclusions

- **Age, Heavy Alcohol Consumption, and Physical Activity had the greatest effects on the target variable (diabetes_012)according to our best predictive model Logistic Regression**
- The findings emphasize the importance of lifestyle variables and health indicators in determining the likelihood of developing Type 2 Diabetes.
- The research findings can be utilized to improve risk-taking individual prevention strategies and create more precise prediction models.

References

- Cowap, N. (2015). *Diabetes*. Mercury Learning & Information. https://eds.s.ebscohost.com/eds/ebookviewer/ebook/bmxlYmtfXzE4MDkxMDVfX0FO0?sid=f4975b1b-055e-427f-91e2-60ccb3be55c6@redis&vid=1&format=EB&rid=1
- Diabetes Research Institute Foundation. (2023, March 31). *Diabetes Statistics.* https://diabetesresearch.org/diabetes-statistics/#:~:text=37.3%20million%20people%2C%20or%2011.3,%2C%20economic%2C%20and%20ethnic%20backgrounds.
- Teboul, A. (2021, November 8). *Diabetes Health Indicators Dataset*. Kaggle. https://www.kaggle.com/datasets/alexteboul/diabetes-health-indicators-dataset
- Ullah, Z., Saleem, F., Jamjoom, M., Fakieh, B., Kateb, F., Ali, A. M., & Shah, B. (2022). Detecting High-Risk Factors and Early Diagnosis of Diabetes Using Machine Learning Methods. *Computational Intelligence & Neuroscience*, *2022*, 1–10. https://doi.org/10.1155/2022/2557795

Do you have any questions?