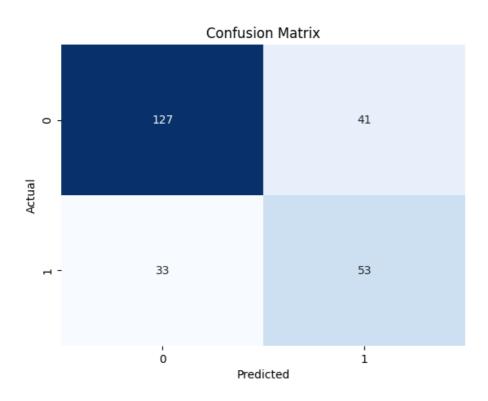
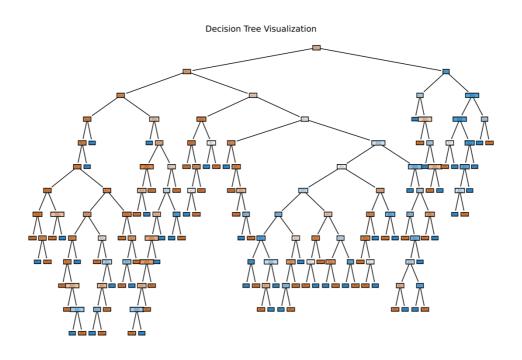
2.1 Tree

Uni-variate Tree Method





Model Performance:

- The univariate Decision Tree achieved a 70.87% accuracy on the test set.
- The confusion matrix details True Positives (53), True Negatives (127),
 False Positives (41), and False Negatives (33).

Tree Complexity:

- The Decision Tree comprises 177 nodes, indicating decision points in the structure.
- A depth of 12 suggests a relatively complex model, potentially prone to overfitting.

Visualizations:

- The confusion matrix provides a clear breakdown of prediction outcomes.
- The Decision Tree visualization visually represents feature usage in predictions.

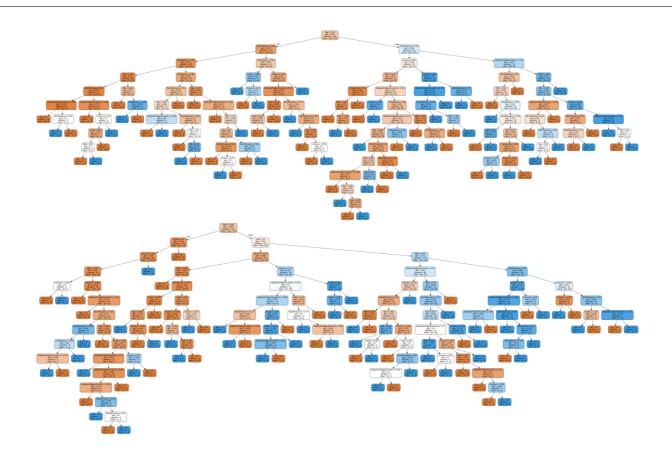
Analysis of Results:

- The model exhibits moderate accuracy, showing promise in diabetes prediction.
- Areas for improvement include addressing false positives and negatives.
- Model complexity may be a concern; tuning hyperparameters or exploring alternatives is advisable.

Personal Considerations:

- Decision Trees offer interpretability but may struggle with data complexities.
- Further experimentation, like hyperparameter tuning, can enhance model robustness.
- Balancing interpretability and complexity is crucial in practical applications.
- In summary, the analysis guides refining the Decision Tree model, emphasizing areas of improvement for more effective diabetes prediction.

Random Forest Programates c 3.121 | Agr c 0.144 | gmi = 0.460 | samples = 136 | samples = 19 | value = [10.15] | | Column | C | Gai-Thickness c 1,022 | Gai-Thickness c 1,022 | Gai-Thickness c 1,534 | Gai-| Open = 0.0 | Ohacose < 0.192 | Open = 0.202 | Open = 0.202 | Open = 0.202 | Open = 12 | Open = 12 | Open = 115 | Open = | SecThebrence 1,20 | gr = 1,34 | gr = 0.50 | secples 1 | secples Diabetimil redgreef function c 0.622 gris = 0.400 samples = 9 value = [3, 4] pair = 60 samples = 3 value = [5, 0] | SM < 0.220 | pp = 0.644 | pp = 0.05 | pp = 0.044 | pp = 0.05 | pp = 0.044 | pp = 0.05 | girs = 0.0 samples = 7 value = [10, 0] $\begin{array}{c} \text{gris} = 0.0 \\ \text{Surption} \equiv 1 \\ \text{subset} = [1,0] \\ \text{subset} = [2,1] \\ \end{array}$ (ME < 1.70 gas + 0.400 samples + 40 value + (25, 36) Ghazare 4-0.272 gra = 0.405 seriptina = 53 union = (54, 23) Age = 0.453 grs = 0.465 samples = 115 value = (76.16) ObsedProcuse s -0.777 grs = 0.466 hergins = 33 solue = [34,33] Programa s 1.554 gri = 0.336 gri = 0.336 samples = 23 sale = (33.7] value = [75, 25] Commercial pre-tot sension: 4 select js 0 lengths - 34 select js 0 Constitution of the consti | App 1 0 000 | Septem 1 0 000 | Septem 1 16 | Septem 1 16 | The C + 1.040 | gis + 0.01 | | Per | CO | Per | | Part | Tage class green and the same a Sair/Swissess < 0.528 gas = 6.5 samples = 2 water = [1, 1] gen of the control of $\begin{array}{c} \text{posito} \\ \text{posito} \\$ Stanffladowno v 1,546 ges + 0,453 samples + 350 water + 550,1115 \$86 x 0.968 ger = 0.325 namples = 10 value = \$10, 10] Apr + 0.764 gar + 3.66 margins + 765 margins + 7 margi Observe c 1.007 gas = 3.407 sangles = 55 value = [41, 47] Embernafridgeseri avction c 2.755 gas + 6.456 samples + 76 sept. (861 + 0.2(3) ga = 0.4(3) samples = 60 sala = (60.30) Apr < 4:000 ga = 5:00 par(5) = 22 selue = 20, 50 selue = 20, 50 | Mark | SME | | Mark | SME | | Mark | SME | Mark | gen = 0.07 semples = 2 selber = (2 ct) (SAR Phalaments = -0.500) semples = 2 value = (2 ct) WATER CONTROL OF THE PROPERTY THE PROPERTY OF THE PROPERTY O Glacone < 1.166 ges = 0.487 samples = 66 value = 60 60 (Bit < 0.219 gri = 0.429 samples = 200 Insulin c - 0.482 glet = 0.482 samples = 143 value + (129, 100) Age c 0.571 gri = 0.490 narrgins = 95 value = \$66, 645 gri = 0.0 samples = 23 satus = [31, 0] satus = [51, 0] Programmers 5 (3.00) gas + 0.002 sergine + 6] subs + (61.2) usbs + (62.2) Parties (0,200) 271 = 9 unker (10, 20) Apr x 2,406 (pin t 0,40) bengin x 2,50 bengin x 2,600 bengin x 2,600 bengin x 2,600 bengin x 2,700 unker (10, 2,70) The state of the s TOTAL | Control of the cont | Piopurcon 6 1.56 | Sonthistories 4 0.150 | pri +0.02 | Sangton 8 | Sonthistories 4 0.150 | pri +0.02 | Sangton 1 0.02 | San | Translation | CONTROL OF THE PROPERTY OF THE water 1 married water 1 marrie ger + 0.0 secptor + 3 secptor + 3 secptor + 2 | per | 0.0 | per SkryTrickness c 0.654 ger = 0.5 samples = 2 samples = [1, 1] Standinouse C 0.307 grs = 0.444 samples = 5 water = 54.21 pri = 0.0 samplin; = 1 salue = (0.1) ges = 60 samples = 1 salar = 31.50 Apr 1 G-400 ges = 0.32 samples = 4 salar = 16.50 CMH ± 0.517 grv : 0.5 samples : 2 when : [1, 1] ger = 10 mengha = 1 mengha = 1 mengha = 1



Model Performance:

- The Random Forest achieved a higher accuracy of 75.20% compared to the Decision Tree.
- The confusion matrix highlights 139 True Positives, 29 True Negatives,
 34 False Positives, and 52 False Negatives.

Random Forest Complexity:

 The ensemble comprises 100 trees with an average of 173.82 nodes per tree.

Visualizations:

- The confusion matrix provides a detailed breakdown of predictions for evaluation.
- Decision Trees within the Random Forest offer insights into individual tree contributions.

Analysis of Results:

- The Random Forest demonstrates improved accuracy, addressing some issues observed in the Decision Tree.
- False positives and negatives are reduced, indicating enhanced predictive performance.
- o The ensemble nature contributes to a more robust and stable model.

Personal Considerations:

- The Random Forest's ensemble approach proves effective in handling complexities.
- Improved accuracy signifies enhanced predictive power for diabetes cases.
- The ensemble's average node count suggests a balance between complexity and generalization.
- In conclusion, the Random Forest outperforms the individual Decision Tree, showcasing improved accuracy and a reduction in misclassifications. The ensemble nature enhances model robustness, making it a promising choice for diabetes prediction. Further analysis and optimization can refine the model's performance.