# 2.1 Tree

## Uni-variate Tree Method

## 

A diagram of a tree

Description automatically generated

**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 ForestA diagram of a flowchart

Description automatically generatedA diagram of a network

Description automatically generated with medium confidenceA blurry image of a diagram

Description automatically generatedA diagram of a person

Description automatically generatedA blurry image of a diagram

Description automatically generatedA diagram of a company

Description automatically generated with medium confidence

**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.
  + 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.