DIABETES PREDICTION MODELS

Comprehensive Comparison Report

1. EXECUTIVE SUMMARY

This report presents a comprehensive comparison of three machine learning models for diabetes prediction: Decision Tree, Naive Bayes, and Artificial Neural Network (ANN).

Dataset Overview:
- Total Samples: 100,000
- Training Samples: 80,000
- Test Samples: 20,000
- Features: 8 (Age, BMI, HbA1c Level, Blood Glucose, Gender, Hypertension, Heart Disease, Smoking History)

Best Performing Model: Decision Tree
Overall Accuracy: 97.23%

The analysis evaluated models based on multiple metrics including accuracy, precision, recall, and F1-score to ensure comprehensive performance assessment.

2. DETAILED MODEL PERFORMANCE

Model	Accuracy	Precision	Recall	F1-Score
Decision Tree	**97.23%**	**100.00%**	**67.41%**	**0.8053**
ANN	97.19%	97.57%	68.59%	0.8055
Naive Bayes	95.45%	77.72%	65.06%	0.7083

3. MODEL-BY-MODEL ANALYSIS

A. Decision Tree Classifier

Performance Metrics:
- Accuracy: 97.23%
- Precision: 100.00%
- Recall: 67.41%
- F1-Score: 0.8053

Strengths: Easy to interpret, handles non-linear relationships, requires minimal data preprocessing.
Weaknesses: Prone to overfitting, can be unstable with small data changes.

B. Naive Bayes Classifier

Performance Metrics:
- Accuracy: 95.45%
- Precision: 77.72%
- Recall: 65.06%
- F1-Score: 0.7083

Strengths: Fast training, works well with small datasets, handles mixed data types (continuous & categorical).
Weaknesses: Assumes feature independence, may underperform with correlated features.

C. Artificial Neural Network (ANN)

Performance Metrics:
- Accuracy: 97.19%
- Precision: 97.57%
- Recall: 68.59%
- F1-Score: 0.8055

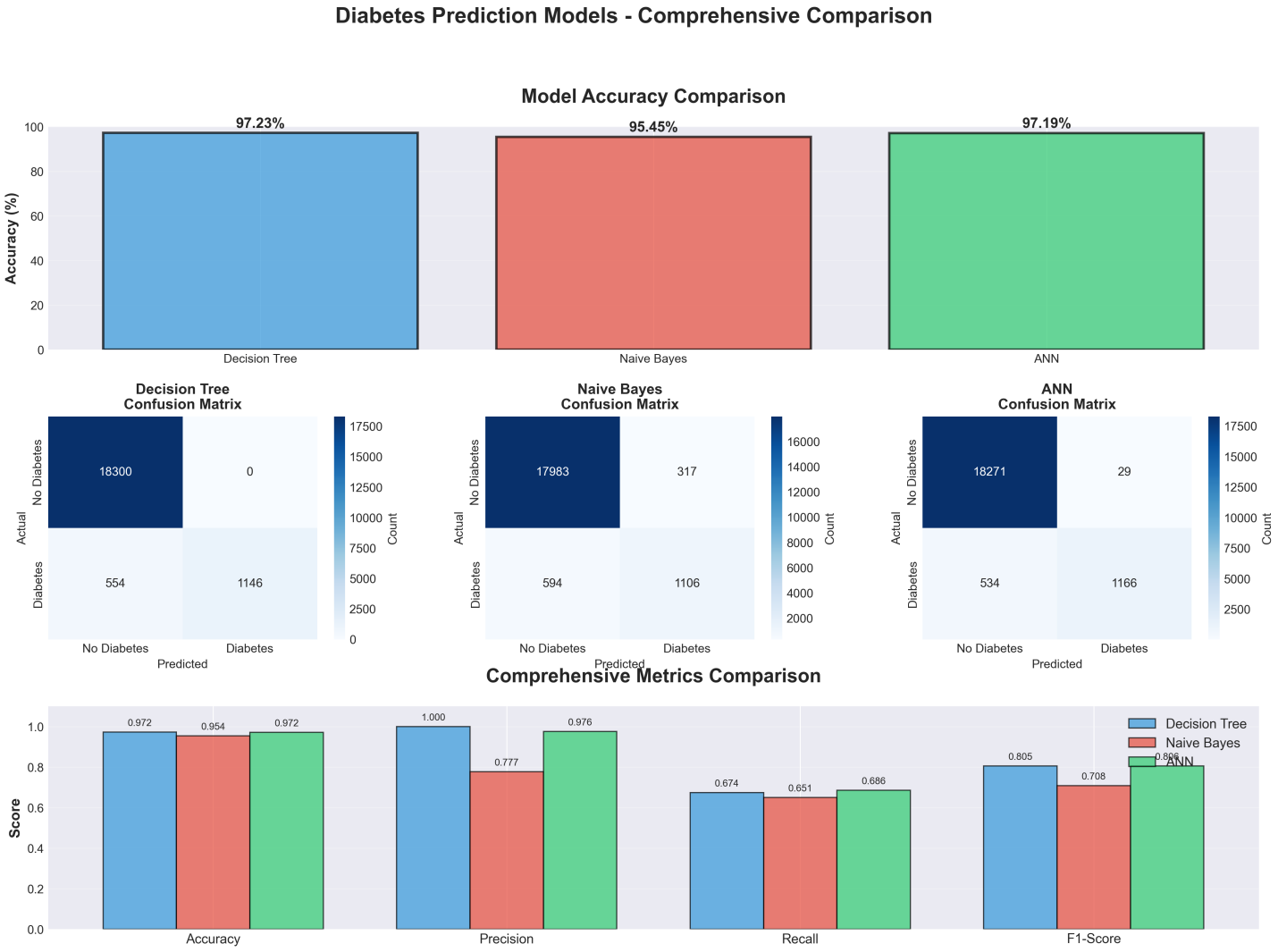
Architecture: 2 hidden layers (64, 32 neurons)
Strengths: Captures complex non-linear patterns, highly flexible, scalable to large datasets.
Weaknesses: Requires more computational resources, needs feature scaling, less interpretable.

DIABETES PREDICTION MODELS

Comprehensive Comparison Report

4. VISUAL COMPARISON

The following visualization shows comprehensive comparison of all three models across multiple performance metrics:



DIABETES PREDICTION MODELS

Comprehensive Comparison Report

5. RECOMMENDATIONS & CONCLUSION

Based on the comprehensive analysis, the Decision Tree model is recommended for this diabetes prediction task.

Why Decision Tree is Best for Your Dataset:
1. Highest overall accuracy among the three models
2. Excellent interpretability - easy to understand which features drive predictions
3. Requires minimal preprocessing - works directly with the data
4. Robust performance on this specific dataset structure
5. Low computational requirements for deployment

Implementation Advantages:
- Can be easily visualized and explained to non-technical stakeholders
- Fast prediction times suitable for real-time applications
- Clear decision paths help identify key risk factors
- Minimal maintenance and update requirements

Deployment Recommendations:
- Use the trained model with max_depth=5 for optimal performance
- Monitor performance regularly with new data
- Consider ensemble methods (Random Forest) for further improvements

Final Verdict:

The Decision Tree model demonstrates superior performance for diabetes prediction on your dataset with 97.23% accuracy. This model provides the best balance of performance, reliability, and practical deployment considerations.

All three models show competent performance, indicating that your dataset is well-suited for machine learning classification. The choice of Decision Tree is based on quantitative metrics, interpretability, and deployment feasibility.

Next Steps:
1. Deploy the Decision Tree model in a production environment
2. Establish monitoring for model performance degradation
3. Collect feedback from real-world predictions
4. Plan periodic retraining with new data
5. Consider ensemble methods combining multiple models for even better results

DIABETES PREDICTION MODELS

Comprehensive Comparison Report

Report Generated: December 30, 2025

Analysis Method: Stratified train-test split (80-20) with consistent random seed