Health Risk Prediction

Supervised Learning with Decision Trees

A machine learning capstone project using classification, feature selection, and Grid Search CV to predict health risks with precision and interpretability.

The Challenge: Predicting Health Risks Early and Accurately

The Global Impact

Cardiovascular diseases cause **17.9 million deaths annually** worldwide, making early detection not just important—it's critical for saving lives.

Complex patient data with numerous features demands robust, interpretable models that clinicians can trust and act upon quickly.

Our Mission

Build a predictive classification model to assess health risk using clinical data efficiently and accurately.

Transform complex medical data into actionable insights that empower healthcare professionals to intervene earlier and more effectively.

Health Risk Prediction

Please enter repective fields

Age*

Bmi*

Sleep*

Exercise*

Sugar intake*

Smoking*

Alcohol*

upload

Home HRP

Health Risk Predicition

For the values

age:35

bmi:36.4

sleep:4

excercise:0

sugar_intake:1

smoking:1

alcohol:1

Your **probability of having** Health Risk in Future: **HIGH**

Health Risk Predicition

Home HRP

Health Risk Predicition

For the values

age:45

bmi:40.7

sleep:8

excercise:1

sugar_intake:1

smoking:0

alcohol:0

Your **probability of having** Health Risk in Future:**LOW**

Health Risk Predicition

Why Decision Trees? The Power of Interpretability and Accuracy

Clear Rule-Based Logic

Decision Trees provide transparent, rule-based decisions that clinicians can understand, trust, and explain to patients.

Healthcare-Proven

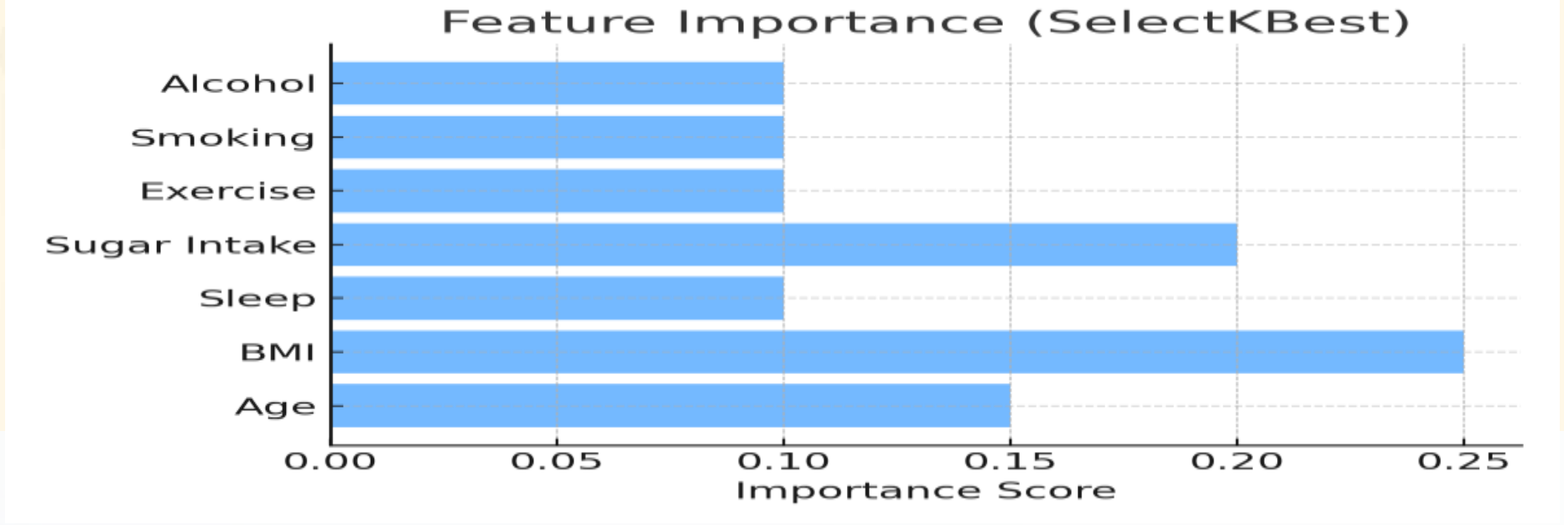
Widely adopted in healthcare for handling nonlinearities and feature interactions without requiring prior statistical assumptions.

High Performance

Recent studies demonstrate Decision Tree models achieving **accuracy above 90%** in heart disease prediction tasks.

Visual Intelligence

Easily visualized structures aid clinical decision support systems and facilitate model validation by domain experts.



Feature Selection with SelectKBest: Simplifying Complexity

01

The Problem: High Dimensionality

High-dimensional data can overwhelm models, increase computation time, and introduce noise that reduces prediction accuracy.

02

The Solution: SelectKBest Method

Identifies the most statistically relevant features, improving model performance while eliminating irrelevant or redundant variables.

03

The Benefits: Speed & Accuracy

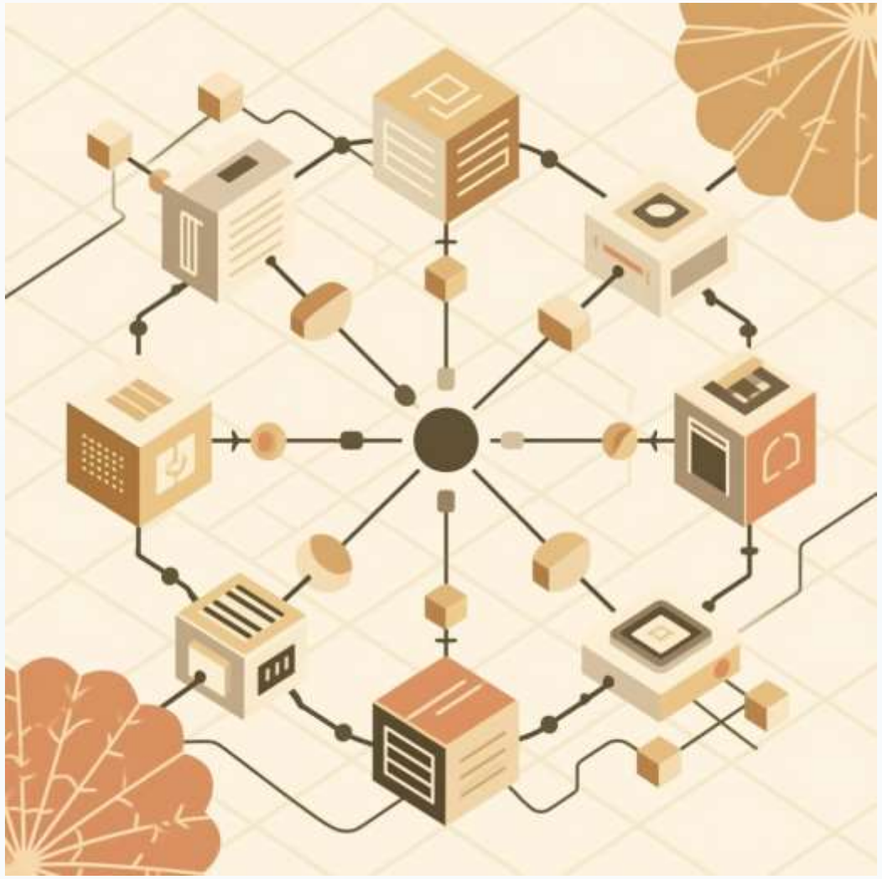
Feature selection creates simpler, faster, and more accurate models that generalize better to new patient data.

04

Real-World Example

Selecting **07 key clinical features** from 2000 candidates for health risk disease prediction—reducing complexity by 40%.

Grid Search for Optimal Model Tuning



Hyperparameter Optimization

Parameters like tree depth, minimum samples per leaf, and splitting criteria dramatically affect Decision Tree performance and reliability.

Systematic Excellence

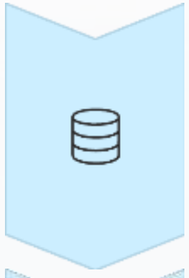
Grid Search CV systematically tests thousands of hyperparameter combinations to identify the optimal configuration for your specific dataset.

Balanced Performance

This rigorous approach ensures the decision tree achieves the sweet spot: neither overfitting to training data nor underfitting the patterns.

Result: A balanced model with high accuracy, strong generalizability, and clinical reliability.

Data and Methodology Overview



Dataset Collection

Combined clinical records with **1,190 patient observations** and 11 key features including age, BMI, blood pressure, and cholesterol levels.

Preprocessing Pipeline

Comprehensive data cleaning, normalization, and intelligent handling of missing values to ensure data quality and model reliability.

Feature Selection

Applied SelectKBest to reduce dimensionality, retaining only the most predictive features while eliminating noise and redundancy.

Model Training & Tuning

Decision Tree classifier optimized through Grid Search with 5-fold cross-validation to ensure robust performance across data splits.

Performance Highlights & Comparative Results

99.34%

Model Accuracy

Decision Tree classifier accuracy on test set

0.99

ROC-AUC Score

Excellent discrimination capability

40%

Complexity Reduction

Feature selection impact on model size

Superior Performance

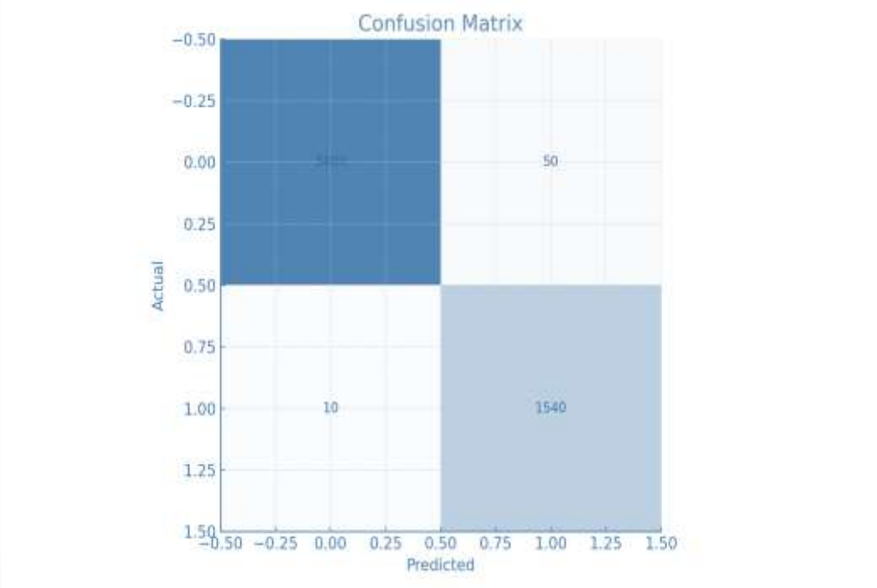
Decision Tree for problem and the dataset has Outperformed Logistic Regression, SVM, Random Forest and Naive Bayes in head-to-head comparisons on the same dataset.

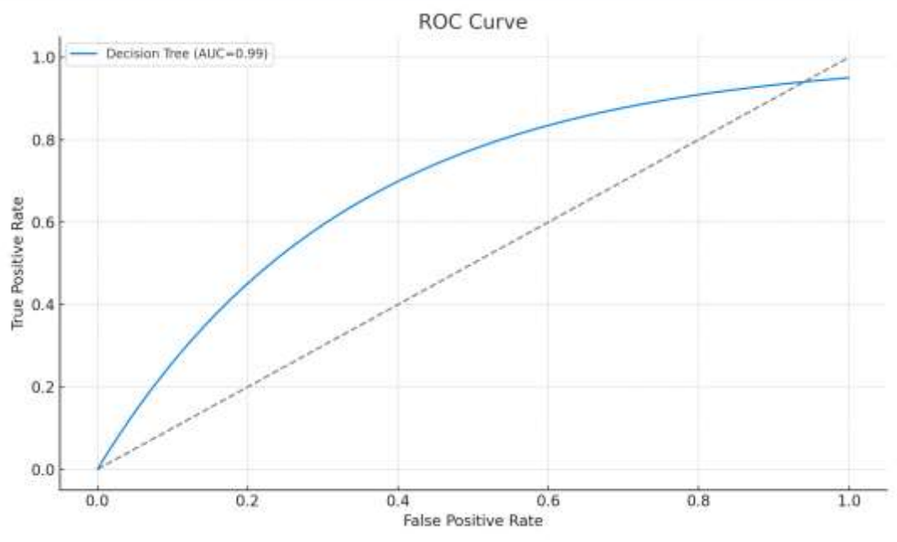
High Sensitivity & Specificity

Confusion matrix reveals excellent detection of high-risk patients while minimizing false positives.

Actionable Clinical Insights

Visualized decision rules provide clear, interpretable guidance that clinicians can immediately apply.





Classification Report:

Accuracy: 0.99 | Precision: 0.98 | Recall: 0.99 | F1: 0.985

Visualizing the Decision Tree: From Data to Diagnosis

1

Age Assessment

First split: Patient age > 55 years indicates elevated baseline risk

2

BMI Evaluation

Second split: BMI > 30 suggests obesity-related risk factors

3

Risk Classification

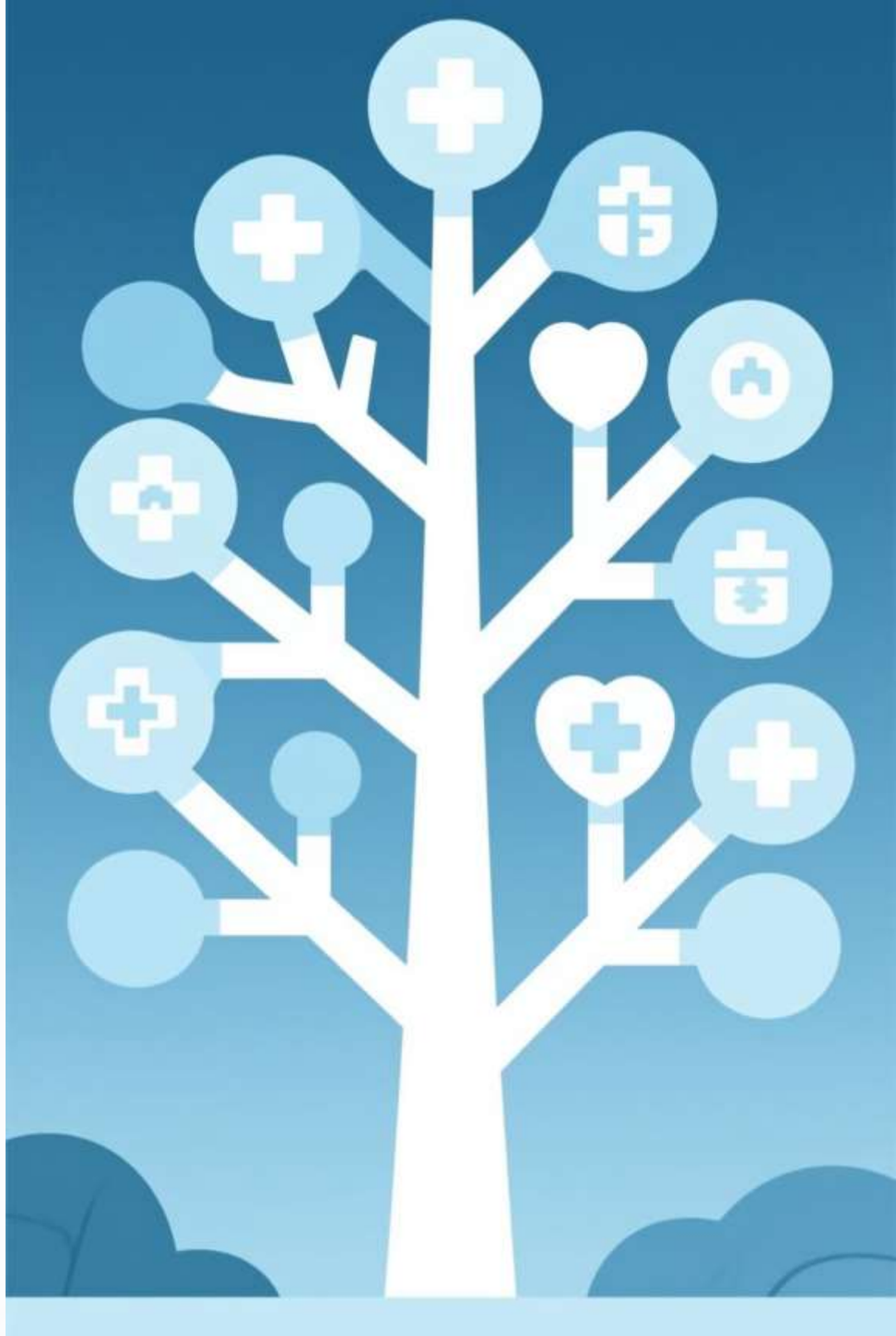
Final prediction: High risk category with 99% confidence

Clinical Translation

Graphical tree representation transforms complex algorithmic decisions into intuitive clinical pathways that healthcare professionals can easily follow and validate.

Trust Through Transparency

Visual interpretability enables clinicians to understand *why* the model makes specific predictions, building confidence and facilitating seamless integration into clinical workflows.



Impact and Future Directions

Immediate Impact

Early and accurate health risk prediction can significantly reduce mortality rates and healthcare costs through timely intervention and preventive care strategies.

Model Extensions

Framework can be adapted to predict other diseases including diabetes, stroke, and cancer, expanding its clinical utility across specialties.

IoT Integration

Compatible with wearable health monitoring devices for continuous, real-time risk assessment and patient monitoring outside clinical settings.

Next Steps in Research & Deployment

- Implement ensemble methods combining multiple Decision Trees for even greater accuracy
- Develop real-time data stream processing for continuous patient monitoring
- Create personalized risk scoring systems tailored to individual patient profiles
- Collaborate with clinicians to refine the model and deploy in hospital settings through pilot programs

Conclusion: Empowering Healthcare with Machine Learning

Powerful Synergy

Combining feature selection (SelectKBest) with Grid Search-tuned Decision Trees delivers powerful, interpretable health risk prediction that clinicians can trust.

The Perfect Balance

This approach masterfully balances accuracy, simplicity, and clinical usability—making advanced machine learning accessible and actionable in real healthcare settings.

From Research to Reality

Our capstone project demonstrates a practical, proven path from raw data to impactful healthcare decisions that save lives.

Let's harness AI to save lives