Researching Machine Learning Methods for Preventing Cardiovascular Diseases

V.V. Mokshin, D.N. Grigoreva, A.R. Faskhutdinova, B.A. Garafutdinov

The purpose of the study

The aim of this work is to develop a model that can quickly and accurately identify the risk of cardiovascular diseases based on a small number of input parameters.

Features	Feature's description
x_1	Age
x_2	Sex
x_3	Chest pain type (4 values)
x_4	Resting blood pressure
x_5	Serum cholestoral in mg/dl
x_6	Fasting blood sugar > 120 mg/dl
x_7	Resting electrocardiographic results (values 0,1,2)
x_8	Maximum heart rate achieved
x_9	Exercise induced angina
x_{10}	Oldpeak = ST depression induced by exercise relative to rest
x_{11}	The slope of the peak exercise ST segment
x_{12}	Number of major vessels (0-3) colored by flourosopy
y_1	CD: 0 = normal; 1 = fixed defect; 2 = reversable defect

Fig. 1. Features for analysis

Logistic regression

The y distribution function for a given:
P{y|x}=f(θ^T*x)^y*(1-f(θ^T*x))^(1-y),y∈{0,1}

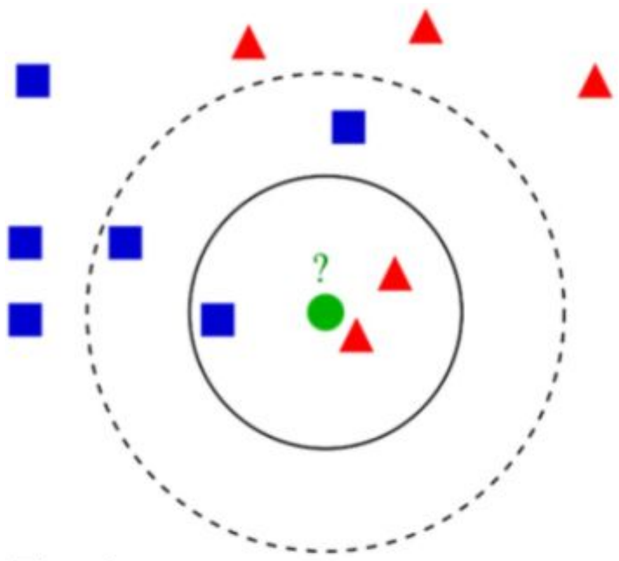

Fig. 2. An example of k-nearest neighbor classification

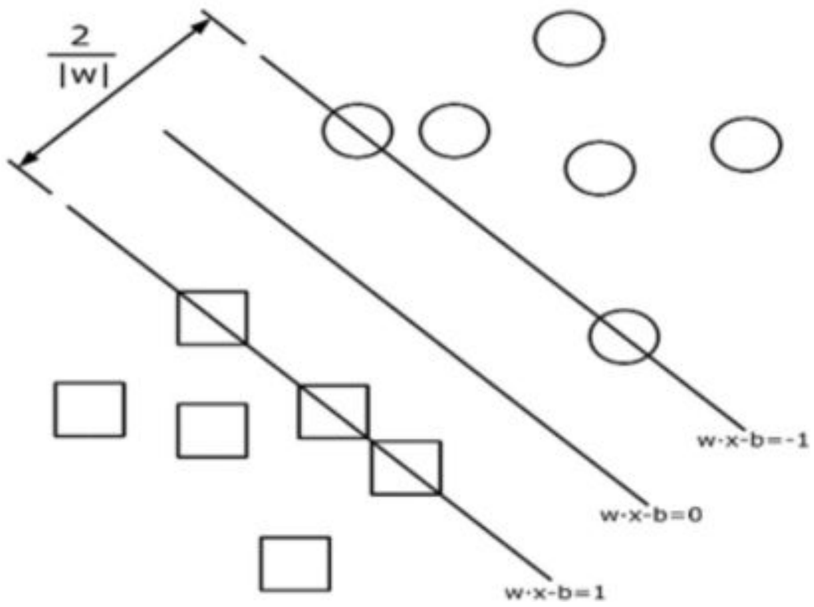

Fig. 3. Optimal separating hyperplane for support vector machines

Bayes' Theorem

$$P(y \mid x_1,...,x_n) = \frac{P(y)P(x_1,...,x_n \mid y)}{P(x_1,...,x_n)}$$

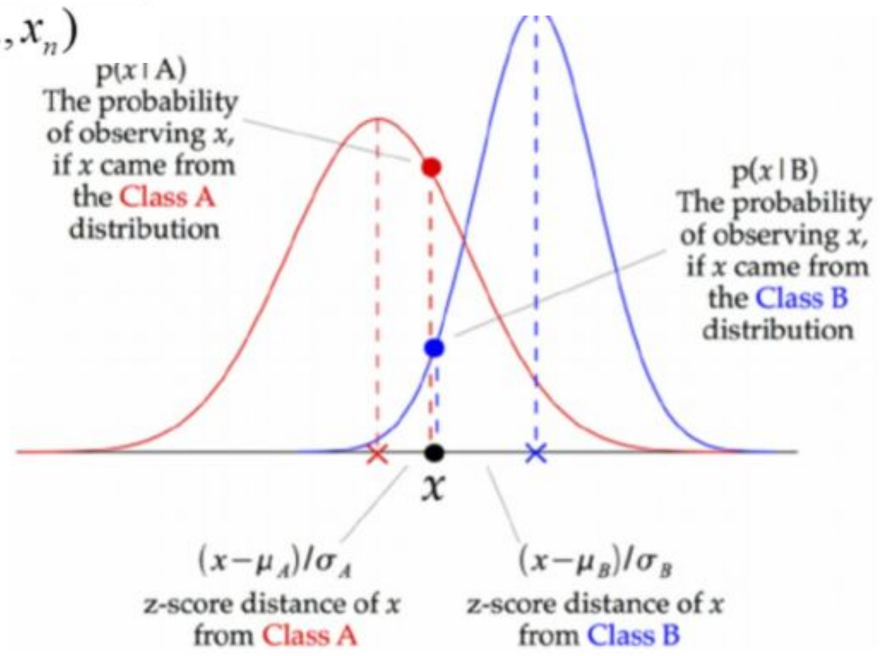

Fig. 4. Gaussian Naive Bayes by Opengenius

Decision tree

$$(x, Y) = (x_1, x_2, x_3, ..., x_k, Y)$$

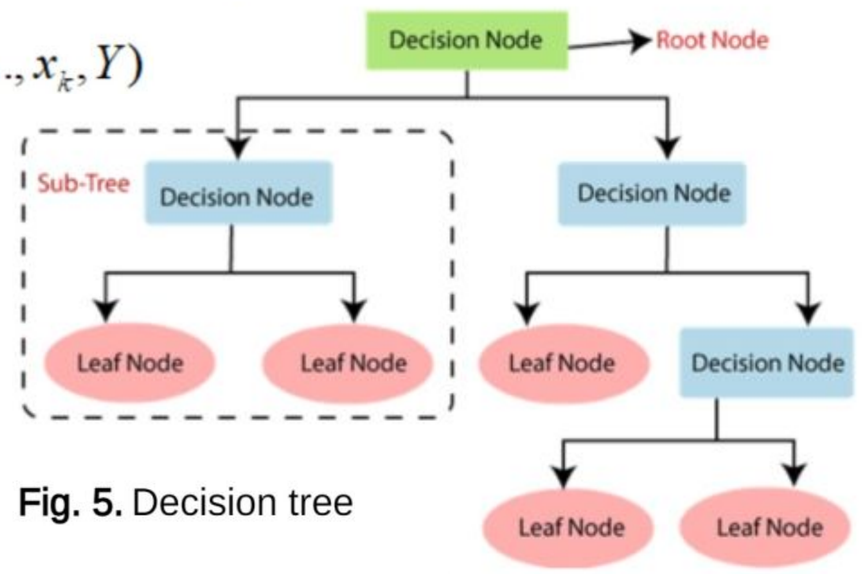

Fig. 5. Decision tree

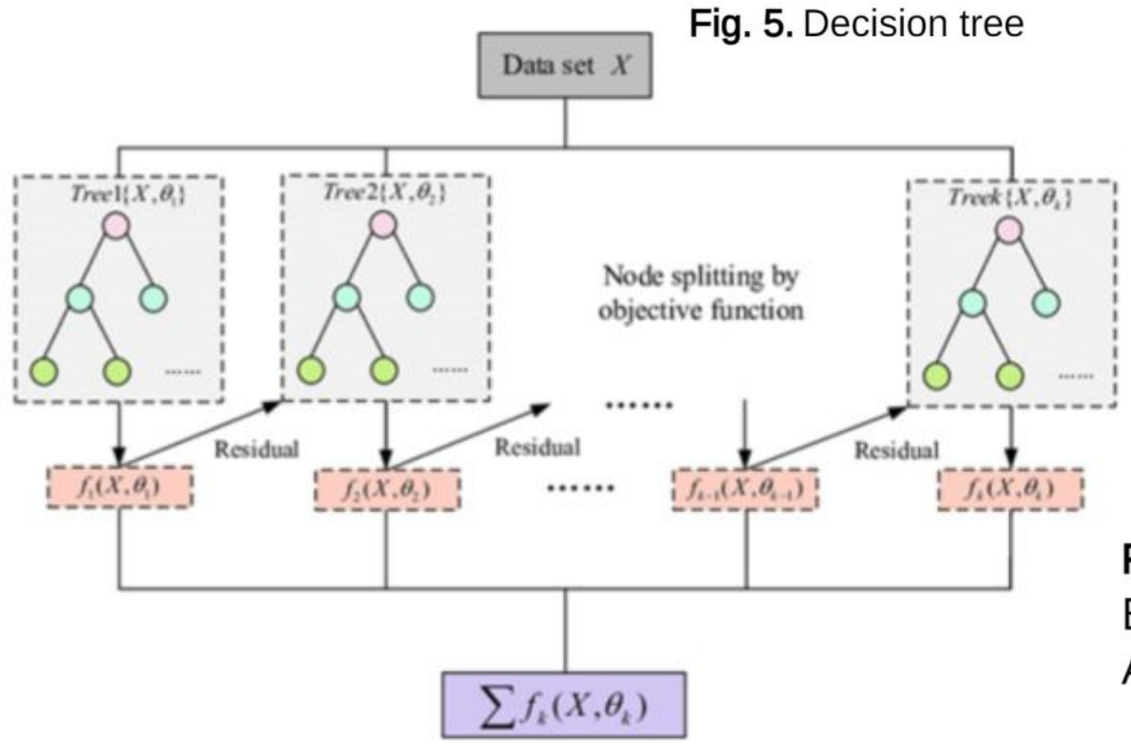

Fig. 6. Boosting Algorithm

Method	Accuracy
K-nearest neighbors	95.61%
random forest	88.78%
Gradient boost	86.83%
Logistic regression	83.9%
Support vectors	83.9%
Decision tree	83.9%
Gaussian Naive Bayesian	82.44%

Fig. 7. Accuracy of the methods used

Conclusion

An analysis of existing methods for predicting heart disease was carried out, various machine learning methods were proposed to build a model for predicting outcome indicators, and the most optimal k-nearest neighbors method was selected. Feature selection was carried out using various methods and learning outcomes were obtained for each of them.

KAZAN NATIONAL RESEARCH TECHNICAL UNIVERSITY
NAMED AFTER A. N. TUPOLEV - KAI
10 K.Marx St., Kazan,
Tatarstan 420111, Russia (www.kai.ru)