

Predicting Cardiovascular Disease with Classification Machine Learning

Lama Ibrahim Alharbi



SDAIA
الهيئة السعودية للبيانات
والذكاء الاصطناعي
Saudi Data & AI Authority



METIS

kaggle

TABLE OF CONTENTS

01

Introduction

02

Data description

03

Feature Engineering

04

Models and Results

INTRODUCTION





Machine learning & Healthcare



Growing medical
demands



Improve
operations



Lower cost

How Machine Learning in Health Informatics Impacts Healthcare



Recordkeeping



Data Integrity



Predictive Analytics



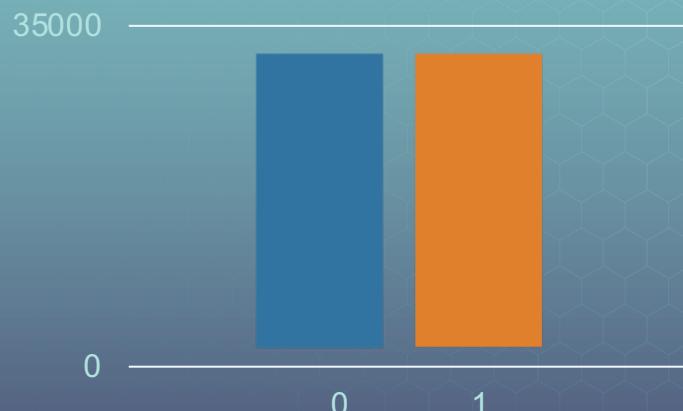
Data





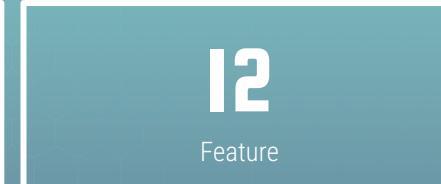
Data

Balanced



70,000

Records



12

Feature



3

Types of input features



1

Target

Data Description

Feature	Feature type	Column name	Feature value type
Age	Objective	age	Int(days)
Height	Objective	height	Int(cm)
Weight	Objective	weight	Float(kg)
Gender	Objective	gender	Object(categorical code)
Systolic blood pressure	Examination	ap_hi	int
Diastolic blood pressure	Examination	ap_lo	int
Cholesterol	Examination	cholesterol	Int (ratio: 1: normal, 2: above normal, 3: well above normal)
Glucose	Examination	gluc	Int (ratio: 1: normal, 2: above normal, 3: well above normal)
Smoking	Subjective	smoke	Int(binary)
Alcohol intake	Subjective	alc	Int(binary)
Physical activity	Subjective	active	Int(binary)
Presence or absence of cardiovascular disease	Target	cardio	Int(binary)

Goal

Predict whether the patient has potential
of cardiovascular disease



Feature Engineering



Feature Engineering



Binary dummies



Converting Age from days to years

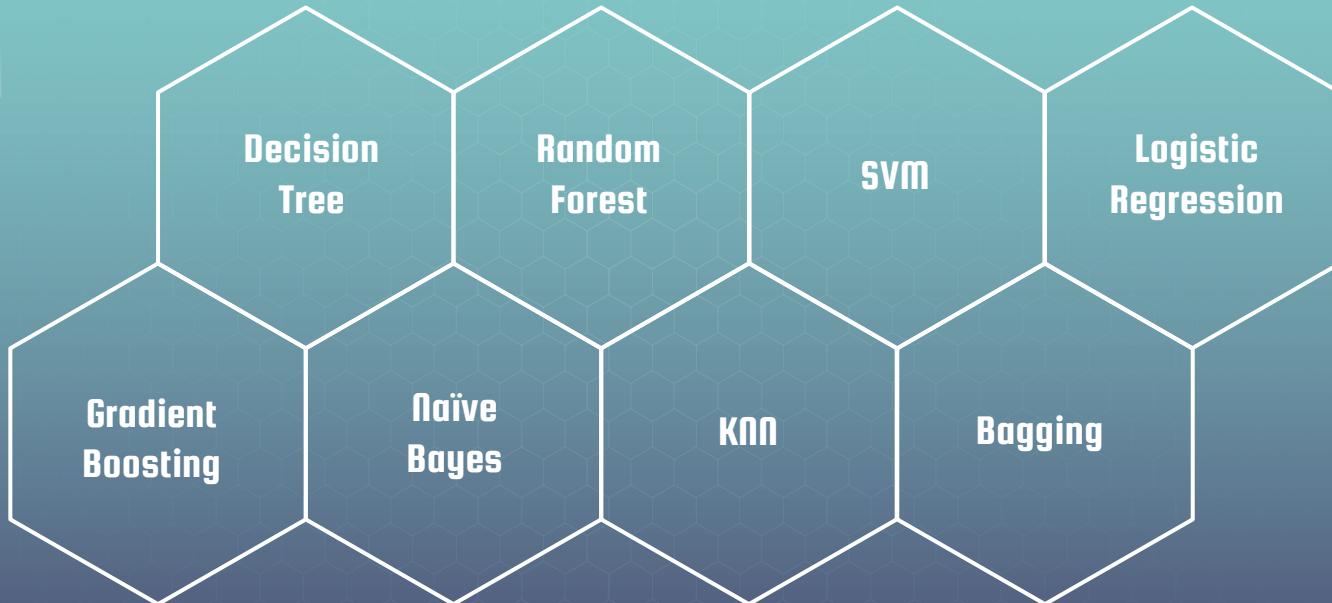
Remove outliers

Models and Results





Classification Algorithms

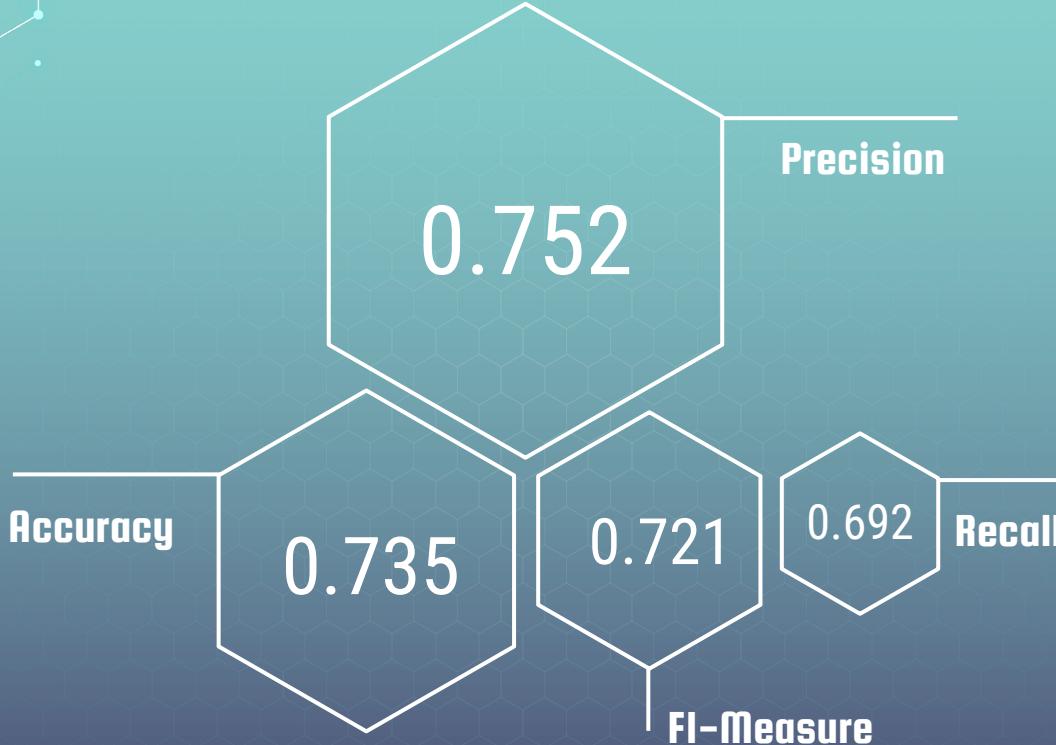


Models Results

	Model	Score	val accuracy
9	GradientBoostingClassifier	73.457143	73.660714
1	RandomForestClassifier	73.250000	73.266071
0	DecisionTreeClassifier	72.978571	73.255357
5	LinearSVC	72.885714	73.389286
8	ExtraTreeClassifier	72.814286	72.605357
7	SupportVectorClassifier	72.664286	72.896429
2	LogisticRegression	72.450000	72.903571
4	XGBoost	72.378571	72.817857
10	AdaBoostClassifier	72.300000	72.432143
12	SGDClassifier	72.171429	72.614286
13	RidgeClassifier	72.092857	72.648214
6	K-NeighborsClassifier	70.807143	71.558929
14	BaggingClassifier	68.157143	68.407143
3	NaiveBayes	67.571429	68.039286
11	perceptron	65.821429	60.110714



Best model





A large, semi-transparent hexagonal grid pattern in light blue and white is overlaid on a dark blue gradient background. In the center, the word "THANKS" is written in a bold, white, sans-serif font.

THANKS