

**P176B101 Introduction to Artificial Intelligence**

**Heart disease prediction**

**Project report**

**2019**

**Project team:**

Andrius Inčiura, IFK-6

Gediminas Jakovlevas, IFF-6/1

Evaldas Kuslevic, IFF-6/1

**Lecturer:** Germanas Budnikas

Content

[1. Project team roles 3](#_Toc9724658)

[2. System program development/application to collect data 3](#_Toc9724659)

[3. Data collection (from text file) 4](#_Toc9724660)

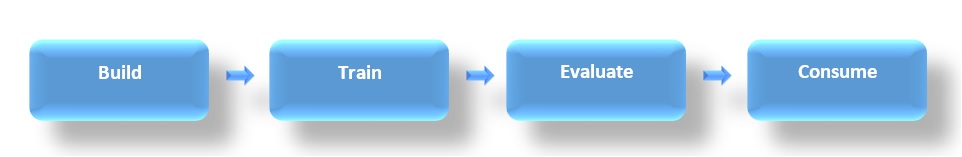
# Project team roles

|  |  |  |
| --- | --- | --- |
|  | System program development/application to collect data | Data collection (from text file) |
| Andrius Inčiura  Monday 11:00 |  |  |
| Gediminas Jakovlevas  Tuesday 09:00 |  |  |
| Evaldas Kuslevic  Tuesday 11:00 |  |  |

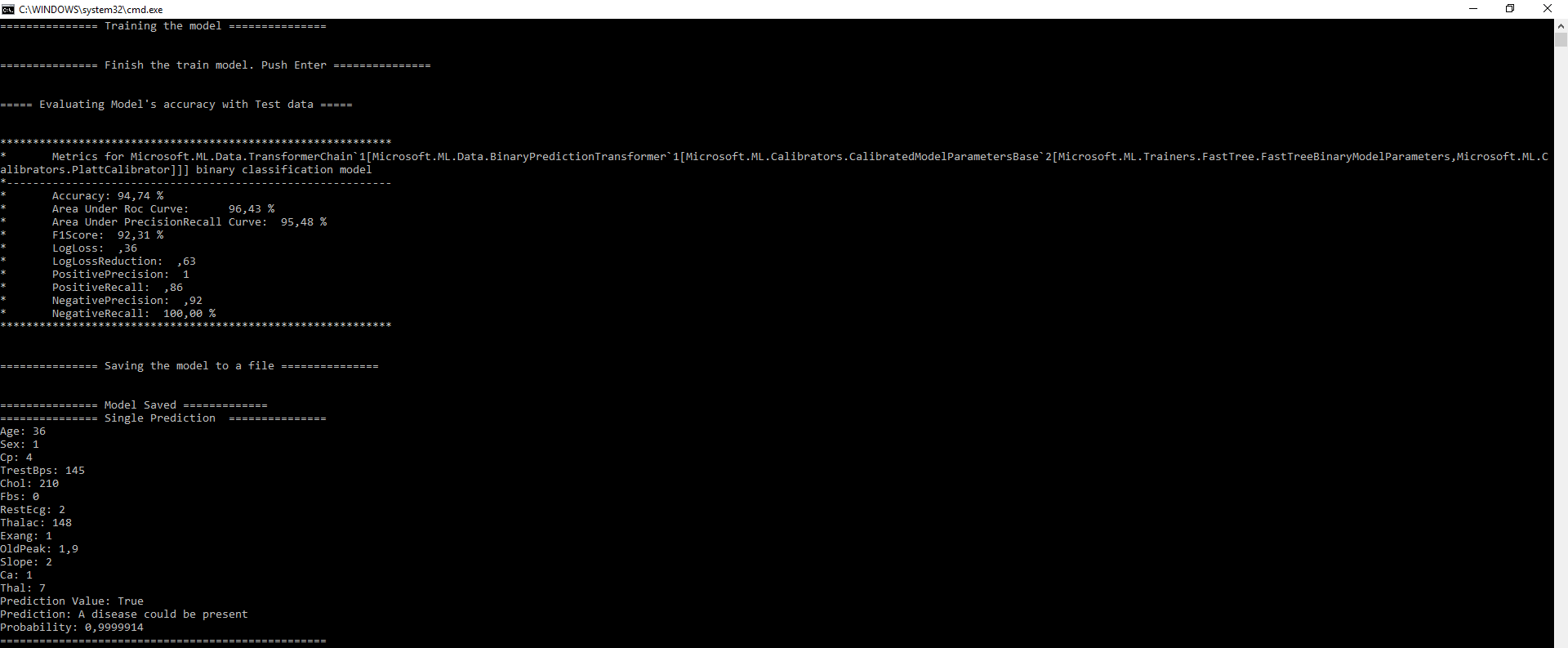
# System program development/application to collect data

In this sample, you'll see how to use [ML.NET](https://www.microsoft.com/net/learn/apps/machine-learning-and-ai/ml-dotnet) to predict type of heart disease. In the world of machine learning, this type of prediction is known as **binary classification**.

To solve this problem, first we will build an ML model. Then we will train the model on existing data, evaluate how good it is, and lastly we'll consume the model to predict a sentiment for new reviews.



GUI is a simple console application:



Data collection code:



# Data collection (from text file)

The dataset used is this: [UCI Heart disease] (<https://archive.ics.uci.edu/ml/datasets/heart+Disease>) This database contains 76 attributes, but all published experiments refer to using a subset of 14 of them.

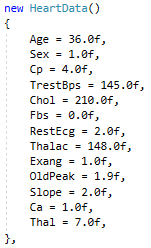
Citation for this dataset is available at [DataSets-Citation](https://github.com/dotnet/machinelearning-samples/blob/master/samples/csharp/getting-started/BinaryClassification_HeartDiseaseDetection/HeartDiseaseDetection/Data/DATASETS-CITATION.txt)

This problem is centered around predicting the presence of hearth disease based on 14 attributes. To solve this problem, we will build an ML model that takes as inputs 4 parameters: Attribute Information:

* (age) - Age
* (sex) - (1 = male; 0 = female)
* (cp) chest pain type -- Value 1: typical angina -- Value 2: atypical angina -- Value 3: non-anginal pain -- Value 4: asymptomatic
* (trestbps) - resting blood pressure (in mm Hg on admission to the hospital)
* (chol) - serum cholestoral in mg/dl
* (fbs) - (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
* (restecg) - esting electrocardiographic results -- Value 0: normal -- Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV) -- Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
* (thalach) - maximum heart rate achieved
* (exang) - exercise induced angina (1 = yes; 0 = no)
* (oldpeak) - ST depression induced by exercise relative to rest
* (slope) - the slope of the peak exercise ST segment -- Value 1: upsloping -- Value 2: flat -- Value 3: downsloping
* (ca) - number of major vessels (0-3) colored by flourosopy
* (thal) - 3 = normal; 6 = fixed defect; 7 = reversable defect
* (num) (the predicted attribute) diagnosis of heart disease (angiographic disease status) -- Value 0: < 50% diameter narrowing -- Value 1: > 50% diameter narrowing (in any major vessel: attributes 59 through 68 are vessels)

and predicts the presence of heart disease in the patient with integer values from 0 to 4: Experiments with the Cleveland database (dataset used for this example) have concentrated on simply attempting to distinguish presence (value 1) from absence (value 0).

Data extry example:



Text