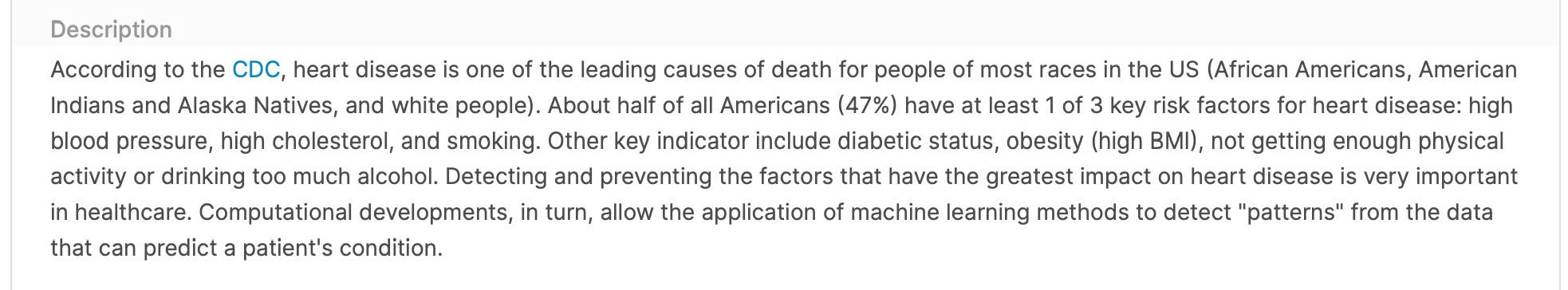
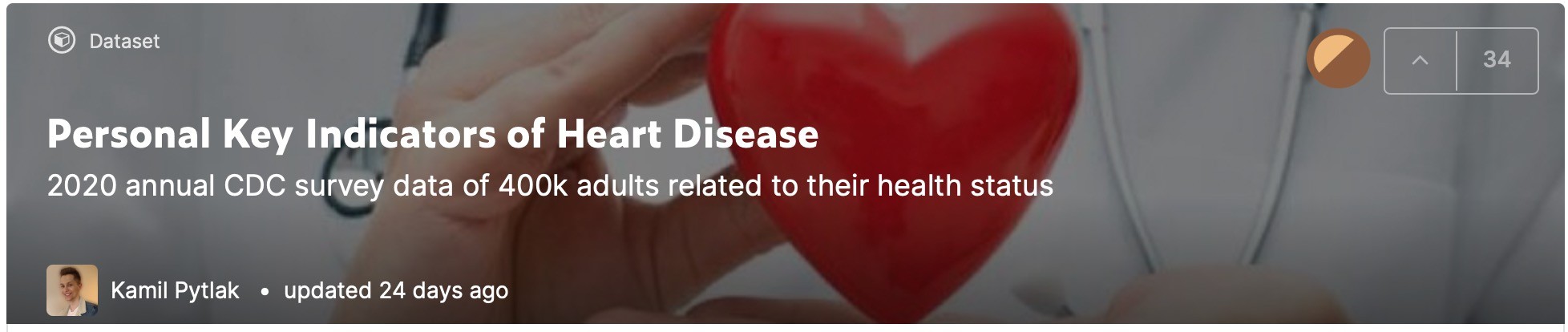
Data Science – Supeívised Leaíning

# END-TO-END MACHINE LEARNING

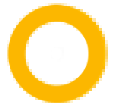
*Evaluation*

Data Science – Supeívised Leaíning



# Predicting the risk of heart disease – Introduction

*https://*[*www.kaggle.com/kamilpytlak/personal-key-indicators-of-heart-disease*](http://www.kaggle.com/kamilpytlak/personal-key-indicators-of-heart-disease)



Data Science – Supeívised Leaíning

**Predicting the risk of heart disease – Assignment**

## THE DATASET

The dataset necessary for the Project is provide in this file.

The dataset has been obtained by sampling and adapting the original dataset from Kaggle. Each row in the dataset corresponds to a different individual. Some columns are the answers to questions asked to respondents about their health status, such as "Do you have serious difficulty walking or climbing stairs?" or "Have you smoked at least 100 cigarettes in your entire life? The target is the HeartDisease variable, which indicates whether respondents that have ever reported having coronary heart disease (CHD) or myocardial infarction.

## THE ASSINGMENT

The implementation of a ML model to **predict a person’s probability of having heart disease.** Your work must include all stages of the ML Checklist (data understanding, data preparation, modelling and evaluation) in a notebook format. Your code is expected to be **properly organized and commented** to facilitate understanding, and it needs to run **without errors**. You’re encouraged to explore different algorithms can be optionally explored as well.