Cardio Case Disease Case

Because some things can't stop!

TABLE OF CONTENTS

01

About the Company!

You can enter here the subtitle if you need it

CARDIO CATCH DISEASE

Its a company founded in 1988, with the purpose of supply cardiovascular diseases diagnostics to its clients, nowadays the company is present in all North American territory and expanding to Brazil.

02

Business Context

You can enter here the subtitle if you need it

Business Context

Because some things can't stop!

Our challenge is to help the **Cadio Catch Diseases** a company that specializes in detecting heart diseases in the early stages.

Its business model is 'service' it means that company offers an early heart diseases diagnostic for a price.

Importants Details:

- Actually the precision of the diagnostic is between 55% and 65%.
- For each 5% of accuracy over 50% the price is R$ 500,00

Main Goal:

Create a tool that increases diagnostic accuracy and that this accuracy is stable for all diagnostics.

Secondary Goals:

Answer the CEO questions,

1. What is the Accuracy and Precision of the tool?
2. How much profit will Cardio Catch Diseases have with the new tool?
3. How Reliable is the result given by the new tool?

03

ML Canvas

Decisions

How are predictions used to make decisions that provide the proposed value to the end-user?

The predictions are going to be used to diagnose if a patient has cardiovascular disease or not.

ML task

Input, output to predict, type of problem.

Input:

Tabular data with characteristics analyzed by the doctors during the diagnose process

Output:

Binary class 0 with the patient has no cardio vascular disease 1 with it does.

Type of problem:

Binary Classification

Value Propositions

What are we trying to do for the end-user(s) of the predictive system? What objectives are we serving?

Create a diagnose predictor this way, the diagnose process can be automated and working-hours of the doctors will be saved.

Objectives:

The objective is to beat the accuracy of the doctors which is 65%. In this way our predictor must achieve:

Accuracy higher than 65%, higher recall possible if at least 65% of precision.

Data Sources

Which raw data sources can we use (internal and external)?

The company supply us the data needed.

Collecting Data

How do we get new data to learn from (inputs and outputs)?

Through the manual process of collecting the data from patients.

Making Predictions

When do we make predictions on new inputs? How long do we have to featurize a new input and make a prediction?

New predictions when the doctors serve the new data to the system and it has to be in real time.

Offline Evaluation

Methods and metrics to evaluate the system before deployment.

Recall,

Accuracy and

Precision

Features

Input representations extracted from raw data sources.

Id, Age, Gender, Height, Weight, Ap_hi, Ap_lo, cholesterol, gluc, smoke, alco, active, cardio

Building Models

When do we create/update models with new training data? How long do we have to featurize training inputs and create a model?

Lgbm_pipe_retrained.pkl

Lgbm_pipe_tunnig_retrained.pkl

Lgbm_pipe_tunning.pkl

lgbm_pipe.pkl

Live Evaluation and Monitoring

We will monitoring the system through recall and accuracy taken any measure necessary.

04

Solutions and Results!

Solution and Results!

Solution!

To our solution we used a LGBM Model with 100 decision trees.

Why we chose to focus on recall?

Because of the nature of the problem, heart diseases, a false diagnosis of an actual sick patient could lead to aggravation of the problem or even death. Bussines wise, it could hurt us as a future legal process. Focusing on recall we make less money but we get away less risk of the legal process.

Actual Performance!

The Actual Accuracy varies between 55% and 65%.

Solution Performance!

Our solution achieved:
Precision of 71%,
Accuracy of 73%
Recall of 78%

Increase of 11%.

*** considering actual as 60%**

05

Impacts in Business.

OK, nice fancy words but how much money it represents?

Financial Impacts!

Scenario Actual!

With 60% accuracy.
The price is $ 1000,00

Profit:
$ 70.000.000,00

*Price times 70000 observations of training dataset

Focusing on Recall!

With 73% accuracy our price would be $ 2000,00

Profit:
$ 140.000.000,00

Focusing on Precision!

With 75% of accuracy our price would be $ 2500,00

Profit:
$ 175.000.000,00

WoW, thats real money. Why to focus on recall agian?

Because of the nature of the problem, heart diseases, a false diagnosis of an actual sick patient could lead to aggravation of the problem or even death. Business wise, it could hurt us as a future legal process. Focusing on recall we make less money but we get away less risk of the legal process.

THANKS

Email: brunovinicius154@gmail.com
Github: https://github.com/bruno154