Today, we will be presenting to you a very important topic: predicting heart diseases with the help of artificial intelligence. This is a topic that affects many people worldwide, and it can truly make a difference in terms of prevention.

Why predict heart diseases?
Heart diseases, such as heart attacks or strokes, are currently the leading cause of death worldwide.
Each year, around 17.9 million people lose their lives to these diseases.

I think we can all agree that is a massive number. And the sad reality is that many of these people didn't even know they were at risk.

So why is it so important to predict these diseases?
Well, one of the keys to saving lives is being able to detect them as early as possible. If these diseases are caught early, it's often possible to prevent complications by adopting a healthier lifestyle or following the appropriate treatment.

This is where artificial intelligence comes into play.
Today, thanks to technological advancements, AI can help doctors spot signs of heart disease in patients even before they show any severe symptoms.
By analysing large amounts of medical data, AI helps detect patterns that are invisible to the human eye and predict risks more accurately and quickly.

To illustrate this, we'll be working with a medical dataset that contains information about 918 patients.
This dataset includes 12 different medical variables, such as age, gender, cholesterol levels, and even the type of chest pain.
Our goal is to use this data to predict whether a patient has heart disease or not, based on the available information.

Throughout this presentation we will be showing a Quick Demonstration of the Dataset and what the dataset looks like.
It will be based upon the information that we will be covering throughout our presentation.

Why is this interesting?
This project shows how AI can transform medicine.
AI doesn't replace doctors, but it helps them make more informed decisions. By predicting heart disease with real data, we hope to save lives.

So, now you have an idea of why and how AI can help predict heart disease.
In the next part of the presentation, we will dive into the data analysis, see how we can leverage this information, and finally create a model that can predict whether a patient ,has heart disease or not.

That's what we're going to try to understand today, with the help of AI.
I'll now hand it over to my colleagues for the data analysis!

Data Analysis - Notebook

Introduction

We will now analyse our dataset to understand what the main trends are. This step is essential, as it allows us to identify the risk factors for heart disease before even training a prediction model. We will use statistics and graphs to visualise the data.

In this section, we will prepare our environment and data analysis. First, we install Seaborn, a library that helps us create graphs. Next, we upload our data file to Google Colab, load it into a DataFrame with Pandas, and look at the first few lines to check that everything is correct.

As you can see, the first 5 lines are displayed correctly.

Partie 1

Here is a table with the basic statistics to help you understand our dataset.

Coding

This table shows the descriptive statistics for the numerical variables in the dataset. In particular, it shows the number of samples (count), the mean (mean), the standard deviation (std), the minimum and maximum values, and the quartiles (, ,). This allows you to quickly grasp the distribution of each variable (example, age, resting blood pressure, cholesterol, etc.) and identify any extreme values.

Here is a brief overview of the variables and their statistics (taken from df.describe().T):

- Age: varies from 28 to 77, with an average of around 53.
- RestingBP (resting blood pressure): average 132 mmHg (a voir), but there was a minimum at 0 (probably an outlier) and a maximum at 200.
- Cholesterol: average close to 199 mg/dl(à voir), ranging up to 603 mg/dl, indicating a wide range.
- FastingBS (fasting blood glucose): the average being ~0.23, this suggests that around 23% of individuals have high blood glucose levels (coded as 1).
- MaxHR (maximum heart rate): average around 137 bpm, with a minimum at 60 and a maximum at 202.
- Oldpeak: measure of ST depression; ranges from 0 to 6.2, with an average of less than 1.
- HeartDisease: binary value (0 or 1), with an average around 0.55, means that around 55% of individuals in the dataset suffer from heart disease.

These statistics make it possible to quickly identify the distribution (minimum, maximum, mean, standard deviation) of each variable and to identify possible unusual values (such as a RestingBP: resting heart rate at 0).

Partie 2

Now let's look at the correlation matrix, which shows the linear links between the different numerical variables in our dataset.

Coding

This correlation diagram shows, for each pair of numerical variables, a correlation coefficient (between -1 and +1) which indicates the extent to which these variables vary together:

- Positive value (close to +1): the more one increases, the more the other tends to increase.
- Negative value (close to -1): as one increases, the other tends to decrease.
- Value close to 0: no marked linear relationship between the two variables.

The colours (ranging from blue/violet for negative correlations to red/orange for positive correlations) help to quickly identify the most highly correlated pairs of variables. In the context of *Heart Failure Prediction*, for example, we can see that :

- FastingBS (fasting blood glucose), Oldpeak (ST segment depression) and MaxHR (maximum heart rate) have stronger correlations with HeartDisease than other variables (even if these correlations remain moderate).
- Age also shows a slight positive correlation with HeartDisease.
- Cholesterol and RestingBP appear to have little correlation with disease, at least according to this linear measure.

This heatmap therefore makes it possible identify which variables might be more influential in detecting or predicting heart disease, even if correlation alone does not prove a causal link.

Partie 3

To visualise the distribution of heart disease according to gender, we're going to create a bar chart comparing the variable HeartDisease (0 or 1) with the variable Sex (M or F).

Coding

In our study, there were more men than women (M=705; F=193). The difference between men and women who have had heart attacks is very large, whereas the difference between men and women who have not had a heart attack is not very large. The number of men who have heart attacks is twice as high as those who have never had a heart attack. For women, however, the difference is not very great.

Partie 4

To better understand the distribution of the different types of chest pain (ChestPainType) according to sex, we will draw a bar chart showing the Sex variable within each ChestPainType category.

Coding

Analysis of the types of chest pain reveals that the presence of asymptomatic chest pain (ASY) is strongly associated heart failure. Even the absence of obvious symptoms, this type of pain could indicate an alteration in the heart's rhythm.
underlying cardiac function, making this criterion an indicator to be monitored closely.

Partie 5

To illustrate the relationship between exercise angina (ExerciseAngina) and the presence of heart disease (HeartDisease), we will draw a clustered bar chart, highlighting the role of fasting blood glucose (FastingBS) in interpreting the results.

Coding

This bar chart compares the presence or absence of heart disease (HeartDisease) with the presence of exercise-induced angina (ExerciseAngina). The red bars indicate individuals without heart disease, and the yellow bars those with heart disease. Here we can see that the majority of people without exercise-induced angina (N) do not have heart disease, while those with exercise-induced angina (Y) have more cases of heart disease. The numbers above each bar correspond to the number individuals in each category. The title refers to FastingBS, suggesting that the analysis is considered in relation to fasting glycaemia, even though this variable is not directly represented on the graph axis.

Conclusion:

This exploratory analysis highlights the importance of variables such as age, fasting blood glucose (FastingBS), exercise angina (ExerciseAngina) and maximum heart rate (MaxHR), while revealing notable differences between men and women. The attention paid to asymptomatic chest pain (ASY) and extreme values will guide the next stages of data preparation and modelling.

Modelisation and Prediction

So, we have explored our data and identified interesting trends, such as the impact of age, gender, or chest pain on heart disease. Now, we will move on to modeling, meaning we will train an artificial intelligence model capable of automatically predicting whether a patient is sick or not."
Since we cannot do this ourselves, we took inspiration from Notebook, and to write the code lines and solve various problems, we used ChatGPT.
We will see how to prepare our data and how to use a machine learning model to make these predictions.

First, we gather and preprocess the data to ensure it is suitable for training. Then, we split it into two sets: one for training and one for testing, allowing us to evaluate the model's performance. Next, we create the machine learning model and train it so that it can learn patterns from the data. Once trained, we test the model on new data and measure its accuracy using performance metrics. Finally, we analyze the results, draw conclusions, and make adjustments if necessary to improve the model's performance.

1. Data Preparation
Before training a model, we must ensure that our data is properly prepared.
Indeed, artificial intelligence algorithms do not accept text or categories like 'Male' or 'Female'. They require numbers. So, we need to convert certain columns into numerical values.
Live Demonstration: Encoding Categories
For example, we will transform the 'Sex' column, which contains 'Male' and 'Female', into 0 and 1.

Coding

Now, instead of having 'Male' or 'Female', we have 0 or 1. Similarly, chest pain types have been transformed into numbers so that the model can understand them.

2. Splitting the Data to Train the Model
Before training our model, we need to divide our data into two parts:

1. 80% of the data for training the algorithm
2. 20% to test and verify its accuracy

It's like studying for an exam: we practice with exercises (80%) and then test our knowledge with a real test (20%).

Live Demonstration: Splitting the Data

Coding

Interpreting the Results:
Now, our model will learn with 80% of the data, and then we will check if it makes accurate predictions on the remaining 20%.

3. Model Creation and Training
For this next step, we will use a model called Random Forest. It is like a forest of decision trees, where multiple trees work together to make the best decision.

To give you a bit more information about the model: it is a very powerful model often used in medicine because it is reliable and interpretable.
Random Forest is an artificial intelligence model used for making predictions.
It works like a forest composed of trees, where each tree makes a decision:
The model combines multiple trees to provide a more reliable final prediction.

Why use it?

- Accurate and powerful – Reduces errors by combining multiple decisions.
- Robust – Less sensitive to variations in data.
- Used in medicine – Can help identify high-risk patients.

In summary:
It's as if each tree in the forest votes, and the majority decides the final result.

Live Demonstration: Training the Model

Coding

Through these lines of code, our model has just learned to make predictions based on the data it was given.
It's as if we showed it thousands of medical cases, and it learned which patients are at risk and which are not thanks to the Random Forest model.

4. Testing the Model and Measuring Its Accuracy
Now that the model is trained, we will ask it to make predictions on **20% of test data** that it has never seen before.

We will then compare its predictions with the real results to check if it is accurate.

To summarize:
We have 80% of the data that has been analyzed and used for training. To ensure the model is correct, we will test it on the remaining 20% without giving it the "answer." Finally, we will check if its responses are correct. It's a bit like the model taking its final exams!

L**ive Demonstration: Testing Accuracy**
Type this code in IDLE:

Coding

Interpreting the results:
Here, our model displays an accuracy of ≈ 89.67%. This means that out of 100 patients, about 90 are correctly predicted.
This is a good accuracy rate, but it can still make mistakes in some cases.

5. Conclusion and Discussion
We now have a model capable of predicting whether a patient is at risk of having heart disease.

We achieved this by:

- Preparing our data.
- Training an artificial intelligence model.
- Testing its accuracy (~90%).

Conclusion

And now to wrap up our presentation.
Today, we explored how artificial intelligence (AI) can be used to predict heart disease. We first analyzed a dataset containing information on 918 patients, identifying key factors such as age, sex, and chest pain type as important indicators for predicting heart disease.

Our machine learning model (Random Forest) that predicts whether a patient has heart disease or not, showed high accuracy, demonstrating that AI can indeed assist doctors in detecting risks more quickly and accurately.

Although the results are promising, there are several areas for improvement:

- Enriching the dataset to improve the model's performance.
- Exploring other AI models, such as neural networks, for even more precise predictions.
- Integrating this model into real-world healthcare systems for real-time detection.

It's important to remember that AI doesn't replace doctors but helps them make more informed decisions. AI in medicine holds immense potential for early detection and prevention of diseases, ultimately helping to save lives.

AI is transforming the healthcare sector, but for it to be truly effective, it must be used in collaboration with healthcare professionals. The future of healthcare will undoubtedly be shaped by AI, offering more accurate and accessible solutions.