Data Science Certification Internship Project

Taste of real-life business problem.

Agenda

- Objective
- Data Information
- General Process
- Presentation Format
- Marking

Objective

- This project will allow you to integrate and apply the knowledge, skills and concepts that you have acquired throughout the academic journey.
- It will allow you to demonstrate your understanding of key concepts, methodologies and best practices.
- It can also help you in professional development by fostering skills such as time management, teamwork and problem solving.

Data Information

- The dataset provided for this project is a subset of data our team worked on for a project of **Parkway Pantai** to understand the influence of various factors on the heart disease.
- The data provided to you contains all the required detail to understand the impact on heart such as cholesterol level, blood pressure, alcohol consumption etc.
- All the information about the features is present in **Data_Dictionary.pdf**

General Process

- Problem Statement
- Research Objective
- Data Understanding and Preparation
- Result Interpretation
- Conclusion

General Process

Problem Statement:

Heart disease is a leading cause of mortality worldwide. Early detection and accurate prediction of heart disease can significantly improve patient outcomes by enabling timely intervention and preventive measures. The objective of this project is to develop a machine learning model that can effectively predict the presence or absence of heart disease based on various medical and lifestyle factors.

General Process

Research Objective:

The goal of this project is to build a predictive model that can accurately classify individuals as either having or not having heart disease. By analysing a comprehensive set of patient attributes, including demographic, clinical, and lifestyle factors, the model aims to identify patterns and features indicative of the presence or absence of heart disease.

General Process

Data Understanding and Preparation:

- First step towards the analysis is to understand and prepare the data you have, once you get the idea about the features given in the dataset and the relationship among the variables then you can start preparing the dataset for the data analysis.
- Data preparation is important because you need to transform the data as per the requirements like imputing missing values, changing the data type, deriving a custom column etc.

General Process

Result Interpretation:

- Once the model is ready, you must test it and check its accuracy.
- In the model creation process try different models and techniques to improve the accuracy.
- Interpret the result of the models.

General Process

Conclusion:

- Overall accuracy of the model.
- The features in the dataset which are important in predicting the target.
- Any recommendations.

Presentation Format

- Create the model using Jupyter Notebook (IDE).
- It would be better if you include the slides with the full process of your analysis so it looks more professional and would be clear to the audience.
- Along with the PowerPoint Presentation file (.ppt), you must include the Jupyter Notebook file (.ipynb) for the submission.

Let’s do the magic with data?