Heart disease prediction

Réalisé Par :

- BENSAID Ahmed Amine
- BENGRAB Ilyas
- BENDAHAN Driss
- BELATIK Yousra
- BENLAMHAIRA Oumaima

Encadré Par :

- Dr. Bentaleb Asmae

Année Universitaire 2022 – 2023

Content table

Liste of figures

INTRODUCTION :

Heart disease is one of the leading causes of death worldwide. Early detection and prevention of heart disease are critical in reducing mortality rates. With the advances in cloud computing and artificial intelligence (AI), it is now possible to predict and diagnose heart disease with high accuracy.

Currently, there are various predictive models and tools available for healthcare professionals to use for the early detection and prevention of heart disease. More advanced techniques, such as machine learning algorithms, have been applied to heart disease prediction and have shown promising results. These algorithms can analyze large amounts of patient data, including medical history, lifestyle factors, and genetic information, to develop highly accurate predictions of heart disease risk. Overall, the current state of heart disease prediction is very promising, with various tools and models available to healthcare professionals to help identify patients at risk and intervene early to prevent heart disease. However, ongoing research and innovation in this field will continue to improve our ability to predict and prevent heart disease.

This report will explore the use of AI in predicting heart disease. We will examine the benefits of using cloud computing and AI in healthcare, and the challenges of implementing these technologies in the healthcare industry. But before that we will give in the first chapters an overview on cloud computing as a devoloping technologie and AI and ofcourse machine learning algorithmes.

Overall, this report aims to provide a comprehensive overview of the potential of cloud computing and AI in heart disease prediction and how it can be applied in the healthcare sector to improve patient care and outcomes.

Chapter 1 : Cloud computing

Introduction

The cloud has revolutionized the way we use technology in today's world. It has become a cornerstone of modern computing, and its impact is felt across industries and disciplines. In this chapter, we will explore what the cloud is, types of cloud Services,deployement models and of characteristics of cloud computing.

What is cloud cloud computing

The cloud refers to a network of servers that are used to deliver computing services over the internet. These services include storage, processing power, and applications that can be accessed by users anywhere in the world. Cloud computing has become popular due to its scalability, reliability, and cost-effectiveness. Rather than owning and maintaining their own hardware and infrastructure, businesses and individuals can rent computing resources from cloud providers and pay only for what they use.

Cloud computing has its roots in the early days of the internet, but the concept of providing computing resources over the internet was not fully realized until the early 2000s.

The term "cloud computing" was coined by Eric Schmidt, the CEO of Google, during a conference in 2006. However, the underlying technologies and ideas that power cloud computing had been developed earlier by companies like Amazon and Salesforce. Amazon Web Services (AWS) launched in 2002, offering cloud-based storage and computing services to developers. AWS started with a few basic services, such as Simple Storage Service (S3) and Elastic Compute Cloud (EC2), and has since grown to become the dominant cloud computing provider in the world.

Salesforce, a cloud-based customer relationship management (CRM) software company, launched in 1999, was one of the first companies to offer enterprise-level cloud services. Salesforce's success proved that cloud computing was a viable business model and paved the way for other companies to follow suit.

Types of cloud services:

The basic cloud computing models include the following:

Infrastructure as a Service (IaaS)

In (Iaas) the cloud service provider provides a set of virtualized computing resources like CPU, Memory, OS, and Application Software etc in the cloud. IaaS uses virtualization technology [6] to convert physical resources into logical resources that can be dynamically provisioned and released by customers as needed. Some of the major companies offering infrastructure as a service include Rackspace Cloud Servers, Google, Amazon EC2, IBM, and Verizon.

Benefits of IaaS Solutions

• Reduces cost of capital expenditures.

• Users pay for the service they want.

• Access to enterprise-grade IT resources and infrastructure.

• Users can scale up and scale down the resources based on their requirements at any time.

Platform as a Service (PaaS)

This is more advanced type of cloud computing service. In PaaS, a cloud service provider offers, runs and maintains both system software (i.e., the operating system) and other computing resources. PaaS services include design, development and hosting of applications. Other services include collaboration, DB integration, security, web service integration, scaling etc. Users don’t need to worry about having their own hardware and software resources or hire experts for management of these resources. This scheme provides flexibility in installing software on system; scalability is another advantage of the PaaS. A downfall of the PaaS is the lack of interoperability and portability among providers. Consumers purchase access to the platforms, enabling them to deploy their own software and applications in the cloud.

Benefits of PaaS Solutions

- Community: Most of the time, many people are involved in building cloud applications in PaaS environments. This creates a strong supportive community that can help your development team along the way.
- No more upgrades: Companies are not required to update or upgrade the infrastructure software. Instead, the PaaS provider handles all upgrades, patches and routine software maintenance.

- Lower cost: Companies face lower risk since they do not have to make upfront investment in hardware and software.
- Simplified deployment : The development team can concentrate on developing the cloud application without having to worry about the testing and deployment infrastructure.

Software as a Service (SaaS)

In this model, cloud service providers are responsible for running and maintaining application software, operating system and other resources. SaaS model appears to the customer as a web-based application interface where internet is used to deliver services that are accessed using a web-browser. The hosted applications like Gmail and Google Docs can be accesses through different devices like smart phones and laptops etc. Unlike traditional software, SaaS has the advantage that the customer does not need to buy licences, install, upgrade, maintain or run software on his own computer. It has also other advantages such as multitenant efficiency, configurability and scalability.

Benefits of SaaS Solutions :

- Rapid Scalability
- Accessibility from any location with Internet
- Eliminates infrastructure concerns
- Custom levels of service offerings
- Bundled maintenance and Support

Recovery as a Service (RaaS)

Recovery as a Service (RaaS) solutions helps companies to replace their backup, archiving, disaster recovery and business continuity solutions in a single, integrated platform. RaaS providers help companies recover entire data centers, servers (OS, applications, configuration and data), and database files. RaaS helps business establishments in reducing the impact of downtime in case of disasters or like situations. RaaS is also called as DRaaS (Disaster Recovery as a Service).

Benefits of RaaS Solutions :

- Prevent temporary or permanent loss of critical company data.
- Prevents permanent loss of physical infrastructure, including IT infrastructure.
- Cost-effective way of recovering data.
- Enables faster recovery while maintaining accuracy.

- Offer greater flexibility on the type of backup required (either primary or secondary backup)

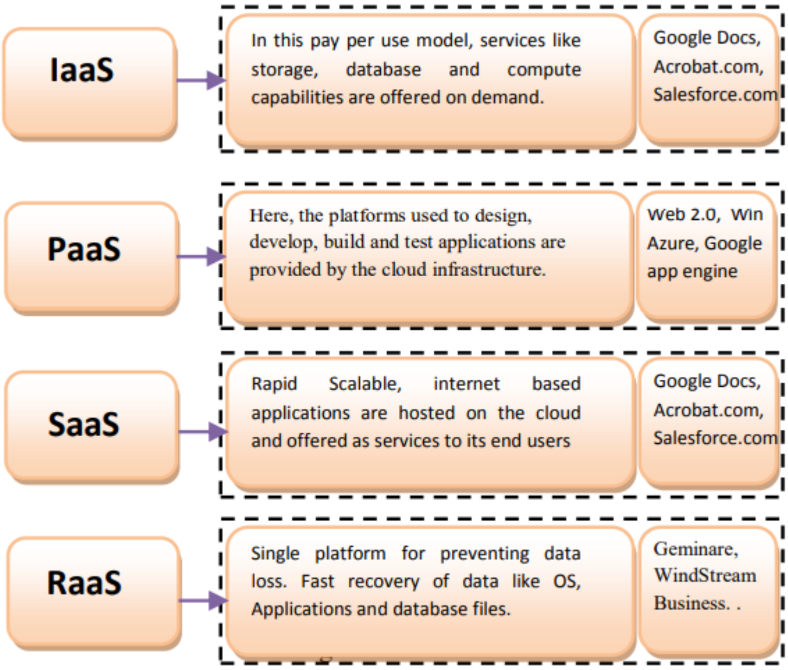

Figure 1:cloud services

Deployment models :

Cloud deployment models indicate how the cloud services are made available to users. The four deployment models associated with cloud computing are as follows:

- **Public cloud** As the name suggests, this type of cloud deployment model supports all users who want to make use of a computing resource, such as hardware (OS, CPU, memory, storage) or software (application server, database) on a subscription basis. Most common uses of public clouds are for application development and testing, non-mission-critical tasks such as file-sharing, and e-mail service.
- **Private cloud** True to its name, a private cloud is typically infrastructure used by a single organization. Such infrastructure may be managed by the organization itself to support various user groups, or it could be managed by a service provider that takes care of it either on-site or off-site. Private clouds are more expensive than public clouds due to the capital expenditure involved in acquiring and maintaining them. However, private clouds are better able to address the security and privacy concerns of organizations today.

- **Hybrid cloud** In a hybrid cloud, an organization makes use of interconnected private and public cloud infrastructure. Many organizations make use of this model when they need to scale up their IT infrastructure rapidly, such as when leveraging public clouds to supplement the capacity available within a private cloud. For example, if an online retailer needs more computing resources to run its Web applications during the holiday season it may attain those resources via public clouds.
- **Community cloud** This deployment model supports multiple organizations sharing computing resources that are part of a community; examples include universities cooperating in certain areas of research, or police departments within a county or state sharing computing resources. Access to a community cloud environment is typically restricted to the members of the community.

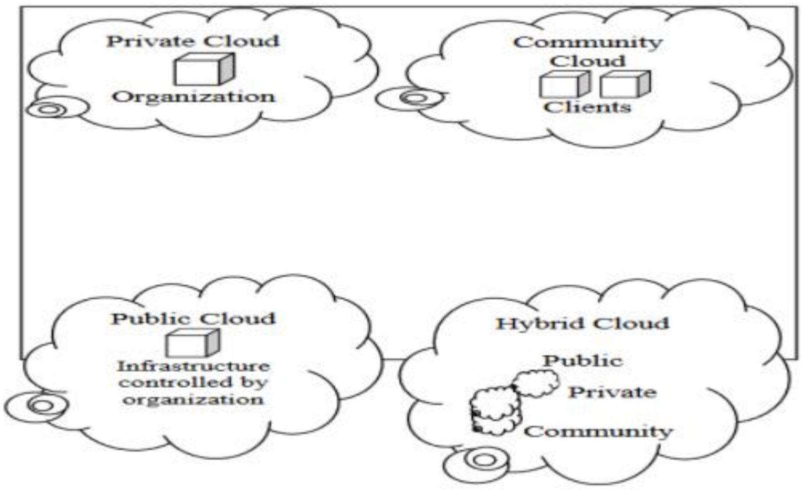

Figure 2:The cloud computing deployement models

Characteristics of cloud computing:

Cloud computing systems satisfy many interesting characteristics that make them promising for future IT applications and services.

- On-demand self-service: cloud services such as CUP time, Storage, network access, server time, web applications etc can be allocated automatically as required by the consumers without any human interaction.
- Cost effectiveness: Services provided by the cloud service providers are very cost effective if not free. The billing model is pay as per usage; there is no need to purchase the infrastructure and therefore lowers maintenance cost.

- Broad Network Access (mobility): consumers can access cloud resources over the Internet all the time and from anywhere (i.e., ubiquitous) through different types of devices (e.g., mobile phones, laptops, and PDAs).
- Scalability: the infrastructure of cloud computing is very scalable. Cloud providers can add new nodes and servers to cloud with minor modifications to cloud infrastructure and software.
- Reliability: is achieved in cloud computing by using multiple redundant sites. High reliability makes the cloud a perfect solution for disaster recovery and business critical tasks.

Conclusion:

Over the past decade, cloud computing has become increasingly popular, with many companies migrating their applications and data to the cloud. Today, cloud computing is an essential technology for businesses of all sizes and across all industries, powering everything from email and file sharing to advanced machine learning and artificial intelligence applications.

Chapter 2: Artificial intelligence and machine learning

Introduction:

In this chapter, we'll discover the importance of artificial intelligence in health care department, in addition of the definition of machine learning and its way of working. Artificial intelligence (AI) and machine learning (ML) are terms that have created a lot of buzz in the technology world, and for good reason. They're helping organizations streamline processes and uncover data to make better business decisions. They're advancing nearly every industry by helping them work smarter, and they're becoming essential technologies for businesses to maintain a competitive edge.

Artificial intelligence in health care department:

ARTIFICIAL intelligence (AI) is already delivering on making aspects of health care more efficient. Over time it will likely be essential to supporting clinical and other applications that result in more insightful and effective care and operations. AI has multiple use cases throughout health plan, pharmacy benefit manager (PBM), and health system enterprises today, and with more interoperable and secure data, it is likely to be a critical engine behind analytics, insights, and the decision-making process. Enterprises that lean into adoption are likely to gain immediate returns through cost reduction and gain competitive advantage over the longer term as they use AI to transform their products and services to better engage with consumers.

Machine learning:

Machine learning (ML) is a form of artificial intelligence (AI) that focuses on creating systems that learn, or improve their performance, based on the data they process. Artificial intelligence is a broad term that refers to systems or machines that simulate a form of human intelligence. Machine learning and AI are often discussed together and these terms are sometimes used interchangeably although they do not refer to exactly the same concept. An important distinction is that while all of machine learning relies on artificial intelligence, artificial intelligence is not limited to machine learning.

How Does Machine Learning Work?

The development of a Machine Learning model is based on four main steps. Typically, a Data Scientist manages and oversees this process.

- The first step is to select and prepare a training data set . This data will be used to feed the Machine Learning model to learn how to solve the problem for which it is designed.

  The data can be labeled, to tell the model which features it should identify. They can also be unlabeled, and the model will have to spot and extract the recurring features from itself.

  In either case, the data must be carefully prepared, organized and cleaned. Otherwise, the training of the machine learning model may be biased. The results of its future predictions will be directly impacted.

- The second step is to select an algorithm to run on the training dataset. The type of algorithm to use depends on the type and volume of training data and the type of problem to be solved.
- The third step is training the algorithm. This is an iterative process. Variables are run through the algorithm, and the results are compared with those it should have produced. The "weights" and bias can then be adjusted to increase the accuracy of the result.

  The variables are then run again until the algorithm produces the correct result most of the time. The algorithm, thus trained, is the Machine Learning model.

- ➢ The fourth and final step is using and improving the model. We use the model on new data, the origin of which depends on the problem to be solved. For example, a Machine Learning model designed to detect spam will be used on emails.

For its part, the Machine Learning model of a robot vacuum cleaner ingests data resulting from the interaction with the real world such as moving furniture or adding new objects in the room. Efficiency and accuracy can also increase over time.

What are the main Machine Learning algorithms?

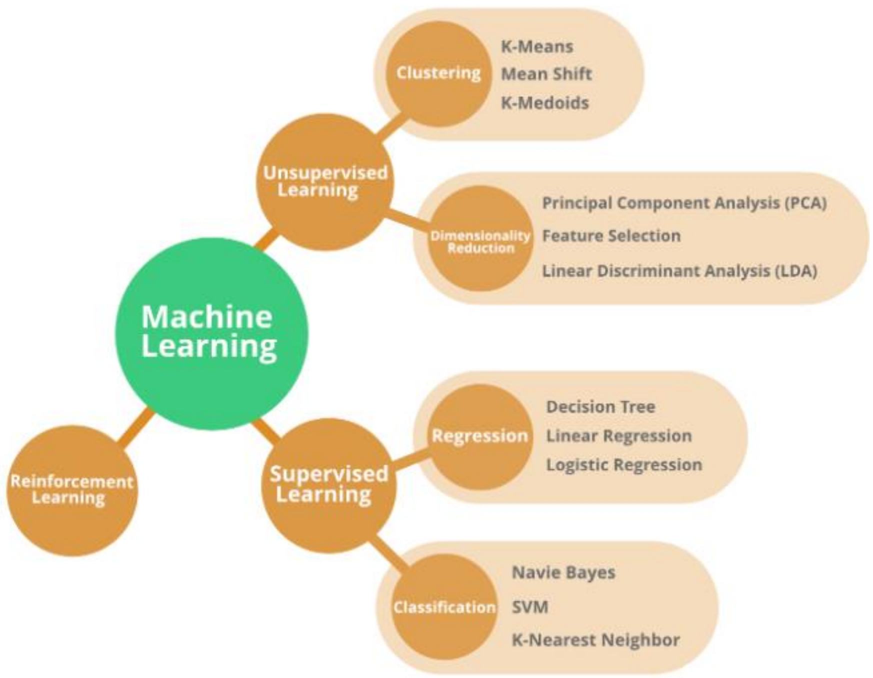

Figure 3:Machine learning algorithmes

There is a wide variety of Machine Learning algorithms. However, some are more commonly used than others. First, different algorithms are used for labeled data.

- Regression algorithms, linear or logistic, help to understand relationships between data. Linear regression is used to predict the value of a dependent variable based on the value of an independent variable. For example, it would be a question of predicting the annual sales of a salesperson according to his level of education or his experience.

- Logistic regression is used when the dependent variables are binary. Another type of regression algorithm called a support vector machine is relevant when the dependent variables are more difficult to classify.
- Another popular ML algorithm is the decision tree . This algorithm makes it possible to establish recommendations based on a set of decision rules based on classified data. For example, it is possible to recommend which football team to bet on based on data such as the age of the players or the winning percentage of the team.
- For unlabeled data, clustering algorithms are often used. This method consists of identifying groups with similar records and labeling these records according to the group to which they belong.
- Previously, the groups and their characteristics are unknown. Among the clustering algorithms, there are the K-means , the TwoStep or the Kohonen.
- Association algorithms, on the other hand, are used to discover patterns and relationships in data, and to identify “if/then” relationships called “association rules”. These rules are similar to those used in the field of Data Mining or data mining.
- Finally, neural networks are algorithms in the form of a multi-layered network. The first layer allows data ingestion, one or more hidden layers draw conclusions from the ingested data, and the last layer assigns a probability to each conclusion.
- A “deep” neural network is composed of multiple hidden layers, each allowing to refine the results of the previous one. It is used in the field of Deep Learning.

Conclusion:

With the unprecedented advancement of data aggregation and deep learning algorithms, artificial intelligence and machine learning are poised to transform the practice of medicine. The field of orthopedics, in particular, is uniquely suited to harness the power of big data, and in doing so provide critical insight into elevating the many facets of care provided by orthopedic surgeons. The purpose of this review is to critically evaluate the recent and novel literature regarding ML in the field of orthopedics and to address its potential impact on the future of musculoskeletal care.

Chapitre3: Project realization

Introduction:

The use of cloud computing and AI in healthcare has the potential to revolutionize the industry by transforming the way healthcare providers deliver care and manage patient data. Cloud computing enables healthcare organizations to store and access large amounts of data securely and efficiently, while AI-powered tools can help healthcare providers analyze this data to make more accurate diagnoses and personalized treatment plans.

Implementation:

Tools and libraries:

Libraries:

Keras: Keras is an open-source neural network library written in Python. It is designed to make it easy to build and experiment with deep neural networks. Keras provides a high-level interface to popular deep learning frameworks, such as TensorFlow and Theano, making it easy to build, train, and deploy deep learning models.

Imblearn: is a Python library for dealing with imbalanced datasets in machine learning. Imbalanced datasets occur when one class is much more prevalent than the other class(es) in the dataset. This can cause problems for machine learning algorithms because they can become biased towards the majority class.

Sklearn: is a popular Python library for machine learning. It provides a wide range of tools for data preprocessing, feature selection, model selection, and model evaluation. sklearn is built on top of other Python libraries, such as NumPy, SciPy, and matplotlib, and provides a consistent interface for various machine learning tasks.

Pandas: is a popular open-source Python library for data manipulation and analysis. It provides powerful data structures for efficiently storing and manipulating large and complex datasets, as well as tools for working with missing data, time-series data, and categorical data.

Numpy: is a Python library for scientific computing and numerical computations. It provides high-performance multidimensional arrays and matrices, as well as tools for working with these arrays. NumPy is an essential library for many scientific and data analysis tasks, and it is widely used in the fields of machine learning, data science, and engineering.

Tools:

IBM cloud: IBM Cloud is a cloud computing platform that provides a wide range of services for developing, deploying, and managing applications in the cloud. It offers a secure, scalable, and flexible environment for building and running cloud-native applications, as well as integrating with existing on-premise infrastructure.

Ibm Watson studio: IBM Watson Studio is an integrated development environment (IDE) for building and deploying AI and machine learning models. It provides a suite of tools and services for data scientists, developers, and business analysts to collaboratively build, train, and deploy machine learning models.

Visual studio: is an integrated development environment (IDE) created by Microsoft for developing a variety of applications such as desktop, web, mobile, and game development. It provides a range of tools and services for developers to build, debug, and deploy applications in a variety of programming languages.

Implementation phase:

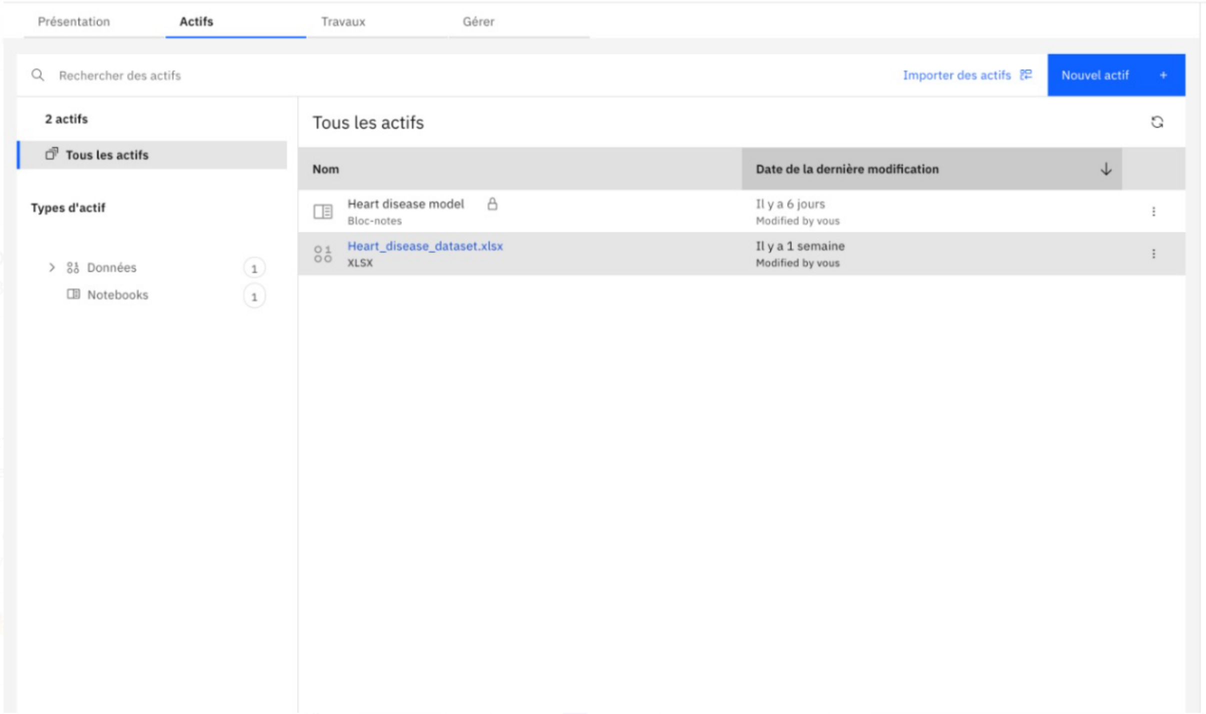

Figure 4:ibm cloud dashboard

We have visualized the data with ibm visualisation tools and have noticed an imbalance

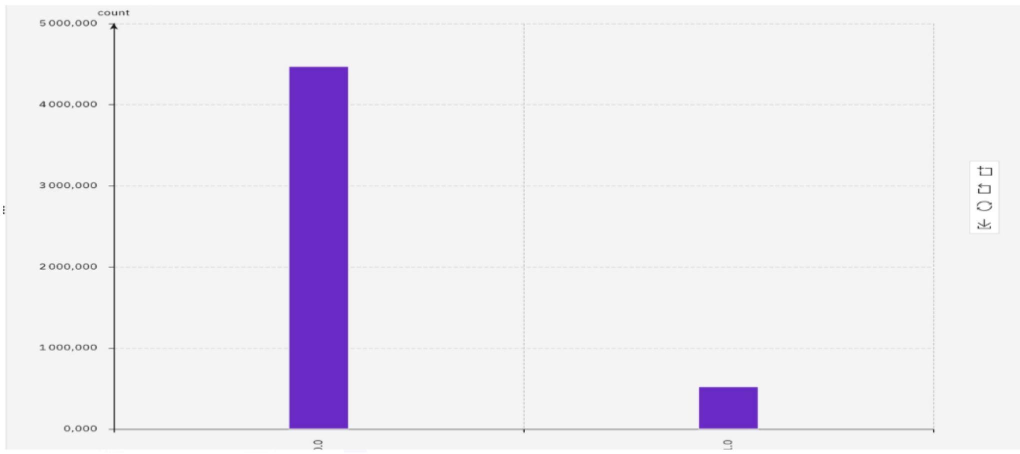

Figure 5:data visualization

So to solve the imbalance problem we need to import the libraries:

```
import pandas as pd
import numpy as np
import keras
import sklearn
from keras.utils.np_utils import to_categorical
from sklearn.metrics import classification_report,accuracy_score
from keras.models import Sequential
from keras.layers import Dense
from keras.optimizers import Adam
from sklearn import model_selection
from sklearn.neighbors import KNeighborsClassifier as knn
import seaborn as sns
from pylab import rcParams
import matplotlib.pyplot as plt
```

Figure 6:libraries importation

To solve the problem we compiled randomOverSample()r and RandomUnderSampler() functions of the library imblearn for a better performance.

Random Oversampling involves supplementing the training data with multiple copies of some of the minority classes. Oversampling can be done more than once (2x, 3x, 5x, 10x, etc.) This is one of the earliest proposed methods, that is also proven to be robust.Instead of duplicating every sample in the minority class, some of them may be randomly chosen with replacement.

Random undersampling randomly remove samples from the majority class, with or without replacement. This is one of the earliest techniques used to alleviate imbalance in the dataset, however, it may increase the variance of the classifier and is very likely to discard useful or important samples

```
ros = RandomOverSampler()

X_ros, y_ros = ros.fit_resample(x, y)
c_combined=RandomUnderSampler()
[214]  ✓ 0.4s
```

Figure 7:Data balancing

The content of our dataset:

	HeartDiseaseorAttack	HighBP	HighChol	CholCheck	BMI	Smoker	Stroke	Diabetes	PhysActivity	Fruits	...	AnyHealthcare	NoDocbcCost	GenHlth	MentHlth	PhysHlth	DiffWalk	Sex	Age	Education	Income
0	0.0	1.0	1.0	1.0	40.0	1.0	0.0	0.0	0.0	0.0	...	1.0	0.0	5.0	18.0	15.0	1.0	0.0	9.0	4.0	3.0
1	0.0	0.0	0.0	0.0	25.0	1.0	0.0	0.0	1.0	0.0	...	0.0	1.0	3.0	0.0	0.0	0.0	0.0	7.0	6.0	1.0
2	0.0	1.0	1.0	1.0	28.0	0.0	0.0	0.0	0.0	1.0	...	1.0	1.0	5.0	30.0	30.0	1.0	0.0	9.0	4.0	8.0
3	0.0	1.0	0.0	1.0	27.0	0.0	0.0	0.0	1.0	1.0	...	1.0	0.0	2.0	0.0	0.0	0.0	0.0	11.0	3.0	6.0
4	0.0	1.0	1.0	1.0	24.0	0.0	0.0	0.0	1.0	1.0	...	1.0	0.0	2.0	3.0	0.0	0.0	0.0	11.0	5.0	4.0

5 rows x 22 columns

Figure 8:five first rows of the data set

In our study case we have decided to work with six different algorithms to see the one with the best performance and accuracy:

Naive Bayes is a classification algorithm that is based on the Bayes theorem. It is a probabilistic algorithm that makes use of conditional probabilities to predict the class label of a given data point.

K-Nearest Neighbors (KNN) is a classification algorithm that is used to predict the class of a data point based on the classes of its K nearest neighbors in the feature space. KNN is a non-parametric and lazy learning algorithm, meaning that it does not make any assumptions about the underlying distribution of the data and it does not have a separate training phase.

Logistic regression is a classification algorithm that is used to predict the probability of a binary outcome (i.e., a class label that can take on one of two values). The algorithm models the relationship between the predictor variables and the probability of the outcome using a logistic function.

Decision tree is a classification algorithm that builds a tree-like model of decisions and their possible consequences. It partitions the feature space into a hierarchical tree structure where each node represents a decision based on a feature, and each branch represents the possible outcome of that decision. The algorithm recursively splits the data into smaller subsets based on the most discriminative features until the subsets are homogeneous enough to be assigned a class label.

Random Forest is an ensemble learning algorithm that combines multiple decision trees to create a more robust and accurate model. It is a popular and powerful algorithm used for both classification and regression tasks.

ALGORITHM	ACCURACY	RECALL	F1_SCORE	PRECISION	TIME
NAÏVE BAYS	72%	74%	73%	74%	2S
LOGISTIC REGRESSION	77%	78%	73%	74%	3.3S
KNN	90.19%	92%	90%	90%	44S
DEEPLEARNING	93%	93%	92%	92%	2.5H
DECISION TREE	95%	95%	95%	94%	2S
RANDOM FOREST	97%	96%	97%	96%	54S

As the results show the algorithm with the best accuracy and the highest performance was random forest.

Conclusion:

After analyzing the collected patient data and applying machine learning algorithms, the goal of our project was to develop a predictive model for heart disease. we have developed a model that can accurately predict the risk of heart disease in patients.it has significant potential to improve patient outcomes and quality of life. By accurately identifying patients who are at high risk , we can take proactive steps to prevent or treat the disease, leading to better health outcomes and a healthier population.

Conclusion:

In conclusion, the use of cloud computing and artificial intelligence in the prediction and management of heart disease is a promising approach that has the potential to revolutionize the healthcare industry. By leveraging the power of cloud-based platforms and sophisticated algorithms, healthcare providers can analyze large amounts of patient data and identify risk factors for heart disease with greater accuracy and efficiency.

Moreover, the use of AI-based predictive models can help healthcare providers develop personalized treatment plans for patients based on their individual health history, genetic makeup, and lifestyle factors. This can lead to better outcomes and improved quality of life for patients with heart disease,since the integration of cloud computing and artificial intelligence into the healthcare system can also reduce costs and increase accessibility to healthcare services, especially for underserved populations.

However, it is important to ensure that patient data is properly protected and that ethical considerations are taken into account when implementing these technologies.

Overall, the integration of cloud computing and AI in the healthcare industry has the potential to revolutionize the way we diagnose, manage, and treat heart disease. With continued research and development in this area, we can look forward to a future where heart disease is more effectively prevented and managed, leading to better health outcomes for patients.

References :

Rashid, A., & Chaturvedi, A. (2019). Cloud computing characteristics and services: a brief review. *International Journal of Computer Sciences and Engineering*, *7*(2), 421-426.

RASHID, Aaqib et CHATURVEDI, Amit. Cloud computing characteristics and services: a brief review. *International Journal of Computer Sciences and Engineering*, 2019, vol. 7, no 2, p. 421-426.

Rashid, Aaqib, and Amit Chaturvedi. "Cloud computing characteristics and services: a brief review." *International Journal of Computer Sciences and Engineering* 7.2 (2019): 421-426.

Diaby, T., & Rad, B. B. (2017). Cloud computing: a review of the concepts and deployment models. *International Journal of Information Technology and Computer Science*, *9*(6), 50-58.

DIABY, Tinankoria et RAD, Babak Bashari. Cloud computing: a review of the concepts and deployment models. *International Journal of Information Technology and Computer Science*, 2017, vol. 9, no 6, p. 50-58.

Sawyer, Jennifer. "Wearable Internet of Medical Things sensor devices, artificial intelligence-driven smart healthcare services, and personalized clinical care in COVID-19 telemedicine." *American Journal of Medical Research* 7.2 (2020): 71-77.

“Machine Learning is a form of artificial intelligence…” [Online]. Available : https://www.oracle.com/fr/artificial-intelligence/machine-learning/what-is-machine-learning/.

“How does Machine Learning work?...” [Online]. Available: https://datascientest.com/machine-learning-tout-savoir.

Helm, J. Matthew, et al. "Machine learning and artificial intelligence: definitions, applications, and future directions." *Current reviews in musculoskeletal medicine* 13 (2020): 69-76.