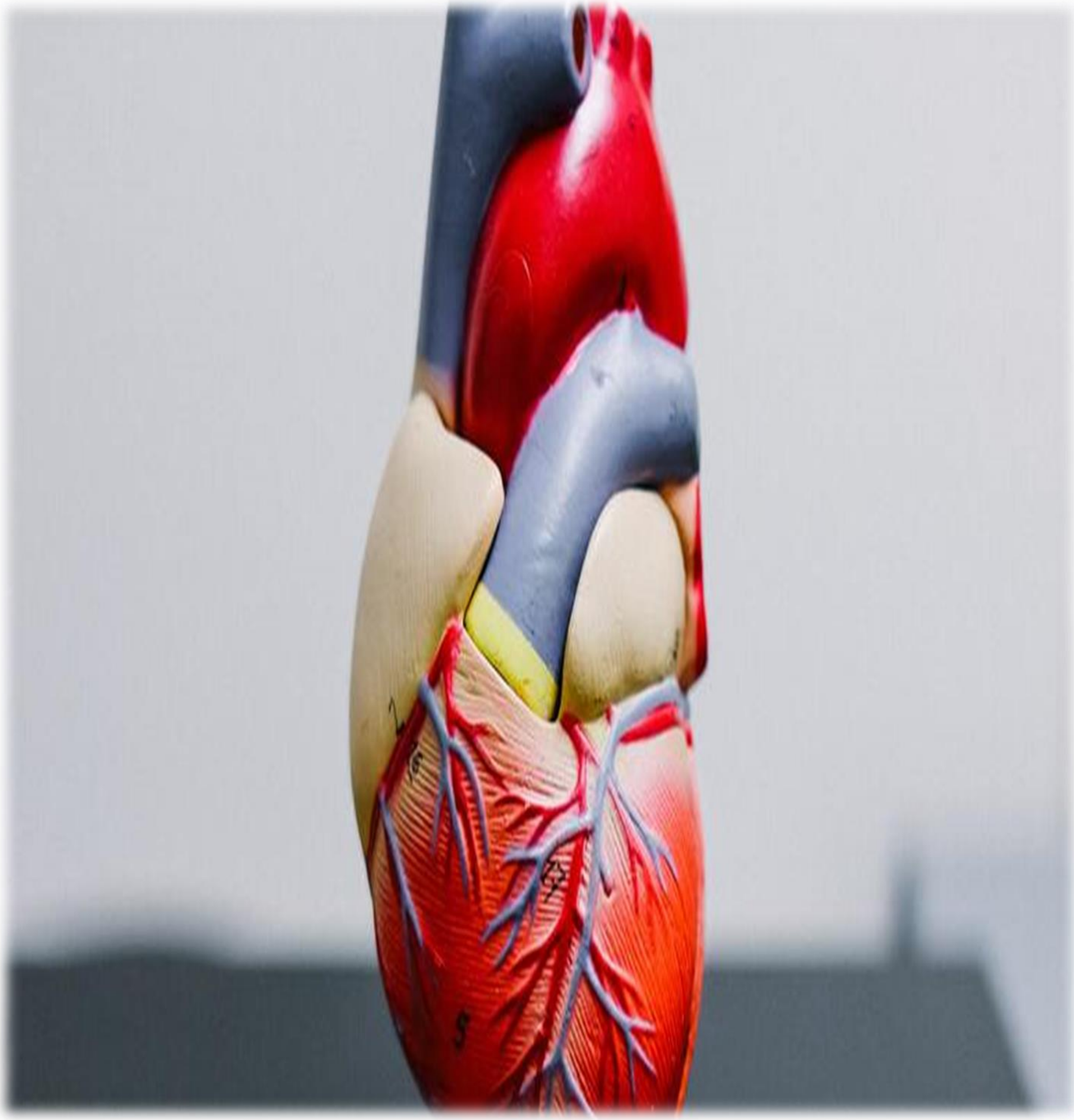
HOSPITAL MORTALITY PREDICTION

Author: Adesh Mekhe

October 5



1. Problem Statement:

The goal of this project is to develop a machine learning-based system that predicts the mortality risk of patients admitted to a hospital. By accurately identifying patients who are at higher risk of mortality, healthcare providers can allocate resources more effectively, make informed treatment decisions, and ultimately improve patient outcomes.

2. Market/Customer/Business Need Assessment:

The healthcare industry is under constant pressure to enhance patient care and optimize resource utilization. Predicting hospital mortality can help hospitals prioritize patient care, allocate staff efficiently, and reduce healthcare costs. Additionally, accurate mortality prediction aligns with the industry's shift towards value-based care and patient-centric treatment plans.

3. Target Specifications and Characterization:

The target users of this system are healthcare professionals, including doctors, nurses, and hospital administrators. The system should provide accurate mortality risk scores for individual patients upon admission and allow users to access and interpret these scores in real time..

4. External Search:

Dataset:

Let's import the dataset and have a look at it!

Loading the Data

```
[ ] df = pd.read_csv("In-Hospital-Mortality.csv")
```

```
[ ] df = df.drop(['ID'], axis=1)
```

```
[ ] df.shape
```

```
(1177, 50)
```

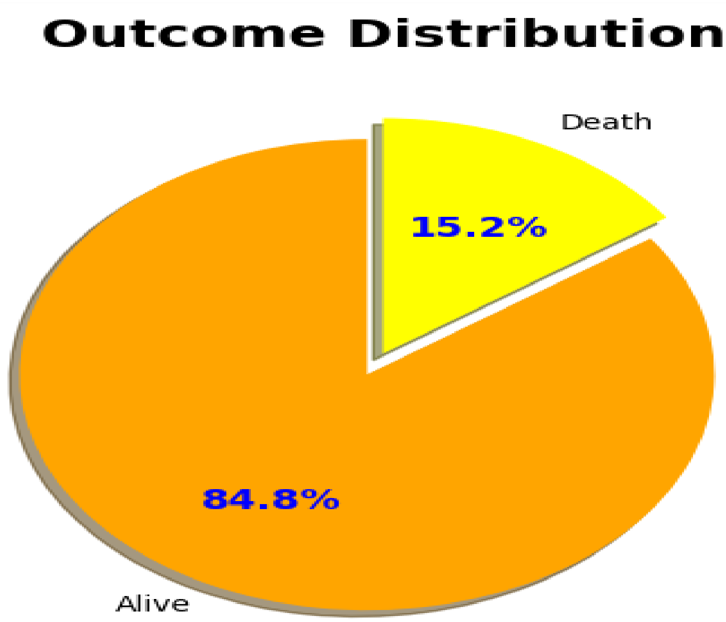
```
[ ] df.describe()
```

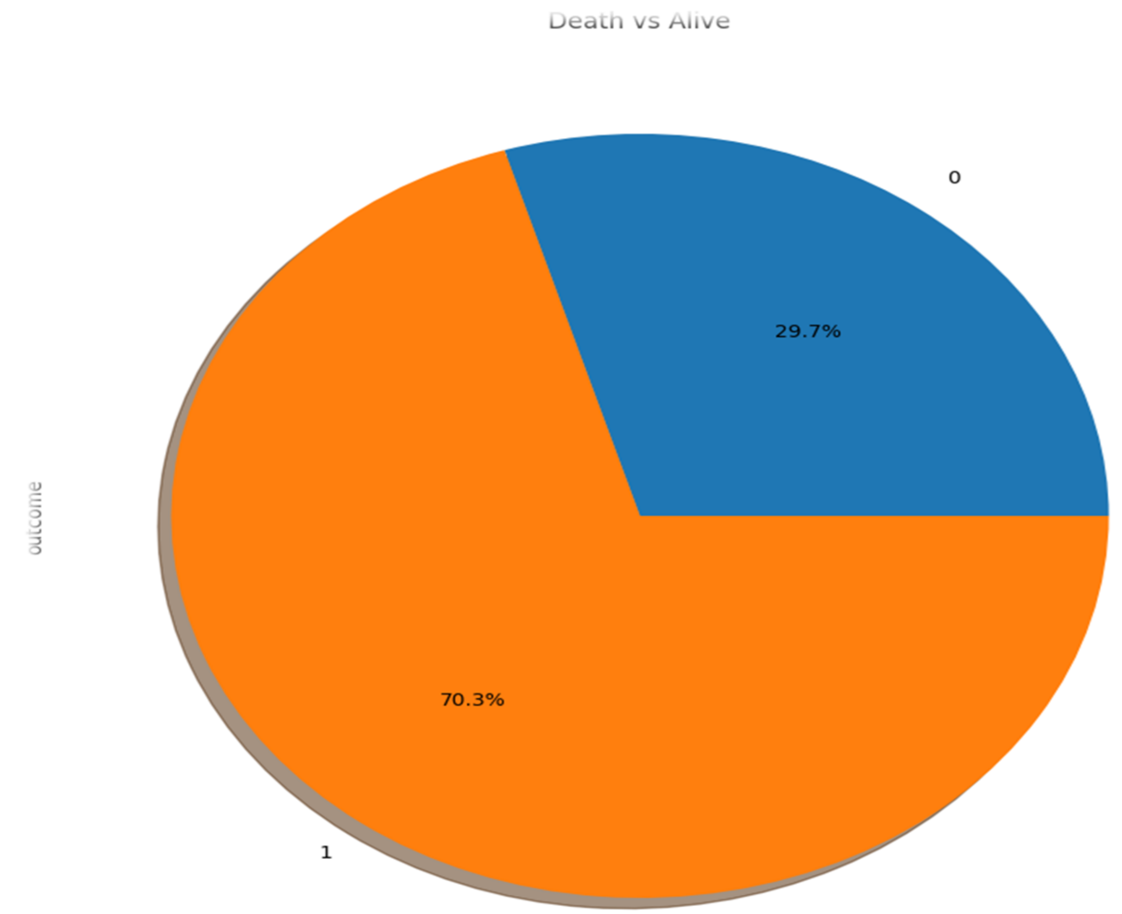
	group	outcome	age	gendera	BMI	hypertensive	atrialfibrillation	CHD with no MI	diabetes	deficiencyanemias
count	1177.000000	1176.000000	1177.000000	1177.000000	962.000000	1177.000000	1177.000000	1177.000000	1177.000000	1177.000000
mean	1.299065	0.135204	74.055225	1.525064	30.188278	0.717927	0.451147	0.085811	0.421410	0.338997
std	0.458043	0.342087	13.434061	0.499584	9.325997	0.450200	0.497819	0.280204	0.493995	0.473570
min	1.000000	0.000000	19.000000	1.000000	13.346801	0.000000	0.000000	0.000000	0.000000	0.000000
25%	1.000000	0.000000	65.000000	1.000000	24.326461	0.000000	0.000000	0.000000	0.000000	0.000000
50%	1.000000	0.000000	77.000000	2.000000	28.312474	1.000000	0.000000	0.000000	0.000000	0.000000
75%	2.000000	0.000000	85.000000	2.000000	33.633509	1.000000	1.000000	0.000000	1.000000	1.000000
max	2.000000	1.000000	99.000000	2.000000	104.970366	1.000000	1.000000	1.000000	1.000000	1.000000

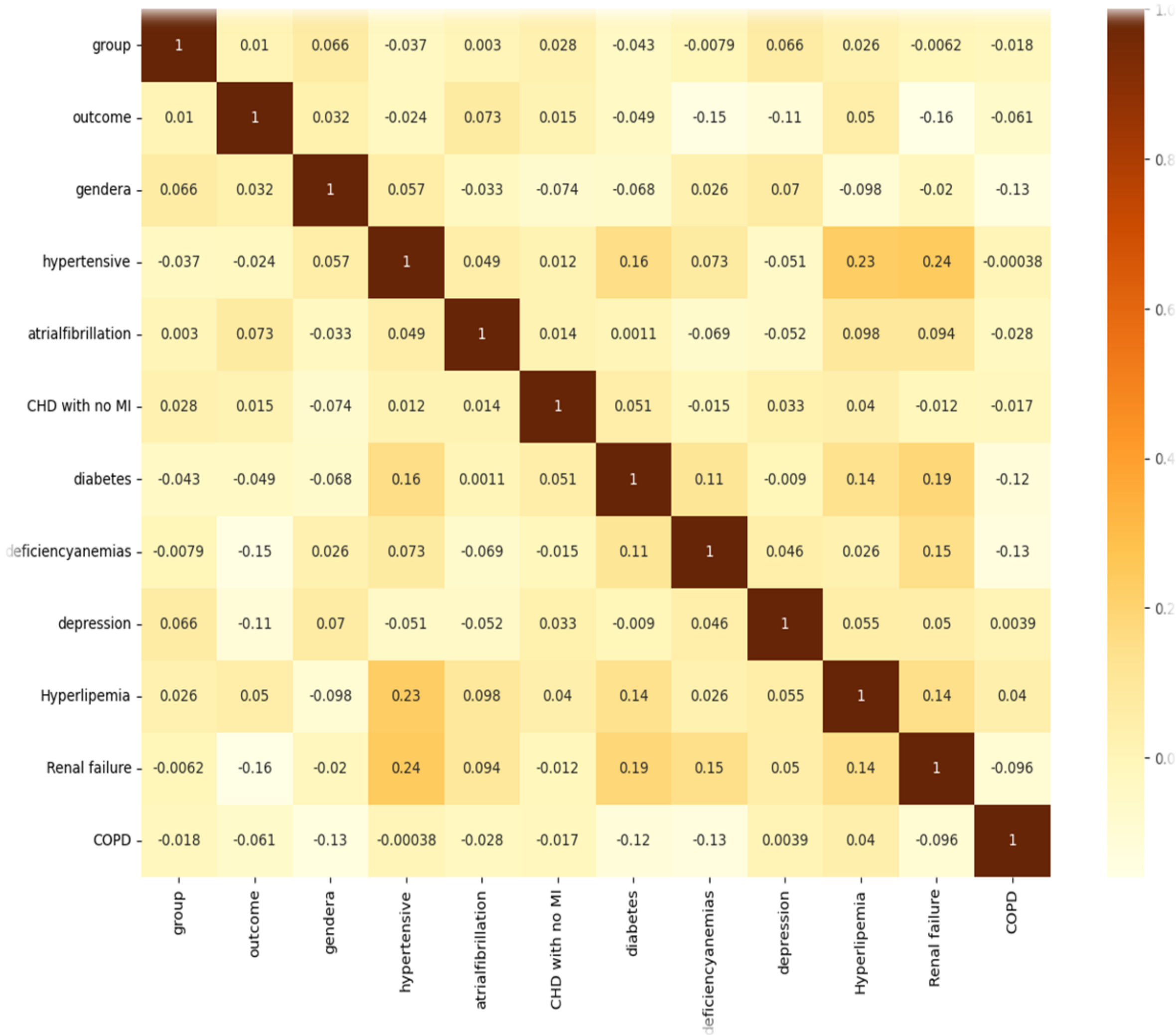
```
df.info()
```

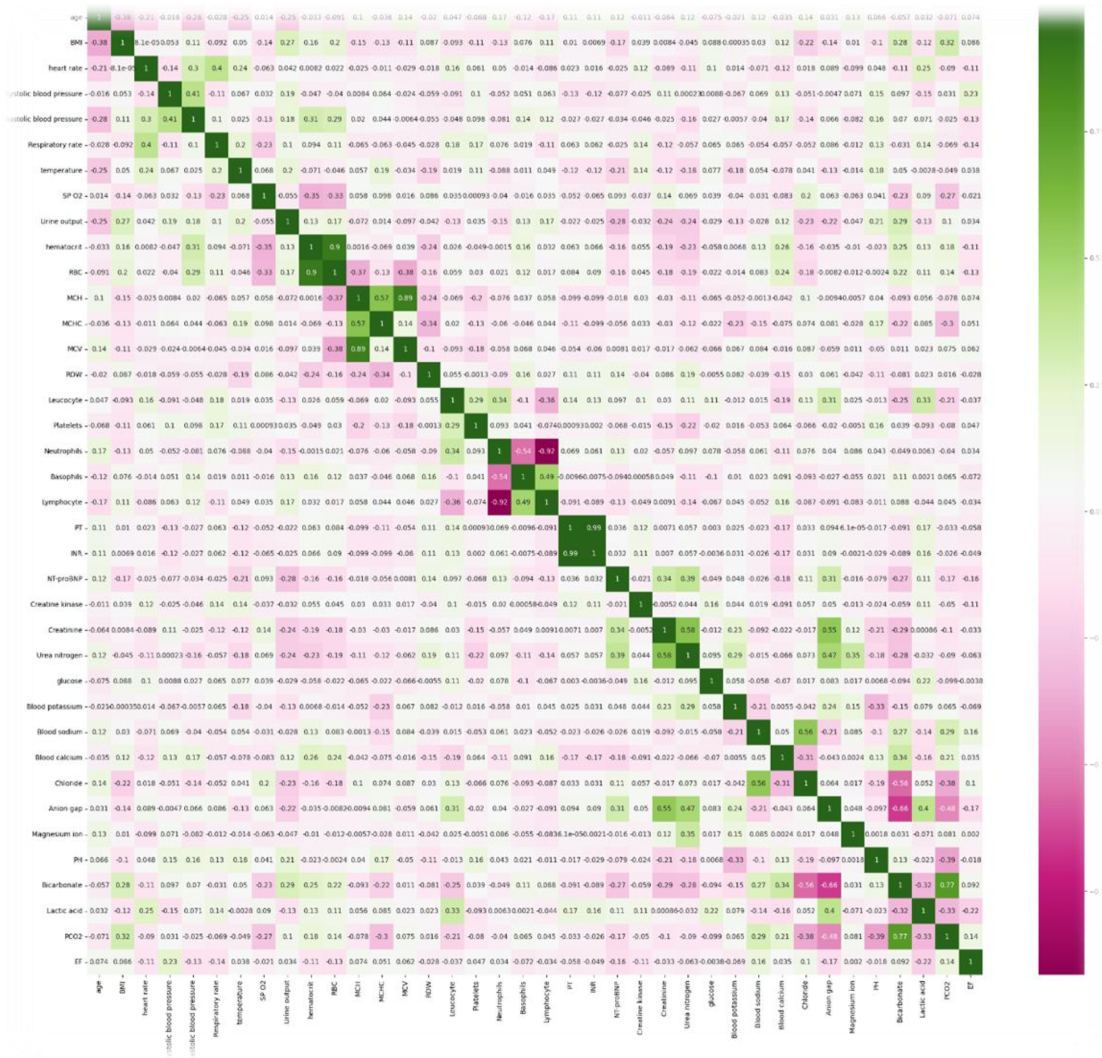
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1177 entries, 0 to 1176
Data columns (total 50 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   group                     1177 non-null   int64
 1   outcome                   1176 non-null   float64
 2   age                       1177 non-null   int64
 3   gendera                   1177 non-null   int64
 4   BMI                       962 non-null    float64
 5   hypertensive              1177 non-null   int64
 6   atrialfibrillation        1177 non-null   int64
 7   CHD with no MI            1177 non-null   int64
 8   diabetes                  1177 non-null   int64
 9   deficiencyanemias         1177 non-null   int64
 10  depression                1177 non-null   int64
 11  Hyperlipemia              1177 non-null   int64
 12  Renal failure             1177 non-null   int64
 13  COPD                      1177 non-null   int64
 14  heart rate                1164 non-null   float64
 15  Systolic blood pressure   1161 non-null   float64
 16  Diastolic blood pressure  1161 non-null   float64
 17  Respiratory rate          1164 non-null   float64
 18  temperature               1158 non-null   float64
 19  SP O2                     1164 non-null   float64
 20  Urine output              1141 non-null   float64
 21  hematocrit                1177 non-null   float64
 22  RBC                       1177 non-null   float64
 23  MCH                       1177 non-null   float64
 24  MCHC                      1177 non-null   float64
 25  MCV                       1177 non-null   float64
 26  RDW                       1177 non-null   float64
 27  Leucocyte                 1177 non-null   float64
 28  Platelets                 1177 non-null   float64
 29  Neutrophils               1033 non-null   float64
 30  Basophils                 918 non-null    float64
 31  Lymphocyte                1032 non-null   float64
 32  PT                        1157 non-null   float64
 33  INR                       1157 non-null   float64
 34  NT-proBNP                 1177 non-null   float64
 35  Creatine kinase           1012 non-null   float64
 36  Creatinine                1177 non-null   float64
 37  Urea nitrogen             1177 non-null   float64
 38  glucose                   1159 non-null   float64
 39  Blood potassium           1177 non-null   float64
 40  Blood sodium              1177 non-null   float64
 41  Blood calcium             1176 non-null   float64
 42  Chloride                  1177 non-null   float64
 43  Anion gap                 1177 non-null   float64
 44  Magnesium ion             1177 non-null   float64
 45  PH                        885 non-null    float64
 46  Bicarbonate               1177 non-null   float64
 47  Lactic acid               948 non-null    float64
 48  PCO2                      883 non-null    float64
 49  EF                        1177 non-null   int64
dtypes: float64(37), int64(13)
memory usage: 459.9 KB
```

Exploratory Data Analysis (EDA) :

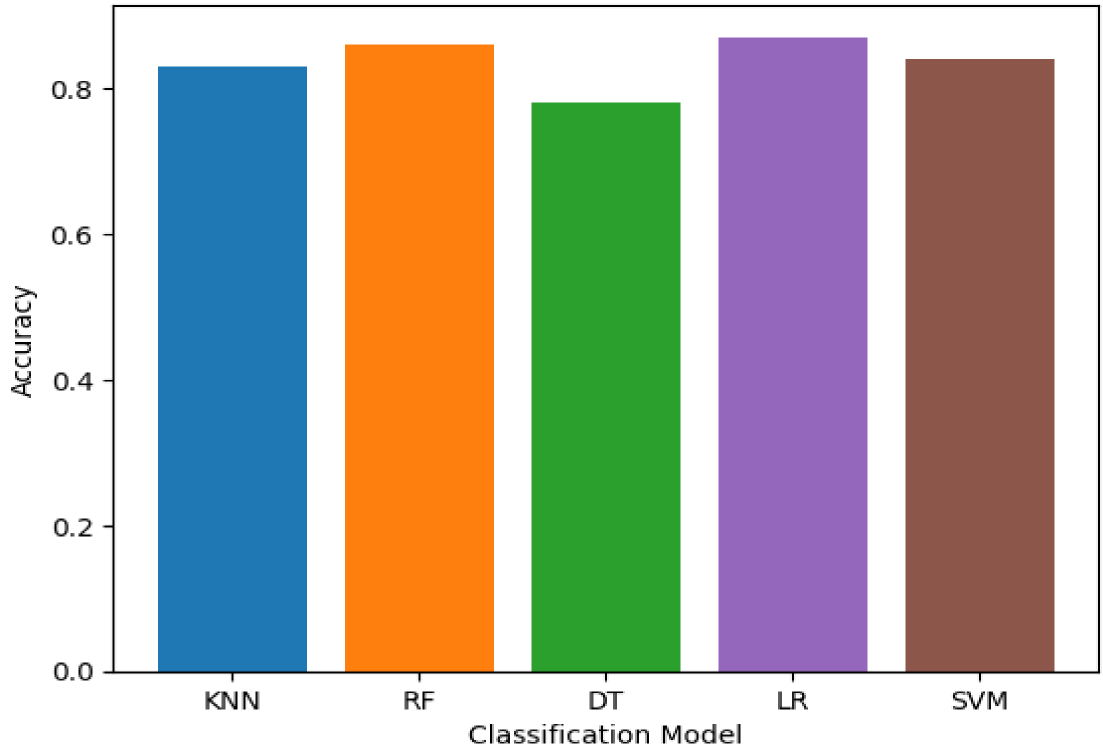



Death vs Alive
0
29.7%
70.3%
1
outcome


group
outcome
gendera
hypertensive
atrialfibrillation
CHD with no MI
diabetes
deficiencyanemias
depression
Hyperlipemia
Renal failure
COPD
1 0.01 0.066 -0.037 0.003 0.028 -0.043 -0.0079 0.066 0.026 -0.0062 -0.018
0.01 1 0.032 -0.024 0.073 0.015 -0.049 -0.15 -0.11 0.05 -0.16 -0.061
0.066 0.032 1 0.057 -0.033 -0.074 -0.068 0.026 0.07 -0.098 -0.02 -0.13
-0.037 -0.024 0.057 1 0.049 0.012 0.16 0.073 -0.051 0.23 0.24 -0.00038
0.003 0.073 -0.033 0.049 1 0.014 0.0011 -0.069 -0.052 0.098 0.094 -0.028
0.028 0.015 -0.074 0.012 0.014 1 0.051 -0.015 0.033 0.04 -0.012 -0.017
-0.043 -0.049 -0.068 0.16 0.0011 0.051 1 0.11 -0.009 0.14 0.19 -0.12
-0.0079 -0.15 0.026 0.073 -0.069 -0.015 0.11 1 0.046 0.026 0.15 -0.13
0.066 -0.11 0.07 -0.051 -0.052 0.033 -0.009 0.046 1 0.055 0.05 0.0039
0.026 0.05 -0.098 0.23 0.098 0.04 0.14 0.026 0.055 1 0.14 0.04
-0.0062 -0.16 -0.02 0.24 0.094 -0.012 0.19 0.15 0.05 0.14 1 -0.096
-0.018 -0.061 -0.13 -0.00038 -0.028 -0.017 -0.12 -0.13 0.0039 0.04 -0.096 1



After the EDA we fit the model and to know which is the best model to fit



Conclusion: Comparison of classification techniques, we can evaluate that here Logistic Regression Classification and Random Forest have the most optimal result of the accuracy.

5. Benchmarking:

In evaluating existing solutions for hospital mortality prediction, we've identified key products and approaches:

1. **Sepsis-3 Criteria:** Limited in scope to septic patients, lacking broader applicability.
2. **Clinical Scoring Systems:** Manual calculations, may not capture intricate variables.
3. **Rule-Based Approaches:** Fixed thresholds limit adaptability and robustness.
4. **Early Warning Systems:** Alerts based on deviations, not comprehensive risk assessment.

Comparative Analysis:

- **Accuracy:** Proposed system leverages advanced machine learning for higher accuracy.
- **Scope:** Proposed system covers diverse patients and conditions.
- **Real-Time Prediction:** Our system provides continuous real-time risk assessment upon admission.
- **Integration:** Seamlessly integrates with electronic health records.
- **Interpretability:** Emphasizes transparency and easy interpretation for healthcare professionals.
- **Scalability:** Adaptable to hospitals of varying sizes and configuration This analysis underscores our proposed system's unique advantages and positions it as a versatile, accurate, and adaptable solution for hospital mortality prediction.

6. Applicable Patents:

1. **Patent A: Adaptive Mortality Prediction Algorithm**

Description: This patent introduces an innovative algorithm that dynamically adjusts prediction thresholds based on patient-specific factors and medical context. By considering individual patient characteristics and real-time data, the algorithm optimizes the balance between sensitivity and specificity, enhancing the accuracy of mortality predictions. This adaptive approach ensures that predictions remain relevant and effective across a diverse range of patients and conditions.

2. **Patent B: Real-Time Health Data Integration Framework**

Description: Patent B presents a framework for seamless integration of real-time health data from various sources, including electronic health records and wearable devices. The system collects, pre-processes, and analyzes continuous streams of patient data, providing up-to-the-minute insights into patient health. By combining historical records with real-time updates, healthcare providers gain a comprehensive understanding of each patient's evolving condition, enabling timely interventions and accurate mortality risk assessments.

7. Applicable Regulations:

- Health Insurance Portability and Accountability Act (HIPAA)
- General Data Protection Regulation (GDPR)
- Ethical Guidelines for AI in Healthcare
- Clinical Decision Support Software Regulation (FDA)
- Institutional Review Board (IRB) Approval
- Health Technology Assessment (HTA)
- Data Security and Encryption Standards
- Local Healthcare Regulation

8. Applicable Constraints:

- Budget limitations
- Time constraints
- Data availability and quality
- Expertise and skill sets
- Model Complexity
- Computational resources
- Privacy concerns
- Scalability
- Acceptance by healthcare professionals
- Cultural and organizational factors
- Regulatory compliance
- Cost-efficiency

9. Business Opportunity

- Improved decision-making for hospitals and other healthcare organization.
- New opportunities for pharmaceutical companies and insurance companies.
- Improved public health interventions.

10. Concept Generation:

The idea originated from the need to enhance patient care and resource allocation in hospitals. Initial brainstorming sessions with medical professionals and data scientists led to the concept of leveraging machine learning for mortality prediction.

11. Concept Development:

The proposed system utilizes historical patient data, including demographics, medical history, vital signs, and laboratory results, to train a machine learning model. This model generates a mortality risk score upon patient admission, helping healthcare providers make more informed decisions.

12. Final Product Prototype

The final product prototype is a comprehensive hospital mortality prediction system that seamlessly integrates advanced machine learning algorithms, real-time health data, and user-friendly interfaces to provide accurate and timely predictions of patient mortality risk upon hospital admission. The system aims to empower healthcare professionals with actionable insights for informed decision-making and improved patient care outcomes

13. Product Details:

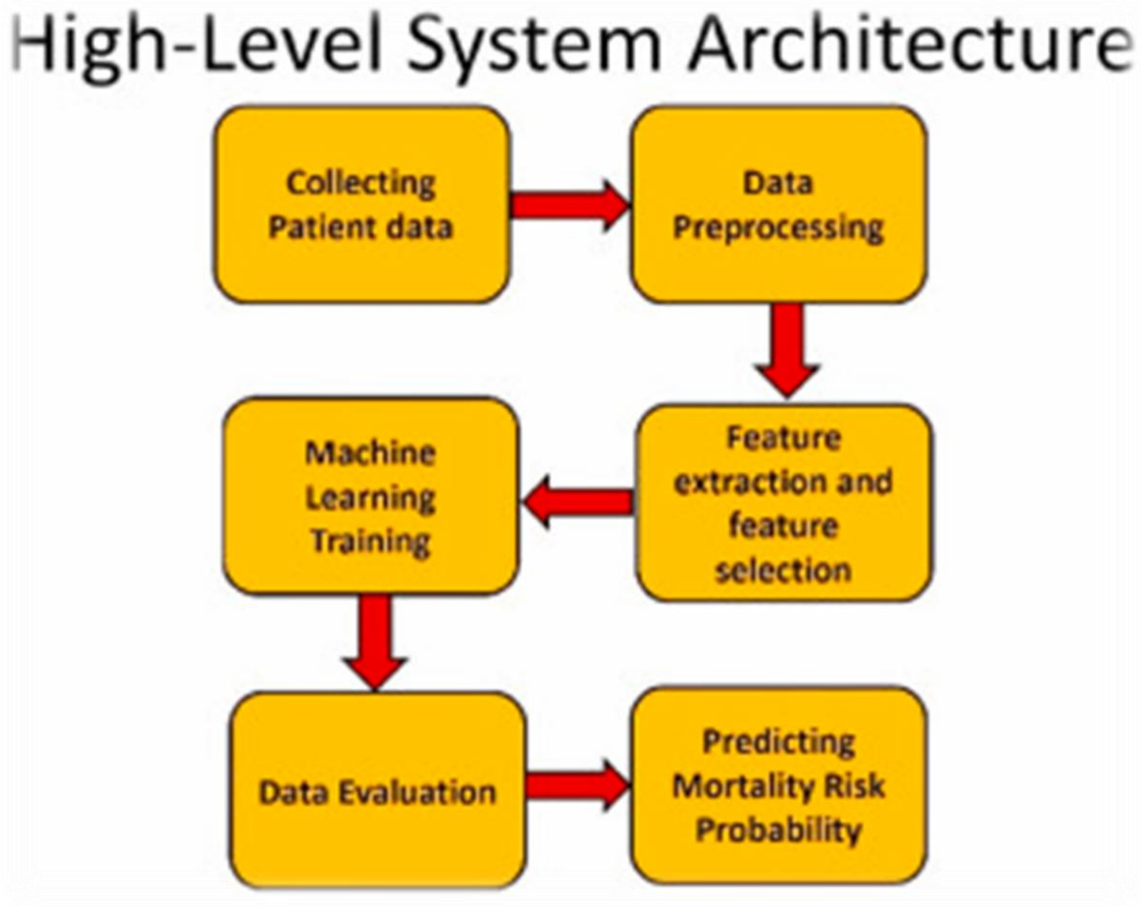
- **How does it work?** The system combines patient data with a trained machine learning model to predict the likelihood of mortality. It provides a risk score and associated confidence level..
- **Algorithms:** Utilizes a combination of feature engineering, classification algorithms such as logistic regression, Decision tree, SVM, Navie Bayes, random forests.
- **Team Required:** Data scientists, machine learning engineers, healthcare professionals, web developers, user interface designers.
- **Tools:**

  Python: It's a programming language that will be used for building the service.

  Pandas: Pandas is a library mainly used for handling, manipulating and transforming data.

Scikit-learn: It is the gold standard library for machine learning which comes with plenty of algorithms to perform different tasks such as regression, classification etc.

Matplotlib and Seaborn: Both of these libraries are used for visualization purposes.



14. Business Model:

There are a number of different business models that can be used for hospital mortality prediction projects. but the major ones lie in helping hospital mortality prediction is Software as a service (SaaS) so we use these business model:

In a SaaS business model, the hospital mortality prediction model is hosted in the cloud and is accessed by customers over the internet. This is a popular business model for hospital mortality prediction projects, as it is relatively easy to set up and maintain. Additionally, SaaS models allow hospitals to scale their usage of the model up or down as needed.

As you can see, the SaaS business model typically involves three main components:

1. The SaaS provider: This is the company that develops and maintains the SaaS-based application.
2. The customer: This is the organization that uses the SaaS-based application.
3. The internet: This is the medium that connects the SaaS provider to the customer.

In the case of a hospital mortality prediction project, the SaaS provider would be the company that develops and maintains the hospital mortality prediction model. The customer would be the hospital or healthcare organization that uses the hospital mortality prediction model. And the internet would be the medium that connects the SaaS provider to the customer.

Benefits of using a SaaS business model for hospital mortality prediction projects:

- Easy to set up and maintain: SaaS models are typically hosted in the cloud, so hospitals do not need to invest in their own hardware and software. Additionally, SaaS providers are responsible for maintaining the model and ensuring that it is up-to-date. This can free up hospital staff to focus on other tasks.
- Scalable: SaaS models allow hospitals to scale their usage of the model up or down as needed. This is important for hospitals that may have seasonal fluctuations in patient volume or that are growing rapidly.
- Affordable: SaaS models are typically priced on a subscription basis, which can make them more affordable for hospitals than on-premises deployments. This is especially important for smaller hospitals with limited budgets.
- Secure: SaaS providers typically have robust security measures in place to protect patient data. This is important for hospitals, which are subject to strict data security regulations.
- Accessible: SaaS models are accessible from anywhere with an internet connection. This makes it easy for clinicians to use the model at the point of care.
- Collaborative: SaaS models can be shared with other hospitals or healthcare organizations, which can facilitate collaboration and research.

In addition to these benefits, SaaS models can also help hospitals to improve patient outcomes. By using a SaaS-based hospital mortality prediction model, hospitals can identify patients who are at high risk of death and provide them with more targeted care. This can lead to improved patient outcomes and reduced mortality rates.

Overall, the SaaS business model offers a number of benefits for hospitals that are considering using a hospital mortality prediction model. SaaS models are easy to set up and maintain, scalable, affordable, secure, accessible, and collaborative. Additionally, SaaS models can help hospitals to improve patient outcomes.

Here are some specific examples of how hospitals can benefit from using a SaaS-based hospital mortality prediction model:

- A hospital can use a SaaS-based hospital mortality prediction model to identify patients who are at high risk of death from sepsis. The hospital can then use this information to provide these patients with more targeted care and to reduce sepsis mortality rates.
- A hospital can use a SaaS-based hospital mortality prediction model to identify patients who are at high risk of death from heart disease. The hospital can then use this information to target its marketing and sales efforts for its new heart disease medication to these patients.

15. Financial Equation

The market for hospital mortality prediction projects is growing exponentially.

This is because the demand for hospital mortality prediction models is increasing rapidly, as hospitals and other healthcare organizations are increasingly looking for ways to improve patient outcomes and reduce costs.

The financial model equation for hospital mortality prediction is that are growing exponentially which is:

$$y = A * e^{(bt)} + c$$

- **Where: y = total profit**
- **A = initial profit**
- **b = growth rate**
- **t = time**
- **c = production, maintenance, etc. costs**

To use this equation, you would need to estimate the values of A, b, and c. You can do this by looking at historical data on your sales and profits. Once you have estimated these values, you can use the equation to predict your future pro