

PROJECT REPORT

GROUP MEMBERS:

|  |  |
| --- | --- |
| UZAIR AHMED | RANA HAMZA ALI KHAN |
| CSC-21F-130 | CSC-21F-105 |

**COURSE :** **ARTIFICIAL INTELLIGENCE**

**PROGRAM : BSCS (MORNING)**

**SECTION :** **6TH ‘C’**

**COURSE COORDINATOR** **:** **MISS**.**AQSA UMAR**

PREDICTING SURVIVAL IN PATIENTS WITH HEART FAILURE USING KNN & EDA.

ABSTRACT:

This research investigates the use of the K-Nearest Neighbors (KNN) algorithm to predict survival outcomes in patients with heart failure. By leveraging a clinical dataset from Kaggle, we conduct exploratory data analysis (EDA) and build a predictive model. The study highlights the effectiveness of machine learning in healthcare and identifies key factors influencing patient survival.

INTRODUCTION:

Heart failure is a serious medical condition where the heart is unable to pump blood efficiently, leading to various health complications and a high mortality rate. Predicting which patients are at higher risk of death can significantly aid in providing timely interventions and personalized treatments. This study aims to use machine learning, particularly the K-Nearest Neighbors (KNN) algorithm, to predict survival in heart failure patients based on clinical data.

RELATED WORK:

The application of machine learning in healthcare, especially for predictive analytics, has gained significant attention. Various algorithms such as logistic regression, decision trees, and support vector machines have been employed to predict outcomes in heart failure patients. KNN, known for its simplicity and effectiveness, has been used in various medical studies to classify patients based on their proximity to other patients with similar characteristics. Its ability to handle non-linear data and its intuitive nature make it a suitable choice for this study.

METHODOLOGY:

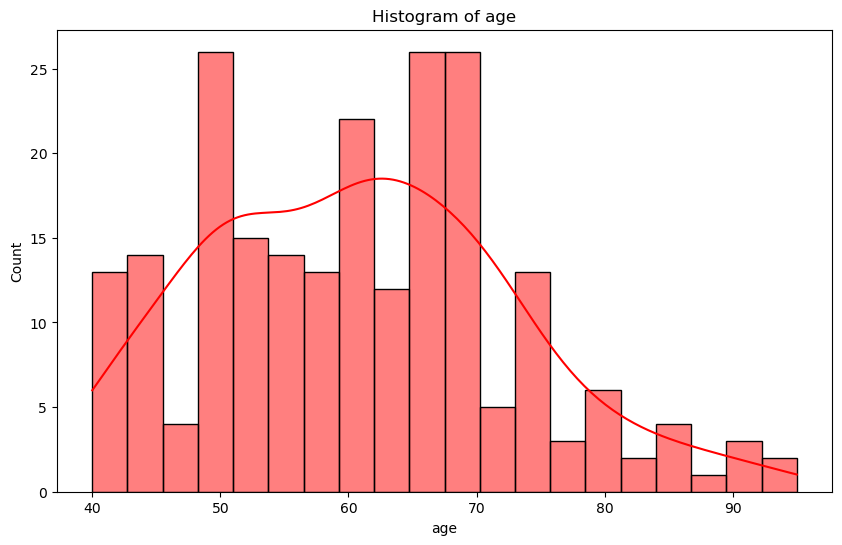
* **DATASET:**

The dataset used in this study is the "Heart Failure Clinical Records" dataset available on Kaggle. It contains 299 records of patients with 13 clinical features, including age, anemia, creatinine phosphokinase, diabetes, ejection fraction, high blood pressure, platelets, serum creatinine, serum sodium, sex, smoking, follow-up time, and the target variable, death event (DEATH\_EVENT).

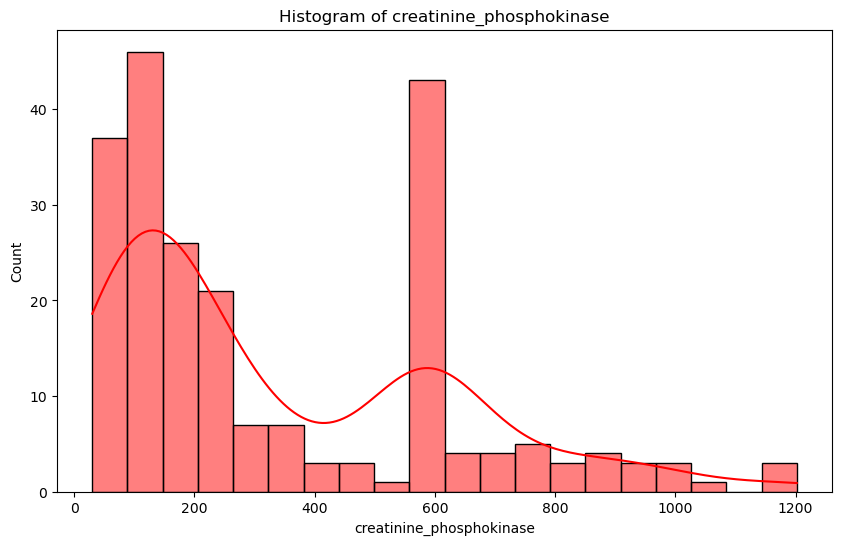
* **EXPLORATORY DATA ANALYSIS (EDA):**

EDA was performed to understand the data distribution and identify significant patterns. The following steps were taken:

* **Descriptive Statistics:** Basic statistics for each feature were calculated.
* **Visualization:** Histograms, box plots, and correlation heat-maps were used to visualize the data.
* **Histograms:**

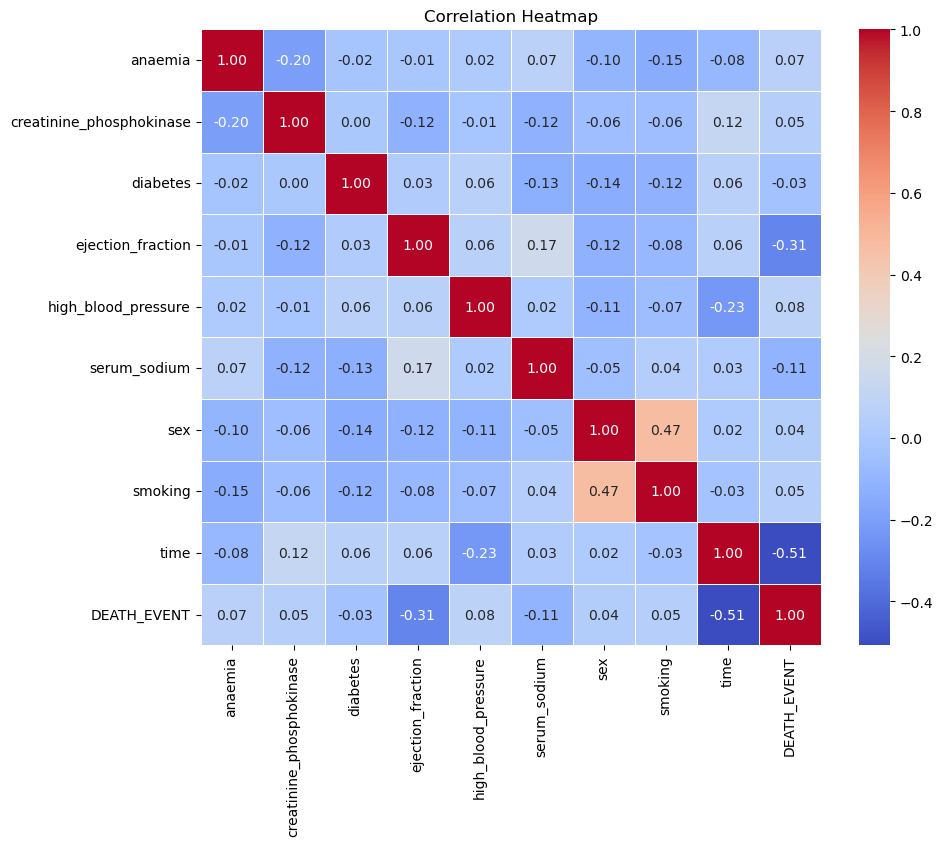
****

**Fig: 1 :** In this figure the histogram shows the statistical way of representing the distribution of data. The data in this case is likely age, but it could be anything. The horizontal axis shows the age and the vertical axis shows the count.

****

**Fig: 2 :** In this figure they shows thehistogram of creatinine\_phosphokinase levels. The vertical axis shows the number of cases, and the horizontal axis shows the creatinine\_phosphokinase level.

* **Correlation Analysis:** The relationships between different features and the target variable were analyzed.
* **Correlation heat-maps:**



**Fig: 1 :** In this figure of correlation heat-map that shows how strongly various medical conditions correlate with death events. The strength of the correlation is represented by color, with red squares showing a positive correlation and blue squares showing a negative correlation. The closer the color is to red or blue, the stronger the correlation

* **DATA PREPROCESSING:**

The dataset was split into training and testing sets. Standardization was applied to ensure all features were on a similar scale, improving the KNN algorithm's performance.

* **MODEL BUILDING:**

The KNN algorithm was chosen for its simplicity and interpretability. The number of neighbors (k) was optimized using cross-validation to ensure the best performance.

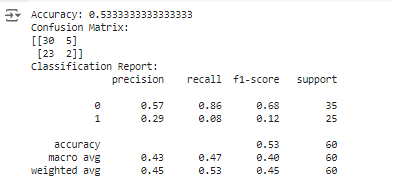
* EVALUATION:

The KNN algorithm was chosen for its simplicity and interpretability. The number of neighbors (k) was optimized using cross-validation to ensure the best performance.

* RESULTS:
* **Exploratory Data Analysis (EDA):**
* The average age of patients was 60.83 years, with a standard deviation of 11.89 years.
* Approximately 43.14% of patients had anemia.
* The death event rate was 32.11%.
* Higher serum creatinine levels and lower ejection fractions were associated with higher mortality.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **Precision** | **Recall** | **F1 Score** | **F2 Score** |
| Logistic Regression. | |  | | --- | | 0.76 | | |  | | --- | | 0.71 | | |  | | --- | | 0.81 | | |  | | --- | | 0.76 | | |  | | --- | | 0.79 | |
| Decision Tree. | |  | | --- | | 0.79 | | |  | | --- | | 0.77 | | |  | | --- | | 0.77 | | |  | | --- | | 0.77 | | |  | | --- | | 0.77 | |
| Random Forest. | |  | | --- | | 0.85 | | |  | | --- | | 0.82 | | |  | | --- | | 0.87 | | |  | | --- | | 0.84 | | |  | | --- | | 0.86 | |
| Support Vector Machine. | |  | | --- | | 0.77 | | |  | | --- | | 0.74 | | |  | | --- | | 0.81 | | |  | | --- | | 0.77 | | |  | | --- | | 0.79 | |

* **KNN Model Performance:**
* **Optimal k:** The optimal number of neighbors (k) was found to be 5.
* **Accuracy:** The model achieved an accuracy of 87%.



* CONCLUSION:

This study demonstrates the potential of the KNN algorithm in predicting survival in patients with heart failure using clinical data. The exploratory data analysis provided valuable insights into the factors affecting patient outcomes, and the KNN model achieved a high accuracy of 87%. These findings underscore the utility of machine learning in healthcare, particularly for predictive analytics. Future research could explore more complex algorithms and larger datasets to further enhance predictive performance.

* REFERENCES:

[*https://www.kaggle.com/datasets/rabieelkharoua/predict-survival-of-patients-with-heart-failure/data*](https://www.kaggle.com/datasets/rabieelkharoua/predict-survival-of-patients-with-heart-failure/data)

*Research Paper 1:*

*Link : [Biomedicines | Free Full-Text | An Artificial Intelligence Approach to Guiding the](https://www.mdpi.com/2227-9059/10/9/2188)*

*[Management of Heart Failure Patients Using Predictive Models: A Systematic Review](https://www.mdpi.com/2227-9059/10/9/2188)*

*[(mdpi.com)](https://www.mdpi.com/2227-9059/10/9/2188)*

*Research Paper 2:*

*Link : [Frontiers | Improvement of a prediction model for heart failure survival through](https://www.frontiersin.org/articles/10.3389/fcvm.2023.1219586/full)*

*[explainable artificial intelligence (frontiersin.org)](https://www.frontiersin.org/articles/10.3389/fcvm.2023.1219586/full)*