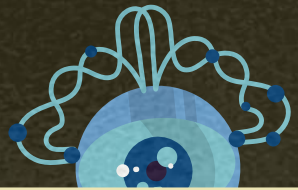
CHRONIC KIDNEY DISEASE DETECTION

GROUP-08

PRESENTED BY
2004173 - TANNU PRIYA
2004190 - KAUSTUBH
2004221 - CHANDRAGUPT VAIBHAV
2004225 - JYOTI RADITYA

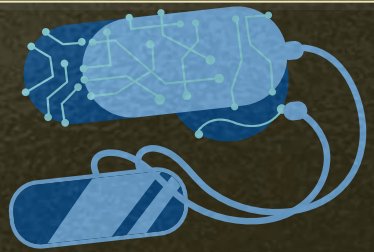
Contents



Problem Statement

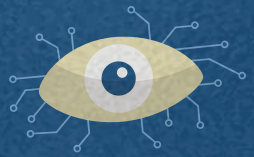
Chronic kidney disease includes conditions that damage our kidneys and decrease their ability to keep us healthy by filtering wastes from our blood. If kidney disease worsens, wastes can build to high levels in your blood and make you feel sick. We may develop complications like:

- High Blood Pressure
- Anaemia (Low Blood Count)
- Weak Bones
- Poor Nutritional Health
- Nerve Damage



INTRODUCTION

The project aims to develop an accurate and fast prediction model for chronic kidney disease (CKD) using machine learning techniques. The model will focus on reducing the number of attributes required for prediction while maintaining high accuracy. The goal is to aid in early detection and management of CKD, potentially improving patient outcomes and reducing the burden on healthcare systems.



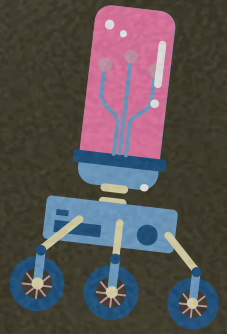
Literature Review

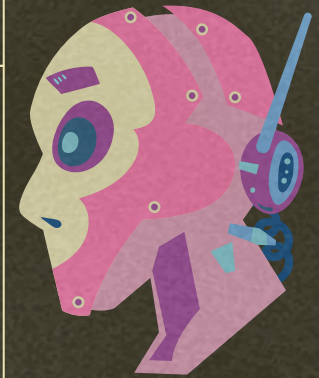
Researcher	Year	Classifier	Accuracy	Remarks
K.R. Lakshmi	2014	ANN	93.8521%	Performed better than Decision Tree and Logical regression classifiers.
Naganna Chetty	2015	Naive Bayes, SMO, IBK	99%, 98.25%, 100%	Attribute Reduction using Wrapper Method.
S.Vijayarani	2015	SVM	76.32%	584 instances and six attributes.
L. Jerlin Rubini	2015	Multilayer Preceptor	99.75%	Performed better than radial basis function network, logistic regression.
Uma N Dulhare	2016	Naive Bayes	97.5%	Attribute Reduction using OneR.
Huseyin Polat	2017	SVM	98.5%	Attribute Reduction
Wala A.	2017	Decision Tree	99%	Missing Values are replaced with mean.

Methodology & Discussion

1. Performing Data-Preprocessing and Prepare your data for analysis & Modelling Purpose as well.

- Import some libraries
- Read data
- Rename Column name
- Knowing data type
- Converting the object data type to numeric
- Differentiating between Categorical column & numerical column
- Finding dirtiness in data (or Unique Categories)
- Replace all misclassified to correct data
- Print data



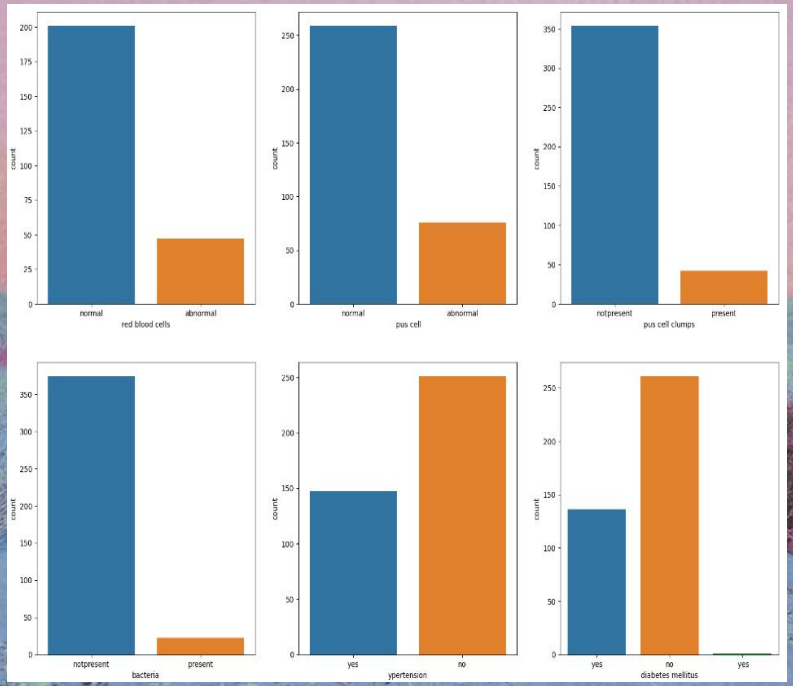


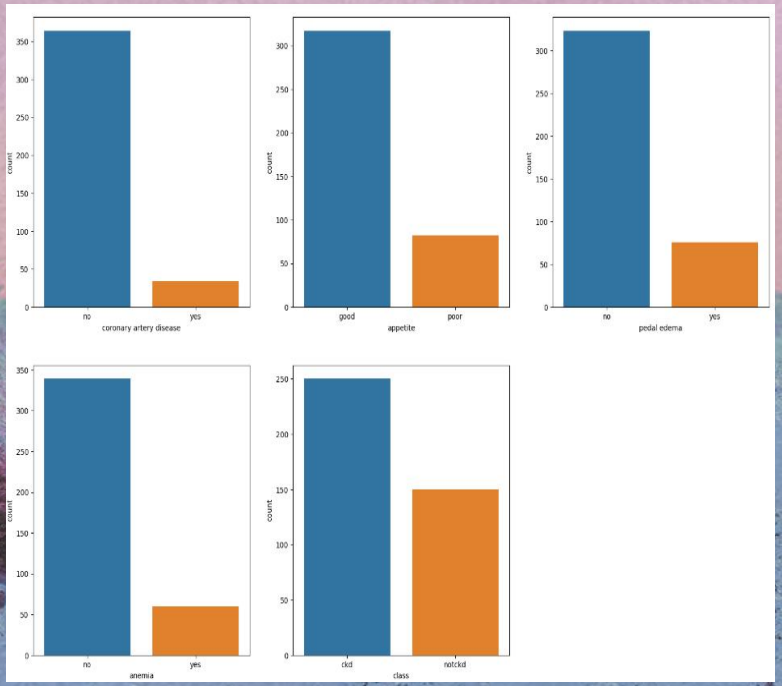
2. Analysing distribution of each & every numerical column.

Distribution of Numeric column in form of subplot

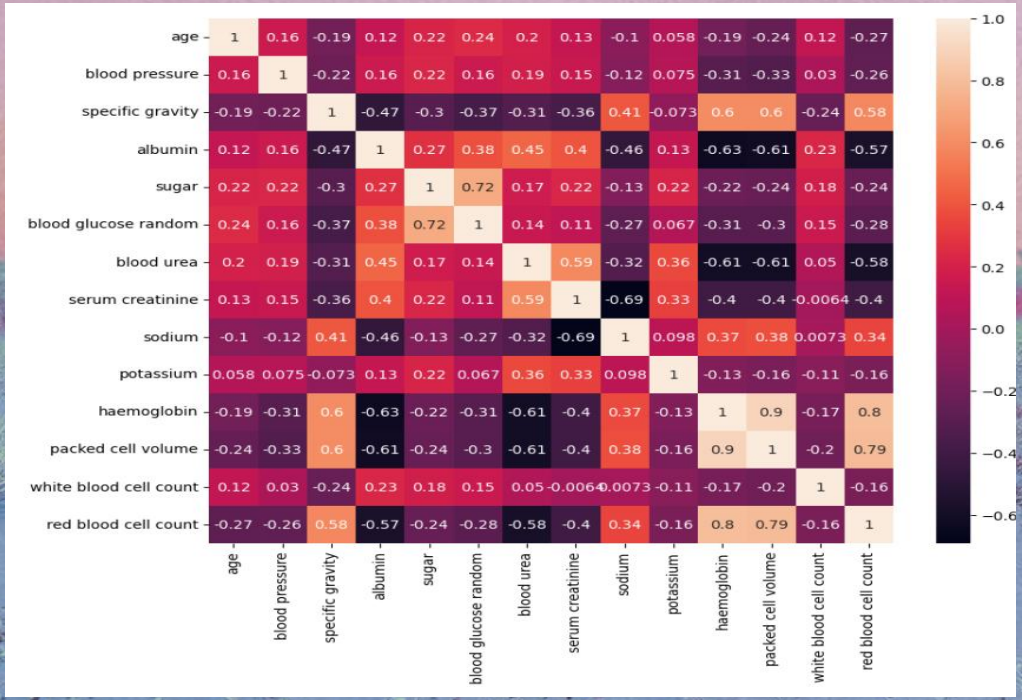
Data Inconsistency and not regular

3. Checking Label Distribution of Categorical data.

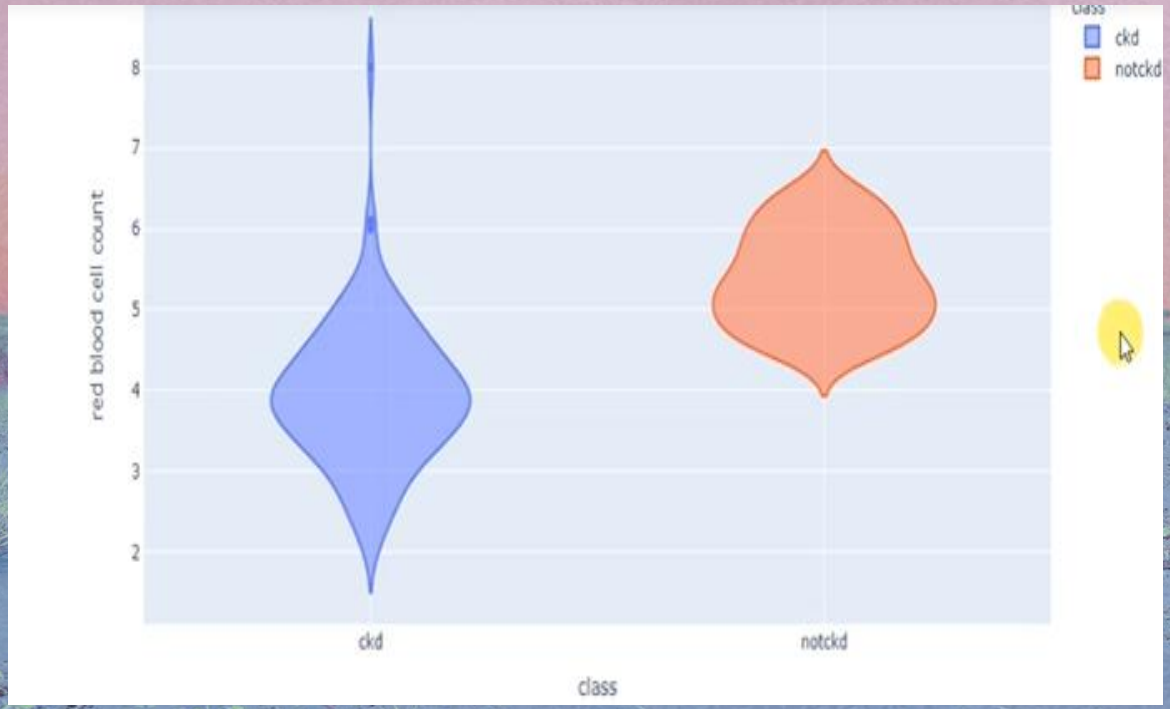




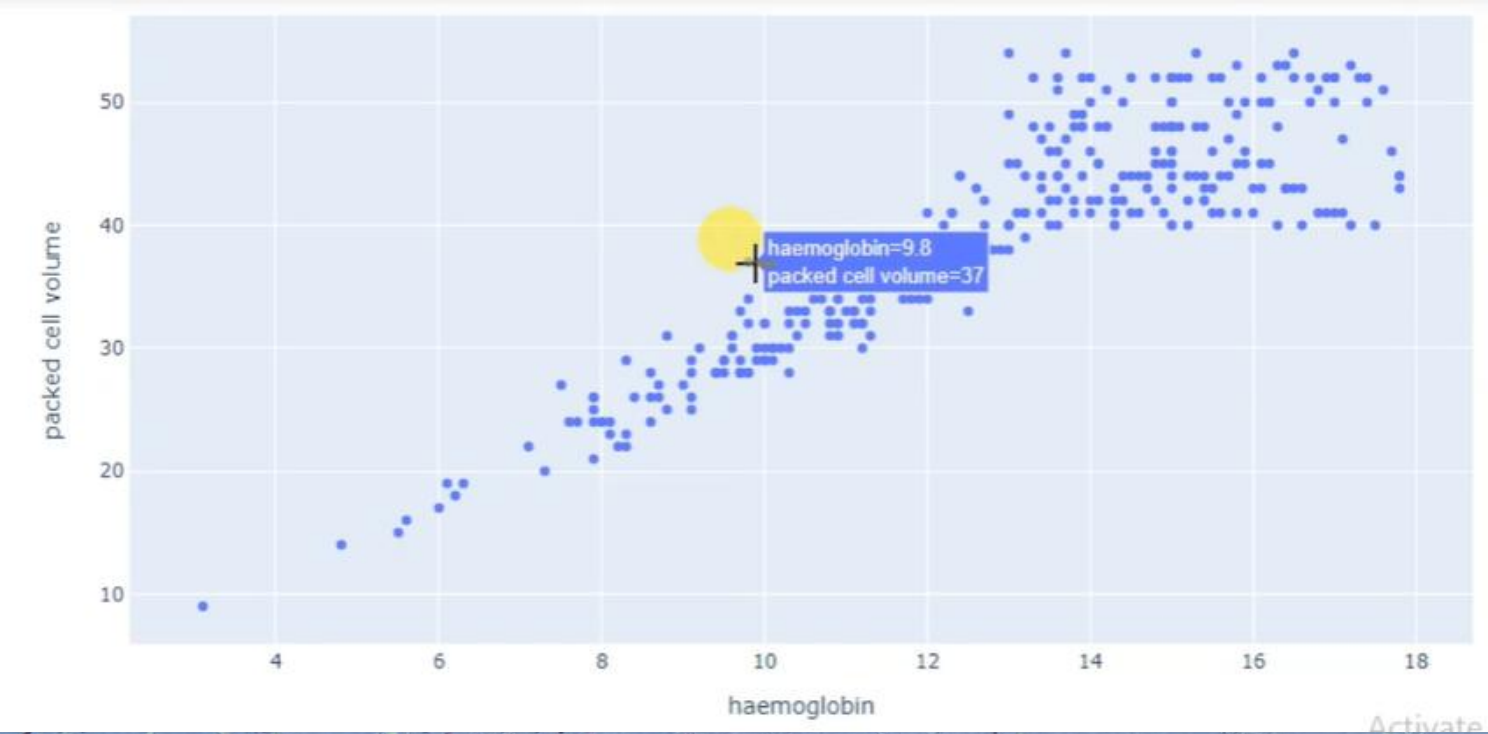
4. Check how Columns are Co-Related with each other & its impact on target feature



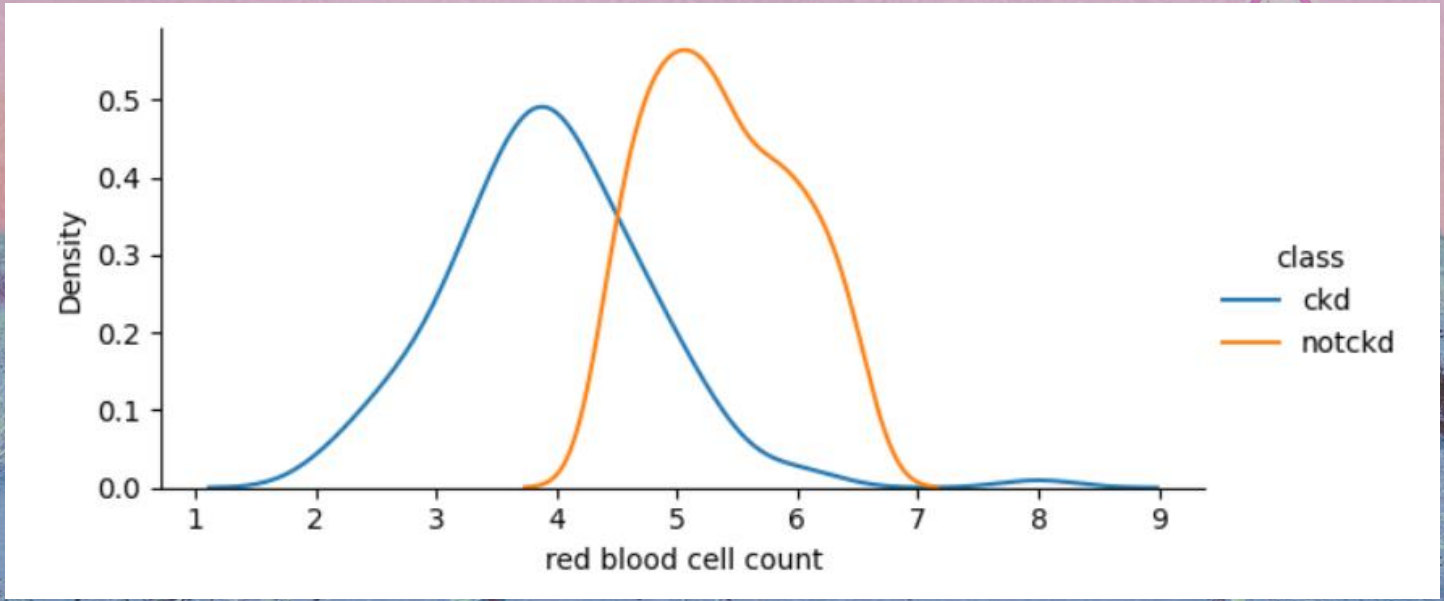
Impact on target feature



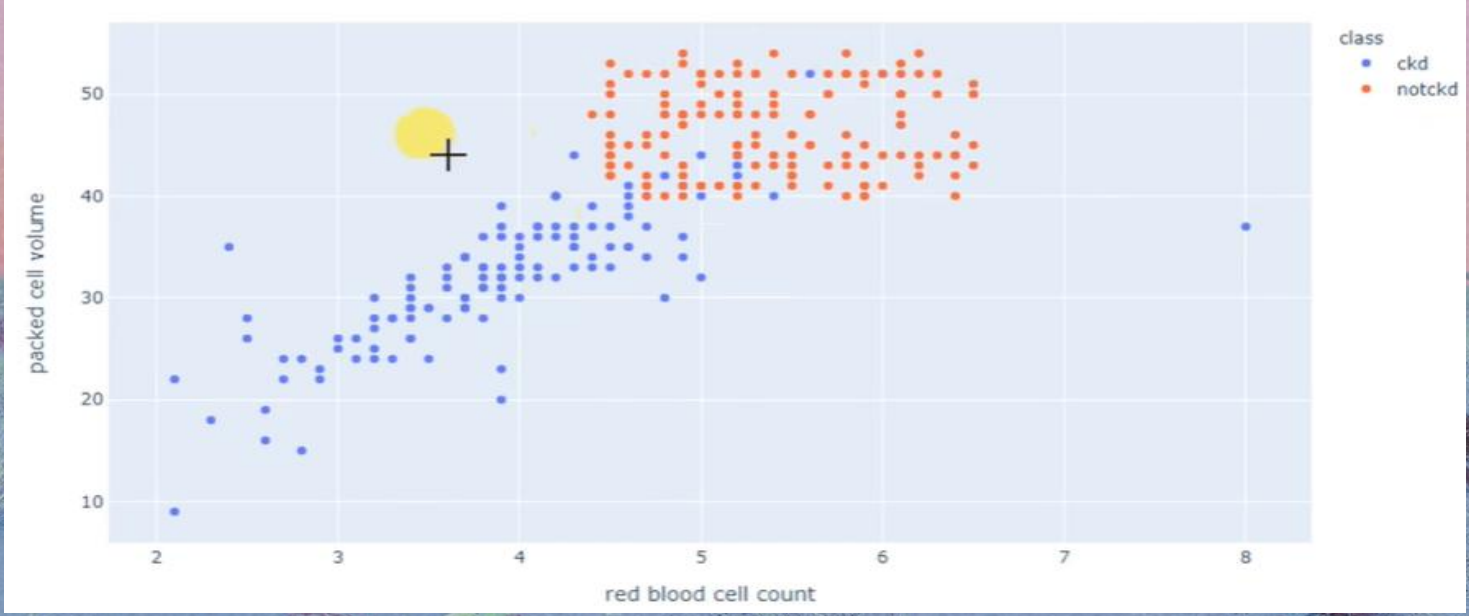
5. Find relationship between haemoglobin & packed cell volume.



6. Analysing distribution of 'red blood cell count' Chronic as well as non-Chronic



7. Performing Exploratory data Analysis on data



According to above graph we can say that if a person is having high haemoglobin level with increasing packed cell volume that person has got high chances of having chronic disease.

8. Perform data cleaning & deal with Missing Values.

- Check for missing Values first
- Fill data randomly in data set

9. Applying Feature encoding technique on data.

- Identifying Categories and go for Label Encoding

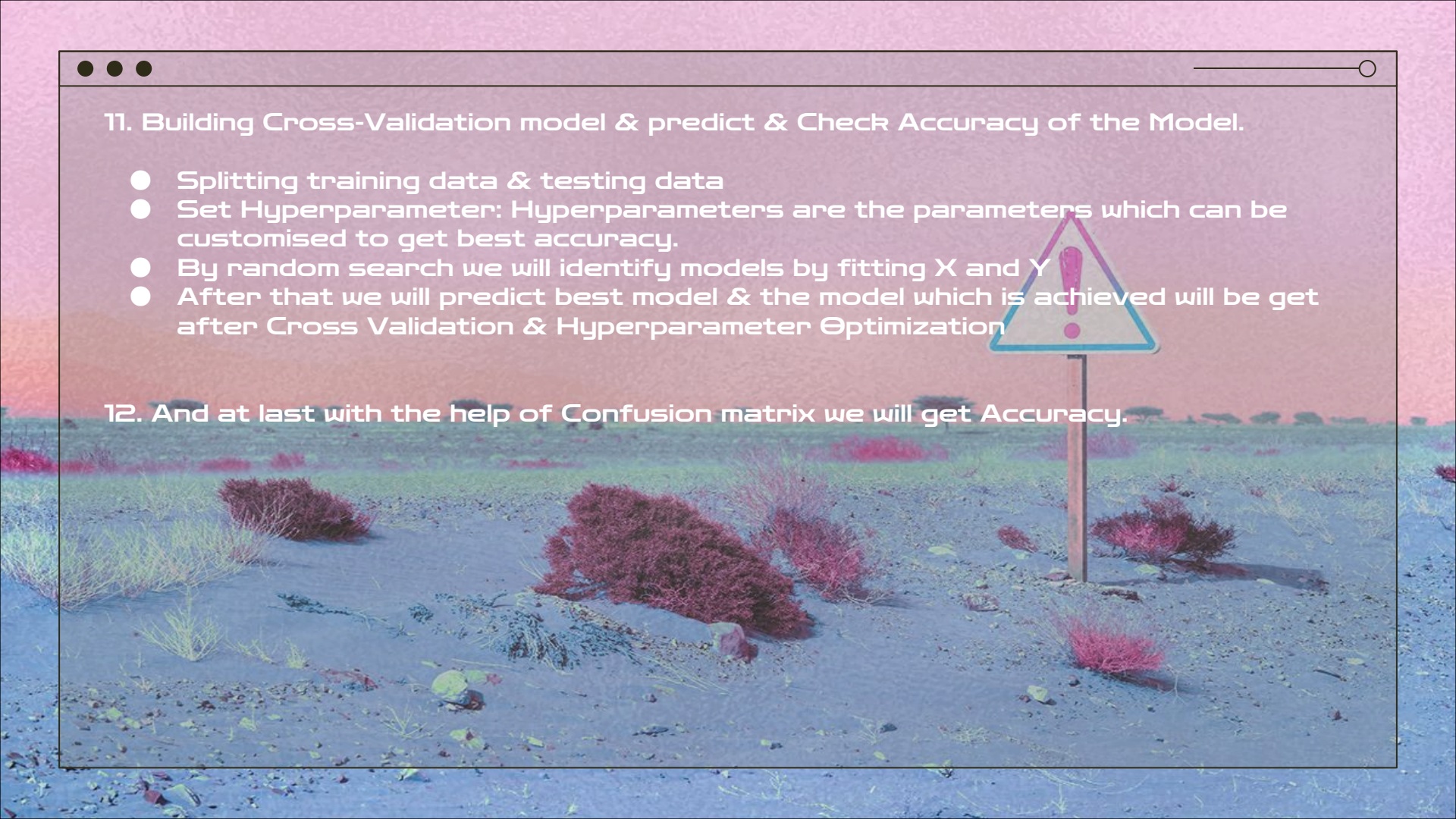
10. Select best features of your model using suitable feature Importance techniques.

- Import Chi2: to check the probability values that are less than 0.5
- We get dependent and undependent feature
- After that we will go for Kbest through which we will get Kbest feature depending upon its probability values
- Then we will extract top 10 data on the basis of score
- And now this 10 data will decide whether the person has CKD or not

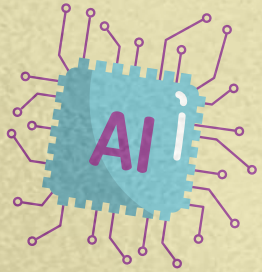
11. Building Cross-Validation model & predict & Check Accuracy of the Model.

- Splitting training data & testing data
- Set Hyperparameter: Hyperparameters are the parameters which can be customised to get best accuracy.
- By random search we will identify models by fitting X and Y
- After that we will predict best model & the model which is achieved will be get after Cross Validation & Hyperparameter Optimization

12. And at last with the help of Confusion matrix we will get Accuracy.



OUTPUT



Confusion Matrix

In [1]:
```
#Checking for accuracy
from sklearn.metrics import confusion_matrix,accuracy_score
```

In [123]:
```
confusion_matrix(y_test,y_pred)
```

Out[123]:
```
array([[59,  3],
       [ 1, 37]], dtype=int64)
```

In [124]:
```
accuracy_score(y_test,y_pred)
```

Out[124]:
```
0.96
```

Accuracy rate is 96%

CONCLUSION

In conclusion, the project aims to use machine learning to accurately detect chronic kidney disease by analyzing patient data. By developing a reliable and accurate predictive model, the project has the potential to contribute significantly to the early identification and intervention of the disease, improving patient outcomes.

REFERENCE

[1] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9874070/

[2] https://link.springer.com/article/10.1007/s44174-022-00027-y

[3]https://www.analyticsvidhya.com/blog/2022/01/predicting-chronic-kidney-disease-using-machine-learning/

[4] https://ieeexplore.ieee.org/document/9185249

[5] https://www.hindawi.com/journals/cmmm/2021/6141470/

[6] https://www.mdpi.com/2504-2289/6/3/98

[7]https://www.researchgate.net/publication/329395701_Prediction_of_Chronic_Kidney_Disease_Using_Machine_Learning_Algorithm

THANK YOU!!!