Chronic Kidney Disease

A Big Zero

Md. Mosfikur Rahman
Akash Ahmed
Mahfuja Ferdousi Mahin

16 April 2021

CSE413: Artificial Intelligence Lab

Afjal H. Sarower
Lecturer
Department of Computer Science & Engineering,
Daffodil International University

Chronic Kidney Disease: An Analytical Prediction Using Machine Learning

Objective

- To predict Chronic Kidney Disease based on each patient's factor report.
- To deal with a wide range of null values.
- To complete data normalization and segmentation for the Test and Training situations.
- To determine the most optimal technique for predicting Chronic Kidney Disease.
- Using the Confusion Matrix, determine which algorithm is the most accurate.

User Benefits

- This dissertation would have the most impact on the medical field.
- Doctors and pathologists would be the most frequent users in this case.
- Using their standard laboratory report study, they will predict a patient's Kidney Disease.
- This project would help save money because the expense of analyzing a Kidney Disease study usually is very high.
- In an improved version of this project, IoT could be used.

Features in Details

By the word, "feature" means a lot in Machine Learning along with Project manners. So, a detailed report on features of this project –

- Algorithms:
  1. Logistic Regression
  2. K-Nearest Neighbor
  3. Random Forest
  4. Gaussian Naïve Bayes
  5. Decision Tree

- The Attributes of Dataset:

Variables	Attribute	Scale	Data Type
age	Age	age in years	Numerical
bp	Blood Pressure	in mm/Hg	Numerical
sg	Specific Gravity	(1.005,1.010,1.015,1.020,1.025)	Nominal
al	Albumin	(0,1,2,3,4,5)	Nominal
su	Sugar	(0,1,2,3,4,5)	Nominal
rbc	Red Blood Cells	(normal,abnormal)	Nominal
pc	Pus Cell	(normal,abnormal)	Nominal
pcc	Pus Cell Clumps	(present,notpresent)	Nominal
ba	Bacteria	(present,notpresent)	Nominal
bgr	Blood Glucose Random	in mgs/dl	Numerical
bu	Blood Urea	in mgs/dl	Numerical
sc	Serum Creatinine	in mgs/dl	Numerical
sod	Sodium	in mEq/L	Numerical
pot	Potassium	in mEq/L	Numerical
hemo	Hemoglobin	in gms	Numerical
pcv	Packed Cell Volume	-	Numerical
wc	White Blood Cell Count	in cells/cumm	Numerical

rc	Red Blood Cell Count	in millions/cmm	Numerical
htn	Hypertension	(yes,no)	Nominal
dm	Diabetes Mellitus	(yes,no)	Nominal
cad	Coronary Artery Disease	(yes,no)	Nominal
appet	Appetite	(good,poor)	Nominal
pe	Pedal Edema	(yes,no)	Nominal
ane	Anemia	(yes,no)	Nominal
classification	Class	(ckd,notckd)	Nominal

- Standard Train-Test Splitting
  We use 75% of Dataset randomly to train the model and 25% to test our approaches. In numerical formation, 300 rows of Dataset were used to prepare the models, and 100 rows were used to test the algorithms.

Development Tools and Technology

We have used software and modules to build this project and packages to make a specified outcome. The table below contains each of these and versions.

Names	Version
Python	3.8.8
Numpy	1.19.2
Pandas	1.2.3
Matplotlib	3.3.4
Cython	0.29.22
Number of CPUs	6
IPython	7.22.0
SciPy	1.6.2
Scikit Learn	0.24.1
Scikit Plot	0.3.7

Table 1: Development Tools and Technology

Team Members Contribution Chart

This project is entirely implemented by team "A Big Zero." The contribution Chart for this project among all members are given below –

Name	Contributions
Md. Mosfikur Rahman	Complete Implementation
Akash Ahmed	Data Collection & Data Preprocessing
Mahfuja Ferdousi Mahin	Crosscheck all theoretical approaches.

Table 2: Contribution Chart

Project Timeline

Since we have been working with Machine Learning for quite a while, the Dataset came from UCI Machine Learning Repository, and the Dataset was processed using panda's data frame. The process we have covered the general population, high-risk persons, and all stages of Chronic Kidney Disease patients with comprehensive strategies such as early screening, practical Chronic Kidney Disease registry, prevention, and Chronic Kidney Disease comprehensive care models including cost-effectiveness analysis. Five best-suited algorithms have been used to train the Dataset and predict an accurate value for creating the best model.

Screenshot of Interface

We worked in Jupyter Notebook and Jupyter Lab. In this environment, the whole work can be done quickly, and the short note can be processed very efficiently. The entire notebook can be converted into a pdf file or any HTML file too. This environment is very much helpful in developing a Machine Learning Project.

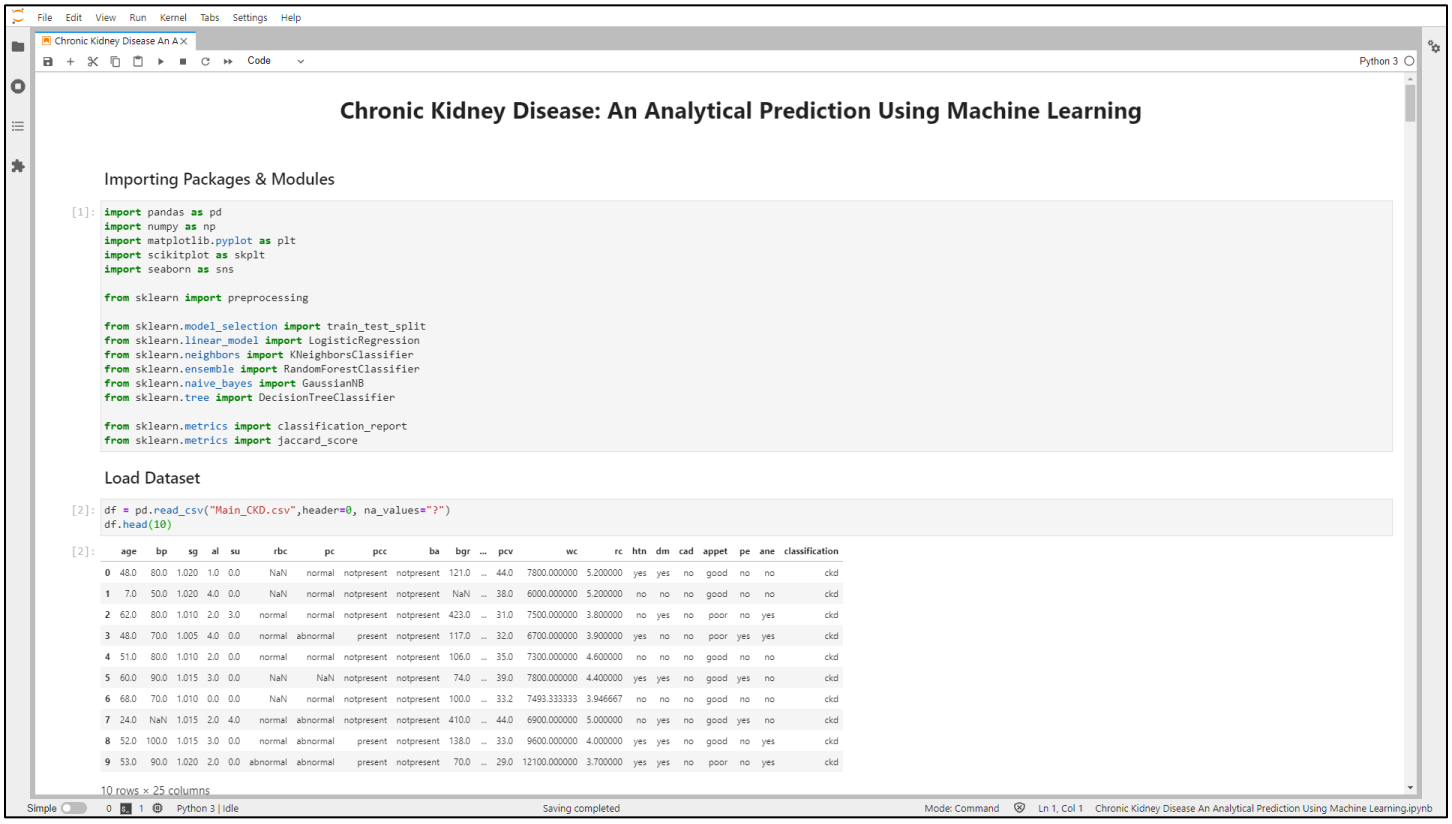

Figure 1: Screenshot of Starting Project

Test Results and Output Screenshot

We have predicted Chronic Kidney Disease with five practical Algorithms. And all of them come up with a Data frame to show output.

- Accuracy Table

Algorithm Name	Accuracy Score (%)	Jaccard Score (%)	Misclassification (%)
Logistic Regression	100	100	0
KNN	100	100	0
Random Forest	100	100	0
Naive Bayes	97	94.64	3
Decision Tree	95	91.38	5

Table 3: Accuracy Table

Source Code

The entire programming file is much larger, and I used hyperlink here. Click Here to access the program file.

References

[1] Ahmed, I., Khan, M. S., Paidi, S., Liu, Z., Zhang, C., Liu, Y., & Lau, C. (2021). Laser-induced breakdown spectroscopy with machine learning reveals lithium-induced electrolyte imbalance in the kidneys—Journal of Pharmaceutical and Biomedical Analysis, 194, 113805.

[2] Chaudhuri, A. K., Sinha, D., Banerjee, D. K., & Das, A. (2021). A novel enhanced decision tree model for detecting chronic kidney disease. Network Modeling Analysis in Health Informatics and Bioinformatics, 10(1), 1-22.

[3] Balusamy, B. (2021). Utilizing schemes for detecting the threat levels for chronic kidney disease. Materials Today: Proceedings.

[4] Dubey, G., Srivastava, Y., Verma, A., & Rai, S. (2021). Chronic Kidney Disease Prediction Using Artificial Neural Network. In Proceedings of International Conference on Big Data, Machine Learning and their Applications (pp. 395-401). Springer, Singapore.

[5] Pourhomayoun, M., & Shakibi, M. (2021). Predicting mortality risk in patients with COVID-19 using machine learning to help medical decision-making. Smart Health, 20, 100178

Conclusion

In terms of data imputation and sample diagnosis, the proposed CKD diagnostic approach is feasible. The number of data samples available is relatively limited, with just 400 pieces available. Furthermore, since the data collection only contains two types of data samples (ckd and notckd), the model cannot diagnose the seriousness of CKD. A considerable amount of more complex and representative data will be obtained in the future to train the model and increase generalization efficiency.