KaggleX- Showcase

Chronic Kidney Disease Prediction Using K-nearest Classifier

Tsagli Klenam Martha

Presentation Overview

Background (A bit about me)

Project Objective & Description

Brief History on ckd

Model Building & Deployment

Background

01 ➤ Aspiring Data Scientist

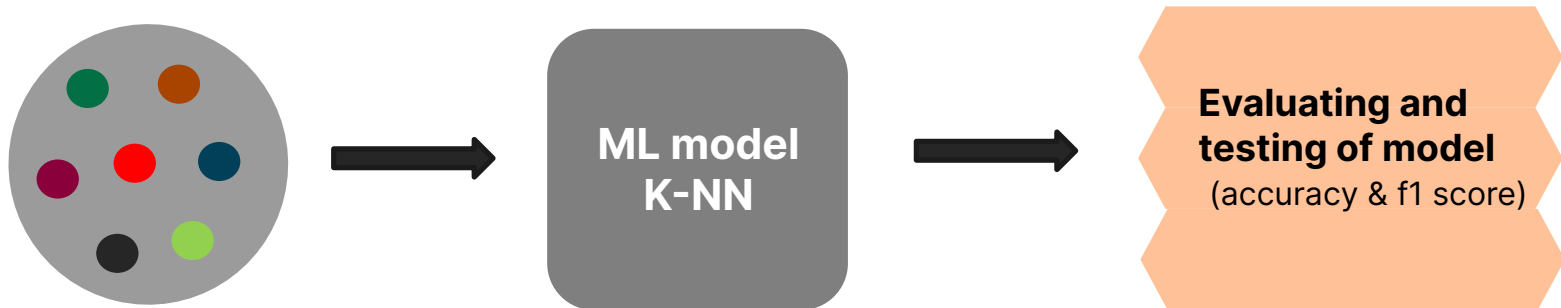**02** ➤ B.A French and Political Science

03 ➤ Love daring, singing and coding

Project Objective & Description

- Project is aimed at building a machine learning model to aid doctors in diagnosing Chronic Kidney Disease in patients.

known data (training set)

- The machine learning model will be trained on the known data (thus recognizing patterns in the labeled data)

Brief history on Chronic Kidney Disease

Inspiration

☐ Chronic Kidney Disease is a **long term** medical condition in which **kidneys fail** to filter **excess fluid** and waste from the body. Diagnosing patients with CKD will aid doctors to offer priority and rapid treatment to affected patients.
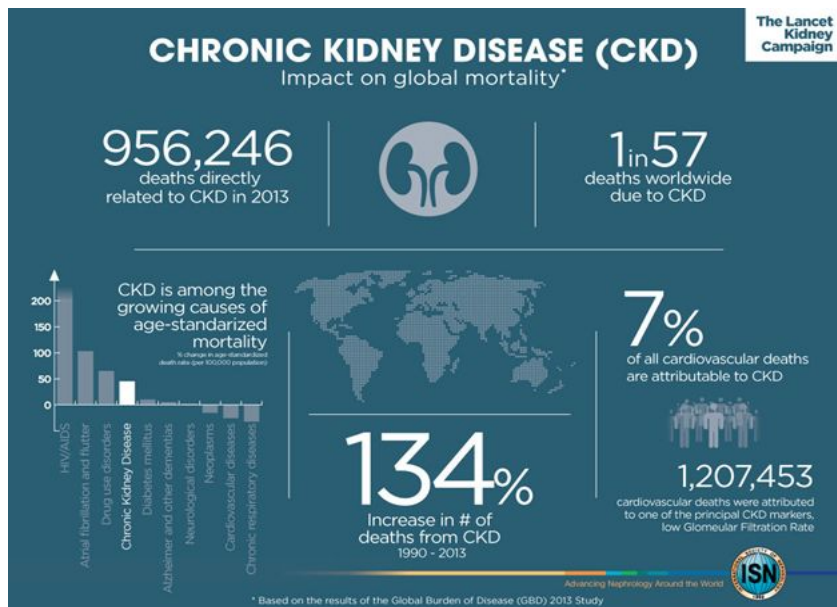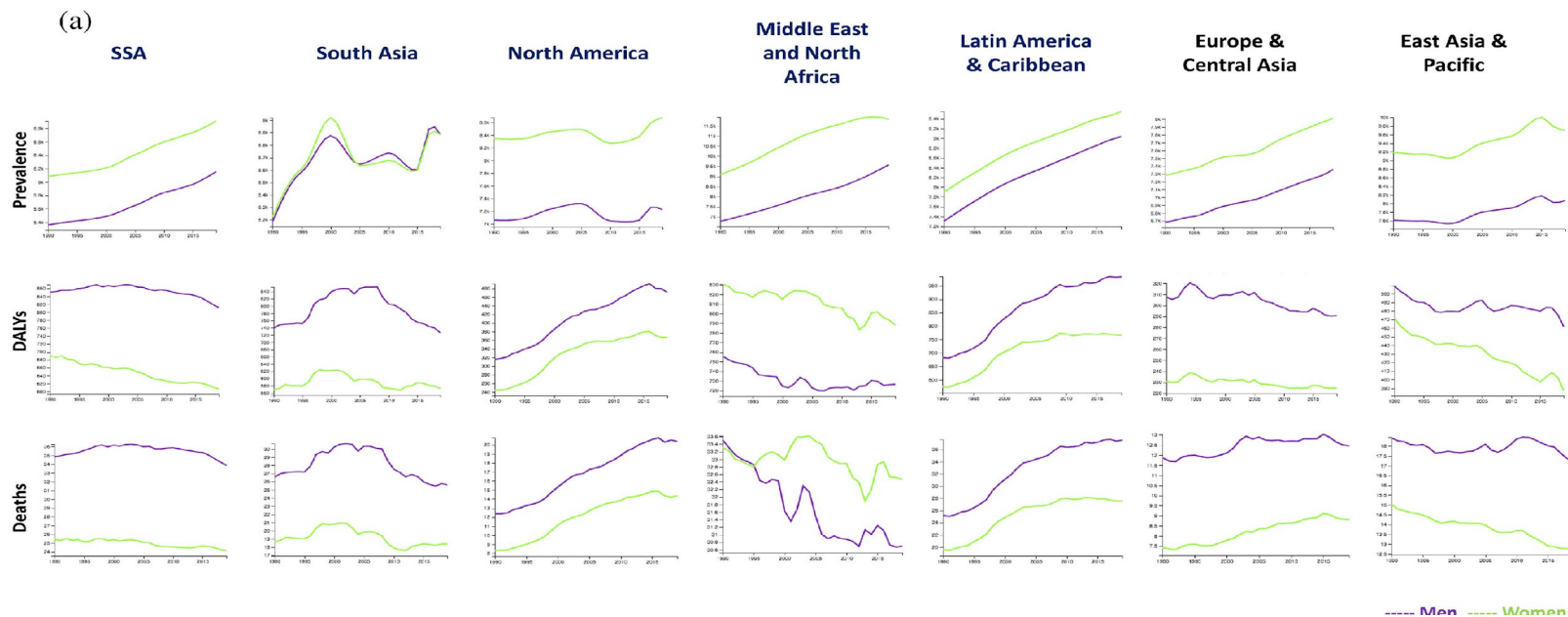

Lancet Kidney Campaign report

☐ Research shows that **10%** of the entire population world wide are affected by Chronic Kidney disease

☐ More **women** are prone to having Chronic Kidney Disease than men

☐ **Glomerular Filtration Rate** (GFR) measures the function of the kidney by **assessing** the flow rate of filtered fluid through the kidney. Usually measured by substances present in **the blood** and **urine**

Gender risk of ckd prevalence (a 2015 line series report)

- More men die of Chronic Kidney Disease (CKD) than women as men reach kidney failure sooner than women

Men ▪ Women ▪

(a)

SSA | South Asia | North America | Middle East and North Africa | Latin America & Caribbean | Europe & Central Asia | East Asia & Pacific

Prevalence

DALYs

Deaths

----- Men ----- Women

Preparatory & Analysis phase

Data Source

- Data set was gotten from UCI machine learning website
  See here Datasets - UCI Machine Learning Repository

- Other dataset of similar kind can be found on the Kaggle website
  See here Chronic Kidney Disease Prediction (98% Accuracy) | Kaggle

- Data was pre-processed and deployed before use (to ensure data quality and reliable model)

- Dataset has **158** individual observations and **25** features with **115** recording "no ckd" and **43** recorded to "have ckd"

kaggle

Learning the attributes of Known data

o An example of a code snippet to view the attributes in your dataset

dset.head()

	Age	Blood Pressure	Specific Gravity	Albumin	Sugar	Red Blood Cells	Pus Cell	Pus Cell clumps	Bacteria	Blood Glucose Random	...
0	48	70	1.005	4	0	normal	abnormal	present	notpresent	117	...
1	53	90	1.020	2	0	abnormal	abnormal	present	notpresent	70	...
2	63	70	1.010	3	0	abnormal	abnormal	present	notpresent	380	...
3	68	80	1.010	3	2	normal	abnormal	present	present	157	...
4	61	80	1.015	2	0	abnormal	abnormal	notpresent	notpresent	173	...

5 rows × 25 columns

o Where dset is the name of our data frame

kaggle

Predicting CKD in patients; what do we consider?

1. **Conducting blood tests:** Blood test is conducted to evaluate how well an individual's kidneys are filtering blood. The Glomerular Filtration Rate(GFR) is measured in this process

2. **Albumin test:** Tests are conducted from urine samples to check for the presence of albumin in urine.
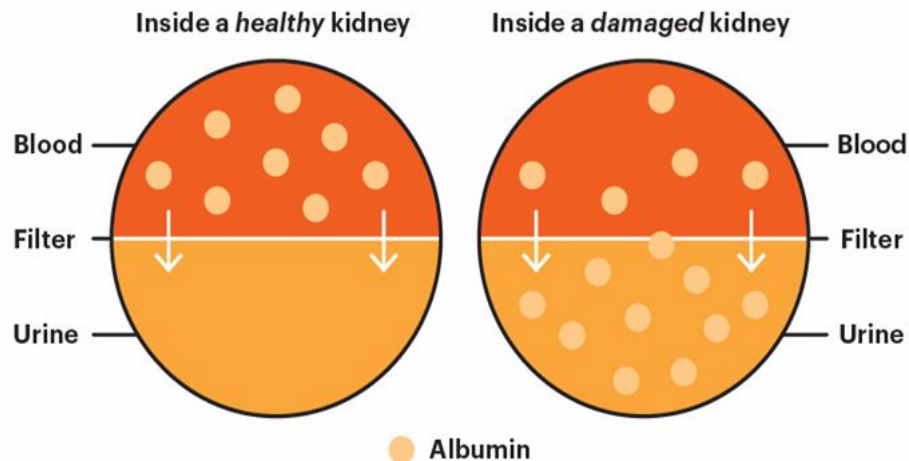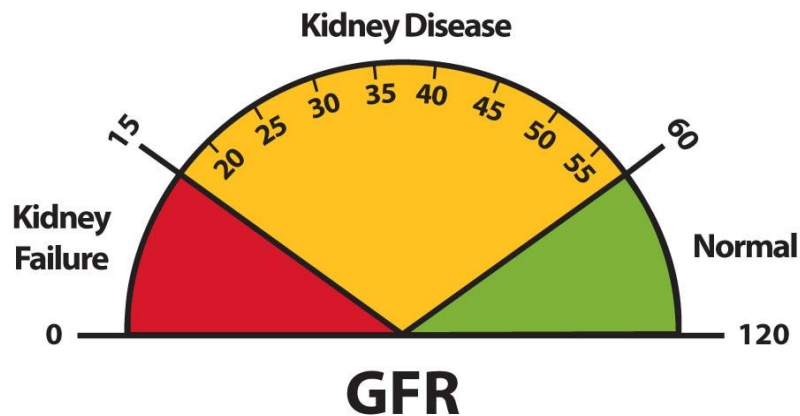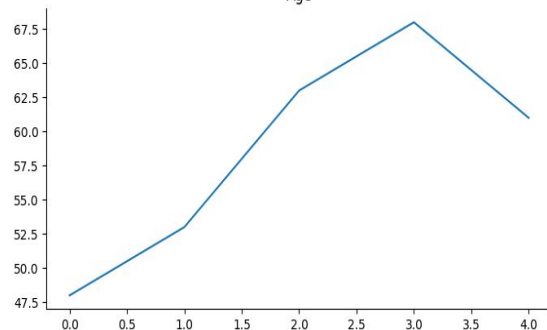
Performing EDA for CKD Analysis

1. Scatterplot of some x and y co-ordinates
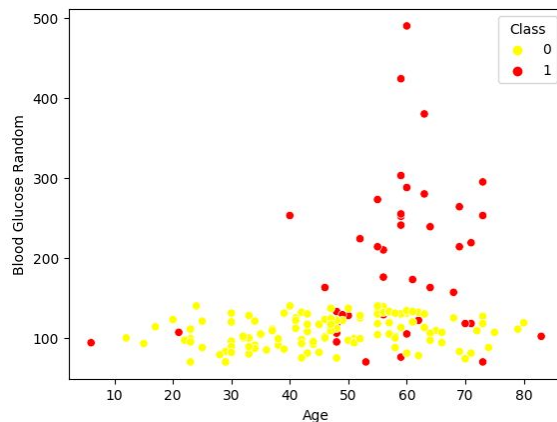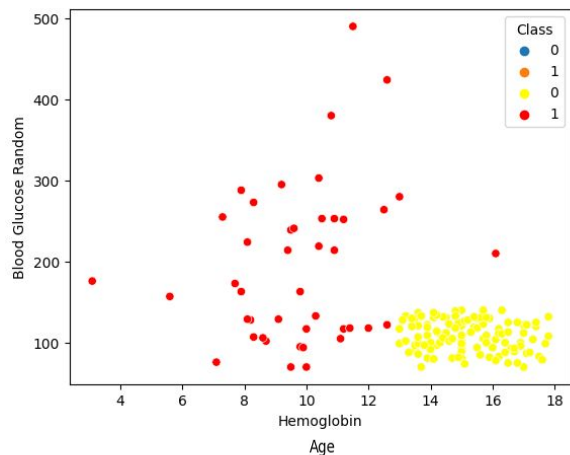
Age	Class
73	1
73	1
73	1
71	1
71	1
70	1
69	1
68	1
63	1
60	1

Sample of Age range having ckd

Research links

- Find links to relevant research materials on the subject of chronic kidney disease

1. https://www.niddk.nih.gov/health-information/kidney-disease

2. https://drive.google.com/file/d/1xZLDPBbAsRHRnIEH_N8Z9tE4DrZyEb1E/view?usp=sharing

3. https://my.clevelandclinic.org/health/diagnostics/21659-kidney-function-tests

4. https://www.thelancet.com/commissions/0by25-acute-kidney-injury

Model building and deployment for ckd prediction

Testing and splitting data

- Introducing ML model to training set

k-nn classifier

Labeled data

CLASS A
CLASS B

K= 3

Dataset

Train split | Test split

Training | Validation | Test

Model Training | Model evaluation | Prediction

Sample model building process
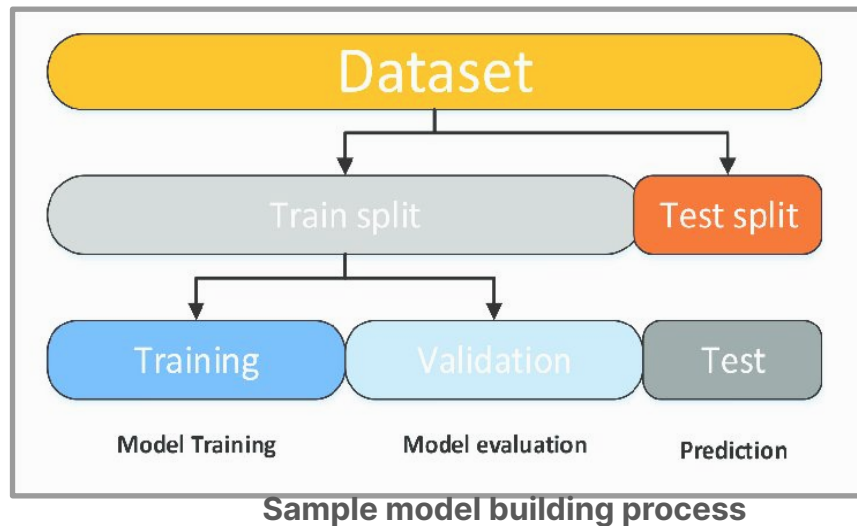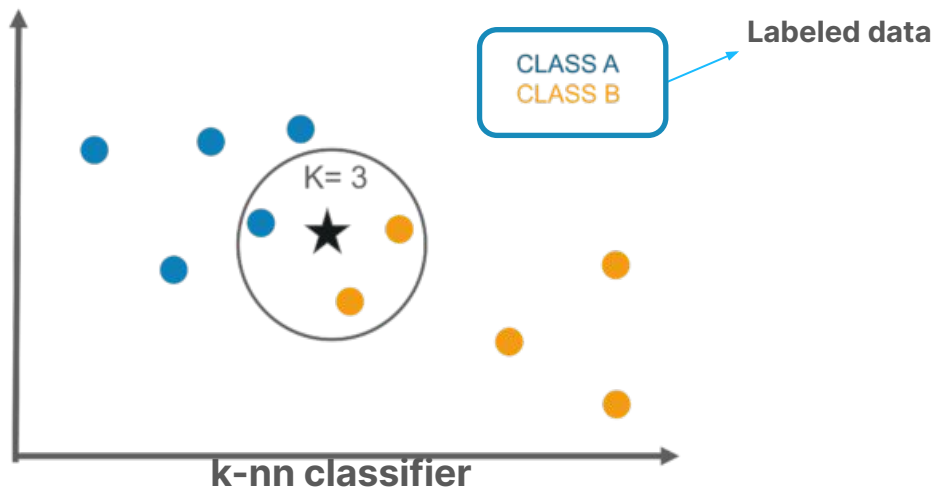
- k-nearest neighbour classier is trained with the labeled data (where it learns attributes of various individual observations) to make inferences on new data points

Performance Metrics

Model metrics is used to evaluate the performance of ml model

- **Accuracy**

Accuracy score for model resulted in **85%** correctly classified labels and **15%** incorrectly classified samples. Meaning, with k-value of 3, model correctly predicted new data points to their classes

- **F1 score**

F1 score achieved a result of **0.625** on testing set and **0.69** on training set. The f1 score proved that the k-nn classifier made quiet effective prediction of new data points to it correct class. This also means that between precision and recall, model achieved a balance.

Note: The k-nearest neighbor algorithm was considered in this analysis because of the moderate size of the dataset as k-nn will be suitable for the dataset

Model deployment

Due to the sensitive nature of our analysis, ckd model could be deployed by;

○ **Building a user-friendly app:** To enable doctors, health practioners etc to input data of patients and get chronic kidney disease predictions

○ **Continuous monitoring and maintenance of model's performance in real time:** Regularly updating model and training model with new data to keep it effective and accurate

kaggle

Project Details

❖ **Medium link**

https://medium.com/@klenamtsagli/classification-machine-learning-a-brilliant-approach-in-health-care-diagnosis-0993fb50e1e6

❖ **GitHub link**

https://github.com/martfem13/KaggleX_Data-Science_Project/blob/main/KaggleX_Project_on_Chronic_Kidney_Disease.ipynb

Category	GFR	Terms	Clinical Presentations
G1	≥ 90	Normal or high	Markers of kidney damage (nephrotic syndrome, nephritic syndrome, tubular syndromes, urinary tract symptoms, asymptomatic urinalysis abnormalities, asymptomatic radiologic abnormalities, hypertension due to kidney disease)
G2	60-89	Mildly decreased*	
G3a	45-59	Mildly to moderately decreased	• Mild to severe complications: ○ Anemia ○ Mineral and bone disorder ▪ Elevated parathyroid hormone ○ Cardiovascular disease
G3b	30-44	Moderately to severely decreased	▪ Hypertension
G4	15-29	Severely decreased	▪ Lipid abnormalities ▪ Low serum albumin
G5	< 15	Kidney failure	• Includes all of the above in addition • Uremia

GFR categories in CKD

GFR = mL/min/1.73 m²
*Relative to young adult level
In the absence of evidence of kidney damage, neither GFR category G1 nor G2 fulfill the criteria for CKD.
Refer to a nephrologist and prepare for kidney replacement therapy when GFR <30 mL/min/1.73m².

Glomerular Filtration Rate test range

❖ **Read article on**

"Biological variation of measured and estimated glomerular filtration rate in patients with chronic kidney disease"

kaggle

Connect with Me

-LinkedIn

www.linkedin.com/in/martha-klenam-tsagli

-GitHub

https://github.com/martfem13

-Medium

https://medium.com/@klenamtsagli

-Email me on

klenamtsagli@gmail.com