Project Report Template

1 INTRODUCTION

1.1 Overview

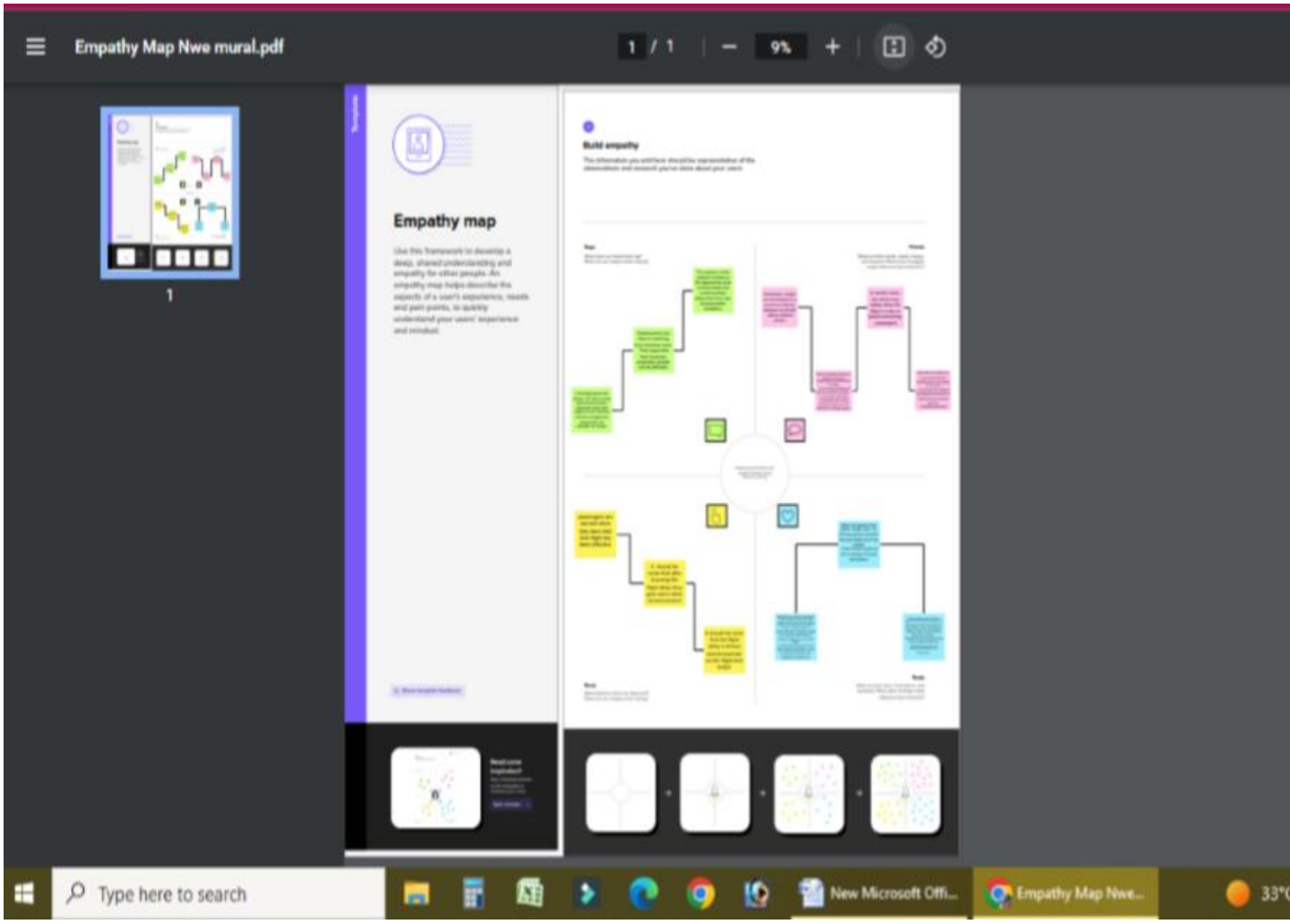
One of the non-communicable diseases with the quickest growth rate is chronic kidney disease (CKD), a significant cause of death and disease. It has affected more than 10% of the world's population, and millions of people die each year [1]. According to the Global Burden of Disease Study, almost 697.5 million cases of all-stage CKD were registered in 2017, resulting in a global prevalence of 9.1%, up 29.3% from 1990. Meanwhile, between 1990 and 2017, the global all-age death rate from CKD grew by 41.5%]. It is predicted that 1 out of every 10 persons has some symptom of kidney disease. It is called a "chronic" disease because kidney disease affects the functioning of the urinary system and develops gradually over time. This illness indirectly impacts global morbidity and mortality rates by increasing the risks of the other main killers (cardiovascular disease, diabetes, hypertension, HIV infection, and malaria). Chronic kidney disease progression is a critical aspect of a person's health. As a result, maintaining excellent kidney function is critical for general health .
Chronic kidney disease treatment is both expensive and ineffective. In contrast, only about 5% of individuals with early CKD are aware of their condition. Glomerular damage has reached over 50% and is usually irreversible once CKD is identified. In this regard, accurate chronic renal disease prognosis can be highly beneficial. As a result, multiple CKD prediction models for various populations have been developed .
In this paper, we proposed three optimized deep learning algorithms, CNN, ANN, and LSTM, to predict CKD at the primary stage. The goal of this classification is to determine whether or not a patient is susceptible to CKD. Multiple layers with linear or nonlinear activation functions are assembled to form the model. These layers are taught to work together to master a complex problem-solving approach .

1.2 Related Work

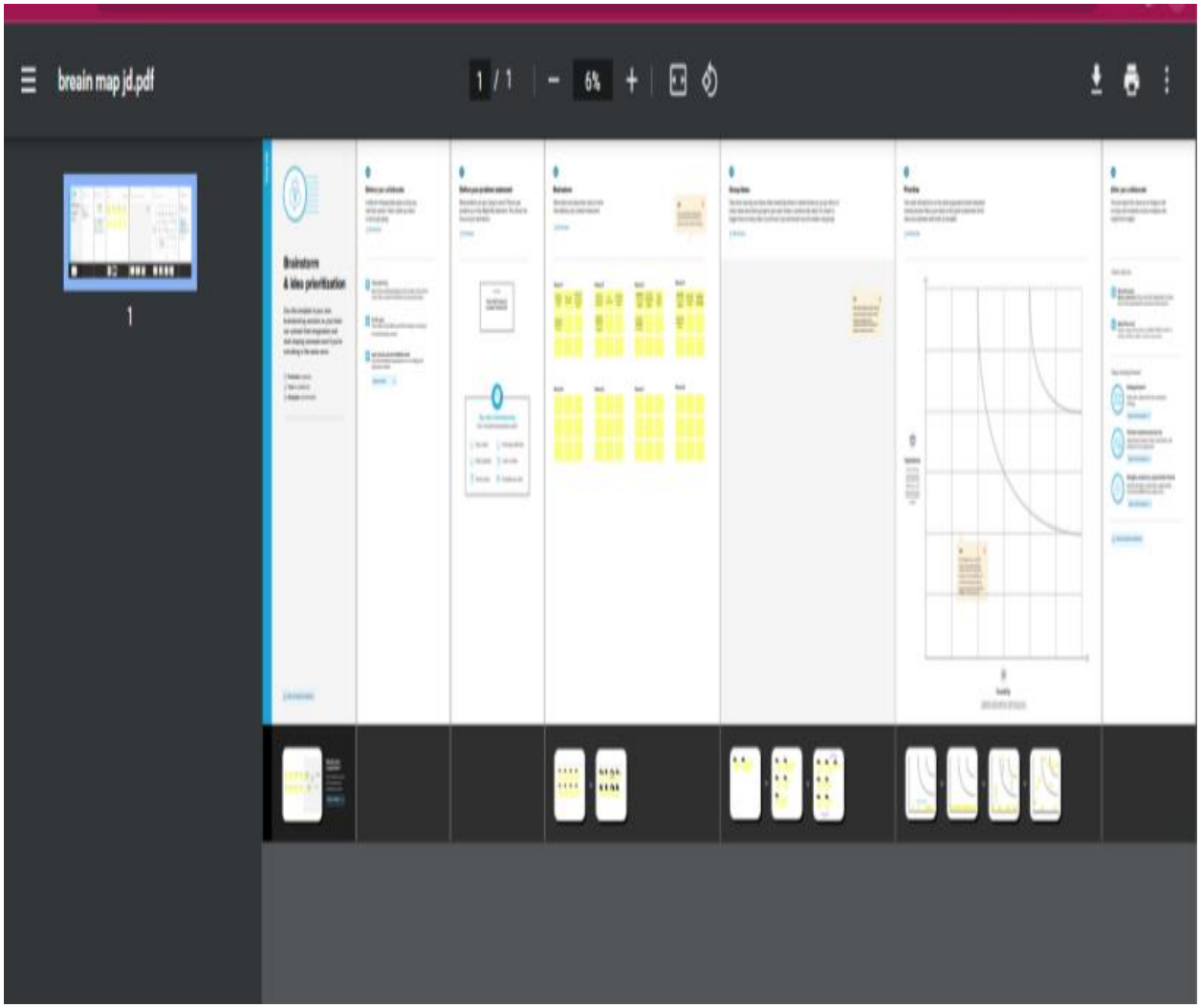
In 2021, Chotimah, et al. [6] introduced a strategy that uses deep learning to identify diagnoses in medical records that contain the most important information about chronic kidney disease. In this study, the first stage is to use the sequential backward feature selection (SBFS) to select a diagnosis that will have a significant impact. Features that do not have as much influence as other features are removed using this algorithm. The top diagnostic features are fed into the artificial neural network (ANN) classification algorithm in the second step. On their study, 18 features are used as input, where 15 features are considered the best features by the system. When 15 diagnostic features were used as inputs, the accuracy of the ANN classification system reached 88%. These results were improved compared to the overall diagnostic features, which have an accuracy of 80%. Alsuhibany, et al. [7] created an IoT-based ensemble of deep-learning-based clinical decision support systems (EDL-CDSS) to diagnose CKD. The adaptive synthetic (ADASYN) technique for the outlier detection process is designed using the EDL-CDSS technique. A convolutional neural network with a gated recurrent unit (CNN-GRU), deep belief network (DBN), and kernel extreme learning machine (KELM) are proposed. They achieved the highest accuracy of 96.91% using the EDL-CDSS approach. Akter, et al. [8], in 2021, deployed seven state-of-the-art deep learning algorithms, ANN, LSTM, GRU, bidirectional LSTM, bidirectional GRU, MLP, and simple RNN, for CKD prediction and classification along with the numerous clinical features of CKD that have been proposed. Their investigation calculated the loss and validation loss in prediction and looked at recall, accuracy, and precision. The ANN, simple RNN, and MLP algorithms performed well in the CKD classification (99%, 96%, and 97% accuracy, respectively). In 2020, Iliyas, et al. [9], using a dataset of 400 patients from Bade General Hospital and 11 features or parameters, used a deep neural network (DNN) to predict chronic kidney disease (CKD). During the preprocessing of the data, several missing cells were simply imputed using the attribute's mean value. They employed the deep neural network (DNN) model to predict if CKD would be present in a patient. The DNN model generated a 98% accuracy rate. Of the 11 variables, creatinine and bicarbonate impact CKD prediction most. In 2020, Ma, et al. [10] suggested chronic kidney illness utilizing a heterogeneous modified artificial neural network based on deep learning. On the Internet of Medical Things (IoMT) platform, they used a heterogeneous modified artificial neural network (HMANN) for the early detection, segmentation, and diagnosis of chronic renal failure. To identify the location of kidney stones, HMANN assists in

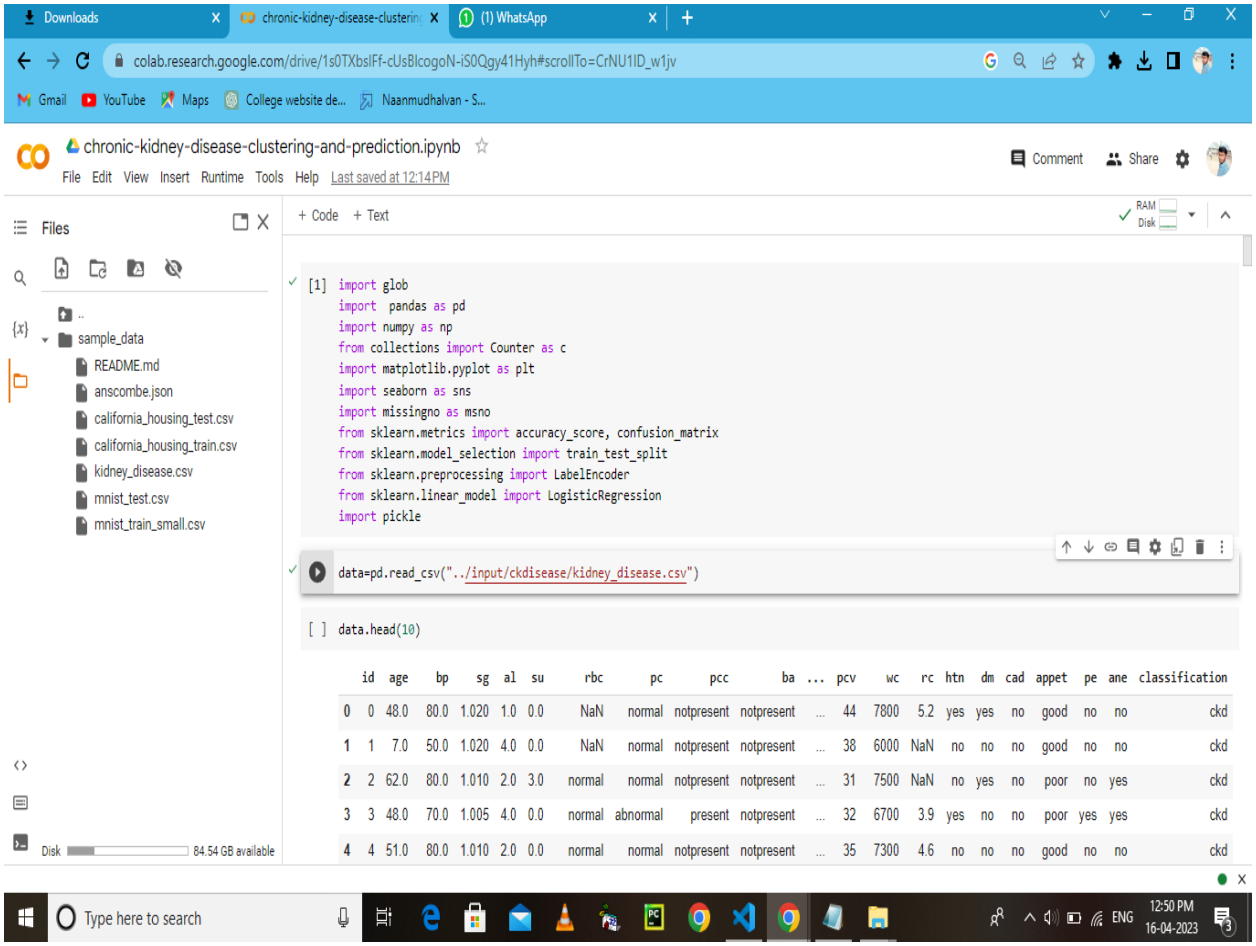
reducing noise and segmenting the kidney image. The accuracy rates for ANN-SVM and HMANN were 92.3% and 97.5%, respectively. A bidirectional long short-term memory (LSTM) network and a one-dimensional correlational neural network (1-D CorrNN) are combined in the deep learning model presented by Bhaskar, et al. [11], in 2020. The suggested model is trained and evaluated using the CKD-sensing module. The accuracy of the suggested model, 1-D CorrNN-LSTM, is 98.08%, and it performed better than other models. In 2019, N. Bhaskar, et al. [12] proposed an effective and novel method for non-invasive chronic renal disease diagnosis. The proposed task involves designing and building a novel sensing module for CKD diagnostics. They used a hybrid deep learning convolution neural network–support vector machine (CNN-SVM) model to make predictions. The proposed model is put to the test in experiments, and its performance is compared to that of a traditional CNN. The accuracy of the suggested hybrid deep learning model (CNN-SVM) was 96.59%. In 2019, Almansour, et al. [13] used classification techniques including a artificial neural network (ANN) and a support vector machine (SVM). Using the mean of the corresponding attributes, they replaced all missing values in the datasets. Additionally, they employed a 10-fold cross-validation procedure to divide the training and test datasets according to the ratio (90:10). In their proposed method, ANN performs better. Using the optimized features, the accuracy is 99.75%, while, from SVM, the accuracy is 97.75%

2.problem definition & design thinking

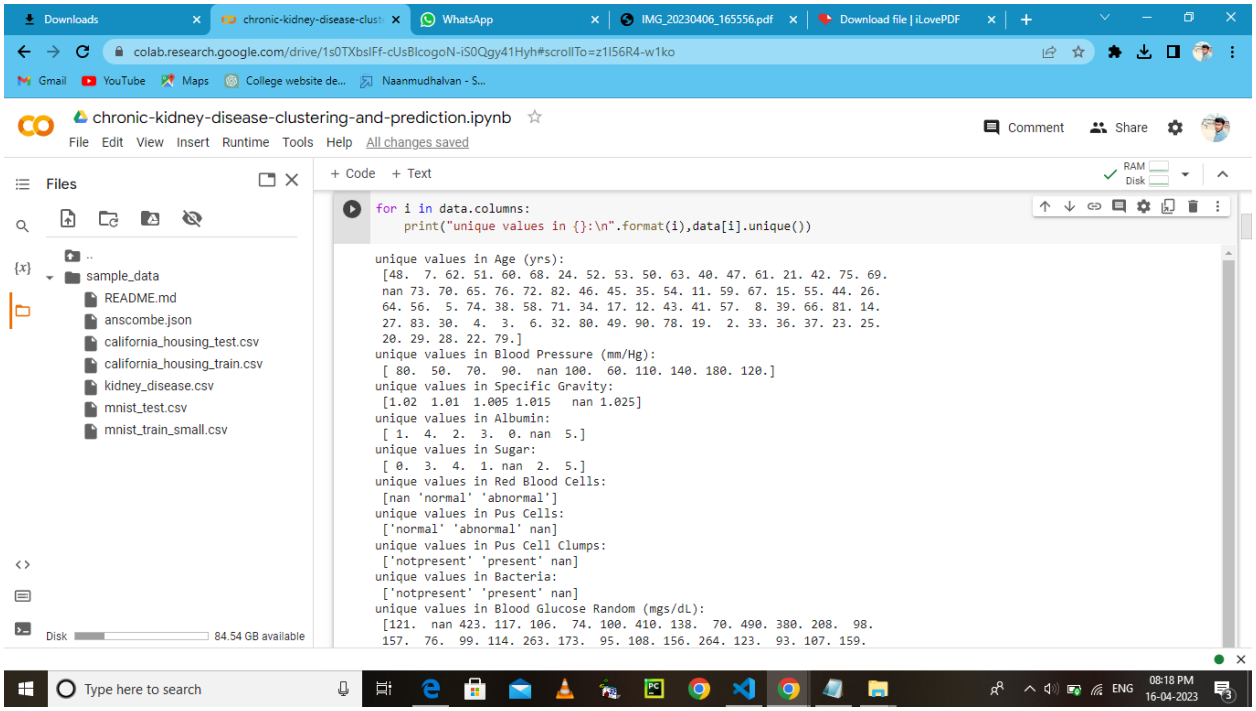


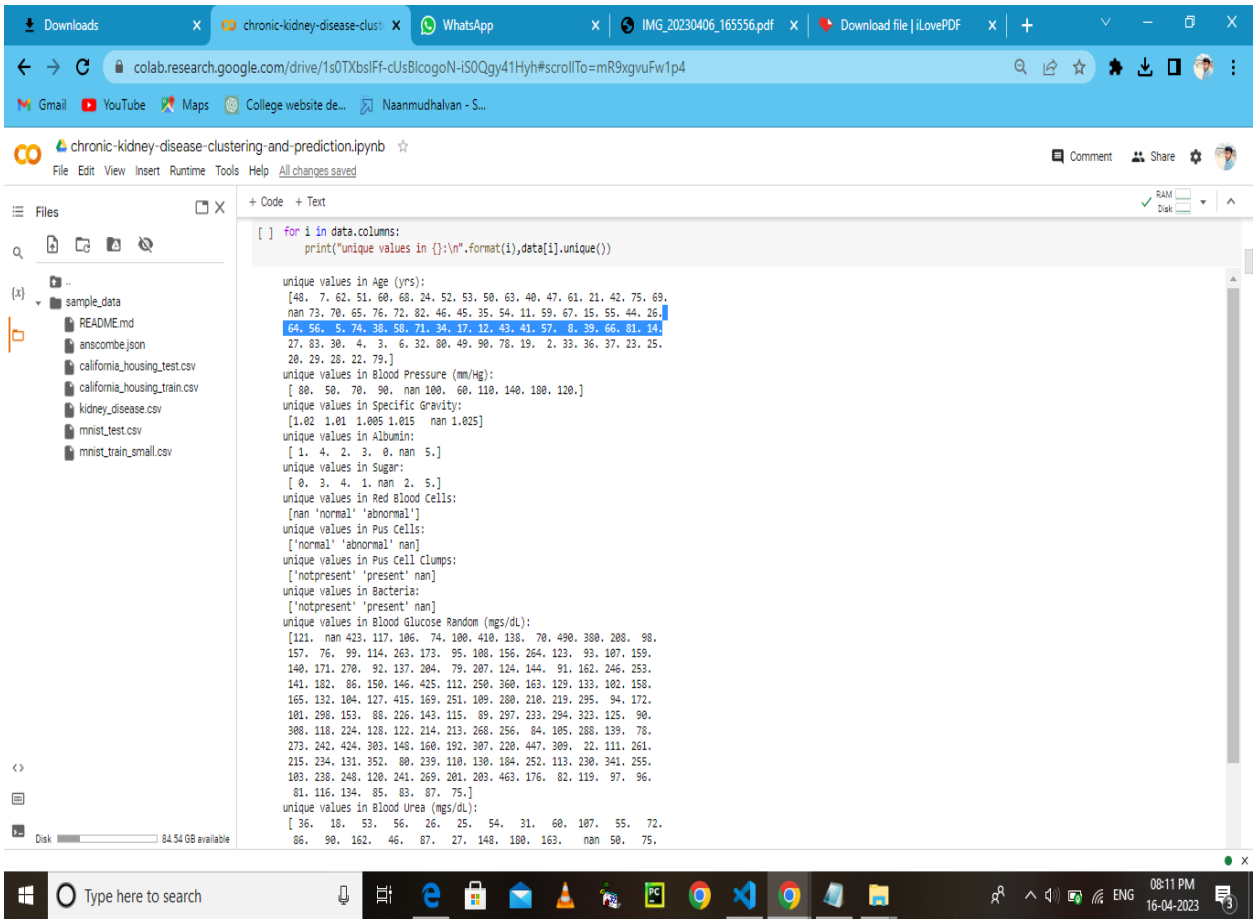
2.2 Ideation & brainstorming map

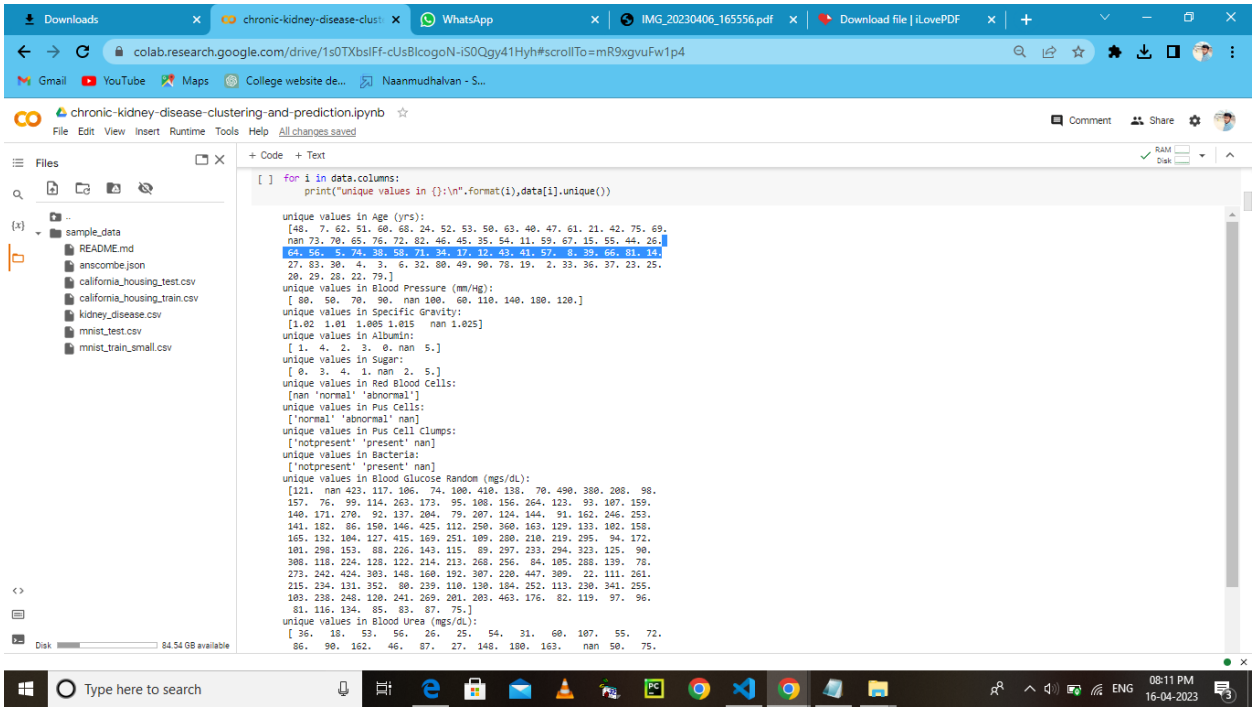


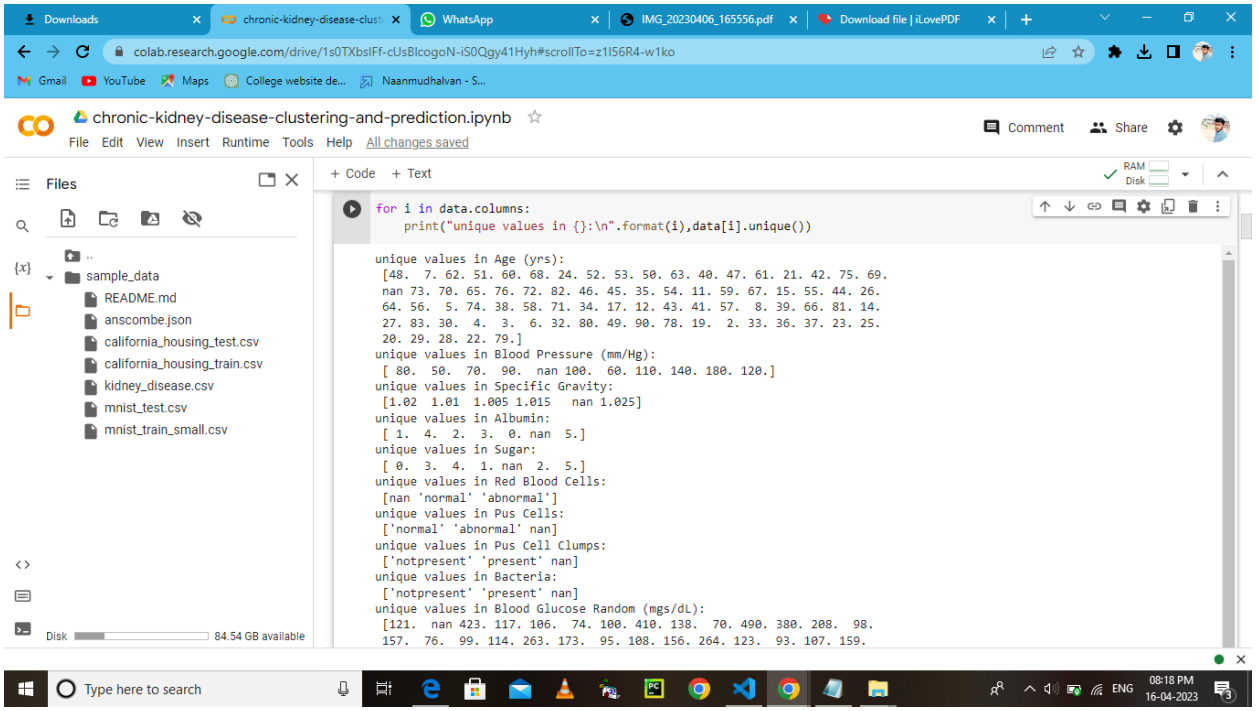


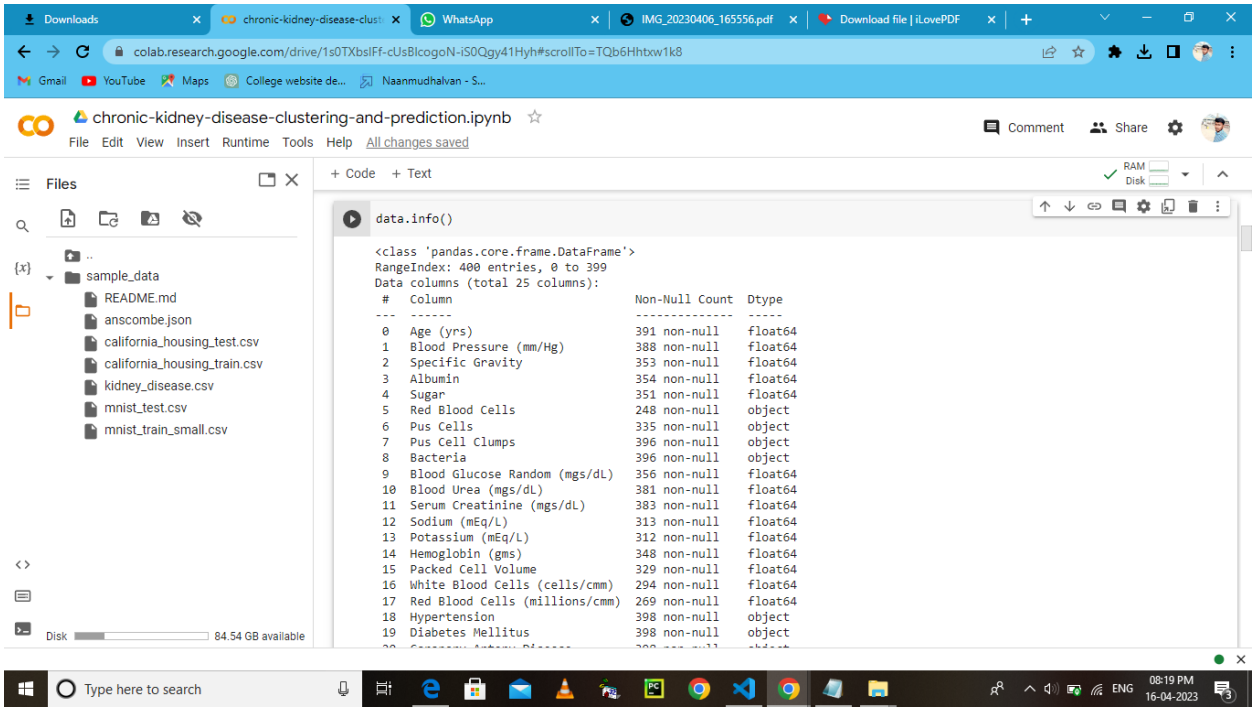
	id	age	bp	sg	al	su	rbc	pc	pcc	ba	...	pcv	wc	rc	htn	dm	cad	appet	pe	ane	classification
0	0	48.0	80.0	1.020	1.0	0.0	NaN	normal	notpresent	notpresent	...	44	7800	5.2	yes	yes	no	good	no	no	ckd
1	1	7.0	50.0	1.020	4.0	0.0	NaN	normal	notpresent	notpresent	...	38	6000	NaN	no	no	no	good	no	no	ckd
2	2	62.0	80.0	1.010	2.0	3.0	normal	normal	notpresent	notpresent	...	31	7500	NaN	no	yes	no	poor	no	yes	ckd
3	3	48.0	70.0	1.005	4.0	0.0	normal	abnormal	present	notpresent	...	32	6700	3.9	yes	no	no	poor	yes	yes	ckd
4	4	51.0	80.0	1.010	2.0	0.0	normal	normal	notpresent	notpresent	...	35	7300	4.6	no	no	no	good	no	no	ckd

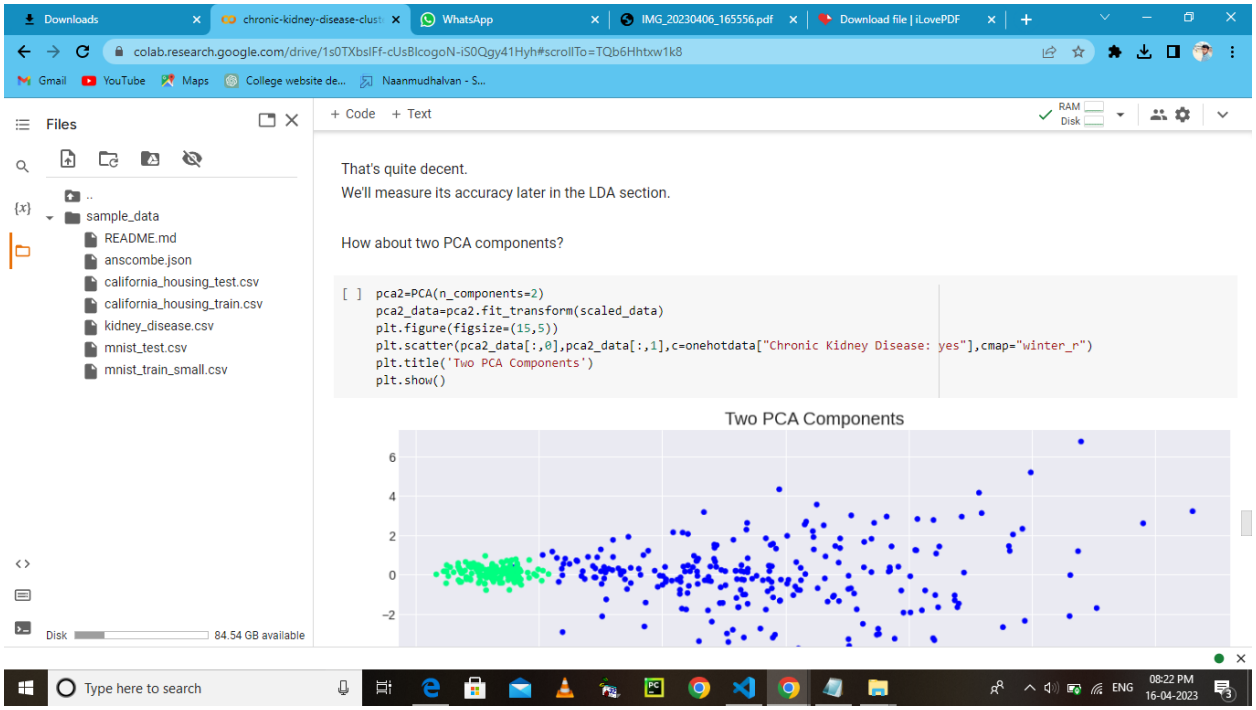

Downloads
chronic-kidney-disease-clust
WhatsApp
IMG_20230406_165556.pdf
Download file | iLovePDF
colab.research.google.com/drive/1s0TXbsIFf-cUsBlcogoN-iS0Qgy41Hyh#scrollTo=z1I56R4-w1ko
Gmail
YouTube
Maps
College website de...
Naanmudhalvan - S...
chronic-kidney-disease-clustering-and-prediction.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code
+ Text
RAM
Disk
..
sample_data
README.md
anscombe.json
california_housing_test.csv
california_housing_train.csv
kidney_disease.csv
mnist_test.csv
mnist_train_small.csv
Disk
84.54 GB available
for i in data.columns:
print("unique values in {}:\n".format(i),data[i].unique())
unique values in Age (yrs):
[48. 7. 62. 51. 60. 68. 24. 52. 53. 50. 63. 40. 47. 61. 21. 42. 75. 69.
nan 73. 70. 65. 76. 72. 82. 46. 45. 35. 54. 11. 59. 67. 15. 55. 44. 26.
64. 56. 5. 74. 38. 58. 71. 34. 17. 12. 43. 41. 57. 8. 39. 66. 81. 14.
27. 83. 30. 4. 3. 6. 32. 80. 49. 90. 78. 19. 2. 33. 36. 37. 23. 25.
20. 29. 28. 22. 79.]
unique values in Blood Pressure (mm/Hg):
[80. 50. 70. 90. nan 100. 60. 110. 140. 180. 120.]
unique values in Specific Gravity:
[1.02 1.01 1.005 1.015 nan 1.025]
unique values in Albumin:
[1. 4. 2. 3. 0. nan 5.]
unique values in Sugar:
[0. 3. 4. 1. nan 2. 5.]
unique values in Red Blood Cells:
[nan 'normal' 'abnormal']
unique values in Pus Cells:
['normal' 'abnormal' nan]
unique values in Pus Cell Clumps:
['notpresent' 'present' nan]
unique values in Bacteria:
['notpresent' 'present' nan]
unique values in Blood Glucose Random (mgs/dL):
[121. nan 423. 117. 106. 74. 100. 410. 138. 70. 490. 380. 208. 98.
157. 76. 99. 114. 263. 173. 95. 108. 156. 264. 123. 93. 107. 159.
Type here to search
ENG
08:18 PM
16-04-2023


Downloads
chronic-kidney-disease-clust
WhatsApp
IMG_20230406_165556.pdf
Download file | iLovePDF
colab.research.google.com/drive/1s0TXbslFf-cUsBlcogoN-iS0Qgy41Hyh#scrollTo=mR9xgvuFw1p4
Gmail
YouTube
Maps
College website de...
Naanmudhalvan - S...
chronic-kidney-disease-clustering-and-prediction.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code + Text
RAM
Disk
[ ] for i in data.columns:
print("unique values in {}:\n".format(i),data[i].unique())
unique values in Age (yrs):
[48. 7. 62. 51. 60. 68. 24. 52. 53. 50. 63. 40. 47. 61. 21. 42. 75. 69.
nan 73. 70. 65. 76. 72. 82. 46. 45. 35. 54. 11. 59. 67. 15. 55. 44. 26.
64. 56. 5. 74. 38. 58. 71. 34. 17. 12. 43. 41. 57. 8. 39. 66. 81. 14.
27. 83. 30. 4. 3. 6. 32. 80. 49. 90. 78. 19. 2. 33. 36. 37. 23. 25.
20. 29. 28. 22. 79.]
unique values in Blood Pressure (mm/Hg):
[80. 50. 70. 90. nan 100. 60. 110. 140. 180. 120.]
unique values in Specific Gravity:
[1.02 1.01 1.005 1.015 nan 1.025]
unique values in Albumin:
[1. 4. 2. 3. 0. nan 5.]
unique values in Sugar:
[0. 3. 4. 1. nan 2. 5.]
unique values in Red Blood Cells:
[nan 'normal' 'abnormal']
unique values in Pus Cells:
['normal' 'abnormal' nan]
unique values in Pus Cell Clumps:
['notpresent' 'present' nan]
unique values in Bacteria:
['notpresent' 'present' nan]
unique values in Blood Glucose Random (mgs/dL):
[121. nan 423. 117. 106. 74. 100. 410. 138. 70. 490. 380. 208. 98.
157. 76. 99. 114. 263. 173. 95. 108. 156. 264. 123. 93. 107. 159.
140. 171. 270. 92. 137. 204. 79. 207. 124. 144. 91. 162. 246. 253.
141. 182. 86. 150. 146. 425. 112. 250. 360. 163. 129. 133. 102. 158.
165. 132. 104. 127. 415. 169. 251. 109. 280. 210. 219. 295. 94. 172.
101. 298. 153. 88. 226. 143. 115. 89. 297. 233. 294. 323. 125. 90.
308. 118. 224. 128. 122. 214. 213. 268. 256. 84. 105. 288. 139. 78.
273. 242. 424. 303. 148. 160. 192. 307. 220. 447. 309. 22. 111. 261.
215. 234. 131. 352. 80. 239. 110. 130. 184. 252. 113. 230. 341. 255.
103. 238. 248. 120. 241. 269. 201. 203. 463. 176. 82. 119. 97. 96.
81. 116. 134. 85. 83. 87. 75.]
unique values in Blood Urea (mgs/dL):
[36. 18. 53. 56. 26. 25. 54. 31. 60. 107. 55. 72.
86. 90. 162. 46. 87. 27. 148. 180. 163. nan 50. 75.
sample_data
README.md
anscombe.json
california_housing_test.csv
california_housing_train.csv
kidney_disease.csv
mnist_test.csv
mnist_train_small.csv
Disk
84.54 GB available
Type here to search
ENG
08:11 PM
16-04-2023


Downloads
chronic-kidney-disease-clust
WhatsApp
IMG_20230406_165556.pdf
Download file | iLovePDF
colab.research.google.com/drive/1s0TXbsIFf-cUsBlcogoN-iS0Qgy41Hyh#scrollTo=mR9xgvuFw1p4
Gmail
YouTube
Maps
College website de...
Naanmudhalvan - S...
chronic-kidney-disease-clustering-and-prediction.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
sample_data
README.md
anscombe.json
california_housing_test.csv
california_housing_train.csv
kidney_disease.csv
mnist_test.csv
mnist_train_small.csv
Disk 84.54 GB available
+ Code + Text
[ ] for i in data.columns:
print("unique values in {}:\n".format(i),data[i].unique())
unique values in Age (yrs):
[48. 7. 62. 51. 60. 68. 24. 52. 53. 50. 63. 40. 47. 61. 21. 42. 75. 69.
nan 73. 70. 65. 76. 72. 82. 46. 45. 35. 54. 11. 59. 67. 15. 55. 44. 26.
64. 56. 5. 74. 38. 58. 71. 34. 17. 12. 43. 41. 57. 8. 39. 66. 81. 14.
27. 83. 30. 4. 3. 6. 32. 80. 49. 90. 78. 19. 2. 33. 36. 37. 23. 25.
20. 29. 28. 22. 79.]
unique values in Blood Pressure (mm/Hg):
[80. 50. 70. 90. nan 100. 60. 110. 140. 180. 120.]
unique values in Specific Gravity:
[1.02 1.01 1.005 1.015 nan 1.025]
unique values in Albumin:
[1. 4. 2. 3. 0. nan 5.]
unique values in Sugar:
[0. 3. 4. 1. nan 2. 5.]
unique values in Red Blood Cells:
[nan 'normal' 'abnormal']
unique values in Pus Cells:
['normal' 'abnormal' nan]
unique values in Pus Cell Clumps:
['notpresent' 'present' nan]
unique values in Bacteria:
['notpresent' 'present' nan]
unique values in Blood Glucose Random (mgs/dL):
[121. nan 423. 117. 106. 74. 100. 410. 138. 70. 490. 380. 208. 98.
157. 76. 99. 114. 263. 173. 95. 108. 156. 264. 123. 93. 107. 159.
140. 171. 270. 92. 137. 204. 79. 207. 124. 144. 91. 162. 246. 253.
141. 182. 86. 150. 146. 425. 112. 250. 360. 163. 129. 133. 102. 158.
165. 132. 104. 127. 415. 169. 251. 109. 280. 210. 219. 295. 94. 172.
101. 298. 153. 88. 226. 143. 115. 89. 297. 233. 294. 323. 125. 90.
308. 118. 224. 128. 122. 214. 213. 268. 256. 84. 105. 288. 139. 78.
273. 242. 424. 303. 148. 160. 192. 307. 220. 447. 309. 22. 111. 261.
215. 234. 131. 352. 80. 239. 110. 130. 184. 252. 113. 230. 341. 255.
103. 238. 248. 120. 241. 269. 201. 203. 463. 176. 82. 119. 97. 96.
81. 116. 134. 85. 83. 87. 75.]
unique values in Blood Urea (mgs/dL):
[36. 18. 53. 56. 26. 25. 54. 31. 60. 107. 55. 72.
86. 90. 162. 46. 87. 27. 148. 180. 163. nan 50. 75.
Type here to search
ENG
08:11 PM
16-04-2023

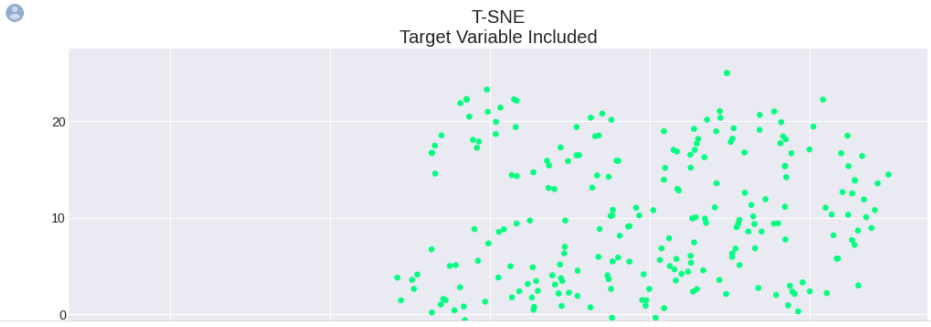

Downloads
chronic-kidney-disease-clust
WhatsApp
IMG_20230406_165556.pdf
Download file | iLovePDF
colab.research.google.com/drive/1s0TXbslFf-cUsBlcogoN-iS0Qgy41Hyh#scrollTo=z1l56R4-w1ko
Gmail
YouTube
Maps
College website de...
Naanmudhalvan - S...
chronic-kidney-disease-clustering-and-prediction.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
sample_data
README.md
anscombe.json
california_housing_test.csv
california_housing_train.csv
kidney_disease.csv
mnist_test.csv
mnist_train_small.csv
Disk 84.54 GB available
+ Code + Text
RAM
Disk
for i in data.columns:
print("unique values in {}:\n".format(i),data[i].unique())
unique values in Age (yrs):
[48. 7. 62. 51. 60. 68. 24. 52. 53. 50. 63. 40. 47. 61. 21. 42. 75. 69.
nan 73. 70. 65. 76. 72. 82. 46. 45. 35. 54. 11. 59. 67. 15. 55. 44. 26.
64. 56. 5. 74. 38. 58. 71. 34. 17. 12. 43. 41. 57. 8. 39. 66. 81. 14.
27. 83. 30. 4. 3. 6. 32. 80. 49. 90. 78. 19. 2. 33. 36. 37. 23. 25.
20. 29. 28. 22. 79.]
unique values in Blood Pressure (mm/Hg):
[80. 50. 70. 90. nan 100. 60. 110. 140. 180. 120.]
unique values in Specific Gravity:
[1.02 1.01 1.005 1.015 nan 1.025]
unique values in Albumin:
[1. 4. 2. 3. 0. nan 5.]
unique values in Sugar:
[0. 3. 4. 1. nan 2. 5.]
unique values in Red Blood Cells:
[nan 'normal' 'abnormal']
unique values in Pus Cells:
['normal' 'abnormal' nan]
unique values in Pus Cell Clumps:
['notpresent' 'present' nan]
unique values in Bacteria:
['notpresent' 'present' nan]
unique values in Blood Glucose Random (mgs/dL):
[121. nan 423. 117. 106. 74. 100. 410. 138. 70. 490. 380. 208. 98.
157. 76. 99. 114. 263. 173. 95. 108. 156. 264. 123. 93. 107. 159.
Type here to search
ENG
08:18 PM
16-04-2023

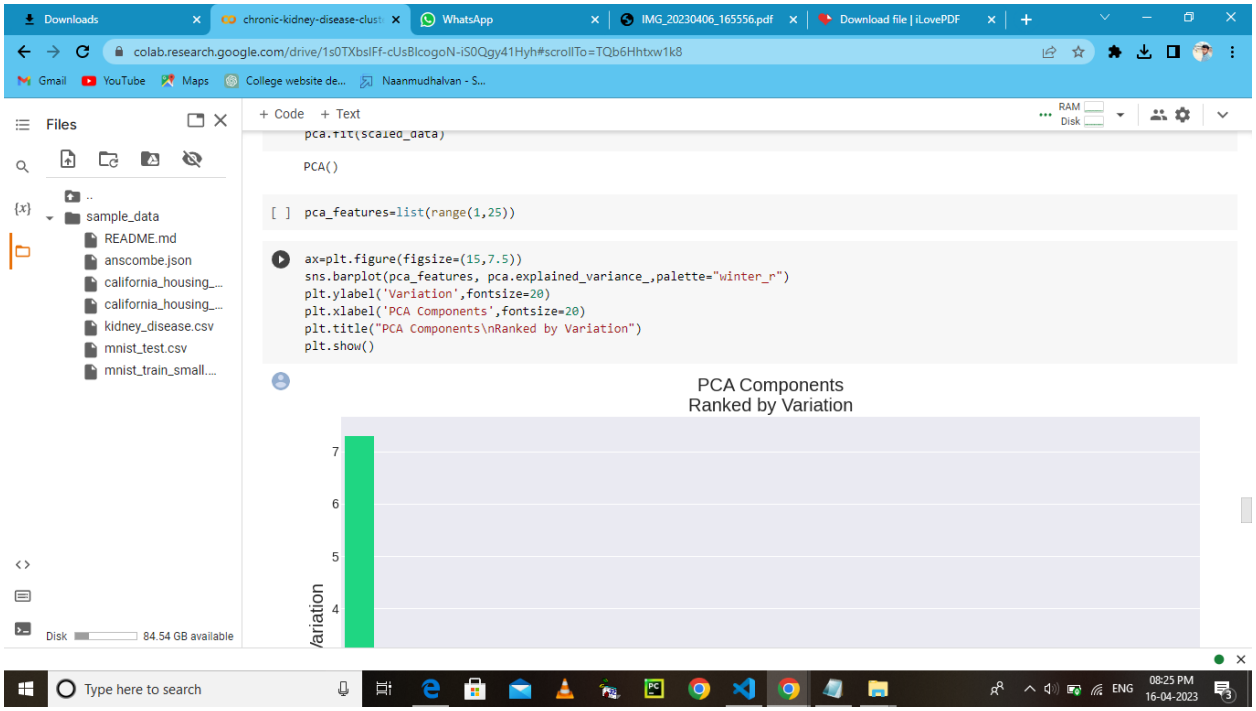

Downloads
chronic-kidney-disease-clust
WhatsApp
IMG_20230406_165556.pdf
Download file | iLovePDF
colab.research.google.com/drive/1s0TXbsIFf-cUsBlcogoN-iS0Qgy41Hyh#scrollTo=TQb6Hhtxw1k8
Gmail
YouTube
Maps
College website de...
Naanmudhalvan - S...
chronic-kidney-disease-clustering-and-prediction.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Files
+ Code
+ Text
RAM
Disk
data.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 400 entries, 0 to 399
Data columns (total 25 columns):
Column Non-Null Count Dtype
--- ------ -------------- -----
0 Age (yrs) 391 non-null float64
1 Blood Pressure (mm/Hg) 388 non-null float64
2 Specific Gravity 353 non-null float64
3 Albumin 354 non-null float64
4 Sugar 351 non-null float64
5 Red Blood Cells 248 non-null object
6 Pus Cells 335 non-null object
7 Pus Cell Clumps 396 non-null object
8 Bacteria 396 non-null object
9 Blood Glucose Random (mgs/dL) 356 non-null float64
10 Blood Urea (mgs/dL) 381 non-null float64
11 Serum Creatinine (mgs/dL) 383 non-null float64
12 Sodium (mEq/L) 313 non-null float64
13 Potassium (mEq/L) 312 non-null float64
14 Hemoglobin (gms) 348 non-null float64
15 Packed Cell Volume 329 non-null float64
16 White Blood Cells (cells/cmm) 294 non-null float64
17 Red Blood Cells (millions/cmm) 269 non-null float64
18 Hypertension 398 non-null object
19 Diabetes Mellitus 398 non-null object
sample_data
README.md
anscombe.json
california_housing_test.csv
california_housing_train.csv
kidney_disease.csv
mnist_test.csv
mnist_train_small.csv
Disk
84.54 GB available
Type here to search
ENG
08:19 PM
16-04-2023


Downloads
chronic-kidney-disease-clust
WhatsApp
IMG_20230406_165556.pdf
Download file | iLovePDF
colab.research.google.com/drive/1s0TXbsIFf-cUsBlcogoN-i50Qgy41Hyh#scrollTo=TQb6Hhtxw1k8
Gmail
YouTube
Maps
College website de...
Naanmudhalvan - S...
Files
+ Code
+ Text
RAM
Disk
sample_data
README.md
anscombe.json
california_housing_test.csv
california_housing_train.csv
kidney_disease.csv
mnist_test.csv
mnist_train_small.csv
Disk
84.54 GB available
That's quite decent.
We'll measure its accuracy later in the LDA section.
How about two PCA components?
[ ]
pca2=PCA(n_components=2)
pca2_data=pca2.fit_transform(scaled_data)
plt.figure(figsize=(15,5))
plt.scatter(pca2_data[:,0],pca2_data[:,1],c=onehotdata["Chronic Kidney Disease: yes"],cmap="winter_r")
plt.title('Two PCA Components')
plt.show()
Two PCA Components
6
4
2
0
−2
Type here to search
ENG
08:22 PM
16-04-2023

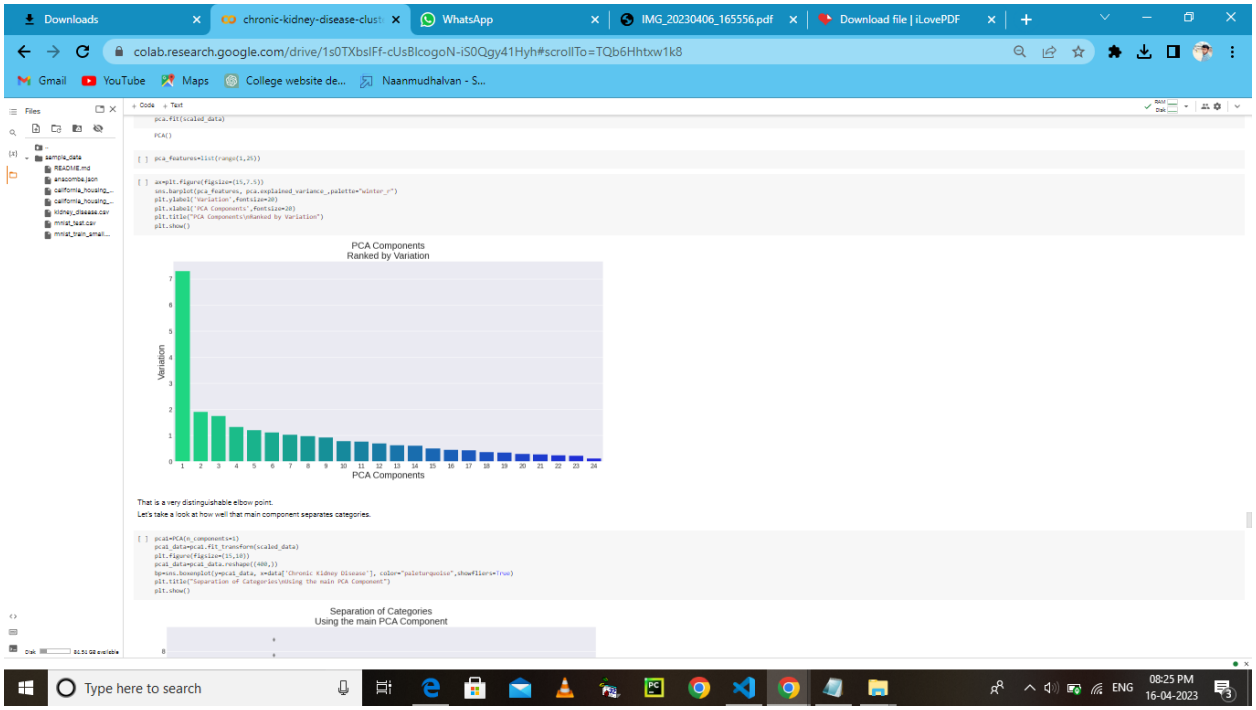
```
from sklearn.manifold import TSNE
```

```
tsne_model = TSNE()
tsne_data= tsne_model.fit_transform(full_scaled_data)
xs=tsne_data[:,0]
ys=tsne_data[:,1]
plt.figure(figsize=(15,10))
plt.scatter(xs,ys,c=pd.get_dummies(onehotdata['Chronic Kidney Disease: yes'],drop_first=True).values,cmap="winter")
plt.title("T-SNE\nTarget Variable Included")
plt.show()
```




Downloads
chronic-kidney-disease-clust
WhatsApp
IMG_20230406_165556.pdf
Download file | iLovePDF
colab.research.google.com/drive/1s0TXbslFf-cUsBlcogoN-iS0Qgy41Hyh#scrollTo=TQb6Hhtxw1k8
Gmail
YouTube
Maps
College website de...
Naanmudhalvan - S...
Files
+ Code
+ Text
RAM
Disk
PCA()
[ ] pca_features=list(range(1,25))
ax=plt.figure(figsize=(15,7.5))
sns.barplot(pca_features, pca.explained_variance_,palette="winter_r")
plt.ylabel('Variation',fontsize=20)
plt.xlabel('PCA Components',fontsize=20)
plt.title("PCA Components\nRanked by Variation")
plt.show()
sample_data
README.md
anscombe.json
california_housing_...
california_housing_...
kidney_disease.csv
mnist_test.csv
mnist_train_small....
Disk
84.54 GB available
PCA Components
Ranked by Variation
7
6
5
4
Type here to search
ENG
08:25 PM
16-04-2023

3.RESULT


Downloads
chronic-kidney-disease-clust
WhatsApp
IMG_20230406_165556.pdf
Download file | iLovePDF
colab.research.google.com/drive/1s0TXbsIFf-cUsBIcogoN-iS0Qgy41Hyh#scrollTo=TQb6Hhtxw1k8
Gmail
YouTube
Maps
College website de...
Naanmudhalvan - S...
PCA()
pca_features=list(range(1,25))
PCA Components
Ranked by Variation
Variation
PCA Components
That is a very distinguishable elbow point.
Let's take a look at how well that main component separates categories.
Separation of Categories
Using the main PCA Component
Type here to search
ENG
08:25 PM
16-04-2023

4.Advantages&Disadvantages

4.1 Advantages

Kidney function tests check how well your kidneys are working. Healthy kidneys assist with removing waste from your body. Conditions such as diabetes or high blood pressure can affect your kidney function. You may also need a kidney function test to diagnose or rule out an infection.

One of the objectives is to **reduce premature mortality from non-communicable disease by third in 2030**. Chronic kidney disease (CKD) is among the significant contributor to morbidity and mortality from non-communicable diseases that can affected 10–15% of the global population

4.2 Disadvantages

They **help your bones stay healthy, tell your body when to make new blood cells, and even help you stay upright when you're walking around all day by taking care of your blood pressure**. With all those important functions, scientist think having two kidneys must be important for our survival

The only way to find out if people have CKD is through simple **blood and urine tests**. The blood test checks for the level of creatinine, a waste product produced by muscles, to see how well the kidneys work. The urine test checks for protein, which may indicate kidney damage.

5.Applications

Your kidneys **remove wastes and extra fluid from your body**. Your kidneys also remove acid that is produced by the cells of your body and maintain a healthy balance of water, salts, and minerals—such as sodium, calcium, phosphorus, and potassium—in your blood.

6.Conclusion

This study aims to develop prediction models for detecting and diagnosing CKD based on predominant features using machine learning techniques. In addition, to help reduce clinical expenses incurred by patients who are prescribed multiple identical tests, fewer mandatory tests sufficient to detect CKD can be performed instead. Several preprocessing steps have been applied to the dataset, such as missing value imputation, normalization, and feature selection. The processed dataset was trained using different prediction models such as KNN, SVM, RF, and bagging. The models' performance was estimated to show higher reliability and significance in terms of accuracy, sensitivity, F-measure, specificity, and AUC score. KNN outperformed the existing state-of-the-art methods used in the literature, showing the efficacy of the model to be used as a decision-making system for detecting and diagnosing CKD in the early stages.

Although the dataset contains all possible attributes that are enough to detect CKD at the early stage, there is a need for additional attributes that can aid in detecting CKD. In the future, the attributes such as GFR and eGFR which are also the main predictors for detecting CKD at the early stage could be added, and the performance of the trained models could be tested.

7.Future Scope

Novel therapeutic alternatives for ESRD include **wearable artificial kidneys, xenotransplantation, stem cell–based therapy, and bioengineered and bio-artificial kidneys**. Of note, one of the main objectives of these novel

therapeutic approaches should be to maintain patients at home and to avoid dialysis centers.

8.APPENDIX

```
import glob
import  pandas as pd
import numpy as np
from collections import Counter as c
import matplotlib.pyplot as plt
import seaborn as sns
import missingno as msno
from sklearn.metrics import accuracy_score, confusion_matrix
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.linear_model import LogisticRegression
import pickle
data=pd.read_csv("../input/ckdisease/kidney_disease.csv")
n_rows, n_cols = (7,2)

figure, axes = plt.subplots(nrows=n_rows, ncols=n_cols,figsize=(20,
 50))
figure.suptitle('\n\nDistributions of Numerical Features', fontsize
=60)

for index, column in enumerate(numeric):

    i,j = (index // n_cols), (index % n_cols)

    miss_perc="%.2f"%(100*(1-
(data[column].dropna().shape[0])/data.shape[0]))

    collabel=column+"\n({}% is missing)".format(miss_perc)

    fig=sns.distplot(data[column], color="g", label=collabel, norm_
hist=True,

    ax=axes[i,j], kde_kws={"lw":4})

    fig=fig.legend(loc='best', fontsize=18)
```

```
    axes[i,j].set_ylabel("Probability Density",fontsize='medium')

    axes[i,j].set_xlabel(None)

plt.show()
for index, column in enumerate(categoricals):

    i,j = index // n_cols, index % n_cols

    miss_perc="%.2f"%(100*(1-
(data[column].dropna().shape[0])/data.shape[0]))

    collabel=column+"\n({}% is missing)".format(miss_perc)

    fig = sns.countplot(x=column, data=data,label=collabel, palette
=sns.cubehelix_palette(rot=-.35,light=0.85,hue=1),

    ax=axes[i,j])

    axes[i,j].set_title(collabel,fontsize=30)

    axes[i,j].set_xlabel(None)

    axes[i,j].set_ylabel("Count",fontsize=20)

    axes[i,j].set_xticklabels(axes[i,j].get_xticklabels(), Fontsize
=28)

plt.show()
```