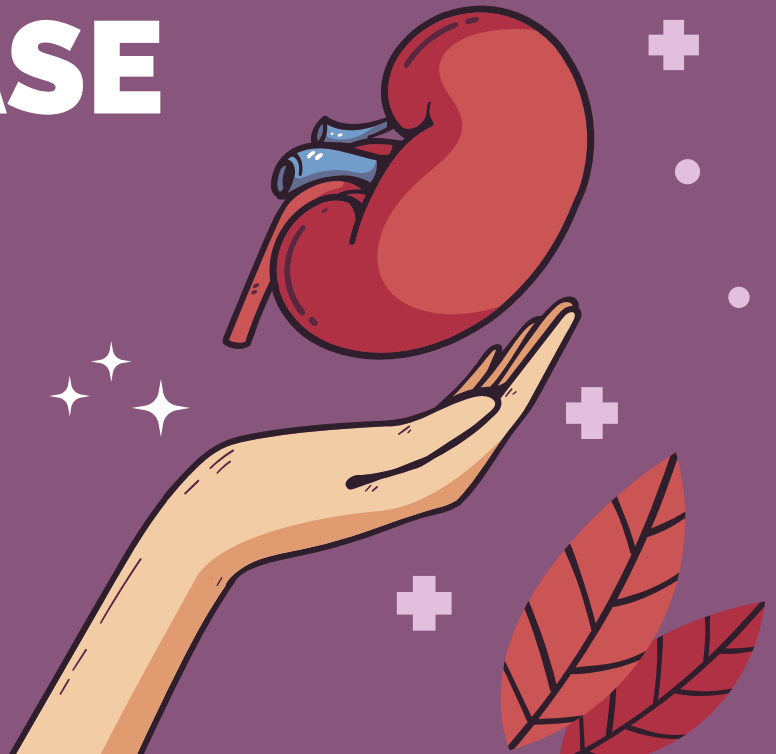
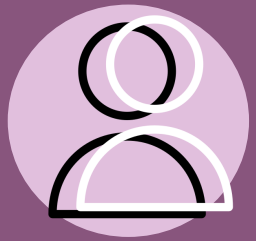
KIDNEYS DISEASE PREDICTION

Data set



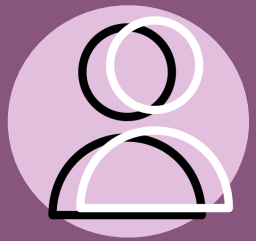
OUR TEAM



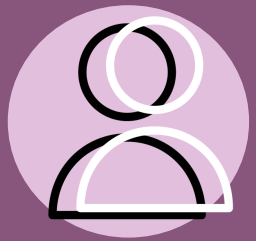
Elaan moazen
2006947

Ghaidaa Almaghrabi
1912386

Raghad alQithmi
2005999



Raneen Alshehri
2005560

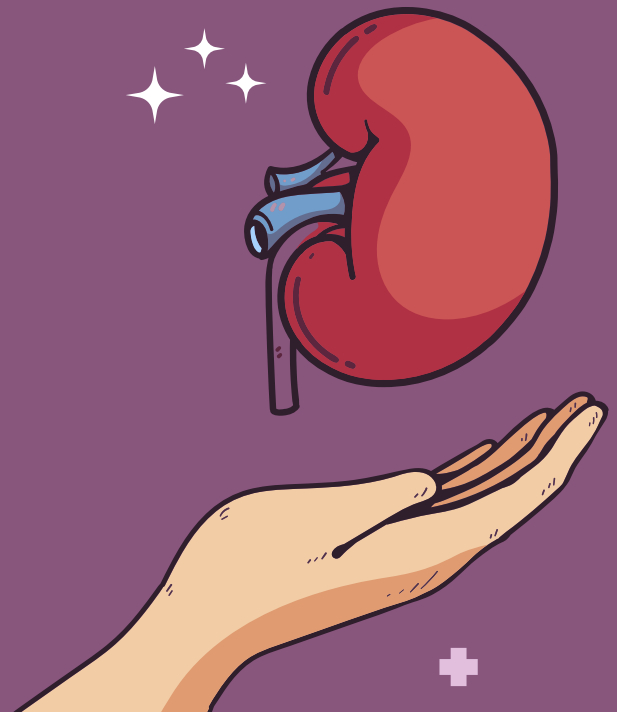


Taif Basheikh
2005890

INTRODUCTION

We use the Chronic Kidney Disease dataset to determine who has kidney disease and who does not have kidney disease and To know the symptoms that increase the incidence of kidney disease

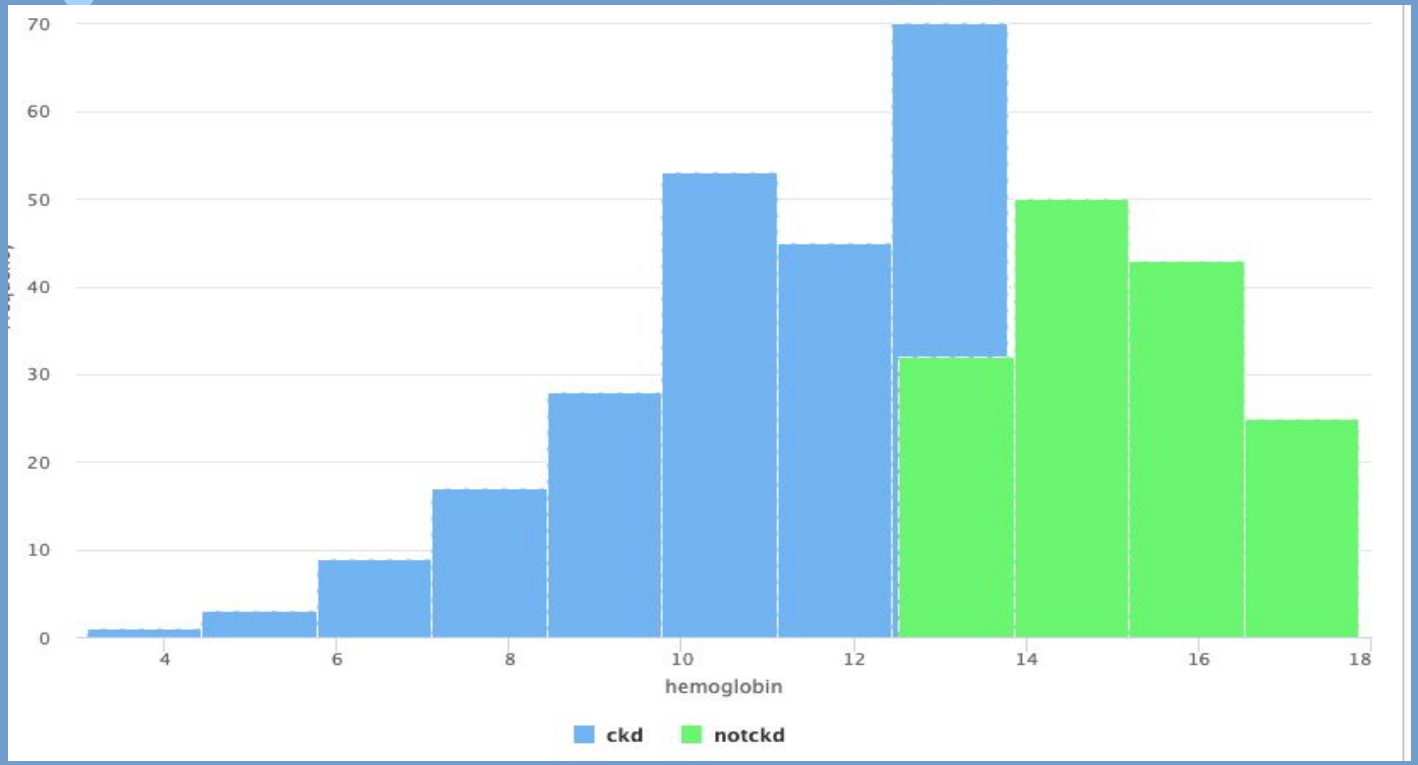
Our data set contains 390 rows and we assign a Classification label to help us divide it into two groups



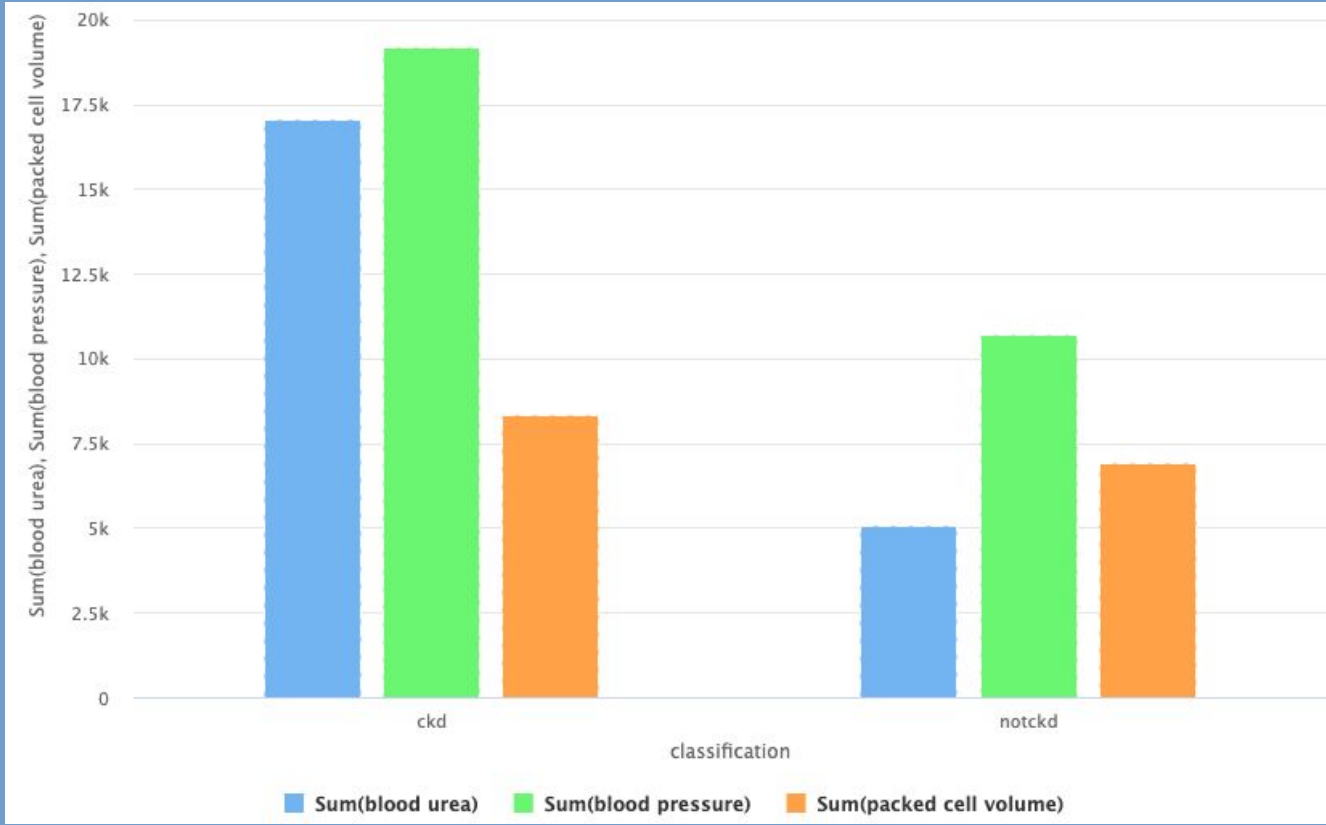
Phase 1

- Visualization Techniques
- Cleaning Data
- Correlated Attributes

Histogram



Bar Chart

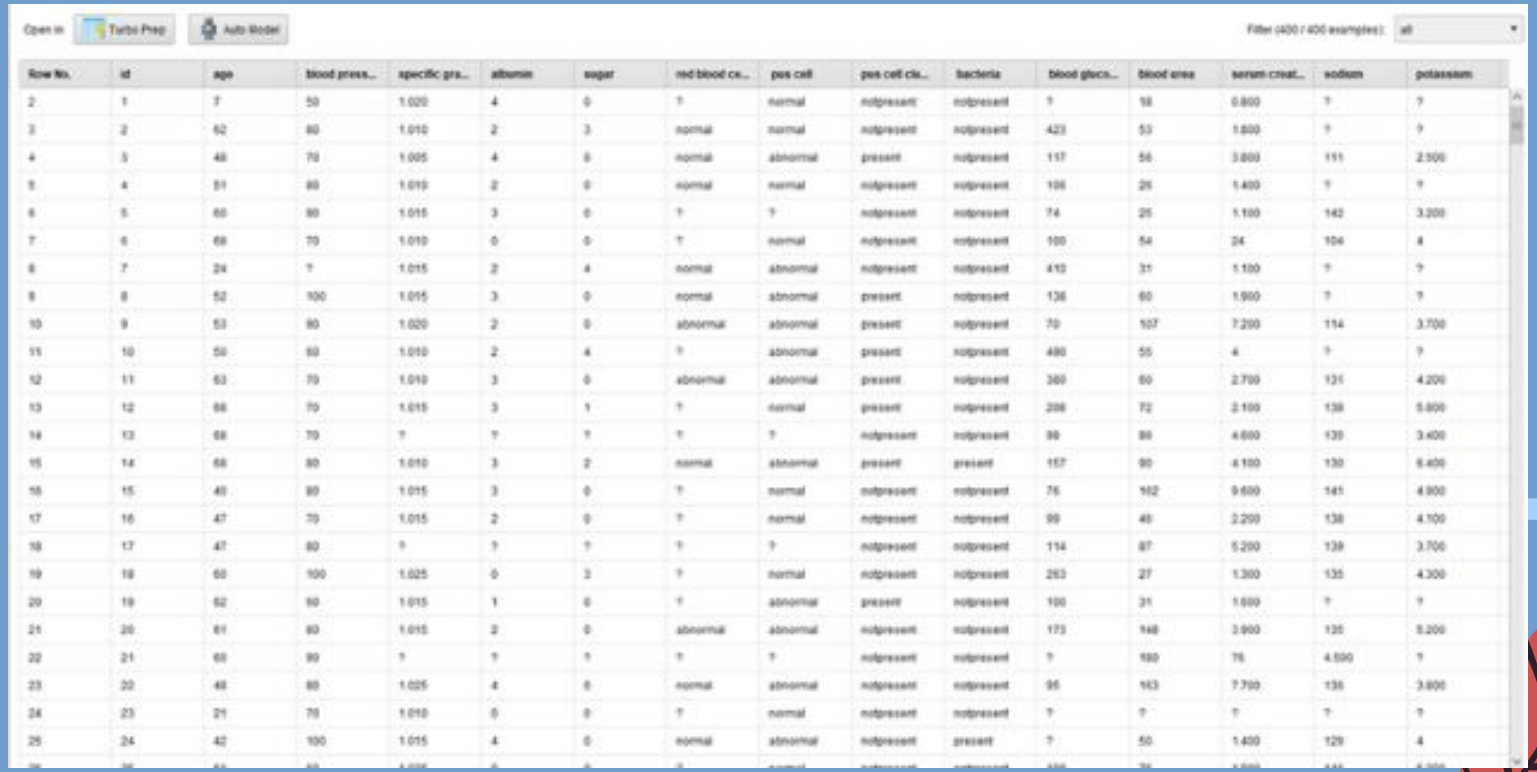

Sum(blood urea), Sum(blood pressure), Sum(packed cell volume)
20k
17.5k
15k
12.5k
10k
7.5k
5k
2.5k
0
ckd
notckd
classification
Sum(blood urea)
Sum(blood pressure)
Sum(packed cell volume)

correlation

The most related attribute

Positive Correlation	Negative Correlation
hemoglobin, packed cell volume	serum creatinine, sodium
0.850	**-0.626**

Data Before Cleaning



Open in Turbo Prep Auto Model

Filter (400 / 400 examples): all

Row No.	id	age	blood press...	specific gra...	albumin	sugar	red blood ce...	pus cell	pus cell cla...	bacteria	blood gluco...	blood urea	serum creat...	sodium	potassium
2	1	7	50	1.020	4	0	?	normal	notpresent	notpresent	?	18	0.800	?	?
3	2	62	80	1.010	2	3	normal	normal	notpresent	notpresent	423	53	1.800	?	?
4	3	48	70	1.005	4	0	normal	abnormal	present	notpresent	117	56	3.800	111	2.500
5	4	51	80	1.010	2	0	normal	normal	notpresent	notpresent	106	26	1.400	?	?
6	5	60	90	1.015	3	0	?	?	notpresent	notpresent	74	25	1.100	142	3.200
7	6	68	70	1.010	0	0	?	normal	notpresent	notpresent	100	54	24	104	4
8	7	24	?	1.015	2	4	normal	abnormal	notpresent	notpresent	410	31	1.100	?	?
9	8	52	100	1.015	3	0	normal	abnormal	present	notpresent	138	60	1.900	?	?
10	9	53	90	1.020	2	0	abnormal	abnormal	present	notpresent	70	107	7.200	114	3.700
11	10	50	60	1.010	2	4	?	abnormal	present	notpresent	490	55	4	?	?
12	11	63	70	1.010	3	0	abnormal	abnormal	present	notpresent	380	60	2.700	131	4.200
13	12	68	70	1.015	3	1	?	normal	present	notpresent	208	72	2.100	138	5.800
14	13	68	70	?	?	?	?	?	notpresent	notpresent	98	86	4.600	135	3.400
15	14	68	80	1.010	3	2	normal	abnormal	present	present	157	90	4.100	130	6.400
16	15	40	80	1.015	3	0	?	normal	notpresent	notpresent	76	162	9.600	141	4.900
17	16	47	70	1.015	2	0	?	normal	notpresent	notpresent	99	46	2.200	138	4.100
18	17	47	80	?	?	?	?	?	notpresent	notpresent	114	87	5.200	139	3.700
19	18	60	100	1.025	0	3	?	normal	notpresent	notpresent	263	27	1.300	135	4.300
20	19	62	60	1.015	1	0	?	abnormal	present	notpresent	100	31	1.600	?	?
21	20	61	80	1.015	2	0	abnormal	abnormal	notpresent	notpresent	173	148	3.900	135	5.200
22	21	60	90	?	?	?	?	?	notpresent	notpresent	?	180	76	4.500	?
23	22	48	80	1.025	4	0	normal	abnormal	notpresent	notpresent	95	163	7.700	136	3.800
24	23	21	70	1.010	0	0	?	normal	notpresent	notpresent	?	?	?	?	?
25	24	42	100	1.015	4	0	normal	abnormal	notpresent	present	?	50	1.400	129	4

Data After Cleaning

Row No.	outlier	id	red blood ce...	pus cell	pus cell clu...	bacteria	hypertension	diabetes me...	coronary art...	appetite	pedal edema	anemia	classification	age	bloo
1	false	0	normal	normal	notpresent	notpresent	yes	yes	no	good	no	no	ckd	48	80
2	false	1	normal	normal	notpresent	notpresent	no	no	no	good	no	no	ckd	7	50
3	false	2	normal	normal	notpresent	notpresent	no	yes	no	poor	no	yes	ckd	62	80
4	false	3	normal	abnormal	present	notpresent	yes	no	no	poor	yes	yes	ckd	48	70
5	false	4	normal	normal	notpresent	notpresent	no	no	no	good	no	no	ckd	51	80
6	false	5	normal	normal	notpresent	notpresent	yes	yes	no	good	yes	no	ckd	60	90
7	false	6	normal	normal	notpresent	notpresent	no	no	no	good	no	no	ckd	68	70
8	false	7	normal	abnormal	notpresent	notpresent	no	yes	no	good	yes	no	ckd	24	76
9	false	8	normal	abnormal	present	notpresent	yes	yes	no	good	no	yes	ckd	52	100
10	false	9	abnormal	abnormal	present	notpresent	yes	yes	no	poor	no	yes	ckd	53	90
11	false	10	normal	abnormal	present	notpresent	yes	yes	no	good	no	yes	ckd	50	60
12	false	11	abnormal	abnormal	present	notpresent	yes	yes	no	poor	yes	no	ckd	63	70
13	false	12	normal	normal	present	notpresent	yes	yes	yes	poor	yes	no	ckd	68	70
14	false	13	normal	normal	notpresent	notpresent	yes	yes	yes	poor	yes	no	ckd	68	70
15	false	14	normal	abnormal	present	present	yes	yes	yes	poor	yes	no	ckd	68	80
16	false	15	normal	normal	notpresent	notpresent	yes	no	no	good	no	yes	ckd	40	80
17	false	16	normal	normal	notpresent	notpresent	no	no	no	good	no	no	ckd	47	70
18	false	17	normal	normal	notpresent	notpresent	yes	no	no	poor	no	no	ckd	47	80
19	false	18	normal	normal	notpresent	notpresent	yes	yes	yes	good	no	no	ckd	60	100
20	false	19	normal	abnormal	present	notpresent	yes	no	yes	good	no	no	ckd	62	60
21	false	20	abnormal	abnormal	notpresent	notpresent	yes	yes	yes	poor	yes	yes	ckd	61	80
22	false	21	normal	normal	notpresent	notpresent	yes	yes	yes	good	no	no	ckd	60	90
23	false	22	normal	abnormal	notpresent	notpresent	yes	no	no	good	no	yes	ckd	48	80
24	false	23	normal	normal	notpresent	notpresent	no	no	no	poor	no	yes	ckd	21	70
25	false	24	normal	abnormal	notpresent	present	yes	no	no	poor	no	no	ckd	42	100

Phase 2

Measure the accuracy

- Decision Tree
- pre-processing steps
- Neural Network
- Random Forest

Decision Tree

Normalizing the data

Range normalizations & Z-score normalizations

Accuracy: 96.67%

accuracy: 96.67% +/- 2.43% (micro average: 96.67%)

	true ckd	true notckd	class precision
pred. ckd	237	10	95.95%
pred. notckd	3	140	97.90%
class recall	98.75%	93.33%	

Decision Tree

Imputing missing values

Accuracy: 95.90%

accuracy: 95.90% +/- 0.73% (micro average: 95.90%)

	true ckd	true notckd	class precision
pred. ckd	231	7	97.06%
pred. notckd	9	143	94.08%
class recall	96.25%	95.33%	

Decision Tree

Discretizing the numeric attribute(bins 3)

Accuracy: 92.56%

accuracy: 92.56% +/- 5.05% (micro average: 92.56%)

	true ckd	true notckd	class precision
pred. ckd	220	9	96.07%
pred. notckd	20	141	87.58%
class recall	91.67%	94.00%	

Decision Tree

Reducing Dimensions (Weight by Relied)

Accuracy: 97.18%

accuracy: 97.18% +/- 2.82% (micro average: 97.18%)			
	true ckd	true notckd	class precision
pred. ckd	231	2	99.14%
pred. notckd	9	148	94.27%
class recall	96.25%	98.67%	

pre-processing steps

Included dimensionality reduction

Accuracy: 96.67%

Best accuracy of them is for Normalized Range(bin 3)

accuracy: 96.67% +/- 3.43% (micro average: 96.67%)			
	true ckd	true notckd	class precision
pred. ckd	233	6	97.49%
pred. notckd	7	144	95.36%
class recall	97.08%	96.00%	

pre-processing steps

excluding dimensionality reduction

Accuracy: 97.95%

Best accuracy of them is for (Normalize Range bin 2)

accuracy: 97.95% +/- 2.36% (micro average: 97.95%)

	true ckd	true notckd	class precision
pred. ckd	232	0	100.00%
pred. notckd	8	150	94.94%
class recall	96.67%	100.00%	

Neural Network

Accuracy: 98.97%

accuracy: 98.97% +/- 1.32% (micro average: 98.97%)

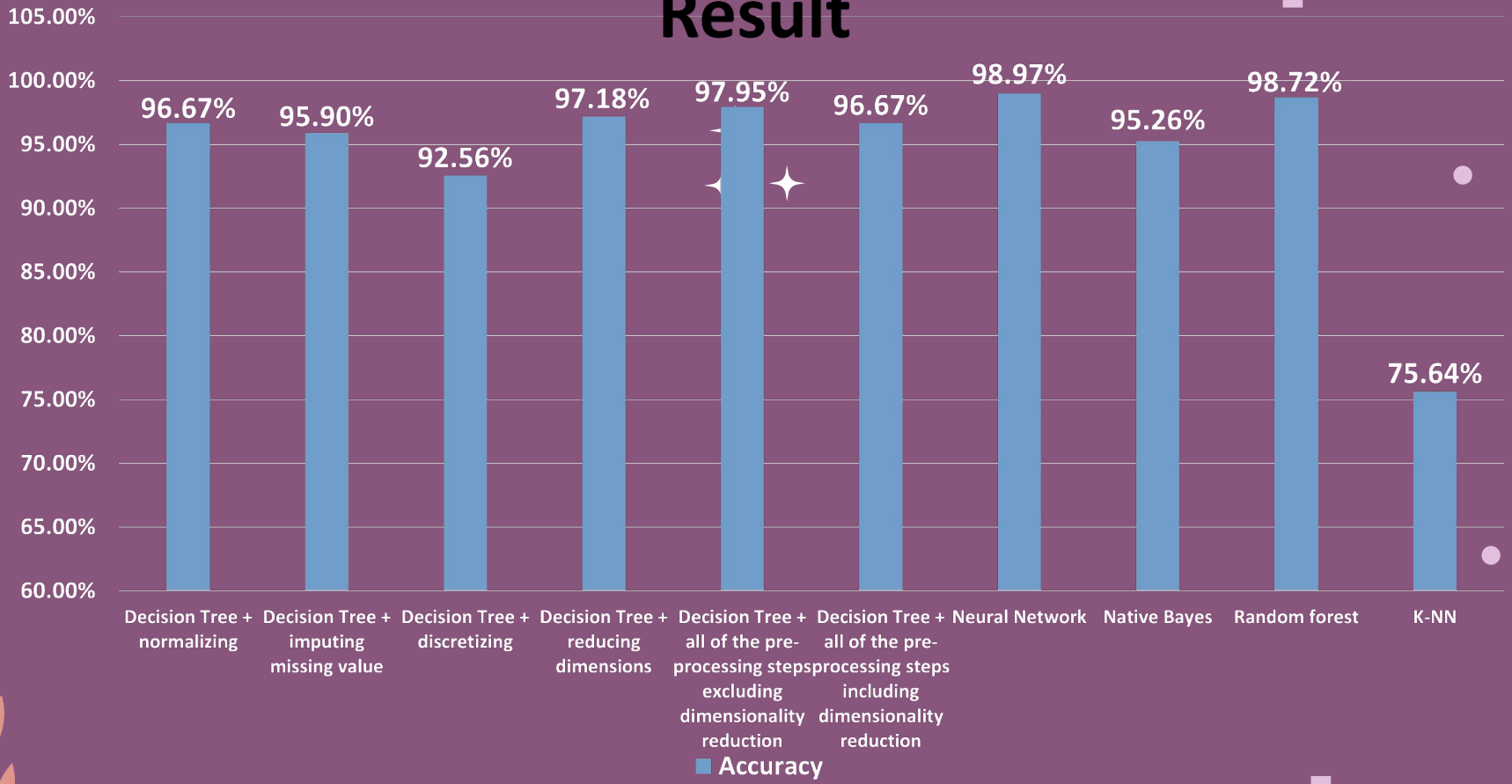
	true ckd	true notckd	class precision
pred. ckd	236	0	100.00%
pred. notckd	4	150	97.40%
class recall	98.33%	100.00%	

Random Forest

Accuracy: **98.72%**

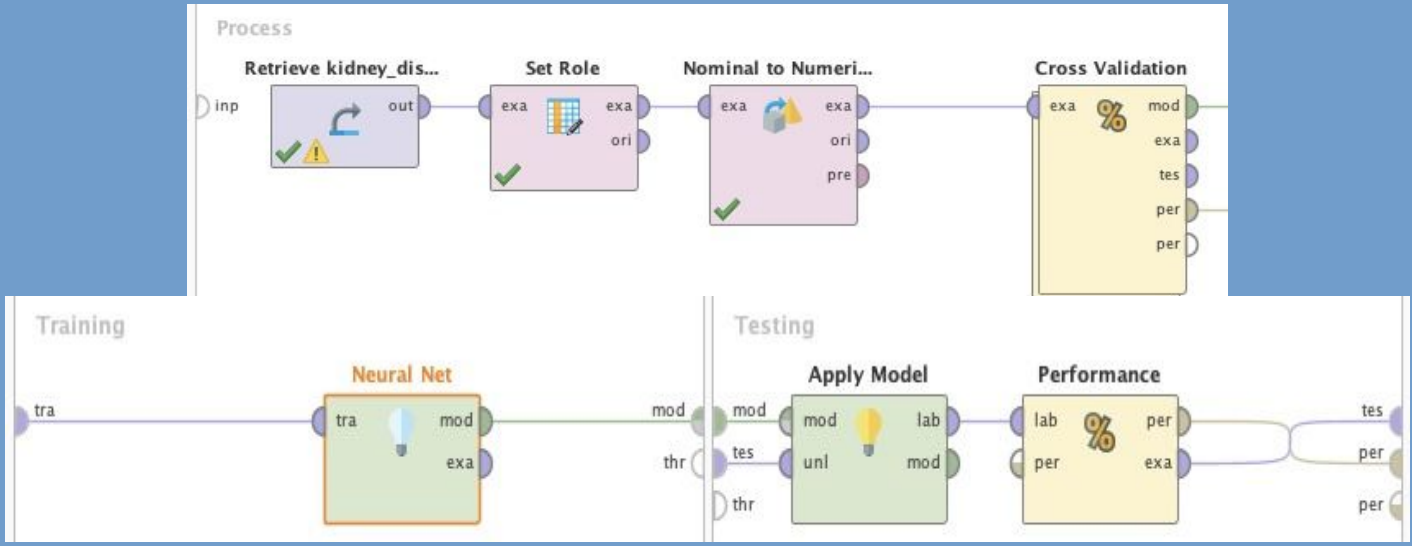
	true ckd	true notckd	class precision
pred. ckd	238	3	98.76%
pred. notckd	2	147	98.66%
class recall	99.17%	98.00%	

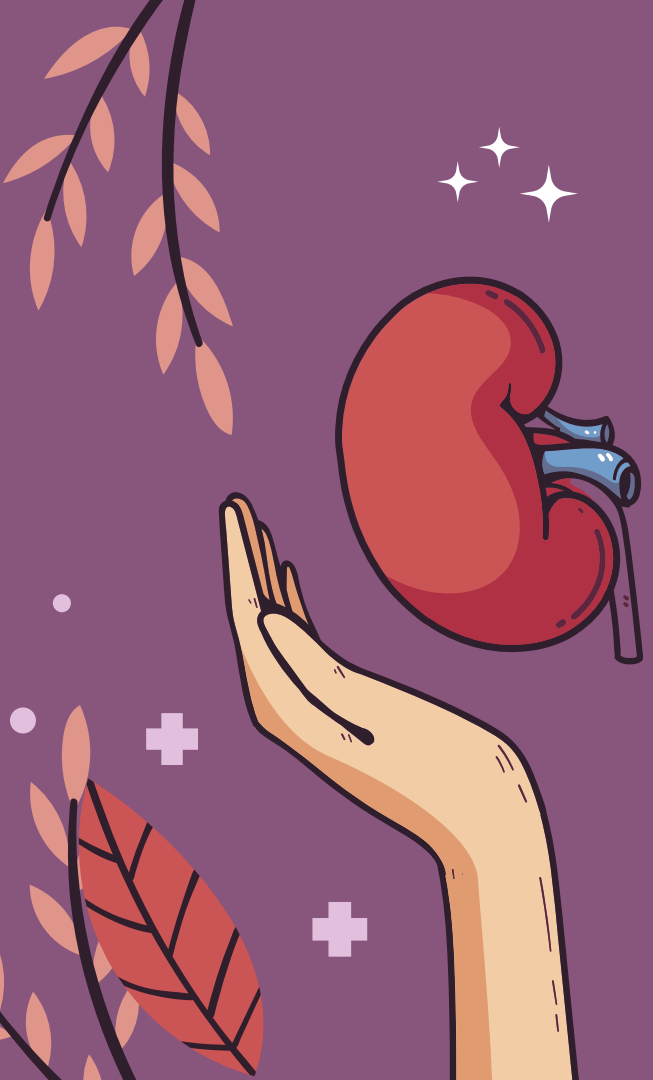
Result


105.00%
100.00%
95.00%
90.00%
85.00%
80.00%
75.00%
70.00%
65.00%
60.00%
96.67%
95.90%
92.56%
97.18%
97.95%
96.67%
98.97%
95.26%
98.72%
75.64%
Decision Tree + normalizing
Decision Tree + imputing missing value
Decision Tree + discretizing
Decision Tree + reducing dimensions
Decision Tree + all of the pre-processing steps excluding dimensionality reduction
Decision Tree + all of the pre-processing steps including dimensionality reduction
Neural Network
Native Bayes
Random forest
K-NN
Accuracy

The Best Combination

Neural Network **98.97%**





THANKS!

DO YOU HAVE ANY QUESTIONS?