



Potability Test of Water



POTABLE WATER



NON-POTABLE WATER

By - Akshay Pokale

PGA- 42

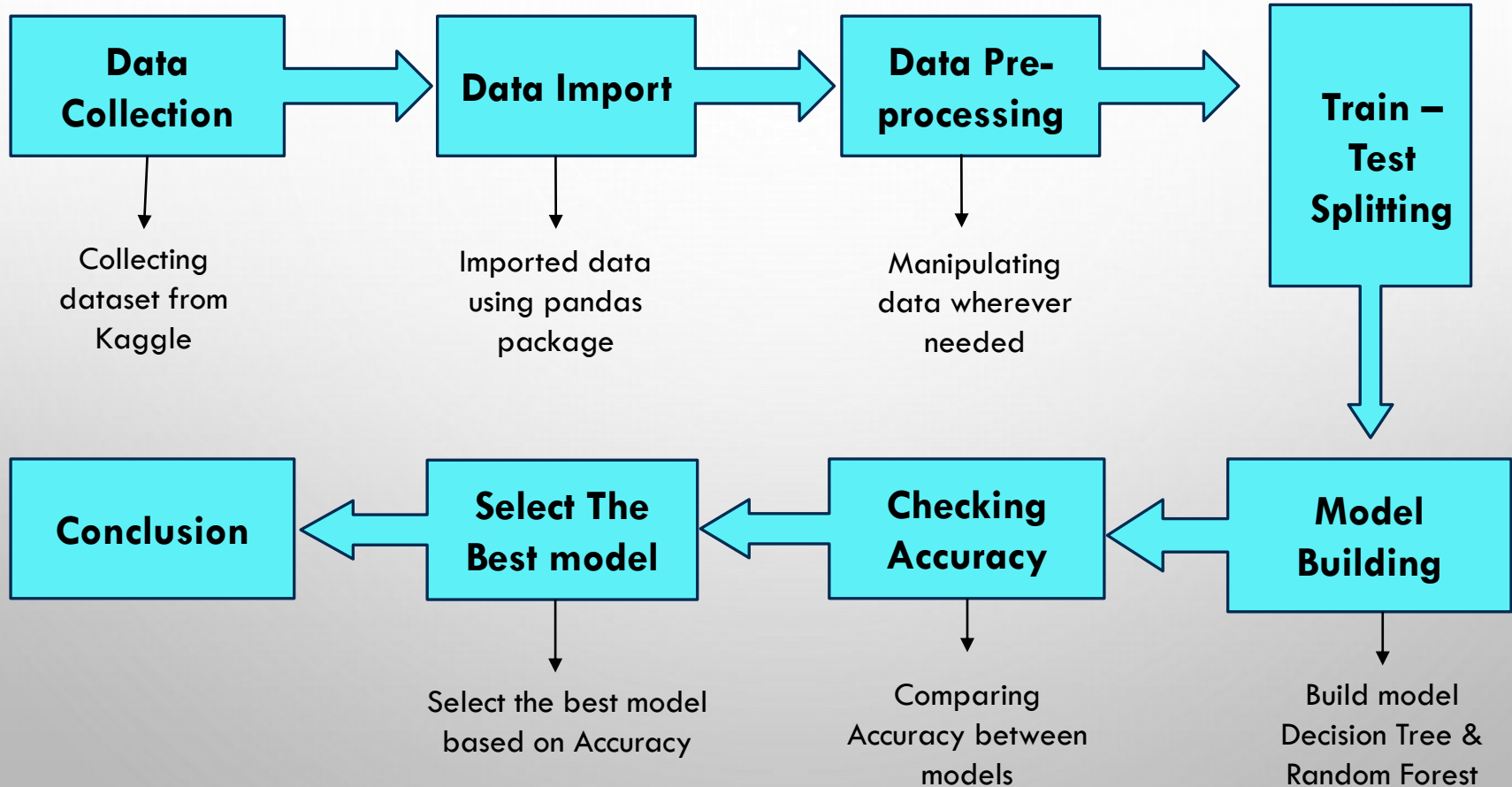
Introduction

- ❖ Water is an inorganic compound with the chemical formula H_2O .
- ❖ Potable water, also known as drinking water, comes from surface and ground sources.
- ❖ Non-potable water is not suitable for drinking.
- ❖ Microbiologically contaminated drinking water can transmit diseases such as **diarrhoea**, **cholera**, **dysentery**, **typhoid** and **polio** and is estimated to cause **4,85,000 diarrhoeal** deaths each year. (According to WHO)

Objective

“Developing A Model To Predict The Potability Of Water Based On Various Parameters Such As Ph Level, Hardness, Turbidity, And Conductivity.”

Process Flow



Tools And Platform Used

❖ Tools : Python



❖ Platform : Jupyter Notebook



❖ Library Used : Numpy, pandas,
Matplotlib, Seaborn, Scikit-Learn



Data Description

Ph	pH is a measure of the acidity or alkalinity of water
Hardness	Presence of salts of calcium and magnesium
Solids(TDS)	"Dissolved solids" refer to any minerals, salts, metals, cations or anions dissolved in water.
Chloramines	disinfectants used to treat drinking water
Sulfate	A substance that occurs naturally in drinking water
Conductivity	A measure of the ability of water to pass an electrical current
Organic_carbon	amount of carbon that is present in the water in an organic form
Trihalomethanes	A byproduct of the water treatment process
Turbidity	The measure of relative clarity of a liquid
Potability	Potability refers to the suitability of water for human consumption (1-potable/ 0- non-potable)

Data Pre-processing

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 3276 entries, 0 to 3275
```

```
Data columns (total 10 columns):
```

#	Column	Non-Null Count	Dtype
0	ph	2785 non-null	float64
1	Hardness	3276 non-null	float64
2	Solids	3276 non-null	float64
3	Chloramines	3276 non-null	float64
4	Sulfate	2495 non-null	float64
5	Conductivity	3276 non-null	float64
6	Organic_carbon	3276 non-null	float64
7	Trihalomethanes	3114 non-null	float64
8	Turbidity	3276 non-null	float64
9	Potability	3276 non-null	object

```
dtypes: float64(9), object(1)
```

```
memory usage: 256.1+ KB
```

Identifying and treatment of missing value

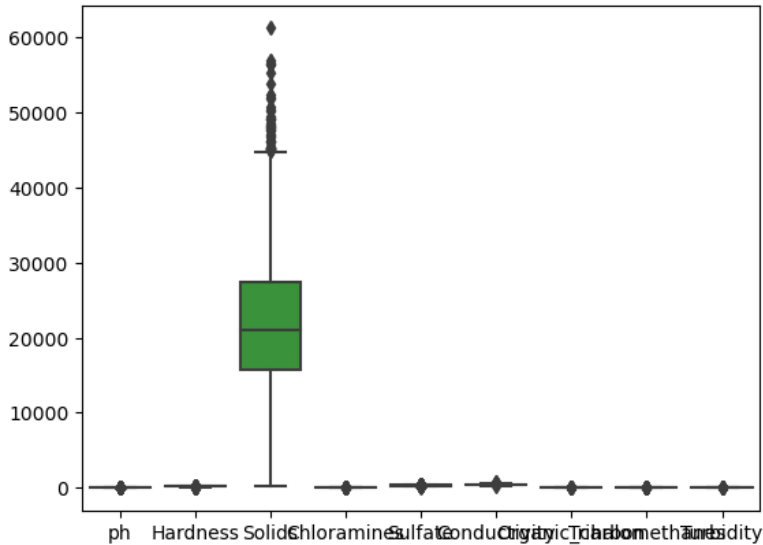
```
ph          491
Hardness    0
Solids      0
Chloramines 0
Sulfate     781
Conductivity 0
Organic_carbon 0
Trihalomethanes 162
Turbidity   0
Potability  0
dtype: int64
```

Replacing Missing
values by Median

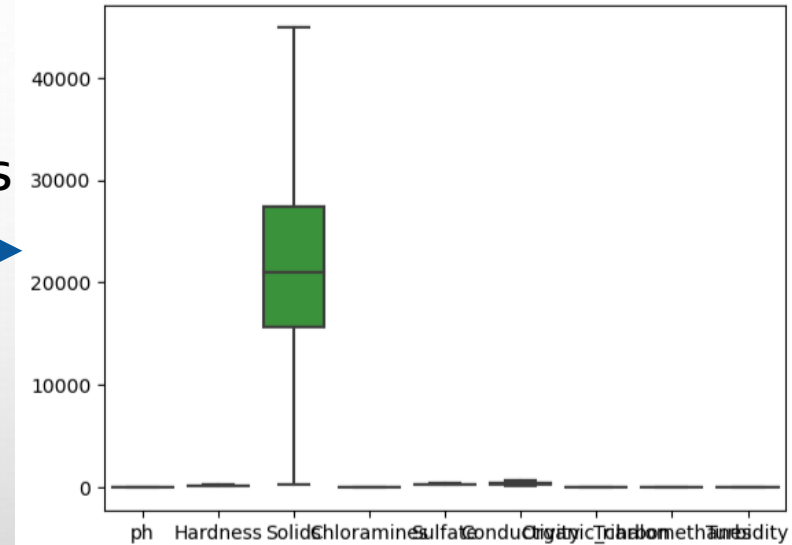


```
ph          0
Hardness    0
Solids      0
Chloramines 0
Sulfate     0
Conductivity 0
Organic_carbon 0
Trihalomethanes 0
Turbidity   0
Potability  0
dtype: int64
```

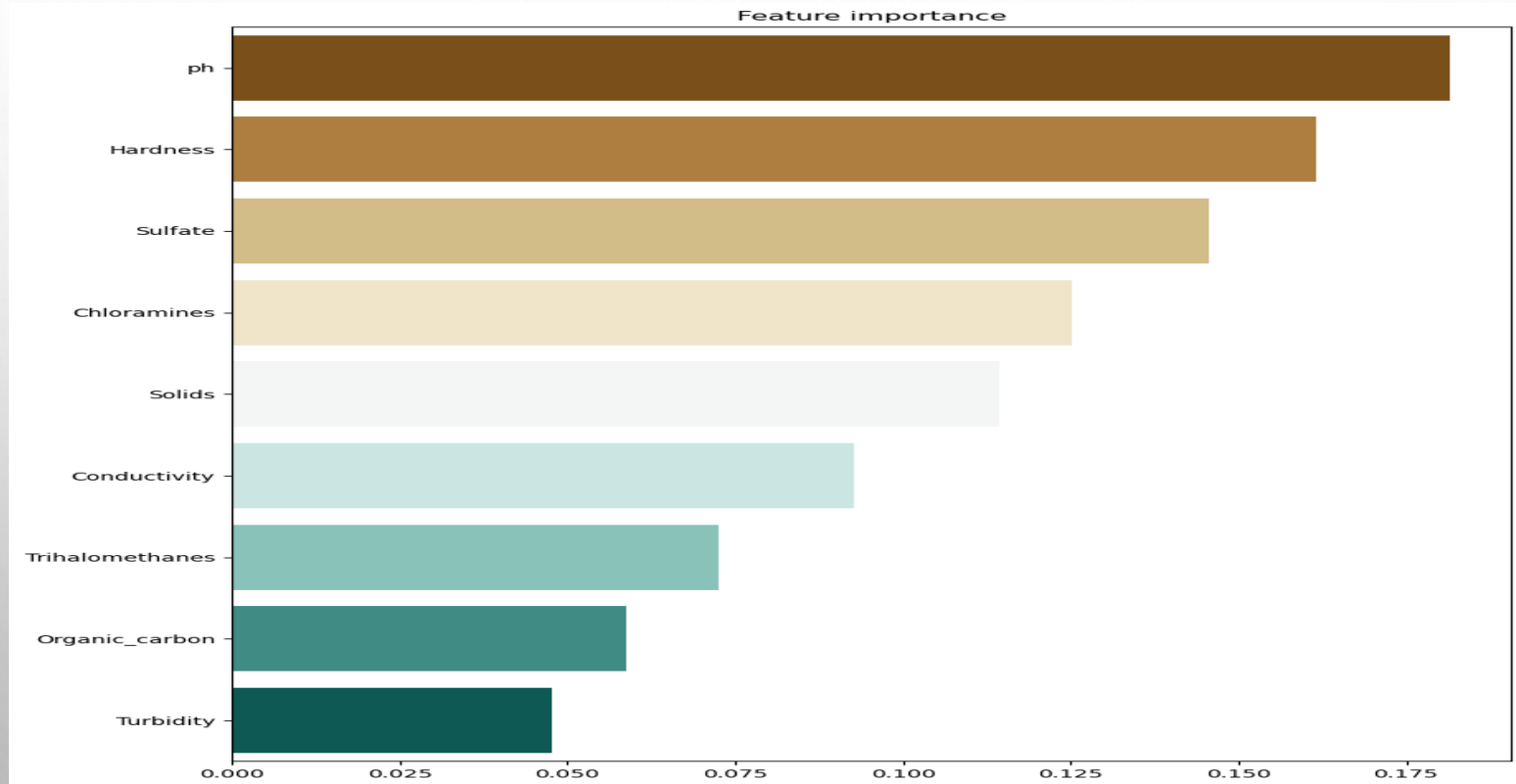

Checking & Treatment of Outliers



TREAT OUTLIERS
USING
WINSORIZING
TECHNIQUE

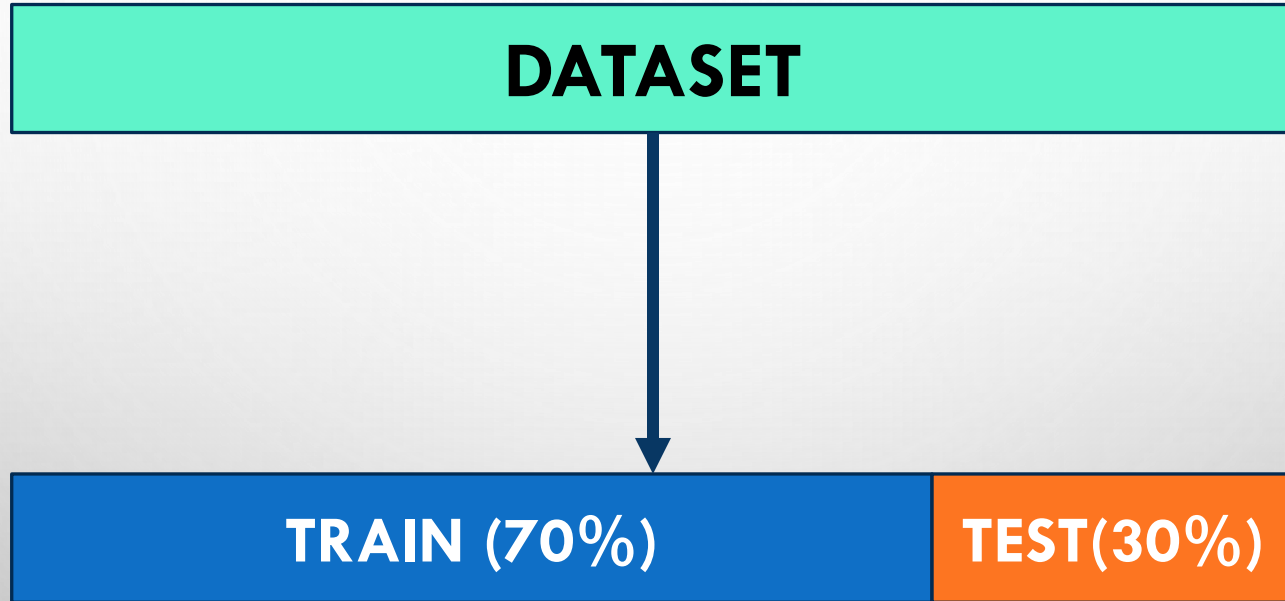


Important Features/Variables

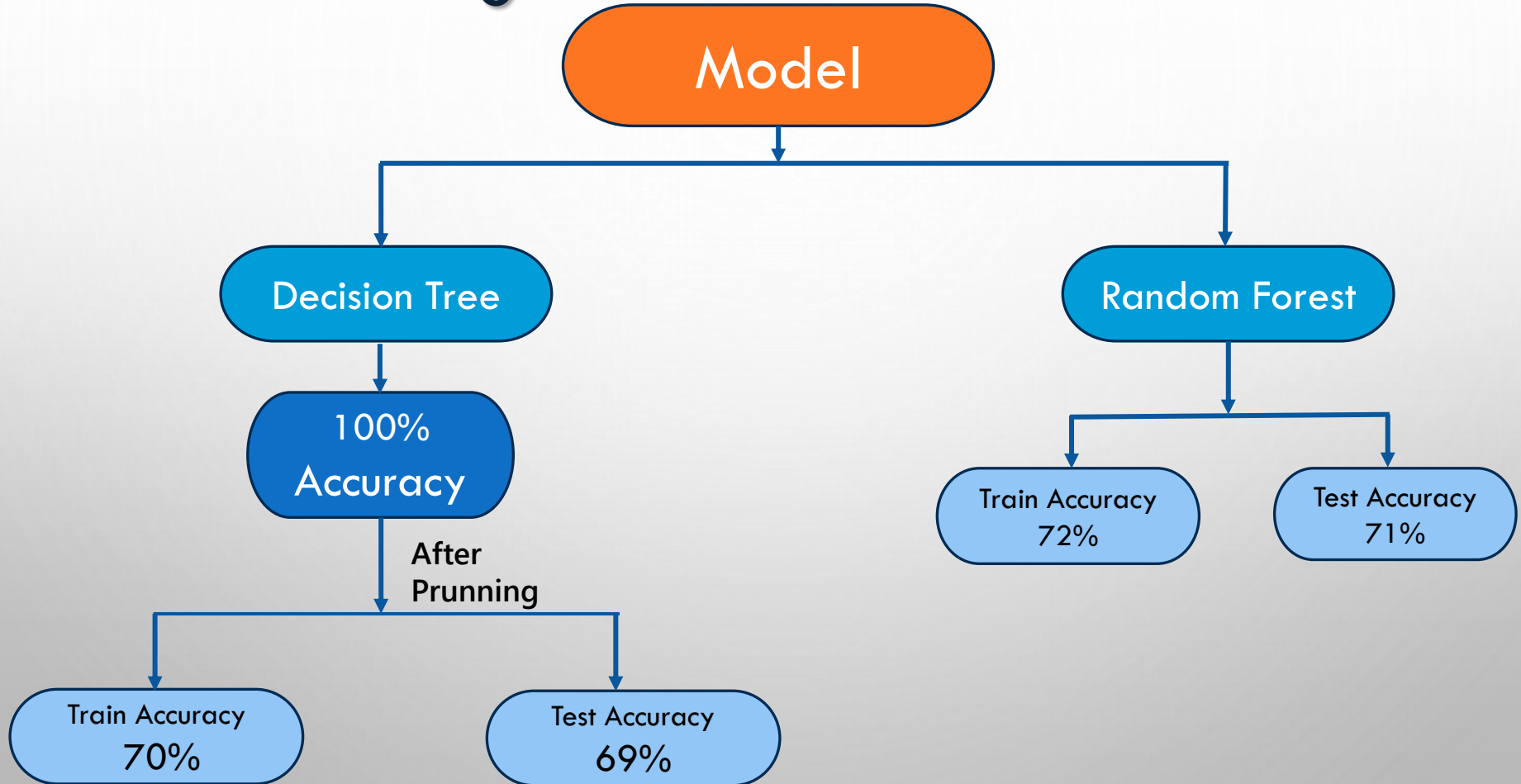


ph>hardness>sulfate>Chloramines>Solids>Conductivity>Trihalomethanes>Organic Carbon>Turbidity

SPLITTING : DIVIDING DATA INTO TRAIN AND TEST



Model Building



Model Selection



Sr No.	Model	Accuracy(Train)	Accuracy(Test)
1.	Decision Tree	70%	69%
2.	Random Forest	72%	71%

❖ Here we select Random Forest as a Best model with High accuracy

Conclusion

❖ The Important Variables In Our Dataset Is Ph, Sulfate, Chloramines, solids.

❖ Water With Ph Value 6.5-8.5, Sulfate Content- 250-500 Mg/L, Chloramines Content- 4 Ppm, Solids(tds)- 500=100 Mg/L Is Drinkable(potable Water).

❖ VARIOUS METHODS TO TREAT WATER ARE:-

1. Coagulation And Flocculation
2. Sedimentation
3. Filtration
4. Disinfection

**Thank
You**