DM

1. Total no of cases - $85+2+4+9 = 100$

True positive - 85
False positive - 4
True Negative - 9
False Negative - 2

$$\text{Accuracy} = \frac{Tp + TN}{\text{total}} = \frac{85+9}{100} = 0.94$$

$$\text{precision} = \frac{TP}{(TP+FP)} = \frac{85}{85+4} = 0.95$$

$$\text{sensitivity} = \frac{TP}{TP+FN} = \frac{85}{85+2} = 0.97$$

$$\text{F-measure} = \frac{2}{\frac{1}{\text{precision}} + \frac{1}{\text{sensitivity}}} = \frac{2}{\frac{1}{0.95} + \frac{1}{0.97}} = 0.96$$

$$\text{Specificity} = \frac{TN}{TN+FP} = \frac{9}{9+4} = 0.692$$

2) Eucledian distance $= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$

Let the value of $k=3$ then according to dataset nearest members are (160, 60, M), (163, 60, M), (163, 61, m) so T-shirt size of customer (161, 61) is M.

if $k=5$ then nearest members are (160, 60, M), (163, 61, M) (163, 60, M) (160, 59, M), (160, 64, L). as the masority voting approach no of M is max so the customers of

hight 161 cms and 61 kgs get t-shirt size of M

Hight (cms)	weight (kgs)	T-shirt size	distance from (161, 61)
158	58	M	4.24
158	59	M	3.6
158	63	M	3.6
160	59	M	2.2
160	60	M	1.4
163	60	M	2.2
163	61	M	2.0
160	64	L	3.16
163	64	L	3.6
165	61	L	4.0
165	62	L	4.1
165	65	L	5.65
168	62	L	7.07
168	63	L	7.28
168	66	L	8.6
170	63	L	9.21
170	64	L	9.48
170	68	L	11.40

3) Data Warehouse :—

Data warehouse is a central repository system in which business store valuable information, such as customer and sales data, for analytics and reporting purpose.

- **Data wersehouse benefits :—**

(i) Provide a stable centralized repository for large amounts of historical data.

(ii) Increse a business's overall return on investment.

(iii) Improve data quality.

(iv) Enhance BI performance and capabilities by drawing on multiple sources.

(v) Provide acess to historical data busines-wide.

(vi) Use AI and machine learning to improve business analytics.

- **key components :—**

(i) **Central Database :—**

If data worehouse integrate data from multiple sources and external data source.

(ii) ~~ETL~~

(ii) ETL :—

This refers to the process of extracting data from multiple source, transforming it into a consistent format and loading it into the data warehouse.

(iii) Metadata :—

It provide the data stored in data warehouse. Including data definitions, linage and transformation rules.

(iv) Access controls :—

Access tools allows users to interact with the data in data werehouse.