a. Sebuah hospital perlu mengenalpasti simptom yang di alami pesakit-pesakit nya sebagai ‘SEVERE’, ‘MILD’ or ‘MODERATE’. Apabila menulis satu program bagi menganalisa data di hospital itu, anda mestilah menetapkan susunan simptom yang mana ‘SEVERE’ adalah lebih tinggi daripada ‘MODERATE’, dan ‘MODERATE’ adalah lebih tinggi daripada ‘MILD’.

A hospital need to identify its patients’ symptoms as ‘SEVERE’, ‘MILD’ or ‘MODERATE’. When writing a program to analyse data at the hospital, you must set the order of the symptoms such that ‘SEVERE’ is higher than ‘MODERATE’, and ‘MODERATE’ is higher than ‘MILD’.

i) Apakah jenis data untuk simptom ?  
What is the data type for symptoms?  
  
Answer: Logical

(1 markah/ marks)

ii) Apakah struktur data R yang paling sesuai bagi menyimpan data simptom? Jelaskan pilihan anda.

What is the most appropriate R data structure to use to store the symptoms data? Explain your choice.

Using data frame as it is able to store different variables in different columns

(2 markah/ marks)

iii) Tulis kod-kod R menunjukkan bahawa anda telah mengumpukkan dan menyusun ‘SEVERE’, ‘MILD’ or ‘MODERATE’ kepada satu objek bernama “Psymptom”. Kemudian uji objek tersebut bagi memastikan nilai mana adalah kurang daripada “SEVERE”. Output anda akan dapat adalah dalam bentuk TRUE atau FALSE bergantung pada syarat tersebut.

Write the R codes to show that you have assigned and ordered ‘SEVERE’, ‘MILD’ and ‘MODERATE’ to an object called “Psymptom”. Then test the object to check which of the values are less than ‘SEVERE’. Your result should return values TRUE or FALSE depending on the condition.

Psymptom <- c ("SEVERE","MILD","MODERATE")  
Psymptomvalue <- c (10,7,5)

Psym.data <- data.frame(Psymptom, Psymptomvalue)

If(Psym.data $ Psymptomvalue < Psym.data $ Psymptomvalue(Psymptom(“SEVERE”))){

Print(Psymptom)

}

(2 markah/ marks)

b. Kanser payudara boleh dirawat terutamanya jika dikesan awal. Automasi boleh membantu meningkatkan kecekapan proses pengesanan. Algoritma k-NN boleh diaplikasikan ke atas ukuran sel-sel biopsi payudara yang tidak normal bagi mengklasifikasi data tersebut sebagai Benign atau Malignant. Benign bermaksud sel-sel adalah bukan kanser, dan Malignant bermaksud sel-sel adalah kanser.

Breast cancer can be treated especially if detected early. Automation can help increase the efficiency of the detection process. k-NN algorithm can be applied to the measurements of biopsied cells of abnormal breast masses to classify the data as either Benign or Malignant. Benign means that the cells are not cancerous, and Malignant means that the cells are cancerous.

i) Setelah data dikumpul, dan dibersihkan, jelaskan langkah-langkah seterusnya dan aktviti-aktiviti yang terlibat.

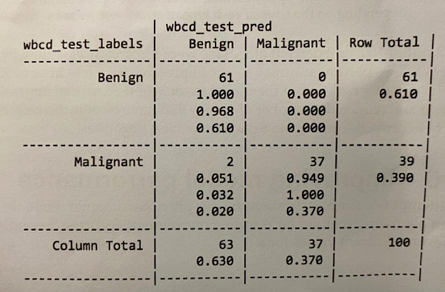
After data has been collected, and cleaned, explain the next step(s) and the activities involved.

Answers:

* Data will be analysed where data will be transformed and model to discover new useful information
* Data will be presented where data will be visualize using table and graphs for better understanding and pattern detection.

(2 markah/ marks)

ii) Analisis menghasilkan jadual di bawah. Tafsirkan jadual tersebut.   
The analysis provides the table below. Interpret the table.



The top-left cell (intersection Benign- Benign) indicates that 61/100 is benign correctly identified by k-NN algorithm.  
The bottom-right cell (intersection Malignant- Malignant) indicates that 37/100 is Malignant correctly identified by k-NN algorithm.  
The cell with 3 value (intersection Malignant- Benign) is the cases where k-NN predicted benign but was actually malignant.