I] Identification of the problem statement:-

1. **Domain Selection** - Machine Learning

   Reason:-
   > Since both input and output data's have numerical values.

2. **Learning Type** – Supervised Learning

   Reason:-
   > Both input and output data's are given in the problem statement.

3. **Implementation Phase:**
   - i)   Data Preprocessing – Input and output data separation
   - ii)   Assigning variables to input and output data
   - iii) Creating and checking Training model set by using various MLR Algorithms and standardization Techniques.
   - iv)   Evaluation by using R square value
   - v)   Saving the best model

4. **Deployment Phase:**
   - i)   Loading the best model
   - ii)   Prediction by giving the desired inputs and finding the output.
   - iii) Call to action

II] Basic Information of Data Set:-

    i)   Rows – 1338, Columns – 6
    ii)   Input columns – 5
    iii)   Output column – 1

III] Pre-processing Method:-

Answer:- We have 2 Categorical Input columns in the given dataset. MLR Algorithms cannot support these data, it supports only numeric data. So by analyzing the giving categorical input it falls under Nominal data, and by using One-Hot Encoding Algorithm we can convert those columns into corresponding numeric data. After converting we can use it for our Model Creation.

IV] Conclusion:-

For this Data set, SVM and Random forest Algorithm gives the same highest
R2 value =0.87