**Statistical Learning Lab**

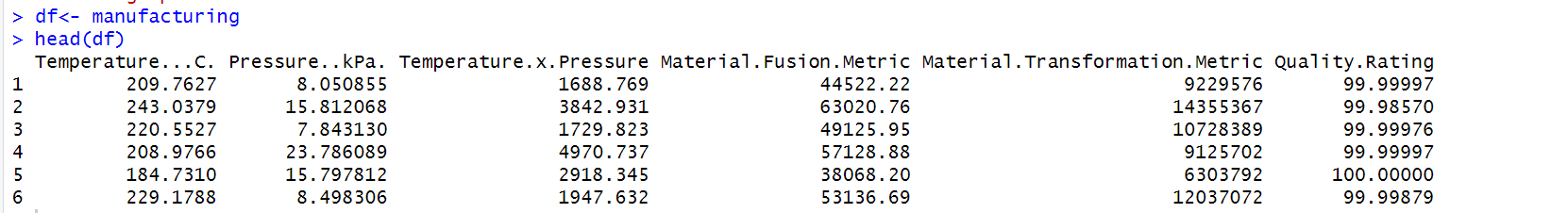
**Assignment - 4**

**Cross-validation and Bootstrapping**

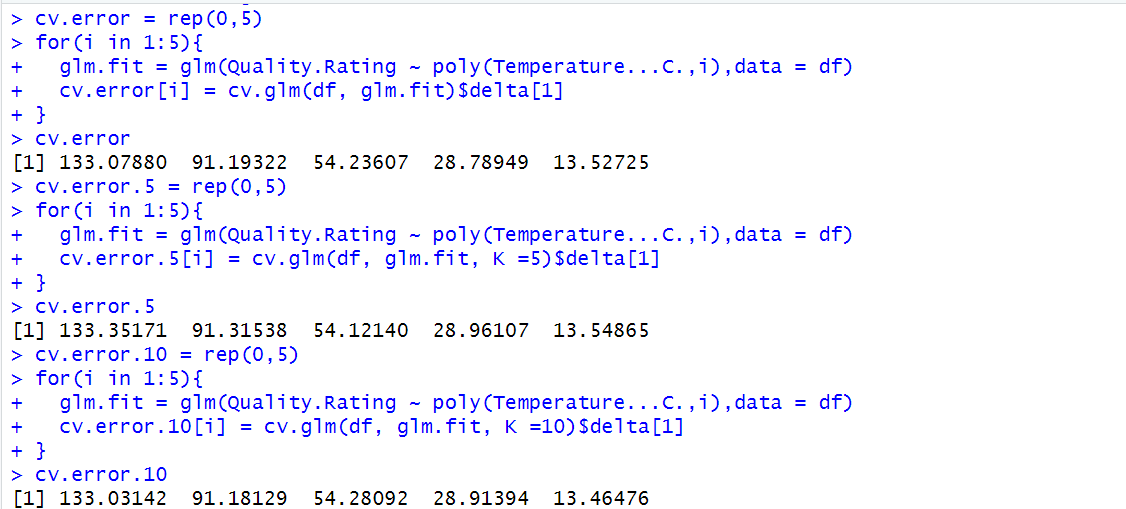
**Name: Sunny Kumar , Roll No: 22IM10040**

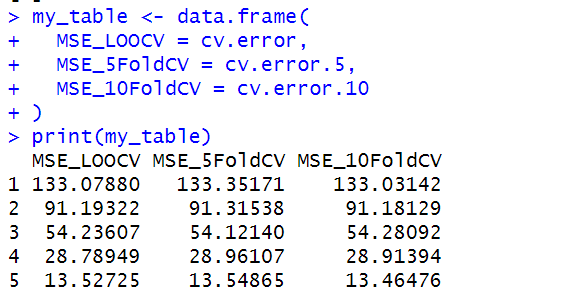
**Show the code snippets and the corresponding output for the following:**

1. **Load the dataset “manufacturing.csv”. Display first few rows of the dataset. Take “Quality Rating” as response variable.**

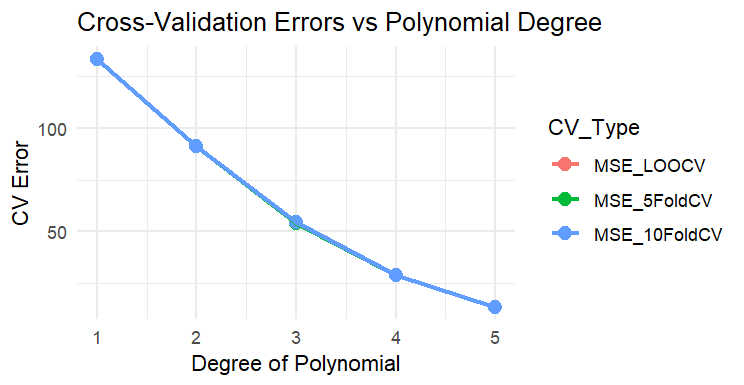


1. **Fit polynomial models between Quality ~ Temp. Vary the degree of polynomial on temperature from 1 to 5 (temp, temp^2, temp^3 etc.). Perform LOOCV, k-fold CV for k=5 and 10 and compare the cross-validation MSE errors for different degrees of polynomials. Create a table showing the CV errors for different degree of polynomials and for different CV techniques. Plot the results. Discuss which degree of polynomial is preferable.**





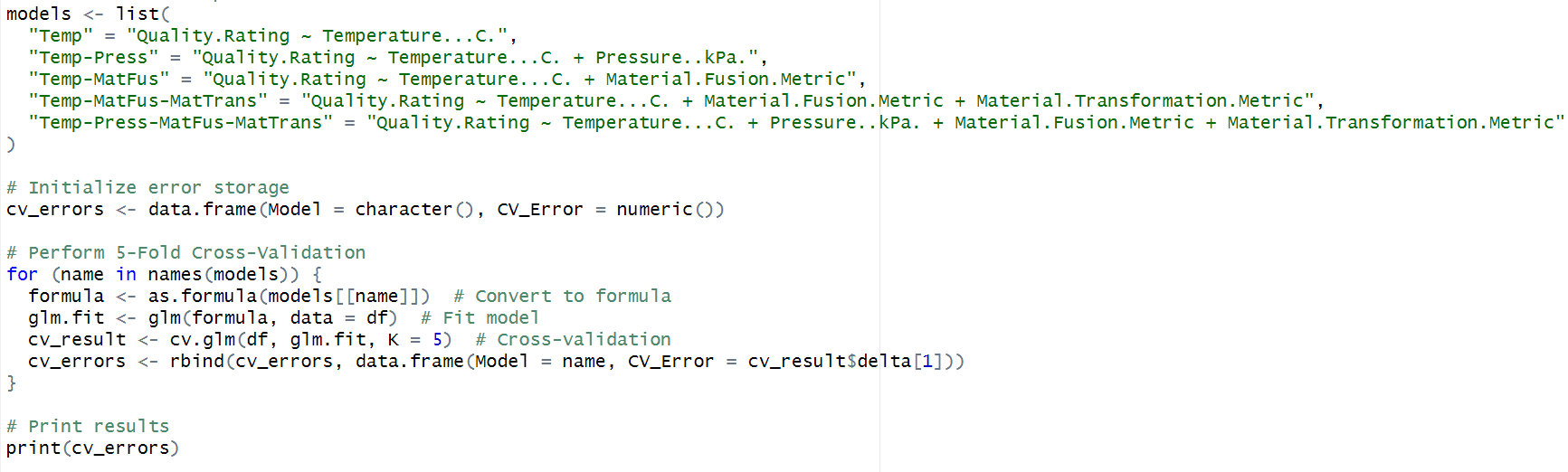
We can clearly see from here that for polynomial degree 1 and 2 , 10- fold Cross Validation gave the least error in these three models and for 3rd degree polynomial , 5-fold CV gives the least error and for 4degree polynomial , LOOCV gives the least error and for 5degree polynomial , 10-fold Cross Validation gives the least error . Graphs are shown below.

From

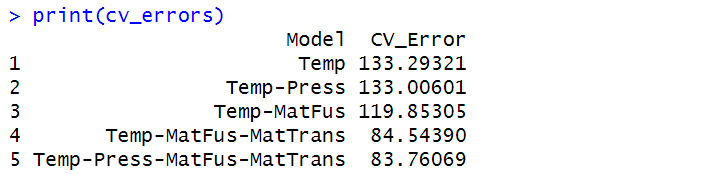
From the graphs , we can clearly say that , Polynomial of degree 5 is favourable because it is giving the least error .

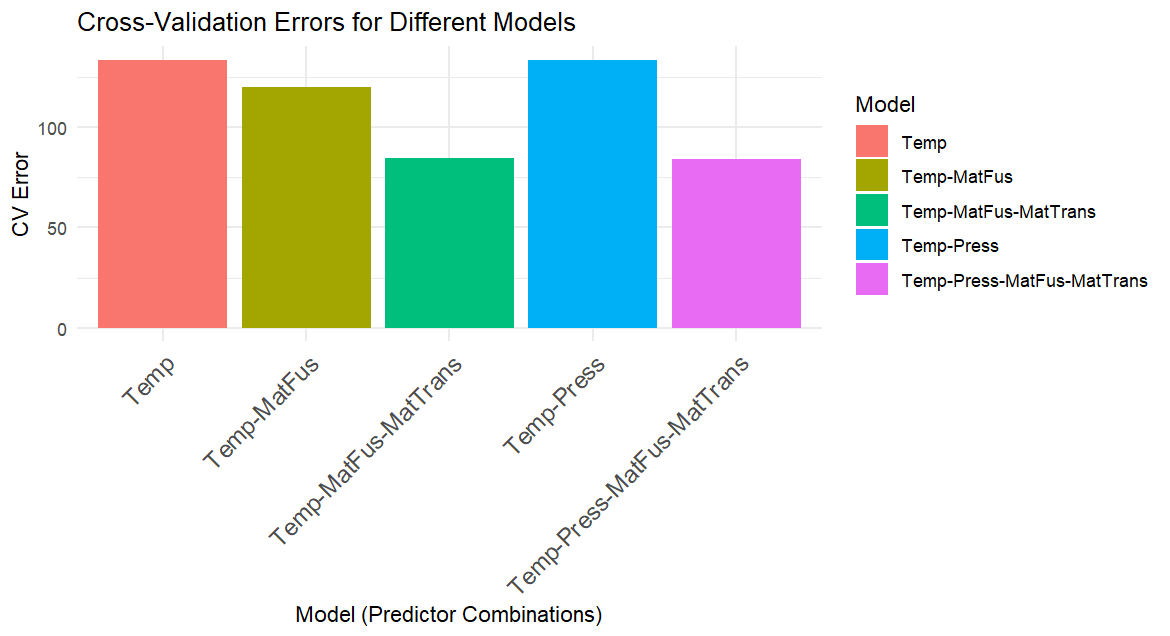
1. **Perform the analysis in problem no. 2, but this time, fit linear models with different combination of X variables, without interaction. Discuss which model is most preferable based on the cross-validation results. Plot the results and on X-axis labels, provide the X-variable combinations used in the model, e.g. (temp, temp-press, temp-matfus, temp-matfus-mattr etc.)**

Code:



Output :





From the above graph we can conclude that the combination of Temperature...C. + Pressure..kPa. + Material.Fusion.Metric + Material.Transformation.Metric" gives the least error .

1. **Generate 50 random numbers from Normal Distribution . Now create 100 bootstrap samples with 20 datapoints each, with replacement. Estimate the mean and variance of the population from the bootstrap samples.**

