Information

Optical Recognition of Handwritten Digits dataset

-Successfully implemented Naïve Bayes classification algorithm obtained an accuracy of 89.04%.
-Used naiveBayes library to implement the algorithm
-Tried to use another e1071 and caret library for implementing the same.
-Challenges faced: It took some time to understand the different parameters used in the naiveBayes and e1071 function.
- It took less time to create the model. It was really fast for creating the model. Therefore, we haven't used high performance cluster for this dataset.

Amazon reviews sentiment analysis dataset

-We are not able to get good libraries which performs the sentiment score calculation of each review in the dataset.
-Since sentiment calculation in python is giving better results as compared to R, we are calculating the sentiment score by using **Vader Sentiment library.**
-Still the accuracy was not improving much in the case of amazon dataset by using Naïve Bayes classification.
- Maximum accuracy obtained was 58.56%
- Challenging part: Misclassification was happening with reviews of rating 2,3,4. We haven't used High Performance cluster for this dataset also since it was working fine with laptop's memory and the execution time for creating the model is very less.