Programming: Question

In this problem, we will use university application data for the purpose of admission classification. Fine data_train and data_test on canvas.

(a) Data Pre-processing: Create a binary label based on the column "*Chance of Admit*". Convert any values bigger than the median to 1 and 0 otherwise. Split the data into training and validation sets. You can use a 80-20 split.

(b) Model Initialization: Initialize 3 different SVM models with the following kernels.

1. Linear kernel
2. RBF kernel
3. Polynomial (degree 3) kernel

(c) Feature Selection and Model Training: Train each SVM Model above with the following feature combinations to predict admission.

1. CGPA and SOP
2. CGPA and GRE Score
3. SOP and LOR
4. LOR and GRE Score

(d) Support Vectors: What are the support vectors for each model and feature combination? How many support vectors does each class have in each case?

(e) Result Visualization: For each kernel - input combination, visualize the predictions on the training set. Color code the points by their labels.

(f) Result Analysis: Just by looking at the figures you generated, answer this question: Which features + kernel combinations give you the best result? Validate your model on the validation set and find the best performing combination with respect to accuracy.

(g) Inference: Use the best model you found in the previous step to predict the label of the test data. Save the prediction in a csv file "FirstName LastName preds.csv"