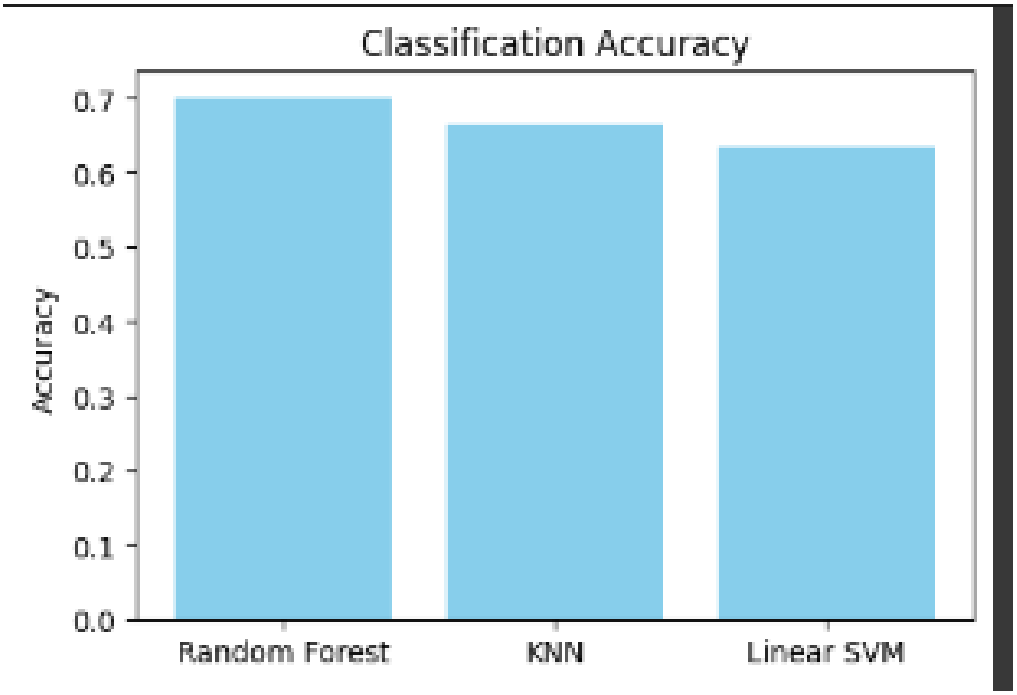
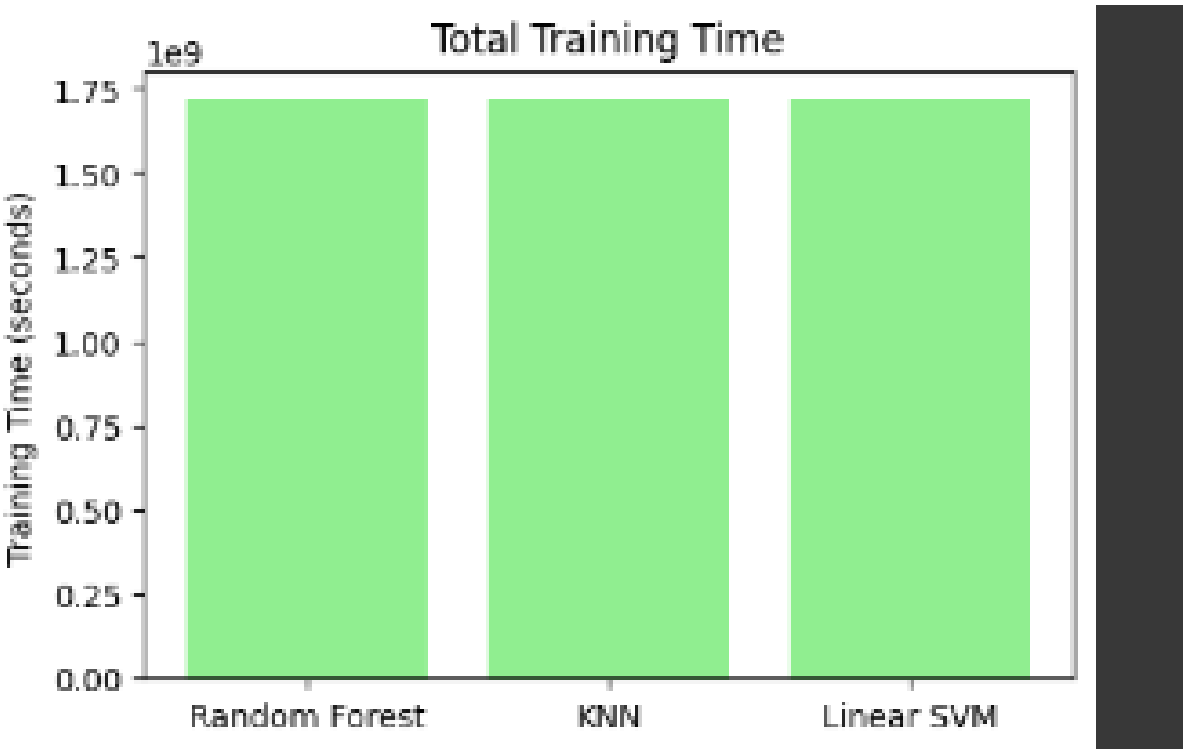
Pattern Recognition
Project(1): Apartment Rent Prediction
Milestone 2 Report
Team_ID: CS_29

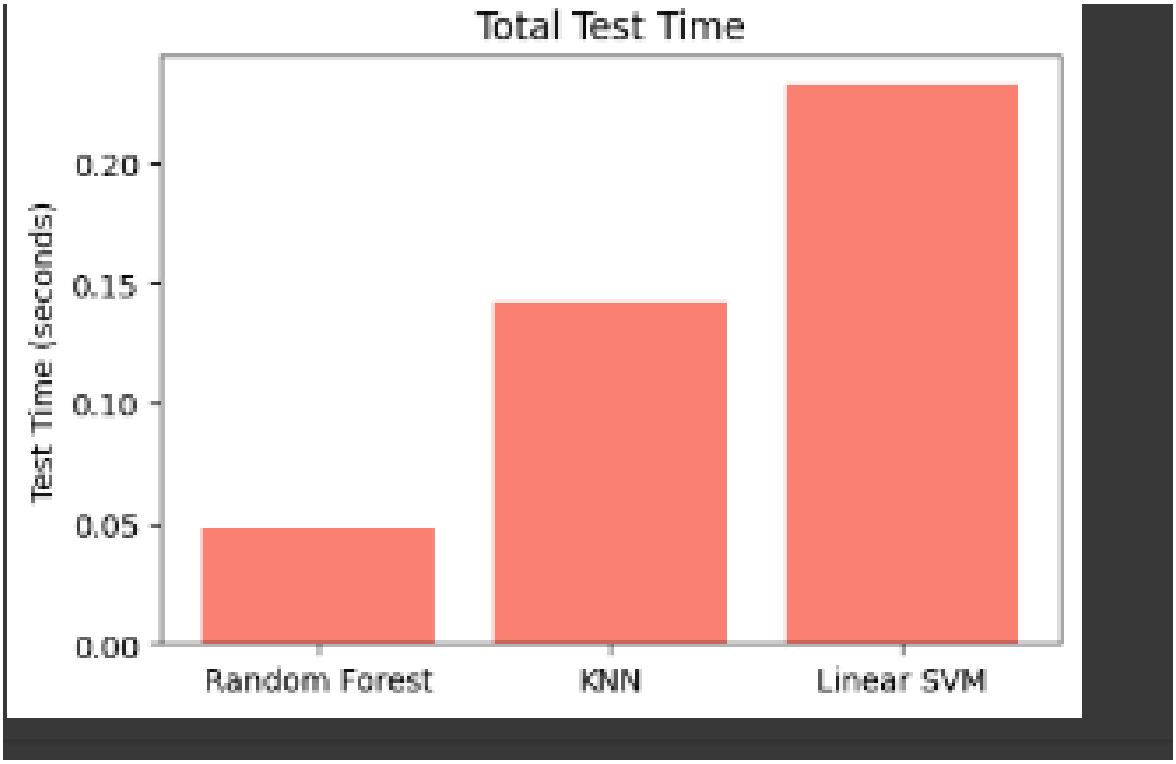
لجين عماد محمد صلاح	2021170413
مارينا نبيل كامل بنيامين	2021170432
جنى محمود مرسى محمود	2021170143
لطيفه احمد حنفى عبدالفتاح	2021170414
مريم مصطفى عبد الاه	2021170514

- **Classification Models**
  We find that these models with best accuracy :
  Random forest accuracy: 0.7
  KNN accuracy: 0.6644444444444444
  Linear SVM accuracy: 0.6333333333333333

- And these are their bar graphs :




Total Training Time
1e9
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Training Time (seconds)
Random Forest
KNN
Linear SVM


Total Test Time
0.20
0.15
0.10
0.05
0.00
Test Time (seconds)
Random Forest
KNN
Linear SVM

- **Feature Selection:**

  We use ‘Anova’ as this is suitable for classification where understanding the relationship between features and the target variable is essential for accurate predictions. ANOVA is commonly used in scenarios where dependent variable (target)(RentCategory) is categorical.

- **Hyperparameter tuning**

in this section we will show how hyperparameter affect the model performance

1- Svm model

we found that hyperparameter are

- Kernel function
- C (regularization parameter)
- Degree of polynomial kernel

Kernel function

In first we use a `linear` kernel, we also considering other kernels like `rbf` (Radial Basis Function), `poly` (polynomial), or `sigmoid why?` might provide better performance with non-linear data distributions.**Effect**.

Results from that we found that applying `linear Svm get training accuracy =` `0.699`

applying polynomial `Svm get training accuracy =` `0.71`

applying rbf `Svm get training accuracy = 0.73`

C (regularization parameter)

If you set C to be a low value the SVM classifier give larger number of misclassifications. When C is set to a high value the classifier try to minimize the misclassifications.

Degree of polynomial kernel

Higher Degrees: As the degree increases, the polynomial kernel allows the SVM to fit more complex patterns **Lower Degrees**: A lower degree (such as 2 or 3) will produce smoother, simpler decision boundaries. So we choose degree 3

2-Randomforest Model

we found that hyperparameter are

- **n_estimators**
- **random_state**

n_estimators

Number of trees in the forest. We Increase this number to 100 to improve model performance because the ensemble's predictions become more stable

random_state

we set random_state to 42 to ensure that every time we run this script, the `Randomforest classifier` will behave identically, given the same inputs and parameters

3-KNN Model

we found that hyperparameter

- **n_neighbors**

n_neighbors

we use value of n =5 we find it is suitable for our complex decision boundaries and after often requires experimentation and careful consideration.

Conclusion

We predict that **Svm(rbf)** will be the best model in accuracy , and we found that **Random Forest** model is the best model in accuracy

- for **Random Forest**:
- Model Test Accuracy: 0.7283333333

- For **Knn**
- Model Test Accuracy: 0.69611111111

- For **logistic regression**:
- Model Test Accuracy: 0.6505555555

- For **Svm(rbf)**
- Model Test Accuracy: 0.5055555555

- For **Svm(poly)**
- Model Test Accuracy: 0.6511111111

- For **Svm(linear)**
- Model Test Accuracy: 0.6511111111

In conclusion

of this phase we learn that Feature Selection process differ in this phase than the previous (regression)

also we learn that shouldn’t make assumption before analysis of the data well and clearly .