Electronic Device Rating Prediction

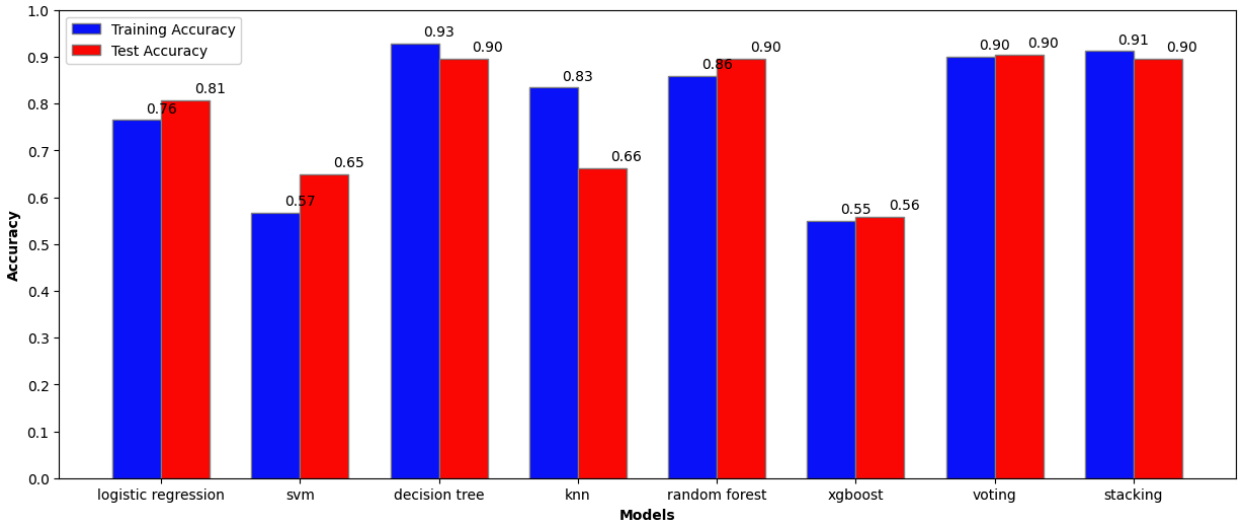
Report

Team: CS_27

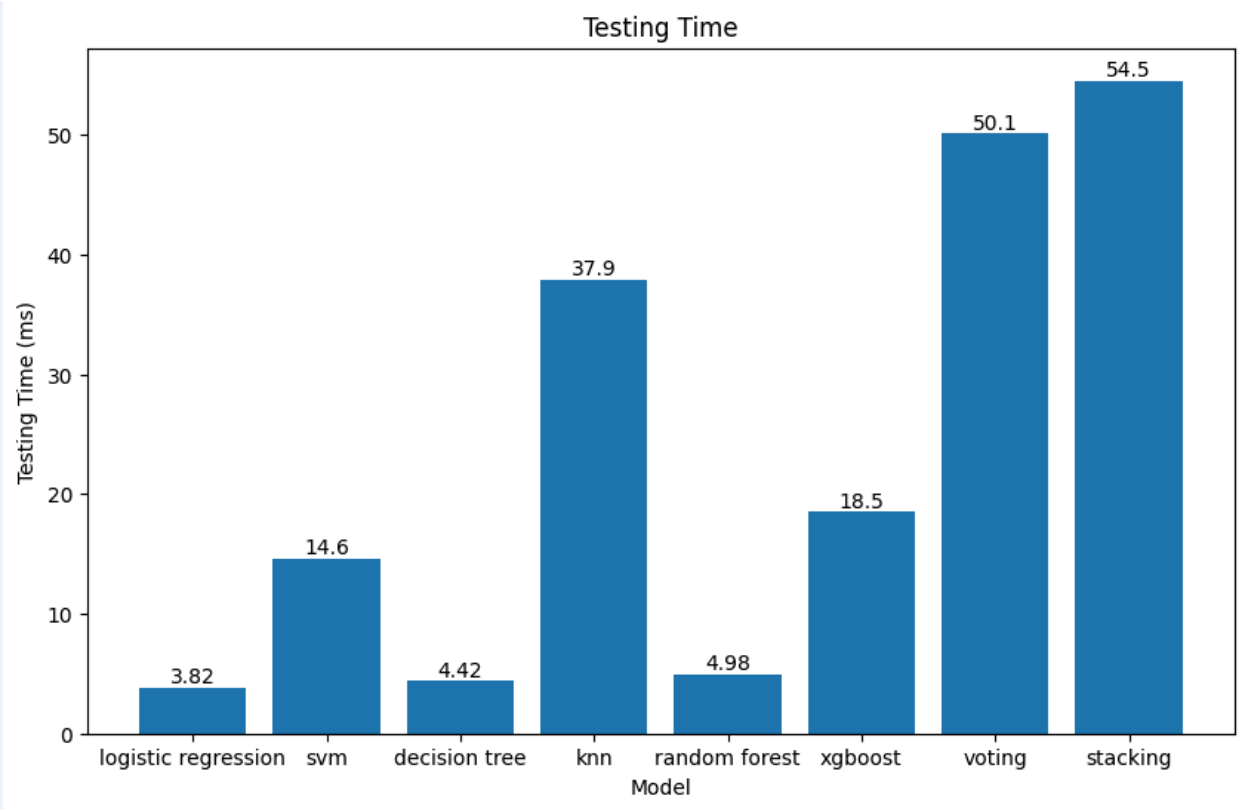
sec 8	2021170642	يوسف محمد سيد محمد	
sec 6	2021170480	محمد عوض طلعت محمد	
sec 8	2021170654	يوسف وجيه وديع ناشد	
sec 8	2021170649	يوسف مصطفى محمد يوسف	
sec 6	2021170483	محمد مبروك فكرى عبدالفتاح	
sec 8	2021170641	يوسف عمرو احمد زهران	

Bar Graphs

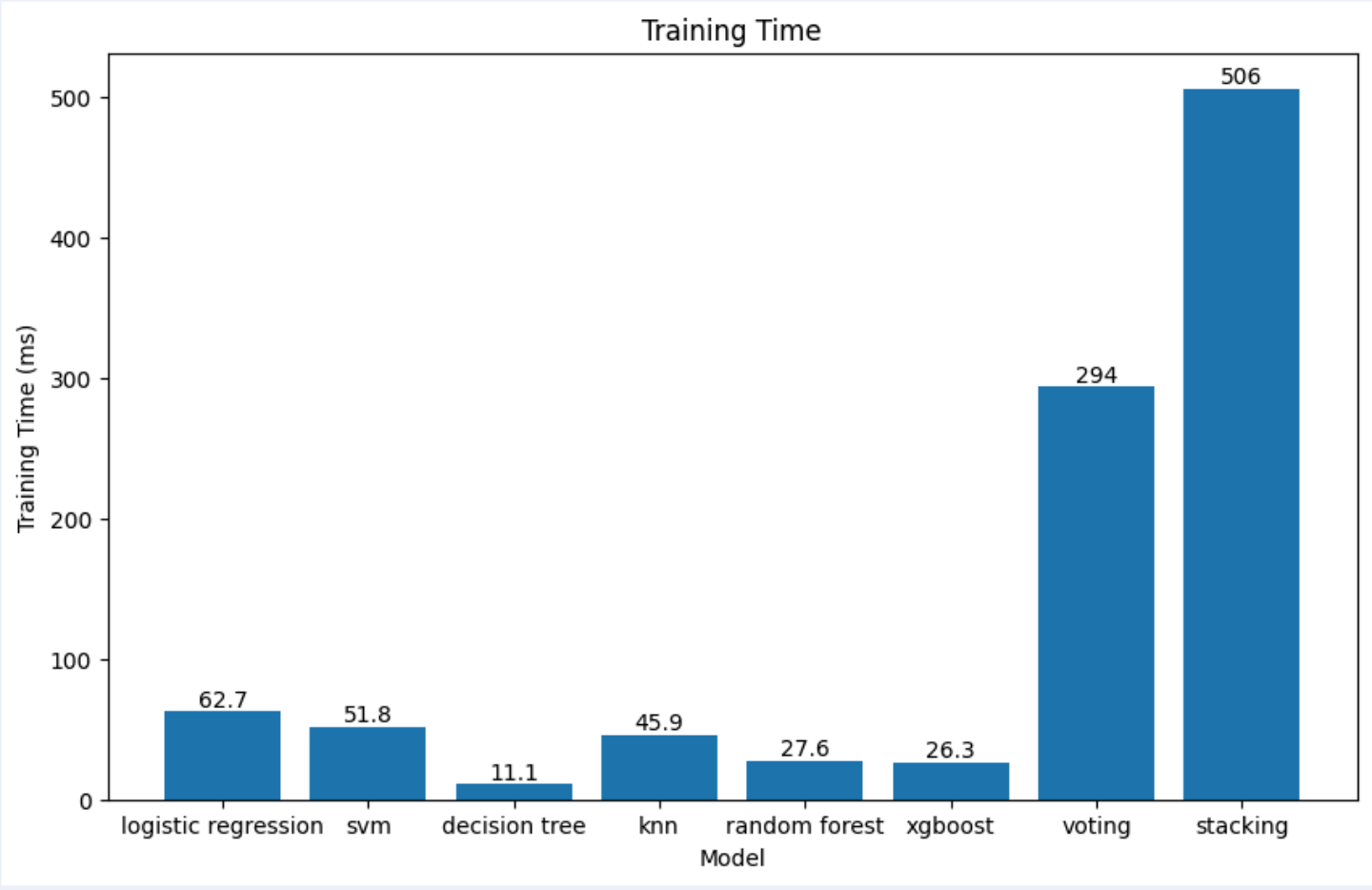
Training & Testing Accuracy



Testing Models Time



Training Models Time



Feature Selection

Chi-Squared

since the input variables and output variables are categorical, Chi-Squared Feature Selection was the best choice and gave the highest accuracy when setting the K hyper parameter to 22

Hyperparameter Tuning

Decision Tree

- when max_depth = 5 – the best
    - Training MSE: 0.0726643598615917
    - Testing MSE: 0.10344827586206896
    - Training Accuracy: 0.9273356401384083
    - Testing Accuracy: 0.896551724137931
- when max_depth = 4
    - Training MSE: 0.0847750865051903
    - Testing MSE: 0.1103448275862069
    - Training Accuracy: 0.9152249134948097
    - Testing Accuracy: 0.8896551724137931
- when max_depth = 10 – overfitting
    - Training MSE: 0.0034602076124567475
    - Testing MSE: 0.10344827586206896
    - Training Accuracy: 0.9965397923875432
    - Testing Accuracy: 0.896551724137931

KNN:

- when n_neighbors = 3– the best
    - Training MSE: 0.16608996539792387
    - Testing MSE: 0.33793103448275863
    - Training Accuracy: 0.8339100346020761
    - Testing Accuracy: 0.6620689655172414
- when n_neighbors = 1 -- overfitting

    - Training MSE: 0.0
    - Testing MSE: 0.27586206896551724
    - Training Accuracy: 1.0
    - Testing Accuracy: 0.7241379310344828

- when n_neighbors = 10

    - Training MSE: 0.2975778546712803
    - Testing MSE: 0.46206896551724136
    - Training Accuracy: 0.7024221453287197
    - Testing Accuracy: 0.5379310344827586

Conclusion:

The project successfully classified electronic device ratings using machine learning models. Among the models tested, the Decision Tree achieved the highest accuracy with efficient training and testing times. The results were challenging because many of them were overfitting and cannot be considered. Overall, the project highlighted the importance of model selection and evaluation in classification tasks.