

The box plot displays the distribution of scores for 14 different models. The y-axis represents the score, ranging from 0.0 to 1.0. The x-axis lists the models: Linear, Logistic, DecisionTree, RandomForest, AdaBoost, GradientBoosting, XGBoost, CatBoost, LightGBM, Stacking, Voting, Ensemble, and a final unlabeled model. For each model, two box plots are shown: a blue one for 'Train_score' and an orange one for 'Test_score'. The 'Train_score' is consistently higher than the 'Test_score' for all models, indicating overfitting. The 'Test_score' is generally higher for the ensemble models (Stacking, Voting, Ensemble) compared to the individual models.

Model	Score Type	Min	Q1	Median	Q3	Max
Linear	Train_score	0.85	0.86	0.87	0.88	0.89
	Test_score	0.75	0.78	0.80	0.83	0.86
Logistic	Train_score	0.88	0.89	0.90	0.91	0.92
	Test_score	0.80	0.82	0.84	0.86	0.88
DecisionTree	Train_score	0.88	0.89	0.90	0.91	0.92
	Test_score	0.78	0.80	0.82	0.85	0.88
RandomForest	Train_score	0.88	0.89	0.90	0.91	0.92
	Test_score	0.80	0.82	0.84	0.86	0.88
AdaBoost	Train_score	0.88	0.89	0.90	0.91	0.92
	Test_score	0.80	0.82	0.84	0.86	0.88
GradientBoosting	Train_score	0.88	0.89	0.90	0.91	0.92
	Test_score	0.80	0.82	0.84	0.86	0.88
XGBoost	Train_score	0.88	0.89	0.90	0.91	0.92
	Test_score	0.80	0.82	0.84	0.86	0.88
CatBoost	Train_score	0.88	0.89	0.90	0.91	0.92
	Test_score	0.80	0.82	0.84	0.86	0.88
LightGBM	Train_score	0.88	0.89	0.90	0.91	0.92
	Test_score	0.80	0.82	0.84	0.86	0.88
Stacking	Train_score	0.88	0.89	0.90	0.91	0.92
	Test_score	0.80	0.82	0.84	0.86	0.88
Voting	Train_score	0.88	0.89	0.90	0.91	0.92
	Test_score	0.80	0.82	0.84	0.86	0.88
Ensemble	Train_score	0.88	0.89	0.90	0.91	0.92
	Test_score	0.80	0.82	0.84	0.86	0.88
Unlabeled	Train_score	0.88	0.89	0.90	0.91	0.92
	Test_score	0.80	0.82	0.84	0.86	0.88

Predictor left out