CLICK THROUGH RATE PREDICTION ASSIGNMENT
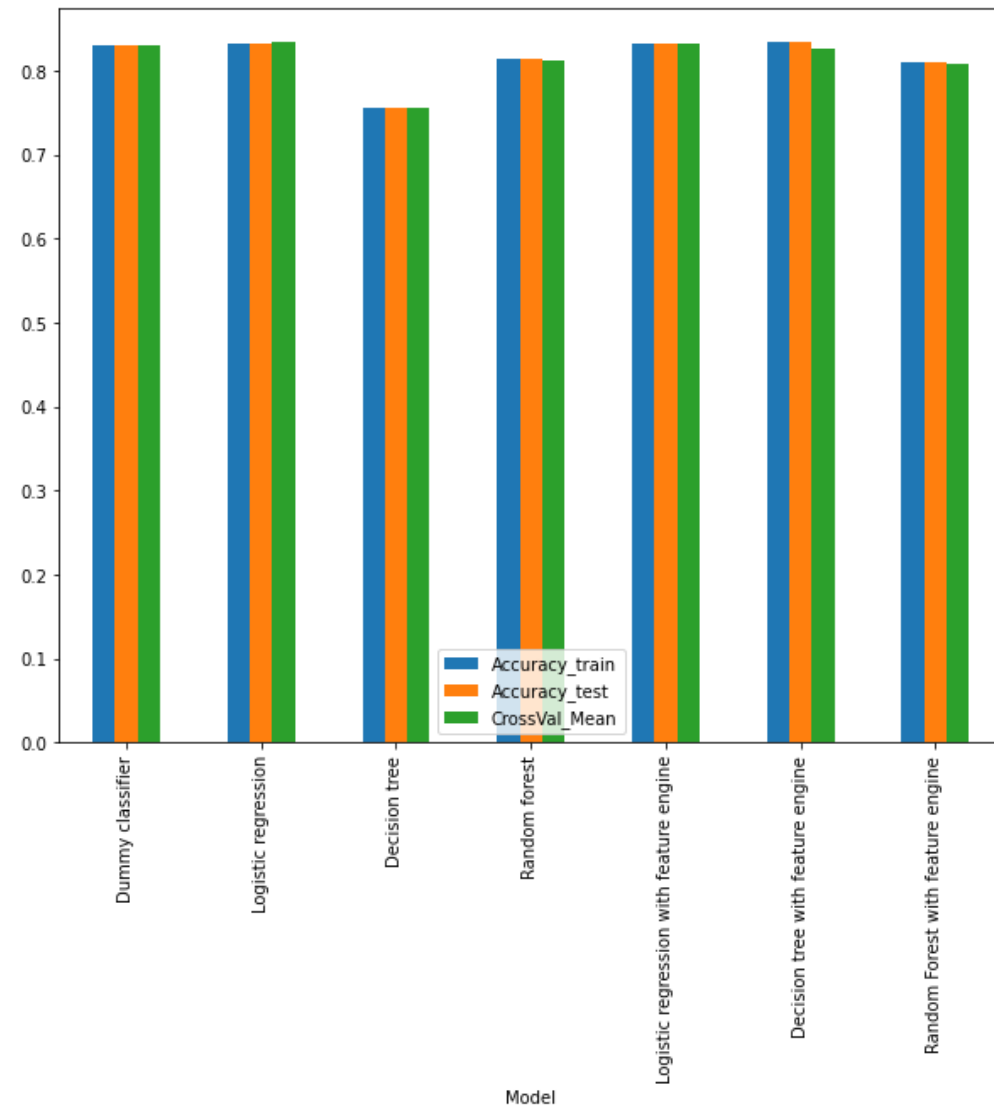
Saptarshi Saikia

Evaluation Metrics of all the Classification methods

Model	Accuracy_train	recall_train	precision_train	Accuracy_test	recall_test	precision_test	CrossVal_Mean	CrossVal1	CrossVal2	CrossVal3	CrossVal4	CrossVal5
Dummy classifier	0.830200	NaN	0.000000	0.830200	NaN	0.000000	0.830198	0.830214	0.830214	0.830214	0.830143	0.830202
Logistic regression	0.832333	0.538462	0.087947	0.832333	0.538462	0.087947	0.833483	0.833071	0.834000	0.833571	0.833357	0.833417
Decision tree	0.755767	0.287049	0.295446	0.755767	0.287049	0.295446	0.755682	0.757429	0.758429	0.748500	0.759143	0.754911
Random forest	0.813400	0.399119	0.195720	0.813400	0.399119	0.195720	0.813297	0.812500	0.815786	0.812357	0.812286	0.813558
Logistic regression with feature engine	0.832367	0.539016	0.088143	0.832367	0.539016	0.088143	0.833398	0.832571	0.834000	0.833714	0.833214	0.833488
Decision tree with feature engine	0.833933	0.541854	0.142324	0.833933	0.541854	0.142324	0.827383	0.827929	0.828429	0.826286	0.827714	0.826559
Random Forest with feature engine	0.810033	0.386321	0.201806	0.810033	0.386321	0.201806	0.808726	0.810286	0.810214	0.807143	0.807929	0.808058

Here we can observe the evaluation metrics used to judge all the classification models before and after feature selection

Graph of the Evaluation metrics

Here is the same data but in a more picaresque form. We can clearly see that Decision tree with feature selection overall boasts the best results closely followed by Logistic regression with feature selection.

Conclusions

- Decision tree with feature engineering is overall the best method but it is very close in performance to logistic regression with feature selection.
- Banner position of the Ads have high correlation with device type and it's important to structure it properly for high number of clicks.
- Because of the randomized nature of the data picked it's better to do this test with several more times with random clusters.
- The errors made in the model ultimately do point to the fact that its virtually impossible to accurately predict what gets clicked but we can gain an understanding on how to increase the likelihood of getting click.

THANK YOU

Saptarshi Saikia