

Peng Shen

1.

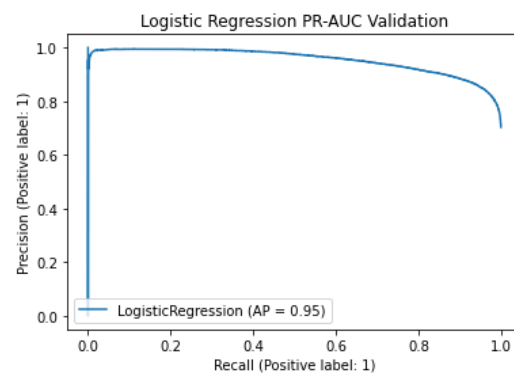
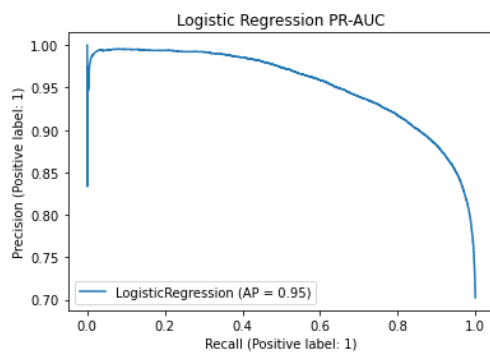
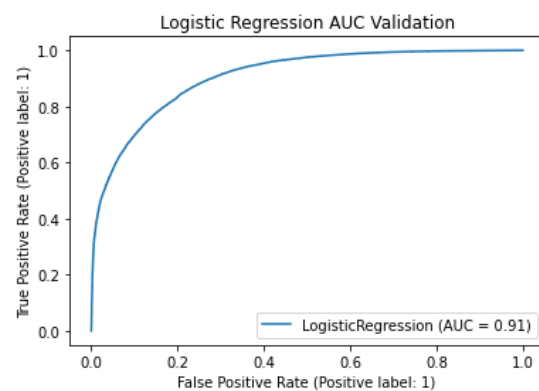
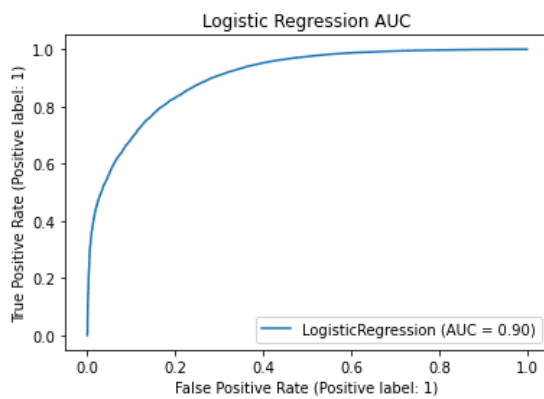
There are a total of 185 features selected. Some of them are 'dep_name', 'esi', 'age', 'gender', 'ethnicity', 'race', 'lang', 'religion', 'maritalstatus', 'employstatus', ... 'meds_vitamins', 'n_surgeries', 'cc_abdominalpain', 'cc_alcoholintoxication', 'cc_alteredmentalstatus', 'cc_chestpain', 'cc_neurologicproblem', 'cc_other', 'cc_psychiatricevaluation', 'cc_shortnessofbreath'.

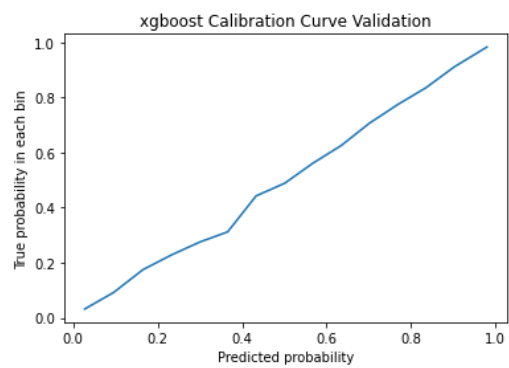
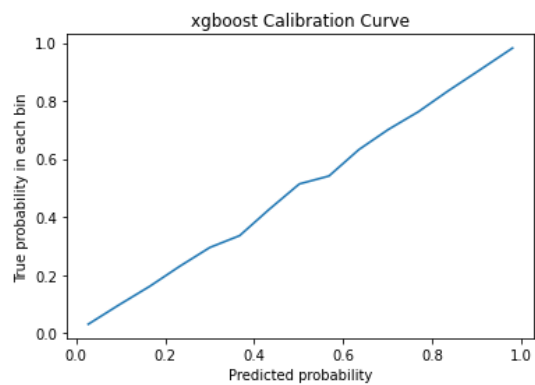
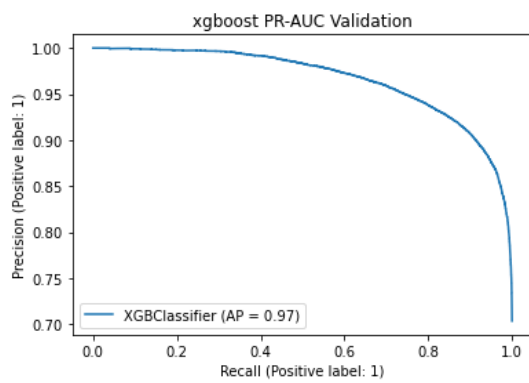
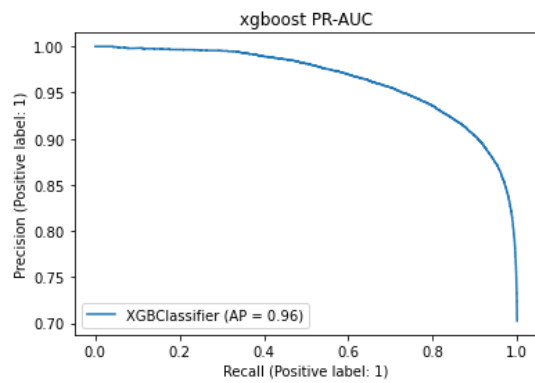
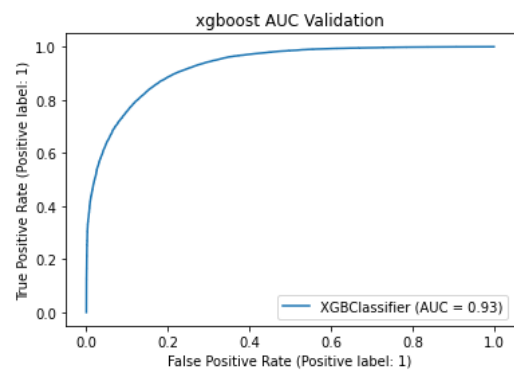
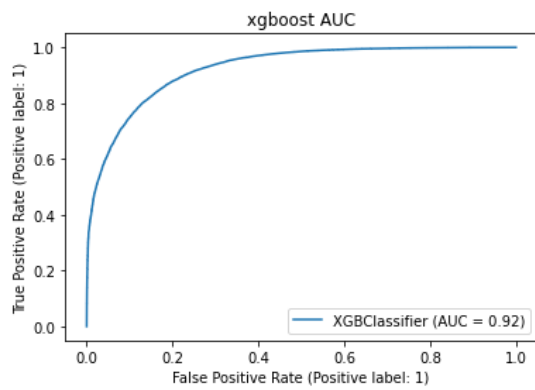
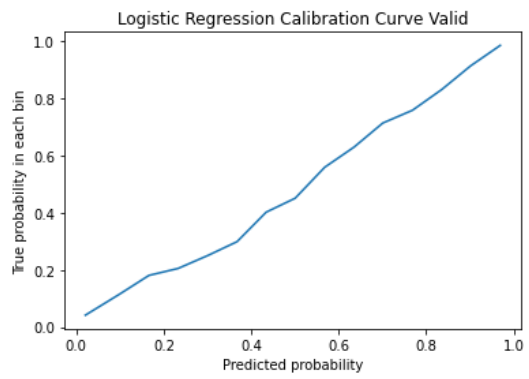
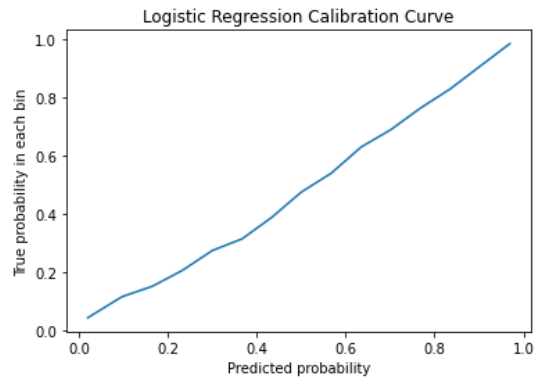
2.

The AUC, PR-AUC for logistic regression in test data are 0.90 and 0.95. The AUC, PR-AUC for logistic regression in validation data are 0.91 and 0.95. The Calibration curve is shown below.

The AUC, PR-AUC for XGBoost in test data are 0.92 and 0.96. The AUC, PR-AUC for logistic regression in validation data are 0.93 and 0.97. The Calibration curve is shown below.

The XGBoost performance is better than logistic regression. The performance of xgboost model for validation dataset is better than in test set, which indicates over-fitting.





3.

The odds ratio of top 10 variables of importance in logistic regression is shown below.

	feature	odds_ratio
1	esi	3.200151
178	cc_alcoholintoxication	1.416240
23	hyperlipidem	1.169736
17	asthma	1.100964
47	rbc_last	1.084050
10	insurance_status	1.074210
21	esophgealidx	1.068772
16	anxietydisorders	1.064619
0	dep_name	1.060855
122	rbc_median	1.060163

The SHAP summary plot of the top 20 variables in XGBoost is shown, along with the force plot and the dependence plot.

In summary plot, patients with higher esi, or lower meds_cardiovascular or lower age is having a higher probability of admission. Vice versa, lower esi, or higher meds_cardiovascular or higher age corresponds to higher probability of discharge.

The force plot of the 5 visits is an interactive plot, thus it would be better to interpret using software.

Looking at the dependence plot of age and systolic blood pressure, besides the information already provided by the summary plot, age and systolic blood pressure are negatively correlated for the patients. Higher age often corresponds to a lower systolic blood pressure.

