

Презентация для кейса компании "Самолет"



Работу выполнил Зотеев А.А.

ИСП-21



самолет

Задача:

На основе больших данных о предыдущем опыте взаимодействия с клиентом, разработать модель, позволяющую прогнозировать вероятность покупки клиентами дополнительных услуг, в частности, приобретения машино-мест в паркинге

Среди клиентов компании необходимо выделить покупателей, наиболее склонных к покупке машино-места



самолет

Проблематика

Ключевую роль в продажах играет эффективная целевая рассылка.

Рассылки позволяют оперативно информировать клиентов об актуальных предложениях и сервисах компании. Однако каждая рассылка сопряжена с издержками:

- Финансовые затраты - на подготовку и доставку сообщений
- Временные затраты маркетологов и продавцов
- Риск раздражения получателей частыми сообщениями



Ход работы



Инструменты

Pandas



Matplotlib



Python



Scikit-learn



[illegible]

1. Анализ данных

```
df_train.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 14456 entries, 0 to 14455  
Columns: 2666 entries, report_date to col2663  
dtypes: float64(2316), int64(3), object(347)  
memory usage: 294.0+ MB
```



2. Предварительная очистка



самолет

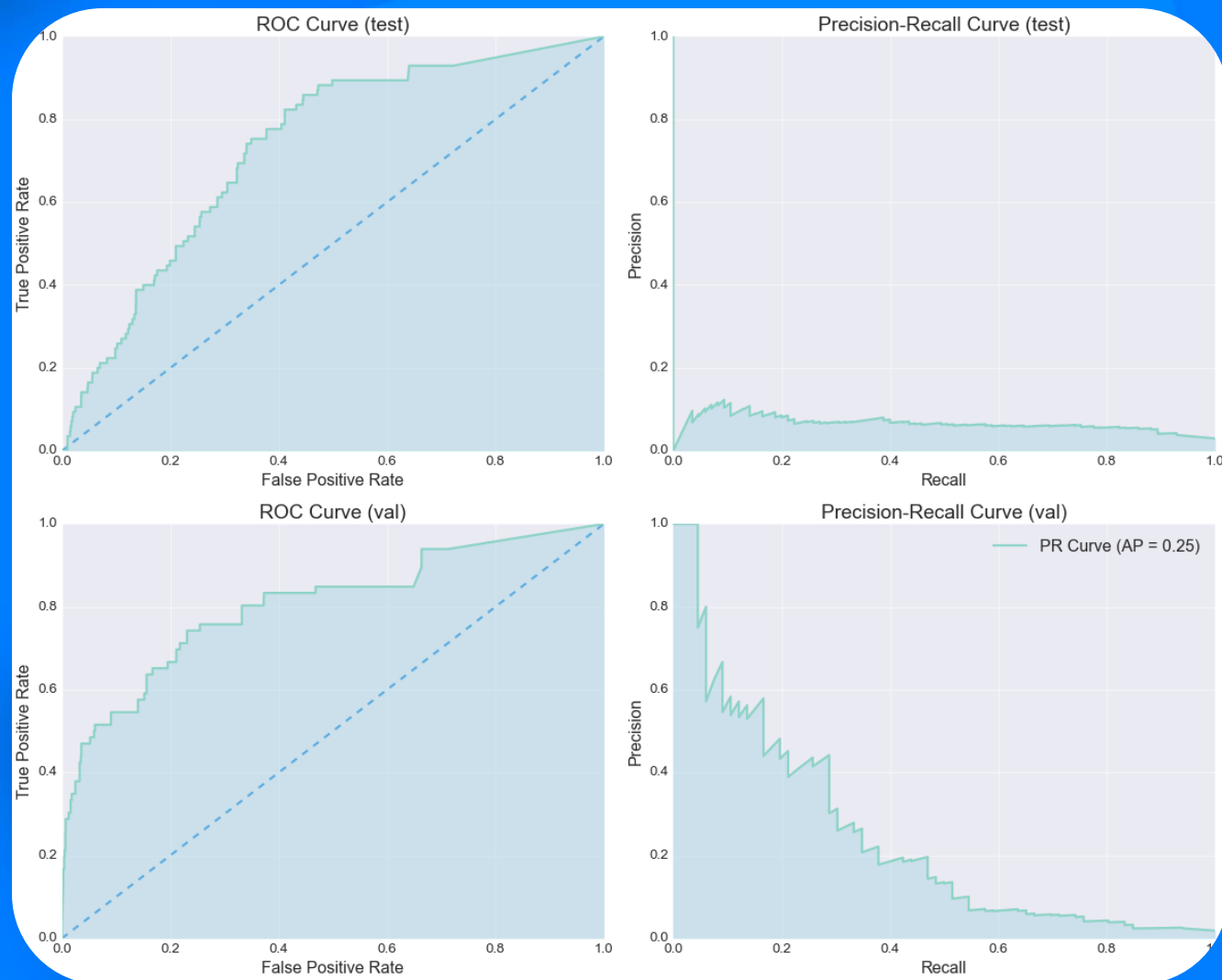
3. Построение baseline модели

```
RandomForestClassifier
```

```
RandomForestClassifier(class_weight='balanced', min_samples_leaf=7,  
                        min_samples_split=18, n_estimators=35, random_state=42)
```



4. Анализ метрик



самолет

5. Углубленная очистка



- ❖ Группировка
- ❖ KNNImputer
- ❖ StandardScaler
- ❖ PCA
- ❖ ADASYN/SMOTE
- O-S



6. Построение модели

- LogisticRegression
- RandomForestClassifier
- Boosting модели
- Ансамблевый метод

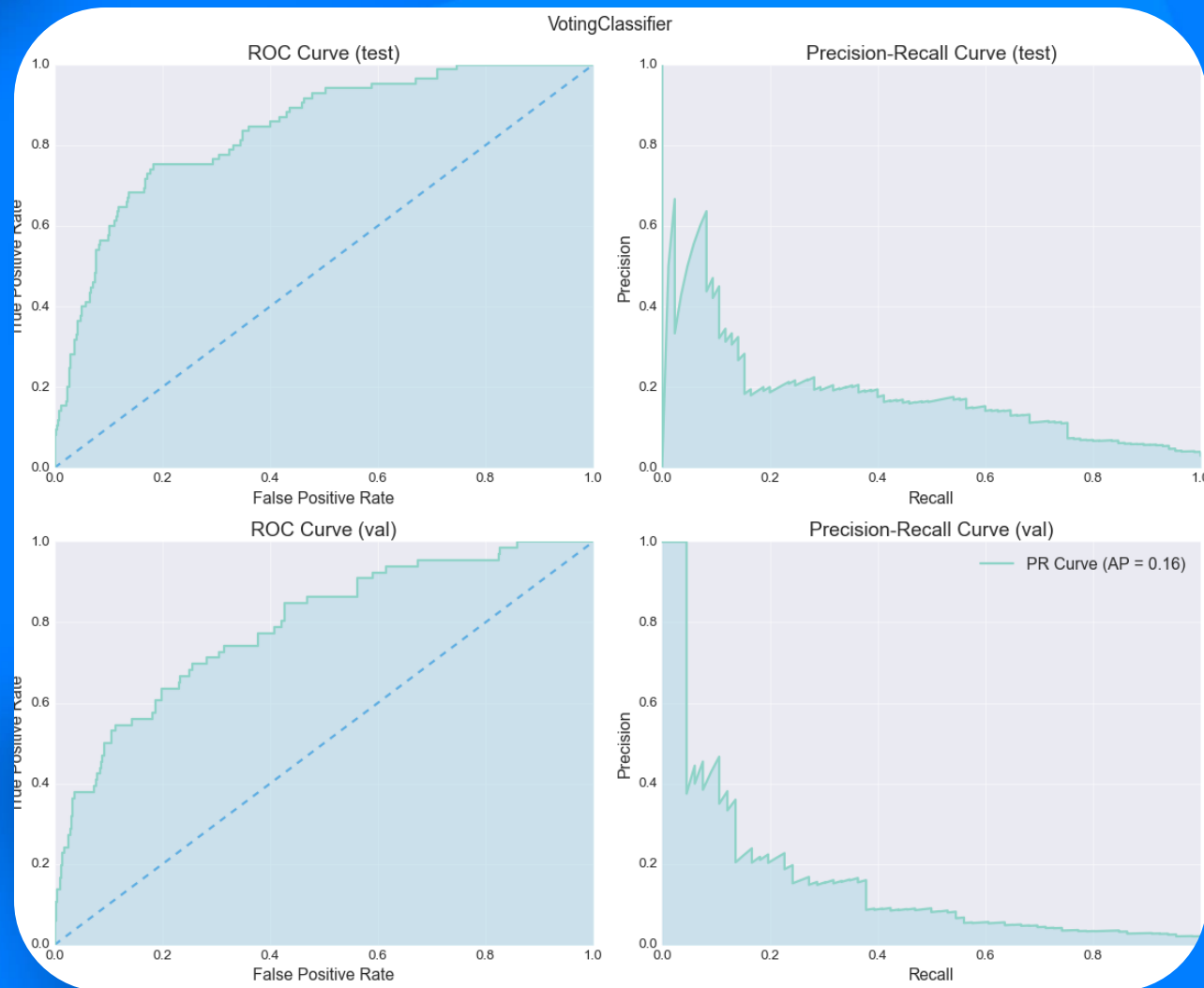


7. Выбор лучшей модели

```
Precision-Recall on test data: 0.17562957589678818
ROC-AUC on validation data: 0.7903722628974308
Precision-Recall on validation data: 0.16733708561455612
-----
ROC-AUC on test data: 0.7992848074481591
Precision-Recall on test data: 0.10514580565713039
ROC-AUC on validation data: 0.6955121686665311
Precision-Recall on validation data: 0.048841343697779364
-----
ROC-AUC on test data: 0.8205268726195515
Precision-Recall on test data: 0.1849712600678276
ROC-AUC on validation data: 0.8054665785370483
Precision-Recall on validation data: 0.1986268964209075
-----
ROC-AUC on test data: 0.7988150655945832
Precision-Recall on test data: 0.1452881177103031
ROC-AUC on validation data: 0.7643570774862722
...
Precision-Recall on test data: 0.15565822249262282
```



8. Вывод графиков



самолет

9. Сохранение предсказаний

```
[ ]
```

```
...      target  prediction  probability
0        0.0         0.0       0.020429
1        0.0         0.0       0.020092
2        0.0         0.0       0.073151
3        0.0         0.0       0.231567
4        0.0         0.0       0.010523
```



Выводы

Для данного набора данных, считаю я добился удовлетворительного результата по заданной метрике и valid датасету



самолет

Потенциальные улучшения

- Более углубленная работа с данными
- Подбор гиперпараметров для моделей
- Разработка API для более удобной работы с моделью, а также автоматизировать рассылку писем

