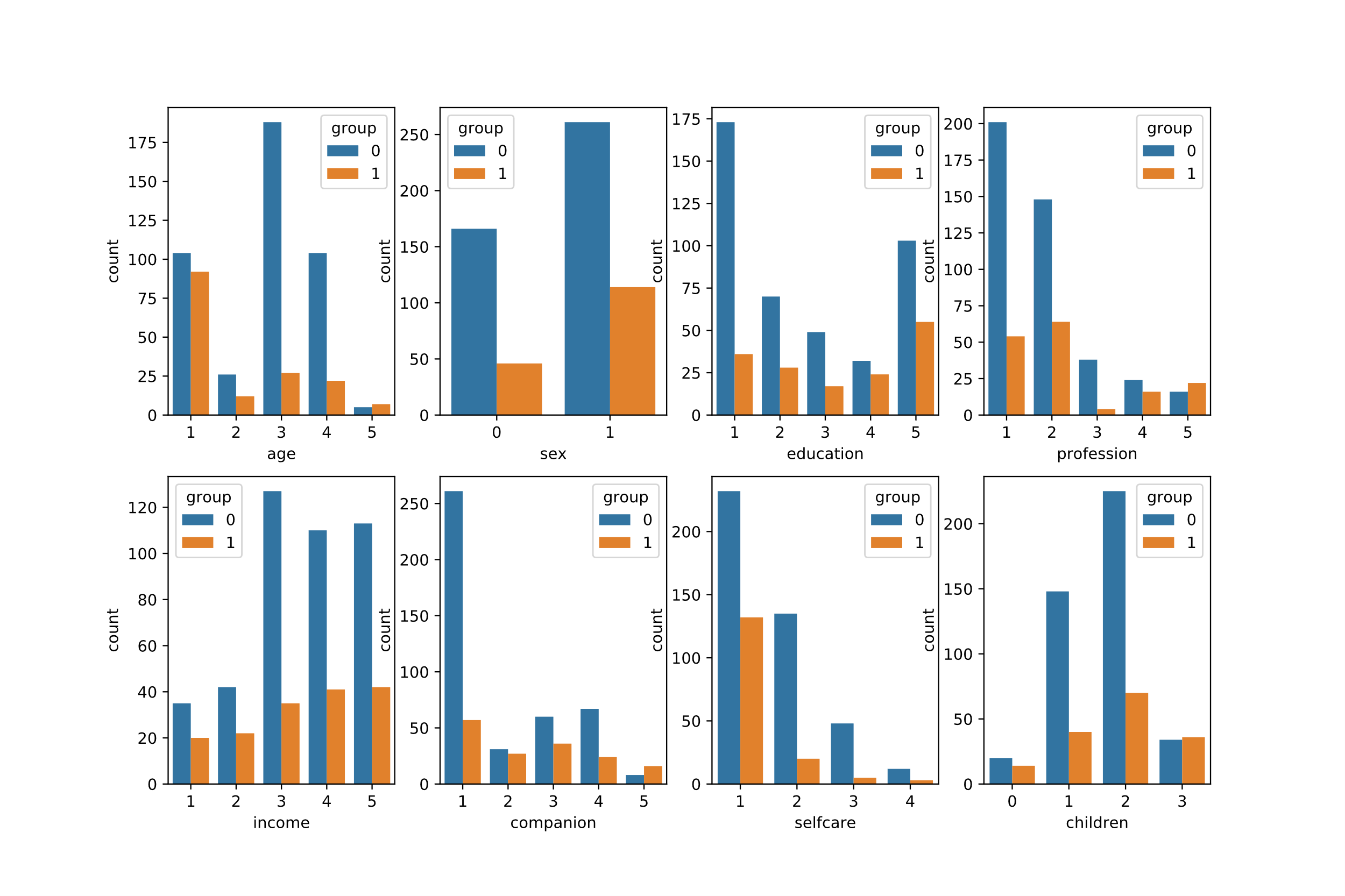
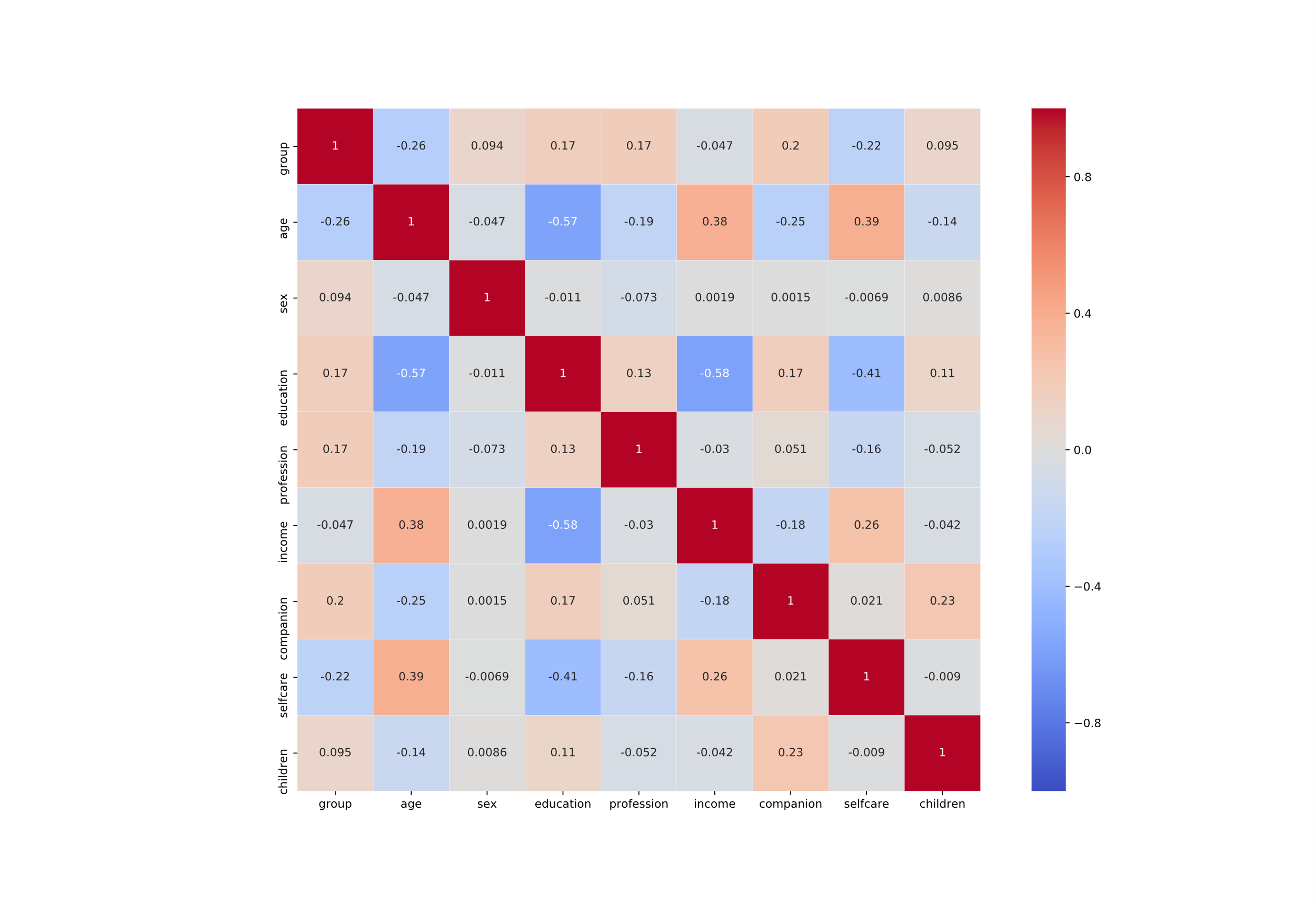
1. 单项分析结果 (Group 0=home, 1=outside）

说明：8个特征与分组的对应图示，可依此单项分析，居家养老427人，家外养老160人，不平衡数据。此图是否放入论文，酌情研究。587人选择70%作为模型训练用，30%作为最后评估用。



2. 相关图例

说明：说明各特征与分组的相关性，适当讨论说明本图。



3. 模型选择

说明：模型筛选选用的分类器，缩写如下，因本数据属于不平衡数据，筛选时以f1\_score作为评分标准，各分类器按照默认参数进行计算。最终选择CART和RFC最高，最终选择便于理解的CART决策树模型进行分析。

LR : LogisticRegression

LDA : LinearDiscriminantAnalysis

QDA : QuadraticDiscriminantAnalysis

SVM : SVC

KNN : KNeighborsClassifier

NB : GaussianNB

CART : DecisionTreeClassifier

RFC : RandomForestClassifier

ABC : AdaBoostClassifier

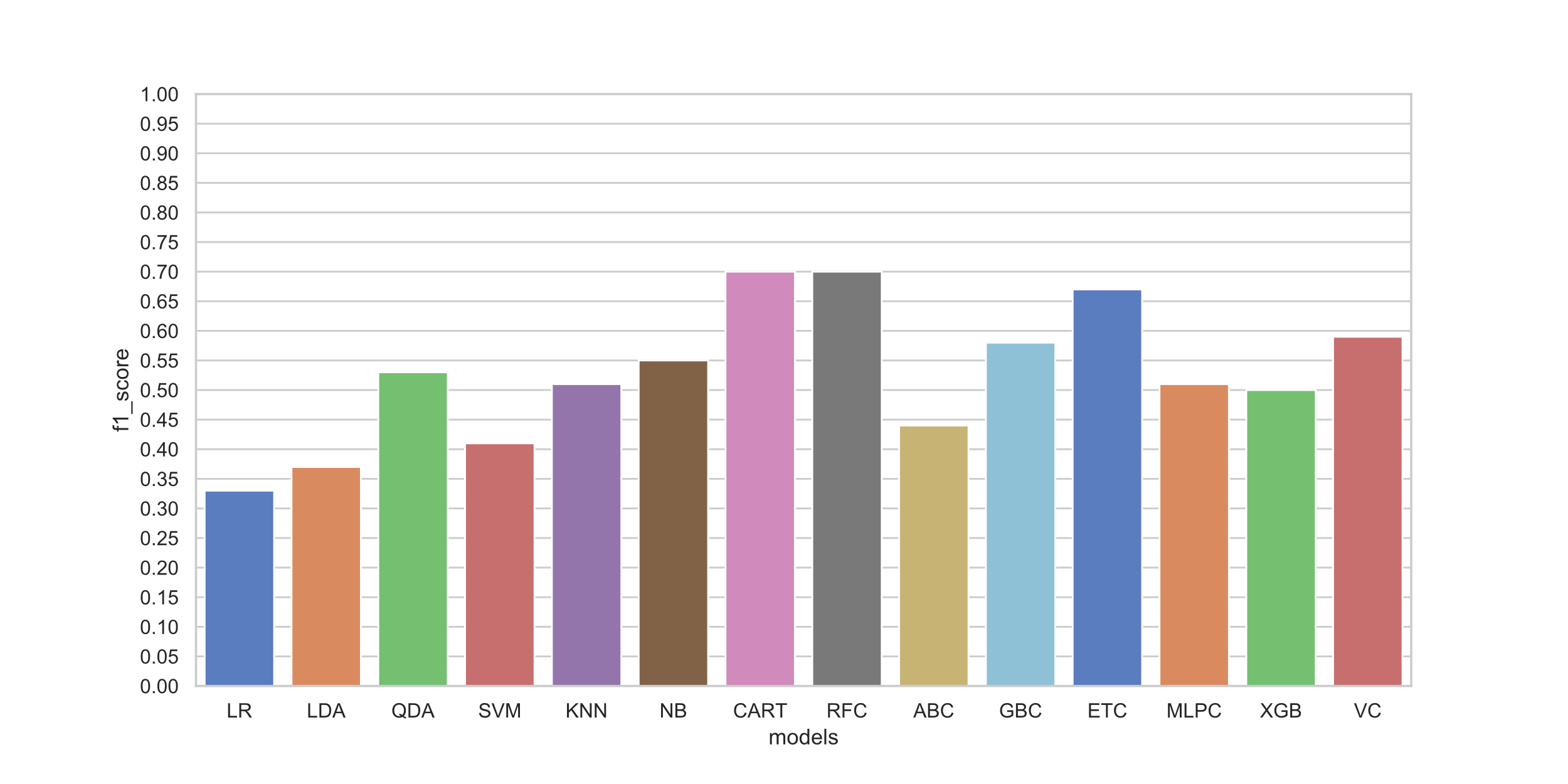
GBC : GradientBoostingClassifier

ETC : ExtraTreesClassifier

MLPC : MLPClassifier

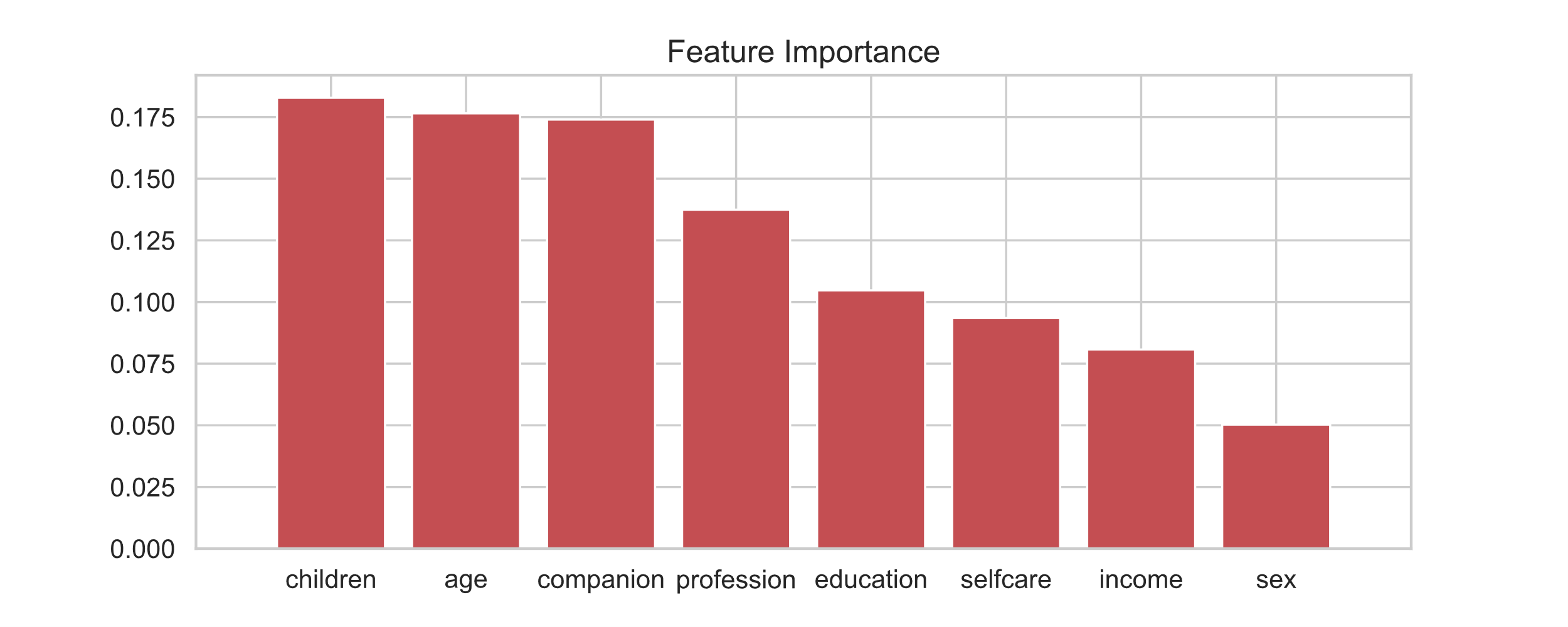
XGB : XGBClassifier

VC : VotingClassifier



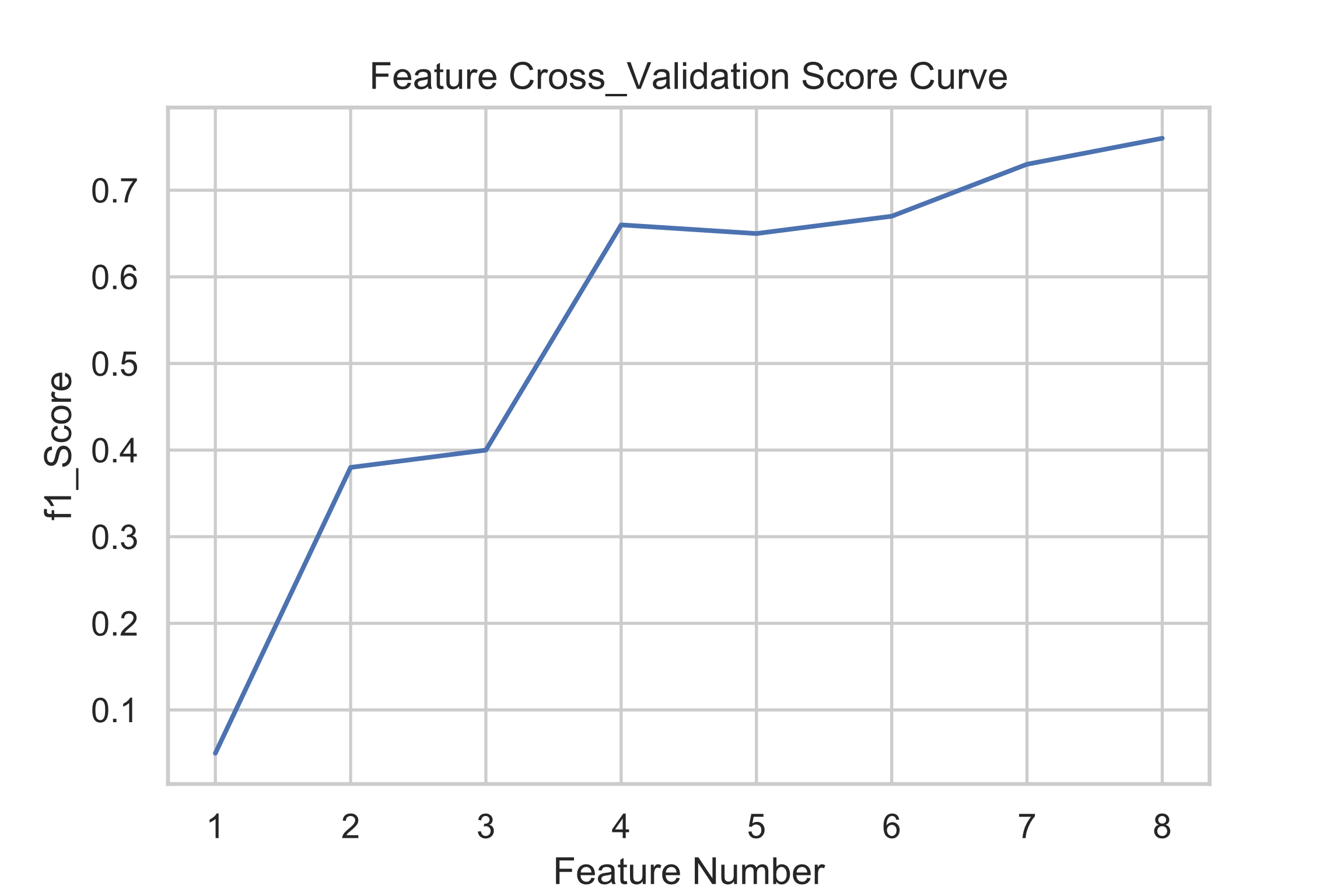
4. 特征重要性分析

说明：children, age, companion对于选择最终家里和家外养老影响更多。



5. 特征数量对模型评分的影响

说明：一共8各特征，按照特征重要性就行f1评分评价，随着特征数量的增加，评分增高，说明8项特征均作为模型的最终特征项。



6. 模型调参

说明：由于模型特征数较少，样本数不大，故仅对最大深度进行调参，通过GridSearchCV最终选择参数如下。

DecisionTreeClassifier(class\_weight=None, criterion='gini', max\_depth=10,

max\_features=None, max\_leaf\_nodes=None,

min\_impurity\_decrease=0.0, min\_impurity\_split=None,

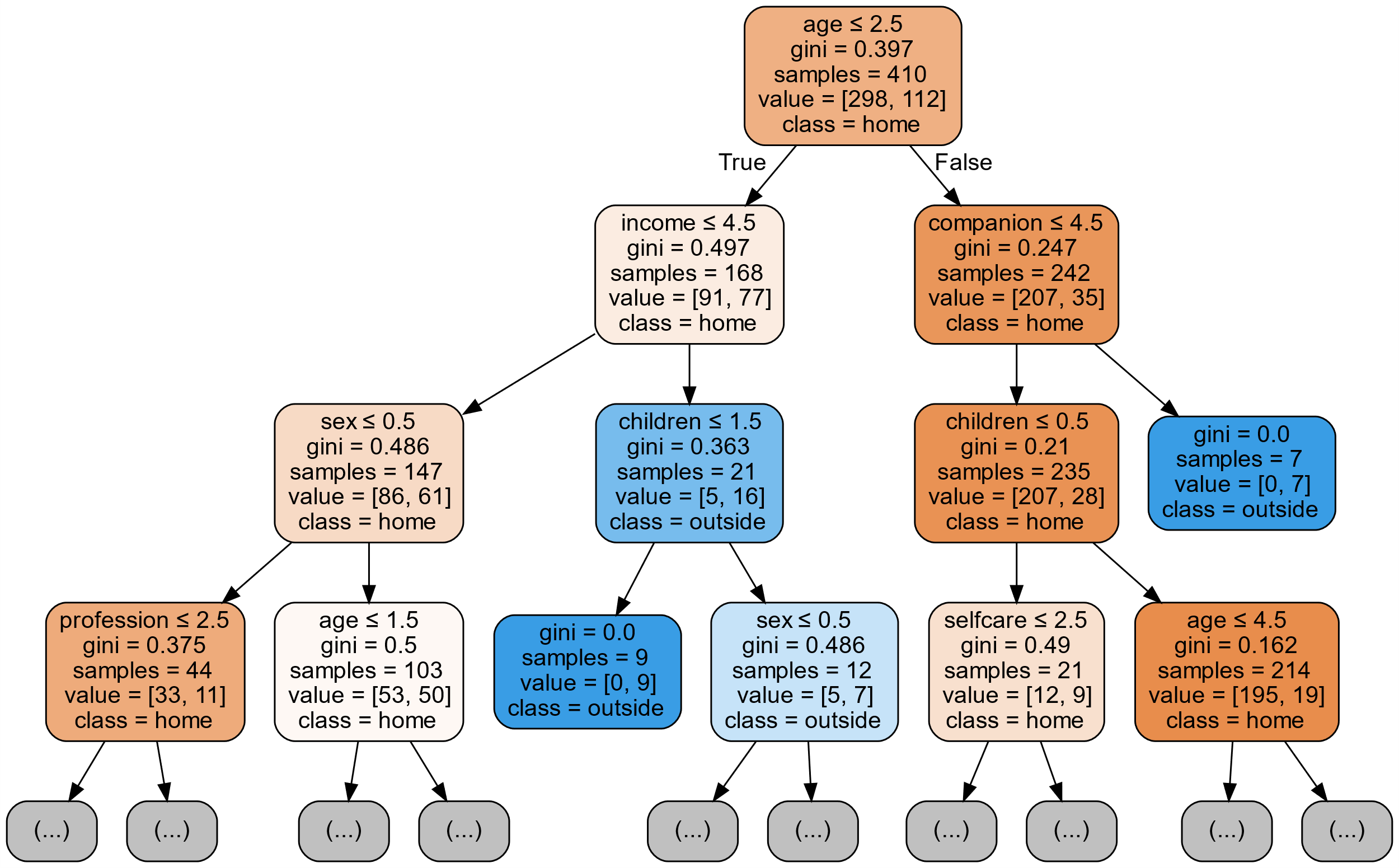
min\_samples\_leaf=1, min\_samples\_split=2,

min\_weight\_fraction\_leaf=0.0, presort=False,

random\_state=0, splitter='best')

7. 决策树图示

说明：通过决策树分析模型的分类过程，本图只做了前3层，试做分析



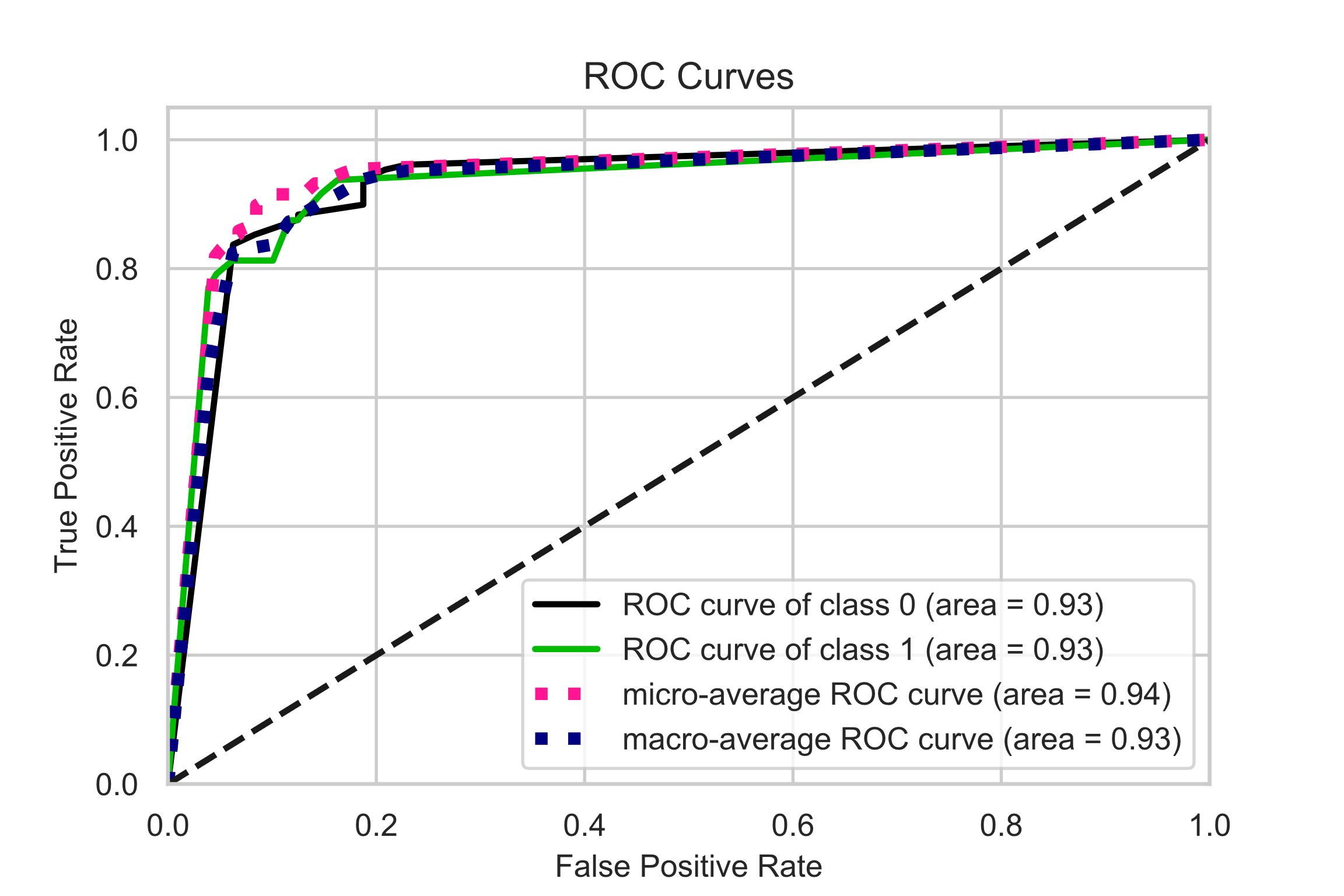
8. 学习曲线

说明：样本例数对模型的样性，如果加大样本例数还可以进一步提到准确率。



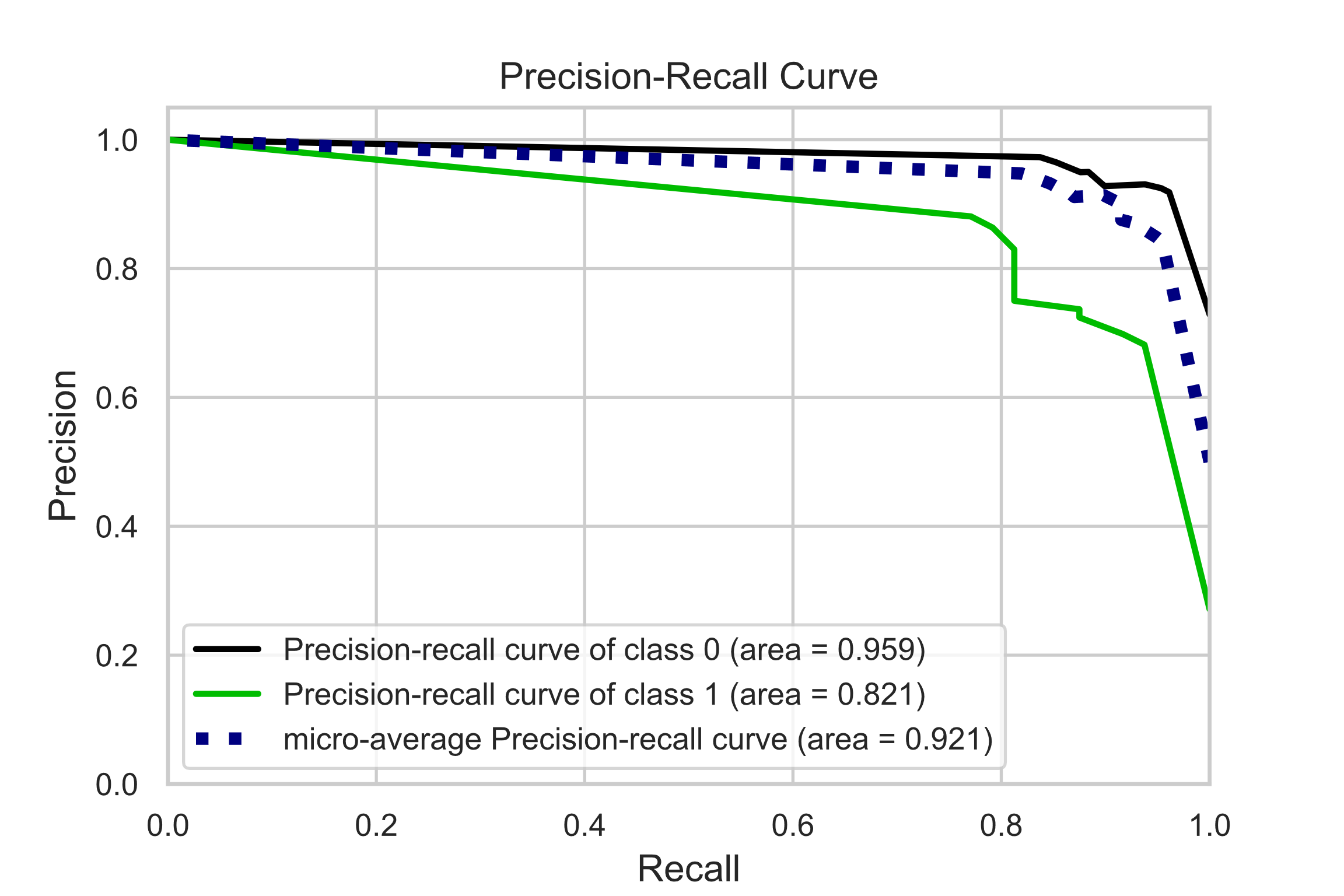
9. ROC曲线

说明：class 1 0.93，整体形状不错。



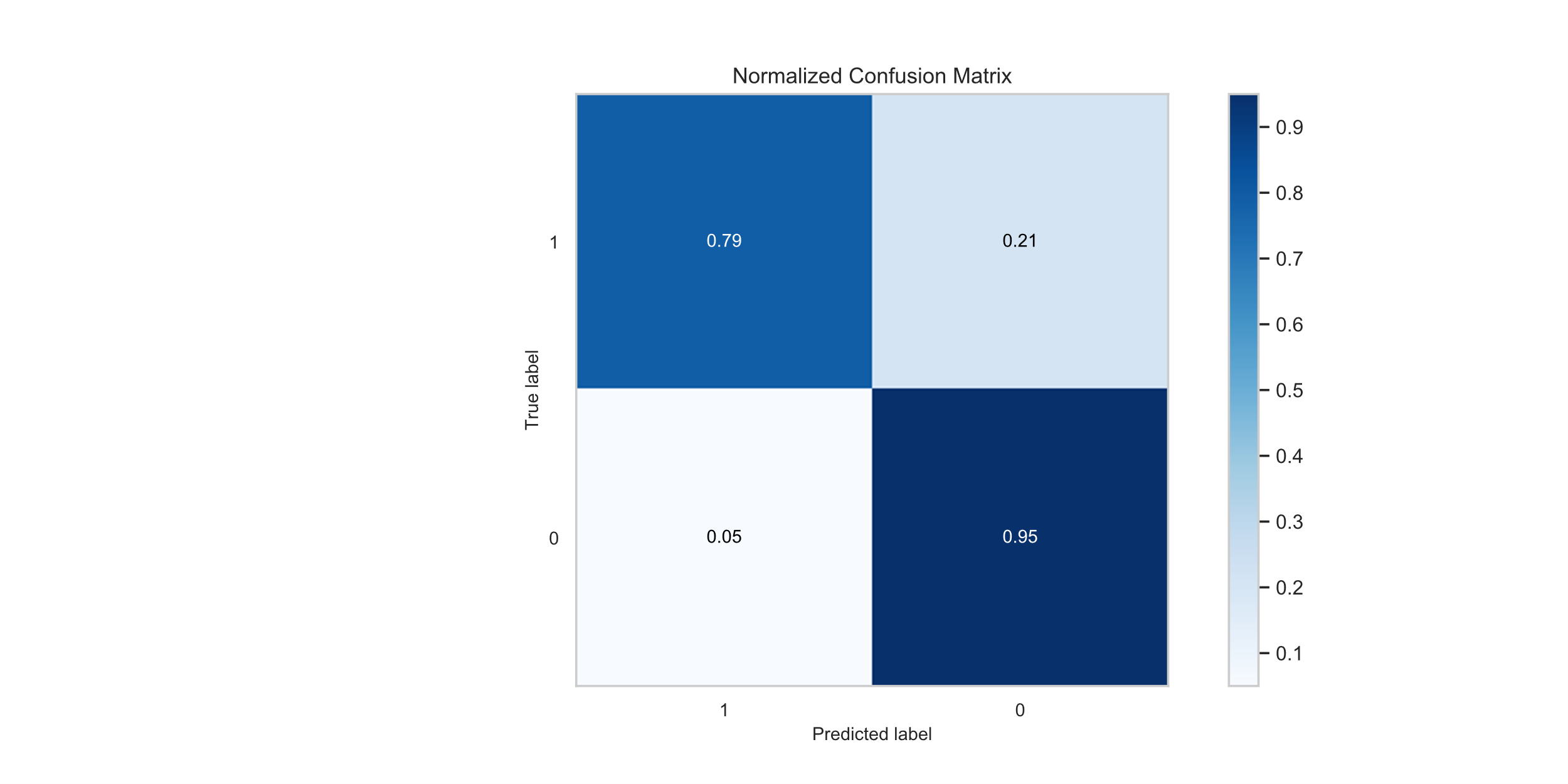
10. PR曲线

说明：到0.8的召回率以后会明显下降



11. 混淆矩阵

说明：网上学习一下，这个表格很重要，主要分析当分类为1时，就是政府评估时尽量将所有为1的分类，就是家外养老的人挑选出来进行计算和决策，所以recall要尽量大一些，同时保证precision更高些，这样f1指标就很重要。本模型对于1分类精确率0.86，recall 0.79，整体准确率0.91，可以为决策起到评估的作用。



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Classification Report** | | | | |
|  | **precision** | **recall** | **f1-score** | **support** |
| **0** | 0.92 | 0.95 | 0.94 | 129 |
| **1** | 0.86 | 0.79 | 0.83 | 48 |
| **accuracy** |  |  | 0.91 | 177 |
| **macro avg** | 0.89 | 0.87 | 0.88 | 177 |
| **weighted avg** | 0.91 | 0.91 | 0.91 | 177 |