## 投诉分析算法过程记录

**目的：**利用python替换原有R语言在“联络数据分析系统”中投诉预测的算法模型，以此消除Java在调用R过程中的不稳定性。

**前言：**在python替换R语言前，借助外部数据集对算法进行评估。投诉预测属于分类算法，因此常用的机器学习方法包括：逻辑回归、朴素贝叶斯、KNN、决策树、SVM支持向量机，附件是python程序设计及测试结果。

**结果分析：**

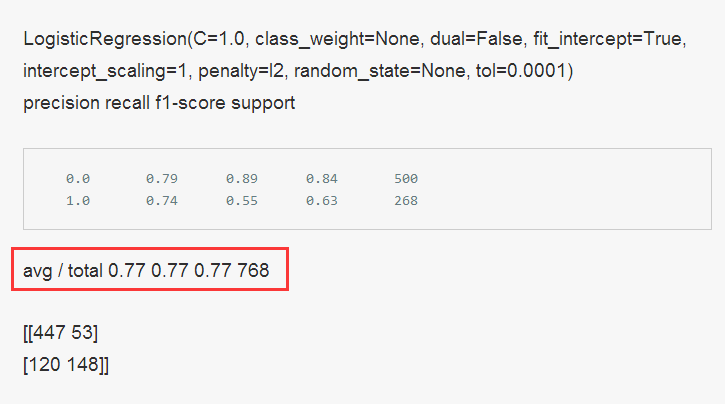
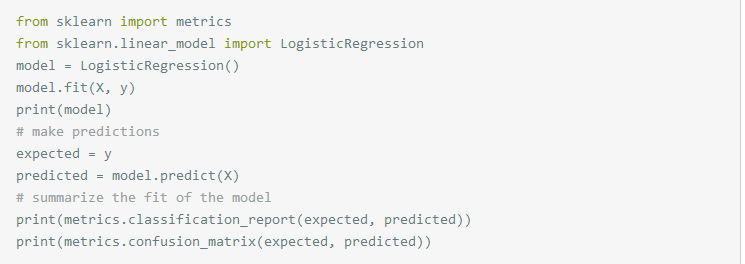
1. 投诉预测属于典型的二分类模型，而专门针对二分类模型的算法包括：逻辑回归和SVM支持向量机，由于业务要求输出一个概率值以供调整，但SVM只能回答是与否，不能满足业务需求，由此算法初步拟定为**逻辑回归**；
2. 现阶段主要目的是能够将系统贯通，对于数据缺失值填充，样本抽取，算法优化，参数调整都会放在今后有了真实实验环境的条件下进行调整；

**现有算法优势：**

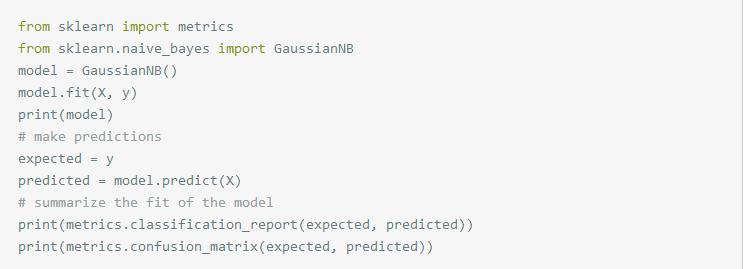
1. 直接编译的算法，解决由于公共包调用导致的算法黑箱问题，使得该算法具有更强的业务针对性；
2. 在保证输出结果完整性的前提下，减少代码多次调用及计算所导致的运行程序速度问题；
3. 由于单个txt文件承载量为百万行，若数据量超过这个范围，现有程序也保留了优化成分布式计算的可能性；

## 附件

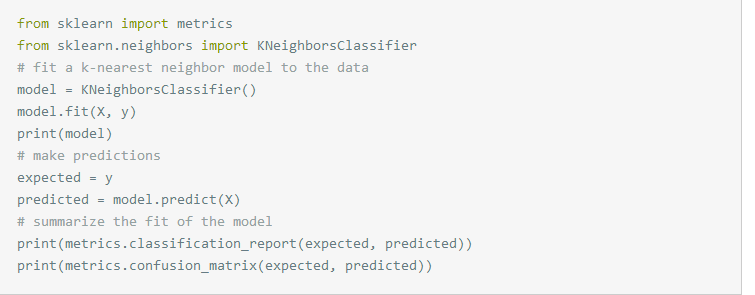
**1.逻辑回归示例及结果**



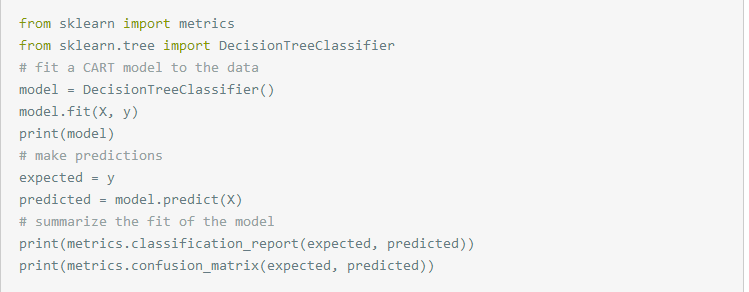
**2.朴素贝叶斯示例及结果**



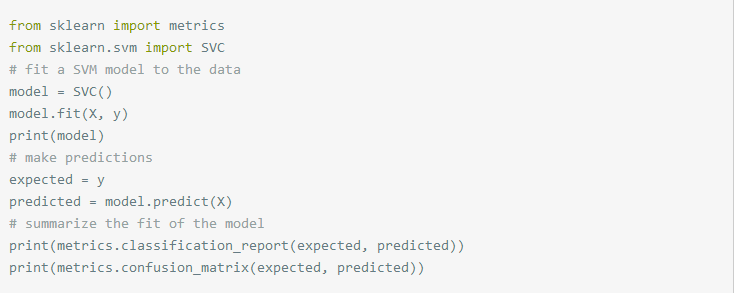
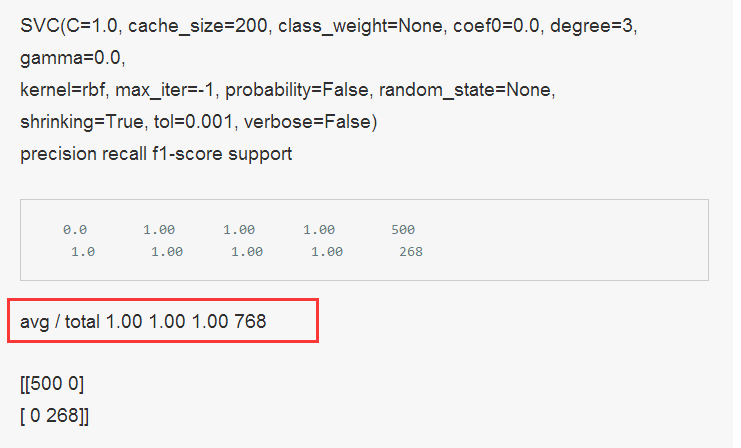
**3.KNN示例及结果**



**4.决策树示例及结果**

**5.SVM示例及结果**

** **