

课程概述

您好！欢迎来到数据分析中级课程，本课程是中级部分的第三个案例：新三板公司

去向分析。

1. 内容简介

1) 问题界定

本案例从公司去向的角度，构造主业务问题。

2) 数据准备

在增加一些新方法的基础上，完成数据预处理工作为数据建模做好准备。但是大体内容和中级课程的前二个案例基本一致。

3) 数据建模与数据可视化

主要包括特征选择和关于朴素贝叶斯等相关理论基础，以及关于特征选择和高斯朴素贝叶斯模型构建等在代码上的实现，并且结合理论知识对输出模型结果进行较为细致的分析。

2. 学习目标

学习完本次案例，你能够达到以下目标：

- 1) 掌握数据重构的新方法；
- 2) 能够独立完成特征选择以及高斯朴素贝叶斯模型的构建；
- 3) 使用流水线进行网格搜索；
- 4) 能够结合相关理论知识对最终模型效果进行评估分析。

业务问题构造

公司挂牌新三板有多种去向，包括极少量公司成功 IPO 上市，以及大部分公司因各种原因包括运营不善、对市场失望、违规违纪等进行退市摘牌。而公司最终去向直接决定了投资人如何退出公司股份以及投资人的最终收益。

1. 业务问题

新三板作为全国性的非上市股份有限公司股权交易平台，主要面向全国中小微公司解决融资难的问题，市场因素繁多，公司质量参差不齐，公司最终去向的不确定性大大提高了投资人的风险，亟需数据分析的思路和结果提供相关参考。因此本案例需要能够帮助投资人分析其去向原因以及相关因素，指导投资人预测公司未来走势，从而帮助投资人更好地制定投资策略。

公司挂牌新三板即接受市场监督，大部分重要事务需要向市场发布公告披露。新三板公司的公告能直接反应公司的运作情况，其所反映的公司特征为数据分析提供了有力的数据支持。本次案例基于已知去向公司的数据，构建公司去向分类模型，从而能够帮助预测后续新公司的未来走势。

因此本次案例的主要内容是：

围绕新三板部分公司数据，对公司的最终去向进行分类分析。