**基于聚类思想的特征提取以及贝叶斯网络的实时股票波动预测模型产品文档**

1. 《产品简介》 -- 包括产品概念、适用范围、产品功能、功能特点、运行要求和联系地址。

产品概念：基于聚类思想，以及贝叶斯网络，计算股票波动的概率

适用范围：帮助投资者配置股票，实现多种投资的最佳组合

功能：输入当前公司财报及资金情况，以及实时的交易单量和换手率，得出股票波动的概率

功能特点：根据部分历史数据以及实时数据做出预测

运行要求：环境python3.7 ，依赖包：pandas、sklearn、SciPy、tushare、pandaSQL、stockstats

1. 《产品演示》

团队简介：青岛大学数据科学与软件工程学院开源三班，张剑维，范景泉

产品背景：在投资方式多样化的今天，希望能结合大数据，给与投资者一定的投资建议

产品描述：基于贝叶斯网络进行波动概率的预测

产品特点：只需要知道当前实时数据，再根据历史交易记录就可以判断上涨的概率了。

产品作用：计算某只股票的波动概率。

适用范围：仅限上市股票A股

功能模块：数据提取，数据清洗，数据聚类，参数学习，贝叶斯网络，预测模块

1. 《功能介绍》

软件介绍：基于贝叶斯网络，通过专家法构建有向无环图，运用聚类方法对特征分类，最终计算股票涨幅的概率

软件结构：数据获取、数据清洗、特征聚类分组， 交易数据获取、交易数据RFM分组，贝叶斯网络构建

1. 《评测报告》

评测目的：验证模型效果

评测环境：win10+anaconda

评测内容：贝叶斯网络概率的输出

实测数据：income=1

结果分析：数据量不大，所以在某些类别会出现概率为0的情况，甚至在某特征出现的概率总和不为一。

5. 《安装手册》

系统环境 win10 & MAC & Linux

运行平台 anaconda

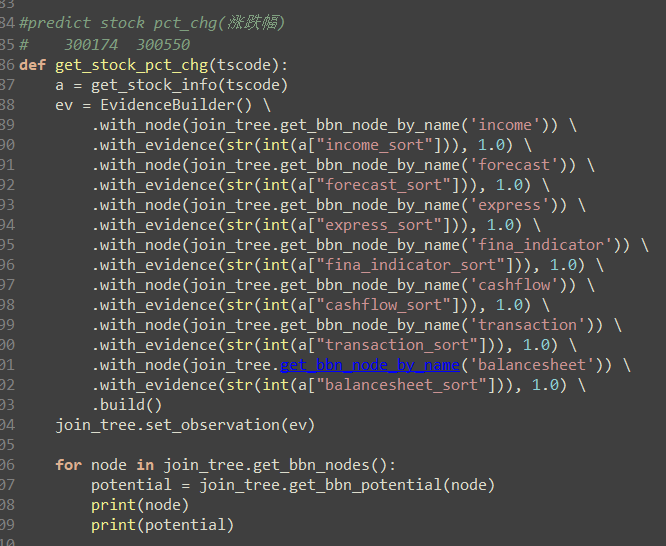
产品安装过程 配置相关包文件

调用 get\_stock\_pct\_chg 函数，可以直接得出涨跌的概率。

如果能知道某股票的确切信息，可以调用之前的数据模型对个股进行重新分类，将更新后的数据添加到贝叶斯网络。计算相应的概率

！需要注意的是，模型无法持久化，所以每次调用get\_stock\_pct\_chg函数都会改变贝叶斯网络的结构，所以每次预测之前，需要重新运行bnn的python文件，以获得最新的贝叶斯网络。

原始数据的更新需要手工操作，所以需要定时运行raw\_data和etl的.py文件。



初始环境设置 直接执行测试即可