# 股票选股监督式学习分类模型

目录

[股票选股监督式学习分类模型 1](#_Toc513198453)

[一、价格模型](#_Toc513198454) **[错误！未定义书签。](#_Toc513198454)**

[二、现金流模型](#_Toc513198455) **[错误！未定义书签。](#_Toc513198455)**

## 业务场景

### 上市公司的财务指标可以清晰的反映出该公司的盈利情况和之后的成长能力。所以，对公司财务指标的统计和分析，是进行投资的重要砝码和度量衡。

### 本模型是基于领域内专业分析结果和大数据相关性筛选验证，选取了直接影响上市公司利润和盈利能力的多个因子和指标，立足于人工智能制定的出的方法和策略。通过该模型和某企业最新数据，可对此企业之后两个季度的盈利能力进行预测，为投资提供准确的依据，进而实现投资的保值或增值。

## 数据获取和清洗

通过数据抓取获得A股所有上市公司一年内企业股价情况和一年内企业的财务分析数据，包括盈利能力，营运能力，成长能力，偿债能力和业绩报表，2018年度上半年企业的股票收益率变化情况。数据库包含2017年所有股票的每日股票情况和财务分析数据，2018年所有股票的每日股票情况。对获得的数据进行以下步骤处理：

1. 数据的规整和清洗  
   对获得的所有财务分析数据通过股票代码进行合并，去除重复项和缺失项。计算2017年内股票的收益率情况和最大回撤率，并和财务分析数据合并，清洗缺失项和重复项。终表和2018年上半年收益率情况合并，一共有1442支股票可进行下一步分析。

合并总表的数据处理  
对合并后的总表进行特征列的重命名，数据格式的转换，数据特征列的新增和删减。总表存入数据库。  
数据特征列表如下：

|  |
| --- |
| 特征向量类别 |
| 净资产收益率(%)前一年:roe\_yearly |
| 毛利率(%)前一年：gross\_profit\_rate\_yearly |
| 总资产增长率(%)前一年：targ\_yearly |
| 每股收益增长率(%)前一年：epsg\_yearly |
| 营业收入净利率(%)前一年：TTM\_yearly |
| 每股收益前一年:eps\_yearly |
| 最大收益率(%)前一年:er\_yearly |

## 分类标准

通过对每只股票后两个季度的收益率逆序排序，对收益率大小由高到低进行二分，定义在所有公司中，股票收益率排名为前百分之二十五的公司为一类，其余为二类。

## 数据划分

对数据库中总表的数据需要划分成训练集和测试集，训练集用来对深度学习模型进行训练测试集用来对训练好的模型进行测试并且评价模型效果。划分原则如下：

1. 训练数据集和测试数据集互斥。
2. 训练数据集占总数据集的百分之七十五，测试数据集占总数据集的百分之二十五。
3. 为了保证划分后数据的一致性，在划分时使用了随机抽样和分层抽样方法。  
   划分后的数据集分为四类：训练特征数据，训练标签数据，测试特征数据，测试标签数据。存入数据库。

## 模型输入和输出

在Tensorflow上进行深度学习模型开发，主要包括优化学习率，神经元个数，神经网络层数等。

步骤为：

1. 训练特征数据和训练特征标签输入进模型
2. 在每轮训练中，模型对特征数据进行训练，计算损失函数
3. 计算模型输出结果和训练特征标签的匹配度，调整参数使得损失函数减小
4. 继续下一轮训练
5. 达到稳定或者完成训练轮数后，跳出训练
6. 打印并输出结果，保存训练模型
7. 用测试标签数据和测试特征数据输入保存好的模型
8. 输出准确度，评价模型
9. 合并训练特征数据和测试特征数据，输入模型，评价结果

## 模型评价

1442支股票特征数据集输入训练模型后，输出的结果与其真实2018年上半年收益率对比，一共有1257支预测正确，正确率为86.6%。符合预期值。



## 总结

该模型预测准确率达到预期，之后可以对以下几点进行优化：

1. 输入数据特征进行扩充和完善，找寻更多财务指标特征。
2. 对股票数目进行扩充和完善，把更多的股票加入训练和预测的股票池。
3. 对模型的参数优化函数进行优化，寻找合适的参数优化函数。
4. 对模型标签分类依据进行优化，寻找适合的阈值对股票进行二分类。

该模型基于对企业重要财务指标的分析和归纳，结合人工智能算法，能够预测出后两个季度的公司收益率变化，给投资者提供了一个自动化预测公司利润变动的平台和参考指南，给与了投资保值增值的可靠依据。