## 金融大数据分析与决策系统（原始需求文档）

**系统开发目的：**

实现几个策略模型，系统根据策略模型，选出相应股票，并给出提示。需要对金融数据（股票）整理，入库，数据挖掘预测，数据统计分析及可视化等工作。

系统基本功能：

1. 二八轮动模型

A股是一个价格投机和资金驱动很强的市场，从日K、周K级别看，大多数股票都是同涨同跌。跷跷板现象：所谓“跷跷板现象”，又叫“二八轮动”，其中的“二”是权重股（以沪深300为代表），“八”是中小市值的股票（以中证500、创业板为代表）。多数时候，权重股强势，小票就会表现得弱一些；反之亦然。单边趋势明显：每隔一段时间，会有一个或大或小的行情，资金越多，趋势越强，最典型的就是2007年和2014-2015年的两波大牛市，技术指标上看不断出现顶背离，整体估值也是“没有最贵只有更贵”。

基于以上几点，选择代表权重股的沪深300，以及代表小盘股的创业板，当二者均处于升势，持有较强的一方。当二者均处于弱势，空仓。

2 多因子模型

多因子模型是应用最广泛的一种选股模型，基本原理是采用一系列的因子作为选股标准，满足这些因子的股票则被买入，不满足的则卖出。

基本概念

各种多因子模型核心的区别第一是在因子的选取上，第二是在如何用多因子综合得到一个最终的判断。

一般而言，多因子选股模型有两种判断方法，一是打分法，二是回归法。

打分法就是根据各个因子的大小对股票进行打分，然后按照一定的权重加权得到一个总分，根据总分再对股票进行筛选。回归法就是用过去的股票的收益率对多因子进行回归，得到一个回归方程，然后再把最新的因子值代入回归方程得到一个对未来股票收益的预判，然后再以此为依据进行选股。

多因子选股模型的建立过程主要分为候选因子的选取、选股因子有效性的检验、有效但冗余因子的剔除、综合评分模型的建立和模型的评价及持续改进等5个步 骤。

候选因子的选取

可以看出这个最简单的多因子模型说明了某个因子与未来一段时间收益率之间的关系。同样的，可以选择其他的因子，例如可能是一些基本面指标，如 PB、PE、EPS 增长率等，也可能是一些技术面指标，如动量、换手率、波动等，或者是其它指标，如预期收益增长、分析师一致预期变化、宏观经济变量等。

选股因子有效性的检验

多因子的模型最重要是两个方面：一个是有效因子，另外一个是因子的参数。例如到底是PE有效还是ROE有效；到底是采用1个月做调仓周期还是3个月做调仓周期。这些因子和参数的获取只能通过历史数据回测来获得。但是在回测过程中，要注意，不能过度优化，否则结果可能反而会不好。

影响股价走势的主要因子包括市场整体走势（市场因子，系统性风险）

估值因子（市盈率、市净率、市销率、市现率、企业价值倍数、PEG等）

成长因子（营业收入增长率、营业利润增长率、净利润增长率、每股收益增长率、净资产增长率、股东权益增长率、经营活动产生的现金流量金额增长率等）

盈利能力因子（销售净利率、毛利率、净资产收益率、资产收益率、营业费用比例、财务费用比例、息税前利润与营业总收入比等）

杠杆因子（负债权益比、资产负债率等）

动量反转因子（前期涨跌幅等）

交易因子（前期换手率、量比等）

规模因子（流通市值、总市值、自由流通市值、流通股本、总股本等）

股价因子（股票价格）

红利因子（股息率、股息支付率）

股价波动因子（前期股价振幅、日收益率标准差等）

市场预期因子（预测净利润增长率、预测主营业务增长率、盈利预测调整等）。