|  |  |  |
| --- | --- | --- |
| 项目名称 | | 密级 |
| 基于网络舆情的股票统计分析系统 | | 仅供收件方查阅 |
| 项目编号 | 版本 | 文档编号 |
| 1 | 1.0.0 | Project ID\_INIT\_002 |

基于网络舆情的股票统计分析系统项目

立项报告

|  |  |  |  |
| --- | --- | --- | --- |
| 拟制 | 陈嘉嘉 王苏卿  殷逢悦 王德勋 | 日期 | 2024-12-23 |
| 评审人 | 王文鑫 | 日期 | 2024-12-23 |
| 批准 | 王文鑫 | 日期 | 2024-12-23 |



**武汉市软酷网络科技有限公司**

版权所有 不得复制

Revision Record

修订记录

| Date  日期 | Revision Version  修订版本 | CR ID /Defect ID  CR/ Defect号 | Sec No.  修改章节 | Change Description  修改描述 | Author  作者 |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

Catalog

目 录

[1 项目提出 7](#_Toc355435251)

[2 开发团队组成和计划时间 7](#_Toc355435252)

[3 项目预计支出 8](#_Toc355435253)

[4 风险评估和规避 9](#_Toc355435254)

# 项目提出

**项目名称：基于网络舆情的股票统计分析系统**

**项目简介：**

我国A股市场是散户为主的市场，众多中小投资者对A股的关注和热情是股市上涨的重要推动力，而投资者对市场的热情往往体现在热门财经网站的新闻阅读量、搜索网站的搜索量等媒体介质上，对这些“网络热度”即时的分析和利用，可以从新的视角上了解投资者情绪，形成投资策略，指导投资。随着市场的发展和竞争的加剧，构建这样一个基于网络舆情的股票统计分析系统显得尤为迫切。

**项目目标：**

1.深度挖掘网络舆情数据与股票市场行情之间的内在关联，为投资者提供具有前瞻性的投资策略建议，有效解决投资者在投资决策过程中缺乏有效数据支持和精准情绪判断的痛点。

2.实现对行业指数舆情数据的实时监测与分析，准确识别行业轮动趋势，助力投资者把握行业投资机会，提高投资组合的收益水平并降低风险。

3.系统具备高可靠性和稳定性，能够处理大规模的网络舆情数据，确保数据的准确性和分析结果的及时性，以满足投资者在快速变化的市场环境中的需求。

**系统边界：**

功能边界：用户中心、数据采集、数据处理与分析、投资策略生成、可视化展示、后台运维

技术边界：前端bootstrap、后端flask，数据库MySQL，数据处理与分析nlp技术

**工作量估计：**

|  |  |  |  |
| --- | --- | --- | --- |
| **模块** | **子模块** | **工作量估计（人天）** | **说明** |
| 用户中心 | 注册、登录 | 5 | 需设计安全的用户认证与授权机制，建立用户信息数据库表 |
| 首页门户 | 10 | 实时展示最新的舆情数据内容。  数据以图表（如折线图、柱状图、饼图）和数字卡片的形式直观呈现。 |
| 数据采集 | 网络爬虫 | 10 | 开发网络爬虫从各大财经网站和社交媒体平台抓取股票舆情数据 |
| 数据清洗 | 15 | 编写数据清洗程序去除重复数据、噪声数据（如广告、无关字符等）以及错误数据，对文本数据进行格式标准化处理，如统一日期格式、编码转换等，同时将清洗后的数据分类存储到合适的数据库表或数据结构中，以便后续处理。 |
| 数据处理与分析 | 话题识别 | 15 | 从收集的舆情数据中提取出话题主要涉及的行业，从海量的舆情数据中快速锁定对投资决策、市场波动等具有影响力的主题。 |
| 情感分析 | 15 | 从收集的舆情数据中识别和提取情感信息，理解公众对热点事件的情感反应，为后续对舆情与股票相关性的分析提供参考。 |
| 可视化展示 | 界面设计 | 15 | 需进行用户需求调研与分析，设计界面布局与交互流程，制作原型图并与用户反复沟通确认，确保界面简洁美观、易用性高。 |
| 总工作量（人天）： | 85 | | |

表 1工作量估算

备注：“人天”即1个人工作8小时的量就是1人天

# 开发团队组成和计划时间

**项目计划：** 2024年 12 月 23 日 - 2025年 1 月 10 日 （计18天）

**项目总监：** 1 人 姓名：王文鑫

**项目经理：** 1 人 姓名： 王文鑫

**项目成员：** 4 人

**人员来源：**

计算机学院 2021214112 陈嘉嘉

计算机学院 2021124075 王苏卿

计算机学院 2021214110 殷逢悦

计算机学院 2021214734 王德勋

# 项目预计支出

**设备,场地占用费**:

无

**本地人员工资(管理费) ：**

(average salary+ management fee)\* number of staff \* months= 20000\*3.8

(平均工资+管理费)\*人员数目\*天= 20000\*3.8

**外协人员工资：**

无

**加班费：**

无

**交通费：**

无

**住宿费：**

无

**其它费用（如业务交往,招待,办公等）：**

无

**总计：**76000

**说明： 无**

# 风险评估和规避

**技术风险：**

1.网络热度数据来源于多种媒体介质，如新闻阅读量、搜索量等，这些数据的收集过程可能受到技术限制或平台算法调整的影响，导致数据不完整或失真。

2.基于网络热度数据构建预测模型时，模型的选择和参数设置直接影响预测结果的准确性。如果模型过于简单，可能无法捕捉到市场的复杂动态；如果模型过于复杂，则可能引发过拟合问题。

解决:提高数据质量和准确性、优化模型构建与验证过程、结合基本面分析进行投资决策、持续跟进技术更新与迭代、以及遵守相关法律法规和监管要求。

**管理风险：**

1.建立数据质量监控体系，定期审查数据来源的可靠性、数据的完整性和准确性。对于异常数据或可疑数据，应进行追溯和验证，确保数据质量。

2.定期对模型进行验证和回测，评估其预测能力和稳定性。一旦发现模型失效或预测能力下降，应及时进行调整和优化。

解决：建立数据质量监控体系，定期审查数据来源的可靠性、数据的完整性和准确性。对于异常数据或可疑数据，应进行追溯和验证，确保数据质量。

建立风险预警机制，当模型预测结果出现异常或市场情况发生重大变化时，及时发出预警信号，以便及时调整投资策略。

**其他风险：**

1.网络热度数据往往反映了市场的短期情绪波动，而非长期基本面变化。如果过度依赖这些数据，可能导致对市场趋势的过度解读，从而做出错误的投资决策。

2.网络热度数据可能存在滞后性，即市场已经发生变化，但数据尚未反映出来。此外，数据还可能受到操纵或误导，如虚假新闻、恶意刷量等。

解决：应结合基本面分析和技术分析，对市场情绪进行理性判断。同时，建立风险预警机制，当市场情绪指标出现异常波动时，及时进行调整和反思。