QuantWave(股浪预) Web APP——基于量化金融因子的实时因子分析与智能股市预测系统 项目需求分析文档

### 1. 引言

* 编写目的
* 背景
* 术语和缩略词

### 2. 目标概述

* 项目概述
* 项目来源及背景
* 项目目标
* 系统功能概述
* 用户特点
* 假定和约束

### 3. 功能需求

* **3.1实时因子分析**
* 数据收集与整合
* 因子计算与分析
* 实时数据更新与处理

* **3.2智能股市预测**
* 预测模型选择与实现
* 预测结果可视化
* 预测结果分析与优化

* **3.3用户界面**
* 用户登录与权限管理
* 数据可视化界面设计
* 用户交互功能

### 4. 性能要求

* 响应时间
* 数据精度
* 灵活性
* 可靠性
* **可维护性**

### 5. 数据需求

* 静态数据
* 动态数据
* 数据字典
* 数据库描述
* 数据安全与隐私保护

### 6. 系统架构

* 技术架构概述
* 模块划分与组件设计
* 第三方集成需求

### 7. 测试策略与计划

* 功能测试
* 性能测试
* 安全测试
* 用户验收测试计划

### 8. 实施计划

* 发展阶段划分
* 发布计划与版本控制

### 9. 风险管理与问题解决

* 风险识别与评估
* 应对策略
* 问题追踪与解决流程

### 10. 附录

* 参考资料

### 一、 引言

#### 1.1 编写目的

  本文档的目的是详细地介绍 QuantWave(股浪预) Web APP所包含的需求，以便客户能够确认产品的确切需求以及开发人员能够根据需求设计编码，以下叙述将结合文字描述、数据流图、ER图等来描述 QuantWave(股浪预) Web APP的功能、性能、用户界面、运行环境、外部接口以及针对用户操作给出的各种响应。本文档的预期读者有客户、项目经理、开发人员以及跟该项目相关的其他竞争人员。

#### 1.2 背景

在当前投资市场愈发复杂化的进程中，普遍面临如下问题：

1、技术和方法落后。国内量化交易市场仍处于初级阶段，大部分交易系统和策略仍采用传统的统计方法和手工处理，缺乏智能化和自动化的解决方案。相比国外，尤其是欧美成熟市场，技术水平和方法论相对落后。

2、缺乏多样化的策略支持。国内量化交易市场的交易策略相对单一，缺乏多样化的策略库支持。这限制了投资者能够选择和比较不同的投资策略，导致市场上的资金规模相对分散。

3、模型复杂性和解释性问题。一些预测模型过于复杂，导致预测结果的解释性不强。这使得投资者难以理解和信任模型的预测能力，从而限制了这些模型在实际投资决策中的应用。

4、市场透明度不足。国内封装的交易系统在展示投资绩效时存在选择性，不够公开透明。这限制了投资者获取足够信息和进行全面分析的能力，增加了投资决策的不确定性和风险。

本项目旨在根据机器学习算法对股市数据进行研究，通过自动化地挖掘有效因子并对其进行全面系统的分析，提出相应的智能量化投资策略，并进行多策略间的对比与排名，最终建立一个可视化的智能股市预测分析系统。

#### 1.3 术语和缩略词

因子：

策略：

sharpe：

QW：QuantWave(股浪预) Web APP

### 目标概述

#### 2.1 项目概述

该项目主要聚焦于开发一套智能化的股市预测分析系统，以提高股票预测的准确性。该系统基于先进的机器学习算法，能够自动化处理海量数据，进行深度因子分析，并构建高效的预测模型。通过因子分析与预测模型的结合，根据已有的模型和股市信息，开发出一套预测效果更优的算法，为投资者提供更好的决策依据。

同时，将提供一个可视化平台，该平台将针对量化从业人员或懂得使用代码进行量化交易的人群设计，使用户能够直观地了解算法的预测情况，易于理解的投资决策依据。平台还支持用户自己上传数据、模型，从而对我们的算法与其他模型的表现进行分析比较。此外，我们还提供定制化的投资咨询服务，根据投资者的需求和风险偏好，提供个性化的投资方案。

#### 2.2 项目目标

1、提供可视化平台进行分析比较

目前用于股票预测算法的可视化平台大多只供相关企业内部参考。我们将开发一个直观易用的可视化平台，使用户能够对我们的算法与其他模型进行比较和分析，从而更好地理解其优缺点，进行模型选择。

2、提高股票预测准确率

通过开发更准确的量化交易算法，我们的目标是提高股票价格预测的准确性，从而为投资者提供更可靠的决策依据。

#### 2.3系统功能概述

该系统不仅仅提供先进的机器学习算法和预测模型，更通过互动性强、可视化友好的平台，满足年轻专业人士和量化投资爱好者对于高效数据分析和智能投资决策的需求。这些特点将帮助他们更好地理解市场趋势、识别风险和机会，实现更优秀的投资表现。

#### 2.4用户特点

该系统主要面向年龄在20到35岁之间的年轻专业人士和量化投资爱好者。他们熟练运用各种科技工具和平台，习惯使用电脑和移动设备进行复杂任务和数据分析，并且对投资和财务决策有浓厚兴趣，希望通过数据和科学方法提高投资的决策准确性和效果。

#### 2.5 假定和约束

1、人力和时间的约束

本APP开发过程中需要考虑到人力和时间的约束,相较于一些软件的开发团队来说人员较少时间较短。

2、技术发展的约束

计算机技术和发展的日新月异，将会给信息处理带来更多手段，同时也会带来更加丰富的信息表达形式，例如现在发展起来的人工智能等等，可能导致我们在搜索问题的时候没有那么智能，这就要求软件在设计时要考虑技术变化的可能性，为可能的变化预留一定的处理能力。

### 功能需求

插图：功能概述

#### 3.1实时因子分析

1、数据收集与整合

首先使用可靠的数据挖掘平台，获取并清洗股票市场的历史数据，构建一百支股票的数据库，包括股票市值、非线性市值、账面市值、动量、残差波动率等，并进行预处理、因子值计算等，为后续的量化金融因子计算和模型构建奠定了基础。

2、因子计算与分析

基于已有的股市数据，我们运用量化金融因子公式和计算模型，对市场因子的走势进行了深入分析和研究。我们会根据分析结果，画出IC/IR图与sharpe图，将显示选择策略的原因量化、可视化，供用户参考。

3、用户数据上传

用户在注册账号并登录经过验证后，可以按照规定格式上传需要预测的股票数据，根据量化金融因子公式和计算模型，进而对市场因子的走势进行了深入分析和研究。

4、算法设计与优化

基于机器学习和深度学习等技术，在现有的股票预测算法的基础上进行改进，更加注重因子之间的关系对于因子权重的影响，并设计和优化因子筛选的过程，从而提高股票预测的准确性和稳定性。在设计过程中，我们将考虑模型的复杂度、泛化能力和实时性等因素。

使用历史数据对设计的算法进行评估和验证，考察其在不同市场情况下的表现。同时，开发一个可视化平台，使用户也能够直观地比较我们的算法与其他模型的预测结果。

#### 3.2智能股市预测

##### 3.2.1 预测模型选择与实现

1. 线性回归模型

线性回归是一种最简单的回归分析方法，假设因变量（股票价格）与一个或多个自变量（股票因子）之间存在线性关系。模型的目标是找到一组系数，使得自变量的线性组合能最准确地预测因变量。具有简单易用、解释性强、低计算成本的优点，但无法捕捉股票市场复杂的非线性关系。

2. 向量自回归模型（VAR）

VAR模型是一种多变量时间序列模型，用于捕捉多个时间序列变量之间的相互影响。每个变量不仅由自己的过去值决定，还受到其他变量过去值的影响，能够分析多个股票或经济指标之间的动态关系。适用于分析多个股票或经济指标之间的动态关系，但需要较长的历史数据进行模型训练

3. XGBoost模型（Extreme Gradient Boosting）

XGBoost是一种集成学习方法，基于梯度提升树（GBT）的实现。通过迭代地构建一系列弱预测模型（决策树），并将它们结合起来来提高整体预测性能。擅长捕捉数据中的非线性模式，适合复杂的股票市场数据，同时还可以提供特征的重要性，有助于理解哪些因素对预测结果影响最大。但模型有很多超参数，需要进行仔细的调优以获得最佳性能。

1. Autoformer模型（属于神经网络模型）

Autoformer 是一个结合了自动深度学习 (AutoML) 和注意力机制 (Transformer) 的模型。它主要由两个部分组成：自动搜索架构的 AutoML 和基于 Transformer 的网络结构。

**AutoML**：

Autoformer 使用了 AutoML 技术来自动搜索适合特定任务的神经网络结构。主要负责搜索最优的网络结构，包括确定模型的深度、宽度以及各个层次的配置。

**Transformer**：

确定了网络结构后，Autoformer 使用 Transformer 结构来构建具体的神经网络模型。Transformer 结构适合处理序列数据，并能利用注意力机制有效地捕捉长距离依赖关系。

**Autoformer模型的优势**

AutoFormer能够通过 AutoML 的自动搜索和优化过程，减少人工调参的需求，提高模型的性能和效率。

##### 3.2.2 预测结果可视化

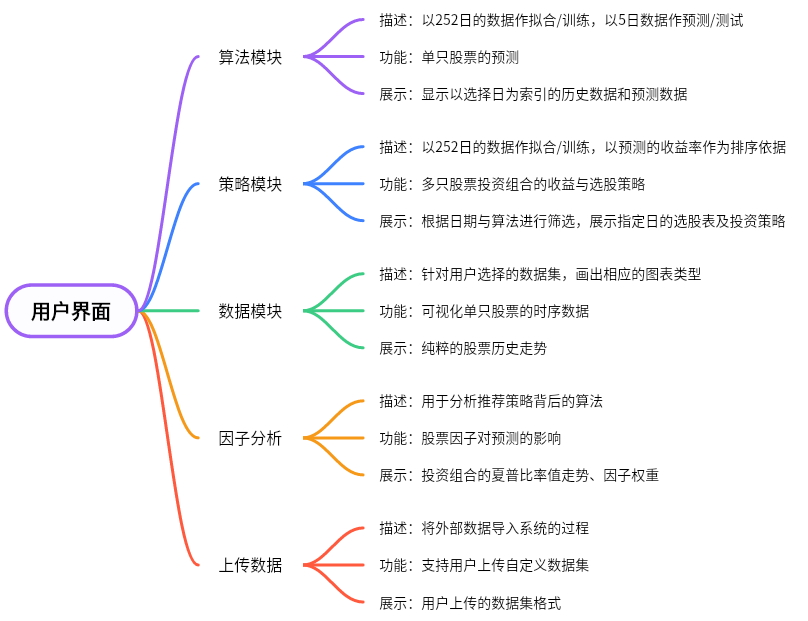
1. 单只股票：假设以252日的数据作拟合/训练，（回看窗口30）以5日数据作预测/测试展示：单击/选择 某只股票，依据日期，可以显示以某日为索引的历史数据和预测数据

2. 根据日期与算法进行筛选，展示指定日的选股表及投资策略（或做成滚动屏）。每日排名前10的股票，画出回看窗口与预测值的折线图。

##### 3.2.3 预测结果分析与优化

针对已有模型得出的预测结果，可以通过分析特征的重要性，了解哪些因子对于股票价格预测影响最大。这有助于提高模型的可解释性，并指导后续的投资决策。根据预测结果，设计和回测不同的投资策略。可以基于历史数据模拟不同的买入卖出决策，评估策略的盈利能力和风险管理效果。最后可以比较不同模型的预测效果，包括线性回归、VAR模型等，以确定在特定情境下哪种模型表现最佳。

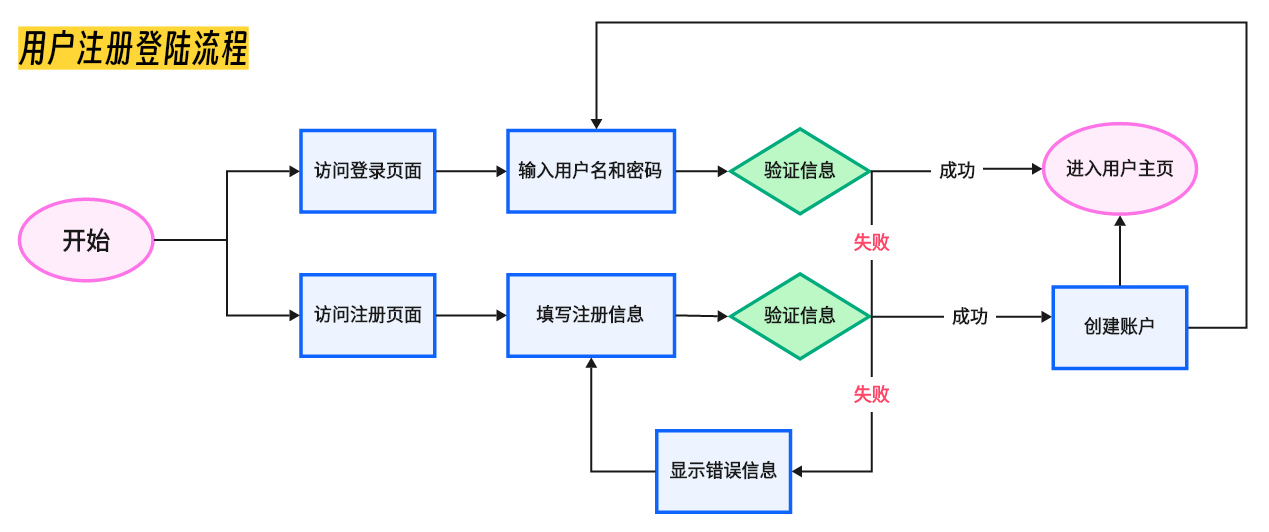
#### 3.3用户界面



##### 3.3.1用户注册登录功能

（1）注册：用户填写该页面的“用户名”、“昵称”、“密码”、“确认密码”信息后点击提交即可成功注册，返回“注册是否成功的消息”。

（2）登录：用户填写该页面的“用户名”、“密码”信息后点击登录即可成功登录，如果用户没有账号可以点击下方的链接进行注册。

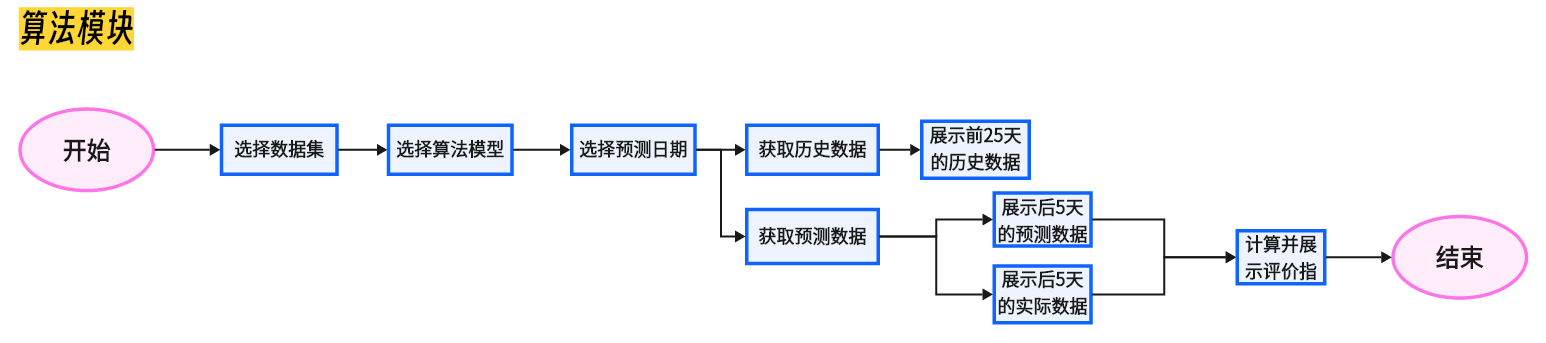


##### 3.3.2数据可视化界面设计

本项目分为五个模块：算法模块、策略模块、数据模块、因子分析和上传数据

（1）算法模块

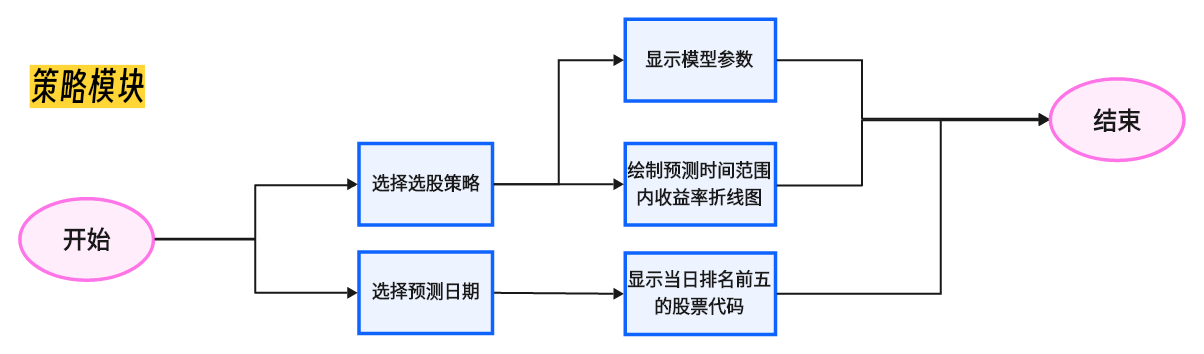
本模块的功能是预测单只股票。用户选择要展示的数据集、对应的算法模型、要预测的日期，平台根据用户的选择展示出绘制出当前日期前25天的历史数据，以及后5天的实际数据与预测数据，并给出评价算法的指标MSE与MAE的值。



（2）策略模块

本模块的功能是展示多只股票投资组合的收益与选股策略。用户依据模型预测的结果，选择选股策略后，平台会显示模型参数，并绘制出预测时间范围内收益率的折线图。

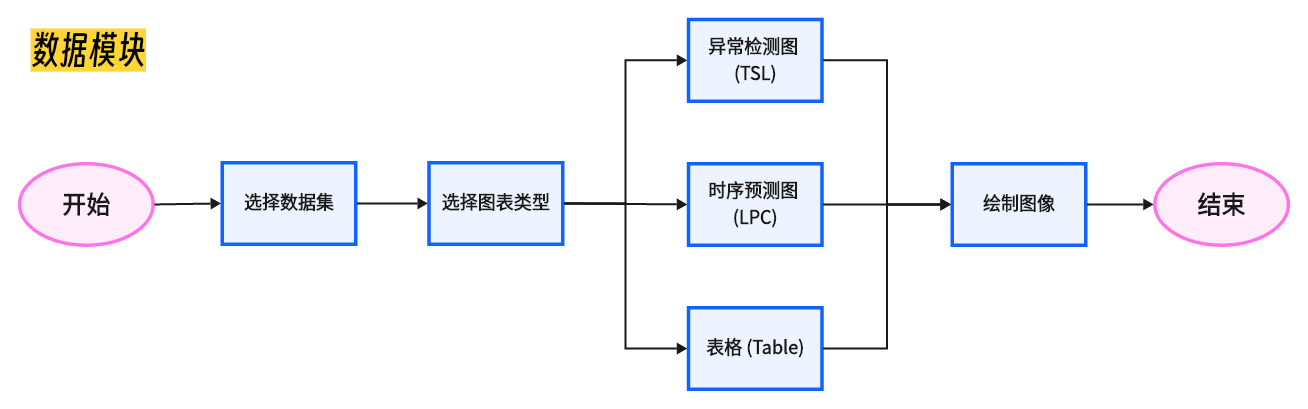
用户选择要预测的日期后，平台会显示出当日排名前五的五只股票代码。



（3）数据模块

本模块的功能是可视化单只股票的时序数据。用户选择需要可视化的数据集，再选择需要可视化的图表类型，平台会绘制出对应的图像。

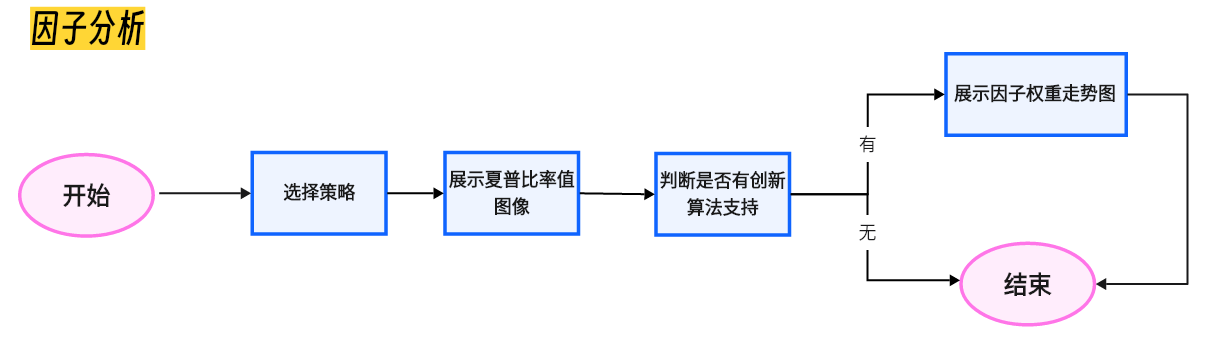
目前支持三种类型的图表：异常检测图（TSL）、时序预测图（LPC）、表格（Table）。



（4）因子分析

本模块的功能是展示股票因子对预测的影响。用户选择策略后，平台会可视化对应的夏普比率值图像。夏普比率（Sharpe Ratio）是一种用于评估投资组合风险调整后收益表现的指标，即投资回报与多冒风险比例。这个比例越高，代表投资组合越佳。

如果后续推出的创新算法能支持有关因子权重的策略选择，则还会画出因子权重的走势图。



（5）上传数据

本模块的功能是支持用户上传自定义数据集。针对不同投资者的需求和偏好，平台提供个性化定制的投资建议和服务，投资者可以上传自己需要的股市信息，帮助他们制定符合自身投资目标的策略。

### 四、性能要求

#### 4.1 响应时间

响应时间指从用户发出请求到系统做出响应的时间间隔。较短的响应时间通常被认为是性能良好的标志，特别是在交互式应用程序中（如网页），响应时间的快慢会直接影响用户体验。

**4.1.1 Web应用程序的响应时间要求**

1. 页面加载时间

网页应该在几秒钟内加载完成（2-3秒内）；

2. 交互响应时间

用户与网页的交互（如点击按钮）应该在几百毫秒内得到反馈，以确保用户的流畅体验。

**4.1.2 测试响应时间的方式**

1. 使用浏览器开发者工具的网页监控功能

可以监测每个请求的响应时间、加载时间以及页面渲染时间。

2. 编写自动化的性能测试脚本：

例如，Python中的requests库和BeautifulSoup可以用来发送请求和分析响应时间。

3. 测试网页的加载速度和响应时间的在线工具/软件：

例如Google PageSpeed Insights提供了网页的性能分析报告，包括加载时间、建议优化等；Pingdom可以监测网站的响应时间、性能统计和报告。

#### 4.2 数据精度

数据精度指系统处理和输出数据的准确程度。软件在数据计算、处理、存储和传输过程中，应保持高精度，确保输出结果符合预期且符合业务需求。数据精度涉及到算法设计、数值计算精度、数据验证和检查等方面。

具体到本项目，可以采取以下的数据精度测试方式：

**1. 历史数据回测**

使用历史数据来模拟实际投资情况，回测系统的预测能力。可以选择一段历史时间范围，通过模拟每日或每周的投资决策，来评估系统的数据精度。

**2. 交叉验证**

将数据分为训练集和测试集，使用训练集来训练模型和因子选择，然后测试集上进行验证。

**3. 指标评估**

通过预测系统性能的具体指标评估性能，如评估收益率的精度、准确率、召回率等指标。

**4. 风险调整的回报率**

还可以考虑投资组合的风险调整回报率（例如夏普比率），这可以更好的评估系统的风险管理能力和整体投资效益。

**5. 与市场基准比较**

将系统的表现与市场基准（如标准普尔500指数）或其他同类投资策略进行比较，以衡量系统相对于市场的超额收益或风险调整的超额收益。

#### 4.3 灵活性

软件的灵活性指其在面对变化、需求调整或扩展时的能力。包括可扩展性、可移植性、模块化等。

**扩展性测试例如：**

测试软件是否能较好地集成新的功能模块或插件。可以通过开发和集成一个简单的插件，并验证其对整体系统的影响来测试。

**可移植性测试例如：**

在不同的硬件平台或操作系统上部署和测试软件，确保软件能够在这些不同的环境中正常运行并保持性能。

**模块化测试例如：**

针对软件的不同模块进行单元测试和集成测试，确保模块之间的耦合度低。

#### 4.4 可靠性

软件的可靠性包括正确性、稳定性、容错性等。

测试软件的可靠性，有以下方式：

**1. 稳定性和负载测试**：

测试软件在长时间运行下的表现，比如性能是否下降、是否存在内存泄漏等问题。

**2. 可恢复性测试**：

模拟软件在面对异常情况（如断电、网络中断）时的行为，评估其是否能处理异常并在恢复后恢复到正常状态。

#### 4.5 可维护性

 软件的可维护性，包括清晰的代码结构、注释、命名规范、版本控制等。

### 数据需求

#### 5.1 静态数据

（在系统运行过程中主要作为参考的数据）

用户信息、用户自己导入的数据、股票数据(是实时的拉下来呢还是用静态的存好的？或者训练用存好的展示用实时的？？）

#### 5.2 动态数据

（在运行中要不断发生变化或者在特定的条件下要发生变化的数据，以及在运行中要输入、输出的数据。）

股票的分析数据、判断用户某一行为后的状态码和提示信息

#### 5.3 数据字典

用户名

新用户注册信息

|  |  |
| --- | --- |
| 名称 | 新用户注册信息 |
| 简述 | 新用户进行注册的信息 |
| 来源 | 用户 |
| 去处 | 注册 |

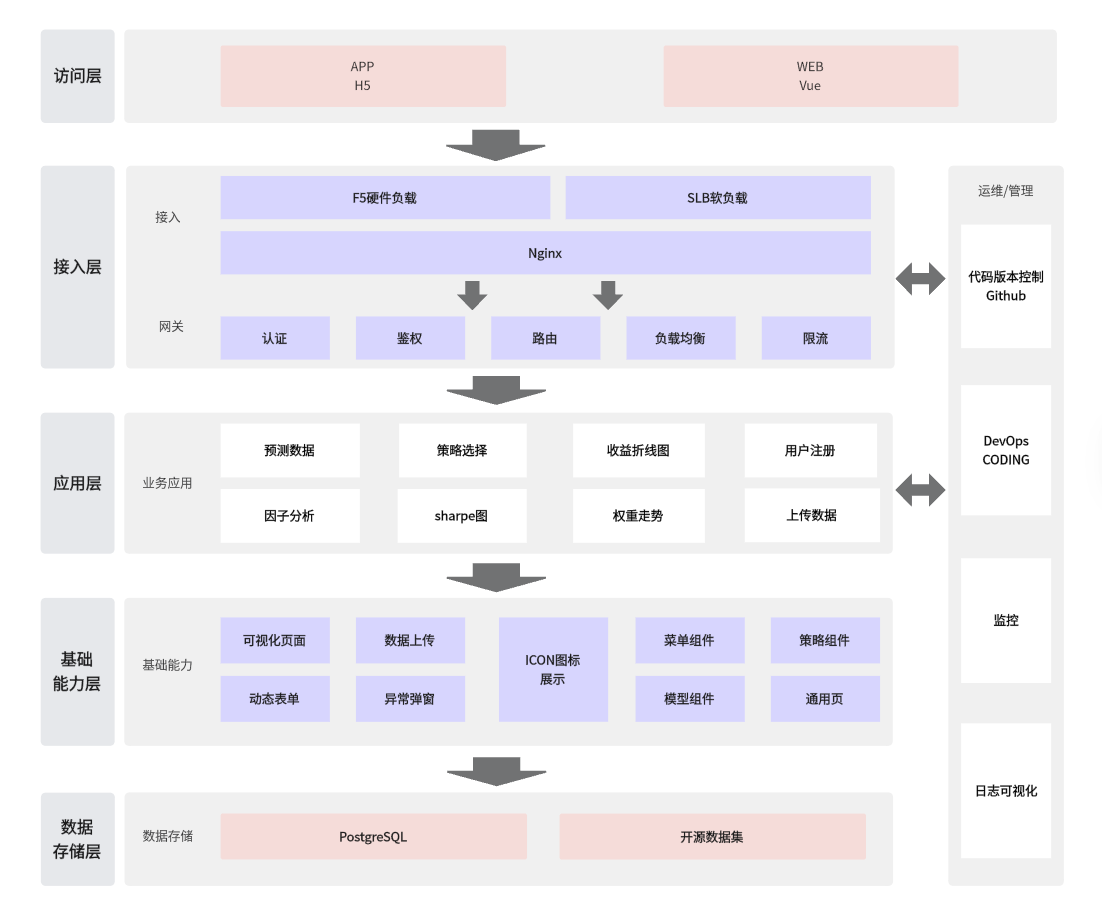
#### 5.4 数据库描述

用户表 用户数据表 股票数据表

#### 5.5 数据安全与隐私保护

### 系统架构

#### 6.1 系统架构图



#### 6.2 系统架构概述

QW是基于HTML5和VUE为前端技术；node.js和PostgreSQL为后端技术的系统。如上图所示，在访问层面，用户可以通过APP（基于H5技术）和WEB界面（采用Vue框架）便捷地访问系统。接入层有效地管理和分发来自用户的请求。网关层包括认证、鉴权、路由、负载均衡和限流功能，确保安全可靠地管理用户权限和流量控制。应用层涵盖了预测数据、策略选择、收益分析图、用户注册、因子分析、Sharpe图、权重走势和数据上传等多项核心功能，为用户提供全面的量化金融分析和决策支持。基础能力层提供了可视化页面、数据上传、动态表单、异常弹窗、ICON图标展示、菜单组件、策略组件和模型组件等通用功能，增强了系统的灵活性和用户体验。数据存储层采用开源的PostgreSQL数据库，支持高效管理和存储系统运行所需的各类数据集。这些组件共同构成了一个强大而稳健的金融分析平台，为用户提供了一体化的数据处理、分析和可视化解决方案。

### 测试策略与计划

#### 7.1 功能测试

#### 7.2 性能测试

#### 7.3 安全测试

#### 7.4 用户验收测试计划

### 实施计划

软件生命周期是指软件产品从计划到软件交付使用，直到最终退出为止的过程。包括需求分析阶段、设计阶段、编码与开发阶段、测试阶段和运行维护阶段。本项目由于时间紧、任务重，遵循高内聚，低耦合的原则将不使用瀑布模型、快速原型模型、增量模型、螺旋模型、喷泉模型等效率低下的传统开发模型，而是使用课上所讲述的敏捷开发方法。

#### 8.1、发展阶段划分

1、需求分析阶段与设计阶段

|  |  |  |
| --- | --- | --- |
| 时间 | 负责人 | 任务 |
| 07.01 | 小组全员 | 需求分析 |
| 07.02-07.03 | 小组全员 | 完善需求分析并设计系统前端页面、数据库以及整体框架 |
|  |  |  |
|  |  |  |
|  |  |  |

2、编码与开发阶段

|  |  |  |
| --- | --- | --- |
| 时间 | 负责人 | 任务 |
| 07.04- | 曹可心 | 数据收集与预处理、服务器部署 |
| 07.04- | 朴祉燕 | 前端页面搭建 |
| 07.05- | 朱陈媛 | 量化金融因子的计算与可视化画图 |
| 07.05- | 陈予曈 | 算法评估与比较 |
|  |  |  |
|  |  |  |
|  |  |  |

3、测试阶段

|  |  |  |
| --- | --- | --- |
| 时间 | 负责人 | 任务 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

4、运行维护阶段

|  |  |  |
| --- | --- | --- |
| 时间 | 负责人 | 任务 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

#### 8.2、发布计划与版本控制

|  |  |  |
| --- | --- | --- |
| 发布计划时间 | 版本号 | 发布内容 |
| 7.10 | v0.1 | 原来的所有内容+服务器部署 |
| 7.14 | v0.2 | 用户注册账号登录功能 |
| 7.19 | v0.3 | 上传数据功能，实时数据？ |
| 7.23 | v1.0 | 最终版本，美化+优化 |

### 风险管理与问题解决

#### 9.1、风险识别与评估

#### 9.2、应对策略

#### 9.3、问题追踪与解决流程

### 附录

#### 10.1、参考开源项目

#### 10.2、参考文献

[1]余湄,吴卫星,邓军,等. 国内综合院校及财经类高校量化金融专业现状和人才培养模式探讨[C]. 2017年中国高等财经教育论坛 论文集. 2017.

[2]Ludkovski M., Statistical Machine Learning for Quantitative Finance [J]. Annual Review of Statistics and Its Application. Volume 10, Issue. 2023.

[3]胡光宏. 量化投资发展分析[J]. 投资与创业,2020.

[4]刘思淼. 市场呼唤量化金融品牌[J]. 中国金融,2019.

[5]杨世林. 基于聚宽量化投资平台的股票多因子策略应用[D]. 浙江:浙江大学,2018.