1. 课题背景及简介

自从改革开放以来，我国经济以惊人的速度飞速增长，居民的收入与生活水平已经不可同日而语。伴随着世界进入科技时代，我国国民收入继续提高，经济金融运行效率也在不断提高，中等收入人群在不断扩大，居民的理财需求不再局限于结算需求和预防需求，还有资产保值增值需要。而由于大多人不懂股票、债券、期货、期权等金融工具以及信息技术的快速进步，使得各种各样的理财产品在互联网上出现，使大多人的理财需求得到满足。但人们因为不懂经济、金融，无法判断一种基金、理财产品的好坏，因此科技金融、智慧金融的概念的近些年被提了出来，主要是将大数据、云计算的技术运用于金融中，将金融的相关数据进行机器学习而发现其中的规律。目前对科技金融最权威的定义是我国《“十三五”国家科技创新规划》中的表述：落脚于金融，利用金融创新，高效可控地服务于科技创新创业的金融业态和金融产品。在互联网背景下，科技金融最重要和典型的形态就是互联网金融或者叫“互联网+金融”。[1]

1. 金融（股票）数据分析现状

（1）智能投顾

近些年，随着互联网、大数据、人工智能等信息技术的快速发展，金融科技也成为金融领域最热门的话题，很多计算机和金融的专业人士都在金融科技上不断探索，其中与财富增值相关的智能投顾也大为人们所关注。智能投顾是指一系列基于客户自身理财需求、资产状况、风险承受能力、风险偏好等因素，运用现代投资组合理论和人工智能技术，通过构建数据模型、搭建网络平台为投资者提供智能化的理财顾问服务。[2]本课题研究的方向是对客户感兴趣的股票进行全面的数据分析并且对这些数据进行基于HTML5的可视化，是智能投顾中的专科。

（2）智能需求

随着我国越来越强大，经济发展也越来越快，我国的居民越来越富有，对自己的财富进行投资的观念也越来越强。但是金融所涉及的知识体系、数据量等维度非常广，即便专业人士也不能非常容易、非常快速对金融市场、对股票进行分析，更何况非金融专业但有理财炒股需求的人士，他们不了解金融所以不知从何下手。为了降低人们投资理财的风险，避免情绪化带来的损失，以及加强经济发展趋势的了解，现在很多理财平台都开始开展金融数据的分析业务。

（3）数据挖掘与分析技术

Python语言作为近年来发展最快的高级编程语言之一，因其拥有强大的计算生态，在数据分析领域越来越受欢迎，其语法非常接近自然语言，学习容易且上手简单，可降低开发难度，过去几年中在数据分析领域非常广泛地被应用。[3]在数据挖掘中广泛运用的机器学习算法有SVM、神经网络、深度学习等用于股票预测，通过聚类、分类、决策树机器学习算法对上市公司的经营分闲进行评估等。[4]所谓的机器学习，就是从大量的可公开的历史数据当中学习到新的模型。过去十年我国银行业发生了巨大的变化，随着互联网金融风起云涌，传统银行的客户被大量分流，业务场景被逐步占领，因此各大银行也纷纷选择数字化转型，在金融数据分析平台等科技金融平台的建设方面取得巨大的进步。[5]

1. 涉及的技术点

（1）python

Python近年来在数据统计这一方面得到广泛的应用，相较于传统数据统计中应用的SAS、SPSS等统计工具，python是一种基于现代化、信息化、网络化、科学化背景下的新型统计工具，相较于传统统计工作，该工具具有简单易学，操作便利，可深入挖掘数据的相关信息等的优点。因此，在金融数据的挖掘、分析中应用python是非常有必要的。[6]

（2）java+hadoop后端大数据处理架构

对python爬虫爬取的数据，将其存储到HDFS集群中，由于Hadoop的HDFS中存储数据种类繁多，可采用spark streaming流数据处理进行数据的实时的数据处理，然后存储与hive中。hive是基于Hadoop的一个数据仓库工具，本身没有专门的数据存储格式，也没有为数据建立索引，所以数据存储在HDFS上。在本项目中，通过hive与mysql的映射，可以用Sqoop使hive与mysql进行数据交互，到此就将复杂非格式化的数据通过预处理、清洗成格式化的数据，然后python和java就可以通过接口对mysql中的数据进行增删改查，然后将数据发到客户端进行呈现。在开发过程中，大数据模块最为复杂，需要协调好各个组件的兼容性，并且需要了解一定的分布式概念，还需要对数据结构、操作系统、计算机网络、数据库比较熟悉。[7]

（3）html5

Html是在制作网页中使用的超文本标记语言，通过它能够构造出各种各样的网页结构。而html5是下一代的html，其中不仅包含的html中老标签，而且增加很多非常实用的新标签与新特性。其中HTML5的Canvas功能，具有强大且高效的绘图特性，被广泛应用于客户端的游戏开发以及图像处理中。在Web应用中通过Canvas元素引入各种图形交互方式或是直接操作的其他交互方式，这不但丰富了信息操作的方式，而且给客户带来更好的用户体验。[8]在本课题中，是通过HTML5的Canvas功能来画动态的数据分析图表，用户可以通过鼠标或键盘交互性的浏览数据。

（4）JavaScript

JavaScript是一种直译式脚本语言，是一种动态类型、基于原型的语言，内置支持类型。它的解释器被称为JavaScript引擎，为浏览器的一部分，广泛用于客户端脚本语言，最早是在HTML（标准通用标记语言下的一个应用）网页上使用，用来给HTML网页增加动态功能。

（5）MySQL数据库

MySQL是一个关系型数据库管理系统，由瑞典MySQL AB公司开发，目前属于Oracle旗下产品。MySQL是最流行的关系型数据库管理系统之一，在WEB应用方面，MySQL是最好的RDBMS（Relational Database Management System，关系型数据库管理系统）应用软件。

1. 结论

在当下金融科技大环境下，智能投顾作为科技金融中新生事物，还在不断地研究突破中。本课题研究基于web的股票全方位数据分析系统是智能投顾中的股票分析专科，如今股票市场正处于蓬勃发展之中，越来越多的股民投身于股票市场，而股票市场的快速发展与国民经济息息相关，所以对股票的数据分析系统的建设和引导股民理性投资具有重要的意义，也非常具有先驱金融科技研究意义。[9]

参考文献

[1]曹金飞,李芸达.互联网发展趋势研判及其对科技金融的影响[J].科技管理研究,2020,40(20):15-21.

[2]薛敬明. 智能金融的关键技术研究[D].国防科技大学,2018.

[3]孙艳华,冯妍,李宏然.基于Python平台的金融数据分析技术研究[J].信息与电脑(理论版),2020,32(15):155-159.

[4]薛敬明. 智能金融的关键技术研究[D].国防科技大学,2018.

[5]徐永桂.金融数据分析平台构建[J].信息技术与标准化,2020(05):42-46.

[6]方骥,谢慧敏.Python在大数据挖掘和分析中的应用研究[J].数字技术与应用,2020,38(09):75-76+81.

[7]焦向雨,黄康辉,卢峥.Hadoop+JavaWeb大数据分析可视化系统[J].中小企业管理与科技(下旬刊),2020(02):151-152.

[8]吴飞燕.基于HTML5 Canvas绘图技术应用[J].电子测试,2018(04):116+118.

[9]徐崤笑. 一种基于大数据分析与移动均线寻找股票买入点的方法[D].沈阳师范大学,2020.