

服务热线: 400-615-1233

★ 配套精品教学资料包

www.huatengedu.com.cn

金融大数据分析

JINRONG DASHUJU FENXI

高等职业教育财经商贸系列创新教材
工学结合、校企合作系列新形态教材

高等职业教育财经商贸系列创新教材

工学结合、校企合作系列新形态教材

金融 大数据分析

金融 大数据分析

主编 刘利 周启运 张淑芬
主审 马毅

主编 刘利 周启运 张淑芬

北京邮电大学出版社



ISBN 978-7-5635-7853-5



9 787563 578535 >

定价: 49.80元

策划编辑: 袁相芬
责任编辑: 张海红
封面设计: 刘文东



北京邮电大学出版社
www.buptpress.com

高等职业教育财经商贸系列创新教材

工学结合、校企合作系列新形态教材

金融 大数据分析

主 编 刘 利 周启运 张淑芬
副主编 罗 婷 王 亮 林 棠
主 审 马 毅



北京邮电大学出版社
www.buptpress.com

内 容 简 介

本书紧扣数字金融的发展趋势,以项目任务为导向,将理论知识、典型案例及实操深度融合,助力学生熟练应用金融大数据解决实际问题。编者将全书内容分为三大单元、九个项目,包括金融大数据概述、数据分析工具 Python、金融大数据的数据处理、大数据在银行业的应用、大数据在证券期货行业的应用、大数据在保险行业的应用、大数据在金融科技公司的应用、大数据金融下的商业模式创新、金融大数据监管。

本书既可作为高等职业院校财经商贸类相关专业的教学用书,又可作为金融领域从业者及对金融大数据感兴趣的读者的参考用书。

图书在版编目(CIP)数据

金融大数据分析 / 刘利, 周启运, 张淑芬主编.
北京: 北京邮电大学出版社, 2026. -- ISBN 978-7-5635-7853-5
I. F830.41
中国国家版本馆 CIP 数据核字第 20262PP861 号

策划编辑: 袁相芬 责任编辑: 张海红 封面设计: 刘文东

出版发行: 北京邮电大学出版社
社 址: 北京市海淀区西土城路 10 号
邮政编码: 100876
发 行 部: 电话: 010-62282185 传真: 010-62283578
E-mail: publish@bupt.edu.cn
经 销: 各地新华书店
印 刷: 三河市骏杰印刷有限公司
开 本: 787 mm×1 092 mm 1/16
印 张: 14.75 插页 1
字 数: 305 千字
版 次: 2026 年 月第 1 版
印 次: 2026 年 月第 1 次印刷

ISBN 978-7-5635-7853-5

定 价: 49.80 元

• 如有印装质量问题,请与北京邮电大学出版社发行部联系 •

服务电话: 400-615-1233

金融大数据以前所未有的速度与规模重塑着金融领域的运作模式、服务形态与竞争格局,深刻影响着全球经济的发展脉络。

党的二十届三中全会审议通过的《中共中央关于进一步全面深化改革 推进中国式现代化的决定》中提出:“积极发展科技金融、绿色金融、普惠金融、养老金融、数字金融,加强对重大战略、重点领域、薄弱环节的优质金融服务。完善金融机构定位和治理,健全服务实体经济的激励约束机制。”在数字技术已成为经济转型升级核心引擎的时代背景下,数字金融顺势而为、蓬勃发展,为经济高质量发展注入强劲动能。

本书共分为三大单元,第一单元为基础篇,内容涵盖金融大数据概述、数据分析工具 Python、金融大数据的数据处理;第二单元为实务篇,内容涵盖大数据在银行业的应用、大数据在证券期货行业的应用、大数据在保险行业的应用、大数据在金融科技公司的应用;第三单元为创新篇,内容涵盖大数据金融下的商业模式创新、金融大数据监管。

本书在编写过程中始终注重理论与实践的有机结合,不仅收录了丰富的行业案例并进行深度分析,还精心设计了多个实战演练环节,助力学生巩固所学知识;而在知识小结部分,又对本项目的核心知识点进行了系统回顾与提炼总结,以帮助学生加深理解与记忆。

本书具有以下特色。

(1) 结构清晰。全书分为基础篇、实务篇和创新篇,从基础知识铺垫到实际场景应用,再到商业模式创新与监管探讨,层次分明,逻辑连贯,便于学生构建系统的知识体系并逐步掌握相关技能。

(2) 从零开始。即使是无编程基础的学生,也能通过本书循序渐进式的引导,从 Python 的基础知识入手,逐步学习 Python 软件的安装、Python 编程中的函数与库,以及 Python 在金融领域的应用,稳步提升利用 Python 进行数据分析的实操能力。

(3) 大数据应用全面。本书不仅覆盖互联网金融、银行、保险等核心金融行业的数据分析与应用场景,还深入探讨了大数据时代下企业战略分析思维的转变及管理模式的变革,为学生提供多维度、全方位的大数据应用视角。

本书旨在为金融领域从业者、相关专业师生及对金融大数据感兴趣的广大读者搭建一个全面且实用的知识体系框架,助力其系统掌握金融大数据的核心

理论与实操技能。

本书由佛山职业技术学院的刘利、周启运、张淑芬担任主编,佛山职业技术学院的罗婷、王亮、林棠担任副主编,佛山职业技术学院的马毅担任主审,深圳市前海泓来资产管理有限公司李祖锋参与编写。具体分工如下:刘利负责项目一、项目二和项目三的编写;周启运负责项目四和项目八的编写;张淑芬负责项目六的编写;罗婷、王亮、林棠共同负责项目五、项目七和项目九的编写;李祖锋负责全书行业前沿资料的整理与汇编。

在本书的编写过程中,编者秉持严谨治学的态度,广泛查阅行业前沿资料,深度结合实际案例,力求使知识讲解准确翔实、逻辑清晰。但鉴于金融大数据发展迅速、知识更新迭代频繁,加之编者的学术水平与实践经验有限,书中难免存在不足之处。恳请广大读者指正,您的宝贵意见将助力本书不断修订完善,以便更好地服务于读者对金融大数据知识的学习与探索,共同推动金融大数据知识不断传播与发展。

编 者

第一单元 基础篇

项目一 金融大数据概述 2

- | | |
|---------------------|----|
| 任务一 认识大数据 | 3 |
| 任务二 认识金融大数据 | 13 |
| 任务三 认识金融大数据的发展趋势及应用 | 17 |

项目二 数据分析工具 Python 26

- | | |
|---------------------|----|
| 任务一 Python 的基础知识 | 27 |
| 任务二 Python 软件的安装 | 31 |
| 任务三 Python 编程中的函数与库 | 40 |
| 任务四 Python 在金融领域的应用 | 63 |

项目三 金融大数据的数据处理 68

- | | |
|--------------|----|
| 任务一 金融大数据的采集 | 69 |
| 任务二 金融大数据的清洗 | 73 |
| 任务三 金融大数据的挖掘 | 81 |

第二单元 实 务 篇

项目四 大数据在银行业的应用 90

任务一	认识商业银行	91
任务二	大数据下的银行信贷业务	99
任务三	业务经营与客户关系管理创新	105
任务四	银行贷款质量的评估	110

项目五 大数据在证券期货行业的应用 117

任务一	认识证券业务和期货业务	118
任务二	证券市场的情绪度量	122
任务三	大数据下的证券投资分析	125
任务四	大数据下的量化投资	130

项目六 大数据在保险行业的应用 136

任务一	认识保险业务	138
任务二	基于大数据的保险风险识别与评估	146
任务三	基于大数据的保险产品设计与定价	153
任务四	基于大数据的保险产品精准营销	159
任务五	基于大数据的保险产品理赔	161

项目七 大数据在金融科技公司的应用 169

任务一	大数据与第三方支付	170
任务二	大数据与 P2P	176
任务三	大数据与互联网消费金融	180
任务四	大数据与供应链金融	183

第三单元 创 新 篇**项目八 大数据金融下的商业模式创新 192**

任务一	认识商业模式	194
任务二	商业模式创新带来的机遇	196
任务三	商业模式创新的特点	197
任务四	基于大数据的商业模式框架	198
任务五	大数据时代商业模式的创新方向	200

项目九 金融大数据监管 203

任务一	监管金融企业对大数据的应用	204
任务二	大数据对金融市场的监管	207
任务三	大数据与金融信息安全	216

参考文献 230

第一单元



基础篇



项目一

金融大数据概述



项目导读

在数字化浪潮席卷之下,大数据已成为推动各行业变革与发展的核心力量,金融领域也不例外。本项目将带领学生走进金融大数据的奇妙世界,了解其在金融行业的无限潜能。



学习目标

【知识目标】

了解大数据的概念,熟悉其发展历程,尤其大数据在中国的发展历程及其标志性事件。

理解大数据的特征、数据结构类型及计算模式。

掌握大数据的处理流程与分析类型。

了解大数据时代的思维变革。

掌握金融大数据的概念、特征及理论基础。

了解金融大数据的发展趋势,熟悉其在银行、证券、保险等行业的应用现状。

【能力目标】

具备大数据识别能力,能够区分数据类型,为数据处理分析做准备。

能够熟练掌握与大数据处理、分析相关的技能,合理选用大数据的计算模式。

能够运用金融大数据理论解决实际问题,为金融机构发展提供建议。

能够培养数据思维与创新能力,创新金融服务模式和产品设计。

【素质目标】

强化数据安全与隐私保护意识,依法依规处理数据。

秉持严谨的科学态度,确保数据真实客观,杜绝造假。

提高团队协作与沟通能力,助力金融大数据项目的开展。

树立可持续发展理念,利用大数据推动绿色金融的发展。

思维导图



课前预习

在正式学习本项目内容之前,请完成以下预习任务,以便更好地理解和掌握本项目的知识点。

(1)查阅资料。查阅权威资料,记录大数据的概念、起源时间及最初受到重视的行业。了解大数据与传统数据在处理方式上可能存在的差异。

(2)梳理事件。绘制大数据在全球及中国发展历程的时间轴,标注出关键时间节点和标志性事件。

(3)思考问题。思考以下问题,带着问题进入本项目的学习。

①什么是大数据? 在我们的日常生活中大数据有哪些作用?

②大数据时代下,金融专业的学生需要具备什么专业知识和技能才能适应行业企业的人才需求?

任务一 认识大数据

任务情景

在互联网购物时,大家是否注意到,电商平台总能精准地向你推荐心仪的商品?例如,你近期浏览了某品牌的运动鞋,下次打开购物 App,页面上就会出现类似款式或同品牌其他产品的推荐。这背后,正是大数据在发挥作用。大数据无处不在,它已经渗透到各个行业和我们生活的方方面面。但你是否想过,大数据究竟是什么?本任务将深入探索大数据的基本知识与发展历程。

任务分析

大数据相关知识繁杂且应用广泛,并且在当今数字时代发挥着极为关键的作用。我们要从理解大数据的基本知识出发,结合其在各领域的实际表现,系统地分析、归纳和总结其特征、类型、计算模式、处理流程等内容,正确运用相关理论和实际案例辅助学习,做到全面掌握。

任务实施



动画

大数据认知初体验

一、大数据的概念与发展历程

(一)大数据的概念

大数据泛指巨量的数据集合,因可从数据集中挖掘出有价值的信息而受到重视。本书认为,大数据是指无法在一定时间范围内用常规软件工具进行捕捉、管理和处理的数据集合,是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产。

(二)大数据的发展历程

1. 大数据在国外的发展历程

大数据的概念起源于美国。20世纪90年代至21世纪初是大数据发展的萌芽时期。21世纪初至2010年,互联网行业迎来了飞速发展的时期,信息技术也不断地推陈出新,大数据最先在互联网行业得到重视。这一阶段是大数据的发展时期,大数据开始展现活力。2011年至今是大数据的兴盛时期。2011年,国际商业机器公司(International Business Machines Corporation, IBM)研制出了沃森超级计算机,以每秒扫描并分析4TB的数据量打破了当时的世界纪录,大数据计算迈向了一个新的高度。紧接着,麦肯锡全球研究院(McKinsey Global Institute, MGI)发布了题为“海量数据,创新、竞争和提高生产率的下一个前沿”(Big Data: The next frontier for innovation, competition, and productivity)的研究报告,详细介绍了大数据在各个领域中的应用情况,以及大数据的技术架构,提醒各国政府为应对大数据时代的到来,应尽快制定相应的战略。2012年,世界经济论坛在瑞士达沃斯召开,论坛将大数据作为专题讨论的主题之一,发布了《大数据、大影响:国际发展新的可能性》等系列报告,向全球正式宣布大数据时代的到来。

2. 大数据在中国的发展历程

大数据在中国的发展也经历了以下四个阶段。

(1)初期阶段(2008年以前)。1997年,中国互联网络信息中心(China Internet Network Information Center, CNNIC)成立,致力于对中国互联网发展进行研究和管理。2007年,

伴随着社交网络用户数量的激增,技术博客和专业人士为大数据的概念注入了新活力。在这一背景下,大数据的概念逐渐被人们了解,但相关技术和应用尚未发展起来。

(2)成长阶段(2009—2012年)。截至2009年12月31日,CNNIC的统计数据显示:中国网民规模达到3.84亿人,互联网普及率达到28.9%。宽带网民规模达到3.46亿人,国际出口带宽达866 367 Mbps。大数据技术开始在国外被广泛应用,中国的大数据也在此时间段进入成长阶段。各个行业都开始尝试运用大数据,同时,一些大数据企业也开始在国内崛起,如华为技术有限公司、北京百度网讯科技有限公司(以下简称“百度”)、阿里巴巴集团控股有限公司(以下简称“阿里巴巴”)、深圳市腾讯计算机系统有限公司(以下简称“腾讯”)等,这些企业在大数据技术方面不断突破,也逐渐将大数据应用到自身业务当中。

2012年10月,中国计算机学会成立了大数据专家委员会,专注研究大数据中的科学与技术问题。科技部印发的《中国云科技发展“十二五”专项规划》和工业和信息化部印发的《物联网“十二五”发展规划》等都将大数据技术作为一项重点工作予以支持。

(3)爆发阶段(2013—2015年)。随着一系列标志性事件的发生,人们越发感受到大数据时代的力量。“BAT”(百度、阿里巴巴、腾讯)各显身手,分别推出创新大数据应用,因此2013年被许多国外媒体和专家称为“中国大数据元年”。

自2014年起,大数据首次被写进我国政府工作报告,大数据产业被提升至国家战略层面。2015年8月,国务院发布《促进大数据发展行动纲要》,其中主要任务之一就是“加快政府数据开放共享,推动资源整合,提升治理能力”,并明确了时间节点。

(4)快速发展阶段(2016年至今)。2016年3月,国家发展改革委、工业和信息化部、中央网信办同意贵州省建设国家大数据(贵州)综合试验区,这是首个国家级大数据综合试验区。

2016年12月,中国信息通信研究院(工业和信息化部电信研究院)发布《大数据白皮书(2016年)》,指出:“大数据是国家基础性战略资源,是21世纪的‘钻石矿’。党中央、国务院高度重视大数据在经济社会发展中的作用,提出‘实施国家大数据战略’,出台《促进大数据发展行动纲要》,全面推进大数据发展,加快建设‘数据强国’。”此后,国家大数据综合试验区逐渐建立起来,相关政策与标准体系不断完善。

2017年跨部门数据资源共享共用格局基本形成。

2018年建成政府主导的数据共享开放平台,打通政府部门、企事业单位间的数据壁垒,并在部分领域开展应用试点。

2019年9月,大数据产业生态联盟联合赛迪顾问发布《2019中国大数据产业发展白皮书》,指出2018年中国大数据产业规模为4 384.5亿元,同比增长23.5%。

2020年实现政府数据集的普遍开放,国务院和国务院办公厅又陆续印发了一系列文件,推进政务信息资源共享管理、政务信息系统整合共享、互联网+政务服务试点、政务服务一网一门一次改革等,推进跨层级、跨地域、跨系统、跨部门、跨业务的政务信息系统整合、互联、协同和数据共享,用政务大数据支撑“放管服”改革落地,建设数字政府和智慧政府。目

前,我国政务领域的数据开放共享已取得了重要进展和明显效果。例如,“最多跑一次”“一网通办”“一网统管”“一网协同”等创新实践不断涌现。

根据国际数据公司(International Data Corporation, IDC)2025年V1版本《全球大数据支出指南》披露,中国大数据市场正以全球领先的速度实现爆发式增长,长期增长动能持续强劲。数据显示,2022年中国大数据市场总体信息技术的投资规模为170亿美元;2024年全球大数据信息技术的总投资达3540亿美元,其中中国市场已明确超越亚太地区(除中、日外)市场总和,成为区域核心增长极。IDC进一步上调对中国大数据市场规模的预测:自2022年中国大数据市场支出为170亿美元,2025年预计突破300亿美元,2026年增至500亿美元以上,2028年达621.7亿美元(约合4486亿元人民币),2030年有望逼近1.2万亿元人民币,较2022年实现十年增长超6倍。此次预测中,2026的数值较2023年V1版本的364.9亿美元的预期提升近70%,充分彰显中国大数据市场的强劲发展潜力。中国大数据市场支出及其预测(2021—2026年)如图1-1所示。

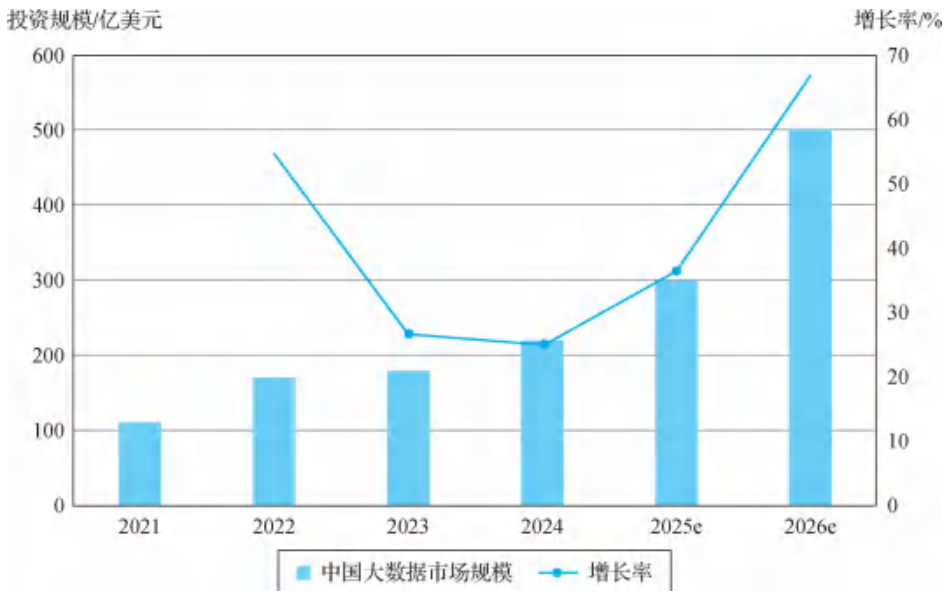


图 1-1 中国大数据市场支出及其预测(2021—2026年)

自提出大数据国家战略以来,中国在大数据领域的发展已在行业及企业应用层面引起广泛关注,并展现出巨大的发展潜力。在大数据、云计算、人工智能等新一代信息技术的驱动下,预计大数据将继续对我国经济发展和产业升级产生深远影响。可以预见,中国大数据领域未来将持续迎来更加繁荣的发展态势。

二、大数据的特征与数据结构类型

(一)大数据的特征

随着大数据技术的不断发展,数据的复杂程度越来越高,企业在尝试分



微课

大数据的特征

析现有海量信息以推动业务价值增值时,必定会采用大数据技术。业界通常用 8 个 V 来概括大数据的特征。大数据的 8V 特征如图 1-2 所示。



图 1-2 大数据的 8V 特征

(1)数量(Volume)。数量是指大规模的数据量。大数据的数据量大,包括采集、存储和计算的量都非常大。目前,大数据的一般起始计量单位至少是 P(1 000 个 T)、E(100 万个 T)或 Z(10 亿个 T)。

数据的最小单位是 bit。按数据的大小顺序列出所有数据单位,即 bit、Byte、KB、MB、GB、TB、PB、EB、ZB、YB、BB、NB、DB,它们按照进率 1 024(2 的十次方)计算。

(2)种类(Variety)。大数据的种类繁多且来源广泛,涵盖了多种不同格式和类型的数据集合。数据的来源多样,包括机器在人与系统交互过程中自动生成的数据。数据来源的多样性导致了数据类型的多样性。数据类型主要分为结构化、半结构化和非结构化数据,具体表现为网络日志、音频、视频、图片、地理位置信息等多种形式。这种多类型的数据对数据处理能力提出了更高的要求。

(3)价值(Value)。数据的价值密度相对较低,可谓是历经筛选而愈发珍贵。随着互联网和物联网技术的广泛应用,信息的感知已经无所不在。尽管信息量庞大,但其价值密度并不高。如何结合业务逻辑,利用强大的机器算法来挖掘数据的价值,成为大数据时代亟须解决的核心问题。此外,数据的价值还与数据的真实性及数据处理的时间紧密相关。

(4)速度(Velocity)。在大数据背景下,数据通常以动态且迅速生成的数据流形式出现。随着数据存储和采集速度的提升,数据处理能力也相应增强,展现出极强的时效性。用户必须精准掌握数据流的管理,以便有效利用这些数据。例如,搜索引擎需要确保几分钟前的新闻内容能够被用户检索到,而个性化推荐算法则力求实现即时的推荐效果。这些特点显著区分了大数据与传统数据挖掘的不同。

(5)准确性(Veracity)。数据的准确性即数据的质量。大数据环境下的数据最好具有较高的信噪比,信噪比越高的数据,准确性越高。

(6)可变性(Variability)。可变性指数据的变化,意味着相同的数据在不同的上下文中可能具有不同的含义。分析算法需要理解上下文,并发现该上下文中数据的确切含义和价值。

(7)波动性(Volatility)。波动性是指数据有效存储的时间,这对于实时分析尤为重要。波动性要求确定数据的目标时间窗口,以便分析人员可以专注于特定问题,并从分析中获得良好的性能。

(8)可视化(Visualization)。可视化使数据易于理解,通过普通的图形或多维视图使大量数据易于理解。可视化需要大数据分析师和业务领域专家之间的大量交互对话和共同努力,以使可视化变得有意义。

(二)大数据的数据结构类型

大数据的数据结构类型包括结构化数据、半结构化数据和非结构化数据。

1. 结构化数据

结构化数据通常是指能够利用关系型数据库进行表示和存储的数据,它们可以通过二维表的形式进行逻辑表达和实现。在结构化数据的范畴内,数据的结构往往先于数据本身存在。这类数据的存储和排列具有高度的规律性,从而便于执行查询和修改等操作。

2. 半结构化数据

半结构化数据是介于完全结构化数据(如关系型数据库、面向对象数据库中的数据)和完全无结构的数据(如音频、图像文件等)之间的数据。半结构化数据[如可扩展标记语言(extensible markup language,XML)、超文本标记语言(hypertext markup language,HTML)文档等]也是有结构的数据,与结构化数据不同的是,半结构化数据是先有数据,再有结构。

3. 非结构化数据

非结构化数据没有固定结构的数据。文本、图片、各类报表、图像、音频和视频等都属于非结构化数据。

根据艾瑞咨询发布的《2022年中国数智融合发展洞察》,随着数字化不断推进,全球数据量以超过59%的年增长率快速增长,中国的数据量增速比全球更快。其中,80%是非结构化和半结构化数据。并且,在企业内部,非结构化数据增速远快于结构化数据,非结构化数据占比也越来越高。企业内结构化数据与非结构化数据占比及使用情况如图1-3所示。

长期以来,由于技术的限制,企业对结构化数据的利用程度一直高于非结构化数据。然而,非结构化数据不仅体量庞大,而且蕴含着更为丰富的信息,实际上是企业尚未充分挖掘的宝贵资源。

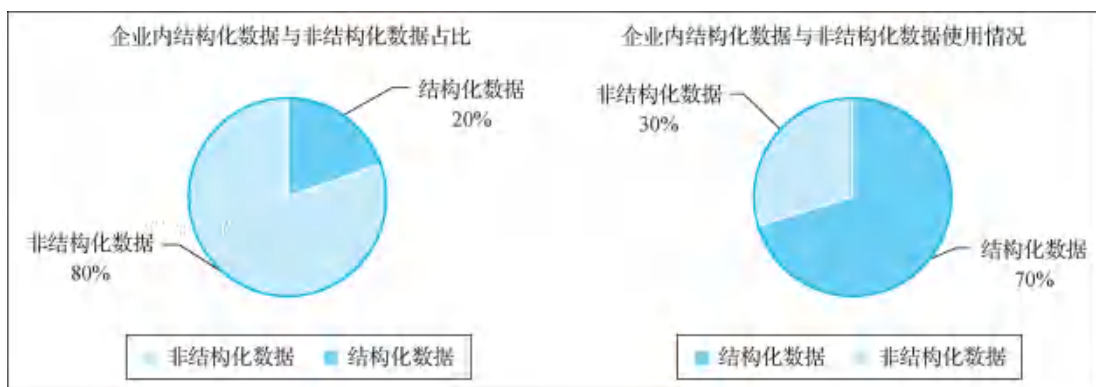


图 1-3 企业内结构化数据与非结构化数据占比及使用情况

结构化数据、半结构化数据及非结构化数据的特点对比如表 1-1 所示。

表 1-1 结构化数据、半结构化数据及非结构化数据的特点对比

类 型	特 点
结构化数据	数据结构字段含义确定、清晰,典型的如数据库中的表结构
半结构化数据	具有一定结构,但语义不够确定,典型的如 HTML 文档,有的字段确定,有的字段不确定
非结构化数据	杂乱无章的数据,很难按照一个概念去进行抽取,无规律性

三、大数据的计算模式

大数据的计算模式主要有批处理计算、流式计算、图计算和查询分析计算。

(一)批处理计算

批处理计算是指对存储的静态数据进行大规模并行批量处理的计算模式,是一种批量、高时延、主动发起的计算。在数据实时生成的情况下,假设存在一种场景:当我们拥有一个非常强大的硬件系统,能够以毫秒级速度处理 Gb 级别的数据,那么批量计算也可以以毫秒级速度得到统计结果。批处理计算的代表产品有 MapReduce、Spark 等。

(二)流式计算

流式数据是指随时间分布和数量上无限的一系列动态数据的集合体,数据价值随时间流逝而降低,必须采用实时计算方式给出响应。流式计算可以实时处理多源、连续到达的流式数据,并实时分析处理。流式计算的代表产品有 Storm、Flink 等。

(三)图计算

图计算是指一类在实际应用中非常常见的计算类型。许多大数据都是以大规模图或网络的形式呈现,如社交网络、传染病传播途径、交通事故对路网的影响。许多非图结构的大

数据,也常常会被转换为图模型后进行分析。图数据结构很好地表达了数据之间的关联性。要处理规模巨大的图数据,传统的单机方式已经无力应对,必须采用大规模机器集群构成的并行数据库。图计算的代表产品有 Pregel、GraphX 等。

(四)查询分析计算

针对超大规模数据的存储管理和查询分析,提供实时或准实时的响应。超大规模数据比大规模数据的量更庞大,多以 PB 级别计量。查询分析计算的代表产品有 Dremel、Hive 等。

上述四种大数据的计算模式的总结对比如表 1-2 所示。

表 1-2 四种大数据的计算模式的总结对比

大数据的计算模式	解决的问题	代表产品
批处理计算	针对静态数据进行大规模并行批量处理	MapReduce、Spark 等
流式计算	针对流式数据的实时处理	Storm、Flink 等
图计算	针对大规模图数据的处理	Pregel、GraphX 等
查询分析计算	针对超大规模数据的存储管理和查询分析	Dremel、Hive 等

四、大数据的处理流程与分析类型

(一)大数据的处理流程

大数据是一项能够对数量巨大、来源分散、格式多样的数据进行采集、存储和关联性分析的新一代信息系统架构和技术。大数据处理流程主要包括数据采集、数据预处理、数据存储、数据分析与挖掘、数据可视化、数据应用等环节,大数据的处理流程示意图如图 1-4 所示。

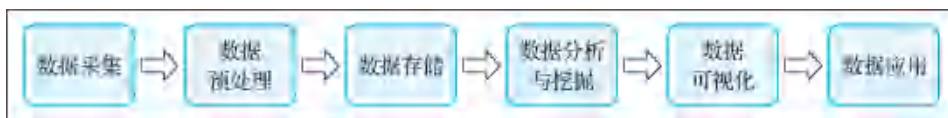


图 1-4 大数据的处理流程示意图

其中,数据质量贯穿整个大数据流程,每一个大数据处理环节都会对大数据的质量产生影响。

1. 数据采集

目前,行业对数据采集有两种解释:一种是将数据从无到有的过程(如 Web 服务器打印的日志、自定义采集的日志等)称为数据采集;另一种是将使用 Flume 等工具把数据采集到指定位置的过程称为数据采集。

在数据采集过程中,数据源会影响大数据质量的真实性、完整性、一致性、准确性和安全性。对于 Web 数据,多采用网络爬虫的方式进行收集,这需要对爬虫软件进行时间设置,以

保障收集到的数据的时效性。

2. 数据预处理

数据预处理主要包括数据清理、数据集成、数据归约与数据转换等内容,可以大大提高大数据的总体质量。

(1)数据清理。数据清理包括对数据的不一致检测、噪声数据识别、数据过滤与修正等工作,有助于提高大数据的一致性、准确性、真实性和可用性等。

(2)数据集成。数据集成是将多个数据源的数据进行整合,形成集中、统一的数据库或数据立方体的过程,有利于提高大数据的完整性、一致性、安全性和可用性。

(3)数据归约。数据归约是在不影响分析结果准确性的前提下,通过维归约、数据抽样等技术缩小数据集规模、简化数据结构的过程,有利于提高大数据的价值密度,即提升数据存储的价值性。

(4)数据转换。数据转换包括基于规则或元数据的转换、基于模型与学习的转换等技术,可通过标准化处理实现数据格式统一,有利于增强大数据的一致性和可用性。

3. 数据存储

数据存储是指通过存储器存储采集到的数据,并建立相应数据库进行管理与调用,主要解决大数据存储、处理及有效传输等关键问题,通常采用高性能、高吞吐率、大容量的基础设施存放数据。

4. 数据分析与挖掘

大数据分析技术主要涵盖对现有数据的分布式统计分析技术,以及对未知数据的分布式挖掘技术和深度学习技术。其中,分布式统计分析技术可通过数据处理工具实现,分布式挖掘技术与深度学习技术则在大数据分析阶段被深度应用。这些技术包括聚类与分类、关联分析和深度学习等,能够揭示大数据集合中数据的内在关联性,构建描述事物特征的模式或属性规则。通过搭建机器学习模型并利用海量训练数据,可进一步提高数据分析与预测的准确性。

5. 数据可视化

数据可视化是指将大数据分析与预测结果以计算机图形或图像的直观形式呈现给用户,并支持用户交互式操作的过程。数据可视化技术有助于发现大量业务数据中隐含的规律性信息,为管理决策提供支撑。因此,数据可视化是影响大数据可用性和易于理解性的关键因素。

6. 数据应用

数据应用是指将经分析处理后得到的大数据结果应用于管理决策、战略规划等实际场景的过程,既是对大数据分析结果的检验与验证,也直接体现了大数据分析处理结果的价值性和可用性。数据应用对分析处理过程具有引导作用。在开展大数据收集、处理等操作前,通过充分调研应用情境、深入分析管理决策需求,明确大数据处理与分析的目标,为数据收

集、存储、处理、分析等环节提供清晰方向,进而保障大数据分析结果的可用性、价值性及对用户需求的满足。

(二)大数据的分析类型

大数据的分析类型主要包括描述性分析、诊断性分析、预测性分析和决策性分析。

1. 描述性分析

描述性分析是最基础、最常见的分析方法,主要通过绘制分布图、聚类分析、算术描述、相关性分析等手段,研究变量之间的关系,考察数据的概率分布特征。例如,某企业收集过去5年每一季度的产品销量数据,经整理与可视化分析后得出结论:“近5年来,该产品市场表现持续低迷,销量呈逐年下滑趋势”,这一过程即属于描述性分析,核心作用是清晰呈现“事件是什么”。

2. 诊断性分析

诊断性分析是描述性分析的延伸,旨在探究现象背后的原因。若描述性分析回答“发生了什么”,诊断性分析则聚焦“为什么会发生”。

延续上述案例,当企业明确“产品销量近5年持续下滑”的现象后,需进一步通过细分维度(如区域、客群、渠道)、交叉维度分析等方法定位原因,最终发现“产品存在重大质量问题”是导致销量持续低迷的核心因素,这一分析过程即为诊断性分析。

3. 预测性分析

预测性分析是在描述性分析、诊断性分析基础上的进阶应用,通过分析历史数据、构建预测模型,对未来数据趋势或结果进行推断。常用的标准机器学习算法(如分类、回归、聚类技术)可用于生成预测模型,支撑未来结果预测。例如,某大型零售企业基于客户历史购买模式,利用预测性分析技术预测客户未来消费行为、商品库存需求、客户潜在关联购买偏好等,据此判断销售趋势,进而提供个性化推荐、预测季度或年度销售总量,为库存管理与营销决策提供依据。

4. 决策性分析

决策性分析是利用数据分析辅助决策者制定明智、有效的策略,以实现最优结果的技术,核心是通过决策树、遗传算法、灰色关联模型等评估算法筛选最优方案,用数据解决实际决策问题。例如,企业利用城市人口分布数据,分析居民时空分布特征,在大规模线上广告投放中精准定位人群聚集区域,以最低投放成本触达最多目标人群,减少不必要的成本支出,这一过程即运用了决策性分析。

五、大数据时代的思维变革

思维是个人职业属性的重要体现,如同新闻传媒从业者具备热点追踪思维、法律从业者具备法律逻辑思维、金融行业从业者具备金融专业思维一样,大数据时代的到来也推动着思

维模式的革新。这种变革并非摒弃原有思维,而是在保留传统优势的基础上,学习并掌握新型思维方式。当前我们处于数据爆炸的时代,具备良好的数据思维,对日常生活、职业工作及未来发展均具有重要意义。大数据时代的思维变革体现在四个方面,即一切皆可量化、全样而非抽样、相关而非因果、概率而非精确。

任务二 认识金融大数据



任务情景

金融大数据已成为金融行业创新发展的核心驱动力,它将如何改变传统的金融模式?有着怎样的理论基础?未来又将朝着哪些方向发展?带着这些问题,本任务将探索金融与大数据深度融合背后的奥秘。



任务分析

金融大数据融合金融与大数据的相关知识,对金融行业发展的意义重大。需从剖析其基本概念着手,结合金融业务特点,理解其特征、理论基础,探究其发展趋势与应用,借助实际案例辅助学习,实现对知识的全面掌握。



任务实施

一、金融大数据的概念

对于金融大数据的概念,不同领域的专家界定可能存在差异。结合不同领域专家的定义,本书认为金融大数据是指运用大数据技术开展金融服务,即集合大规模结构化、半结构化、非结构化数据,通过互联网、云计算和数据挖掘等信息处理的方式进行实时分析,向客户提供全方位信息的技术。同时,通过分析和挖掘客户交易记录、客户消费习惯等信息,预测客户行为,进而结合传统金融服务开展资金融通、创新金融服务的模式。

二、金融大数据的特征

大数据技术与金融领域相结合,将彻底改变传统金融服务模式,重构金融产业价值链。与传统金融相比,大数据背景下的金融行业具有显著不同的特征,金融行业将呈现数字化、开放性、富有极高的生产力、决策更具科学性。



动画

认知金融大数据



微课

金融大数据的特征

（一）数字化

从长远来看,数字化和网络化的深入发展将极大改变金融行业,大数据的应用将改变传统金融机构的资金中介职能,使其呈现电子化、虚拟化的交易特征。整个金融行业未来的发展方向将是虚拟化,全面变革当前的金融服务形态,具体体现在以下三个方面。

1. 产品的虚拟化

传统资金流将逐渐转化为数据信号的交换,电子货币等数字化金融产品将成为经济生活的主流。

2. 服务的虚拟化

传统人工服务将逐渐被移动互联网、全息仿真技术等科技手段替代,银行等金融机构可通过完全虚拟的渠道,更广泛地向客户提供金融服务。

3. 流程的虚拟化

银行业务流程中的各类单据、凭证等,将逐渐从传统纸质形式转变为数字文件形式处理,大幅提升工作效率与操作便利性。

可以合理预见,在大数据时代,传统金融机构将在管理理念、运营方式等多个方面面临挑战;未来金融机构的整体运作将围绕数据“洪流”展开,“数字金融”将全面落地。

（二）开放性

在大数据时代背景下,金融机构已不再天然占据社会经济信息中心的地位,企业也不再单纯依赖向金融机构提供信息以获取信用。社交网络、物联网、搜索引擎、移动互联网、大数据、云计算等新兴技术,已改变信息的产生、传播、处理和应用方式。这些变化对传统金融机构的业务发展构成了重大挑战。因此,金融机构必须摒弃以往“自然、被动的社会经济信息收集中心”的角色,转而以开放姿态与客户平等交流,主动收集客户信息。

（三）富有极高的生产力

在未来的经济活动中,大数据将与物质资本、人力资本共同成为生产过程中的重要生产要素,可转化为现实生产力并创造巨大的经济价值。随着大数据的广泛应用,开放、数字化的金融机构将实现更高生产力,具体体现在以下四个方面。

1. 降低经营成本

信息技术推动金融产品交易虚拟化,使金融供应链向外延伸,降低全社会融资成本与财务费用,提升整个市场的生产效率。

2. 提高营销的精确度和转化率

通过分析客户画像与行为数据,精准定位目标客群,实现个性化营销,提升营销的转化率。

3. 提高风险控制能力

依托海量数据构建风险评估模型,实时监测交易风险,提前识别潜在风险点,降低违约概率。

4. 促进业务产品创新

基于客户需求数据挖掘,开发适配市场的新型金融产品(如个性化理财产品、场景化信贷产品),拓展业务边界。

(四)决策更具科学性

大数据的客观性和价值性将彻底重塑传统金融机构的决策机制,为金融机构的运营管理提供全面、及时的决策支持信息。金融机构可从每一个经营环节中挖掘数据价值,通过大数据分析深入了解客户行为特征、客户群体网络行为模式,优化运营流程并推动业务创新。

在传统金融机构中,决策模式主要依赖样本数据分析与高层管理者的经验。然而,在大数据时代,全量数据分析的应用使分析结果更客观,决策支持性更强。金融机构的决策过程将逐渐以数据为核心,实现精准判断。通过运用大数据分析技术,金融机构能够对庞大的结构化数据与非结构化数据进行深入分析、评估和提取,从而及时且准确地识别业务与管理领域中的潜在机会与风险,为业务拓展和风险控制提供关键的决策支持。

三、金融大数据的理论基础

(一)信息经济学

信息经济学主要研究信息不对称对经济活动的影响。美国经济学家、诺贝尔经济学奖得主约瑟夫·斯蒂格利茨(Joseph Stiglitz)、乔治·阿克洛夫(George Akerlof)等学者指出,在市场经济活动中,市场参与者对信息的了解存在差异,掌握充分信息的市场参与者往往处于有利地位,而信息匮乏的一方则处于不利境地。市场中卖方比买方更了解商品信息,信息不对称决定了竞争是不完全的,私有信息对交易具有重要作用。如同旧车市场中存在的劣币驱逐良币现象,正是信息不对称导致的典型结果。

金融行业的核心是经营与化解风险,而金融业的风险主要源于客户信息的不完全与不对称,因此金融业的演进与信息技术创新密切相关。众多研究表明,金融市场并非完全有效市场,信息不对称广泛存在,如股票发行方与认购方对企业信息的掌握程度不同,银行借贷双方对借款者信用状况的了解存在差异。信息不对称会导致事前的逆向选择(如高风险客户更倾向于申请贷款)与事后的道德风险(如借款者获得贷款后可能违规使用资金)。以借贷业务为例,申请固定利率贷款的群体中,信用状况较差的人占比较高,且这类人群在获得贷款后违约概率也更高。

金融大数据为解决金融市场信息不对称问题提供了有效路径:通过对大量跨领域、跨时期数据的分析,挖掘数据规律以优化决策。例如,银行通过分析全部借贷数据,识别违约率最高人群的共同特征(如教育程度、婚姻状况、所在地区发达程度等),有针对性地减少对这类人群的放贷额度或增加风险控制附加条款,以降低信贷风险。

(二)金融中介理论

金融中介理论认为,金融中介的主要作用是生产、传递和处理信息。在过去,主要依靠银行、券商、保险等传统金融中介机构收集信息。但在大数据时代,信息来源渠道大幅拓展,掌握用户大数据的企业具备了成为新型金融中介的基础条件。

从信息传递与处理角度来看,大数据企业与传统金融中介形式存在显著差异:传统金融中介沉淀了大量信息,但未能有效传递至社会层面,也未进行深度挖掘;而基于大数据的新型中介有望改进这一缺陷,大数据依托互联网发展,互联网本身具有自由、开放的属性,使数据的获取与传输更加便捷。与传统金融中介相比,这些新型中介在数据分析与运用方面更具动力,数据处理能力也更强。

大数据的出现将改变人们对金融中介的传统认知,丰富金融中介的内涵。基于大数据的金融中介,有可能取代以银行为主的传统金融中介,形成金融再脱媒现象(资金供需双方绕过传统金融中介,通过新型中介直接完成交易)。

(三)金融功能理论

金融功能理论是相对于金融机构理论存在的。金融机构理论的缺陷在于未能认识到当社会环境与技术条件发生快速变化时,金融机构组织会随之调整,跨行业金融产品与模式将不断涌现,而监管往往难以跟上这种变化速度。

知识拓展

美国经济学家、麻省理工学院教授罗伯特·默顿(Robert Merton)与美国经济学家兹维·博迪(Zvi Bodie)于1993年提出的金融功能理论认为:第一,金融功能比金融机构具有更强的稳定性;第二,金融功能比金融机构更加重要。该理论将金融的功能总结为六大类,具体如图1-5所示。

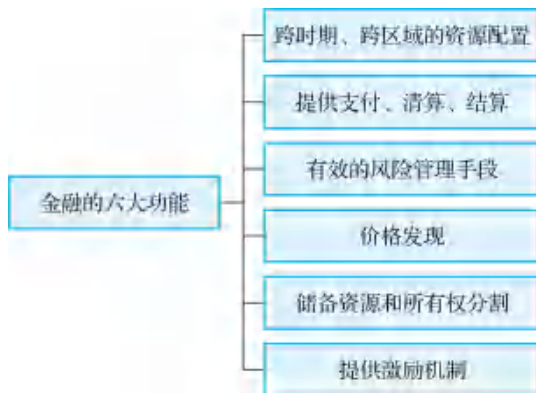


图 1-5 金融的六大功能

任务三 认识金融大数据的发展趋势及应用

任务情景

大数据可以帮助金融机构更准确地评估绿色项目的风险和收益,推动资金流向环保、可持续发展的领域。那么,这些令人期待的变化究竟是如何发生的?金融大数据未来还有哪些发展方向和应用场景?

任务分析

金融大数据的发展趋势及应用关乎金融行业的未来走向。要基于行业现状,结合技术发展趋势,分析跨机构数据融合、技术融合创新、与绿色金融的关系,通过研究银行业等实际案例,把握其应用现状与潜力,助力金融行业科技化升级。

任务实施

一、金融大数据的发展趋势

(一)跨机构金融大数据融合

金融机构大数据融合的核心意义在于解决金融大数据领域的关键问题,包括数据分布不均衡与信息不对称。金融行业作为数据密集型、科技驱动型领域,通过大数据融合,金融机构能够全面掌握客户信息与需求,制定更全面且精准的智能风险控制、智能营销及智能运营策略,最终实现风险管理成本降低与经济效益提升。然而,大数据融合面临多方面的挑战,如数据所有权界定、定价标准缺失、激励机制不完善、隐私保护难题及安全隐患等。因此,建立合理的数据共享规范,解决数据确权、安全防护等问题,对促进数据安全共享与金融行业高质量发展至关重要。

在金融领域,跨机构大数据融合的应用场景丰富多样:在管理提升层面,可通过统计与分析集团客户信息,制定精准的智能营销方案;在智能风控层面,借助大数据融合实现实时监控,维护优质客户关系并管理高风险客户,提升资源配置效率;在智能营销层面,通过用户画像与产品画像的匹配,实现个性化推荐,提高转化效率;在产品画像层面,联合分析金融产品的资产配置与盈利情况,构建完善的产品画像体系,实现产品与客户需求的精准匹配。

(二)大数据与其他新一代信息技术的融合创新

新一代信息技术的快速发展正加速推动全球产业分工深化与经济结构调整,重塑全球

经济竞争格局。大数据与其他新一代信息技术的融合示意图如图 1-6 所示。

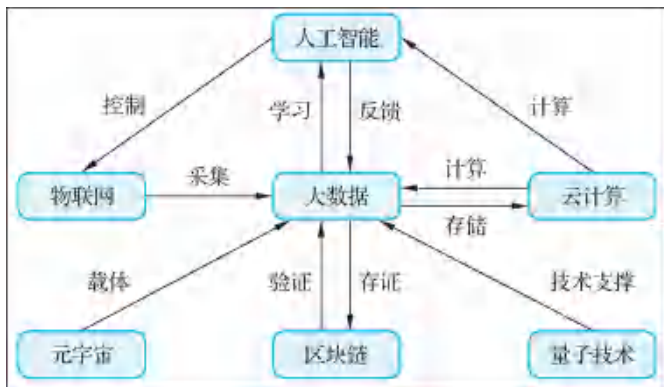


图 1-6 大数据与其他新一代信息技术的融合示意图

1. 大数据与区块链技术的融合

大数据与区块链技术的融合具有多重优势,包括保障数据隐私性、安全性与完整性,降低交易成本,推动信息互联网向价值互联网转型,支持数据预测分析。基于区块链技术的大数据系统技术架构如图 1-7 所示。

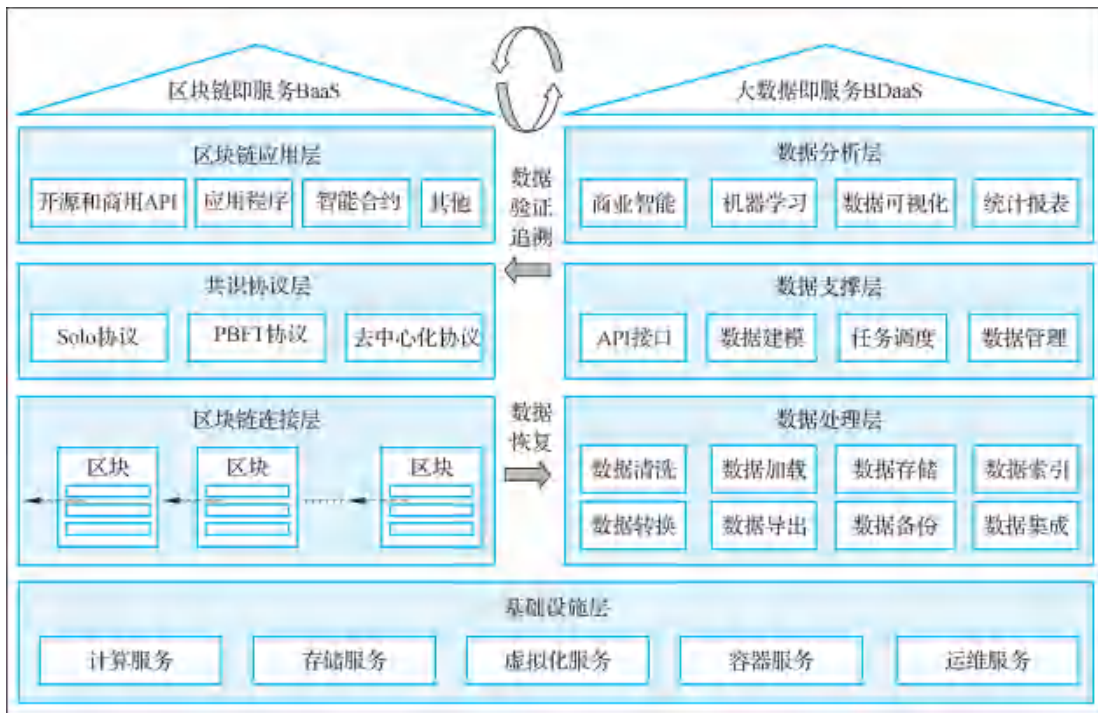


图 1-7 基于区块链技术的大数据系统技术架构

随着数字技术的不断发展,金融行业需要不断创新,开发更加高效、安全的区块链和大数据应用。区块链与大数据技术可以帮助金融机构优化业务流程,提高运营效率。同时,二

者的结合能加强数据安全防护,适用于敏感数据处理,助力金融机构强化数据安全与隐私保护。

2. 大数据与量子技术的融合

大数据应用的安全性需求包括数据合法使用、安全存储、传输安全及隐私保护,这对密钥质量提出了更高的要求。而量子信息技术中的量子通信技术,已初步具备保障大数据信息安全的能力。量子通信不仅能实现原理上无条件安全的密钥分发,在密钥快速更新、大范围分发方面也具有显著优势。

金融大数据的应用依赖庞大的数据库,而量子比特的叠加特性使量子服务器的存储能力远超现有的经典服务器;通过优化传统机器学习算法,量子计算能大幅提升数据处理效率,节约服务器资源。在智能营销、智能风控等对计算速度与精度要求较高的领域,量子技术可支持开发更智能、更灵敏的机器学习系统,在算法能力与速度上形成显著优势,彻底解决银行等金融机构的“计算能力瓶颈”问题,为金融大数据发展提供有力支撑。

3. 大数据与云计算技术的融合

随着云计算的发展,大数据已成为云计算面临的重要考验,而云计算能为大数据解决方案提供三项核心支撑,即外部数据集接入、可扩展性处理能力与大容量存储。技术人员借助“云”的技术优势,可实现对多格式、多模式大数据的统一管理、高效流通与实时分析,挖掘大数据价值,释放其核心效能。

云计算与大数据作为相互独立却紧密关联的技术,在金融行业的结合能发挥更大优势:首先,云计算可提供强大的计算与存储能力,支撑金融机构高效处理海量数据;其次,二者结合可推动金融机构业务创新与服务升级,云计算的高速网络连接与广域数据交换能力可实现不同分支机构间数据与信息的实时共享;最后,结合后的技术体系可提升金融机构风险管理与预测能力,云计算的高可用性与容灾能力,保障业务连续性与数据安全性,降低金融风险。

4. 大数据与人工智能技术的融合

当前大数据研究的核心方向之一是深度学习技术的应用,为提升计算机操作质量提供支撑。人工智能、大数据与深度学习的关系如图 1-8 所示。

金融科技是金融领域发展的核心支撑力量,其生态圈涵盖支付清算、融资、金融基础设施、大数据、交易、投资管理等领域。基于人工智能大数据平台的金融科技应用包括智能投顾、保险科技、数字银行、支付清算等,其中智能投顾与保险科技的发展速度最为迅猛。

人工智能与大数据在金融领域的融合应用既能带来机遇,也伴随着挑战:二者能提升金融机构运营效率与客户满意度,但同时需重视数据隐私保护、安全合规等问题,确保应用的可靠性与合规性。随着技术的持续进步,金融领域的人工智能化进程将进一步加快,为行业带来更多创新与变革。

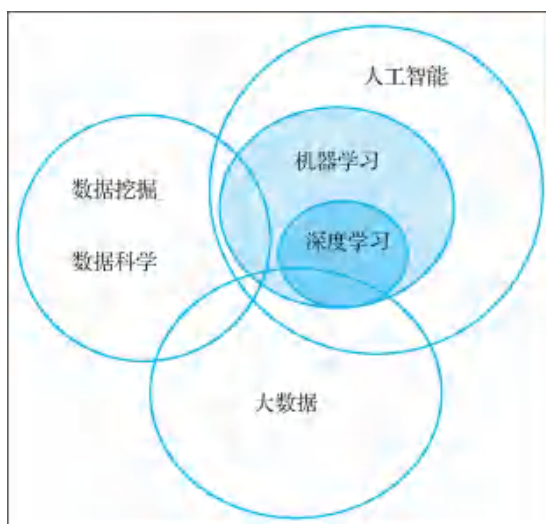


图 1-8 人工智能、大数据与深度学习的关系

5. 大数据与物联网技术的融合

物联网和大数据是相互依存的关系,物联网是大数据的核心数据来源,大数据则承担物联网数据的分析任务。随着数据量的快速增长,通过大数据分析得出的信息含金量与可靠性同步提升;而物联网应用范围的广泛性,使得数据来源更加多元。

物联网正重塑金融大数据的发展路径,助力银行及其他金融机构提升运营效率、降低运营成本;通过实时监测与监管重大风险、预测市场动态、洞察新兴趋势,以物联网增强金融机构在业务领域的竞争力;二者的结合将显著加速自动化金融与移动金融服务的发展,对金融科技等新兴领域产生深远影响。随着技术创新与应用拓展,物联网在金融大数据领域的重要性将日益凸显。

6. 大数据与元宇宙的融合

元宇宙是人类运用数字技术构建的虚拟世界,既能映射现实世界,也能超越现实世界,还能与现实世界交互,是具备新型社会体系的数字生活空间。

在元宇宙的虚拟世界中,数字化程度远高于现实世界,其空间结构、场景、主体等均以数据形式存在,而大数据是构建元宇宙的基础技术之一。元宇宙需实现虚拟世界与现实世界的连接交互,为支撑大规模用户同时在线,需应用增强现实技术,同时依赖大量传感器、智能终端等物联网设备实时收集与处理数据,因此强大的大数据处理能力至关重要,技术层面需实现计算能力升级。构建元宇宙的基础技术如图 1-9 所示。

金融行业是元宇宙技术的重点应用领域之一。随着现实世界业务逐步向元宇宙迁移,元宇宙金融将迎来广阔的应用空间。例如,元宇宙中的虚拟银行可以通过虚拟助理跟踪与记录虚拟金融交易数据、目标客户数据等,进一步推动金融大数据的发展。

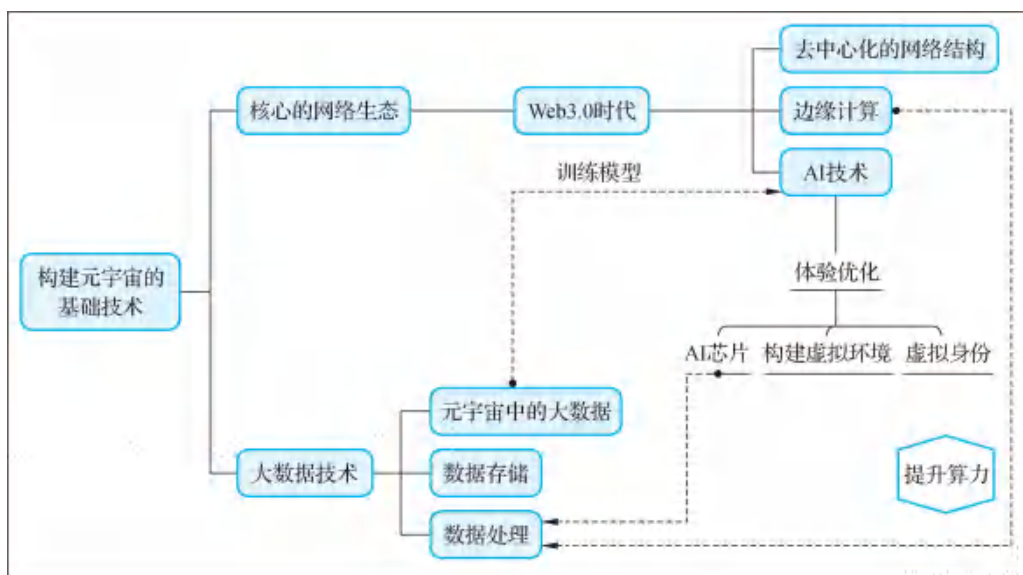


图 1-9 构建元宇宙的基础技术

(三) 大数据与绿色金融

1. 绿色金融的概念和现状

绿色金融是指支持环境改善、应对气候变化和资源高效利用的经济活动的金融服务。自我国 2020 年提出碳达峰、碳中和重大战略决策以来，绿色金融的发展进入加速阶段。绿色金融涵盖绿色信贷、绿色证券、绿色发展基金、绿色保险等领域，其核心特点是关注生态环境平衡，需政府引导与支持。然而，绿色金融领域仍面临挑战，即信息共享不足导致资金匹配困难，数据造假问题增加监管难度，绿色金融产品种类有限，且大数据基础设施不完善。这些问题制约绿色金融发展，需政府、金融机构与技术机构协同发力，推动行业突破瓶颈。

2. 大数据支持绿色金融发展的方式

大数据在绿色金融领域发挥关键支撑作用，推动行业快速发展，具体体现在以下三个方面。

(1) 破除“数据孤岛”，提供科学数据支持。大数据技术能实时抓取各类标准化与非标准化数据，整合为信用信息或绿色行为信息，为金融机构提供全面数据支撑。通过综合分析，金融机构可更准确地判断客户信用状况与绿色行为，降低绿色识别成本，优化绿色信贷决策，提升对符合绿色金融标准项目的投资意愿，推动绿色金融业务拓展；其实时数据处理与分析能力可进一步加速绿色金融产品的推广应用。

(2) 支撑绿色金融信用保障体系建设。绿色金融的可持续发展需要稳定、完善的征信系统，而大数据等新技术为构建大数据征信系统提供技术支持。该系统可将企业环境违法违规信息等环境数据纳入金融信用信息基础数据库，建立企业环境信息共享机制，为金融机构贷款与投资决策提供依据；同时，大数据技术能推动金融信息基础设施升级，促进信息与统

计数据共享,构建分析预警机制,强化对绿色金融资金运用的监督与评估,提升绿色金融可持续性,为投资者提供安全可靠的选择。

(3)提升绿色金融服务水平与效率。大数据、云计算、人工智能等技术的应用,实现金融机构业务系统与绿色信息系统底层数据平台的对接,构建绿色项目融资方的社会关系网络,生成环境效益评估报告,大幅缩短金融机构对绿色企业或项目的识别、认证审查时间,提升服务效率与客户体验;同时,技术应用简化服务流程、加速审批速度,为企业提供便捷融资渠道。此外,基于大数据的金融科技平台具有较强的网络外部性,能降低金融机构服务单个客户的成本,使传统金融机构未覆盖的长尾客户、中小微企业也能享受绿色金融服务,并推动绿色金融普惠化发展。

二、金融大数据的应用

随着信息技术与移动互联网的发展,金融业务与服务日趋多样化,金融市场规模持续扩大,金融行业数据收集能力逐步提升,积累了大量时间连续、动态变化的数据。与其他行业相比,大数据对金融业的价值潜力更为突出。麦肯锡全球研究院的研究显示,金融业在大数据价值潜力指数中排名第一。当前,传统金融机构数字化转型已成为行业趋势,云计算、大数据、区块链、人工智能等技术逐步与金融业务融合,金融大数据应用覆盖多个场景,有力推动行业科技化升级,提升了风控能力与资源配置效率,催生创新金融业务模式。目前,银行业、证券业、保险业等金融机构已在大数据应用领域开展深入探索。金融大数据的应用场景示例如图 1-10 所示。

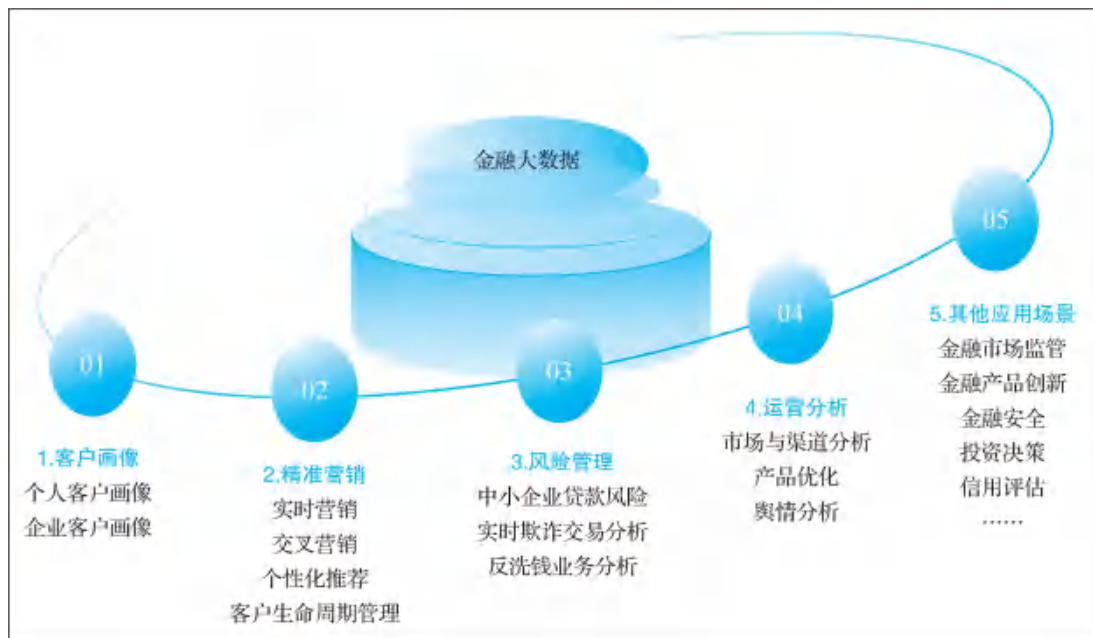


图 1-10 金融大数据的应用场景示例

（一）大数据在银行业的应用现状

近年来,银行业的客户数据、交易数据、管理数据等均呈爆炸式增长。中国人民银行发布的公开数据显示,2024年年末在用发卡数量进一步提升至99.13亿张,全年共发生银行卡交易5642.29亿笔,其中消费业务3571.78亿笔;2025年第三季度末,银行卡在用发卡量突破100亿张至101.49亿张,当季银行卡交易1652.78亿笔,其中消费业务1035.94亿笔。若叠加开户信息、银行网点与在线交易及金融系统自身运营数据后,国内银行每年产生的数据规模可达数十PB。数据的海量增长为银行业带来机遇与挑战,其服务与管理模式也逐步发生变革。

目前,银行业大数据商业应用的核心是挖掘内部交易数据,同时引入外部机构数据(如社交网络数据、商务经营信息、征信数据、收支消费信息、社会关联信息等),开展描述性数据分析与预测性数据建模;常见应用场景集中在精准营销、风险管理等领域,如通过客户画像实现个性化产品推荐,依托风险模型实时监测信贷风险。

（二）大数据在证券行业的应用现状

证券业务对数据的实时性、准确性与安全性要求较高,在国内证券行业政策逐步开放的背景下,证券业对数据应用的竞争日趋激烈。证券数据涵盖金融行情数据、历史交易数据、上市公司资料、宏微观经济数据、交易者行为规律数据等,证券公司的投资决策、业务营销、客户关系管理等均需以这些数据为支撑。证券行业常见的大数据应用场景包括智能投顾(基于客户风险偏好与市场数据提供投资建议)、市场预测(通过数据分析预判市场走势)等。

（三）大数据在保险行业的应用现状

大数据与保险业具有天然适配性,保险经营的核心基础是“大数法则”,如保险生命表需以十万人作为样本进行测算。无论是财产险的概率性事件,还是寿险的生命周期分析,均依赖大量数据分析得出规律。长期以来,保险业通过上门服务、柜面交互、信函、电话、短信、微信等多种渠道,积累了大量客户交互数据。2013年7月,经国务院批准,中国保险保障基金有限责任公司出资20亿元成立中国保险信息技术管理有限责任公司,搭建起中国保险行业首家数据共享平台。

目前,多家保险公司已布局大数据应用,核心领域包括产品创新(如基于用户行为数据开发场景化保险产品)、风险控制(如通过数据分析评估投保标的风险)、运营优化(如简化理赔流程)等。近年来兴起的互联网保险,也成为保险业收集数据的新平台。据统计,国内大型保险公司每年新增数据量可达PB级别。在全球保险大数据应用市场中,核心领域包括客户行为分析、产品定价、联网数据分析、市场渠道分析、风险建模、预测分析、商业决策、欺诈侦测等。

数据处理流程解析与实践

【实训背景】

在数字经济蓬勃发展的当下,大数据技术深度融入金融领域,推动金融行业发生深刻变革。金融机构依托海量数据开展风险评估、产品创新、客户服务优化等工作,对从业人员的数据处理能力与业务理解水平提出了更高的要求。为使学生深入了解大数据金融场景,掌握金融大数据处理的核心流程与方法,专门开展本次实训。

【实训目的】

(1)通过查阅资料和案例制作,系统梳理大数据金融场景下的典型数据处理流程,加深对金融大数据从采集、清洗、存储到分析、应用全生命周期的理解。

(2)以小组协作形式制作 PPT 案例并展示,锻炼团队协作能力、资料整合能力、PPT 制作与演示能力,以及面对问题时的沟通表达能力和应变能力。

【实训步骤】

步骤一:资料查阅与小组分工。

(1)以小组为单位,利用图书馆数据库、学术期刊网站(如中国知网、万方数据知识服务平台等)、行业报告平台(如艾瑞咨询、易观分析等)及金融机构官方发布的资料,收集大数据金融相关的政策法规、行业动态、技术应用案例等资料。

(2)小组成员进行明确分工,如指定专人负责数据采集与清洗流程的资料收集,专人负责数据分析与应用流程的资料梳理等,确保资料收集全面且高效。

步骤二:案例制作与 PPT 设计。

(1)结合收集到的资料,选取具有代表性的大数据金融场景,如智能信贷审批、精准营销、风险预警等,制作详细的案例。案例内容需涵盖场景描述、涉及的数据类型、数据处理的具体流程、应用效果及典型案例分析等。

(2)将案例内容以 PPT 的形式呈现,注重 PPT 的美观度与逻辑性。合理运用图表、图片等可视化元素,对数据处理流程进行直观展示;统一字体、配色和排版风格,确保整体视觉效果协调。

步骤三:小组展示与问答环节。

(1)各小组依次进行 PPT 展示,展示时间控制在 10~15 分钟,小组成员需清晰、有条理地阐述案例内容,重点突出大数据金融场景下的金融大数据处理流程。

(2)展示结束后,进入问答环节,由教师或其他小组成员针对展示内容进行提问,展示小组需准确、流畅地回答问题,展示对相关知识的掌握程度和应变能力。

知识小结

本项目构建了金融大数据知识体系,学习路径如下:首先,了解大数据的基础概念(定义、特征、处理流程等)及大数据时代的思维变革;其次,掌握金融大数据的概念、特征与理论基础(信息经济学、金融中介理论、金融功能理论);最后,探讨金融大数据的发展趋势(跨机构金融大数据融合、与新一代信息技术融合、与绿色金融结合等),以及在银行业、证券业、保险业等金融领域的应用现状。

通过本项目的学习,学生能系统地掌握金融大数据的基础知识,把握其在金融领域的价值和应用场景,为解决实际问题 and 洞察行业发展提供助力。

思考练习

一、填空题

1. 数据的最小单位是_____。
2. 大数据的数据结构类型包括_____、_____和_____。
3. 大数据分析的类型包括_____、_____、_____和_____。
4. 大数据的_____和_____将彻底重塑传统的金融机构决策机制,为金融机构的经营管理提供了全面、及时的决策支持信息。

二、简答题

1. 请简述结构化数据的概念。
2. 大数据与区块链技术融合有哪些优势?