

# 本科毕业论文开题报告

国际工商管理学院 院(系)

日期 2025.11.9

姓名	刘蕊	性别	女	学号	0221147010
专业	信息管理与信息系统		指导老师姓名	宋艳	
论文题目	基于 LLM 智能体的医药市场分析系统设计与实现				

论 文 大 纲  及 撰 写 计 划	<h2>一、选题背景</h2> <p>当前，以大数据与人工智能为核心的数字化浪潮正深刻重塑全球医疗健康产业格局，医药行业的市场竞争越来越激烈。专利保护期限的逼近、仿制药市场的冲击以及日益严格的合规监管环境，共同促使企业告别以往粗放式的增长模式，转而寻求更为精细化和高效的运营策略。与此同时，“以患者为中心”的理念已从行业倡导转变为市场共识，这意味着企业的市场营销活动不再是简单的药品推广，而是需要基于对医生处方决策背后复杂因素的深刻理解，提供更具价值的学术支持和个性化服务。</p>
	<p>在这一转型过程中，数据成为了企业最宝贵的战略资产。医药企业在长期运营中积累了包括医生的处方行为、药品流通记录、销售代表拜访活动、学术会议参与信息在内的海量多维数据。这些数据理论上能够为市场趋势预测、目标医生群体画像和营销资源分配等关键决策提供坚实的依据。然而，这些数据往往散落在不同的业务系统或报表中，形成“数据孤岛”，其内在的价值难以被有效发掘。如何将这些原始、分散的数据转化为能够直接赋能一线市场与销售团队、可指导行动的商业活动策略，已成为提升企业竞争力的瓶颈。</p> <p>为了应对这一挑战，传统的 BI (Business Intelligence, 商业智能) 系统作为第一代解决方案应运而生，它通过标准化的数据报表和可视化看板，在一定程度上提升了数据的透明度。然而，这类系统在应对现代商业分析的深度需求时，其局限性也日益凸显，BI 系统善于描述性分析，却无法深入探究原因以及未来策略。分析师和决策者仍需投入大量精力进行手动的数据分析、交叉验证和报告撰写，整个决策链冗长且高度依赖个人经验，且效率较为低下。</p> <p>在此背景下，以大语言模型 (Large Language Model) 为基础的智能体 (Agent) 技术为破解上述困境提供了新的技术路径，即人机交互与信息处理范式的又一次革新。特别是当智能体具备调用数据库、分析模型等外部工具并自主执行任务的能力时，便能以结构化的报告形式生成蕴含实际业务内容的结论。将这一前沿 AI (Artificial Intelligence, 人工智能) 技术与经典的商业数据分析方法相结合，构建一个能够深度分析并自动化生成策略的智能化信息系统，已成为信息管理与信息系统领域一个具有研究与应用价值的方向。</p> <p>然而，将前沿的 AI 能力转化为企业内部可稳定运行、可便捷使用、可产生实际业务价值的信息系统，是一个待解决的软件工程与系统开发方面的挑战。单纯的算法研究或 AI 模型探索，无法替代系统工程中数据集成、人机交互、业务流程融合以及系统可靠性等现实问题。</p> <p>本研究设计的系统，其目标用户是医药企业的市场部、BI 分析师以及一线销售与区域管理团队——他们通常不具备专业的数据分析背景，目标是帮助他们从海量数据中快速获得可执行的医生画像和策略建议。因此，本研究的立足点在于，不仅要探索 K-Means 聚类等经典数据挖掘算法与前沿智能体技术的理论融合，更要设计并实现一个融合二者的“医药市场智能分析与策略生成系统”。本研究与某医药企业（后文简称“XX 企业”）合作，以该企业内部的部分脱敏多维业务数据为研究样本，旨在通过一个完整的信息系统工程实践，为医药行业所面临的挑战提供一个兼具技术创新性与可行性的解决方案。</p>

本研究针对当前医药市场分析中海量数据到有效商业策略转化效率低下的问题，将设计并实现一个基于 XX 企业的业务数据集的 Web 应用系统——“医药市场智能分析与策略生成平台”，为以下两大关键问题提供系统性的解决方案：

数据层问题：如何科学地识别并定义目标医生群体。

应用层问题：如何将复杂的分析模型与业务流程相结合，高效智能地将分析结果转化为一线运营团队可理解、可执行的市场策略。

围绕上述总体目标，本研究设定了四个层层递进的具体目的：

#### （一）构建医生群体科学画像与分群的数据挖掘模型

综合运用数据清洗、特征工程及 K-Means 聚类算法，对原始数据进行加工与分析，构建一个能够客观反映医生处方行为和市场潜力的医生分群模型，超越传统基于经验的主观划分，为后续 Agent 的自动化分析提供结构化的、量化的事实依据。

#### （二）设计混合智能的系统与分析架构

融合传统机器学习算法与新兴的多智能体系统，设计一种混合智能分析架构，打造一个既能发挥机器学习在处理结构化数据、发现统计规律方面的严谨性，又能利用智能体在自然语言理解、逻辑推理与任务编排方面灵活性的新平台。

#### （三）实现自然语言驱动的报告生成功能

基于 Dify 平台的工作流，设计一个模拟真实分析团队的多智能体协作系统，用户通过自然语言指令来驱动系统自动完成数据提取、模型调用、可视化及报告撰写全过程，赋能营销策略报告的全流程自动化生成。

#### （四）开发可交互的信息系统原型

遵循软件工程的原型法与敏捷开发思想，采用前后端分离的技术栈，将上述的数据模型、系统架构与核心功能进行实现，开发出一个功能多样、界面友好、可交互的信息系统。

### 三、研究价值及创新点

本系统旨在赋能一线业务人员，通过 LLM 智能体技术，将复杂的数据分析能力封装为人可用的 AI 分析助理。其价值与创新点体现在以下几个方面：

#### （一）学术价值与创新点

##### 1. 架构创新——探索统计学习与生成式 AI 的混合智能（Hybrid-AI）架构

传统的知识发现（Knowledge Discovery in Databases）流程强调从数据中通过算法提取模式，但模式的解读和知识的转化仍高度依赖分析师。Wang 等人(2024)在关于 LLM 智能体的综述中提到描述的架构，智能体通过调用外部数据库作为其“符号记忆”（Symbolic Memory）来进行决策（Wang et al., 2024）。本研究在理论上探索一种混合智能架构，该架构将统计学模型（K-Means）的严谨性与 LLM 智能体的认知推理能力相结合。具体而言，K-Means 模型被封装为 AI 可调用的工具，为 LLM 提供进行逻辑推理所必需的依据。通过此设计来解决 LLM 在专业分析上容易产生幻觉的问题。

##### 2. 范式创新——推动决策支持系统向认知型决策支持的范式演进

传统的信息系统理论将 DSS 的发展划分为提供静态报告的 MIS 阶段与支持数据钻取的 BI 阶段。传统 BI 模式如下：用户必须学习工具、定义指标、拖拽报表等。本系统则通过引入具备推理、规划和工具调用能力的多智能体系统，探索 DSS 向更高阶的认知智能阶段演进：用户仅需通过简单的自然语言下达一个模糊的任务，例如“帮我分析一下 A 药品的潜力医生”，

多智能体工作流（包含数据分析师、BI、报告撰写员等）即可自主规划、执行并交付一份完整的策略报告。Poszler 与 Lange(2024)在其系统性综述中指出，21 世纪以 AI 技术为核心的智能决策支持系统 (IDSS) 的出现，在“能力、复杂性和自主性”上已经超越了传统 DSS (Poszler & Lange, 2024)。在此新范式下，系统不应再是被动响应数据查询的工具，而是能够主动参与知识发现全过程的分析助手，它能够理解高层级的分析任务、自主编排分析流程并生成蕴含真实业务分析结果的知识产物。

### 3. 工程创新——基于真实业务场景地领域专用智能体实践

通用大语言模型虽然知识广博，但在专业领域的应用中常因缺乏深度和精确性而受限。Kumar (2024)在其综述中指出，通用 LLM 在应用于领域特定问题时面临重大挑战，因为它们缺乏“深入、实时和准确的领域知识” (Kumar, 2024)。本研究通过为智能体设计一套面向医药市场分析的专用工具集（包括医生分群查询、销售数据对比等），并利用 Dify 工作流进行编排，探索如何将通用 AI 能力锚定在特定业务领域，使其在保证专业性的前提下发挥效能。

## （二）行业与企业价值

### 1. 企业价值——普及数据分析能力

本系统致力于将专业的数据挖掘能力赋能给企业内最广大的非技术用户，降低高级数据分析的应用门槛，同时提升决策的便捷性、实用性和时效性。在传统的模式下，数据挖掘项目高度依赖具备编程和统计学背景的专业数据科学人员，一线销售想获得一份医生画像报告，需要经历“提需求——BI 排期——数据分析师建模——交付报表”的漫长流程。而且，一位数据分析师完成一份深度报告，往往需要经历数据提取、清洗、建模、可视化及报告撰写流程。本系统通过将这一系列复杂任务封装于自动化的工作流中，压缩为一次自然语言对话，一线销售人员只需通过发起指令表达自己的业务需求，即可获得相关的分析报告，最终解决实际问题，提升整个组织的运营智慧。

### 2. 行业价值——提供可复制的“AI+营销”解决方案

本研究基于 XX 企业的脱敏业务数据集，验证该混合智能架构在医药行业的有效性。本系统主要包括 K-Means 模型、Agent 工作流以及前后端架构，是一个可复制、可扩展、具备商业应用价值的项目。它不仅可以提供医生画像与决策建议，未来还能扩展到销售团队拜访医生策略的优化、市场部门推广资源的聚焦、合规审计、患者行为分析等更多场景，为整个行业应对数字化转型、提升精细化运营能力，提供一个具有高实用价值的工程范例。

## 四、拟解决的问题和采用的方法

### （一）主要研究内容

本研究的主要内容将遵循信息系统开发的标准流程，分为四个层次，从底层的原始数据处理到上层的应用界面实现，具体如下：

#### 1. 数据层：数据源的获取与处理

本研究将使用合作的 XX 企业提供的客户关系管理数据、药品处方数据及销售代表拜访数据为主要数据源，其可以在一定程度上真实地反映企业面临实际业务挑战。

由于这些数据在结构、格式和编码上存在差异，本阶段的首要任务是对异构数据进行系统性的预处理，包括但不限于：数据清洗、数据格式统一、缺失值补充，以及基于企业内部医生的主数据编码进行跨数据集的特征关联，最终构建一个可用于后续分析的、统一规范的数据集。

#### 2. 模型层：医生画像 K-Means 聚类模型构建

基于预处理后的数据集，本研究将借鉴客户关系管理 (CRM) 的经典模型，并结合医药

行业特性，主要从两个维度构建特征体系：一是代表其产出能力的处方特征，包括年度处方总金额、年度处方总次数等；二是代表其市场价值与投入潜力的投入特征，包括年度从药企获取的支付总额、支付次数、以及其本身的职称级别等。

随后，对特征数据进行标准化处理，采用肘部法（Elbow Method）和轮廓系数（Silhouette Coefficient）确定最优 K 值，并构建 K-Means 聚类模型。模型构建完成后，将对形成的各个医生群体进行画像解读与业务命名，并将该模型持久化，通过 Python 函数或 API 端点封装为可供智能体调用的、标准化的分析工具。

### 3. 智能层：基于 Dify 的多智能体协作设计

Dify (dify.ai) 是一个行业领先的开源 LLMops（大型语言模型运维）平台，它提供了一套可视化的界面，可以快速构建、部署和运营具备生产力水平的 AI 应用，可以进行可视化的 Prompt 编排、对 RAG（Retrieval-Augmented Generation，检索增强生成）的内置支持、以及应用监控与运营。

本研究将充分利用这一特性，将构建的 K-Means 模型、数据库查询等功能封装为 Dify 可调用的 API 工具。在此基础上，设计一个模拟真实业务分析团队的多智能体协作工作流。在此过程中，Dify 充当本系统的智能中台，将复杂的 AI 分析逻辑与前端应用和后端业务服务实现解耦（Decoupling）。该工作流主要包含三个职责明确的核心智能体角色：

#### （1）数据分析师 Agent

深度理解用户的高级自然语言指令，通过调用预设的 Python 代码执行器、数据库查询工具或前述封装的 K-Means 模型 API，从数据层查询所需数据或调用模型的分析结果。

#### （2）可视化专家 Agent

接收数据分析师 Agent 输出的结构化数据，并调用 ECharts 等图表库的 API，将其自动生成为符合分析逻辑的动态、直观的可视化图表。

#### （3）报告撰写员 Agent

整合前两个 Agent 产出的所有文本分析和可视化图表，结合在 Dify 知识库中预先存储的医药领域知识库，编排并生成一份结构清晰、逻辑严谨且包含策略建议的分析报告。

### 4. 应用层：系统原型开发与集成

在系统架构上，本项目将采用主流的前后端分离模式，遵循原型法开发思想，将上述数据、模型与智能层进行工程化的集成，最终交付一个可交互的 Web 信息系统原型。

后端将采用 Python 的 FastAPI 框架，封装所有核心业务逻辑，包括数据处理、模型调用、触发并管理 Dify 智能体工作流等，并以 RESTful API 的形式为前端提供标准化服务。

前端将采用 Vue.js 框架，利用其组件化的开发模式，高效构建用户交互界面（UI），包括数据看板、医生画像分析结果展示以及智能报告的自然语言操作界面等。

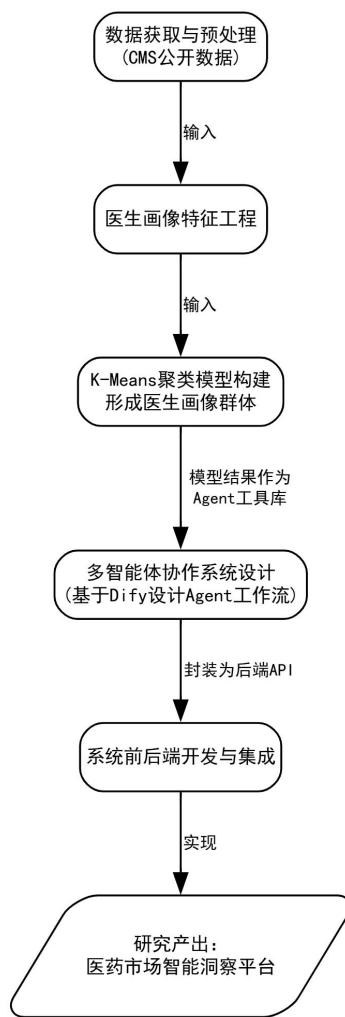


图 1 研究内容逻辑流程图

## （二）研究方法

### 1. 文献研究法

为奠定本研究的理论基础、指导系统架构设计与技术选型，在项目初期及研究过程中，系统性查阅国内外在决策支持系统（DSS）、客户关系管理（CRM）、数据挖掘（KDD）、大语言模型与智能体（LLM & Agent），以及软件工程与信息系统开发方法论等相关领域的学术论文、行业报告和技术文档。

### 2. 实证分析法

本研究的产出之一是医生画像模型，将采用定量的实证分析方法，基于数据集，采用 K-Means 聚类算法等方法，对数据进行客观的客户分群与画像构建。

### 3. 原型开发法

遵循现代软件工程思想，采用原型开发法作为核心开发策略。鉴于本系统所集成的 AI 智能体分析能力具有探索性，初期需求难以一次性完全定义，因此将通过快速构建核心功能原型，进行模拟交互测试，并根据结果进行多轮次的迭代开发，最终交付一个功能闭环、可验证的 Web 应用系统。

## （三）技术路线

本研究将采用一套主流、前后端分离的技术栈来实现系统原型，具体的技术路线如下：

	<p><b>1. 数据层</b></p> <p>技术选型：MySQL 数据库</p> <p>实现内容：存储由 XX 企业提供、经清洗和预处理后的医生处方、拜访、支付等数据，为上层分析提供统一的数据来源。</p> <p><b>2. 模型层</b></p> <p>技术选型：Pandas, NumPy, Scikit-learn</p> <p>实现内容：基于 Python，使用 Pandas 进行特征工程处理，使用 Scikit-learn 库实现 K-Means 聚类算法，构建医生画像模型，并将其持久化。</p> <p><b>3. 智能层</b></p> <p>技术选型：Dify 平台，Google Gemini LLM</p> <p>实现内容：利用 Dify 平台进行 AI 工作流的编排与智能体管理，将负责解析用户指令、定义和调用工具，并接入 Google Gemini 大语言模型进行推理和内容生成。</p>
表 1 多智能体协作角色、工具与工作流设计	

## 5. 部署层

技术选型: Docker, Docker Compose

实现内容: 为确保开发与最终运行环境的一致性, 将使用 Docker 对上述所有服务进行器化封装与一键部署。

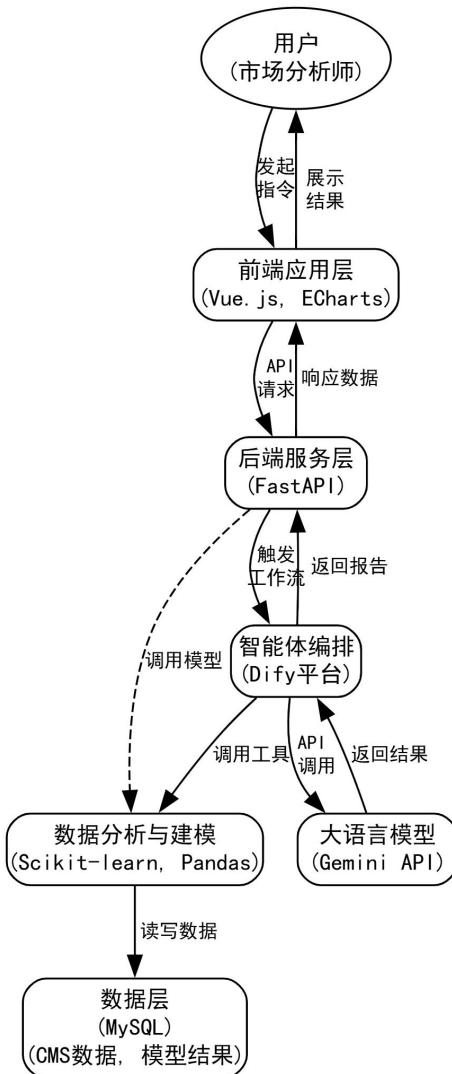


图 2 系统架构图

## 五、进度安排

第七学期第 8 周前: 确定选题、撰写开题报告

第七学期第 9~10 周: 提交开题报告

第七学期第 11~14 周: 开题答辩

第七学期末: 中期检查

第八学期初: 初稿

第八学期第 5~6 周: 定稿

## 六、文献综述

### (一) 基于数据挖掘的医生画像与分群研究

客户分群是现代商业智能与客户关系管理的核心技术之一，其目的是通过数据分析将异质的客户群体划分为具有相似特征的同质群组，以实现精准营销和个性化服务。

在这一领域，K-Means 聚类算法因其高效、易于理解和实施的特性，得到了极为广泛的应用。Anitha & Patil (2022) 的研究将经典的 RFM (Recency, Frequency, Monetary) 模型与 K-Means 算法相结合，对零售行业用户的购买行为进行了有效分割，并通过轮廓系数 (Silhouette Coefficient) 验证了聚类的有效性。国内学者的研究也印证了这一路径的有效性，蔡少霖 (2022) 详细阐述了利用 K-Means 算法对消费者数据进行数据清洗、分析、聚类，并最终构建“消费群体画像”的完整流程，其研究明确指出，这种技术可以帮助企业准确地定义目标客户、核心产品和营销模式。何帆 (2025) 的研究将目光投向了在线医疗平台，通过对医生的个人特征、信息行为、患者评价乃至医患交流特征等四个维度进行层次聚类，成功构建了精细化的医生画像。该研究证明了对医生群体进行数据驱动的画像分析是可行的，并能为优化平台推荐策略、发掘非核心医生潜力提供重要指导。无论是针对普通消费者还是特定的医生群体，利用 K-Means 等聚类算法构建画像，均是业界和学界公认的、成熟的数据分析方法论。

## (二) 大语言模型驱动的智能体技术现状

上述传统的数据挖掘方法虽然有效，但其产出的聚类结果、模型和报表通常是静态的，高度依赖专业分析师的手动解读和转化，决策链条长且效率低下。大语言模型特别是智能体技术的崛起，为解决这一“最后一公里”的自动化难题提供了全新的思路。

Wang et al. (2024) 在其对 LLM 驱动的自主智能体的全面综述中，为本研究的设计提供了关键的理论框架，他们提出了一个包含 Profile、Memory、Planning 和 Action 四大核心模块的统一智能体架构。在此架构中，智能体通过 Action 模块与外部环境交互，其核心能力之一就是工具调用。

本研究拟采用 Dify 平台构建的多智能体系统，正是对这一理论框架的工程实践。其中，数据分析师 Agent 调用数据库和 K-Means 模型的能力，即是 Wang et al. (2024) 所定义的“工具调用”的具体实现。这一理念也与 Hong et al. (2024) 提出的 Data Interpreter 项目相印证，该项目证实了 LLM Agent 有能力自动生成和验证代码，端到端地解决复杂的数据科学问题。

因此，利用 LLM 构建一个具备规划、记忆和工具调用能力的智能体系统，使其自动化地编排和执行数据分析任务，是当前人工智能领域的前沿发展方向。

## (三) 医药市场的数据分析与 AI 应用

医药行业的数字化转型正在加速，Verma et al. (2024) 的研究指出，AI 在重塑制药企业数字化外联 (Digital Outreach) 方面发挥着开拓性作用。他们强调，AI 能够帮助企业进行精准定位与细分 (Precision targeting and segmentation) 以及增强客户互动，具有实践意义。Cuomo et al. (2021) 的研究展示了利用行业支付透明度公开数据，对医生群体的行业互动情况进行深入分析的可行性。他们通过数据分析，揭示了不同科室、不同地区医生从行业获取报酬的差异性。这表明，基于真实世界数据的深度分析是可行的，并能为市场策略提供重要指导。

## (四) 信息系统开发方法论研究

本研究不仅是数据分析项目，更是一个完整的信息系统设计与实现项目。因此，选择合适的开发方法论至关重要。传统的瀑布模型 (Waterfall Model) 虽然结构清晰，但其流程僵化，难以适应现代信息系统，特别是 AI 应用系统需求快速变化和探索性的特点。

现代软件工程和信息系统开发领域更推崇迭代和增量的方法，原型法是软件工程中应对需求不确定性的经典策略 (Pressman & Maxim, 2020)。该方法主张快速构建一个可运行的系统原型，通过早期交付和用户反馈，不断迭代修正需求和设计，尤其适用于本研究这类需要

将复杂 AI 能力与用户交互相结合的创新型系统。

同时，敏捷开发作为一种更宏观的理念，强调以人为本、快速响应变化和持续交付价值（Conboy, 2009）。汪应洛与郭上（2000）在国内的研究也指出，信息系统的开发必须与其所处的环境和组织需求相适应，采用灵活的方法论是系统成功的关键。本研究采用 Dify 平台和前后端分离的架构，天然适合采用敏捷的迭代开发模式，通过短周期的开发与测试，确保最终系统功能与研究目标高度一致。

### （五）文献评述

综上所述，本研究的理论与实践基础是清晰且多学科交叉的。

首先，在数据分析层面，利用 K-Means 对医生进行画像分析是成熟且有效的数据挖掘方法（Anitha & Patil, 2022；蔡少霖, 2022；何帆, 2025）。然而，其核心局限在于产出的静态结果高度依赖人工解读。

其次，在技术前沿层面，医药行业正在积极拥抱 AI 进行精准营销（Verma et al., 2024）。大语言模型的崛起为此提供了新路径。正如 Poszler & Lange (2024) 的系统性综述所指出的，技术正推动传统决策支持系统向智能决策支持系统演进，即系统从被动的数据提供者转变为自动的分析助理。

实现这一转变的核心技术是自主智能体（Autonomous Agents）。Wang et al. (2024) 的最新综述为这类智能体提供了统一的架构。其关键在于智能体能够通过工具调用执行复杂任务（Hong et al., 2024），这使得智能体调用 K-Means 这类传统数据挖掘模型在理论上成为可能。

最后，在系统工程层面，要将上述 AI 能力落地，必须遵循严谨的开发方法论。原型法和敏捷思想被证明是开发此类需求不确定、需快速迭代的 AI 信息系统的关键策略（Pressman & Maxim, 2020；Conboy, 2009）。

目前，虽然 K-Means 挖掘、Agent 架构、医药数据、开发方法等领域均有独立研究，但仍缺乏一个能将 K-Means 聚类分析的严谨性、多智能体系统的自动化能力与信息系统开发的工程规范性相结合，并应用于医药市场分析领域的端到端解决方案。因此，本研究的创新之处正是在于填补这一集成的空白：在理论上，构建一个融合经典数据挖掘与前沿智能体架构的混合智能分析架构；在工程上，将该架构封装为 Dify 智能体可调用的工具，并遵循原型法（Pressman & Maxim, 2020），设计并实现一个从用户自然语言指令到自动化生成深度分析报告的、可用的信息系统。

## 七、论文大纲

摘要

ABSTRACT

目录

第 1 章 绪论

    1.1 研究背景

    1.2 研究意义

        1.2.1 理论意义

        1.2.2 实践意义

    1.3 研究内容

    1.4 研究方法与技术路线

    1.5 论文的创新点

    1.6 论文的组织架构

	<p>第 2 章 相关研究与理论基础</p> <ul style="list-style-type: none"><li>2.1 客户关系管理与数据挖掘研究<ul style="list-style-type: none"><li>2.1.1 医生分群理论与 RFM 模型</li><li>2.1.2 聚类算法在商业智能中的应用</li></ul></li><li>2.2 大语言模型智能体应用研究现状</li><li>2.3 决策支持系统与智能体研究现状</li><li>2.4 信息系统开发方法论<ul style="list-style-type: none"><li>2.4.1 敏捷开发思想</li><li>2.4.2 软件工程原型法</li></ul></li></ul>
	<p>第 3 章 系统关键技术与数据源</p> <ul style="list-style-type: none"><li>3.1 K-Means 聚类算法原理</li><li>3.2 Dify 智能体工作流平台</li><li>3.3 前后端技术栈<ul style="list-style-type: none"><li>3.3.1 Python FastAPI 后端框架</li><li>3.3.2 Vue.js 前端框架</li></ul></li><li>3.4 XX 企业业务数据集介绍</li></ul>
	<p>第 4 章 系统总体设计</p> <ul style="list-style-type: none"><li>4.1 系统需求分析<ul style="list-style-type: none"><li>4.1.1 功能性需求分析</li><li>4.1.2 非功能性需求分析</li></ul></li><li>4.2 系统总体架构设计<ul style="list-style-type: none"><li>4.2.1 技术架构</li><li>4.2.2 混合智能分析架构</li></ul></li><li>4.3 系统功能模块设计<ul style="list-style-type: none"><li>4.3.1 数据处理与画像建模模块</li><li>4.3.2 智能报告生成模块</li><li>4.3.3 可视化看板与系统管理模块</li></ul></li><li>4.4 数据库设计<ul style="list-style-type: none"><li>4.4.1 概念结构设计</li><li>4.4.2 逻辑结构设计</li></ul></li></ul>
	<p>第 5 章 医生画像分析模块设计与实现</p> <ul style="list-style-type: none"><li>5.1 数据处理<ul style="list-style-type: none"><li>5.1.1 数据清洗与整合</li><li>5.1.2 异常值处理</li></ul></li><li>5.2 特征工程<ul style="list-style-type: none"><li>5.2.1 关键特征提取</li><li>5.2.2 特征标准化</li></ul></li><li>5.3 K-Means 聚类模型构建</li><li>5.4 画像分析与可视化界面实现</li><li>5.5 画像模块后端 API 接口实现</li></ul>
	<p>第 6 章 智能报告生成模块设计与实现</p> <ul style="list-style-type: none"><li>6.1 Dify 平台集成与 Agent 工作流编排</li><li>6.2 核心工具的开发与集成<ul style="list-style-type: none"><li>6.2.1 数据库查询工具</li><li>6.2.2 聚类结果调用工具</li></ul></li></ul>

	<p>6.3 知识库构建</p> <p>6.3 智能报告交互界面实现</p> <p>6.4 智能报告后端 API 接口实现</p> <p>第 7 章 系统测试</p> <p>7.1 测试环境与用例设计</p> <p>7.2 模块功能测试</p> <p>7.3 智能报告生成效果评估</p> <p>第 8 章 结论与展望</p> <p>8.1 研究工作总结</p> <p>8.2 研究不足与展望</p> <p>参考文献</p> <p>致谢</p> <p>附录</p>
主要参考书目	<p>[1] 蔡少霖,吴丽文,郑东荣.基于 K-means 聚类分析的农产品消费群体画像构建及精准营销策略研究[J].农村经济与科技,2022,33(22):251-254+262.</p> <p>[2] 何帆,赵梦石,刘思源,等.基于医患互动的在线医疗平台医生的画像构建与优化策略[J].现代情报,2025,45(10):125-137.</p> <p>[3] 陈永伟.智能体经济的崛起: AI 智能体对商业世界的重塑[J].财经问题研究,2025,(07):15-31.</p> <p>[4] Anitha P, Patil M M. RFM model for customer purchase behavior using K-Means algorithm[J]. Journal of King Saud University - Computer and Information Sciences, 2022, 34(2): 1785-1792.</p> <p>[5] Hong S, Lin Y, Liu B, et al. Data interpreter: An LLM agent for data science[J]. arXiv preprint, 2024, 2402.18679v4.</p> <p>[6] Li X, Wang S, Zeng S, et al. A survey on LLM-based multi-agent systems: workflow, infrastructure, and challenges[J]. Vicinagearth, 2024, 1(9).</p> <p>[7] Cuomo R E, Cai M, Shah N, et al. Physicians payment in the United States between 2014 and 2018: An analysis of the CMS Open Payments database[J]. PLOS ONE, 2021, 16(6): e0252656.</p> <p>[8] Verma S, Tiwari R K, Singh L. Integrating technology and trust: Trailblazing role of AI in reframing pharmaceutical digital outreach[J]. Intelligent Pharmacy, 2024, 2(3): 435–440.</p> <p>[9] Pressman R S, Maxim B R. Software Engineering: A Practitioner's Approach (9th ed.)[M]. New York: McGraw-Hill, 2020.</p> <p>[10] Conboy K. Agility from first principles: Reconstructing the concept of agility in information systems development[J]. Information Systems Research, 2009, 20(3): 329-354.</p> <p>[11] Kumar P. Large language models (LLMs): survey, technical frameworks, and future challenges[J]. Artificial Intelligence Review, 2024, 57: 260.</p> <p>[12] Poszler F, Lange B. The impact of intelligent decision-support systems on humans' ethical decision-making: A systematic literature review and an integrated framework[J]. Technological Forecasting &amp; Social Change, 2024, 204: 123403.</p> <p>[13] Wang L, Ma C, Feng X, et al. A survey on large language model based autonomous agents[J]. Frontiers of Computer Science, 2024, 18(6): 186345.</p>
审核老	同意提交

师 意 见	审核人（签名）：  2025年11月11日
-------------	--