








数眸 - 智能数据分析平台 - 产品说明文档

产品概述

数眸是一款基于Streamlit构建的现代化数据分析应用，专为不同技能水平的数据分析师和研究人员设计。该平台集成了数据上传、清洗、分析、可视化和机器学习等完整功能，并配备了AI智能助手，为用户提供专业的数据分析指导。

核心特色

-  **数眸品牌**：让数据洞察如眸般清澈明亮
-  **多模式设计**：新手模式、普通模式、专业模式，满足不同用户需求
-  **AI智能助手**：集成大语言模型，提供智能分析和建议
-  **丰富可视化**：支持20+种图表类型，包括3D图表和交互式图表
-  **专业统计分析**：提供完整的描述性和推断性统计功能
-  **机器学习集成**：支持分类、回归、聚类等多种ML算法
-  **现代化UI**：采用Material Design 3设计系统，界面美观易用

技术架构

核心技术栈

| 技术组件 | 版本 | 用途 |

|-----|-----|-----|

| **Streamlit** | ≥1.28.0 | Web应用框架 |

| **Pandas** | ≥2.0.0 | 数据处理 |

| **NumPy** | ≥1.24.0 | 数值计算 |

| **Plotly** | ≥5.15.0 | 交互式可视化 |

| **Scikit-learn** | ≥1.3.0 | 机器学习 |

| **SciPy** | ≥1.11.0 | 科学计算 |

| **LangChain** | ≥0.1.0 | AI助手集成 |

架构设计

智能数据分析平台

└─ 🖥️ 用户界面层 (Streamlit)

└─ 🧠 业务逻辑层 (Python模块)

└─ ☯️ AI智能层 (LangChain + 大语言模型)

└─ 📊 数据处理层 (Pandas + NumPy)

└─ 📈 可视化层 (Plotly + Matplotlib)

└─ 🔍 分析算法层 (Scikit-learn + SciPy)

模块化设计

src/

└─ config/ # 配置管理

| └─ settings.py # 应用配置和样式

└─ modules/ # 页面模块

```
|   |— pages.py          # 页面渲染

|   |— beginner_mode.py  # 新手模式

|   |— intermediate_mode.py # 普通模式

|— utils/                # 工具模块

    |— data_processing.py    # 数据处理

    |— visualization_helpers.py # 可视化工具

    |— ml_helpers.py         # 机器学习

    |— ai_assistant.py       # AI助手
```

功能模块详解

1. 模式选择系统

新手模式 ()

- **目标用户：**数据分析初学者
- **核心功能：**
 - 引导式学习流程
 - 示例数据集演示
 - 基础概念解释
 - AI智能指导
 - 简化操作界面

普通模式 (🚀)

- **目标用户：** 有一定经验的数据分析师
- **核心功能：**
 - 完整的数据处理功能
 - 多种可视化选项
 - 基础统计分析
 - AI智能建议
 - 中等复杂度操作

专业模式 (⚡)

- **目标用户：** 专业数据分析师
- **核心功能：**
 - 高级数据处理工具
 - 专业可视化图表
 - 完整统计分析
 - 机器学习算法
 - AI深度分析
 - 完整工具集

2. 📁 数据上传模块

支持格式

- **CSV文件** (.csv)
- **Excel文件** (.xlsx, .xls)
- **JSON文件** (.json)
- **Parquet文件** (.parquet)

核心功能

- **智能数据检测**：自动识别数据类型和格式
- **数据质量评估**：实时计算数据质量评分
- **基础统计分析**：自动生成描述性统计
- **缺失值分析**：可视化缺失值分布
- **AI智能建议**：基于数据特征提供分析建议

3. 🧹 数据清洗模块

清洗功能

- **缺失值处理**：
 - 删除行/列
 - 均值/中位数/众数填充
 - 前向/后向填充
 - 插值填充
- **重复值处理**：
 - 删除重复行
 - 保留首次/末次出现
- **异常值处理**：
 - IQR方法
 - Z-score方法
 - 百分位法
- **数据类型转换**：
 - 自动类型推断

- 手动类型转换
- 格式标准化

整洁数据指南

- 遵循"每行一个观测值，每列一个变量"原则
- 提供数据格式最佳实践建议
- 可视化数据质量改进效果

4. 自动数据分析模块

分析功能

- **描述性统计**：均值、中位数、标准差、分位数等
- **相关性分析**：Pearson、Spearman相关系数
- **数据分布分析**：直方图、密度图、Q-Q图
- **异常值检测**：多种检测算法
- **数据质量评分**：综合评估数据质量

AI智能解释

- 自动分析结果解释
- 数据特征识别
- 潜在问题预警
- 进一步分析建议

5. 高级可视化模块

图表类型

图表类型	适用场景	特点
柱状图	分类数据比较	直观对比

| **折线图** | 趋势分析 | 时间序列 |

| **散点图** | 相关性分析 | 变量关系 |

| **饼图** | 占比分析 | 比例展示 |

| **直方图** | 分布分析 | 数据分布 |

| **箱线图** | 异常值检测 | 统计特征 |

| **热力图** | 相关性矩阵 | 多维关系 |

| **3D散点图** | 三维数据 | 立体展示 |

| **小提琴图** | 分布比较 | 密度分布 |

| **雷达图** | 多维评估 | 综合评价 |

交互功能

- **缩放和平移**: 支持图表交互操作
- **悬停信息**: 显示详细数据点信息
- **图例交互**: 可选择性显示数据系列
- **导出功能**: 支持PNG、SVG格式导出

6. 统计分析模块

描述性统计

- **集中趋势**: 均值、中位数、众数
- **离散程度**: 方差、标准差、IQR、变异系数
- **分布特征**: 偏度、峰度
- **分位数**: Q1、Q3、百分位数

假设检验

- **正态性检验**: Shapiro-Wilk检验
- **t检验**: 独立样本t检验、配对t检验

- **方差分析：**单因素ANOVA
- **相关性检验：**Pearson、Spearman检验
- **卡方检验：**独立性检验

结果解释

- 自动生成检验报告
- 显著性水平说明
- 效应量计算
- 统计功效分析

7. 🤖 机器学习模块

任务类型

分类任务

- **算法支持：**
 - 逻辑回归
 - 随机森林
 - 支持向量机
 - 决策树
 - K近邻
- **评估指标：**
 - 准确率 (Accuracy)
 - 精确率 (Precision)
 - 召回率 (Recall)
 - F1分数
 - 混淆矩阵

回归任务

- **算法支持：**

- 线性回归
- 随机森林回归
- 支持向量回归
- 多项式回归

- **评估指标：**

- R^2 分数
- 均方误差 (MSE)
- 平均绝对误差 (MAE)
- 均方根误差 (RMSE)

聚类任务

- **算法支持：**

- K-means聚类
- 层次聚类
- DBSCAN
- 高斯混合模型

- **评估指标：**

- 轮廓系数
- 肘部法则
- 轮廓图

高级功能

- **特征工程**：特征选择、特征缩放、特征变换
- **交叉验证**：K折交叉验证
- **超参数调优**：网格搜索、随机搜索
- **学习曲线**：模型性能分析
- **特征重要性**：变量重要性排序

8. 报告生成模块

报告内容

- **数据概览**：数据集基本信息
- **数据质量评估**：质量评分和问题分析
- **描述性统计**：数值型变量统计摘要
- **可视化图表**：关键图表展示
- **分析结论**：主要发现和建议

报告格式

- **HTML报告**：交互式网页报告
- **可下载**：支持本地保存
- **自定义样式**：专业美观的排版
- **响应式设计**：适配不同设备

9. AI智能助手

核心功能

- **智能分析建议**：基于数据特征提供分析策略
- **问题解答**：回答数据分析相关问题
- **方法推荐**：推荐合适的统计方法和算法
- **结果解释**：解释分析结果和图表含义
- **学习指导**：提供学习建议和资源

技术实现

- **大语言模型**：集成通义千问等先进LLM
- **上下文理解**：结合数据特征和用户问题
- **专业领域知识**：针对数据分析场景优化
- **多轮对话**：支持连续交互和深入讨论

用户界面设计

Material Design 3 设计系统

色彩系统

- **主色调**：深蓝紫渐变 (#6366F1)
- **次要色调**：青绿渐变 (#10B981)
- **第三色调**：橙金渐变 (#F59E0B)
- **语义色彩**：成功、警告、错误、信息

组件设计

- **卡片组件**：现代化阴影和圆角设计
- **按钮组件**：渐变背景和悬停效果
- **输入组件**：统一的表单控件样式
- **导航组件**：清晰的层级结构

响应式布局

- **桌面端**：完整功能展示
- **平板端**：适配中等屏幕
- **移动端**：简化界面布局

部署和使用

环境要求

- **Python版本**: 3.8+
- **内存要求**: 4GB+ RAM
- **存储空间**: 2GB+ 可用空间
- **网络连接**: 用于AI助手功能

安装步骤

1. 克隆项目

```
git clone [项目地址]

cd DataAnalysis_Core
```

2. 安装依赖

```
pip install -r requirements.txt
```

3. 配置环境变量

```
# Windows

set DASHSCOPE_API_KEY=your_api_key_here


# Linux/Mac

export DASHSCOPE_API_KEY=your_api_key_here
```

4. 启动应用

```
streamlit run app.py
```

使用流程

1. **选择模式**: 根据技能水平选择合适的分析模式
2. **上传数据**: 支持多种格式的数据文件上传
3. **数据清洗**: 处理缺失值、异常值和重复值
4. **探索分析**: 进行描述性统计和可视化分析
5. **深度分析**: 执行统计检验和机器学习建模
6. **生成报告**: 创建专业的分析报告

配置和定制

配置文件

- **settings.py**: 应用配置和样式设置
- **环境变量**: API密钥和外部服务配置
- **自定义CSS**: 界面样式定制

扩展开发

- **模块化架构**: 易于添加新功能
- **插件系统**: 支持第三方扩展
- **API接口**: 提供程序化访问

性能优化

缓存机制

- **数据缓存**: 使用Streamlit缓存减少重复计算
- **模型缓存**: 机器学习模型结果缓存

- **图表缓存：**可视化图表生成缓存

内存管理

- **数据分块处理：**大文件分块读取
- **及时释放：**自动清理临时数据
- **内存监控：**实时显示内存使用情况

安全和隐私

数据安全


- **本地处理：**数据在本地处理，不上传服务器
- **临时存储：**会话结束后自动清理数据
- **加密传输：**API调用使用HTTPS加密

隐私保护


- **无数据收集：**不收集用户个人信息
- **匿名使用：**支持完全匿名使用
- **数据控制：**用户完全控制自己的数据



版本历史

v3.0.0 (当前版本)




-  重构为模块化架构
-  新增多模式设计
-  集成AI智能助手
-  升级到Material Design 3
-  优化性能和用户体验

v2.0.0

-  增加机器学习功能





-  改进可视化系统
-  添加统计分析模块

v1.0.0





-  基础数据分析功能
-  简单可视化图表
-  数据上传和清洗

未来规划





短期计划 (3-6个月)

-  增加更多机器学习算法
-  支持更多数据格式
-  优化AI助手性能
-  添加更多可视化图表

中期计划 (6-12个月)

-  开发移动端应用
-  集成数据库连接
-  支持云端部署
-  添加协作功能

长期计划 (1-2年)

-  构建插件生态系统
-  开发企业版本
-  支持实时数据分析
-  集成更多AI模型

技术支持

联系方式

- 邮箱: tjn.chaos@qq.com
- 技术负责人: LarryTang
- 项目地址: [GitHub仓库链接]

文档资源

- 用户手册: 详细的使用指南
- API文档: 开发者接口文档
- 视频教程: 操作演示视频
- FAQ: 常见问题解答

社区支持

- GitHub Issues: 问题反馈和功能建议
- 讨论区: 用户交流和经验分享
- 贡献指南: 参与项目开发

数眸致力于为数据分析师和研究人员提供专业、易用、智能的数据分析工具，让数据科学变得更加简单和高效。

版本: v3.0.0 / 更新时间: 2025年