

飞算全自动数据建模平台

专业版

白皮书



飞算云创
AI_INNOVATION



AI Modeler Pro

目录

1	背景介绍	2
2	平台概览	3
3	产品特点	4
3.1	简化数据准备	4
3.2	降低大数据处理复杂性	4
3.3	可扩展、可重用的模块库	4
3.4	易用的自动建模	5
3.5	自动模型生产发布	5
3.6	多方位安全保障	5
3.7	多维度可靠性支撑	5
4	产品优势	6
4.1	数据管理能力	6
4.2	开放建模高级配置	7
4.3	解释性建模日志	8
4.4	标准评分卡	9
4.5	工程化能力	10
4.5.1	模型生产全流程覆盖	10
4.5.2	Auto ML	10
4.6	模型生命周期管理	11
4.7	团队协作	11
5	合作案例	12
5.1	合作客户	12
5.2	合作背景	12
5.3	客户需求	13
5.4	解决方案	13
5.5	合作优势	13
5.6	合作成果	14
6	联系方式	15

1 背景介绍

作为国内领先的金融科技应用平台，飞算云创自成立以来，坚持科技驱动，不断探索人工智能、云计算、区块链等前沿技术在金融领域的场景应用，赋能金融机构数字化创新和转型。飞算云创自主研发出了“AI.Modeler Pro”（飞算全自动数据建模平台专业版），它搭载了最新的 AutoML 技术，能更好地解决目前 AI 落地应用中的难点和痛点，为银行和其他金融机构提供全方位的数据建模支持。

AutoML 技术主要是将模型训练中的各种环节自动化，只需给它必要的数据就能实现无人工参与建模。相比同类产品，飞算云创 AutoML 技术凝聚了许多营销、风控专家的建模经验，沉淀了在实际业务中验证下来较优的建模流程和方法论。例如变量编码、最优分箱、可视化调整、逐步回归等。

2 平台概览

飞算全自动数据建模平台专业版 AI.Modeler Pro 是面向数据科学团队的一站式数据分析平台，它是集数据准备、特征工程、算法实现、模型开发、模型发布、模型生产化管理于一体的机器学习平台，能够帮助企业快速构建数据分析应用。

AI.Modeler Pro 支持数据科学家、数据工程师、AI 开发者快速构建 AI 应用，实现端到端的 AI 开发服务。平台包含智能数据融合管理、自动化机器学习、智能特征和算法组件，为用户提供从数据接入到 AI 应用开发的全生命周期工具支持，赋能企业级 AI 应用开发与管理。

3 产品特点

基于 AutoML 的设计和开发理念，飞算 AI Modeler Pro 自动化建模平台针对数据模型开发到应用的全流程均提供了简洁易用的自动化、流程化的前端操作界面以及必要的后端支持能力。

3.1 简化数据准备

支持多种数据连接器，轻松获取各类数据源的数据，包括本地数据、关系型数据库的数据、Hadoop 大数据平台数据（HDFS、Hive 等）。

3.2 降低大数据处理复杂性

平台支持分布式任务，可自动完成 Hadoop、Spark 的环境调用和任务分发执行，对使用者屏蔽了大数据技术组件的复杂性，使数据分析人员轻松获得大数据处理能力。

3.3 可扩展、可重用的模块库

基于 Docker 容器的自定义模块封装发布，一次编程，多处使用，可积累的模块库成为公司重要的智力资产，提升新模型的开发效率。

3.4 易用的自动建模

通过界面配置基础数据和业务目标，实现一键式建模，自动完成算法建模、模型评估、模型选择，降低建模门槛。

3.5 自动模型生产发布

建模完成后平台支持自动选择最优模型，实现一键自动模型发布功能，面向生产系统提供标准 REST API 调用服务。

3.6 多方位安全保障

平台支持快速企业私有环境部署，确保生产环境安全；面向用户、角色、工作空间的多层级访问控制，保障数据安全；用户的访问、编辑、操作等任何行为都可追溯，实现责任认定。

3.7 多维度可靠性支撑

平台服务采用负载均衡和高可用设计；完善的容灾备份机制，提供数据的备份与恢复方案，以及针对异常情况的自动化处理。

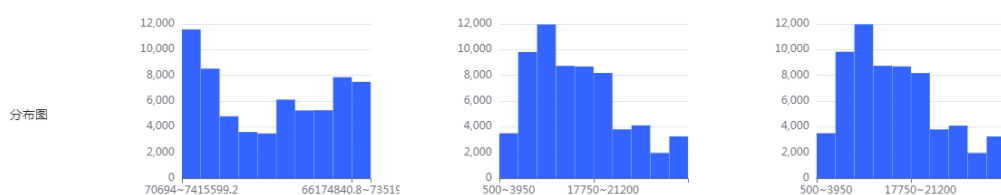
4 产品优势

飞算 AI.Modeler Pro 自动化建模平台深挖数据建模全流程的痛点难点，特别是数据处理、特征工程、算法调参以及信用风险评估领域特有的标准评分卡开发等系列功能做出了优化和整合，使得产品整体差异化更加明显。

4.1 数据管理能力

导入数据后，可快速预览和查看统计指标，自动完成数据分布检测、统计指标计算、自动对异常或缺失值检测、删除低信息密度变量等来实现必要的探索性数据分析，并支持对数据进行增删查改。

平台支持交互式的数据字典上传以增强解释性，可以自定义字段级别的特殊值为后续建模服务。



列名	member_id	loan_amnt	funded_amnt
注释	用户ID	申请贷款金额	承诺给贷款人的总金额
类型	int	int	int
平均值	34830098.9618	14780.4489	14767.8936
最小值	70694.0	500.0	500.0
最大值	7.3519746E7	35000.0	35000.0

4.2 开放建模高级配置

针对建模过程中重要的特征工程和算法调参部分，开放参数配置项，支持建模人员自定义控制。包含以下功能点：

- 自动化数据清洗
- 自定义编码转换
- 交叉特征衍生
- 特征聚合统计
- 自研特征选择算法/手动筛选特征宽表
- **导入导出特征处理配置文件**
- 自研超参搜索框架
- 自动算法选择
- 过拟合约束
- 支持超参数配置
- **导入导出超参配置文件**

每个关键步骤均支持“智能配置”和“手动配置”，可兼顾不同水平建模人员的使用体验。

特征工程 算法选择 保存配置

特征工程 智能配置 手动配置 不配置

整体处理

推荐配置 导入配置 导出配置

特征预处理 全部离散特征 全部连续特征

特征筛选 关闭

处理详情

特征名称: 清空

名称	特征类型	Y	标准预处理	异常值处理	缺失值处理	衍生方法	处理
loan_amnt	连续		最小最大值标准化(MinMax)	不处理		关联log.abs.sqrt.exp 交叉 聚合	⌵
funded_amnt	连续		最小最大值标准化(MinMax)	不处理		关联log.abs.sqrt.exp 交叉 聚合	⌵
funded_amnt_inv	连续		最小最大值标准化(MinMax)	不处理		关联log.abs.sqrt.exp 交叉 聚合	⌵
term	离散		默认		默认	交叉 聚合	⌵
int_rate	连续		最小最大值标准化(MinMax)	不处理		关联log.abs.sqrt.exp 交叉 聚合	⌵
grade	离散		默认		默认	交叉 聚合	⌵

特征工程 [算法选择](#)

推荐配置 导入配置 导出配置

传统机器学习

参数搜索方式: 自研方式(推荐) 随机搜索 网格搜索 贝叶斯优化 不搜索

设置训练轮次: 训练轮次预估:

LR

LightGBM

XGBoost

RandomForest

* 深度: - 输入值是1-100的正整数

* 提升树策略:

* 行采样比例: - 输入值是0-1的数

* 列采样比例: - 输入值是0-1的数

* L1正则: - 输入值是0-1的数

* L2正则: - 输入值是0-1的数

* 学习率: - 输入值是0-1的数

* 提升迭代次数: - 输入值是1-1000的正整数

* 单棵树最大叶子数: - 输入值是1-100的正整数

4.3 解释性建模日志

业务解释性日志贯穿建模全流程，使模型开发过程充分透明。具体日志内容包含数据清洗、编码转化、归一化、缺失异常处理逻辑、特征工程、特征选择、模型迭代的细节，全部以业务语言呈现。保证建模全流程可解释，支持对具体步骤追溯调整。

- 2021-08-16 16:18:18 ○ 完成多重共线性检测，总共筛选出0个特征。
- 2021-08-16 16:18:18 ○ 正在进行多重共线性检测 (13/13)。
- 2021-08-16 16:18:20 ○ 最终特征筛选完成。
- 2021-08-16 16:26:13 ○ 正在提交模型训练任务
- 2021-08-16 16:26:26 ○ 开始进行逻辑回归模型训练。
- 2021-08-16 16:26:56 ○ 开始对特征 (funded_amnt_inv,term,batch_enrolled,verification_status,purpose,dti,inq_last_6mths,open_acc,revol_bal,revol_util,initial_list_status,total_rec_int,last_week_pay) 做 woe编码。
- 2021-08-16 16:28:13 ○ 完成对特征 (funded_amnt_inv,term,batch_enrolled,verification_status,purpose,dti,inq_last_6mths,open_acc,revol_bal,revol_util,initial_list_status,total_rec_int,last_week_pay) 的 woe编码。
- 2021-08-16 16:28:26 ○ 正在计算特征重要性。

4.4 标准评分卡

飞算云创 AI.Modeler Pro 集成专属标准评分卡开发模块，凝聚了飞算风控专家的建模经验，是在实际业务中验证下来较优的建模方式。

评分卡开发的全流程线性完成，复杂的指标计算和分箱编码全部可由平台自动完

成，并且支持在平台计算结果上的可视化调整，所见即所得，使建模人员专注于数据业务理解。



4.5 工程化能力

4.5.1 模型生产全流程覆盖

平台提供的功能模块包括数据准备、特征工程、模型训练、模型评估、模型对比、模型发布，涵盖模型生产的标准流程，实现图形化、流程化工程建模。

4.5.2 Auto ML

通过数据集选择和简单配置学习目标就可以实现自动化建模，支持自动算法选择、

自动超参数优化、自动模型评估、自动模型选择。

4.6 模型生命周期管理

通过模型仓库模块实现对模型的版本管理，支持基于版本的模型生产上线、生产下线以及模型删除等操作。

4.7 团队协作

支持项目成员管理，团队不同角色(专家级数据科学家、数据分析师、编码人员、业务人员) 成员之间协同开发，提升开发效率。

5 合作案例

5.1 合作客户

某大型国有信托公司

5.2 合作背景

随着银保监会一系列监管政策的落地，持牌金融机构被要求必须有自己的风控能力，监管强调“金融机构应强化风险控制主体责任，独立开展互联网贷款风险管理，并自主完成对贷款风险评估和风险控制具有重要影响的风控环节，严禁将贷前、贷中、贷后管理的关键环节外包”。

同时，在原有业务基础上，信托公司都在积极拓展消费金融和小微信贷业务。但是针对互联网客群的反欺诈和风险运营经验以及相关样本的积累相对较少，既有风控系统无法满足互联网业务的高实时性和快速迭代能力，导致金融机构在搭建消费金融业务时面临很多挑战。在面对大量进件时，原有风控系统无法准确识别和控制风险。

具体而言，在发展新业务模块的过程中，例如对消费金融方面的尝试，往往会缺乏相关的经验、人才以及数据积累，从零开始搭建非常困难。

因此，客户虽有进行自主风控的意愿，但却往往因技术，资源，数据等积累不足，

需要外部资源的配合与支持。

5.3 客户需求

掌握自主风控能力是持牌金融机构发展的趋势，客户期望能够在这方面做更多地尝试，提升自身的整体竞争地位，成为行业的发展标杆，因此整体合作态度也更为积极开放。

5.4 解决方案

飞算团队以咨询方式进入后辅导客户将数据清理和规整完毕后，再进行科学的主题分类以及适当的特征衍生，有针对性地搭建数据集市；之后再在集市的基础上利用 AI Modeler Pro 建模平台搭建相应的决策模型，用以实施风险管理。

合作过程中，飞算科技能够通过制定反欺诈策略集的制定，在业务前期将高危客群进行过滤，提高客户的资产质量，进一步完善整体风控体系。同时提出数据治理相关建议，为完善数据体系提供有力支持。根据客户基本信息、交易行为、购买历史、足迹信息、渠道偏好、风险偏好、产品偏好等构建完备的客户画像标签体系，快速完成模型开发全流程。

5.5 合作优势

在 2017 年开始，飞算已经与银行、消金等大型客户进行金融科技整体方案的输

出合作，具有成熟和丰富的行业工作经验。

在合作过程中，飞算还能够根据客户现状将客群进行科学合理分类，针对不同客群做不同的规则，一方面为之后的业务进展打好基础，提升工作效率；另一方面能够有效识别坏客，降低坏账率，为客户决策提供有力支持。

5.6 合作成果

飞算云创 AI.Modeler Pro 平台完成欺诈检测模型或信用风险评分卡的开发上线，提高模型分类区分度，减少客户贷款违约损失。

6 联系方式

公司官网：<https://www.feisuanyc.com>

电话联系方式：18948313772

邮箱联系方式：contact@feisuanyc.com



飞算云创
AI_INNOVATION



AI.Modeler