

# Kaleido 特征工程平台

打开数据密码的钥匙, 开启最具创造力的数据科学

## 产品介绍 Product Introduction

Kaleido 是一款基于分布式存储架构和分布式计算框架的特征工程平台, 可以最大限度地从原始数据中提取特征以供算法和模型使用。数据科学家的工作中, 有 80% 的时间都在获取、清洗和特征处理, Kaleido 通过分布式存储架构和分布式计算框架, 快速完成这 80% 的工作量, 将数据转换为能更好的表示业务逻辑的特征, 从而提高机器学习的性能。

## 核心功能 Core Function

- **通用板块**  
主要实现常规的数据处理分析, 包括统计分析、数据清洗、特征变换、特征衍生、特征提取等功能;
- **文本分析板块**  
主要实现文本数据的处理分析, 包括文本分词、文本量化、主题抽取、情感分析等功能;
- **网络分析板块**  
主要实现网络数据的处理分析, 包括连通图、静态特征分析、社团发现等功能;
- **时序分析板块**  
主要实现时序数据的处理分析, 包括移动平均、指数平滑、GARCH、ARIMA 等功能;
- **量化分析板块**  
主要实现基金、股票等数据的转换处理, 包括包括模式识别、交易量指标、波动率指标等;
- **金融板块**  
主要实现金融领域常用数据处理, 包括 WOE 变换、评分卡模型等功能;
- **勘探板块**  
主要应用于石油勘探领域, 包括特定地震数据格式的读取及信息提取等功能。

## 市场痛点 Market Pain Points

### 人才供需失衡

市场上同时具备数据科学知识和编程经验的人才难以培养, 成本高昂, 供需严重失衡。

### 运行效率低下

通过单机与服务器方式运行数据处理脚本, 执行质量与工程师经验息息相关, 难以持续保证高效运行。

### 上线流程漫长

IT 部门与模型部门沟通困难, 重构代码容易引入错误。

## 产品价值 Product Value

### 分布式计算框架

分布式计算框架能够增强计算可扩展性, 轻松应对海量数据处理;

### 高效构建有效特征

拖拉拽免代码方式实现特征构建, 直线缩短时间、降低成本;

### 特征可追溯

可对建模历史的特征工程版本进行追溯, 流程可回溯, 降低人才流失风险;

### 模型快速上线

生产流程即上线流程, 无需代码重构, 大大提高模型时效。



### 天云融创数据科技(北京)有限公司

北京市朝阳区东四环中路78号大成国际A1座9A08-9A09

上海市黄浦区河南南路16号中汇大厦二层

商务合作: websales@beagledata.com 媒体合作: media@beagledata.com

联系电话: 010-59625658