

易明智能建模

一键式智能数据建模



一键式智能数据建模

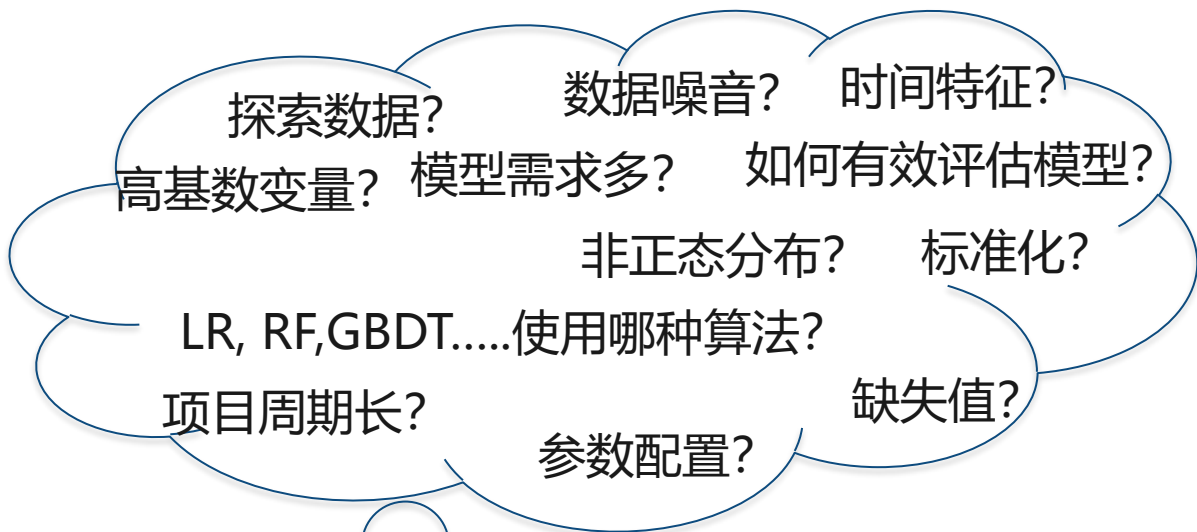


手工建模

VS



易明智能建模



探索数据?

数据噪音? 时间特征?

高基数变量? 模型需求多? 如何有效评估模型?

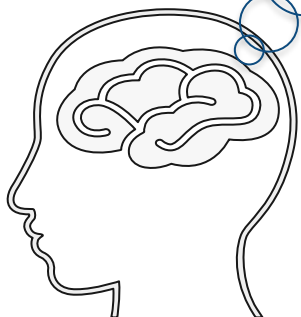
非正态分布? 标准化?

LR, RF, GBDT.....使用哪种算法?

项目周期长?

缺失值?

参数配置?



人工?



人工智能

智能!



建模流程完全自动进行

一键式建模, 又快又好!

无需数据科学家即可建模

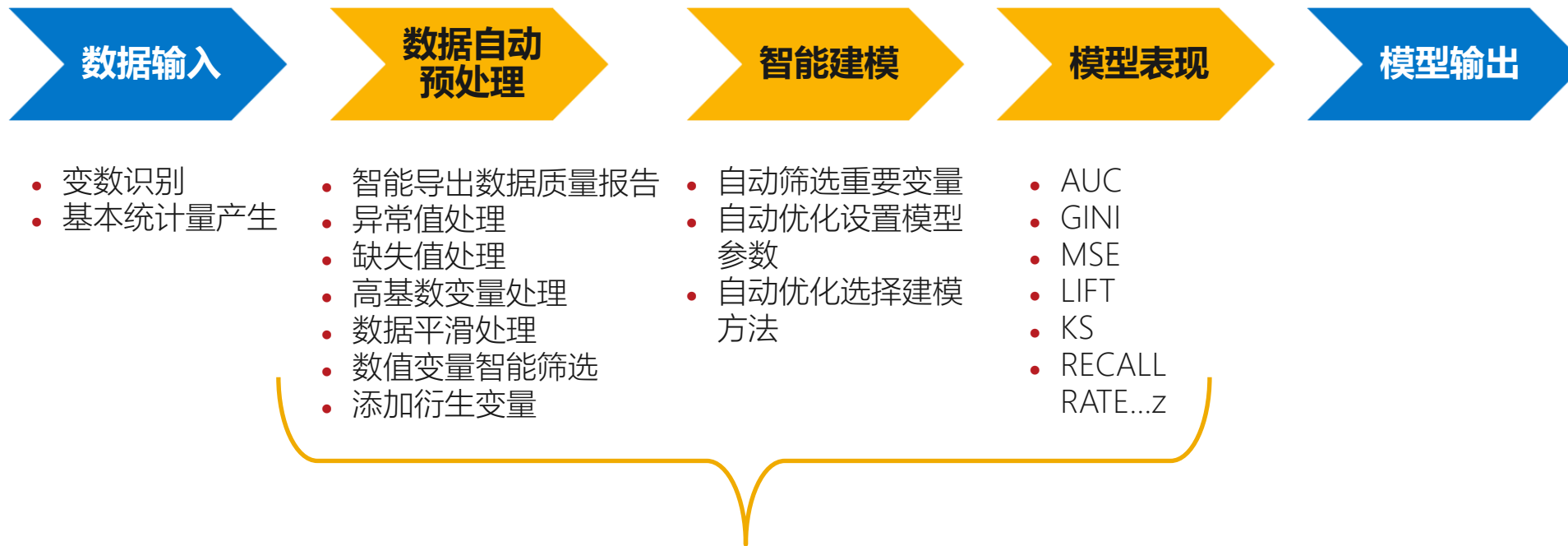
传统建模过程



需要数据科学家完成的众多人工任务



智能建模过程



原本要**数据科学家**完成的人工任务，智能建模工具**一键完成**，并确保模型质量和稳定性



易明智能建模架构



易明智能建模过程

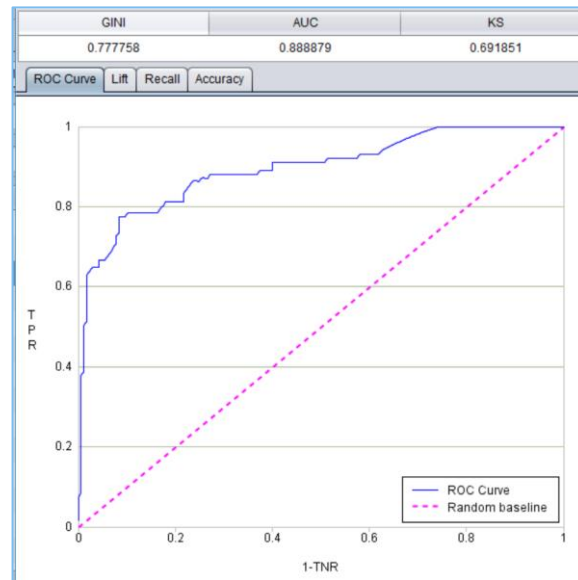
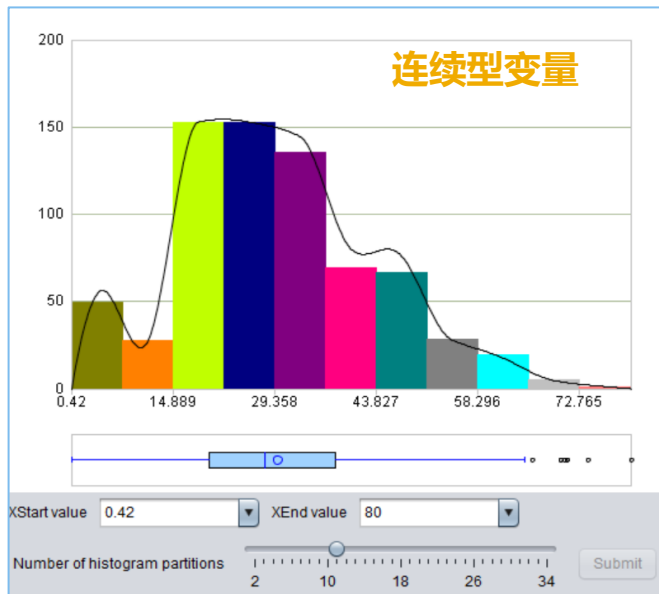
Target variable: **Survived** Set Variable filter

NO.	Variable name	Type	Date format	Select
1	PassengerId	ID		<input checked="" type="checkbox"/>
2	Survived	Binary variable		<input checked="" type="checkbox"/>
3	Pclass	Categorical variable		<input checked="" type="checkbox"/>
4	Name	ID		<input type="checkbox"/>
5	Sex	Binary variable		<input checked="" type="checkbox"/>
6	Age	Numerical variable		<input checked="" type="checkbox"/>
7	SibSp	Categorical variable		<input checked="" type="checkbox"/>
8	Parch	Categorical variable		<input checked="" type="checkbox"/>
9	Ticket	Categorical variable		<input checked="" type="checkbox"/>
10	Fare	Numerical variable		<input checked="" type="checkbox"/>
11	Cabin	Categorical variable		<input checked="" type="checkbox"/>
12	Embarked	Categorical variable		<input checked="" type="checkbox"/>
13	title	Categorical variable		<input checked="" type="checkbox"/>

Search variable: _____ Import 891 rows, 13 variables

自动识别变量类型

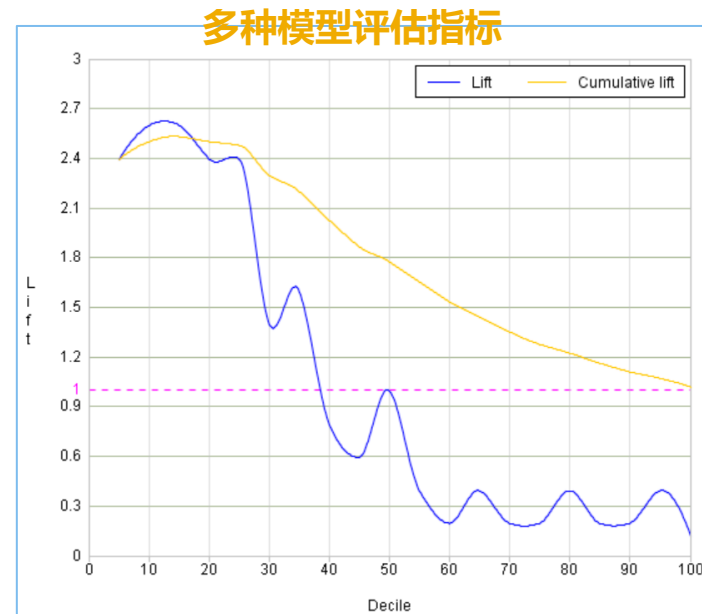
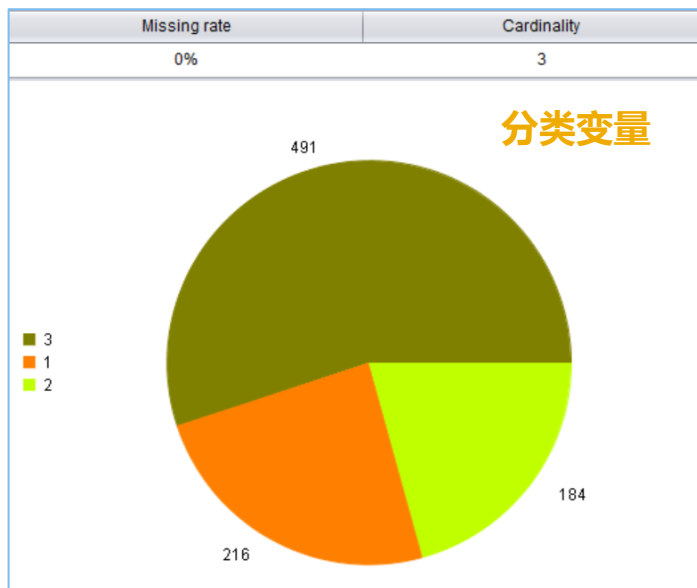
数据集大小统计



自动预处理+建模

```

Build model
INFO: Modeling data preparing...10%
[2020-03-21 15:55:50]
INFO: Modeling data preparing...20%
[2020-03-21 15:55:50]
INFO: Modeling data preparing...30%
[2020-03-21 15:55:50]
INFO: Modeling data preparing...40%
[2020-03-21 15:55:52]
INFO: 2020-03-21 15:55:52.126315: W tensorflow/stream_executor/platform/default/dso_loader.cc:55] Could not load dynamic library 'cudart64_101.dll'; dlerror: cudart64_101.dll not found
[2020-03-21 15:55:52]
INFO: 2020-03-21 15:55:52.126785: I tensorflow/stream_executor/cuda/cudart_stub.cc:29] Ignore above cudart dlerror if you do not have a GPU set up on your machine.
    
```



为什么是我们？



1 统计学专家毕生追求的呕心力作

数十年数据挖掘建模的实战经验，参与和主持过多次国内外银行、保险业数据挖掘项目，屡次带队在国际SAS大赛中获奖

2 研发团队的扛鼎力作

深刻的数学理解能力、超强的软件实现能力、业界领先的高性能大数据技术

案例：个人信贷违约预测

目标

- 建立信贷违约模型，给出用户信贷违约概率
- 给出用户合理的信贷额度
- 让业务人员根据经验选择数据建模，帮助业务人员接受模型的应用与普及
- 提高违约客户捕获率

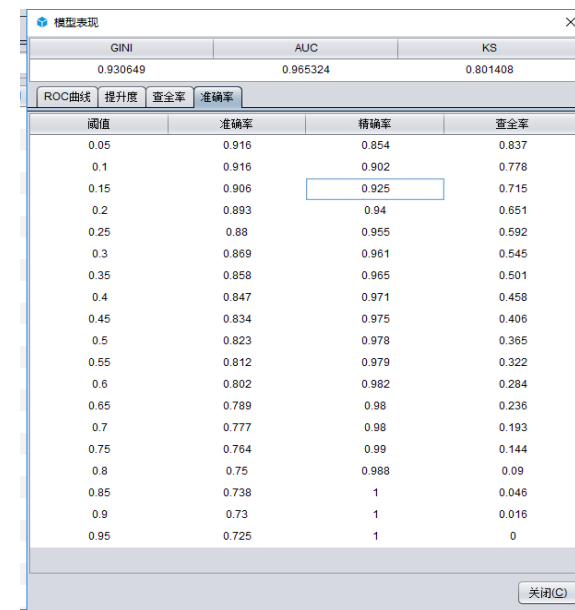
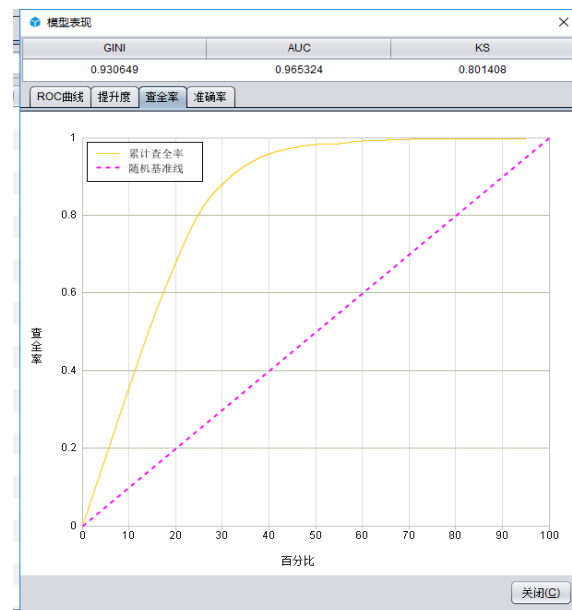
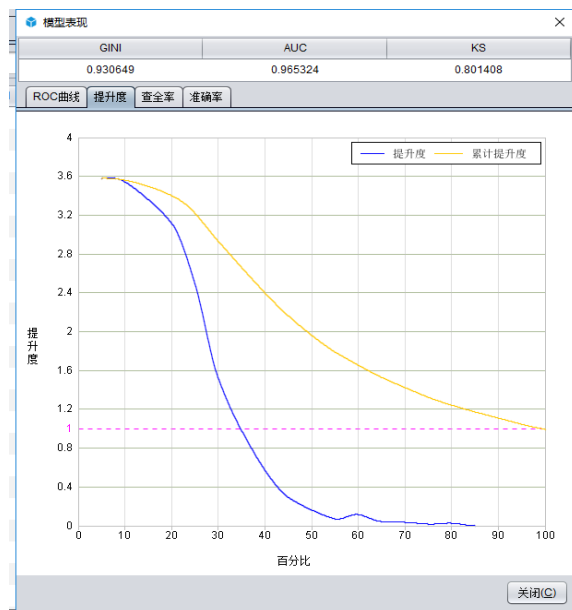
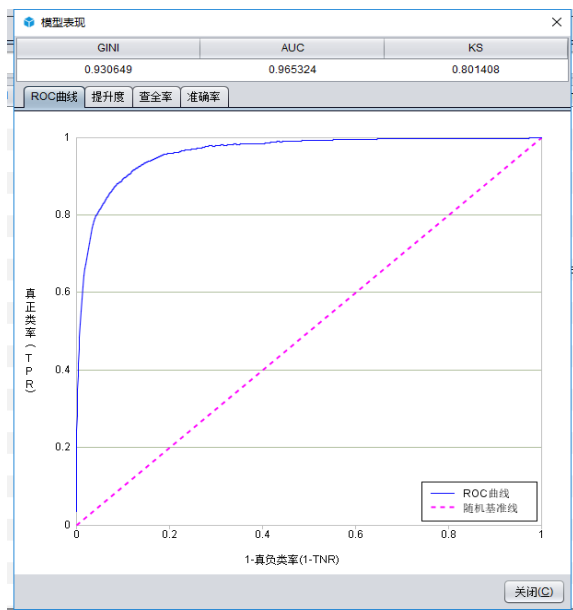
痛点

- 找到合理的数据维度
- 高基数与非线性问题对模型的影响
- 挑选合理的模型或者模型组合
- 阳性样本少，避免模型过拟合

建模结果对比

	智能建模	传统建模
人数	1	1
建模时间	5分钟 (数据预处理+建模)	2个月
建模数量	1	1
数据规模	100000+ / 28MB	100000+ / 28MB
模型AUC	0.9728 (测试集0.965)	0.957

模型表现 (测试集)



案例：理财产品精准营销

	客群1	客群2	客群3	客群4
建模人数	1	1	1	1
模型数量	13	13	13	13
建模时间	1.5小时/个	1.5小时/个	1分钟/个	2分钟/个
数据量	134万	155万	6400	1.2万

	累计提升度	累计捕获率	AUC值
前5%	14.4	72%	>0.9
前10%	9.4	94%	
前15%	6.3	94.5%	
前20%	4.8	96%	
当前该产品购买率为1.72%			

智能建模 VS 手工建模：

	模型数量	时间	项目参与人数
智能建模	50-60个	2周	1人
手工建模	不适合大量建模	1周~2个月/个 (实际取决于模型复杂程度和建模人员水平, 时间不可控)	数人

1. 使用模型前5%的数据比不使用模型效果提升**14.4倍**购买率。即每一百名精选的客户, 可以完成**24.77**次交易, 远高于平均每百名客户成交1.72次。
2. 由模型抓取的前5%的数据中, 可以捕获**72.0%**的目标客户。由模型抓取的前20%的数据中, 可以捕获**96.0%**的目标客户。

易明智能建模特点

人工智能少人工

AI



自动建模
效率高



非数据科学家
成本低



模型完善
准确率高

智能建模改变应用模式：业务用户主导，应用过程中随时随地建模

易明智能建模

挖掘数据价值

THANKS

