



2018

## 全球（南京）人工智能应用大赛

——人工智能社会保险反欺诈分析

团队：海尔炮  
成绩：A榜第一，B榜第二





# 目录

## CONTENTS

1

团队介绍

2

产品价值

3

产品优势

4

产品技术

5

商业应用





01

Part Four

# 团队介绍

来自北京中关村科技园周边高校的研究生组成的团队



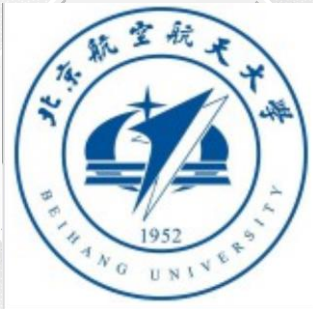


# 团队成员



崔灏

中央财经大学  
金融学硕士



李宇杰

北京航空航天大学  
软件工程AI方向



罗时超

四川大学  
统计学硕士



赵超杰

清华大学  
自动化学院机器学习方向



张鹏

微软亚洲研究院  
机器学习方向



# 团队成就



美国大学生数学建模竞赛  
(MCM/ICM)

一等奖



中诚信征信“风云杯”建模大赛

冠军



全国社会保险大数据应用  
创新大赛

季军

CCF大数据计算智能大赛

冠军

北京市校园高校大数据竞赛

亚军

腾讯社交广告高校算法大赛季

军

...





02

Part One

# 产品价值

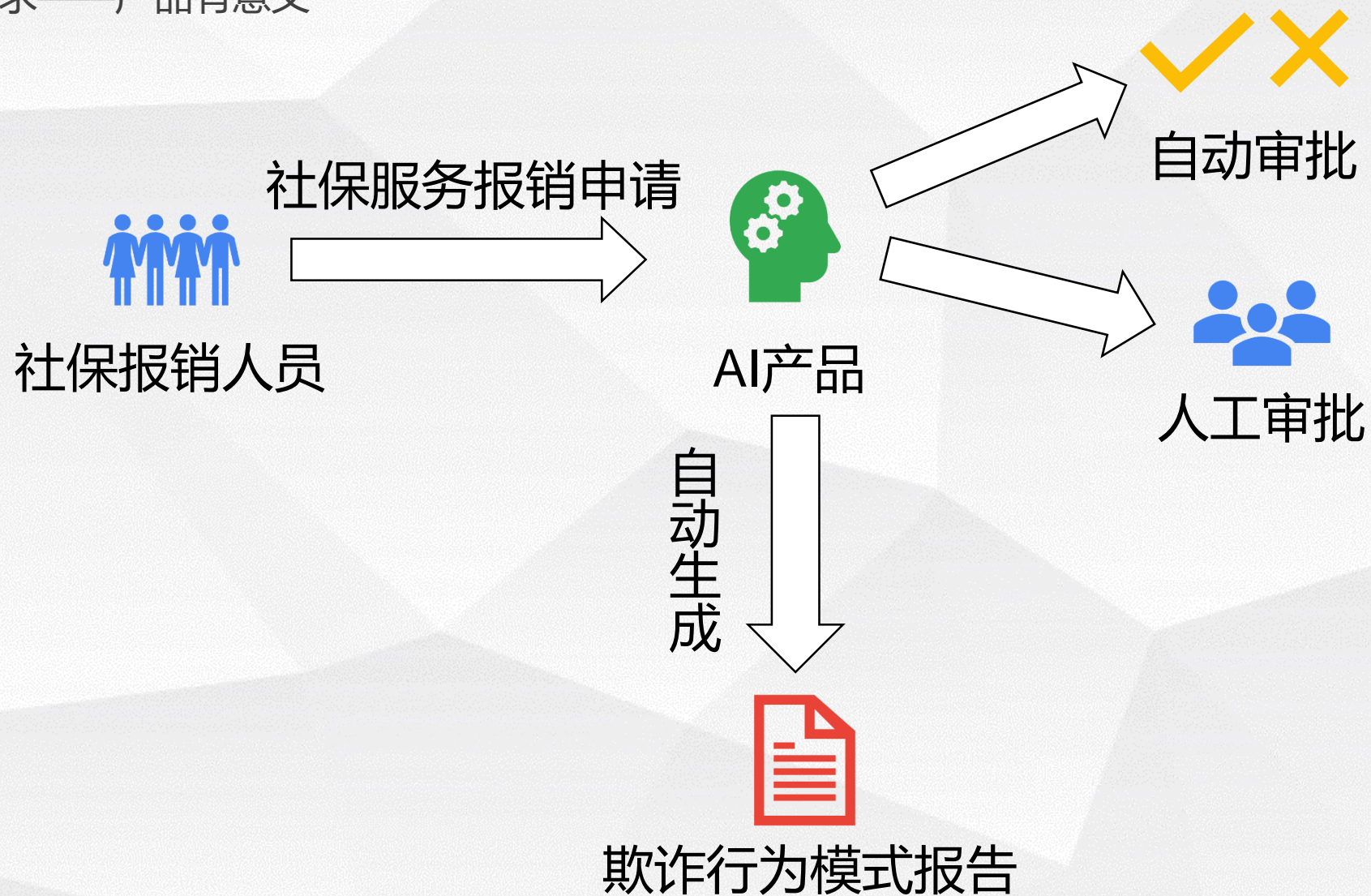
本AI产品创造的经济价值以及社会意义





# 产品功能

基本需求——产品有意义

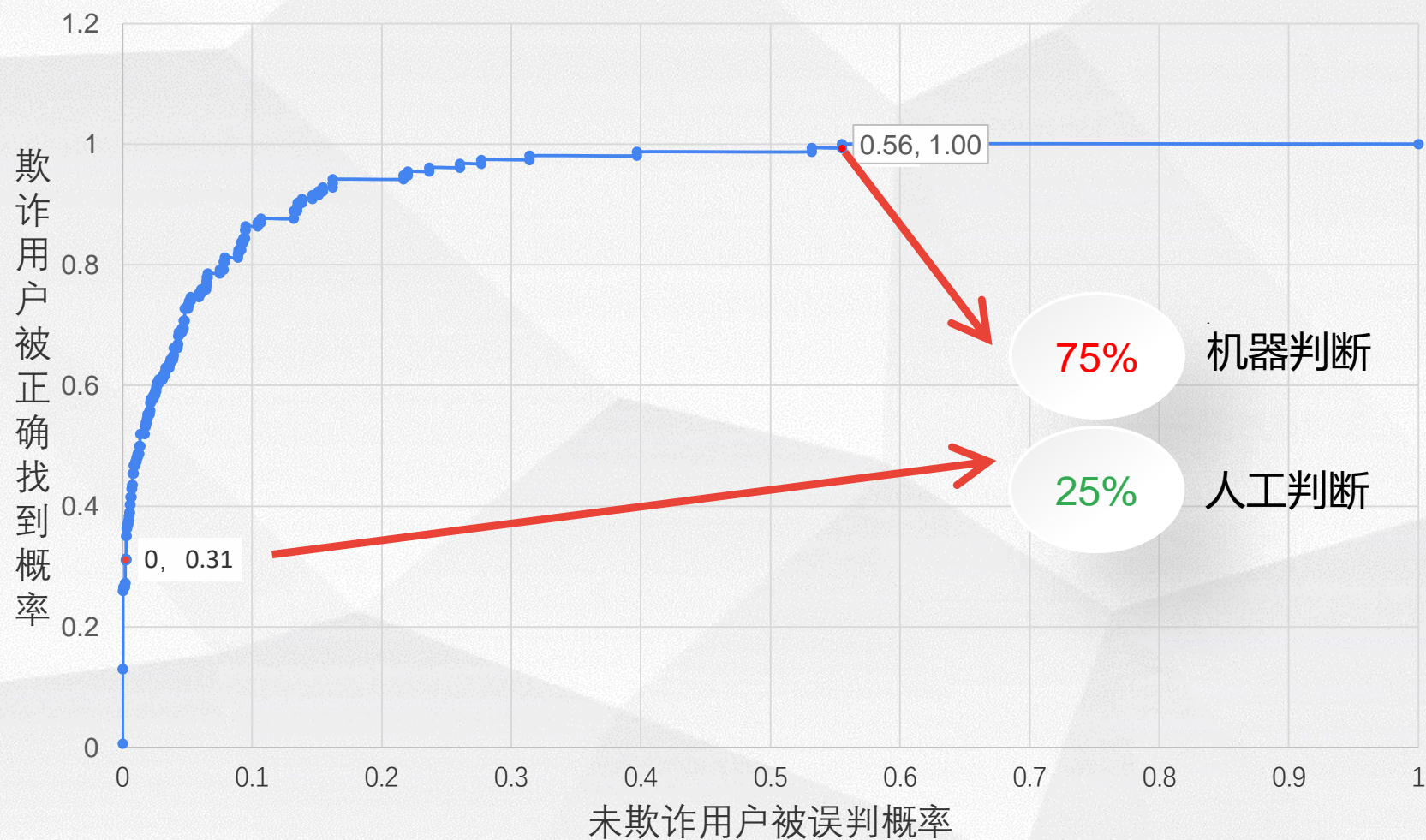




# 产品效果

可行性——产品能落地

AI产品预测效果





# 产品效果

可行性——产品能落地

## 用户检测报告

	欺诈概率	原因
用户1	0.76	涉及医疗项目过多
用户2	0.04	
用户3	0.27	单笔报销金额过高
用户4	0.02	
用户5	0.32	行为异常
用户6	0.01	
用户7	0.73	重复报销
用户8	0.05	
用户9	0.17	周期可疑
用户10	0.09	
用户11	0.03	
.....	.....	.....

## 欺诈模式报告

	重要度	影响方向
总金额	149	0.95
药品费用	130	0.67
自负比例	97	-0.41
金额c3统计量	68	0.34
金额偏度	45	0.23
金额峰度	43	-0.37
傅里叶系数	33	0.42
.....	.....	.....



# 产品效果

可行性——产品能落地



注：数据量：130w \* 50

**超低成本的模型开发成本**

## 速度

- 在kesci平台上，训练针对该赛题AI模型所用的全部时长。



## 系统开销

- 笔记本电脑利用开源环境就可以快速完成训练



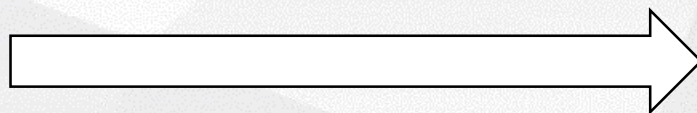


# 商业价值

市场广——产品能存活



全民医保推进



医保政策更加便捷惠民



国家统计局2017年公布数据



# 商业价值

市场广——产品能存活

调查方法不健全

“病医不相投” 调查方法争议大

无法满足未来医保情景

全民医保，即时报销

抽查覆盖范围小，助长  
个人投机主义，损害社  
会诚信



现有医疗监管体系与  
医保发展不相适应

造成社保基金严重流失

人力成本高昂

调查人员必须对医药病理知识有经验

调查方式漏网之鱼多

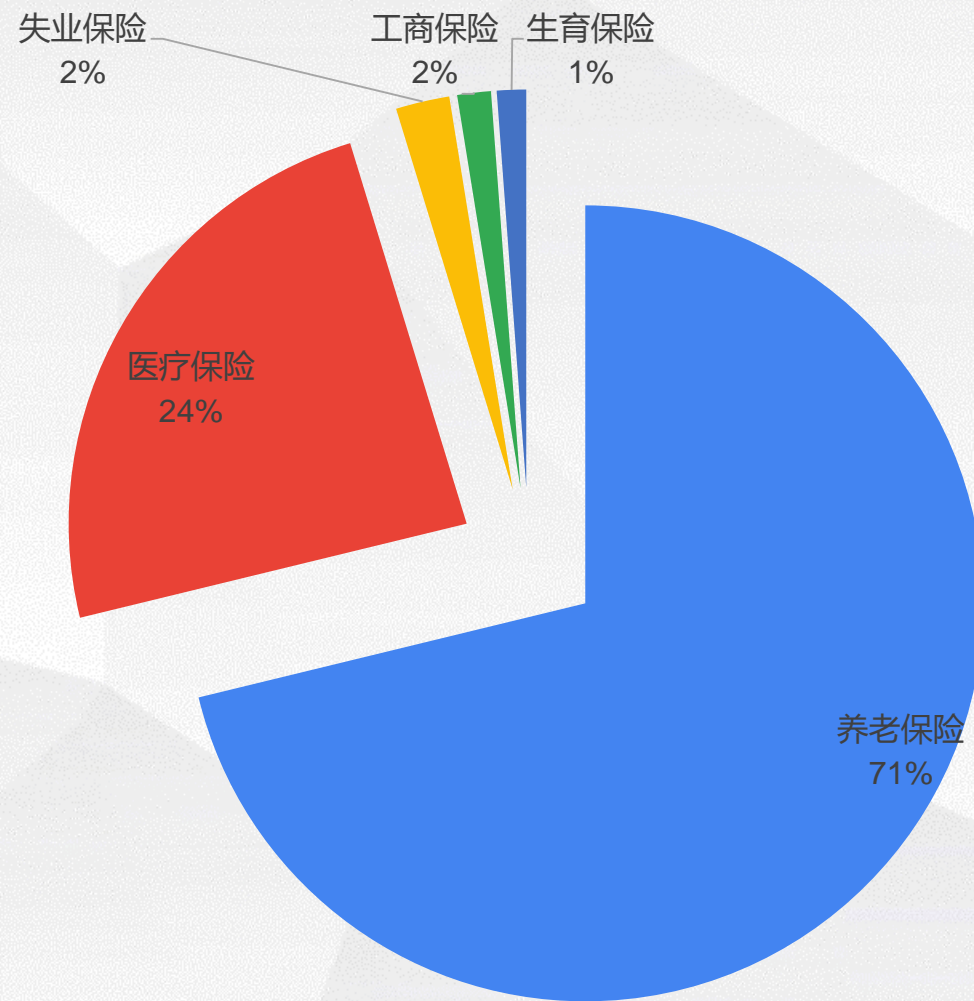
碍于人力有限，只能小范围抽查，重  
点筛查



# 商业价值

市场广——产品能存活

- 各社会保险支出均逐年递增
- 公布5大社会保险下47种欺诈行为



国家统计局2016年社会保险支出情况统计数据,  
总额=46888.4亿



# 社会意义

---

社会价值——产品能发展

AI智慧助力社保各项政策推进

AI监控保障社会诚信



# 产品价值

总结——产品能落地能发芽

应用于医疗保险监控  
产生经济价值巨大

01

泛化到社会保险监控  
产生更大经济价值

02

与医保社保政策发展  
相适应，助力其发展

03

保障社会诚信

04





03

Part Two

# 产品优势

与同类产品对比，我们的优势





# 与市场现有产品对比

相似产品少，竞争力强

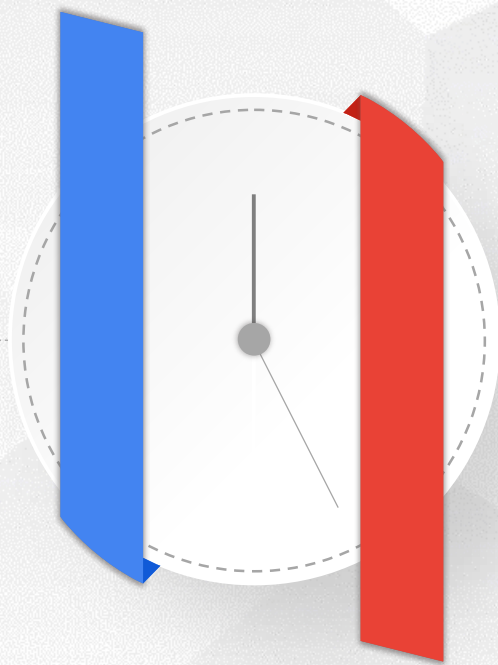
根据行业经验规则预测

## 市场现有产品

行业规则判断方式争议大

完全暴露个人隐私信息

应用场景在于预防而非监控



预测结果更准确可靠

隐私数据更安全

实时监控

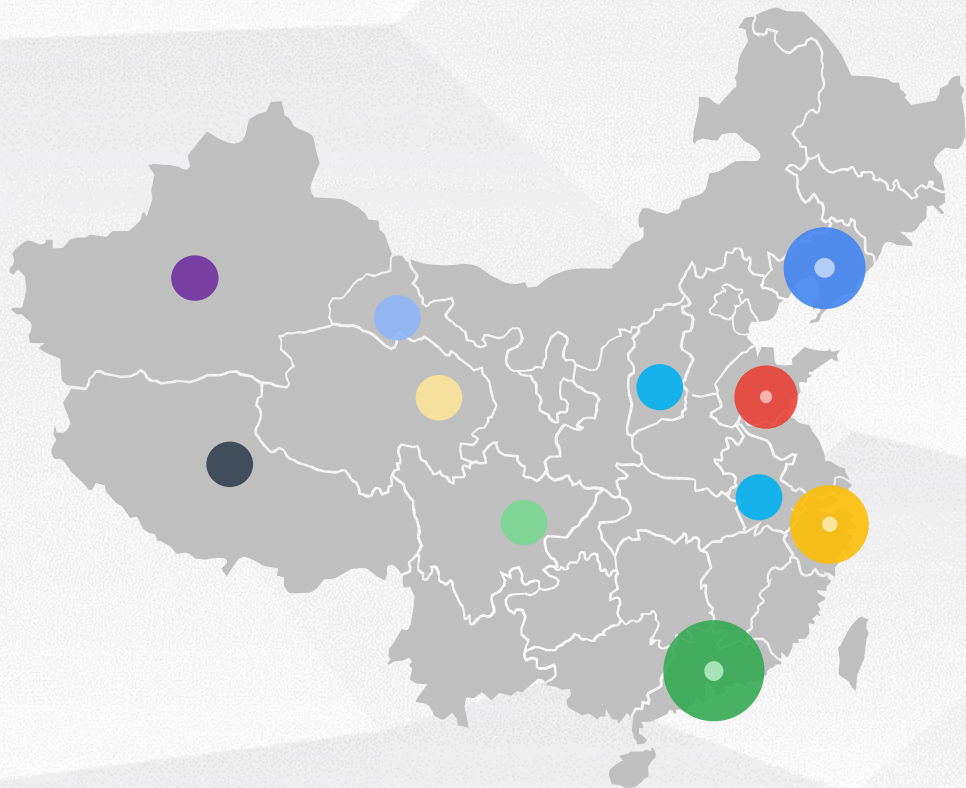
## 我们的AI智能产品

根据脱敏结算数据智能预测



# 与赛题下其他产品对比

数据驱动开发，理念创新



社保数据随地区政策具有多样性，  
产品以数据为驱动开发多种模型

训练  
速度

数据  
处理

产品技术支撑

快

全





04

Part Three

# 产品技术

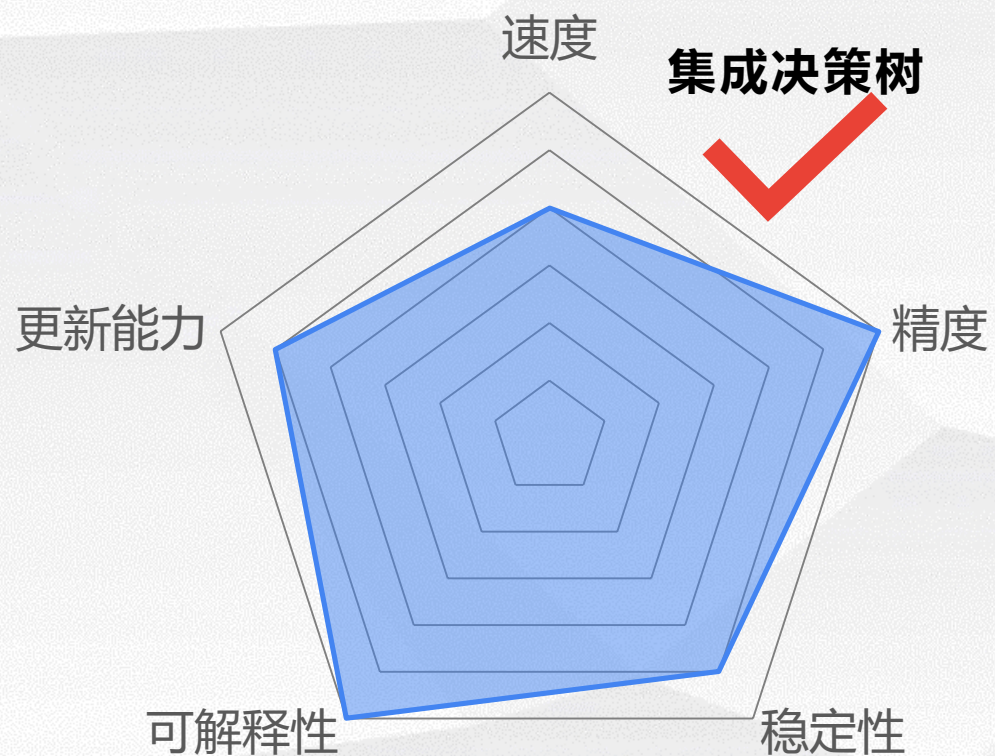


围绕以数据驱动开发理念，创新性提出新技术

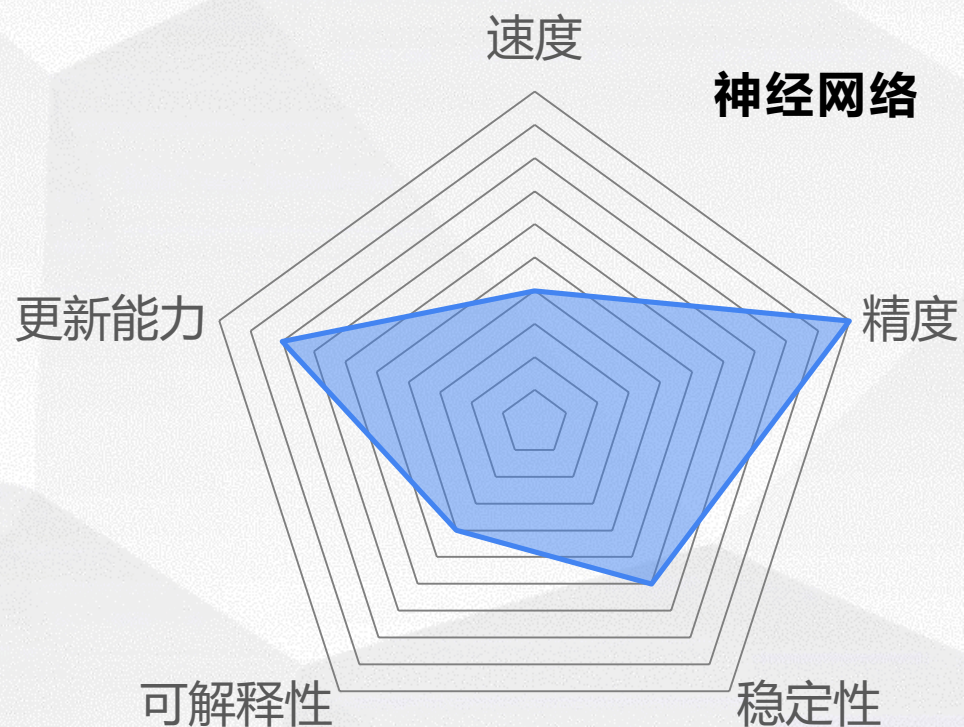


# 建模算法

模型训练速度快



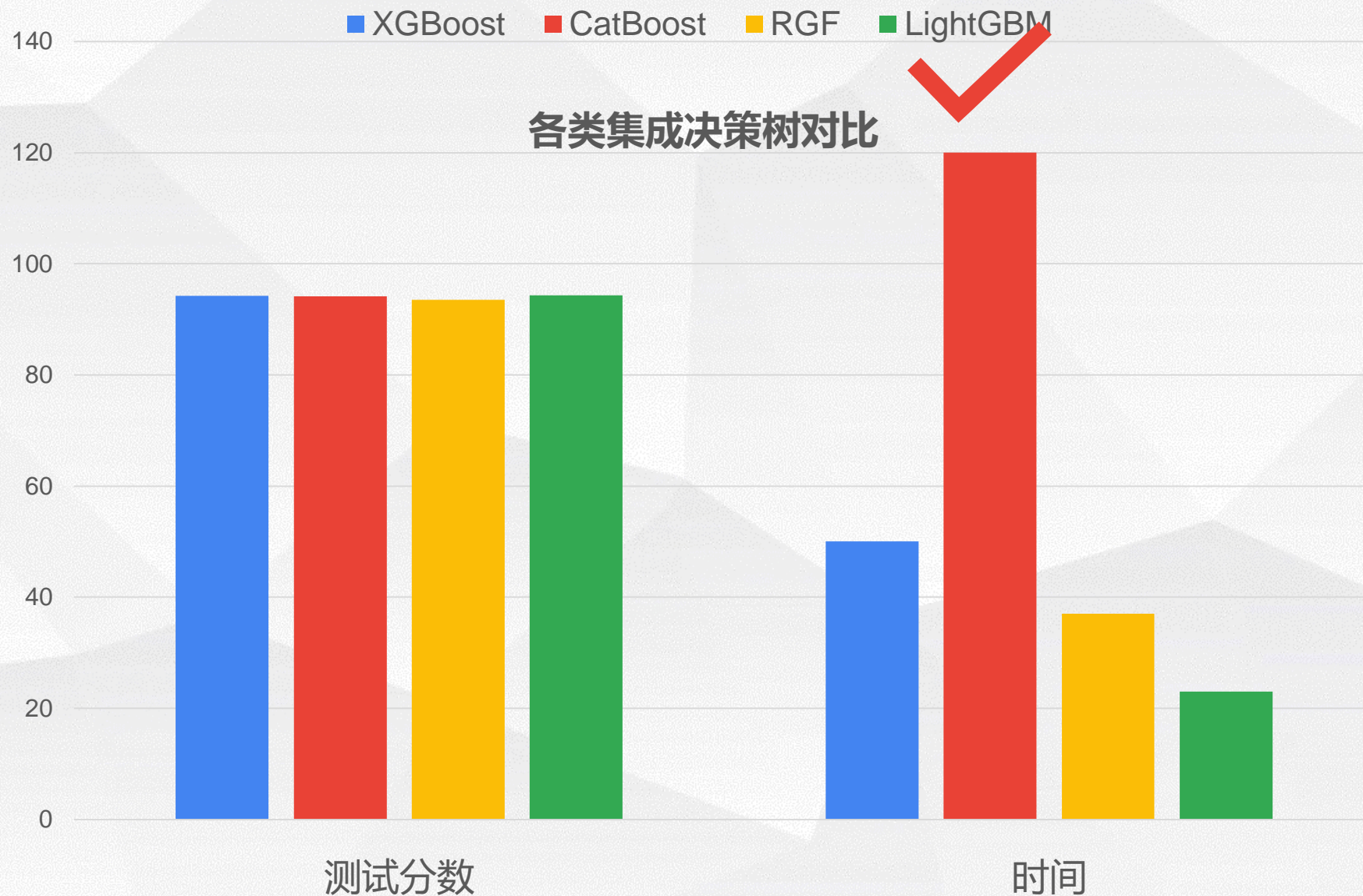
VS





# 建模算法

模型训练速度快





# 数据处理

挖掘数据特性全

$$\frac{1}{n-2lag} \sum_{i=0}^{n-2lag} x_{i+2lag}^2 \cdot x_{i+lag} \cdot x_i$$

偏度

孤立森林

自回归系数

峰度

$$\sqrt{\sum_{i=0}^{n-2lag} (x_i - x_{i+1})^2}$$

偏自相关系数

小波变换

CID

滑动窗口

关联分析

创新性引入近千种统计特征  
精确刻画用户

$$\sum_{i=1, \dots, n-1} |x_{i+1} - x_i|$$

C3统计量

近似熵

分箱熵

线性趋势

傅里叶系数

ADF检验量

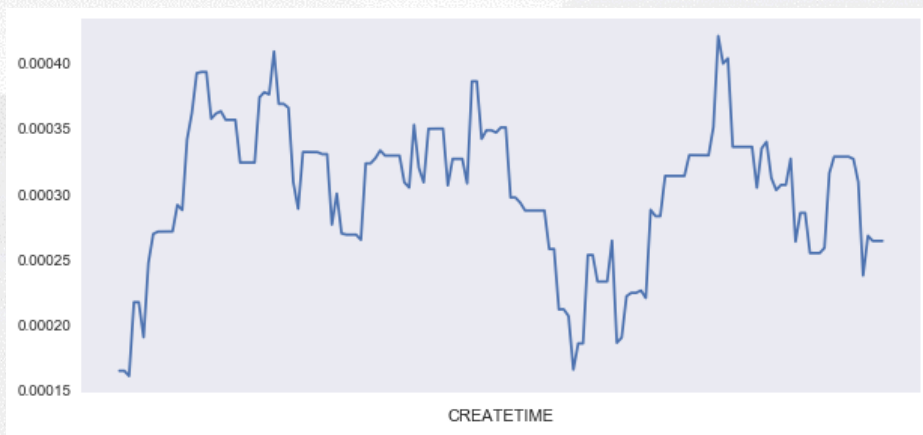
$$\alpha_k = \frac{Cov(x_t, x_{t-k} | x_{t-1}, \dots, x_{t-k+1})}{\sqrt{Var(x_t | x_{t-1}, \dots, x_{t-k+1}) Var(x_{t-k} | x_{t-1}, \dots, x_{t-k+1})}}$$

$$A_k = \sum_{m=0}^{n-1} a_m \exp \left\{ -2\pi i \frac{mk}{n} \right\}, \quad k = 0, \dots, n-1.$$



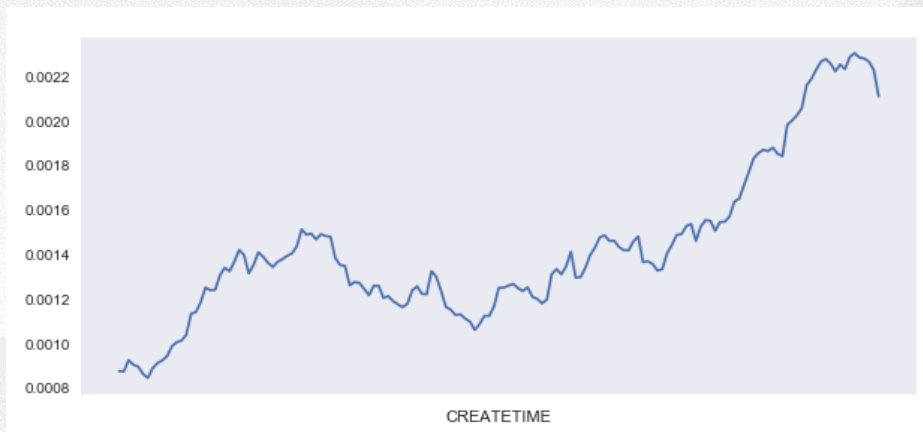
# 结果呈现

## 一个例子



### 正常就诊者

金额规律性变化  
时间呈现周期性  
与就诊项目相符



### 涉嫌欺诈者

金额不断增大  
时间没有规律  
与治疗项目严重不符



# 产品效果

可行性——产品能落地

## 用户检测报告

	欺诈概率	原因
用户1	0.76	药病不符
用户2	0.04	
用户3	0.27	金额过高
用户4	0.02	
用户5	0.32	行为异常
用户6	0.01	
用户7	0.83	重复报销
用户8	0.05	
用户9	0.17	周期可疑
用户10	0.09	
用户11	0.03	
.....	.....	.....

## 欺诈模式报告

	重要度	影响方向
总金额	149	0.95
药品费用	130	0.67
自负比例	97	-0.41
金额c3统计量	68	0.34
金额偏度	45	0.23
金额峰度	43	-0.37
傅里叶系数	3	0.42
.....	.....	.....



# 辅助算法

---

算法应用创新——确保预测效果



## 团伙诈骗识别

通过聚类、相似度分析等算法，锁定高风险人群，识别团伙诈骗。



## 异常行为监测

使用孤立森林模型，实时监测可疑的异常行为。



## 就诊项目分析

Apriori关联分析，就诊项目与欺诈行为关联



# 技术创新

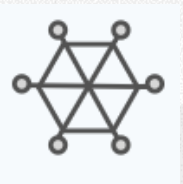
---





# 技术优势

---



## 实时快速响应

对每笔报销行为达到毫秒级响应



## 预测精准稳定

A榜第一，B榜第二（差距小）



## 敏感数据保护

支持数据脱敏



## 快速更新迭代

建立模型速度快





05

Part Three

# 商业应用

进一步打造社会保险反欺诈智能云平台——如何落地





# 产品战略分析

总结——市场无同质化产品，产品开发理念更适应于社保数据特点

在不同地区，不同政策下，产品效果稳定

为了产品实际可行，牺牲了少量准确性



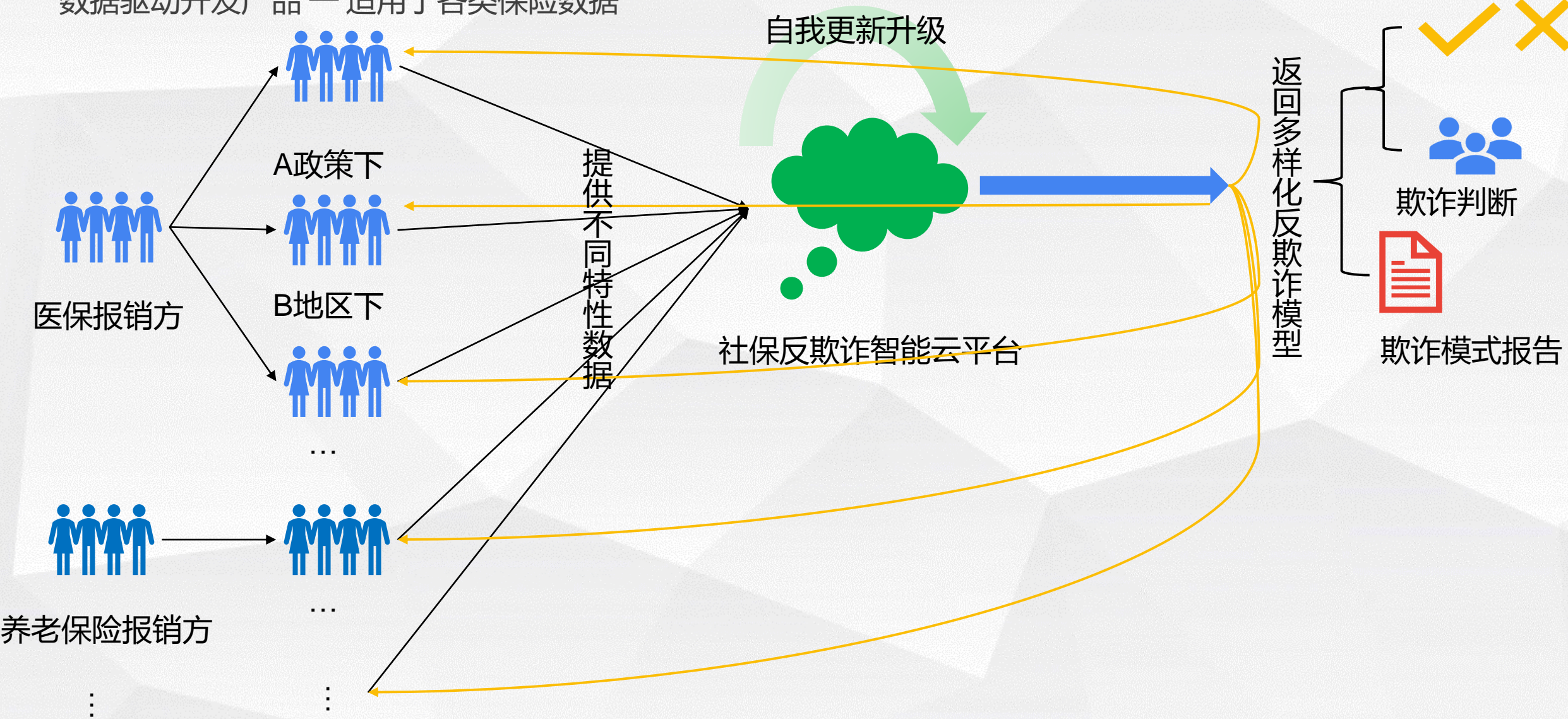
更适应未来医保管理集中  
更容易发展到全国各地

准确性可能略逊专门  
针对一地区产品一筹



# 保险反欺诈智能云平台

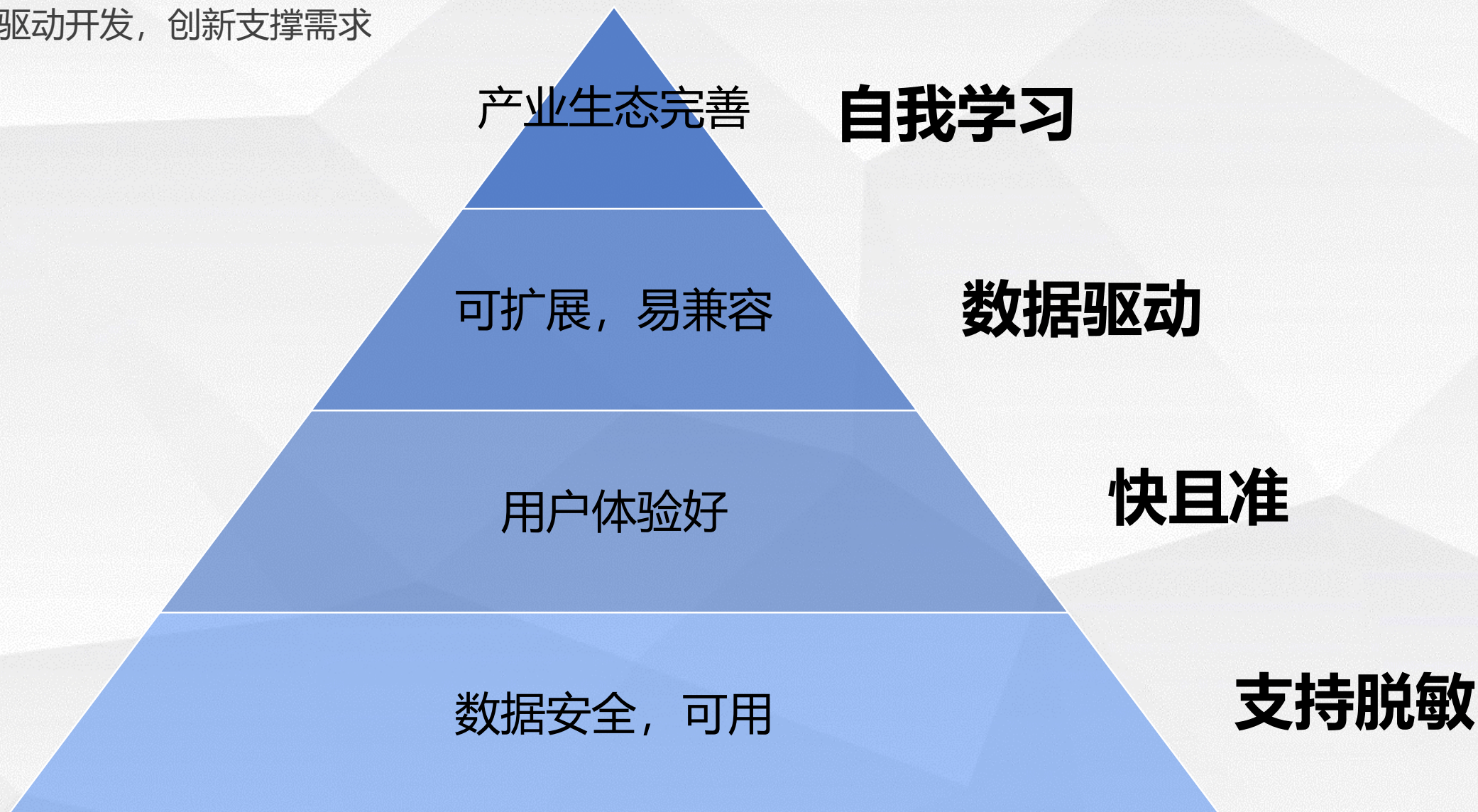
数据驱动开发产品 — 适用于各类保险数据





# 总结

数据驱动开发，创新支撑需求





# 产业生态闭环

---





感谢观看