

法律声明

■ 本课件包括演示文稿、示例、代码、题库、视频和声音等内容，北风网和讲师拥有完全知识产权；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或者机构不得盗版、复制、仿造其中的创意和内容，我们保留一切通过法律手段追究违反者的权利。

■ 课程详情请咨询

◆ 微信公众号：北风教育

◆ 官方网址：<http://www.ibeifeng.com/>



AI人工智能之机器学习项目

金融风险控制中的反欺诈项目

主讲人：Joe Chen

上海育创网络科技有限公司



课程要求

■ 课上课下 “九字” 真言

- ◆ 认真听，善摘录，勤思考
- ◆ **多温故，乐实践**，再发散

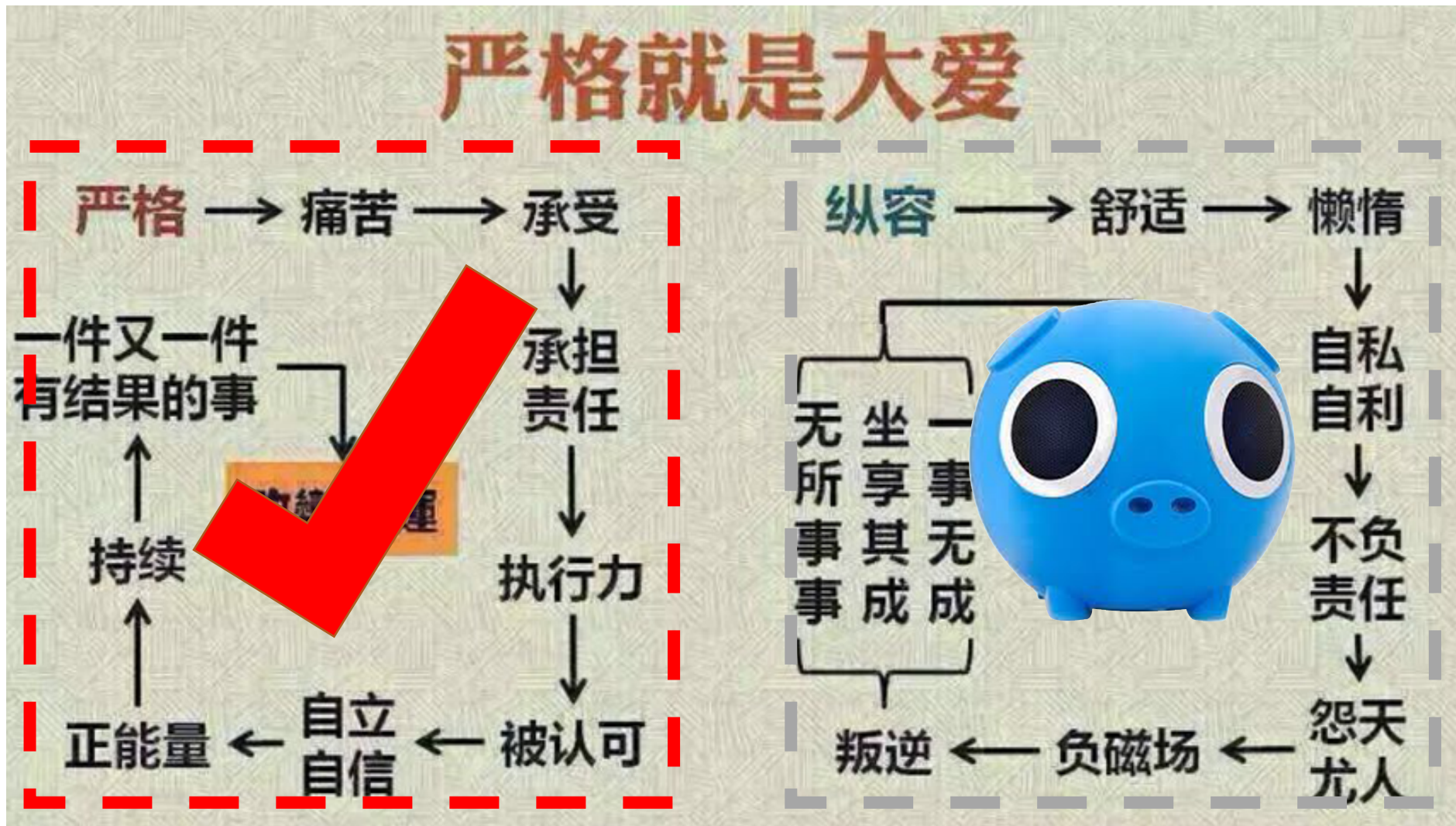
■ 四不原则

- ◆ **不懒散惰性，不迟到早退**
- ◆ **不请假旷课，不拖延作业**

■ 一点注意事项

- ◆ 违反 “四不原则”，不包就业和推荐就业

严格是大爱



寄语



做别人不愿做的事，
做别人不敢做的事，
做别人做不到的事。

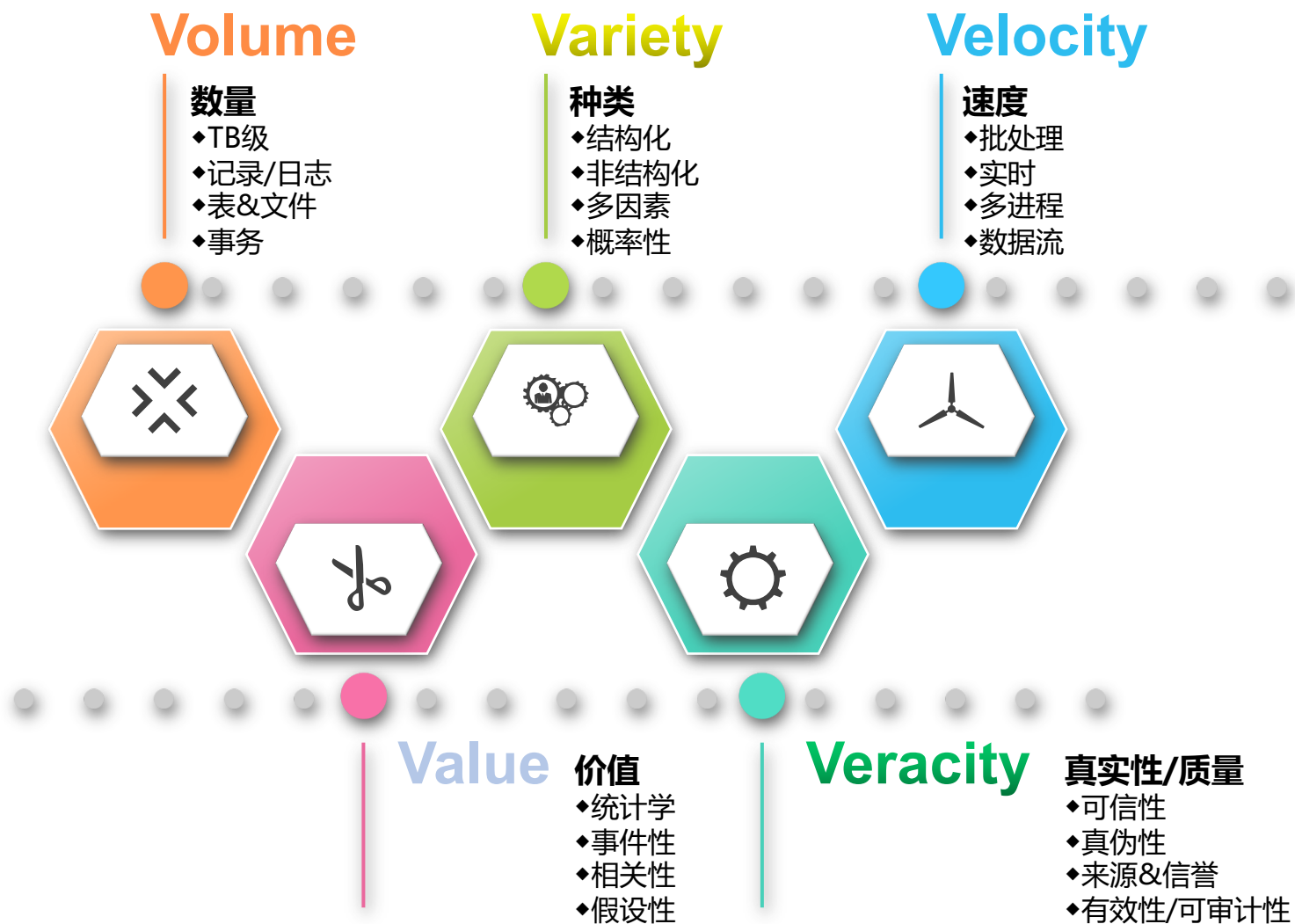
课程内容

- 垃圾邮件过滤
- 音乐系统文件分类
- 金融反欺诈项目

金融与科技的结合

- 金融的本质：资源的最合理利用
- 互联网技术：交易的边界成本趋向“零”
- 金融科技：用大数据、云计算等技术实现资金的融通、支付、投资和信息中介服务；利用大数据、机器学习等相关领域技术对大量的数据进行处理分析操作，并得出一些对于业务有一定促进能力的结论。

金融与科技的结合



金融产品借贷产品排名

发展排名	平台	发展指数	上线时间	所在城市	成交	人气	合规	品牌	分散度	透明度	关注	对比
1	宜人贷 	78.04	2012.07	北京.朝阳	93.46	85.98	85.12	85.49	94.73	51.42	关注	对比
2	陆金服	77.87	2012.01	上海.浦东	97.66	90.8	90.65	73.03	93.47	54.24	关注	对比
3	人人贷	77.83	2010.10	北京.海淀	89.31	77.54	94.63	75.89	94.18	71.24	关注	对比
4	拍拍贷	76.42	2007.06	上海.浦东	91.36	75.39	77.43	86.82	95.26	62.13	关注	对比
5	点融网	76.39	2013.03	上海.黄浦	77.4	73.44	88.08	85.96	83.83	73.06	关注	对比
6	爱钱进	73.91	2014.05	北京.东城	93.2	88.14	85.75	58.45	93.99	56.16	关注	对比
7	微贷网 	73.34	2011.07	浙江.杭州	88.77	83.18	69.7	72.54	94.66	61.94	关注	对比
8	有利网	73.26	2013.02	北京.海淀	96.77	85.18	82.11	58.17	91.71	65.01	关注	对比
9	团贷网  	72.43	2012.07	广东.东莞	72.68	78.91	71.32	74.13	94.76	73.54	关注	对比
10	积木盒子	72.28	2013.08	北京.朝阳	72.37	73.77	85.44	59.52	91.27	76	关注	对比
11	投哪网	70.62	2012.05	广东.深圳	74.11	72.65	77.04	68.28	87.32	66.29	关注	对比
12	翼龙贷 	69.3 ↑ 2	2011.04	北京.海淀	77.86	68.36	82.71	66.09	94.99	63.3	关注	对比
13	搜易贷	68.85 ↓ 1	2014.09	北京.海淀	63.27	61.39	81.28	67.76	78.41	63.6	关注	对比

中国信用贷款领域产品分级

APR: Annual Percentage Rate – 年化利率

Prime

$APR < 10\%$

$10\% \leq APR < 20\%$

省呗 还呗

Near prime

$20\% \leq APR < 40\%$

宜人贷 给你花 小花钱包 我来贷
 平安I贷 投哪网 卡卡贷 久融金融
 功夫贷 小赢卡贷 名校贷 读秒 信通贷

$40\% \leq APR < 80\%$

马上金融 拍拍贷 小企鹅
 贷款王 飞贷 叮当贷 你我贷

Sub-prime

$APR \geq 80\%$

闪银 闪电借款 信用钱包 发薪贷 借钱快
 用钱宝 现金巴士 手机贷 现金白卡

金融行业面临的欺诈风险

风险	遇到的问题	业界通常解决方法	业界的方法为什么无效
信用风险	还款能力	收集收入水平、消费水平、负债情况等对用户进行风险评分	无权威数据、数据收集难度大、传统评分卡有效特征挖掘难度大
欺诈风险	伪冒申请和欺诈交易	人工审查、信用黑名单、基于规则	人工效率低、无权威黑名单、无法自动发现异常、欺诈手段更新快

人群团体化



地区集中化



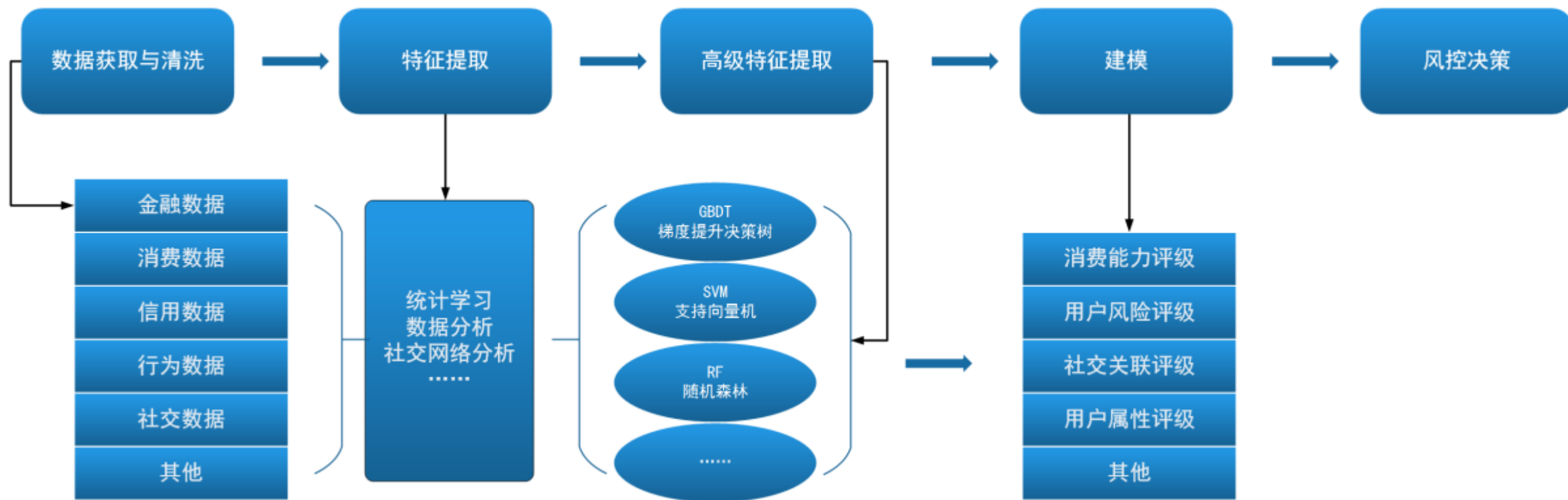
方式多样化



工具智能化

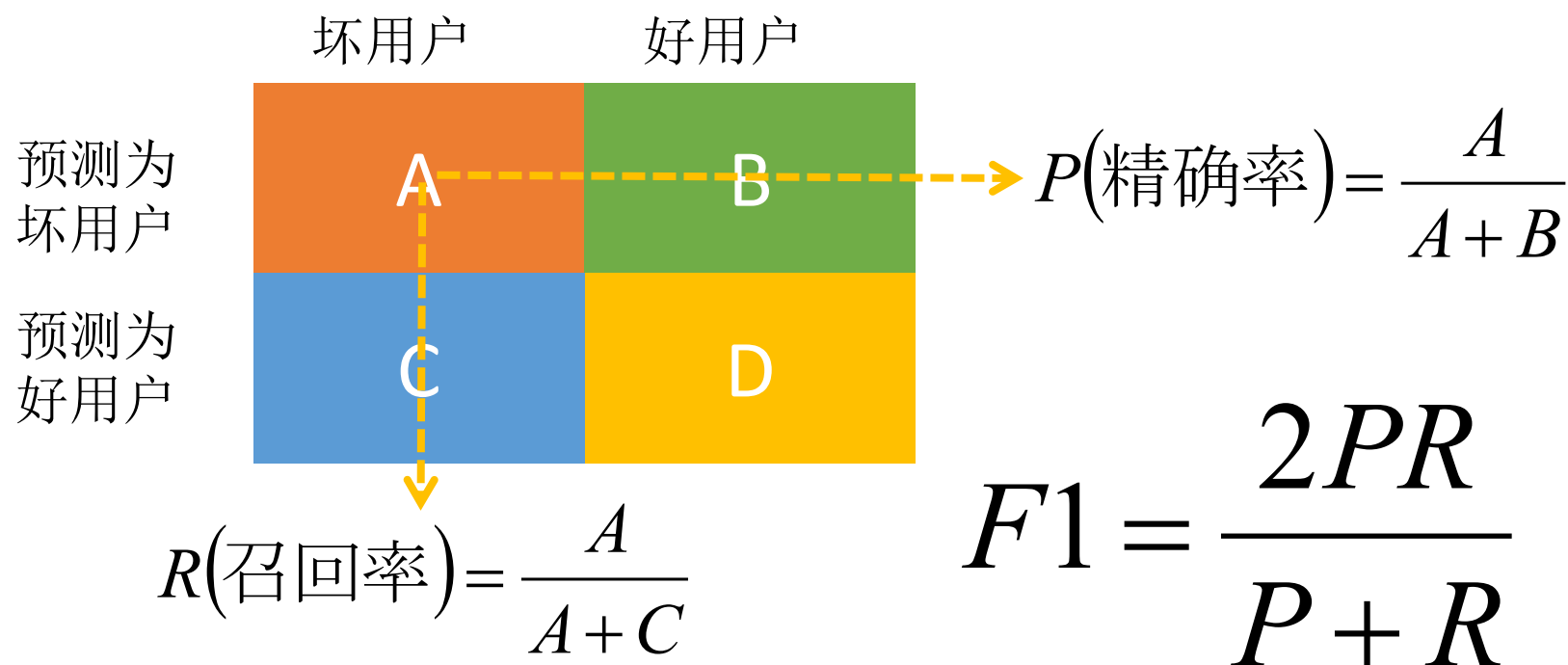


反欺诈建模过程



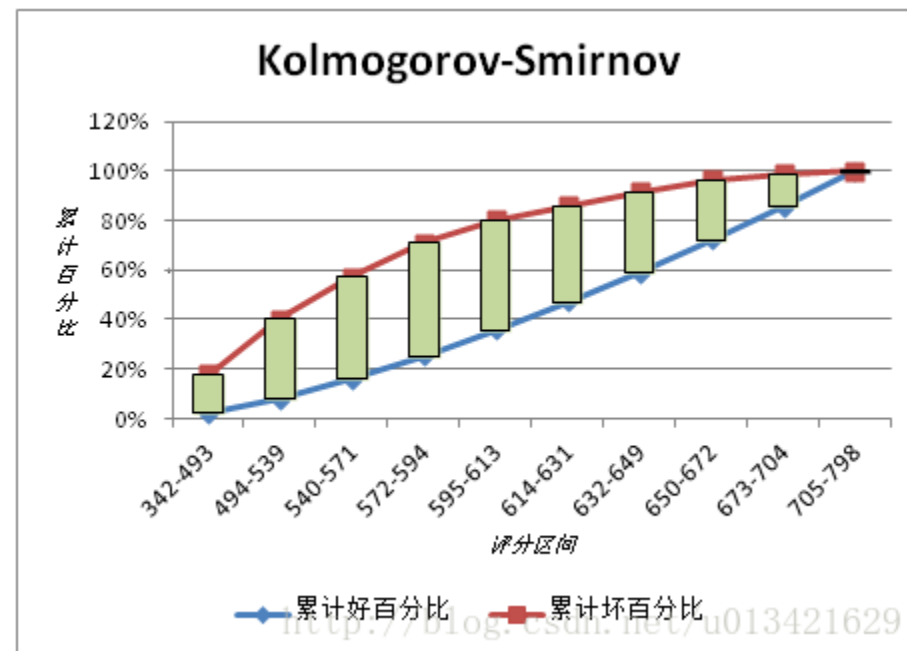
模型效果评估

- 在风险控制这种类型的项目中，只要有一笔坏用户的借贷申请没有被拦截掉，那么就有可能产生很大的损失，所以在这种类型的项目中，对应算法的精准度要求，可能没有那么严格，但是对于算法模型的召回率要求相对来讲比较严格。但是从借贷的业务上来讲，又期望具有较高的精确率，以保证足够的利润空间，所以一般可以考虑采用F1值作为评价指标



模型效果评估

- 在金融风险控制领域，除了使用F1值作为指标外，经常还使用KS指标作为风控中评分卡的效果评估指标参数，KS指标的主要是对模型的风险区分能力进行评估，直白来讲，就是模型将好坏样本分隔开这种能力。
- KS指标计算过程：
 - ◆ 计算每个评分区间的好坏账户数
 - ◆ 计算每个评分区间的累计好账户数占总好账户数比率(good%)和累计坏账户数占总坏账户数比率(bad%)
 - ◆ 计算每个评分区间累计坏账户占比与累计好账户占比差的绝对值（累计good%-累计bad%），然后对这些绝对值取最大值即得此评分卡的K-S值



模型效果评估

- 一般情况下，当KS百分比值大于20的时候，我们就可以认为模型的区分效果相对来讲可以接受，KS值一般满足如下要求：
 - ◆ 小于20，此模型没有区分能力
 - ◆ 20~40之间，模型勉强可以接受
 - ◆ 41~50之间，模型具有区分能力
 - ◆ 51~60之间，模型具有很好的区分能力
 - ◆ 61~75之间，模型具有非常好的区分能力
 - ◆ 大于75，此模型可能存在异常，有问题

算法项目实施

- 下载LendingClub上的借贷数据集，下载地址为：
<https://www.lendingclub.com/info/download-data.action>
- 对数据进行特征工程处理，并将处理的结果数据以csv格式进行输出
- 构建模型进行训练



THANK YOU

上海育创网络科技有限公司