

僵尸企业画像及分类

通过对企业从业人数、成立年限、注册资本、营业收入、风险信息、行政处罚、纳税信用等级、黑名单、上市信息、电商信息等数据关联处理，按需进行权重分割，并对这些数据进行综合分析，构建出企业全息画像，整体评估一个企业综合价值，更好的勾勒出企业的经营变化情况。

如何根据企业全息画像，面向不同的应用场景对企业分类识别，分析挖掘法人、自然人、物品（包括特种设备、产品、食品等）、案事件之间的显性和隐性关系，建立市场主体信用管理与风险评估模型，对企业的信用风险、经营风险、质量风险、安全风险、关联风险等进行全面评估及分级预警，为系统监管、精准监管、高效监管提供数据支撑。

僵尸企业是指缺乏盈利能力却能够以低于市场最优利率成本获得信贷资源，依靠外界输血而缺乏自生能力的企业。

5. 项目说明

企业数据一部分为通用的标签信息，其他的新增标签由参赛人员根据经验和数据本身特点，针对具体问题制订；然后通过数据治理，并采用统计、分析、挖掘等技术，完善企业画像的各维度标签；最后采用人工智能算法建立僵尸企业的分类模型，达到对企业进行合理的分类识别的目的。

【用户期望】

要求：

- （1）多表数据融合的能力；
- （2）数据特征提取的能力；
- （3）数据预处理的能力；
- （4）数据建模的能力。

目标：追求模型分类识别的精确性与高效性。

6. 任务要求

【开发说明】

需要对僵尸企业的表现形式有深入的了解和阐述。

需要依据用户期望的内容对数据的处理过程进行详细的描述。

【任务清单】

- （1）模型的概述与简介，对算法模型有充分的认识理解，切不可只会套用；
- （2）模型的参数调优过程（如果包含调优过程）；
- （3）模型的效率（包括数据预处理的效率与打标签的效率）。

【提交材料】

- (1) 项目概要介绍;
 - (2) 项目简介 PPT;
 - (3) 项目详细方案;
 - (4) 项目演示视频;
 - (5) 企业要求提交材料:
- Ø 所使用的分类方法原理与参数调优过程概述
 - Ø 训练后的分类方法, 以及对测试集打标签结果
 - Ø 可视化 web 端开发, 支持单个与批量输入查询分类结果
 - (5) 团队自愿提交的其他补充材料。

中国工业部门僵尸企业的全貌。

(1) 分年份来看, 2000-2013 年, 中国工业部门的僵尸企业比例最高时 (2000 年) 大约 30%, 此后呈下降趋势, 并在 2004 年之后保持稳定。2005-2013 年的工业部门僵尸企业比例大约为 7.51%。这说明中国僵尸企业的问题是在逐渐缓解的。

(2) 分行业来看, 利用 2013 年中国上市公司数据, 人大课题组发现: 僵尸企业比例最高的五个行业是: 钢铁 (51.43%)、房地产 (44.53%)、建筑装饰 (31.76%)、商业贸易 (28.89%) 和综合类 (21.95%); 僵尸企业比例最低的五个行业是: 银行 (0.00%)、传媒 (4.12%)、非银金融 (4.65%)、计算机 (5.23%) 和休闲服务 (5.88%)。

(3) 分地区来看, 经济发展水平较高的东部南部地区僵尸企业比例比较低, 而经济发展水平较低的西南、西北和东北地区僵尸企业比例较高。**将地区划分一下做可视化可能会有所帮助。**

(4) 分所有制来看, 国有和集体企业中僵尸企业的比例最高, 民营企业和港澳台及外商企业中僵尸企业的比例相近, 且远低于国有和集体企业中僵尸企业的比例。**注意企业类型**

(5) 分规模来看, 大型企业和中型企业的僵尸企业比例最高, 但由于基数比较小, 大部分僵尸企业还是小型企业。

(6) 分年龄来看, 随着企业年龄的增长, 僵尸企业的比例越来越高。成立 1-5 年的企业中, 只有约 3%的企业是僵尸企业; 而在成立超过三十年的“老”企业中, 约有 23%的企业都是僵尸企业。注册时间也很重要

导致僵尸企业的五个主要原因。

(1) 地方政府和企业之间的政企合谋。为了政绩和维稳, 地方政府不断给濒临破产的僵尸企业进行各种形式的“输血”, 或者给非僵尸企业施加就业压力和产量扩张压力, 然后再通过**补贴和贷款**来维持其局面。这使得本来不是僵尸的企业变成了“僵尸”、已经是“僵尸”的企业更加难以清理。**我们发现, 如果一个企业获得了更多补贴、或者冗员更多、或者是国企, 那么成为僵尸企业的概率就更高。 所以融资当中项目获得政策得补贴过多得应当予以重视**

(2) 地方政府之间和国企之间的恶性竞争。一旦国家将某个行业列入重点扶持范围，地方政府就一拥而上支持这个行业的发展，造成重复建设、产能过剩。在该行业出现大量僵尸企业之后，地方政府又纷纷提供优惠政策和补贴来支持当地企业，希望能够通过自己的“扶持”来“挤垮”其他地区的企业。导致僵尸企业与地区和产业相关

(3) 大规模刺激的后遗症。2008 年 11 月，为了应对国际金融危机给我国经济带来的压力，中国政府推出了投资总量约四万亿的经济刺激计划。然而，四万亿投资计划在一些行业引起了过度投资、盲目扩张，埋下了产生僵尸企业的隐患。

(4) 外部需求冲击。2008 年全球金融危机后，世界主要经济体增长放缓、需求减少，使得出口依赖型行业和企业在短时间内受到了巨大的冲击，订单不足、商品滞销、资金周转困难、投资方撤资，许多原本发展良好的企业纷纷陷入困境，甚至沦为僵尸企业。企业类型应当予以关注

(5) 银行的信贷歧视。2008 年以后，国有和集体企业的利润率在波动中下降，但负债率却逐年增高；与此同时，民营企业利润率一直比较稳定，但负债率却有所下降。也就是说，2008 年经济危机之后，虽然国有和集体企业盈利能力有所下降，但却更容易获得贷款；虽然民营企业盈利能力非常稳定，但却更难获得贷款。集体企业应当予以重视

企业的活动综合下来有三种：经营、投资、融资。怎样在企业三种活动中发现机会、识别风险，这就是企业画像存在的价值。

基本情况、经营情况、财务情况、融资情况

1.Base_train_run（基本信息）

字段大类类型	字段名	解释说明
企业基本信息	ID	一个企业身份凭证
	注册时间	企业注册生效
	注册资本	认缴制 注册不需要缴清 破产时清算要
	行业	企业所经营的领域
	区域	企业所在地
	企业类型	<p>集体所有制：这是指一定范围内的劳动群众出资举办的企业。</p> <p>股份制企业：企业的财产由两个或两个以上的出资者共同出资，并以股份形式而构成的企业。我国的股份制企业主要是指股份有限公司和有限责任公司（包括国有独资公司）两种组织形式。</p> <p>有限责任制公司：由 50 个以下的股东出资设立，每个股东以其所认缴的出资额对公司承担有限责任，公司法人以其全部资产对公司债务承担全部责任的组织。</p> <p>股份有限制公司：由 2 人以上 200 人以下的发起人组成，公司全部资本为等额股份，股东以其所持股份为限对公司承担责任。</p> <p>合作社：合作社是指根据合作原则建立的以优化社员（单位或个人）经济利益为目的的非盈利企业形式。</p> <p>合伙企业：合伙企业是由两个或两个以上的自然人通过订立合伙协议，共同出资经营、共负盈亏、共担风险的企业组织形式。我国合伙组织形式仅局限于私营企业。合伙企业一般无法人资格，不缴纳所得税。合伙企业可以由部分合伙人经营，其他合伙人仅出资并共负盈亏，也可以由所有合伙人共同经营。</p>
	控制人类型	实际控制人是指“虽不是公司的股东, 但通过投资关系、协议或者其他安排，能够实际支配公司行为的人。”

		<p>企业法人是指具有符合国家法律规定的资金数额、企业名称、组织章程、组织机构、住所等法定条件，能够独立承担民事责任，经主管机关核准登记取得法人资格的社会经济组织。</p> <p>自然人具有权利能力，即获有在法律上享受权利，负担义务之资格。</p>
	控制人持股	<p>举个例子</p> <p>实际控制人持股比例如何计算，比如实际控制人张三持有上市公司 13%股份，张三持有 A 公司 56%股份，A 持有上市公司 23%股份。张三持有 A 公司 56%，导致张三绝对控制了 A 公司</p> <p>A 持有上市公司 23%股份，导致张三间接持有上市公司 23%股份</p> <p>张三持有上市公司的股份是 $13\%+56\%*23\%$ (张三享有的权益为这么多)，但是他能控制的股份为 $13\%+23\%=36\%$</p>
	Flag	是否为僵尸企业 1 是 na 不确定

2.Money information (融资信息)

字段大类类型	字段名	解释说明
融资额度	债权融资额度	债权融资是指企业通过举债的方式进行融资，债权融资所获得的资金，企业需要支付利息，并在借款到期后向债权人偿还本金。最常见的即银行贷款
	股权融资额度	企业通过出让部分企业所有权，通过企业增资的方式引进新的股东的融资方式。股权融资所获得的资金，企业无须还本付息，但新股东将与老股东同样分享企业的赢利与增长。
	内部融资和贸易融资额度	<p>内部融资：公司利用内部自有资金，即累积未分配利润筹集投资项目所需要的资金的行为。内源融资主要指企业提取的折旧基金、无形资产摊销和企业的留存收益。</p> <p>贸易融资：指银行对进口商或出口商提供的与进出口贸易结算相关的短期融资或信用便利。</p>
	项目融资和政策融资额度	<p>项目融资是近些年兴起的一种融资手段，是以项目的名义筹措一年期以上的资金，以项目营运收入承担债务偿还责任的融资形式。</p> <p>以政府信用为担保的，政策性银行或其他银行对一定的项目提供的金融支持。主要以低利率甚至无息贷款的形式，其针对性强，发挥金融作用强。</p>

融资成本	债权融资成本	连本带息付给债权人
	股权融资成本	显性： 隐性： 部分股权出让
	内部融资和贸易融资成本	
	项目融资和政策融资成本	

3.year_report_train_sum (年报)

Year	经营年份
ID	公司编号
从业人数	公司经营规模大小
资产总额	资产总额是指企业资产负债表的资产总计项。公司过去的交易、事项形成并由企业拥有或控制的资源，资产总额即是指企业拥有或控制的全部资产，这些资产包括流动资产、长期投资、固定资产、无形及递延资产、其他长期资产等。
负债总额	负债总额是指过去的交易、事项形成的现时义务，履行该义务预期会导致经济利益流出企业，包括流动负债和长期负债。
营业总收入	营业收入是指企业在从事销售商品，提供劳务和让渡资产使用权等日常经营业务过程中所形成的经济利益的总流入。分为主营业务收入和其他业务收入。
主营业务收入	主营业务收入是指企业从事本行业生产经营活动所取得的营业收入。主营业务收入根据各行业企业所从事的不同活动而有所区别，如工业企业的主营业务收入指“产品销售收入”；建筑业企业的主营业务收入指“工程结算收入”；交通运输业企业的主营业务收入指“主营业务收入”；批发零售贸易业企业的主营业务收入指“商品销售收入”；房地产业企业的主营业务收入指“房地产经营收入”；其他行业企业的主营业务收入指“经营(营业)收入”。
利润总额	利润总额是“亏损总额”的对称。企业在一定时期内通过生产经营活动所实现的最终财务成果。企业纯收入构成内容之一。工业企业的利润总额，主要由销售利润和营业外净收支（营业外支出抵减利润）两部分构成。按规定缴纳资源税的企业，应交的资源税也是利润总额的组成部分之一（抵减利润）。利润总额应按规定在国家与企业之间进行分配。
净利润	净利润是指企业当期利润总额减去所得税后的金额，即企业的税后利润。所得税是指企业将实现的利润总额按照所得税法规定的标准向国家计算缴纳的税金。它是企业利润总额的扣减项目。
纳税总额	累计纳税总额是指增值税 17%+销售税金及附加 12%+所得税 25%的总计数据。
所有者权益合计	所有者权益合计是指企业投资人对企业净资产的所有权。企业净资产等于企业全部资产减去全部负债后的余额，其中包括企业投资人对企业的最初投入以及资本公积金、盈余公积金和未分配利润。对股份制企业，所有者权益即为股东权益。

4.knowledge_train_sum (知识产权)

ID	公司编号
专利	在现代，专利一般是由政府机关或者代表若干国家的区域性组织根据申请而颁发的一种文件，这种文件记载了发明创造的内容，并且在一定时期内产生这样一种法律状态，即获得专利的发明创造在一般情况下他人只有经专利权人许可才能予以实施。在我国，专利分为发明、实用新型和外观设计三种类型。
商标	商标 (trade mark) 是一个专门的法律术语。品牌或品牌的一部分在政府有关部门依法注册后，称为“商标”。商标受法律的保护，注册者有专用权。
著作权	即版权