



基于集成算法融合模型的电商订单异常检测

1. 项目介绍

本案例是针对中国某500强企业在2013年的部分电商交易数据进行的异常预测。该案例数据来源于某500强企业流出的非脱敏数据，最初在2018年时首次在Gitee被公开，是**货真价实来自企业、带有企业业务信息的现实数据**。这份数据披露了某500强企业（国美）在电商领域的战略部署、业务逻辑以及2013年全年完成的成绩，从中我们可以稍微复原电商战争兴起之时电商领域的竞争概况。对任意线上购物平台而言，平台的关键职责之一就是准确判断每笔订单背后的购买意图、并以此为依据保护买家权益，这有助于维护平台环境，有利于给买家、卖家带来更好的交易体验。因此，在交易完成之后迅速对交易进行异常检测（风险评估）是很重要的一环。在这个案例里，**我们将利用该电商平台在2013年内进行的部分交易数据进行建模，并实现对异常订单的识别检测。**

2. 项目背景

2.1 电商的核心价值：更高的交易效率

电商是国民经济的重要产业之一，也是互联网时代下我们生活中关键的组成部分。**在2021年，全国大约有8.12亿人参与过线上购物，占网民总体的80.3%**。在电商用户中，许多人都可以轻松理解电商交易的具体流程，并对线上购物驾轻就熟，但我们却对于电商行业本身却知之甚少。即便作为互联网工作者/技术工作者，我们理解电商的关键在于“人货场”，却也很少能够触达电商行业本质与核心价值，比如：

- 为什么电商行业能够发展如此迅速呢？整个行业的核心价值在哪里？
- 假设电商平台提供的核心价值是“链接”（让买家能够买到线下买不到的产品），那为什么在提供同一价值的情况下，会有淘宝、京东、苏宁、唯品会等看起来高度相似的产品形态？
- 电商平台盈利吗？电商平台的收入从哪里来？

要回答这一系列问题，我们需要对电商的核心价值进行深入的剖析。

要理解电商，首先得理解**零售**。零售是一系列商业模式的统称。如果一种商业模式是通过某种交易结构，让消费者和商品之间产生链接、并最终实现成交，那这种商业模式就是零售的一种形态。

用阿里系的话来说，零售就是把最终付费的“人”（消费者）和“货”（商品）链接在一起的场。当这种场是实际存在的店铺时，零售发生在商场里，当这种场是线上构建的商城时，零售就成为了电商。毫无疑问的，电商平台的核心作用是链接供给侧与需求侧并促成交易，因此——

电商本质上是一种零售方式。

零售的本质是把人和货链接在一起的场。

因此，几乎所有的电商分析都离不开“人货场”三要素，人货场本身就是零售业永恒的概念。

- **人** - 消费者以及与消费者行为相关的一切业务（如销售、转化、复购等）
- **货** - 商品和服务，包括提供商品的卖家、品牌及供应链
- **场** - 交易场所，包括交易场景、交易服务、交易结构等

现在问题来了，如果说电商本质就是一种零售的话，为什么线上电商会挤占线下零售的生存空间呢？在线上购物时，人们会损失线下购物时的体验、会对产品品质、商家信用等产生不信任，理论上来说，如果同样的商品可以在线下以同样的价格买到，人们就不会线上购物。但事实完全相反。**电商并不只是单纯地将交易的“场”搬到了线上，而是在“人货场”这三大零售要素中取得了比线下购物更高的交易效率。**

在我国独特的社会和经济条件下，在零售业中——

- 消费者总是渴望更快、更便捷地买到符合需求的好产品，因此会（无意识地）追求更高的交易效率
- 商家会有意识地追求更高的交易效率并且愿意为此付费，因为更高的效率会带来交易优势与利润

在整个行业的发展历程中，无论模式、商品、交易场景如何变化，**更高的效率都是整个电商行业的核心价值之一**。我们来看看具体是怎么回事：

电商在“人货场”三大要素中的交易效率比线下购物更高。

消费者/商家总是在追求更高的交易效率，并且愿意为高效付费。

- **场的效率：信息流、资金流、物流**

- **【优势】信息流：**消费者了解商品信息的成本越小，效率就越高

线下 - 腾出时间、找到相关门店、看商品、咨询门店工作人员、人力比价

线上 - 足不出户，信息触手可得，商品包罗万象、又多又全，还自动比价

互联网提供的信息流异常高效

- **【优势】资金流：**消费者付费方式越便捷，效率越高

线下 - 现金支付（专人收款、点清现金、退还零钱、检验真伪），刷卡支付

线上 - 扫码支付，无需找零、无需专人、无需验伪

互联网的资金流便捷性很高

- **【----】物流：**商品越快到达消费者手中，效率越高

线下 - 近，即买即得，容易辨别真伪优劣，但货品范围有限、难以比价

线上 - 快，世界商品触手可得，价格也会被拉平，但质量参差不齐

线下的“近”与线上的“快”形成了相似的效果

不难发现，线上交易在许多方面都具有效率优势，以往需要徒步一整天才能完成的工作，通过电商只需要点点手机就可以完成，交易效率大幅提升，对用户的吸引力极大，侧面降低了商家的获客成本。

- **人的效率：流量、转化、客单价、复购**

- **【优势】流量**

线下 - 人流，将店铺开在人流量多、目标人群集中的地段可以吸引更多的人流

线上 - 流量，忽略距离、时间、地域等限制，上百种渠道触达比线下多数倍的用户

互联网的流量优势会极力扩大每个店铺的潜在消费者

- **【----】转化**

线下 - 成交，依赖于线下销售人员、体验感进行成交

线上 - 转化，依赖于店铺宣传材料、商品评价、推荐系统、社群进行成交

根据商品类别不同，线上线下成交/转化率会有巨大区别。在需要体验感、或金额巨大的交易中，线下更有优势。在需求明确、金额较小的交易中，线上更有优势。

- **【----】客单价**

线下 - 主要依赖于线下销售人员的业务水平

线上 - 推荐系统、满减优惠、提高连带率来增加客单价

相同价格的产品在线上客单价会更有优势，因为用户对线上连带接受度更高。

但相同的产品在线上会有更低的价格。

- **【优势】复购**

线下 - 回头客，必须先获取用户联系方式/或品牌已具备口碑，才能二次触达

线上 - 复购率，触达方式丰富（推送、私信、短信等），可大幅提高复购率

在商品质量有保证的前提下，线上二次触达效率大幅领先

“人”的效率要素其实涉及到多方面、多因素的影响，因此不像“场”的效率要素那样容易比较。然而，线上的优势是显而易见的：

- 无论在与消费者相关的哪个交易环节上，电商都能使用多种方法提升其交易效率、甚至以很小的成本快速改变其转化或销售策略，因此**对商家的吸引力极高**。相对的，线下环境中的则没有这么多选择、也没有这么灵活——一家线下店不太可能在短时间内随人流的移动而改变地址，但线上店则可以在任何流量聚集的地方进行交易。
- 除此之外，线上的每个环节都能够捕获到精准的数字，但线下却很难精确监控“人”相关的一系列数据，因此线下可以优化的程度就更小。

从各个角度来说，电商在“人”相关的效率点上都是更有优势的。

- **货的效率：定倍率**

定倍率是商业世界最基础的逻辑之一，它衡量一个产品/一种服务从产出到售出过程中的效率。

定倍率 = 商品零售价/商品成本

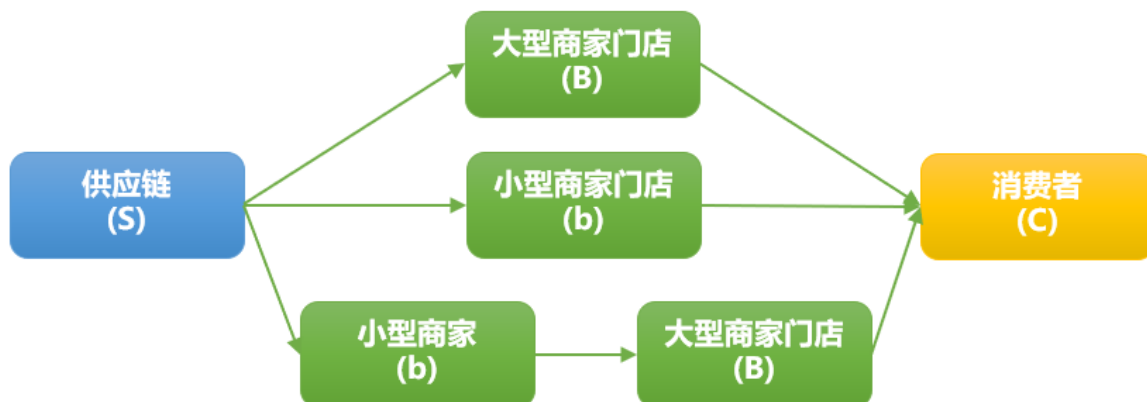
100元成本的商品卖500元，则该商品的定倍率是5。定倍率越高，则说明商品从产出到售出之间的效率越低。为什么定倍率可以衡量产出到售出过程中的效率呢？我们先来看看从商品产出到售出需要经手多少流程：



一个产品被设计(D)出来，交与制造商进行制造(M)，再通过供应链进入大小商家或线上平台(B,b,P)的货架，最终被卖给消费者(C)。在这个流程中，制造、供应商、平台、商家都需要赚取利润，因此每经手一轮，就会多一个“中间商”赚差价，也就多了交易成本。对商家而言，交易成本就是商品被送到消费者手中的成本，该成本越高，商品的价格也就必须相应地越高，定倍率也会高，商品的市场竞争力就会越弱。**如果能够缩短/改善商品交换的流程，就能够提升货的效率，从而降低商品价格，获得更高的市场竞争力。**

在传统线下的B2C(b2C)模式中，小型店铺或大型店铺直接面向消费者售卖产品，因此店铺需要支付的租金、仓储、水电费等基础费用都会被包括在产品价格中。除此之外，有许多小型店铺没有单独的线下店，他们是在大型商场或超市中占据一个门面，此时的模式是b2B2c，用户付出的价格会更贵，定倍率更高。

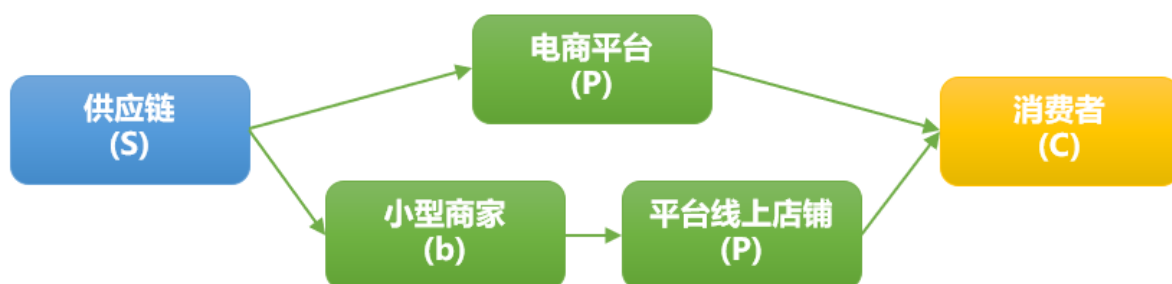
线下零售模式



线下模式类比到线上，会出现两个明显的不同：

- 在线上，小型商家大多无法直接接触到消费者，因此很少有不借助平台实现的b2C
- 在借助平台的模式中，供应商先给小型商家供货，小型商家再将商品上架到平台，形成b2P2C模式（这也是当前电商领域最普遍的模式，由于平台的存在实在是太理所当然了，因此我们常常把这种模式也称之为b2C）。在这种模式下，上线平台本身是免费的（相当于中间商“平台”并没有赚差价），线上装饰店铺、制作宣传图像的成本与线下店铺租金比起来完全不值一提，货品成列的成本可以大幅度被消减。同时，由于商家的成本变低、上线后竞争环境更加透明、商品价格也随之降低，将大幅降低定倍率。

电商模式



【优势】天然的，大部分商品/服务的线上定倍率是低于线下的，这说明电商在“货”领域的效率常常能够更高。这是为什么我们在美团团购时总能拿到比门店更低的价格，也是“网购”看起来总是比线下更加便宜的核心原因。在这一前提下，巨头们还纷纷执行缩减更多流程的M2B模式、M2C模式、C2M模式，再结合“新零售”概念中的线上结合线下模式，电商行业中“货的效率”还在继续提升。

不难发现，电商在“人货场”三大要素中的交易效率比线下购物高很多，毫无疑问**更高的效率是电商行业的核心价值之一**。毫不夸张地说，正是**线上信息流天然的效率优势 + 中国飞速增长的国民购买力**催生了最初的C2C、B2C电商。2005年后**国家基础设施建设的飞速发展 + 本土电商企业持续提升各环节交易效率**，让电商行业生机勃勃、机遇不断。毫无疑问，电商是为效率而生，也因效率才壮大的。

2.2 辨析电商交易模式：开放平台vs价值链整合

电商的核心价值就是更高的交易效率。正是因为高于线下购物的效率优势，电商才得以蓬勃发展、成为国民经济的支柱产业之一。但电商的发展并非一蹴而就，这一行业并不是天生就带有各方面难以战胜的效率优势，相反，**电商在1990年代初生时，不仅在人和货的效率上存在巨大问题，在场的效率方面更是处于绝对劣势的**。在过去的三十年中，中国本土电商平台创意不断、艰难探索、持续提升各环节的交易效率、克服了一个个效率上的难题，才成就了今日全球最大规模的电商市场。

特别的，本土电商的两大巨头淘宝和京东，在发展历程中从不同的角度提升了不同的交易效率，从而形成了不同的模式，我们来看看——

• 淘宝：打造电商效率优势的根基

回到2003年，中国的电子商务市场还是一片混沌，除了遭受全球互联网危机的牵连之外，互联网配套的基础设施、以及民众对于电子商务的认知都严重不足。那时候电商唯一的优势，就是**信息流**：商家和用户可以跨越千里进行交易，人们终于不受地域限制进行购物了。但在其他方面的劣势，会让信息流上天然的优势变得不再重要。比如说——

场的效率：资金流

- **【劣势】最初的线上支付方式是带身份证去银行转账、去邮局汇款，是一件很重大的事儿**
- **【劣势】必须开通网上银行，人们对网银的安全问题十分担忧**
- **【劣势】无法“一手交钱一手交货”，用户和商家对彼此有强烈的“不信任感”**

为此，**支付宝**及平台配套的信任体系被发明出来，成为了影响当代互联网发展的关键产品之一。在淘宝的体系中，支付宝有以下四大优势：

- 作为淘宝收款的第三方工具，在商品邮递期间暂时保管款项，只有消费者收到货并且满意后商家才能够拿到钱，搭配**用户14天自动收货、无物流信息课申请退款**等功能，完美解决商家和用户之间的信任问题
- 接入各大银行系统，无需再登录网银/银行转账/邮局汇款，**可直接在PC端使用密码支付**为进一步提升资金流效率，密码支付进化为**扫码支付**
- 有了信任，淘宝得以拉长现金流的时间差，反哺支付宝相关的金融业务
- 作为支付工具自带的流量在提升支付效率的同时，也提升了流量的效率

得益于支付宝的出现，线上资金流效率大幅提升，甚至在线下也几乎替代了现金支付（从支付工具的角度来看，支付宝可是微信的大前辈了）。

货的效率：定倍率

- **【劣势】受海外成熟电商体系影响，本土电商平台刚诞生时都是采取收费制，商家入驻平台、发布商品都需要缴费，成交金额还需要按比例上交平台，这是因为b2P2C过程中平台(P)需要盈利**

新世纪之初，国内最大的电商平台是美资eBay控制的易趣。易趣在全球都使用以上的方式、作为中间平台赚取差价。淘宝在2003年成立，刚一诞生就与易趣开始了长达3年的“商家争夺战”，淘宝为争取更多商家入驻平台，宣布执行“**平台费用全免**”政策，马云甚至严令淘宝在3年内不许盈利、只争商家和用户。全免政策在当时引起了巨大的轰动，并最终成为淘宝战胜易趣、独霸国内电商市场的关键因素。

淘宝的全免政策就是线上b2C模式的雏形，**虽然平台(P)处于小型商家(b)与消费者(C)之间，但商家可以选择支付额外的费用**，这就吸引了众多商家入驻。同时，在平台上陈列商品并不占据物理空间，因此不需要支付店铺租金，可以极大地降低商家的成本。由于**商家的成本变低，商品价格也随之降低**，消费者接踵而来。如此，淘宝通过全免政策，在短期之内形成了对商家、对用户都更有吸引力的格局。淘宝战胜易趣之后，几乎没有本土电商再按“入场费”方式对商家进行收费了，电商的盈利模式也发生了巨大的改变。

不难发现，淘宝的方式几乎奠定了整个中国电商模式的基础，在许多具体的产品设计、产品逻辑上，京东、国美、苏宁、唯品会等知名电商平台都大规模接入或参考了淘宝的交易系统。从03年建立、06年击退美资易趣、成为本土霸主、08年实现盈利的历程来看，阿里巴巴毫无疑问是一家了不起的互联网公司。

支付宝作用于资金流效率，全免政策降低了定倍率

结合互联网天然的信息流优势，淘宝形成了交易基本流畅的开放平台模式（C2C/b2C）

● 京东：站在巨人的肩膀上，解决巨人的问题

淘宝虽然为本土电商平台进行了基本的打样，但还是有许多难以克服的效率问题存在。例如——

场的效率：信息流

- 【劣势】商家与消费者之间存在信息差，假冒伪劣商品横行，商品质量堪忧
- 【劣势】线上购物缺失“体验感”，无法获得尺寸、颜色、版型等直观信息，掉坑几率较大

场的效率：物流

- 【劣势】平邮邮寄到货时间14天，快递邮寄3-7天
- 【劣势】快递服务质量不高，价格昂贵

存在信息差、缺乏体验感是电商天生的缺陷，这种情况降低了信息传递效率，削弱了消费者的实际利益，是假冒伪劣商品的温床。同时，又昂贵又慢的物流进一步增加了消费者下单时的心理障碍。在中国本土经济和社会条件下，任意环节上严重的低效都可能演变成巨大的问题——

信息流和物流上的劣势导致——

- 【劣势】用户对网络购物缺乏信任，电商一时成为低质、假货的代表
- 【劣势】整体环境不支持大件/精品件/特殊件交易，许多日常需求难以被满足
- 【劣势】低质小件的客单价难以突破，快递费看起来异常昂贵
- 【劣势】售后服务参差不齐，退货难，退货邮费有时候比商品还贵
- 更多复杂的用户心理……

对任何以C2C（b2C）为主流交易方式的电商平台而言，以上问题都会存在，哪怕是亚马逊也不能免俗。由于商品由商家提供，平台本身难以控制商品质量，因此亚马逊、淘宝都尝试使用一套增强用户

体验、增加用户反馈机制的系统来缓解相关的问题。例如，**用户评价系统、商家评分系统、按评价/评分的排序来确定商家在搜索引擎中的权重、与快递公司合作、给与邮费补贴等**。但道高一尺、魔高一丈，面对疯狂刷单、索要好评等行为，系统只能隔靴搔痒，无法解决根本问题。

2004年6月，京东在自己的3C光碟业务蒸蒸日上的时候，决定全面放弃线下、进军电商。在数年的探索中，京东依赖于自身的基因，以电商模式中存在的问题为突破口，进行了多种模式创新，并取得了斐然的成就——

- **京东自营：变C2C为P2C/B2C，平台升级为一个“大商家”，亲自把控进货渠道和商品质量**

在传统平台上，商品由商家提供，平台要通过商家控制商品质量成本高、难度大、效果差。京东一改大部分平台“先吸引商家，再用商家吸引用户”的循环，直接与工厂/品牌/高品质商家对接，筛选出高质量商品售卖给用户。**这种模式在京东成立初期就在采用，后来才被命名为“京东自营”**。这个方法有一石多鸟的效果——

- 对完全自营商品，将b2P2C流程缩减为P2C，降低定倍率，无需抽成即可获得商品售卖的收入
- 对半自营商品，引入知名商家的品牌为自己站台，扩大平台影响力
- 平台得以亲自把控商品质量和供应链的一部分，平均商品质量大幅提升
- 平台小商家(b)为与自营店铺竞争，会主动提升商品质量
- 大件/精品件交易成为可能，客单价大幅提升

- **京东仓储&京东物流：免邮+半日达，京东物流效率直逼线下购物**

作为一个“大商家”，京东决定亲自把握了供应链（进货、储存、配送）的一部分。自2007年起，京东开始在全国建立自己的仓储中心用于存货，并针对这些仓库的位置提供京东配送/合作配送服务，此举当时在国内引起了巨大的震动，行业对此非常不看好，但京东还是力排众议走上了发展物流和仓储的道路，今天京东的物流效率已是全球前列。通过将物流规范化、工业化，京东在物流上的效率得以全面提升：

- 免邮利益让渡给用户，自营店免邮，半自营店/非自营店满99免邮，用户的心理成本大幅降低
- 几乎形成了“全场免邮”噱头，平台竞争力大幅提升
- 京津冀、长三角、成渝等数十个主要城市实现“半日达”，“次日达”城市覆盖率也接近95%
- 售后服务+退货系统全面升级，提供上门取件/免退费等服务，大幅降低体验感缺失带来的心理损失

在淘宝的基础上，京东的模式增加了商品质量高、服务质量高、物流费用低、物流速度快这四大优势，这些优势很快就帮助京东占据了一块市场，让京东成为用户心中和淘宝功能不同、印象不同的电商平台。

自营政策提升产品质量与客单价，自建物流系统大幅提升了物流效率

不难发现，京东与淘宝在不同的效率上分别取得了重大突破，几乎已经将所有线上的效率劣势转化为了效率优势，这才得以在2015年前稳坐本土电商的前两把交椅。两个企业在竞争市场份额的同时，也会模仿对方的方式来提升自身平台上的效率，但往往很难颠覆从0到1时创造出的辉煌成果。例如——

- **京东金融**：其声势在支付宝与微信两个支付巨头的斗争中显得微不足道。大部分用户只知道京东白条、金条，对金东金融的其他业务不熟悉也不关心。
- **天猫商城与菜鸟网络**：天猫商城解决淘宝低质问题，通过平台为高品质商家/商品背书的方式引入品牌商家。菜鸟网络包括菜鸟裹裹、菜鸟驿站等5大品牌，提供次日达、上门取件、包邮免邮等服务，主要用于提升淘宝/天猫的物流效率。但京东“正品保证”、“次日达”的名声已经太深入人心，大部分用户根本不熟悉菜鸟裹裹，也无法辨别天猫商城与淘宝有什么区别。

可见，虽然可以将他人从0到1的模式复制或借鉴过来，但品牌的影响力并不能在一朝一夕之间建立。

2.3 淘宝vs京东：迥然不同的盈利逻辑

淘宝和京东有着表面相似、但实际完全不同的商业模式，在不同商业模式下，两大巨头的盈利模式、盈利状态自然也有巨大的差异。平台的盈利模式、以及是否能够盈利本身很大程度上取决于平台在整个交易关系中的地位，我们来看看——

电商模式



在电商的交易环境中，存在三个关键的角色：消费者，商家与平台。三个关键角色的目标不尽相同，甚至有些矛盾：

- 消费者的目标：以尽量低的金钱/时间成本满足对物品/服务的需求
- 商家的目标：以尽量高的利润（低成本高价格）出售尽量多的商品
- 平台的目标：帮助商家和消费者达成目标，并在交易中找出盈利模式

不难发现，消费者渴望的“低价”与商家渴望的“高价”相矛盾，平台渴望从商家/消费者处收费，但商家/消费者都渴望尽量低的成本。因此，消费者、商家、平台三大角色之间是存在制衡与博弈关系的。在这个博弈关系中，占据主导地位就能获得最大的利益，相反，就要被其他角色裹挟着付出更高

的成本。从企业的角度来看，电商平台要盈利，就必须在三方博弈中占据主导地位。那谁是最有话语权的角色呢？在不同的商业模式中，话语权属于不同的角色，我们来看看——

淘宝开放平台模式中，平台最有话语权

京东价值链整合模式中，用户最有话语权

• 淘宝：平台>用户>>>>商家

在b2C/C2C模式中，平台相当于中介，实际的交易发生在商家和消费者之间。在这个过程中，几乎所有线上的效率优势（信息、支付、物流、转化、定倍率）都是来自于平台提供的功能和服务，因此商家想要享受电商行业的红利、想要从线上更高的交易效率中获益，就必须与平台合作。

以当今本土电商的环境来看，大部分个人商家(C)或小型商家(b)更换平台的成本很高，京东、国美、苏宁、亚马逊等商城对商家入驻条件较为苛刻，唯品会、Shopee、毒等商城对商品品类有严格要求，拼多多是C2B模式无法轻易入驻、闲鱼没流量，大部分的小型商家只能在淘宝上求生存。因此商家在平台面前是较为弱势的，商家必须顺从平台的规则才能进行交易，能做出的最大反抗也就是退出淘宝、不再做电商。

同时，商家又面临需要从消费者处获取现金、需要迎合消费者需求的状况。对于消费者来说，切换到淘宝上另一个店铺、或者切换到其他平台进行购物，都是轻而易举、几乎没有成本的行为。所以商家一面要顺从平台的规则、承担平台的成本，一面要哄住用户，是淘宝的模式中话语权最弱的角色。

淘宝的文化本身是“重用户”多过“重商家”。商家可能是吸引用户的关键，但用户才是具有真正消费需求的对象。淘宝相信，要让解决方案跟着需求走（商家来找用户），而不是让需求自己来找解决方案（用户去找商家）。在很长一段时间内，淘宝都专注于提升用户粘性，设计了众多的功能让用户在APP中逛逛逛（这也是为什么内容电商、直播电商会从淘系先开始的根本原因）。在这种文化熏陶下，淘宝上商家就更加弱势了。

商家的弱势、平台的强势构成了淘宝这样C2C/b2C平台能够盈利的关键点。帮助商家提升“人”的效率、享受“场”的效率是淘宝的主要收入来源：

- **【核心收入来源】广告竞价服务：**帮助商家获取流量。在互联网“信息流”优势下，用户只需搜索关键字便可轻易获得品质相似、价格相似的数以万计的商品信息，商家想要在如此多的信息中脱颖而出，就必须购买淘宝提供的广告竞价服务。
 - **直通车：**关键词竞价，用户搜索关键词时，对该关键词出价越高的卖家越优先出现
 - **钻石展位：**在首页/搜索页等页面放置展位，商家可以设置图片广告吸引点击
 - **聚划算：**付费参与活动，拿到“热卖标”、提升关键词排名、配合平台大促一起爆发
- **卖家店铺页面/商品页面设计：**帮助商家提升转化。淘宝提供专业创作团队设计的店铺和商品界面模板，提供专业技术团队开发的拓展服务（如直播带货），这些模板和服务需要付费使用。

- **商家店铺抽成(佣金)及加盟保证金**：按照销售额对店铺抽成，保证金无利息且永不退还（除非关店）
 - **天猫商家**：销售额抽成5.5%，保证金5w-15w+平台年费6w元
 - **淘宝个体户**：销售额不抽成，保证金1000+旺铺使用费
- **支付宝相关收入**：平台使用支付宝所产生的相关收入，严格来说不算淘宝、而算是支付宝的收入。
 - 商家收款时需要支付通道使用费，和第三方支付手段相似，收取一定百分比的手续费
 - 消费者付款与商家收款之间有较大时间差，支付宝可在短时间内囤积大量现金流，利用现金流对外借贷或投资、实现金融方面大规模盈利
 - 收取的店铺保证金也可以作为现金流的部分，对外借贷或投资实现盈利

你发现了吗？**几乎所有的收入都是对商家进行的收费**，这与我们之前分析的、商家在平台模式中处于绝对弱势地位有很强的联系。参考淘宝的模式，几乎所有开放平台电商的盈利值得一提的是，淘宝还提供淘宝大学提供电子商务在线培训服务，卖家可以付费接受培训，学习如何借助淘宝的一系列工具将店铺交易额带飞，相当于让商家付费学习如何更好地为淘宝付费（当然了，也是学习如何更好地为自己赚钱）。

- **京东：用户>平台≈商家**

在P2C/B2C模式中，京东本身就是一个非常大的商城（相比自营店，京东平台上的入驻商家并不是很多）。因此不同于淘宝模式中“平台创造效率、商家哄住用户”的环境，**京东要自己创造效率优势，还要自己哄住用户**。虽然京东的产品质量和服务效率能够极大程度地把握住用户，但在其他电商平台也在飞速发展的前提下，用户跳转到其他平台进行购物的成本还是相对较小，因此用户在京东的模式中话语权要强过平台。

面对这种情况，京东不得不付出较多的成本来提升服务质量和产品质量，深入培养用户对品牌的忠诚，增强用户粘性。京东上的商家虽然也依赖于平台创造的效率完成交易、也要分担平台的成本，但由于京东自己作为“大商家”有较多销售额，因此不会在商家价值上榨取得那么深入（或许他们应该榨取得更深入），因此商家在京东的环境中话语权不是太弱。但如果京东想要加强对商家价值的挖掘，那商家也会变成整个平台上最弱势的存在。在这种环境下，京东的主要收入来源有：

- **【核心收入来源】自营商品的销售额**：京东通过从供应商、采购商进货，在自身平台上交易，争取差价、覆盖自身成本来创造利润。为扩大自身影响力，京东在很长一段时期内都采取薄利多销方式增加用户粘性，近年来京东价格逐渐回升，利润空间逐渐变大。
- **【核心收入来源】与淘宝相似的广告竞价服务、流量推广服务**：主要是焦点图售卖、站内付费推广、站内团购机会等。除了向商家开放，京东的流量推广服务还向其他门户网站、线下商店、线下活动开放，更倾向于通用广告。通用广告曾一度成为京东核心的收入来源之一。
- **店铺抽成及加盟保证金**：京东店铺几乎只接受企业入驻，且必须是自带影响力的品牌。入驻京东一般需要支付3w-10w之间的保证金，平台年费6000-1w2，以及3-10%的交易佣金。考虑到这些店铺都是自带影响力和流量的品牌，低保证金、高交易佣金是合理的选择。
- **京东金融相关收入**：与支付宝类似，卖家提现需支付手续费，另外京东依赖白条、以及京东金融的现金代管服务争取到消费者付费与商家收款之间的时间差，可以产生外贷和投资收益。

可以看到，**京东同时从用户和商家处获得收入**，整体的盈利策略会更温和，消费者和商家的利益都一定程度上得到了保障，是一种多方共赢的局面。在京东文化熏陶下，京东对于员工的薪资也从不吝啬，这一切看起来非常和谐且多元化的收入却不能让京东成为国内电商霸主。

在开放平台模式下，阿里巴巴可以获得利润以及市值

在价值链整合模式中，京东可以获得营收与现金

仁不带兵，义不行贾，对用户和商家过于温和的电商平台自然无法在盈利上有很好的保障。我们可以对比一下阿里巴巴与京东在过去三年的年营业收入、盈利以及利润率：

• 年营收

	A	B	C
1		京东	阿里巴巴
2	2018	4620亿	3452亿
3	2019	5769亿	3594亿
4	2020	7458亿	5097亿

根据数据统计，京东2020年活跃用户大约是淘宝活跃用户的1/3，然而：

- 京东的用户消费力更强、客单价更高
- 京东的商品价格更贵，用户在京东平台容易接受更高的价格
- 京东P2C/B2C业务可以直接从售卖商品获得收入、京东对商家的佣金也普遍高于阿里巴巴

因此京东的年营收是持续高于阿里巴巴的，这与京东的模式一脉相承。同时，由于京东可以直接从商品售卖业务中获得收入，因此京东有固定的现金流收入，其一直很健康，以至于让股东们对亏损忍耐了多年。在2018、2019年时，互联网上充斥着关于“为什么京东一直在亏损，而刘强东夫妇还如此富有”、“京东是世界上最有钱的亏损公司”的讨论，这都是京东在现金流占据优势引发的话题。

• 年盈利

	A	B	C
1		京东	阿里巴巴
2	2018	亏损25亿	614.12亿
3	2019	121亿	934亿
4	2020	494.05亿	1404亿

• 利润率

	A	B	C
1		京东	阿里巴巴
2	2018	-0.54%	17.79%
3	2019	2.10%	25.99%
4	2020	6.62%	27.55%

然而利润上，阿里巴巴就全面占优了，其理由也很容易理解：

- 在淘宝的模式中，营业成本主要集中在为商家提供服务、维持平台稳定这些方面，实际服务用户、给用户让利、与用户博弈的都是商家，因此淘宝在实际交易中承担的成本较低。
- 在价值链整合模式中，京东为了保障好的服务和体验，必须对仓储、物流、售后服务等环节付出高昂的成本，包括各环节的人力、基础设施、管理、服务，而且京东还必须注意这些成本不能过多地转移到商品价格上，否则会极大程度影响平台在用户心中的竞争力。

不难看出，不同的商业模式会反馈到企业的财务状况中，看似很赚钱的淘宝其实不从用户处赚钱、质量很高的京东也不能理所当然地争取到更多的用户。淘宝与京东的比拼为我们展示了藏在“人货场”分析框架背后复杂且深入的电商业务逻辑，也为我们后续深入理解其他电商平台有巨大的帮助。

• 纵横三十年：本土电商行业的风云变幻

◦ 90s - 2003：萌芽阶段遭遇剧变

- 上世纪90年代，互联网交易破土萌芽，网络零售开始起步
1999年，国内首家C2C电商8488正式成立，**同年易趣网、当当网成立**
首批本土电商平台如雨后春笋般冒出，开始进入市场
2000-2003年，互联网泡沫破灭，全球互联网业务遭受重创
海外巨头趁机抢占市场，大批电商平台速生速死
8488没落，eBay收购易趣，亚马逊入驻中国，首批本土电商几乎全灭

◦ 2003 - 2007：外资水土不服，本土企业艰难探索

- 2002年底非典爆发，随即**淘宝网成立**，一年后京东进军电商领域
自诞生起，淘宝与易趣展开了长达2年半的流量战争
放弃盈利、只争用户的免费策略+**支付宝**为淘宝带来了最终的胜利
彼时电商盈利模式尚不明确

◦ 2008 - 2014：巨头竞争格局初显，中国电商行业爆发式增长

- 2008年我国成为全球网民最多的国家，2009年网购人数突破1亿
物流行业飞速发展，奥运为国家基础设施建设格局带来改变
电商巨头逐渐摸索出各自的盈利模式，**双11活动**正式诞生
线下零售业受到严峻挑战
传统线下巨头开始发展线上业务，苏宁、国美开始发起激烈价格战
2014年底，中国电商交易体量已全面超越欧盟、日本

◦ 2015 - 2017：互联网流量危机爆发，电商市场由蓝海变红海

- 流量成本疯狂增长，店铺增速远高于新用户增速
竞争逐渐结束，淘宝京东坐稳行业头名，行业整体增长速度放缓
有流量即有收入的蛮荒时代过去，巨头开始**重用推荐系统**与"猜你喜欢"
大批个人商家难以盈利，退出市场

◦ **2018至今：电商新模式遍地开花，即将迎来更多新变化**

- **拼多多**杀入市场，“百亿补贴”又一次引发电商行业的流量大战
短视频迅速崛起，**抖音/快手**入场电商，直播电商成时代宠儿
电商垂直领域快速发展，社交电商、内容电商、跨境电商成新风口

2.4 订单异常检测：平台的抓手，商家的帮手

在分析电商业务逻辑的过程中，我们已经发现平台、商家、用户的利益诉求并不相同，并且三种角色存在着利益上的博弈关系。从宏观的角度来说，平台/用户可以利用在博弈关系中取得的优势达成自己的目的、商家并无太多反抗之力，但从微观的角度来看，每个商家、每个用户都可能做出扩大自己的利益的行为。

在千万商品、过亿用户的电商平台上，商家和用户的某些行为可能严重破坏平台环境、拉低线上交易效率，因此平台必须要对违规行为进行管控。幸运的是，无论商家和用户采用何种行为令自己获利，最终这些行为都会体现在交易信息上，因此**平台管控商家/用户异常行为的工具就是——订单异常检测系统**。

在不了解电商业务的状况下，大部分人很难理解订单究竟为什么会“异常”。在复杂的电商业务状况下，大量情况都可能是异常订单——

• **A类：明显违背平台效率追求、不遵守平台规则、阻碍正常交易的订单**

- 商家大量刷单（阻碍信息流效率）
- 虚假宣传，被多次投诉/差评的商品相关的订单（阻碍信息流效率）
- 付款后不发货、多次无视用户发货请求的订单（阻碍物流效率）
- 显示发货、但实际并不发出货品、拖延至确认收货日期的订单（阻碍物流效率）

这类订单基本都是商家主动发起，平台会对商家行为进行监控，若多次发现该类异常订单则可能降低商家的权重、实施惩罚，从而影响商家在搜索引擎中的排名，进而严重影响流量。

• **B类：恶意的、滥用买家权益、非真实购买意愿的订单**

- 因商家之间的纠纷（盗图、知识产权）被报复拍下
- 因用户与商家的纠纷被报复拍下
- 同行出于攻击/刷差评目的拍下
- 锁单后一直不付款
- 付款后立即退款、刚发货立即退款

- 满减凑单后申请部分退款
- 打假订单、以广告法、食品法、版权为由的敲诈订单
- 退货后发现货品被调包、或退回商品件数与实际下单数量不符

这类订单基本都是商家对其他商家、或用户发起。平台若检测到该类订单，会对发起恶意攻击的用户或商家进行降权，用户会被限制参与某些满减或营销活动，重者可能面临封号风险。这类检测可以保护商家不受恶意用户和商家侵扰。

• C类：物流/交易过程中出现问题，无法完成正常交易流程

- 平台标记有恶意购物嫌疑的顾客下单，不能走正常发货流程
- 设错价格/设错优惠被大量拍下，库存不足，不能走正常发货流程
- 物流单号被重复使用
- 发货商品无走件记录、货物丢失
- 发货商品因物流过程损坏
- 退货物流无走件记录、货物丢失
- 退回商品与退货订单商品不一致
- 退货物流地址有误、非店铺退货地址
- 用户提出超出正常售后范围的诉求（不接受退款、换货），不能走正常售后流程

这类订单是正常交易流程中出现的意外状况，需要平台出面作为第三方进行裁决，由商家/用户分担责任，解决纠纷。

另外还可能存在：

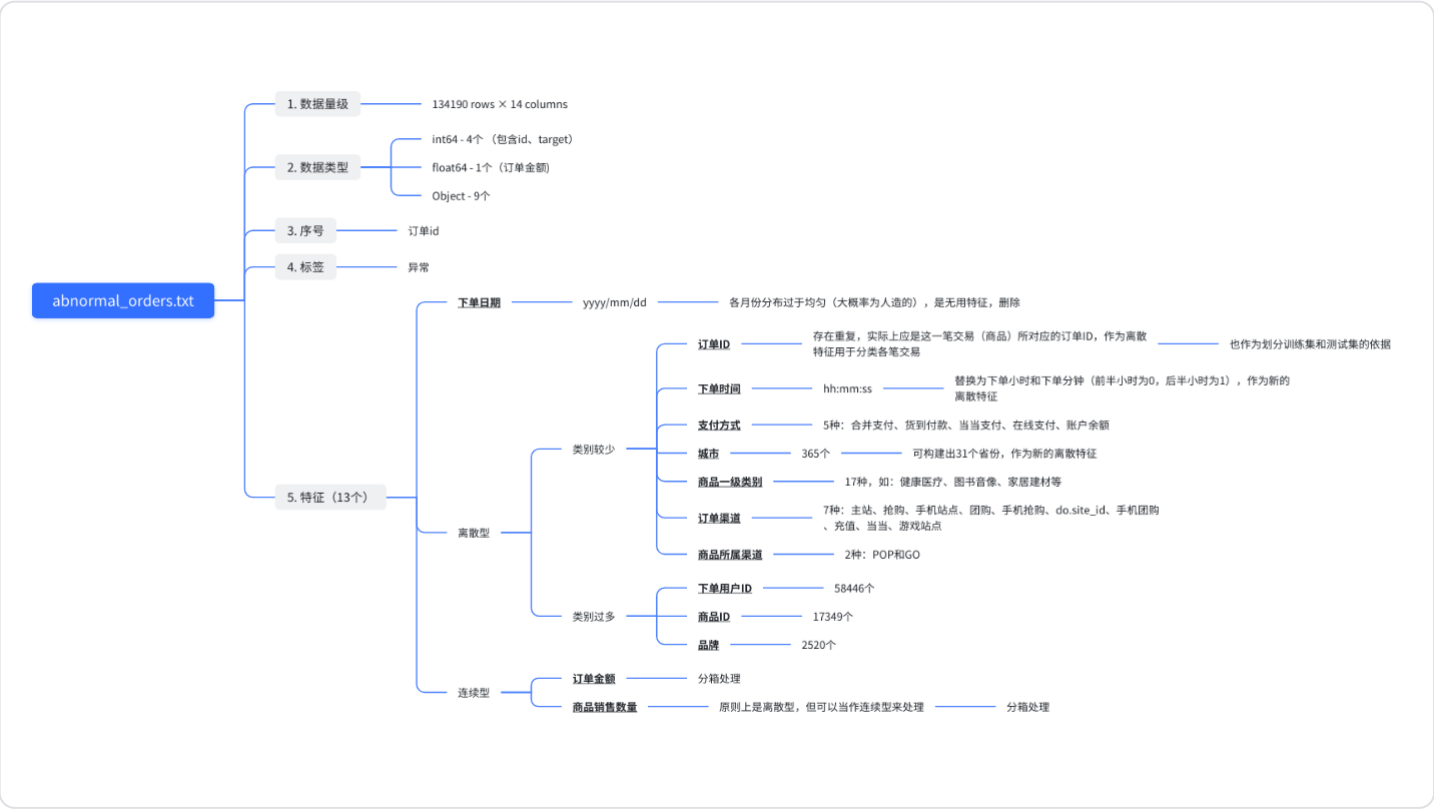
• 明显违背国家法律法规、违反道德约束的订单

- 违法违禁物品/内容的传播
- VPN的售卖
- 珍稀动物售卖、宠物盲盒等

该类订单一旦被平台检测到，会立刻被下架并给与警告，多次无视警告则直接封号处理，商家还可能承担相应法律责任。

对电商平台而言，每日检测异常订单能够维护平台交易环境，保护商家和用户处于相对公平的竞争环境中。对商家而言，异常订单检测系统能够快速筛选出非真实购买意愿、以及无法正常完成交易流程的订单，提升订单的处理效率。当然了，**鉴于电商“交易”的自然属性，在明确阻碍交易流程的事件发生之前，即便订单被系统认为是异常，商家可能也不会主动叫停交易。**因此异常检测系统更多是监督商家和用户行为、并针对高风险订单预警。

3. 数据介绍



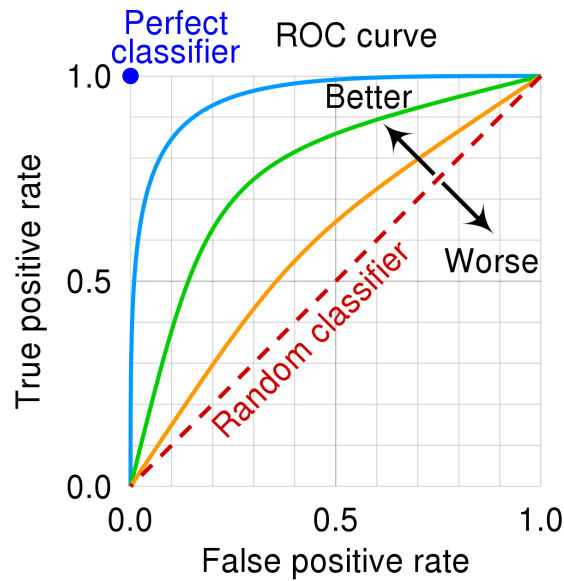
4. 评估方法

4.1 评估指标 (Scoring Metric)

- **AUC (Area under the Curve of ROC)** , 即ROC曲线下方的面积, 用于评估模型的分类能力。
 - 因为是在1x1的方格里求面积(TPR和FPR的范围都在0到1), AUC必在0~1之间。
 - **AUC值越大的分类器, 正确率越高。**

	A	B	C
1	Confusion Matrix	Positive Class	Negative Class
2	Positive Prediction	True Positive (TP)	False Positive (FP)
3	Negative Prediction	False Negative (FN)	True Negative (TN)

- $TPR = TP / (TP + FN)$, 代表所有正例中, 预测为正例的比例, 越大越好
- $FPR = FP / (FP + TN)$, 代表所有负例中, 预测为正例的比例, 越小越好



5. 项目代码

基于集成算法融合模型的国美电商订单异常检测



1. 粗粒度数据探索：建立benchmark