

**内蒙古师范大学计算机科学技术学院**

**毕业设计（论文）开题报告**

**题 目： 京东电子商品数据的分析与新商品销售量的预测**

**专 业 信息管理与信息系统（云计算大数据分析）**

**学 生 陈磊**

**学 号 20161104515**

**指导教师 梁政欣 苗忠义**

**日 期 2019/11/24**

**计算机科学技术学院制**

**说 明**

1.选题目的及意义

在很久之前零售业是一个地区乃至整个国家的最主要的购物方式。它是一个地区或国家经济运行状况的晴雨表，是一个国家最主要最古老的行业之一。当下，我国零售业已经开启了一个全新时代，标志着我国零售业的一个全新起点。零售业至今已有160多年的发展史，可以追溯到19世纪50年代。发展的进程中，零售业不断创新升级，新兴零售形式也不断地出现。近年来，随着我国Internet技术不断发展，网络零售这一新兴行业呈现良好的发展态势，我国加入WTO后，网购已是现代购物中不可或缺的方式之一。当前，通过网络可以实现的商务功能越来越多，如基本的信息发布功能、对外沟通展示功能、网上采购功能、商品展示功能、网上交易功能等。几乎传统购物功能都可以在互联网上得以实现，并且高效运作。

现在看来，把网上交易的，不管是企业还是消费者越来越喜欢如此选择，把网络购物当作是一种时尚。如今对于各种节日的广告效应，人们选择在方便快捷的网购的同时进行了对应的选购策略，不得不说是这个时代造就了人们的消费心理。因此众多的商品选择就成为了一种需要去研究的问题。当我们收集了这些信息之后，对它进行分析和预测，就能够了解大众的消费心理和选购趋势，对于商品的选购模式是怎样的。

2.国内外在该方向的研究现状及分析

厦门举行的2019全球电子商务大会上，《中国电子商务发展报告2018-2019》对外发布。该报告指出，2018年中国电子商务交易总额31.63万亿元人民币，同比增长8.5%。

该报告从现状规模、发展特点、趋势走向等各方面，多维度对中国电子商务的发展进行了研究和分析。该报告是反映2018年中国电子商务发展状况的综合性报告，从中可全面了解到中国电子商务的发展水平、发展特点和发展趋势。从《中国电子商务发展报告2018-2019》报告可见，中国电子商务规模持续扩大，自2016年开始从超高速增长期进入到相对稳定的发展期，但2019年上半年，实物商品网上零售额同比增长仍高达21.6%，电子商务继续承担国民经济发展的强大源动力。

该报告指出，电子商务在壮大数字经济、共建“一带一路”、助力乡村振兴、带动创新创业等方面均发挥了积极作用。中国正把握全球电子商务发展的历史性机遇，结合中国实际创新发展，不断完善电子商务发展环境，提升开放发展水平，加快传统企业利用跨境电商转型升级；培育和发展电子商务新领域与新优势，积极扩大电子商务产业和市场规模，推动电子商务企业积极融入国际市场。当天举办的2019全球电子商务大会上，电子商务企业、跨境电商寄递企业、产业互联网平台、行业组织等业界人士近1200人参会，就“全球视野下的电商发展新机遇”“人工智能、大数据在电子商务的实践和应用”等议题进行研讨。

互联网和电子商务加快全球市场效率提升，商品信息更加对称，贸易门槛逐步降低。跨境电商给全球贸易格局带来新的变化，世界各国、各地区的经济联系愈发紧密。数字经济、人工智能对人类社会和经济产生深刻影响。全球电子商务市场在过去十多年中快速增长，并且这种势头仍在持续。当前，全球网民人数已达41.57亿人，互联网普及率达54.4%，亚洲网民数占全球网民数的比重最高，达48.7%；全球已有七个国家网购用户数量过亿，中国是全球最大的互联网用户市场，网民规模达7.72亿人，普及率达到55.8%；从网购人数增长区域来看，未来几年增长最快的地区将是中东和非洲。

**3．主要研究内容**

（1）收集数据

尽量获得完整、真实、准确的数据，做好数据的预处理工作，是网站数据分析之前要做的重要工作。

网站后台的数据：网站的注册用户数据（注册时间、用户性别、所属地域、来访次数、停留时间等）、订单数据（下单时间、订单数量、商品品类、订单金额、订购频次等）反馈数据（客户评价、退货换货、客户投诉等）。

（2）量化分析

分析的目的是从表面的数据找到问题的本质，然后针对特定的主题进行归纳和总结。常用的分析方法有以下几种：

趋势分析：确定变化趋势和变化规律。趋势分析是将实际达到的结果与不同时期同类指标的历史数据进行比较，从而确定变化规律的一种分析方法。具体的分析方法包括定比和环比两种方法，定比是某一时期数据为基数，把其他各期数据均与该期的基数进行比较；而环比是分别以上一时期数据为基数，把下一时期与上一时期的基数进行比较。

对比分析：把两个相互联系的指标数据进行比较，从数量上展示和说明对象规模的大小、水平的高低、速度的快慢，以及各种关系是否协调。再对比分析中，选择合适的对比标准是十分关键的步骤，这样才能做出客观的评价；若选择不合适，可能得出错误的结论。

关联分析：通过一个事物来预测另一个事物。如果两个或多个事物之间存在一定的关联，那么其中一个事物就能通过其他事物进行预测；他的目的是挖掘隐藏在数据间的相互关系。

因果关系：按事物之间的因果关系，知因测果或倒果查因。因果分析是为了确定引起某一现象变化原因的分析，主要解决“为什么”的问题；因果分析就是在研究对象的先行情况中，把作为它的原因的现象与其他非原因现象区别开来，或者是对研究对象的后行情况中，把作为它的结果的现象与其他的现象区别开来。

**4. 研究方案**

* 文献研究法：本课题前期工作主要是采取文献调研的方法，充分利用图书馆资源和网络资源，对有关期刊、相关专著等学术资料和最新的数据进行查找、搜集、阅读、分析寻找获得课题科学回答的线索使得所研究的内容理论更为准确，为研究提供坚实可靠的文献基础。
* 电子问卷调查法：本课题拟对人们日常采购的行为趋向进行调查，获得第一手数据和资料并进行分析整理。
* 实时分析法：对整个数据的真实性进行调查和深入剖析。
* 比较研究法：本课题拟采用比较法，对不同人们的消费趋向进行横向比较，同时对同一行业竞争网站进行纵向对比，从中寻找可以借鉴的经验。

**5．进度安排，预期达到的目标**

2019年11月16日——11月20日 确定选题、收集相关资料

2019年11月21日——11月25日 撰写开题报告与开题

2019年12月1日——12月31日 收集资料，开展研究，形成写作提纲

2020年1月1日——2月29日深入研究，形成论文初稿

2020年4月1日——5月1日 论文修改、定稿、打印、答辩

**6．课题已具备和所需的条件、经费**

本人现在的专业就是大数据分析，再加上这个时代给予了我们互联网思维，不论是信息的搜集还是处理方式出现的问题，都能够有足够的保障和解决措施，我对电子商务的数据分析有更深层次的认识并有了自己独到的见解，在图书馆的大量和文献资料充实了我的大脑，使我的研究更加方便。

本人由于能力有限，对京东商品数据的现状有待更深一步的探索和研究学习，更多的分析项目和分析思维有待我去亲身经历和提高。

**7．研究过程中可能遇到的困难和问题，解决的措施**

对于很多的专业性的分析方法还是欠缺专业知识，其他时间多进行研读和学习。可能抓取数据的同时会出现很多行不通的情况，但是还是需要学更多知识进行破解和渗透。对于以下两种编程语言（R、Python），操作性分析的软件（SPSS）的熟练度还需要进行提升。

**8．主要参考文献**

[1]房家梁.《中国电子商务发展报告2018-2019》在厦门发布. 中国新闻网. [2019.9.8]. <http://www.chinanews.com/cj/2019/09-08/8951031.shtml>

[2] [美] robert i. kabacoff. 《R语言实战(第2版）》.人民邮电出版社

[3] [美] ryan Mitchell.《Python网络数据采集》. 人民邮电出版社

[4] 武松.潘发明.《SPSS统计分析大全》. 清华大学出版社