**基于情感分析的商品评价模型**

**Abstract**

With the rapid development of the information age, the Internet and modern information technology are increasingly developed. The network has become the most convenient platform for data sharing, which also marks the arrival of a new era - the era of big data. In the era of big data, everyone is the creator and user of network data. Through big data, we can easily describe everyone's behavior habits and reflect the laws of things. E-commerce is the product of big data era. Founded in 1995 in Seattle, Washington, Amazon is the largest online e-commerce company in the United States. In the online marketplace it created, Amazon provides customers with an opportunity to rate and review purchases. In order to use these data to gain insights into the markets in which they participate, the timing of that participation, and the potential success of product design feature choices，it’s very meaningful to build a model to analyze user comments and ratings.

firsly, 为了帮助阳光公司提供的三个产品在亚马逊商城的销售取得成功，本文利用机器学习中的LightGBM算法建立模型, 通过自然语言情感分析技术将评论文字映射到数值空间, 同时对预处理之后的数据进行特征工程, 以帮助模型更好的分析数据, 建立模型后通过模型评价得到模型的F1值均在0.9左右, 表明模型建立的非常成功, 同时我们得出了数据中各指标的重要程度, 最终发现特征review\_body(评论), rate(好评率), review\_date(评论日期)和total\_votes(总投票数)对一个产品的评级影响很大, 故公司应注重相应的数据信息来成功运营产品.

secondly, 针对第二问的a小问, 在第一问时建立的模型就已经综合考虑了各种相关数据, 一旦阳光公司的产品开始销售, 根据反馈的评价信息, 将相关数据输入到模型中得到预测评分, 再根据预测评分和真实评分来分析数据得到相关结论. 对于b小问, 本文先通过熵权法确定了综合评级和评论的评价score结果来描述产品的声誉, 并通过月份销售量, 好评率, 差评率的统计数据绘制统计图进行基于时间模式的分析, 发现xxxxxxx; 由于前边两个模型都结合了相关的文本和评级的度量, 故c小问通过模糊综合评价方法利用问题一模型结论得出相关指标和b小问得出的score进行综合评判得到相应指标的权重, 利用权重计算综合得分, 评判一个产品的潜在成功与否; d小问要求分析特定的评级是否会引发某种类型的评论, 通过计算二者的斯皮尔曼相关系数和实际数据发现, 特定的评级会引发某种类型的评论; 最后, 通过构建特定描述符词典, 统计相关描述符的词频, 并利用热力图和词云可视化分析得出, “never”等否定词汇与一星级密切相关, 而”love”,”like”,”great”等积极词汇常出现在4星或5星评级中.

finally, 本文将以上数据挖掘和分析结果汇总成报告, 以书信的方式向阳光公司的市场总监提供, 真诚的希望本次结果能够为公司提供最好的帮助.