《基于pyspark的社交平台情感分析系统》测试报告

**一、测试目标**

验证系统模块的覆盖范围、数据处理能力、模型性能及交互功能，确保其能够有效支撑社交平台情感分析的核心服务功能。测试重点包括数据多样性处理、模型训练稳定性、情感分析准确性，以及图形化界面交互体验。

**二、测试员及测试环境**

测试员：黄少伟、许世烜

硬件环境：分布式服务器集群（2节点），存储容量1TB，带宽1Gbps。

软件环境：基于Python的Scrapy爬虫框架、NLP情感分析模型（支持向量机与LSTM结合）。

**三、数据采集**

使用包含微博评论及实时广场信息的综合数据集，数据来源于微博评论以及实时广场。该数据集涵盖“美食分享”、“旅游推荐”、“娱乐八卦”以及“生活日常”的标签，全面反映了微博用户多样化的话题讨论。数据集总量为7.8万条记录，按70%、15%、15%的比例划分为训练集、验证集和测试集，以确保模型训练、验证和测试的有效性和准确性。存储格式为CSV，这种格式便于数据的读取、存储和处理，能够满足后续数据分析和模型训练的需求。

**四、测试内容与结果分析**

**1.多源数据采集覆盖测试**

**测试员：许世烜**

**测试方法**：收集关键词为“奥运会”，在12小时内采集多条微博评论数据。

结果：

**覆盖范围**：成功抓取社交媒体(微博)评论78805万条。

**验证结论**：系统支持多维度、多格式数据抓取，满足舆情监测的全面性要求。

**2.数据清洗与预处理测试**

**测试员：许世烜**

**测试方法：**对原始数据进行数据处理，包括读取数据（每一行数据以' #### '为分隔符，拆分为标签和文本内容两部分）、文本清洗（保留文本中的中文字符，过滤掉其他非中文字符）、分词与去除停用词、特征工程（对标签进行 one - hot 编码，标准化数值特征）以及数据拆分、文本数字化与序列填充。

**结果：**

数据清洗效果：成功去除数据中的非中文字符、停用词，将文本数据转化为便于模型处理的格式，有效减少了噪声数据对模型训练的干扰。经清洗后，文本数据更加简洁、规范，。

特征工程成果：通过Tokenizer对文本进行编码，将文本数据转化为数字序列，并进行了序列填充，使所有文本序列长度一致，便于模型处理。同时，对标签数据进行了one - hot编码，将分类标签转化为适合模型输入的向量形式，有助于模型更好地学习和识别数据中的模式和特征。

数据拆分准确性：训练集、验证集和测试集的记录数量分别为78805条、11821条和11821条，符合70%、15%、15%的预期比例。

验证结论：数据预处理模块能够有效清理和转换原始数据，为后续模型训练提供高质量数据，为基于LSTM的情感分析模型的训练和准确预测奠定了坚实基础，确保了模型能够从处理后的数据中学习到有效的特征和模式，进而实现对微博文本情感的准确分类。

**3.模型训练测试**

**测试员：黄少伟**

**测试方法：**利用model\_train.py脚本，采用随机森林分类器进行模型训练，设置超参数网格并通过交叉验证进行调优。记录训练时间、最佳超参数以及模型在验证集上的评估指标。

**结果：**

训练时间：完成模型训练（含超参数调优）耗时约 15 分钟。

最佳超参数：Best Epoch: 2，Best Dev Accuracy: 0.9067。

测试集评估指标：AUC 为 0.9076，Precision (Macro)为 0.9078，精确率为 0.87，召回率为0.8941，F1 分数为 0.9000。

**验证结论：**模型训练过程稳定，超参数调优有效提升模型性能，在验证集和测试集上达到较好的预测效果。

**4.文本情感分析测试**

**测试员：黄少伟**

**测试方法：**基于模型训练后的结果，在图形化界面中，输入不同情感倾向的测试文本，点击“开始”按钮，检查界面显示的预测结果是否正确。同时，测试“清空”按钮的功能，确保点击后文本输入框和预测结果显示区域能正确清空。

结果：在图形化界面的测试中，输入的测试文本大部分能得到正确的预测结果，“开始”按钮和“清空”按钮的功能均正常工作，界面响应及时，没有出现卡顿或报错现象。

**验证结论：**文本情感分类功能在测试中表现良好，模型能够对微博文本的情感倾向进行较为准确的分类，各项评估指标均达到了一定的水平，说明模型具有较好的泛化能力。图形化界面的设计和功能也符合预期，能够方便用户进行文本情感预测操作。

**五、测试总结与建议**

通过本次测试，系统模块在数据处理、模型训练与评估、图形化界面交互等方面均达到设计要求，能够为社交平台情感分析提供可靠的支撑。

**1.优势总结：**

**数据处理高效合理：**系统在数据处理阶段，具备完善的文本清洗机制，有效剔除了非中文字符、去除停用词。在数据集划分上将数据按合理的比例（训练集 70%、验证集 15%、测试集 15%）进行划分，为模型的训练、验证和测试提供了合适的数据样本，有助于提升模型的泛化能力。同时，对标签数据进行 one - hot 编码，以及对文本数据进行数字化和序列填充。

**模型性能表现良好：**选用 LSTM 模型进行情感分析。从模型在验证集上的评估指标来看，准确率、精确率、召回率和 F1 值都达到了一定的水平，表明模型对微博文本情感分类具有较高的准确性和可靠性，能够较好地识别文本的情感倾向。

**界面交互简单易用：**基于tkinter搭建的图形化界面。用户只需在文本框中输入待分析的文本，点击 “开始” 按钮即可快速获得情感分析结果，点击 “清空” 按钮能够方便地清除输入和输出内容。这种简单的交互设计，方便在社交平台场景下快速进行情感分析。

**2.优化建议：**

**数据层面：**当前数据仅来源于单一文件，数据量和多样性可能有限。可以考虑扩大数据收集范围，增加数据的多样性和规模。此外，还可以进一步优化数据预处理流程，比如尝试更先进的中文分词算法或自定义停用词表，以提高文本处理的精度。

**模型层面：**可以尝试在模型结构上进行改进，例如增加更多的隐藏层，或者尝试使用其他先进的神经网络结构（如 GRU、Transformer 等），对比不同结构在情感分析任务上的效果，选择性能更优的模型。

**功能拓展层面：**当前系统仅能判断文本的积极或消极情感，功能较为单一。可以考虑拓展功能，比如增加情感强度分析，不仅判断情感的正负，还能评估情感的强烈程度；或者增加多分类功能，识别出更多的情感类别。同时，可以考虑与其他自然语言处理任务相结合，为社交平台用户提供更全面的文本分析服务。

**系统性能层面：**在处理大规模数据或高并发请求时，当前系统的性能可能会受到影响。可以考虑对代码进行优化；优化数据加载和预处理流程，提高数据处理效率；在图形化界面方面，优化界面布局和响应机制，提升用户操作的流畅性。