基于机器学习的中文情感识别系统是对大量文本数据进行分析，将文本数据与情绪类别进行关联（正性、中性、负性），在关联情感类型后再与地区关联，最后对各种数据分析的结果进行可视化，供使用人员针对具体问题分析并采取必要措施。该系统具有前端和后端，前端给用户提交数据集以及展示数据可视化结果，后端的情感分析主要分为如下流程：

1. 文本获取：通过爬虫爬取数据或使用用户提交的csv等格式的语料数据。
2. 文本预处理：对收集的数据进行预处理，包括去除无关字符、分词、去除停用词等。
3. 数据标注：对预处理后的数据进行标注，标注的内容一般是情感极性，如正性、中性、负性。python
4. 特征提取：从预处理后的数据中提取出有用的特征。比如词频、词性、情感词汇、N-gram等。
5. 模型训练：选择SVM、决策树、朴素贝叶斯等机器学习算法对上面提取的特征和标注的数据进行模型的训练。

模型评估：使用测试集来测试模

李彩玲.网络健康社区的主题识别与情感分析