* **项目名称：Sentiment Analysis on Movie Reviews(Kaggle)**

**开发工具：**gensim, nltk, sklearn, word2vec, TF-IDF, SVM, TextCNN, TextRNN

**项目描述：**根据带情感标签的影评数据，训练影评自动情感分析模型。

**主要工作：**基于影评数据训练自动情感分析模型，具体工作内容如下：

1. 读入数据，删除html符号，标点符号等，去停用词
2. 对数据做补齐或截断处理
3. 模型选择：

bow/TF-IDF+LR/NB/SVM

word2vec+TextCNN

1. 模型融合：概率加权融合
2. 模型使用：用训练好的模型对未标记数据进行自动情感分析

比赛地址：<https://www.kaggle.com/c/word2vec-nlp-tutorial>

TF-IDF + LR/NB/SVM 解决方案：

1. <https://www.jianshu.com/p/859c0bd726c7>
2. <https://blog.csdn.net/laozhaokun/article/details/42807241>
3. <http://nbviewer.jupyter.org/github/jmsteinw/Notebooks/blob/master/NLP_Movies.ipynb>

word2vec+TextCNN 解决方案：

<https://machinelearningmastery.com/predict-sentiment-movie-reviews-using-deep-learning/>

涵盖知识点：**词袋模型、TF-IDF、逻辑回归、朴素贝叶斯、支持向量机、word2vec(CBOW、skip-gram)、卷积神经网络**