NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

School of Computer Science and Engineering
College of Engineering

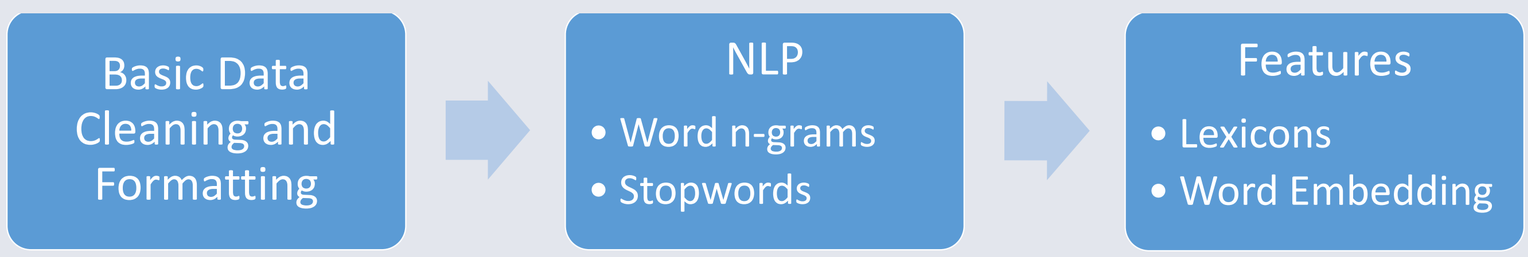
Developing a Hybrid Affect Intensity Tool

Student: Derrick Peh Jia Hao **Supervisor: Dr Althea Liang**

Existing datasets are mainly annotated categorically without an indication of intensity. This makes analysis of sentiments very subjective and prone to opinion spam.

By creating an affect analysis tool that measures sentiments based on intensity, the degree to which affect is expressed in text could be easily understood.

Preprocessing and Feature Generation



Post Processing

Lexicon-Based Method

Texts are being evaluated by various Lexical Resources and a sentiment score is returned. The output of a sentiment score can vary as it is dependent on the type of lexical resource used.
E.g.

- Counting the number of positive and negative words
- Using weighted average of sentiment distributions of word occurrence
- Aggregating positive and negative word scores provided by the resource

Learning-Based Method

Texts are trained into a model by machine learning algorithms.
SVM Regression from LIBLINEAR is used for this model as provides efficiency for training large-scale problems at O(n) cost.

For Logistic regression, LIBLINEAR implements a trust region Newton method. It is accompanied with regularization and loss functions which will help to reduce the amplitude of the model's coefficients, reducing overfitting/multicollinearity.

Hybrid Method

Instead of manual labelling of data which can be time consuming, features generated from lexicon-based method are used as labels. Coupled with derived features from the processed data, the dataset is being fed into the machine learning algorithm.