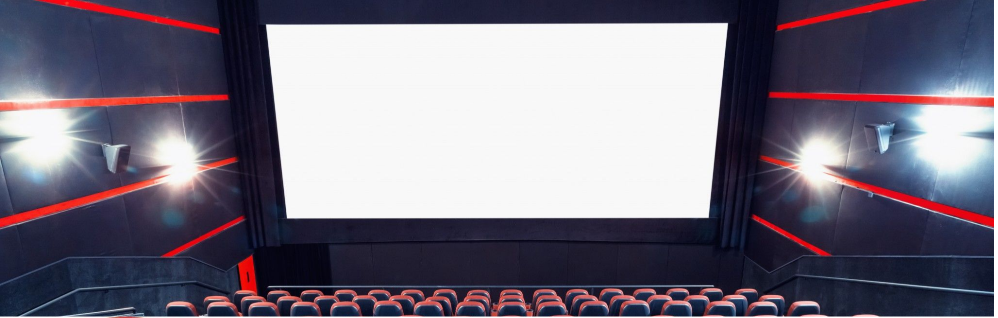
# Sentiment Analysis of Movie Reviews

G1: Riki Saito, Sijie He, Weiwen Leung

# Motivation

Sentiment Analysis is a hugely popular field

We are interested in applying sentiment analysis to movie reviews.



# Datasets

**Cornell Movie Reviews** (tagged as negative or positive)

**IMDB** (tagged with one of 8 emotions: Anger, Fear, Interest, Joy, Love, None, Sadness, Surprise)

# Sample Movie Review.

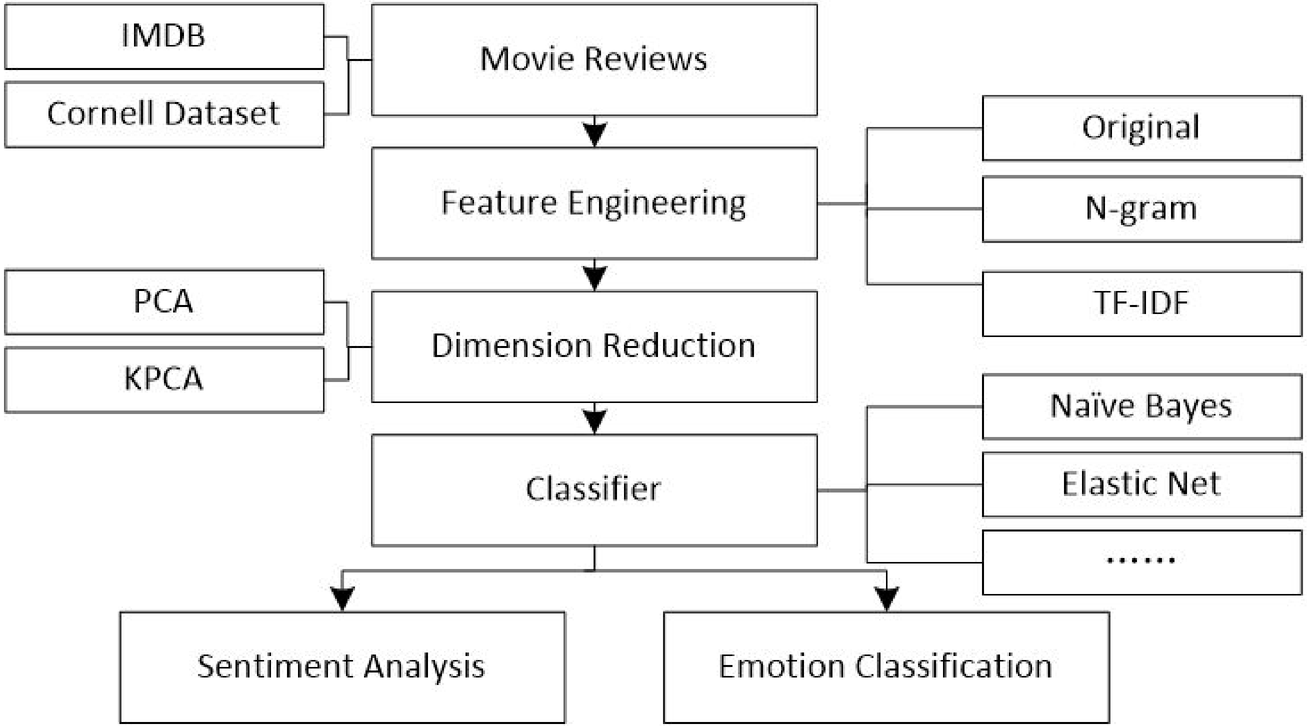
… terrific no-holds-barred performances from the diverse cast .

see it , see it again and when the dvd comes out , buy it , because a movie this hilarious will surely have outtakes to die for . …

*Each word forms a feature*

# Research Questions

- ➔ Which classifiers work best for movie reviews?
- ➔ What is the best set of features (Single Words, N-gram, TF-IDF)?
- ➔ Can dimension reduction improve the performance of classifiers?



# Experimental Design

- ➔ **Feature Engineering**
  n-gram, TF-IDF (more on this later)
- ➔ **Dimensionality Reduction**
  None, PCA, Kernel PCA
- ➔ **Classifiers**
  Regularized Logistic Regression (L1, L2, Elastic Net), Naive Bayes, SVM

# Results

# Sentiment Analysis

| Model | Features | | Training Error | Testing Error | Precision | Recall | F1 |
|---|---|---|---|---|---|---|---|
| Naïve Bayes | Original | | 0.030 | 0.337 | 0.655 | 0.690 | 0.672 |
| | 2-gram | | 0.001 | 0.267 | 0.695 | 0.832 | 0.757 |
| | TFIDF | | 0.008 | 0.351 | 0.638 | 0.690 | 0.663 |
| | KPCA | Original | 0.067 | 0.082 | 0.920 | 0.916 | 0.918 |
| | | 2-gram | 0.121 | 0.187 | 0.785 | 0.862 | 0.822 |
| | | TFIDF | 0.254 | 0.361 | 0.731 | 0.442 | 0.551 |
| L2 Regularized Logistic Regression | Original | | 0.000 | 0.160 | 0.843 | 0.835 | 0.839 |
| | 2-gram | | 0.000 | 0.148 | 0.855 | 0.848 | 0.851 |
| | TFIDF | | 0.028 | 0.160 | 0.831 | 0.853 | 0.842 |
| | KPCA | Original | 0.000 | 0.499 | 0.501 | 0.800 | 0.616 |
| | | 2-gram | 0.000 | 0.499 | 0.501 | 0.800 | 0.616 |
| | | TFIDF | 0.000 | 0.486 | 0.517 | 0.432 | 0.471 |
| L1 Regularized Logistic Regression | Original | | 0.000 | 0.186 | 0.816 | 0.812 | 0.814 |
| | 2-gram | | 0.000 | 0.174 | 0.824 | 0.829 | 0.827 |
| | TFIDF | | 0.209 | 0.247 | 0.726 | 0.812 | 0.767 |
| | KPCA | Original | 0.005 | 0.007 | 0.987 | 1.000 | 0.994 |
| | | 2-gram | 0.041 | 0.041 | 0.987 | 0.930 | 0.958 |
| | | TFIDF | 0.333 | 0.428 | 0.571 | 0.577 | 0.574 |
| Elastic Net | KPCA | Original | 0.000 | 0.044 | 0.946 | 0.968 | 0.957 |

# Emotion (multi-label)

| | | | Training error | Test error | Precision | Recall | F1 |
|---|---|---|---|---|---|---|---|
| L2 Regularized Logistic Regression | Original | | 0.0369 | 0.171 | 0.400 | 0.220 | 0.284 |
| | 2-gram | | 0.0174 | 0.176 | 0.365 | 0.197 | 0.256 |
| | TFIDF | | 0.1155 | 0.157 | 0.433 | 0.075 | 0.128 |
| Naive Bayes | Original | | 0.1043 | 0.219 | 0.321 | 0.380 | 0.348 |
| | 2-gram | | 0.0582 | 0.197 | 0.342 | 0.304 | 0.322 |
| | TFIDF | | 0.1009 | 0.217 | 0.319 | 0.366 | 0.341 |
| | PCA | Original | 0.1788 | 0.197 | 0.351 | 0.330 | 0.340 |
| | | 2 gram | 0.1765 | 0.196 | 0.340 | 0.292 | 0.314 |
| | | TFIDF | 0.1524 | 0.175 | 0.314 | 0.119 | 0.173 |
| Elastic Net | Original | | 0.0702 | 0.211 | 0.308 | 0.296 | 0.302 |
| | 2-gram | | 0.2047 | 0.234 | 0.255 | 0.270 | 0.262 |
| | TFIDF | | 0.186 | 0.177 | 0.375 | 0.232 | 0.286 |

# Best Results

| Problem | Classifier | Dimension Reduction | Feature Engineering | Training Error | Testing Error | Precision | Recall | F1 |
|---|---|---|---|---|---|---|---|---|
| Sentiment | Naïve Bayes | KPCA | Original | 0.067 | 0.082 | 0.920 | 0.916 | 0.918 |
| | L1 Logistic Regression | KPCA | Original | 0.005 | 0.007 | 0.987 | 1.000 | 0.994 |
| | Elastic Net | KPCA | Original | 0.000 | 0.044 | 0.946 | 0.968 | 0.957 |
| Emotion | Naïve Bayes | None | Original | 0.104 | 0.219 | 0.321 | 0.38 | 0.348 |
| | Naïve Bayes | PCA | Original | 0.179 | 0.197 | 0.351 | 0.33 | 0.34 |
| | Elastic Net | None | TF-IDF | 0.186 | 0.177 | 0.375 | 0.232 | 0.286 |

# Sentiment vs Emotion Classification

***- Why was Emotion Classification much more difficult than Sentiment Analysis?***

| ***Sentiment Analysis*** | ***Emotion Classification*** |
|---|---|
| 2 classes, 2000 reviews X 13290 word features | 8 classes, 629 sentences X 728 word features |
| Words generally describe only one of two sentiment: positive or negative | Words can describe multiple emotions = emotions highly correlated |

# Next Steps:

**Wish List**

I want:

- ...
- ...
- ...

# Domain Similarity Measure

### Horror Movies

"It was a **shocking** and **great** movie."

### Comedy Movies

"It was **great**, but **shocking** that the actor said vulgarities."

### History Movies

"A **shocking** depiction of the war."

# Calculating domain similarity

$$R_{i,j} \equiv \frac{n_{i,j}^{S} - n_{i,j}^{O}}{n_{i,j}^{S} + n_{i,j}^{O}}$$

$n_{i,j}^{S}, n_{i,j}^{O}$ = frequency of words i and j sharing the same and opposite sentiment

$$DomainSimilarity_{m,n} = \frac{\sum_{w=1}^{D} \sum_{v \neq w} \left| R_{w,v}^{m} + R_{w,v}^{n} \right| \cdot \min\{N_{w,v}^{m}, N_{w,v}^{n}\}}{\sum_{w=1}^{D} \sum_{v \neq w} \left( \left| R_{w,v}^{m} \right| \cdot N_{w,v}^{m} + \left| R_{w,v}^{n} \right| \cdot N_{w,v}^{n} \right)}$$

(adapted from Wu and Huang, 2016)

# Domain Adaptation

With our existing movie review classifier, what can we do if we had an unlabelled dataset of documentary reviews?

Answer:

Adapt the classifier to make better predictions by using *domain adaptation.*

# Steps

**1**

Build a global classifier across all movie reviews

**2**

Build a local classifier for each movie genre

**3**

Calculate domain similarity between each movie genre and documentaries

**4**

Build a classifier which combines the global and local classifiers, but weighs movie genres that are similar more heavily