

Countries

Nguyen, Annie Zou

Government	Percentage
Current government	85%
Previous government	15%

Related Work

- # Regression

The figure consists of two vertically stacked scatter plots, each representing a linear regression model for a different social media platform.

Top Plot: Facebook
 Title: Linear Regression: Number of views vs. the word 'Facebook'
 Y-axis: Number of Views (scaled by $1e8$)
 X-axis: Word Count
 Legend: Actual (red dots), Predicted (blue dots)
 The plot shows a dense cluster of data points at low word counts (0-4), with actual values ranging from approximately 0 to 2.0 billion views. Predicted values are generally lower, mostly below 0.5 billion views. As word count increases, both actual and predicted values drop significantly, approaching zero.

Bottom Plot: Twitter
 Title: Linear Regression: Number of views vs. the word 'Twitter'
 Y-axis: Number of Views (scaled by $1e8$)
 X-axis: Word Count
 Legend: Actual (red dots), Predicted (blue dots)
 The plot shows a similar distribution to the Facebook plot. Actual view counts are highest at low word counts, reaching up to 2.0 billion views. Predicted values are consistently lower than actual values at low word counts, with most predictions falling below 0.5 billion views. The data points for both actual and predicted values converge towards zero as the word count increases beyond 6.

Figure 1: **Visualizations of fitting multiple linear regression to the US dataset, plotting as number of views versus word counts for the top weighted words from US LDA.**

Comparison of United States' Top Features with United Kingdom's and Canada's

Country	Facebook	Instagram	Twitter	Music	Trailer	Food	Makeup	News	War	Funny
United States(US)	6800	5600	6500	5200	2500	2500	1500	4300	1800	1600
United Kingdom(UK)	12800	13200	5100	2100	3700	4000	3900	2700	3300	2800
Canada (CA)	3600	2800	4000	2500	3200	4200	2500	2500	1400	2200

Classification

The figure displays three Precision-Recall curves for different machine learning models: Multinomial Naive Bayes, Linear SVM, and Logistic Regression. Each plot shows Precision (Y-axis) versus Recall (X-axis) for 15 categories. The curves are color-coded by category, and iso-f1 curves are shown as gray lines. The area under each curve is labeled, representing the f1 score for that category.

Precision-Recall Curve for Multinomial Naive Bayes

- iso-f1 curves
- Film & Animation (area = 0.89)
- Autos & Vehicles (area = 0.98)
- Music (area = 0.98)
- Pets & Animals (area = 0.96)
- Sports (area = 0.95)
- Travel & Events (area = 0.97)
- Gaming (area = 0.98)
- People & Blogs (area = nan)
- Comedy (area = 0.08)
- Entertainment (area = 0.06)
- News & Politics (area = 0.20)
- Howto & Style (area = 0.05)
- Education (area = 0.07)
- Science & Technology (area = 0.05)
- Nonprofits & Activism (area = 0.07)
- Shows (area = 0.00)

Precision-Recall Curve for Linear SVM

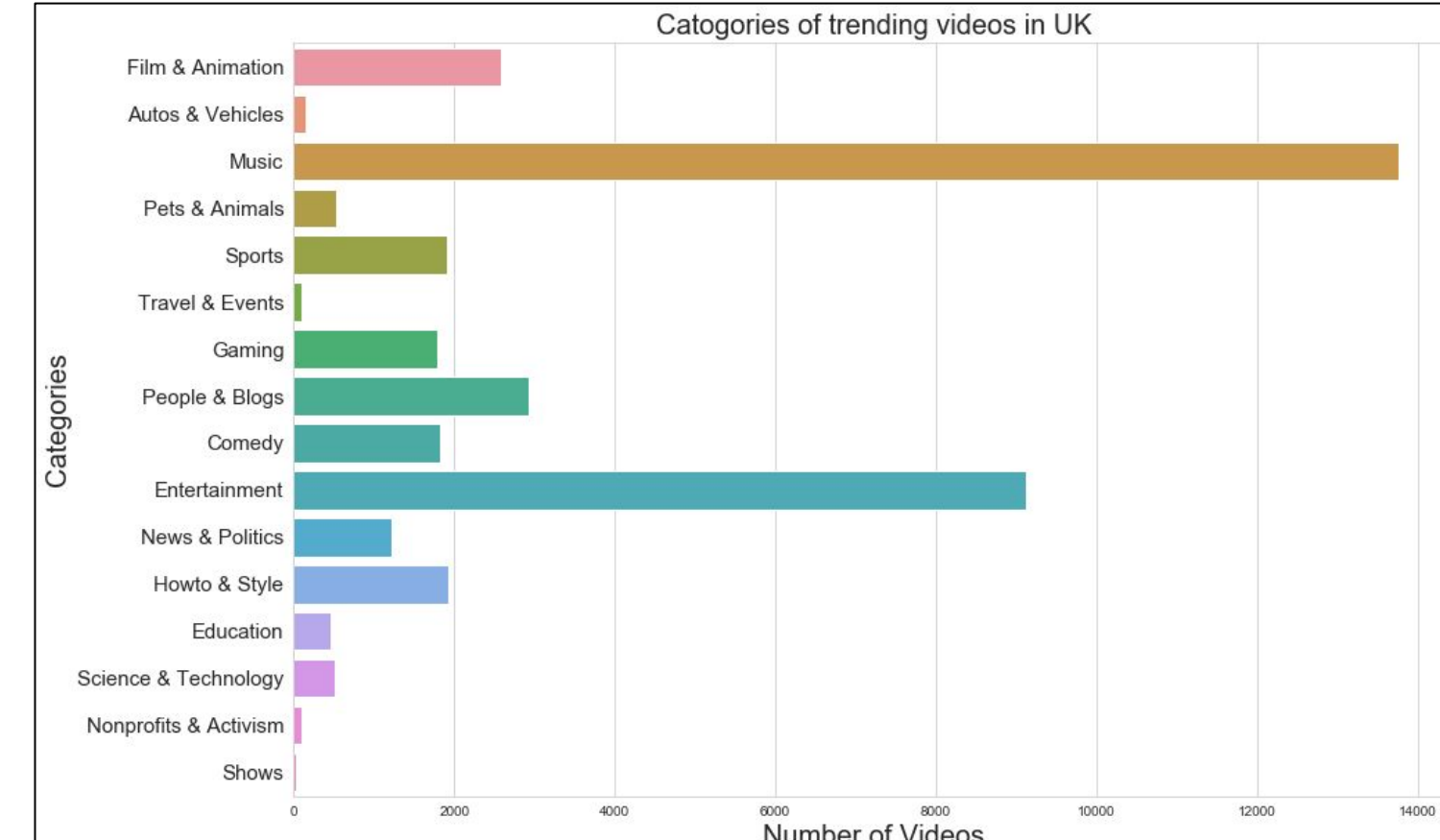
- iso-f1 curves
- Film & Animation (area = 1.00)
- Autos & Vehicles (area = 0.97)
- Music (area = 1.00)
- Pets & Animals (area = 0.99)
- Sports (area = 1.00)
- Travel & Events (area = 0.99)
- Gaming (area = 0.99)
- People & Blogs (area = nan)
- Comedy (area = 0.09)
- Entertainment (area = 0.06)
- News & Politics (area = 0.20)
- Howto & Style (area = 0.05)
- Education (area = 0.08)
- Science & Technology (area = 0.05)
- Nonprofits & Activism (area = 0.08)
- Shows (area = 0.01)

Precision-Recall Curve for Logistic Regression

- iso-f1 curves
- Film & Animation (area = 0.52)
- Autos & Vehicles (area = 0.21)
- Music (area = 0.94)
- Pets & Animals (area = 0.85)
- Sports (area = 0.92)
- Travel & Events (area = 0.69)
- Gaming (area = 0.52)
- People & Blogs (area = nan)
- Comedy (area = 0.07)
- Entertainment (area = 0.07)
- News & Politics (area = 0.22)
- Howto & Style (area = 0.04)
- Education (area = 0.11)
- Science & Technology (area = 0.09)
- Nonprofits & Activism (area = 0.05)
- Shows (area = 0.00)

Figure 4: **Precision-Recall curves for multi-class classifiers on the US dataset**

Categories	Number of Videos
Film & Animation	2500
Autos & Vehicles	500
Music	6500
Pets & Animals	1000
Sports	2200
Travel & Events	500
Gaming	1000
People & Blogs	3500
Comedy	3800
Entertainment	10000
News & Politics	2800
Howto & Style	4000
Education	1800
Science & Technology	2500
Nonprofits & Activism	100
Shows	100



Citations

-
- Precision-Recall Curve for Linear SVM**
- iso-f1 curves
- Film & Animation (area = 1.00)
 - Autos & Vehicles (area = 0.97)
 - Music (area = 1.00)
 - Pets & Animals (area = 0.99)
 - Sports (area = 1.00)
 - Travel & Events (area = 0.99)
 - Gaming (area = 0.99)
 - People & Blogs (area = nan)
 - Comedy (area = 0.09)
 - Entertainment (area = 0.06)
 - News & Politics (area = 0.20)
 - Howto & Style (area = 0.05)
 - Education (area = 0.08)
 - Science & Technology (area = 0.05)
 - Nonprofits & Activism (area = 0.08)
 - Shows (area = 0.01)
- Precision-Recall Curve for Logistic Regression**
- iso-f1 curves
- Film & Animation (area = 0.52)
 - Autos & Vehicles (area = 0.21)
 - Music (area = 0.94)
 - Pets & Animals (area = 0.85)
 - Sports (area = 0.92)
 - Travel & Events (area = 0.69)
 - Gaming (area = 0.52)
 - People & Blogs (area = nan)
 - Comedy (area = 0.07)
 - Entertainment (area = 0.07)
 - News & Politics (area = 0.22)
 - Howto & Style (area = 0.04)
 - Education (area = 0.11)
 - Science & Technology (area = 0.09)
 - Nonprofits & Activism (area = 0.05)
 - Shows (area = 0.00)