

# Identifying Media Bias in Online News

Noa Glaser, John Sholar

CS 229: Machine Learning – Final Project

## Overview

Cognitive science defines bias as deviation from the norm, or true value. We explore applications of supervised (Naive Bayes, SVM) and unsupervised (K-Nearest-Neighbors) machine learning techniques to analyze media bias.

**Objective:** Identify trends in the selection and presentation of online news

## Dataset and Processing

Queried Event Registry, a news/event aggregator.

### Keyword Analysis

- 160,000 (News Org., Article Title) pairings for analysis of 20 multinational news sources.
- 1000 pairings focused specifically on Israel-Palestine conflict, from Israeli, Palestinian, American, and International sources.
- Data preprocessed using Scikit-Learn vectorizer and TF-IDF transformer.

### Event Analysis

- Events are clusters of news articles by event registry by time, location and key words
- Searched for 600 events with location Israel from June 1<sup>st</sup> to September 30<sup>th</sup> 2014
- Used all fetched 1143 news sources to normalize articles, clustered 100 news sources with most captured articles

## Title Keyword Analysis

### Predicting News Sources

- Trained Multinomial Naive Bayes (NB) and Multiclass SVM models to discriminate between 8 news sources based on article title tokens.
- Accuracy of classifiers validates keyword analysis.

Model	Precision	Recall	F <sub>1</sub> Score
Naive Bayes	.48	.40	.38
SVM	.47	.48	.48

### Keyword Analysis

- NB yields  $P(\text{token} | \text{news outlet})$  for every (word, news outlet) pair
- Informally, model how indicative a token is of a given news outlet as:

$$\log \frac{P(\text{token} | \text{news outlet})}{P(\text{token} | \text{NOT news outlet})}$$

Tokens that maximize the above value for selected news sources:

Source	Reputation	Keyword Tokens
Haaretz	Lib. Israeli	jewish, bid, live, watch, world, lawmakers, probe, vote
Israel Hayom	Con. Israeli	israel, turkey, jews, mind, caught, european, hamas, blackmail, pm
Palestine Chronicle	Con. Palestinian	palestine, book, apartheid, nakba, zionist, media, bds, struggle
Palestine News Agency (WAFA)	Lib. Palestinian	review, dailies, focus, newspapers, highlight, killing, premier, international
Fox News	Con. American	claim, site, holy, nations, western, muslim, prepares, holocaust
Reuters	Mod. International	treaty, solidifying, update, kill, vatican, agrees, relationship, troops

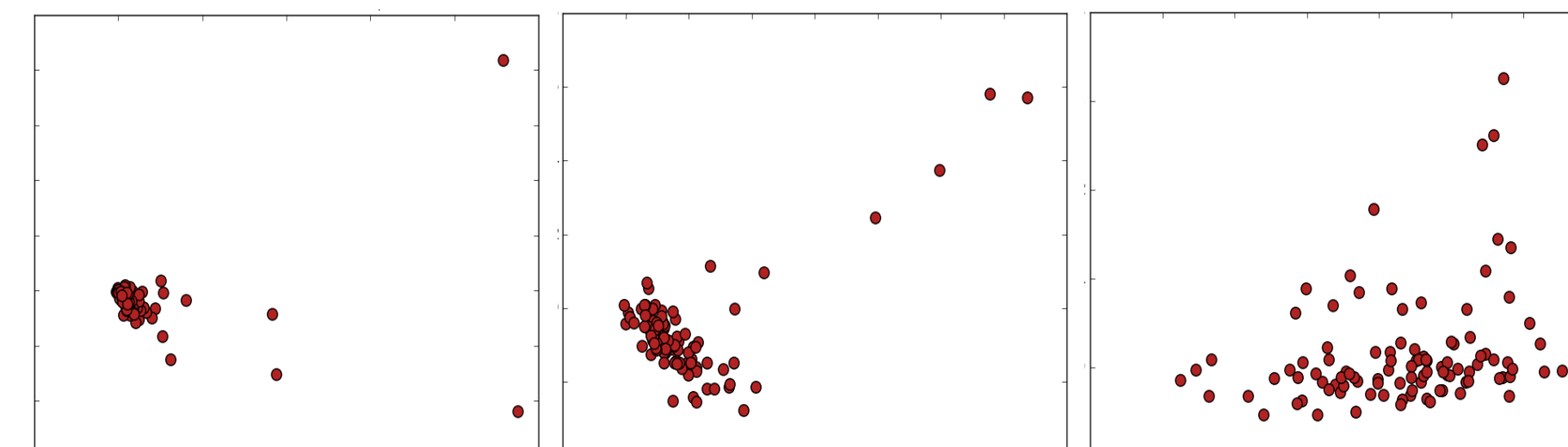
## Event Selection Bias

### Data

To study event selection bias, we examined the news outlets by:

- Frequency: Number of articles on each event
- Binomial: Whether the source reported on each event
- Normalized: Frequency norm2 normalized and centered about zero

**Figure 1:** Frequency [left], binomial [center] normalized [right] vectors of 600 events. PCA units omitted. Representing article frequency makes all local news sources outliers due to report propensity. Event binomials represent cherry picking. Events normalized show good spread.



### Results

KNN, Mixture of Gaussians, Hierarchical Clustering had similar results.

Most popular clustering based on poll of 20 Stanford students, K-Nearest Neighbors of the normalized vectors with 6 clusters

www.presstv.ir, Irish Sun, International Business Times UK, www.aa.com.tr, www.albawaba.com, www.palestine-info.co.uk, Naharnet, News From Antivar.com, english.wafa.ps	GULF NEWS, The Huffington Post, The Daily Star Lebanon, ABC News, CNN International, english.farsnews.com, Republika Online, Star Tribune, ReliefWeb	TIME, Los Angeles Times, Truthdig, DIE WELT, The Japan Times, San Francisco Gate, Ad Hoc News, The National, N24.de, Thomson Reuters Foundation, The Irish Times, www.greenpeace-magazin.de
The Jerusalem Post, Arutz Sheva, Haaretz.com, www.israelhayom.com, ynet, The New York Times, Jewish Journal, The Washington Post, The Guardian, The Independent	BBC News, NDTV, CBC News, The Globe and Mail, Telegraph.co.uk, Xinhuanet.com, Reuters, Mail Online, Miami Herald, Fox News, news24	Yahoo News, DW.com, Independent.ie, Economic Times, The Wall Street Journal, Sify, El Economista, Daily News and Analysis (DNA) India

\* Due to poster space constraints, abridged to outlets with most articles in each cluster, full clusters can be found in our paper