Predicting Wikipedia Article Translation Counts

by Jay Park

EN

Outline

Contents [hide]

Wikipedia is the 13th most visited website worldwide

Wikipedia is the 13th most visited website worldwide

- Founded in 2001
- Available in 285 languages
- The English Language Wikipedia is the largest
- The official song of Wikipedia is “Hotel Wikipedia”

Articles are often translated/written in other languages

Themaverick7 Talk Sandbox Preferences Beta Watchlist Contributions Log out

Wikipedia
The Free Encyclopedia

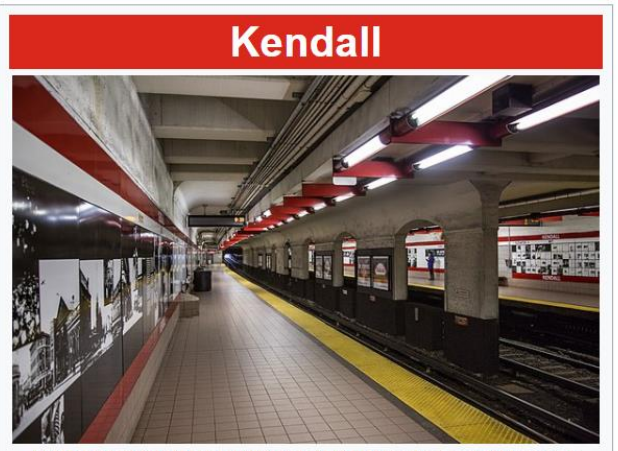
Article Talk | Read Edit View history More | Search Wikipedia

Kendall/MIT station

From Wikipedia, the free encyclopedia
(Redirected from Kendall-MIT (MBTA station))

Coordinates: 42.3623°N 71.0862°W

Kendall/MIT is an underground rapid transit station in Cambridge, Massachusetts. It serves the MBTA Red Line, Located at the intersection of Main Street and Broadway, it is named for the primary areas it serves - the Kendall Square business district and the Massachusetts Institute of Technology (MIT). Opened in March 1912 as part of the original Cambridge Subway, Kendall/MIT has two side platforms serving the line's two tracks. The *Kendall Band*, a public art installation of hand-operated musical sculptures, is located between the tracks in the station with controls located on the platforms. Kendall/MIT station is accessible. With 17,018 weekday boardings by a FY2019 count, Kendall/MIT has the fourth highest ridership among MBTA subway stations.[2]



Outbound platform with historic timeline and images from nearby Massachusetts Institute of Technology

Location Main Street at Broadway

Contents [hide]

Main page
Contents
Current events
Random article
About Wikipedia
Contact us
Donate

Contribute

Help
Learn to edit
Community portal
Recent changes
Upload file

Tools

Articles are often translated/written in other languages

Cite this page
Wikidata item

Print/export
Download as PDF
Printable version

In other projects
Wikimedia Commons

Languages
Deutsch
Español
中文
Edit links

3 References
4 External links

History [edit]

Kepler, one section of *Kendall Band*, in 2011

The Cambridge subway opened from Park Street Under to Harvard on March 23, 1912, with intermediate stops at Central and **Kendall**.[1] From the early 20th century through the 1970s, the MBTA operated a powerhouse above ground in Kendall Square, including rotary converters (also called cycloconverters) to transform incoming AC electrical power to 600 volts DC power fed to the third rail to run the subway. An old-fashioned cycloconverter consisted of an AC motor coupled to a huge, slowly rotating flywheel coupled to a DC generator, hence the name.
Despite the development of compact low-maintenance semiconductor-based power rectifiers, the long-obsolete electromechanical technology still occupied prime real estate in the heart of Kendall Square. The MBTA powerhouse was demolished, and replaced with an office building located at the convergence of Broadway and Main Street.[*citation needed*]

Connections	MBTA bus: CT2, 64, 68, 85 EZRide
Construction	
Bicycle facilities	58 spaces
Disabled access	Yes
History	
Opened	March 23, 1912[1]
Passengers	
FY2019	17,018 (weekday average boardings)[2]

Services

Preceding station	MBTA	Following station
Central *toward Alewife*	Red Line	Charles/MGH *toward Ashmont or Braintree*

Articles are often translated/written in other languages

CAN WE PREDICT THIS?

History [edit]

Kepler, one
Band, in 2011

The Cambridge subway opened from Park Street Under to Harvard on March 23, 1912, with intermediate stops at Central and **Kendall**.[1] From the early 20th century through the 1970s, the MBTA operated a powerhouse above ground in Kendall Square, including rotary converters (also called cycloconverters) to transform incoming AC electrical power to 600 volts DC power fed to the third rail to run the subway. An old-fashioned cycloconverter consisted of an AC motor coupled to a huge, slowly rotating flywheel coupled to a DC generator, hence the name.
Despite the development of compact low-maintenance semiconductor-based power rectifiers, the long-obsolete electromechanical technology still occupied prime real estate in the heart of Kendall Square. The MBTA powerhouse was demolished, and replaced with an office building located at the convergence of Broadway and Main Street.[*citation needed*]

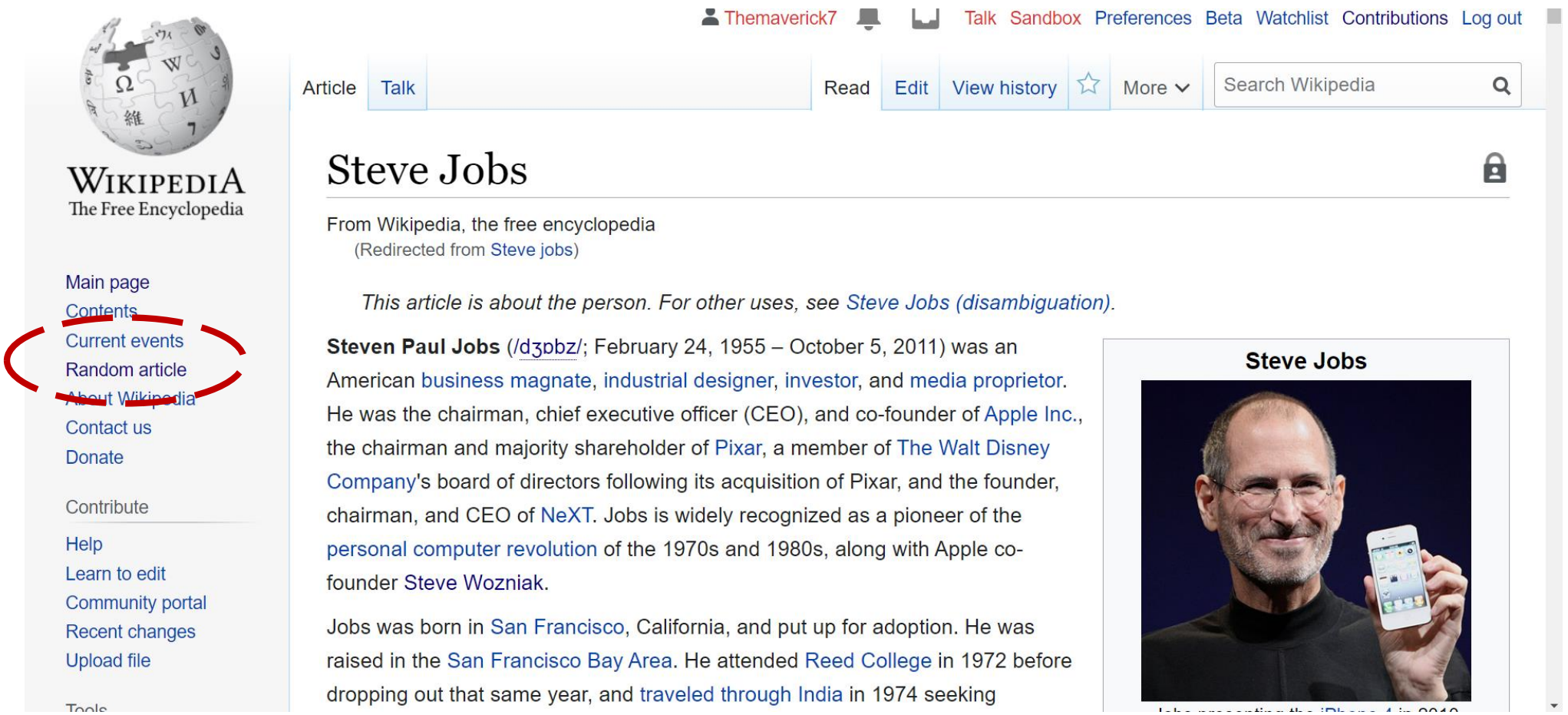
Connections	MBTA bus: CT2, 64, 68, 85 EZRide
Construction	
Bicycle facilities	58 spaces
Disabled access	Yes
History	
Opened	March 23, 1912[1]
Passengers	
FY2019	17,018 (weekday average boardings)[2]
Services	

Preceding station	MBTA	Following station
Central *toward Alewife*	Red Line	Charles/MGH *toward Ashmont or Braintree*

Outline

Contents [hide]

3655 Articles were scraped using the 'Random article' feature



Themaverick7 Talk Sandbox Preferences Beta Watchlist Contributions Log out

Wikipedia
The Free Encyclopedia

Article Talk | Read Edit View history | More | Search Wikipedia

Main page
Contents
Current events
Random article
About Wikipedia
Contact us
Donate

Contribute
Help
Learn to edit
Community portal
Recent changes
Upload file

Tools

Steve Jobs

From Wikipedia, the free encyclopedia
(Redirected from Steve jobs)

This article is about the person. For other uses, see Steve Jobs (disambiguation).

Steven Paul Jobs (/dʒɒbz/; February 24, 1955 – October 5, 2011) was an American business magnate, industrial designer, investor, and media proprietor. He was the chairman, chief executive officer (CEO), and co-founder of Apple Inc., the chairman and majority shareholder of Pixar, a member of The Walt Disney Company's board of directors following its acquisition of Pixar, and the founder, chairman, and CEO of NeXT. Jobs is widely recognized as a pioneer of the personal computer revolution of the 1970s and 1980s, along with Apple co-founder Steve Wozniak.

Jobs was born in San Francisco, California, and put up for adoption. He was raised in the San Francisco Bay Area. He attended Reed College in 1972 before dropping out that same year, and traveled through India in 1974 seeking

Steve Jobs

Jobs presenting the iPhone 4 in 2010

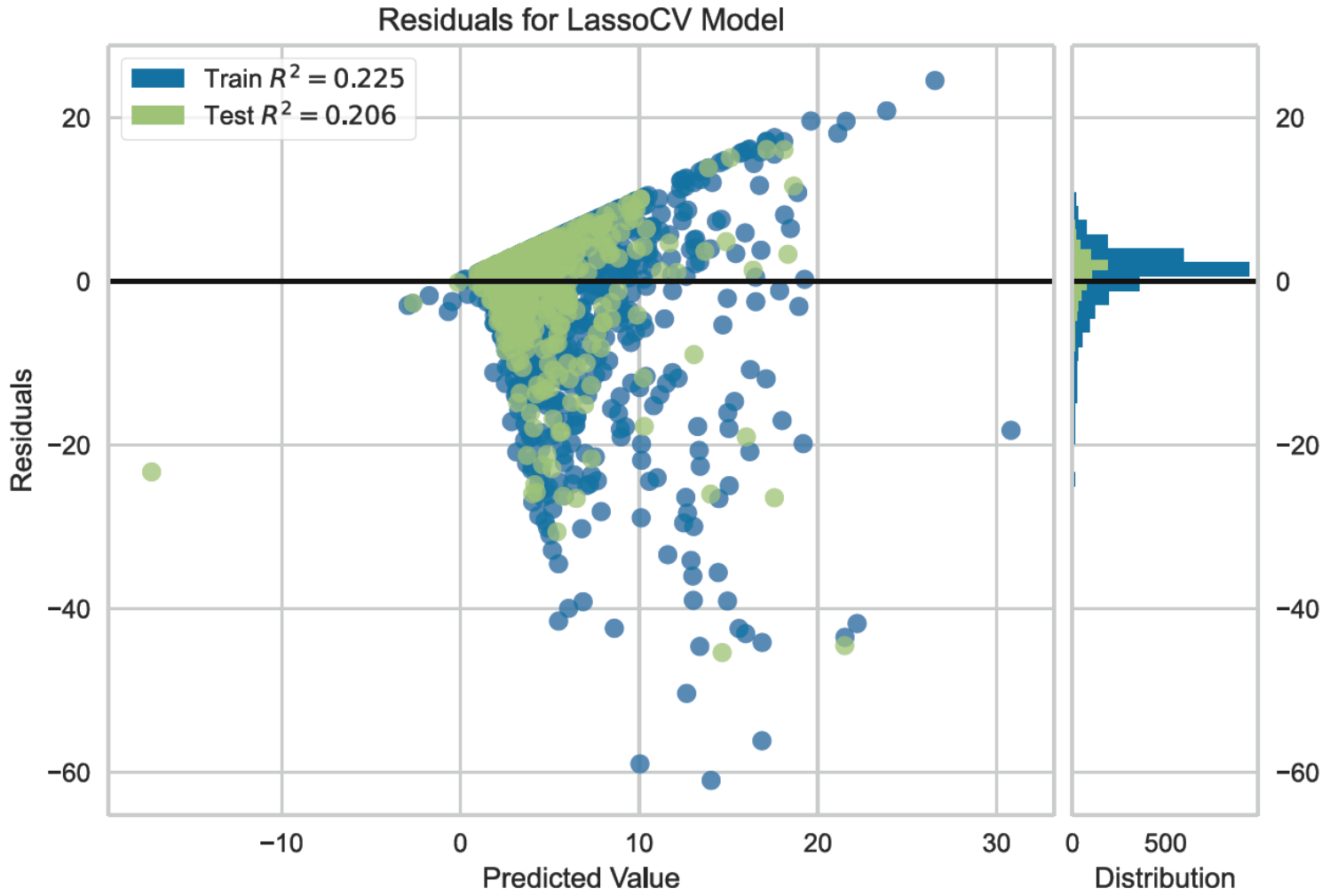
Selected features for regression

- Body text length
- Days since article creation
- # of references

Selected features for regression

- Body text length
- Days since article creation
- # of references
- Total length of text in tables
- # of “Main article” links
- # of “See also” links
- # of “Further information” links
- # of links to other wiki articles
- Is sports?

Regression using manually-curated features were weakly predictive



NLP method: Bag of Words

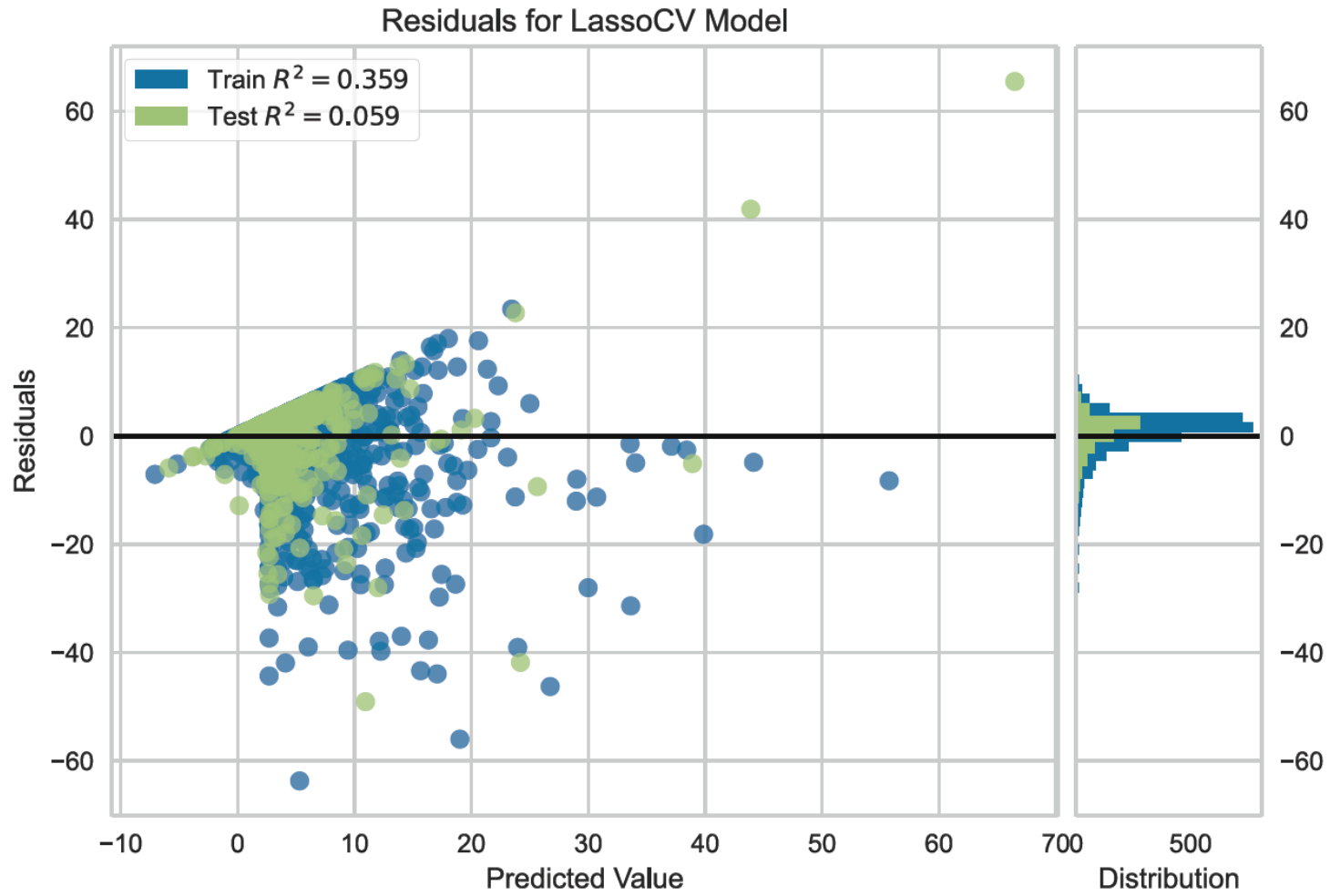
This is a article about the future of our country. It looks quite grim. However, we believe we can make it through all the hardships like we have before. Our country is strong.The author of this article approves this message.

Cleanup

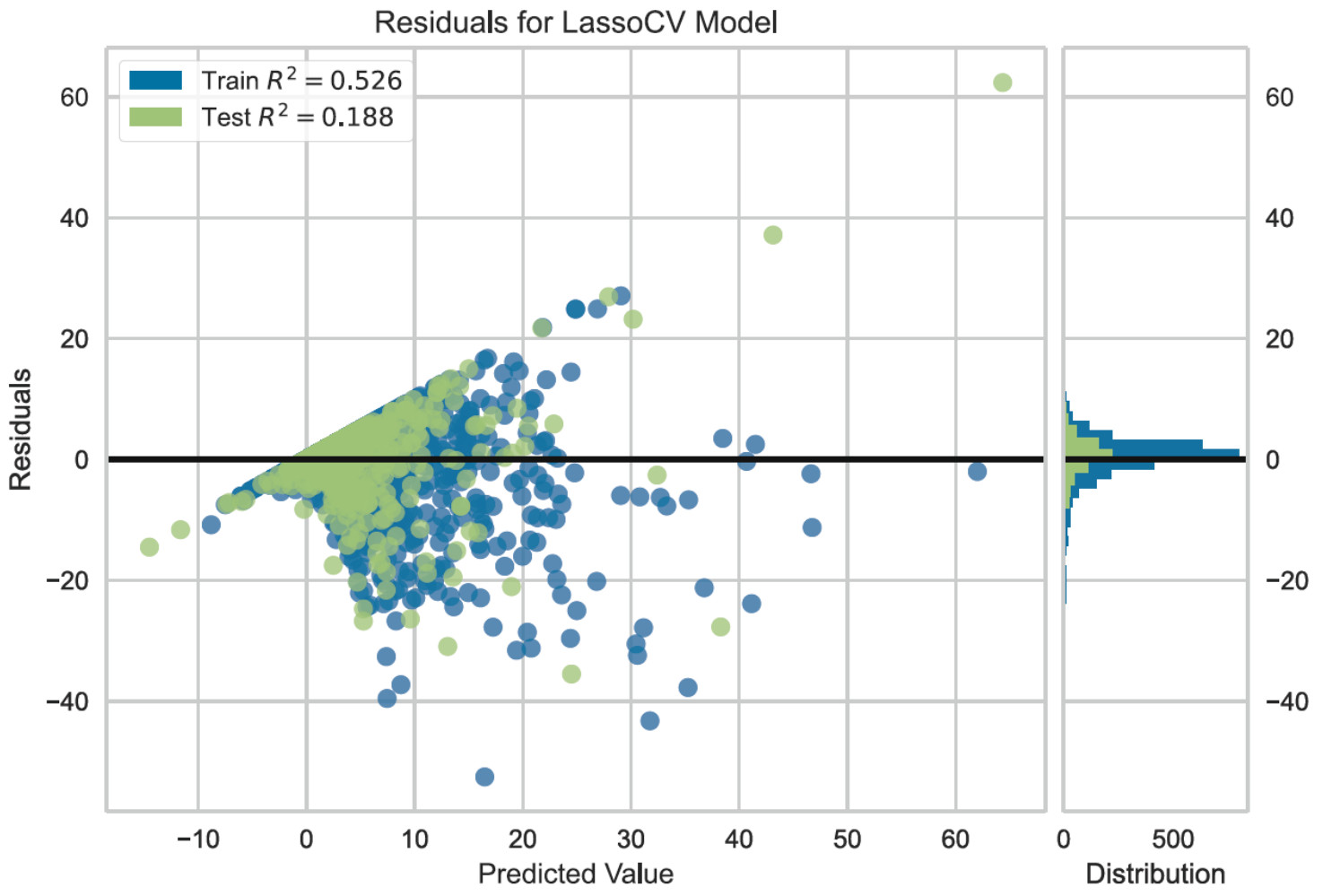
'article'	'future'	'country'	'comfort'	'author'	'message'
2	1	2	0	1	1

. . .

Regression using basic NLP was less predictive



Combination of the two approaches did not yield an noticeable improvement



Outline

Contents [hide]

Conclusions and future directions

- # of languages can be weakly predicted
- The most under-translated articles tend to be regional articles in the United States, often featuring sports
- Future improvements:
  - More refined NLP methods, or neural networks
  - Try incorporating PageRank
  - Frequency of editing per year

Thank you Wikipedia!

Appendix: List of coefficients for the most impactful words

population	1.212857
county	0.814047
roman	0.779921
also	0.703669
le	0.666219
...	...
united	-0.249402
well	-0.271809
show	-0.279516
date	-0.377511
economic	-0.519919

Appendix: Histogram of language distributions