

Vandalism Detection (WSDM Cup 2017)

1. Project Objectives

Wikidata is the new, large-scale knowledge base of the Wikimedia Foundation which can be edited by anyone. Its knowledge is increasingly used within Wikipedia as well as in all kinds of information systems, which imposes high demands on its integrity. Nevertheless, Wikidata frequently gets vandalized, exposing all its users to the risk of spreading vandalized and falsified information.

The goal of the vandalism detection task is to detect vandalism nearly in real time as soon as it happens. Hence, the following rules apply:

- Use of any additional data that is newer than the provided training data is forbidden. In particular, you may not scrape any Wikimedia website, use the API, the dumps, or any related data source to obtain data that is newer than February 29, 2016.
- You may use sources of publicly available external data having to do with geographical information, demographic information, natural language processing, etc. This data must not relate to the specific revision label (vandalism vs regular).

2. Dataset and Evaluation

2.1 Dataset

To develop your software, we provide you with a training, validation and test corpus that consists of Wikidata revisions and whether they are considered vandalism. For detailed information and downloading the data, please check this link:
<http://www.wsdm-cup-2017.org/vandalism-detection.html>.

For each Wikidata revision in the test corpus, your software shall output a vandalism score in the range $[0,1]$. The output shall be formatted as a CSV file in the format RFC4180 and consist of two columns: The first column denotes Wikidata's revision id as an integer and the second column denotes the vandalism score as a float32.

2.2 Evaluation

ROC-AUC as primary evaluation measure. For informational purposes, you might compute further evaluation measures such as PR-AUC and the runtime of the software.

3 Grading Policy

Final grade of this project will be given based on your team's evaluation scores in experiments, implementation and the final report. TAs will read your instructions on the report to run your code and test your program. Please make sure you write clear instructions.