School of Computing and Information Systems
The University of Melbourne
COMP90049 Knowledge Technologies, Semester 2 2018

Project 2: Twitter Trolls and the Tweeters who Love them

Due:	Stage I: 2pm (14h00 UTC+11), Thu 11 Oct 2018 Stage II: 2pm (14h00 UTC+11), Thu 18 Oct 2018
Submission materials:	Stage I: Predictions; PDF Report Stage II: Reviews (via Turnitin PeerMark)
Assessment criteria:	Method, Critical Analysis, Report Quality; Reviews
Marks:	The Project will contribute 20% of your overall mark for the subject.

Overview

In this Project, you will be tasked with applying some Machine Learning methods over a dataset comprised of a number of "troll" tweets, generated by Twitter users associated with the Internet Research Agency. The goal of this Project is to gain some knowledge about the problem of automatically identifying troll users based on the text of their tweets. This aims to reinforce concepts in (supervised) Machine Learning, and to strengthen your skills in data analysis and problem solving.

Although maximising the performance of a Machine Learning system on the given dataset is occasionally an interesting question, here we are only using the evaluation in service to help us find knowledge. In all likelihood, this will mean grappling with questions along the lines of: "Can we use tweet text to help us to identify trolls on Twitter? If so, how? If not, why not?"

Deliverables

1. The predicted labels of the test tweets.
2. An **anonymous** technical report, of 1100–1350 words, which must:
   - Give a short description of the problem and data set
   - **Briefly** summarise some relevant literature
   - Present the results, in terms of evaluation metric(s) and, ideally, illustrative examples
   - Contextualise the system's behaviour, based on the (admittedly incomplete) understanding from the subject materials
   - **Clearly** demonstrate some knowledge about the problem
3. In Stage II, reviews of three reports written by other students, 250–350 words each, which:
   - Briefly summarise what the author has done
   - Indicate what you think the author has done well, and why
   - Indicate what you think could have been improved, and why

Terms of Use

By using the data, you absolutely must cite its curators, preferably as the working paper which discusses its creation:

> Linvill, Darren and Patrick Warren (2018) *Troll factories: The Internet Research Agency and state-sponsored agenda building* (working paper). Clemson University.

You may alternatively cite the FiveThirtyEight publication which accompanies the available dataset:

> Roeder, Oliver (2018) Why we're sharing 3 million Russian troll tweets. In *FiveThirtyEight*. 31 Jul 2018,`https://fivethirtyeight.com/features/why-were-sharing-3-million-russian-troll-tweets/`.

Due to the very nature of the dataset, you should be cautious in reading the content contained within. Much of the information was probably written with the intent to manipulate or deceive the reader; some tweets could also be reasonably construed as offensive. Although the data contains hypertext links, we strongly recommend not following or dereferencing them.

Needless to say, the opinions expressed within the tweets in no way express the official views of the University of Melbourne or any of its employees. In addition, Linvill and Warren (2018) and Roeder (2018) make certain judgements that could be construed as normative (including the very classification of the users as trolls); using the data in this teaching capacity does not constitute endorsement of the underlying opinions of these authors, nor of the corresponding legal actions taken by the U.S. Justice Department. The University accepts no responsibility for offence caused by any content connected with this data.

If you object to these Terms, please contact us (`nj@unimelb.edu.au`) as soon as possible.

Assessment Criteria

Report (16 marks out of 20)

Method: (20% of the report mark)
You will identify a knowledge problem, and design experiments using one or more Machine Learning methods, which could plausibly be used to gain knowledge about that problem. You will describe your method(s) in a manner which would make your work reproducible from your report. You will produce the predicted labels of the test tweets.

Critical Analysis: (50% of the report mark)
You will explain the practical behaviour of your system(s), referring to the theoretical behaviour of the Machine Learning methods where appropriate. You will support your observations with evidence, in terms of evaluation metrics, and, ideally, illustrative examples. You will derive some knowledge about the underlying problem of identifying a Twitter "troll" user, based on the text of their tweets.

Report Quality: (30% of the report mark)
You will produce a formal report, which is commensurate in style and structure with a (short) research paper. You must express your ideas clearly and concisely, and remain within the word limits. You will include a short summary of related research.

We will post a marking rubric to indicate what we will be looking for in each of these categories when marking.

Reviews (4 marks out of 20)

You will write a review for each of three reports written by other students; you will follow the guidelines as stated above.

1 mark will be assigned to each completed review, and 1 mark will be assigned for overall effort. Completing the reviews is expected to take about 3–4 hours in total.

Changes/Updates to the Project Specifications

If we require any (hopefully small-scale) changes or clarifications to the project specifications, they will be posted on the LMS. Any addendums will supersede information included in this document.

Academic Misconduct

For most people, collaboration will form a natural part of the undertaking of this project. However, it is still an individual task, and so reuse of ideas or excessive influence in algorithm choice and development will be considered cheating. We will be checking submissions for originality and will invoke the University's Academic Misconduct policy (`http://academichonesty.unimelb.edu.au/policy.html`) where inappropriate levels of collusion or plagiarism are deemed to have taken place.

Late Submission Policy

You are strongly encouraged to submit by the time and date specified above, however, if circumstances do not permit this, then the marks will be adjusted as follows:

- Each business day (or part thereof) that the report is submitted after the due date (and time) specified above, 10% will be deducted from the marks available, up until 5 business days (1 week) has passed, after which regular submissions will no longer be accepted.
- Due to the end of semester, and the inherent inconvenience caused by late submission of reviews, any submission after the reviewing system closes will incur a flat 50% penalty (i.e. 2 of the 4 marks available); reviews submitted more than 5 business days (1 week) after the deadline will not be assessed.

Note that submitting the report late will mean that you may lose the opportunity for your report to participate in the reviewing process, which means that you will receive less feedback.