

# MLMI13 Practical: Sentiment Detection of Reviews

Kate Knill, Adian Liusie  
(developed by Milica Gašić, Kevin Heffernan, Simone Teufel, Andreas Vlachos)  
Michaelmas 2021

# Task: Sentiment Classification (1)

**ars** TECHNICA

BIZ & IT TECH SCIENCE POLICY CARS GAMING & CULTURE

"MOM, I JUST WOKE UP" —

## *Dune* 2021 film review: The spice must flow, but then it stops abruptly

"Must-watch" for classic *Dune* fans does not mean the same thing as "must-love."

SAM MACHKOVECH - 10/21/2021, 2:00 PM



# Task: Sentiment Classification (2)

- Explore different approaches to automatically classify a given movie review as Negative or Positive
  - Without looking at the star rating (obviously) 😊
- Investigate:
  - Symbolic approach – sentiment lexicon
  - Machine learning-based classifiers:
    - Naïve Bayes
    - Support Vector Machine
  - Input features – including words, part-of-speech tags, stemming
    - Embedding: doc2vec, optional BERT
  - Statistical significance testing

# Task: Sentiment Classification (3)

- Write code to implement the approaches
  - Python3 (iPython) code supplied: fill in the “TODO” gaps
  - <https://github.com/adianliusie/MLMI13>
- Data:
  - Part 1: 1000 positive + 1000 negative example texts
  - Part 2: 100,000 IMDB movie review database

# Timeline

- Practical session **Today**:
  - Develop the baseline – symbolic approach
  - [Start on baseline machine learning approach - Naïve Bayes]
- Practical session **Oct 28**:
  - Check on your NB baseline systems
  - Develop second baseline system (SVM)
  - [Start training doc2vec system]
- Practical session **Nov 11**:
  - Check systems
  - Analysis doc2vec systems
- **Noon on Dec 7: submit 2000 word report**
- Report deadline is fixed – progress through sessions can vary

# If unable to attend in person

- Demonstrators will be available for online support over Teams during sessions
  - To contact:
    - demonstrator chat - see bottom of MLMI13 Moodle page for the link
    - or Teams: kmk1001, al826
- Not essential that you attend all the practical sessions
  - they are there to provide support