University of St Andrews

Sentiment Analysis of Comments in MOOCs

Adriana Wilde

agw5@st-andrews.ac.uk

University of St Andrews

UNIVERSITY OF Southampton

Sentiment Analysis of Comments in a FutureLearn MOOC

Adriana Wilde and Jing Wang

agw5@st-andrews.ac.uk

jw10y15@ecs.soton.ac.uk

Sentiment (according to Google)

sentiment

/ˈsɛntɪm(ə)nt/

noun

noun: **sentiment**; plural noun: **sentiments**

1. a view or opinion that is held or expressed.
   "I agree with your sentiments regarding the road bridge"
   synonyms: view, point of view, way of thinking, feeling, attitude, thought, opinion, belief, idea
   "the comments in today's Daily Telegraph echo my own sentiments"
   - general feeling or opinion.
     "the council sought steps to control the rise of racist sentiment"
   - a feeling or emotion.
     "an intense sentiment of horror"
     synonyms: feeling, emotion
     "overpowered by an intense sentiment of horror, I leapt up"

Sentiment (according to Google)

sentiment

/ˈsɛntɪm(ə)nt/

noun

noun: **sentiment**; plural noun: **sentiments**

1. a view or opinion that is held or expressed.
   "I agree with your sentiments regarding the road bridge"
   synonyms: view, point of view, way of thinking, feeling, attitude, thought, opinion, belief, idea
   "the comments in today's Daily Telegraph echo my own sentiments"
   - general feeling or opinion.
     "the council sought steps to control the rise of racist sentiment"
   - a feeling or emotion.
     "an intense sentiment of horror"
     synonyms: feeling, emotion
     "overpowered by an intense sentiment of horror, I leapt up"

Sentiment analysis

- Sentiment analysis, or opinion mining, seeks to classify subjective feelings or emotions hidden in source texts using natural language processing techniques.

- The complexity of emotion recognition is reduced by using sentiment polarity with three categories: positive, neutral, and negative.

Why sentiment analysis of comments in MOOCs?

- Educators and learning designers can gauge at a glance whether the learning activities are well received or not ☹
- Sentiment analysis may offer additional insights for understanding attrition and learners taxonomy 😐
- We have the data!
  And the technology!
  We can do cool visualisations! ☺

Related work

- Shi Min Chua, Caroline Tagg, Mike Sharples and Bart Rienties, “Discussion Analytics: Identifying Conversations and Social Learners in FutureLearn MOOCs”

Future Learn

10,003 comments

- Devendra Singh Chaplot, Eunhee Rhim, and Jihie Kim, "Predicting Student Attrition in MOOCs using Sentiment Analysis and Neural Networks." AIED Workshops, 2015.

coursera

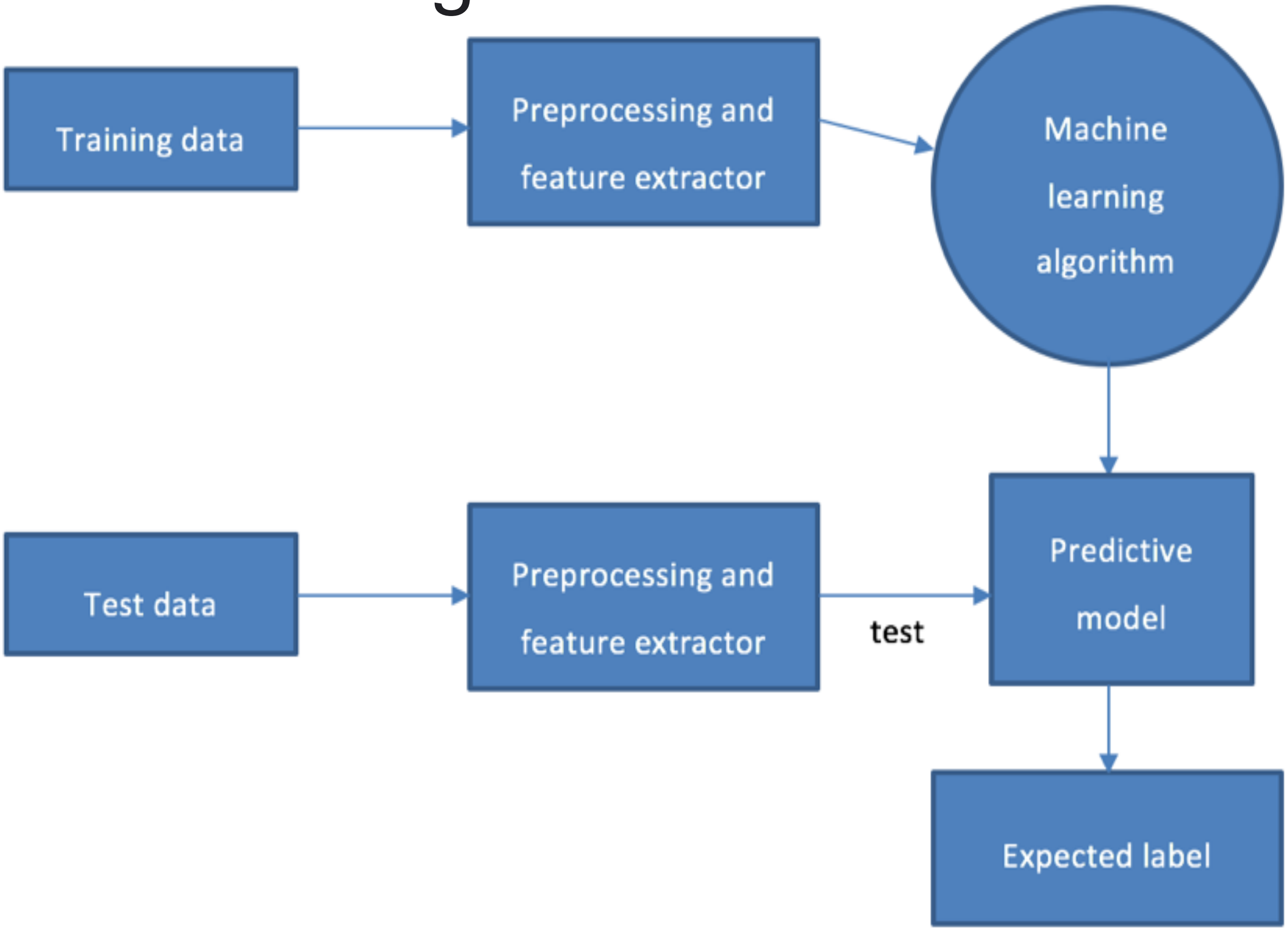
> 5,000 posts

- This!

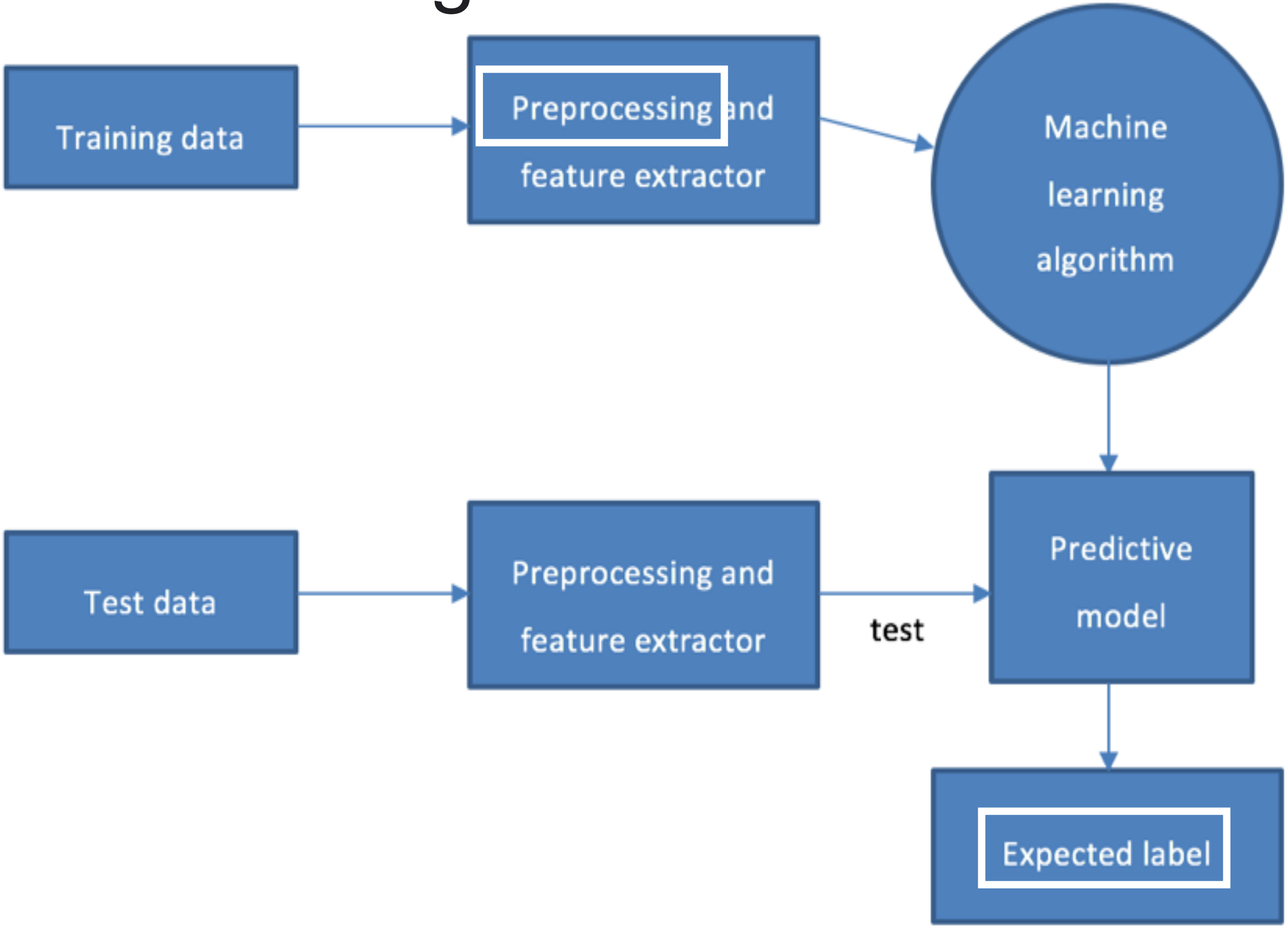
Future Learn

18,414 comments

Machine learning



Machine learning



Preprocessing

- Source texts are noisy! (e.g. scripts, HTML tags, and punctuation)

- Preprocessing reduces the noise which could affect the performance of the sentiment analysis algorithms and helps preparing the texts for classification

- Two approaches: Lexicon-based and machine learning methods.

Preprocessing

Tokenization: divides sentences into signal words. (e.g. “I like this course.” becomes a list [“I”, “like”, “this”, “course”, “.”])

Stemming: reduces all variations of words to their common root (e.g. “course” and “courses” have the same root)

Punctuation and stop word removal: minimizes the total number of words by removing segments that do not provide useful sentiment information (i.e. emoticons are kept!)

Term Weighting: for text representation in machine learning. Words, terms or phrases which could show the sentiment of the content are common features, presented in vector form. (e.g. TF-IDF)

TD-IDF

TF or term frequency represents the frequency of each term presented in a content;

IDF stands for inverse document frequency.

Together, TF-IDF is a weight matrix that describes the importance of particular terms with respect to the whole content. Each document or content could be shown in a form of numeric vector.

Dataset

- Comments in the University of Southampton Web Science course (all runs to date combined) delivered in FutureLearn

- There are **18414 records** in this dataset

Key	id	author_id	parent_id	step	text	timestamp	moderated	likes
Type	int	string	string	string	string	string	string	int

- For our work, we are only interested in the “text” field

Dataset structure

"id":	id of that comment
"author_id":	id of user who has written this comment
"parent_id":	id of comment with subsequent reply/ replies
"step":	instructed step of this course
"text":	**content of comment**
"timestamp":	time of the comment being posted
"moderated":	id of moderator
"likes":	number of users who liked this comment

Manual labelling sentiment polarity

Comments’ dataset was preprocessed, and a subset of it (2,000 comments) was manually labelled

The shortcomings of it being a “small” subset were addressed by using an online API for sentiment analysis (ParallelDots)

Manual labelling criteria for sentiment polarity

- Comments with positive words or attitudes are labelled as **positive** (e.g. “I like this course and it is very interesting”)

- Comments with negative words or attitudes are labelled as **negative** (e.g. “This video is not clear and hard to understand.”)

- Comment not fitting into any of the previous criteria are labelled as **neutral**. There are two cases:
  - the comment has <u>both</u> positive and negative wordings or attitudes. (e.g.. “This course is very helpful, while it is boring.”)
  - the comment has <u>no obvious</u> attitude or emotion wording. (e.g. “I use the web to contact my friends and do online shopping.”)

A note on emoticons

Emoticons	Example	score
Positive emoticons	:-) :) (: (-: :-D :D X-D XD xD <3	1
Negative emoticons	:-(:(:((:’-(>:-(	-1

Testing sentiment polarity on the API

1. “This course is interesting.”
2. “I do not like it due to the boring structure.”
3. “Though I’m not interested in this project, the professor is nice.”

ParallelDots API scores:

1. 99%
2. 4%
3. 85%

Testing sentiment polarity on the API

1. “This course is interesting.”
2. “I do not like it due to the boring structure.”
3. “Though I’m not interested in this project, the professor is nice.”

ParallelDots API scores:

1. 99%
2. 4%
3. 85%

Testing sentiment polarity on the API

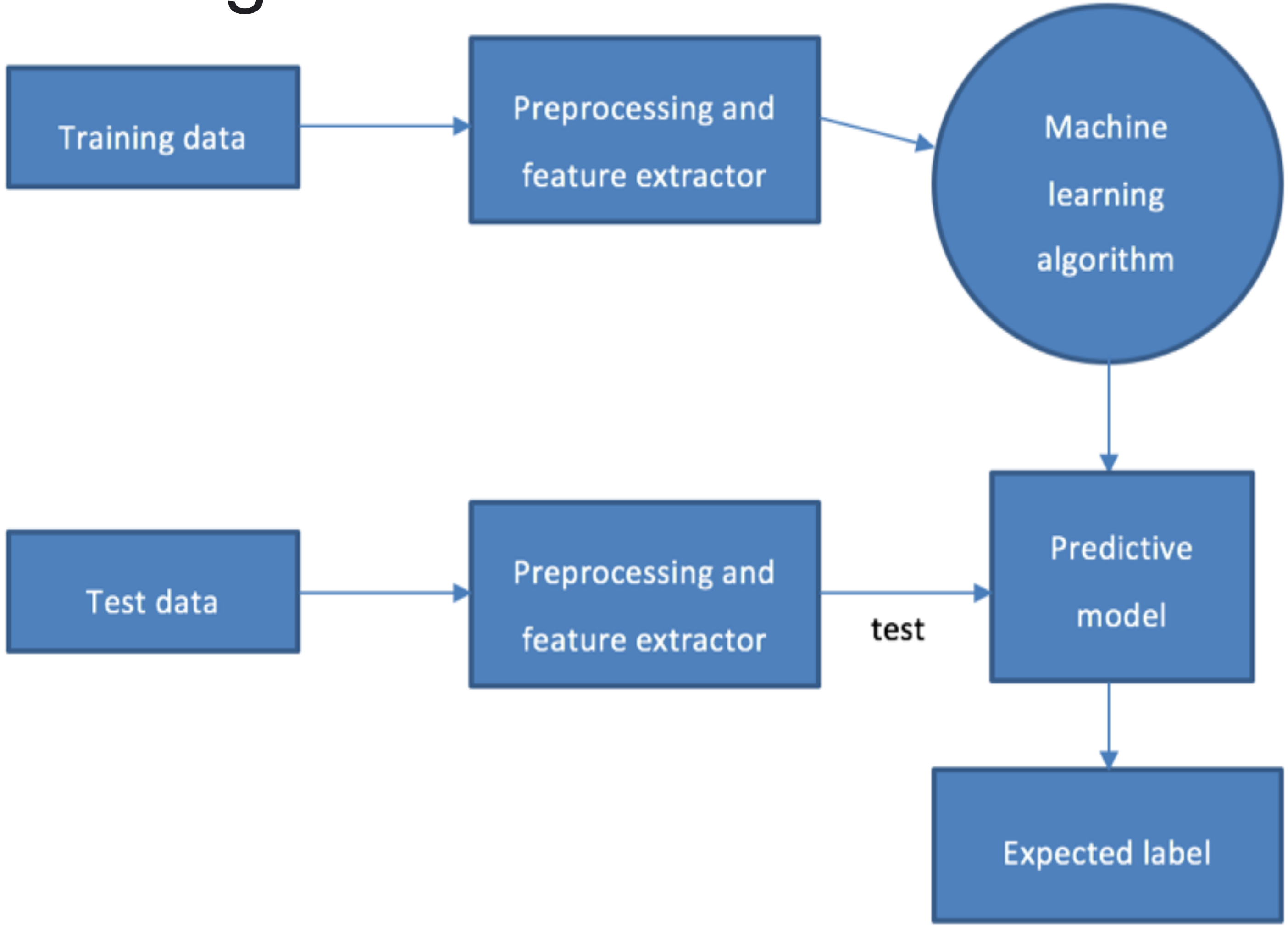
1. “This course is interesting.”
2. “I do not like it due to the boring structure.”
3. “Though I’m not interested in this project, the professor is nice.”

ParallelDots API scores:

1. 99%
2. 4%
3. 85%

Ambiguity makes it difficult to pre-define precise boundaries for the scale transformation. A nested iterative loop is applied to find the two appropriate boundaries with the range of 0-50 (the boundary between negative and neutral classes) as the outer loop and the range of 50-100 (the boundary between neutral and positive classes) as the inner loop. The resulting boundaries have produce transformed API scores that are more similar to the manual labels.

We didn't get to talk about features…

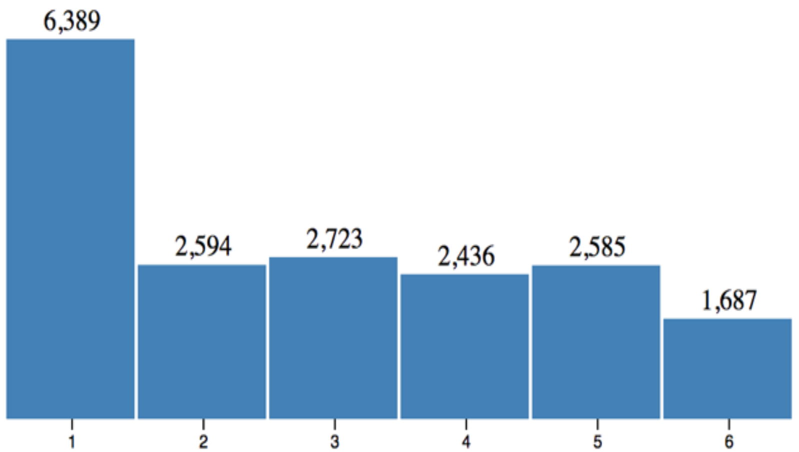


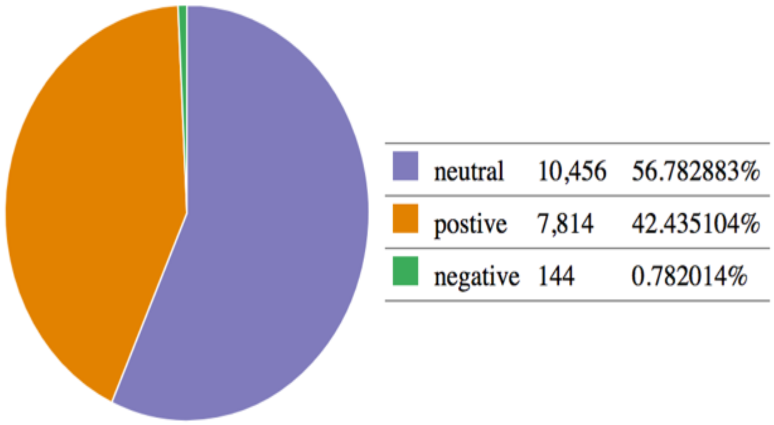
Model	Feature	MLP layer	alpha	validation data	
				accuracy%	f1-score%
1	TF-IDF matrix (ngram_range= (1,1)) with new tokenizer function	100	23	71.25	69.25
2	TF-IDF matrix (ngram_range= (1,2)) with new tokenizer function	100	22	72.08	70.08
3	TF-IDF matrix (ngram_range= (1,1))	100	20	69.58	67.65
4	TF-IDF matrix (ngram_range= (1,2))	90	21	73.75	71.7
5	TF-IDF matrix (ngram_range= (1,3))	90	21	73.33	71.3
6	TF-IDF matrix (ngram_range= (1,4))	90	20	72.08	70.08
7	TF-IDF (1,2) +Amount of words	100	21	73.75	71.7
8	TF-IDF (1,2) +positive words	100	23	74.17	72.12
9	TF-IDF (1,2) +negative words	100	21	73.75	71.7
10	TF-IDF (1,2) +positive words +likes	100	23	**74.6**	**72.52**
11	TF-IDF (1,2) +positive words +likes +reply	100	22	72.5	70.5

Comparison of various ML models

model	Methods	accuracy	F1-score
1	Naïve Bayes	72.08	69.86
2	SVM	70	68.06
3	Maximum Entropy	74.17	72.1
4	Neural Network	74.6	72.52
5	ensemble model	75	72.92

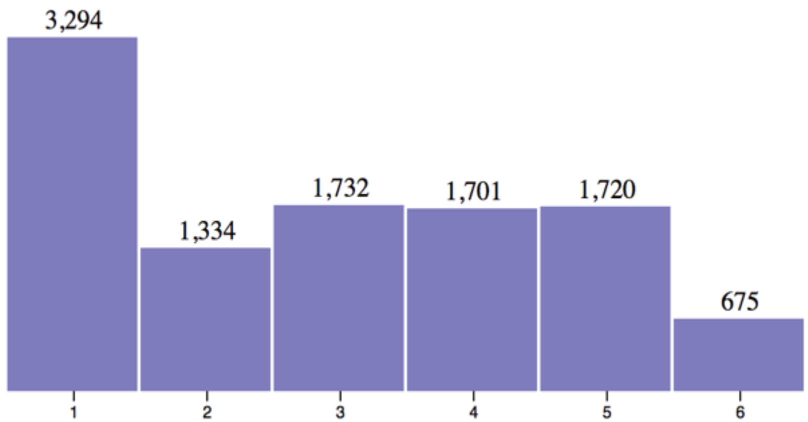
Visualisations

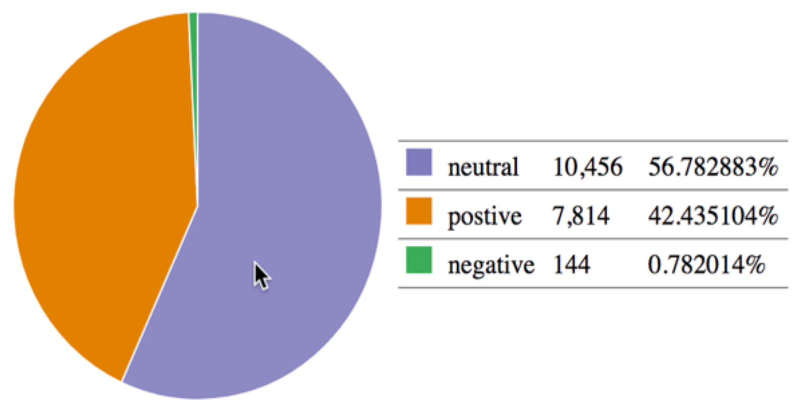




At first glance we can see there are not many negative comments, and there are many more comments on the first week than in any other individual week.

Visualisations

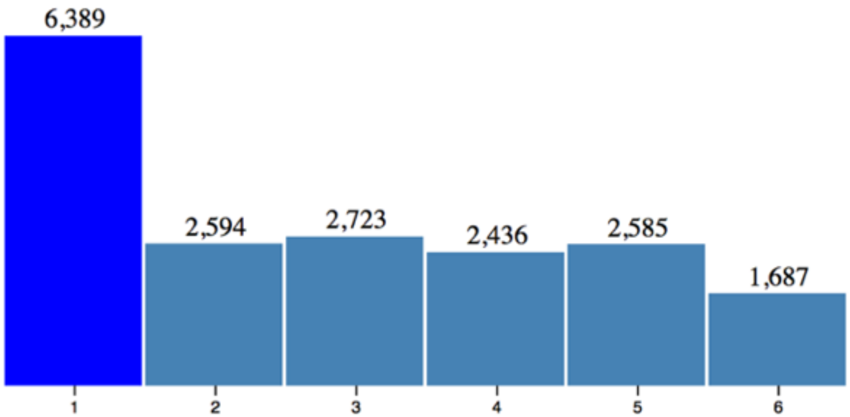


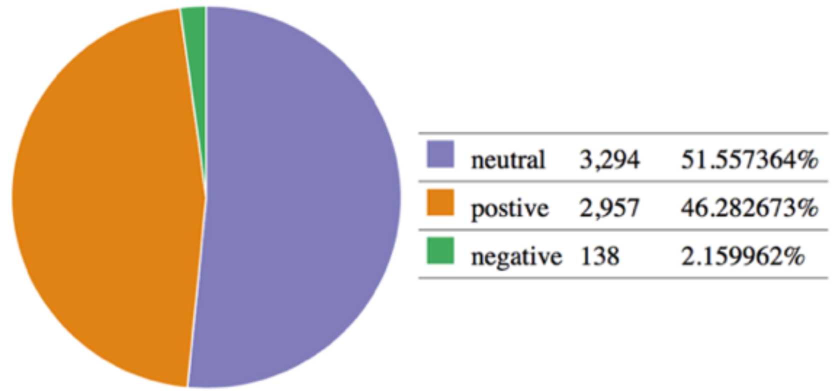


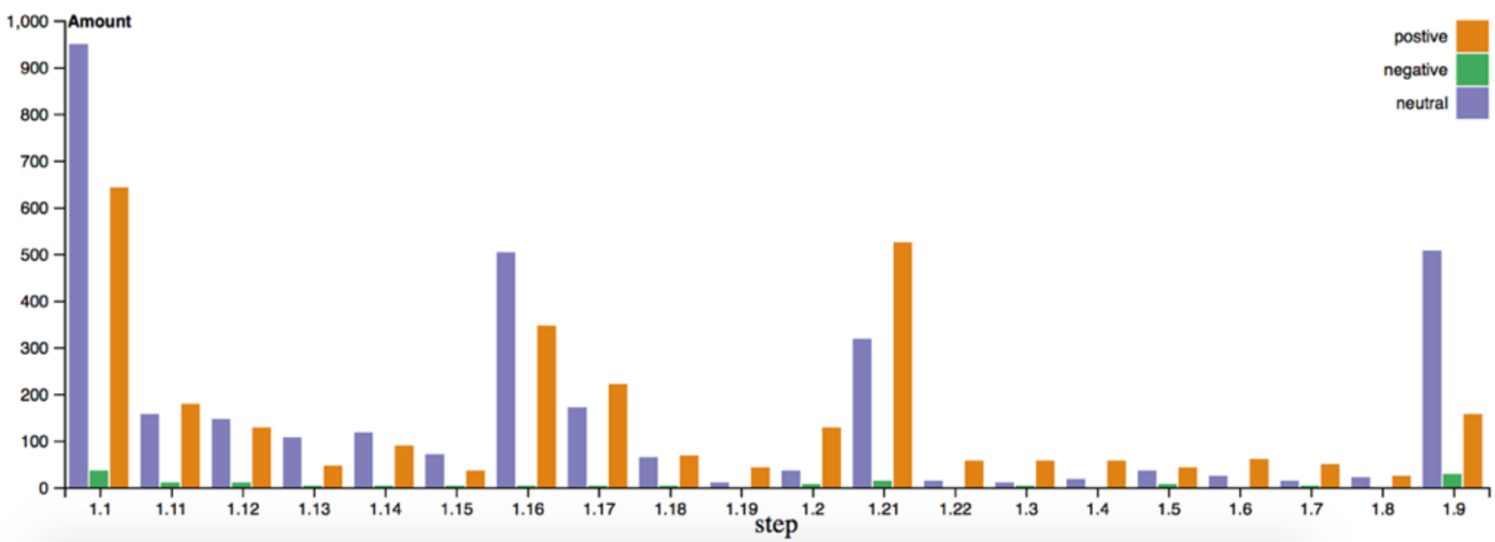
Hovering the mouse over the pie chart for the course, it gives the breakdown per week for the selected sentiment.

Selecting any given week, results on detailed information about the sentiment for each step in that week.

Visualisations







Conclusions and future work

- Great care needs to be put into the preprocessing phase as this could affect the performance of the ML algorithms
- For instance, the final performance of the best model is not perfect with the 75% accuracy, it still needs to be improved but this is not a trivial problem.
- Being able to visualise the sentiment of an ongoing course is valuable for educators and learning designers. (i.e. it could be a useful addition to the MOOC dashboard)