

# Fundamentals of Machine Learning

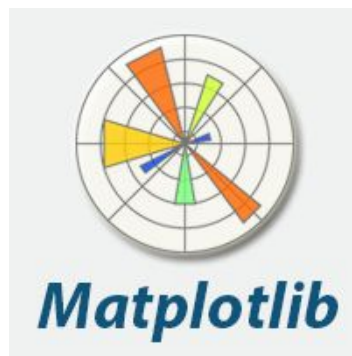
## Sentiment Analysis on IMDb Movie Reviews

---

Simone Murari

University of Verona  
A.Y. 2023-2024

## Libraries used




# Dataset: IMDb Movie Reviews

<https://paperswithcode.com/dataset/imdb-movie-reviews>

- Binary sentiment analysis dataset
- 50000 reviews from IMDb: 25000 labeled as *positive* and 25000 as *negative*
- Negative reviews: score  $\leq 4$  out of 10
- Positive reviews: score  $\geq 7$  out of 10

|   | review  | sentiment |
|---|---|-----------|
| 0 | I wonder how many MINI Cooper automobiles were... | positive  |
| 1 | Aparna Sen's 15 Park Avenue is a film about na... | positive  |
| 2 | Although it's definitely an enjoyable way to s... | positive  |
| 3 | Unlike what one reviewer said this is NOT a ri... | negative  |
| 4 | While it's true that the movie is somewhat int... | negative  |

Usage 



# Preprocessing and feature extraction

---

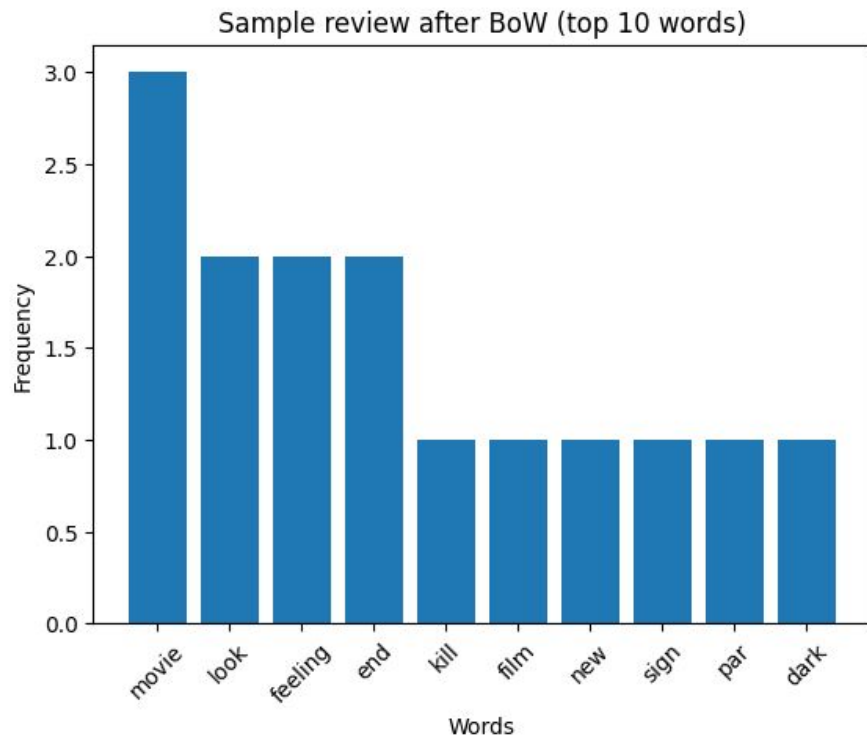
# Dataset sample

I love cheesy horror flicks. I don't care if the acting is sub-par or whether the monsters look corny. I liked this movie except for the bewildered feeling all the way from the beginning of the film to the very end. Look, I don't need a 10 page dissertation or a sign with big letters explaining a plot to me. But Dark Floors takes the "what is this movie about?" thing to a whole new (annoying) level. What IS this movie about?  
  
This isn't exceptionally scary or thrilling but if you have an hour and a half to kill and/or you want to end up feeling frustrated and confused, rent this winner.

# Dataset sample after regex

I love cheesy horror flicks. I don't care if the acting is sub-par or whether the monsters look corny. I liked this movie except for the bewildered feeling all the way from the beginning of the film to the very end. Look, I don't need a 10 page dissertation or a sign with big letters explaining a plot to me. But Dark Floors takes the "what is this movie about?" thing to a whole new (annoying) level. What IS this movie about? This isn't exceptionally scary or thrilling but if you have an hour and a half to kill and/or you want to end up feeling frustrated and confused, rent this winner.

# Dataset sample after Bag of Words



- To create the bag of words I used Sklearn CountVectorizer with the english stopwords from NLTK to remove unnecessary words and at the same time create the bag of words
- I keep only the 5000 most frequent words from all the reviews
- Features shape: 50000x5000

# Classification & results evaluation

---



# Qualitative results: correctly classified reviews sample

Review:

This is by far one of the best films that India has ever made. Following are the plus points of the film...Wonderful direction, cinematography and editing, the editing is very smooth and the timing of changeovers is excellent. Even though the film shows the life of Mumbai Policemen and their hardships, it never gets boring or sympathetic. Mind-blowing acting by lead actor Nana Patekar. One can surely hope that he gets nominated for the Best actor for the academy awards. Controlled violence. The violence is controlled and the film doesn't become a bloody mess. No stupid songs as in usual Indian movies.

Predicted sentiment: 1, Real sentiment: 1

Review:

In watching Enterprise for the first time, as we all no doubt do with all shows, I went into it with an open mind, enjoying about half of the past Star Trek efforts and disliking the other half. Enterprise has fallen short, but this episode "A Night In Sickbay" made me seriously question why I bother tivoing the shows from Monday night on Sci Fi. Masking some idea that it is one of those 'A Day In The Life' episodes, in which we learn about what makes certain characters operate as humanoids, the writers seemed to forget that this was supposed to be a starship vessel, not the Ricardoes and the Mertzes. A planet, especially one whose people had been offended previously by the Enterprise crew (for eating in public), was no place for a dog. As an animal lover myself, I would have never taken one of my pets into an environment that had proved in the past to be tense. But what made this episode even more ridiculous was the endless problem with all of these ST shows, constantly depicting things that are sacred and insulting to other cultures, as tho they are offering some insight into American religious zealots. The aliens were now offended when the dog urinated on a sacred tree, yet the aliens were quite capable of taking the dog urinating as an insult. Strange how the dog's urinating wasn't regarded as some form of worship. I wonder how that one got by the show's writers. From there, we are subjected to a captain who was misguided by his duties. In watching the episode, I found it very easy to forget that Bakula was supposed to be the ship's captain. He chose to sleep in sickbay and from there we are given more inanities of behavior (sigh) that we aren't supposed to understand and that causes us to furrow brows. The doctor non-chalantly clips long, hairy toenails and feeds them to hungry caged animals. Ewwwww! Then a white bat creature escapes. Oh, how is anyone supposed to sleep with all of this going on! Toenail clipping, for crying out loud! I was waiting for something that feeds on vomit to be presented. Then we were inexplicably given some idea that the captain was in love with T'pol, and that perhaps he was masking those feelings with his concern over the dog. Endless amounts of rubbish. "your dog is ill, so go have sex. You'll feel better." And of course, the captain had to apologize and we humans had to regard his apology as completely ridiculous, because we are so (everyone smile very sarcastically) narrow-minded! Saw depictions such as this endlessly droned out on TNG. Oddly enough, the only thing missing from this awful episode was that Trip person offering his smirks and downhome boyisms, tho Bakula seemed to be covering all of that with the silly dog. Oh, the dog survived, so now go play fetch.

Predicted sentiment: 0, Real sentiment: 0

# Qualitative results: misclassified reviews sample

Review:

If you're looking for a not-so-serious mob movie, with a female as the lead, you're in the right place. Pfeiffer has acted much better than this. You can see she has matured beyond this picture. When I first picked this movie up, I expected Pfeiffer was poorly miscast, however, she plays her mob wife role to the hilt. Not a bad performance from Baldwin, either. If you don't pay attention to the hair, you might enjoy this movie. But don't take it too seriously...

Predicted sentiment: 0, Real sentiment: 1

Review:

East Palace, West Palace reminded me somewhat of The Detective, with Frank Sinatra in the role of the cop, and William Windom is the boy. It's a progressive film for China, I guess, but it also perpetuates myths about the femininity of gay men: much is made of Chinese myths in which men take on female roles. The movie focuses on an effeminate man who wants desperately to be dominated and hurt by a macho guy. He cruises the park without fear--he hopes to be taken into the stationhouse by the officer. And that in fact happens. Then he tells the officer his entire life story while being subjected to mild torture: made to squat for a period of time, handcuffed, slapped. This is what the gay man wants, and, implicitly, the gay man is challenging the cop's self-image as a manly man. The story's about the gay man's life (which include flashbacks) are tolerable, but when he starts describing old Chinese myths and dramatic works, the movie becomes unbearable. It becomes a cry of pity for China's gays, who only want to fulfill a traditional role in Chinese society. Sorry, I can't relate.

Predicted sentiment: 1, Real sentiment: 0

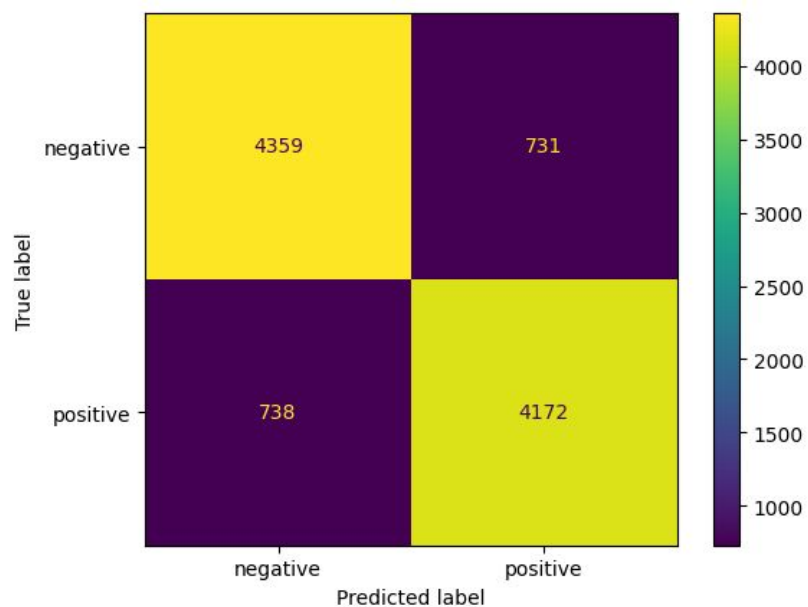
# Another misclassified review

Review:

I hadn't heard of this film until I read an article about it on the Unknown Movies website, which made me curious. As a cartoonist and illustrator myself, I'm an admirer of Richard Williams's work - I rate Ziggy's Gift as one of the finest Christmas specials of all time, and even though Who Framed Roger Rabbit stopped being one of my favourite films when I got past the age of sixteen, I still have the highest of regard for the amount of work, care and attention to detail that went into creating the visuals - but it seems the man has his faults, most notably a propensity for going over budget and over schedule, and this film is a testament of just how far wrong even a super-talented individual like Williams can go, given the right circumstances. Raggedy Ann and Andy is a strange confection that tries to be weird and experimental and off the wall within the confines of a children's cartoon. It tries also to be a musical. It tries to be a thousand and one other things as well - is it a freakout? Is it a mind-blower? Is it a paean to the innocence of childhood imagination? - until it finally collapses under the weight of its own limitless ambitions and aspirations. It's beautifully animated, for the most part, though the bland backgrounds could have used a little more attention, but even that doesn't count for much when you're confronted with the hallucinogenic absurdity that constitutes much of the 'action' here. There are a number of problems with the film, but let's start with Raggedy Ann and Andy themselves. They're the stars of the show, yet they have no personalities. Actually, we get the message that Andy is a wannabe tough guy ("I'm no girl's toy", he sings) and that Ann has a unique perspective on things because her owner, a little girl called Marcella, carries her upside down, but that's all we get to find out about these dull-as-mud characters because the overwhelming weirdness of this film kicks in not long after. I use the word "weirdness" advisedly, because some weird films can be hugely entertaining, but this is just flat-out strange. The toys and dolls in the playroom are supposed to be cute and lovable, but they're actually bizarre and disturbing. The two marionettes who do and say everything in sync are a prime example of this. But even they're relatively normal compared to the constantly sneezing pirate captain, whose moustache becomes erect and whose groin visibly swells when he first catches sight of a glamorous French doll. Yes, this is supposed to be a children's film! Then there's the music, none of which is memorable, and all of which is sung by actors who can't sing. And to add insult to injury, there's a lot of singing in this film. When Ann and Andy finally make it out of the playroom, the first thing they do is sing a LONG number in the woods about how scared they are, about how they'll always have each other and...yes, we get the message. This seems to go on forever, but at least it brings some semblance of normality back to the film. Not for long, though, because the Camel with the wrinkled knees leads us into a bizarre world where everything looks like it's made from worn and faded denim, and - bad enough that he's clearly a paranoid schizophrenic - he also starts hallucinating. But this is nothing compared to the scenes that follow. The Greedy, a living, breathing, belching, farting, constantly eating pool of taffy, is so trippy, creepy and ultimately disturbing, you'll hardly believe what you're seeing - this is as close to a drug-free psychedelic experience as I've ever seen on film. Then, after a l-o-n-g time spent with the Greedy, along comes the psychotic Sir Leonard Looney and his master King KooKoo, whose throne resembles a urinal. I can't believe I'm actually writing a capsule description of a real film here - I just had to rub my eyes and remind myself that I'm not blogging about an overwhelmingly whacked-out nightmare I had. Part of the sequence in Looney Land resembles one of the old Winsor McCay / Little Nemo cartoons, for no good reason other than somebody felt like doing it, probably. All this would be fine if there was some kind of rhyme or reason behind it, but there isn't. These scenes are just strange, and very, very long. Surrealism only works when there's a strong idea behind it, or takes place against some semblance of reality. But NOTHING in these scenes points towards any kind of reality. Take away this element, and you're left with pure self indulgence. As the final scenes unravel, even the animation begins to look less impressive (the pirate ship, ludicrously detailed, jerks about on the water in a manner that suggests some of the cels went missing during the production) and there's a non-event of an ending that simply suggests money ran out. Even at a meagre 86 minutes in length, the film feels like a never-ending ordeal, and it's understandable why it flopped on its original release. Animation buffs will probably scratch their heads and wonder just how Williams managed to flub this one so spectacularly, but he did, and there's nothing anyone can do about it.

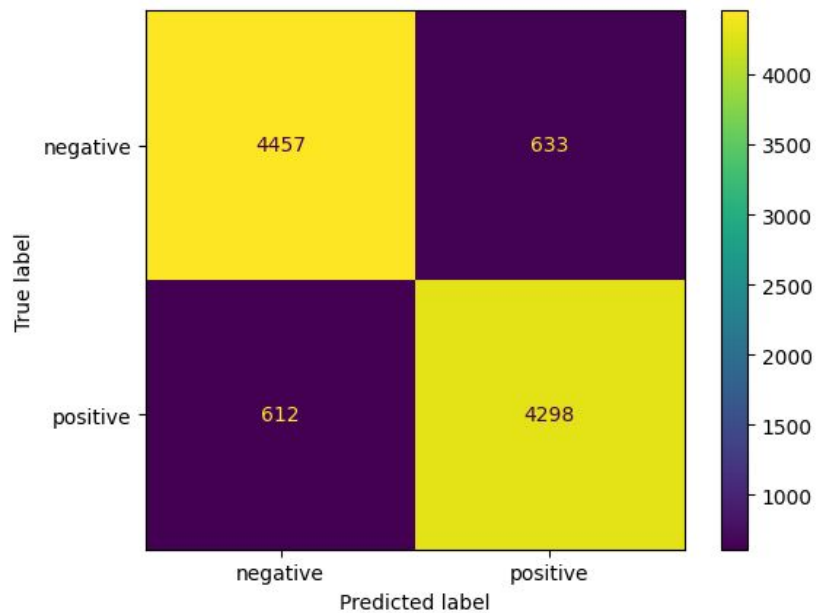
Predicted sentiment: 1, Real sentiment: 0

# Quantitative results: MultinomialNB



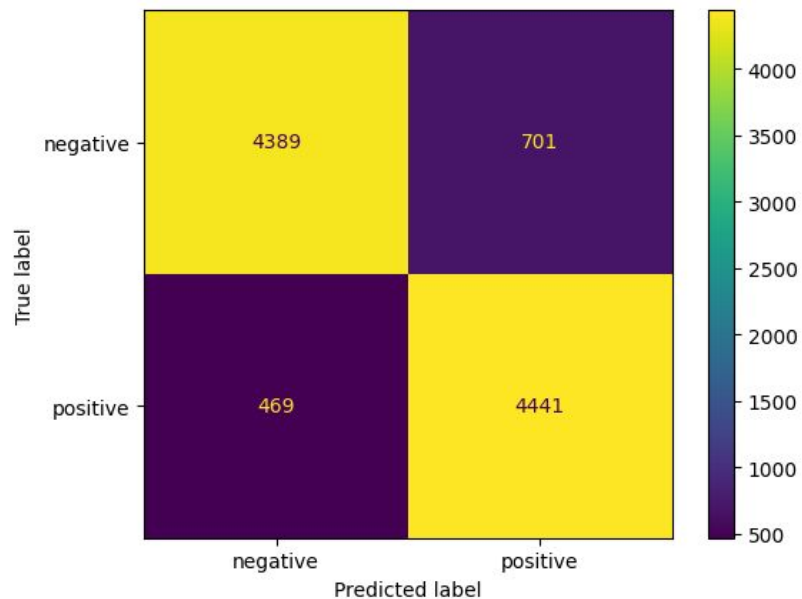
|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| negative     | 0.86      | 0.86   | 0.86     | 5090    |
| positive     | 0.85      | 0.85   | 0.85     | 4910    |
| accuracy     |           |        | 0.85     | 10000   |
| macro avg    | 0.85      | 0.85   | 0.85     | 10000   |
| weighted avg | 0.85      | 0.85   | 0.85     | 10000   |

# Quantitative results: Logistic Regression



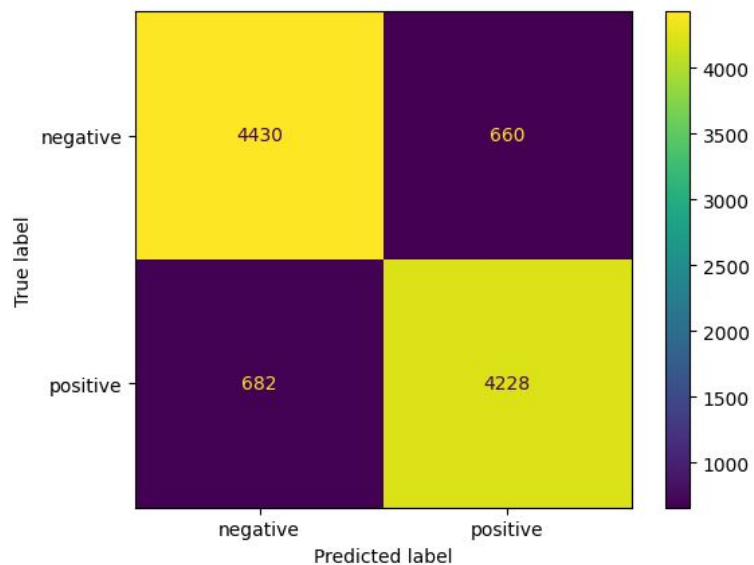
|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| negative     | 0.88      | 0.88   | 0.88     | 5090    |
| positive     | 0.87      | 0.88   | 0.87     | 4910    |
| accuracy     |           |        | 0.88     | 10000   |
| macro avg    | 0.88      | 0.88   | 0.88     | 10000   |
| weighted avg | 0.88      | 0.88   | 0.88     | 10000   |

## Quantitative results: SGDClassifier loss='log\_loss'



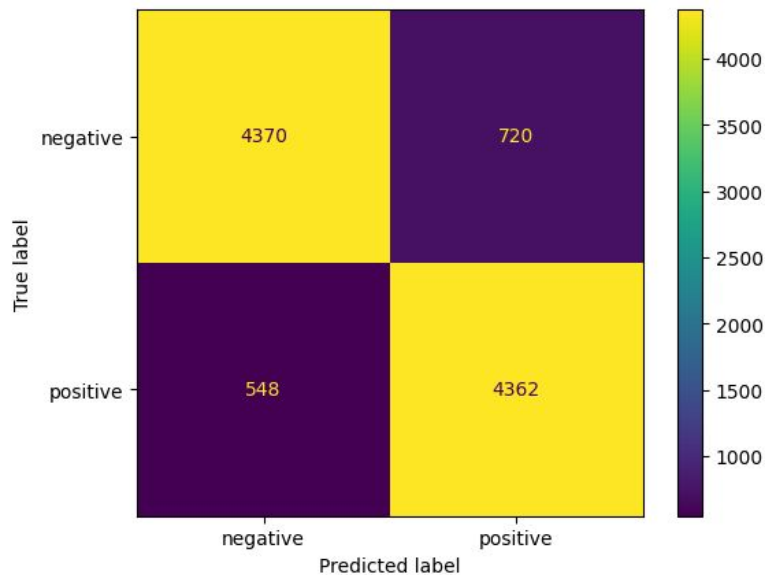
|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| negative     | 0.90      | 0.86   | 0.88     | 5090    |
| positive     | 0.86      | 0.90   | 0.88     | 4910    |
| accuracy     |           |        | 0.88     | 10000   |
| macro avg    | 0.88      | 0.88   | 0.88     | 10000   |
| weighted avg | 0.88      | 0.88   | 0.88     | 10000   |

# Quantitative results: LinearSVC



|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| negative     | 0.87      | 0.87   | 0.87     | 5090    |
| positive     | 0.86      | 0.86   | 0.86     | 4910    |
| accuracy     |           |        | 0.87     | 10000   |
| macro avg    | 0.87      | 0.87   | 0.87     | 10000   |
| weighted avg | 0.87      | 0.87   | 0.87     | 10000   |

## Quantitative results: SGDClassifier loss='hinge'



|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| negative     | 0.89      | 0.86   | 0.87     | 5090    |
| positive     | 0.86      | 0.89   | 0.87     | 4910    |
| accuracy     |           |        | 0.87     | 10000   |
| macro avg    | 0.87      | 0.87   | 0.87     | 10000   |
| weighted avg | 0.87      | 0.87   | 0.87     | 10000   |



Thank you!

---