

## Dataset (I took)

# IMDB Dataset of 50K Movie Reviews

Large Movie Review Dataset

Data Card Code (1770) Discussion (9) Suggestions (1)

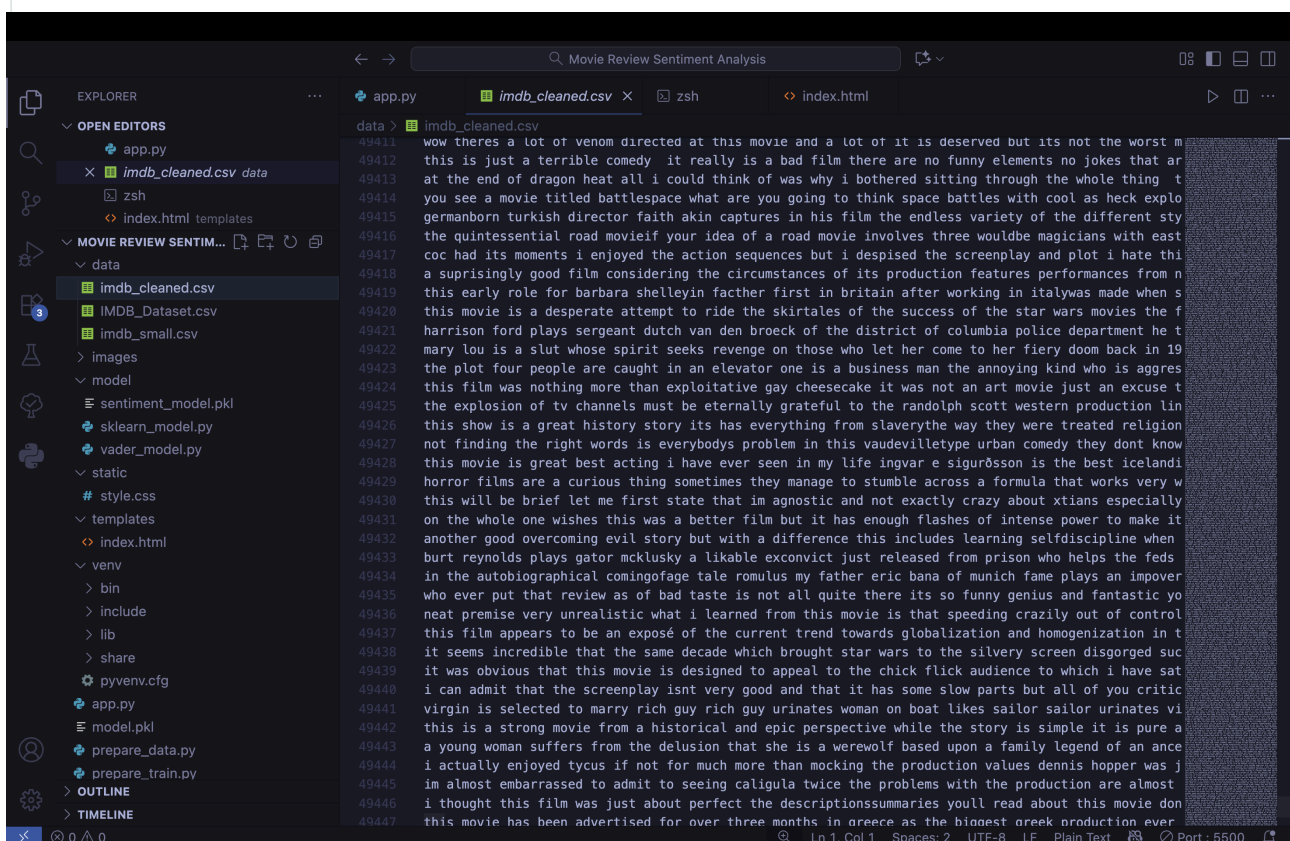
## About Dataset

IMDB dataset having 50K movie reviews for natural language processing or Text analytics.

This is a dataset for binary sentiment classification containing substantially more data than previous benchmark datasets. We provide a set of 25,000 highly polar movie reviews for training and 25,000 for testing. So, predict the number of positive and negative reviews using either classification or deep learning algorithms.

For more dataset information, please go through the following link,

<http://ai.stanford.edu/~amaas/data/sentiment/>



The screenshot shows a code editor with a file explorer on the left and a code editor on the right. The file explorer shows a project structure with files like `app.py`, `imdb_cleaned.csv`, `IMDB_Dataset.csv`, `imdb_small.csv`, `images`, `model`, `sentiment_model.pkl`, `sklearn_model.py`, `vader_model.py`, `static`, `style.css`, `templates`, `index.html`, `venv`, `bin`, `include`, `lib`, `share`, `pyvenv.cfg`, `prepare_data.py`, and `prepare_train.py`. The code editor shows a file named `imdb_cleaned.csv` with a list of movie reviews. The reviews are numbered from 49411 to 49447. The reviews are as follows:

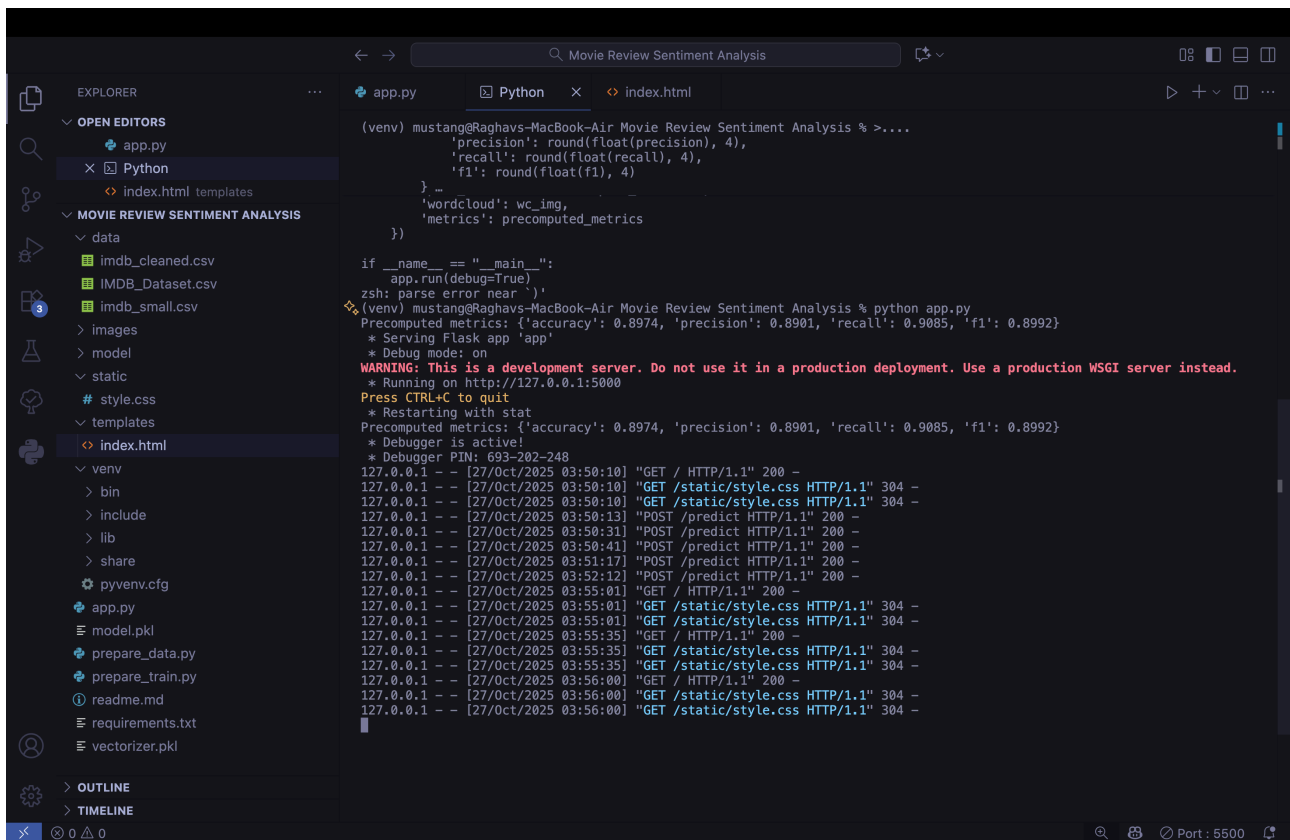
- 49411: wow theres a lot of venom directed at this movie and a lot of it is deserved but its not the worst m
- 49412: this is just a terrible comedy it really is a bad film there are no funny elements no jokes that ar
- 49413: at the end of dragon heat all i could think of was why i bothered sitting through the whole thing t
- 49414: you see a movie titled battlespace what are you going to think space battles with cool as heck explo
- 49415: germanborn turkish director faith akin captures in his film the endless variety of the different sty
- 49416: the quintessential road movieif your idea of a road movie involves three wouldbe magicians with east
- 49417: coc had its moments i enjoyed the action sequences but i despised the screenplay and plot i hate thi
- 49418: a suprisingly good film considering the circumstances of its production features performances from n
- 49419: this early role for barbara shelleyin facther first in britain after working in italywas made when s
- 49420: this movie is a desperate attempt to ride the skirtales of the success of the star wars movies the f
- 49421: harrison ford plays sergeant dutch van den broeck of the district of columbia police department he t
- 49422: mary lou is a slut whose spirit seeks revenge on those who let her come to her fiery doom back in 19
- 49423: the plot four people are caught in an elevator one is a business man the annoying kind who is aggres
- 49424: this film was nothing more than exploitative gay cheesecake it was not an art movie just an excuse t
- 49425: the explosion of tv channels must be eternally grateful to the randolph scott western production lin
- 49426: this show is a great history story its has everything from slaverythe way they were treated religion
- 49427: not finding the right words is everybodys problem in this vaudevilletype urban comedy they dont know
- 49428: this movie is great best acting i have ever seen in my life ingvar e sigurðsson is the best icelandi
- 49429: horror films are a curious thing sometimes they manage to stumble across a formula that works very w
- 49430: this will be brief let me first state that im agnostic and not exactly crazy about xtians especially
- 49431: on the whole one wishes this was a better film but it has enough flashes of intense power to make it
- 49432: another good overcoming evil story but with a difference this includes learning selfdiscipline when
- 49433: burt reynolds plays gator mcklusky a likable exconvict just released from prison who helps the feds
- 49434: in the autobiographical comingofage tale romulus my father eric bana of munich fame plays an impover
- 49435: who ever put that review as of bad taste is not all quite there its so funny genius and fantastic yo
- 49436: neat premise very unrealistic what i learned from this movie is that speeding crazily out of control
- 49437: this film appears to be an exposé of the current trend towards globalization and homogenization in t
- 49438: it seems incredible that the same decade which brought star wars to the silvery screen disgorged suc
- 49439: it was obvious that this movie is designed to appeal to the chick flick audience to which i have sat
- 49440: i can admit that the screenplay isnt very good and that it has some slow parts but all of you critic
- 49441: virgin is selected to marry rich guy rich guy urinates woman on boat likes sailor sailor urinates vi
- 49442: this is a strong movie from a historical and epic perspective while the story is simple it is pure a
- 49443: a young woman suffers from the delusion that she is a werewolf based upon a family legend of an ance
- 49444: i actually enjoyed tykus if not for much more than mocking the production values dennis hopper was j
- 49445: im almost embarrassed to admit to seeing caligula twice the problems with the production are almost
- 49446: i thought this film was just about perfect the descriptionsummaries youll read about this movie don
- 49447: this movie has been advertised for over three months in greece as the biggest greek production ever

# Performance analysis

After training and testing 2k dataset

Metric	Value (approx)
Accuracy	80–83%
Precision	0.80–0.83
Recall	0.78–0.82
F1-score	0.79–0.82

After training and testing 50k dataset



```
(venv) mustang@Raghavs-MacBook-Air Movie Review Sentiment Analysis % python app.py
Precomputed metrics: {'accuracy': 0.8974, 'precision': 0.8901, 'recall': 0.9085, 'f1': 0.8992}
* Serving Flask app 'app'
* Debug mode: on
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
* Running on http://127.0.0.1:5000
Press CTRL+C to quit
* Restarting with stat
Precomputed metrics: {'accuracy': 0.8974, 'precision': 0.8901, 'recall': 0.9085, 'f1': 0.8992}
* Debugger is active!
* Debugger PIN: 693-202-248
127.0.0.1 - - [27/Oct/2025 03:50:10] "GET / HTTP/1.1" 200 -
127.0.0.1 - - [27/Oct/2025 03:50:10] "GET /static/style.css HTTP/1.1" 304 -
127.0.0.1 - - [27/Oct/2025 03:50:10] "GET /static/style.css HTTP/1.1" 304 -
127.0.0.1 - - [27/Oct/2025 03:50:13] "POST /predict HTTP/1.1" 200 -
127.0.0.1 - - [27/Oct/2025 03:50:31] "POST /predict HTTP/1.1" 200 -
127.0.0.1 - - [27/Oct/2025 03:50:41] "POST /predict HTTP/1.1" 200 -
127.0.0.1 - - [27/Oct/2025 03:51:17] "POST /predict HTTP/1.1" 200 -
127.0.0.1 - - [27/Oct/2025 03:52:12] "POST /predict HTTP/1.1" 200 -
127.0.0.1 - - [27/Oct/2025 03:55:01] "GET / HTTP/1.1" 200 -
127.0.0.1 - - [27/Oct/2025 03:55:01] "GET /static/style.css HTTP/1.1" 304 -
127.0.0.1 - - [27/Oct/2025 03:55:01] "GET /static/style.css HTTP/1.1" 304 -
127.0.0.1 - - [27/Oct/2025 03:55:35] "GET / HTTP/1.1" 200 -
127.0.0.1 - - [27/Oct/2025 03:55:35] "GET /static/style.css HTTP/1.1" 304 -
127.0.0.1 - - [27/Oct/2025 03:55:35] "GET /static/style.css HTTP/1.1" 304 -
127.0.0.1 - - [27/Oct/2025 03:56:00] "GET / HTTP/1.1" 200 -
127.0.0.1 - - [27/Oct/2025 03:56:00] "GET /static/style.css HTTP/1.1" 304 -
127.0.0.1 - - [27/Oct/2025 03:56:00] "GET /static/style.css HTTP/1.1" 304 -
```

## Performance metrics

```
(venv) mustang@Raghavs-MacBook-Air Movie Review Sentiment Analysis % python app.py
Precomputed metrics: {'accuracy': 0.8974, 'precision': 0.8901, 'recall': 0.9085, 'f1': 0.8992}
* Serving Flask app 'app'
* Debug mode: on
```