

# Implementasi passive aggressive classifier untuk mendeksi berita fake atau real

Muhammad Dafa Wardana<sup>1\*</sup>, Muhammad Faisal<sup>2</sup>

<sup>1,2</sup> Program Studi Teknik Informatika, Universitas Islam Negeri Maulana Malik Ibrahim Malang

e-mail: \*[200605110045@student.uin-malang.ac.id](mailto:200605110045@student.uin-malang.ac.id)<sup>1</sup>, [mfaisal@ti.uin-malang.ac.id](mailto:mfaisal@ti.uin-malang.ac.id)<sup>2</sup>

## Kata Kunci:

machine learning; passive aggressive classifier; algoritma klasifikasi; klasifikasi berita; deteksi berita palsu

## Keywords:

machine learning; passive aggressive classifier; fake news detection; news classification

dengan cepat dan akurat.

## ABSTRAK

Penyebaran berita palsu di era informasi digital mendorong kebutuhan akan algoritma deteksi yang efisien. Dalam penelitian ini, diajukan penggunaan Passive Aggressive Classifier untuk mengenali berita palsu dan real. Passive Aggressive Classifier adalah algoritma pembelajaran mesin adaptif yang bisa mengklasifikasikan data online dengan menyesuaikan diri pada pola baru. Dataset terdiri dari berita palsu dan real digunakan dalam pelatihan model deteksi. Proses melibatkan vektorisasi teks via TF-IDF dan pelatihan model menggunakan algoritma Passive Aggressive Classifier. Evaluasi melibatkan metrik standar seperti akurasi, presisi, recall, dan f1-score. Hasilnya tunjukkan bahwa model mampu mendeksi berita palsu secara akurat. Implikasinya, pendekatan ini bisa efektif untuk mengatasi penyebaran berita palsu

## ABSTRACT

The proliferation of fake news or hoaxes has become an increasingly pressing issue in today's digital information era. To address this problem, an effective and efficient detection algorithm is necessary. In this study, we propose the implementation of the Passive Aggressive Classifier to identify fake or real news. The Passive Aggressive Classifier is an adaptive machine learning algorithm capable of classifying data online by adjusting to emerging patterns. A dataset consisting of fake and real news is used to train the detection model. The implementation process involves text vectorization using the TF-IDF scheme and model training using the Passive Aggressive Classifier algorithm. The model's performance is evaluated using standard metrics such as accuracy, precision, recall, and F1-score. The research findings demonstrate that the implementation of the Passive Aggressive Classifier is capable of accurately detecting fake news. Consequently, this approach can be employed as an effective solution to swiftly and accurately counteract the spread of fake news.

## Pendahuluan

Pada era informasi yang semakin maju, penyebaran berita dan informasi telah menjadi sangat luas dan cepat melalui platform digital. Dalam konteks pendidikan, dampak perkembangan teknologi tercermin dalam penggunaan dokumen, salah satunya yaitu dokumen digital. Dokumen digital telah muncul sebagai hasil evolusi teknologi di era digital saat ini. Dalam bidang ini, informasi yang terkait di dalam jurnal



This is an open access article under the [CC BY-NC-SA](#) license.

Copyright © 2023 by Author. Published by Universitas Islam Negeri Maulana Malik Ibrahim Malang.

online menjadi aspek yang signifikan dan sangat diminati oleh berbagai kalangan. Jurnal online menjadi format yang sangat dibutuhkan dalam berbagai sektor. Jurnal memiliki berbagai bentuk terbitan dan transmisi, mengangkat berita dan hasil-hasil penelitian di berbagai bidang, termasuk media (Pratama et al., 2019). Namun, bersamaan dengan penyebaran informasi yang cepat, juga muncul tantangan dalam membedakan berita yang benar dari berita palsu (Rahmatullah, 2018). Berita juga dikenal sebagai informasi, yang merujuk pada pengetahuan yang dihasilkan dari proses pengolahan data yang saling terhubung menjadi suatu kesimpulan. Beberapa kumpulan data dapat dikategorikan sebagai informasi apabila dari sejumlah data tersebut sudah mampu menarik suatu kesimpulan yang bermakna. Kita juga diminta untuk melakukan pemeriksaan yang cermat terhadap kebenaran suatu berita atau informasi. Kita juga dapat menerapkan sebuah sistem untuk memeriksa kebenaran informasi tersebut (Hanani, 2008). Berita palsu atau hoaks dapat menyebabkan kerugian besar, termasuk menyebabkan ketidakpercayaan masyarakat, mengganggu stabilitas politik, dan merugikan reputasi individu atau organisasi (Chumairoh, 2020).

Dalam konteks penelitian ini, mekanisme operasinya mirip dengan cara pemeriksaan plagiarisme, di mana penelitian mengadopsi algoritma Manber-Winnowing sebagai solusi untuk mengenali tindakan plagiarisme. Teknik ini berlandaskan pada pemrosesan urutan karakter dengan menghilangkan unsur yang tidak relevan, kemudian menciptakan representasi khas yang memungkinkan perbandingan antara dokumen-dokumen. Proses ini melibatkan beberapa tahap, termasuk segmentasi teks menjadi fragmen yang lebih kecil dan penyingkiran elemen yang tak penting, seperti spasi atau karakter spesifik (Faisal et al., 2020).

Untuk mengatasi masalah ini, pendekatan yang menggunakan teknik-teknik kecerdasan buatan dan pembelajaran mesin telah dikembangkan. Salah satu metode yang dapat digunakan adalah Passive Aggressive Classifier (PAC), sebuah algoritma pembelajaran mesin yang dapat digunakan untuk mendeteksi berita palsu atau real (Popek et al., 2014). PAC adalah algoritma yang adaptif dan efisien dalam mengklasifikasikan data dengan cepat (Rathi et al., 2018).

Dalam penelitian ini, kami bertujuan untuk mengimplementasikan Passive Aggressive Classifier untuk mendeteksi berita palsu atau real. Kami akan menggunakan teknik pemrosesan bahasa alami (natural language processing) untuk menganalisis teks berita dan mengekstrak fitur-fitur penting yang dapat membedakan antara berita palsu dan real (Mandical et al., 2020). Selanjutnya, kami akan melatih model PAC menggunakan data latih yang telah diberi label dan menguji keakuratannya dengan menggunakan data uji yang belum pernah dilihat sebelumnya (Imbwaga et al., 2022).

Diharapkan bahwa implementasi PAC dalam mendeteksi berita palsu atau real ini dapat memberikan kontribusi dalam mengatasi masalah penyebaran berita palsu dan membantu masyarakat untuk memperoleh informasi yang benar dan dapat dipercaya (Riadi Silitonga, 2019).

## Pembahasan

Pembahasan terkait paragraf di bawah ini akan membahas mengenai dua komponen penting dalam implementasi metode "TfidfVectorizer" dan "Passive Aggressive Classifier."

### TfidfVectorizer

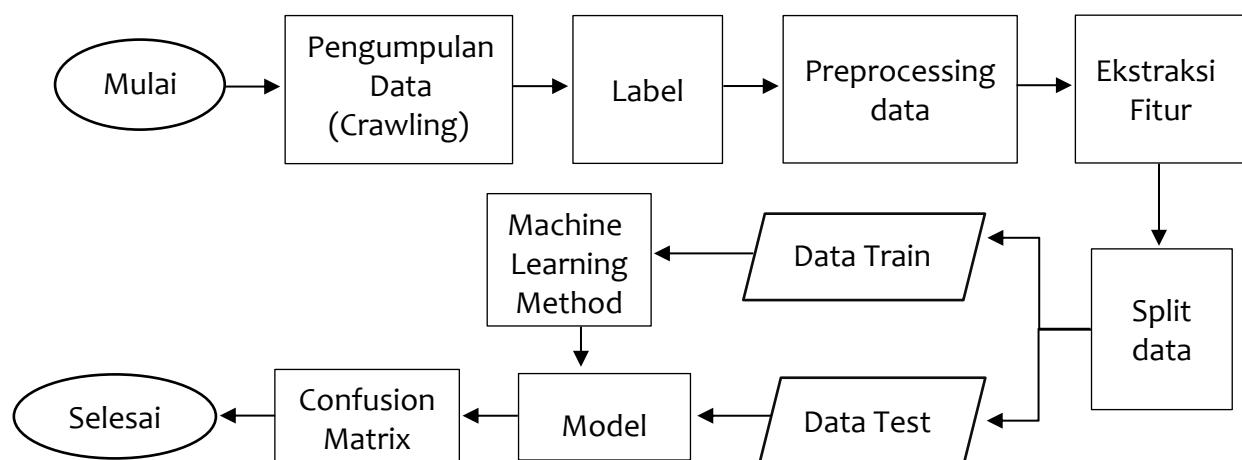
TfidfVectorizer adalah sebuah metode dalam pemrosesan bahasa alami (Natural Language Processing/NLP) yang digunakan untuk mengubah teks menjadi representasi numerik yang dapat digunakan oleh algoritma machine learning (Villagracia Octaviano, 2021). TfidfVectorizer menggabungkan dua konsep utama, yaitu Term Frequency (TF) dan Inverse Document Frequency (IDF) (Abdulrahman & Baykara, 2020). Term Frequency (TF) mengukur sejauh mana sebuah kata muncul dalam sebuah dokumen. TF mengasumsikan bahwa kata-kata yang sering muncul dalam sebuah dokumen memiliki tingkat penting yang tinggi (Bounabi et al., 2019). Inverse Document Frequency (IDF) mengukur sejauh mana sebuah kata umum atau jarang muncul di seluruh koleksi dokumen. Kata-kata yang jarang muncul di koleksi dokumen memiliki tingkat penting yang lebih tinggi dibandingkan dengan kata-kata yang sering muncul (Tripathy et al., 2015). TfidfVectorizer menggabungkan kedua konsep ini untuk memberikan representasi numerik dari teks yang memperhitungkan pentingnya sebuah kata dalam suatu dokumen dan dalam keseluruhan koleksi dokumen.

### Passive Aggressive Classifier

Passive Aggressive Classifier adalah algoritma machine learning yang digunakan untuk pemodelan klasifikasi pada data yang berubah seiring waktu. Algoritma ini digunakan dalam konteks "online learning", di mana model diperbarui secara iteratif saat data baru masuk (Ahmed et al., 2022). PassiveAggressiveClassifier merupakan algoritma yang adaptif dan efisien dalam menangani masalah klasifikasi biner dan multi-kelas. Ia memanfaatkan konsep "passive" dan "aggressive" untuk melakukan pembaruan model saat menghadapi data baru (University of Peradeniya. Department of Electrical and Electronics Engineering et al., 2019)

### Alur Penelitian

Alur dari penelitian ini dapat dilihat pada gambar di bawah ini :



## **Pengumpulan Data (Crawling)**

Proses pengumpulan data berita melibatkan penggunaan perangkat lunak otomatis, disebut "crawler" atau "spider," untuk mengambil informasi dari berbagai sumber berita di internet. Langkah-langkahnya meliputi inisiasi crawler, penentuan sumber berita, ekstraksi data berdasarkan kata kunci atau kategori tertentu, dan penyimpanan hasil dalam format yang sesuai. Dalam beberapa kasus, perlu mengatasi hambatan keamanan pada situs web tertentu. Setelah data terkumpul, pemfilteran dan analisis data dapat dilakukan untuk memahami tren berita atau sentimen. Proses ini memfasilitasi pengumpulan data besar-besaran untuk analisis lebih lanjut tanpa melibatkan intervensi manual.

## **Labeling**

Labeling adalah memberi tanda pada data, memisahkan berita palsu dan real. Dalam konteks "Implementasi Passive Aggressive Classifier untuk Mendeteksi Berita Palsu atau Real," labeling menunjukkan status kebenaran berita dalam dataset. Ini membantu algoritma seperti Passive Aggressive Classifier belajar pola dan memprediksi kebenaran berita baru. Label diberikan setelah penilaian manusia, memungkinkan model belajar dan meningkatkan klasifikasi berita di masa depan.

## **Preprocessing**

Preprocesssing atau Pra-pemrosesan pada "Implementasi Passive Aggressive Classifier untuk Mendeteksi Berita Tidak Benar atau Benar" adalah rangkaian langkah yang dilakukan sebelum menerapkan model klasifikasi. Ini melibatkan pembersihan, normalisasi, dan transformasi data teks. Tujuannya adalah untuk memastikan kualitas data sebelum analisis lebih lanjut. Ini termasuk menghilangkan karakter tidak diperlukan, mengubah teks menjadi format seragam, menghapus kata-kata umum, memecah teks menjadi unit-unit kecil, dan mengonversi teks menjadi bentuk numerik dengan metode seperti TF-IDF. Hal ini mempersiapkan data untuk model klasifikasi dan memungkinkan analisis lebih fokus dan akurat.

## **TF-IDF (Term Frequency-Inverse Document Frequency)**

Dalam judul "Implementasi Passive Aggressive Classifier Untuk Mendeteksi Berita Palsu atau Real," istilah "TF-IDF" mengacu pada metode pengolahan teks yang digunakan dalam analisis dan pengklasifikasian teks. "TF-IDF" adalah singkatan dari teknik yang disebut "Term Frequency-Inverse Document Frequency."

## **Split Data**

Split data berjutuan untuk Tujuannya adalah untuk membagi dataset menjadi dua bagian: data pelatihan (train) dan data pengujian (test). Pada source code yang dibawah, proses split data terjadi dengan menggunakan fungsi `train_test_split` dari pustaka `sklearn.model_selection`.

**Gambar 1.** Split dataset

```
#DataFlair - Split the dataset
x_train,x_test,y_train,y_test=train_test_split(df['text'], labels, test_size=0.2, random_state=7)
```

Data pelatihan (`x_train` dan `y_train`) digunakan untuk melatih model, sementara data pengujian (`x_test` dan `y_test`) digunakan untuk menguji performa model pada data yang belum pernah dilihat sebelumnya. Dalam kode tersebut, pengaturan `test_size=0.2` mengindikasikan bahwa data pengujian akan terdiri dari 20% dari total data. Penggunaan `random_state=7` bertujuan untuk memastikan hasil pembagian data yang konsisten jika kode dijalankan ulang.

## Hasil dan Pembahasan

Dalam implementasi model "Passive Aggressive Classifier" untuk mendeteksi berita palsu atau nyata, metode "TfidfVectorizer" memainkan peran penting dalam mengubah teks menjadi representasi numerik. Metode ini menggabungkan konsep Term Frequency (TF) dan Inverse Document Frequency (IDF). TF mengukur seberapa sering kata muncul dalam sebuah dokumen, sedangkan IDF mengukur keunikan kata dalam seluruh koleksi dokumen.

Sementara itu, algoritma "Passive Aggressive Classifier" digunakan sebagai metode pembelajaran mesin. Ini memiliki adaptabilitas terhadap perubahan data seiring waktu dalam paradigma "online learning." Algoritma ini memanfaatkan pendekatan "passive" dan "aggressive" untuk memperbarui model saat menghadapi data baru.

### Gambar 2. Hasil Pengujian

```
Out[8]: array([[590,  48],
   [ 45, 584]], dtype=int64)
```

1. Di baris pertama, kolom pertama (590) adalah True Negative (TN), yang menunjukkan jumlah berita yang benar-benar terdeteksi sebagai berita palsu.
2. Di baris pertama, kolom kedua (48) adalah False Positive (FP), yang menunjukkan jumlah berita yang sebenarnya benar tetapi salah terdeteksi sebagai berita palsu.
3. Di baris kedua, kolom pertama (45) adalah False Negative (FN), yang menunjukkan jumlah berita yang sebenarnya palsu tetapi salah terdeteksi sebagai berita real.
4. Di baris kedua, kolom kedua (584) adalah True Positive (TP), yang menunjukkan jumlah berita yang benar-benar terdeteksi sebagai berita real.

Dalam konteks deteksi berita palsu atau nyata, penggabungan "TfidfVectorizer" dan "Passive Aggressive Classifier" menghasilkan representasi teks yang adaptif dan model yang mampu mengklasifikasikan dengan akurat. Hasil array konfusi matriks yang dihasilkan merupakan indikator performa model dalam memprediksi klasifikasi berita. Kesimpulannya, implementasi ini berpotensi menjadi solusi efektif untuk mengatasi penyebaran berita palsu di era informasi digital.

## Kesimpulan dan Saran

Adapun kesimpulan yang dapat kita ambil dari penjelasan di atas ialah sebagai berikut:

1. Dengan menggunakan algoritma seperti PassiveAggressiveClassifier dan metode vektorisasi TF-IDF, kita dapat mengembangkan model yang dapat memprediksi apakah suatu berita adalah asli atau palsu berdasarkan fitur-fitur teks yang relevan.
2. Penggunaan TfidfVectorizer membantu dalam mengubah teks menjadi representasi numerik menggunakan skema pembobotan TF-IDF. Hal ini memungkinkan model Machine Learning untuk bekerja dengan data teks.
3. Dalam kode yang diberikan, akurasi model dihitung sebagai metrik evaluasi untuk mengukur seberapa baik model dalam memprediksi berita palsu. Akurasi merupakan persentase jumlah prediksi yang benar terhadap total jumlah data uji.

## Daftar Pustaka

Abdulrahman, A., & Baykara, M. (2020). Fake News Detection Using Machine Learning and Deep Learning Algorithms. *3rd International Conference on Advanced Science and Engineering, ICOASE 2020*, 18–23.  
<https://doi.org/10.1109/ICOASE51841.2020.9436605>

Ahmed, S., Hinkelmann, K., & Corradini, F. (2022). Development of Fake News Model using Machine Learning through Natural Language Processing. *14(12)*, 454–460.  
<http://arxiv.org/abs/2201.07489>

Bounabi, M., El Moutaouakil, K., & Satori, K. (2019). Text classification using Fuzzy TF-IDF and Machine Learning Models. *ACM International Conference Proceeding Series*. <https://doi.org/10.1145/3372938.3372956>

Chumairoh, H. (2020). Ancaman Berita Bohong di Tengah. *Vox Populi*, *3(1)*, 22–30.  
<http://journal.uin-alauddin.ac.id/index.php/voxpupuli/article/view/14395>

Faisal, M., Nugroho, F., El Sulthan, M. M., Amini, F., Hariyadi, M. A., & Sedayu, A. (2020). Plagiarism detection using manber and winnowing algorithm. *International Journal of Advanced Science and Technology*, *29(6 Special Issue)*, 2130–2136.

Hanani, A. (2008). Rancang Bangun Sistem Informasi Akademik Online Universitas Islam Negeri Malang. *140*.

Imbwaga, J. L., Chittaragi, N., & Koolagudi, S. (2022). Fake News Detection Using Machine Learning Algorithms. *ACM International Conference Proceeding Series*, *9(3)*, 271–275. <https://doi.org/10.1145/3549206.3549256>

Mandical, R. R., Mamatha, N., Shivakumar, N., Monica, R., & Krishna, A. N. (2020). Identification of Fake News Using Machine Learning. *Proceedings of CONECCT 2020 - 6th IEEE International Conference on Electronics, Computing and Communication Technologies*.  
<https://doi.org/10.1109/CONECCT50063.2020.9198610>

Popek, M., Swiatek-Najwer, E., Majak, M., Jaworowski, J., & Szram, D. (2014). Navigated surgical tools calibration issue in computer aided tumor resection procedures. *Computational Vision and Medical Image Processing IV - Proceedings*

of Eccomas Thematic Conference on Computational Vision and Medical Image Processing, VIPIMAGE 2013, 7, 331–336. <https://doi.org/10.1201/b15810-63>

Pratama, R. P., Faisal, M., & Hanani, A. (2019). Deteksi Plagiarisme pada Dokumen Jurnal Menggunakan Metode Cosine Similarity. *SMARTICS Journal*, 5(1), 22–26. <https://doi.org/10.21067/smartics.v5i1.2848>

Rahmatullah, T. (2018). Hoax dalam Perspektif Hukum Indonesia. *Jurnal Hukum Media Justitia Nusantara*, 8(2), 103–111. <http://ojs.uninus.ac.id/index.php/MJN/article/view/673/457>

Rathi, M., Malik, A., Varshney, D., Sharma, R., & Mendiratta, S. (2018). Sentiment Analysis of Tweets Using Machine Learning Approach. *2018 11th International Conference on Contemporary Computing, IC3 2018*, 2–4. <https://doi.org/10.1109/IC3.2018.8530517>

Riadi Silitonga, Y. (2019). Sistem Pendekripsi Berita Hoax di Media Sosial dengan Teknik Data Mining Scikit Learn. *Jurnal Ilmu Komputer*, 4, 173. [www.beritasatu.com](http://www.beritasatu.com),

Tripathy, A., Agrawal, A., & Rath, S. K. (2015). Classification of Sentimental Reviews Using Machine Learning Techniques. *Procedia Computer Science*, 57, 821–829. <https://doi.org/10.1016/j.procs.2015.07.523>

University of Peradeniya. Department of Electrical and Electronics Engineering, IEEE Sri Lanka Section. Central Region Subsection, IEEE Sri Lanka Section, Institute of Electrical and Electronics Engineers. Kharagpur Section, & Institute of Electrical and Electronics Engineers. (2019). *Sentiment Analysis in Tamil Texts: A Study on Machine Learning Techniques and Feature Representation*. 18–20.

Villagracia Octaviano, M. (2021). Fake News Detection Using Machine Learning. *ACM International Conference Proceeding Series*, 177–180. <https://doi.org/10.1145/3485768.3485774>