



JURNAL PUSTAKA ILMIAH

p-ISSN 2477-2070 | e-ISSN 2685-8363



Universitas Sebelas Maret
(UNS) Library,
Jl. Ir. Sutami 36 A Kentingan,
Jebres, Surakarta 57126

<https://jurnal.uns.ac.id/jurnalpustakailmiah>

Submitted : 25-01-2025

Accepted : 03-03-2025

Published : 30-06-2025

Diajukan : 25-01-2025

Diterima : 03-03-2025

Diterbitkan : 30-06-2025



Jurnal Pustaka Ilmiah is licensed under
a [Creative Commons Attribution-
NonCommercial-ShareAlike 4.0
International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

Analisis Visual dan Machine Learning untuk Mengukur Validitas Dokumen Akademik Perpustakaan: Studi pada Data Turnitin

Hesti Ari Wardani*, Imam Yuadi

Universitas Airlangga – Kampus Dharmawangsa (B)
Jl. Airlangga No.4 – 6, Airlangga, Kec. Gubeng, Surabaya,
Jawa Timur, 60115

*Korespondensi: hesti.ari.wardani-2024@fisip.unair.ac.id

ABSTRACT

The issue of plagiarism in academic works is a major concern in higher education as it can undermine academic integrity. This study aims to analyze the distribution of Turnitin results in academic library documents from various study programs at Universitas Nasional, Jakarta. Using a data visualization approach and machine learning algorithms, this research explores the relationship between Turnitin scores and document validity status. The methods used include data visualization through Scatter Plots, Violin Plots, and Box Plots, with these visualizations utilizing Orange Data Mining as the data processing method. Additionally, a logistic regression algorithm is applied to classify documents based on Turnitin scores. Furthermore, the Chi-Square statistical test is implemented to evaluate the significance of the relationship between Turnitin results and document validity status. The findings of this study indicate that Turnitin scores exhibit significant distribution differences among study programs, with notable disparities between valid and invalid documents. Documents from certain study programs tend to have dominant scores in the 20-50% range, which serves as a critical threshold in determining document validity. This study provides in-depth insights into the patterns of academic document validity and offers a data-driven approach to improving the quality of academic evaluation in higher education. Additionally, this research is expected to serve as a foundation for academic policies in strengthening plagiarism detection systems, increasing transparency in evaluation, and promoting the development of more accurate and efficient document validation methods. However, this study has limitations regarding the scope of data used, particularly in terms of study program representation and external factors that may influence Turnitin scores. Further research is needed to examine pedagogical aspects and academic policies that contribute to variations in Turnitin scores across different study programs.

Keywords: plagiarism; turnitin; data visualization; machine learning; document validity; higher education

ABSTRAK

Masalah plagiarisme dalam karya akademik menjadi perhatian utama di dunia pendidikan tinggi karena dapat merusak integritas akademik. Penelitian ini bertujuan untuk menganalisis distribusi hasil Turnitin pada dokumen akademik perpustakaan dari berbagai program studi di Universitas Nasional Jakarta.

Dengan pendekatan visualisasi data dan algoritma *machine learning*, penelitian ini mengeksplorasi hubungan antara skor turnitin dan status validitas dokumen. Metode yang digunakan meliputi visualisasi data melalui *Scatter Plot*, *Violin Plot*, dan *Box Plot*, visualisasi tersebut menggunakan orange data mining sebagai metode olah data, serta penerapan algoritma regresi logistik untuk mengklasifikasi dokumen berdasarkan skor Turnitin. Selain itu, uji statistik *Chi-Square* diterapkan untuk mengevaluasi signifikansi hubungan antara hasil Turnitin dengan status validitas dokumen. Hasil dari penelitian menunjukkan bahwa skor Turnitin memiliki distribusi yang signifikan antar prodi, dengan perbedaan mencolok antara dokumen valid dan tidak valid. Dokumen dari beberapa prodi cenderung memiliki skor yang dominan pada rentang 20-50%, yang menjadi titik kritis dalam pengambilan keputusan mengenai validitas dokumen. Penelitian ini memberikan wawasan yang mendalam mengenai pola validitas dokumen akademik dan menawarkan pendekatan berbasis data dalam meningkatkan kualitas evaluasi akademik perguruan tinggi. Penelitian ini juga diharapkan dapat menjadi dasar bagi kebijakan akademik dalam memperkuat sistem deteksi plagiarisme, meningkatkan transparansi dalam evaluasi, serta mendorong pengembangan metode validasi dokumen yang lebih akurat dan efisien. Namun, penelitian ini memiliki keterbatasan dalam cakupan data yang digunakan, terutama terkait representasi program studi dan faktor-faktor eksternal yang dapat mempengaruhi skor Turnitin. Penelitian lebih lanjut diperlukan untuk mengkaji aspek pedagogis dan kebijakan akademik yang berkontribusi terhadap variasi skor Turnitin di berbagai program studi.

Kata Kunci: plagiarisme; turnitin; visualisasi data; *machine learning*; validitas dokumen, pendidikan tinggi

PENDAHULUAN

Plagiarisme telah menjadi tantangan serius dalam dunia pendidikan tinggi, terutama dengan meningkatnya akses terhadap informasi digital yang memudahkan aktivitas penyalinan tanpa izin (Park, 2017). Alat seperti Turnitin banyak digunakan oleh institusi pendidikan khususnya di Perpustakaan karena sebagai alat untuk mendeteksi kesamaan teks dalam dokumen akademik sebagai langkah pencegahan plagiarisme (Sinaga, 2018). Namun alat pendeteksi plagiarisme seperti Turnitin memiliki keterbatasan, terutama dalam membedakan antara plagiarisme murni dan kutipan yang sah, sehingga diperlukan pendekatan analisis data yang lebih mendalam (Picorari, 2008). Oleh karena itu diperlukan pendekatan analisis data yang mendalam untuk memahami pola plagiarisme secara komprehensif (Baayen, 2008). Turnitin telah menjadi salah satu alat utama dalam mendeteksi plagiarisme di lingkungan akademik (Ramlan et al, 2017). Alat ini dirancang untuk membandingkan teks dokumen akademik dengan berbagai sumber, seperti database akademik, artikel jurnal, dan publikasi daring. Turnitin memberikan persentase kesamaan sebagai indikator awal untuk mendeteksi kemungkinan plagiarisme. Namun, nilai kesamaan ini sering kali disalahartikan oleh pengguna, terutama ketika tidak disertai dengan interpretasi yang memadai, Turnitin lebih cocok digunakan sebagai alat pendukung dalam menilai keaslian karya akademik, bukan sebagai penentu tunggal (Permana et al, 2022).

Plagiarisme dalam dunia akademik menjadi perhatian utama, terutama dengan semakin luasnya akses terhadap informasi digital yang dapat mempermudah aktivitas penyalinan tanpa izin (Park, 2017). Berbagai studi telah meneliti peran perangkat lunak deteksi plagiarisme,

seperti Turnitin, dalam menjaga integritas akademik, misalnya penelitian oleh Sinaga (2018) menunjukkan bahwa Turnitin efektif dalam mendeteksi kesamaan teks, tetapi belum sepenuhnya dapat membedakan antara plagiarisme dan kutipan yang sah. Studi lain oleh Witten (2011) menunjukkan bahwa analisis berbasis algoritma dapat meningkatkan akurasi dalam menilai keaslian dokumen akademik. Menurut James et al. (2021) pendekatan statistik seperti regresi logistik digunakan untuk mengklasifikasi validitas dokumen berdasarkan parameter yang relevan. Selain itu, penelitian Creswell & Creswell (2018) memberikan landasan tentang pentingnya kombinasi metode kuantitatif dan visualisasi data dalam analisis akademik.

Namun studi-studi tersebut masih belum secara spesifik mengkaji bagaimana faktor akademik mahasiswa, seperti program studi atau pemahaman mereka tentang etika akademik berkontribusi terhadap skor Turnitin. Penelitian ini bertujuan untuk mengisi kesenjangan tersebut dengan menganalisis hubungan antara tingkat kesamaan dokumen berdasarkan skor Turnitin dan faktor akademik mahasiswa, khususnya di Universitas Nasional. Diharapkan penelitian ini dapat memberikan rekomendasi kebijakan akademik yang lebih efektif dalam meningkatkan kesadaran mahasiswa terhadap pentingnya integritas akademik.

Di Universitas Nasional, masih ditemukan kasus-kasus dimana mahasiswa mendapatkan skor kesamaan yang tinggi dalam Turnitin, baik karena kurangnya pemahaman tentang penggunaan sumber atau praktik plagiarisme yang disengaja. Hal ini menunjukkan perlunya strategi yang lebih komprehensif, tidak hanya dalam penggunaan alat pendeteksi plagiarisme tetapi juga dalam membangun budaya akademik yang menjunjung tinggi integritas ilmiah (Sozon et al, 2024). Dengan demikian penelitian ini berfokus pada analisis pola plagiarisme berdasarkan hasil Turnitin di Universitas Nasional yang diharapkan dapat menjadi dasar bagi pengembangan kebijakan akademik yang lebih efektif dalam menangani permasalahan ini. Penelitian ini bertujuan untuk menganalisis hubungan antara tingkat kesamaan dokumen berdasarkan skor Turnitin (variabel independen) dan faktor akademik mahasiswa, seperti program studi dan tingkat pemahaman terhadap etika akademik (variabel dependen). Dengan memahami pola hubungan ini, penelitian ini diharapkan dapat memberikan rekomendasi kebijakan akademik yang lebih efektif dalam meningkatkan kesadaran mahasiswa terhadap pentingnya integritas akademik.

Dalam konteks ilmu perpustakaan, penelitian ini dapat memberikan wawasan bagi pustakawan dalam mendukung pengelolaan dokumen akademik di perpustakaan. Pustakawan memainkan peran penting dalam memastikan bahwa mahasiswa dan akademisi memiliki pemahaman yang baik mengenai etika penulisan dan sitasi yang benar. Hasil penelitian ini

diharapkan dapat membantu pustakawan dalam mengidentifikasi pola plagiarisme, memberikan edukasi kepada mahasiswa, serta menyusun kebijakan yang lebih efektif dalam memastikan integritas akademik (Howard, 1995). Penelitian ini menawarkan solusi berupa kombinasi visualisasi data dan algoritma *machine learning* untuk mengukur validitas dokumen akademik berdasarkan skor Turnitin. Dengan menggunakan *scatter plot*, *violin plot*, dan *box plot*, penelitian ini menunjukkan distribusi data secara visual (Saputra, 2023). Selanjutnya, algoritma regresi logistik digunakan untuk mengklasifikasi dokumen menjadi valid dan tidak valid, memberikan hasil yang lebih objektif dan akurat dibandingkan analisis manual. Selain itu, uji statistik seperti *Chi-Square* memberikan validasi yang mendukung interpretasi hasil.

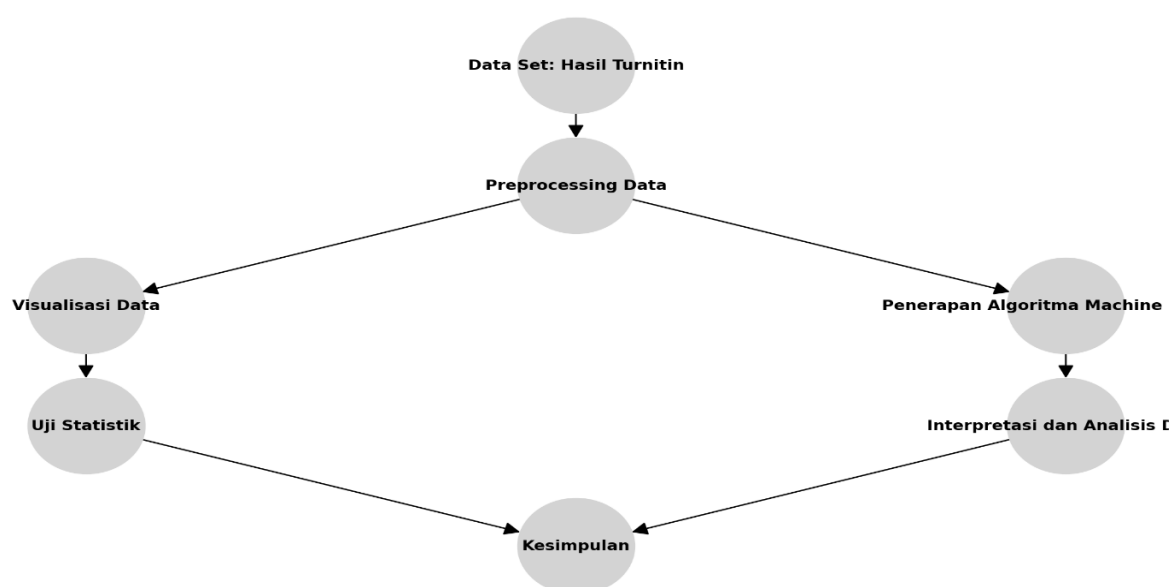
Dalam konteks metode yang digunakan, studi oleh Hastie et al. (2009) menggarisbawahi keunggulan regresi logistik sebagai alat untuk mengklasifikasi data kategorikal. Penelitian oleh Ghozali (2018) mendukung penggunaan uji Chi-Square untuk mengevaluasi hubungan antara variabel, menilai signifikansi hubungan antara variabel kategorikal, serta memungkinkan evaluasi data yang lebih objektif dan berbasis bukti dalam konteks akademik. Risparyanto (2020) menegaskan bahwa algoritma *machine learning* seperti regresi logistik memiliki potensi tinggi untuk diterapkan dalam mendeteksi pola plagiarisme. Selain itu, pentingnya visualisasi data seperti *scatter plot* dan diagram batang untuk mengidentifikasi distribusi data dan pola anomali. Visualisasi data menjadi alat yang semakin penting untuk memahami pola hasil Turnitin. Studi oleh Witten (2011) menunjukkan bahwa *scatter plot* dan *violin plot* sangat efektif dalam mengidentifikasi distribusi data yang kompleks. Sebagai contoh visualisasi dapat membantu mengungkap program studi tertentu yang cenderung memiliki skor Turnitin tinggi atau distribusi skor yang berbeda signifikan antara dokumen valid dan tidak valid. Pentingnya menggunakan visualisasi dalam mengkomunikasikan pola data kepada pengambil keputusan akademik, terutama untuk menyoroti area yang memerlukan intervensi (Hartama, 2018). Pendekatan berbasis *machine learning* telah digunakan untuk meningkatkan akurasi dalam mendeteksi plagiarisme. Algoritma seperti regresi logistik memungkinkan klasifikasi dokumen akademik menjadi valid dan tidak valid berdasarkan parameter tertentu, seperti skor Turnitin (James et al, 2021). Penerapan *machine learning* dalam deteksi plagiarisme menghasilkan prediksi yang lebih konsisten dibandingkan dengan evaluasi manual. Kombinasi metode ini memungkinkan analisis yang lebih komprehensif dan berbasis bukti, menghindari bias dan memastikan bahwa interpretasi data dilakukan secara valid (Santoso, 2020).

Penelitian sebelumnya juga membahas pentingnya mengintegrasikan temuan analisis data ke dalam kebijakan institusi. Edukasi mengenai etika akademik harus menjadi fokus utama

untuk mencegah plagiarisme (Sutrisno et al, 2024). Selain itu pentingnya pelatihan penggunaan alat seperti Turnitin, agar penggunaannya tidak hanya sebatas mendeteksi plagiarisme, tetapi juga mendidik mahasiswa tentang pentingnya keaslian karya akademik (Risparyanto, 2020). Kombinasi teknologi dan kebijakan berbasis data dapat menciptakan akademik yang lebih sehat (Witten, 2011).

Penelitian ini disusun dengan struktur yang pertama pendahuluan, memberikan latar belakang, fenomena, dan tujuan penelitian. Kemudian yang kedua adalah metode, menjelaskan teknik analisis yang digunakan. Ketiga adalah hasil dan pembahasan, menyajikan temuan utama dalam bentuk visualisasi dan klasifikasi data, serta mengeksplorasi implikasi dari temuan. Terakhir adalah kesimpulan.

METODE



Gambar 1. Flowchart Metodologi
Sumber: Penulis (2024)

Flowchart ini secara sistematis menjelaskan bagaimana data hasil Turnitin diproses, dianalisis, dan diinterpretasikan menggunakan pendekatan berbasis data dan *machine learning*, alasan penulis menggunakan *machine learning* didasarkan pada keunggulannya dalam mengolah data dalam jumlah besar, mengenali pola serta meningkatkan efisiensi analisis. Dalam konteks deteksi plagiarisme akademik, *machine learning* memungkinkan pemrosesan data secara otomatis, mengurangi ketergantungan pada metode manual yang lebih lambat dan rentan terhadap subjektivitas. *Machine learning* sangat efektif dalam menemukan pola

tersembunyi dalam data set yang kompleks sehingga dapat memberikan analisis yang lebih mendalam dibandingkan metode statistik konvensional (Bishop, 2006).

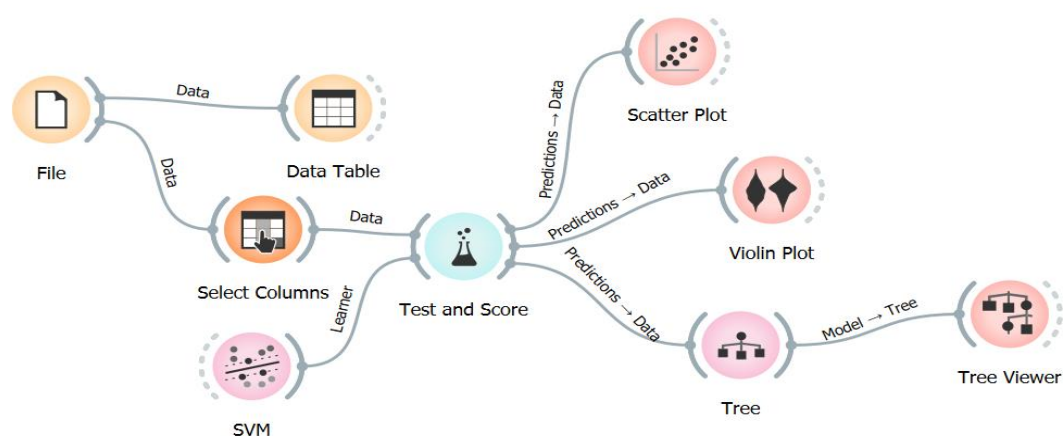
Setiap langkah saling terhubung untuk menghasilkan analisis yang mendalam dan mendukung pengambilan keputusan yang berbasis bukti di lingkungan akademik. Penelitian dimulai dengan data set dari hasil Turnitin dengan jumlah 339 dokumen akademik yang dikumpulkan dari berbagai program studi di Universitas Nasional, Jakarta. Dokumen-dokumen ini mencakup tugas akhir mahasiswa berbentuk skripsi dan artikel yang berasal dari 16 Program Studi yaitu Ilmu Hukum, Keperawatan, Kebidanan, Profesi Ners, Profesi Bidan, Sastra Korea, Sastra Inggris, Sastra Indonesia, Informatika, Sistem Informasi, Ilmu Komunikasi, Administrasi Publik, Hubungan Internasional, Manajemen, Biologi, dan Fisika. Pemilihan data set ini dilakukan dengan mempertimbangkan representasi yang seimbang antara berbagai program studi, sehingga memungkinkan analisis yang lebih menyeluruh terhadap pola distribusi skor Turnitin dan validitas dokumen.

Meskipun pendekatan ini memberikan wawasan yang spesifik terhadap lingkungan akademik di universitas tersebut, keterbatasan utama penelitian ini adalah kemungkinan bias yang muncul akibat penggunaan data dari satu institusi saja. Oleh karena itu, hasil penelitian ini perlu ditafsirkan dengan mempertimbangkan konteks spesifik universitas tersebut dan tidak secara langsung digeneralisasi ke institusi lain. Langkah berikutnya adalah *preprocessing* data bertujuan untuk membersihkan data agar siap diproses ke data analisis selanjutnya, prosesnya antara lain seperti menghapus data yang sama atau duplikat dan tidak relevan, memformat data agar lebih relevan untuk dianalisis, mengatasi nilai kosong dalam data (*missing values*).

Selanjutnya data divisualisasikan menggunakan *scatter plot*, *violin plot* maupun *box plot*. *Scatter plot* adalah jenis diagram yang digunakan untuk menunjukkan hubungan atau korelasi antara dua variabel numerik, sedangkan *violin plot* yaitu variasi dari *box plot* yang menggabungkan elemen distribusi probabilitas dari data. Selain menampilkan median dan kuartil, *violin plot* juga menunjukkan kepadatan data di berbagai nilai dengan menggunakan *kernel density estimation* dan *box plot* merupakan diagram statistik yang memberikan ringkasan lima nilai utama dari suatu distribusi data, minimum, kuartil pertama (Q1), median (Q2), kuartil ketiga (Q3), dan maksimum. Diagram ini juga sering digunakan untuk mendeteksi *outlier* dalam data. Seluruh proses ini memungkinkan untuk mengidentifikasi pola distribusi skor Turnitin. Semua visualisasi tersebut menggunakan perangkat lunak data *open source* (terbuka) berbasis data komponen untuk data mining yang dikembangkan oleh Universitas Ljubljana yaitu Orange Data Mining atau lebih dikenal dengan Orange (Saputra, 2023). Tujuan

visualisasi data antara lain mendeteksi anomali dalam data, memahami perbedaan distribusi antara dokumen valid dan tidak valid dan yang terakhir adalah mengidentifikasi tren dalam skor Turnitin berdasarkan program studi.

Pada penerapan algoritma *machine learning*, langkah ini menggunakan *Algoritma Decision Tree* atau SVM (*Support Vector Machine*) untuk membuat model prediktif. Algoritma ini digunakan untuk memprediksi status validitas dokumen berdasarkan skor Turnitin dan fitur lainnya, model dievaluasi menggunakan metrik seperti akurasi, sensitivitas dan presisi. Berikut hasil dari olah data menggunakan Orange Data Mining.

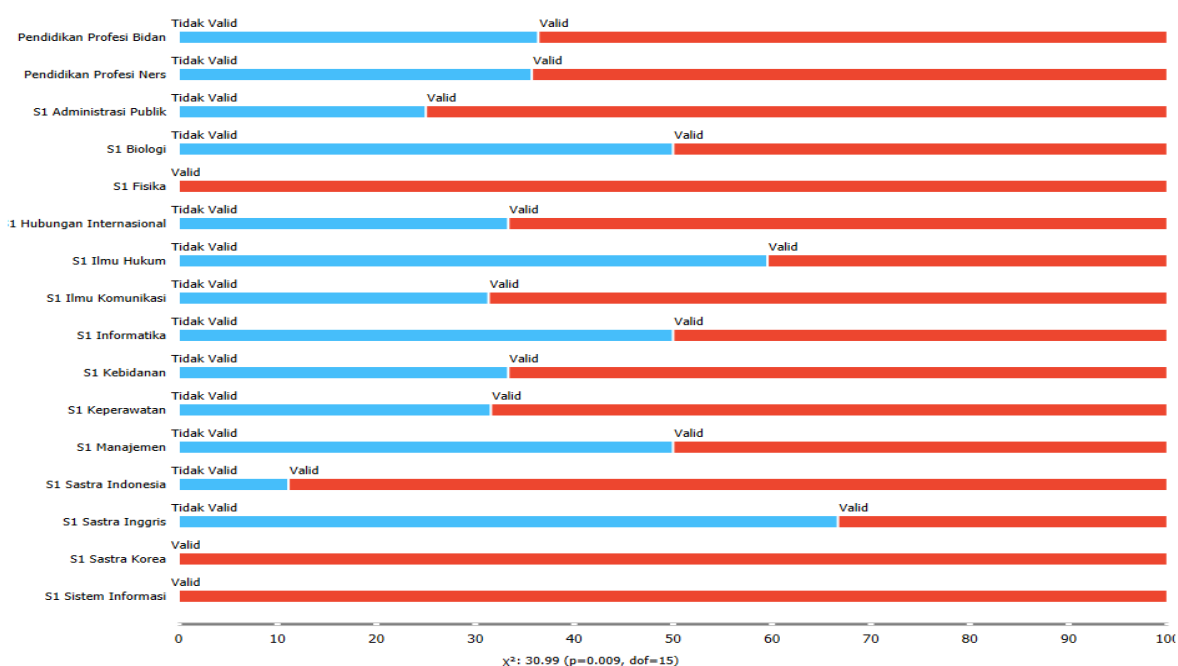


Gambar 2. Visualisasi menggunakan Orange Data Mining
Sumber: Penulis (2024)

Langkah berikutnya adalah uji statistik menggunakan uji *Chi Square* bertujuan untuk mengevaluasi hubungan antara skor Turnitin dan status validitas dokumen dan memberikan validasi yang berbasis bukti terhadap hasil dari model prediktif dan visualisasi data. Kaitannya untuk memastikan keandalan prediksi, model telah diuji menggunakan teknik validasi *K-Fold Cross-Validation*, bertujuan untuk mengurangi *overfitting* dan meningkatkan akurasi hasil prediksi. Data yang telah dianalisis diinterpretasikan untuk memahami implikasi dari temuan termasuk pembahasan mengenai faktor-faktor yang telah mempengaruhi skor Turnitin, penyebab dokumen yang dianggap valid dan tidak valid, serta implikasi bagi kebijakan akademik dan edukasi mahasiswa. Tahap terakhir adalah kesimpulan, merupakan hasil akhir penelitian mencakup temuan utama dari analisis visual, prediktif, dan statistik lalu rekomendasi untuk meningkatkan validitas dokumen akademik serta pentingnya pendekatan berbasis data untuk pengambilan keputusan terkait plagiarisme.

HASIL DAN PEMBAHASAN

Grafik *box plot* di bawah ini menggambarkan distribusi status dokumen akademik (valid dan tidak valid) berdasarkan hasil Turnitin dari berbagai program studi di Universitas Nasional. Setiap program studi ditampilkan dalam dua kategori yaitu dokumen valid (ditandai dengan warna merah) dan dokumen tidak valid (ditandai dengan warna biru). *Outlier* di *box plot* membantu mendeteksi dokumen dengan skor yang mencurigakan, yang memerlukan evaluasi manual lebih lanjut (Witten, 2011).

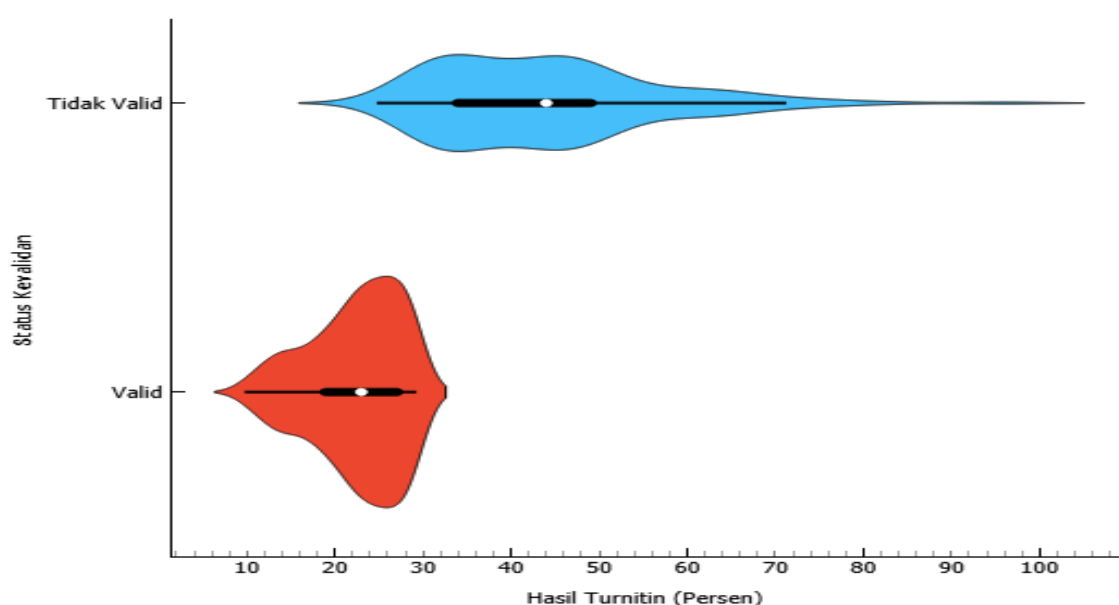


Gambar 3. Grafik Blox Plot
Sumber: Penulis (2024)

Distribusi skor Turnitin menunjukkan variasi antar program studi yang dapat dikaitkan dengan beberapa faktor utama. Program studi seperti Ilmu Hukum dan Sastra Inggris cenderung memiliki tingkat dokumen tidak valid yang lebih tinggi. Hal ini dapat disebabkan oleh tingginya jumlah kutipan langsung dalam penulisan akademik, kurangnya pemahaman mahasiswa tentang teknik parafrase, atau standar akademik yang lebih ketat dalam mendeteksi kesamaan teks. Di sisi lain program studi dengan tingkat validitas tinggi, seperti S1 Fisika, S1 Sastra Korea dan S1 Sistem Informasi, memiliki distribusi dokumen valid yang lebih besar dibandingkan program studi lain. Faktor yang mungkin berkontribusi adalah penggunaan metode analisis kuantitatif, teknis, atau berbasis pemrograman, yang mengurangi potensi kesamaan teks dengan sumber lain. Program studi dengan Distribusi Seimbang seperti S1 Informatika, S1 Biologi dan S1 Manajemen menunjukkan keseimbangan antara dokumen valid

dan tidak valid, meskipun masih ada dominasi dari kategori tidak valid. Perbedaannya tidak terlalu mencolok dibandingkan dengan kelompok sebelumnya. Hal ini bisa mengindikasikan bahwa dalam program studi ini terdapat variasi metode penulisan yang lebih luas dengan beberapa tugas yang bersifat teknis dan lainnya berbasis esai atau laporan penelitian. Analisis ini menunjukkan bahwa variasi distribusi skor Turnitin tidak hanya bergantung pada individu mahasiswa, tetapi juga pada karakteristik disiplin ilmu, metode penulisan yang umum digunakan, serta kebijakan akademik yang berlaku di masing-masing program studi.

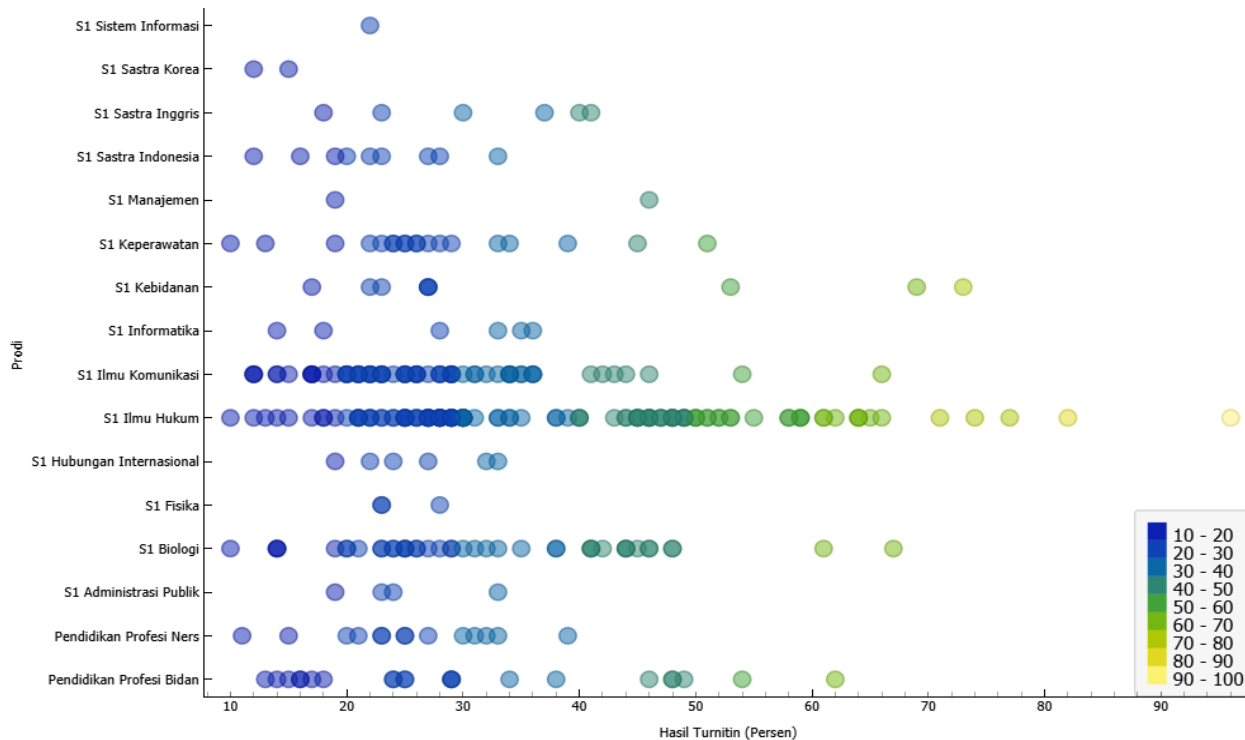
Statistik Uji *Chi-Square*, grafik ini juga menampilkan hasil uji statistik *Chi-Square* dengan nilai $\chi^2 = 30.99$, $p = 0.009$, dan derajat kebebasan (df) = 15. Hal ini menunjukkan adanya hubungan yang signifikan antara program studi dan status validitas dokumen. Dengan p -value kurang dari 0.05, hasil ini menyiratkan bahwa distribusi dokumen valid dan tidak valid berbeda secara signifikan antar program studi. Visualisasi seperti *box plot* sangat efektif untuk menggambarkan pola distribusi dan membantu interpretasi data yang kompleks (James et al, 2021).



Gambar 4. Diagram Violin
Sumber: Penulis (2024)

Gambar 4 menunjukkan diagram violin yang membandingkan distribusi hasil Turnitin (dalam persen) antara dua kategori status validitas dokumen yaitu Valid (diwakili oleh warna merah) dan Tidak Valid (diwakili oleh warna biru). Diagram ini memberikan informasi mengenai median, sebaran data, dan kepadatan data dalam setiap kategori. *Violin plot* membantu mengidentifikasi pola yang tidak terlihat dalam box plot biasa, seperti distribusi bimodal atau kepadatan tinggi di skor tertentu (Witten, 2011).

Pada kategori Tidak Valid, distribusi hasil Turnitin untuk dokumen tidak valid lebih lebar, dengan mayoritas nilai berada pada rentang 30-60%. Median untuk kategori ini berada pada sekitar 40%, yang menunjukkan bahwa dokumen dengan status tidak valid cenderung memiliki nilai Turnitin yang lebih tinggi. Hal ini mengindikasikan bahwa dokumen dengan kemiripan tinggi cenderung dikategorikan sebagai tidak valid. Temuan ini mendukung penelitian Permana et al. (2022) yang menyebutkan bahwa skor Turnitin dapat menjadi indikator awal plagiarisme. Distribusi hasil Turnitin untuk dokumen valid lebih sempit dibandingkan kategori tidak valid, dengan rentang dominan sekitar 10-30%. Median untuk kategori ini lebih rendah, sekitar 20%. Hal ini menunjukkan bahwa dokumen valid cenderung memiliki skor Turnitin rendah, yang mengindikasikan orisinalitas konten lebih tinggi. Perbedaan rentang skor antara dokumen valid dan tidak valid menunjukkan pola yang jelas, di mana skor tinggi lebih sering dikaitkan dengan status tidak valid. Sedangkan untuk pola kepadatan, bisa dilihat bahwa dokumen tidak valid memiliki kepadatan tinggi di skor Turnitin sekitar 40-50%, sementara dokumen valid lebih terkonsentrasi di sekitar 15-25%. Hal ini menunjukkan bahwa skor Turnitin dapat menjadi indikator awal untuk membedakan validitas dokumen. Temuan ini konsisten dengan studi James et al. (2021) yang menegaskan bahwa visualisasi data seperti violin plot sangat efektif untuk membandingkan distribusi dalam kelompok yang berbeda.



Gambar 5. Scatter Plot
Sumber: Penulis (2024)

Grafik *scatter plot* pada Gambar 5 menampilkan distribusi hasil Turnitin dalam persentase untuk berbagai program studi di Universitas Nasional. Sumbu horizontal menunjukkan skor Turnitin dalam persentase, sementara sumbu vertikal menunjukkan daftar program studi. Warna titik pada grafik menggambarkan rentang skor Turnitin, dengan warna biru untuk skor rendah (10–20%) dan warna kuning untuk skor tinggi (90–100%). *Scatter plot* menunjukkan pola distribusi yang dapat membantu mengidentifikasi program studi dengan tingkat plagiarisme lebih tinggi, yang memerlukan intervensi khusus (James et al, 2021).

Hasil skor Turnitin tersebar luas antara 10% hingga lebih dari 90%, dengan distribusi yang bervariasi untuk setiap program studi. Kebanyakan skor berada dalam rentang 20–50% (warna biru hingga hijau muda). Skor sangat tinggi (di atas 80%) relatif jarang terjadi, terutama di program studi tertentu. *Scatter plot* efektif untuk mendeteksi *outlier* dalam data skor Turnitin, sehingga dapat digunakan untuk mengevaluasi dokumen secara manual (Witten, 2011). Program studi seperti S1 Sistem Informasi, S1 Sastra Korea, dan S1 Fisika cenderung memiliki skor Turnitin yang rendah (10–30%), menunjukkan tingkat orisinalitas yang relatif baik pada dokumen mereka. Pada program studi dengan Sebaran Lebar seperti S1 Ilmu Komunikasi dan S1 Ilmu Hukum menunjukkan distribusi skor Turnitin yang lebih beragam, dari rendah hingga tinggi. Hal ini menunjukkan adanya perbedaan signifikan dalam kualitas dokumen di antara mahasiswa di program studi ini. Sedangkan prodi dengan Konsentrasi Tinggi di Skor Turnitin Menengah, seperti S1 Keperawatan dan Pendidikan Profesi Ners memiliki banyak dokumen dengan skor Turnitin di kisaran 30–50% (warna hijau), yang dapat menunjukkan kemiripan dengan karya lain yang cukup tinggi. Pendidikan Profesi Bidan, sebagian besar dokumen memiliki skor rendah hingga sedang (10–40%), menunjukkan bahwa dokumen mereka relatif orisinal.

Tabel 1. Analisis Distribusi Skor Turnitin Berdasarkan Program Studi dan Implikasinya

Kategori	Program Studi	Analisis	Rekomendasi
Identifikasi <i>Outlier</i>	S1 Biologi, S1 Hukum, S1 Ilmu Komunikasi	Terdapat beberapa Skor Turnitin sangat tinggi (>80%) berpotensi indikasi plagiarisme	Evaluasi lebih lanjut terhadap dokumen dengan skor tinggi untuk mendeteksi plagiarisme atau kesalahan sitasi
Program studi dengan variasi skor tinggi	S1 Hukum, S1 Ilmu Komunikasi	Sebaran skor yang luas menunjukkan perbedaan pemahaman dalam penulisan akademik	Perlu pengawasan lebih ketat dan bimbingan terkait sitasi dan penyusunan dokumenn akademik

Kebijakan anti plagiarisme yang spesifik	Semua Program Studi	Variasi skor antar prodi menunjukkan kebijakan umum mungkin tidak efektif	Institusi perlu merancang kebijakan anti-plagiarisme yang disesuaikan untuk masing-masing prodi.
Edukasi dan Pelatihan	S1 Keperawatan dan S1 Administrasi Publik	Mayoritas dokumen memiliki skor tinggi, mengindikasikan perlunya peningkatan literasi akademik	Pelatihan tambahan bagi mahasiswa dalam Teknik parafrase, sitasi yang benar dan standar keaslian akademik, melalui <i>user education</i> .

Sumber: Penulis (2024)

Hasil dari penelitian ini menunjukkan distribusi skor Turnitin yang beragam di berbagai program studi Universitas Nasional, dengan implikasi signifikan terhadap validitas dokumen akademik sebagaimana ditunjukkan pada Tabel 1. Beberapa temuan utama adalah Program studi seperti S1 Sistem Informasi dan S1 Sastra Korea memiliki skor Turnitin yang cenderung rendah, mengindikasikan tingkat orisinalitas yang lebih tinggi. Sebaliknya, program seperti S1 Ilmu Hukum dan S1 Ilmu Komunikasi menunjukkan distribusi skor yang lebih lebar, yang dapat mencerminkan variasi kemampuan akademik mahasiswa dalam menyusun karya asli. Pada Korelasi Validitas dan Skor Turnitin, *violin plot* menunjukkan pola yang jelas antara skor Turnitin dan validitas dokumen. Dokumen dengan skor Turnitin yang lebih rendah cenderung dianggap valid, sementara dokumen dengan skor tinggi sering dikategorikan tidak valid. Penelitian ini sejalan dengan temuan (Risparyanto, 2020) yang menunjukkan bahwa skor Turnitin dapat berfungsi sebagai indikator awal validitas dokumen akademik, meskipun skor tinggi pada Turnitin seringkali menandakan potensi plagiarisme, interpretasi skor tersebut memerlukan konteks tambahan, seperti pemahaman terhadap jenis dokumen dan penerapan sitasi yang tepat. Hal ini juga tercermin dalam temuan penelitian ini, di mana skor tinggi pada Turnitin ditemukan pada beberapa program studi yang memerlukan evaluasi lebih lanjut untuk mengidentifikasi plagiarisme atau kesalahan sitasi. Selain itu, hasil penelitian ini konsisten dengan pandangan (Witten, 2011), yang menekankan pentingnya visualisasi data, khususnya *scatter plot*, dalam menganalisis pola dalam dataset besar. Dalam konteks penelitian ini, visualisasi digunakan untuk menggambarkan sebaran skor Turnitin, mengidentifikasi variasi antar program studi, serta mendeteksi outlier. Pendekatan ini sejalan dengan pendekatan yang diusulkan oleh (Witten, 2011), yang bertujuan untuk menyederhanakan dan mempermudah interpretasi data yang kompleks.

Hasil penelitian ini memiliki beberapa implikasi praktis yang penting untuk pengembangan kebijakan akademik. Pertama, program studi dengan skor tinggi atau distribusi

skor yang lebar, seperti S1 Ilmu Hukum, memerlukan pendekatan yang lebih intensif dalam edukasi mahasiswa tentang penulisan akademik dan etika sitasi. Hal ini bertujuan untuk meningkatkan pemahaman mahasiswa mengenai teknik penulisan yang benar serta pentingnya menjaga keaslian dalam karya akademik. Selain itu, hasil penelitian ini menunjukkan bahwa institusi tidak seharusnya hanya mengandalkan skor Turnitin sebagai satu-satunya indikator validitas dokumen. Mengingat keterbatasan yang ada dalam interpretasi skor Turnitin, diperlukan pendekatan berbasis data yang lebih komprehensif, yang terintegrasi dengan analisis manual atau algoritma tambahan, guna memberikan wawasan yang lebih tepat tentang keaslian dokumen akademik. Selain itu, institusi juga harus menyediakan pelatihan yang sistematis bagi mahasiswa dan dosen mengenai penggunaan Turnitin dan pentingnya menjaga keaslian dokumen akademik. Pelatihan ini dapat mencakup teknik parafrase yang tepat, penggunaan kutipan yang sah, serta cara yang benar dalam memanfaatkan alat pengecekan plagiarisme.

Meskipun demikian, penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan. Pertama, data yang digunakan hanya berasal dari satu institusi sehingga generalisasi hasil penelitian ini ke institusi lain perlu dilakukan dengan hati-hati (Cyber Library Universitas Nasional, 2023). Penelitian yang melibatkan berbagai universitas akan memberikan pemahaman yang lebih luas mengenai penggunaan Turnitin dan validitas dokumen akademik di tingkat yang lebih global. Kedua, penelitian ini tidak dapat menganalisis semua faktor yang mempengaruhi skor Turnitin secara mendalam, seperti kesalahan teknis dalam penggunaan alat atau penggunaan kutipan yang sah. Faktor-faktor ini dapat memengaruhi skor Turnitin tetapi tidak sepenuhnya tercakup dalam analisis. Selain itu, penelitian ini menggunakan algoritma *machine learning* terbatas pada regresi logistik. Penggunaan algoritma lain, seperti Random Forest atau Neural Networks, dapat memberikan wawasan yang lebih mendalam dan akurat mengenai pola plagiarisme dan validitas dokumen.

Penelitian selanjutnya, terdapat beberapa arah penelitian yang perlu dipertimbangkan adalah analisis multidimensional yang lebih mendalam mengenai faktor-faktor lain yang memengaruhi validitas dokumen akademik, seperti gaya penulisan, pilihan bahasa, atau tingkat pendidikan penulis. Analisis yang lebih luas ini dapat memberikan pemahaman yang lebih komprehensif terkait plagiarisme dan kesalahan dalam penulisan. Penelitian pada skala yang lebih besar, yang melibatkan berbagai universitas, juga dapat memperluas wawasan tentang pola penggunaan Turnitin dan efektivitas kebijakan anti-plagiarisme yang diterapkan di berbagai institusi. Terakhir, integrasi teknologi lanjutan seperti Natural Language Processing

(NLP) dapat meningkatkan akurasi analisis dokumen akademik, memungkinkan analisis yang lebih dalam terhadap isi dokumen untuk memvalidasi hasil Turnitin dan mendeteksi plagiarisme secara lebih tepat, termasuk dalam kasus parafrase yang kurang efektif atau kutipan yang tidak sesuai standar akademik.

KESIMPULAN

Penelitian ini menunjukkan bahwa skor Turnitin memiliki hubungan signifikan dengan status validitas dokumen akademik di Universitas Nasional. Temuan utama mencakup distribusi skor Turnitin yang bervariasi di berbagai program studi. Sebagian besar program studi menunjukkan skor rendah (10–30%), yang lebih sering dikategorikan sebagai dokumen yang valid, sementara skor tinggi (>50%) cenderung dianggap sebagai indikasi dokumen yang tidak valid. Analisis statistik juga mengungkapkan adanya korelasi yang signifikan antara skor Turnitin dan status validitas dokumen, dengan nilai $\chi^2 = 30.99$ dan $p = 0.009$, yang menunjukkan bahwa skor Turnitin dapat berfungsi sebagai indikator awal untuk menilai keaslian dokumen akademik. Penggunaan pendekatan visualisasi data, seperti *scatter plot* dan *violin plot*, serta penerapan algoritma regresi logistik, berhasil memberikan wawasan yang lebih mendalam tentang pola validitas dokumen. Temuan ini menekankan pentingnya pemahaman yang lebih baik mengenai skor Turnitin, serta menunjukkan perlunya kebijakan edukasi dan pelatihan yang lebih intensif untuk meningkatkan kesadaran dan pemahaman tentang keaslian akademik di kalangan mahasiswa dan dosen.

DAFTAR PUSTAKA

- Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge University Press.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer Science+Business Media, LLC.
- Creswell, J. W., & Creswell, J. D. (2018). *Research Design Fifth Edition: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications.
- Cyber Library Universitas Nasional. (2023). Layanan Uji Turnitin Universitas Nasional. <https://perpustakaan.unas.ac.id/Panduan-Uji-Turnitin/>
- Ghozali, I. (2018). *Aplikasi analisis multivariate dengan program IBM SPSS 25* (Edisi 9). Badan Penerbit Universitas Diponegoro.
- Sozon, M., Mohammad Alkharabsheh, O. H., Fong, P. W., & Chuan, S. B. (2024). Cheating and plagiarism in higher education institutions (HEIs): A literature review. *F1000Research*, 13, 788. <https://doi.org/10.12688/f1000research.147140.2>
- Hartama, D. (2018). Analisa Visualisasi Data Akademik Menggunakan Tableau Big Data. *Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)*, 3, 46–55. <http://dx.doi.org/10.30645/jurasik.v3i0.65>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.

- Howard, R. M. (1995). *Plagiarisms, Authorships, and the Academic Death Penalty*. <https://doi.org/https://doi.org/10.2307/378403>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning* (Second). Springer.
- Park, C. (2017). In other (people's) words: Plagiarism by university students-literature and lessons. *Academic Ethics*, 28(5), 525–542. <https://doi.org/10.4324/9781315263465-42>
- Permana, R. A., Priharsari, D., & Perdanakusuma, A. R. (2022). *Analisis Penggunaan Software Turnitin Sebagai Alat Pendeteksi Plagiarisme*. 6(11), 5446–5453. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/11903>
- Picorari, D. (2008). *Academic Writing and Plagiarism*. Internetarchivebooks.
- Ramlan, Erwinsyahbana, T., & Hakim, N. (2017). *Buku Malu Menjadi Plagiator*. Intelgensia Media.
- Risparyanto, A. (2020). Turnitin Sebagai Alat Deteksi Plagiarisme. *UNILIB : Jurnal Perpustakaan*, 11(2). <https://doi.org/10.20885/unilib.vol11.iss2.art5>
- Santoso, S. (2020). *Panduan lengkap SPSS 26*. Elex Media Komputindo.
- Saputra, I. (2023). *Belajar Mudah Data Mining untuk Pemula*. Informatika.
- Sinaga, S. (2018). Penggunaan Aplikasi Turnitin Sebagai Sarana Cek Plagiarisme dalam Layanan Perpustakaan Universitas Ukrida. *BIBLIOTIKA: Jurnal Kajian Perpustakaan dan Informasi*, 2(2), 123-131. doi:<http://dx.doi.org/10.17977/um008v2i22018p123>
- Sutrisno, E., Syafrizal, Ritnawati, Erdawaty, Rochmatika, E., Mahyuni, E. T., Soetijono, I. K., Mayasari, E., Widodo, M. L., Yuniarti, E., Ma'rief, A. A., Hernovianty, F. R., & Simarmata, J. (2024). *Plagiarisme dan Integritas Akademik*. Yayasan Kita Menulis.
- Witten, I. H. et al. (2011). *Data Mining : Practical Machine Learning Tools and Techniques, Third Edition* (Third). Morgan Kaufmann Publishers is an imprint of Elsevier.