Pendeteksi Akun Palsu di Instagram Menggunakan Model RuleFit dan *Gaussian Process Classifier*

Florentina Yuni Arini[1], Muhammad Azzam Fadhlullah[2], Diva Satria[3], Afrilza Daffa Naryapramono[4], Tyto Rinandi[5], Fawwaz Haryolukito Pambudi[6]
Universitas Negeri Semarang
e-mail: [1]floyuna@mail.unnes.ac.id [2]azzamfadli247@students.unnes.ac.id,
[3]divasatria885@students.unnes.ac.id, [4]afrilzadaffa01@students.unnes.ac.id,
[5]tytorinandi130904@students.unnes.ac.id, [6]fawwazhp00@students.unnes.ac.id

Diajukan: 27 Mei 2025; Direvisi: 19 Juli 2025 ; Diterima: 24 Juli 2025

Abstrak

Deteksi akun palsu di platform media sosial menjadi tantangan krusial dalam upaya mitigasi penyebaran informasi palsu dan penipuan secara online. Penelitian ini mengusulkan pendekatan dengan menggabungkan model RuleFit dan Gaussian Process Classifier (GPC) melalui kombinasi feature engineering, di mana RuleFit digunakan untuk menghasilkan rule-based features yang kemudian dilatih dengan model GPC. Dataset penelitian terdiri dari 576 akun Instagram dengan berbagai fitur seperti karakteristik profil, pola aktivitas, dan interaksi pengguna yang kemudian diproses menggunakan One Hot Encoding dan standarisasi. Penggunaan model RuleFit dan GPC untuk deteksi akun palsu Instagram menunjukkan performa yang mampu bersinergi. Hal ini ditunjukkan dengan hasil eksperimen bahwa model RuleFit dan GPC mencapai performa menjanjikan dengan akurasi 92,2%, precision 98%, dan recall 86%, sehingga mendominasi hasil model individual RuleFit (akurasi 91,38%) dan model GPC (akurasi 90,52%). Penelitian ini memberikan kontribusi praktis berupa pengembangan sistem deteksi akun palsu yang lebih andal untuk meningkatkan keamanan platform media sosial.

Kata kunci: *akun palsu, pembelajaran mesin, deteksi akun palsu.*

Abstract

Fake accounts detection on social media platforms is a crucial challenge in mitigating the spread of false information and online fraud. This research proposes a approach by combining RuleFit and Gaussian Process Classifier (GPC) models through feature engineering, where RuleFit is used to generate rule-based features that are then trained with the GPC model. The research dataset consists of 576 Instagram accounts with various features such as profile characteristics, activity patterns, and user interactions, which are then processed using One Hot Encoding and standardization. Implementation RuleFit and GPC models for detecting fake Instagram accounts shows synergistic performance. Thus, experimental results support that the RuleFit and GPC models achieve promising performance with 92.2% accuracy, 98% precision, and 86% recall. These percentage dominate the individual results of RuleFit model (91.38% accuracy) and the GPC model (90.52% accuracy). This research provides a practical contribution in the form of developing a more reliable fake account detection system to enhance security on social media platforms.

Keywords*: fake accounts, machine learning, account detection.*

1. Pendahuluan

Akun media sosial yang menyamar sebagai orang lain atau entitas lain dengan tujuan tertentu, seperti menyebarkan informasi palsu, penipuan, atau manipulasi opini publik, dikenal sebagai akun palsu [1]. Dalam beberapa situasi, akun media sosial ini dapat menjadi alat yang bermanfaat bagi masyarakat. Contohnya, mereka dapat memberikan Informasi yang bermanfaat kepada pengguna sosial media lain melalui Instagram, Twitter, dan social media lainnya. Namun, hal yang paling mencemaskan bahwa akun media sosial ini akan memperparah kondisi yang sama atau bahkan lebih buruk seperti mempengaruhi atau

menipu banyak pengguna dengan menyebarkan pesan politik, berita palsu, atau tautan berbahaya [2]. Seiring dengan berkembangnya teknologi digital, muncul berbagai peluang dan metode yang dapat dimanfaatkan untuk mendeteksi keberadaan akun palsu di dunia maya. Teknologi seperti kecerdasan buatan, analisis data, dan pelacakan digital kini memungkinkan peneliti untuk mengidentifikasi aktivitas mencurigakan yang dilakukan oleh akun-akun tersebut. Selain itu, jejak digital yang ditinggalkan, seperti pola interaksi, alamat IP, hingga metadata unggahan, dapat dianalisis lebih lanjut untuk mengetahui asal-usul dan tujuan dari akun palsu tersebut. Dengan demikian, kemajuan teknologi turut memberikan kontribusi besar dalam menjaga keamanan dan keaslian identitas di ranah digital [3]. Oleh karena itu, adanya pendeteksi akun palsu yang bertujuan untuk mengidentifikasi akun tidak autentik di platform media sosial, yaitu akun yang dikendalikan oleh bot atau pengguna yang menyamar dengan identitas palsu. Proses ini biasanya dilakukan dengan menganalisis fitur profil, pola aktivitas, hubungan sosial, dan konten yang diposting menggunakan algoritma pembelajaran mesin atau teknik berbasis jaringan [4].

Pemanfaatan teknik machine learning menjadi opsi terbaik untuk membangun alat yang dapat mendeteksi akun-akun palsu tersebut. Pada penelitian ini, algoritma yang digunakan yaitu RuleFit, serta *Gaussian Process* dan kombinasi RuleFit dan GPC. Metode RuleFit menggabungkan pohon keputusan dan regresi linear untuk membangun model yang akurat dan mudah diinterpretasikan. Dengan mengekstrak aturan dari pohon keputusan dan mengubahnya menjadi fitur biner dalam model linear, RuleFit mampu menghasilkan model yang tidak hanya kuat dalam prediksi, tetapi juga transparan dalam menjelaskan keputusan yang diambil. RuleFit dapat digunakan pada platform media sosial untuk menemukan akun palsu dengan menganalisis pola-pola yang ada, seperti perilaku pengguna yang mencurigakan, aktivitas yang tidak biasa, atau interaksi yang tidak wajar. Ini dapat membantu platform media sosial menemukan akun yang mungkin berbahaya [5]. Metode klasifikasi probabilistik *Gaussian Process*, yang didasarkan pada teori Bayesian, digunakan untuk memodelkan sekumpulan variabel acak dengan distribusi Gaussian multivariat. Metode ini digunakan dalam penelitian ini untuk mendeteksi akun Instagram yang tidak asli karena kemampuan untuk menangkap ketidakpastian dalam prediksi serta memberikan distribusi probabilistik atas kelas, sehingga sangat efektif dalam mengurangi kesalahan deteksi, terutama yang berkaitan dengan prediksi positif palsu [6]. Metode RuleFit dan GPC menggabungkan dua metode berbeda untuk melengkapi kelebihan dan kekurangan masing-masing sehingga proses pengembangan sistem menjadi lebih fleksibel, efisien, dan tetap terstruktur. Metode ini cocok digunakan ketika dibutuhkan kecepatan pengembangan tanpa mengorbankan kualitas dan ketertiban proses kerja [7].

2. Metode Penelitian

1. Penelitian ini bertujuan untuk mengidentifikasi akun palsu di platform Instagram dengan memanfaatkan pendekatan machine learning yang menggabungkan akurasi dan interpretabilitas. Model RuleFit dipilih karena kemampuannya dalam menghasilkan aturan keputusan yang mudah dipahami, yang diperoleh dari ensemble pohon keputusan dan diintegrasikan ke dalam model linear yang sparsity-nya dikendalikan oleh regularisasi L1 [8]. Sementara itu, *Gaussian Process Classifier* (GPC) digunakan karena sifatnya yang non-parametrik dan kemampuannya dalam memberikan prediksi probabilistik, memungkinkan estimasi ketidakpastian yang bermanfaat dalam konteks klasifikasi [9].
2. Kombinasi kedua metode ini diharapkan dapat memberikan hasil klasifikasi yang tidak hanya akurat tetapi juga dapat diinterpretasikan dengan baik, sehingga memudahkan dalam memahami karakteristik akun palsu di Instagram. Setiap tahapan dalam metode ini dirancang secara sistematis untuk memastikan proses klasifikasi berjalan secara efektif dan efisien.

2.1. RuleFit

RuleFit merupakan model regresi dan klasifikasi yang secara umum dibangun sebagai kombinasi linear dari aturan-aturan sederhana yang dihasilkan dari data. Setiap aturan merupakan gabungan dari beberapa pernyataan sederhana mengenai nilai-nilai variabel input tertentu. Kumpulan aturan ini mampu memberikan akurasi prediksi yang sebanding dengan metode terbaik lainnya. Namun, keunggulan utamanya terletak pada kemudahannya untuk dipahami dan diinterpretasikan [10]. Meskipun demikian, dalam praktiknya, model rule ensemble seperti RuleFit memerlukan sejumlah besar aturan berbobot untuk mempertahankan performa generalisasi yang baik. Oleh karena itu, sering kali terdapat kompromi antara interpretabilitas dan akurasi; semakin banyak aturan yang digunakan untuk menjaga akurasi, semakin sulit pula model untuk dipahami secara keseluruhan [11].

RuleFit merupakan metode klasifikasi yang menggabungkan aturan-aturan (rules) hasil ekstraksi dari decision tree dengan model linier biasa. Secara umum, prediksi dalam RuleFit dinyatakan dengan

rumus pada persamaan (1), dimana α_0 sebagai bias (intercept), $r_d(x)$ sebagai aturan ke-d yang bernilai 0 atau 1, α_d sebagai bobot untuk aturan tersebut, $l(x_j)$ sebagai transformasi linier fitur ke-j, dan β_j sebagai bobot dari fitur linier tersebut. Dengan kata lain, model ini membangun prediksi sebagai kombinasi dari keputusan berbasis aturan dan hubungan linier dari fitur input [12].

$$F(x) = \alpha_0 + \sum_{d=1}^{D}\big(\alpha_0 * r_d(x)\big) + \sum_{j=1}^{p}(\beta_j * l(x_j)) \quad \text{..... (1)}$$

RuleFit mampu menghasilkan model yang akurat sekaligus mudah diinterpretasi dengan menggabungkan aturan-aturan dari pohon keputusan ke dalam regresi linier. RuleFit juga secara otomatis melakukan seleksi fitur penting melalui penalti regularisasi, sehingga menghasilkan model yang lebih sederhana dan efisien. Kemampuannya untuk menangkap interaksi kompleks dan pola non-linier menjadikan RuleFit sangat efektif, terutama dalam konteks data klinis. Selain itu, interpretabilitas yang tinggi membuatnya cocok digunakan dalam pengambilan keputusan yang membutuhkan transparansi [13].

2.2. *Gaussian Process Classifier* (GPC)

Secara umum, *Gaussian Process Classifier* (GPC) merupakan metode klasifikasi berbasis Bayesian non-parametrik yang memodelkan sebuah fungsi laten $f(x)$, yang kemudian dipetakan ke dalam probabilitas kelas melalui fungsi probit $\Phi(f(x))$ [14]. Dalam kasus klasifikasi biner, probabilitas keanggotaan suatu data x terhadap kelas positif dinyatakan sebagai $p(y = 1|x) = \Phi(f(x))$. Proses klasifikasi dengan Gaussian Process dilakukan dengan menempatkan prior *Gaussian Process* atas fungsi laten tersebut, yang ditentukan oleh fungsi mean $m(x) = E[f(x)]$ dan fungsi kovariansi $k(x, x') = Cov\big(f(x), f(x')\big)$. Konsep dasar *Gaussian Process Classifier* beserta formulasi umum ini dijelaskan oleh Rodrigues. Sebelum kemudian dikembangkan lebih lanjut untuk menangani skenario klasifikasi dengan multiple annotators yang memiliki tingkat keahlian berbeda [15]. Untuk melakukan prediksi pada data uji x^* (yaitu data baru yang belum pernah dilihat atau digunakan dalam proses pelatihan), distribusi posterior dari nilai fungsi laten f^* (nilai dari fungsi laten yang diprediksi untuk data uji x^*) dihitung melalui rumus persamaan (2).

$$p(f^*|x^*, X, y) = \int p(f^*|x^*, X, y)p(f|X, y)\, df \quad \text{..... (2)}$$

dan probabilitas prediksi kelas diberikan oleh rumus persamaan (3).

$$p(y^* = 1|x^*, X, y) = \int \Phi(f^*)p(f^*|x^*, X, y)\, df^* \quad \text{..... (3)}$$

Gaussian Process Classification (GPC) menawarkan sejumlah keunggulan, seperti kemampuan menghasilkan prediksi berbasis probabilitas lengkap dengan estimasi ketidakpastian, bekerja efektif pada data kecil, serta fleksibel karena bersifat non-parametrik. GPC juga secara alami menghindari overfitting melalui regularisasi kernel dan mampu menangkap hubungan non-linear antar fitur. Selain itu, model ini menyediakan interpretasi melalui struktur kovariansi, menjadikannya akurat dan transparan dalam berbagai aplikasi klasifikasi [16].

2.3. Dataset

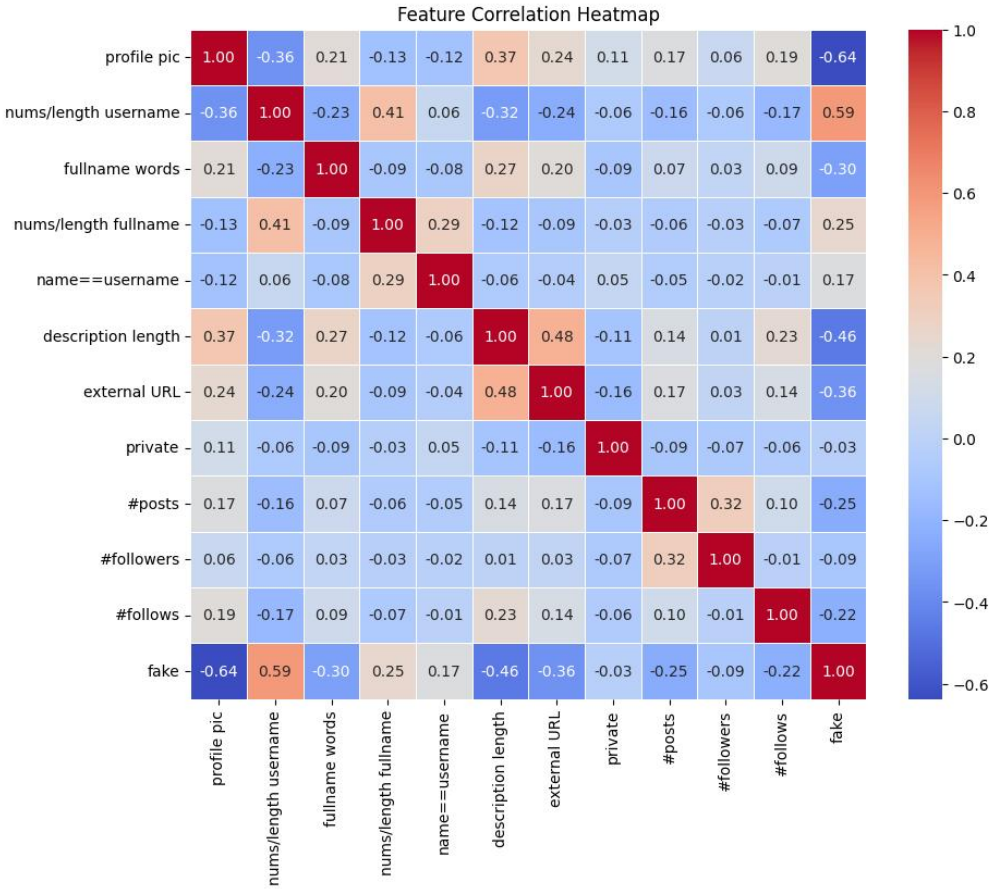
Kumpulan data yang digunakan dalam penelitian dikenal sebagai dataset, yang terdiri dari sejumlah data dengan karakteristik tertentu yang digunakan untuk proses klasifikasi dan analisis kinerja algoritma [17]. Pada penelitian ini menggunakan dataset *Instagram Detecting fake accounts* [18]. Dataset ini mengandung berbagai fitur yang menunjukkan karakteristik akun, seperti foto profil dan tautan eksternal yang diwakili oleh indikator biner, rasio jumlah karakter terhadap panjang nama lengkap dan username, dan kecocokan antara nama lengkap dan username. Selain itu, data numerik seperti panjang nama lengkap, jumlah postingan, jumlah pengikut, dan status akun (publik atau privat) juga termasuk dalam dataset ini [19].

3. Hasil dan Pembahasan

3.1. *Preprocessing* Data

Preprocessing merupakan langkah awal dalam proses klasifikasi yang bertujuan untuk mempersiapkan data agar siap untuk diproses pada tahap berikutnya [20]. Proses *preprocessing* data dilakukan untuk pemenuhan tahap *text preprocessing*, yaitu proses konversi data dari bentuk yang tidak terstruktur atau acak menjadi format yang terstruktur, seperti mengubah kumpulan teks menjadi indeks

term [21]. Setelah melakukan tahapan pembersihan data, tidak ditemukan *missing value* pada dataset. Kemudian langkah yang dilakukan pada tahap berikutnya adalah memisahkan fitur (x) dan label (y). Data tersebut kemudian dikonversi menjadi format numerik menggunakan teknik One Hot Encoding dengan menghapus salah satu kategori untuk menghindari multikolinearitas. Proses ini dilakukan agar algoritma Machine Learning dapat memproses data kategorikal dengan lebih efektif. Setelah itu, data dibagi menjadi data latih dan data uji dengan proporsi 80:20 menggunakan metode *train-test split*.



Gambar 1. Heatmap Korelasi Antar Fitur pada Dataset Deteksi Akun Palsu

Pada gambar 1 menunjukkan beberapa hubungan penting antara berbagai fitur dan label "fake" pada akun. Korelasi negatif terkuat ditemukan pada fitur "profile pic" (-0.64), mengindikasikan bahwa akun palsu umumnya tidak menggunakan foto profil. Sementara itu, "nums/length username" memiliki korelasi positif signifikan (0.59), menunjukkan bahwa akun tidak autentik sering menggunakan nama pengguna dengan banyak angka atau karakter tertentu. Beberapa fitur menunjukkan korelasi negatif menengah terhadap akun palsu, seperti "description length" (-0.46), "external URL" (-0.36), dan "fullname words" (-0.30). Ini berarti akun palsu cenderung memiliki deskripsi singkat, tidak menyertakan tautan eksternal, dan menggunakan nama yang lebih sederhana. Metrik keterlibatan seperti "#followers", "#follows", dan "#posts" memperlihatkan korelasi lemah dengan kisaran (-0.09) hingga (-0.25), menandakan bahwa jumlah aktivitas tidak secara kuat mencerminkan keaslian akun. yang berarti jumlah aktivitas atau keterlibatan akun tidak sepenuhnya mencerminkan keaslian akun. Terdapat juga korelasi internal dengan hubungan positif seperti korelasi antara "description length" dan "external URL" (0.48), di mana akun dengan deskripsi lebih panjang cenderung menyertakan tautan eksternal. Temuan-temuan ini memberikan dasar penting untuk seleksi fitur dalam pengembangan model deteksi akun palsu, dengan prioritas pada fitur-fitur yang memiliki korelasi tinggi terhadap target.

3.2. Evaluasi Model *Gaussian Process Classification*

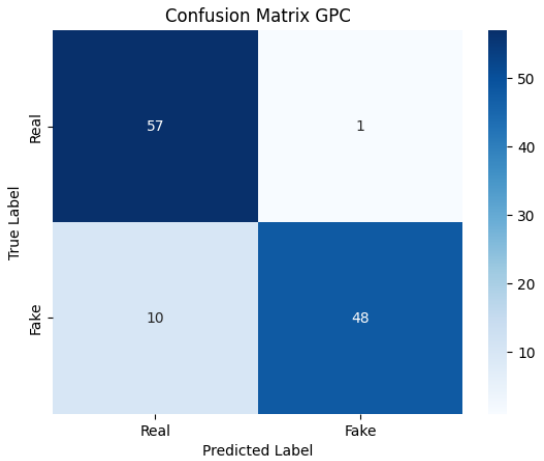
Untuk mengevaluasi performa dari algoritma *Gaussian Process Classification* (GPC), dilakukan pengujian menggunakan metrik evaluasi klasifikasi seperti *precision*, *recall*, *F1-score*, dan *accuracy*. Metrik-metrik ini digunakan untuk mengukur seberapa baik model dalam mengklasifikasikan data ke dalam

dua kelas, yaitu akun asli (*Real*) dan akun palsu (*Fake*). Hasil evaluasi model GPC ditunjukkan pada Tabel 1 berikut.

Tabel 1. Hasil klasifikasi *Gaussian Process Classification*

	Precision	Recall	F1-Score	Support
Real	0.85	0.98	0.91	58
Fake	0.98	0.83	0.90	58
Accuracy	0.9052			

Pada tabel 1 dapat dilihat bahwa algoritma tersebut Model GPC menghasilkan akurasi sebesar 0.9052, yang berarti sekitar 90.5% dari seluruh data berhasil diklasifikasikan dengan benar. Kelas 0 memiliki precision sebesar 0.85, menunjukkan bahwa dari semua prediksi akun asli, 85% benar-benar normal, sedangkan precision untuk kelas 1 yang berupa akun palsu mencapai 0.98, yang sangat tinggi dan menunjukkan bahwa model ini jarang salah dalam memprediksi akun palsu. Namun, dari sisi recall, GPC menunjukkan kelemahan pada kelas 1 dengan skor 0.83, yang artinya model masih melewatkan sejumlah kasus akun palsu. F1-score untuk kedua kelas masing-masing 0.91 dan 0.90, yang menunjukkan keseimbangan antara precision dan recall. Secara keseluruhan, GPC cukup akurat untuk mendeteksi akun palsu dan akun asli namun terdapat kelebihan serta kekurangan dalam hal precision maupun recall untuk kedua jenis akun.



Gambar 2. Confusion Matrix Gaussian Process Classification

Pada gambar 2 Model GPC menunjukkan performa yang solid dengan akurasi 90,8%. Kelas "Real" diklasifikasikan sangat akurat dengan 57 dari 58 sampel terdeteksi benar, menghasilkan recall tinggi 98,3%. Sementara itu, kelas "Fake" memiliki precision sebesar 97,9%, namun recall-nya sedikit lebih rendah di 82,8% karena 10 data "Fake" diklasifikasikan sebagai "Real". F1-score untuk kelas Real sebesar 0.96, dan untuk kelas Fake 0.89, mencerminkan bahwa GPC sangat kuat dalam mengenali berita asli, namun sedikit kurang sensitif dalam mendeteksi berita palsu.

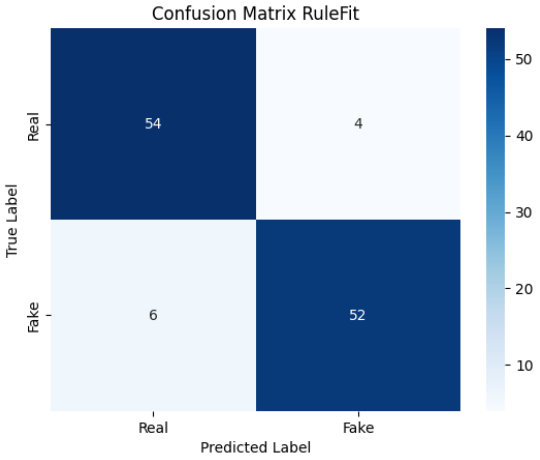
3.3. Evaluasi Model Rule Fit

Evaluasi terhadap model Rule Fit dilakukan untuk mengetahui sejauh mana model ini mampu mengklasifikasikan data dengan akurat dan seimbang antara dua kelas yang diamati, yaitu akun asli dan akun palsu. Evaluasi dilakukan menggunakan metrik klasifikasi seperti precision, recall, F1-score, dan accuracy. Hasil evaluasi dari model Rule Fit disajikan pada Tabel 2. Model RuleFit pada tabel 2 menunjukkan performa yang sangat baik dengan akurasi mencapai 91,38%, bersaing ketat dengan model lainnya. Pada kelas "Real", nilai recall sebesar 93% menunjukkan kemampuan model dalam mendeteksi berita nyata secara efektif, sementara precision sebesar 90% menunjukkan bahwa sebagian besar prediksi

berita nyata memang benar. Untuk kelas "Fake", model ini mencatat precision sebesar 93% dan recall 90%, memperlihatkan keandalan yang seimbang dalam mengidentifikasi berita palsu. Nilai F1-Score yang konsisten, yakni 0.92 untuk Real dan 0.91 untuk Fake, mengindikasikan keseimbangan yang stabil antara ketepatan dan cakupan klasifikasi. Dengan performa yang cukup merata pada kedua kelas, RuleFit menjadi pilihan yang solid untuk skenario klasifikasi yang membutuhkan hasil yang andal namun tetap seimbang dalam menangani kedua jenis data.

Tabel 2. Hasil klasifikasi Rule Fit

	Precision	Recall	F1-Score	Support
Real	0.90	0.93	0.92	58
Fake	0.93	0.90	0.91	58
Accuracy	0.9138			



Gambar 3. Confusion Matrix Rule Fit

Pada gambar 3 Model RuleFit menunjukkan performa yang cukup seimbang dengan akurasi 88,3%. Kelas "Real" diklasifikasikan dengan baik, di mana 54 dari 58 sampel diklasifikasikan dengan benar, menghasilkan recall 93,1%. Untuk kelas "Fake", model berhasil mengidentifikasi 52 dari 58 sampel dengan benar, memberikan precision tinggi 92,9%, meskipun recall-nya sedikit lebih rendah di 89,7%. F1-score untuk kelas Real dan Fake masing-masing 0.93 dan 0.91, menunjukkan bahwa RuleFit tergolong model yang cukup andal dan seimbang untuk klasifikasi berita palsu.

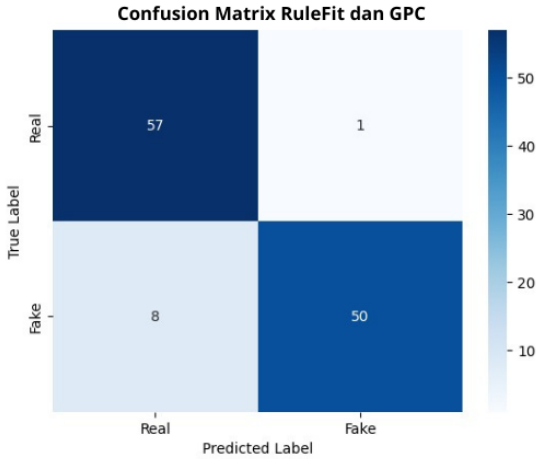
3.4. Evaluasi Model RuleFit dan GPC

Setelah mengevaluasi performa masing-masing model secara terpisah, langkah selanjutnya adalah menggabungkan model RuleFit dan GPC untuk melihat apakah kombinasi keduanya dapat meningkatkan kinerja klasifikasi. Evaluasi dilakukan menggunakan metrik yang sama, yaitu precision, recall, F1-score, dan accuracy.

Tabel 3. Hasil klasifikasi RuleFit dan GPC

	Precision	Recall	F1-Score	Support
Real	0.88	0.98	0.93	58
Fake	0.98	0.86	0.92	58
Accuracy	0.922			

Pada tabel 3 Model Rulefit dan GPC mencatat akurasi tertinggi (92,2%), mengungguli performa model RuleFit dan GPC tunggal. Meskipun precision untuk akun asli (88%) sedikit lebih rendah dibanding RuleFit, precision untuk akun palsu (98%) sangat tinggi, menunjukkan bahwa model ini hampir tidak pernah salah mengklasifikasikan akun asli sebagai palsu (false positive rendah). Dalam hal recall, model RuleFit dan GPC sangat efektif dalam mengenali akun asli (98%), artinya hampir semua akun genuine terdeteksi dengan benar. Namun, recall untuk akun palsu (86%) sedikit lebih rendah dibanding RuleFit (90%), yang berarti model kombinasi RuleFit dan GPC mungkin melewatkan beberapa akun palsu (false negative lebih tinggi). Meskipun demikian, F1-score kombinasi RuleFit dan GPC tetap yang terbaik, menunjukkan keseimbangan optimal antara precision dan recall. Kekuatan utamanya terletak pada kemampuannya meminimalkan kesalahan klasifikasi (precision tinggi untuk akun palsu) sekaligus mempertahankan sensitivitas yang baik untuk akun asli.

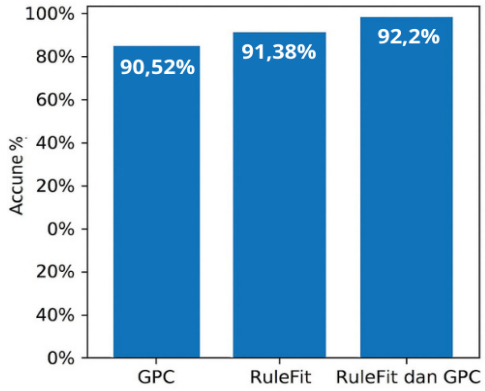


Gambar 4. Confusion Matrix RuleFit dan GPC

Pada gambar 4 Model RuleFit dan GPC menunjukkan performa kuat dengan akurasi 92,2%. Kelas "Real" diklasifikasikan sangat baik (57 dari 58 benar), menghasilkan recall 98%. Sementara kelas "Fake" memiliki precision tinggi 98%, meskipun recall-nya sedikit lebih rendah di 86%. F1-score seimbang antara kedua kelas (Real: 0.93, Fake: 0.92), menjadikan kombinasi RuleFit dan GPC yang andal dan seimbang untuk deteksi berita palsu.

3.5. Perbandingan Hasil Model

Setelah dilakukan evaluasi pada masing-masing model dan kombinasi model, langkah selanjutnya adalah membandingkan performa dari ketiganya secara visual. Perbandingan ini bertujuan untuk menunjukkan perbedaan tingkat akurasi yang dicapai oleh model GPC, RuleFit, dan kombinasi RuleFit-GPC dalam mendeteksi akun palsu menggunakan dataset *Instagram Detecting fake accounts* [18]. Gambar 5 berikut menyajikan perbandingan akurasi dari ketiga pendekatan (model GPC, RuleFit, dan kombinasi RuleFit-GPC) dalam prosentase.



Gambar 5. Diagram Akurasi RuleFit dan GPC

Hasil pengujian pada gambar 5 menunjukkan bahwa model kombinasi RuleFit dan GPC memberikan performa terbaik dengan akurasi mencapai 92,24%, presisi sebesar 98,00%, recall 86,00%, dan F1-score 92,00% untuk memprediksi akun palsu. Dibandingkan model individu, RuleFit memperoleh akurasi sebesar 91,38%, presisi 93,00%, recall 90,00%, dan F1-score 91,00%. Adapun model GPC mencatatkan akurasi 90,52%, presisi 98,00%, recall 83,00%, dan F1-score 90,00%. Dari hasil tersebut, terlihat bahwa model kombinasi mampu memberikan hasil yang lebih stabil dan akurat, terutama dalam meningkatkan presisi deteksi akun palsu. Hal ini mengindikasikan bahwa pemanfaatan fitur berbasis aturan dari RuleFit yang dikombinasikan dengan pendekatan klasifikasi berbasis probabilistik dari GPC mampu meningkatkan kinerja sistem secara keseluruhan dalam mengidentifikasi akun palsu secara lebih efektif.

4. Kesimpulan

Penelitian ini menggunakan pendekatan *combined features* yang menggabungkan model RuleFit dan Gaussian Process Classifier (GPC) untuk membuat sistem pendeteksi akun palsu di Instagram yang lebih baik. Dataset penelitian terdiri dari 576 akun dengan berbagai ciri seperti panjang username, kesamaan nama, jumlah pengikut, dan pola aktivitas pengguna. Hasil pengujian menunjukkan bahwa model *combined features* (RuleFit dan GPC), yang menggunakan fitur berbasis aturan dari RuleFit dan dilatih dengan GPC, menghasilkan akurasi tertinggi sebesar 92,2%, lebih baik dibandingkan model individual RuleFit (akurasi 91,38%) dan model GPC (akurasi 90,52%). Keunggulan utama pendekatan kombinasi ini terlihat pada tingkat presisi untuk deteksi akun palsu (98%), yang berarti sangat jarang terjadi kesalahan mengidentifikasi akun asli sebagai palsu, sekaligus mempertahankan tingkat recall tinggi (98%) untuk akun asli. Ini membuktikan bahwa penggabungan fitur berbasis aturan dengan klasifikasi probabilistik berhasil menciptakan sistem deteksi yang lebih seimbang dan dapat diandalkan. Hasil penelitian menunjukkan bahwa metode *combined features* ini memberikan solusi praktis untuk meningkatkan keamanan media sosial dengan memperbaiki ketepatan dan keakuratan sistem pendeteksi akun palsu secara otomatis.

Daftar Pustaka

[1] Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., & Tesconi, M. (2015). Fame for sale: Efficient detection of fake Twitter followers. Decision Support Systems, 80, 56-71.

[2] S. Lopez-Joya, J. A. Diaz-Garcia, M. D. Ruiz, and M. J. Martin-Bautista, 'Bot Detection in Twitter: An Overview', in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2023. doi: 10.1007/978-3-031-42935-4_11.

[3] Refalia, Salsa, Boedi P. (2024). Pembuktian Akun Palsu terhadap Selebgram yang Diduga Melakukan Promosi Judi Online. Rewang Rencang: Jurnal Hukum Lex Generalis. 5(7), 1-14

[4] Kudugunta, S., & Ferrara, S. (2018). Deep Neural Networks for Bot Detection. arXiv preprint arXiv:1802.04289, 1-10.

[5] H2O.ai. (2023). RuleFit Algorithm. Retrieved from https://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science/rulefit.html

[6] Mahesa, A. A., Wisesty, U. N., & Adiwijaya. (2019). *Klasifikasi Keadaan Mata berdasarkan Sinyal Electroencephalography Menggunakan Gaussian Process*. e-Proceeding of Engineering, 6(2), 9069–9077.

[7] Ramadhan, J. A., Haniva, D. T., & Suharso, A. (2023). Systematic Literature Review Penggunaan Metodologi Pengembangan Sistem Informasi Waterfall, Agile, dan Hybrid. JIEET: Journal Information Engineering and Educational Technology, 7(1), 36–42. ISSN: 2549-869X.

[8] Kanamori, K. (2023). Learning Locally Interpretable Rule Ensemble. Artificial Intelligence Laboratory, Fujitsu Ltd.

[9] Rasmussen, C., E., & Williams, C., K., I. (2016). Gaussian Processes for Machine Learning. The MIT Press

[10] Friedman, J. H., & Popescu, B. E. (2008). Predictive learning via rule ensembles. The Annals of Applied Statistics, 2(3), 916–954.

[11] Kanamori, K. (2023). Learning Locally Interpretable Rule Ensemble. arXiv preprint arXiv:2306.11481, 1–15.

[12] Ebner, L., Nalenz, M., ten Teije, A., van Harmelen, F., & Augustin, T. (2021). *Expert RuleFit: Complementing Rule Ensembles with Expert Knowledge*. Vrije Universiteit Amsterdam dan University of Munich.

[13] Luo, C., Li, S., Zhao, Q., Ou, Q., Huang, W., Ruan, G., Liang, S., Liu, L., Zhang, Y., & Li, H. (2022). RuleFit-Based Nomogram Using Inflammatory Indicators for Predicting Survival in Nasopharyngeal Carcinoma, a Bi-Center Study, Journal of Inflammation Research.

[14] C. E. Rasmussen dan C. K. I. Williams, Gaussian Processes for Machine Learning, MIT Press, 2006.

[15] Rodrigues, F., Pereira, F. C., & Ribeiro, B. (2014). *Gaussian Process Classification and Active Learning with Multiple Annotators*. Proceedings of the 31st International Conference on Machine Learning (ICML), JMLR: W&CP Volume 32, Beijing, China.

[16] Frohlich, B., Rodner, E., Kemmler, M., & Denzler, J. (2010). Efficient Gaussian Process Classification Using Random Decision Forests. Mathematical Theory of Pattern Recognition.
[17] Apriliyani, E., & Salim, Y. (2022). *Analisis Performa Metode Klasifikasi Naïve Bayes Classifier pada Unbalanced Dataset*. Indonesian Journal of Data and Science (IJODAS), 3(2), 47–54.
[18] Notra, J. (2023). Instagram_Detecting fake accounts. kaggle. https://www.kaggle.com/datasets/jasvindernotra/instagram-detecting-fake-accounts
[19] Kaviya, P., Sudharsana, I., & Hariesh, B. B. C. (2025). Detecting deceptive identities: A machine learning approach to unveiling fake profiles on social media. *SN Computer Science*, *6*(16).
[20] Alfiana, R. (2020). Penerapan Naïve Bayes Classifier Untuk Klasifikasi Akun Online Shop Instagram Yang Dicurigai Penipuan. Tugas Akhir. Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau.
[21] Herga, M. R. (n.d.). Implementasi Text Mining Sistem Klasifikasi dan Pencarian Naive Bayes Classifier.