

Identifikasi Arsip Digital Dengan Pendekatan *Machine Learning*

I N T I S A R I

Perkembangan teknologi khususnya kamera memberikan banyak keuntungan dalam pengelolaan arsip. Sebaliknya, kemudahan dalam melakukan manipulasi arsip digital untuk kepentingan tertentu merupakan permasalahan dalam otentikasi arsip digital. Tujuan dari penelitian ini adalah identifikasi sumber digital terutama kamera sebagai salah satu alat yang digunakan dalam proses digitalisasi arsip. Metodologi penelitian dengan mengkombinasikan klustering dan klasifikasi *Logistic Regression* dalam mesin pembelajaran (*machine learning*) untuk menentukan 6 merek kamera *handphone*. Total data yang digunakan dalam eksperimen ini sebanyak 2400 arsip citra digital. Hasil eksperimen menunjukkan bahwa tingkat identifikasi akurasi klasifikasi sebesar 99%. Hal ini menunjukkan bahwa studi ini sangat efektif untuk menentukan otentikasi arsip digital terutama menentukan *metadata* sumber kamera.

A B S T R A C T

Technological developments, especially cameras, provide many advantages in archive management. On the other hand, the ease of manipulating digital archives for certain purposes is a problem in authenticating digital archives. The purpose of this study is to identify digital sources, especially cameras as one of the tools used in the process of digitizing archives. The research methodology combines clustering and classification of Logistic Regression in machine learning to determine 6 brands of cellphone cameras. The total data used in this experiment is 2400 digital image files. The experimental results show that the identification rate of classification accuracy is 99%. This shows that this study is very effective for determining digital archive authentication, especially determining camera source metadata.

PENULIS

Danish Faiq Ibad Yuadi

SMAN 2 Jombang
danishfaiq170@gmail.com

KATA KUNCI

arsip digital, otentikasi arsip,
logistic regression, mesin

KEY WORDS

*electronic records digital,
records authentication,
logistic regression, machine
learning*

PENGANTAR

Latar Belakang Masalah

Kecanggihan teknologi digitalisasi dan *editing* telah membuat publik sadar bahwa betapa mudahnya dokumen digital dapat dimanipulasi tanpa terdeteksi. Para ahli digital forensik sangat yakin bahwa dalam dekade ini, baik publik maupun pengadilan akan memiliki sedikit kepercayaan pada arsip digital yang tidak memiliki jaminan autentikasi terbaik yang dapat diberikan oleh teknologi. Maka dari itu, arsiparis dapat menerapkan teknik-teknik digitalisasi arsip yang aman dari perubahan. Berbagai macam teknik digunakan untuk memberikan pengamanan pada hasil digitalisasi arsip. Pengamanan hendaknya juga diberikan terhadap jenis dan alat digitalisasi misalnya kamera, *scanner* dan alat digital lainnya. Meskipun teknik yang sama dapat diterapkan pada arsip digital, namun sifat arsip digital memaksa kita untuk menggunakan teknik tambahan untuk memastikan keaslian dan integritasnya. Teknik digital dilakukan dengan memberikan jaminan yang jauh lebih kuat daripada teknik yang digunakan untuk arsip media kertas. Selain itu, arsip digital berpotensi lebih rentan terhadap pemalsuan dan gangguan berupa penambahan, penghapusan, atau pengubahan arsip secara sengaja yang sulit dideteksi.

Kamera sebagai salah satu alat penciptaan arsip untuk digitalisasi arsip akhir dekade ini mendapat perhatian luas karena perannya yang penting dalam masalah keamanan dan hukum. Masalah penetapan asal usul media digital yang diperoleh melalui perangkat pencitraan menjadi penting setiap kali konten digital arsip yang disajikan dan digunakan sebagai alat bukti di pengadilan. Identifikasi sumber kamera adalah proses menentukan perangkat atau model kamera mana yang telah digunakan untuk menangkap citra.

Untuk menentukan jenis atau merek kamera dilakukan dengan berbagai macam metode. Teknik identifikasi sensor kamera dapat diwujudkan dengan menganalisis *photo response non uniformity* (PRNU) yang dipahami sebagai sidik jari kamera yang unik. PRNU juga sering disebut sebagai *sensor pattern noise* (SPN) atau *residual noise* Lukas dkk (2006: 205-214). Perbedaan *noise* inilah yang didapat dipakai untuk membedakan citra gambar arsip digital yang dihasilkan oleh suatu kamera. Demikian juga perbedaan *pixel* yang dimiliki oleh setiap kamera dapat dijadikan sebagai pedoman untuk membedakan hasil citra digital. Selain itu, pendekatan aktif seperti tanda tangan digital Friedman (1993:905–910) dan *watermarking digital* Podilchuk (2001:33–46) memiliki beberapa

keterbatasan karena informasi khusus harus disematkan selama pembuatan gambar, yang meningkatkan biaya produksi kamera digital, dan kredibilitas informasi yang tertanam di dalamnya. Analisis forensik hanya memiliki citra tersangka yang mereka miliki dan harus mengeksplorasi informasi yang berguna dari gambar itu untuk mengumpulkan bukti forensik, melacak perangkat dan mendeteksi setiap tindakan manipulasi. Pendekatan pasif didasarkan pada jejak internal yang ditinggalkan oleh kamera dalam gambar yang diberikan. Oleh karena itu, untuk menentukan sumber kamera atau merek kamera yang menghasilkan arsip digital, penulis melakukan pendekatan digital forensik dengan menggunakan kombinasi metodologi *machine learning*. Implementasi dari kombinasi klasifikasi dan klustering dalam *Orange Data Mining* diharapkan mampu mengatasi problem identifikasinya.

Rumusan Masalah

Berdasarkan penjelasan yang dipaparkan pada pendahuluan tersebut, yaitu (1) adanya kesulitan untuk melakukan identifikasi sumber gambar digital dari suatu arsip untuk menentukan arsip asli atau telah dimodifikasi, dan (2) banyaknya jenis kamera ataupun *scanner* yang dapat digunakan untuk proses digitalisasi suatu arsip, maka pertanyaan

penelitian yang dapat dirumuskan berupa: bagaimana identifikasi arsip digital untuk mengetahui sumber kamera atau jenis kamera yang digunakan dalam proses digitalisasi arsip?

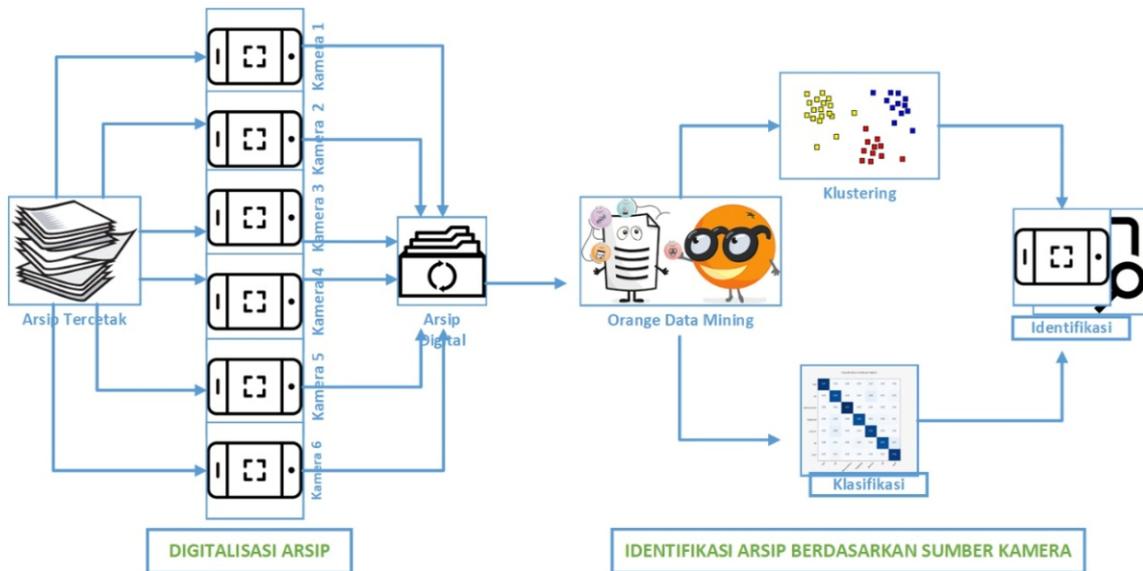
Tujuan Penelitian

Penelitian ini memfokuskan pada penerapan ide mutakhir dan solusi berupa forensik digital yang mengkombinasikan algoritma klustering dan klasifikasi dalam mesin pembelajaran berdasarkan studi sebelumnya. Studi ini juga bertujuan membuat terobosan baru dalam membangun solusi yang disesuaikan dengan fungsionalitas yang kuat dan kemampuan kreatif dalam identifikasi sumber digital terutama kamera sebagai salah satu alat yang digunakan dalam proses penciptaan arsip.

Metodologi Penelitian

Studi ini menggunakan *Orange Data Mining* untuk identifikasi arsip digital. Metode ini bekerja sebagaimana seperti yang dipaparkan pada gambar 1. Untuk *file* foto, diproses terlebih dahulu sebelum ekstraksi fitur, dan fitur yang diekstraksi kemudian dianalisis menggunakan klustering dan diklasifikasikan berdasarkan *Logistic Regression*.

Proses identifikasi arsip digital dilakukan dengan beberapa langkah. Pertama, arsip tercetak yang diambil dari



Gambar 1. Proses identifikasi arsip digital berdasarkan sumber kamera

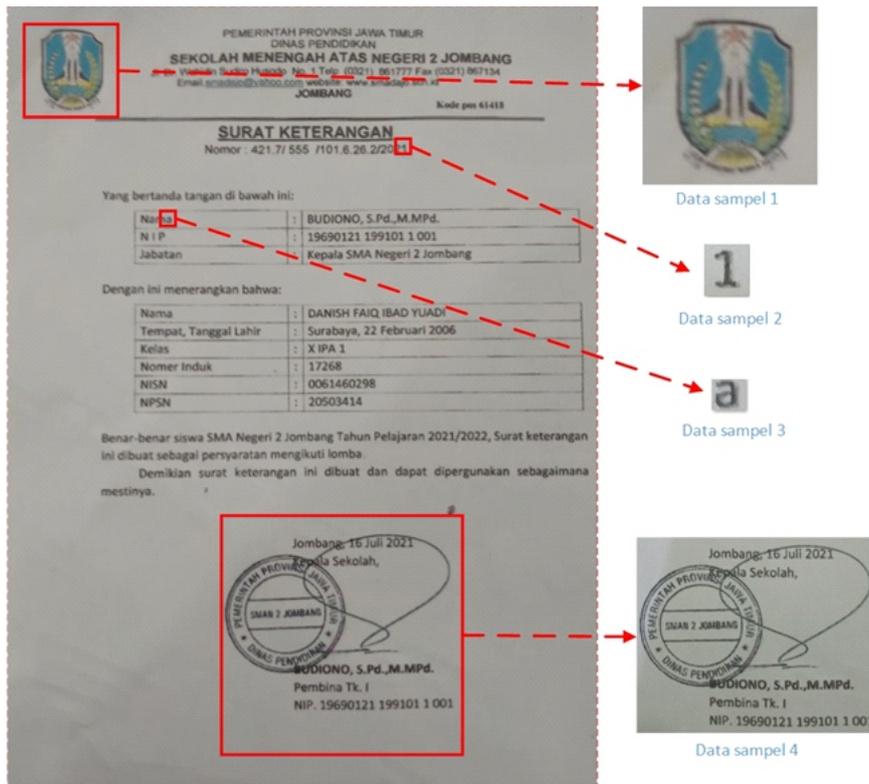
Sumber: koleksi peneliti, 2022

arsip SMAN 2 Jombang seperti pada gambar 2 didigitalisasikan dengan menggunakan kamera *handphone* dengan merek yang berbeda. Peneliti menggunakan merek *handphone* Oppo A12, Oppo A74, Oppo F1s, Realme 5, Vivo V17 Pro, dan Vivo Y65. Jumlah data yang diproses sebanyak 2.400 data sampel arsip digital dari 6 kamera *handphone* yang berbeda. Selanjutnya, arsip yang sudah dalam bentuk digital diproses dengan *Orange Data Mining*. Untuk arsip digital, diproses terlebih dahulu dengan melakukan *cropping* gambar sebelum ekstrak. Kemudian citra yang telah diekstraksi diunggah ke dalam *Orange* pada *Image Embedding* dan dianalisis menggunakan *Test and Score* dengan *Logistic Regression* sebagai *learner* untuk klasifikasi. Setelah berhasil diproses, hasil citra tersebut dikelompokkan

menggunakan *Clustering* dan diklasifikasikan menggunakan *Confusion Matrix*. Pada langkah terakhir, *scatter plot* digunakan sebagai *widget* yang menampilkan hasil variabel yang telah dianalisis dalam bentuk diagram.

Kerangka Pemikiran

Dengan semakin mudahnya untuk memperoleh perangkat murah dan mudah digunakan dalam akuisisi data visual, maka hampir setiap orang saat ini memiliki kemungkinan untuk merekam, menyimpan, dan berbagi sejumlah besar citra digital. Pada saat yang sama, ketersediaan besar perangkat lunak pengedit citra mempermudah pengguna untuk mengubah konten citra, atau membuat yang baru, sehingga merusak dan memalsukan konten visual dapat dilakukan oleh orang yang bukan ahli.



Gambar 2. Data sampel arsip digital
Sumber: arsip SMAN 2 Jombang

Akhirnya, perangkat lunak saat ini memungkinkan untuk membuat grafik komputer foto realistis yang tidak dapat dibedakan oleh pemirsa dari gambar fotografi Meyer (1986:30-50) atau juga menghasilkan konten visual lainnya.

Forensik digital secara sederhana dapat didefinisikan sebagai disiplin ilmu yang menggabungkan unsur-unsur hukum dan ilmu komputer untuk mengumpulkan dan menganalisis data dari sistem komputer, jaringan, komunikasi nirkabel, dan perangkat penyimpanan dengan cara yang dapat diterima sebagai bukti di pengadilan. Forensik digital adalah cabang ilmu forensik yang muncul dari komunitas penegak hukum pada 1980an

sebagai seperangkat metode yang diterapkan untuk mengumpulkan, mengambil, menganalisis, dan melaporkan informasi yang disimpan di perangkat digital, seringkali dalam kaitannya dengan penyelidikan hukum Reith, Carr & Gunsch (2002: 1-12). Kriteria penilaian untuk menentukan validitas dan akurasi metode atau alat forensik adalah dibangun di atas standar Ryan & Shpantzer (2002). Ini sama pentingnya untuk dokumentasi yang tepat, pengembangan, dan peningkatan koleksi yang ada dan membuatnya tersedia untuk kurator yang ingin menggunakannya sesuai dengan standar klasifikasi (I. Yuadi, M. D. Artaria, S. & A. Taufiq Asyhari,

2021:2). Secara khusus, ilmu forensik digital muncul dalam satu dekade terakhir sebagai respons terhadap eskalasi kejahatan yang dilakukan dengan menggunakan perangkat elektronik sebagai instrumen yang digunakan untuk melakukan kejahatan atau sebagai gudang bukti yang terkait dengan kejahatan (Amerini, 2010: 21-23).

Setiap hari, semakin banyak kamera digital ditemukan oleh perusahaan teknologi. Akibatnya, kamera digital menjadi lebih terjangkau bagi konsumen. Saat ini *handphone* pun sudah dilengkapi dengan kamera digital. Hal ini semakin meningkatkan jumlah individu yang memiliki perangkat pengambilan gambar. Akibatnya, ribuan gambar dibuat setiap hari dengan beberapa di antaranya menangkap momen kritis dalam waktu seperti kejahatan. Gambar-gambar ini dapat digunakan di pengadilan sebagai bukti untuk menunjukkan hubungan antara tersangka dan penjahat (Wen, 2006:1-11). Namun, masalah utama dalam menggunakan gambar digital sebagai bukti di pengadilan adalah bahwa gambar digital mudah dibuat dan dimanipulasi tanpa meninggalkan jejak modifikasi yang jelas. Manipulasi bukti menyebabkan kredibilitas dan keaslian citra digital dipertanyakan Piva (2013:1-22). Oleh karena itu, kami membutuhkan lebih banyak alat dan aplikasi untuk mengatasi masalah verifikasi keaslian suatu gambar (Wen, 2006:1-11).

Logistic Regression adalah metode klasifikasi statistik yang menyesuaikan data dengan fungsi logistik. *Orange* menyediakan berbagai peningkatan metode, seperti pemilihan variabel secara bertahap dan penanganan variabel konstan dan singularitas. Algoritma pembelajaran untuk regresi logistik yang mengimplementasikan seleksi subset fitur bertahap seperti yang dijelaskan dalam Regresi Logistik Terapan (Hosmer & Lemeshow, 2000). Ini adalah algoritma yang dapat digunakan untuk regresi dan klasifikasi tetapi banyak digunakan untuk tugas klasifikasi. Variabel respon biner dari salah satu kelas digunakan untuk memprediksi variabel kategori dengan bantuan variabel dependen. Pertimbangan ada dua kelas dan titik data baru harus diperiksa kelas mana yang akan menjadi miliknya. Kemudian algoritma menghitung nilai probabilitas yang berkisar dari 0 dan 1. Dalam studi ini, melakukan prediksi pada data uji berdasarkan model yang telah dilatih menggunakan data latih. Sebagian besar waktu, pemodelan data dan bagian prediksi adalah yang paling menarik karena memerlukan pemikiran dan penyesuaian parameter yang mendasarinya untuk meningkatkan hasil. Toles menyebutkan (2016: 533-534) keluaran dengan lebih dari dua nilai dimodelkan dengan regresi logistik multinomial dan jika beberapa kategori

diurutkan dengan regresi logistik ordinal (misalnya model logistik ordinal odds proporsional).

PEMBAHASAN

Pembahasan menjelaskan hasil yang dilakukan berdasarkan jenis tipe *handphone* dalam eksperimen identifikasi surat digital. Hasil tersebut disajikan dalam bentuk pendekatan algoritma *machine learning*.

Hasil Eksperimen

Pada bagian ini, penulis menjelaskan langkah-langkah eksperimen dan hasilnya. Dari mulai *pre-processing* hingga hasil klasifikasi untuk identifikasi arsip digital dengan kinerja yang komprehensif. Data yang kami kumpulkan berasal dari surat digital SMAN 2 Jombang. Berikut kami jelaskan langkah-langkah serta hasil eksperimen. Langkah-langkah dalam identifikasi surat adalah sebagai berikut:

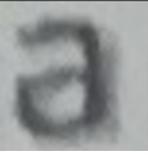
1. Pre-processing

Pre-processing adalah proses mengubah data mentah menjadi format yang dapat dipahami seperti yang tertera pada Tabel 1. Proses ini sangat penting dalam *data mining* karena kualitas data yang akan diproses harus diperiksa sebelum diterapkan pada algoritma *data mining* untuk mencapai hasil analisis berkualitas lebih baik.

2. Orange Data Mining

Untuk memproses data yang telah melalui proses *cropping* maka diperlukan *software orange data mining*. *Orange* merupakan aplikasi yang dirancang untuk pembelajaran mesin dan *data mining*. Pada Gambar 3, *Orange* menawarkan program yang mudah dipakai bagi peneliti yang mempelajari pembelajaran mesin untuk membuat prototipe algoritma baru dan prosedur eksperimental. *Orange* juga menyediakan kerangka kerja pemrograman visual dengan penekanan pada interaksi dan kombinasi kreatif dari komponen visual untuk analisis data yang eksploratif.

Tabel 1. Surat Digital yang Telah Melalui Proses *Cropping*

Data sampel 1	Data sampel 2	Data sampel 3	Data sampel 4
			

Sumber: data peneliti, 2022

3. Image Embedding

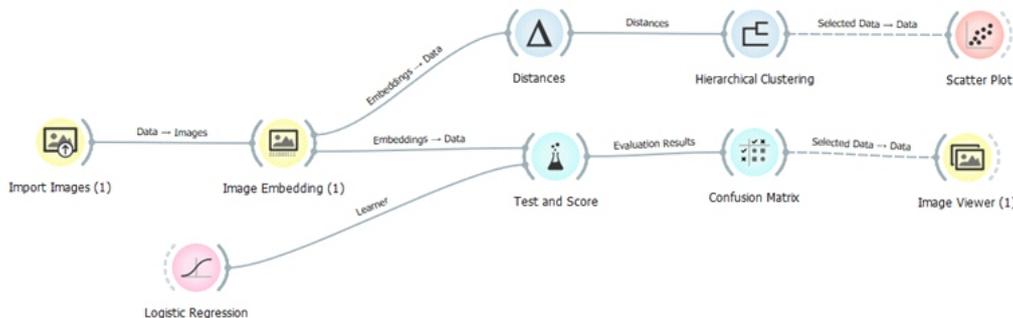
Widget ini membaca serta menyaring input untuk mempelajari cara memprediksi dan mengklasifikasikan data yang telah dipilih. Pada Gambar 4, *SqueezeNet* digunakan sebagai *deep learning* model yang dapat cepat mengenali gambar yang dilatih di *ImageNet*. *Widget Image Embedding* akan mengeluarkan sebuah *enhanced data table* dengan tambahan kolom (*image descriptor*). Gambar dapat di import menggunakan *widget Import Images* atau sebagai *path* ke *image* di sebuah file *spreadsheet*.

4. Klastering

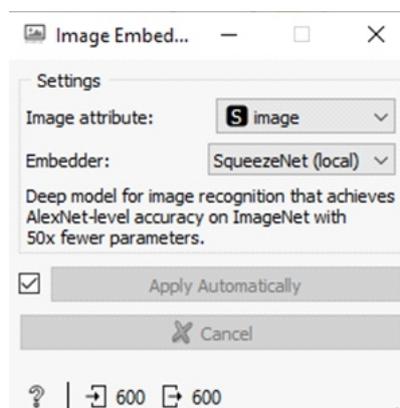
Proses ini dilakukan dengan mempartisi sekumpulan data (atau objek) menjadi satu set sub-kelas bermakna, menjadi kluster seperti yang ditunjukkan pada Gambar 5, dapat membantu memahami pengelompokan atau struktur alami dalam kumpulan data.

5. Klasifikasi

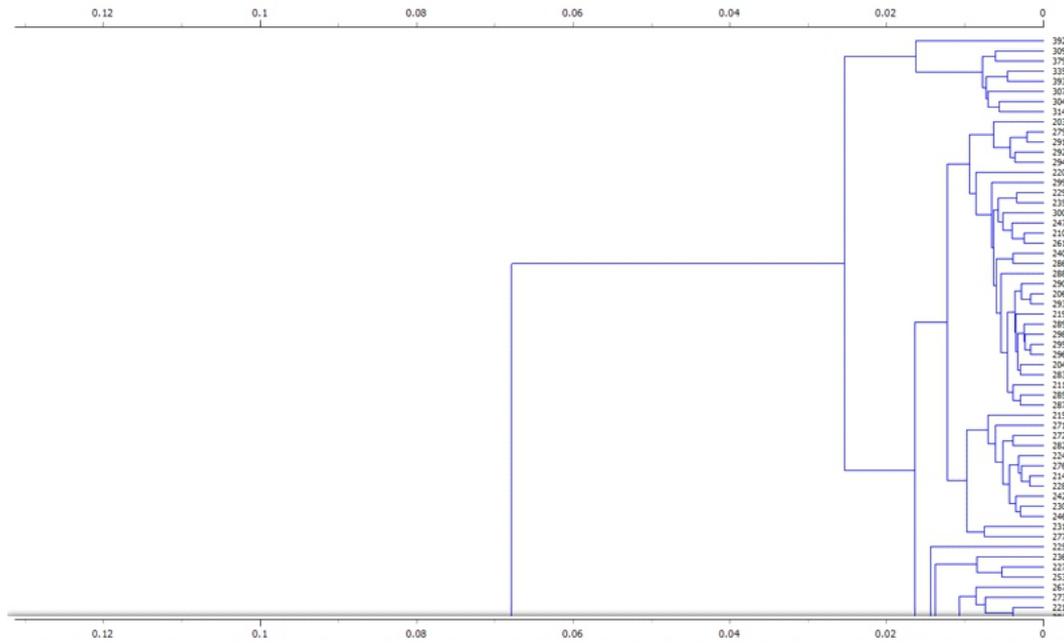
Logistic Regression digunakan sebagai input metode klasifikasi statistik untuk menganalisis data yang telah diproses. *Test and Score* bekerja sebagai elemen untuk menunjukkan hasil analisis dari input dan ditunjukkan sebagai tabel evaluasi



Gambar 3. Metode *data mining* untuk identifikasi arsip digital



Gambar 4. *Image Embedding* untuk gambar yang dipotong
Sumber: data peneliti, 2022



Gambar 5. Hierarchical Clustering untuk arsip digital
Sumber: data peneliti, 2022

hasil pengujian algoritma klasifikasi. Tabel *Confusion Matrix* biasanya mendapatkan hasil evaluasinya dari *Test and Score* dan memperlihatkan seberapa efektif model *machine learning* yang digunakan. Dengan memberikan jumlah/proporsi instansi antara kelas yang diprediksi dan yang sebenarnya, orang dapat mengamati data spesifik manakah yang salah diklasifikasikan dan bagaimana caranya.

Hasil evaluasi yang ditunjukkan pada Tabel 2 menandakan bahwa semua data sampel 1 telah dianalisis dengan tingkat akurasi 100% untuk semua jenis *handphone*. Pada Tabel 3, Sebagian besar klasifikasi data sampel menunjukkan hasil yang benar, akan tetapi beberapa data yang seharusnya diklasifikasikan sebagai Vivo Y65 diprediksi sebagai citra dari

Oppo A12 sebanyak 5 data dan Realme 5 sebanyak 1 data. Sedangkan pada hasil data Oppo A12 mengungkapkan sebanyak 2 data diprediksi sebagai citra dari Vivo Y65. Hasil analisis klasifikasi data sampel 3 pada Tabel 4, hampir semua sampel menunjukkan hasil yang akurat, hanya 1 data yang sebenarnya sebagai citra dari Vivo V17 Pro diprediksi sebagai citra dari Realme 5. Sedangkan Tabel 5 yang memproses data sampel 4, hasil klasifikasi hampir sepenuhnya akurat. *Logistic Regression* berhasil mengidentifikasi dengan benar 99 citra dari Vivo Y65 dan 1 citra lainnya tidak berhasil teridentifikasi dengan benar sebagai Oppo A12.

6. Scatter Plot

Hasil *scatter plot* menunjukkan data yang telah dianalisis sebelumnya dan hasil tersebut ditunjukkan di *Image Viewer* pada *Orange*.

Tabel 2. Proporsi Prediksi *Confusion Matrix* Data Sampel 1

		Predicted						Σ
		Oppo A12	Oppo A74	Oppo F1s	Realme 5	Vivo V17 Pro	Vivo Y65	
Actual	Oppo A12	100.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	100
	Oppo A74	0.0 %	100.0 %	0.0 %	0.0 %	0.0 %	0.0 %	100
	Oppo F1s	0.0 %	0.0 %	100.0 %	0.0 %	0.0 %	0.0 %	100
	Realme 5	0.0 %	0.0 %	0.0 %	100.0 %	0.0 %	0.0 %	100
	Vivo V17 Pro	0.0 %	0.0 %	0.0 %	0.0 %	100.0 %	0.0 %	100
	Vivo Y65	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	100.0 %	100
Σ		100	100	100	100	100	100	600

Sumber: data peneliti, 2022

Tabel 3. Proporsi Prediksi *Confusion Matrix* Data Sampel 2

		Predicted						Σ
		Oppo A12	Oppo A74	Oppo F1s	Realme 5	Vivo V17 Pro	Vivo Y65	
Actual	Oppo A12	95.1 %	0.0 %	0.0 %	0.0 %	0.0 %	2.1 %	100
	Oppo A74	0.0 %	100.0 %	0.0 %	0.0 %	0.0 %	0.0 %	100
	Oppo F1s	0.0 %	0.0 %	100.0 %	0.0 %	0.0 %	0.0 %	100
	Realme 5	0.0 %	0.0 %	0.0 %	99.0 %	0.0 %	0.0 %	100
	Vivo V17 Pro	0.0 %	0.0 %	0.0 %	0.0 %	100.0 %	0.0 %	100
	Vivo Y65	4.9 %	0.0 %	0.0 %	1.0 %	0.0 %	97.9 %	100
Σ		103	100	100	101	100	96	600

Sumber: data peneliti, 2022

Tabel 4. Proporsi Prediksi *Confusion Matrix* Data Sampel 3

		Predicted						Σ
		Oppo A12	Oppo A74	Oppo F1s	Realme 5	Vivo V17 Pro	Vivo Y65	
Actual	Oppo A12	100.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	100
	Oppo A74	0.0 %	100.0 %	0.0 %	0.0 %	0.0 %	0.0 %	100
	Oppo F1s	0.0 %	0.0 %	100.0 %	0.0 %	0.0 %	0.0 %	100
	Realme 5	0.0 %	0.0 %	0.0 %	99.0 %	0.0 %	0.0 %	100
	Vivo V17 Pro	0.0 %	0.0 %	0.0 %	1.0 %	100.0 %	0.0 %	100
	Vivo Y65	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	100.0 %	100
Σ		100	100	100	101	99	100	600

Sumber: data peneliti, 2022

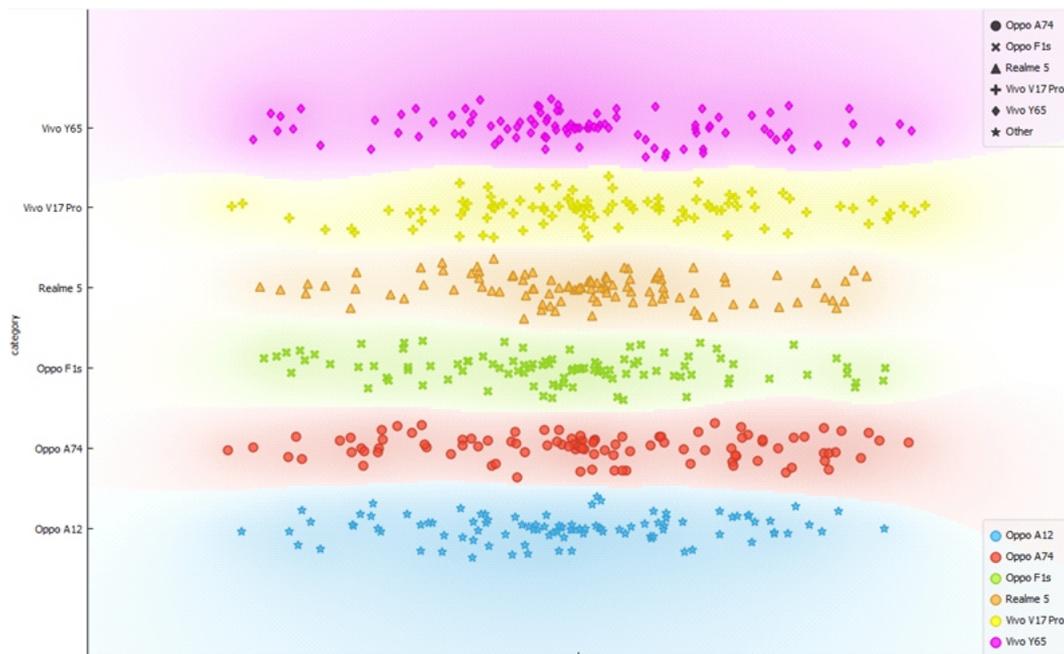
Tabel 5. Proporsi Prediksi *Confusion Matrix* Data Sampel 4

		Predicted						Σ
		Oppo A12	Oppo A74	Oppo F1s	Realme 5	Vivo V17 Pro	Vivo Y65	
Actual	Oppo A12	99.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	100
	Oppo A74	0.0 %	100.0 %	0.0 %	0.0 %	0.0 %	0.0 %	100
	Oppo F1s	0.0 %	0.0 %	100.0 %	0.0 %	0.0 %	0.0 %	100
	Realme 5	0.0 %	0.0 %	0.0 %	100.0 %	0.0 %	0.0 %	100
	Vivo V17 Pro	0.0 %	0.0 %	0.0 %	0.0 %	100.0 %	0.0 %	100
	Vivo Y65	1.0 %	0.0 %	0.0 %	0.0 %	0.0 %	100.0 %	100
Σ		101	100	100	100	100	99	600

Sumber: data peneliti, 2022

Berdasarkan Gambar 6, pola titik memungkinkan penentuan apakah ada hubungan atau korelasi antara dua variabel kontinu. Jika ada hubungan, *Scatter Plot* menunjukkan arahnya dan apakah itu hubungan linier atau melengkung. *Scatter Plot* mengandung elemen-elemen sumber arsip digital yang dihasilkan dari 6 kamera berbeda yang

menunjukkan bahwa sumbu X yang mewakili nilai variabel kontinu yaitu variabel independen ketika sumber kamera dapat diklasifikasikan. Sedangkan sumbu Y mewakili nilai variabel kontinu di mana secara tradisional sebagai variabel dependen.



Gambar 6. *Scatter Plot* berdasarkan kelompok kamera *handphone*

Sumber: data peneliti, 2022

Diskusi

Tingkat akurasi dari hasil eksperimen yang dilakukan dalam studi ini mencapai rata-rata tingkat akurasi 99 %. Hasil ini lebih baik 1 % dari penelitian identifikasi merek kamera yang dilakukan oleh di Freire-Obregon dkk. (2017: 86-91) yang menggunakan metode *Convolutional neural networks* (CNN) dengan lapisan konvolusional disusun oleh satu set filter lolos tinggi. Mereka menggunakan obyek kamera *smartphone* modern yaitu: iPhone 5, Samsung Galaxy S4 dan Samsung Galaxy Tab II digunakan untuk klasifikasi. Meskipun akurasi klasifikasi sangat tinggi dan mencapai sekitar 98% untuk setiap sumber merek kamera, kumpulan data gambar hanya berdasarkan tiga perangkat yang disebutkan tidak dapat dianggap memiliki akurasi tinggi. Demikian juga, dibandingkan dengan pendekatan berbasis CNN lainnya, metode yang diusulkan oleh Bondi dkk. (2017: 259-263) melakukan identifikasi beberapa model kamera dalam satu perbandingan. Inti dari metode ini mirip dengan gambar yang diproses pada penelitian sebelumnya dimana tingkat akurasi identifikasi merek kamera juga memiliki akurasi yang tinggi mencapai sekitar 98%.

Namun demikian, penelitian yang dilakukan oleh Chen dkk. (2017: 4337-4341) menggunakan *residual neural* (ResNet) dimana sifat-sifat sistem lensa

digunakan untuk melatih jaringan menghasilkan akurasi yang lebih rendah, sedangkan Yang dkk. (2019: 195-204) menggunakan konsep perpaduan *content-adaptive fusion residual networks* yang setiap kategori *fusion residual network* dilatih dengan *transform learning* akurasi klasifikasi identifikasi kamera adalah tidak lebih rendah dari 98%.

SIMPULAN

Otentikasi digital arsip yang dihasilkan dari suatu teknologi penciptaan arsip yaitu kamera dapat mewakili pengetahuan penting dalam investigasi dan pemeriksaan keaslian suatu arsip. Dalam dua dekade terakhir, banyak algoritma telah diusulkan untuk memperoleh informasi tentang sumber kamera yang lebih akurat. Model klasifikasi dengan pendekatan klasifikasi *logistic regression* dalam studi ini menghasilkan tingkat akurasi hingga 99 % bahkan ada sebagian jenis kamera dapat mencapai akurasi 100%. Perolehan dari suatu dataset dan besarnya jumlah data memungkinkan menyebabkan tingkat akurasi dalam penelitian ini. Oleh karena itu, penelitian kedepan dapat dilakukan dengan melakukan penelitian pada objek yang berbeda dengan jumlah kamera yang lebih banyak. Demikian juga dapat dilakukan untuk jenis alat digitalisasi yang lain.

DAFTAR PUSTAKA

- Bondi, L., Baroffio, L., Guera, D., Bestagini, P., Delp, E.J., Tubaro, S., (2017). First Steps Toward Camera Model Identification with Convolutional Neural Networks. *IEEE Signal Processing Letters*, 24(3), 259-263.
- Chen, Y., Huang, Y., Ding, X. (2017). Camera Model Identification with Residual Neural Network. *IEEE International Conference on Image Processing, ICIP 2017*, Beijing, China, September 17-20, 2017, 4337-4341.
- Freire-Obregon, D., Narducci, F., Barra, S., Santana, M.C. (2019). Deep Learning for Source Camera Identification on Mobile Devices. *Pattern Recognition Letters*, 126, 86-91.
- Friedman, G. L. (1993) The Trustworthy Digital Camera: Restoring Credibility to the Photographic Image. *IEEE Transactions on Consumer Electronics*, 39(4), 905-910.
- Hosmer, D.W. & Lemeshow, S. (2000). *Applied Logistic Regression. 2nd Edition*, New York: John Wiley & Sons, Inc.
- Irene Amerini, Roberto Caldelli, Vito Cappellini, Francesco Picchioni & Alessandro Piva. (2010). Estimate of PRNU Noise Based on Different Noise Models for Source Camera Identification. *International Journal of Digital Crime and Forensics*, 2(2), 21-33.
- Lukas, J., Fridrich, J.J., Goljan, M. (2006). Digital Camera Identification from Sensor Pattern Noise. *IEEE Transactions on Information Forensics and Security*, 1(2), 205-214.
- Meyer, G. W., H. E. Rushmeier, M. F. Cohen, D. P. Greenberg, & K. E. Torrance. (1986). An Experimental Evaluation of Computer Graphics Imagery. *ACM Transactions on Graphics*, 5(1), 30-50.
- Yuadi, I., M. D. Artaria, S. & A. Taufiq Asyhari. (2021). Digital Forensics for Skulls Classification in Physical Anthropology Collection Management. *Computers, Materials & Continua*, 68(3), 3979-3995.
- Piva A. (2013). An Overview on Image Forensics. *ISRN Signal Process*, 2013, 1-22.
- Podilchuk, C. I., E. J. Delp. (2001). Digital Watermarking: Algorithms and Applications. *IEEE Signal Processing Magazine*, 18(4), 33-46.
- Reith, M., Carr C. & Gunsch, G. (2002). An Examination of Digital Forensic Models. *International Journal of Digital Evidence*, 1(3), 1-12.
- Ryan, D.J. & Shpantzer, G. (2002). *Legal aspects of digital forensics*. Diakses dari
- Tolles, Juliana dan Meurer, William J. (2016). Logistic Regression Relating Patient Characteristics to Outcomes. *JAMA*, 316(5), 533-534.

- Wen C, Yang K. (2006). Image Authentication for Digital Image Evidence. *Forensic Science Journal*, 5, 1–11.
- Yang, P., Ni, R., Zhao, Y., Zhao, W. (2019). Source Camera Identification Based on Content-Adaptive Fusion Residual Networks. *Pattern Recognition Letters*, 119, 195-204.