

DETEKSI KONTEN PORNO PADA AKUN TWITTER MELALUI NIPPLE DETECTION

Fahmi Auliya Tsani⁽¹⁾, Imam Riadi⁽²⁾, Abdul Fadlil⁽³⁾

⁽¹⁾⁽³⁾Program Studi Magister Teknik Informatika, Universitas Ahmad Dahlan

⁽²⁾Program Studi Sistem Informasi, Universitas Ahmad Dahlan

e-mail: imam.riadi@is.uad.ac.id⁽¹⁾, fahmi1807048017@webmail.uad.ac.id⁽²⁾, fadlil@mti.uad.ac.id⁽³⁾

Abstrak

Perkembangan internet yang begitu pesat memunculkan banyak efek negatif disamping manfaat positifnya yang begitu luas. Salah satu aspek negatifnya adalah konten pornografi dari website berbasis User Generated Content yaitu media sosial. Konten akan selalu berkembang, satu akun pornografi diblok maka bisa dengan segera muncul akun serupa. Disinilah letak kelemahan software porn blocker yang menggunakan deteksi berdasarkan alamat website, karena akun media sosial bisa dengan mudah dibuat. Pada penelitian ini akan menggunakan algoritma Viola-Jones untuk mendeteksi organ intim dalam hal ini adalah puting payudara sehingga citra bisa diklasifikasikan sebagai citra porno. Hasil dari penelitian ini diharapkan bisa mendapatkan tingkat akurasi lebih dari 85%.

Kata Kunci : Object Detection, Porn Blocker, Twitter Porn Content, Viola and Jones, Web Service

1. PENDAHULUAN

Belakangan ini akses internet semakin mudah dan meluas. Namun, dibalik manfaatnya yang begitu luas masih banyak dijumpai efek negatif yang ditimbulkan. Kementerian Komunikasi dan Informatika (Kemkominfo) mengaku telah memblokir 773 ribu situs selama tahun 2016. Dari data sebanyak itu ternyata situs bermuatan konten pornografi menduduki peringkat pertama. Selain pornografi, beberapa kategori lain yang terblokir diantaranya SARA, penipuan/dagang ilegal, narkoba, perjudian, radikalisme, kekerasan, anak, keamanan internet, dan Hak Kekayaan Intelektual (HKI) (Jamaludin, 2016).

Pemerintah sendiri dalam hal ini Kementrian Komunikasi dan Informatika melakukan operasi "bersih-bersih" dunia maya dari berbagai macam konten negatif salah satunya konten pornografi dengan menggunakan mesin mirip dengan crawler. Mesin yang dinamai Ais dan berharga 194 miliar rupiah ini, mampu mendeteksi sekitar 120 ribu situs porno dari Indonesia dan merupakan hasil crawling dari sekitar 1,2 juta alamat internet. (Putri, 2018). Mesin Ais ini bekerja dengan cara memasukkan terlebih dahulu keyword pencarian yang diinginkan, kemudian mesin bisa menarik jutaan konten sekaligus. Konten-konten tersebut akan dianalisis sedemikian rupa untuk ditentukan apakah situs tersebut termasuk kategori yang wajib diblokir atau tidak dengan melibatkan peran manusia agar hasilnya lebih humanis (Agung, 2018).

Akan tetapi, sistem crawling ini belum menjamin efektif sepenuhnya dalam melakukan pemblokiran konten negatif di internet, sebab ada beberapa kelemahan dari mesin Ais tersebut. Salah satu kelemahannya yaitu hanya bisa mendeteksi konten porno berupa teks, untuk konten media jenis lainnya seperti gambar ataupun video sepertinya belum bisa dideteksi secara otomatis (Ngazis, 2017).

Padahal konten porno tak hanya diproduksi oleh situs yang memang dari awal memperkenalkan diri sebagai situs berkonten "dewasa", namun juga dihasilkan dari situs bertipe *User Generated Content* dimana pengunjung bisa mendaftar sebagai *content creator* dan tentu saja bisa membuat konten. Salah satu contoh situs bertipe *User Generated Content* adalah media sosial. Mahbub (2017) menyatakan bahwa Twitter, menurut data Kementerian Komunikasi dan Informatika, merupakan media sosial utama penyebaran konten negatif di internet Indonesia.

Atas dasar permasalahan tersebut, penulis bermaksud membuat tools berupa aplikasi berbasis *web service* sehingga bisa dimanfaatkan oleh berbagai macam device, baik laptop, komputer, smartphone, dll. Tools ini akan mendeteksi konten porno dari sosial media, dalam hal ini adalah twitter, dengan mengunduh terlebih dahulu beberapa sample gambar dari suatu akun yang sedang diakses. Sample gambar tersebut kemudian akan dianalisa dengan menggunakan algoritma Viola-Jones. Jika tools ini mendeteksi adanya konten bermuatan pornografi, maka akses pengguna ke akun tersebut akan langsung diblok.

Web service sendiri diartikan oleh Booth *et al.* (2004) sebagai sistem perangkat lunak yang didesain untuk mendukung komunikasi/interaksi antar mesin melalui jaringan internet. Desain ini memiliki *interface*/antar muka yang dideskripsikan dalam format yang bisa diproses oleh mesin seperti XML dan JSON.

2. TINJAUAN PUSTAKA

Penelitian tentang pengolahan citra yang berfokus pada pendeteksian objek sudah banyak dilakukan, baik untuk deteksi objek pada umumnya maupun spesifik pada deteksi citra porno. Penelitian dari Surya et al. (2017) berfokus pada deteksi pola batik sehingga bisa diklasifikasikan sebagai pola batik Pekalongan. Penelitian tersebut menggunakan metode ekstraksi ciri untuk mendapatkan fitur tekstur ciri khas dari batik Pekalongan, sehingga bisa dilakukan proses klasifikasi untuk pendataan citra pola batik Pekalongan. Metode ekstraksi ciri yang digunakan yaitu *Gray Level Cooccurrence Matrix* (GLCM) dan filter Gabor.

Saifullah et al. (2017) juga melakukan penelitian yang berinti pada penggunaan *Gray Level Cooccurrence Matrix* (GLCM), namun objek yang diteliti adalah telur ayam kampung. GLCM ini digunakan untuk identifikasi ciri fertilitas telur ayam kampung dengan menggunakan enam parameter yaitu Angular Second Moment (ASM), kontras, korelasi, varians, Inverse Difference Moment (IDM), dan entropi. Hasil akhir yang didapatkan yaitu parameter-parameter GLCM ini mampu membedakan dan mengelompokkan telur ayam kampung berdasarkan fertilitasnya.

Penelitian yang dilakukan Muwardi & Fadlil (2017) menerapkan pengolahan citra untuk mengenali jenis-jenis bunga. Citra bunga dikumpulkan dari kamera handphone lalu melakukan pre-processing yaitu proses grayscale dan cropping sehingga bisa melakukan ekstraksi ciri menggunakan histogram warna citra dan pengklasifikasi jarak (Manhattan dan Euclidean). Penelitian ini menggunakan sembilan sampel pola dasar bentuk bunga antara lain bunga kamboja, bunga lili putih, bunga lidah mertua, bunga alamanda, bunga matahari, bunga kenanga, bunga sepatu, bunga mawar, dan bunga melati.

Penelitian selanjutnya yaitu pengolahan citra digital untuk mendeteksi uang palsu. Penelitian ini berangkat dari solusi dari pemerintah untuk membedakan uang palsu secara manual dengan 3d (dilihat, diraba, dan diterawang) yang menurut Umar et al. (2018) belum bisa dikatakan sempurna. Sehingga mereka menawarkan solusi lain untuk membedakan uang palsu melalui pengolahan citra digital. Penelitian tersebut hanya berfokus pada uang kertas pecahan 50.000 dan 100.000 rupiah, lalu melakukan fotokopi warna pada uang tersebut sehingga diperoleh citra digitalnya. Deteksi uang palsu dilakukan dengan cara mengenali fitur-fitur yang terdapat pada citra seperti tekstur permukaan uang serta diklasifikasikan menggunakan algoritma K-Means Cluster dengan menghitung nilai centroid dari data pada setiap cluster memakai rumus persamaan Euclidean Distance.

Spesifik pada deteksi citra porno, Sugiantoro et al. (2016) memilih menggunakan library/pustaka pemrograman yang sudah ada untuk melakukan deteksi citra porno. Library yang digunakan yaitu *nude.js* yang dibuat oleh Patrick Wied yang juga sudah diuji oleh Rigan Ap-apid. Hasil pengujian tersebut menunjukkan bahwa library tersebut memiliki kinerja yang lebih baik dibandingkan SVM, WIPE, Jones-Rehg, dan Fleck-Forsyth dalam mendeteksi citra porno. Library ini kemudian diimplementasikan ke dalam bentuk add-on untuk web browser Mozilla Firefox.

Hidayattullah & Hapsari (2013) memilih melakukan penelitian untuk deteksi citra porno dengan menggunakan algoritma Viola-Jones yang memang sudah terkenal handal untuk mendeteksi objek. Pemilihan algoritma ini didasari pada adanya kelemahan algoritma-algoritma terdahulu dalam mendeteksi citra porno yang berdasarkan pada perhitungan luas kulit pada objek. Kelemahan yang dimaksud salah satunya adalah sering terjadi salah deteksi pada saat luas area kulit lebih besar dibanding luas area background. Algoritma Viola-Jones ini kemudian di-training untuk mengenali bentuk nipple (puting payudara), sehingga jika suatu citra memiliki ciri-ciri sebagai puting, maka akan dikelompokkan sebagai citra porno. Algoritma ini diimplementasikan ke dalam bahasa pemrograman C++ dan memanfaatkan framework computer vision OpenCV 2.1, serta objek citra yang akan dideteksi hanya fokus pada citra payudara frontal.

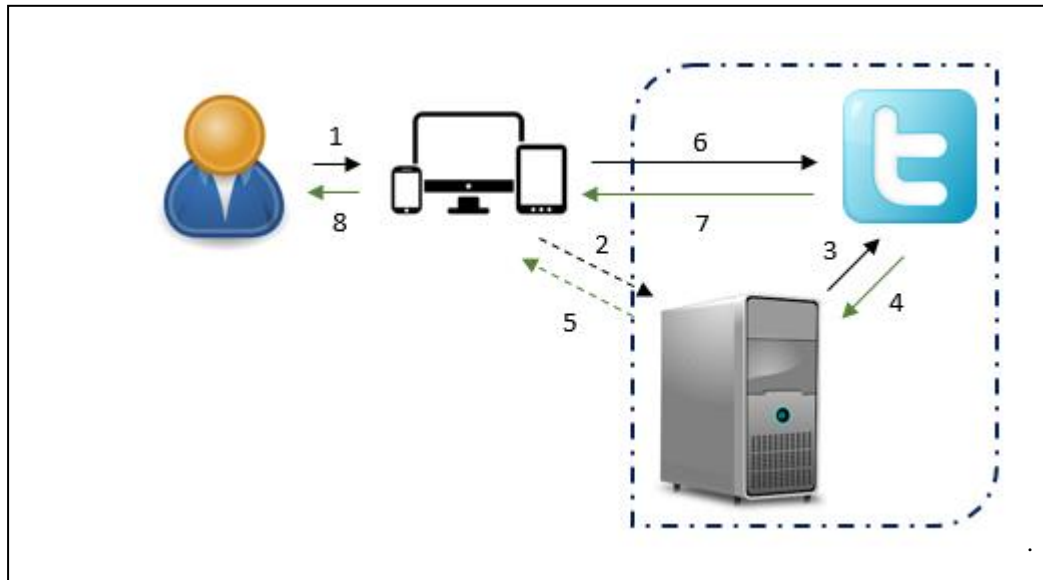
Penelitian terkait selanjutnya yaitu deteksi citra porno dengan memakai shape descriptor berbasis model warna YcbCr dan klasifikasi menggunakan algoritma C4.5. Citra digital yang pada dasarnya menggunakan sistem warna RGB ditransformasikan menjadi YCbCr untuk selanjutnya dihitung nilai skin threshold-nya. Range nilai skin threshold yang dianggap optimal yaitu $77 \leq Cb \leq 127$ dan $133 \leq Cr \leq 173$, sedangkan nilai Cg dianggap kurang berpengaruh secara signifikan pada warna kulit sehingga tidak dipakai. Citra hasil deteksi kulit ini kemudian diolah ke dalam shape descriptor menggunakan Seven Normal Moment Variants agar objek di dalam citra bisa diambil. Setelah bentuk objek bisa diambil, kemudian citra diklasifikasikan menggunakan C4.5 (Ariyanto et al., 2016).

Nugroho et al. (2016) melakukan penelitian untuk deteksi citra porno menggunakan ekstraksi ciri GLCM dan pengklasifikasian kulit menggunakan Support Vector Machine (SVM). Sebelum melakukan ekstraksi ciri menggunakan GLCM, citra digital terlebih dahulu dilakukan deteksi wajah menggunakan algoritma Viola-Jones. Jika citra mengandung area wajah, maka perhitungan segmentasi kulit tidak termasuk pada area wajah tersebut. Setelah melalui 6 jenis ekstraksi fitur menggunakan GLCM (kontras, homogenitas, entropi, korelasi, cluster shade, dan cluster prominence), citra digital kemudian diklasifikasikan menjadi citra yang mengandung kulit dan citra yang tidak mengandung kulit menggunakan SVM. Berdasarkan metode yang

digunakan, program mampu mengenali objek yang mengandung kulit dan yang tidak mengandung kulit dengan sangat baik, sehingga sangat potensial digunakan untuk filter konten negatif.

3. METODE PENELITIAN

Penelitian ini berfokus pada penggunaan algoritma Viola-Jones untuk mengklasifikasikan citra porno dan diimplementasikan ke dalam aplikasi web service yang dikerjakan dalam bahasa pemrograman PHP memakai framework Yii 2. Gambar 1 berikut ini merupakan diagram rancangan program yang kami gunakan untuk penelitian ini, sedangkan penjelasan atas simbol-simbol kami jabarkan pada tabel 1 dan area yang kami batasi dengan garis putus-putus merupakan ranah fokus penelitian kami.



Gambar 1. Diagram proses pendeteksian citra porno pada akun twitter

Tabel 1. Arti simbol untuk gambar 1

No.	Simbol	Arti
1.	→	alur proses pengiriman data dari client ke server
2.	- - - →	alur proses data yang dikirim di background dari device client ke server API
3.	←	alur proses pengirim data yang dikirim dari server ke client
4.	← - - -	alur proses data dari server API menuju device client

Urutan proses dari Gambar 1 di atas yaitu sebagai berikut:

1. pengguna mengakses suatu akun twitter dari device yang dia gunakan
2. aplikasi dari device yang digunakan untuk mengakses akun twitter mengirim informasi alamat akun (username) twitter yang akan diakses ke server API secara background, sedangkan akses ke alamat akun yang dituju tersebut akan di-interrupt sementara sambil menunggu pengecekan oleh server API.
3. server API mengambil beberapa sampel data media/citra dari akun yang akan dicek melalui API yang sudah disediakan oleh twitter
4. twitter mengirim beberapa sampel media/citra dari akun yang akan dicek ke server API untuk dicek apakah konten tersebut merupakan konten porno.
5. setelah dilakukan pengecekan, server API mengembalikan status pengecekan ke device client yaitu berupa data berformat JSON
6. jika data JSON yang dikembalikan oleh server API menyatakan bahwa akun twitter yang diakses mengandung konten porno, maka aplikasi pada device tersebut akan memblokir akun twitter tersebut, sehingga user tidak bisa mengaksesnya, sebaliknya jika tidak terindikasi mengandung konten porno maka akses akan dilanjutkan

7. karena akses disetujui, maka server twitter akan mengirimkan data username yang diakses
8. data diolah oleh aplikasi pada device sehingga bisa ditampilkan ke pengguna.

4.HASIL DAN PEMBAHASAN

Setelah dilakukan simulasi menggunakan tool Postman for Chrome versi 5.5.4 untuk melakukan pengiriman data username twitter yang akan dicek, diharapkan server API berhasil mengklasifikasikan citra yang diunduh dari akun twitter yang diakses serta mengembalikan data status pengecekan citra yang dianalisis dan dengan tingkat akurasi lebih dari 85%. Sehingga, data status pengecekan ini bisa dimanfaatkan oleh aplikasi dari berbagai device, baik untuk melakukan pemblokiran maupun meloloskan akses ke akun twitter tersebut.

5.KESIMPULAN

Dengan menggunakan algoritma Viola-Jones, diharapkan aplikasi dapat mengklasifikasikan citra digital yang diunduh dari suatu akun twitter menjadi citra porno dan citra non porno.

DAFTAR PUSTAKA

- Agung, B. 2018. Cara Kerja Sensor Konten Dua Lapis Ala Kominfo. <https://www.cnnindonesia.com/teknologi/20180102155546-185-266123/cara-kerja-sensor-konten-dua-lapis-ala-kominfo>. diakses 18 Oktober 2018.
- Ariyanto, E. R., Wijanarto, & Sudaryanto. 2016. Klasifikasi Citra Porno dengan Algoritma C 4.5 Berbasis Model Warna YCbCr dan Shape Detector. *Jurnal Techno.COM* Vol. 15, No. 2, pp. 92–98.
- Booth, D., Haas, H., McCabe, F., Newcomer, E., Champion, M., Ferris, C., & Orchard, D. 2004. *Web Services Architecture*. <https://www.w3.org/TR/ws-arch/wsa.pdf>. diakses 15 November 2018.
- Hidayattullah, M. F., & Hapsari, Y. 2013. Automatic Nipple Detection Pada Citra Pornografi Menggunakan Algoritma Viola And Jones Berbasis Adaboost Untuk Feature Selection. *Proceeding of Seminar Nasional Teknologi Informasi & Komunikasi Terapan*. pp. 238–245
- Jamaludin, F. 2016. 773 Ribu Situs Diblokir Kemkominfo Setahun, Pornografi Paling Banyak. <https://www.merdeka.com/teknologi/773-ribu-situs-diblokir-kemkominfo-setahun-pornografi-paling-banyak.html>. diakses 18 Oktober 2018.
- Muwardi, F., & Fadlil, A. 2017. Sistem Pengenalan Bunga Berbasis Pengolahan Citra Dan Pengklasifikasi Jarak. *Jurnal Ilmu teknik Elektro Komputer dan Informatika (JITEKI)* Vol. 3, No. 2, pp. 124–131.
- Ngazis, A. N. 2017. Kelemahan Mesin Sensor Internet Rp200 Miliar Incaran Kominfo. <https://www.viva.co.id/arsip/969804-kelemahan-mesin-sensor-internet-rp200-miliar-incaran-kominfo>. diakses 18 Oktober 2018.
- Nugroho, H. A., Rahadian, F., Adji, T. B., Oktoeberza, W. K. Z., & Buana, R. L. B. 2016. Texture Analysis for Skin Classification in Pornography Content Filtering Based on Support Vector Machine. *Journal of Engineering and Technological Sciences* Vol. 48, No. 5, pp. 584–596.
- Saifullah, S., Sunardi, & Yudhana, A. 2017. Analisis Ekstraks Ciri Fertilitas Telur Ayam Kampung dengan Grey Level Cooccurance Matrix. *JNTE* Vol. 6, No. 2, pp. 66–75.
- Sugiantoro, B., Mafrur, R., & Fajri, N. S. 2016. Pengembangan Perangkat Lunak Deteksi Gambar Porno Di Browser Pengguna Internet, *Jurnal Teknik Informatika* Vol. 9, No. 1, pp. 40–45.
- Surya, R. A., Fadlil, A., & Yudhana, A. 2017. Ekstraksi Ciri Metode Gray Level Co-Occurrence Matrix (GLCM) dan Filter Gabor untuk Klasifikasi Citra Batik Pekalongan. *JURNAL INFORMATIKA : Jurnal Pengembangan IT* Vol. 2, No. 2, pp. 23–26.

Umar, R., Riadi, I., & Miladiah. 2018. Sistem Identifikasi Keaslian Uang Kertas Rupiah Menggunakan Metode K-Means Clustering. *Techno.COM* Vol. 17, No. 2, pp. 179–185.