

Algorithmic Bias Dalam Restorative Justice: Ancaman Bagi Marginal

Franky Ariyadi

Universitas Muhammadiyah Tangerang; cisadane@yahoo.com

*Correspondensi: Franky Ariyadi

Email: cisadane@yahoo.com



Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY NC) license (<http://creativecommons.org/licenses/by/4.0/>).

Abstrak: Integrasi kecerdasan buatan (AI) dalam sistem keadilan restoratif telah membawa paradigma baru sekaligus tantangan kompleks dalam penegakan hukum pidana. Penelitian ini mengkaji secara kritis dampak bias algoritmik terhadap kelompok marginal dalam konteks keadilan restoratif di Indonesia. Dengan menggunakan pendekatan yuridis normatif-kritis dan analisis kualitatif, penelitian ini menganalisis berbagai dokumen hukum, laporan kebijakan, dan studi kasus internasional seperti sistem COMPAS di AS yang terbukti memiliki bias rasial. Hasil penelitian mengungkap beberapa temuan krusial: pertama, algoritma cenderung mereproduksi ketidakadilan struktural yang tertanam dalam data historis; kedua, implementasi sistem prediktif seperti SPPT-TI di Indonesia berpotensi memperparah ketimpangan tanpa mekanisme pengawasan yang memadai; ketiga, prinsip-prinsip dasar keadilan restoratif seperti partisipasi dan kesetaraan terancam tergeser oleh logika teknis yang kaku. Penelitian ini menawarkan rekomendasi kebijakan komprehensif yang meliputi: (1) audit etik berkala terhadap sistem algoritma, (2) pengembangan model hybrid yang memadukan teknologi dengan kearifan lokal, (3) pelatihan etika digital bagi aparat penegak hukum, (4) penguatan regulasi berbasis HAM, serta (5) mekanisme partisipasi publik dalam pengembangan sistem. Temuan penelitian ini tidak hanya relevan bagi pengembangan kebijakan hukum pidana di Indonesia, tetapi juga memberikan kontribusi signifikan dalam diskursus global tentang etika AI dalam sistem peradilan. Penelitian lanjutan diperlukan untuk menguji efektivitas model hybrid dalam konteks lokal yang spesifik.

Keywords: Bias Algoritmik, Keadilan Restoratif, Kelompok Marginal, Kecerdasan Buatan, Hukum Pidana Indonesia

PENDAHULUAN

Perkembangan teknologi informasi dan kecerdasan buatan (AI) telah membawa perubahan signifikan dalam cara sistem peradilan pidana modern beroperasi. Dari penyelidikan digital, pemantauan perilaku, hingga sistem prediksi kriminal, teknologi menjanjikan efisiensi, objektivitas, dan kecepatan proses hukum. Namun, di balik janji tersebut, muncul kekhawatiran mendalam tentang keadilan, terutama ketika algoritma yang digunakan justru mereproduksi bias yang merugikan kelompok rentan atau marginal.

Penelitian terdahulu telah mengkaji bagaimana algorithmic decision-making dalam sistem hukum pidana dapat menciptakan ketidakadilan struktural. Misalnya, laporan investigatif oleh ProPublica menunjukkan bahwa sistem COMPAS—sebuah alat prediksi risiko residivisme di Amerika Serikat—menunjukkan bias rasial terhadap warga kulit hitam (ProPublica, 2016). Studi ini menggugah kesadaran bahwa teknologi bukanlah entitas netral; ia membawa serta nilai, data historis, dan asumsi pembuatnya (Angwin, Larson, Mattu, & Kirchner, 2022) Lihat juga (Eubanks, 2017)

Beberapa penelitian fokus pada tantangan transparansi algoritma (Citron & Pasquale, 2014), serta potensi pelanggaran terhadap prinsip-prinsip keadilan prosedural dalam konteks penegakan hukum berbasis data (Garvie, Bedoya, & Frankle, 2016). Namun, masih sedikit kajian yang secara khusus menyoroti bagaimana bias algoritmik

berdampak dalam konteks restorative justice—yakni suatu pendekatan keadilan yang mengedepankan pemulihan relasi dan keterlibatan korban, pelaku, dan masyarakat.

Padahal, dalam prinsip restorative justice, nilai partisipatif dan kesetaraan menjadi fondasi utama. Ketika teknologi—yang seharusnya mendukung proses tersebut—malah memperkuat diskriminasi data, maka cita-cita keadilan berbasis masyarakat sipil bisa terganggu. Oleh karena itu, penelitian ini berfokus pada analisis kritis terhadap algorithmic bias dalam praktik restorative justice, dengan tujuan untuk mengkaji potensi ancamannya terhadap kelompok marjinal dan mencari alternatif regulatif berbasis keadilan sosial serta prinsip due process dalam KUHAP (Indonesia, 1981) dan instrumen hak asasi manusia seperti International Covenant on Civil and Political Rights (ICCPR).

Kebaruan penelitian ini terletak pada pendekatan interdisipliner yang menggabungkan perspektif hukum pidana, kajian teknologi, dan keadilan sosial, serta analisis empiris terhadap kasus-kasus konkret di Indonesia. Temuan penelitian diharapkan dapat menjadi referensi bagi pembuat kebijakan dalam merumuskan regulasi teknologi di sektor peradilan yang lebih inklusif dan berkeadilan

METODE

(11pt, Times New Roman) Bagian ini menyajikan desain penelitian, populasi, teknik sampel dan sampling, serta metode analisis. Penulis harus menjelaskan bagaimana penelitian dilakukan secara rinci. Semua artikel telah disetujui oleh komite etik dan artikel yang melibatkan subyek manusia telah mendapatkan persetujuan. Etika, validitas dan reliabilitas, atau kepercayaan harus menjadi bagian dari bagian ini.

Penulis dapat menjabarkan poin-poin utama menjadi sub bab tanpa penomoran, misal:

Desain Penelitian

Jelaskan desain penelitian yang digunakan. Bagian ini tidak perlu disampaikan definisi dari desain tersebut menurut teori. Cukup sampaikan secara ringkas desain penelitian yang digunakan.

Populasi, Sampel, Sampling

Penulis harus menjelaskan secara rinci, detail, bagaimana proses seleksi populasi menjadi sampel, sertakan kriteria inklusi dan eksklusi jika ada, berapa jumlah populasi, bagaimana prosesnya hingga terpilih menjadi sampel, berapa yg dieksklud, berapa yg tidak sesuai kriteria, dll. Penulis dapat menyajikan proses seleksi ini dengan bagan serta dinarasikan.

Prosedur Intervensi

Apabila desain penelitian yang digunakan eksperimen, maka penulis harus menyampaikan secara rinci bagaimana prosedur intervensi, berapa kali, berapa lama, diberikan seperti apa, dll. Penulis dapat menggunakan tabel maupun narasi. Apabila eksperimen dilakukan bertahap, misal selama 4 minggu, tiap minggu memberikan intervensi, maka akan lebih baik dijelaskan dalam bentuk tabel, minggu 1 intervensi yang diberikan berupa apa saja, selama berapa lama, diberikan terhadap siapa, dll.

Instrumen

Penulis harus menjelaskan secara rinci instrumen yang digunakan dalam penelitian, apakah hasil modifikasi dari riset sebelumnya, bagaimana validitas dan reliabilitas instrumen yang digunakan, bagaimana cara skoring atau menilai, cara mengukur, dll.

Persetujuan Etik

Penelitian yang berhubungan dengan manusia maka harus melalui persetujuan etik dari komisi etik penelitian kesehatan (KEPK). Peneliti harus menyampaikan bagaimana persetujuan etik dari penelitian, nomor surat persetujuan etik, dan diberikan persetujuan oleh komisi etik mana. Jumlah halaman minimum 6 halaman dan maksimum 10 halaman ukuran A4 **Desain Penelitian**

Penelitian ini menggunakan pendekatan yuridis normatif-kritis yang dikombinasikan dengan analisis kualitatif deskriptif, guna menjawab persoalan normatif dan etis terkait penggunaan algoritma dalam sistem keadilan restoratif. Fokus kajian tidak bersifat eksperimental, namun mendalami secara struktural hubungan antara algoritmik bias dan implikasinya terhadap kelompok marjinal dalam kerangka keadilan pidana.

Pendekatan ini dipilih karena objek penelitian tidak hanya melibatkan norma hukum positif seperti KUHAP (KUHP, 1981) dan ICCPR (ICCPR, 1967), tetapi juga teknologi non-hukum berupa sistem kecerdasan buatan yang digunakan dalam proses prediksi risiko dan seleksi kebijakan peradilan. Model analisis yang digunakan adalah analisis isi dokumen (*document analysis*) dan hermeneutika hukum digital, yang memadukan penafsiran teks hukum dengan hasil studi empiris terkait praktik sistem algoritma peradilan di beberapa yurisdiksi.

Desain ini memungkinkan pengujian normatif atas dua hal:

1. Apakah penggunaan algoritma dalam keadilan restoratif bertentangan dengan prinsip *due process* dan *nondiskriminasi*, dan
2. Apakah terdapat kerangka etik dan hukum yang memadai untuk mencegah terjadinya *algorithmic injustice* terhadap masyarakat rentan.

Populasi, Sampel, dan Sampling

Populasi dalam penelitian ini adalah semua praktik, regulasi, dan kebijakan yang menggunakan algoritma atau sistem prediktif dalam sistem keadilan pidana, khususnya yang relevan dengan prinsip *restorative justice*. Karena sifat kajian adalah kualitatif dokumen, maka populasi tidak berupa individu, tetapi berupa dokumen hukum, laporan penelitian, kebijakan publik, white paper, jurnal ilmiah, serta studi kasus yudisial.

Sampel dipilih menggunakan teknik *purposive sampling*, yaitu pemilihan sumber yang relevan secara langsung dengan tema bias algoritma dan keadilan restoratif, baik nasional maupun internasional. Dokumen yang dianalisis antara lain:

- Putusan pengadilan di AS dan Eropa yang melibatkan penggunaan sistem COMPAS dan COMPAS-like tools.
- Laporan investigatif dari ProPublica (2016) yang membongkar ketimpangan skor algoritmik terhadap kelompok kulit hitam.
- Laporan UNODC dan OECD AI Principles yang membahas etika penggunaan AI di sektor keadilan.
- Regulasi dan dokumen hukum seperti RUU KUHP (draft 2025), dan Peraturan MA No. 1 Tahun 2016 (tentang mediasi sebagai bentuk pendekatan restoratif).

Prosedur Intervensi

Karena penelitian ini tidak menggunakan metode eksperimental atau perlakuan langsung terhadap subjek manusia, maka prosedur intervensi dalam konteks penelitian ini bersifat normatif-dokumentatif. Fokus intervensinya terletak pada analisis kritis terhadap sistem algoritma yang digunakan dalam proses keadilan restoratif, serta penelusuran potensi biasnya terhadap kelompok rentan.

Adapun bentuk intervensi dilakukan melalui tiga tahapan kajian sebagai berikut:

1. Identifikasi dan klasifikasi sistem algoritma yang digunakan dalam proses keadilan pidana berbasis prediksi, khususnya sistem seperti *COMPAS*, *HART*, dan inisiatif digital di Indonesia (seperti *e-Justice* dan Sistem Peradilan Pidana Terbaru berbasis TI atau SPPT-TI).
2. Analisis normatif dan empiris terhadap regulasi dan praktik, termasuk instrumen hukum internasional (ICCPR), peraturan nasional (RKUHAP), serta kebijakan lembaga penegak hukum (SE Kapolri No. SE/2/II/2021; Peraturan Kejaksaan Agung No. 15/2020).
3. Evaluasi prinsip *restorative justice*, lalu menilai apakah sistem algoritmik yang digunakan dalam penilaian risiko atau keputusan intervensi keadilan sesuai dengan nilai-nilai seperti partisipasi, kesukarelaan, pemulihan, dan kesetaraan.

Meskipun tidak ada intervensi dalam bentuk eksperimen, pendekatan penelitian ini tetap menyentuh dimensi intervensi secara intelektual dan normatif, karena memberikan *intervensi argumentatif* terhadap kebijakan berbasis AI yang sedang berkembang di sektor peradilan pidana.

Instrumen Penelitian

Instrumen dalam penelitian ini berupa alat bantu analisis data kualitatif dokumen, termasuk kerangka kerja yang mengacu pada:

1. Kerangka Etik dan Hak Asasi Manusia

Seperti Kelompok Hak Asasi Manusia (Now, 2018) dan prinsip-prinsip AI OECD *Principles* (2020), yang digunakan untuk menilai apakah sistem algoritmik sesuai dengan prinsip keadilan, transparansi, dan non-diskriminasi.

2. Kerangka Hukum Nasional

Termasuk KUHAP, RUU KUHAP (KemenHumham, 2025), Peraturan Jaksa Agung RI (Kejagung, 2020), dan Surat Edaran Kapolri (Kapolri, 2021), sebagai basis menilai bagaimana proses restoratif dipraktikkan dan diatur di Indonesia.

3. Analisis Kasus Empiris

Seperti *ProPublica's* (ProPublica, 2016) dan kasus penggunaan algoritma COMPAS dalam sistem peradilan AS, sebagai *benchmark* perbandingan dalam menilai risiko *algorithmic injustice* terhadap kelompok marginal.

Validitas instrumen diuji melalui proses triangulasi dokumen, yaitu dengan membandingkan hasil dari:

1. Sumber hukum primer (undang-undang, peraturan)
2. Sumber akademik (jurnal, laporan, buku)

3. Praktik lapangan dan dokumentasi empirik (laporan ProPublica, UNODC, OECD)

Reliabilitas dijaga dengan pengulangan analisis silang oleh dua peneliti pendamping, guna mencegah subjektivitas tunggal dan memastikan koherensi antara data dan interpretasi.

Validitas dan Reliabilitas

Dalam konteks penelitian kualitatif-normatif yang berbasis dokumen dan analisis kebijakan, validitas tidak hanya dimaknai sebagai ketepatan data, tetapi juga ketepatan penafsiran hukum dan kontekstualisasi sosial dari sumber hukum yang digunakan (Creswell & Poth, 2018). Validitas dalam studi ini dijaga melalui strategi triangulasi dokumen, yaitu dengan membandingkan hasil temuan dari:

1. Sumber hukum primer – seperti KUHAP, RKUHAP, ICCPR, Peraturan Jaksa Agung RI (Kejagung, 2020), dan Kepolisian Republik Indonesia (Kapolri, 2021).
2. Sumber empirik dan praktis – termasuk laporan investigasi *Machine Bias* (Angwin, Larson, Mattu, & Kirchner, 2022) laporan *Georgetown Law* (Garvie, Bedoya, & Frankle, 2016) dan rekomendasi UNODC tentang AI dalam sistem peradilan pidana (UNODC, 2021).
3. Literatur akademik terkini – seperti tulisan (Citron & Pasquale, 2014) dan (Eubanks, 2017), yang menyuarakan tantangan terhadap sistem *scoring* berbasis algoritma.

Sementara itu, reliabilitas atau kepercayaan terhadap data diperoleh dengan cara:

1. Memastikan bahwa setiap sumber yang dikutip memiliki otoritas akademik, hukum, atau praktis yang jelas, baik berasal dari jurnal bereputasi, undang-undang resmi, maupun lembaga internasional.
2. Proses analisis dilakukan oleh penulis utama dengan *peer-review internal* dari dua kolega bidang hukum dan teknologi, untuk menghindari bias penafsiran serta memperkuat konsistensi argumentasi.

Dalam konteks hukum, reliabilitas juga dipastikan dengan mengacu pada dokumen resmi berbahasa asli (ICCPR, 1967) dan menghindari interpretasi hukum yang bersifat spekulatif.

Persetujuan Etik

Penelitian ini tidak melibatkan subjek manusia secara langsung, karena seluruh data berasal dari dokumen hukum, laporan publik, dan sumber terbuka. Oleh karena itu, tidak diperlukan izin etik dari Komite Etik Penelitian Kesehatan (KEPK).

Namun demikian, prinsip etika akademik dan integritas keilmuan tetap dijunjung tinggi dalam seluruh proses penelitian, antara lain:

1. Tidak melakukan plagiat, manipulasi data, atau pencatutan sumber.
2. Menghormati kerahasiaan jika data yang diakses bersifat semi-terbuka atau mengandung informasi sensitif (misalnya identitas terdakwa dalam laporan algoritmik).

3. Menyajikan temuan dengan *nuansa adil*, tidak menghakimi secara sepihak sistem atau lembaga, serta selalu mencantumkan kerangka hukum dan HAM internasional sebagai rujukan normatif utama (ICCPR, 1967) pasal 14.

Secara substansial, penelitian ini tetap tunduk pada prinsip *do no harm* dalam penelitian sosial (Israel & Hay, 2006), yakni tidak menimbulkan dampak negatif bagi kelompok rentan yang menjadi objek kajian—dalam hal ini masyarakat marginal yang potensial terdampak bias algoritmik dalam *restorative justice*.

HASIL DAN PEMBAHASAN

Penelitian ini menemukan bahwa penggunaan sistem algoritmik dalam konteks keadilan restoratif masih menimbulkan **ketidakesesuaian antara prinsip due process dan nilai keadilan substantif**, khususnya terhadap kelompok masyarakat marginal. Temuan ini diperoleh melalui sintesis dokumen hukum, studi kasus internasional, dan kerangka HAM yang relevan.

A. Hubungan Hasil Penelitian dengan Tujuan Awal

Sebagaimana dijelaskan dalam Pendahuluan, tujuan penelitian ini adalah menganalisis potensi **bias algoritma** (*algorithmic bias*) dalam proses *restorative justice* dan menilai dampaknya terhadap kelompok rentan. Dari analisis dokumen seperti laporan *Machine Bias* (Angwin et al., 2016), dan laporan Georgetown Law (Garvie, Bedoya, & Frankle, 2016)), ditemukan bahwa sistem algoritma yang digunakan untuk prediksi risiko residivisme atau penyaringan intervensi keadilan justru **memperkuat ketidaksetaraan yang telah ada**, terutama berdasarkan ras, status sosial, atau wilayah tempat tinggal.

Dalam konteks Indonesia, meskipun sistem prediktif belum secara luas digunakan dalam *restorative justice* formal, beberapa inisiatif digitalisasi peradilan seperti **SPPT-TI dan e-Justice** mulai mengarah pada penggunaan data berbasis sistem informasi. Risiko bias muncul ketika data input berasal dari rekam jejak penegakan hukum yang sendiri sudah memiliki jejak ketimpangan sosial. Artinya, **ketidakadilan lama bisa diwariskan ke dalam sistem digital melalui data historis yang tidak netral** (Eubanks, 2017).

B. Apakah Hasil Memberikan Interpretasi Ilmiah?

Hasil penelitian ini menunjukkan bahwa algoritma yang diklaim netral pada praktiknya **tidak sepenuhnya bebas nilai**. Algoritma bekerja berdasarkan data masa lalu, dan jika data itu mencerminkan praktik diskriminatif (misalnya penangkapan berlebih terhadap kelompok tertentu), maka **output sistem akan mereproduksi bias yang sama** (O'Neil, 2016).

Dalam konteks *restorative justice*, yang seharusnya memprioritaskan pemulihan dan rekonsiliasi, penggunaan algoritma yang menilai risiko pelaku bisa **menghambat partisipasi kelompok marginal** dalam proses keadilan yang setara. Misalnya, jika pelaku dari komunitas miskin diberi skor risiko tinggi, mereka bisa ditolak masuk ke program pemulihan, meskipun secara substantif mereka memenuhi syarat.

Implikasinya sangat serius: *restorative justice* yang seharusnya mengoreksi ketimpangan sistem justru bisa **dipakai sebagai alat eksklusif**, hanya bisa diakses oleh mereka yang sistem nilai sosialnya dianggap “rendah risiko” oleh sistem algoritma.

C. Apakah Konsisten dengan Penelitian Lain?

Hasil ini konsisten dengan berbagai studi internasional yang telah menyuarakan bahaya penggunaan AI dalam peradilan pidana. (Citron & Pasquale, 2014) mengkritik ketidakjelasan proses penilaian otomatis yang disebut sebagai *black box justice*. Begitu pula dalam laporan UNODC (2021) tentang AI dalam sistem keadilan pidana, ditekankan pentingnya prinsip transparansi, akuntabilitas, dan non-diskriminasi dalam pengembangan teknologi untuk keperluan hukum.

Namun, di sisi lain, ada penelitian yang menunjukkan bahwa algoritma justru dapat mengurangi bias manusia dalam beberapa kasus, selama desainnya diawasi dan diuji secara adil (Berk et al., 2018). Oleh karena itu, pembahasan ini menyimpulkan bahwa **masalah bukan pada teknologi itu sendiri, tetapi pada desain, data input, dan konteks penggunaannya.**

D. Implikasi Hasil terhadap Sistem Hukum Pidana Nasional

Hasil penelitian ini tidak hanya bersifat konseptual, tetapi memiliki **dampak langsung terhadap tata cara penegakan hukum pidana di Indonesia**. Keadilan restoratif yang kini mulai diarusutamakan melalui **Peraturan Jaksa Agung** dan **SE Kapolri** mengandung semangat pemulihan, kesetaraan, dan partisipasi korban-pelaku. Namun demikian, ketika prinsip-prinsip tersebut dioperasikan melalui teknologi algoritmik, terjadi dilema baru: **bagaimana menjamin bahwa proses digital tetap mengedepankan nilai-nilai keadilan sosial dan tidak bersifat diskriminatif secara sistemik?**

Dalam sistem keadilan pidana nasional, nilai *due process of law* sebagaimana tertuang dalam **Pasal 50 hingga 68 KUHAP** menegaskan hak tersangka/terdakwa atas perlindungan hukum yang adil, termasuk hak untuk diperlakukan sama di hadapan hukum. Ini sejalan dengan prinsip *non-discrimination* dalam **ICCPR**, yang menyatakan bahwa “setiap orang berhak untuk diadili secara adil dan terbuka oleh pengadilan yang berwenang, tidak memihak, dan independen.”

Namun, apabila sistem algoritma dipakai untuk membantu aparat penegak hukum memutuskan siapa yang layak memperoleh pemulihan, dan sistem tersebut **mengeluarkan hasil prediktif yang bias terhadap kelompok tertentu**, maka secara hukum telah terjadi **pelanggaran hak asasi dan prinsip keadilan prosedural**.

Contoh konkret dalam konteks Indonesia dapat dilihat dari kemungkinan penggunaan sistem *e-integrated justice* (SPPT-TI), yang mengintegrasikan data kejaksaan, kepolisian, dan pengadilan. Jika data kriminal dari wilayah padat penduduk miskin secara statistik lebih tinggi, dan sistem tidak melakukan pembobotan ulang (*reweighting*), maka individu dari wilayah tersebut **berisiko lebih tinggi diklasifikasikan sebagai “berisiko tinggi”**, sehingga dikeluarkan dari opsi keadilan restoratif.

Artinya, **ketimpangan sosial yang ada di dunia nyata disalin mentah-mentah ke dalam sistem digital**, tanpa penyaringan nilai keadilan. Ini bertentangan dengan semangat *restorative justice*, yang justru lahir untuk memperbaiki relasi kuasa dan ketimpangan dalam sistem hukum.

E. Penafsiran Hasil Berdasarkan Prinsip-Prinsip Restorative Justice

Restorative justice menempatkan *dialog, pemulihan, tanggung jawab bersama, dan kesukarelaan* sebagai inti proses. Prinsip ini tidak menekankan penghukuman, tetapi **rekonsiliasi antara pelaku, korban, dan komunitas**. Ketika sistem prediktif digunakan untuk *menyaring* siapa yang boleh dan tidak boleh memasuki proses restoratif, maka peran nilai kemanusiaan bisa direduksi menjadi *angka dan skor*.

Dalam praktiknya, nilai-nilai tersebut sulit diwakili oleh sistem algoritmik, karena:

1. Algoritma bekerja berdasarkan statistik, bukan pada *nilai-nilai kultural dan spiritual* yang melekat dalam konsep keadilan di masyarakat.
2. AI tidak memahami konteks lokal, seperti tradisi musyawarah, perdamaian berbasis adat, atau ekspresi empati dalam dialog sosial.
3. Sistem tidak selalu memberi ruang pada *narasi pribadi* yang penting dalam memahami latar belakang pelaku maupun korban.

Maka dari itu, dalam konteks restorative justice, **teknologi harus menjadi alat bantu, bukan alat seleksi**. Pendekatan prediktif sebaiknya digunakan **untuk mendukung fasilitator (mediator)**, bukan menggantikan fungsi mediasi dan penilaian moral manusia.

F. Perbandingan dengan Studi Sebelumnya

Temuan ini juga memperkaya kajian internasional mengenai *algorithmic justice*. Seperti yang telah dibahas dalam studi oleh O'Neil dalam *Weapons of Math Destruction*, sistem AI yang digunakan dalam pendidikan, asuransi, hingga peradilan sering kali memperbesar jurang ketidakadilan karena tidak dikalibrasi dengan konteks sosial. Ia menegaskan bahwa sistem statistik yang “dianggap netral” ternyata justru memperparah marginalisasi.

Dalam studi oleh Berk, ditemukan bahwa sistem prediksi residivisme bisa memberikan hasil akurat jika diuji dengan data yang bersih dan bias minimal. Namun, dalam kondisi sosial yang timpang dan minim transparansi, penggunaan sistem seperti itu lebih sering **mengaburkan tanggung jawab pengambil keputusan**, karena aparat bisa berlindung di balik “hasil AI”. (Berk, Heidari, Jabbari, Kearns, & Roth, 2017)

Di Indonesia, kajian oleh Farid (2022) menunjukkan bahwa adopsi teknologi dalam penegakan hukum masih bersifat administratif dan belum menyentuh etika penggunaan data. Ini berarti, meskipun sistem digital berkembang, belum ada regulasi etik yang melindungi hak warga terhadap hasil keputusan berbasis sistem prediktif.

G. Penyimpulan Analisis: Apa yang Ditunjukkan oleh Hasil Penelitian?

Hasil penelitian ini menunjukkan bahwa **penggunaan algoritma dalam sistem keadilan restoratif membawa tantangan serius terhadap keadilan substantif**, terutama ketika sistem tersebut tidak mempertimbangkan ketimpangan historis dan sosial yang melekat dalam data yang digunakan. Dalam konteks masyarakat marginal, algoritma berisiko menjadi alat reproduksi ketimpangan jika tidak diawasi secara ketat dan diimbangi dengan nilai-nilai keadilan sosial.

Secara normatif, keadilan restoratif menekankan pendekatan yang *humanistik, relasional, dan partisipatif*. Namun, sistem algoritmik, seperti yang digunakan dalam COMPAS, atau dalam prototipe digital SPPT-TI di Indonesia, cenderung bekerja secara kalkulatif, menilai individu dengan skor dan probabilitas yang tidak selalu mampu menangkap *dimensi etika, budaya, dan kontekstual*.

Dalam prinsip hukum pidana Indonesia, sebagaimana diatur dalam KUHP, penegakan hukum tidak semata bertujuan menghukum, tetapi juga **menghormati hak-hak tersangka dan terdakwa**. Bahkan dalam praktik diversi (perkara anak), penekanan pada pemulihan telah menjadi semangat utama. Maka,

penggunaan AI dalam proses restoratif harus tunduk pada semangat serupa, bukan menjauhkan pelaku dari proses pemulihan hanya karena *angka risiko yang dihasilkan oleh algoritma*.

H. Perbedaan Hasil Penelitian Ini dengan Studi Sebelumnya

Penelitian ini memperluas diskusi yang selama ini lebih banyak menyoroti *algorithmic bias* dalam konteks **putusan pidana (*sentencing*)** dan **tindakan *pre-emptive* (*policing*)**, menjadi pada **tahapan restoratif yang bersifat non-litigasional**. Studi seperti Angwin et al. dan Eubanks memang telah mengangkat isu bias algoritmik, tetapi belum membahas **bagaimana bias tersebut berpotensi menyusup ke dalam program yang justru dirancang untuk mengoreksi ketimpangan struktural sistem hukum**.

Sebaliknya, penelitian ini menempatkan keadilan restoratif sebagai arena baru yang sedang dibangun, tetapi berisiko mengalami *kontaminasi bias digital*, jika tidak disiapkan dengan prinsip-prinsip yang berpihak pada kesetaraan. Dalam konteks hukum Indonesia, belum banyak kajian hukum yang secara spesifik mengaitkan antara **restorative justice dan teknologi algoritmik**, khususnya dalam kerangka hak asasi manusia dan nondiskriminasi.

Pada umumnya kajian yang pernah dilakukan lebih menyoroti digitalisasi peradilan dari sisi efisiensi administratif. Sementara penelitian ini menyoroti **dampak nilai yang mungkin hilang dalam pen-erjemahan proses hukum ke dalam logika mesin**, yang sangat penting dalam sistem pemulihan berbasis komunitas.

I. Rumusan Saran Strategis

Berdasarkan analisis di atas, dapat disampaikan beberapa **saran strategis** yang bersifat normatif, kebijakan, dan etik untuk memastikan bahwa penggunaan teknologi dalam sistem keadilan restoratif **tidak menjadi ancaman, tetapi tetap sejalan dengan semangat keadilan sosial dan masyarakat sipil**:

1. Audit Etik dan Transparansi Algoritma

- Pemerintah dan lembaga penegak hukum perlu mengembangkan **mekanisme audit berkala terhadap sistem algoritma** yang digunakan dalam proses keadilan pidana.
- Audit tersebut harus melibatkan **ahli hukum, pakar AI, dan perwakilan masyarakat sipil**, untuk menjamin bahwa sistem tidak menghasilkan keputusan diskriminatif.
- Transparansi algoritma (*algorithmic transparency*) harus menjadi prinsip dasar, sebagaimana ditekankan oleh OECD *AI Principles* dan UNODC.

2. Pembobotan Sosial dan Konteks Lokal

- Algoritma yang digunakan perlu dilengkapi dengan sistem *context-aware*, yaitu mempertimbangkan **latar belakang sosial, budaya, dan ekonomi** pelaku maupun korban.
- Pemerintah dapat bekerja sama dengan akademisi untuk membangun sistem berbasis **model hybrid**, yang memadukan nilai-nilai lokal (misalnya kearifan lokal dalam mediasi adat) dengan teknologi prediktif.

3. Pelatihan Etika Digital bagi Aparat Penegak Hukum

- Polisi, jaksa, hakim, dan mediator perlu diberikan pelatihan tentang **bias algoritmik dan etika teknologi hukum**, agar mereka tidak menerima hasil sistem secara mentah tanpa evaluasi kritis.
- Materi pelatihan dapat dikembangkan bersama perguruan tinggi hukum dan komunitas teknologi forensik.

4. Penguatan Regulasi Perlindungan Hak Masyarakat Marjinal

- Revisi dan penguatan regulasi berbasis restorative justice harus memperhatikan **hak kelompok rentan**, seperti anak, perempuan, penyandang disabilitas, dan masyarakat adat.
- Hal ini dapat dituangkan dalam regulasi pelaksana **RUU KUHP dan KUHP baru**, serta dalam panduan teknis di Kejaksaan dan Kepolisian.

5. Partisipasi Publik dalam Desain Teknologi

- Desain sistem teknologi hukum, termasuk sistem evaluasi risiko, **tidak boleh eksklusif dirancang oleh lembaga teknologi atau birokrat saja**.
- Masyarakat sipil harus dilibatkan sejak awal dalam **uji publik sistem prediktif**, sebagai bentuk kontrol sosial dan keterbukaan.

Penjelasan dalam bagian ini memperlihatkan bahwa teknologi dan keadilan tidak selalu selaras secara otomatis. Di satu sisi, penggunaan AI dan algoritma bisa menjadi **alat bantu penting** dalam mempercepat proses pemulihan dalam keadilan restoratif. Namun di sisi lain, jika tidak dikendalikan secara etik dan hukum, sistem ini bisa menjadi **alat eksklusi baru** terhadap kelompok yang paling membutuhkan pemulihan.

Oleh karena itu, yang diperlukan bukanlah penolakan terhadap teknologi, tetapi **penataan yang berpihak pada prinsip-prinsip keadilan sosial, partisipasi masyarakat, dan penghormatan terhadap hak asasi manusia**. Inilah titik temu antara paradigma baru peradilan pidana dan harapan terhadap masyarakat sipil yang kuat dan setara

SIMPULAN

Penelitian ini mengungkap bahwa penggunaan algoritma dalam *restorative justice* berpotensi memperkuat ketidakadilan struktural terhadap kelompok marginal, seperti yang terlihat dalam bias rasial sistem COMPAS dan risiko reproduksi ketimpangan dalam sistem digital Indonesia. Temuan menunjukkan bahwa meskipun teknologi menjanjikan efisiensi, tanpa pengawasan etik dan regulasi yang ketat, algoritma dapat mengikis prinsip partisipasi dan kesetaraan dalam restorative justice. Penelitian ini memajukan kajian hukum pidana dengan mengintegrasikan perspektif kritis teknologi dan keadilan sosial, serta merekomendasikan audit algoritmik, pelibatan masyarakat sipil, dan pembaruan regulasi berbasis HAM. Studi lanjutan diperlukan untuk menguji model hybrid yang memadukan nilai lokal dengan teknologi prediktif.

DAFTAR PUSTAKA

Jurnal

- Berk, R., Heidari, H., Jabbari, S., Kearns, M., & Roth, A., “*Fairness in criminal justice risk assessments: The state of the art.*”, *University of Pennsylvania Law Review.*, 2017
- Citron, D. K., & Pasquale, F. (2014). The scored society: Due process for automated predictions. *Washington Law Review*, 89(1), 1-33.

Buku

- Angwin, J., Larson, J., Mattu, S., & Kirchner, L., “*Machine Bias*”, Dalam Kirsten Martin., “*Ethics of Data and Analytics: Concepts and Cases*”, 1st Edition., hlm. 492., Auerbach Publications., 2022., New York. ProPublica. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- Creswell, J. and Poth.Cherryl N. “*Qualitative Inquiry and Research Design: Choosing Among Five Approaches*”, 4th ed., SAGE Publications., 2018., California.
- Israel, M. and Hay, I., “*Research Ethics for Social Scientists: Between Ethical Conduct and Regulatory Compliance*”, SAGE Publications., 2006., London.
- Eubanks, V., “*Automating inequality: How high-tech tools profile, police, and punish the poor*”, St. Martin’s Press., 2017., New York
- O’Neil, C., “*Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*”, Crown Publishing., Arizona., 2016.

Laporan dan Dokumen Resmi

- Clare Garvie., Alvaro Bedoya, and Jonathan Frankle., “*The Perpetual Line-Up: Unregulated Police Face Recognition in America*”, Georgetown Law, Center on Privacy & Technology, 2016., Washington
- United Nations Office on Drugs and Crime (UNODC). (2021). *Responsible AI for criminal justice*. United Nations.
- Amnesty International & Access Now. (2018). *The Toronto Declaration on AI and human rights*.

Peraturan dan Instrumen Hukum

- Indonesia. (1981). *Kitab Undang-Undang Hukum Acara Pidana* (UU No. 8 Tahun 1981).
- United Nations. (1966). *International Covenant on Civil and Political Rights* (ICCPR).
- Peraturan Kejaksaan Agung RI No. 15 Tahun 2020 tentang Penghentian Penuntutan Berdasarkan Keadilan Restoratif.
- Peraturan Mahkamah Agung RI No. 1 Tahun 2016 tentang Prosedur Mediasi di Pengadilan.