

Analisis Spasial Daerah Rawan Kejahatan di Sulawesi Selatan berdasarkan Data Media Online

Nadia Indriani Sumardi¹, Angelica Osianto², Meilany Olyvia Rosari³, Tyanita Puti Marindah Wardhani⁴

¹Fakultas Teknik Universitas Hasanuddin, Gowa, 92171, email: sumardini23d@student.unhas.ac.id

²Fakultas Teknik Universitas Hasanuddin, Gowa, 92171, email: osianto23d@student.unhas.ac.id

³Fakultas Teknik Universitas Hasanuddin, Gowa, 92171, email: rosarimo23d@student.unhas.ac.id

⁴Fakultas Teknik Universitas Hasanuddin, Gowa, 92171, email: tyanitaputi@unhas.ac.id

Corresponding Author: Nadia Indriani Sumardi

INTISARI — Kriminalitas atau tindak kejahatan adalah perilaku yang melanggar hukum, peraturan, dan norma serta nilai sosial dalam masyarakat. Akibat dari kejahatan ini bisa sangat merugikan dan mengancam keselamatan bahkan nyawa korbannya. Tujuan dari penelitian ini adalah untuk menyajikan analisis klasifikasi kejahatan di wilayah Sulawesi Selatan. Penelitian ini mengidentifikasi jenis kejahatan yang ada, pola waktu kejadian, lokasi dan frekuensi kejahatan. Dengan memahami karakteristik kriminalitas di daerah ini, diharapkan dapat membantu pihak berwenang dalam pengembangan strategi pencegahan dan penanggulangan kejahatan yang lebih efektif dan tepat sasaran. Metodologi penelitian ini melibatkan pengumpulan data dari berbagai sumber, pembuatan model CNN, analisis statistik dan visualisasi data dengan ArcGIS untuk menciptakan gambaran pemetaan kejahatan di Sulawesi Selatan secara komprehensif.

KATA KUNCI — Analisis Spasial, CNN, Kriminalitas, Sulawesi Selatan

I. PENDAHULUAN

A. LATAR BELAKANG

Dalam Undang-Undang Nomor 2 Tahun 2002 Pasal 13, Kepolisian Negara Republik Indonesia (POLRI) memiliki tiga tugas pokok, yaitu untuk memelihara keamanan dan ketertiban masyarakat, menegakkan hukum, dan memberikan perlindungan, pengayoman, dan pelayanan bagi masyarakat. Dengan terciptanya rasa aman dan nyaman melalui penekanan angka kejahatan oleh pihak yang berwenang diharapkan dapat menciptakan situasi yang kondusif serta nyaman untuk tetap produktif dalam menjalankan aktivitas setiap hari. Namun kondisi nyata saat ini justru berlawanan dari kata kondusif, tindak kriminalitas masih sangat marak terjadi di Indonesia.

Menurut data dari Badan Pusat Statistik, terdapat 372.965 kasus kejahatan yang terjadi pada tahun 2022, angka tersebut meningkat drastis dari tahun sebelumnya, yaitu berjumlah 239.481 kasus kejahatan. Dari keseluruhan tindak kejahatan yang terjadi, Polda Sulawesi Selatan masuk kedalam 10 besar dengan total kasus kejahatan terbanyak se-Indonesia dan menempati peringkat ke-enam dengan total kasus kejahatan sebanyak 28.679 pada tahun 2022.

Salah satu jenis kejahatan yang sering terjadi di negara-negara berkembang adalah *blue collar crime*, yang juga dikenal sebagai *street crime* atau kejahatan jalanan. Kejahatan ini biasanya menyasar fasilitas-fasilitas umum seperti jalan raya, dan pelakunya seringkali menggunakan senjata untuk melakukan kekerasan [5]. Menurut Hart, terdapat dua jenis kejahatan dalam *street crimes*, yaitu kejahatan terhadap barang

dan kejahatan terhadap orang. Contoh kejahatan terhadap barang meliputi pencurian, perampokan, dan sejenisnya. Sementara itu, contoh kejahatan terhadap orang mencakup pelecehan, penganiayaan, penculikan, perampokan, dan lainnya [6].

Selain jenis-jenis kejahatan di atas, terdapat pula fenomena begal yang akhir-akhir ini semakin marak di masyarakat. Oleh karena itu, kejahatan ini perlu mendapatkan perhatian khusus dari pemerintah dan juga aparat penegak hukum [7].

Padatnya penduduk di daerah perkotaan dan ketimpangan sosial akibat dari adanya kesenjangan ekonomi serta tingginya tingkat pengangguran merupakan pemberi kontribusi utama dalam peningkatan angka kriminalitas di suatu daerah [2]. Oleh sebab itu perlu adanya tindakan preventif untuk meminimalisir terjadinya teror kejahatan di tengah-tengah masyarakat.

Pada saat ini terdapat berbagai macam upaya telah dilakukan oleh kepolisian sebagai pihak yang berwenang untuk mencegah terjadinya tindak kejahatan yaitu dengan kegiatan patroli. Kegiatan patroli dimaksudkan untuk mengamati serta mengawasi dari potensi terjadinya suatu tindak kejahatan, patroli tersebut juga dilaksanakan pada tempat-tempat yang secara umum dianggap sebagai daerah rawan [3]. Efektifitas dari metode di atas masih belum akurat karena hanya mengandalkan sumber data berupa laporan berdasarkan kasus yang ditangani sebelumnya. Perlu adanya metode pengklasifikasian yang modern untuk memetakan daerah-daerah yang rawan, oleh sebab itu diperlukannya data yang banyak serta dalam rentang waktu yang cepat. Media sosial merupakan salah satu tempat pemberi laporan mengenai kasus kejahatan yang banyak digunakan oleh masyarakat, tidak adanya proses administrasi yang berbelit-belit, akses yang cepat dan murah, serta jangkauan yang luas menjadi sebuah pertimbangan untuk digunkannya media sosial tersebut.

Dalam penelitian ini, penggunaan Convolutional Neural Network (CNN) dan analisis spasial muncul sebagai pendekatan yang menjanjikan untuk mengenali dan mengevaluasi pola kejahatan di Sulawesi Selatan. CNN dapat digunakan untuk mengolah data media online, seperti berita dan laporan kejahatan, untuk menarik informasi tentang lokasi, jenis kejahatan, dan konteks yang terkait.

Sementara itu, analisis spasial bisa membantu dalam memetakan sebaran kejahatan di Sulawesi Selatan dan mengidentifikasi zona-zona yang rawan kejahatan. Dengan demikian, kombinasi CNN dan analisis spasial memiliki potensi untuk memberikan pemahaman yang lebih mendalam mengenai masalah kejahatan di wilayah ini, serta membantu

pihak berwenang dalam merancang langkah-langkah pencegahan yang efektif. Dengan menggabungkan teknologi dan metodologi terkini, penelitian ini bertujuan untuk memberikan kontribusi yang signifikan dalam upaya pencegahan dan penanganan kejahatan di wilayah ini.

B. TUJUAN PENELITIAN

Penelitian ini dilakukan dengan tujuan untuk :

1. Mengidentifikasi tingkat kriminalitas di Sulawesi Selatan.
2. Memprediksi pola kejahatan di Sulawesi Selatan berdasarkan data media online.
3. Memetakan hasil prediksi wilayah yang rawan kejahatan di Sulawesi Selatan berdasarkan lokasi dan waktu kejadiannya.

C. MANFAAT PENELITIAN

Penelitian ini dilakukan dengan harapan agar dapat bermanfaat untuk :

1. Membantu pihak berwenang dalam melakukan investigasi dan merencanakan strategi pencegahan kejahatan
2. Membantu dalam mengalokasikan sumber daya yang efisien sesuai kebutuhan daerah masing-masing berdasarkan tingkat kerawanan kejadiannya.
3. Meningkatkan kewaspadaan masyarakat terhadap kejahatan yang rawan terjadi di daerahnya.
4. Meningkatkan keamanan dan kesejahteraan masyarakat di Sulawesi Selatan melalui tindakan preventif oleh pihak berwenang yang lebih efektif dan tepat waktu.

D. BATASAN MASALAH

Batasan masalah dalam makalah ini yaitu :

1. Dataset yang digunakan bersumber dari X, Instagram, serta website Tribun Timur dan Fajar
2. Jenis kejahatan yang kami teliti terbatas pada pencurian, perampokan, pelecehan, pembunuhan, penculikan, penganiayaan, dan begal.
3. Batas waktu data yang diteliti, yaitu sejak 1 Maret 2019 hingga 29 Februari 2024.

II. KAJIAN TERKAIT YANG RELEVAN

Stec dan Klabjan menggunakan data kejahatan di Chicago dan Portland beserta data cuaca, kependudukan, dan transportasi publik untuk membuat model prediksi kejahatan. Dalam penelitian ini, terdapat 4 model yang digunakan, yaitu Feed Forward Network (FFN), Recurrent Neural Network (RNN), Convolution Neural Network (CNN), serta gabungan RNN dan CNN. Akurasi model terbaik didapatkan dari model gabungan RNN dan CNN dengan akurasi 75,6% untuk data Chicago dan 65,3% untuk data Portland [10].

Kim melakukan penelitian untuk mengaitkan CNN pada bidang NLP dengan bereksperimen pada beberapa model dan dataset. Temuan dari penelitian tersebut adalah bahwa model sederhana dengan menggunakan pre-trained word vector bekerja dengan sangat baik. Selain itu, dropout terbukti merupakan alat regularisasi yang baik karena dapat meningkatkan akurasi sebanyak 2% - 4% [12].

Yudha dkk menggunakan algoritma CNN dengan jumlah filter 50, 100, 200, dan 400, serta seleksi fitur Chi-Square dengan rasio 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, dan 100 % untuk membuat model yang bisa

mengklasifikasikan artikel berita sesuai kategorinya. Dari penelitian tersebut, model CNN terbaik didapatkan dengan menggunakan 200 filter dan rasio seleksi fitur 40% dan menghasilkan akurasi 96,704%, presisi 96,079%, recall 96,074%, dan F1 score 96,070% [11].

III. SOLUSI USULAN

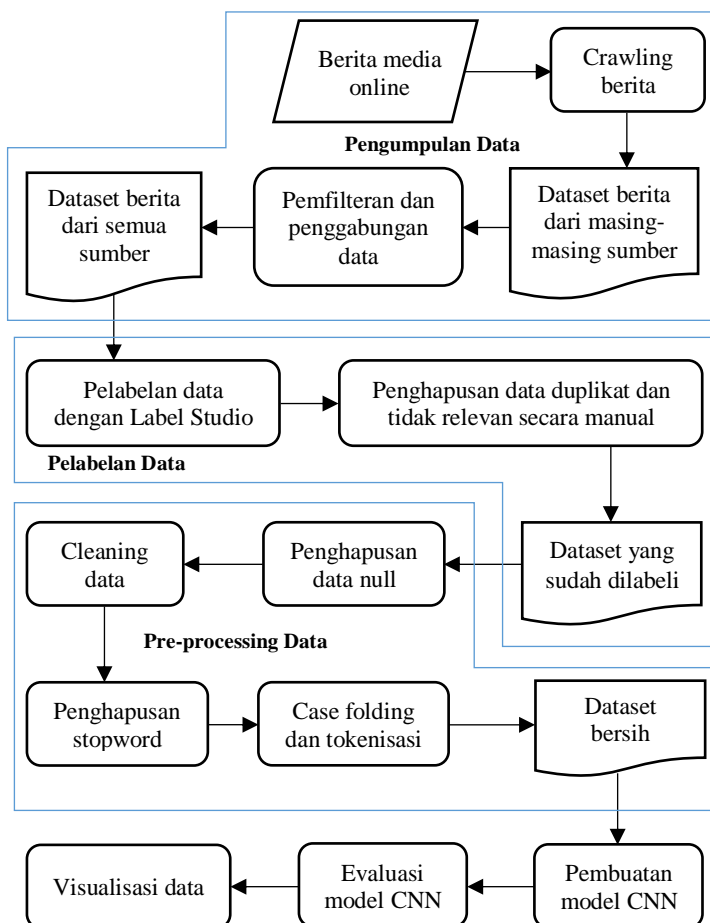
A. DESKRIPSI SOLUSI

Untuk meningkatkan pelayanan publik dalam hal keamanan dan kesejahteraan masyarakat, kami mengajukan solusi berupa pembangunan model CNN dan analisis spasial untuk mengenali dan mengevaluasi tindak kejahatan yang terjadi di Sulawesi Selatan berdasarkan data dari media online.

B. DATASET

Dataset penelitian ini merupakan data teks yang dicrawling dari website makassar.tribunnews.com, fajar.co.id, akun Instagram @makassarinfoku, dan X.

C. METODE YANG DIGUNAKAN



Gambar 1. Diagram alur metode penelitian

Tahapan penelitian ini terdiri dari pengumpulan data, pelabelan data, *pre-processing* data, pembuatan dan evaluasi model CNN serta visualisasi data dengan pemetaan.

1) PENGUMPULAN DATA

Crawling data dari X dilakukan dengan menggunakan *library* Tweet Harvest dengan filter kata kunci jenis kejahatan batas waktu penelitian. Data Instagram diperoleh menggunakan *library* Instaloader, kemudian

difilter untuk hanya menyimpan postingan terkait jenis kejahatan. Data dari situs web dikumpulkan menggunakan library BeautifulSoup 4, lxml, dan requests. Dataset dari semua sumber digabungkan ke dalam satu file dengan total 1303 dataset.

2) PELABELAN DATA

Pelabelan data dimulai dengan memasukkan nama-nama label yang dibutuhkan untuk konfigurasi pelabelan, yaitu nama 7 jenis kejahatan, 307 nama kecamatan di Sulawesi Selatan, 24 nama kabupaten/kota di Sulawesi Selatan, dan 5 pembagian waktu (pagi, siang, sore, malam, dini hari). Dataset yang digunakan untuk pelabelan data sebanyak 90% dari total data. Setelah dilakukan pelabelan serta penghapusan data duplikat dan tidak relevan, hasilnya adalah 843 *training* dan *validation set*, dan 94 *testing set*.

3) PRE-PROCESSING DATA

Proses ini menggunakan library NLTK dan Sastrawi. Pertama, menghapus data *null* yang mungkin ada pada dataset. Selanjutnya, *cleaning data* dilakukan dengan menghapus url, html, emoji, dan simbol. Kemudian, menghapus *stopword*, *case folding* untuk mengubah semua huruf menjadi huruf kecil, dan tokenisasi.

4) PEMBUATAN DAN EVALUASI MODEL CNN

Dataset yang telah melalui proses pelabelan dan *pre-processing* sudah siap untuk digunakan dalam proses training untuk pembuatan model CNN. Berikut dataset bersih untuk penelitian ini:

	text	category
0	fajar co id makassar diduga melakukan begal pa...	[Pelecehan, Kecamatan Tamalanrea, Malam]
1	fajar co id makassar kasat reskrim polrestabes...	[Penganiayaan, Kecamatan Manggala, Pagi]
2	fajar co id makassar heboh pencabulan kota mak...	[Pelecehan, Kecamatan Rappocini, Dini hari]
3	fajar co id makassar tiga pria kota makassar t...	[Pencurian, Kecamatan Rappocini]
4	fajar co id luwu timur lanjutan kasus peleceha...	[Pelecehan, Kecamatan Wotu]
...
839	fajar co id makassar seorang pria berinisial a...	[Penganiayaan, Kecamatan Panakukkang, Malam]
840	fajar co id makassar kasus penganiayaan seoran...	[Penganiayaan, Kota Makassar, Dini hari]
841	fajar co id makassar hasriani daeng caya resmi...	[Penganiayaan, Kota Makassar]
842	fajar co id makassar sekretariat mahasiswa ber...	[Penganiayaan, Kecamatan Rappocini, Dini hari]
843	fajar co id makassar kematian dn pemuda tewas ...	[Penganiayaan, Kecamatan Tinggimoncong]

844 rows × 2 columns

Gambar 2. Pratinjau dataset bersih

Pemodelan CNN dilakukan dengan menggunakan library Pandas, Scikit-learn, Numpy, dan Keras. Pertama, dilakukan *one-hot encoding* pada kolom “category” menggunakan ‘MultiLabelBinarizer’ untuk mengubahnya menjadi representasi numerik. Data teks kemudian dikonversi menjadi daftar dan dibagi menjadi *training* dan *validation set* dengan rasio 82:18, menghasilkan 692 *training set* dan 152 *validation set*. Pembagian ini bertujuan mencapai rasio 70:20:10 antara *training set*, *validation set*, dan *testing set*. Data kemudian ditokenisasi menggunakan ‘tokenizer’ dari Keras dan dipadatkan agar semua data memiliki panjang yang sama. Model CNN dibangun menggunakan library Keras dengan beberapa percobaan mengubah filter dan *dropout* untuk mendapatkan akurasi terbaik. Proses ini menggunakan ‘binary_crossentropy’ sebagai loss function dan ‘adam’ sebagai optimizer.

5) VISUALISASI DATA (PEMETAAN WILAYAH RAWAN KEJAHATAN DAN TIDAK RAWAN KEJAHATAN)

Dataset yang telah melewati proses pre-processing, kemudian digunakan untuk menentukan letak titik koordinat (latitude dan longitude). Data tersebut memuat nama daerah, seperti kabupaten dan kota, maupun kecamatan, tempat terjadinya tindak kejahatan. Proses penentuan letak titik koordinat ini disebut dengan proses geocoding. Untuk geocoding, digunakan metode Nominatim yang menggunakan data dari OpenStreetMap untuk emenentukan titik koordinata.

Setelah titik koordinat dari setiap daerah ditentukan, langkah berikutnya adalah menentukan apakah suatu daerah masuk dalam kategori rawan atau tidak rawan. Parameter yang digunakan untuk menentukan kategori ini adalah frekuensi kemunculan daerah tersebut dalam berita online yang sudah dikumpulkan. Metode kuartil ke-75 (Q3) digunakan untuk menghitung frekuensi tindak kejahatan dari setiap daerah, dengan memisahkan 75% data terkecil dari 25% data terbesar dari suatu set data yang telah disusun dari yang terkecil ke terbesar.

$$Q3 = (n+1) \times 0.75$$

(dengan n adalah jumlah data)

Kemudian data dikategorikan ke dalam daerah rawan atau tidak rawan, data tersebut dipisahkan sesuai dengan kategorinya. Langkah selanjutnya adalah proses visualisasi pemetaan menggunakan ArcGIS. Kedua kategori data tersebut dibedakan berdasarkan warna: warna merah untuk daerah rawan kejahatan dan warna hijau untuk daerah tidak rawan kejahatan. Metode clustering pada ArcGIS juga digunakan untuk mengelompokkan titik-titik data spasial berdasarkan kedekatan geografis. Metode clustering ini membantu untuk menentukan pola distribusi spasial melalui pengelompokan dan analisis daerah rawan dan tidak rawan kejahatan, sehingga berguna untuk pengambilan keputusan yang lebih baik.

D. PERBEDAAN DENGAN SOLUSI SEBELUMNYA DARI KAJIAN TERKAIT

Berbeda dengan dataset yang digunakan oleh Stec dkk., penelitian ini menggunakan dataset dari media online untuk membangun model yang mampu memprediksi pola kejahatan dari berita atau laporan kejahatan yang tersebar di media online, khususnya di Sulawesi Selatan. Penelitian ini menggunakan metode extreme multi-label text classification untuk mengkategorikan artikel kejahatan. Extreme multi-label text classification merujuk pada masalah penentuan label untuk suatu dataset dengan jumlah label yang sangat banyak, mencapai ratusan, ribuan, hingga jutaan [13]. Menurut Kim, penggunaan pre-trained word vector seperti word2vec efektif dalam membangun model CNN. Dalam penelitian ini, ‘tokenizer’ dari Keras digunakan untuk melakukan tugas tersebut. Berdasarkan hasil penelitian dari Kim (2014) dan Yudha dkk (2023), penelitian dilakukan dengan eksperimen berbagai nilai dropout (0.1, 0.2, 0.3, 0.4, 0.5) serta jumlah filter (100, 200, 300, 400, 500) untuk menemukan konfigurasi terbaik yang meningkatkan performa model.

E. METRIK EVALUASI YANG DIGUNAKAN

Metrik evaluasi yang digunakan dalam penelitian ini adalah F1 score dan ROC-AUC. F1 score adalah rata-rata harmonis

dari presisi dan *recall*, yang memberikan ukuran yang lebih akurat dari performa model ketika terdapat ketidakseimbangan antara jumlah sampel positif dan negatif. Nilai F1 score berkisar antara 0 hingga 1, semakin tinggi nilainya menandakan performa klasifikasi yang tinggi juga [14].

ROC AUC adalah metrik yang digunakan untuk mengevaluasi kinerja model klasifikasi biner. ROC (Receiver Operating Characteristic) *curve* adalah plot yang menunjukkan kemampuan model untuk membedakan antara kelas positif dan negatif pada berbagai *threshold*. Sementara, AUC (Area Under Curve) mengukur area di bawah kurva ROC, memberikan nilai tunggal yang merangkum kinerja model. Adapun interpretasi nilai evaluasi ini, yaitu [15] :

TABEL I INTERPRETASI NILAI ROC-AUC	
ROC AUC (RA)	Interpretasi
$0.9 \leq RA$	Sangat Baik
$0.8 \leq RA < 0.9$	Baik
$0.7 \leq RA < 0.8$	Cukup
$0.6 \leq RA < 0.7$	Buruk
$0.5 \leq RA < 0.6$	Gagal

IV. HASIL EKSPERIMEN DAN PENGUJIAN

A. SKENARIO PENELITIAN PERTAMA (SEMUA LABEL)

Dari hasil penelitian pertama didapatkan nilai F1 score tertinggi yaitu 0,58 pada filter 300 dan 500 dropout 0.1, filter 500 dropout 0.2 dan filter 500 dropout 0.3 serta nilai ROC-AUC tertinggi dengan nilai 0,76 pada filter 500 dropout 0.1, filter 500 dropout 0.2 dan filter 500 dropout 0.3. Angka ini termasuk nilai yang masih rendah F1 Score dan masuk di kategori “Cukup” untuk ROC-AUC. Untuk meningkatkan akurasi model CNN yang diuji, ditemukan bahwa salah satu penyebab rendahnya performa adalah ketidakseimbangan jumlah label dalam data. Oleh karena itu, dilakukan percobaan lanjutan dengan memisahkan setiap kategori label, yaitu jenis kejahatan, lokasi kejahatan (dengan memperluas cakupannya menjadi tingkat kabupaten), dan waktu kejadian. Setelah pemisahan, diperoleh 841 dataset untuk label jenis kejahatan, 794 dataset untuk label lokasi kejadian, dan 262 dataset untuk label waktu kejadian.

B. SKENARIO PENELITIAN KEDUA (LABEL JENIS KEJAHATAN)

Dari hasil penelitian kedua didapatkan nilai F1 Score tertinggi yaitu 0,89 pada filter 300 dropout 0.2 serta nilai ROC-AUC tertinggi dengan nilai 0,96 pada filter 200 dropout 0.1 untuk label jenis kejahatan. Kombinasi optimal ditemukan pada dropout 0.1 hingga 0.3 dan filter 200 hingga 500. Dropout yang lebih tinggi, seperti 0.4 dan 0.5 tetap memberikan hasil yang baik, tetapi sedikit menurunkan nilai F1 Score dibandingkan dengan dropout yang lebih rendah. Namun, nilai evaluasi tersebut sudah cukup tinggi dan sudah masuk kategori ‘Sangat Baik’ untuk evaluasi ROC-AUC.

C. SKENARIO PENELITIAN KETIGA (LABEL LOKASI KEJADIAN)

Dari hasil penelitian ketiga didapatkan nilai F1 Score tertinggi yaitu 0,89 pada filter 500 dropout 0.2 serta nilai ROC-AUC tertinggi dengan nilai 0,97 pada filter 300

dropout 0,1 dan filter 500 dropout 0.2. Nilai ini juga sudah cukup tinggi dan masuk dalam kategori ‘Sangat Baik’ untuk evaluasi ROC-AUC. Secara keseluruhan, hasil evaluasi skenario ketiga merupakan yang terbaik diantara semua skenario. Tabel 2 menunjukkan hasil penelitian skenario ketiga.

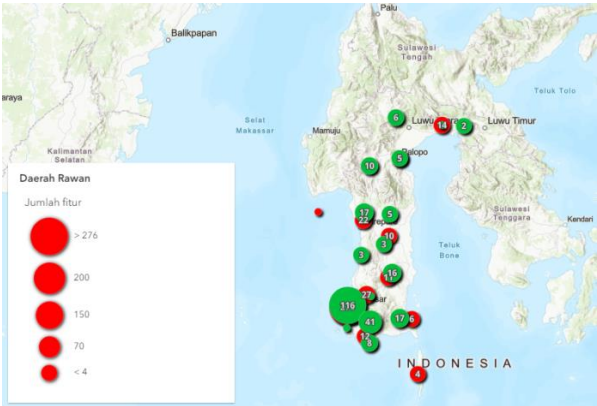
TABEL II
HASIL SKENARIO PENELITIAN KETIGA (LABEL LOKASI KEJADIAN)

Filter	Dropout	F1 Score	ROC-AUC
100	0,1	0,84	0,95
200		0,82	0,95
300		0,85	0,97
400		0,86	0,96
500		0,87	0,96
100	0,2	0,82	0,94
200		0,84	0,94
300		0,84	0,95
400		0,86	0,96
500		0,89	0,97
100	0,3	0,77	0,92
200		0,81	0,94
300		0,85	0,95
400		0,84	0,95
500		0,86	0,95
100	0,4	0,83	0,95
200		0,80	0,96
300		0,82	0,94
400		0,84	0,96
500		0,84	0,96
100	0,5	0,69	0,91
200		0,74	0,92
300		0,77	0,94
400		0,78	0,93
500		0,80	0,94

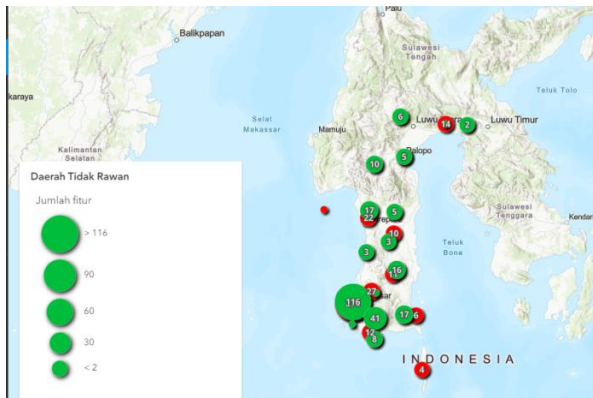
D. SKENARIO PENELITIAN KEEMPAT (LABEL WAKTU KEJADIAN)

Dari hasil penelitian keempat didapatkan nilai F1 Score tertinggi yaitu 0,58 pada filter 100 dan 400 dropout 0.1, filter 200 dropout 0.2 dan filter 300 dropout 0.4 serta nilai ROC-AUC tertinggi dengan nilai 0,81 pada filter 100 dropout 0.1 dan filter 300 dropout 0.3.

E. VISUALISASI DATA (HASIL PEMETAAN WILAYAH RAWAN KEJAHATAN DAN TIDAK RAWAN KEJAHATAN)



Gambar 3. Hasil pemetaan daerah rawan kejahatan



Gambar 4. Hasil pemetaan daerah tidak rawan kejahatan

V. ANALISIS

A. SKENARIO PENELITIAN PERTAMA (SEMUA LABEL)

Pada skenario penelitian pertama, nilai evaluasi yang tinggi didapatkan dari model dengan filter 500 dan dropout 0.1-0.3. Namun, hasil evaluasi yang masih rendah menunjukkan bahwa model kesulitan menangani dataset dengan jumlah label yang tidak seimbang, sehingga model tidak mampu mempelajari semua kelas label dengan baik.

B. SKENARIO PENELITIAN KEDUA (LABEL JENIS KEJAHATAN)

Pemisahan label jenis kejahatan memberikan hasil evaluasi yang jauh lebih baik, menunjukkan bahwa fokus pada satu jenis label dapat membantu model belajar dengan lebih efektif.

C. SKENARIO PENELITIAN KETIGA (LABEL LOKASI KEJADIAN)

Hasil skenario penelitian ketiga yang berfokus pada label lokasi kejadian menghasilkan nilai evaluasi yang terbaik diantara semua skenario. Hal ini dapat mengindikasikan bahwa informasi geografis memiliki pola yang lebih mudah dipelajari oleh model.

D. SKENARIO PENELITIAN KEEMPAT (LABEL WAKTU KEJADIAN)

Pemisahan label waktu kejadian menghasilkan nilai evaluasi yang masih sama dengan skenario pertama yang tergolong rendah. Nilai evaluasi yang masih rendah ini kemungkinan disebabkan karena jumlah dataset yang sangat sedikit dan jauh lebih rendah daripada dataset lain, yaitu hanya sebanyak 262 dataset yang mengandung label waktu kejadian

VI. KESIMPULAN

Penelitian ini membangun model CNN untuk mengkategorikan artikel kejahatan dari media online di Sulawesi Selatan dan memvisualisasikannya dengan pemetaan. Hasil penelitian menunjukkan bahwa dropout 0.1 hingga 0.3 memberikan performa yang stabil dan optimal pada hampir semua percobaan. Dropout yang lebih tinggi (0.4 hingga 0.5) tidak selalu meningkatkan performa dan bisa mengurangi nilai F1 Score, menunjukkan bahwa nilai dropout yang lebih tinggi dapat menyebabkan overfitting. Sementara itu, jumlah filter 300 hingga 500 menunjukkan hasil yang paling konsisten dan terbaik dalam meningkatkan nilai F1 Score dan ROC-AUC. Filter yang lebih kecil, seperti 100 dan 200, cenderung

memberikan nilai performa yang lebih rendah. Pemisahan data berdasarkan jenis dan lokasi kejahatan terbukti sangat efektif dalam meningkatkan performa model. Sementara itu, pemisahan berdasarkan waktu kejadian tidak memberikan peningkatan signifikan dan hasilnya tetap rendah karena jumlah dataset yang masih sangat kurang.

Secara keseluruhan, kombinasi dropout rendah (0.1 hingga 0.3) dan jumlah filter sedang hingga tinggi (300 hingga 500) adalah yang paling efektif untuk meningkatkan performa model CNN. Dataset yang lebih banyak diharapkan untuk penelitian selanjutnya agar dapat lebih meningkatkan akurasi model.

REFERENSI

- [1] BPS. 2023. Statistik Kriminal 2023. Badan Pusat Statistik. Jakarta.
- [2] Bindosano K.A., Yudono A., Hasyim A.W. 2022. 'Arahan Pengurangan Kriminalitas di Kota Jayapura'. *Planning for Urban Region and Environment*, 11(4), pp. 231-239. <https://purejournal.ub.ac.id/index.php/pure/article/view/464/372>. Diakses 14 Maret 2024.
- [3] Saupa A.M., Pioh N.R., Pangemanan F.N. 2023. 'Efektivitas Patroli Pihak Kepolisian Dalam Upaya Pencegahan Dan Penanggulangan Kejahatan', *JURNAL GOVERNANCE*, 3(2), 1-7. <https://ejournal.unsrat.ac.id/v3/index.php/governance/article/view/5002>. Diakses 14 Maret 2024.
- [4] Arif, M., F. dan Choiriyah, D. 2021. PEMANFAATAN MEDIA SOSIAL SEBAGAI SARANA KOMUNIKASI KEPADA MASYARAKAT. <https://badilag.mahkamahagung.go.id/artikel/publikasi/artikel/pemanfaatan-media-sosial-sebagai-sarana-komunikasi-kepada-masyarakat-oleh-muhammad-fajar-arief-s-h-m-h-dan-daryatul-choiriyah-s-t-28-10>. Diakses 14 Maret 2024.
- [5] Listiawati, Nora. 2022. *Memahami Berbagai Macam Bahaya Kejahatan Jalanan (Street Crime) Pada Saat Masa Pandemi Covid-19 di Indonesia*. Diakses dari <https://pid.kepri.polri.go.id/memahami-berbagai-macam-bahaya-kejahatan-jalanan-street-crime-pada-saat-masa-pandemi-covid-19-di-indonesia/#:~:text=Kejahatan%20jalanan%20ini%20pada%20umumnya,%2C%20perampokan%2C%20begal%2C%20dll>.
- [6] Hart, N. T. 2022. *Street Crime vs. White-Collar Crime*. Diakses dari <https://www.harrisonhartlaw.com/street-crime-vs-white-collar-crime/#:~:text=Street%20crime%2C%20also%20called%20E%80%9Cblue,they%20don't%20have%20to>.
- [7] Sianturi, C., Silaen, F. C., Pardosi, T. K., & Yunita, S. 2023. 'Hukum Tentang Begal Sepeda Motor di Kota Medan'. *JRP: Jurnal Relasi Publik*, 1(4), pp. 248-255
- [8] Rahmawati, L., Febrian, W. D., Fachruzzaki, Mardiyati, S., Lengam, R., Dody, I. P., & Suarnatha. 2024. 'Pengembangan Sistem Informasi Geografis (Sig) untuk Analisis Spasial dalam Pengambilan Keputusan'. *Jurnal Review Pendidikan dan Pengajaran*, 7(2), pp. 4058-4068.
- [9] Ridwan, M., & Muzakir, A. 2022. 'Model Klasifikasi Ujaran Kebencian pada Data Twitter dengan Menggunakan CNN-LSTM'. *Jurnal Teknologi dan Informatika: TEKNOMATIKA*, 12(1), pp. 209-218.
- [10] Stec, Alexander dan Klabjan, Diego. 2018. 'Forecasting Crime with Deep Learning'. arXiv:1806.01486.
- [11] Yudha, I. G. L., Sanjaya, N. A., Karyawati, E. A. I. N., Dwidasmara, I. B. G., Putra, I. G. N. A. C., & Mahendra, I. B. M. 2023. 'Article Classification Using Convolutional Neural Network (CNN) And Chi-Square Feature Selection'. *Jurnal Elektronik Ilmu Komputer Udayana*, 11(3), pp. 529-538.
- [12] Kim, Yoon. 2014. 'Convolutional Neural Networks for Sentence Classification', *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha (pp. 1746-1751). arXiv: 1408.5882. <https://doi.org/10.3115/v1/D14-1181>.
- [13] Liu, J., Chang, W., Wu, Y., & Yang, Y. 2017. 'Deep Learning for Extreme Multi-label Text Classification', *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17)*. Association for Computing Machinery, New York, NY, USA, (pp. 115-124). <https://doi.org/10.1145/3077136.3080834>.
- [14] Tharwat, Alaa. 2021. 'Classification assessment methods'. *Applied Computing and Informatics*, Vol. 17 No. 1, pp. 168-192.
- [15] Nahm, F. S. 2022. 'Receiver operating characteristic curve: overview and practical use for clinicians'. *Korean Journal of Anesthesiology*, 75(1), pp. 25-36.