

IMPLEMENTASI DATA MINING NAÏVE BAYES DAN SVM (SUPPORT VECTOR MACHINE) DALAM PENGUMPULAN BERITA KRIMINAL DI INDONESIA PADA MASA PANDEMI COVID-19

Studi Kasus : Situs Berita Detik.com

Ufil Hidayatul Laili

¹ Magister Informatika, Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang
email: uviellaili96@gmail.com

Abstract

The intensity of crime news is relatively high, making the public familiar with criminal cases ranging from corruption, drugs, gambling, immoral acts, theft, fraud, and kidnapping to murder. The media are not only reporting or informing of criminal acts that occur, but the media also acts as a tool of social control. Currently, many news sites provide information related to events and phenomena. On the other hand, the Covid-19 pandemic has created multidimensional crises and problems. One of them is the emergence of crime in the community. This study aims to collect criminal news from news sites during the Covid-19 pandemic. It can be analyzed regarding the possibility of a trend of criminal events in tandem with the direction of the Covid-19 pandemic in Indonesia. The news site used in this research is detik.com.

Keywords: News, Crime, Covid 19

1. Pendahuluan

Saat ini, masyarakat Indonesia cukup menaruh perhatian terhadap isu maupun kejadian yang terjadi di tengah mereka. Pandemi Covid 19 telah mendeterminasi berbagai perubahan signifikan pada berbagai sektor kegiatan masyarakat. Masyarakat dunia seolah dipaksa untuk menyesuaikan diri dengan perubahan yang terjadi di sekitar [1] . *Social distancing* (physical distancing) untuk mengantisipasi persebaran dan perluasan pandemi Covid 19 telah mendorong perubahan dan berbagai penyesuaian di berbagai bidang, baik perdagangan, peribadatan, pendidikan, pemerintahan dan politik. Kenyataan tersebut ternyata beriringan dengan beragamnya situs berita, terutama di era kemajuan teknologi informasi dan komunikasi seperti saat ini, Situs berita menyediakan informasi terkait kejadian maupun fenomena di tengah masyarakat dalam suatu situs web secara dalam jaringan (daring), sehingga situs berita selalu tersedia setiap saat. Adanya situs berita yang beragam membuat masyarakat dapat memperoleh informasi dengan cepat dan mudah. Di sisi lain, timbulnya pandemi Covid-19 sepanjang tahun 2020 ini tidak hanya menimbulkan permasalahan kesehatan masyarakat

di Indonesia dan dunia.

Pandemi Covid-19 juga telah memunculkan permasalahan di berbagai bidang. Adanya pandemi membuat aktivitas perekonomian global melambat dan turut memicu perlambatan perekonomian di Indonesia. Kebijakan pemerintah yang melakukan intervensi dalam meminimalisasi penyebaran virus Covid-19, seperti kebijakan Pembatasan Sosial Berskala Besar (PSBB) juga berdampak pada lemahnya perekonomian di Indonesia. Ketidakpastian perekonomian di Indonesia di masa pandemi Covid-19 dikhawatirkan memunculkan permasalahan lain. Salah satunya adalah timbulnya kriminalitas di tengah masyarakat. Menurut Prof. Dr. H. Dey Ravena, secara sederhana, kebijakan kriminal (*criminal policy*) adalah upaya rasional dari suatu negara untuk menanggulangi kejahatan. Upaya ini pada hakikatnya merupakan bagian integral yang tidak terpisahkan dari upaya perlindungan masyarakat (*social defence planning* atau *protection of society*) yang tujuannya tidak lain adalah untuk mencapai kesejahteraan masyarakat[2].

Permasalahan ekonomi seperti kemiskinan ternyata memiliki pengaruh yang positif dan signifikan terhadap kriminalitas di Indonesia. Di samping itu, dilansir oleh situs berita megapolitan.kompas.com, Kabid Humas Polda

Metro Jaya, Kombes Pol Yusri Yunus menyebutkan bahwa selama pandemi Covid-19, tepatnya pada rentang waktu September 2021, kejadian kriminal di wilayah hukum Polda Metro Jaya tidak terjadi pada kejahatan pencurian dengan kekerasan (curas) atau pencurian kendaraan bermotor (curanmor). Peningkatan hanya terjadi pada jenis tindak kejahatan lain. Polda Metro Jaya membentuk tim khusus untuk mengantisipasi kejahatan yang meningkat pada pandemi Covid-19. Tim tersebut berada di bawah Direktorat Reserse Kriminal Umum (Ditreskrim).

Hal tersebut yang menjadi latar belakang diadakannya penelitian ini, yakni untuk mengumpulkan berita kriminal dari situs berita, khususnya yang terjadi pada masa pandemi Covid-19. Pengumpulan berita adalah salah satu penerapan dari Big data, dimana Big Data dapat menjadi salah satu sumber data alternatif untuk menghasilkan statistik resmi (official statistics). Statistik resmi yang terkait dengan kriminal sendiri rutin dipublikasikan oleh Badan Pusat Statistik Republik Indonesia setiap tahunnya. Publikasi tersebut bernama "Statistik Kriminal". Selanjutnya, kemunculan berita kriminal pada situs berita juga dapat digunakan sebagai proksi (pendekatan) kejadian kriminal yang sesungguhnya terjadi dalam lingkungan masyarakat.

2. Metode Penelitian

Dalam proses penelitian ada hal penting yang harus diperhatikan, yaitu metode penelitian, karena dalam proses penelitian, metode penelitian merupakan bagian terpenting dari proses penelitian, tujuannya agar suatu proses penelitian bisa sesuai dengan objek yang akan diteliti. Pada penelitian ini, peneliti menjelaskan mengenai desain penelitian, langkah analisa, melakukan implementasi metode Naïve Bayes dan melakukan implementasi metode SVM berkenaan dengan penelitian yang dilakukan.

Penelitian ini menggunakan pendekatan kualitatif dengan metode penelitian yang digunakan adalah metode penelitian Deskriptif. Metode analisis deskriptif kualitatif yaitu suatu metode penelitian yang dilakukan untuk menggambarkan proses atau peristiwa yang sedang berlaku pada saat ini di lapangan yang di jadikan objek penelitian, kemudian data atau informasinya di analisis sehingga di peroleh suatu pemecahan masalah[3].

Sedangkan pendekatan kualitatif dipandang

lebih relevan dan cocok karena bertujuan menggali dan memahami apa yang tersembunyi dibalik kriminal pada masa pandemi covid – 19. Penelitian deskriptif kualitatif merupakan penelitian yang berusaha mendeskripsikan dan mengintreprestasikan sesuatu, misalnya kondisi atau hubungan yang ada, pendapat yang berkembang, proses yang sedang berlangsung, akibat yang terjadi, atau tentang kecenderungan yang tengah berlangsung[4]. Dalam penelitian ini, Langkah – Langkah yang diperoleh sebagai berikut :

- a. Pengumpulan data, yaitu langkah yang paling strategis dalam penelitian, karena tujuan utama dari penelitian adalah mendapatkan data.
- b. Melakukan Web Scraping berita yaitu sebagai proses pengambilan data dari sebuah website.
- c. Reduksi Data, yaitu penyederhanaan, penggolongan, dan membuang yang tidak perlu data sedemikian rupa sehingga data tersebut dapat menghasilkan informasi yang bermakna dan memudahkan dalam penarikan kesimpulan. Banyaknya jumlah data dan kompleksnya data, diperlukan analisis data melalui tahap reduksi. Tahap reduksi ini dilakukan untuk pemilihan relevan atau tidaknya data dengan tujuan akhir.
- d. Display Data, yaitu Penyajian data merupakan kegiatan saat sekumpulan data disusun secara sistematis dan mudah dipahami, sehingga memberikan kemungkinan menghasilkan kesimpulan, bisa berupa teks naratif (berbentuk catatan lapangan), matriks, grafik, jaringan ataupun bagan.
- e. Kesimpulan dan Verifikasi, yaitu bertujuan untuk mencari makna data yang dikumpulkan dengan mencari hubungan, persamaan, atau perbedaan untuk ditarik kesimpulan sebagai jawaban dari permasalahan yang ada.
- f. Melakukan klasifikasi metode Naïve Bayes
 - i. Melakukan Inisialisasi Data yaitu membaca data training
 - ii. Menentukan jenis data / memfilter data numerik atau bukan
 - iii. Menentukan mean tiap parameter
 - iv. Menentukan standar devisiasi tiap parameter
 - v. Menentukan tabel mean dan standar devisiasi
- g. Melakukan klasifikasi menggunakan SVM

- i. Melakukan proses penyeleksian data / pemilihan data
- ii. Melakukan pre proses / pembersihan data
- iii. Melakukan Transformasi
- iv. Melakukan Data Mining
- v. Evaluasi

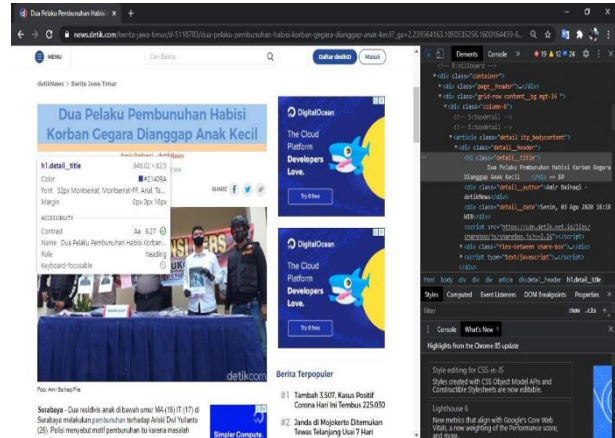
Penelitian ini juga melakukan pencarian data melalui sumber-sumber tertulis untuk memperoleh informasi mengenai objek penelitian ini. Dalam studi literatur ini penulis menganut sistem kepustakaan terbuka dimana dengan mengumpulkan data atau keterangan melalui bahan bacaan mengenai masalah-masalah yang diteliti.

Internet searching merupakan teknik pengumpulan data melalui bantuan teknologi yang berupa alat / mesin pencari di internet [5] dimana segala informasi dari detik.com tersedia didalamnya. Internet searching sangat memudahkan dalam rangka membantu peneliti menemukan suatu file / data dimana kecepatan, kelengkapan dan ketersediaan data dari berbagai tahun tersedia. Mencari data di internet bisa dilakukan dengan cara searching, browsing, surfing ataupun downloading.

3. Hasil Dan Pembahasan

Teknik Web Scraping

Pada penelitian ini yang menggunakan Teknik Web Scraping diperoleh dari situs berita detik.com dimana proses dari web scraping secara umum dilakukan yaitu Pertama, identifikasi terlebih dahulu tag HTML pada situs berita yang memuat berita dan informasi penyerta. Identifikasi dokumen HTML bertujuan untuk mengapit informasi yang akan diambil. Setelah identifikasi dilakukan, kemudian web scraper dapat dibuat. Web scraper dalam package scrapy dibentuk dalam file python berekstensi .py yang biasa disebut dengan istilah “spider”. Hasilnya akan ditampilkan pada gambar dibawah ini :

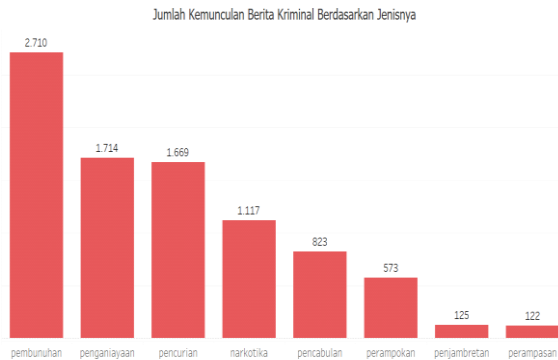


Gambar 1 Web Scraping

Web scraping dalam penelitian ini dilakukan sebanyak dua kali dengan menggunakan dua buah spider yang berbeda. Spider yang pertama digunakan untuk mengekstrak tautan (link) berita kriminal. Output dari spider yang pertama ini berupa kumpulan tautan berita kriminal berformat .json. Sedangkan spider yang kedua digunakan untuk mengekstrak informasi

yang ada pada artikel berupa judul, tanggal, tempat kejadian dan isi berita, dimana artikel-artikel tersebut diakses menggunakan kumpulan tautan yang telah diekstrak oleh spider yang pertama. Jumlah beritanya yang berhasil dikumpulkan sebanyak 9.310 buah artikel. Output dari spider yang kedua berupa artikel yang mengandung isi berita dan informasi penyerta seperti yang terlihat pada gambar dibawah ini :

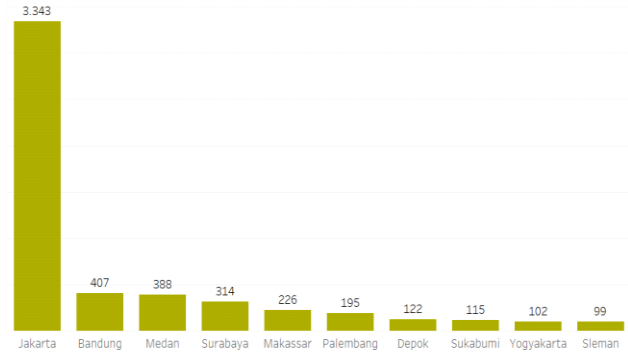
analisis, juga diteliti mengenai kemungkinan adanya tren jumlah kejadian kriminal beriringan dengan tren jumlah kasus terkonfirmasi Covid-19 di Indonesia. Hal pertama yang dapat ditelusuri adalah mengenai jumlah kemunculan berita kriminal berdasarkan jenis kriminal, seperti yang terlihat pada gambar di bawah ini.



Gambar 4 Jumlah Kemunculan Berita Kriminal Berdasarkan Jenisnya

Dapat dilihat pada gambar diatas bahwa pembunuhan adalah jenis kriminal yang paling banyak muncul dalam situs detik.com dibanding jenis kriminal lainnya dengan jumlah kemunculan sebanyak 2.710 buah artikel, kemudian disusul penganiayaan dengan 1.714 buah artikel, pencurian dengan 1.669 buah artikel hingga jenis kriminal yang paling sedikit muncul yakni perampasan dengan 122 buah artikel. Lalu, selain berdasarkan jenis kriminal, dapat diketahui juga mengenai jumlah kemunculan berita kriminal berdasarkan tempat kejadian, seperti yang terlihat pada gambar di bawah ini.

10 Besar Jumlah Kemunculan Berita Kriminal Terbanyak Berdasarkan Tempat



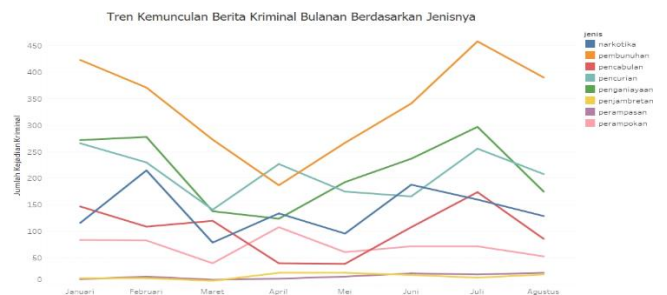
Gambar 5 10 Besar Jumlah Kemunculan Berita Kriminal Terbanyak Berdasarkan Tempat

Insight selanjutnya yang dapat ditelusuri adalah mengenai keterkaitan antara pandemi Covid-19 dengan kemunculan berita kriminal pada masa tersebut. Dapat dilihat pada gambar dibawah ini :



Gambar 6 Tren Perkembangan Jumlah Kasus Terkonfirmasi Covid-19

Selanjutnya akan ditampilkan gambar tren Kemunculan Berita Kriminal Bulanan Berdasarkan Jenisnya, dibawah ini :



Gambar 7 Kemunculan Berita Kriminal Bulanan Berdasarkan Jenisnya

Metode Klasifikasi Teorema Naïve Bayes

Naïve Bayes Classifier atau bisa disebut sebagai multinomial Naïve Bayes merupakan model penyederhanaan dari algoritma bayes [6] yang ocook dalam pengklasifikasian teks atau dokumen Dasar dari teorema Naïve Bayes yang digunakan dalam pemrograman adalah sebagai berikut.

Setelah *preprocessing*, tahapan selanjutnya pada penelitian yang dilakukan adalah menentukan dataset berdasarkan nilai *k* pada *kfold cross validation* yang ditentukan. Metode *cross validation* membagi semua dokumen ke dalam sekumpulan *k* dataset, dimana setiap dataset tersebut memiliki kesempatan untuk menjadi *testing set* [7]. Saat dataset ke-*i* menjadi *testing set*, maka dataset lainnya digunakan untuk menentukan frekuensi seluruh kata pada kategori *j* (), frekuensi kata ke-*i* pada kategori *j* () dan probabilitas setiap dokumen pada kategori *j* terhadap seluruh dokumen (*P(v)*). Dapat dikatakan dataset lainnya tersebut menjadi *training set* pada *classifier* NBC. Setiap pengujian yang dilakukan pada *testing set* diukur presisi, sensitivitas (*recall*) dan akurasi dengan menggunakan *confusion matrix*. Berikut gambar Ilustrasi *k-fold cross validation* dibawah ini :



Gambar 8 Ilustrasi K-fold Cross Validation

Dataset yang telah dilakukan preprocessing tersebut terbagi ke dalam lima kategori, yaitu korupsi, narkoba, pencurian, pemerkosaan dan pembunuhan, sehingga terdapat 100 dokumen di setiap kategori. Pengujian dilakukan dengan teknik *K-Fold Cross Validation* dengan nilai *K* dari 3 sampai 10.

Hasil Pengujian dengan k = 3

Dengan menggunakan *3-fold*, perbandingan jumlah dokumen pada *training set* dan *testing set* adalah sebesar 66,67 : 33,33. Jumlah dokumen untuk setiap *fold* adalah sebanyak 170 dokumen. Pengujian dilakukan sebanyak 3 iterasi.

Tabel 1 Hasil Pengujian dengan k = 3

Iteration	Precision	Recall	Accuracy
1	94.38 %	98.29 %	98.48 %
2	100 %	98.65 %	99.73 %
3	100 %	98.37 %	99.67 %
rata-rata	98.13 %	98.44 %	99.29 %

Dapat dilihat pada Tabel diatas, nilai rata-rata *precision* sebesar 98,13 %, nilai rata-rata *recall* sebesar 98,44 % dan nilai rata-rata *accuracy* sebesar 99,29 %. Dengan demikian nilai rata-rata dari *precision*, *recall* dan *accuracy* untuk nilai *k* = 3 adalah sebesar 98,62 %

Hasil Pengujian dengan k = 4

Dengan menggunakan *4-fold*, perbandingan jumlah dokumen pada *training set* dan *testing set* adalah sebesar 75 : 25. Jumlah dokumen untuk setiap *fold* adalah sebanyak 125 dokumen. Pengujian dilakukan sebanyak 4 iterasi. Dapat dilihat pada tabel dibawah ini :

Tabel 2 Hasil Pengujian dengan k = 4

Iteration	Precision	Recall	Accuracy
1	98.41 %	98.29 %	99.34 %
2	98.43 %	99.47 %	99.57 %
3	96.81 %	96.47 %	98.65 %
4	98.41 %	98.46 %	99.37 %
rata-rata	98.02 %	98.17 %	99.23 %

Pada Tabel diatas, nilai rata-rata *precision* sebesar 98,02 %, nilai rata-rata *recall* sebesar 98,17 % dan nilai rata-rata *accuracy* sebesar 99,23 %. Dengan demikian nilai rata-rata dari *precision*, *recall* dan *accuracy* untuk nilai *k* = 4 adalah sebesar 98,47 %.

Hasil Pengujian dengan k = 5

Dengan menggunakan *5-fold*, perbandingan jumlah dokumen pada *training set* dan *testing set*

adalah sebesar 80 : 20. Jumlah dokumen untuk setiap *fold* adalah sebanyak 100 dokumen. Pengujian dilakukan sebanyak 5 iterasi.

Tabel 3 Hasil Pengujian dengan $k = 5$

Iteration	Precision	Recall	Accuracy
1	98.05 %	99.62 %	99.52 %
2	96.09 %	97.41 %	98.68 %
3	100 %	98.42 %	99.68 %
4	98.02 %	98.06 %	99.21 %
5	96.03 %	95.92 %	98.38 %
rata-rata	97.64 %	97.89 %	99.09 %

Dapat dilihat pada Tabel diatas, nilai rata-rata *precision* sebesar 97,64 %, nilai rata-rata *recall* sebesar 97,89 % dan nilai rata-rata *accuracy* sebesar 99,09 %. Dengan demikian nilai rata-rata dari *precision*, *recall* dan *accuracy* untuk nilai $k = 5$ adalah sebesar 98,21 %.

Hasil Pengujian dengan $k = 6$

Dengan menggunakan *6-fold*, perbandingan jumlah dokumen pada *training set* dan *testing set* adalah sebesar 83,33 : 16,67. Jumlah dokumen untuk setiap *fold* adalah sebanyak 85 dokumen. Pengujian dilakukan sebanyak 6 iterasi. Sebagaimana tabel dibawah ini :

Tabel 4 Hasil Pengujian dengan $k = 6$

Iteration	Precision	Recall	Accuracy
1	100 %	96.58 %	99.31 %
2	100 %	100 %	100 %
3	97.	97.98 %	99.12 %
4	100 %	100 %	100 %
5	95.45 %	97.48 %	98.55 %
6	100 %	100 %	100 %
rata-rata	98.86 %	98.67 %	99.5 %

Dapat dilihat pada Tabel diatas, nilai rata-rata *precision* sebesar 98,86 %, nilai rata-rata *recall* sebesar 98,67 % dan nilai rata-rata *accuracy* sebesar 99,5 %. Dengan demikian nilai rata-rata dari *precision*, *recall* dan *accuracy* untuk nilai $k = 6$ adalah sebesar 99,01 %

Hasil Pengujian dengan $k = 7$

Dengan menggunakan *7-fold*, perbandingan jumlah dokumen pada *training set* dan *testing set* adalah sebesar 85,71 : 14,29. Jumlah dokumen untuk setiap *fold* adalah sebanyak 75 dokumen. Pengujian dilakukan sebanyak 7 iterasi. Sebagaimana pada tabel dibawah ini :

Tabel 5 Hasil Pengujian dengan $k = 7$

Iteration	Precision	Recall	Accuracy
1	97.42 %	99.23 %	99.31 %
2	100 %	98.28 %	99.65 %
3	97.43 %	99.93 %	99.45 %
4	100 %	96.46 %	99.28 %
5	100 %	96.28 %	99.25 %
6	100 %	99.09 %	99.82 %
7	100 %	100 %	100 %
rata-rata	99.26 %	98.47 %	99.54 %

Pada Tabel diatas, nilai rata-rata *precision* sebesar 99,26 %, nilai rata-rata *recall* sebesar 98,47 % dan nilai rata-rata *accuracy* sebesar 99,54 %. Dengan demikian nilai rata-rata dari *precision*, *recall* dan *accuracy* untuk nilai $k = 7$ adalah sebesar 99,09 %.

Hasil Pengujian dengan $k = 8$

Dengan menggunakan *8-fold*, perbandingan jumlah dokumen pada *training set* dan *testing set* adalah sebesar 87,5 : 12,5. Jumlah dokumen untuk setiap *fold* adalah sebanyak 65 dokumen. Pengujian dilakukan sebanyak 8 iterasi. Sebagaimana pada tabel dibawah ini :

Tabel 6 Hasil Pengujian dengan $k = 8$

Iteration	Precision	Recall	Accuracy
1	100	96.97 %	99.39 %
2	97.05 %	99.72 %	99.33 %
3	100	96.32 %	99.25 %
4	100	96.69 %	99.33 %
5	97.05 %	99.72 %	99.33 %
6	97.04 %	99.21 %	99.23 %
7	100	100 %	100 %
8	97	98	98.98 %
rata-rata	98.52 %	98.33 %	99.36 %

Pada Tabel diatas, nilai rata-rata *precision* sebesar 98,52 %, nilai rata-rata *recall* sebesar 98,33 % dan nilai rata-rata *accuracy* sebesar 99,36 %. Dengan demikian nilai rata-rata dari *precision*, *recall* dan *accuracy* untuk nilai $k = 8$ adalah sebesar 98,74 %.

Hasil Pengujian dengan $k = 9$

Dengan menggunakan 9-fold, perbandingan jumlah dokumen pada *training set* dan *testing set* adalah sebesar 88,89 : 11,11. Jumlah dokumen untuk setiap *fold* adalah sebanyak 60 dokumen. Pengujian dilakukan sebanyak 9 iterasi. Sebagaimana pada tabel dibawah ini :

Tabel 7 Hasil Pengujian dengan $k = 9$

Iteration	Precision	Recall	Accuracy
1	100 %	100 %	100 %
2	100 %	100 %	100 %
3	100 %	98.14 %	99.62 %
4	96.63 %	93.93 %	98.11 %
5	100 %	100 %	100 %
6	100 %	100 %	100 %
7	93.66 %	96.83 %	98.03 %
8	100 %	100 %	100 %
9	100 %	100 %	100 %
rata-rata	98.92 %	98.77 %	99.53 %

Pada Tabel diatas, nilai rata-rata *precision* sebesar 98,92 %, nilai rata-rata *recall* sebesar 98,77 % dan nilai rata-rata *accuracy* sebesar 99,53 %. Dengan demikian nilai rata-rata dari *precision*, *recall* dan *accuracy* untuk nilai $k = 9$ adalah sebesar 99,07 %.

Hasil Pengujian dengan $k = 10$

Dengan menggunakan 10-fold, perbandingan jumlah dokumen pada *training set* dan *testing set* adalah sebesar 90 : 10. Jumlah dokumen untuk setiap *fold* adalah sebanyak 50 dokumen. Pengujian dilakukan sebanyak 10 iterasi. Sebagaimana pada tabel dibawah ini :

Tabel 8 Hasil Pengujian dengan $k = 10$

Pada Tabel diatas, nilai rata-rata *precision* didapatkan kesimpulan bahwa klasifikasi

sebesar 98,86 %, nilai rata-rata *recall* sebesar

Iteration	Precision	Recall	Accuracy
1	100	100 %	100 %
2	100	100 %	100 %
3	100	100 %	100 %
4	100	100 %	100 %
5	96.21 %	99.45 %	99.09 %
6	100	94.59 %	98.9 %
7	92.42 %	95.61 %	97.51 %
8	100	100 %	100 %
9	100	98.12 %	99.62 %
10	100	100 %	100 %
rata-rata	98.86 %	98.78 %	99.51 %

98,78 % dan nilai rata-rata *accuracy* sebesar 99,51 %. Dengan demikian nilai rata-rata dari *precision*, *recall* dan *accuracy* untuk nilai $k = 10$ adalah sebesar 99,05 %

Rata-Rata Hasil Pengujian

Berikut disajikan tabel mengenai rata-rata nilai *precision*, *recall* dan *accuracy* dari hasil pengujian dengan k sama dengan 3 sampai 10.

Tabel 9 Rata-Rata Hasil Pengujian

K-Fold	Precision	Recall	Accuracy
3	98.13 %	98.44 %	99.29 %
4	98.02 %	98.17 %	99.23 %
5	97.64 %	97.89 %	99.09 %
6	98.86 %	98.67 %	99.5 %
7	99.26 %	98.47 %	99.54 %
8	98.52 %	98.33 %	99.36 %
9	98.92 %	98.77 %	99.53 %
10	98.86 %	98.78 %	99.51 %
rata-rata	98.53 %	98.44 %	99.38 %

Dapat dilihat pada Tabel diatas, dari seluruh pengujian menggunakan *K-Fold Cross Validation* dengan nilai k dari 3 sampai 10 menghasilkan rata-rata nilai *precision*, *recall* dan *accuracy* yang saling berdekatan dengan nilai yang sangat tinggi. Jika diambil nilai rata-rata,

berita kriminal dengan menggunakan NBC memiliki nilai *precision* sebesar 98,53 %, nilai *recall*

sebesar 98,44 % dan nilai *accuracy* sebesar 99,38 %.

Metode SVM (Support Vector Machine)

Pada penelitian klasifikasi berita kriminal secara *online* digunakan metode *support vector machine*. Fungsi kernel yang akan digunakan adalah kernel linier dan polynomial. Berikut merupakan pembahasan dari kernel linier dan kernel polynomial pada data *training* dan data *testing*.

SVM Menggunakan Kernel Linier pada Data Training

Data *training* kelas pada artikel berita telah diketahui. Dimana tujuan data *training* adalah untuk menghasilkan ketepatan klasifikasi. Dalam penelitian ini akan dilakukan beberapa percobaan *word vector* untuk memeberikan hasil yang paling optimum. Tiap *word vector* dicari ketepatan klasifikasi yang paling baik dengan menggunakan kernel linier. Dengan menggunakan parameter *c* dengan rentang nilai 10^{-3} sampai 10^3 didapatkan hasil seperti pada Tabel dibawah ini :

Ketepatan Klasifikasi (%)
Tabel 10 Ketepatan Klasifikasi

C	0.001	0.01	0.1	1	10	100	1000
1009	98.95	100	100	100	100	100	100
1595	99.04	100	100	100	100	100	100
2220	99.4	100	100	100	100	100	100
2595	99.49	100	100	100	100	100	100
3038	99.56	100	100	100	100	100	100
3784	99.62	100	100	100	100	100	100

Berdasarkan Tabel diatas dapat diketahui bahwa dengan menggunakan kernel linier untuk setiap *word vector* dengan menggunakan *k-fold cross validation* sebesar 10 *fold* didapatkan nilai ketepatan paling besar 100% pada semua *word vector* dengan menggunakan *c*=0.01 sampai *c*=1000. Pada *c*=0.001 didapatkan hasil ketepatan klasifikasi yang berbeda-beda. Parameter *c*=1 akan digunakan pada data *testing*. Selanjutnya akan melihat pengukuran performansi pada data *training word vector* 3784 untuk tiap *fold*.

Dari hasil pengukuran performansi kernel linier tiap *fold* pada data *training* didapatkan hasil rata-rata dari 10 *fold* untuk akurasi total, *recall* total, *precision* total dan *F-Measure* total pada tiap *fold* dalam data *training* menghasilkan nilai 100%. Selanjutnya hasil pengukuran performansi tiap kategori untuk data *training* dengan *word vector* 3784.

Hasil yang didapatkan untuk pengukuran performansi tiap kategori pada data *training* dengan menggunakan *word vector* 3784 menghasilkan rata-rata dari 10 *fold* untuk akurasi total, *recall* total, *precision* total, dan *F-Measure* total sebesar 100%. Selanjutnya akan dilakukan SVM dengan menggunakan kernel *polynomial* pada data *training*.

SVM Menggunakan Kernel Polynomial pada Data Training

Kernel polynomial merupakan pengembangan dari kernel linier dengan menambahkan parameter γ , r dan p . Selama percobaan didapatkan hasil yang paling baik adalah dengan parameter $\gamma = 1$, $r = 6$, dan $p = 3$. Didapatkan hasil seperti pada Tabel dibawah ini:

Tabel 11 SVM Menggunakan Kernel Polynomial pada Data Training

Ketepatan Klasifikasi (%)

C	0.001	0.01	0.1	1	10	100	1000
1009	93.42	99.28	100	100	100	100	100
1595	85.08	98.64	100	100	100	100	100
2220	82.35	98.84	100	100	100	100	100
2595	81.02	98.93	100	100	100	100	100
3038	79.13	98.95	100	100	100	100	100
3784	79.8	99.08	100	100	100	100	100

Tabel diatas menunjukkan bahwa pada *c*=0.1 sampai dengan *c*=1000 didapatkan hasil ketepatan klasifikasi sebesar 100%. Setelah mengetahui ketepatan klasifikasi pada data *training* dengan menggunakan kernel polynomial dilakukan pengukuran performansi pada *word vector* 3784. Selanjutnya parameter *c*=0.1 akan digunakan untuk melihat akurasi dari data *testing*.

Hasil pengukuran performansi untuk tiap *fold* pada data *training* dengan menggunakan *word vector* sebesar 3784 menghasilkan rata-rata dari 10 *fold* untuk akurasi total, *recall* total, *precision* total dan *F-Measure* total sebesar

100% Selanjutnya dilakukan pengukuran performansi tiap kategori. Dari hasil yang didapat diperoleh dari data *training* untuk tiap kategori dengan rata-rata 10 *fold* menghasilkan 100% untuk pengukuran performansi pada *word vector* 3784.

4. Kesimpulan Dan Saran

Dari hasil penelitian ini dapat disimpulkan sebagai berikut:

Telah dilakukan beberapa metode dari data yang diperoleh dari situs berita secara online yaitu detik.com dengan menggunakan Teknik web scraping. Dari hasil web scraping kemudian akan menghasilkan datasets kalsifikasi berita yang diolah beberapa metode yaitu metode teorema Naïve Bayes dan SVM (Support Vector Machine)

Validation,” *Jurnal Sains Dan Informatika*, Vol. 5, No. 2, Pp. 108–117, Dec. 2019, Doi: 10.34128/Jsi.V5i2.177.

M. Ramdhan, *Metode Penelitian*. Cipta Media Nusantara, 2021.

M. Linarwati *Et Al.*, “Studi Deskriptif Pelatihan Dan Pengembangan Sumberdaya Manusia Serta Penggunaan Metode Behavioral Event Interview Dalam Merekrut Karyawan Baru Di Bank Mega Cabang Kudus,” 2016.

Daftar Pustaka

- B. T. Akhirfiarta, “Literasi Digital Pada Pegawai RSUD Dr. Soetomo Surabaya,” *Jurnal Fisip Universitas Airlangga*, 2017, Accessed: Aug. 05, 2022. [Online]. Available: [Http://Journal.Unair.Ac.Id/Download-Fullpapers-Lna3e16b8d81full.Pdf](http://journal.unair.ac.id/download-fullpapers-lna3e16b8d81full.pdf)
- D. Darwis, N. Siskawati, And Z. Abidin, “Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional,” *Jurnal Tekno Kompak*, Vol. 15, No. 1, Pp. 131–145, Feb. 2021, Doi: 10.33365/Jtk.V15i1.744.
- F. Silaen, D. Syawal, And A. Siregar, “Hubungan Kebijakan Kriminal Dengan Kebijakan Hukum Pidana,” *Jurnal Darma Agung*, Vol. 28, No. 1, Pp. 8–16, May 2020, Doi: 10.46930/Ojsuda.V28i1.455.
- G. A. R. Maria And S. T. Raharjo, “Adaptasi Kelompok Usia Produktif Saat Pandemi Covid-19 Menggunakan Metode Reality Therapy,” *Jurnal Kolaborasi Resolusi Konflik*, Vol. 2, No. 2, Pp. 142–149, Aug. 2020, Doi: 10.24198/Jkrk.V2i2.29124.
- H. Rhomadhona, J. Permadi, P. A. Studi Teknologi Informasi Politeknik Negeri Tanah Laut Jl Yani Km, D. Panggung Pelaihari, And K. Selatan, “Klasifikasi Berita Kriminal Menggunakan Naïve Bayes Classifier (Nbc) Dengan Pengujian K-Fold Cross