



## 9<sup>th</sup> Applied Business and Engineering Conference

---

### PREDIKSI TINGKAT KEJAHATAN BERDASARKAN ARTIKEL BERITA NASIONAL MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE (SVM)*

Ajulio Pady Sembiring<sup>1)</sup>, Sharfina Faza<sup>2)</sup>

<sup>1,2</sup> Teknik Komputer dan Informatika, Politeknik Negeri Medan, Jalan Almamater No. 1,  
Kampus USU Padang Bulan, Medan, 20155

E-mail: [ajuliosembiring@polmed.ac.id](mailto:ajuliosembiring@polmed.ac.id)

#### Abstract

Crime is any act or act that is economically and psychologically detrimental, violates the laws and regulations in force in the Indonesian province, as well as social and religious norms. Society opposes it because it can be interpreted as a crime that violates laws and social norms. Research This study aims to predict criminal acts which include narcotics, murder, theft, robbery and maltreatment. The data was taken from online media websites such as detik.com, okezone.com, tribunnews.com and kompas.com. The crime data was taken using a scraping technique. The method used in this research is the Support Vector Machiner algorithm by utilizing the python programming language with jupyter tools. From testing data testing as much as 30%, it shows a dataset with trend, seasonal and residual variables using the support vector machine algorithm with the Radial Basis Fauction kernel resulting in RMSE of 3,132 on murder crime data, 0,32 on prediction of robbery crimes, 2,017 on crime prediction cases. theft, 2.49 on the prediction of the crime of assault cases and 2.49 on the prediction of the crime of narcotics cases. Thus, from the results of model testing, it can be concluded that the trend, seasonal and residual variables are accurate and have the lowest RMSE value

**Keywords:** *Crime, Support Vector Machiner, Radial Basis Fauction*

#### Abstrak

Kejahatan adalah setiap perbuatan atau perbuatan yang merugikan secara ekonomi dan psikis, melanggar peraturan perundang-undangan yang berlaku di provinsi Indonesia, serta norma-norma sosial dan agama. Masyarakat menentangnya karena dapat diartikan sebagai kejahatan yang melanggar hukum dan norma sosial. Penelitian Penelitian ini bertujuan untuk melakukan prediksi tindakan kriminal yang meliputi narkoba, pembunuhan, pencurian, perampokan dan penganiayaan. Data diambil dari website website media online seperti detik.com, okezone.com, tribunnews.com dan kompas.com data kejahatan tersebut diambil menggunakan teknik scraping. Metode yang digunakan dalam penelitian ini adalah metode algoritma Support Vector Machiner dengan memanfaatkan bahasa pemrograman python dengan tools jupyter. Dari pengujian data testing sebanyak 30%, menunjukkan dataset dengan variabel tren, seasonal dan residual menggunakan algoritma support vector machine dengan kernel Radial Basis Fauction menghasilkan RMSE sebesar 3.132 pada data kejahatan pembunuhan, 0,32 pada prediksi kejahatan kasus perampokan, 2.017 pada prediksi kejahatan kasus pencurian, 2.49 pada prediksi kejahatan kasus penganiayaan dan 2.49 pada prediksi kejahatan kasus narkoba. Dengan demikian dari hasil pengujian model



# 9<sup>th</sup> Applied Business and Engineering Conference

dapat disimpulkan bahwa variabel trend, seasonal dan residual akurat dan memiliki nilai RMSE terendah.

**Kata Kunci:** *Kejahatan, Support Vector Machiner, Radial Basis Fauction*

## PENDAHULUAN

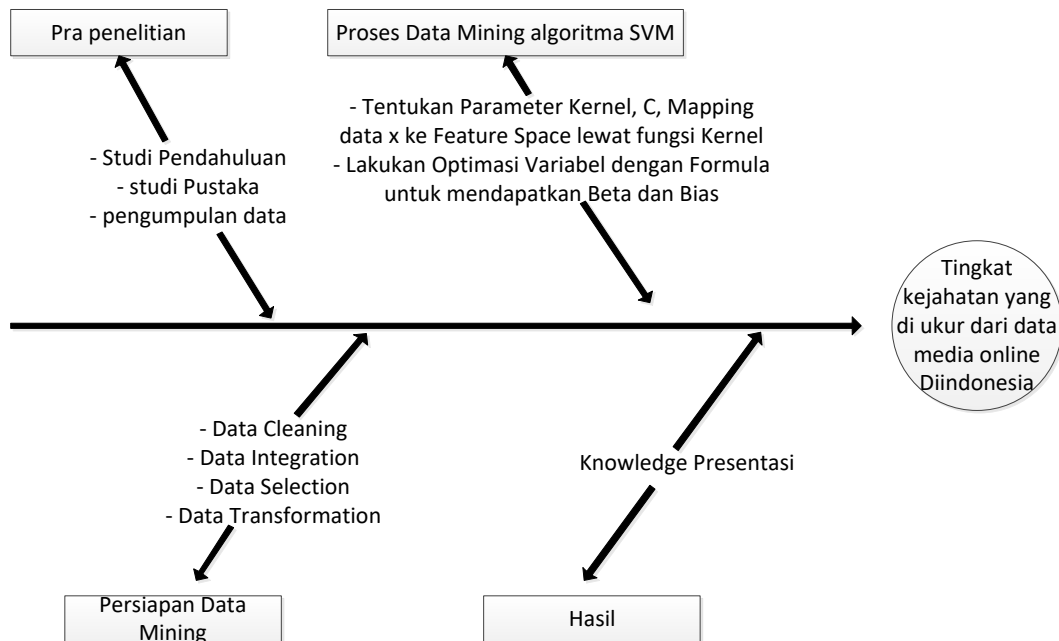
Tindakan kejahatan meningkat pesat dari hari ke hari dan memiliki efek sosial yang signifikan. Sehingga kejahatan dijadikan satu isu utama yang terus berkembang (Dulkiah & Nurjanah, 2018). Kejahatan adalah bentuk perbuatan yang melanggar hukum baik itu perdata maupun pidana yang merupakan salah satu isu yang dominan dan membuat gelisah dalam masyarakat (Putra, 2016). Lembaga penegak hukum mendapatkan informasi data kejahatan dari perkembangan teknologi informasi seperti media-media online yang ada di Indonesia (Damayanti et al., 2016). Kejadian kejahatan tidak dapat diprediksi dan dari pendataan ditemukan bahwa banyak faktor yang mempengaruhi tindakan kejahatan seperti pengangguran, kemiskinan, dan narkoba (Taram, 2019). Media informasi online merupakan sumber informasi harian yang terdapat artikel tindakan kejahatan yang dapat dibaca oleh warga atau masyarakat di Indonesia. Beberapa portal media informasi online seperti okezone.com, Tribunews.com, kompas.com dan detik.com menjadi media media informasi online yang sering diakses berdasarkan penilaian dari situs alexa.com. Penulis melakukan eksplorasi terhadap media online di Indonesia yang fokus terhadap kategori berita kejahatan. Menurut data kepolisian persentase kejahatan di Indonesia semakin meningkat dari tahun ke tahun sehingga perlu ada nya eksplorasi data pada media media di Indonesia dengan melakukan teknik crawling dan scraping untuk mendata persentase kategori berita kejahatan yang ada di Indonesia

## METODE PENELITIAN

Pada Bab ini akan melakukan pembahasan proses prediksi tingkat kejahatanitas berdasarkan dari berita online dengan melihat tingkat kejahatan dari berita tersebut dengan menggunakan teknik crawling dan memanfaatkan metode *support vector machine*.

## 1. Tahapan-Tahapan Penelitian

Proses evaluasi penelitian meliputi proses implementasi dari awal sampai akhir sebagai berikut.



**Gambar 1. Tahapan Penelitian**

Metodologi penelitian yang digunakan yaitu fishbone diagram. Penelitian ini dibagi menjadi tiga tahapan utama yaitu:

1. Pra penelitian dimulai dengan melakukan studi pendahuluan untuk mengidentifikasi masalah, literature review dan pengumpulan data
2. Persiapan proses data mining terdiri dari, Data Cleaning, Data Authentication, Data Integration, Data Selection dan Data Transformation

## 2. Pra Penelitian

### a) Studi Pendahuluan

Studi pendahuluan dilakukan untuk memperoleh informasi tentang penelitian yang akan dikerjakan. Langkah awal yang dilakukan yaitu menentukan masalah penelitian. Studi



## 9<sup>th</sup> Applied Business and Engineering Conference

---

pendahuluan berfungsi untuk memperjelas kedudukan masalah peneliti dengan keadaan lingkungan atau bidang yang sesungguhnya.

### b) Studi Pustaka

Studi pustaka merupakan sebuah usaha yang dilakukan untuk mendapatkan bahan pengetahuan dan informasi yang relevan, informasi tersebut dapat diperoleh dari buku, laporan penelitian, peraturan-peraturan, ketetapan-ketetapan, jurnal dan sumber tertulis lainnya baik tercetak maupun elektronik.

### c) Pengumpulan Data

Di dalam memperoleh data Tindakan kejahatan pada media media online yang akan digunakan pada proses data mining, diawali dengan melalui Teknik crawling pada media media online diindonesia

### 3. Lokasi Penelitian

Lokasi yang peneliti ambil untuk melakukan penelitian ini adalah media media informasi online di Indonesia yaitu [www.okezone.com](http://www.okezone.com), [www.tribunnews.com](http://www.tribunnews.com), [www.kompas.com](http://www.kompas.com), [www.detik.com](http://www.detik.com). Alasan peneliti memilih tempat penelitian tersebut berdasarkan situs-situs berita yang sering diakses oleh masyarakat umum

### 4. Parameter Pengukuran dan Pengamatan

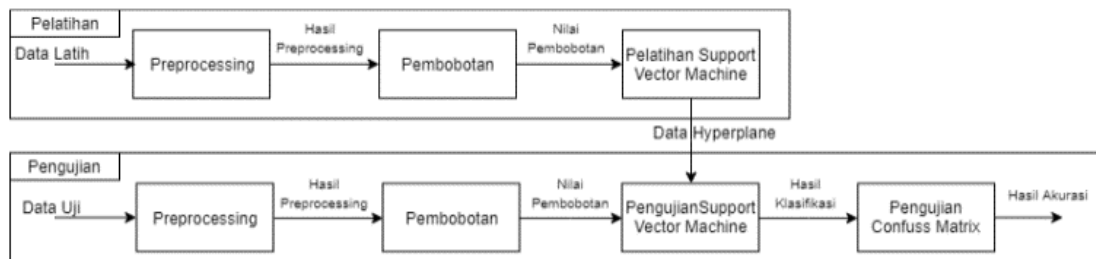
Pada proses pelatihan menggunakan data yang sudah disiapkan sebagai data latih akan melewati tahap 3 proses yaitu proses pre-processing yang menggunakan data data dari hasil crawling berita online diIndonesia dengan kategori kejahatan yang sudah ditentukan pembobotan tf-idf, dan klasifikasi Support Vector Machine ditahap pelatihan. Proses preprocessing meliputi Case Folding, Cleaning, Filtering, Tokenizing, Normalisasi Bahasa dan Stopwords Removal. Pada tahap pembobotan tf-idf akan dilakukan pembobotan pada tiap kata yang sudah melalui tahap preprocessing. Pada tahap klasifikasi SVM terdapat proses melatih data yang dilakukan untuk mendapatkan data hyperlane yang didapatkan dari data training yang telah diinput kemudian data hyperlane tersebut dimasukkan ke database.

## 5. Model Penelitian

Pada penelitian yang akan dilaksanakan dibagi kedalam beberapa tahapan yaitu

1. Pre- Processing
2. Pembobotan
3. Pelatihan
4. Pengujian

Adapun model dari penelitian ini secara keseluruhan digambarkan seperti berikut :



**Gambar 2. Model Penelitian**

## HASIL DAN PEMBAHASAN

### 1. Data yang digunakan

Data yang digunakan merupakan data hasil dari crawling dan scraping yang terdapat pada media media online diindonesia seperti detik.com, kompas.com, okezone.com, teribunnews.com dari tahun 2019 sampai 2020. Data hasil carwling dan scraping dibagi berdasarkan kategori kejahatan seperti, pencurian, narkoba, penganiayaan, pembunuhan, perampokan dengan jumlah 23.919 data kemudian data tersebut dilakukan pembagian sesuai dengan kategori kejahatan. Berikut ini jumlah data dari setiap media online sebagai berikut:



## 9<sup>th</sup> Applied Business and Engineering Conference

**Tabel 1. Total Kejahatan pada website okezzzone.com**

No	Kejahatan	Total kasus
1	Narkotika	2076
2	Penganiayaan	1533
3	Pembunuhan	2383
4	Perampokan	370
5	Pencurian	1464
	Jumlah	7826

**Tabel 2. Total Kejahatan pada website kompas.com**

No	Kejahatan	Total kasus
1	Narkotika	1876
2	Penganiayaan	1785
3	Pembunuhan	2973
4	Perampokan	546
5	Pencurian	1563
	Jumlah	8743

**Tabel 3. Total Kejahatan pada website tribunnews.com**

No	Kejahatan	Total kasus
1	Narkotika	2391
2	Penganiayaan	1589
3	Pembunuhan	2839

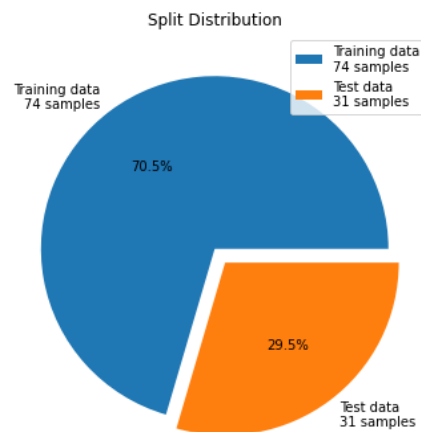
No	Kejahatan	Total kasus
4	Perampokan	643
5	Pencurian	1875
	Jumlah	9337

**Tabel 4. Total Kejahatan pada website detik.com**

No	Kejahatan	Total kasus
1	Narkotika	1982
2	Penganiayaan	923
3	Pembunuhan	1783
4	Perampokan	768
5	Pencurian	1876
	Jumlah	7332

### 2. Data training dan testing

Data training merupakan bagian dataset yang kita latih untuk membuat prediksi sesuai tujuannya masing-masing. Pada penelitian ini data *training* dan *testing* yang digunakan pada setiap kejahatan berbeda beda namun dengan proporsi yang sama yaitu 70% untuk data latih dan 30% untuk data tes. Berikut ini adalah salah satu data *training* dan *testing* yang digunakan untuk melakukan prediksi kejahatan berdasarkan kategori yang sudah ditentukan sebagai berikut.



**Gambar 3. Data training**

Data yang digunakan pada sampel *training* dan *testing* merupakan data *resample* dengan interval mingguan. *Resample* merupakan *sampling* berulang berdasarkan interval waktu tertentu. Dari hasil tersebut didapatkan 105 sampel dari 731 sampel untuk data narkoba.

### SIMPULAN

Setelah Penelitian yang dilakukan dengan jumlah input data kejahatan menunjukkan bahwa penggunaan algoritma support vector machine dengan pemaksimalan nilai parameter untuk prediksi tingkat kejahatan berdasarkan kategori berita kejahatan pada media media online seperti okezone, kompas, tribun dan detik dapat menghasilkan nilai yang cukup baik. Dari pengujian data testing sebanyak 30%, menunjukkan dataset dengan variabel tren, seasonal dan residual menggunakan *Algoritma support vector machine dengan kernel Radial Basis Fauction* menghasilkan



## 9<sup>th</sup> Applied Business and Engineering Conference

RMSE sebesar 0.127 pada prediksi kejahatan dari kasus Penganiayaan 1.759 kemudian dilakukan *smoothing* data dengan metode *moving average* terjadi peningkatan kualitas model yang ditunjukkan pada menurunnya tingkat error RMSE yaitu 2.49.

Pada prediksi kejahatan kasus pencurian, dataset dengan variabel tren, seasonal dan residual menggunakan *Algoritma support vector machine dengan kernel Radial Basis Fauction* menghasilkan RMSE sebesar 2.451 kemudian dilakukan *smoothing* data dengan metode *moving average* terjadi peningkatan kualitas model yang ditunjukkan pada menurunnya tingkat error RMSE yaitu 2.017.

Pada prediksi kejahatan kasus perampokan dataset dengan variabel tren, seasonal dan residual menggunakan *Algoritma support vector machine dengan kernel Radial Basis Fauction* menghasilkan RMSE sebesar 1.275 pada prediksi kejahatan kasus penganiayaan, dataset dengan variabel tren, seasonal dan residual menggunakan *Algoritma support vector machine dengan kernel Radial Basis Fauction* menghasilkan RMSE sebesar 0,32.

Pada prediksi kejahatan kasus pembunuhan Dataset dengan variabel tren, seasonal dan residual menggunakan *Algoritma support vector machine dengan kernel Radial Basis Fauction* menghasilkan RMSE sebesar 3.793 kemudian dilakukan *smoothing* data dengan metode *moving average* terjadi peningkatan kualitas model yang ditunjukkan pada menurunnya tingkat error RMSE yaitu 3.123.

Pada prediksi kejahatan kasus Narkotika, dataset dengan variabel tren, seasonal dan residual menggunakan *Algoritma support vector machine dengan kernel Radial Basis Fauction* menghasilkan RMSE sebesar 3.882 kemudian dilakukan *smoothing* data dengan metode *moving average* terjadi peningkatan kualitas model yang ditunjukkan pada menurunnya tingkat error RMSE yaitu 2.49. Sehingga dapat disimpulkan bahwa dari semua hasil testing menunjukkan nilai RMSE yang baik dengan menggabungkan *smoothing* data dengan metode *moving average*. Dengan demikian dari hasil pengujian model diatas dapat disimpulkan bahwa variabel trend, seasonal dan residual akurat dan memiliki nilai RMSE terendah.

### DAFTAR PUSTAKA

- Annur, H. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes. *Ilkom Jurnal Ilmiah*, 10(2), 160–165.
- Damayanti, F. N., Piarsa, I. N., & Sukarsa, I. M. (2016). Sistem Informasi Geografis Pemetaan Persebaran Kriminalitas Di Kota Denpasar. *Merpati*, 4(1), 22–32. <https://doi.org/10.24843/Jim>





## 9<sup>th</sup> Applied Business and Engineering Conference

- Dewi, S. N., Cholissodin, I., & Santoso, E. (2018). Prediksi Jumlah Kriminalitas Menggunakan Metode Extreme Learning Machine ( Studi Kasus Di Kabupaten Probolinggo ). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-Ptiik) Universitas Brawijaya*, 2(11), 4687–4693.
- Dulkiah, M., & Nurjanah. (2018). Pengaruh Kemiskinan Terhadap Tingkat Tindak Kriminalitas Di Kota Bandung. *Jurnal Ilmu Sosial Dan Ilmu Politik*, 8(2), 57.
- Halilah, M. (2019). *Perbandingan Peramalan Jumlah Penumpang Bus Damri Yogyakarta Menggunakan Metode Artificial Neural Network Dan Support Vector Regression*. Muhammadiyah University, Semarang.
- Panggabean, D. S. O., Buulolo, E., & Silalahi, N. (2020). Penerapan Data Mining Untuk Memprediksi Pemesanan Bibit Pohon Dengan Regresi Linear Berganda. *Jurikom (Jurnal Riset Komputer)*, 7(1), 56–62.
- Pratibha, Gahalot, A., Uprant, Dhiman, S., & Chouhan, L. (2020). Crime Prediction And Analysis. *2nd International Conference On Data, Engineering And Applications, Idea 2020*. <https://doi.org/10.1109/Idea49133.2020.9170731>
- Putra, R. S. (2016). Kriminalitas Di Kalangan Remaja (Studi Terhadap Remaja Pelaku Pencabulan Di Lembaga Pemasyarakatan Anak Kelas Ii B Pekanbaru). *Jurnal Ilmu Komunikasi*, 3(2), 1–15. <https://media.neliti.com/media/publications/127491-id-pengaruh-bauran-promosi-terhadap-minat-b.pdf>
- Putri, R. A. (2016). *Klasifikasi Laporan Kriminalitas Pada Situs Laport Dengan Naïve Bayes Classifier (Nbc) Menggunakan Bahasa Pemrograman*. 196. <http://repository.its.ac.id/63233/1/5212100072-undergraduate-thesis.pdf>
- Rahutomo, F., Pratiwi, I. Y. R., & Ramadhani, D. M. (2019). Naïve Bayes's Experiment On Hoax News Detection In Indonesian Language. *Jurnal Penelitian Komunikasi Dan Opini Publik*, 23(1).
- Sivanagaleela, B., & Rajesh, S. (2019). Crime Analysis And Prediction Using Fuzzy C-Means Algorithm. *Proceedings Of The International Conference On Trends In Electronics And Informatics, Icoei 2019, Icoei*, 595–599. <https://doi.org/10.1109/Icoei.2019.8862691>
- Sumari, A. D. W., Pratama, R. Y. A., & Triswidrananta, O. D. (2021). Sistem Prediksi Tingkat Kriminalitas Menggunakan Metode Triple Exponential Smoothing: Studi Kasus Pada Polres Kabupaten Probolinggo. *Jurnal Teknik Informatika*, 13(2), 171–178. <https://doi.org/10.15408/jti.v13i2.18128>
- Suriani, L. (2020). Pengelompokan Data Kriminal Pada Poldasu Menentukan Pola Daerah Rawan Tindak Kriminal Menggunakan Data Mining Algoritma K-Means



## 9<sup>th</sup> Applied Business and Engineering Conference

---

Clustering. *Jurnal Sistem Komputer Dan Informatika (Jsn)*, 1(2), 151–157.

Taram, Nyoman Gde. (2019). Pengelompokan Tingkat Kriminalitas Dengan Metode Agglomerative Dan K-Means Serta Peubah Pencirinya. *E-Jurnal Matematika*, 8(2), 102. <https://doi.org/10.24843/Mtk.2019.V08.I02.P241>

Xia, J., Wan, W., Liu, R., Chen, G., & Feng, Q. (2015). Distributed Web Crawling: A Framework For Crawling Of Micro-Blog Data. *Iet Conference Publications*, 2015(Cp672), 62–68. <https://doi.org/10.1049/Cp.2015.0255>