

# Penerapan *Support Vector Machine* dengan *SMOTE* Untuk Klasifikasi Sentimen Pemberitaan *Omnibus Law* Pada Situs *CNNIndonesia.com*\*

Widiananda Putri Hutami<sup>1</sup>, Hari Wijayanto<sup>2‡</sup>, Itasia Dina Sulvianti<sup>3</sup>

<sup>1</sup>Indekstat Indonesia, Indonesia

<sup>2,3</sup>Department of Statistics, IPB University, Indonesia

<sup>‡</sup>corresponding author: [hrwijayanto@gmail.com](mailto:hrwijayanto@gmail.com)

Copyright © 2022 Widiananda Putri Hutami, Hari Wijayanto, and Itasia Dina Sulvianti. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Abstract

The declaration of the omnibus law reaped the pros and cons in the community. In a situation like this, the media should be neutral. One of the media that still maintains neutrality is Detik (Rumata 2017). Detik owns several channels such as detikNews, detikFinance, and CNN Indonesia. In this study, the neutrality of the CNN Indonesia media as part of Detik will be studied based on the tendency of sentiment on the omnibus law-related news. Sentiment analysis is used to examine the trend of opinion on news headlines. In conducting sentiment analysis, a method that supports classification is needed. The classification method that will be used in this research is the Support Vector Machine (SVM). There is an imbalance of data in the three categories of sentiment so that the Synthetic Minority Oversampling Technique (SMOTE) method is used to overcome this imbalance. The omnibus law tends to be reported neutrally by *CNNIndonesia.com* site. The one vs all method has a better classification result than the one vs one method. The application of SMOTE only gives slightly better results than data classification without the application of SMOTE because the imbalance in the data is not too extreme. Modeling using the one vs all method with SMOTE and distribution of data 90% train data 10% test data gives the best classification results with a macro average f1-score of 60,33%.

**Keywords:** media neutrality, omnibus law, sentiment analysis, SMOTE, support vector machine.

## 1. Pendahuluan

*Omnibus Law* merupakan salah satu topik yang ramai dibicarakan media sepanjang tahun 2020. Istilah *Omnibus Law* pertama kali muncul dalam pidato pertama Joko Widodo setelah dilantik sebagai Presiden RI pada Minggu, 20 Oktober 2019. *Omnibus Law* merupakan suatu peraturan dalam membuat satu UU baru yang dapat mengamandemen beberapa UU sekaligus (Suriadinata 2019). Semenjak

---

\* Received: Aug 2021; Reviewed: Sep 2021; Published: Jan 2022

dicanangkan, gelombang penolakan terhadap *Omnibus Law* terus mencuat. Berbagai lapisan masyarakat menyuarakan sikap keberatannya terhadap *Omnibus Law*. Pada situasi ini, media diharapkan dapat menjalankan peran normatifnya dengan baik, yaitu harus mampu mewadahi percakapan publik dalam isu *Omnibus Law*.

Belakangan ini tren media mulai bergeser dari media cetak ke media *online*. Kehadiran media *online* mempermudah masyarakat untuk lebih cepat mendapat informasi. Namun, kemudahan ini menghasilkan persoalan baru. Juditha (2013) menyebutkan bahwa media *online* seringkali menjadi sorotan karena kurang mengedepankan objektivitas dan netralitas dalam pemberitaannya. Objektivitas dan netralitas telah menjadi standar baku sekaligus sebagai suatu tolak ukur tingkat profesionalitas kinerja media (McQuail 1992). Dalam pemberitaan *Omnibus Law*, media diharapkan untuk tidak menjadi humas pemerintah dan dituntut untuk bersikap objektif dan netral dalam setiap pemberitaannya.

Detik termasuk salah satu sumber berita *online* utama bagi pengguna internet di Indonesia. Detik memiliki beberapa kanal-kanal seperti detikNews, detikSport, detikFinance, Wolipop, CNN Indonesia, dll. Rumata (2017) menyatakan bahwa Detik merupakan portal berita *online* yang masih menjaga netralitas dalam pemberitaannya. Selain itu, Wardani dan Indrayani (2018) juga menjelaskan bahwa portal berita Detik cenderung menjaga stabilitas pemberitaan sehingga tidak menimbulkan spekulasi berlebihan yang belum tentu sesuai fakta. Pada penelitian ini, penulis mencoba untuk mengkaji objektivitas dan netralitas media *online* CNN Indonesia sebagai bagian dari media *online* Detik dalam pemberitaan mengenai *Omnibus Law*.

Analisis sentimen dilakukan untuk melihat kecenderungan opini terhadap sebuah objek oleh seseorang. Dalam penelitian ini, analisis sentimen dilakukan pada judul berita. Judul berita mampu mengantarkan pembaca ke arah ideologi tertentu yang diuraikan pada isi berita serta dinilai dapat mewakili ideologi media bersangkutan terhadap apa yang diberitakan (Rizam 2019). Oleh karena itu, analisis sentimen dapat dilakukan pada judul berita untuk memeriksa objektivitas dan netralitas dari suatu artikel berita. Dalam melakukan analisis sentimen diperlukan metode yang menunjang klasifikasi. Metode klasifikasi yang digunakan pada penelitian kali ini adalah *Support Vector Machine* (SVM).

Penelitian sebelumnya yang berkaitan adalah oleh Vidya (2015) tentang *twitter sentiment analysis* terhadap *brand reputation*, menunjukkan bahwa SVM memiliki performansi yang lebih baik dibanding algoritma klasifikasi *Naïve Bayes* dan *Decision Tree* dengan nilai akurasi 82.40%. Selain itu penelitian oleh Amelia (2018) dengan judul “Pemodelan *Support Vector Machine* Data Tidak Seimbang Keberhasilan Studi Mahasiswa Magister IPB” menunjukkan bahwa penanganan data tidak seimbang menggunakan SMOTE berhasil meningkatkan kinerja klasifikasi SVM.

## 2. Metodologi

### 2.1 Data

Data diperoleh dengan proses penarikan data judul berita dari portal berita *online* CNNIndonesia.com menggunakan aplikasi ParseHub. Penarikan data dilakukan dengan menggunakan kata kunci “*Omnibus Law*” dan “UU Ciptaker”. Penarikan data dilakukan dalam 3 periode, yaitu 2 bulan sebelum pengesahan (Agustus – September 2020), masa pengesahan (Oktober 2020), dan 2 bulan setelah pengesahan (November – Desember 2020).

## 2.2 Prosedur Analisis Data

Prosedur analisis data yang akan dilakukan adalah sebagai berikut :

1. Mengumpulkan data.
2. Melakukan *Pre-processing*.  
Tahapan *pre-processing* yang akan dilakukan pada keseluruhan data teks meliputi penghapusan data tidak unik, *cleaning*, *casefolding*, *stopword removing*, *stemming*, *tokenizing*, dan pelabelan data.
3. Melakukan Pembobotan TF-IDF.  
Dalam penelitian ini dilakukan pembobotan dengan menggunakan metode *Term Frequency – Inverse Document Frequency* (TF-IDF). Pembobotan kata diperlukan untuk proses klasifikasi dengan algoritma *Machine Learning*.
4. Melakukan Eksplorasi data.  
Tahapan eksplorasi data yang akan dilakukan meliputi pembentukan awan kata, eksplorasi pada periode pemberitaan, dan eksplorasi pada kategori sentimen positif dan negatif.
5. Melakukan Pemodelan Klasifikasi.
  - a. Membagi data menjadi data latih dan data uji masing-masing 80:20 dan 90:10. Pembagian data diatur sedemikian rupa sehingga perbandingan kelas minoritas dan mayoritas pada data latih dan data uji relatif sama dengan data asli. Hal ini dilakukan agar dapat merepresentasikan kondisi ketidakseimbangan pada data awal.
  - b. Melakukan klasifikasi pada data latih menggunakan SVM kernel linear (membandingkan metode *one vs one* dan *one vs all*) dan mengevaluasi kinerja klasifikasi pada data uji.
  - c. Menerapkan SMOTE pada data latih. Metode SMOTE digunakan untuk melakukan *oversampling* pada kelas minoritas.
  - d. Melakukan klasifikasi SVM pada data latih baru yang telah melalui tahap SMOTE menggunakan SVM kernel linear (membandingkan metode *one vs one* dan *one vs all*) serta mengevaluasi kinerja klasifikasi pada data uji.
  - e. Membandingkan kinerja klasifikasi SVM sebelum dan sesudah SMOTE.
6. Kesimpulan.

## 3. Hasil dan Pembahasan

### 3.1 Eksplorasi Data

Data judul berita dengan kata kunci “*Omnibus Law*” dan “UU Ciptaker” yang telah melalui *pre-processing* menghasilkan total data sebanyak 1940 data. Berdasarkan sentimen nya, jumlah data terbagi menjadi 135 data sentimen positif, 1557 data sentimen netral, dan 248 data sentimen negatif.

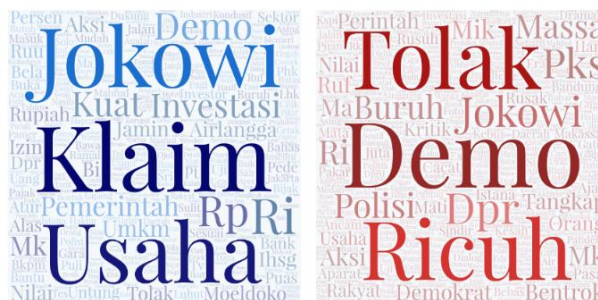
#### 3.1.1 Pembentukan Awan Kata

Awan kata yang dihasilkan dari keseluruhan data judul berita dapat dilihat pada Gambar 1. Berdasarkan awan kata pada Gambar 1, kata “Demo”, “Buruh”, dan “Tolak” adalah tiga kata dengan frekuensi kemunculan paling banyak pada keseluruhan data judul berita.



Gambar 1 Awan kata dari keseluruhan data judul berita

Awan kata untuk data judul berita dengan sentimen positif dan negatif dapat dilihat pada gambar 2. Tiga kata dengan frekuensi kemunculan terbanyak untuk data judul berita bersentimen positif adalah kata “Klaim”, “Jokowi”, dan “Usaha”. Tiga kata dengan frekuensi kemunculan terbanyak untuk data judul berita bersentimen negatif adalah kata “Demo”, “Tolak”, dan “Ricuh”.



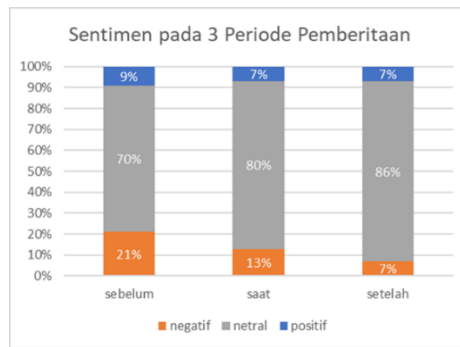
(a)

(b)

Gambar 2 Awan kata untuk data judul berita (a) sentimen positif (b) sentimen negatif

### 3.1.2 Eksplorasi pada 3 Periode Pemberitaan

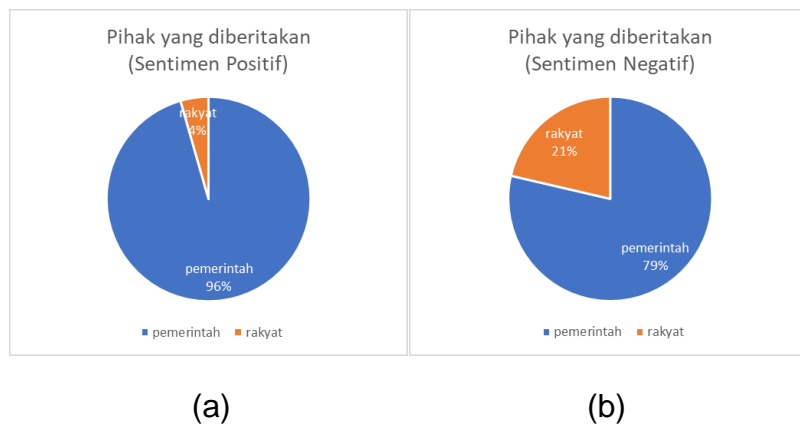
Proporsi masing-masing sentimen pada tiap periode pemberitaan dapat dilihat pada Gambar 3. Berdasarkan informasi yang terdapat pada Gambar 3, dapat dikatakan bahwa sentimen netral cenderung mendominasi di setiap periode pemberitaan. Sentimen positif cenderung memiliki jumlah persentase yang sama pada tiap periode pemberitaan sedangkan pada sentimen negatif, proporsi nya cenderung semakin menurun pada tiap periode pemberitaan.



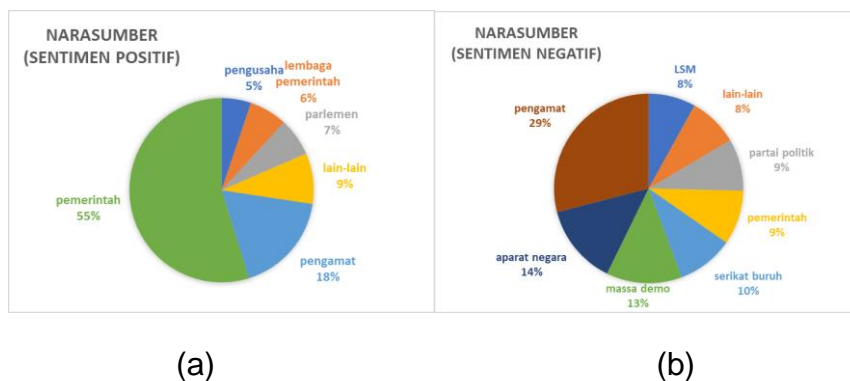
Gambar 3 Proporsi masing-masing sentimen pada tiap periode pemberitaan

### 3.1.3 Eksplorasi pada 3 Kategori Sentimen Berita

Persentase pihak yang diberitakan pada sentimen positif dan negatif dapat dilihat pada Gambar 4. Proporsi antara rakyat dan pemerintah sebagai pihak yang diberitakan pada kategori sentimen negatif cenderung lebih besar daripada kategori sentimen positif. Pemerintah mendominasi pihak yang diberitakan pada kategori sentimen positif dengan persentase sebesar 96%.



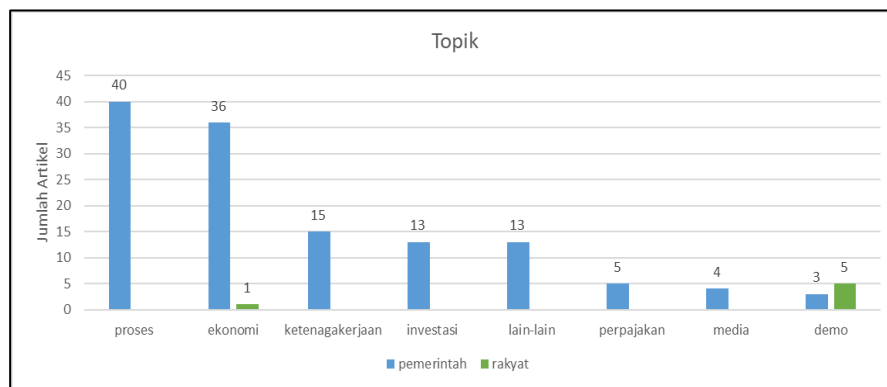
Gambar 4. Persentase pihak yang diberitakan pada judul berita (a) sentimen positif (b) sentimen negatif



Gambar 5. Persentase narasumber pada judul berita (a) sentimen positif (b) sentimen negatif

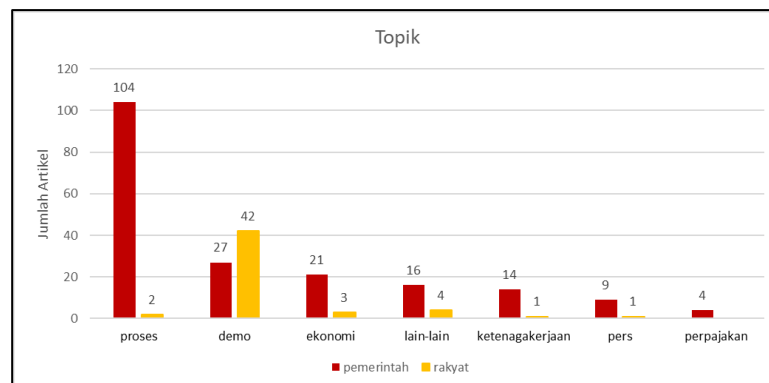
Gambar 5 menunjukkan persentase narasumber pada judul berita bersentimen positif dan negatif. Narasumber pada berita sentimen positif didominasi oleh pemerintah dengan persentase 55%. Pada sentimen negatif, narasumber didominasi oleh pengamat dengan persentase 29%.

Gambar 6 menunjukkan banyaknya artikel pada berita bersentimen positif berdasarkan topik yang dibicarakan. Pada berita sentimen positif dengan pemerintah sebagai pihak yang diberitakan, pujian terkait proses penyusunan hingga pengesahan *Omnibus Law* dan dampak positif *Omnibus Law* pada sektor ekonomi adalah dua topik yang paling banyak dibicarakan. Sedangkan pada berita sentimen positif dengan rakyat sebagai pihak yang diberitakan, kondusifnya situasi demo menjadi topik yang paling banyak dibicarakan.



Gambar 6 Banyaknya artikel pada berita sentimen positif berdasarkan topik yang dibicarakan

Gambar 7 menunjukkan banyaknya artikel pada berita bersentimen negatif berdasarkan topik yang dibicarakan. Pada berita sentimen negatif dengan pemerintah sebagai pihak yang diberitakan, topik yang paling banyak dibicarakan adalah kritikan kepada pemerintah mengenai proses penyusunan hingga pengesahan *Omnibus Law*. Sedangkan pada berita sentimen negatif dengan rakyat sebagai pihak yang diberitakan, topik yang paling banyak dibicarakan adalah kritikan kepada rakyat mengenai situasi demo *Omnibus Law*.



Gambar 7 Banyaknya artikel pada berita sentimen negatif berdasarkan topik yang dibicarakan

### 3.2 Support Vector Machine

Pada penelitian ini, dilakukan empat kali pemodelan klasifikasi menggunakan dua tipe pembagian data dan dua metode SVM yang berbeda. Pembagian data pertama adalah 90% data latih dan 10% data uji dan pembagian data kedua adalah 80% data latih dan 20% data uji. Selanjutnya metode SVM yang digunakan adalah metode *one vs one* dan metode *one vs all*. Berdasarkan hasil klasifikasi yang didapatkan dari empat pemodelan tersebut, diperoleh bahwa hasil klasifikasi pada model dengan penggunaan metode *one vs all* memiliki hasil klasifikasi terbaik. Pemodelan dengan metode *one vs all* memiliki nilai *macro average f1-score* di atas 50% dan nilai AUC pada masing-masing kelas sentimen di atas 0,70 (tingkat klasifikasi cukup baik), sedangkan pada penggunaan metode *one vs one* nilai AUC pada kelas sentimen positif bernilai kurang dari 0,70 (tingkat klasifikasi kurang baik).

### 3.3 Synthetic Minority Oversampling Technique (SMOTE)

Penanganan data tidak seimbang pada penelitian ini dilakukan dengan teknik Synthetic Minority Oversampling Technique (SMOTE). Metode SMOTE digunakan untuk melakukan oversampling pada kelas minoritas data latih. Kelas minoritas pada penelitian ini adalah kategori sentimen negatif dan positif. Data sintesis dibangkitkan pada kategori sentimen negatif dan positif sehingga banyaknya data pada tiga kelas sentimen menjadi sama banyak dengan proporsi akhir 33,33% pada masing-masing kelas sentimen.

### 3.4 Support Vector Machine dengan SMOTE

Pemodelan klasifikasi SVM yang dibentuk dengan data latih setelah dilakukan proses SMOTE dapat dilihat pada Tabel 8. Pemodelan klasifikasi dilakukan sebanyak empat kali dengan menggunakan dua tipe pembagian data dan dua metode SVM yang berbeda, sama seperti sebelum dilakukan SMOTE. Berdasarkan hasil klasifikasi yang didapatkan dari empat pemodelan tersebut, diperoleh bahwa penggunaan metode *one vs all* memiliki hasil klasifikasi terbaik. Pemodelan dengan metode *one vs all* memiliki nilai *macro average f1-score* di atas 50% dan nilai AUC pada masing-masing kelas sentimen di atas 0,70 (tingkat klasifikasi cukup baik), sedangkan pada penggunaan metode *one vs one* nilai AUC pada kelas sentimen positif bernilai kurang dari 0,70 (tingkat klasifikasi kurang baik). Pada kedua hasil klasifikasi sebelum dan sesudah SMOTE, metode *one vs all* memberikan hasil klasifikasi yang cenderung lebih baik daripada metode *one vs one*.

### 3.5 Perbandingan SVM One vs All Dengan dan Tanpa SMOTE

Perbandingan hasil klasifikasi SVM *one vs all* yang dibentuk dengan dan tanpa SMOTE dapat dilihat pada Tabel 1. Berdasarkan hasil klasifikasi pada Tabel 1, dapat dilihat bahwa klasifikasi dengan data hasil penanganan SMOTE menghasilkan hasil klasifikasi yang cenderung lebih baik jika dibandingkan dengan hasil klasifikasi pada

data tanpa penanganan SMOTE. Perbedaan nilai macro average f1-score pada hasil klasifikasi dengan dan tanpa SMOTE tidak terlalu signifikan karena ketidakseimbangan yang terdapat pada data tidak terlalu ekstrim.

Tabel 1 Hasil pemodelan klasifikasi SVM pada data hasil proses SMOTE

Data	Pembagian Data	Kelas	AUC	Akurasi	F1-Score	Macro Average
Tanpa SMOTE	90% training, 10% test	Negatif	0.82	79%	0.28	55%
		Netral	0.91		0.89	
		Positif	0.73		0.48	
	80% training, 20% test	Negatif	0.77	79%	0.22	52.33%
		Netral	0.89		0.89	
		Positif	0.71		0.46	
SMOTE	90% training, 10% test	Negatif	0.82	79%	0.40	60.33%
		Netral	0.90		0.88	
		Positif	0.73		0.53	
	80% training, 20% test	Negatif	0.76	77%	0.30	55%
		Netral	0.89		0.87	
		Positif	0.70		0.48	

Metode one vs all dengan penanganan SMOTE dan pembagian data 90% data latih 10% data uji memiliki hasil klasifikasi terbaik dengan nilai macro average f1-score sebesar 60,33% dan nilai akurasi 79%. Nilai AUC pada hasil klasifikasi tersebut termasuk ke dalam rentang tingkat klasifikasi cukup baik dengan nilai di atas 0,7 untuk masing-masing kategori sentimen.

## 4. Simpulan dan Saran

### 4.1 Simpulan

Situs CNNIndonesia.com cenderung memberitakan omnibus law secara netral pada setiap periode pemberitaan. Persentase sentimen negatif cenderung menurun pada periode saat pengesahan dan setelah pengesahan omnibus law. Proporsi berita negatif dan positif cenderung berimbang pada periode setelah pengesahan omnibus law, hal ini menandakan situs CNNIndonesia.com sudah cenderung bersikap netral



dalam pemberitaan terkait omnibus law pada periode tersebut. Pada sentimen positif, narasumber didominasi oleh pemerintah dengan persentase 55% dan topik yang paling banyak dibicarakan adalah terkait proses dan dampak ekonomi omnibus law, sedangkan pada sentimen negatif, narasumber didominasi oleh pengamat dengan persentase 29% dan topik yang paling banyak dibicarakan adalah terkait proses dan situasi demo omnibus law.

Pada penelitian ini, performa hasil klasifikasi diukur berdasarkan nilai AUC dan macro f1-score yang dihasilkan model. Berdasarkan hal tersebut, metode one vs all memiliki hasil klasifikasi yang lebih baik dibandingkan dengan metode one vs one pada klasifikasi sentimen pemberitaan omnibus law di situs CNNIndonesia.com. Penerapan SMOTE hanya memberikan hasil yang sedikit lebih baik daripada klasifikasi data tanpa penanganan SMOTE karena ketidakseimbangan yang terdapat pada data penelitian ini tidak terlalu ekstrim.

## 4.2 Saran

Pelabelan sentimen sebaiknya dilakukan oleh para ahli sehingga hasilnya bisa lebih akurat dan tidak bersifat subjektif. Selain itu, perlu dicoba berbagai kombinasi tahapan pre-processing untuk mendapatkan performa hasil klasifikasi yang lebih baik.

## Daftar Pustaka

- Akbari MIHAD, Novianty A, Setianingsih C. 2017. Analisis sentimen menggunakan metode *learning vector quantization*. e-Proceeding of Engineering. 4(2):2283–2292.
- Amelia OD, Soleh AM, Rahardiantoro S. 2018. Pemodelan *support vector machine* data tidak seimbang keberhasilan studi mahasiswa magister IPB. *Xplore: Journal of Statistics*. 2(1):33–40.
- Busroh FF. 2017. Konseptualisasi *omnibus law* dalam menyelesaikan permasalahan regulasi pertahanan. *Arena Hukum*. 10(2):227–250.
- Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. 2002. SMOTE : Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*. 16 (2002):321–357.
- Gorunescu F. 2011. *Data mining : concepts, models, and techniques*. India: Springer.
- Han H, Wang WY, Mao BH. 2005. *Borderline-SMOTE : a new over-sampling method in imbalance data sets learning*. Berlin (DE): Springer-Verlag.
- Joachims T. 1997. *Probabilistic analysis of the rocchio algorithm with TFIDF for text categorization*. Tennessee (US): Morgan Kaufmann Publishers.

- Juditha C. 2013. Akurasi berita dalam jurnalisme *online* (kasus dugaan korupsi mahkamah konstitusi di portal berita Detiknews). *Jurnal Pekommas*. 16(3):145–154.
- Klyueva I. 2019. Improving quality of the multiclass SVM classification based on the feature engineering. *Proceedings – 2019 1<sup>st</sup> International Conference on Control Systems, Mathematical Modelling, Automation and Energy Efficiency, SUMMA 2019*. IEEE. 491–494.
- McQuail D. 1992. *Media performance: mass communication and the public interest*. London(UK): Sage Publication.
- Pan S, Ni X, Sun J, Yang Q, Chen Z. 2010. Cross-domain sentiment classification via spectral feature alignment. *International World Wide Web Conference Committee*. 751–760.
- Rizam MM. 2019. Perspektif kompas *online* dalam judul berita *entertainment* : kajian linguistik kritis. *Jurnal Komposisi*. 4(2).
- Rumata VM. 2017. Objektivitas berita pada media dalam jaringan (analisis isi berita pemilihan gubernur DKI Jakarta pada Detiknews selama masa kampanye periode I). *Jurnal Penelitian Komunikasi dan Opini Publik*. 21(2).
- Schlkopf B, Smola AJ. 2001. *Learning with kernels : support vector machines, regularization, optimization, and beyond*. Cambridge : MIT Press.
- Suriadinata V. 2019. Penyusunan Undang-Undang di bidang investasi : kajian pembentukan *omnibus law* di Indonesia. *Refleksi Hukum : Jurnal Ilmu Hukum*. 4(1):115–132.
- Tan PN, Steinbach M, Kumar V. 2005. *Introduction to data mining*. New York(US): Addison Wesley.
- Vapnik V, Lerner A. 1963. Pattern recognition using generalized portrait method. *Automation and Remote Control*. 24:774–780.
- Vidya NA. 2015. *Twitter sentiment analysis terhadap brand reputation* : studi kasus PT XL Axiata Tbk [tesis]. Jakarta : Universitas Indonesia.
- Wang S, Yao X. 2013. Using class imbalance learning for software defect prediction. *IEEE Transactions on Reliability*. 62(2):434–443.
- Wardani AD, Indyarani H. 2018. Netralitas konten berita *online* (analisis *framing* : berita reuni alumni 212 di detik.com). *Interaksi : Jurnal Ilmu Komunikasi*. 7(1):1–7.