

ANALISIS PREDIKSI MAHASISWA MENGUNDURKAN DIRI DARI UNIVERSITAS XYZ DENGAN METODE *SUPPORT VECTOR MACHINE*

Brina Miftahurrohmah¹⁾, Catur Wulandari²⁾

^{1), 2)} Program Studi Sistem Informasi, Fakultas Teknologi Informasi dan Kreatif, Universitas Internasional Semen Indonesia

Kompleks PT. Semen Indonesia (Persero) Tbk.

Jl. Veteran, Gresik Jawa Timur, 61122

Email : brina.miftahurrohmah@uisi.ac.id¹⁾, catur.wulandari@uisi.ac.id²⁾

Abstrak

Fenomena mahasiswa mengundurkan diri bukan hal yang asing lagi bagi Perguruan Tinggi swasta maupun negeri. Alasan mahasiswa mengundurkan diri pun sangat beragam, diantaranya merasa salah masuk jurusan, tidak mampu membayar uang kuliah, orang tua yang pindah tempat kerja dan lain-lain. Bahkan ada pula mahasiswa yang mengundurkan diri tanpa alasan yang jelas. Kasus pengunduran diri tersebut dapat mempengaruhi citra jurusan bahkan universitas yang bersangkutan. Universitas XYZ pun tidak luput dari kejadian tersebut. Salah satu cara untuk mengatasi hal tersebut adalah dengan melakukan prediksi mahasiswa yang mengundurkan diri dari Universitas XYZ. Dengan demikian, mahasiswa yang terindikasi berpotensi mengundurkan diri akan diberi pengarahannya dan motivasi agar tetap melanjutkan kuliah. Prediksi dilakukan dengan metode *Support Vector Machine (SVM)* yang dianggap mampu melakukan analisis prediksi dengan baik. Implementasi metode *imbalance sampling* juga diterapkan untuk mengatasi kelas yang tidak seimbang antara mahasiswa yang mengundurkan diri dan tidak. Metode *imbalance sampling* tersebut adalah metode *undersampling*, *oversampling* dan gabungan dari kedua metode tersebut. Hasil analisis menunjukkan bahwa model yang dihasilkan dari implementasi metode *SVM* dengan kernel *Radial Basis Function (RBF)* dan *undersampling* mampu melakukan prediksi dengan akurasi 84,47%, sensitivitas 79,67%, spesifisitas 84,80% dan *AUC* 82,23%.

Kata kunci: Pengunduran diri, Pemodelan, *Support Vector Machine*, *Imbalance Sampling*

Abstract

The phenomenon of students resigning is a common thing in private and public universities. The reasons for students resigning are also very diverse, including feeling wrong entering the department, unable to pay tuition, parents who move to work and others. There are even students who resign for no apparent reason. The case of resignation can affect the image of the department and the university concerned. The XYZ University, which is a private university, did not escape the incident. One way to overcome this problem is to predict students who resign from that University. Thus, students who are indicated as having the potential to resign will be given direction and motivation to continue their studies. Prediction is done by the *Support Vector Machine (SVM)* method which is considered capable of performing predictive analysis well. Implementation of the *imbalance sampling* method is also applied to unbalanced classes between students who resign and not resign. The *imbalance sampling* method is the *undersampling* method, *oversampling* and a combination of the two methods. The results of the analysis show that the model resulting from the implementation of the *SVM* method with the kernel *Radial Basis Function (RBF)* and *undersampling* is able to predict with 84.47% accuracy, 79.67% sensitivity, 84.80% specificity and *AUC* 82.23%.

Keywords: Resign, Modelling, *Support Vector Machine*, *Imbalance Sampling*

1. PENDAHULUAN

Kasus mahasiswa mengundurkan diri dari suatu Perguruan Tinggi bukan hal asing lagi. Hampir semua perguruan tinggi baik negeri maupun swasta (PTN/ PTS) mengalami hal tersebut. Kejadian pengunduran diri mahasiswa bahkan dialami oleh PTN besar seperti Institut Teknologi Bandung (ITB). Hal itu dibuktikan oleh ulasan harian tempo yang menyatakan bahwa sedikitnya 35 calon mahasiswa baru 2017 melepaskan kursinya di PTN tersebut. Selain itu, Sistem Informasi Perguruan Tinggi (SIPT) Mikroskil menyatakan bahwa pengunduran diri mahasiswa termasuk *dropout* di PTS bahkan cukup banyak, seperti pada STMIK Mikroskil mencapai 88 orang rata-rata per tahun dalam periode tahun 2007 hingga 2010 [1]. Alasan pengunduran diri pun sangat beragam, mulai dari merasa salah masuk jurusan, tidak mampu membayar uang kuliah, orang tua

yang dipindah kerja dan lain-lain. Bahkan ada pula mahasiswa yang mengundurkan diri tanpa alasan yang jelas. Kejadian-kejadian pengunduran diri tersebut jelas dapat memberikan dampak negatif pada jurusan dan kampus yang bersangkutan. Salah satu dampak yang mungkin terjadi adalah berkurangnya kuantitas lulusan dan hal itu merupakan simbol kegagalan jurusan pada universitas tersebut. Kasus-kasus seperti itu mendorong peneliti untuk mengulas hal-hal yang berkaitan dengan pengunduran diri mahasiswa.

[2] dalam [1] telah melakukan penelitian mengenai faktor-faktor yang mempengaruhi berhentinya studi mahasiswa yang dilakukan terhadap seluruh mahasiswa jurusan Tadris Matematika yang terdaftar dari tahun 2008 sampai dengan tahun 2012, hasil penelitian menunjukkan jurusan pada sekolah menengah dan Indeks Prestasi pada semester pertama terbukti secara signifikan berpengaruh terhadap daya tahan kelanjutan studi mahasiswa. [3] juga menganalisis faktor-faktor yang mempengaruhi berhentinya mahasiswa studi dengan metode *data mining add-ins* dan menyimpulkan bahwa nilai NEM yang sangat rendah dan nilai STTB yang sedang mempengaruhi pengunduran diri mahasiswa. [2] kemudian melanjutkan kembali analisisnya dengan memasukkan variabel Kalkulus I dan menyimpulkan bahwa nilai pada mata kuliah Kalkulus I merupakan variabel mempunyai asosiasi paling kuat dengan daya tahan mahasiswa dalam melanjutkan studi [1].

Berdasarkan paparan yang telah dijelaskan, dalam penelitian ini akan dilakukan analisis untuk memprediksi potensi mahasiswa yang kemungkinan mengundurkan diri dari Universitas XYZ. Universitas XYZ digunakan sebagai obyek penelitian karena Universitas XYZ juga merupakan Perguruan Tinggi swasta yang tidak luput dari kasus pengunduran diri oleh mahasiswanya. Metode yang akan digunakan adalah *Support Vector Machine* (SVM) yang dianggap paling baru dan paling baik untuk melakukan prediksi.

2. DASAR TEORI

Support Vector Machine (SVM) merupakan metode salah satu metode *machine learning* yang masuk dalam kategori *supervised learning*. SVM terbukti memberi hasil yang lebih baik dibandingkan ANN dalam hal implementasi, terutama dalam hal solusi yang dicapai [4]. Dalam SVM, tujuan terakhir yang ingin dicapai adalah menemukan fungsi pemisah (klasifier) yang optimal yang *bias* memisahkan dua set data dari dua kelas yang berbeda [5]. Dalam fungsi pemisah yang perlu dicari adalah fungsi linier yang dapat didefinisikan

$$g(x) := \text{sgn}(f(x)), \tag{2}$$

dengan

$$f(x) := w^T x + b, \tag{3}$$

dimana $x, w \in R^n$ dan $b \in R$. Secara matematik, formulasi problem optimasi SVM untuk kasus klasifikasi linier adalah

$$\min \frac{1}{2} \|w\|^2 \tag{4}$$

dengan

$$y_i(wx_i + b) \geq 1, i = 1, 2, \dots, l \tag{5}$$

dimana adalah x_i data input, y_i adalah data output dan w, b merupakan parameter yang akan dicari. Namun menurut [6], jika beberapa data tidak dapat dikelompokkan dengan benar, formulasi matematikanya menjadi

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l t_i \tag{6}$$

dengan

$$y_i(wx_i + b) + t_i \geq 1, t_i \geq 0, i = 1, 2, \dots, l \tag{7}$$

dimana t_i adalah variabel *slack*.

Formulasi SVM yang telah dipaparkan, merupakan formulasi yang dikembangkan dengan asumsi kelinieran, sehingga tidak cocok untuk diterapkan pada kasus non linier. Metode

kernel merupakan metode yang sesuai untuk mengatasi kasus tersebut adapun fungsi kernel yang biasanya dipakai dalam literature SVM adalah linier: $x^T x$,

1. *polynomial*: $(x^T x_i + 1)^P$,
2. *radial basis function* (RBF): $\exp\left(\frac{-1}{2\sigma^2} \|x - x_i\|^2\right)$,
3. *tangent hyperbolic (sigmoid)*: $\tanh(\beta x^T x_i + \beta_1); \beta, \beta_1 \in R$ [4].

Cross validation merupakan satu dari banyak metode yang digunakan untuk mengevaluasi model prediksi [7]. Salah satu metode *cross validation* yang banyak digunakan adalah *k-folds cross validation*. Cara kerja metode ini adalah dengan membagi data menjadi *k* bagian yang sama banyak. Satu bagian dijadikan sebagai data *testing* sedangkan bagian lainnya dijadikan sebagai data *testing*.

Sebagian besar masalah klasifikasi dunia nyata memperlihatkan kondisi kelas yang tidak seimbang (*imbalanced sampling*) yaitu kondisi dimana jumlah data dari setiap kelas memiliki jumlah data yang sangat jauh berbeda. Hal tersebut akan menurunkan tingkat kebaikan kelas minoritas dalam melakukan klasifikasi. Dengan demikian perlu dilakukan analisis dengan metode yang dapat mengatasi kasus tersebut. Beberapa metode telah diperkenalkan untuk mengatasi kasus *imbalanced sampling*, diantaranya *undersampling*, *oversampling* dan gabungan kedua metode tersebut. *Undersampling* merupakan salah satu metode *imbalance sampling* yang bekerja dengan mengurangi kelas mayoritas sehingga didapatkan sampel yang sama di setiap kelas. Kerugian yang didapat dari metode ini adalah sebagian besar data yang didapat dari kelas mayoritas akan terbuang sehingga banyak informasi berharga yang akan terbuang juga. *Oversampling* merupakan salah satu metode yang paling umum digunakan dalam *deep learning* [8]. Berbanding terbalik dengan *undersampling*, metode ini bekerja dengan melakukan replikasi kelas minoritas sehingga didapatkan jumlah sampel sebanyak kelas minoritas. Metode ini terbukti efektif dalam menangani kasus data yang tidak seimbang, namun dapat menyebabkan *overfitting* [9]. Meskipun mungkin tidak tampak intuitif, menurut [10] dalam [11] ada beberapa bukti bahwa dalam beberapa situasi *undersampling* bisa lebih baik daripada *oversampling*. Metode *undersampling* dan *oversampling* juga dapat dilakukan secara bersamaan dengan cara melakukan *resampling* dengan menambah kelas minoritas dan mengurangi kelas mayoritas. Dengan demikian jumlah data yang digunakan dalam analisis tersebut tidak berubah [12].

Ketepatan hasil prediksi sangat diperlukan dalam melakukan analisis prediksi agar didapatkan parameter yang akurat untuk mendapatkan hasil prediksi yang sesuai. Beberapa kriteria yang umum digunakan untuk mengetahui tingkat ketepatan prediksi adalah akurasi, sensitivitas dan spesifisitas. Untuk mengukur akurasi, sensitivitas dan spesifisitas biasanya digunakan *confusion matrix*. *Confusion matrix* adalah alat ukur berbentuk matrix yang digunakan untuk mendapatkan jumlah ketepatan klasifikasi terhadap kelas dengan algoritma yang dipakai [13]. Berikut akan disajikan bentuk *confusion matrix* pada Tabel 1.

Tabel 1. *Confussion Matrix*

<i>Confussion Matrix</i>		Aktual	
		<i>True</i>	<i>False</i>
Prediksi	<i>True</i>	TP (<i>True Positive</i>) Correct result	FP (<i>False Positive</i>) Unexpected result
	<i>False</i>	FN (<i>False Negative</i>) Missing result	TN (<i>True Negative</i>) Correct absence of result

Tabel 1 nilai TP (*true positive*) dan TN (*true negative*) menunjukkan tingkat ketepatan klasifikasi. Umumnya semakin tinggi nilai TP dan TN semakin baik pula tingkat klasifikasi dari akurasi, sensitivitas dan spesifisitas. Jika label prediksi keluaran bernilai benar (*true*) dan nilai sebenarnya bernilai salah (*false*) disebut sebagai *false positive* (FP). Sedangkan jika prediksi label keluaran bernilai salah (*false*) dan nilai sebenarnya bernilai benar (*true*) maka hal ini disebut sebagai *false negative* (FN). Berikut formulasi untuk menghitung akurasi, sensitivitas, dan spesifisitas.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{8}$$

$$Sensitivitas = \frac{TP}{TP + FN} \times 100\% \tag{9}$$

$$Spesifisitas = \frac{TN}{TN + FP} \times 100\% \tag{10}$$

Akurasi juga dapat diukur oleh *Area Under Curva* (AUC) yaitu suatu kurva yang menggambarkan probabilitas dengan variabel sensitivitas dan spesifisitas. Area seluas 100% merepresentasikan pengujian yang sempurna sedangkan 50% dianggap sebagai pengujian yang tidak berguna [14]. Panduan untuk mengklasifikasikan keakuratan tes diagnostik adalah sebagai berikut.

90% - 100% = *Excellent Classification*

80% - 90% = *Good Classification*

70% - 80% = *Fair Classification*

60% - 70% = *Poor Classification*

50% - 60% = *Fail Classification*

3. METODOLOGI PENELITIAN

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diambil dari Direktorat Akademik dan Kemahasiswaan Universitas XYZ. Data yang diambil adalah data mengenai profil mahasiswa Universitas XYZ baik yang masih berstatus mahasiswa maupun yang tidak berstatus mahasiswa. Variabel yang digunakan dalam penelitian ini adalah status mahasiswa yang terdiri dari dua kategori dengan skala data nominal yaitu mengundurkan diri dan tidak mengundurkan diri yang bertindak sebagai variabel *output*, sedangkan variabel *input* terdiri dari pendapatan, IPK dan IPS terakhir yang masing-masing memiliki skala numerik. Analisis dimulai dengan tahap *preprocessing data*, kemudian data dibagi menjadi data *training* dan *testing* menggunakan metode validasi silang. Setelah data dibagi menjadi data *training* dan *testing*, data kemudian dianalisis menggunakan SVM Linear dan RBF. Setiap analisis SVM dilakukan dengan menerapkan metode *imbalance sampling* dan tanpa menerapkan metode tersebut. Pada akhirnya, hasil analisis dengan berbagai skenario tersebut dibandingkan dengan mengevaluasi berdasarkan akurasi, sensitivitas, spesifisitas dan AUC.

4. HASIL DAN PEMBAHASAN

Pada analisis ini, penentuan model dilakukan dengan dua kernel yaitu linier dan RBF. Dari kedua kernel tersebut akan dipilih model terbaik dimana kriteria model terbaik diukur dari tingkat akurasi, sensitivitas, spesifisitas dan AUC. Analisis dilakukan menggunakan dua skenario yaitu dengan tanpa menerapkan *imbalance sampling* dan dengan menerapkan *imbalance sampling*. Hasil analisis dengan tanpa menerapkan *imbalance sampling* ditunjukkan pada Tabel 2 dan Tabel 3.

Tabel 2. Perbandingan Hasil Analisis dengan SVM Linier tanpa *imbalance sampling*

Kriteria	Nilai (%)
Akurasi	93.61
Sensitivitas	0.00
Spesifisitas	100.00
AUC	50.00

Tabel 2 menunjukkan bahwa hasil prediksi memiliki akurasi yang tinggi yaitu sebesar 93.61%. Namun, hasil tersebut tidak dapat dikatakan baik karena nilai sensitivitas sangat kecil yaitu 0% dan spesifisitas sangat tinggi yaitu 100%. Hal itu menunjukkan bahwa model yang didapatkan cenderung mengklasifikasikan semua unit penelitian ke salah satu kelas saja, dalam

hal ini adalah kelas tidak mengundurkan diri. Selain itu, nilai AUC yang didapat juga sangat kecil yaitu 50% yang artinya gagal dalam melakukan pengklasifikasian.

Tabel 3. Perbandingan Hasil Analisis dengan SVM RBF tanpa *imbalance sampling*

C	Kriteria (%)			
	Akurasi	Sensitivitas	Spesifisitas	AUC
1	93.29	11.50	98.88	55.19
5	92.80	22.92	97.58	60.25
10	92.64	24.17	97.32	60.75
15	92.32	25.33	96.89	61.11

Analisis SVM RBF dilakukan dengan melakukan simulasi nilai C, yaitu C = 1, C = 5, C = 10, dan C = 15. Selain itu, parameter σ yang digunakan adalah 1. Hasil analisis menggunakan SVM RBF parameter $\sigma = 1$ dan simulasi beberapa nilai C dapat dilihat pada Tabel 3. Dari hasil analisis tersebut diketahui bahwa model terbaik adalah model dengan C = 5 dimana AUC yang didapatkan lebih besar dibandingkan AUC dari model yang menggunakan C yang lain sehingga keakuratan hasil analisis dapat digolongkan sebagai *poor classification*. Akurasi dan spesifisitas yang didapatkan juga sangat tinggi yaitu 92.32% dan 96.89%. Meskipun akurasi dan sensitivitas paling kecil dibandingkan dengan analisis menggunakan nilai C yang lain, namun model yang didapatkan lebih mampu memprediksi kelas minoritas. Setelah dilakukan analisis tanpa menerapkan *imbalance sampling*, selanjutnya analisis dilakukan dengan menerapkan *imbalance sampling*. Analisis tersebut dilakukan dengan tiga metode *imbalance sampling* yaitu *undersampling*, *oversampling* dan gabungan dari *oversampling* dan *undersampling* (selanjutnya disebut gabungan). Hasil analisis menggunakan SVM Linier dengan ketiga metode *imbalance sampling* tersebut ditunjukkan oleh Tabel 4 sedangkan SVM RBF ditunjukkan oleh Tabel 5.

Tabel 4. Perbandingan Hasil Analisis dengan SVM Linier dengan *imbalance sampling*

Metode <i>imbalance sampling</i>	Kriteria (%)			
	Akurasi	Sensitivitas	Spesifisitas	AUC
Undersampling	88.35	70.83	89.54	80.19
Oversampling	87.78	73.33	88.77	81.05
Gabungan	87.87	73.33	88.85	81.09

Tabel 4 menunjukkan bahwa akurasi tertinggi adalah SVM Linier *undersampling* yaitu sebesar 88.35% pengamatan diprediksi dengan tepat. Begitu pula dari sisi spesifisitas, *undersampling* mempunyai spesifisitas yang paling tinggi dibanding metode *imbalance sampling* yang lain. Namun dari sisi sensitivitas dan AUC metode gabungan memiliki nilai yang paling tinggi dibandingkan metode lain yaitu secara berturut – turut sebesar 73.33% dan 81.09%. Metode *oversampling* juga menghasilkan sensitivitas yang sama tinggi dengan gabungan akan tetapi untuk kriteria lain metode ini tidak menghasilkan nilai yang tinggi. Berdasarkan nilai AUC, metode gabungan dapat dinyatakan mempunyai kemampuan pengklasifikasian yang baik (*Good Classification*).

Tabel 5. Perbandingan Hasil Analisis dengan SVM RBF *imbalance sampling*

Metode <i>Imbalance Sampling</i>	C	Kriteria (%)			
		Akurasi	Sensitivitas	Spesifisitas	AUC
<i>Undersampling</i>	1	84.47	79.67	84.80	82.23
	5	83.58	77.17	84.02	80.59
	10	81.23	73.42	81.77	77.59
	15	79.13	74.67	79.43	77.05
<i>Oversampling</i>	1	88.76	73.25	89.80	81.53

	5	87.38	63.17	89.03	76.10
	10	86.81	54.42	89.03	71.72
	15	86.33	54.42	88.51	71.46
Gabungan	1	89.08	73.25	90.15	81.70
	5	87.38	68.17	88.68	78.42
	10	87.38	61.92	89.11	75.51
	15	88.03	59.42	89.98	74.70

Berdasarkan Tabel 5 dapat disimpulkan bahwa model terbaik didapat dari analisis menggunakan SVM RBF *undersampling*. Hal itu dapat dilihat dari nilai AUC sebesar 82.23% yang menyatakan bahwa model yang didapatkan merupakan model yang baik (*Good Classifier*). Selain itu, nilai AUC yang didapatkan merupakan AUC tertinggi dibandingkan SVM RBF dengan *oversampling* dan gabungan. Hal yang sama juga dapat dilihat dari sisi sensitivitas yang menunjukkan bahwa nilai sensitivitas terbaik adalah sensitivitas yang didapatkan dari SVM RBF *undersampling*. Meskipun dari segi akurasi dan spesifisitas paling kecil, akan tetapi rata-rata kemampuan prediksi *undersampling* cenderung tinggi. Hasil akhir yang diharapkan dari analisis ini adalah mendapatkan metode terbaik yang dapat menghasilkan model terbaik sehingga dapat dilakukan prediksi mahasiswa Universitas XYZ yang berpotensi mengundurkan diri dengan tepat. Dari hasil evaluasi menggunakan SVM Linier dan SVM RBF tanpa dan dengan *imbalance sampling* yang telah dilakukan maka didapatkan hasil seperti yang diperlihatkan oleh Tabel 6.

Tabel 6. Perbandingan Metode Analisis Prediksi

Metode <i>imbalance sampling</i>	Kernel	Kriteria (%)			
		Akurasi	Sensitivitas	Spesifisitas	AUC
SVM tanpa <i>imbalance sampling</i>	Linier	93.61	0	100	50
	RBF	92.32	25.33	96.89	61.11
SVM dengan <i>imbalance sampling</i>	Linier	87.87	73.33	88.85	81.09
	RBF	84.47	79.67	84.80	82.23

Hasil analisis pada Tabel 6 menunjukkan bahwa model yang terbaik didapatkan dari metode SVM RBF *imbalance sampling* dengan nilai AUC sebesar 82.23%. Metode *imbalance* tersebut adalah metode *undersampling* seperti yang ditunjukkan pada Tabel 5. Selain menghasilkan nilai AUC yang paling tinggi, nilai sensitivitas yang didapatkan juga paling tinggi yaitu sebesar 79.67%, yang artinya model yang dihasilkan dianggap paling mampu memprediksi kelas minoritas dibandingkan metode lain. Kelas mayoritas juga mampu dijelaskan oleh model tersebut sebesar 84.88%. Selain itu, Tabel 6 juga menunjukkan bahwa penggunaan metode *imbalance sampling* dapat meningkatkan kebaikan prediksi mencapai lebih dari 20%.

5. KESIMPULAN

Berdasarkan hasil analisis, dapat disimpulkan bahwa metode analisis terbaik untuk memperkirakan model prediksi siswa di Universitas XYZ yang mengundurkan diri adalah metode SVM dengan kernel RBF dengan menerapkan *undersampling*. Penerapan metode *sampling* tidak seimbang dapat meningkatkan akurasi prediksi hingga 20% hingga 30%.

Daftar Pustaka

- [1] L. Kurnia, "Analisis Daya Tahan Mahasiswa dalam Melanjutkan Studi dengan Chi-Squared Automatic Interaction Detection," *Journal of Sainstek*, vol. 14, no. 2, pp. 56-63, 2016.

- [2] L. Kurnia, *Pemodelan Faktor yang Mempengaruhi Mahasiswa Berhenti Studi pada Jurusan Tadris Matematika STAIN Batusangkar. (Penelitian tidak dipublikasikan).*, Batusangkar: P3N STAIN Batusangkar, 2013.
- [3] A. P. U. Sembiring and M. Ginting, "Analisis Faktor-Faktor Yang Mempengaruhi Pengunduran Diri Mahasiswa Dengan Aplikasi Data Mining Add-Ins – Studi Kasus Pada STMIK Mikroskil," / *JSM STMIK Mikroskil*, vol. 14, no. 2, pp. 139-146, 2013.
- [4] B. Santosa, *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*, Yogyakarta: Graha Ilmu, 2007.
- [5] V. Vapnik, *The Nature of Statistical Learning Theory*, Verlag: Springer, 1995.
- [6] S. Haykin, *Neural Network: A Comprehensive Foundation*, New Jersey: Prentice Hall, 1999.
- [7] F. Zhang, *Cross validation and Regression Analysis in Hight-Dimentional Sparse Linear Model. Dissertation*, United States: Creative Commons, 2011.
- [8] N. Jaccard, T. W. Rogers, E. J. Morton and L. D. Griffin, "Detection of Concealed Cars in Complex Cargo X-ray Imagery using Deep Learning," *Journal of X-Ray Science and Technology*, vol. 25, no. 3, pp. 1-17, 2016.
- [9] K.-J. Wang, B. Makond, K.-H. Chen and K.-M. Wang, "Hybrid Classifier Combining Smote With Pso To Estimate 5-Year Survivability Of Breast Cancer Patients," *Applied Soft Computing*, vol. 20, no. 1, pp. 15-24,, 2014.
- [10] C. Drummond and R. C. Holte, "C4.5, Class Imbalance, and Cost Sensitivity: Why Under-Sampling beats Over-Sampling," in *Workshop on Learning from Imbalanced Datasets II, ICML*, Washington DC, 2003.
- [11] M. Buda, A. Maki and M. Mazurowski, "A Systematics Study of Class Imbalance Problem In Convolutional Neural Network," *Neural Networks*, vol. 106, no. 1, pp. 249-259, Oktober 2018.
- [12] N. Lunardon, G. Menardi and N. Torelli, "ROSE: A Package For Binary Imbalanced Learning," *R Journal*, vol. 6, no. 1, pp. 82-92, 2014.
- [13] T. B. Sasongko, "Komparasi dan Analisis Kinerja Model Algoritma SVM dan PSO-SVM (Studi Kasus Klasifikasi Jalur Minat SMA)," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 2, no. 2, pp. 2443-2229, 2016.
- [14] F. Gorunescu, *Data Mining: Concepts, Models, and Techniques*, Verlag Berlin Heidelberg: Springer, 2011.