

# Klasifikasi Pelanggan Deposito Potensial menggunakan Ensemble Least Square Support Vector Machine

<sup>1</sup>Firman Aziz, <sup>2</sup>Jeffry

<sup>1,2</sup>Fakultas Matematika dan Ilmu Pengetahuan Alam

<sup>1,2</sup>Universitas Pancasakti

Makassar, Indonesia

<sup>1</sup>firman.aziz@unpacti.ac.id

<sup>2</sup>jeffry@unpacti.ac.id

## Abstrak

Abstrak Jumlah data yang sangat banyak pada industri perbankan sangat susah dianalisis secara manual untuk mendapatkan suatu informasi yang berguna agar dapat menentukan suatu kebijakan. Oleh karena itu, penggunaan data mining diharapkan dapat memberikan kontribusi dalam mengolah data tersebut. Berbagai metode telah banyak digunakan untuk mengklasifikasi suatu data, salah satunya adalah metode support vector machine. Penelitian ini bertujuan untuk melakukan klasifikasi terhadap nasabah yang berpotensi berlangganan deposito pada bank marketing dataset. Fokus penelitian ini mengusulkan pengembangan dari metode support vector machine yaitu metode least square support vector machine kemudian di ensemble menggunakan boosting. Data yang akan diolah adalah bank marketing dataset. Hasil menunjukkan bahwa metode yang diusulkan yakni ensemble least square support vector machine lebih baik dibandingkan dengan metode lainnya dengan persentase tingkat accuracy, sensitivity, specificity masing-masing adalah 95.15%, 92.93%, 97.61% dengan total rata-rata hasil klasifikasi sebesar 95.23%.

Kata Kunci: *SVM, LS-SVM, Ensemble, AdaBoost, Bank Marketing.*

## Abstract

*Abstract The vast amount of data in the banking industry is very difficult to analyze manually to obtain useful information in order to determine a policy. Therefore, the use of data mining is expected to contribute in processing the data. Various methods have been widely used to classify data, one of which is the support vector machine method. This study aims to classify customers who have the potential to subscribe to deposits in the bank marketing dataset. The focus of this study proposes the development of the support vector machine method, which is the least square support vector machine method, then ensemble using boosting. The data to be processed is the bank marketing dataset. The results show that the proposed method that is ensemble least square support vector machine is better than other methods with a percentage of accuracy level of 95.15%, sensitivity 92.93%, specialization of 97.61%. The average total classification results is 95.23%.*

Keyword: *SVM, LS-SVM, Ensemble, AdaBoost, Bank Marketing.*

## 1. Pendahuluan

Istilah Business Intelligence saat ini sangat familiar dikalangan perusahaan yang digunakan untuk menganalisis data bisnis agar menghasilkan suatu informasi yang terdiri dari strategi dan teknologi (Dedic & Stanier, 2016). Teknologi Business Intelligence memberikan suatu pandangan historis, terkini, dan prediksi kedepan suatu bisnis yang meliputi reporting, online analytical processing, analytics, data mining, process mining, complex event processing, business performance management, benchmarking, text mining, predictive analytics, dan prescriptive analytics. Teknologi Business Intelligence dapat menangani jumlah data skala besar untuk membantu mengidentifikasi, mengembangkan, dan menciptakan peluang bisnis strategis baru dengan tujuan memudahkan interpretasi data, mengidentifikasi peluang baru, dan menerapkan strategi yang efektif berdasarkan wawasan untuk memberikan keunggulan bisnis yang kompetitif dan memiliki stabilitas jangka Panjang (Rud, 2009).

Sektor perbankan mulai menerapkan teknologi Business Intelligence dengan mengolah informasi nasabah terutama pada bidang pemasaran. Secara umum pendekatan perbankan dalam memperkenalkan barang atau jasanya yaitu dengan iklan melalui televisi, radio, surat kabar, dan lain-lain atau dengan menargetkan nasabah secara khusus atau disebut dengan bank direct marketing (Moro et al., 2011). penerapan bank direct marketing dalam memperkenalkan suatu produk baik berupa barang maupun jasa kepada nasabah melalui telepon, email dan lain-lain dianggap sangat efektif dalam menggaet nasabah. Akan tetapi, banyak nasabah yang merasa terganggu sehingga menimbulkan penilaian negatif bagi bank serta biaya dan waktu yang dikeluarkan oleh telemarketer (Elsalamony & Elsayad, 2013), (Elsalamony, 2014), (Vajiramedhin & Suebsing, 2014). mengatasi permasalahan tersebut maka perlu dilakukan pengolahan data dengan melakukan klasifikasi untuk menemukan suatu informasi yang berfungsi untuk membantu menemukan nasabah potensial (Abbas, 2015). Dalam klasifikasi terdapat proses menganalisis satu set data dan menghasilkan seperangkat aturan pengelompokan yang dapat digunakan untuk mendapatkan informasi masa depan.

Penelitian (Zhuang et al., 2018) mengadopsi teknik data mining melalui SPSS Modeler untuk memprediksi perilaku langganan deposito berjangka pelanggan dan memahami fitur-fitur pelanggan untuk meningkatkan efektivitas dan keakuratan pemasaran bank. Penelitian ini terbatas hanya pada penarikan kesimpulan tanpa memiliki pengukuran kinerja dalam teknik data mining. Kemudian, penelitian (Parlar, 2017) menganalisis dua metode pemilihan fitur yaitu information gain dan Chi-square pada dataset bank direct marketing untuk mengetahui fitur-fitur yang berpengaruh. Hasil menunjukkan metode pemilihan fitur Information Gain dan Chi-square sangat dekat, meskipun keduanya berbeda untuk lima fitur teratas. Penelitian (Ruangthong & Jaiyen, 2015) mengusulkan metode untuk menganalisis informasi asimetris menggunakan algoritma SMOTE dan Rotation Forest (PCA) -J48 untuk menyelesaikan klasifikasi set data yang tidak seimbang. dimana algoritma SMOTE digunakan untuk memodifikasi data dan meningkatkan akurasi prediksi. Hasil menunjukkan algoritma SMOTE secara efektif menyelesaikan ketidaksetaraan data dan Rotation Forest (PCA) -J48 dapat mencapai nilai akurasi dan spesifisitas tertinggi. Namun, Rotation Forest (PCA) -J48 memiliki sensitivitas yang tinggi. Penelitian (Grzonka et al., 2016) membandingkan beberapa metode seperti decision trees, bagging, boosting, and random forests untuk melakukan klasifikasi data direct marketing campaigns. Hasil menunjukkan bahwa faktor yang sangat penting untuk hasil yang diperoleh adalah keacakan sampel bootstrap yang

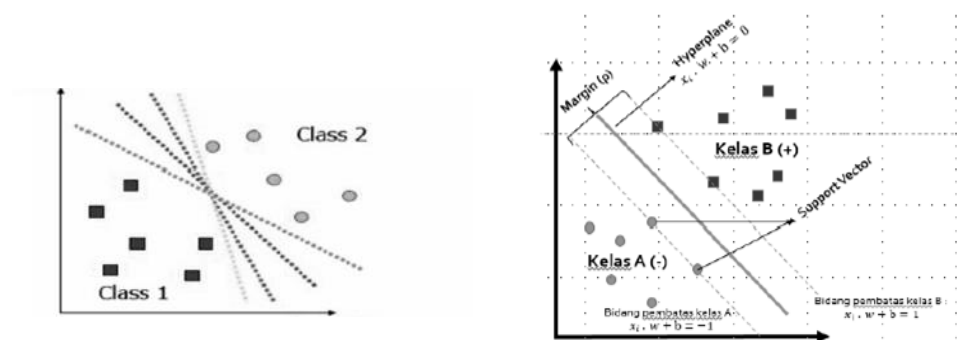
digunakan untuk membangun model. Kemudian, Penelitian (Lawi et al., 2017) mengusulkan metode Support Mesin Vektor menggunakan algoritma Adaboost untuk mengklasifikasikan pelanggan potensial untuk pemasaran langsung. Hasil menunjukkan bahwa metode Support Mesin Vektor menggunakan algoritma Adaboost menghasilkan akurasi 95,07% dan sensitivitasnya 91,65% lebih tinggi dari pendekatan SVM biasa. Penelitian ini berhasil mengatasi permasalahan keacakan sample bootstrap Tetapi penelitian ini masih terbatas pada satu partisi data menggunakan data tidak seimbang dan menurut (Zhou et al., 2010) meskipun support vector machine tunggal menunjukkan kinerja yang baik dalam klasifikasi, akan tetapi sensitif terhadap pengaturan sampel dan parameter. Oleh karena itu untuk mengatasi masalah tersebut digunakan Least Square Support Vector Machine (LSSVM) yaitu dengan menyelesaikan fungsi kendala berbentuk persamaan linear untuk mendapatkan solusi dari permasalahan quadratic programming.

Penelitian ini mengusulkan metode Ensembel Least Square Support Vector Machine menggunakan Algoritma Boosting dimana algoritma boosting yang digunakan adalah Adaboost. Penelitian ini juga akan menyelidiki diberbagai partisi data untuk melihat kinerja dari metode yang diusulkan.

## 2. Metodologi

### 2.1. Support Vector Machine

Support Vector Machine digunakan untuk memecahkan masalah klasifikasi pola dan sangat cocok digunakan pada data yang dapat dipisahkan secara linear dengan memaksimalkan batas bidang pemisah (Hyperplane) untuk memisahkan data ke dalam dua class pada sebuah ruang fitur (Borges, 1998).



Gambar 1. Konsep Support Vector Machine.

Gambar 1 menunjukkan pilihan hyperplane yang mungkin untuk set data dan bidang pemisah terbaik dengan margin maksimal yang tertelak di antara kedua kelas.

### 2.2. Least Square Support Vector Machine

Least Squares Support Vector Machine merupakan pengembangan SVM yang dipelopori oleh Suykens dan Vandewalle pada tahun 1999 (Suykens & Vandewalle, 1999). Least Squares Support Vector Machine menghasilkan tingkat error yang lebih rendah dibandingkan dengan Support Vector Machine karena melakukan klasifikasi lanjutan yaitu dengan mengolah kembali data yang salah terklasifikasi menggunakan

quadratic hyperplane. Bentuk Least Squares Support Vector Machine ditunjukkan dengan fungsi objektif berikut :

$$\min \frac{1}{2} \|\mathbf{w}\|^2 + \frac{C}{2} \sum_{i=1}^n \xi_i^2 \quad (1)$$

Dengan kendala

$$y_i(\varphi(\mathbf{x}_i) \cdot \mathbf{w}^T + b) = 1 - \xi_i, k = 1, 2, \dots, N \quad (2)$$

Dimana  $\xi$  adalah variabel slack yang menentukan tingkat kesalahan klasifikasi (misclassification) sampel data.  $C > 0$  adalah parameter yang menentukan besar penalti akibat kesalahan dalam klasifikasi data.  $K$  adalah fungsi kernel yang digunakan.

### 2.3. Adaptive Boosting (AdaBoost)

Secara umum algoritma AdaBoost melatih pengklasifikasian dasar secara sekuensial dalam setiap iterasi menggunakan data latih dengan koefisien bobot yang bergantung dari performa pengklasifikasian pada iterasi sebelumnya untuk memberikan bobot yang lebih besar pada data yang salah terklasifikasi (Schapire & Freund, 2013), (Schwenker, 2013).

langkah-langkah algoritma Adaboost adalah (Freund & Schapire, 1999):

1. Input: Data pelatihan beserta labelnya  $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)\}$ , Component Learn, dan jumlah perputaran  $T$
2. Inisialisasi bobot data pelatihan:

$$D_i^1 = \frac{1}{N}, i = 1, \dots, N \quad (3)$$

3. Untuk iterasi  $t=1, \dots, T$ 
  - a. Gunakan algoritma Component Learner untuk melatih suatu komponen klasifikasi,  $h_t$ , pada bobot pelatihan.
  - b. Hitung bobot kesalahan klasifikasi pada  $h_t$

$$\epsilon_t = \sum_{i=1}^N (D_i^t (y_i \neq h_t(x))) \quad (4)$$

Indeks kepercayaan pembelajaran dihitung sebagai:

$$c_t = \frac{1}{2} \ln \left( \frac{(1-\epsilon_t)}{\epsilon_t} \right) \quad (5)$$

- c. Perbaharui bobot sampel pelatihan

$$D_i^{t+1} = \frac{D_i^t \exp(c_t \times y_i \neq y_i^*)}{\sum_{i=1}^N D_i^t} \quad (6)$$

- d. Testing model dengan data uji
4. Keluaran pembelajaran terakhir  
Kombinasi semua klasifikasi

$$h_j = \text{sign}(\sum_{t=1}^T c_t h_t(x)) \quad (7)$$

## 2.4. Evaluasi Kinerja

Evaluasi hasil Kinerja dari setiap klasifikasi di hitung berdasarkan 3 pengukuran yaitu : Accuracy, Sensitivity, dan Specificity berdasarkan nilai True Positive, False Negative, False Positive, dan True Negative.

- True Positive (TP) : Jumlah data yang teridentifikasi sebagai 'Pelanggan Potensial' dan hasil prediksinya 'Pelanggan Potensial'.
- True Negative (TN) : Jumlah data yang teridentifikasi sebagai 'Pelanggan Non-Potensial' dan hasil prediksinya 'Pelanggan Non-Potensial'.
- False Positive (FN) : Jumlah data yang teridentifikasi sebagai 'Pelanggan Non-Potensial' tetapi hasil prediksinya 'Pelanggan Potensial'.
- False Negative (FN) : Jumlah data data yang teridentifikasi sebagai 'Pelanggan Potensial' tetapi hasil prediksinya 'Pelanggan Non-Potensial'.

Tabel 1. Confusion Matrix.

Prediksi	Pelanggan Potensial	Pelanggan Non-Potensial
Pelanggan Potensial	True Positive (TP)	False Negative (FN)
Pelanggan Non-Potensial	False Positive (FP)	True Negative (TN)

$$Akurasi = \frac{TP+TN}{TN+FP+FN+TP} \quad (8)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (9)$$

$$Specificity = \frac{TN}{TN+FP} \quad (10)$$

## 3. Eksperimental

### 3.1. Dataset

Data yang diolah adalah data bank direct marketing dataset yang dapat diakses melalui University of California at Irvine (UCI) Machine Learning Repository. Data tersebut berasal dari bank Portugal, dari bulan mei 2008 sampai bulan juni 2013. deskripsi dataset ditunjukkan pada Tabel 2.

Tabel 2. Deskripsi Dataset.

Atribut	Tipe data	Keterangan
Age	Numeric	Umur Pelanggan

<b>Atribut</b>	<b>Tipe data</b>	<b>Keterangan</b>
Job	Categorical :	
	1. Admin	
	2. Blue-collar	
	3. Entrepreneur	
	4. Housemaid	
	5. Management	
	6. Retired	
	7. Self-employed	
	8. Services	
	9. Student	
	10. Technician	
	11. Unemployed	
12. Unknown		Tipe Pekerjaan
Marital	Categorical :	
	1. Divorced	
	2. Married	
	3. Single	
4. Unknown		Status Pernikahan
Education	Categorical :	
	1. Basic.4y	
	2. Basic.6y	
	3. Basic.9y	
	4. High.school	
	5. Illiterate	
	6. Professional.course	
	7. University.degree	
8. Unknown		Tingkat Pendidikan
Default	Categorical :	
	1. No	
	2. Yes	
3. Unknown		Apakah memiliki kredit ?
Housing	Categorical :	
	1. No	
	2. Yes	
3. Unknown		Apakah memiliki pinjaman perumahan ?
Loan	Categorical :	
	1. No	
	2. Yes	
3. Unknown		Apakah memiliki pinjaman pribadi ?
Contact	Categorical :	
	1. Celular	
2. Telephone		Dihubungi melalui ?

Atribut	Tipe data	Keterangan
Month	Categorical :	
	1. Jan	
	2. Feb	
	3. Mar	
	4. Apr	
	5. May	
	6. Jun	
	7. Jul	
	8. Aug	
	9. Sep	
	10. Oct	
	11. Nov	
	12. Dec	
Day_of_week	Categorical :	
	1. Mon	
	2. Tue	
	3. Wed	
	4. Thu	
	5. Fri	
Duration	Numeric	lama durasi selama dihubungi.
Campaign	Numeric	Jumlah kontak yang dilakukan selama pengiklanan untuk nasabah termasuk kontak terakhir
Pdays	Numeric	Selisih hari setelah nasabah terakhir kali dikontak untuk pengiklanan sebelumnya
Previous	Numeric	Jumlah kontak yang dilakukan sebelum pengiklanan ini dan untuk nasabah ini
Poutcome	Categorical :	
	1. Failure	
	2. Nonexistent	
	3. Success	
y	Binary :	
	1. Yes	
	2. No	

### 3.2. Normalisasi Data

Dari total jumlah data keseluruhan sebanyak 41.188 pelanggan, dinormalisasi menjadi 9.280 untuk menyeimbangkan anatar pelanggan deposito berjangka dan bukan pelanggan deposito berjangka. Jadi data yang diolah sebanyak 4,640 data pelanggan deposito berjangka dan 4,640 data bukan pelanggan deposito berjangka.

### 3.3. Implementasi

Fokus dari penelitian ini untuk menerapkan metode ensemble least squares support vector machine menggunakan AdaBoost. Tahapan metode yang diusulkan adalah sebagai berikut :

- a. Load Dataset
- b. Identifikasi Label Atribut, class dan jumlah data.
- c. Menentukan jumlah data latih dan data uji.
- d. Membentuk klasifikasi
  - Menentukan Nilai Gamma.
  - Menentukan Nilai Sigma.
  - Inisialisasi Variabel, dan bobot awal.
- e. Menentukan jumlah perulangan (Iterasi)
  - Menggunakan algoritma component learner (least squares support vector machine) untuk membentuk suatu model klasifikasi.
  - Menghitung bobot kesalahan klasifikasi
  - Memperbaharui bobot sampel data latih
  - Pengujian model dengan data uji
  - Hasil dari Pengujian model.

## 4. Hasil

Proses pengujian menggunakan software Matlab untuk melihat kinerja dari metode-metode yang diusulkan. Berdasarkan hasil confusion matrix seperti yang terlihat pada Tabel 3, hasil kinerja dari masing-masing metode ditampilkan pada Tabel 4. Hasil dari total rata-rata yang didapatkan metode ensemble least square support vector machine menggunakan adaboost mendapatkan persentase yang paling tinggi diantara metode support vector machine dan least square support vector machine. Hal ini membuktikan bahwa metode ensemble least square support vector machine menggunakan adaboost lebih baik dibandingkan dengan metode lainnya. Kinerja terbaik dihasilkan ketika partisi data 80:20 yakni 80% data latih dan 20% data uji.

Tabel 3. Hasil confusion matrix.

Metode	Data Latih / Data Uji %	TP	TN	FP	FN
Support Vector Macine	10 / 90	4149	1943	27	2233
	20 / 80	3670	2272	42	1440
	30 / 70	3155	2054	93	1194
	40 / 60	2690	1622	94	1162
	50 / 50	2217	1585	103	735
	60 / 40	1771	1385	85	471
	70 / 30	1271	1198	121	194
	80 / 20	908	841	20	87
	90 / 10	445	434	19	30



Metode	Data Latih / Data Uji %	TP	TN	FP	FN
Least Square Support Vector Machine	10 / 90	3924	3365	252	811
	20 / 80	3556	3029	156	683
	30 / 70	3069	2742	179	506
	40 / 60	2642	2073	142	711
	50 / 50	2152	1970	168	350
	60 / 40	1756	1645	100	211
	70 / 30	1260	1291	132	101
	80 / 20	906	852	22	76
	90 / 10	442	433	22	31
Ensemble Least Square Support Vector Machine Menggunakan Adaboost	10 / 90	3910	3389	266	787
	20 / 80	3543	3051	169	661
	30 / 70	3056	2757	192	491
	40 / 60	2621	2125	163	659
	50 / 50	2124	2033	196	287
	60 / 40	1743	1671	113	185
	70 / 30	1260	1291	132	101
	80 / 20	907	859	21	69
	90 / 10	443	434	21	30

Tabel 4. Hasil Klasifikasi.

Metode	Data Latih / Data Uji %	Accuracy %	Sensitivity %	Specificity %	Rata-Rata %
Support Vector Macine	10 / 90	72,94	65,01	98,63	78,86
	20 / 80	80,04	71,82	98,18	83,35
	30 / 70	80,19	72,55	95,67	82,80
	40 / 60	77,44	69,83	94,52	80,60
	50 / 50	81,94	75,10	93,90	83,65
	60 / 40	85,02	78,99	94,22	86,08
	70 / 30	88,69	86,76	90,83	88,76
	80 / 20	94,23	91,26	97,68	94,39
	90 / 10	94,72	93,68	95,81	94,74
Least Square Support Vector Machine	10 / 90	87,27	82,87	93,03	87,73
	20 / 80	88,70	83,89	95,10	89,23
	30 / 70	89,46	85,85	93,87	89,72
	40 / 60	84,68	78,80	93,59	85,69
	50 / 50	88,84	86,01	92,14	89,00
	60 / 40	91,62	89,27	94,27	91,72
	70 / 30	91,63	92,58	90,72	91,64
	80 / 20	94,72	92,26	97,48	94,82
	90 / 10	94,29	93,45	95,16	94,30
Ensemble Least Square Support Vector Machine Menggunakan Adaboost	10 / 90	87,39	83,24	92,72	87,79
	20 / 80	88,82	84,28	94,75	89,28
	30 / 70	89,49	86,16	93,49	89,71
	40 / 60	85,24	79,91	92,88	86,01
	50 / 50	89,59	88,10	91,21	89,63

Metode	Data Latih / Data Uji %	Accuracy %	Sensitivity %	Specificity %	Rata-Rata %
	60 / 40	91,97	90,40	93,67	92,01
	70 / 30	91,63	92,58	90,72	91,64
	80 / 20	95,15	92,93	97,61	95,23
	90 / 10	94,50	93,66	95,38	94,52

## 5. Kesimpulan

Dalam penelitian ini diusulkan metode ensemble least squares support vector machine menggunakan AdaBoost untuk mengklasifikasikan pelanggan deposito potensial. Fokus dari penelitian ini adalah meningkatkan kinerja support vector machine yang sensitive terhadap pengaturan sampel dan parameter dengan mengusulkan metode least squares support vector machine. Tetapi, metode least squares support vector machine memiliki masalah pada keacakan sample maka dilakukan pengembangan dengan melakukan teknik ensemble.

Dengan membandingkan hasil kinerja dari metode SVM, LSSVM, dan Ensemble LSSVM menggunakan Adaboost didapatkan hasil bahwa metode yang diusulkan yakni Ensemble LSSVM menggunakan AdaBoost berhasil meningkatkan kinerja dari metode SVM dan LSSVM dengan persentase tingkat accuracy, sensitivity, specificity masing-masing adalah 95.15%, 92.93%, 97.61% dengan total rata-rata hasil klasifikasi sebesar 95.23%. dapat disimpulkan bahwa metode Ensemble LSSVM berhasil mengatasi kelemahan dari metode SVM dan LSSVM. Untuk kedepannya akan diteliti menggunakan data seimbang.

## Daftar Pustaka

- Dedic, N., & Stanier, C. (2016, December). Measuring the success of changes to existing business intelligence solutions to improve business intelligence reporting. In *International Conference on Research and Practical Issues of Enterprise Information Systems* (pp. 225-236). Springer, Cham.
- Rud, O. P. (2009). *Business intelligence success factors: tools for aligning your business in the global economy* (Vol. 18). John Wiley & Sons.
- Moro, S., Laureano, R., & Cortez, P. (2011). Using data mining for bank direct marketing: An application of the crisp-dm methodology. In *Proceedings of European Simulation and Modelling Conference-ESM'2011* (pp. 117-121). EUROSIS-ETI.
- Elsalamony, H. A., & Elsayad, A. M. (2013). Bank direct marketing based on neural network and C5.0 Models. *Int. J. Eng. Adv. Technol. IJEAT*, 2(6).
- Elsalamony, H. A. (2014). Bank direct marketing analysis of data mining techniques. *International Journal of Computer Applications*, 85(7), 12-22.
- Vajiramedhin, C., & Suebsing, A. (2014). Feature selection with data balancing for prediction of bank telemarketing. *Applied Mathematical Sciences*, 8(114), 5667-5672.
- Abbas, S. (2015). Deposit subscribe prediction using data mining techniques based Real marketing dataset. *arXiv preprint arXiv:1503.04344*.

- Zhuang, Q. R., Yao, Y. W., & Liu, O. (2018). Application of data mining in term deposit marketing. In *Proceedings of the International MultiConference of Engineers and Computer Scientists* (Vol. 2).
- Parlar, T. (2017). Using data mining techniques for detecting the important features of the bank direct marketing data. *International journal of economics and financial issues*, 7(2), 692.
- Ruangthong, P., & Jaiyen, S. (2015, July). Bank direct marketing analysis of asymmetric information based on machine learning. In *2015 12th International Joint Conference on Computer Science and Software Engineering (JCSSE)* (pp. 93-96). IEEE.
- Grzonka, D., Suchacka, G., & Borowik, B. (2016). Application of selected supervised classification methods to bank marketing campaign. *Information Systems in Management*, 5(1), 36-48.
- Lawi, A., Velayaty, A. A., & Zainuddin, Z. (2017, August). On identifying potential direct marketing consumers using adaptive boosted support vector machine. In *2017 4th International Conference on Computer Applications and Information Processing Technology (CAIPT)* (pp. 1-4). IEEE.
- Zhou, L., Lai, K. K., & Yu, L. (2010). Least squares support vector machines ensemble models for credit scoring. *Expert Systems with Applications*, 37(1), 127-133.
- Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2), 121-167.
- Suykens, J. A., & Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural processing letters*, 9(3), 293-300.
- Schapire, R. E., & Freund, Y. (2013). Boosting: Foundations and algorithms. *Kybernetes*.
- Schwenker, F. (2013). Ensemble methods: Foundations and algorithms [book review]. *IEEE Computational Intelligence Magazine*, 8(1), 77-79.