

# PREDIKSI MINAT KLIEN PADA PRODUK DEPOSITO MENGUNAKAN ALGORITMA SVM PARAMETER KERNEL POLYNOMIAL

Dani Saputra<sup>1</sup>, Maxsi Ary<sup>2</sup>

<sup>1</sup> Universitas Adhirajasa Reswara Sanjaya  
Jl. Sekolah Internasional No. 1-2 Antapani, Bandung, 022-7100124  
e-mail: [dantra1207@gmail.com](mailto:dantra1207@gmail.com)

<sup>2</sup> Universitas Adhirajasa Reswara Sanjaya  
Jl. Sekolah Internasional No. 1-2 Antapani, Bandung, 022-7100124  
e-mail: [maxsi@ars.ac.id](mailto:maxsi@ars.ac.id)

## Abstrak

Departement Telemarketing Bank dituntut mampu membuat target klien, karena kinerja telemarketing bank dinilai dari target klien yang berminat pada produk bank yang ditawarkan untuk mendukung telemarketing bank meningkatkan keberhasilannya dalam promosi, maka dilakukan prediksi untuk menentukan keputusan klien berminat atau tidak pada produk deposito, akan tetapi kepada nasabah atau klien yang memiliki potensial untuk menerima penawaran dari telemarketing bank, yang dimana dapat meminimalisir kerugian biaya dalam promosi kepada klien. Maka dapat dilakukan penerapan dan pemanfaatan teknik data mining guna memprediksi minat klien menggunakan *Support Vector Machine* yang merupakan salah satu teknik yang cocok untuk melakukan prediksi. Dengan menerapkan *K-Fold Cross Validation* untuk mengevaluasi kinerja model atau algoritma dimana data dipisahkan mejadi dua subset data proses pembelajaran dan data validasi. Setelah penelitian selesai dilakukan, maka muncul nilai akurasi dari metode *Support Vector Machine K-Fold Cross Validation* sebesar 94,91% dengan nilai AUC 0.979 dengan menggunakan *10-fold K Validation*. Kemudian dilakukan kembali perhitungan lebih dengan parameter *Kernel Polynomial* mendapatkan hasil akurasi lebih besar dengan nilai akurasi sebesar 95,18% dengan nilai AUC 0,974.

**Kata Kunci:** Prediksi, Bank Marketing Data Set, *Support Vector Machine*, *Kernel Polynomial*, Klasifikasi

## Abstract

*The Telemarketing Department of the Bank is demanded to be able to make the target client, because the performance of the bank telemarketing is assessed from the target of clients who are interested in the bank products offered to support the telemarketing bank to increase its success in promotion, then predictions are made to determine whether or not the client's decision is interested in deposit products. customers or clients who have the potential to receive offers from telemarketing banks, which can minimize cost losses in promotions to clients. So the application and utilization of data mining techniques can be done to predict client interests using Support Vector Machine which is one technique that is suitable for making predictions. By applying K-Fold Cross Validation to evaluate the performance of a model or algorithm where the data is separated into two subsets of learning process data and validation data. After the research is completed, the accuracy of the Support Vector Machine K-Fold Cross Validation method is 94,91% with AUC value of 0.979 using 10-fold K Validation. Then do more calculations with Polynomial Kernel parameters to get greater accuracy with an accuracy value of 95,18% with an AUC value of 0,974.*

**Keywords:** Prediction, Bank Marketing Data Set, *Support Vector Machine*, *Polynomial Kernel*, Classification

## 1. Pendahuluan

Percepatan transaksi keuangan pada zaman *modern* sangatlah diperlukan oleh setiap orang, sebab dengan kemudahan dalam bertransaksi maka semua hal tentang kebutuhan pun akan lebih mudah dilakukan. Oleh karena itu teknologi menjadi faktor terpenting dalam percepatan transaksi, terutama pada bidang teknologi komunikasi yang akan membawa konsekuensi pada manusia untuk selalu meningkatkan kualitas hidupnya. Seiring berkembangnya teknologi, kebutuhan tenaga kerja pada dibidang industri komunikasi akan semakin dibutuhkan, seperti dalam bidang kerja telemarketing yang akan mengalami perkembangan seiring dengan berkembangnya teknologi (Ghani et al., 2019).

Kemajuan terhadap teknologi sangatlah mempengaruhi pemasaran *domain* dengan permintaan yang sangat *spesifik*, jadi banyak perusahaan yang lebih memilih kampanye pemasaran target dengan media telekomunikasi dari pada kampanye pemasaran massal secara langsung dilapangan. Untuk melakukan kampanye pemasaran target ialah dengan mempelajari *subdomain* dan penambangan data, yang dapat dipelajari dari teknik otomatis untuk pembelajaran membuat prediksi yang akurat dan spesifik berdasarkan pengamatan masa lalu. Pembelajaran mesin menggunakan dua jenis teknik: pembelajaran terbimbing (klasifikasi dan regresi), yang melatih model input dan output yang telah diketahui datanya sehingga dapat memprediksi keluaran di masa depan, dan selanjutnya ialah tanpa pengawasan *learning (clustering)*, yang menemukan pola-pola yang tersembunyi atau *intrinsik* terstruktur dalam inputan data (Tekouabou et al., 2019).

Dengan memanfaatkan teknologi yang sudah canggih, bank dapat menggunakan alat komunikasi jarak jauh seperti telepon atau media elektronik lainnya. Pihak bank biasanya melakukan strategi untuk mempromosikan produk dan layanan jasanya kepada klien, dari beberapa produk yang diamati ialah deposito. Deposito adalah produk investasi dengan menyimpan uang dengan kurun waktu tertentu untuk penarikannya yang telah dijanjikan oleh pihak bank dengan persetujuan nasabah (WK & Adani, 2018). Bank pun memiliki data yang cukup sangat

besar kapasitasnya disimpan sebagai *database* dan diolah kembali untuk menghasilkan informasi yang berkaitan tentang nasabah, data tersebut pula dapat digunakan untuk keperluan menjaga hubungan bank dengan nasabah, maka salah satu caranya ialah dengan ketentuan individual dalam target pemasaran produk bank (Aqham & Hartomo, 2019). Untuk mendapatkan kelayakan dalam suatu kehidupan banyak orang maka perlu bantuan dari beberapa masyarakat yang memiliki dana berlebih dan menyumbangkan dana tersebut kepada masyarakat lain dalam bentuk kredit, biasanya badan usaha terbesar yang menghimpun dana simpanan dan memberikan bantuan dana kepada yang membutuhkannya disebut dengan bank (Sulaehani, 2016).

Pada masa sekarang, beberapa perusahaan besar seperti operator jaringan seluler biasanya menggunakan cara melalui panggilan untuk menawarkan sebuah produk atau layanannya, karena sifatnya memang secara pribadi langsung kepada konsumen, dan biasanya cara ini cenderung berhasil karena dapat dengan mudah diterima oleh klien, serta telemarketing sendiri merupakan sebuah cara baru dalam bidang pemasaran yang menggunakan teknologi komunikasi sebagai bagian dari sarana pemasaran yang teratur dan terstruktur (Saputra, 2017). Pemesanan jarak jauh diidentikan sebagai pengguna telepon atau sering disebut pusat panggilan untuk menarik prospek klien, mempromosikan kepada pelanggan tentang apa yang dimiliki oleh pihak bank, serta menjawab pertanyaan dan mengambil pesanan klien melalui alat telekomunikasi. *Departement telemarketing bank* dituntut mampu membuat target klien, karena kinerja telemarketing bank dinilai dari target klien yang berminat pada produk bank yang ditawarkan. Untuk mendukung telemarketing bank meningkatkan keberhasilannya dalam promosi, maka dilakukan prediksi untuk menentukan keputusan klien berminat atau tidak pada produk deposito, yang dimana dapat meminimalisir kerugian biaya dalam panggilan promosi kepada klien. Dalam suatu proses penawaran produk deposito, bank tidak melakukan penawarannya ke semua nasabah atau kliennya, akan tetapi kepada nasabah atau klien yang memiliki potensial untuk menerima penawaran dari

telemarketing bank. Untuk menentukan potensial terhadap nasabah atau klien yang akan di tawarkan, maka bank memerlukan sebuah system perhitungan agar tepat sasaran. Salah satu yang dapat diterapkan adalah dengan menerapkan metode Data Mining (Agustina, 2019).

Penulis melakukan perhitungan dengan algoritma *Support Vector Machine* dengan basis perhitungan data mining, yang dimana membuat klasifikasi perhitungan tentang prediksi klien menggunakan data-set public dari *UCI Dataset*. Dalam data-set tersebut memiliki setidaknya 17 atribut yang dapat dilakukan sebuah ketentuan perhitungan untuk menghasilkan nilai akurasi dari minat klien, serta untuk menentukan atribut mana yang dapat mengurangi akurasi dan menambahkan akurasi pada perhitungan klasifikasi minat klien. Maka dapat dilakukan penerapan dan pemanfaatan teknik data mining guna memprediksi minat klien menggunakan *SVM* yang merupakan salah satu teknik yang cocok untuk melakukan prediksi, karena pemisah dua buah class pada input space dijabarkan dengan sederhana untuk mencari hyper terbaik untuk mendapatkan nilai terbaik dalam akurasi (Sulaehani, 2016).

Kelebihan dari metode *SVM* adalah *SVM* dapat menentukan *hyperplane* atau bidang pemisah dengan memilih bidang dengan optimal margin maka generalisasi pada *SVM* dapat terjaga dengan sendirinya, tingkat generalisasi pada *SVM* tidak dipengaruhi oleh jumlah data latih, dengan menentukan parameter soft margin, noise dapat dikontrol sehingga makin besar parameter soft margin, makin besar pula pinalti yang dikenakan pada kesalahan pada klasifikasi sehingga proses pelatihan semakin ketat. (Nugroho, Witarto, & Handoko, 2003).

Dari jurnal yang disusun oleh (Amalia, 2018) dengan judul Perbandingan Metode Data Mining *SVM* dan *NN* Untuk Klasifikasi Penyakit Ginjal Kronis, *SVM* memiliki tingkat akurasi yang lebih tinggi yaitu 95.16% sedangkan akurasi dari metode *Neural Network* adalah 93.36%.

Dari penelitian kedua yang disusun oleh (Budianto, Maryono, & Ariyuana, 2018) yang berjudul Perbandingan *K-Nearest Neighbor (KNN)* Dan *Support Vector Machine (SVM)* Dalam Pengenalan Karakter Plat Kendaraan Bermotor mempunyai tingkat akurasi 95% sedangkan dengan

menggunakan metode *KNN* memiliki tingkat akurasi sebesar 80%.

Dari penelitian ketiga yang disusun oleh (Ferawaty, Zarlis, & Nababan, 2016) dengan judul Analisis Perbandingan Akurasi dalam Identifikasi Autism dengan *SVM* dan *Naive Bayes* hasil akurasi dengan menggunakan metode *SVM* lebih besar yaitu sebesar 93.12% sedangkan menggunakan metode *Naive Bayes* yaitu sebesar 73.34%.

Dari ketiga jurnal yang telah disebutkan dapat disimpulkan bahwa metode *SVM* mempunyai kelebihan dari segi akurasi dibanding metode *Naive Bayes*, *KNN* dan *Neural Network*, hal itu disebabkan karena *svm* memiliki kelebihan pada *Generalisasi, Curse Of Dimensionality, Feasibility*.

*SVM* merupakan salah satu metode unggulan dibidang *pattern recognition*, sebagai algoritma yang terhitung masih baru (1992), *SVM* berkembang pesat karena berprinsip *Structural Risk Minimization (SRM)* dengan tujuan menemukan *hyperplane* terbaik untuk memisahkan dua buah kelas pada suatu space sehingga algoritma ini cocok untuk diimplementasikan sebagai clasificator. Berbeda dengan strategi neural network yang berusaha mencari hyperplane pemisah antar class, *SVM* berusaha menemukan hyperplane yang terbaik pada input space. Prinsip dasar *SVM* adalah *linear classifier*, dan selanjutnya dikembangkan agar dapat bekerja pada problem non-linear. dengan memasukkan konsep kernel trick pada ruang kerja berdimensi tinggi (Aulia, Hadiyoso, & Ramadan, 2015). Selain itu alasan yang kedua *SVM* memiliki tahapan-tahapan seperti transormasi data, matrik k, pendugaan koefisien *weight (w)* dan *bias (b)*.

Pada penelitian dengan judul Perbandingan Metode Data Mining *SVM* dan *NN* Untuk Klasifikasi Penyakit Ginjal Kronis (Amalia, 2018) disimpulkan bahwa Penelitian ini melakukan pengolahan dataset penyakit ginjal kronis yang diperoleh dari uci repository. Peneliti melakukan pengolahan data untuk memperoleh metode mana yang memiliki nilai akurasi yang lebih tinggi bagi dataset penyakit ginjal kronis. Diketahui dari hasil penelitian yaitu dengan metode *Neural Network* diperoleh nilai akurasi 93,36% dan metode *Support Vector Machine (SVM)* diperoleh nilai akurasi 95.16%. Hasil yang diperoleh termasuk

dalam jenis klasifikasi sangat baik. Sehingga dapat disimpulkan bahwa SVM dan NN memiliki performa kinerja yang baik untuk pengolahan dataset penyakit ginjal kronis. Dan dari hasil penelitian diketahui untuk dataset ginjal kronis bahwa metode SVM menghasilkan nilai akurasi yang lebih tinggi dari metode *Neural Network*.

Pada penelitian dengan judul Perbandingan *K-Nearest Neighbor (KNN)* Dan *Support Vector Machine (SVM)* Dalam Pengenalan Karakter Plat Kendaraan Bermotor (Budianto, Maryono, & Ariyuana, 2018) disimpulkan bahwa dari hasil pengenalan karakter plat nomor dengan metode SVM dan KNN didapatkan hasil bahwa Prosentase akurasi pengenalan dengan metode SVM adalah 95%. Sedangkan Pengenalan dengan menggunakan metode KNN adalah 80%. Terjadi beberapa kesalahan dalam mengenali karakter yang hampir mirip seperti 8 dengan B, 0 dengan O dan Q dan beberapa karakter yang lain.

Pada penelitian dengan judul Analisis Perbandingan Akurasi dalam Identifikasi Autism dengan SVM dan *Naive Bayes* (Ferawaty, Zarlis, & Nababan, 2016). Berdasarkan pengujian dengan menggunakan banyak data yang berbeda – beda saat percobaan dapat diambil kesimpulan bahwa Metode SVM menghasilkan tingkat akurasi rata - rata klasifikasi 93,12%. Hasil 93,12% ini merupakan total dari hasil rata – rata dari beberapa percobaan yang telah dilakukan dengan menggunakan metode SVM. Metode *Naive Bayes* menghasilkan rata – rata klasifikasi 73,34%. Hasil 73,34% ini merupakan total dari hasil rata – rata dari beberapa percobaan yang telah dilakukan dengan menggunakan metode *Naive Bayes*.

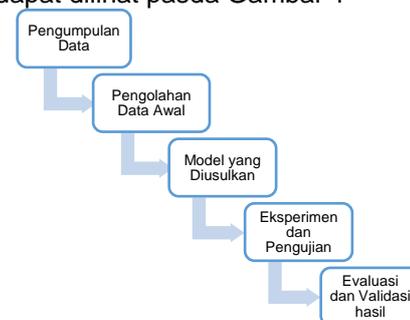
## 2. Metode Penelitian

Penelitian atau *research* dapat diartikan mencari kembali. *Re* yang berarti kembali dan *To Search* ialah mencari. Dikatakan mencari kembali karena teori atau hasil penelitian sebelumnya belum tentu benar selamanya. Oleh sebab itu, waktu, objek dan lokasi yang berbeda dapat dilakukan penelitian kembali Obyek yang diteliti berdasarkan tujuan dan tingkat kealamian (*natural setting*) yang diklasifikasikan disebut dengan metode penelitian. Berdasarkan tujuannya, metode penelitian pun dapat diklasifikasikan menjadi penelitian terapan (*applied research*) dan

penelitian pengembangan (*research and development*). Berdasarkan tingkat kealamiannya, metode penelitian dapat dikelompokkan menjadi metode penelitian eksperimen, *survey* dan naturalistik (Suhardjono et al., 2019).

Menurut Nurdin & Hartati (2019: 13) menyimpulkan bahwa "penelitian (*research*) pada hakikatnya merupakan salah satu cara untuk menyelesaikan suatu masalah atau mencari jawaban dari persoalan yang dihadapi secara ilmiah. Penelitian menggunakan cara berpikir reflektif, berpikir keilmuan dengan prosedur yang sesuai dengan tujuan dan sifat penyelidikan".

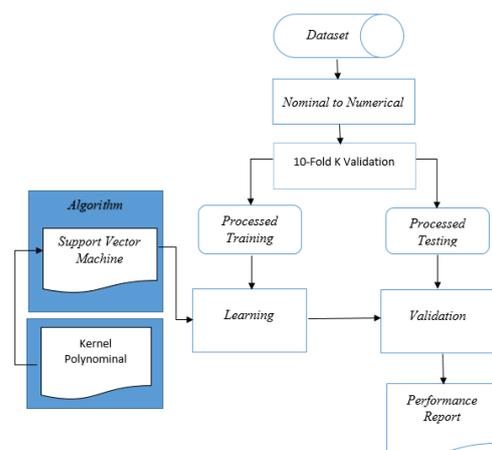
Metode penelitian secara umum, memiliki 4 metode penelitian yaitu: penelitian tindakan, eksperimen, studi kasus, dan *survey* (Dawson, 2009). Pada metode penelitian yang digunakan adalah eksperimen dan tahapan dalam penelitian yang dapat dilihat pasda Gambar 1



Gambar 1. Rancang Penelitian  
Sumber : Dawson (2009)

### 2.1 Pemodelan yang diusulkan

Setelah menentukan metode yang akan digunakan selanjutnya menentukan pemodelan yang diusulkan, pemodelan yang diusulkan dapat dilihat pada Gambar 2



Gambar 2. Pemodelan yang diusulkan

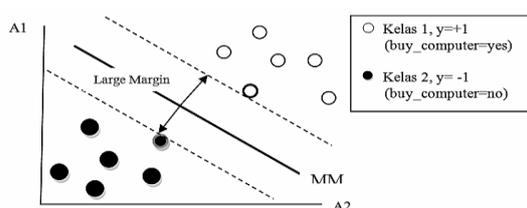
## 2.2 Prediksi

Prediksi ialah salah satu cara atau proses yang memperkirakan secara urut dan sistematis dalam mengenai sesuatu yang mungkin dapat terjadi pada masa depan berdasarkan adanya informasi pada suatu kejadian yang sudah terjadi dan yang sekarang diperoleh, agar tingkat kegagalan dan kesalahan yang terjadi dapat diperkecil. Dalam suatu prediksi mungkin saja jawabannya tidak ada pada suatu kepastian, tentang apa yang nantinya terjadi pada kejadian yang akan mendatang, akan tetapi nanti ada upaya untuk mencari jawaban yang akurat saat sesuatu terjadi (Kusumodestoni & Sarwido, 2017).

Prediksi pula sama artiannya dengan ramalan atau perkiraan. Prediksi bisa berdasarkan metode ilmiah ataupun *subjektif* belaka. Kesimpulannya ialah prediksi memiliki pengertian tergantung pada *konteks* atau permasalahannya. Berbeda dengan pengertian prediksi secara bahasa yang berarti ramalan atau perkiraan pada suatu yang menjadi pengertian yang baku (Amak Yunus EP, 2017).

## 2.3 Metode SVM

Algoritma *Support Vector Machine* (SVM) dikembangkan oleh *Boser, Guyon, Vapnik*, pertama kali dipresentasikan pada tahun 1992 di *Annual Workshop on Computational Learning Theory*. Suatu teknik yang relatif baru dilakukan untuk memprediksi, baik dalam kasus klasifikasi maupun regresi, yang sangat populer belakangan ini dengan selalu mencapai solusi yang sama untuk setiap running dan berusaha untuk menemukan fungsi pemisah (*klasifier*) yang optimal dan mampu memisahkan dua set data dari dua kelas yang berbeda, performansinya meyakinkan dalam memprediksi kelas suatu data baru (Abbas, 2016). Proses cara kerja SVM dapat dilihat pada Gambar 5



Gambar 5. SVM Hyperlane Terbaik Yang Memisahkan Antara Class +1 dan -1 (Sumber Prima Wijaya & Muslim, 2016)

Menurut Noor (2018), SVM dapat diartikan sebagai salah satu proses yang paling berpengaruh dan kuat untuk memecahkan sebuah *type* klasifikasi. SVM sendiri ialah seperangkat metode yang terkait untuk suatu metode pembelajaran, untuk kedua masalah klasifikasi dan regresi. Dengan berorientasi pada suatu tugas, hasil yang kuat, sifat komputerisasi yang mudah dikerjakan, SVM mampu mencapai sukses besar hingga dianggap sebagai *state-of-the-art classifier* pada saat ini. Dua kelas data yang digambarkan sebagai lingkaran dan pada titik-titik yang disajikan pada sebuah angka. Secara intuitif diamati, ada banyak keputusan *hyperplanes* yang dapat digunakan untuk memisahkan kedua kelompok data. Namun, yang digambarkan dengan angka ini dipilih sebagai yang menguntungkan memisahkan bidang, karena mengandung margin bermuatan maksimal antara dua kelas. Karena itu, dalam tujuan fungsi svm, sebuah istilah *regularization* mewakili margin muncul. Apalagi seperti yang terlihat di angka ini, hanya mereka yang penuh poin disebut mendukung vektor terutama menentukan memisahkan bidang, sementara poin lain tidak memberi kontribusi untuk margin di semua. Dalam kata lain, hanya menjadi sejumlah titik penting untuk melakukan klasifikasi tujuan dalam kerangka SVM dan dengan demikian harus diambil hasilnya.

Banyak teknik data *mining* atau *machine learning* yang dikembangkan dengan asumsi kelinieran. Sehingga algoritma yang dihasilkan terbatas untuk kasus-kasus yang linier. Oleh sebab itu, apa bila suatu kasus klasifikasi memperlihatkan ketidaklinieran, seperti *perceptron* misalnya, maka tidak bisa diatasi. Secara umum, kasus-kasus di dunia nyata adalah kasus yang tidak linier (Budi Santosa, 2018: 198).

Data pada suatu *dataset* diberikan variable  $x_i$ , sedangkan kelas pada *dataset* diberikan variable  $y_i$ . Metode SVM membagi *dataset* menjadi 2 kelas. Kelas pertama yang akan dipisah oleh *hyperplane* bernilai 1 jika  $X_i.W + b \geq 1$  untuk  $y_i = 1$ , sedangkan kelas lainnya bernilai -1 jika  $X_i.W + b \leq -1$  untuk  $y_i = -1$ . Bobot vector ( $w$ ) adalah garis *vector* yang tegak lurus antara titik pusat koordinat dengan garis *hyperplane*. *Bias* ( $b$ ) merupakan garis relative terhadap titik koordinat. merupakan persamaan untuk menghitung nilai  $b$  dan persamaan mencari nilai  $w$  (Parapat et al., 2018). sebagai berikut:

$$b = -\frac{1}{2} (w \cdot x + w \cdot x) \quad (1)$$

$$w = \sum_{i=1}^n \alpha_i y_i x_i \quad (2)$$

Banyak teknik data mining atau *machine learning* yang dikembangkan dengan asumsi kelinieran, sehingga algoritma yang dihasilkan terbatas untuk kasus-kasus yang linier (Santosa, 2007). SVM dapat bekerja pada data non-linier dengan menggunakan pendekatan kernel pada fitur data awal himpunan data. Fungsi kernel yang digunakan untuk memetakan dimensi awal (dimensi yang lebih rendah) himpunan data ke dimensi baru (dimensi yang relatif lebih tinggi). Menurut Prasetyo (2012) macam fungsi kernel diantaranya:

1. Kernel *Gaussian Radial Basic Function* (RBF)

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2a^2}\right)$$

2. Kernel *Polynomial*

$$K(x_i, x_j) = (x_i \cdot x_j + C)^d$$

dan adalah pasangan dua data training. Parameter merupakan konstanta. Fungsi kernel mana yang harus digunakan untuk *substitusi dot product* di *feature space* sangat tergantung pada data karena fungsi kernel ini akan menentukan fitur baru di mana *hyperplane* akan dicari (Santosa, 2007).

## 2.4 Pengumpulan Data

Pada penelitian ini penulis menggunakan data publik dari *UCI Machine Learning Repository* dengan nama data *Bank Marketing Data Set*, yang mudah diakses dan tersedia untuk publik karena bersifat terbuka. *Dataset* tersebut dipublikasikan pada tahun 2014. *Dataset* terdiri dari 45211 *record*, serta terdiri dari 17 atribut. Dimana setiap atribut memeberikan informasi tentang profil klien dan pendekatan seorang *telemarketing* dalam mempromosikan sebuah produk, serta satu atribut kelas yang memberikan informasi klien berminat atau tidak pada deposito yang ditawarkan. Informasi yang dibawakan pada *Bank Marketing Data Set* seperti: *age, job, marital, education, default, housing, loan, contact, month, day of week, duration, campaign, pdays, previous, poutcome* dan y (S. Moro, P. Cortez dan P. Rita).

Dari jumlah data yang begitu besar tidak dapat diproses, karena terlalu lama untuk dilakukan perhitungan pada aplikasi yang digunakan. Maka pada tahapan selanjutnya data akan diperkecil agar lebih cepat dalam pemrosesan perhitungannya.

## 2.5 Pengolahan Data Pertama

Data yang akan diolah menggunakan data yang berbentuk numerik pada atribut data bank marketing dataset, dari 17 atribut pada dataset bank marketing ada 7 atribut yang memiliki data numerik ialah sebagai berikut: (x1) age, (x2) balance, (x3) day, (x4) duration, (x5) campaign, (x6) pdays, (x7) previous, dan satu atribut class dengan label Y (1) untuk Yes dan (-1) untuk No.

Dari atribut yang telah diambil maka selanjutnya adalah menentukan jumlah data yang diolah sebesar 36 data yang dimana terdiri dari 6 data latih untuk menentukan pemodelan dan 30 data uji yang akan ditentukan hasil dalam prediksi untuk mengklasifikasikan klien berminat atau tidak pada produk deposito. seperti pada table data berikut.

Tabel 1. Data Awal

No	X1	X2	X3	X4	X5	X6	X7	Y
1	30	1787	19	79	1	-1	0	-1
2	33	4789	11	220	1	339	4	-1
3	35	1350	16	185	1	330	1	-1
4	30	1476	3	199	4	-1	0	-1
5	59	0	5	226	1	-1	0	-1
6	35	747	23	141	2	176	3	-1
7	36	307	14	341	1	330	2	-1
8	39	147	6	151	2	-1	0	-1
9	41	221	14	57	2	-1	0	-1
10	43	-88	17	313	1	147	2	-1
11	39	9374	20	273	1	-1	0	-1
12	43	264	17	113	2	-1	0	-1
13	36	1109	13	328	2	-1	0	-1
14	31	360	29	89	1	241	1	-1
15	40	194	29	189	2	-1	0	-1
16	56	4073	27	239	5	-1	0	-1
17	37	2317	20	114	1	152	2	-1
18	25	-221	23	250	1	-1	0	-1
19	55	1613	3	296	1	270	3	1
20	34	1337	17	330	2	127	3	1
21	63	3904	4	250	2	187	2	1
22	48	1147	5	389	1	64	2	1
23	36	1049	14	224	4	-1	0	1
24	54	2206	12	104	1	99	2	1
25	38	1988	27	164	2	130	2	1
26	33	661	18	968	1	-1	0	1
27	59	1	15	159	2	378	3	1
28	37	480	22	344	2	182	8	1
29	37	189	1	238	1	107	2	1
30	57	0	28	648	1	-1	0	1
31	60	366	3	593	1	-1	0	1
32	44	205	3	289	1	-1	0	1
33	32	5514	22	319	1	182	1	1
34	33	273	16	803	1	-1	0	1
35	28	4579	12	409	2	-1	0	1
36	37	5106	30	244	1	-1	0	1

Tabel 2. Data Latih

X1	X2	X3	X4	X5	X6	X7	Y
35	747	23	141	2	176	3	-1
43	264	17	113	2	-1	0	-1
25	-221	23	250	1	-1	0	-1

54	2206	12	104	1	99	2	1
57	0	28	648	1	-1	0	1
37	5106	30	244	1	-1	0	1

Tabel 3. Data Uji

X1	X2	X3	X4	X5	X6	X7	Y
30	1787	19	79	1	-1	0	-1
33	4789	11	220	1	339	4	-1
35	1350	16	185	1	330	1	-1
30	1476	3	199	4	-1	0	-1
59	0	5	226	1	-1	0	-1
36	307	14	341	1	330	2	-1
39	147	6	151	2	-1	0	-1
41	221	14	57	2	-1	0	-1
43	-88	17	313	1	147	2	-1
39	9374	20	273	1	-1	0	-1
36	1109	13	328	2	-1	0	-1
31	360	29	89	1	241	1	-1
40	194	29	189	2	-1	0	-1
56	4073	27	239	5	-1	0	-1
37	2317	20	114	1	152	2	-1
55	1613	3	296	1	270	3	1
34	1337	17	330	2	127	3	1
63	3904	4	250	2	187	2	1
48	1147	5	389	1	64	2	1
36	1049	14	224	4	-1	0	1
38	1988	27	164	2	130	2	1
33	661	18	968	1	-1	0	1
59	1	15	159	2	378	3	1
37	480	22	344	2	182	8	1
37	189	1	238	1	107	2	1
60	366	3	593	1	-1	0	1
44	205	3	289	1	-1	0	1
32	5514	22	319	1	182	1	1
33	273	16	803	1	-1	0	1
28	4579	12	409	2	-1	0	1

### 3. Hasil dan Pembahasan

#### 3.1. Hasil Penelitian

Hasil penelitian penerapan algoritma *Support Vector Machine* bertujuan untuk membantu telemarketing dalam memprediksi minat klien. Hasil prediksi yang dilakukan secara maksimal atau tidaknya dapat dilihat setelah melakukan eksperimen. Eksperimen ini dilakukan menggunakan algoritma *Support Vector Machine* dengan menggunakan kernel *polynomial*.

#### 3.2. Perhitungan Manual SVM

Pada tahapan ini dilakukan perhitungan manual dengan menggunakan *software excel 2013* dengan ketentuan perhitungan dimulai dari pemilihan data berjumlah 36 *record*, yang dimana data diambil dengan 18 *record* dengan atribut *class* bernilai *yes* (1) dan 18 *record* dengan *class* bernilai *no* (-1). Dan untuk data yang diuji menggunakan 7 atribut yang dimana merupakan faktor utama untuk mengetahui hasil dari prediksi. Berikut tahapan pada perhitungan manual yang dilakukan :

#### 1. Transformasi Data

Lakukan normalisasi data sehingga data tersebut bernilai antara 0 dan 1 yang dapat dilihat pada tabel 1 dengan menggunakan persamaan :

$$X_n = \frac{08 * (X - a)}{b - a} = 0.1$$

Dimana :

Xn = Nilai ke-n

A = Nilai angka Terendah

B = Nilai angka Tertinggi

0.8 dan 0.1 = Ketetapan

Dari rumus transformasi data yang sudah dihitung, maka menghasilkan nilai data yang dapat dilihat pada Tabel 4

Tabel 4. Transformasi Data

X1	X2	X3	X4	X5	X6	X7	Y
0.21	0.27	0.60	0.12	0.10	0.10	0.10	-1
0.27	0.52	0.38	0.24	0.10	0.82	0.50	-1
0.31	0.23	0.51	0.21	0.10	0.80	0.20	-1
0.21	0.24	0.16	0.22	0.70	0.10	0.10	-1
0.82	0.12	0.21	0.25	0.10	0.10	0.10	-1
0.31	0.18	0.71	0.17	0.30	0.47	0.40	-1
0.33	0.14	0.46	0.35	0.10	0.80	0.30	-1
0.39	0.13	0.24	0.18	0.30	0.10	0.10	-1
0.44	0.14	0.46	0.10	0.30	0.10	0.10	-1
0.48	0.11	0.54	0.32	0.10	0.41	0.30	-1
0.39	0.90	0.62	0.29	0.10	0.10	0.10	-1
0.48	0.14	0.54	0.15	0.30	0.10	0.10	-1
0.33	0.21	0.43	0.34	0.30	0.10	0.10	-1
0.23	0.15	0.87	0.13	0.10	0.61	0.20	-1
0.42	0.13	0.87	0.22	0.30	0.10	0.10	-1
0.75	0.46	0.82	0.26	0.90	0.10	0.10	-1
0.35	0.31	0.62	0.15	0.10	0.42	0.30	-1
0.10	0.10	0.71	0.27	0.10	0.10	0.10	-1
0.73	0.25	0.16	0.31	0.10	0.67	0.40	1
0.29	0.23	0.54	0.34	0.30	0.37	0.40	1
0.90	0.44	0.18	0.27	0.30	0.50	0.30	1
0.58	0.21	0.21	0.39	0.10	0.24	0.30	1
0.33	0.21	0.46	0.25	0.70	0.10	0.10	1
0.71	0.30	0.40	0.14	0.10	0.31	0.30	1
0.37	0.28	0.82	0.19	0.30	0.38	0.30	1
0.27	0.17	0.57	0.90	0.10	0.10	0.10	1
0.82	0.12	0.49	0.19	0.30	0.90	0.40	1
0.35	0.16	0.68	0.35	0.30	0.49	0.90	1
0.35	0.13	0.10	0.26	0.10	0.33	0.30	1
0.77	0.12	0.84	0.62	0.10	0.10	0.10	1
0.84	0.15	0.16	0.57	0.10	0.10	0.10	1
0.50	0.14	0.16	0.30	0.10	0.10	0.10	1
0.25	0.58	0.68	0.33	0.10	0.49	0.20	1
0.27	0.14	0.51	0.76	0.10	0.10	0.10	1
0.16	0.50	0.40	0.41	0.30	0.10	0.10	1
0.35	0.54	0.90	0.26	0.10	0.10	0.10	1

#### 2. Mencari Matrik K

Untuk mencari nilai SVM terlebih dahulu kita harus mencari nilai matrik *K* dengan dengan parameter kernel polinomial persamaan sebagai berikut :

$$K(x, xi) = \phi(xi)$$

$$K(1,1) = (x1 * x1 + 1)^2$$

$$= ((0.31 * 0.31) + (0.18 * 0.18) + (0.71 * 0.71) + (0.17 * 0.17) + (0.30 * 0.30) + (0.47 * 0.47) + (0.40 * 0.40) + 1)^2 = 4.5509$$

Setelah menghitung nilai matrik kernel  $K$ , maka didapatkan matrik  $K$  dengan ukuran  $N * N$

$$K(x, xi) = \phi(xi) = K (N * N)$$

Dari hasil rumus diatas maka di dapat hasil yang di dapat pada Tabel 5.

Tabel 5. Hasil Dari Mencari Matrik K

NO	1	2	3	4	5	6
1	4,5509	3,0979	2,9344	3,5433	4,3421	4,0293
2	3,0979	2,8038	2,3557	2,9356	3,9478	3,3196
3	2,9344	2,3557	2,6319	2,2370	3,5465	3,3382
4	3,5433	2,9356	2,2370	3,9037	4,3469	3,5593
5	4,3421	3,9478	3,5465	4,3469	7,5048	5,2494
6	4,0293	3,3196	3,3382	3,5593	5,2494	5,4301

### 3. Pendugaan Kefesien weight (w) dan bias (b)

Pendugaan koefisien *weight* dan *bias* dapat menggunakan persamaan berikut:

Pada tahapan ini dilakukan perhitungan manual *Support Vector Machine* untuk mencari nilai  $W$  dengan langkah- langkah sebagai berikut:

$$w = \sum_{i=1}^N a_i y_i K(x, x_1)$$

$$= 0.28444 \begin{pmatrix} 4.551 \\ 3.098 \\ 2.934 \\ 3.543 \\ 4.342 \\ 4.029 \end{pmatrix} + 0.28444 \begin{pmatrix} 3.098 \\ 2.804 \\ 2.356 \\ 2.936 \\ 2.948 \\ 3.320 \end{pmatrix}$$

$$+ 0.28444 \begin{pmatrix} 2.934 \\ 2.356 \\ 2.632 \\ 2.237 \\ 3.547 \\ 3.338 \end{pmatrix} - 0.28444 \begin{pmatrix} 3.543 \\ 3.936 \\ 2.237 \\ 3.904 \\ 4.347 \\ 3.559 \end{pmatrix}$$

$$- 0.28444 \begin{pmatrix} 4.342 \\ 3.948 \\ 3.547 \\ 4.347 \\ 7.505 \\ 5.249 \end{pmatrix} - 0.28444 \begin{pmatrix} 4.029 \\ 3.320 \\ 3.338 \\ 3.559 \\ 5.249 \\ 5.430 \end{pmatrix}$$

$$= \begin{pmatrix} 6.399 \\ 5.251 \\ 4.848 \\ 5.838 \\ 8.231 \\ 7.090 \end{pmatrix}$$

Pada proses ini terdapat nilai  $w$  dimana nilai  $w$  didapat dari rumus  $a_i y_i K(x, x_1)$ , dimana untuk mencari nilai  $a = 6 / (\sum_{i=1}^N (K * z))$ . Pencarian nilai  $z$  dilakukan dengan perkalian  $Y_i$  dikali  $Y_j$ . Dimana nilai atribut class  $Y_i$  sama dengan jumlah hasil pada kernel matrix ( $Y_j$ ). Setelah didapat nilai  $w$ , selanjutnya pilih salahsatu support vector dari kelas "+1" dan "-1" untuk menghitung nilai  $b$  dengan persamaan :

$$b = -\frac{1}{2} (w \cdot x^+ + w \cdot x^-)$$

$$= \frac{1}{2} * \begin{pmatrix} 6.399 \\ 5.251 \\ 4.848 \\ 5.838 \\ 8.231 \\ 7.090 \end{pmatrix} \begin{pmatrix} 4.551 \\ 3.098 \\ 2.934 \\ 3.543 \\ 4.342 \\ 4.029 \end{pmatrix} + \begin{pmatrix} 6.399 \\ 5.251 \\ 4.848 \\ 5.838 \\ 8.231 \\ 7.090 \end{pmatrix} \begin{pmatrix} 3.543 \\ 3.936 \\ 2.237 \\ 3.904 \\ 4.347 \\ 3.559 \end{pmatrix}$$

$$= -138.67375489$$

Jadi  $b$  sudah diketahui dengan nilai sebesar

$$= -138.674$$

### 4. Menentukan Hyperplane

Maka dari pemodelan yang telah dilakukan oleh data latih ditentukan hyperplane pada perhitungan SVM dengan kernel *polynomial* dengan ketentuan:

$$F(x) = w_1 \cdot x_1 + w_2 \cdot x_2 + w_3 \cdot x_3 + w_4 \cdot x_4 + w_5 \cdot x_5 + w_6 \cdot x_6 + b = 0$$

maka nilai *hyperplane* nya sebagai berikut:

$$F(x) = 6,399 x_1 + 5,251 x_2 + 4,848 x_3 + 5,838 x_4 + 8,231 x_5 + 7,090 x_6 + (-138,674) = 0$$

### 5. Prediksi SVM

Setelah mendapatkan nilai  $w$  (*weight*) dan  $b$  (*bias*), maka model SVM dapat digunakan *Kernel Trick* yang dimana  $c = 1$  dan  $d = 2$  pada data uji dengan persamaan :

$$f(\phi(x)) = \text{sign}(w \cdot \phi(x) + b)$$

$$= \text{sign} \left( \sum_{i=1}^N a_i y_i \phi(xi)^T \cdot \phi(x) + b \right)$$

Dari hasil rumus diatas dapat menghasilkan nilai data untuk hasil prediksi yang telah dihitung, yang dapat dilihat pada Tabel 6.

Tabel 6. Hasil Prediksi

X1	X2	X3	X4	X5	X6	X7	Y	Prediksi
0.21	0.27	0.60	0.12	0.10	0.10	0.10	-1	-1
0.27	0.52	0.38	0.24	0.10	0.82	0.50	-1	1
0.31	0.23	0.51	0.21	0.10	0.80	0.20	-1	-1
0.21	0.24	0.16	0.22	0.70	0.10	0.10	-1	-1
0.82	0.12	0.21	0.25	0.10	0.10	0.10	-1	-1
0.33	0.14	0.46	0.35	0.10	0.80	0.30	-1	-1
0.39	0.13	0.24	0.18	0.30	0.10	0.10	-1	-1
0.44	0.14	0.46	0.10	0.30	0.10	0.10	-1	-1
0.48	0.11	0.54	0.32	0.10	0.41	0.30	-1	-1
0.39	0.90	0.62	0.29	0.10	0.10	0.10	-1	-1
0.33	0.21	0.43	0.34	0.30	0.10	0.10	-1	-1
0.23	0.15	0.87	0.13	0.10	0.61	0.20	-1	-1
0.42	0.13	0.87	0.22	0.30	0.10	0.10	-1	-1
0.75	0.46	0.82	0.26	0.90	0.10	0.10	-1	1
0.35	0.31	0.62	0.15	0.10	0.42	0.30	-1	-1
0.73	0.25	0.16	0.31	0.10	0.67	0.40	1	-1
0.29	0.23	0.54	0.34	0.30	0.37	0.40	1	-1
0.90	0.44	0.18	0.27	0.30	0.50	0.30	1	1
0.58	0.21	0.21	0.39	0.10	0.24	0.30	1	-1
0.33	0.21	0.46	0.25	0.70	0.10	0.10	1	-1
0.37	0.28	0.82	0.19	0.30	0.38	0.30	1	1
0.27	0.17	0.57	0.90	0.10	0.10	0.10	1	-1
0.82	0.12	0.49	0.19	0.30	0.90	0.40	1	1
0.35	0.16	0.68	0.35	0.30	0.49	0.90	1	-1
0.35	0.13	0.10	0.26	0.10	0.33	0.30	1	-1
0.84	0.15	0.16	0.57	0.10	0.10	0.10	1	-1
0.50	0.14	0.16	0.30	0.10	0.10	0.10	1	-1
0.25	0.58	0.68	0.33	0.10	0.49	0.20	1	1
0.27	0.14	0.51	0.76	0.10	0.10	0.10	1	-1
0.16	0.50	0.40	0.41	0.30	0.10	0.10	1	-1

Pada Tabel 6 ditentukan bahwa hasil pemodelan dan dilakukannya sebuah prediksi dengan 30 data yang dihitung secara manual mendapatkan hasil klasifikasi sebesar 24 data klien mempunyai nilai prediksi (-1) dan 6 data klien mempunyai nilai prediksi (1), dimana nilai prediksi (-1) adalah klien yang tidak berminat terhadap produk deposito dan nilai (1) adalah klien yang berminat terhadap produk deposito.

#### 4. Kesimpulan

Pada penelitian ini dilakukan eksperimen terhadap algoritma *Support Vector Machine* untuk mengklasifikasi data dari *UCI Machine Learning Repository* menggunakan *dataset Bank Marketing*. Dengan sumber data terbaru yang berisi 17 atribut dan 45211 data. Dan saat ini dilakukan perhitungan manual menggunakan *Microsoft Excel 2013* dengan jumlah data 36 data yang dibagi menjadi data latih sebesar 6 data dan 30 data untuk diuji kebenaran dari pemodelan data yang diambil dan ternyata dari 30 data hasil prediksi menggambarkan 24 data yang tidak berminat pada deposito yang ditawarkan, sedangkan 6 data yang lainnya berminat kepada deposito yang ditawarkan. Dari 30 data maka hanya 20% yang menerima atau berminat kepada produk yang ditawarkan oleh *telemarketing bank*.

Maka dalam hasil yang didapatkan ternyata hasil prediksi dan data yang sudah ada memiliki nilai yang berbeda terutama

pada data yang bernilai (1) atau Yes, dalam hasil prediksi untuk data Yes hanya 4 data yang benar pada perhitungan prediksinya daripada hasil prediksi (-1) atau No yang memiliki 2 data yang salah dalam perhitungan prediksinya. Jadi dapat disimpulkan untuk True Positif pada dataset ialah data (-1) atau No karena data hasil perhitungannya lebih besar dari (1) atau Yes yang dapat disebut dengan True Negatif.

Bila diberikan sebuah hasil atau nilai accuracy yang digunakan pada perhitungan manual saat ini sebesar 56,67% untuk data yang tepat dalam perhitungannya.

Selanjutnya melakukan penelitian menggunakan algoritma *Support Vector Machine* untuk mengklasifikasi data dari *UCI Machine Learning Repository* menggunakan *Rapidminer* dengan *dataset Bank Marketing*. Dengan sumber data yang berisi 17 atribut yang dimana 9 atribut *nominal* dirubah menjadi *numeric* dengan operator *nominal to numeric* dengan perbandingan 1:0 merubah *nominal* yang dipilih menjadi satu (1) dan yang tidak terpilih menjadi nol (0), serta jumlah data sebesar 45211 data menjadi 4521 data yang mengikuti data penelitian terbaru dari *Bank Marketing Dataset*. Selanjutnya dihitung menggunakan Metode 10 *K-Fold Cross Validation Support Vector Machine* dan dengan perbandingan perhitungan lebih *Kernel Polynomial* mendapatkan nilai yang optimal dan terbaik dari kernel *dot product*. Kesimpulan yang dapat diambil dari hasil penelitian yang sudah dilakukan sebagai berikut:

1. Telah diterapkan algoritma SVM pada Bank Marketing Data Set menggunakan metode 10-Fold Cross Validation performa klasifikasi dengan akurasi sebesar 94,91% dan AUC 0.979.
2. Telah diketahui nilai hasil akurasi dari algoritma SVM untuk klasifikasi Bank Marketing Data Set dengan parameter kernel polynominal hasil akurasi optimal sebesar 95,18% dan AUC 0.974 menggunakan 10-Fold Cross Validation.

#### Referensi

- Abbas, I. (2016). Penerapan Metode Moving Average (MA) Berbasis Algoritma Support Vector Machine (SVM) untuk Membandingkan Pola Kurva dengan Trend Kurva pada Trading Forex Online. *ILKOM Jurnal Ilmiah*.
- Agustina, C. (2019). Analisa Nasabah Potensial Tabungan Deposito

- Berjangka Menggunakan Teknik Klasifikasi Data Mining. *Jurnal Teknologi Informasi Dan Terapan*, 5(2), 105–112.
- Amak Yunus EP, A. (2017). Sistem Prediksi Penentuan Jenis Tanaman Sayuran Berdasarkan Kondisi Musim Dengan Pendekatan Metode Trend Moment. *Jurnal Mahasiswa Fakultas Sains Dan Teknologi*, 1(5).
- Aqham, A. A., & Hartomo, K. D. (2019). Data Mining untuk Nasabah Bank Telemarketing Menggunakan kombinasi Algoritma Naïve Bayes Dan Algoritma Genetik. *InfoTekJar (Jurnal Nasional Informatika Dan Teknologi Jaringan)*, 4(1), 47–56.
- Budi Santosa, A. U. (2018). *Data Mining dan Big Data Analytics: Teori dan Implementasi Menggunakan Python & Apache Spark* (Isa (ed.); 2nd ed.). Penebar Media Pustaka.
- Dawson, S. (2009). Introduction to electroweak symmetry breaking. *AIP Conference Proceedings*.
- Ghani, A. D., Salman, N., & Mustikasari. (2019). Algoritma k-Nearest Neighbor Berbasis Backward Elimination Pada Client Telemarketing. *Prosiding Seminar Ilmiah Sistem Informasi Dan Teknologi Informasi*, VIII(2), 141–150.
- Kusumodestoni, R. H., & Sarwido, S. (2017). Komparasi Model Support Vector Machines (Svm) Dan Neural Network Untuk Mengetahui Tingkat Akurasi Prediksi Tertinggi Harga Saham. *Jurnal Informatika Upgris*, 3(1).
- Noor, A. (2018). Perbandingan Algoritma Support Vector Machine Biasa dan Support Vector Machine berbasis Particle Swarm Optimization untuk Prediksi Gempa Bumi. *Jurnal Humaniora Teknologi*, 4(1), 31–37.
- Parapat, I. M., Furqon, M. T., & Sutrisno. (2018). Penerapan Metode Support Vector Machine ( SVM ) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(10), 3163–3169.
- Pratama, A., Wihandika, R. C., & Ratnawati, D. E. (2018). Implementasi Algoritme Support Vector Machine (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(March), 1704–1708.
- Saputra, E. P. (2017). Prediksi Keberhasilan Telemarketing Bank Untuk. *Jurnal Ilmu Pengetahuan Dan Teknologi Komputer*, 2(2), 66–72.
- Siregar, A. M., & Pusphabuana, A. (2016). *DATA MINING: Pengolahan Data Menjadi Informasi dengan RapidMiner*. CV Kekata Group.
- Suhardjono, Wiajaya, G., & Abdul, H. (2019). Prediksi Waktu Kelulusan Mahasiswa Menggunakan Svm Berbasis Pso. *Bianglala Informatika*, 7(2), 97–101.
- Sulaehani, R. (2016). Prediksi Keputusan Klien Telemarketing Untuk Deposito Pada Bank Menggunakan Algoritma Naive Bayes Berbasis Backward Elimination. *ILKOM Jurnal Ilmiah*, 8(3), 182–189.
- Tekouabou, S. C. K., Cherif, W., & Silkan, H. (2019). A data modeling approach for classification problems: Application to bank telemarketing prediction. *ACM International Conference Proceeding Series, Part F1481*, 1–7.
- Thohir, M. (2013). Metodologi Penelitian Sosial Budaya. In *CV Andi Offset*. Yogyakarta. MEDIA SAHABAT CENDEKIA.
- WK, W. N., & Adani, Y. (2018). Penerapan Algoritma Naïve Bayes Untuk Memprediksi Keputusan Calon Nasabah Dan Nasabah Tetapbank Bri Syariah Menerima Penawaran Program Deposito Berjangka. *Jurnal Teknologi Dan Informasi*, 8(1), 13–24.