Klasifikasi Pasien Pengidap Diabetes menggunakan Metode *Support Vector Machine*

Diniyal Amru Agatsa[1], Rita Rismala[2], Untari Novia Wisesty[3]

[1,2,3]Fakultas Informatika, Universitas Telkom, Bandung
[1]diniyalamruagatsa@students.telkomuniversity.ac.id,
[2]ritaris@telkomuniversity.ac.id, [3]untarinw@telkomuniversity.ac.id

Abstrak

Diabetes merupakan penyakit kronis yang terjadi ketika pankreas tidak menghasilkan cukup insulin(hormone yang mengatur gula darah) atau ketika tubuh tidak dapat menggunakan insulin secara efektif.

Rata-rata global angka pengidap diabetes pada orang dewasa meningkat hingga dua kali lipat sejak tahun 1980 dari 4,7% menjadi 8,5%. Diabetes menyebabkan 1,5 juta kematian pada tahun 2012. Glukosa darah yang lebih tinggi dari angka optimal menyebabkan 2,2 juta kematian, dengan meningkatkan risiko penyakit kardiovaskular dan lainnya. 43% dari 3,7 juta kematian tersebut terjadi sebelum usia 70 tahun [1]**.**

Untuk itu diperlukan suatu sistem klasifikasi data pasien yang dapat membantu penanggulangan diabetes, dengan menggunakan dataset Pima Indians Diabetes Database yang diperoleh dari *UC Irvine Machine Learning Repository* dan klasifikasi menggunakan *Support Vector Mechine* (SVM).

Kata kunci: diabetes, klasifikasi, SVM(support vector machine).

Abstract

Diabetes is a chronic disease that occurs because the pancreas does not produce enough insulin or hormones that cannot be used by insulin.

The number of people with global diabetes in adults has doubled since 1980 from 4.7% to 8.5%. Diabetes caused 1.5 million deaths in 2012. Blood glucose higher than the optimal number caused 2.2 million deaths, by increasing the risk of cardiovascular and other diseases. 43% of the 3.7 million deaths occurred before the age of 70 years [1].

For this reason, a patient data classification system is needed to help with diabetes management, using the Pima Indians Diabetes Database dataset obtained from the UC Irvine Machine Learning Repository and collected using Support Vector Machine (SVM).

Keywords: diabetes, classification, SVM(support vector machine).

1. Pendahuluan

1.1 Latar Belakang

Diabetes merupakan penyakit kronis yang terjadi ketika pankreas tidak menghasilkan cukup insulin(hormone yang mengatur gula darah) atau ketika tubuh tidak dapat menggunakan insulin secara efektif. Semua jenis diabetes dapat menyebabkan komplikasi dan meningkatkan risiko penyakit, beberapa penyakit yang mungkin menjadi komplikasi adalah serangan jantung, stroke, gagal ginjal, amputasi kaki, kehilangan penglihatan dan kerusakan saraf. Pada masa kehamilan diabetes yang tidak terkontrol dapat meningkatkan risiko kematian janin.

Secara global diperkirakan 422 juta orang dewasa mengiap diabetes pada tahun 2014 sedangkan pada tahun 1980 hanya 108 juta. Rata-rata global angka pengidap diabetes pada orang dewasa meningkat hingga dua kali lipat sejak tahun 1980 dari 4,7% menjadi 8,5%. Selama dekade terakhir, prevalensi diabetes meningkat cepat di Negara-negara berpenghasilan rendah dan menengah daripada di Negara-Negara berpenghasilan tinggi. Diabetes menyebabkan 1,5 juta kematian pada tahun 2012. Glukosa darah yang lebih tinggi dari angka optimal menyebabkan 2,2 juta kematian, dengan meningkatkan risiko penyakit kardiovaskular dan lainnya. 43% dari 3,7 juta kematian tersebut terjadi sebelum usia 70 tahun [1].

Dengan angka perkembangan diabetes yang meningkat, diperlukan teknologi yang dapat membantu penanggulangan diabetes, maka penulis membangun sebuah sistem klasifikasi pasien pengidap diabetes menggunakan metode *support vector machine* (SVM) yang dapat mengelompokan pasien kedalam

kelompok positif diabetes dan negatif diabetes.

Pada penelitian sebelumnya [2] dilakukan penelitian mengenai diabetes menggunakan dataset *Pima Indians Diabetes Database* menggunakan metode klasifikasi SVM menghasilkan nilai akurasi mencapai 83%. Oleh karna itu penulis melakukan penelitian dengan metode dan *dataset* serupa namun dilakukan proses k-fold untuk mengantisipasi kemungkinan perolehan model dengan akurasi yang lebih baik.

1.2 Topik dan Batasannya

Topik yang dibahas pada penelitian ini adalah bagaimana mengimplementasikan sistem klasifikasi pasien diabetes agar kemudian dapat dimanfaatkan untuk menunjang kemajuan ilmu kesehatan. Data yang digunakan merupakan *Pima Indians Diabetes dataset* yang mana merupakan data yang diperoleh dari 768 pasien dari sumber *UCI Repository of Machine Learning Databases*.

1.3 Tujuan

Membangun sistem klasifikasi data pasien yang dapat melakukan klasifikasi kedalam dua kelas yaitu positif diabetes atau negatif diabetes dengan menggunakan metode *support vector machine*(SVM) kemudian menganalisa kinerja sistem klasifikasi pasien diabetes yang telah dibangun.

1.4 Organisasi Tulisan

Bab dua akan menelaskan studi terkait yang berisi teori yang mendukung penelitian ini. Bab tiga menjelaskan rancangan sistem yang dibangun. Bab empat akan menjelaskan pengujian dan analisis penelitian. Bab lima akan menjelaskan kesimpulan yang diperoleh dari penelitian serta saran untuk penelitian selanjutnya.

2. Studi Terkait

2.1 Diabetes

Diabetes merupakan penyakit kronis yang terjadi ketika pankreas tidak menghasilkan cukup insulin(hormone yang mengatur gula darah) atau ketika tubuh tidak dapat menggunakan insulin secara efektif. Diabetes yang tidak terkontrol dapat mengakibatkan kerusakan serius pada jantung, pembuluh darah mata, ginjal dan saraf. Lebih dari 400 juta orang mengidap diabetes. Diabetes tipe 1 diakibatkan oleh kurangnya produksi insulin dalam tubuh sehingga orang-orang dengan diabetes tipe 1 memerlukan masukan insulin setiap harinya untuk dapat mengatur kadar glokosa dalam darah. Penyebab diabetes tipe 1 tidak diketahui dan tidak dapat dicegah. Gejala dari diabetes tipe 1 meliputi buang air kecil dan haus berlebihan, rentan merasa lapar, penurunan berat badan, perubahan penglihatan dan kelelahan. Diabetes tipe 2 diakibatkan oleh kurangnya kemampuan tubuh untuk memanfaatkan insulin dengan efektif. Gejalanya kurang lebih sama dengan diabetes tipe 1 namun kadang kala lebih tidak terdeteksi sehingga tidak terdiagnosis dan baru terlihat ketika muncul komplikasi. Secara umum disepakati bahwa diabetes tipe 1 adalah hasil dari suatu kompleks interaksi antar gen dan factor lingkungan, sedangkan diabetes tipe 2 ditentukan dari interaksi genetic dan metabolisme. Etnisitas, riwayat diabetes keluarga, kelebihan berat badan, usia saat hamil, dan usia saat ini menjadi beberapa dari penentu diabetes. Diagnosis didiagnosis dengan mengukur glukosa dalam sampel darah yang diambil saat pasien puasa, atau dua jam setelah 75g oral beban glokosa diambil [1].

2.2 Support Vector Machine(SVM)

Support Vector Machine (SVM) adalah pengklasifikasi *stock* terbaik. SVM memiliki ciri *low generation error*, secara komputasi tidak mahal, mudah untuk diinterpretasikan, namun sangat sensitif pada turning parameter dan pilihan kernel [3]. SVM cocok digunakan untuk data berdimensi tinggi [4]. *Support Vector Machine (SVM)* menerima input kontinu dan mengeluarkan output diskrit yang hanya dua kelas. Konsep *Support Sector Machine (SVM)* yaitu menemukan *hyperplane* yang memisahkan himpunan data ke dalam dua kelas secara linier [5]. *Hyperplane* dibangun oleh SVM pada dimensi ruang tinggi atau bahkan dimensi ruang tak terhingga [6]. *Hyperplane* pemisah untuk data berdimensi satu adalah sebuah titik, untuk data berdimensi dua adalah sebuah garis, untuk data berdimensi tiga adalah sebuah bidang datar, untuk data berdimensi lima *hyperplanenya* berdimensi empat. SVM berusaha menemukan *hyperplane* yang paling optimum. *Hyperplane* dikatakan optimum jika berada tepat di tengah-tengah kedua kelas sehingga memiliki jarak paling jauh ke data-data terluar di kedua kelas. Hal ini dapat terlihat dari margin yang dimiliki *hyperplane* [5]. Margin adalah jarak dari *hyperplane* ke data terdekat di kelas yang berbeda. Dengan demikian *hyperplane* yang optimal adalah yang memiliki margin paling besar [7]. Ada dua langkah yang dilakukan SVM untuk mendapatkan *hyperplane* yang optimum, yaitu menemukan data-data terluar pada

kedua kelas yang berada di perbatasan lalu dengan memperhitungkan data-data terluar tersebut dan tanpa memperhitungkan data lainnya ditentukanlah *hyperplane* paling optimum [5].

Proses pembelajaran SVM untuk menemukan support vector hanya bergantung pada *dot product* dari data pada ruang fitur yang dilambangkan dengan Φ. Perhitungan *dot product* dapat digantikan dengan fungsi kernel karena pada umumnya transformasi Φ tidak diketahui dan sulit dipahami. Fungsi kernel mendefinisikan fungsi transformasi Φ secara implisit, fungsi kernel ini disebut dengan *kernel trick*. Untuk mengetahui support vector hanya perlu diketahui fungsi kernel yang dipakai tanpa perlu mengetahui wujud dari fungsi non linear Φ. Terdapat empat fungsi kernel yang biasa digunakan, yaitu kernel linier, kernel polynomial, kernel gausian (*radial basis functional*, *RBF*), dan kernel sigmoid. Untuk menemukan hyperplane optimum yang berbasis metode sekuensial Algoritma 2.1 digunakan sebagai pembelajaran SVM [5].

Algoritma 3.1 SVM

1. Initialization, $\alpha_i = 0$
   Hitung matriks $D_{ij} = y_i y_j (K(x_i, x_j) + \lambda^2$
2. Lakukan tiga langkah dibawah ini untuk i = 1, 2, …, l
   a. $E_i = \sum_{j=1}^{l} \alpha_j D_{ij}$
   b. $\delta\alpha_i = \min \{\max[\gamma(1 - E_i) - \alpha_i], C - \alpha_i\}$
   c. $\alpha_i = \alpha_i + \delta\alpha_i$
3. Kembali ke langkah 3 sampai α konvergen

Keterangan:

α	= Variabel penyimpan untuk memeriksa kekonvergenan.
i, j	= Indeks.
λ	= Skalar perluasan.
D	= Matriks hasil modifikasi.
x	= Vector masukan.
y	= Klasifikasi objek data.
C	= Kontrol optimasi antara margin dan kesalahan klasifikasi.
γ	= Kontrol kecepatan belajar.
δ	= Perbedaan atau selisih.

2.3 Min-max Normalization

Min-max normalization digunakan untuk menyamakan skala antar atribut. Jika $\min_A$ dan $\max_A$ adalah nilai minimum dan maksimum dari atribut A. *Min-max normalization* memetakan nilai V_i pada A menjadi V'_i dalam kisaran [new_$\max_A$, new_$\max_A$] dengan komputasi [8] .

$$V'_i = \frac{V_i - \min_A}{\max_A - \min_A}(new\_max_A - new\_min_A) + new\_min_A \quad (1)$$

Keterangan:

A	= Sebuah atribut A.
V_i	= Nilai awal suatu atribut sebelum dilakukan *Min-max normalization.*
V'_i	= Nilai akhir suatu atribut setelah dilakukan *Min-max normalization.*
$\min_A$	= Nilai terkecil dari atribut A.
$\max_A$	= Nilai terbesar dari atribut A.
new_min_A	= Nilai terkecil yang akan dijadikan batas bawah pada atribut A.
new_max_A	= Nilai terbesar yang akan dijadikan batas atas pada atribut A.

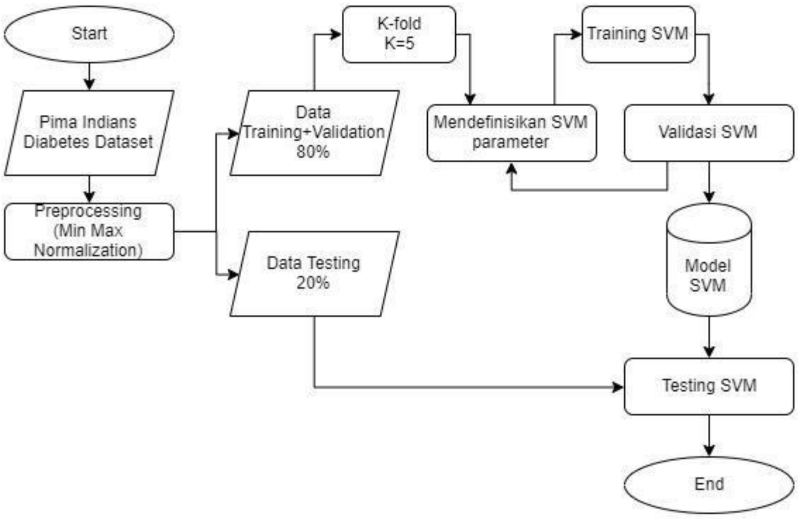
2.4 K-Fold Cross Validation

K-fold memecah data menjadi k bagian dimana masing-masing bagiannya memiliki jumlah data yang

seimbang. Satu bagian akan disimpan sebagai data validasi sementara data lainnya dapat digunakan sebagai data latih. Satu bagian data yang digunakan sebagai data validasi digantikan dengan bagian lain dan terus dilakukan hal yang sama sampai semua bagian data sudah diberlakukan sebagai data validasi. Akurasi yang didapat dari masing-masing iterasi kemudian dirata-rata untuk mendapatkan akurasi model [9].

3. Sistem Klasifikasi Pasien Pengidap Diabetes

Sistem yang dibangun pada penelitian ini bertujuan untuk melakukan klasifikasi terhadap pasien berdasarkan data yang diperoleh dari pasien. Untuk gambaran sistem dapat dilihat pada Gambar 3.1



Gambar 3. 1 Gambaran umum Sistem Klasifikasi

3.1 Dataset

Dataset yang digunakan adalah *Pima Indians Diabetes Database* yang diperoleh dari *UCI Machine Lerning Repository* yang terdiri dari delapan atribut dengan nilai numerik dan memiliki dua kelas data yaitu *tested_positive* untuk kelas pasien teridentifikasi diabetes dan *tested_negative* untuk kelas pasien teridentifikasi bebas diabetes, data tersebut diambil dari 768 pasien. Delapan atribut yang dimaksud adalah jumlah pengalaman hamil (preg), konsentrasi glukosa plasma 2 jam dalam tes toleransi glukosa oral (plas), tekanan darah diastolik (pres), ketebalan lipatan kulit trisep (skin), insulin serum 2 jam (insu), indeks masa tubuh (mass), fungsi silsilah diabetes (pedi) dan umur (age). Sejumlah 614 data digunakan sebagai data latih berikut data validasi dan 154 data digunakan sebagai data uji.

3.2 Preprocessing menggunakan *Min-max normalization*

Min-max normalization dilakukan terhadap seluruh data, dalam penelitian ini new_min_A dan new_max_A masing masing berniai 0 dan 0,9 dimana ini berarti semua data akan berkisar antara 0 sampai 0,9 setelah dilakukan normalisasi. Batas bawah ditetapkan 0 karena memang data terkecil pada atribut bernilai 0 dan tidak mungkin terdapat data yang lebih kecil, batas atas 0,9 untuk mengantisipasi adanya data yang lebih besar dari data maksimum yang terdapat dalam data set. Dalam Tabel 3.1 terdapat beberapa data yang belum dinormalisasi, sedangkan pada Tabel 3.2 terdapat beberapa data yang telah dinormalisasi.

Tabel 3. 1 Data sebelum preprocessing

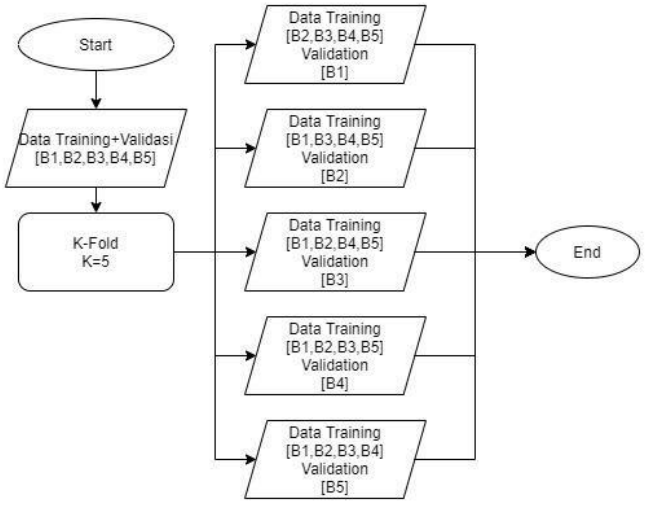
preg	plas	pres	skin	insu	mass	pedi	age	class
6	148	72	35	0	33.6	0.627	50	tested_positive
1	85	66	29	0	26.6	0.351	31	tested_negative
8	183	64	0	0	23.3	0.672	32	tested_positive
1	89	66	23	94	28.1	0.167	21	tested_negative
0	137	40	35	168	43.1	2.288	33	tested_positive

Tabel 3. 2 Data setelah preprocessing

preg	plas	pres	skin	insu	mass	pedi	age	class
0.423529	0.850251	0.57541	0	0	0.642474	0.120154	0.393333	tested_positive
0.370588	0.687437	0.64918	0.4	0	0.670641	0.188471	0.3	tested_positive
0.105882	0.447739	0.383607	0.136364	0.1	0.329955	0.290948	0.1	tested_negative
0.052941	0.492965	0.413115	0.190909	0.143617	0.338003	0.357899	0.126667	tested_negative
0.105882	0.39799	0.545902	0.172727	0.056383	0.388972	0.15158	0.113333	tested_negative

3.3 K-fold validation

Pada umumnya, metode K-fold menggunakan nilai k=10 karena akurasi yang didapatkan akan memiliki variansi relatif rendah [5], berarti data yang digunakan untuk testing adalah 10% dari data semula. Pada penelitian ini sejumlah 80% dari jumlah *dataset* digunakan untuk menjadi data latih dan data validasi dengan menggunakan bantuan K-fold, hal ini bertujuan untuk mendapatkan akurasi yang lebih meyakinkan, karena jika meggunakan nilai k tinggi tentu saja akan menghasilkan akurasi yang baik, namun jika jumlah data uji memiliki porsi yang lebih banyak maka nilai akurasi akan lebih terpercaya meskipun umumnya akan lebih kecil, tetapi ketika diperoleh nilai akurasi yang baik meski menggunakan nilai k yang kecil pemodelan tersebut akan dapat dikatakan baik dan terpercaya. Dalam penelitian ini data latih dan data validasi yang masih belum terpisah dibagi menjadi lima bagian, yang mengartikan bahwa nilai k=5, dan masing-masing bagiannya diwakilkan dengan B1,B2,B3,B4,B5. Nilai K=5 diambil karena pengambilan 20% data sebagai data validasi menjadi porsi yang cukup besar untuk penggunaan data validasi namun tidak banyak mengurangi porsi data latih. Gambaran kinerja K-fold kurang lebih dapat digambarkan pada Gambar 3.2

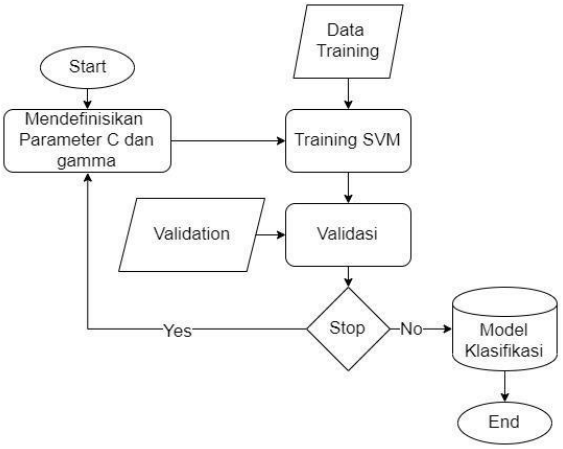


Gambar 3. 2 Alur K-fold

Langkah ini melakukan lima iterasi yang akan membentuk lima set data *training* dan lima set data validasi. Sebagai contoh pada iterasi pertama saat bagian data B1 diberlakukan sebagai data validasi, bagian data lainnya yaitu B2, B3, B4 dan B5 diberlakukan sebagai data *training* hal ini dilakukan secara bergantian sampai semua bagian data telah diberlakukan sebagai data validasi.

3.4 *Training* dan Validasi SVM

Training SVM dilakukan dengan lima set data latih dan lima set data validasi yang diperoleh dari K-fold, pencarian model terbaik dilakukan dengan mengganti-ganti parameter SVM. Model dengan akurasi rata-rata terbaik akan disimpan dalam suatu variabel dan jika ditemukan akurasi yang lebih baik model tersebut akan digantikan. Gambaran kinerja *training* dan validasi SVM kurang lebih dapat digambarkan pada Gambar 3.3



Gambar 3. 3 Alur training dan validasi SVM

Parameter C memiliki keterkaitan dengan margin pada SVM yang dimodelkan, semakin besar nilai C margin dalam model SVM akan semakin kecil, dalam penelitian ini nilai C yang diterapkan adalah C=[0.01, 0.01, 0.1, 1, 10, 100, 1000, 10000], sedangkan parameter gamma yang diterapkan pada penelitian ini adalah gamma(γ)=[0.0001, 0.001, 0.01] parameter gamma sangat berpengaruh terhadap model yang dibangun, karena jika gamma terlalu besar akan banyak area yang terpengatuh oleh gamma sehingga memicu adanya *overfitting* sedangkan jika gamma terlalu kecil model yang dihasilkan akan terlalu terbatas sehingga tidak dapat mencakup kompleksitas atau bentuk data [10]. Penelitian ini mengkombinasikan parameter gamma dan C yang berbeda-beda, kombinasi yang dilakukan dapat dilihat pada Tabel 3.3.

Tabel 3. 3 Kombinasi parameter yang diterapkan

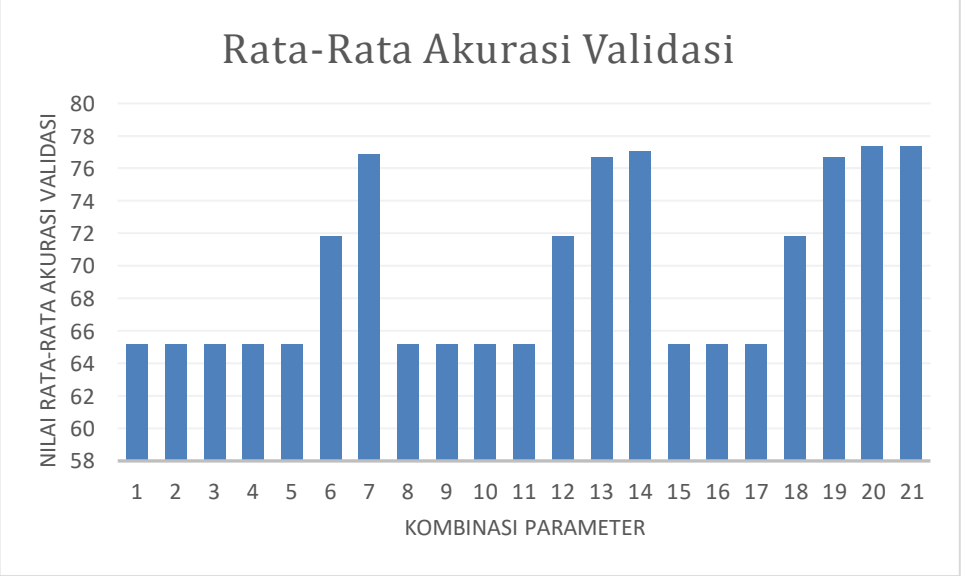
Kombinasi	Gamma	C
1	0.0001	0.01
2	0.0001	0.1
3	0.0001	1
4	0.0001	10
5	0.0001	100
6	0.0001	1000
7	0.0001	10000
8	0.001	0.01
9	0.001	0.1
10	0.001	1
11	0.001	10
12	0.001	100
13	0.001	1000
14	0.001	10000
15	0.01	0.01
16	0.01	0.1
17	0.01	1
18	0.01	10
19	0.01	100
20	0.01	1000
21	0.01	10000

Training dan validasi dilakukan dengan menggunakan lima set data latih dan lima set data validasi yang sebelumnya diperoleh dari hasil K-fold terhadap gamma dan C yang sama, setelah dilakukan training dan validasi oleh kelima data diperoleh akurasi rata-rata untuk gamma dan C yang diaplikasikan. Langkah ini dilakukan secara berulang hingga semua parameter pilihan sudah dicoba. Model dengan akurasi rata-rata terbaik akan disimpan kedalam suatu variabel sebagai hasil dari proses ini.

4. Evaluasi

4.1 Hasil Pengujian

Perolehan rata-rata akurasi validasi divisualisasikan kedalam grafik pada Gambar 4.1. Model terbaik yang diperoleh mencapai akurasi rata-rata sebesar 77.36 dengan gamma: 0,01 dengan C: 1000,0 yang merupakan kombinasi parameter 20. Model ini selanjutnya diuji dengan sejumlah 20% data dari data semula yang disimpan sebagai *data test*(data uji). Dari 154 *data test*, 120 data berhasil diklasifikasi dengan tepat sehingga model ini menghasilkan akurasi sebesar 77,92 %. Hasil observasi lengkap dapat dilihat di lampiran.



Gambar 4. 1 Grafik rata-rata akurasi validasi tiap kombinasi parameter

4.2 Analisis Hasil Pengujian

Langkah-langkah yang dilakukan dalam membangun sistem klasifikasi dapat memberikan akurasi yang baik. Pada tahap *preprocessing* normalisasi data dengan menggunakan *min-max normalization* membantu data untuk memiliki skala teratur pada setiap atributnya, sehingga pemetaan data saat pemodelan akan lebih mudah. Lalu pembentukan data latih dan data validasi yang dilakukan dengan K-fold dapat memberikan referensi data latih yang berbeda dan kemudian ketika data tersebut digunakan saat training SVM dan validasi, peluang memperoleh model klasifikasi bang baik menjadi lebih besar daripada hanya melakukan *training* terhadap satu set data latih saja, selain itu juga perhitungan akurasi rata-rata lebih dapat dipercayai jika dibandingkan dengan hanya membandingkan akurasi menggunakan satu set data latih. Dan untuk memberikan kepastian sebaik apa model yang didapat tahap *testing* diakir dapat memberikan informasi akurasi yang dapat dipercaya.

5. Kesimpulan

Diabetes merupakan penyakit kronis yang serius, dengan membangun sistem klasifikasi yang dapat mengklasifikasi data pasien kedalam kelas positif diabetes atau negatif diabetes, pasien yang belum mengetahui bahwa dirinya termasuk pasien diabetes dapat lebih berhati-hati dan memperoleh pengobatan lebih awal. Namun sistem klasifikasi yang di bangun dalam penelitian ini belum cukup sempurna untuk diaplikasikan secara langsung dalam kehidupan masyarakat secara nyata karena untuk mengatasi kasus nyata terutama dibidang kesehatan diperlukan akurasi yang sangat tinggi agar tidak menyebakan hal-hal yang tidak

diinginkan sementara sistem klasifikasi dalam penelitian ini baru mencapai akurasi 77,92%. Kendati demikian penelitian ini masih dapat dilanjutkan dengan melakukan kiat-kiat yang belum dilakukan pada penelitian ini seperti halnya melakukan *feature selection* agar *training* model lebih berfokus pada atribut tertentu saja yang memang dianggap penting, melakukan penyetaraan kelas pada data *training* sebelum dilakukan proses *training* model atau bahkan mengganti metode klasifikasi.

Daftar Pustaka

[1] World Health Organization, "Global Report on Diabetes," World Health Organization, Geneva, 2016.

[2] S. Karatsiolis and C. N. Schizas , "Region based Support Vector Machine Algorithm for Medical Diagnosis on Pima Indian Diabetes DataSet," in *International Conference on Bioinformatics & Bioengineering (BIBE)*, Larnaca, 2012.

[3] W. Budiharto, Machine Learning & Computational Intelligence, Yogyakarta: Penerbit Andi, 2016.

[4] B. Han and L. S. Davis, "Density-Based Multifeature Background Subtraction with Support Vector Machine," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012.

[5] Suyanto, Machine Learning Tingkat Dasar dan Lanjut, Bandung: Informatika Bandung, 2018.

[6] D. Singh, M. A. Khan, A. Bansal and N. Bansal, "An application of SVM in character recognition with chain code," in *2015 Communication, Control and Intelligent Systems (CCIS)*, Mathura, 2015.

[7] E. Alpaydın, Introduction to Machine Learning, Cambridge: Massachusetts Institute of Technology, 2010.

[8] J. Han, M. Kamber and J. Pei, in *Data Mining Concepts and Techniques*, Morgan Kaufmann, 2012, p. 114.

[9] S. Yadav and S. Shukla, "Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification," in *IEEE*, Bhimavaram, 2016.

[10] scikit-learn, "RBF SVM parameters," scikit-learn developers (BSD License), [Online]. Available: https://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html#sphx-glr-auto-examples-svm-plot-rbf-parameters-py. [Accessed 8 January 2020].

Lampiran

Hasil observasi:

Kombinasi	K=1		K=2		K=3		K=4		K=5		Rata-rata Akurasi Validasi
	Akurasi Train	Akurasi Validasi	Akurasi Train	Akurasi Validasi	Akurasi Train	Akurasi Validasi	Akurasi Train	Akurasi Validasi	Akurasi Train	Akurasi Validasi	
1	66.802	58.537	65.988	61.789	63.544	71.545	63.136	73.171	66.26	60.656	65.13928
2	66.802	58.537	65.988	61.789	63.544	71.545	63.136	73.171	66.26	60.656	65.13928
3	66.802	58.537	65.988	61.789	63.544	71.545	63.136	73.171	66.26	60.656	65.13928
4	66.802	58.537	65.988	61.789	63.544	71.545	63.136	73.171	66.26	60.656	65.13928
5	66.802	58.537	65.988	61.789	63.544	71.545	63.136	73.171	66.26	60.656	65.13928
6	70.061	66.667	72.912	70.732	75.764	76.423	71.487	80.488	72.358	64.754	71.81261
7	78.208	74.797	79.837	73.984	76.986	78.049	76.986	82.114	78.862	75.41	76.87059
8	66.802	58.537	65.988	61.789	63.544	71.545	63.136	73.171	66.26	60.656	65.13928
9	66.802	58.537	65.988	61.789	63.544	71.545	63.136	73.171	66.26	60.656	65.13928
10	66.802	58.537	65.988	61.789	63.544	71.545	63.136	73.171	66.26	60.656	65.13928
11	66.802	58.537	65.988	61.789	63.544	71.545	63.136	73.171	66.26	60.656	65.13928
12	70.061	66.667	72.912	70.732	75.764	76.423	71.487	80.488	72.358	64.754	71.81261
13	78.208	74.797	79.837	73.984	76.986	78.049	76.782	82.114	78.659	74.59	76.70665
14	78.615	75.61	79.022	72.358	78.208	78.862	76.782	82.114	79.675	76.23	77.03452
15	66.802	58.537	65.988	61.789	63.544	71.545	63.136	73.171	66.26	60.656	65.13928
16	66.802	58.537	65.988	61.789	63.544	71.545	63.136	73.171	66.26	60.656	65.13928
17	66.802	58.537	65.988	61.789	63.544	71.545	63.136	73.171	66.26	60.656	65.13928
18	70.061	66.667	72.912	70.732	75.764	76.423	71.487	80.488	72.154	64.754	71.81261
19	78.208	74.797	79.633	73.984	76.986	78.049	76.782	82.114	78.862	74.59	76.70665
20	**78.615**	**75.61**	**79.43**	**73.984**	**78.208**	**79.675**	**76.578**	**82.114**	**80.285**	**75.41**	**77.35839**
21	78.004	77.236	79.43	74.797	79.022	78.862	77.393	82.114	80.691	73.77	77.35572
Akurasi rata-rata tertinggi adalah: 77.3583899773424 dengan gamma: 0.01 dengan C: 1000.0											