

Deteksi Dini Anak Disleksia dengan metode Support Vector Machine

Ardhian Ekawijana, Akhmad Bakhrun*, Zulkifli Arsyad

Teknik Komputer dan Informatika, Teknik Informatika, Politeknik Negeri Bandung, Bandung, Indonesia

Email: ¹ardhian.ekawijana@polban.ac.id, ^{2,*}abakhrun@polban.ac.id, ³zulkifli.arsyad@polban.ac.id

Email Penulis Korespondensi: abakhrun@polban.ac.id

Submitted: 26/09/2022; Accepted: 30/09/2022; Published: 30/09/2022

Abstrak—Disleksia adalah kelainan pada otak diakibatkan genetika. Penderita disleksia dapat hidup normal bahkan memiliki kelebihan tertentu jika mendapatkan pendidikan yang sesuai. Penderita disleksia sering mendapatkan predikat bodoh karena guru yang tidak tahu kasus anak didiknya. Deteksi dini terhadap anak yang disleksia dapat dilakukan dengan serangkaian uji sehingga sistem dapat menyimpulkan bahwa data tersebut masuk disleksia atau tidak. Support Vector Machine adalah metode klasifikasi data untuk membagi hasil uji disleksia atau bukan. Sistem ini dilatih dengan data hasil uji yang sudah tersedia dengan menggunakan metode SVM. Penelitian kali ini menggunakan data gamifikasi untuk mendeteksi anak disleksia atau bukan. SVM membuktikan tingkat akurasi yang bagus dalam prediksi sampai 94 %.

Kata Kunci: Disleksia; SVM; Deteksi; Klasifikasi; Akurasi

Abstract—Dyslexia is a brain disorder caused by genetics. People with dyslexia can live a normal life and even have certain advantages if they get the correct education. People with dyslexia often get the predicate stupid because teachers do not know the case of their students. Early detection of dyslexic children can be done with a series of tests so that the system can conclude that the data is dyslexic or not. Support Vector Machine is a data classification method to share dyslexia test results or not. This system is trained with test results data that are already available using the SVM method. This study uses gamification data to detect dyslexic children or not. SVM proves a good level of accuracy in predictions up to 94%.

Keywords: Dyslexia; Support Vector Machine; Detection; Classification; Accuracy

1. PENDAHULUAN

Artificial Intelligence (AI) adalah salah satu cabang keilmuan di bidang informatika. Artificial intelegent adalah sebuah sistem untuk menjadikan sebuah program cerdas dan dapat meniru kemampuan manusia. Beberapa cabang dalam Artificial intelligence adalah berupa Searching, Reasoning, Planning, dan Learning [1]. AI menjadi hal yang mulai digunakan dalam kehidupan sehari-hari. Keilmuan ini mulai memasuki ke ranah kesehatan, keamanan, kemudahan dan ranah lainnya. Bidang kesehatan dipakai Ilmu ini untuk mendeteksi penyakit-penyakit seperti kanker [2], [3], dan kelainan sperma [4], [5]. Bidang keamanan dipakai untuk pengamanan sistem informasi, deteksi pencuri pada video [6], dan pengamanan kota. Rumah dan gedung juga memakai AI. Rumah dapat memberikan petunjuk menghemat energi dan kemudahan untuk menggantikan aktivitas yang bersifat rutin [7]. Aktifitas rutin tersebut seperti mematikan lampu, menyalakan televisi, dan aktivitas rumah lainnya.

Learning menjadi salah satu cabang dalam pembahasan Artificial Intelligence. Learning menjadi pembahasan yang cukup menarik. Pembahasan ini menjadi sangat aktual karena kebanyakan yang dipakai adalah pembahasan Learning. Salah satu penggunaan learning pada Artificial Intelligence ialah pada Data Mining. Data mining adalah sebuah metode untuk mengubah data menjadi informasi [8]. Pembahasan penelitian ini berkuat tentang data mining dalam menggali informasi pada kasus disleksia. Disleksia adalah kelainan pada otak diakibatkan keturunan. Kelainan ini bisa membawa berkah jika dirawat dengan sesuai. Penanganan yang tidak sesuai akan membuat objeknya menjadi menderita dan tidak keluar kemampuan terbaiknya. Contoh penderita disleksia seperti Einstein, dan Thomas Alpha Edison adalah contoh penderita disleksia dengan penanganan yang baik. Deteksi awal dalam kasus disleksia diperlukan untuk mencegah salah langkah dalam penanganannya.

Deteksi awal untuk kasus disleksia dapat membantu perkembangan anak agar mampu menuju lebih baik. Kesalahan identifikasi dapat menyebabkan anak tertekan dan merasa tidak memiliki potensi seperti anak normal lainnya. Klasifikasi anak apakah disleksia atau tidaknya menggunakan teknik Data Mining agar memiliki akurasi yang baik. Akurasi yang baik menjadikan anak tidak salah penanganan kedepannya [9].

Kasus yang digali adalah klasifikasi informasi pada disleksia dengan menggunakan teknik data mining, dalam hal ini menggunakan Support Vector Machine (SVM). Support Vector Machine dipilih karena memiliki akurasi yang sangat tinggi untuk memodelkan data non linear [8]. Harapan penelitian ini adalah dapat mengklasifikasikan data disleksia menjadi informasi yang berguna untuk penanganan anak-anak sejak dini. Penelitian ini menggunakan data dari penelitian sebelumnya oleh Luz Rello dkk. Data tersebut adalah data rekap hasil gamifikasi yang kemudian dibuat hasil prediksinya menggunakan metode Random Forests. Penelitian kali ini hendak mencari solusi dengan data sama sedangkan metode yang berbeda.

2. METODOLOGI PENELITIAN

Penelitian kali ini dibagi menjadi beberapa tahapan. Tahapan ini menunjang untuk kebutuhan penelitian agar tercapainya tujuan penelitian. Tahapan – tahapan tersebut dapat dilihat pada :



Gambar 1. Proses Mendapatkan Model

Tahapan menemukan pola atau model yang diinginkan adalah tahapan baku pada setiap metode data mining khususnya metode klasifikasi. *Preprocessing* adalah tahapan pertama dalam setiap penelitian. Tahapan ini memastikan keadaan data untuk proses selanjutnya. Kepastian data yang diperlukan harus dipastikan keberadaannya, keadaannya, cara mendapatkannya, dan dipastikan sesuai format yang diinginkan.

2.1 Preprocessing

Proses sebelum melakukan training adalah proses yang bisa berpengaruh terhadap keberhasilan laju training. Proses ini dilakukan sebelumnya tergantung kepada data yang tersedia. Adapun proses yang biasa dilakukan adalah [18]:

- Pembuangan *Outlier***, *Outlier* adalah data yang berbeda dari data lainnya. Data *outlier* biasanya dibuang untuk mempersiapkan proses training karena terletak sangat jauh dengan data kebiasaannya, sehingga akan mempengaruhi kualitas model yang akan terbentuk. *Outlier* yang dibuang biasanya adalah yang memiliki standar deviasi menjangkau hingga 99%. *Outlier* biasanya sedikit namun tetap akan berbahaya terhadap pelatihan. Pendeteksian *outlier* bisa dilakukan dengan metode K-Nearest Neighbor, DBSCAN, *Outlier Removal Clustering* dan lainnya. Situasi *outlier* tidak selamanya dibuang, namun bisa juga digunakan. Contoh penggunaan *outlier* adalah pada fraud detection dan mahasiswa paling berprestasi.
- Normalisasi Data**, Data yang didapatkan pada setiap fitur biasanya terdapat perbedaan. Perbedaan yang dimaksud adalah mengenai rentang nilai yang terlalu jauh. Misal data X memiliki rentang nilai [0..20], namun pada data Y memiliki rentang nilai [2000..10000]. Fitur dengan rentang nilai tinggi bisa jadi lebih mendominasi fitur lainnya sehingga proses klasifikasi tidak terlalu menggambarkan yang sebenarnya.
- Data yang Salah**, Dalam kenyataannya data tertentu biasanya selalu ada nilai salah, kosong dan data yang berbeda formatnya. Masalah ini sering terjadi karena kesalahan input, data kuisisioner yang diisikan tidak lengkap, basis data dan antarmuka aplikasi yang tidak integritasnya.

Adapun yang dilakukan pada penelitian ini adalah melakukan 3 hal tersebut. Pembuangan *outlier* dilakukan setelah mengecek dengan DBSCAN metode. Normalisasi dilakukan karena ada beberapa data yang berbeda sebaran datanya. Data yang salah dilakukan dengan menginput data yang kosong, dan perbedaan format yang menggunakan karakter atau string diubah menjadi format numerik.

2.2 Training

Training adalah proses mencari model yang diinginkan. Akurasi menjadi tolak ukur untuk mencari model yang terbaik dalam pelatihannya. Data *dyslexia* yang dipakai berjumlah 5039 baris dengan 197 column. Klasifikasi yang digunakan adalah 'yes' dan 'no'. Data yang sudah benahi dan dihilangkan sejumlah *outlier*. Data yang sudah baik dipakai untuk proses training. Data training dibagi-bagi menjadi 10 skenario training untuk mencari akurasi terbaik. Data training dipisahkan menjadi 10 cara dengan mengubah data test diambil setiap 500 baris.

2.3 Testing

Testing adalah data diluar data *training* yang sejumlah 500. Data *testing* ini akan dimasukkan ke model yang sudah didapat. Hasil dari input data *testing* kemudian dibandingkan dengan nilai data *testing* yang sebenarnya. Hasil perbandingan tersebut menghasilkan akurasi.

2.4 Dataset

Dataset yang dipakai adalah data yang diambil dari penelitian Luz Rello dkk [11] mengenai *Predicting risk of dyslexia with an online gamified test*. Dataset tersebut memuat data hasil permainan games. Jumlah partisipan dalam penelitian tersebut sejumlah 3644 dengan jumlah dataset yang terbentuk sejumlah 5039 didalamnya terdapat pasien yang sudah terdiagnosa disleksia sebanyak 392. Data pada setiap peserta ada sekitar 196 atribut. Atribut 1 - 4 adalah demografi peserta, sedangkan 5 - 196 adalah performansi. Data tersebut diambil dari peserta saat menjawab 32 pertanyaan. Setiap 1 pertanyaan terdapat informasi *Clicks*, *Hits*, *Misses*, *Score*, *Accuracy*, dan *Missrate*.

Data mining adalah sebuah metode untuk menggali informasi dari data yang ada. Proses data mining erat hubungannya seperti menambang emas. Layaknya penambangan emas, Informasi didapat dari proses panjang dengan berbagai metode yang cocok untuk menemukan informasi yang berguna. Adapun proses dalam data mining adalah [8] :

- Data Cleaning**, adalah proses untuk menghapus data noise, data duplikat, dan data tak berguna.
- Data Integration**, adalah proses menyatukan data dari berbagai sumber dan format yang berbeda.
- Data Selection**, adalah pemilihan data yang relevan terhadap kebutuhan analisisnya.
- Data Transformation**, adalah data diubah menjadi format standar yang disepakati.

- e. **Data Mining**, adalah proses utama dengan memasukan sistem cerdas untuk menghasilkan model atau pattern yang berguna.
- f. **Pattern Evaluation**, adalah proses mengevaluasi model yang dihasilkan apakah mendekati dengan nilai akurasi tertentu.
- g. **Knowledge Presentation**, adalah bagaimana sebuah informasi divisualisasikan dalam sebuah tampilan yang dapat diterima oleh pengguna informasi.

Proses data mining meliputi hal-hal yang telah disebutkan diatas namun dalam pelaksanaannya kita juga melihat data yang ada. Data yang ada sekiranya sudah bagus dan bersih dapat langsung ke langkah berikutnya. Fungsi data mining memiliki berbagai lingkup dalam penelitian ini kita akan menggunakan classification dalam memprediksi kasus disleksia.

2.5 Classification

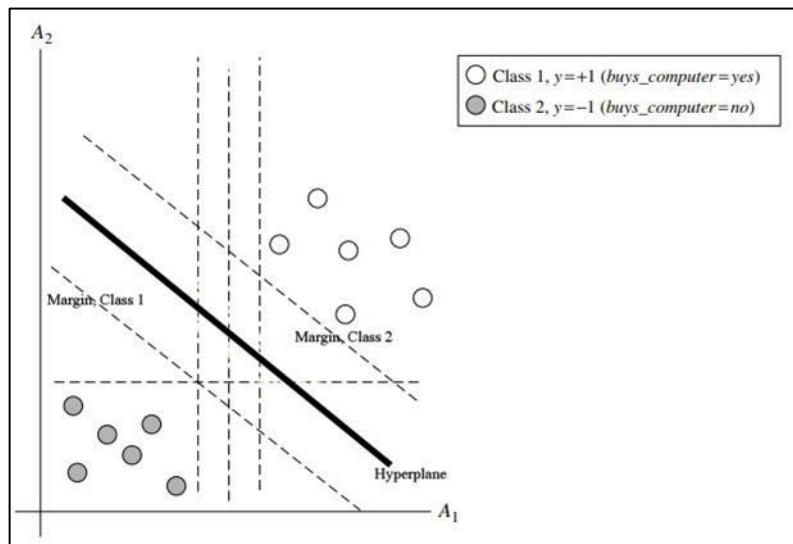
Classification adalah proses untuk menemukan sebuah model atau fungsi yang dapat mendeskripsikan dan menjadi pembeda antara kelas satu dengan kelas yang lainnya. Model didapatkan dengan melakukan proses pelatihan dari data-data yang ada yang telah ada label kelasnya. Model yang didapat lalu digunakan untuk memprediksi kelas tertentu pada data-data yang belum ada label kelasnya [8]. Salah satu metode klasifikasi yang diterapkan pada penelitian ini adalah Support Vector Machine.

2.6 Support Vector Machine

Support Vector Machine adalah sebuah metode untuk klasifikasi baik data linier maupun nonlinier. Metode ini digunakan untuk memetakan data yang ada kepada dimensi yang lebih tinggi. Dalam dimensi baru ini mencari hyperplane pemisah yang optimal linier. SVM menemukan pemisah hyperplane menggunakan vektor pendukung.

2.7 Support Vector Machine dengan data linear

Setiap data latih dinyatakan oleh (X_i, Y_i) dengan $i = 1, 2, \dots, N$, dan $x_i = \{x_{i1}, x_{i2}, \dots, x_{iq}\}^T$ merupakan atribut (fitur) set untuk data latih ke- i . Untuk $y_i \in \{-1, +1\}$ menyatakan label kelas. Hyperplane klasifikasi linear SVM seperti gambar berikut :



Gambar 2. Margin Hyperlane [8]

Gambar 1 menjelaskan pemisahan antara 2 data yang berbeda dengan menggunakan *Hyperplane*. Jarak yang besar antara *Hyperplane* dan setiap kelompok data menyatakan semakin mampu data di klasifikasikan dengan baik. *Class 1* adalah kelompok data yang membeli komputer, sedangkan *Class 2* adalah kelompok yang tidak membeli komputer. Rumus untuk memisahkan *hyperplane* dapat dilihat sebagai berikut :

$$W \cdot X + b = 0 \tag{1}$$

W adalah *weight* vektor dinamakan $W = \{w_1, w_2, \dots, w_n\}$; dimana n adalah jumlah dari atribut. Dan b adalah skalar, biasa disebut bias. Agar lebih memahami persamaan tersebut, kita gunakan contoh 2 atribut. A_1 dan A_2 . Baris dari *training* adalah 2-D (e.g., $X = (x_1, x_2)$), dimana x_1 dan x_2 adalah atribut dari A_1 dan A_2 untuk X . Jika kita melihat ada bias maka kita bisa sebut sebagai w_0 , maka persamaannya menjadi sebagai berikut :

$$w_0 + w_1x_1 + w_2x_2 = 0 \tag{2}$$

sehingga semua titik yang ada berada diatas *hyperplane* adalah sebagai berikut :

$$w_0 + w_1x_1 + w_2x_2 > 0 \tag{3}$$

hampir sama, semua titik yang ada dibawah *hyperplane* adalah :

$$w_0 + w_1x_1 + w_2x_2 < 0 \tag{4}$$

Bobot dapat disesuaikan sehingga *hyperplane* didefinisikan sisi sebagai margin adalah :

$$H1 : w_0 + w_1x_1 + w_2x_2 \geq 1 \text{ for } y_i = +1 \tag{5}$$

$$H2 : w_0 + w_1x_1 + w_2x_2 \leq -1 \text{ for } y_i = -1 \tag{6}$$

Artinya adalah setiap data yang diatas H1 maka termasuk kelas +1 dan setiap data yang dibawah H2 adalah termasuk kelas -1. Kombinasi dari kedua persamaan diatas adalah:

$$y_i(w_0 + w_1x_1 + w_2x_2) \geq 1, \forall i. \tag{7}$$

Setiap data yang berada tepat pada H1 atau H2 yang memenuhi persamaan tersebut adalah disebut **Support Vector**.

2.8 Disleksia

Membaca dan menulis adalah pekerjaan yang kompleks. Otak kita membutuhkan untuk menghubungkan antara suara dan huruf, menempatkan suara pada tempat yang benar sesuai dengan tulisan, menyatukan kata-kata menjadi suatu kalimat dan paragraf yang dapat kita baca, tuliskan dan sama-sama memahami.

Orang dengan disleksia mengalami kesulitan dalam mencocokkan antara huruf yang mereka lihat dan suara yang harus keluar dari mereka. Ada masalah dengan mencocokkan tersebut menjadikan kesannya bisa lebih sulit. Anak-anak dan orang dewasa yang kesulitan dalam membaca dan berjuang untuk melakukan hal tersebut bukan berarti mereka bodoh. Kecerdasan mereka tidak ada hubungannya dengan kesulitan membaca dan menulis. Disleksia ada kekurangan dalam membaca yang lambat namun mereka adalah pemikir yang cepat dan kuat serta memiliki kreatifitas dan penalaran yang kuat.

Disleksia adalah penyakit turunan dan tidak akan bisa disembuhkan. Disleksia ada dalam 20% populasi manusia dan mewakili 80-90 % dari semua orang yang mengalami kesulitan dalam belajar di sekolah. Disleksia dengan segala kelemahan, namun juga memiliki kelebihan lainnya bahkan bisa sukses di masa depan [10].

2.9 Pentingnya Deteksi Awal Disleksia

Disleksia merupakan gangguan dalam komponen fonologi bahasa yang sering tak terduga. Otaknya seringkali tidak mampu mengidentifikasi antara kata dan suara yang harus keluar secara akurat. Jika seseorang mengetahui lebih dini akan disleksianya maka dia akan mengambil strategi yang tepat dalam pembelajarannya, namun jika orang tersebut tidak diberikan dukungan yang tepat maka seringkali bisa terjadi gagal sekolah, 35% putus sekolah, dan diperkirakan kurang dari 2 % sarjana yang disleksia tidak bisa lulus tepat waktu 4 tahun [11].

Beberapa penelitian memperlihatkan dan mencoba mendeteksi anak-anak dengan disleksia. Penelitian juga menentukan tingkat kognitif siswa tersebut dan menentukan defisit kemampuan yang terjadi pada usianya. Ada setidaknya 3 tahapan dalam penanganan siswa yang mengalami disleksia. Tahapan tersebut adalah [12] :

- a. Deteksi siswa yang terkena dampak,
- b. Penilaian kemampuan kognitif
- c. Penerapan program bantuan.

2.10 Penelitian Terdahulu

Penelitian tentang disleksia sudah banyak dilakukan oleh peneliti lainnya. Penelitian ini diinspirasi oleh penelitian Luz Rello dkk. Penelitian tersebut untuk memprediksi disleksia dengan gamifikasi. Penelitian ini menggunakan data dari Luz Rello, namun metode diganti dengan SVM. Adapun penelitian sebelumnya mengenai penelitian sejenis ialah ; Penggunaan Gamifikasi untuk memperoleh data test pada anak. Dataset tersebut yang digunakan untuk prediksi menggunakan metode Random Forests [11]. Hasil prediksi model random forest memperoleh akurasi 89,2%, karena 89,2% peserta tidak menderita disleksia. Gambaran tentang manfaat dan tantangan yang akan ada terkait dengan identifikasi awal disleksia [13]. Identifikasi awal saja tidak cukup-menangani anak disleksia. Oleh karena itu, pemantauan kemajuan yang berkelanjutan dan akses ke program intervensi yang sesuai sangat penting. Peninjauan algoritma pembelajaran Machine learning dan Deep Learning yang ada yang telah diterapkan untuk disleksia dan deteksi biomarker [14]. Hasil dari semua metode pembelajaran mesin yang ditinjau adalah menjanjikan dan membuktikan fakta bahwa, disleksia adalah gangguan heterogen yang disebabkan oleh perubahan karakteristik jaringan otak. Penelitian selanjutnya ini fokus pada membantu guru di sekolah dasar upaya untuk mengembangkan desain pembelajaran dan mengembangkan beberapa adaptasi sumber edukatif untuk mengajar membaca bagi anak dengan ketidakmampuan belajar seperti disleksia. Selain itu, aspek pedagogis tidak eksplisit, ini menjadi sulit bagi seorang guru yang mencoba untuk memecahkan masalah disleksia siswa mereka [15].

Kontribusi utamanya adalah klasifikasi pola pendidikan menurut empat tingkat membaca tata bahasa utama seperti: pra-suku kata, suku kata, pra-abjad dan abjad. Tingkat pra-suku kata mengelompokkan kembali praktik terbaik tentang penggunaan aplikasi pendidikan untuk membantu siswa mengidentifikasi setiap alfabet. Penelitian ini fokus pada hubungan antara penderita disleksia mampu menjadi programmer yang baik [16]. Keluaran dari penelitian ini adalah adanya hubungan antara disleksia dan kemampuan menjadi programmernya.

Penelitian ini fokus pada kebaruan metode *Sparse Balanced Support Vector Machine* (SB-SVM) yang diterapkan pada kasus Diabetes [17]. Pendekatan SB-SVM yang diusulkan menunjukkan kestabilan terbaik antara kinerja prediktif dan waktu komputasi. SB-SVM mampu mengelola data dimensi tinggi, denganmeningkatkan interpretasi model dan menemukan yang paling banyak fitur yang relevan saat berhadapan dengan kelas tidak seimbang yang biasa distribusi.

3. HASIL DAN PEMBAHASAN

Tahapan pertama dari proses penelitian ini adalah *preprocessing*. *Preprocessing* data dilakukan agar data yang digunakan sesuai yang diharapkan dan lebih homogen sehingga memudahkan proses *training*.

3.1 Preprocessing

Preprocessing adalah proses awal yang dilakukan sebelum melakukan training model. Preprocessing kali ini akan dilakukan tiga tahapan yaitu konversi nilai dari *string* ke numerik, normalisasi dan membuang data pencilan / *Outlier*.

a. Perbaikan Data Dari Data Salah

Data yang ada beberapa ada yang masih bernilai *String* dan perlu konversi ke bentuk decimal atau numerik. Data *String* diisi diperbaiki dengan cara mengisi nilai 0 dan 1 berikut contoh perbaikannya :

Gender	Nativelang	Otherlang		Gender	Nativelang	Otherlang
Male	No	Yes	➔	1	0	1
Female	Yes	Yes		0	1	1
Female	No	Yes		0	0	1
Female	No	Yes		0	0	1
Female	No	Yes		0	0	1
Female	No	Yes		0	0	1
Female	No	Yes		0	0	1
Female	No	No		0	0	0
Female	No	Yes		0	0	1

Gambar 1. Konversi Data Penelitian

Gambar 3 menjelaskan konversi data dari bentuk *string* kedalam bentuk numerik. Atribut *gender* mengubah nilai *Male* menjadi 1 dan *Female* menjadi 0. Atribut *Nativelang* diubah menjadi *Yes* bernilai 1 dan *No* bernilai 0 dan terakhir atribut *Otherlang* asalnya *Yes* menjadi 1 dan *No* menjadi 0.

b. Normalisasi Data

Normalisasi data dilakukan agar setiap nilai memiliki nilai rentang yang sama. Normalisasi yang dipilih adalah pada rentang [0,1] dikarenakan nilai yang ada rata-rata terjadi direntang tersebut. Berikut data hasil normalisasinya :

Age	Clicks1	Hits1	λ	Age	Clicks1	Hits1
7	10	10	➔	0	0.119047619	0.5
13	12	12		0.6	0.142857143	0.6
7	6	6		0	0.071428571	0.3
7	0	0		0	0	0
8	4	4		0.1	0.047619048	0.2
8	9	9		0.1	0.107142857	0.45
8	6	6		0.1	0.071428571	0.3
9	4	4		0.2	0.047619048	0.2
9	4	1		0.2	0.047619048	0.05

Gambar 2. Normalisasi Data

Gambar 4 menunjukkan adanya normalisasi data dari data yang memiliki jangkauan data yang berbeda-beda menjadi data yang memiliki jangkauan data antara 0 – 1. Perubahan jangkauan data dilakukan agar tidak ada ketimpangan jangkauan data antara atribut satu dan lainnya.

c. Membuang Data Pencilan atau Outlier

Mencari data pencilan dilakukan dengan menggunakan DBSCAN. Beberapa scenario pengujian pencarian outlier dapat dilihat pada :

Tabel 1. Data deteksi Outlier

No	Min Data	Persen	Epsilon	Data_Clean	Outlier	Total Data
1	5	0,02%	4	5038	1	5039
2	5	0,16%	3	5031	8	5039
3	5	1,01%	2	4988	51	5039
4	5	4,54%	1.5	4810	229	5039
5	5	5,04%	1.474	4785	254	5039
6	5	30,92%	1	3481	1558	5039
7	5	90,93%	0.5	457	4582	5039

Tabel 1 digunakan untuk mencari nilai outlier sebanyak 5 % atau yang mendekati. Pembuangan data outlier biasanya dibuang sebanyak 5 persen. Pembuangan data dilakukan agar model yang didapat lebih memahami seharusnya data. Normalnya data akan memandu sistem yang didapat lebih mudah

3.2 Training

Proses Training adalah proses untuk mendapatkan model yang baik untuk digunakan. Proses Training kali ini digunakan proses pembagian data training dan testing yang berbeda-beda untuk melihat seberapa efektif data yang berbeda mempengaruhi bagus tidaknya suatu model data. Berikut pembagian data training untuk mendapatkan model data :

Tabel 2. Pembagian Data Training dan Testing

No	Data Training	Data Test
1	Baris 501 - 4785	Baris 1 - 500
2	Baris 1 - 500 dan Baris 1001- 4785	Baris 501 - 1000
3	Baris 1 - 1000 dan Baris 1501- 4785	Baris 1001 - 1500
4	Baris 1 - 1500 dan Baris 2001- 4785	Baris 1501 - 2000
5	Baris 1 - 2000 dan Baris 2501- 4785	Baris 2001 - 2500
6	Baris 1 - 2500 dan Baris 3001- 4785	Baris 2501 - 3000
7	Baris 1 - 3000 dan Baris 3501- 4785	Baris 3001 - 3500
8	Baris 1 - 3500 dan Baris 4001- 4785	Baris 3501 - 4000
9	Baris 1 - 4000 dan Baris 4501- 4785	Baris 4001 - 4500

Tabel 2 menjelaskan pembagian data Training dan data Testing. Pembagian data yang berbeda-beda tersebut untuk mengetahui adanya perubahan akurasi dari setiap data training yang akan diujikan.

3.3 Testing

Testing adalah proses terakhir dari data mining. Proses ini menguji model yang didapat apakah memiliki kemampuan prediksi yang baik atau tidak. Proses ini dilakukan dengan membandingkan data testing aktual yang tidak dimasukan ke data training dengan data hasil prediksi model yang menggunakan metode SVM atau *Support Vector Machine*. Berikut gambar yang menjelaskan hasil dari percobaan tersebut :

```

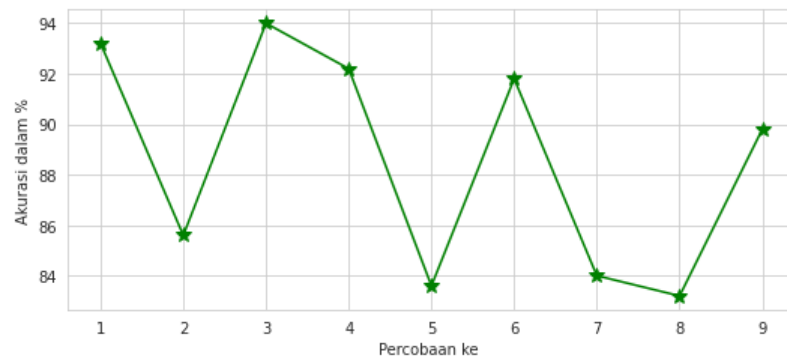
x = Data Attribut test
  Gender  Nativelang  Otherlang  Age  Clicks1  Hits1  Misses1  Score1  \
4225      1          1           0  0.3  0.095238  0.40      0.0  0.16
4226      1          1           0  0.3  0.119048  0.50      0.0  0.20
4227      0          1           0  0.3  0.071429  0.30      0.0  0.12
4228      1          1           0  0.3  0.083333  0.35      0.0  0.14
4229      0          1           0  0.3  0.095238  0.40      0.0  0.16

y_pred = Hasil data x dimasukan ke model
['No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No'
 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No'
 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No'
 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No'
 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No']

Data y yang sebenarnya
['No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No'
 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No'
 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No'
 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No'
 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No' 'No']
  
```

Gambar 3. Hasil Testing Model

Gambar 5 adalah hasil dari testing data dengan atribut sejumlah 196. Atribut digambar tidak dimasukan semua karena keterbatasan tempat. x adalah data matriks yang berisi data yang sudah dibersihkan dari *outlier* sejumlah 196 atribut kesamping. Baris yang ada sejumlah 4785. Data y_pred adalah data hasil prediksi hasil memasukan data x ke metode SVM di pemrograman pyhton. Data y_pred berisi yes jika disleksia dan no jika bukan. Data y_pred berisi 4785 data Yes atau No. Data y adalah data sebenarnya dari kondisi tersebut. Akurasi adalah membandingkan antara kesamaan data y_pred dengan data y . Gambar diatas menunjukkan akurasi 100 persen karena tidak memuat semua informasi dari y_pred dan y . Kasus diatas sebenarnya memiliki akurasi sebesar 93.2. Penelitian kali ini kami mencoba melakukan prediksi dengan 9 kasus percobaan seperti pada tabel 4. Adapun hasil dari akurasi percobaannya adalah :



Gambar 4. Hasil Akurasi Pada Beberapa Percobaan

Gambar 6 adalah hasil dari pengukuran akurasi setelah dibandingkan antara data prediksi dengan data aktualnya. Tidak ada korelasi antara data keberapa yang menjadi data testingnya, namun tetap ketika data yang digunakan untuk testing dan training berbeda ada perbedaan nilai akurasi yang terjadi pada model yang menggunakan SVM. Model yang terbaik adalah yang memiliki akurasi tertinggi yaitu pada percobaan ke 3 yang memiliki akurasi 94 %, sedangkan model dengan akurasi terendah terdapat pada percobaan ke 8 dengan nilai akurasi 83,2 %. Akurasi jaungkauannya ada pada 83,2 % sampai 94 %, dengan rata-rata sebesar 88,6 %.

4. KESIMPULAN

Penelitian kali ini adalah untuk memprediksi atau mendeteksi anak disleksia sedini mungkin. Penelitian kali ini menggunakan data dari PLUS ONE berupa data hasil rekup uji anak dengan gamifikasi. Penelitian sebelumnya menggunakan data tersebut untuk membuat model prediksi dengan metode Random Forest. Penelitian kali ini kami mencoba menggunakan metode Support Vector Machine. Kami mencoba dengan metode SVM mendapatkan akurasi hasil tertinggi pada 94 % dengan menggunakan percobaan ketiga, sedangkan memiliki rata-rata dari 9 percobaan sebesar 88,6%. Penelitian kami memperoleh perbaikan akurasi pada kasus tertentu sampai 94%.

REFERENCES

- [1] S. J. Russell and P. Norvig, *Artificial intelligence*, vol. 4. Pearson Education, Inc., 2010.
- [2] N. Lee, G. Márquez, J. M. Levsky, and J. K. Gohagan, "Potential of Computer-Aided Diagnosis to Improve CT Lung Cancer Screening," *IEEE Rev. Biomed. Eng.*, vol. 2, no. 2008, pp. 136–146, 2009, doi: 10.1109/RBME.2009.2034022.
- [3] Y. Wang *et al.*, "Deeply-Supervised Networks with Threshold Loss for Cancer Detection in Automated Breast Ultrasound," *IEEE Trans. Med. Imaging*, vol. 39, no. 4, pp. 866–876, 2020, doi: 10.1109/TMI.2019.2936500.
- [4] P. Hidayatullah, T. L. E. R. Mengko, R. Munir, and A. Barlian, "Bull Sperm Tracking and Machine Learning-Based Motility Classification," *IEEE Access*, vol. 9, pp. 61159–61170, 2021, doi: 10.1109/ACCESS.2021.3074127.
- [5] S. Qi, T. Nie, Q. Li, Z. He, D. Xu, and Q. Chen, "A Sperm Cell Tracking Recognition and Classification Method," *Int. Conf. Syst. Signals, Image Process.*, vol. 2019-June, pp. 163–167, 2019, doi: 10.1109/IWSSIP.2019.8787312.
- [6] R. Kakadiya, R. Lemos, S. Mangalan, M. Pillai, and S. Nikam, "AI Based Automatic Robbery/Theft Detection using Smart Surveillance in Banks," *Proc. 3rd Int. Conf. Electron. Commun. Aerosp. Technol. ICECA 2019*, pp. 201–204, 2019, doi: 10.1109/ICECA.2019.8822186.
- [7] F. Fanitabasi and E. Pournaras, "Appliance-Level Flexible Scheduling for Socio-Technical Smart Grid Optimization," *IEEE Access*, vol. 8, pp. 119880–119898, 2020, doi: 10.1109/ACCESS.2020.3001763.
- [8] J. Han, M. Kamber, and J. Pei, *Data mining: Data mining concepts and techniques*. 2012.
- [9] Loeziana, "Urgensi Mengenal Ciri Disleksia," *J. Pendidik. Kegur.*, vol. 3, no. 2, pp. 42–58, 2017, [Online]. Available: <http://jurnal.ar-raniry.ac.id/index.php/bunayya/article/download/1698/1235>.
- [10] S. Shaywitz, "What is Dyslexia?," *The Yale Center For Dyslexia & Creativity*, 2022. <https://dyslexia.yale.edu/dyslexia/what-is-dyslexia/> (accessed Sep. 23, 2022).
- [11] L. Rello, R. Baeza-Yates, A. Ali, J. P. Bigham, and M. Serra, "Predicting risk of dyslexia with an online gamified test," *PLoS One*, vol. 15, no. 12 December, pp. 1–15, 2020, doi: 10.1371/journal.pone.0241687.
- [12] C. Mejia, B. Florian, R. Vatrapu, S. Bull, S. Gomez, and R. Fabregat, "A Novel Web-Based Approach for Visualization and Inspection of Reading Difficulties on University Students," doi: 10.1109/TLT.2016.2626292.



- [13] D. Colenbrander, J. Ricketts, and H. L. Breadmore, “Early identification of dyslexia: Understanding the issues,” *Lang. Speech. Hear. Serv. Sch.*, vol. 49, no. 4, pp. 817–828, 2018, doi: 10.1044/2018_LSHSS-DYSLC-18-0007.
- [14] O. L. Usman, R. C. Muniyandi, K. Omar, and M. Mohamad, “Advance Machine Learning Methods for Dyslexia Biomarker Detection: A Review of Implementation Details and Challenges,” *IEEE Access*, vol. 9, pp. 36879–36897, 2021, doi: 10.1109/ACCESS.2021.3062709.
- [15] T. S. Kuswiyanti, S. Santoso, and F. Indriyani, “Aplikasi Pengenalan Profesi pada Anak Usia Dini Berbasis Android,” *Acad. J. Comput. Sci. Res.*, vol. 2, no. 2, pp. 2–6, 2020, doi: 10.38101/ajcsr.v2i2.288.
- [16] J. Stein, “Why dyslexics make good coders,” *Itnow*, vol. 60, no. 3, pp. 58–60, 2018, doi: 10.1093/itnow/bwy081.
- [17] M. Bernardini, L. Romeo, P. Misericordia, and E. Frontoni, “Discovering the Type 2 Diabetes in Electronic Health Records Using the Sparse Balanced Support Vector Machine,” *IEEE J. Biomed. Heal. Informatics*, vol. 24, no. 1, pp. 235–246, 2020, doi: 10.1109/JBHI.2019.2899218.
- [18] E. Prasetyo, *Data Mining, Mengolah Data Menjadi Informasi Menggunakan Matlab*, 1st ed. Yogyakarta: Andi Yogyakarta, 2014.