



## Prediksi Kualitas Air Sungai Menggunakan Metode Pembelajaran Mesin: Studi Kasus Sungai Ciliwung

### Prediction of River Water Quality Using Machine Learning Methods: Ciliwung River Case Study

MOHAMMAD HAEKAL\*, WAHYU CATUR WIBOWO

Program Magister Ilmu Komputer, Fakultas Ilmu Komputer, Universitas Indonesia, Depok, 16424

\*mohammad.haekal02@ui.ac.id

#### ARTICLE INFO

##### Article history:

Received 7 June 2023

Accepted 24 July 2023

Published 31 July 2023

##### Keywords:

Online Monitoring

Artificial Neural Networks

Support Vector Machines

Random Forests

Naive Bayes

#### ABSTRACT

The Ciliwung River is the largest river in Jakarta area, and its water quality tends to decline. To determine river water quality from time to time, real-monitoring is carried out in real time using Online Monitoring (ONLIMO) technology by installing a multiprobes sensor in the Ciliwung River. Eight parameters were monitored, including pH, Dissolved Oxygen, Nitrate, Turbidity, Total Dissolved Solids, Salinity, Electrical Conductivity, and Temperature. In this study, a data science approach using four Machine Learning models, namely Artificial Neural Networks (ANN), Support Vector Machine (SVM), Random Forest, and Naive Bayes, was used to process monitoring data for one year, from January 1 until December 31, 2018, to predict the Ciliwung River water quality. The ANN results showed that the 5-Hidden Layers model has the highest value of Confusion Matrix (Accuracy, Recall, Precision, Specificity, and F1 Score) compared with all tested Hidden Layer of architecture models. In addition, the Random Forest method has the highest confusion matrix value, followed by the ANN method with 5 Hidden Layers, the Naive Bayes method, and the SVM method. The accuracy value of the first three methods is very high, above 89%. Thus, the first three methods work well to predict the quality of Ciliwung River water.

#### INFORMASI ARTIKEL

##### Histori artikel:

Diterima 7 Juni 2023

Disetujui 24 Juli 2023

Diterbitkan 31 Juli 2023

##### Kata kunci:

Online Monitoring

Jaringan Syaraf Tiruan

Support Vector Machine

Random Forest

Naive Bayes

#### ABSTRAK

Sungai Ciliwung merupakan sungai terbesar di wilayah Jakarta, kualitas airnya cenderung menurun. Untuk mengetahui kualitas air sungai dari waktu ke waktu dilakukan pemantauan secara *real time* menggunakan teknologi *Online Monitoring* (ONLIMO) dengan memasang *multiprobes sensor* di Sungai Ciliwung. Delapan parameter yang dipantau meliputi pH, Oksigen Terlarut, Nitrat, Kekeruhan, Total Padatan Terlarut, Salinitas, Konduktivitas Listrik, dan Suhu. Pada penelitian ini, pendekatan sains data menggunakan empat model *Machine Learning* yaitu Jaringan Syaraf Tiruan (JST), *Support Vector Machine* (SVM), *Random Forest*, dan *Naive Bayes* digunakan untuk mengolah data pemantauan selama satu tahun, dari 1 Januari hingga 31 Desember 2018 untuk memprediksi kualitas air sungai Ciliwung. Hasil metode JST menunjukkan bahwa model dengan 5 *Hidden Layer* memiliki nilai *Confusion Matrix* (*Accuracy*, *Recall*, *Precision*, *Specificity*, dan *F1 Score*) tertinggi dibandingkan dengan semua model arsitektur *Hidden Layer* yang diuji. Selain itu, metode *Random Forest* memiliki nilai *Confusion Matrix* tertinggi, diikuti metode JST dengan 5 *Hidden Layer*, metode *Naive Bayes*, dan terakhir metode SVM. Tingkat akurasi ketiga metode pertama sangat tinggi, yaitu di atas 89%. Dengan demikian, ketiga metode pertama bekerja dengan baik untuk memprediksi kualitas air Sungai Ciliwung.

## 1. PENDAHULUAN

### 1.1 Latar Belakang

Sungai Ciliwung memiliki arti penting bagi kota Jakarta karena keberadaannya membentang dari selatan sampai ke utara kota Jakarta, bermuara ke teluk Jakarta. Berdasarkan Peraturan Pemerintah No. 82 Tahun 2001, sungai Ciliwung termasuk sungai kelas dua yang peruntukannya antara lain untuk sarana rekreasi air, budidaya perikanan, dan mengairi pertamanan. Selama ini cara untuk mengetahui kualitas air sungai masih dilakukan secara konvensional yaitu dengan mengambil air sungai langsung ke lokasi, kemudian dibawa ke laboratorium untuk dianalisis (Standar Nasional Indonesia, 2008). Cara ini memerlukan waktu cukup lama dan data yang dihasilkan juga terbatas pada waktu tertentu saja.

Pada tahun 2018, untuk melihat perubahan kualitas air Sungai Ciliwung dari waktu ke waktu, Pusat Teknologi Lingkungan (PTL), Badan Pengkajian dan Penerapan Teknologi (BPPT) – sekarang Badan Riset dan Inovasi Nasional (BRIN), melakukan penelitian pemantauan kualitas air Sungai Ciliwung menggunakan teknologi “*Online Monitoring*”, disingkat dengan “ONLIMO”. Ada delapan parameter yang dipantau yaitu pH, Oksigen Terlarut, Nitrat, Kekeruhan, Total Padatan Terlarut, Salinitas, Konduktivitas Listrik, dan Suhu. Hasil pemantauan dikirim ke pusat data setiap satu jam, selama 24 jam dalam satu hari. Jumlah data pemantauan dalam 1 tahun, dari 1 Januari sampai 31 Desember 2018 ada 5.476 data. Data hasil pemantauan ini belum dimanfaatkan untuk memprediksi kualitas air sungai. Damayanti *et al.*, (2022) menggunakan data hasil pemantauan ONLIMO untuk menganalisis status mutu air danau Toba di Sumatera Utara, mengacu kepada Keputusan Menteri Negara Lingkungan Hidup No. 115 Tahun 2003 tentang Status Mutu Air.

Untuk memprediksi kualitas air Sungai Ciliwung, data hasil pemantauan diolah dengan pendekatan Sains Data menggunakan model *Machine Learning* dengan metode Jaringan Syaraf Tiruan (JST), *Support Vector Machine* (SVM), *Random Forest*, dan *Naive Bayes*. Han *et al.*, (2012), dalam penelitiannya menjelaskan bahwa Sains Data dapat digunakan untuk mengolah data dalam jumlah besar menjadi lebih sederhana, terstruktur, dan mudah dipahami.

Di kota besar seperti di Jakarta, buangan air limbah domestik memberi kontribusi utama pada pencemaran air sungai (JICA, 1990). Trofisa (2011) menyebutkan sumber pencemaran air sungai disamping air limbah domestik juga air limbah industri. Disamping itu, perubahan pola tutupan lahan di daerah aliran sungai (DAS) juga menjadi penyebab turunnya kualitas air sungai (Alphayo *et al.*, 2018). Hasil penelitian Patricia *et al.* (2018) menjelaskan konsentrasi Nitrat dan Fosfat dalam air sungai Ciliwung cukup tinggi, masing-masing bervariasi antara 2,28 hingga 5,66 mg/l dan 1,03 hingga 10,87 mg/l. Untuk mengatasi masalah pencemaran, pemerintah telah menerbitkan Keputusan Menteri Lingkungan Hidup Nomor 5 Tahun 2014 tentang kewajiban industri untuk mengolah air limbah sampai memenuhi Baku Mutu yang ditetapkan. Khusus di wilayah provinsi DKI Jakarta, Gubernur Pemerintah Provinsi DKI Jakarta melalui Keputusan Gubernur Nomor 122 Tahun 2005 juga mewajibkan pihak usaha kegiatan untuk mengolah air limbah domestik sebelum dibuang ke badan air seperti sungai.

Najah *et al.*, (2011) menggunakan metode JST yaitu *Multilayer Perceptron Artificial Neural Networks* (MLP-ANN) dan *Ensemble Artificial Neural Networks* (E-ANN) untuk memprediksi kualitas air sungai dan membandingkannya dengan metode SVM dengan mengambil parameter Oksigen Terlarut (DO), Kebutuhan Oksigen Biokimia (BOD), dan Kebutuhan Oksigen Kimia (COD). Mereka melaporkan metode SVM mengungguli metode yang lain, di mana persentase distribusi kesalahan SVM tidak melebihi 5% untuk DO, BOD dan COD, sementara yang lain di atas 15%. Rizal *et al.*, (2022) dalam penelitiannya mengembangkan model JST untuk memprediksi kualitas air Sungai Langat, Malaysia menggunakan parameter Kebutuhan Oksigen Biokimia (BOD), Total Padatan Terlarut (TDS), Padatan Terlarut (DS), Total Padatan (TS), Nitrat, dan Fosfat. Pada penelitian tersebut diperoleh nilai koefisien ( $R^2$ ) yang tinggi yaitu antara 0,9906 sampai 0,9998, dengan nilai *training data* dan *testing data* berkisar antara 0,8797 sampai 0,9972. Al-Adhaileh *et al.*, (2021) melakukan pemodelan dalam memprediksi kualitas air dengan menggunakan Kecerdasan Buatan. Parameter yang digunakan ada tujuh, yaitu Oksigen Terlarut (DO), pH, *Conductivity*, Kebutuhan Oksigen Biokimia (BOD), Nitrat, *Fecal Coliform*, dan *Total Coliform*. Hasil penelitian menunjukkan prediksi kesalahan yang sangat rendah yaitu 0,00336 untuk nilai *Mean Square Error* (MSE), dan nilai korelasi regresi (R) yang tinggi yaitu 96,17%. Ben Hassen *et al.*, (2020) melakukan kajian prediksi parameter kualitas air irigasi menggunakan model *Machine Learning*. Dari 574 data yang dipakai, persentase yang benar untuk memprediksi kualitas air irigasi menggunakan model *Machine Learning* sebesar 73,34%, sedangkan presentasi yang salah mencapai 26,66%.

### 1.2 Tujuan Penelitian

Tujuan dari penelitian ini adalah membangun model prediksi kualitas air Sungai Ciliwung menggunakan data pemantauan ONLIMO memakai beberapa model Pembelajaran Mesin, yaitu Jaringan Syaraf Tiruan (JST), *Support Vector Machine* (SVM), *Random Forest*, dan *Naive Bayes*. Parameter yang digunakan ada 8, yaitu pH, Oksigen Terlarut (DO), Nitrat, Kekeruhan, Jumlah Padatan Terlarut (TDS), Konduktivitas Listrik (DHL), Salinitas, dan Suhu. Di Indonesia, studi prediksi kualitas air sungai Ciliwung menggunakan metode-metode di atas belum pernah dilakukan.

## 2. METODE

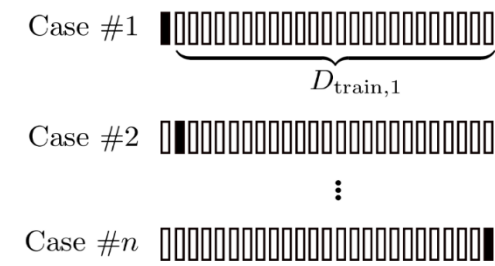
### 2.1 Lokasi dan Pengumpulan Data

Data pemantauan kualitas air Sungai Ciliwung diperoleh dari Pusat Teknologi Lingkungan (PTL), Badan Pengkajian dan Penerapan Teknologi (BPPT) - sekarang Badan Riset dan Inovasi Nasional (BRIN), yaitu data hasil pemantauan dalam satu tahun dari tanggal 1 Januari sampai 31 Desember tahun 2018. Jumlah data hasil pemantauan ada 5.476 data. Pemantauan dilakukan dengan teknologi “*Online Monitoring*” atau disingkat “ONLIMO”. Metode pengukuran menggunakan *multiprobes sensor* yang dipasang di aliran sungai Ciliwung di depan Masjid Istiqlal, Jakarta Pusat (Wahjono, 2016). Hasil pengukuran dikirim dengan

menggunakan teknologi *Short Message Service* (SMS) ke pusat data di Serpong, Tangerang Selatan.

**2.2 Pemrosesan Data**

*K-fold cross-validation* adalah cara untuk menguji dataset melakukan evaluasi algoritma Pembelajaran Mesin dengan jumlah data terbatas (Karal, 2020). Tujuan *K-fold cross validation* adalah untuk membagi kumpulan data menjadi “k” bagian, salah satu bagian data akan dipisahkan menjadi Data Pengujian (*Testing Data*) dan sisanya adalah Data Pelatihan (*Training Data*) dan mengevaluasi algoritma pada setiap sampel, serta memberikan akurasi rata-rata untuk kumpulan data tersebut. Gambar 1 merupakan contoh penerapan *K-Fold Cross Validation*.



Gambar 1. Cara kerja dari *K-fold Cross Validation*

Pada *Case #1*, sampel yang diberi warna hitam merupakan data pengujian, sedangkan sampel yang berwarna putih merupakan data pelatihan. Setelah selesai pengujian, selanjutnya masuk ke sampel data ke 2 dimana sampel data lainnya termasuk sampel data yang sebelumnya digunakan sebagai data pengujian menjadi data pelatihan. *Case #n* merupakan sampel data terakhir yang digunakan sebagai data pengujian, karena sampel data sebelumnya sudah digunakan sebagai data pengujian, sehingga sampel data lainnya menjadi data pelatihan. Pada penelitian ini, data yang sudah dikumpulkan, kemudian dibersihkan terlebih dahulu dari data anomali atau *outlier*. Cara membersihkan data anomali adalah menghapus data anomali dari *database* sehingga tidak ikut masuk dalam analisa perhitungan. Data anomali terjadi diduga karena *probe sensor* kotor tertutup sampah atau lumpur sehingga sensor tidak bisa mengukur

konsentrasi parameter yang dipantau dengan benar. Sekitar 0,67% data dikeluarkan karena terindikasi sebagai data anomali. Perbandingan distribusi data pelatihan dan data pengujian, digunakan 80% untuk data pelatihan dan 20% untuk data pengujian, tujuannya agar bisa melatih model dengan baik dan mendapatkan akurasi yang bagus dan sesuai dengan model yang digunakan. Untuk pengembangan model, data yang digunakan adalah sama, di mana data telah diproses dan dibagi menjadi dua kelompok, baik untuk data pelatihan maupun data pengujian.

**2.3 Pembangunan Model**

Model Pembelajaran Mesin yang digunakan pada penelitian ini adalah Jaringan Saraf Tiruan (JST), *Support Vector Machine* (SVM), *Random Forest*, dan *Naive Bayes*. Pada Tabel 1 disajikan sekilas kelebihan dan kekurangan model *Machine Learning* yang digunakan pada penelitian ini. Bahwa keunggulan dan kelemahan dari masing-masing model *Machine Learning* yang disebutkan di dalam Tabel 1, tergantung kepada kasus yang dihadapi.

**2.3.1 Jaringan Saraf Tiruan (JST)**

Jaringan Syaraf Tiruan (JST) adalah bagian dari Jaringan Saraf Biologis yang dipantau dengan komputer dan dapat diterapkan menggunakan perangkat elektronik atau diduplikasi dalam aplikasi yang ada pada sebuah komputer (Setiabudi, 2015, Solikhun *et al.*, 2017, Monika *et al.*, 2019).

Menurut Madontang *et al.*, (2013) dan Zola *et al.*, (2018), arsitektur Jaringan Syaraf Tiruan dapat di ketahui dari jumlah lapisan (*layer*) dan jumlah *neuron*, kemudian dikelompokkan menjadi lapisan *neuron* yang dibagi menjadi tiga bagian, yaitu:

1. *Input Layer*, yaitu *node* pada lapisan awal yang disebut *Input unit*.
2. *Hidden Layer*, yaitu *node* pada lapisan tengah disebut *Hidden unit*.
3. *Output Layer*, yaitu *node* pada lapisan terakhir disebut *Output unit*.

Penjelasan tentang cara kerja Jaringan Syaraf Tiruan dapat dilihat pada Gambar 2.

Tabel 1. Kelebihan dan kekurangan model *Machine Learning* yang digunakan

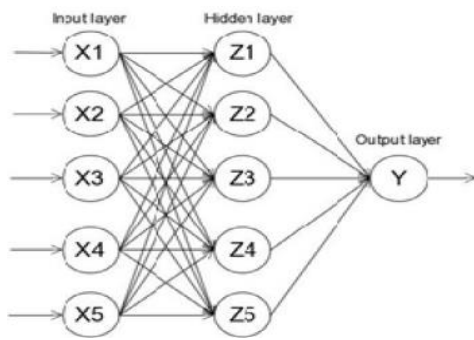
No.	Metode	Kelebihan	Kekurangan
1.	JST*	Memiliki toleransi yang tinggi terhadap data yang mengganggu, kemampuan baik untuk mengklasifikasikan pola, cocok untuk input dan output yang bersifat kontinyu	Membutuhkan sejumlah parameter yang terbaik secara empiris, waktu percobaan yang lama, interpretabiliti kurang baik
2.	SVM**	Dapat menghasilkan model klasifikasi yang baik walupun himpunan data relatif sedikit	Sulit diaplikasikan pada himpunan dengan jumlah sampel sangat besar
3.	<i>Random Forest</i> ***	Dapat mengolah data dalam jumlah yang besar	Interprestasi sulit dan membutuhkan <i>tuning model</i> yang tepat untuk mengolah data
4.	<i>Naive Bayes</i> ****	Model lebih sederhana, implementasi mudah, kecepatan pelatihan tinggi	Tingkat probabilitas rendah, tingkat estimator rendah

\* Gupta, A. *at al.*, (2015), Sudarsono, A. (2016), Cynthia *at al.*, (2017).

\*\* Parapat *et al.*, (2018), Otchere *et al.*, (2021).

\*\*\* Breiman, L. (2001).

\*\*\*\* Praningki *at al.*, (2018), Wickramasinghe *at al.*, (2021).



Gambar 2. Arsitektur Jaringan Syaraf Tiruan

Pada Gambar 2, arsitektur Jaringan Saraf Tiruan mempunyai tiga bagian, yaitu *input layer*, *hidden layer*, dan terakhir *output layer*. *Input layer* memiliki lima *node*, yaitu  $x_1$ ,  $x_2$ ,  $x_3$ ,  $x_4$ , dan  $x_5$ . Pada lapisan tersembunyi (*hidden layer*) juga terdapat lima *node*, yaitu  $z_1$ ,  $z_2$ ,  $z_3$ ,  $z_4$ , dan  $z_5$ . Pada lapisan keluaran (*output layer*) terdapat satu *node* yaitu nilai akhir dari *Hidden Layer*.

Ramadhona *et al.*, (2018) melakukan penelitian Prediksi Produktivitas Padi menggunakan Jaringan Syaraf Tiruan. Jumlah *Hidden Layer* digunakan mulai dari 1 sampai 10. Mereka melaporkan bahwa pada *Hidden Layer* berjumlah 3 diperoleh hasil tertinggi dibandingkan dengan jumlah *Hidden Layer* lain, yaitu nilai *Root Mean Square Error* (RMSE) mencapai 9,4156.

Di dalam studi ini, delapan parameter dibuat beberapa percobaan dengan menggunakan jumlah *Hidden Layer* yang berbeda dimulai dari 1, 3, 5, 7 sampai 9. Untuk jumlah *node* yang digunakan pada setiap *Hidden Layer* adalah 3, seperti pada Tabel 2. Parameter Jaringan Syaraf Tiruan yang digunakan pada ke 5 arsitektur pembuatan model dibuat sama, tujuannya agar tidak terjadi perbedaan hasil, tulisan atau cara pada saat melakukan percobaan model di *Google CoLab*.

Tabel 2. Jumlah *Hidden Layer* dan Simpul (*Node*)

No.	Jumlah <i>Hidden Layer</i>	Jumlah <i>Node</i>
1.	1	3
2.	3	3
3.	5	3
4.	7	3
5.	9	3

Tabel 3 menjelaskan parameter Jaringan Syaraf Tiruan (JST) yang digunakan selama proses pembuatan model. Fungsi Aktivasi (*Activation Function*) yang digunakan adalah *Rectified Linear Unit* (ReLU) karena dapat mengurangi waktu *Training Data* dan *Testing Data* secara signifikan (Lu *et al.*, 2019). *Optimizer* atau *Solver* yang digunakan adalah *Adam* karena dapat diimplementasikan dengan mudah, mempunyai memori lebih sedikit karena tidak perlu menyimpan data banyak, dan cocok diterapkan pada data atau parameter dengan jumlah yang besar (Kingma *et al.*, 2014). *Epoch* yang digunakan adalah 300 dan *Learning Rate* yang dipakai adalah 0,001 agar mendapatkan hasil yang lebih

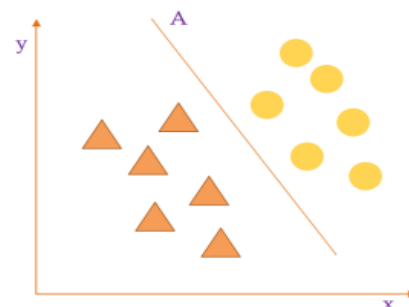
bagus dibandingkan memakai *Learning Rate* yang lebih besar ataupun yang lebih kecil (Wilson *et al.*, 2001).

Tabel 3. Ringkasan parameter Jaringan Syaraf Tiruan (JST)

Jumlah <i>Hidden Layer</i>	Fungsi Aktivasi	<i>Learning Algorithm Optimizer/Solver</i>	<i>Iterasi Max (Epoch)</i>	<i>Learning Rate</i>
1	ReLU	Adam	300	0,001
3	ReLU	Adam	300	0,001
5	ReLU	Adam	300	0,001
7	ReLU	Adam	300	0,001
9	ReLU	Adam	300	0,001

### 2.3.2 Support Vector Machine (SVM)

Menurut Chouhan *et al.*, (2020), *Support Vector Machine* (SVM) merupakan Algoritma Pembelajaran Mesin yang termasuk ke dalam kategori *Supervised Machine Learning* yang digunakan untuk tujuan klasifikasi dan regresi. Selain itu, SVM juga bisa digunakan dalam kasus analisis data, termasuk teknik-teknik yang digunakan dalam *Data Mining* seperti *Classification*, *Association Rules* dan *Clustering*. Pada metode ini, setiap item data ditempatkan dalam ruang fitur n-dimensi dengan memisahkannya di antara koordinat yang ditentukan. Kemudian, dengan bantuan *hyperplane*, yaitu garis dalam 2D, SVM mengklasifikasikan kedua bidang tersebut. Gambar 3 merupakan tampilan dari *Support Vector Machine* (SVM).



Gambar 3. *Support Vector Machine* (Chouhan *et al.*, 2019)

Sheykhmousa *et al.*, (2020) melakukan penelitian *remote sensing image classification* menggunakan metode SVM dan *Random Forest*. Pada SVM, kumpulan data dibagi menjadi dua bagian. Bagian pertama dipisahkan oleh *hyperplane*, memiliki nilai 1, sedangkan bagian lainnya yang tidak dipisahkan oleh *hyperplane*, memiliki nilai -1, seperti berikut.

$$X_i. W + b \geq 1 \text{ untuk } Y_i = 1 \dots\dots\dots (1)$$

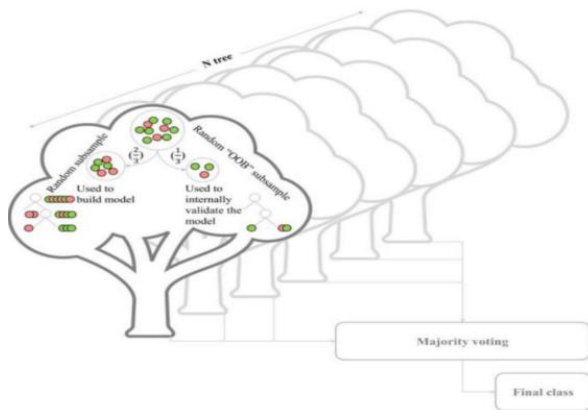
$$X_i. W + b \leq -1 \text{ untuk } Y_i = -1 \dots\dots\dots (2)$$

Dimana :  
 $X_i$  = data ke  $i$   
 $W$  = nilai berat  
 $b$  = nilai bias  
 $Y_i$  = kelas data ke  $-i$

### 2.3.3 Random Forest

Breiman (2001) menjelaskan *Random Forest* merupakan algoritma yang termasuk ke dalam pendekatan pembelajaran ensemble (*Ensemble Learning*), kemudian dikembangkan oleh

Primartha, R. (2018), untuk memecahkan masalah klasifikasi dan regresi. *Ensemble Learning* adalah skema pembelajaran mesin untuk meningkatkan akurasi dengan mengintegrasikan beberapa model untuk memecahkan masalah yang sama. Struktur *Random Forest* dengan *Struktur Tree* lainnya berbeda satu sama lain dan variabel pemisahan mendorong berbagai contoh *overfitting* dan *outlier* di antara model *Ensemble Tree*. Gambar 4 merupakan tampilan dari *Random Forest*.



Gambar 4. Cara kerja *Random Forest* (Sheykhmousa et.al, 2020)

Pada Gambar 4, terlihat bahwa dengan memperluas nilai *Random Forest* hingga jumlah pohon yang ditentukan oleh user (*N-tree*), dapat menghasilkan nilai (pohon) dengan *varians* yang tinggi dan bias yang rendah. Oleh karena itu, *input set data* yang tidak berlabel akan dihitung terhadap semua Pohon Keputusan (*Decision Tree*) yang dihasilkan dalam *Ensemble Learning*, dan setiap pohon memilih keanggotaan kelasnya masing-masing. Priantama at al., (2022) menjelaskan bahwa *Random Forest* sering digunakan untuk menyelesaikan masalah klasifikasi dan regresi. Model *Random Forest* yang digunakan, merupakan model turunan yang berasal dari pengembangan model *Classification And Regression Tree* (CART) dengan metode *bootstrap aggregating* (*bagging*) dan *random feature selection*. Rumus *Random Forest* yang terdiri atas beberapa (*n*) pohon ditulis sebagai berikut:

$$I(y) = \operatorname{argmax}_c \left( \sum_{n=1}^N I h_n(y) = c \right) \dots\dots\dots (3)$$

Dengan:  
*I* = variabel fungsi indikator  
*ffi<sub>n</sub>* = pohon ke sekian (*n*)

Pohon pada algoritma *Random Forest Classifier* dibangun dengan *Classification and Regression Tree* (CART), menggunakan ketentuan *Gini Impurity* dalam menentukan pecahan pohon keputusan. Mula-mula dihitung nilai *Gini Index*, kemudian dilanjutkan menghitung *Gini Impurity*. Indeks Gini dihitung memakai rumus berikut:

$$Gini = \sum_{i=1}^n P_{i^2} \dots\dots\dots (4)$$

Dengan:  
*n* = Jumlah kelas sasaran  
*i* = Kelas objek  
*P* = proporsi kelas sasaran

Gini Impurity dihitung dengan persamaan berikut:

$$Gini\ Impurity = 1 - \sum_{i=1}^n P_{i^2} \dots\dots\dots (5)$$

2.3.4 *Naive Bayes*

Menurut Rish (2001), metode *Bayesian* adalah metode untuk melakukan klasifikasi fenomena berdasarkan kemungkinan terjadi atau tidaknya fenomena tersebut. Metode *Naive Bayes* akan memberikan hasil yang lebih baik karena adanya tahap pelatihan (*Training*). Metode *Naive Bayes* menjelaskan bagaimana cara menghitung *posterior probability*. Berikut ini adalah persamaan untuk menghitung *posterior probability* (Vembandasamy et al., 2015).

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \dots\dots\dots (6)$$

Labels in diagram: Likelihood (P(x|c)), Class Prior Probability (P(c)), Posterior Probability (P(c|x)), Predictor Prior Probability (P(x)).

Dengan:  
*P(c|x)* = Data yang dipakai  
*P(c)* = *Class Prior Probability*  
*P(x|c)* = *Posterior Probability*  
*P(x)* = *Predictor Prior Probability*

2.3.5 Evaluasi Model

Pengukuran hasil terdiri atas 4 klasifikasi yang berbeda, yaitu *True Positive*, *True Negative*, *False Positive*, dan *False Negative*. Tabel 4 menampilkan *Confusion Matrix* beserta jumlah label yang dihitung dan seberapa banyak jumlah data pada bagian label tersebut, yang artinya semakin banyak datanya maka label tersebut semakin terang dan berwarna kuning, sebaliknya semakin sedikit datanya maka label tersebut semakin gelap dan berwarna ungu.

Tabel 4. *Confusion Matrix* beserta jenis kelasnya

		Actual Class	
		Positive (P)	Negative (N)
Predicted Class	Positive (P)	True Positive (TP)	False Positive (FP)
	Negative (N)	False Negative (FN)	True Negative (TN)

Gambar 5 memperlihatkan hasil perhitungan *Confusion Matrix* dari model pada satu *Hidden Layer*. Terdapat tulisan *True Label* dan *Predicted Label*. *True Label* merupakan label yang benar yang diwakili dengan indikator 0 dan 1, di mana 0 merupakan kondisi air memenuhi Baku Mutu dan 1 merupakan kondisi air tidak memenuhi Baku Mutu. Sedangkan *Predicted Label* adalah hasil prediksi model. Pada Gambar 5 terdapat tulisan *True Label* dan *Predicted Label*,

dimana *True Label* berisi *Positive* dan *Negative*, yang artinya label tersebut berisi benar dan salah, dan sudah berupa fakta. Sedangkan *Predicted Label* berisi *Positive or Negative*, yang artinya label tersebut benar atau salah, dan dalam bentuk pernyataan atau opini, karena belum pasti apakah label itu benar atau salah. Pada Gambar tersebut, label berwarna kuning di sebelah kanan bawah merupakan *True Negative* yang nilai prediksi negatif dan nilai aktual salah, yang artinya prediksi dan hasil sudah sesuai dalam bentuk negatif. Label berwarna ungu di sebelah kiri bawah merupakan *False Negative* yang artinya nilai prediksi negatif namun nilai aktual benar, yang artinya prediksinya salah, tetapi hasilnya benar. Label berwarna ungu di sebelah kanan atas merupakan *False Positive*, yang artinya nilai prediksi positif namun nilai aktual salah, yang artinya prediksinya benar, tetapi hasilnya salah. Dan label berwarna biru di sebelah kiri atas merupakan *True Positive* yang artinya nilai prediksi positif dan nilai aktual benar, yang artinya prediksi dan hasil sudah sesuai dalam bentuk positif.



Gambar 5. Contoh tampilan *Confusion Matrix*

Berdasarkan hasil *Confusion Matrix*, kemudian dihitung tingkat akurasi (*Accuracy*), *Recall*, *Precision*, *Specificity*, dan *F1 Score*. Berdasarkan hasil perhitungan ini dapat diketahui metode mana yang memiliki potensi yang baik untuk memprediksi kualitas air sungai.

*Accuracy* merupakan rasio untuk memprediksi nilai *True Positive* dan *True Negative* dari keseluruhan data. *Accuracy* menjawab pertanyaan “Berapa persen data yang benar diprediksi tidak memenuhi baku mutu dan memenuhi baku mutu dari keseluruhan data”. Untuk mencari tingkat akurasi digunakan persamaan 7.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \dots\dots\dots (7)$$

Dengan:

TP = *True Positive*

TN = *True Negative*  
 FP = *False Positive*  
 FN = *False Negative*

*Recall* merupakan rasio untuk memprediksi nilai *True Positive* dari *True Positive* dan *False Negative*. *Recall* menjawab pertanyaan “Berapa persen data yang tidak memenuhi baku mutu dari keseluruhan data yang benar dan yang sebenarnya tidak memenuhi baku mutu”. Untuk mencari nilai *Recall* digunakan persamaan 8.

$$Recall = \frac{TP}{TP + FN} \dots\dots\dots (8)$$

*Precision* adalah rasio untuk memprediksi nilai *True Positive* dari keseluruhan nilai positif yaitu *True Positive* dan *False Positive*. *Precision* menjawab pertanyaan “Berapa persen data yang tidak memenuhi baku mutu dari keseluruhan data yang sebenarnya dan juga yang di prediksi tidak memenuhi baku mutu”. Untuk mencari nilai *Precision* digunakan persamaan 9.

$$Precision = \frac{TP}{TP + FP} \dots\dots\dots (9)$$

*Specificity* adalah rasio untuk memprediksi nilai *True Negative* dengan keseluruhan nilai negatif yaitu *True Negative* dan *False Negative*. *Specificity* menjawab pertanyaan “Berapa persen data yang memenuhi baku mutu dari keseluruhan data yang yang sebenarnya memenuhi baku mutu dan juga yang salah di prediksi”. Untuk mencari nilai *Precision* digunakan persamaan 10.

$$Specificity = \frac{TN}{TN + FN} \dots\dots\dots (10)$$

*F1 Score* adalah perbandingan nilai *Precision* dan *Recall* yang di bobotkan. Untuk mencari nilai *Precision* digunakan persamaan 11.

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \dots\dots\dots (11)$$

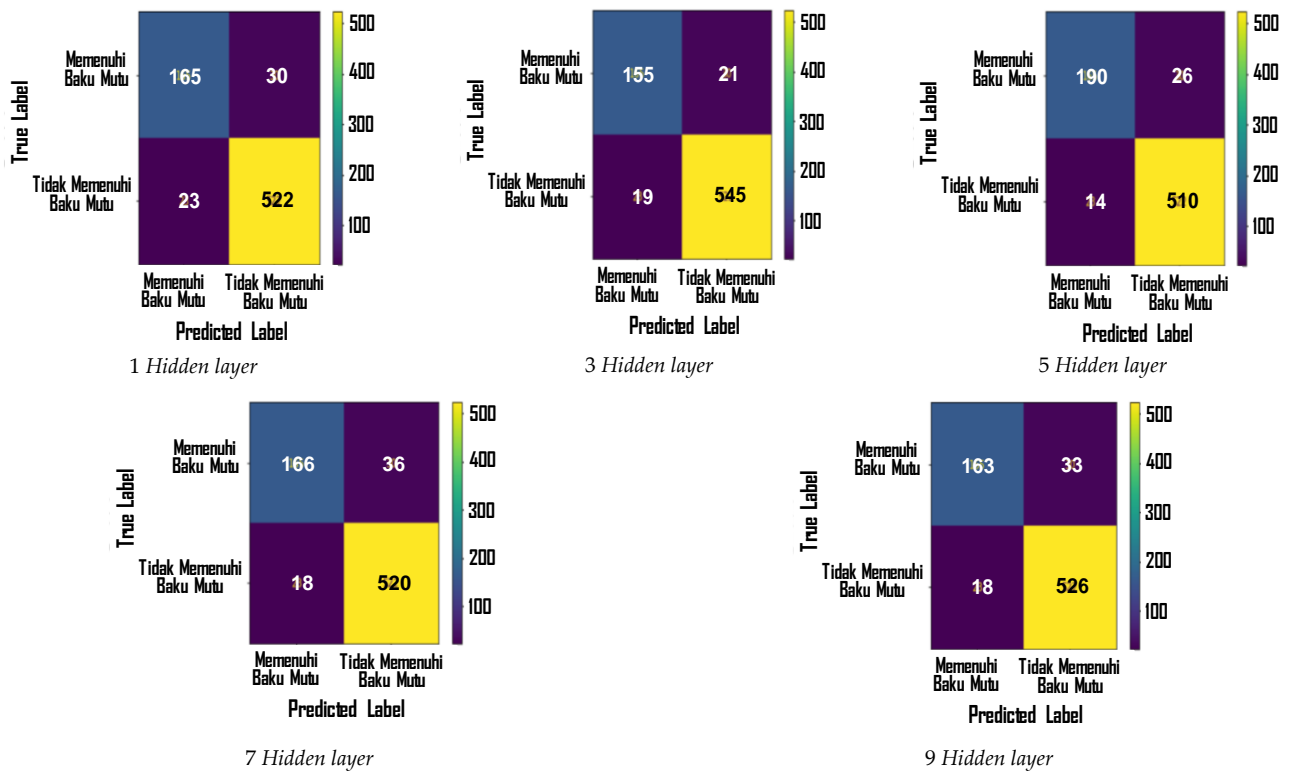
### 3. HASIL DAN PEMBAHASAN

#### 3.1 Data Hasil Pemantauan

Data yang digunakan untuk memprediksi kualitas air sungai Ciliwung adalah data pemantauan ONLIMO selama satu tahun, mulai tanggal 1 Januari sampai tanggal 31 Desember tahun 2018. Interval waktu pemantauan dilakukan satu kali dalam satu jam selama satu tahun, pemantauan dilakukan sebanyak 5.476 kali. Parameter yang dipantau adalah DO, pH, Nitrat, Suhu, Salinitas, TDS, DHL, dan Kekeruhan. Pada Tabel 5 disajikan contoh hasil pemantauan.

Tabel 5. Hasil pemantauan kualitas air sungai Ciliwung

Tgl.	Jam	DO (mg/l)	pH (-)	Nitrat (mg/l)	DHL (µS/cm)	TDS (mg/l)	Salinitas (mg/l)	Turbiditas (Ntu)	Suhu °C	Keterangan
1/1/2018	00	3,27	8,88	7,66	26,7	0,5	0,3	29,1	29	Tidak memenuhi baku mutu
1/1/2018	01	3,35	8,89	7,67	27,2	0,5	0,3	32,5	29	Tidak memenuhi baku mutu
1/1/2018	02	3,44	8,89	7,68	28	0,5	0,3	47,8	28,9	Tidak memenuhi baku mutu
30/3/2018	01	6,54	7,88	9,98	40,5	0,2	0,1	85,5	29,5	Memenuhi baku mutu
30/3/2018	02	6,53	7,83	9,51	41,7	0,2	0,1	71,8	29,8	Memenuhi baku mutu
30/3/2018	03	6,54	7,81	9,43	44	0,3	0,1	68,7	30,2	Memenuhi baku mutu



Gambar 6. Hasil Confusion Matrix metode JST pada beberapa Hidden Layer

### 3.2 Hasil Prediksi Model

Data hasil pemantauan selama satu tahun yang sudah dibersihkan dari data anomali ada sebanyak 5.441 data, dipakai sebagai Data Pelatihan (*Training Data*) sebesar 80%, sedangkan sisanya sebesar 20% digunakan sebagai Data Pengujian (*Testing Data*). Pengambilan data pelatihan dan data pengujian dilakukan secara acak. Selanjutnya akan dijelaskan hasil yang diperoleh pada setiap metode yang digunakan dengan *input* 8 parameter yaitu DO, pH, Nitrat, DHL, TDS, Salinitas, Turbiditas dan Suhu.

#### 3.2.1 Hasil Prediksi Model Menggunakan Metode Jaringan Saraf Tiruan (JST)

Seperti yang telah dijelaskan di atas, model prediksi kualitas air sungai dengan metode JST dilakukan dengan 5 variasi jumlah *Hidden Layer* yaitu 1, 3, 5, 7 dan 9. Di mana simpul (*node*) yang dipakai di setiap *Hidden Layer* ada 3. Gambar 6 memperlihatkan *Confusion Matrix* dari model pada lima macam *Hidden Layer* yang digunakan dan Tabel 6 merupakan rangkuman besaran *true positive*, *true negative*, *false positive* dan *false negative* untuk masing-masing *Hidden Layer*. Berdasarkan hasil evaluasi, didapati bahwa 5 *Hidden Layer* mempunyai nilai *True Positive* yang paling tinggi yang menandakan bahwa model yang menggunakan 5 *Hidden Layer* paling sedikit yang tidak tepat di dalam menentukan kategori kualitas air yang “memenuhi Baku Mutu” jika dibandingkan dengan *Hidden Layer*nya. Selain itu, *Hidden Layer* dengan jumlah 7 didapati paling rendah di dalam menentukan kategori kalitas air yang “memenuhi Baku Mutu” dimana nilai *False positive* yang paling tinggi jika dibandingkan dengan model lain.

Tabel 6. Rangkuman nilai *True Positive*, *True Negative*, *False Positive*, dan *False Negative* untuk 5 macam *Hidden Layer*

Hidden Layer	True Positive	True Negative	False Positive	False Negative
1	165	522	30	23
3	155	545	21	19
5	190	510	26	14
7	166	520	36	18
9	163	526	33	18

Berdasarkan data yang tertera dalam Tabel 5, selanjutnya dengan persamaan (7), (8), (9), (10) dan (11) dihitung tingkat Akurasi, *Recall*, *Precision*, *Specificity*, dan *F1 Score* pada setiap *Hidden Layer* yang diuji. Tabel 7 adalah rangkuman hasil perhitungan *Confusion Matrix* metode JST menggunakan jumlah *Hidden Layer* tipe 1, 3, 5, 7 dan 9. Untuk jumlah *node* yang dipakai pada setiap *Hidden Layer* adalah 3.

Hasil evaluasi metode JST dari 5 jenis arsitektur yaitu 1, 3, 5, 7 dan 9 *Hidden Layer* memperlihatkan hasil *Confusion Matrix* yang cukup tinggi, semua di atas angka 82%. Nilai *Accuracy* tertinggi pada *Hidden Layer* 3 dan 5 yaitu 94,6%, terendah pada *Hidden Layer* 7 yaitu 92,7%. Nilai *Recall* cukup bervariasi, tertinggi pada *Hidden Layer* 5 yaitu 93,1% dan terendah pada *Hidden Layer* 1 yaitu 87,8%. Nilai *Precision* tertinggi pada *Hidden Layer* 3 dan 5 yaitu 88,1%. Nilai *Specificity* pada semua *Hidden Layer* sangat tinggi, yaitu di atas 95%, nilai tertinggi pada *Hidden Layer* 5 yaitu 97,3%, terendah pada *Hidden Layer* 1 yaitu 95,8%. Terakhir, nilai *F1 Score* tertinggi pada *Hidden Layer* 5 yaitu 90,5%, dan terendah pada *Hidden Layer* 7 yaitu 85,7%. Pada Tabel 7 terlihat bahwa model JST 5 *Hidden Layer* memiliki tingkat akurasi, *Recall*, *Precision*, *Specificity* dan *F1*

Score paling tinggi dari semua *Hidden Layer* yang diuji. Berdasarkan hasil juga terlihat bahwa semakin meningkatnya *Hidden Layer* yang digunakan maka nilai akurasi semakin tinggi. Namun, perlu diperhatikan juga jika terlalu banyak *Hidden Layer* yang digunakan ada kemungkinan akurasi menurun. Ini terlihat ketika setelah *Hidden Layer* 5, nilai akurasi mulai menurun. Hal ini mungkin dapat disebabkan karena kompleksitas dari arsitektur JST sehingga sulit untuk mencapai kondisi maksimal. Menurut Ramadhona *et. al.*, (2018), variasi *Hidden Layer* sangat berpengaruh pada proses *backpropagation*. Semakin banyak *Hidden Layer* yang digunakan, maka nilai yang dihasilkan akan semakin rendah. Sebaliknya jika *Hidden Layer* yang digunakan terlalu sedikit, maka akan mendapatkan hasil yang kurang maksimal. Selanjutnya, metode JST dengan 5 *Hidden Layer* ini digunakan mewakili metode SJT dalam melakukan evaluasi model bersama metode Pembelajaran Mesin yang lain.

Tabel 7. Rangkuman Hasil *Confusion Matrix* metode JST pada 5 macam *Hidden Layer* (%)

<i>Hidden Layer</i>	Accuracy	Recall	Precision	Specifity	F1Score
1	92,8	87,8	84,6	95,8	86,2
3	94,6	89,1	88,1	96,6	88,6
5	94,6	93,1	88,1	97,3	90,5
7	92,7	89,3	82,3	96,7	85,7
9	93,1	90,1	83,2	96,7	86,5

### 3.3.2 Hasil Prediksi Model Menggunakan Motode Lainnya

Gambar 7 adalah hasil *Confusion Matrix* dari model, masing-masing metode SVM, *Random Forest*, dan *Naive Bayes*.

Untuk model yang menggunakan metode SVM didapati bahwa angka *True Positive*, *False Positive*, *True Negative*, dan *False Negative* masing-masing adalah 114, 87, 473 dan 66, seperti yang diperlihatkan di dalam Tabel 8. Hasil di atas mengindikasikan bahwa kesalahan di dalam mengklasifikasi kualitas air berdasarkan kategori memenuhi baku mutu dan tidak memenuhi baku mutu cukup besar. Manakala, untuk metode lainnya seperti *Random Forest* mempunyai angka *True Positive*, *False Positive*, *True Negative*, dan *False Negative* masing-masing adalah 201, 2, 537, dan 0 dan metode *Naive Bayes* masing-masing adalah 142, 77, 520, dan 1. Jika dibandingkan dengan semua model yang telah dikembangkan pada studi ini, diketahui bahwa metode *Random Forest* mempunyai

kesalahan yang kecil di dalam mengklasifikasikan kualitas air berdasarkan kategori memenuhi baku mutu dan tidak memenuhi baku mutu di mana jumlah dari *false positive*, dan *false negative* adalah paling kecil.

Tabel 8. Rangkuman *True Positive*, *True Negative*, *False Positive*, dan *False Negative* pada metode JST (5 *Hidden Layer*), SVM, *Random Forest*, dan *Naive Bayes*

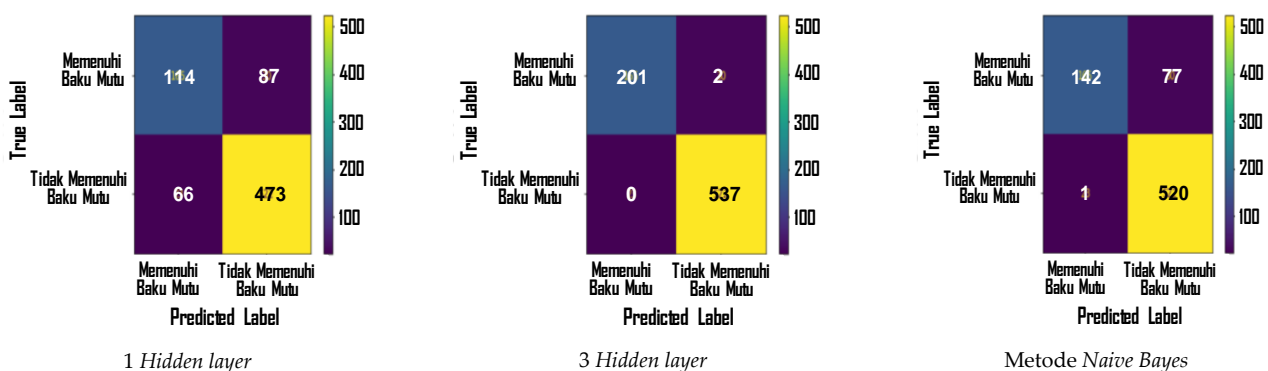
Metode	<i>True Positive</i>	<i>True Negative</i>	<i>False Positive</i>	<i>False Negative</i>
<i>Support Vector Machine</i> (SVM)	114	473	87	66
<i>Random Forest</i>	201	537	2	0
<i>Naive Bayes</i>	142	520	77	1

Dari data dalam Tabel 8, kemudian dihitung tingkat Akurasi, *Recall*, *Precision*, *Specifity* dan *F1 Score* untuk metode SVM, *Random Forest* dan *Naive Bayes* dengan persamaan (7), (8), (9), (10), dan (11). Tabel 9 adalah rangkuman hasil *Confusion Matrix* dari 4 macam metode berbasis pembelajaran mesin yang diuji yaitu metode JST dengan 5 *Hidden Layer*, SVM, *Random Forest* dan *Naive Bayes*.

Dari Tabel 9 dapat dilihat bahwa metode *Random Forest* memiliki nilai *Confusion Matrix* paling baik, angka-angka hasil evaluasi paling tinggi dari semua metode yang diujikan. Tingkat akurasi, *Recall*, *Precision*, *Specifity* dan *F1 Score* masing-masing adalah 99,7%, 100%, 99,0%, 100% dan 99,5%. Setelah metode *Random Forest* adalah metode Jaringan Syaraf Tiruan (JST) dengan 5 *Hidden Layer*, kemudian *Naive Bayes*, dan terakhir adalah metode SVM.

Tabel 9. Rangkuman Hasil *Confusion Matrix* pada 4 macam Metode *Machine Learning* (%)

Metode	Accuracy	Recall	Precision	Specificity	F1 Score
JST (5 <i>hidden layer</i> )	94,6	93,1	88,1	97,3	90,5
SVM	79,3	63,3	56,7	87,8	59,8
<i>Random Forest</i>	99,7	100	99,0	100	99,5
<i>Naive Bayes</i>	89,5	99,3	64,8	99,8	78,4



Gambar 7. Hasil *Confusion Matrix* pada metode SVM, *Random Forest*, dan *Naive Bayes*



Berdasarkan hasil penelitian yang dilakukan oleh Supriyadi *et al.* (2020) diperoleh *Confusion Matrix* pada metode *Random Forest* lebih tinggi dibanding metode *Decision Tree* dan metode SVM. Selain itu, Bawono *et al.*, (2019) menemukan bahwa tingkat akurasi metode *Random Forest* lebih tinggi dari metode *Naïve Bayes*. Salah satu alasan adalah karena *Random Forest* bekerja secara efisien pada kumpulan data yang besar. Sementara hasil *Confusion Matrix* yang didapat pada metode SVM tidak terlalu tinggi (Najah *et al.*, 2011), karena SVM tidak begitu sesuai jika digunakan untuk mengolah data yang jumlahnya sangat banyak. Beberapa hasil penelitian seperti dijelaskan diatas, mendukung hasil yang diperoleh pada penelitian ini.

#### 4. KESIMPULAN

Hasil evaluasi metode JST dari 5 jenis arsitektur yang digunakan, yaitu terdiri atas 1, 3, 5, 7, dan 9 *hidden layer* memperlihatkan hasil *Confusion Matrix* yang cukup tinggi, semua di atas angka 82%. Dari hasil penelitian ini terlihat bahwa model JST dengan 5 *Hidden Layer* memiliki tingkat Akurasi (*Accuracy*), *Recall*, *Precision*, *Specificity* dan *F1 Score* tertinggi dari semua *Hidden Layer* lain yang diuji.

Berdasarkan hasil evaluasi model dari 4 jenis metode *Machine Learning* diteliti, yaitu JST, SVM, *Random Forest*, dan *Naïve Bayes*, diketahui hasil *Confusion Matrix* yaitu tingkat Akurasi, *Recall*, *Precision*, *Specificity*, dan *F1 Score* pada *Random Forest* adalah yang tertinggi, mendekati 100%, disusul metode JST dengan 5 *Hidden Layer*, *Naïve Bayes*, dan terakhir SVM. Berdasarkan hasil studi ini dapat diketahui bahwa metode *Random Forest*, JST, dan *Naïve Bayes* mempunyai potensi yang baik untuk memprediksi kualitas air sungai, khususnya untuk kasus sungai Ciliwung.

#### PERSANTUNAN

Pertama, ucapan terima kasih disampaikan kepada Kepala Pusat Riset Lingkungan dan Teknologi Bersih (PRLTB), Badan Riset dan Inovasi Nasional (BRIN) yang telah memberi fasilitas untuk melakukan penelitian ini. Kedua kepada Heru Dwi Wahyono, B. Eng., M. Kom., Peneliti Ahli Utama pada PRLTB, BRIN yang telah memberi bimbingan selama melakukan penelitian.

#### DAFTAR PUSTAKA

- Al-Adhaileh, M. H., & Alsaade, F. W. (2021). Modelling and prediction of water quality by using artificial intelligence. *Sustainability (Switzerland)*, 13(8). <https://doi.org/10.3390/su13084259>.
- Alphayo, S. M., & Sharma, M. P. (2018). Impact of Land Use on Water Quality in Ruvu River Basin. Tanzania: Bloomsbury India.
- Bawono, B., dan Wasono, R. (2019). Perbandingan Metode Random Forest Dan Naïve Bayes Untuk Klasifikasi Debitur Berdasarkan Kualitas Kredit. Seminar Nasional Edusaintek. FMIPA, Universitas Muhammadiyah Semarang.
- Ben Hassen, T., El Bilali, H., & Allahyari, M. S. (2020). Impact of COVID-19 on Food Behavior and Consumption in Qatar (p. 319). Qatar University. <https://doi.org/10.29117/quarfe.2020.0296>.
- Breiman, L. (2001). Random Forests. *Machine learning*, 45(1), 5-32. Springer.
- Chouhan, S. S., Singh, U. P., & Jain, S. (2020). Applications of computer vision in plant pathology: a survey. *Archives of computational methods in engineering*, 27(2), 611-632.
- Cynthia, E. P., & Ismanto, E. (2017). Jaringan syaraf tiruan algoritma backpropagation dalam memprediksi ketersediaan komoditi pangan provinsi riau. *Rabit: Jurnal Teknologi dan Sistem Informasi Univrab*, 2(2), 83-98.
- Damayanti, A. A., Wahjono, H. D., & Santoso, A. D. (2022). Pemantauan Kualitas Air Secara Online dan Analisis Status Mutu Air di Danau Toba, Sumatera Utara. *Jurnal Sumberdaya Alam dan Lingkungan*, 9(3), 113-120.
- Gupta, A., Gupta, S., & Singh, D. (2015). A systematic review of classification techniques and implementation of ID3 Decision Tree Algorithm. In 4th International Conference on System Modeling & Advancement in Research Trends (SMART) College of Computing Sciences and Information Technology (CCSIT), Teerthanker Mahaveer University, Moradabad.
- Han, J., Kamber, M., Pei, J. (2012). *Data Mining: Concepts and Techniques*. United States of America: Morgan Kaufmann Publisher.
- Japan International Cooperation Agency (1990). *The Study On Urban Drainage and Wastewater Disposal Project in The City of Jakarta*, JICA.
- Karal, O. (2020). Performance comparison of different kernel functions in SVM for different k value in k-fold cross-validation. In *Proceedings - 2020 Innovations in Intelligent Systems and Applications Conference, ASYU 2020*. Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ASYU50717.2020.9259880>.
- Kingma, D. P., Ba, J., Adam. (2014). A method for stochastic optimization. arXiv preprint arXiv: 1412.6980.
- Lu, L., Shin, Y., Su, Y., & Karniadakis, G. E. (2019). Dying relu and initialization: Theory and numerical examples. arXiv preprint arXiv:1903.06733.
- Madontang Z. A. (2013). Jaringan Syaraf Tiruan dengan Algoritma Backpropagation untuk menentukan kelulusan Sidang Skripsi.
- Monika, D., Ahmad, A., & Wardani, S. (2019). Model Jaringan Syaraf Tiruan Dalam Memprediksi Ketersediaan Cabai Berdasarkan Provinsi. *Teknika*, 8 (1), 17-24.
- Najah, A., El-Shafie, A. H., Karim, O. A., Jaafar, O., El-Shafie, A. H. (2011). An application of different artificial intelligences techniques for water quality prediction. *Int J Phys Sci* 6.22: 5298-5308.
- Otchere, D. A., Ganat, T. O. A., Gholami, R., & Ridha, S. (2021). Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models. *Journal of Petroleum Science and Engineering*, 200, 108182.

- Parapat, Monika. I., (2018). Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak. Diss. Universitas Brawijaya.
- Patricia, Conchita, Astono, W., Hendrawan, D.I. (2018). Kandungan nitrat dan fosfat di sungai ciliwung. Prosiding Seminar Nasional Cendekiawan.
- Peraturan Menteri Lingkungan Hidup Republik Indonesia Nomor 5 Tahun 2014 tentang Baku Mutu Air Limbah.
- Peraturan Menteri Lingkungan Hidup dan Kehutanan Republik Indonesia Nomor: P. 56/Menlhk-Setjen/2015 Tentang Tata Cara dan Persyaratan Teknis Pengelolaan Limbah Bahan Berbahaya dan Beracun dari Fasilitas Pelayanan Kesehatan.
- Peraturan Pemerintah Republik Indonesia Nomor 82 Tahun 2001 tentang Pengelolaan Kualitas Air dan Pengendalian Pencemaran Air. Lampiran 1.
- Praningki, T., & Budi, I. (2018). Sistem Prediksi Penyakit Kanker Serviks Menggunakan CART, Naive Bayes, dan k-NN. *Creative Information Technology Journal*, 4(2), 83-93.
- Priantama, Y. (2022). Optimasi Correlation-Based Feature Selection Untuk Perbaikan Akurasi Random Forest Classifier Dalam Prediksi Performa Akademik Mahasiswa. *JIKO (Jurnal Informatika dan Komputer)*, 6(2), 251-260.
- Ramadhona, G., Setiawan, B. D., & Bachtiar, F. A. (2018). Prediksi Produktivitas Padi Menggunakan Jaringan Syaraf Tiruan Backpropagation. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(12), 6048-6057.
- Rish, I. (2001). An empirical study of the Naive Bayes Classifier. *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, 22230, 41-46. <https://doi.org/10.1039/b104835j>
- Rizal, N. N. M., Hayder, G., & Yusof, K. A. (2022). Water quality predictive analytics using an artificial neural network with a graphical user interface. *Water*, 14(8), 1221.
- Setiabudi, D. (2015). Sistem Informasi Peramalan Beban Listrik Jangka Panjang di Kabupaten Jember Menggunakan Jaringan Syaraf Tiruan Backpropagation. ISSN: 2476-9754.
- Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., & Homayouni, S. (2020). Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/JSTARS.2020.3026724>.
- Solikhun, Safii, M. & Trisno, A. (2017). Jaringan Saraf Tiruan Untuk Memprediksi Tingkat Pemahaman Siswa Terhadap Matapelajaran Dengan Menggunakan Algoritma Backpropagation. *Jurnal Sains Komputer & Informatika*, Vol. 1, No.1, pp. 24-36.
- Standar Nasional Indonesia Nomor 698957 Tahun 2008 tentang Metoda Pengambilan Contoh Sampel Air Permukaan.
- Sudarsono, A. (2016). Jaringan Syaraf Tiruan Untuk Memprediksi Laju Pertumbuhan Penduduk Menggunakan Metode Backpropagation (Studi Kasus Di Kota Bengkulu). *Jurnal Media Infotama*, 12(1).
- Supriyadi, R., Gata, W., Maulidah, N., dan Fauzi, A. (2020). Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah. *Jurnal Ilmiah Ekonomi dan Bisnis*, Vol.13, No.2, pp. 67-75.
- Trofisa, D. (2011). Kajian Beban Pencemar dan Daya Tampung Pencemaran Sungai Ciliwung di Segmen Kota Bogor. Bogor: Institut Pertanian Bogor.
- Vembandasamy, K., R. Sasipriya, and Deepa, E. (2015) "Heart Diseases Detection Using Naive Bayes Algorithm. *International Journal of Innovative Science, Engineering & Technology* 2.9: 441-444.
- Wahjono, H. D. (2016). Penerapan Teknologi Online Monitoring Kualitas Air Untuk Das Prioritas di Sungai Ciliwung Dan Sungai Cisadane. *Jurnal Air Indonesia*, 9(1).
- Wickramasinghe, I., & Kalutarage, H. (2021). Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation. *Soft Computing*, 25(3), 2277-2293.
- Wilson, D., Martinez, Tony. (2001). The need for small learning rates on large problems. 1. 115 - 119 vol.1. 10.1109/IJCNN.2001.939002.
- Zola, F. (2018). Jaringan Syaraf Tiruan Menggunakan Algoritma Backpropagation Untuk Memprediksi Prestasi Siswa. *Jurnal Teknologi Dan Open Source*, 1(1), 58-72.