



## Optimasi Prediksi Bencana Banjir menggunakan Algoritma SVM untuk penentuan Daerah Rawan Bencana Banjir

Saruni Dwiasnati<sup>1</sup>, Yudo Devianto<sup>2</sup>

<sup>1</sup>Fakultas Ilmu Komputer, Teknik Informatika, Universitas Mercu Buana

<sup>2</sup>Fakultas Ilmu Komputer, Sistem Informasi, Universitas Mercu Buana

Email: [saruni.dwiasnati@mercubuana.ac.id](mailto:saruni.dwiasnati@mercubuana.ac.id), [yudo.devianto@mercubuana.ac.id](mailto:yudo.devianto@mercubuana.ac.id)

### Abstract

*Flooding is one factor that can cause an economic slowdown in the affected area. Bandung is known as the "City of Flowers" and "City of Fashion," since the term refers to the city's many different modes that sprout up in various locations to facilitate the purchasing and selling process. Bandung not only gave rise to fashion trends that would become popular year after year, but it also had a plethora of traditional cuisine dishes that were both unique and exciting. Because of this, Bandung Regency is now one of the flood-prone areas in the country. It will be easier to deliver information to communities in the Bandung Regency that are included in Flood-Prone Areas or Non-Flood-Prone Areas if a model of flood-prone areas is created. The SVM method is a technique that may be used in classification and regression and has recently become quite popular. In terms of functions and conditions of problems that can be solved, SVM is similar to Artificial Neural Network (ANN), and to improve its accuracy, it uses what can be optimized with PSO (Particle Swarm Optimization), where the test data used is BNPB official website data, BPS Bandung District, and BMKG that has been processed. Data included in the classification criteria for Flood Prone Areas are Rainfall Intensity, Water Discharge, Area, Length of Rain, and Population Density. The accuracy rate generated by using the SVM algorithm is 85.71% and the AUC generated is 0.841 while the accuracy rate generated by using the PSM-based SVM algorithm is 97.62%. and AUC produced at 1,000.*

Keywords: SVM Algorithm, PSO, Flood

### Abstrak

Banjir merupakan salah satu hal yang dapat memperlambat laju ekonomi di daerah yang terkena dampaknya. Bandung di juluki kota Kembang dan kota Mode, karena julukan tersebutlah yang membuat Bandung menjadi kota dengan berbagai macam mode yang tumbuh di beberapa tempat sebagai titik untuk menjalankan proses jual beli tersebut. Bandung tidak hanya melahirkan mode-mode yang akan menjadi hitz di setiap tahunnya namun banyak pula menjadi kiblat dari olahan makanan tradisional yang luar biasa unik dan menarik. Karena hal itu lah yang membuat Kabupaten Bandung saat ini menjadi salah satu daerah menjadi Daerah Rawan Banjir. Dengan membuat model daerah rawan banjir, dapat mempermudah memberikan informasi kepada masyarakat yang ada di daerah Kabupaten Bandung yang termasuk pada Zona Daerah Rawan Banjir atau Daerah Tidak Rawan Banjir. Algoritma SVM merupakan suatu teknik yang dapat digunakan dalam kasus klasifikasi maupun regresi, yang sangat populer belakangan ini. SVM berada dalam satu kelas dengan Artificial Neural Network (ANN) dalam hal fungsi dan kondisi permasalahan yang bisa diselesaikan dan untuk dapat meningkatkan nilai akurasi maka menggunakan yang dapat dioptimasi dengan PSO (Particle Swarm Optimization) dimana data uji yang digunakan adalah Data website resmi BNPB, BPS Kabupaten Bandung dan BMKG yang telah diolah. Data yang masuk dalam kriteria klasifikasi Daerah Rawan Banjir yaitu Intensitas Curah Hujan, Debit Air, Luas Wilayah, Lamanya Hujan, dan Kepadatan Penduduk. Tingkat akurasi yang dihasilkan dengan menggunakan algoritma SVM adalah 85,71% dan AUC yang dihasilkan sebesar 0.841 sedangkan tingkat akurasi yang dihasilkan dengan menggunakan algoritma SVM berbasis PSO adalah 97,62%. dan AUC yang dihasilkan sebesar 1.000.

Kata kunci: Algoritma SVM, PSO, Banjir

### 1. Pendahuluan

Dalam beberapa tahun terakhir ini, banjir merupakan bencana yang sering terjadi di Indonesia ditinjau dari frekuensi debit hujan dan lama dari turunnya hujan tersebut. Berdasarkan data Badan Nasional Penanggulangan Bencana (BNPB), banjir merupakan bencana yang paling sering terjadi di Indonesia dengan 464 kejadian banjir setiap tahunnya. Banjir yang disertai longsor menjadi bencana ke-6 yang paling sering terjadi

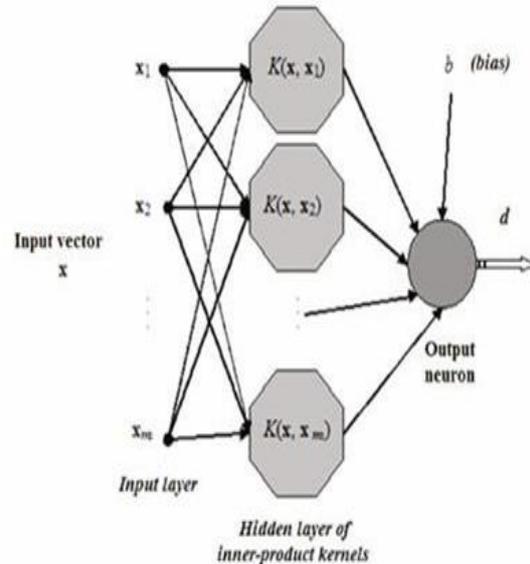
di Indonesia dengan 32 kejadian setiap tahunnya. Ada beberapa faktor utama yang mengakibatkan bencana banjir tidak dapat dihindari lagi saat ini, antara lain yaitu berkurangnya tutupan pohon, cuaca ekstrem, dan kondisi topografis Daerah Aliran Sungai (DAS). Untuk mengetahui daerah mana yang termasuk zona rawan banjir di Kabupaten Bandung diperlukan analisis daerah rawan banjir.

Pada penelitian ini dilakukan analisa daerah rawan banjir melalui pendekatan data mining untuk mengetahui daerah mana saja yang masuk dalam daerah rawan bencana banjir di daerah Kabupaten Bandung. Dalam pengolahan data yang digunakan dalam penelitian ini adalah Rapidminer versi 10.

Data Mining adalah suatu bidang interdisipliner yang berpatokan pada ilmu komputer (data base, kecerdasan buatan, pembelajaran mesin, grafis dan visualisasi model), statistik dan teknik (pengenalan pola, jaringan saraf). Penambangan data sering juga disebut Knowledge Discovery in Databases (KDD) yaitu suatu aktivitas yang meliputi pengumpulan, penggunaan data historis untuk menemukan pola reguler, pola hubungan dalam kumpulan data yang besar [1]. SVM adalah suatu teknik yang baru untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi yang sangat populer pada saat ini. Svm berada dalam satu kelas dengan ANN dalam hal fungsi dan kondisi permasalahan yang biasa diselesaikan. Kedua algoritma tersebut masuk dalam kelas supervised learning. Baik yang digunakan untuk para ilmuwan ataupun praktisi yang mempunyai kesukaan dalam bidang Machine Learning, Data Mining atau Big Data telah banyak menerapkan teknik ini dalam menyelesaikan masalah-masalah yang terlihat real dalam kehidupan sehari-hari yang digunakan untuk memprediksi sesuatu hal pada tahun ke depan nya. Algoritme machine learning belajar dari data, sehingga penting menyiapkan data secara tepat untuk menyelesaikan suatu masalah [2]. Disisi lain [3] menyatakan pada banyak algoritme machine learning langkah standar sebelum pelatihan adalah menghapus rata-rata dari data, yang dikenal dengan zero-mean atau standarisasi. Penerapan unsupervised learning juga telah dilakukan untuk clustering data history dari log transaksi pengguna internet di website microsoft dan MSNBC dengan menggunakan kaidah pada soft set theory [4],[5]. Sedangkan penerapan supervised learning juga telah dilakukan untuk klasifikasi batuan berjenis igneous [6], klasifikasi dan predikti kualitas air sungai [7], klasifikasi data citra pada bidang medis [8], seleksi atribut pada dataset yang berpengaruh secara signifikan pada proses pengambilan keputusan [9], prediksi penggunaan minyak dalam kurun waktu tertentu [10], dan peramalan kejadian penyakit demam berdarah [11]. Confusion Matrix adalah model yang akan membentuk matrix yang terdiri dari true positif atau tupel positif dan true negatif atau tupel negative [12]. Analisis pola frekuensi tertinggi menghasilkan kombinasi item yang memenuhi syarat minimum support yang sudah ditentukan untuk pembentukan pola association rule dilakukan untuk menemukan aturan assosiasi yang memenuhi minimum confidence dari frekuensi tertinggi yang telah ditemukan [13].

Klasifikasi adalah satu bentuk analisis data yang menghasilkan model untuk mendeskripsikan kelas data yang penting [14]. ANN menemukan solusi berupa local

optimal sedangkan SVM menemukan solusi yang global optimal [15]. Support Vector Machines (SVM) telah menjadi metode klasifikasi dan regresi yang sering digunakan untuk masalah linear dan nonlinear. Kelebihan dari algoritma Support Vector Machines ini adalah dari kemampuannya untuk menerapkan pemisahan linear pada input data non linear berdimensi tinggi, dan ini diperoleh dengan menggunakan fungsi kernel yang diperlukan. Efektivitas Support Vector Machines sangat dipengaruhi oleh jenis fungsi kernel yang dipilih dan diterapkan berdasarkan karakteristik data [16].



Gambar 1. Arsitektur Support Vector Machine (SVM)

Beberapa algoritma yang termasuk dalam metode metaheuristik antara lain Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Firefly Algorithm dan masih banyak lagi. Ada beberapa permasalahan yang umum diselesaikan menggunakan metode metaheuristik, misalnya NP-Hard problem, yaitu masalah yang memiliki solusi non-deterministic, solusi ini berjalan dalam waktu polinomial. Algoritma Genetik, dan Particle Swarm Optimization (PSO) merupakan metode pencarian acak yang sering digunakan untuk mencari parameter dari sistem persamaan tak linier tersebut [17]. Algoritma PSO telah banyak digunakan untuk pencarian parameter model tak linier. Menurut penelitian [18,19] yang menggunakan PSO untuk menentukan parameter model banjir. Sedangkan pada penelitian lainnya [20] mengimplementasikan algoritma PSO sebagai bagian dari algoritma kecerdasan buatan. PSO juga telah digunakan dalam riset model curah hujan oleh penelitian [21]. Penelitian lain yang dilakukan [22] menghasilkan informasi yang memberikan gambaran tentang pola kemiripan dan klasifikasi dari nilai atribut biodata CPMI sehingga diperoleh informasi yang berarti tentang nilai atribut biodata tersebut dan dapat berguna untuk melakukan prediksi klasifikasi data CPMI bagi keperluan pemasaran PT. SAM.

Populasi dalam penelitian ini merupakan dari dataset yang dikeluarkan oleh website resmi BNPB, BPS Kabupaten Bandung dan BMKG. Data yang diambil berkisar pada tahun 2016-2017 dan wilayah yang menjadi pusat penelitian adalah Kabupaten Bandung. Kriteria variabel yang digunakan pada penelitian saat ini untuk klasifikasi Daerah Rawan Banjir yaitu Intensitas Curah Hujan, Debit Air, Luas Wilayah, Lamanya Hujan, dan Kepadatan Penduduk.

Penelitian ini bertujuan untuk membantu memberikan informasi kepada masyarakat yang ada di daerah Kabupaten Bandung apakah daerah yang mereka tempati/tinggali termasuk pada Zona Daerah Rawan Banjir atau Daerah Tidak Rawan Banjir, informasi tersebut dapat diakses pada halaman resmi BNPB, BPS Kabupaten Bandung dan BMKG.

Metode yang digunakan dalam penelitian ini adalah Support Vector Machine (SVM) dengan tambahan feature selection yakni Particle Swarm Optimization (PSO). Penelitian sebelumnya yang membahas mengenai optimasi pada algoritma klasifikasi adalah dengan menggunakan algoritma Naïve Bayes dan optimasi feature selection Information Gain yang hanya menaikan sebesar accuracy 0.85% untuk sentimen analis pada movie review[4], diharapkan penelitian kali ini menggunakan algoritma Support Machine Vector (SVM) dan optimasi feature selection ini dapat menaikan hasil accuracy nya walaupun objek penelitiannya berbeda. Penelitian yang dilakukan ini akan membandingkan apakah metode SVM saja dengan SVM yang menggunakan feature selection dapat meningkatkan nilai accuracy dan juga membandingkan nilai AUC (Area Under Curve) untuk kedua metode tersebut. Pemilihan algoritma SVM ini didasarkan pada banyaknya penelitian yang menggunakan algoritma tersebut untuk penentuan Daerah Rawan Banjir dan hasil dari algoritma tersebut akan dioptimasikan agar menjadi akurat.

## 2. Metode Penelitian

Machine learning saat ini sedang sangat ramai dibicarakan banyak orang, karena kemampuannya hampir setara dengan proses yang manusia jalankan, hal tersebut bisa menjalankan kemampuan manusia oleh sebuah mesin. Machine Learning yang merupakan cabang dari Ilmu Komputer berhubungan dengan Artificial Intelligent atau ilmu kecerdasan buatan berfokus pada pembuatan/pengembangan dan studi suatu sistem dengan tujuan agar mampu belajar dari data-data yang diperoleh dari objek yang dituju. Menurut[23] Machine learning adalah suatu bidang studi yang memberikan kemampuan pada program komputer untuk belajar tanpa secara eksplisit diprogram. Maka dataset yang diberikan kepada mesin terdiri dari 2 jenis ,yaitu[24]:

### a. Data Training.

Yakni sekumpulan data yang digunakan untuk melatih mesin untuk mengetahui dan menerapkan algoritma yang digunakan. Sekumpulan data tersebut berguna untuk menemukan model yang baik dari data tersebut agar bisa diterapkan pada data testing.

### b. Data Testing.

Sedangkan data testing merupakan sekumpulan data yang digunakan sebagai referensi dan mengevaluasi dari model yang telah ditemukan oleh mesin. Apakah model yang ada mempunyai nilai akurasi yang baik atau tidak.

Penelitian ini merupakan penelitian berbasis eksperimental. Pada bagian ini menjelaskan tentang metode penelitian yang digunakan yaitu teori Support Vector Machine(SVM), dan metrik pengukuran performa algoritme seperti Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), dan confusion matrix.

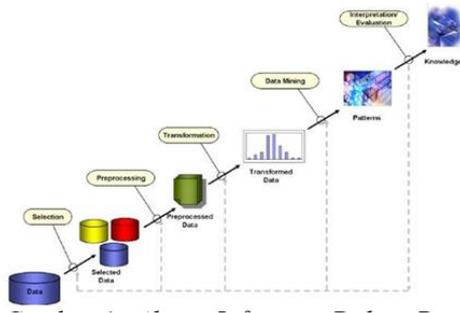
### 2.1. Knowledge Discovery in Database (KDD)

Knowledge discovery in database (KDD) adalah kegiatan yang meliputi Penyeleksian data dapat disimpan dalam berbagai format seperti flat file, spreadsheet, atau tabel-tabel relasional, dan dapat menempati tempat penyimpanan data terpusat atau terdistribusi pada banyak tempat. Preprocessing data merupakan sebuah langkah untuk mentransformasikan data input mentah ke dalam format yang sesuai untuk analisis selanjutnya. Langkah-langkah yang terlibat dalam preprocessing data meliputi menggabungkan data dari berbagai sumber, membersihkan (cleaning) data untuk membuang noise dan observasi duplikat, dan menyeleksi record dan fitur yang relevan untuk pekerjaan data mining. Pembentukan model dari data set. Hasil dari data mining sering kali diintegrasikan dengan decision support system (DSS), Integrasi demikian memerlukan langkah postprocessing yang menjamin bahwa hanya hasil yang valid dan berguna yang akan digabungkan dengan DSS. Salah satu pekerjaan dan postprocessing adalah visualisasi yang memungkinkan analyst untuk mengeksplor data dan hasil yang di dapat dari data mining dari berbagai sudut pandang. Membandingkan dua hasil dari pemodelan yang berupa accuracy, precision, recall maupun AUC antara model dengan Feature Seleccion (Particle Swarm Optimization) maupun tidak.

### 2.2.Feature Selection (Particle Swarm Opmization)

Masalah yang sering timbul dalam klasifikasi khusus, dan machine learning pada umumnya, adalah menemukan cara untuk mengurangi dimensi  $n$  dari ruang fitur  $F$  untuk mengatasi risiko "Overfitting", overvitting sendiri adalah nilai yang disebabkan karena sedikitnya data training dibanding dengan jumlah  $n$  sebagai attribut[25]. Maka dibutuhkan sebuah teknik untuk menghindari risiko overfitting, digunakanlah feature selection. Particle Swarm Optimization adalah salah satu metode yang digunakan untuk menaikan nilai

akurasi dari model machine learning yang dikembangkan.

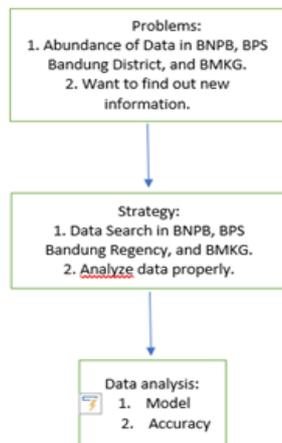


Gambar 2. Aliran Informasi Dalam Data Mining

### 2.3. Kerangka Berpikir

Dalam sebuah penelitian dibutuhkan kerangka berpikir, agar dalam pelaksanaan penelitian berjalan sesuai dengan apa yang telah direncanakan. Selain itu kerangka berpikir merupakan alur logika dalam suatu penelitian. Baik tidaknya suatu penelitian bisa dilihat dalam kerangka berpikir.

Dikerangka berpikir juga dapat dilihat apakah penelitian yang dilakukan merupakan suatu hal yang masuk logika, bila dilihat dari kerangka berpikir sudah merupakan hal yang tidak masuk dalam nalar, maka dapat dipastikan penelitian tersebut merupakan hal yang tidak mungkin berhasil. Adapun kerangka berpikir dalam penelitian ini dapat dilihat pada Gambar 3.



Gambar 3. Kerangka Kerja

Pada kerangka berpikir terlihat bahwa optimasi yang digunakan pada Prediksi Bencana Banjir dalam penelitian ini diletakkan sebelum algoritma SVM.

#### 1. Pengambilan Data.

Pengambilan data untuk penelitian ini termasuk dalam proses Data Understanding. Pengambilan dilakukan pada website resmi BNPB, BPS Kabupaten Bandung dan BMKG.

#### 2. Data Cleaning

Sebelum dataset dimasukkan dalam model maka dilakukan proses data cleaning yang meliputi pengisian missing value (data yang kosong), menghaluskan noisy data, mengidentifikasi atau menghilangkan outlier, dan menghilangkan inkonsistensi. Data perlu dilakukan pembersihan sebelum diproses dengan Teknik data mining. Data yang didapat dari kasus nyata (real worlds) biasanya tidak siap digunakan, dalam arti mengandung data yang tidak benar. Hal ini bisa jadi dikarenakan adanya kesalahan instrument faulty, kesalahan yang dilakukan oleh manusia atau kesalahan transmisi. Setelah data di rasa sudah baik untuk digunakan dalam pengolahan data maka di save dalam bentuk CSV.

#### 3. Transformasi

Data yang telah di save, maka data tersebut di import untuk dapat menghasilkan model yang ingin di dapat. Model yang dihasilkan tersebut menggunakan tools yaitu Rapidminer versi 10. Setelah model di dapatkan maka hasil Accuracy dan ROC juga akan di dapatkan.

#### 4. Data Mining

Pada tahapan ini dilakukan 10 fold cross validation yakni proses membagi dataset menjadi 10 bagian yang mana 1 diantara bagian lainnya menjadi data testing, dan yang lainnya menjadi data training. Kemudian dimasukkan kedalam model algoritma Support Vector Machine (SVM). Ini dilakukan bergantian pada tiap bagian data sampai mendapat nilai terbaik dari model ini.

#### 5. Interpretation/Evaluation

Setelah tahapan data mining selesai maka selanjutnya adalah melakukan evaluasi terhadap hasil dari pemodelan tersebut. Membandingkan dua hasil dari pemodelan yang berupa accuracy, precision, recall maupun AUC antara model dengan Feature Seleccion (Particle Swarm Optimization) maupun yang hanya menggunakan algoritma SVM saja.

### 3. Hasil dan Pembahasan

Pada bab ini akan dijelaskan mengenai proses yang dijalankan dalam penelitian ini. Ada beberapa langkah yang diterapkan pada penelitian ini.

#### 3.1. Pengambilan Data

Pengambilan data pada penelitian ini menggunakan data resmi website dari pihak yang terkait dalam objek penelitian ini. Penggunaan Data Miner sangat memudahkan bagi peneliti yang ingin mendapatkan data dari sebuah halaman website. Layanan yang diberikan oleh Data Miner juga memungkinkan pengambilan data di export kedalam bentuk file yang diinginkan.

#### 3.2. Data Cleaning

Setelah pengambilan data dari website, data tersebut tidak bisa langsung dimasukan dalam pengolahan untuk

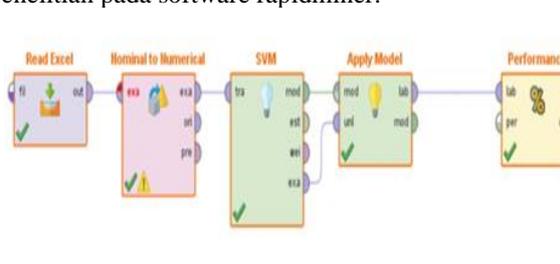
memprediksi bencana banjir, maka dilanjutkan dengan tahapan Data Cleaning.

### 3.3 Feature Selection (Particle Swarm Optimization)

Untuk langkah selanjutnya digunakan operator Particle Swarm Optimization yang berfungsi untuk menghitung secara terus menerus calon solusi dengan menggunakan suatu acuan kualitas. Algoritma ini mengoptimasi permasalahan dengan cara menggerakkan partikel / calon solusi di dalam ruang permasalahan menggunakan fungsi tertentu untuk posisi dan kecepatan dari partikel yang diketahui. Pergerakan partikel dipengaruhi oleh solusi terbaik partikel tersebut, dan solusi terbaik secara umum yang didapatkan dari partikel lainnya. Sekumpulan partikel ini dinamakan swarm, dan pada akhirnya swarm ini akan bergerak menuju kepada solusi terbaik.

### 3.4 Data Mining

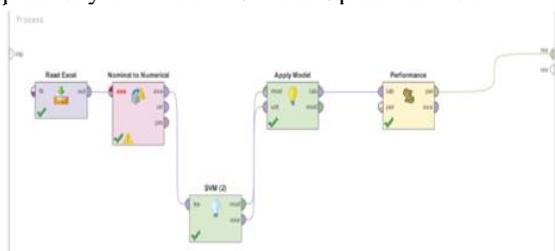
Adapun pemodelan untuk penelitian ini digambarkan secara lengkap terlihat pada gambar 4. Yang merupakan keseluruhan operator yang digunakan dalam penelitian pada software rapidminer.



Gambar 4. Pemodelan SVM tanpa PSO

Pada pemodelan diatas diketahui bahwa penggunaan algoritma SVM tanpa Optimasi.

Dan Gambar 5 merupakan algoritma SVM dengan Optimasi yaitu Particle Swarm Optimatization.



Gambar 5. Pemodelan SVM dengan PSO

### 3.5 Interpretation/Evaluation

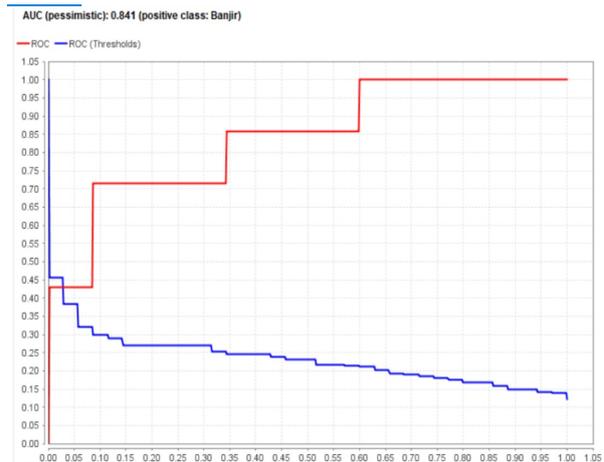
Dan untuk mengevaluasi penelitian ini menggunakan Accuracy dan AUC (Area Under Curve). Hasil SVM tanpa dan dengan seleksi fitur akan dibandingkan untuk mengukur seberapa besar peningkatan yang dihasilkan. Untuk mendapatkan nilai yang akan dievaluasi, Accuracy SVM (Gambar 6).

accuracy: 85.71%

	true Tidak Banjir	true Banjir	class precision
pred. Tidak Banjir	35	6	85.37%
pred. Banjir	0	1	100.00%
class recall	100.00%	14.29%	

Gambar 6. Accuracy SVM

Diketahui bahwa dari Accuracy SVM tanpa feature selection hanya mendapatkan 85.71%.



Gambar 7. AUC SVM

Dan nilai AUC yang di dapat tanpa feature selection hanya mendapatkan sebesar 0,841.

accuracy: 97.62%

	true Tidak Banjir	true Banjir	class precision
pred. Tidak Banjir	34	0	100.00%
pred. Banjir	1	7	87.50%
class recall	97.14%	100.00%	

Gambar 8. Accuracy SVM dengan PSO

Diketahui bahwa dari Accuracy SVM dengan feature selection mendapatkan 97.62%.



Gambar 9. AUC SVM dengan PSO

#### 4. Kesimpulan

Setelah melakukan penelitian dengan model algoritma SVM tanpa feature selection dan dibandingkan dengan model SVM dengan feature selection dapat kita simpulkan bahwa penggunaan feature selection Weight by Correlation dapat meningkatkan nilai dari Accuracy dan AUC. Peningkatan yang didapatkan sangat signifikan yang sebelumnya model SVM tanpa feature selection hanya menghasilkan 66.49% dan nilai AUC 0.716 setelah ditambahkan penggunaan feature selection menjadi 81.18% untuk accuracy dan nilai AUC 0.943.

Jelas hal ini memberikan peningkatan yang sangat tajam dengan nilai selisih untuk accuracy sebesar 14.69 dan 0.227 untuk AUC. Sehingga dapat disimpulkan penggunaan feature selection Weight by Correlation sangat baik apabila digunakan pada algoritma Support Vector Machine (SVM).

Untuk penelitian selanjutnya dapat disarankan menggunakan feature selection Weight lainnya selain feature selection Weight by Correlation. Karena masih ada beberapa operator untuk feature selection contoh seperti feature selection Weight by SVM yang menghitung pembobotan dengan metode SVM. Kemudian pengambilan data juga dapat dilakukan dengan lebih banyak jumlahnya untuk mendapatkan nilai akurasi lebih baik.

#### Ucapan Terimakasih

Terima Kasih kepada LPPM Universitas Mercu Buana yang telah mendanai penelitian ini.

#### Daftar Rujukan

- [1] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2012.
- [2] Agus Ambarwari, Qadhli Jafar Adrian, Yeni Herdiyeni. 2020. Analisis Pengaruh Data Scaling Terhadap Performa Algoritma Machine Learning untuk Identifikasi Tanaman. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*. Vol. 4 No. 1. 117–122. ISSN Media Elektronik: 2580-0760.
- [3] Y. Tang and I. Sutskever, "Data normalization in the learning of restricted Boltzmann machines," in Department of Computer Science, University of Toronto, Technical Report UTML-TR-11-2, 2011.
- [4] E. Sutoyo, I. T. R. Yanto, R. R. Saedudin, and T. Herawan, "A soft set-based co-occurrence for clustering web user transactions," *Telkomnika (Telecommunication Comput. Electron. Control)*, vol. 15, no. 3, 2017.
- [5] E. Sutoyo, I. T. R. Yanto, Y. Saadi, H. Chiroma, S. Hamid, and T. Herawan, "A Framework for Clustering of Web Users Transaction Based on Soft Set Theory," in Springer, 2019, pp. 307–314.
- [6] I. T. R. Yanto, E. Sutoyo, A. Apriani, and O. Verdiansyah, "Fuzzy Soft Set for Rock Igneous Classification," in 2018 International Symposium on Advanced Intelligent Informatics (SAIN), 2018, pp. 199–203.
- [7] E. Sutoyo, R. R. Saedudin, I. T. R. Yanto, and A. Apriani, "Application of adaptive neuro-fuzzy inference system and chicken swarm optimization for classifying river water quality," in *Electrical, Electronics and Information Engineering (ICEEIE)*, 2017 5th International Conference on, 2017, pp. 118–122.
- [8] M.-L. Antonie, O. R. Zaiane, and A. Coman, "Application of data mining techniques for medical image classification," in *Proceedings of the Second International Conference on Multimedia Data Mining*, 2001, pp. 94–101.
- [9] R. R. Saedudin, E. Sutoyo, S. Kasim, H. Mahdin, and I. T. R. Yanto, "Attribute selection on student performance dataset using maximum dependency attribute," in *Electrical, Electronics and Information Engineering (ICEEIE)*, 2017 5th International Conference on, 2017, pp. 176–179.
- [10] H. Chiroma et al., "An intelligent modeling of oil consumption," *Adv. Intell. Syst. Comput.*, vol. 320, 2015.
- [11] A. R. Muhajir, E. Sutoyo, and I. Darmawan, "Forecasting Model Penyakit Demam Berdarah Dengue Di Provinsi DKI Jakarta Menggunakan Algoritma Regresi Linier Untuk Mengetahui Kecenderungan Nilai Variabel Prediktor Terhadap Peningkatan Kasus," *Fountain Informatics J.*, vol. 4, no. 2, pp. 33–40, Nov. 2019.
- [12] N. Iriadi and N. Nuraeni, "Kajian Penerapan Metode Klasifikasi Data Mining Algoritma C4.5 Untuk Prediksi Kelayakan Kredit Pada Bank Mayapada Jakarta," *J. Tek. Komput. AMIK BSI*, vol. 2, 201.
- [13] R. Riszky, M. Sadikin. 2019. Data Mining Menggunakan Algoritma Apriori untuk Rekomendasi Produk bagi Pelanggan. *Jurnal Teknologi dan Sistem Komputer*. 103-108.
- [14] RA Pangestu, S. Rudiarto, D. Fitriana. 2018. Aplikasi Web berbasis Algoritma K-NEAREST NEIGHBOUR untuk Menentukan Klasifikasi Barang STUDI KASUS: PERUM PERURI. *Jurnal Ilmu Teknik dan Komputer*. Vol. 2 No. 1 Januari. ISSN 2548-740X E-ISSN 2621-1491.
- [15] Lukman. 2016. Penerapan Algoritma Support Vector Machine (SVM) dalam Pemilihan Beasiswa: STUDI KASUS SMK YAPIMDA. *Faktor Exacta* 9(1): 49-57, 2016 ISSN: 1979-276X.
- [16] Haddi, E., Liu, X., & Shi, Y., 2013. The Role of Text Pre-processing in Sentiment Analysis. *First International Conference on Information Technology and Quantitative Management*, 17, 26–32. <https://doi.org/10.1016/j.procs.2013.05.05>
- [17] Nahriyatunnur Hidayatus Solihah1), Muliadi1), Arie Antasari Kushadiwijayanto2\*). 2018. Estimasi Parameter Model Curah Hujan Menggunakan Particle Swarm Optimization (PSO): Studi Kasus Ketapang dan Melawi. *Jurnal Fisika FLUX*. 13-19. Volume 15, Nomor 1. ISSN : 2514-1713.
- [18] Mauliana, P., 2016, Prediksi Banjir Sungai Citarum dengan Logika Fuzzy Hasil Algoritma Particle Swarm Optimization. *INFORMATIKA*, 3, 269-276.
- [19] Ary, M., 2017. Aplikasi Prediksi Banjir Metode Fuzzy Logic, Hasil Algoritma Spade dan Algoritma PSO. In: *Konferensi Nasional Ilmu Sosial & Teknologi (KNiST)*, 342-348.
- [20] Nurmahaludin., 2013. Perancangan Algoritma Belajar Jaringan Syaraf Tiruan Menggunakan Particle Swarm Optimization (PSO). *Jurnal POROS TEKNIK*, 5(1), 18-23.
- [21] Factmawati, M., Widodo, B., and Wahyuningsih, N., 2014. Estimasi Autoregressive Integrated Average (ARIMA) Menggunakan Algoritma Particle Swarm Optimization (Studi Kasus: Peramalan Curah Hujan DAS Brangkal, Mojokerto). Surabaya: Skripsi ITS.
- [22] Dedy dan Anis Cherid. "Data Mining Pengolahan Data Calon Pekerja Migran Indonesia (PMI) dengan Penerapan Metode Klustering K-Means dan Metode Klasifikasi K-Nearest Neighbor (KNN): STUDI KASUS PT. SAM". *Jurnal Format Volume 9 Nomor 2 Tahun 2020* :: ISSN : 2089 – 5615 :: E-ISSN : 2722 – 7162.
- [23] Fikriya, Zulfa Afiq; Irawan, Mohammad Isa; Soetrismo, 2017. "Implementasi Extreme Learning Machine untuk Pengenalan Objek Citra Digital", *Jurnal Sains dan Seni ITS*, Vol.6, No. 1. 2337-3520
- [24] Rahmansyah A., Dewi O., Andini P., Hastuti PN, Triana and Eka Suryana, Muhammad. 2016, Membandingkan Pengaruh Feature Selection Terhadap Algoritma Naïve Bayes dan Support Vector Machine. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, 2018 p. A1 - A7.
- [25] Guyon, I., Weston, J., and Barnhill, S. (2002), *Machine Learning, Gene Selection for Cancer Classification using Support Vector Machines*, Netherland, Kluwer Academic Publishers.