

**PREDIKSI JUMLAH PENDERITA DEMAM BERDARAH DENGUE DI
KOTA BANDUNG MENGGUNAKAN SUPPORT VECTOR REGRESSION
DAN GAUSSIAN PROCESS REGRESSION**

***PREDICTION OF DENGUE HEMORRHAGIC FEVER SUFFERERS IN
BANDUNG USING SUPPORT VECTOR REGRESSION AND GAUSSIAN
PROCESS REGRESSION***

Ardhya Chaeruna Salim¹, Dr. Tito Waluyo Purboyo, S. Si., S.T., M.PMat², Dr. Meta
Kallista S.Si., M.Si.³

¹Prodi S1 Teknik Komputer, Fakultas Teknik Elektro, Universitas Telkom

²Prodi S1 Teknik Komputer, Fakultas Teknik Elektro, Universitas Telkom

³Prodi S1 Teknik Komputer, Fakultas Teknik Elektro, Universitas Telkom

¹ardhyachaerunas@student.telkomuniversity.ac.id, ²titowaluyo@telkomuniversity.ac.id,

³metakallista@telkomuniversity.ac.id

Abstrak

Dalam jurnal tugas akhir ini mencakup suatu rancangan sistem untuk memprediksi penyebaran infeksi penyakit demam berdarah berbasis web. Demam berdarah hingga saat ini masih menjadi permasalahan dalam kesehatan masyarakat di Indonesia karena hampir seluruh provinsi terjangkit penyakit demam berdarah. Maka dari itu dibutuhkan prediksi penderita demam berdarah pada masa mendatang untuk menanggulangi meningkatnya infeksi DBD pada masyarakat. Rancangan sistem ini menerapkan metode klasifikasi dan regresi dengan algoritma Support Vector Regression (SVR) dan Gaussian Process Regression (GPR) dengan menggunakan bahasa pemrograman python dan berbasis web.

Kata kunci : Demam berdarah, *Support Vector Regression, Gaussian Process Regression*

Abstract

In this final project journal includes a system design to predict the spread of web-based dengue fever infections. Dengue fever is still a problem in public health in Indonesia because almost all provinces have dengue fever. Therefore, it takes predictions of dengue fever in the future to cope with increasing DHF infection in the community. This system design employs classifications and regression methods with the Support Vector Regression (SVR) and Gaussian Process Regression (GPR) algorithms using a Python and web-based programming language.

Keywords: *Dengue Hemorrhagic Fever, Support Vector Regression, Gaussian Process Regression*

1. Pendahuluan

Penyakit demam berdarah (DBD) merupakan penyakit yang memiliki tingkat kematian yang tergolong cukup tinggi di Indonesia [1]. Penyakit ini disebabkan oleh gigitan nyamuk *Aedes aegypti* dan penyakit ini masih termasuk dalam kategori penyakit berbahaya. Penyakit ini disebabkan oleh gigitan nyamuk *Aedes aegypti* dan penyakit ini masih termasuk dalam kategori penyakit berbahaya. Selain disebabkan oleh gigitan nyamuk penyakit DBD ini bisa disebabkan dari cuaca dan kurangnya kebersihan di lingkungan masyarakat.

Menurut kementerian kesehatan kasus demam berdarah dengue (DBD) dari 29 Januari 2019 mencapai 13.683 kasus dengan 133 meninggal dunia, terus bertambah hingga tanggal 3 Februari 2019 ada 16.692 kasus dengan 169 orang meninggal dunia. Berdasarkan data diatas maka

diperlukannya penanganan khusus terhadap penyebaran penyakit demam berdarah, dengan cara memprediksi penyebarannya dimasa yang akan datang.

Diharapkan dengan pembuatan sistem ini dapat membantu Dinas Kesehatan Kota Bandung dalam menangani penyebaran penyakit demam berdarah yang terjadi di Kota Bandung serta dapat mengetahui dan memprediksi kenaikan angka korban demam berdarah di Kota Bandung, dari hasil tersebut dapat digunakan sebagai referensi atau acuan oleh dinas kesehatan untuk mengantisipasi penyakit demam berdarah. Data yang digunakan jumlah kasus korban terinfeksi DBD dan Jumlah Korban Meninggal Akibat DBD pada rentang tahun 2010 hingga 2020 di Kota Bandung.

2. Dasar Teori

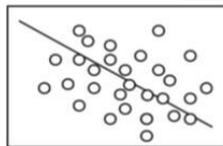
Demam berdarah dengue (DBD) merupakan penyakit menular yang cukup berbahaya diseluruh dunia [1], terutama pada daerah daerah yang beriklim tropis dan subtropis. Di Indonesia penyakit ini merupakan salah satu penyakit yang memiliki angka kesakitan yang cukup tinggi dan cenderung meningkat dari tahun ke tahun, serta daerah yang terjangkau daerah yang terjangkau semakin meluas mencapai 400 kabupaten/kota di Indonesia [2].

Penyakit DBD ini biasa ditularkan melalui nyamuk *Aedes Aegypti*, nyamuk ini mirip dengan nyamuk *Anophelini* yang memiliki metamorfosis sempurna yaitu dimulai dari telur-larva-kepompong-nyamuk. Pada fase telur, larva, dan kepompong didalam air [4]. Nyamuk ini biasanya berkembangbiak di tempat-tempat yang tidak langsung menyentuh tanah, yaitu pada penampungan air seperti, drum, kamar mandi, dan ember.

Selama musim hujan tempat berkembangbiak nyamuk akan dipenuhi dengan air sehingga perkembangbiakan nyamuk akan meningkat. Maka dari itu, harus adanya penanggulangan untuk mencegah perkembangbiakan nyamuk, dengan cara membersihkan setiap bak penampungan supaya tidak menjadi tempat perkembangbiakan nyamuk

2.1. Regression

Regression merupakan suatu metode statistik untuk mengidentifikasi hubungan antara variabel [3]. Hubungan variabel tersebut diwujudkan dalam suatu proses matematis yang bersifat fungsional.

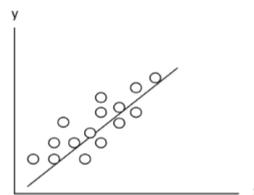


Gambar 2.1 Metode Analisis Regresi [3]

Regression dibagi menjadi beberapa jenis, yaitu:

1. Regression Sederhana

Merupakan regresi secara linear antara satu variabel independen (X) dengan variabel independen (Y).

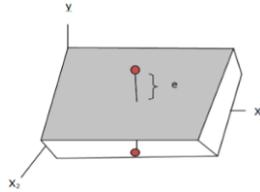


Gambar 2.2 Regression Sederhana [3]

Analisis ini digunakan untuk mengetahui arah hubungan antara variabel independen dengan variabel dependen, apakah positif atau negatif serta digunakan untuk memprediksi nilai dari variabel dependen apabila nilai dari variabel independen mengalami kenaikan atau penurunan nilai.

2. Regression Berganda

Merupakan model regresi atau prediksi yang melibatkan lebih dari satu variabel bebas atau prediktor. Istilah regresi berganda bisa disebut juga *multiple regression*.

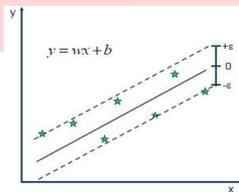


Gambar 2.3 *Regression* Berganda [3]

Prediksi adalah suatu perkiraan sistematis mengenai suatu proses yang akan terjadi di periode selanjutnya berdasarkan sumber informasi dari periode sebelumnya dan periode yang diperoleh saat ini untuk memperkecil suatu kesalahan.

2.2. Support Vector Regression

Support Vector Regression (SVR) merupakan pengembangan dari SVM untuk kasus regresi dalam output berupa bilangan riil (nyata) atau data sekuensial [3]. SVR merupakan metode yang dapat mengatasi *overfitting*. Artinya suatu kondisi dimana model dibangun dengan memperhitungkan seluruh ciri yang ada, termasuk *noise* [4].

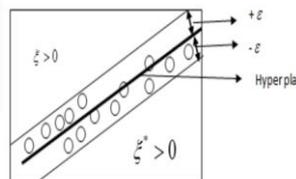


Gambar 2.4 Fungsi SVR [5].

Tujuan SVR adalah untuk membuat semua dataset masuk kedalam satu zona dengan tetap meminimalisir nilai epsilon (ε). Semakin kecil nilai epsilon (ε) maka hasil prediksi mendekati nilai nyatanya dengan berbagai faktor lainnya.

2.2.1 Linear Support Vector Regression

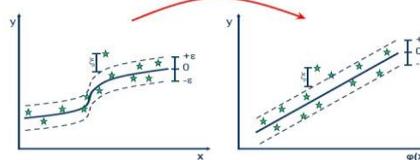
Support Vector Regression menggunakan fungsi linear kernel untuk regresi yang mirip untuk mendukung mesing vektor tetapi SVR menetapkan margin toleransi (ε) untuk pendekatan [3].



Gambar 2.5 Support Vector Regression Linear [3]

2.2.2. Non Linear Support Vector Regression

Non-linear Support Vector Regression, fungsi kernel non-linear yang digunakan untuk pengolahan data pelatihan dalam ruang fitur [1].

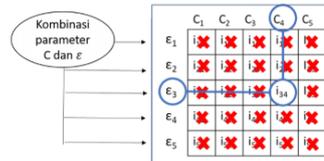


Gambar 2.6 Non-Linear SVR [4]

Fungsi kernel mengubah data menjadi ruang fitur dimensi yang lebih tinggi untuk memungkinkan dilakukannya pemisahan linear, artinya penyebaran data mengikuti garis *hyperline* [6].

2.3. Grid Search Cross-Validation

Algoritme grid search yaitu salah satu Algoritme umum yang sering digunakan untuk estimasi parameter, dengan prinsip kerjanya dengan menentukan beberapa nilai parameter pada rentang tertentu, kemudian memilih parameter pada nilai terbaik pada rentang tersebut dan melakukan pencarian berulang pada grid (rentang nilai) yang lebih kecil, dst [7]. Metode grid search digunakan untuk mendapatkan model SVR dengan parameter C , ϵ dan γ yang optimal.

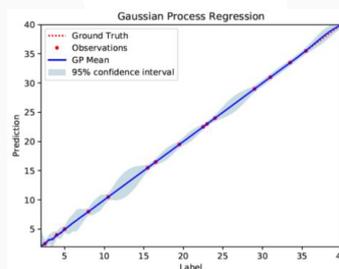


Gambar 2.7 Ilustrasi Algoritma Grid Search Cross Validation [8]

. Pada gambar 2.5.1 merupakan ilustrasi 2 parameter SVR yang akan dicari kombinasi parameter yang optimal. Parameter yang optimal adalah pasangan parameter yang memiliki akurasi terbaik yaitu mendekati nilai 1 yang didapat dari hasil uji *cross-validation*, selanjutnya digunakan untuk model SVR terbaik.

2.4. Gaussian Process Regression

Gaussian Process Regression adalah suatu proses stokastik dimana himpunan berhingga manapun dari himpunan peubah acak Y mempunyai sebaran bersama *Gaussian* ganda.



Gambar 2.8 Gaussian Process Regression [9]

Gaussian Process Regression dapat diturunkan dari sudut pandang regresi nonparametric Bayesian yaitu dengan menempatkan secara langsung sebaran prior Gaussian bagi fungsi-fungsi regresi $f(x)$.

3. Pembahasan

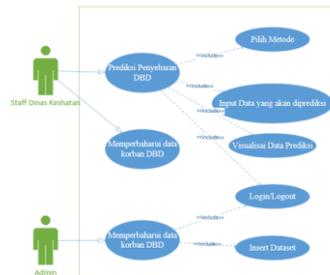
Aplikasi yang dibuat merupakan aplikasi berbasis web dengan menggunakan bahasa pemrograman *python* yang digunakan untuk pengolahan data dan menggunakan bahasa *php* untuk tampilan web. Pada penelitian ini akan dilakukan sebuah perbandingan antara dua metode *machine learning* yaitu *Support Vector Regression* dan *Gaussian Process Regression* dalam pengolahan data infeksi DBD Kota Bandung.

Tujuan dari perancangan aplikasi berbasis web ini untuk memprediksi jumlah korban penyakit DBD di Kota Bandung yang dimana akan digunakan oleh pihak dinas kesehatan Kota Bandung melakukan strategi dan pencegahan untuk menangani penyebaran infeksi DBD secara efektif supaya korban yang terkena DBD tidak akan bertambah dalam jumlah besar.

3.1. Perancangan Sistem

3.1.1. Use Case Diagram

Use case diagram ini digunakan untuk menggambarkan sebuah interaksi antar *user* dengan sistem. Berikut ini adalah *Use Case Diagram* dari Sistem informasi Prediksi Penyebaran Penyakit DBD di Kota Bandung :



Gambar 3.1 Use Case Diagram Sistem yang Diusulkan

Pada sistem ini memiliki dua aktor yaitu Admin dan Staff Dinas Kesehatan. Sistem ini dimulai dari staff Dinas Kesehatan yang mengumpulkan dan memperbarui data korban DBD setiap bulan.

3.2. Pengujian

Pada penelitian ini, peneliti melakukan pengujian pada sistem yaitu pengujian *alpha*. Pengujian *alpha* dilakukan dengan cara mengamati hasil eksekusi melalui data uji dan memeriksa fungsional dari perangkat lunak. Setelah melakukan pengujian dari perangkat lunaknya, selanjutnya akan dilakukan hasil perbandingan prediksi dari setiap metode.

3.2.1. Pengujian Alpha

Dilakukannya pengujian alpha bertujuan untuk menguji fungsional pada aplikasi, apakah aplikasi yang dirancang dapat berjalan sesuai dengan perancangan yang telah dibuat. Dilakukannya pengujian ini dapat mengidentifikasi kesalahan yang terjadi selama perancangan aplikasi sebelum digunakan oleh pengguna. Pengujian alpha terdiri dari pengujian *white box*, dan *black box*.

3.2.2. Pengujian Beta

Pengujian beta melibatkan Teknik Kotak Hitam (*Black Box*) atau uji fungsionalitas aplikasi oleh admin untuk dapat memastikan apakah aplikasi dapat digunakan baik oleh Staff Dinas Kesehatan Kota Bandung.

4. Kesimpulan

Berdasarkan perbandingan performa akurasi kedua metode, ternyata dari kedua metode memiliki performa akurasi yang baik setelah melakukan validasi data hasil prediksi dengan data *real* memiliki selisih 1 angka dari data real. Jadi bisa dikatakan kedua metode ini layak digunakan.

Daftar Pustaka

- [1] V. S. R. R. Kavitha S, "A comparative analysis of linear regression and Support of Vector Regression," *International Conference*, p. 2, 2016.
- [2] S. Rutgers, "Support Vector Machine - Regression (SVR)," 2019. [Online]. Available: https://www.saedsayad.com/support_vector_machine_reg.htm. [Accessed 11 February 2019].
- [3] R. K. Mariette Awad, *Support Vector Regression : Efficient Learning Machines, Efficient Learning Machines*, Apress, Berkeley, CA. IEEE, 2015.
- [4] T Sutoyo, E Mulyanto, V Suhartono, OD Nurhayati, "Kecerdasan Buatan," *ANDI*, 2011.
- [5] H. Sit, "A quickQuick Start to Gaussian Process Regression," *A quick guide to understanding Gaussian process regression (GPR) and using scikit-learn's GPR package*, pp. [Online]. Available:<https://towardsdatascience.com/quick-start-to-gaussian-process-regression-36d838810319>. [Accessed 26 Febuary,2020]., 19 June 2019.
- [6] M. M. M. d. F. Nhita, "The Spreading Prediction of Dengue Hemorrhagic Fever (DHF) In Bandung Regency Using K-Means Clustering and Support Vector Machine Algorithm," *International Conference on Information and Communication Technology (ICoICT)*, vol. 6, 2018.
- [7] K. K. R. Indonesia, *Situasi Penyakit Demam Berdarah di Indonesia*, Indonesia: Kementerian Republik Indonesia, 2017.
- [8] B. A. d. H. I. Mustakim, "Peformance Comparison Between Support Vector Regression and Artificial Neural Network for Prediction," *Jurnal Ilmu Komputer dan Informasi (Journal of Computer Science and information)*, vol. 9, p. 1, 2016.
- [9] R. A. d. S. A. Divya Tomar, "Prediction of Profitability of Industries using Weighted SVR," *International Journal on Computer Science and Engineering (IJCSE)*, vol. 3, p. 5, 2011.
- [10] H. Y. d. A. W. B. Rezzy Eko Caraka, "Peramalan Crude Palm Oil (CPO) Menggunakan Support Vector Regression Kernel Radial Basis," *E-Journal Universitas Udayana*, vol. 7, 2007.

