Dari table 6 kita dapat melihat rata-rata akurasi masing-masing data set sebesar 0.9826 data latih sedangkan untuk data uji sebesar 0.9863. Kernel terbaik untuk memprediksi turunan senyawa DPP-IV adalah kernel RBF dengan akurasi sebesar 0.9869, *precision* 0.9745, *recall* 1.0, dan *F1-Score* 0.9871 hasil dari data uji. Di urutan selanjutnya adalah kernel poly lalu di ikuti oleh kernel linear dengan akurasi sebesar 0.9851, *precision* 0.9744, *recall* 0.9963, dan *F1-Score* 0.9853. Berdasarkan hasil dari data latih dan data uji, setiap kernel menunjukan konsistensi dalam memprediksi senyawa turunan DPP-IV

D. Kesimpulan

Senyawa inhibitor DPP-IV diketahui membawa kandidat obat anti diabetes baru menurut beberapa studi QSAR. Pada penelitian ini tahap seleksi fitur dilakukan dengan menghitung 100 korelasi tertinggi masing-masing deskriptor terhadap pIC50 menggunakan PCC dan menggunakan metode Algortima Genetika (AG) untuk mendapatkan kombinasi deskriptor terbaik dari dataset tersebut. Model prediksi berhasil dibangun menggunakan metode AG dan SVM. AG berhasil mendapatkan kombinasi fitur yang baik dan SVM dapat digunakan untuk memprediksi aktifitas inhibitor DPP-IV dengan akurasi rata-rata sebesar 0.983 untuk data latih dan 0.986 untuk data uji. Hasil juga menunjukan bahwa kernel RBF merupakan kernel terbaik dalam memprediksi senyawa DPP-IV dengan akurasi sebesar 0.9869, *precision* 0.9745, *recall* 1.0, dan *F1-Score* 0.9871.

Daftar Pustaka

[1] UCLA Center for Health Policy Research, “Health Impact of Diabetes,” *Public Health Advocacy*, no. May, 2014.

[2] U. Saqib and M. I. Siddiqi, “3D-QSAR studies on triazolopiperazine amide inhibitors of dipeptidyl peptidase-IV as anti-diabetic agents,” *SAR and QSAR in Environmental Research*, vol. 20, no. 5–6, pp. 519–535, 2009, doi: 10.1080/10629360903278677.

[3] A. M. Al-Fakih, Z. Y. Algamal, M. H. Lee, M. Aziz, and H. T. M. Ali, “A QSAR model for predicting antidiabetic activity of dipeptidyl peptidase-IV inhibitors by enhanced binary gravitational search algorithm,” *SAR and QSAR in Environmental Research*, vol. 30, no. 6, pp. 403–416, 2019, doi: 10.1080/1062936X.2019.1607899.

[4] D. P. P. Iv, “Studi Qsar Senyawa Turunan Triazolopiperazin Amida Sebagai Inhibitor Enzim Dipeptidil Peptidase-IV (DPP IV) Menggunakan Metode Semiempirik AM1,” *Bimipa*, vol. 23, no. 3, pp. 288–296, 2013.

[5] B. D. Green, P. R. Flatt, and C. J. Bailey, “Dipeptidyl peptidase IV (DPP IV) inhibitors: A newly emerging drug class for the treatment of type 2 diabetes,” *Diabetes and Vascular Disease Research*, vol. 3, no. 3, pp. 159–165, 2006, doi: 10.3132/dvdr.2006.024.

[6] I. M. Al-masri, M. K. Mohammad, and M. O. Taha, “Discovery of DPP IV inhibitors by pharmacophore modeling and QSAR analysis followed by in silico screening,” *ChemMedChem*, vol. 3, no. 11, pp. 1763–1779, 2008, doi: 10.1002/cmdc.200800213.

[7] P. Bharatam, D. Patel, L. Adane, A. Mittal, and S. Sundriyal, “Modeling and Informatics in Designing Anti-Diabetic Agents,” *Current Pharmaceutical Design*, vol. 13, no. 34, pp. 3518–3530, 2007, doi: 10.2174/138161207782794239.

[8] “Top 4 advantages and disadvantages of Support Vector Machine or SVM | by Dhiraj K | Medium.” https://dhirajkumarblog.medium.com/top-4-advantages-and-disadvantages-of-support-vector-machine-or-svm-a3c06a2b107 (accessed Feb. 07, 2021).

[9] M. Lorca *et al.*, “Structure-activity relationships based on 3D-QSAR CoMFA/CoMSIA and design of aryloxypropanol-amine agonists with selectivity for the human β3-adrenergic receptor and anti-obesity and anti-diabetic profiles,” *Molecules*, vol. 23, no. 5, 2018, doi: 10.3390/molecules23051191.

[10] A. Allouche, “Software News and Updates Gabedit — A Graphical User Interface for Computational Chemistry Softwares,” *Journal of computational chemistry*, vol. 32, pp. 174–182, 2012, doi: 10.1002/jcc.

[11] U. Maulik and S. Bandyopadhyay, “Genetic algorithm-based clustering technique,” *Pattern Recognition*, vol. 33, no. 9, pp. 1455–1465, 2000, doi: 10.1016/S0031-3203(99)00137-5.

[12] S. Forrest, “Genetic algorithms: Principles of natural selection applied to computation,” *Science*, vol. 261, no. 5123, pp. 872–878, 1993, doi: 10.1126/science.8346439.

[13] C. Shen, L. Wang, and Q. Li, "Optimization of injection molding process parameters using combination of artificial neural network and genetic algorithm method," *Journal of Materials Processing Technology*, vol. 183, no. 2–3, pp. 412–418, 2007, doi: 10.1016/j.jmatprotec.2006.10.036.

[14] J. Benesty, J. Chen, and Y. Huang, "On the importance of the pearson correlation coefficient in noise reduction," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 4, pp. 757–765, 2008, doi: 10.1109/TASL.2008.919072.

[15] J. Shapiro, "Genetic algorithms in machine learning," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 2049 LNAI, pp. 146–168, 2001, doi: 10.1007/3-540-44673-7_7.

[16] S. K. Shevade, S. S. Keerthi, C. Bhattacharyya, and K. R. K. Murthy, "Improvements to the SMO algorithm for SVM regression," *IEEE Transactions on Neural Networks*, vol. 11, no. 5, pp. 1188–1193, 2000, doi: 10.1109/72.870050.

[17] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya, and K. R. K. Murthy, "A fast iterative nearest point algorithm for support vector machine classifier design," *IEEE Transactions on Neural Networks*, vol. 11, no. 1, pp. 124–136, 2000, doi: 10.1109/72.822516.

[18] S. H. Min, J. Lee, and I. Han, "Hybrid genetic algorithms and support vector machines for bankruptcy prediction," *Expert Systems with Applications*, vol. 31, no. 3, pp. 652–660, 2006, doi: 10.1016/j.eswa.2005.09.070.

[19] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004, doi: 10.1023/B:STCO.0000035301.49549.88.

[20] J. T. Informasi, D. Kurniawan, P. Pascasarjana, M. Teknik, I. Universitas, and D. Nuswantoro, "Optimasi Algoritma Support Vector Machine (Svm)," vol. 9, no. April, pp. 38–49, 2013.

[21] A. Kulkarni, D. Chong, and F. A. Batarseh, *Foundations of data imbalance and solutions for a data democracy*. Elsevier Inc., 2020.

[22] A. Luque, A. Carrasco, A. Martín, and A. de las Heras, "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognition*, vol. 91, pp. 216–231, 2019, doi: 10.1016/j.patcog.2019.02.023.