IMPLEMENTASI METODE *MODIFIED K-NEAREST NEIGHBOR* UNTUK PREDIKSI HASIL *TREATMENT* PENYAKIT HEPATITIS C

Mohammad Reza Destyar[1], Deni Arifianto[2], Qurrota A'yun[3]
Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember
Email : rezadestyar21@gmail.com

ABSTRAK

Virus Hepatitis C (HCV) merupakan penyakit peradangan yang menyerang hati. Akibat atau terinfeksi penyakit Virus Hepatitis C bisa menimbulkan sirosis pada hati. Penyakit Virus Hepatitis C sering tidak menimbulkan gejala awal, sehingga menyebabkan penyakit Virus Hepatits C menjadi kronis (bisa menyebabkan kematian) karena keterlambatan dalam penanganannya, juga menyebabkan tingkat penularannya semakin tinggi. Dengan memanfaatkan perkembangan teknologi, memprediksi Hasil *Treatment* gejala penyakit Virus Hepatitis C dapat dilakukan lebih awal menggunakan sistem Prediksi Hasil *Treatment* penyakit Virus Hepatitis C berbasis web. Metode *Modified K-Nearest Neighbor* merupakan pengembangan dari metode konvensional *K-Nearest Neighbor*. Tujuan dari penelitian ini untuk mengetahui nilai akurasi, presisi, recall pada metode *Modified K-Nearest Neighbor* (MKNN) untuk Prediksi Hasil *Treatment* penyakit Virus Hepatitis C. Hasil pengujian dari tujuan penelitian Tugas Akhir ini yaitu untuk mengukur nilai akurasi, presisi dan recall, didapatkan hasil nilai akurasi tertinggi ada pada *5-fold* skenario 5 dengan akurasi = 97%. Nilai presisi tertinggi yaitu menghasilkan 100%. Ada pada *4-fold* skenario 1 dan skenario 2 pada kelas kronis, *10-fold* skenario 1, skenario 2, skenario 3, skenario 5, skenario 7, skenario 9, dan skenario 10 pada kelas sedang, *10-fold* skenario 6 pada kelas berat. Nilai recall tertinggi ada pada *5-fold* skenario 5.

Kata Kunci : Klasifikasi, Hepatitis, Hepatitis C Virus, *Metode Modified K-Nearest Neighbor.*

.

ABSTRACT

Hepatitis C virus (HCV) is an inflammatory disease that attacks the liver. As a result or infection with Hepatitis C virus can cause cirrhosis of the liver. Hepatitis C virus disease often does not cause early symptoms, causing Hepatits C virus to become chronic (can cause death) due to delays in handling, also causing a higher transmission rate. By utilizing technological developments, diagnosing the symptoms of the Hepatitis C Virus can be done earlier using a web-based Hepatitis C Virus disease diagnosis system. The Modified K-Nearest Neighbor method is a development of the conventional K-Nearest Neighbor method. The purpose of this study is to determine the value of accuracy, precision, recall in the Modified K-Nearest Neighbor (MKNN) method for the diagnosis of Hepatitis C virus. The test results of the research objectives of this final project are to measure the value of accuracy, precision and recall. The highest accuracy is in 5-fold scenario 5 with accuracy = 97%. The highest precision value is to produce 100%. There are 4-fold scenario 1 and scenario 2 chronic class, 10-fold scenario 1, scenario 2, scenario 3, scenario 5, scenario 7, scenario 9, and scenario 10 in medium class, 10-fold scenario 6 in heavy class. The highest recall value is in the 5-fold scenario 5

Keywords : Classification, Hepatitis, Hepatitis C Virus, Modified K-Nearest Neighbor Method

1. PENDAHULUAN

Hepatitis didefinisikan sebagai suatu penyakit yang ditandai dengan adanya peradangan pada hati. Hepatitis merupakan suatu proses terjadinya inflamasi atau nekrosispada jaringan hati yang dapat disebabkan oleh infeksi, obat-obatan, toksin, gangguan metabolik, maupun kelainan sistem antibodi. Infeksi yang disebabkan virus merupakan penyebab paling banyak dari Hepatitis akut. Terdapat beberapa jenis virus penyebab utama infeksi akut, yaitu virus Hepatitis A, B, dan C. Penyakit Hepatitis yang disebabkan oleh virus, menduduki tempat pertama dalam hal jumlah dan penyebarannya yang di akibatkan oleh virus (Arief, 2012).

Menurut (Nasr et al., 2018) Mesir memiliki jumlah kasus penyakit tertinggi Virus Hepatitis C di dunia dengan tingkat jumlah kasus penyakit yang mencapai 13% - 15% dan menjadikan Virus Hepatitis C sebagai masalah kesehatan masyarakat yang utama dan juga sebagai masalah ekonomi.

Dalam klasifikasi *data mining* terdapat banyak metode yang dapat diterapkan untuk mengklasifikasi suatu data. Salah satunya adalah *Modified K-Nearest Neighbor* (MKNN). Penerapan metode *Modified K-Nearest Neighbor* (MKNN) sebagai proses yang bertujuan untuk prediksi hasil *treatment* pasien penyakit Hepatitis C. Oleh karena itu, penulis memutuskan menggunakan metode *Modified K-Nearest Neighbor* (MKNN), dikarenakan pada penelitian sebelumnya yang dilakukan oleh (Muhammad, 2015)

2. PENELITIAN TERKAIT

A. Hepatitis C Virus

Hepatitis C merupakan peradangan yang terjadi pada hati, yang dapat disebabkan oleh suatu proses infeksi, seperti virus, bakteri, jamur, dan parasit,atau bukan infeksi, seperti alkohol, obat, autoimun, dan metabolik. Proses peradangan yang terjadi pada hati dapat bersifat akut maupun kronis. Pada keadaan kronis, terjadi peradangan dalam kurun waktu minimal enam bulan dengan tingkat kerusakan yang bervariasi (Blachier et al., 2013).

Salah satu bentuk penularan infeksi HCV adalah melalui transfusi darah, sehingga diperlukan pemeriksaan yang cepat, aman dan teliti. Pemeriksaan darah donor di PMI cabang Surabaya saat ini menggunakan alat otomatisasi dengan metode Enzyme Linked Immunosorbent Assay (ELISA) dan Chemiluminescent Microparticle Immunoassay (CMI A). Pemeriksaan ini memiliki nilai diagnostik yang baik, tetapi memerlukan waktu lebih lama dan tata langkah lebih rumit (Permatasari et al., 2018).

B. *Data Mining*

Data mining adalah istilah yang digunakan menguraikan penemuan pengetahuan di dalam database. Menggunakan teknik statistik, matematika, kecerdasan tiruan, dan *machine learning* untuk mengestraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* yang berukuran besar (Prasetyo, 2013)

C. Klasifikasi

Klasifikasi merupakan suatu teknik yang diterapkan untuk memprediksi kelas pada setiap instance data. Teknik ini dilakukan dengan memanipulasi data yang tersedia dan telah diklasifikasikan sehingga dapat menghasilkan sejumlah aturan (Oktaviana, 2016)

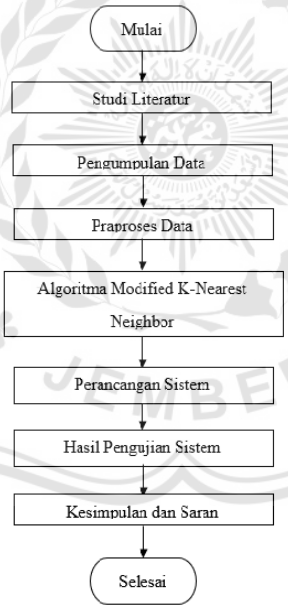
D. *K-fold Cross Validation*

K-Fold Cross Validation merupakan teknik untuk melakukan validasi pada dataset untuk menemukan akurasi yang baik. Teknik ini membagi dataset sebanyak k subset. Satu dari subset ini akan dijadikan sebagai data uji dan $k - 1$ subset sisanya digunakan untuk proses data latih (Haltuf, 2011).

E. Metode *Modified K-Nearest Neighbor* (MKNN)

Metode *Modified K-Nearest Neighbor* merupakan pengembangan dari metode *K-Nearest Neighbor* yang diusulkan oleh (Parvin et al., 2010) dengan penambahan 2 proses didalamnya yaitu perhitungan nilai validitas dan perhitungan *weighted voting* (pembobotan). Sedangkan untuk algoritma *K-Nearest Neighbor* sendiri merupakan algoritma klasifikasi yang sederhana untuk mengelompokkan data baru dengan k tetangga terdekat.

F. Tahapan Penelitian



Gambar 1. Diagram Alur Penelitian

G. Studi Literatul

Studi Literatur digunakan untuk mencari referensi teori yang berkaitan dengan kasus atau permasalahan yang ditemukan. Studi literatur pada penelitian ini dengan melakukan pencarian jurnal-jurnal dari penelitian sebelumnya yang nantinya dapat mendukung dan dapat dijadikan rujukan atau referensi yang akan memperkuat argumentasi-argumentasi yang ada. Maka penulis perlu mempelajari beberapa literatur yang digunakan, kemudian

literatur tersebut diseleksi untuk dapat ditentukan literatur mana yang akan digunakan dalam penelitian.

H. Pengumpulan Data

Pengumpulan data diperoleh dari link *UCI Machine Learning Repository* yaitu berupa dataset Hepatitis C Virus *For Egyptian Patient*, dengan jumlah data sebanyak 1000 data. Dataset tersebut akan diolah menggunakan algoritma *Modified K-Nearest Neighbor*. Dan dataset ini terdapat 28 atribut.

I. Implementasi Modifed K-Nearst Neighbor

Gambar 2. Flowchart MKNN

3. HASIL DAN PEMBAHASAN

Implementasi sistem adalah lanjutan dari perancangan yang sudah dibahas pada Bab III. Pada bab hasil dan pembahasan akan dijelaskan cara kerja dan hasil dari sistem (web) yang telah dibuat. Dalam pembuatan sistem **Implementasi *Modified K-Nearest Neighbor* Untuk Prediksi Hasil Treatment Penyakit Hepatitis C** berbasis web ini dilakukan menggunakan bahasa pemrograman *PHP*, *MySQL* sebagai *database management system* dan *xampp* sebagai *web server*.

A. Pengujian Akurasi, Presisi, dan Recall

1. Skenario dan Hasil Pengujian

Pengujian terhadap *k-fold* dan nilai k = 3 yang dilakukan yaitu menggunakan *2-fold*, *4-fold*, *5-fold*, dan *10-fold* dengan jumlah data yang berjumlah 1000. Berikut nilai akurasi, presisi, dan recall pengujian *k-fold*.

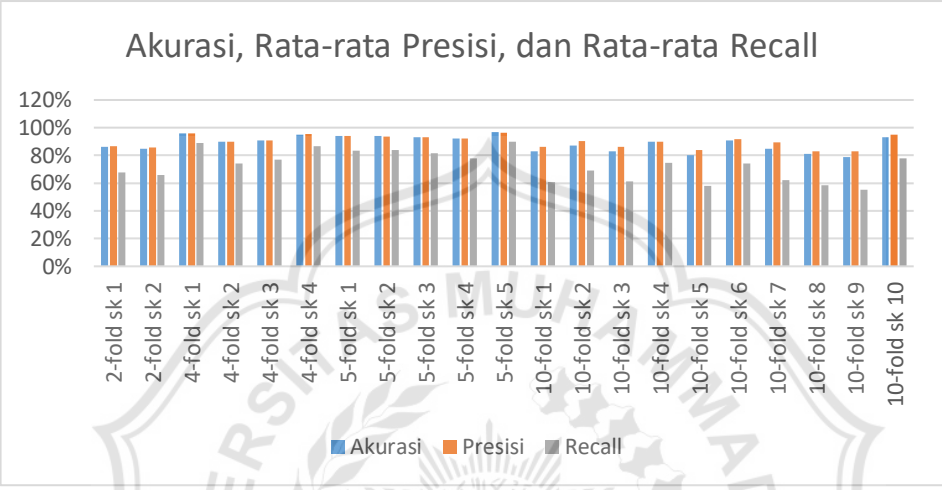
Berikut nilai akurasi, presisi, dan recall pengujian *k-fold* dapat dilihat pada Tabel 1.

Tabel 1. Nilai Akurasi, Presisi, dan Recall Pengujian *k-fold*

k-values	*k-fold*	Akurasi	Presisi				Recall
			ringan	sedang	berat	kronis/ akut	
3	2-fold skenario 1	86%	81%	81%	86%	98%	67,75%
	2-fold skenario 2	85%	89%	79%	75%	99%	65,75%
	4-fold skenario 1	96%	97%	94%	93%	100%	89%
	4-fold skenario 2	90%	81%	90	88	100%	74,25%
	4-fold skenario 3	91%	89%	88%	90%	97%	76,75%
	4-fold skenario 4	95%	92%	96%	95%	98%	79,25%
	5-fold skenario 1	94%	95%	95%	98%	89%	83,50%
	5-fold skenario 2	94%	86%	95%	97%	96%	84%
	5-fold skenario 3	93%	88%	92%	98%	95%	81,50%
	5-fold skenario 4	92%	86%	93%	95%	95%	77,75%
	5-fold skenario 5	97%	96%	95%	98%	97%	89,75%
	10-fold skenario 1	83%	74%	100%	88%	83%	61%
	10-fold skenario 2	87%	89%	100%	95%	72%	69%
	10-fold skenario 3	83%	66%	100%	96%	82%	61,25%
	10-fold skenario 4	90%	81%	95%	96%	87%	74,75%
	10-fold skenario 5	80%	64%	100%	89%	82%	58,25%
	10-fold skenario 6	91%	91%	90%	100%	86%	74%
	10-fold skenario 7	85%	80%	100%	94%	83%	62,25%
	10-fold skenario 8	81%	70%	91%	89%	81%	58,50%
	10-fold skenario 9	79%	68%	100%	71%	92%	55,25%

	10-fold skenario 10	93%	89%	100%	85%	90%	77,75%

Berdasarkan hasil dari pengujian pada tabel 1 menunjukan bahwa pengujian yang dilakukan menggunakan nilai k = 3 dengan *2-fold, 4-fold, 5-fold, dan 10-fold*. Menunjukan hasil akurasi tertinggi sebesar 97% pada *5-Fold Skenario 5*, hasil presisi tertinggi sebesar 100% pada kelas penyakit sedang dan berat, dan hasil Recall tertinggi sebesar 89,75% pada *5-Fold skenario 5*. Berikut grafik hasil pengujian yang ditunjukkan pada Gambar 3.



Gambar 3. Diagram Akurasi, Presisi dan Recall Pada Pengujian k-fold

4. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan penelitian melakukan prediksi hasil treatment pasien penyakit hepatitis c menggunakan metode *Modified K-Nearest Neighbor*, maka dapat diambil kesimpulan sebagai berikut :

1. Nilai akurasi yang diperoleh dari menggunakan metode *Modified K-Nearest Neighbor* dengan pengujian *k-fold*, nilai akurasi tertinggi ada pada *5-fold* skenario 5 dengan akurasi = 97%.
2. Nilai presisi yang diperoleh dari menggunakan metode *Modified K-Nearest Neighbor* dengan pengujian *k-fold*, nilai presisi tertinggi yaitu menghasilkan 100%. Pada *4-fold* skenario 1 dan skenario 2 pada kelas kronis, *10-fold* skenario 1, skenario 2, skenario 3, skenario 5, skenario 7, skenario 9, dan skenario 10 pada kelas sedang, *10-fold* skenario 6 pada kelas berat.
3. Nilai recall yang diperoleh dari menggunakan metode *Modified K-Nearest Neighbor* dengan pengujian *k-fold*, nilai recall tertinggi ada pada *5-fold* skenario 5.

B. Saran

Pada penelitian ini masih sangat jauh dari kata sempurna, disarankan bagi penelitian selanjutnya agar :

1. Menggunakan *k* dan *k-fold* yang lebih beragam.
2. Mengembangkan dengan menggunakan metode optimasi seperti algoritma genetika.

5. DAFTAR PUSTAKA

Arief. (2011). Pemograman web dinamis menggunakan php dan mysql. Yogyakarta. *Yogyakarta: C.V ANDI OFFSET*.

Arief, S. (2012). *Hepatitis Virus. In: Juffrie, M., et al., ed. Buku Ajar Gastroenterologi-Hepatologi. 3rd ed*. IDAI.

Blachier, M., Leleu, H., Peck-Radosavljevic, M., Valla, D. C., & Roudot-Thoraval, F. (2013). The burden of liver disease in Europe: A review of available epidemiological data. In *Journal of Hepatology*. https://doi.org/10.1016/j.jhep.2012.12.005.

Crivat, J. M. Z. T. B. (2011). Microsoft Data Mining with SQL Server ® ® 2008. In *Molecular ecology resources*.

Elisa, E. (2017). Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti. *Jurnal Online Informatika*. https://doi.org/10.15575/join.v2i1.71.

Fitri, A. D., Eka, R. D., & Wahyu, W. A. (2017). Deteksi Penyakit Kucing dengan Menggunakan Modified K-Nearest Neighbor Teroptimasi (Studi Kasus : Puskeswan Klinik Hewan dan Satwa Sehat Kota Kediri). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, *1*(11), 1295–1301.

Haltuf, M. (2011). *UNIVERSITY OF ECONOMICS IN PRAGUE Faculty of Business Administration*.

Hanifa, T. T., & Al-faraby, S. (2017). Analisis Churn Prediction pada Data Pelanggan PT . Telekomunikasi dengan Logistic Regression dan Underbagging. *Universitas Telkom*.

Imanda, A. C., Hidayat, N., & Furqon, M. T. (2018). Klasifikasi Kelompok Varietas Unggul Padi Menggunakan Modified K- Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*.

Irani, J., Pise, N., & Phatak, M. (2016). Clustering Techniques and the Similarity Measures used in Clustering: A Survey. *International Journal of Computer Applications*. https://doi.org/10.5120/ijca2016907841.

Muhammad, B. L. (2015). Modified Nearest Neighbor Untuk Prediksi Curah Hujan. *Konferensi Nasional Sistem dan Informatika 2015*, *1*, 272–277.

http://ejournal.stikom-bali.ac.id/index.php/knsi/article/view/463/115.

Nasr, M., El-Bahnasy, K., Hamdy, M., & Kamal, S. M. (2018). A novel model based on non invasive methods for prediction of liver fibrosis. *ICENCO 2017 - 13th International Computer Engineering Conference: Boundless Smart Societies*. https://doi.org/10.1109/ICENCO.2017.8289800.

Oktaviana, A. R. (2016). Penerapan data mining klasifikasi pola nasabah menggunakan algoritma c4.5 pada bank bri batang. In *Fik, UDINUS*.

Parvin, H., Alizadeh, H., & Minati, B. (2010). A Modification on K-Nearest Neighbor Classifier. *Global Journal of Computer Science and Technology*, *10*(14), 37–41.

Permatasari, R., Aryati, A., & Arifah, B. (2018). DETEKSI ANTIBODI MULTIPEL HEPATITIS C DALAM DARAH DONOR (Multiple Antibody Detection of Hepatitis C in Donor Blood). *Indonesian Journal of Clinical Pathology and Medical Laboratory*, *21*(3), 261. https://doi.org/10.24293/ijcpml.v21i3.1278.

Prasetyo, E. (2013). Data Mining : Konsep Dan Aplikasi Menggunakan Matlab. In *Journal of Chemical Information and Modeling*. https://doi.org/10.1017/CBO9781107415324.004.

Purbadian. (2016). XAMPP merupakan suatu software yang bersifat open source yang merupakan pengembangan dari LAMP (Linux, Apache, MySQL, PHP dan Perl). *Journal of Chemical Information and Modeling*.

Riadi, M. (2017). *Pengertian, Fungsi, Proses dan Tahapan Data Mining*. Kajianpustaka.Com.

Ridwan, M., Suyono, H., & Sarosa, M. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Jurnal EECCIS*.

Riyanto. (2010). xamp. *(Choliviana, Triyono, & Sukadi, 2012)*.

Supono dan Putratama. (2015). Effendi. *Journal of Chemical Information and Modeling*.