

# ANALISIS METODE *K NEAREST NEIGHBOR* TERHADAP KLASIFIKASI DATA PASIEN PENDERITA GAGAL JANTUNG

Jurusan Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jember

<sup>1</sup>Umi Nur Hidayah, <sup>2</sup>Hardian Octavianto, <sup>3</sup>Lutfi Ali Muharom

<sup>1</sup>uminurhidayah13@gmail.com, <sup>2</sup>hardian@unmuhjember.ac.id, <sup>3</sup>lutfi.muharom@unmuhjember.ac.id

## ABSTRAK

Penelitian ini menjelaskan penggunaan metode *K Nearest Neighbor* terhadap klasifikasi data pasien penderita gagal jantung. Dalam penelitian ini dilakukan beberapa tahapan dan metode. Pada tahapan *Preprocessing* penelitian ini menggunakan metode *MinMax Scaler* untuk normalisasi serta metode *LogX* untuk mengurangi tingkat *skewness* dan *outlier*. Pada tahap skenario uji dalam penelitian ini menggunakan metode *Cross Fold Validation* dengan nilai  $k = 3, 5, 7$  dan  $9$  dan sebagai pembeda dalam penelitian ini dilakukan dua proses yang berbeda yaitu proses implementasi *K-NN* menggunakan Teknik *SMOTE* sebelumnya dan implementasi *K-NN* tanpa Teknik *SMOTE*. Pada pengukuran hasil klasifikasi berfokus pada tingkat akurasi dan presisi. Dalam penelitian ini data yang digunakan berjumlah 299 pasien yang berasal dari *UCI Machine Learning*. Pengukuran jarak pada *K-NN* dalam penelitian ini menggunakan *vector Euclidean Distance* dengan nilai ketetangaan  $3, 5$  dan  $9$ . Dari implementasi yang dilakukan diperoleh hasil yaitu klasifikasi *K Nearest Neighbor* terhadap data pasien gagal jantung tanpa *SMOTE* diperoleh hasil validasi terbaik terdapat pada *Nearest Neighbor*  $k = 9$  dengan rata-rata akurasi sebesar  $71.59\%$  dan akurasi tertinggi yaitu sebesar  $83.33\%$  pada *fold* kedua. Sedangkan pada data uji akurasi tertinggi diperoleh pada *Nearest Neighbor*  $k = 3$  dengan nilai akurasi  $71.66\%$ . Klasifikasi *K Nearest Neighbor* terhadap data pasien gagal jantung dengan menggunakan *SMOTE* diperoleh hasil validasi terbaik terdapat pada *Nearest Neighbor*  $k = 3$  dengan rata-rata akurasi sebesar  $77.88\%$  dengan akurasi tertinggi yaitu sebesar  $80.14\%$  pada *fold* keenam dan delapan. Sedangkan pada data uji akurasi tertinggi diperoleh pada *Nearest Neighbor*  $k = 3$  dengan nilai akurasi  $63.33\%$ .

Keyword : Klasifikasi, *K-NN*, Gagal Jantung, *SMOTE*, *Cross Fold Validation*

## PENDAHULUAN

### 1. Latar Belakang

Gagal jantung kongestif atau *Congestive Heart Failure* (CHF) merupakan penyebab kematian nomor satu di dunia (Maulida, 2018). Jumlah gagal jantung di Amerika Serikat kira-kira  $5.7$  juta orang dewasa dan  $550.000$  kasus baru didiagnosis setiap tahunnya diagnosis (Mozaffarian, et al., 2016).

Seiring perkembangan zaman dalam dunia kesehatan terjadi perkembangan pula. Salah satunya penggunaan *Machine Learning* dan *Data Mining* dalam dunia kesehatan. *Data Mining* merupakan suatu proses penggalian data atau penyaringan data dengan memanfaatkan kumpulan data yang cukup besar melalui serangkaian proses untuk mendapatkan informasi yang berharga dari data tersebut, selain itu *Data Mining* merupakan salah satu teknik yang dapat melakukan klasifikasi dengan teknik pengelompokan data. Dalam dunia kesehatan *Data Mining* dapat dimanfaatkan untuk mendapatkan informasi seperti penentuan kriteria suatu penyakit, misalnya tingkat kebutuhan transfusi darah dari penderita *Thalassaemia*. Kebutuhan waktu transfusi yang berbeda-beda dari

tiap penderita menjadi sebuah masalah dalam mempersiapkan pemberian jumlah obat terafi kelasi besi dan kesiapan pendonor darah (Sulastri & Curroni, 2017).

Dari dataset atau sampel latihan pada penelitian sering muncul permasalahan-permasalahan di antaranya *imbalanced data* atau ketidakseimbangan data. *Imbalanced data* (ketidakseimbangan data) adalah salah satu masalah utama yang muncul dalam deteksi anomali pada dataset yang bersifat real time. Beberapa teknik digunakan untuk mengatasi permasalahan data yang tidak seimbang, salah satunya adalah teknik *Synthetic Minority Oversampling Technique* (*SMOTE*). *Synthetic Minority Oversampling Technique* (*SMOTE*) adalah salah satu metode yang digunakan untuk menangani kasus ketidakseimbangan data pada dataset (Santoso, et al., 2017). Dari rangkaian fakta terhadap metode *K Nearest Neighbor* dan penyakit gagal jantung di atas maka penulis akan melakukan penelitian dengan judul Analisis Metode *K Nearest Neighbor* Terhadap Klasifikasi Data Pasien Penderita Gagal Jantung.

2. Rumusan Masalah

Dari rangkaian pada latar belakang yang telah dipaparkan, penulis melakukan penelitian yang menitik beratkan pada beberapa permasalahan yaitu:

- a. Berapakah nilai akurasi yang diperoleh metode *K Nearest Neighbor* dalam mengklasifikasi data pasien gagal jantung?
- b. Berapakah nilai akurasi yang diperoleh metode *K Nearest Neighbor* dalam mengklasifikasi data pasien gagal jantung setelah melalui proses *SMOTE*?

3. Tujuan

Dari rangkaian titik permasalahan yang dijelaskan pada rumusan masalah berikut tujuan yang ingin diperoleh dalam penelitian ini yaitu:

- a. Mengetahui nilai akurasi yang diperoleh metode *K Nearest Neighbor* dalam mengklasifikasi data pasien gagal jantung.
- b. Mengetahui nilai akurasi yang diperoleh metode *K Nearest Neighbor* dalam mengklasifikasi data pasien gagal jantung melalui proses *SMOTE*.

4. Manfaat

Hasil dalam penelitian ini diharapkan bermanfaat terhadap penulis, pembaca dan masyarakat luas. Berikut manfaat yang diharapkan dalam penelitian ini:

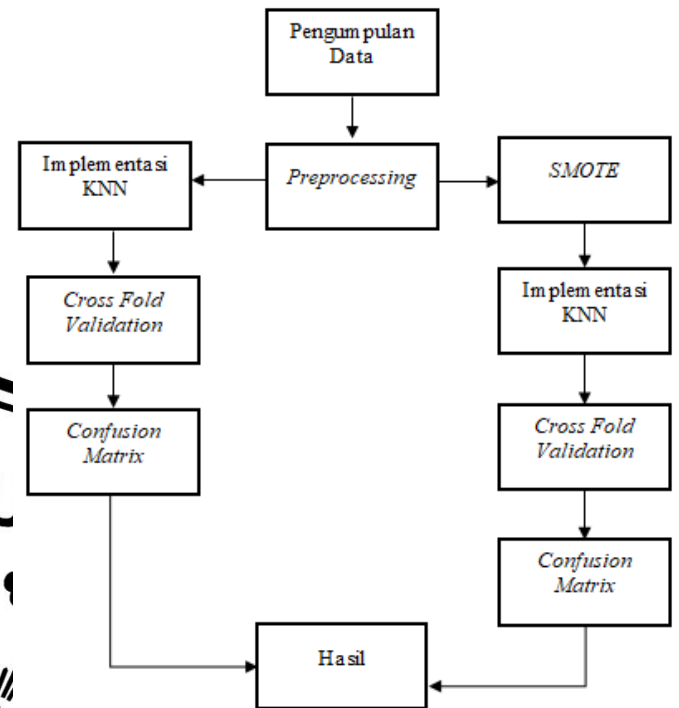
- a. Pengembangan Data Mining dalam dunia kesehatan.
- b. Menjadi bahan pertimbangan kesehatan dalam perawatan pasien gagal jantung.

5. Batasan Masalah

Dalam pokok masalah yang telah dijabarkan pada rumusan permasalahan, penulis membatasi penelitian ini agar tetap dalam topik permasalahan yang telah ditentukan, berikut batasan-batasan masalah yang diterapkan:

- a. Data yang dipakai dalam penelitian ini bersumber dari *UCI Machine Learning* pada laman <https://archive.ics.uci.edu/ml/datasets/Heart+failure+clinical+records#>.
- b. Jumlah data yang digunakan pada penelitian ini sebanyak 299 data pasien.
- c. *Vector* pengukuran jarak yang digunakan dalam metode *K Nearest Neighbor* adalah *Euclidean Distance*.
- d. Skenario uji yang digunakan dalam penelitian ini yaitu *Cross Fold Validation* dengan nilai  $k = 2, 5, 7$  dan  $10$ .
- e. Nilai *K Nearest Neighbor* yang dipakai adalah  $3, 5, 7$  dan  $9$ .

Penelitian terhadap analisis metode *K Nearest Neighbor* terhadap data penderita gagal jantung memiliki beberapa proses di dalamnya. Berikut alur tahapan proses yang akan dilakukan.



Gambar 1 Tahapan Penelitian

1. Pengumpulan Data

Pengumpulan data dimulai dengan mengumpulkan data dari sumber data yaitu *UCI Machine Learning*. Selanjutnya menganalisis kebutuhan pengolahan lanjutan seperti kajian pustaka sebagai dasar dari penelitian ini dan teknik pengolahan.

2. Preprocessing

*Preprocessing* dilakukan guna menghindari jika pada data terdapat *missing value* atau data yang dianggap *noise* oleh metode *K Nearest Neighbor* pada saat implementasi. *Preprocessing* yang dilakukan pada penelitian ini adalah mengurangi tingkat *skewness* dan *outlier* serta normalisasi data menggunakan *MinMax Scaler*. Untuk mengurangi tingkat *skewness* menggunakan persamaan *LogX* sedangkan untuk mengurangi tingkat *outlier* menggunakan persamaan median yaitu nilai-nilai *outlier* akan diganti dengan nilai median pada atribut tertentu.

3. Implementasi KNN tanpa SMOTE

Pada penelitian ini dilakukan dua proses yang berbeda untuk mendapatkan perbandingan hasil yang lebih baik. Tahap ini akan melakukan proses *K Nearest Neighbor* yang sebelumnya tanpa melakukan *SMOTE* pada dataset. Sehingga data setelah melalui proses *Preprocessing* akan dilakukan partisi data dan skenario uji menggunakan metode *Cross Fold Validation*.

METODE PENELITIAN

- Implementasi KNN menggunakan *SMOTE*  
Tahap ini akan menjadi pembandingan terhadap tahap implementasi tanpa *SMOTE*. Pada proses ini data hasil *Preprocessing* akan dilakukan *SMOTE*. Setelah data diproses *SMOTE*, data akan dipartisi dan dibagi sesuai skenario uji. Skenario uji menggunakan metode *Cross Fold Validation*.

## IMPLEMENTASI DAN HASIL

### 1. Pengumpulan Data

Dataset penelitian ini berasal dari Universitas Faisalabad Pakistan yang dihimpun oleh Tanvir Ahmad, Assia Munir, Sajjad Haider Bhatti, Muhammad Aftab, dan Muhammad Ali Raza dan di unggah pada *FigShare* di bawah hak cipta *Attribution 4.0 Internasional*. Dataset ini didonasikan dan dipublikasikan pada *UCI Machine Learning* pada tanggal 5 Februari 2020. Jumlah data pada penelitian ini 299 data dan 13 atribut. Berikut atribut pada dataset penelitian ini.

### 2. Data Partition

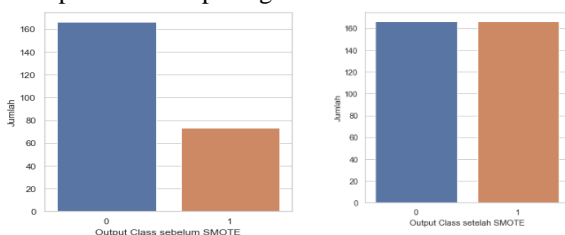
Kondisi data setelah dilakukan normalisasi adalah melakukan partisi atau memecahan data. Data akan dipecah menjadi dua yaitu data *training* atau latihan data dan data testing atau uji. Berikut diagram perbandingan data dari total 299 *Heart Failure Clinical Records Data Set* akan dibagi menjadi dua yaitu, 80% data *training* atau latihan dan 20% data testing atau uji.

### 1. Implementasi KNN

Dalam implementasi *K Nearest Neighbor* pada penelitian ini menggunakan *Euclidean Distance* untuk pengukuran jaraknya dan *Nearest Neighbor*  $k = 3, 5, 7$  dan  $9$  untuk tetangga terdekatnya serta skenario uji *K Fold Cross Validation* dengan nilai  $k$  yaitu  $10$  *fold*. Dalam penelitian ini juga diterapkan *SMOTE* sebagai pembandingan untuk implementasi *K Nearest Neighbor* tanpa menggunakan *SMOTE* terlebih dahulu dan *K Nearest Neighbor* menggunakan *SMOTE*.

$$D_{xy} = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad (\text{Persamaan Euclidean Distance}).$$

Dari implementasi metode *SMOTE* untuk mengatasi *imbalanced* data pada penelitian ini diperoleh hasil pada gambar 4.1 berikut.



Gambar 2 Perbandingan *data training* sebelum dan sesudah *SMOTE*

Gambar 2 dapat dijelaskan bahwa pada gambar 4.28 (kiri) merupakan *data training* awal dengan perbandingan *output class* yaitu *ouput 0* sebanyak 166 dan *output 1* sebanyak 73 sedangkan pada gambar 4.28 (kanan) menunjukkan *data training* setelah proses *SMOTE* dengan perbandingan *output class* yaitu *output 0* sebanyak 166 dan *output 1* sebanyak 166, hal ini menandakan kedua *output* berimbang jumlahnya.

### 3. K fold Cros Validation

Dalam implementasi *K Nearest Neighbor* pada penelitian ini menggunakan *Euclidean Distance* untuk pengukuran jaraknya dan *Nearest Neighbor*  $k = 3, 5, 7$  dan  $9$  untuk tetangga terdekatnya serta skenario uji *K Fold Cross Validation* dengan nilai  $k$  yaitu  $10$  *fold*. Dalam penelitian ini juga diterapkan *SMOTE* sebagai pembandingan untuk implementasi *K Nearest Neighbor* tanpa menggunakan *SMOTE* terlebih dahulu dan *K Nearest Neighbor* menggunakan *SMOTE*, berikut adalah hasil perbandingan akan di sajikan dalam bentuk tabel.

Fold	Tanpa SMOTE				Menggunakan SMOTE			
	K3	K5	K7	K9	K3	K5	K7	K9
akurasi (%)								
1	75	70.83	70.83	70.83	73.53	64.71	58.52	67.65
2	66.67	62.5	70.83	70.83	79.41	76.47	73.53	64.71
3	70.83	70.83	79.17	79.17	75.76	69.7	60.61	54.55
4	75	70.83	75	83.33	81.82	75.76	66.67	69.7
5	79.17	75	70.83	75	72.73	78.79	72.73	69.7
6	70.83	70.83	75	70.83	87.88	84.85	66.67	60.61
7	62.5	62.5	66.67	66.67	78.79	72.73	63.64	63.64
8	75	66.67	62.5	62.5	87.88	84.85	75.76	78.79
9	62.5	62.5	70.83	70.83	81.82	72.73	66.67	69.7
10	73.91	73.91	69.57	69.57	81.82	84.85	78.79	78.79
rata-rata akurasi (%)	71.141	68.64	71.123	71.956	80.144	76.544	68.389	67.784

Tabel 1 Perbandingan *K fold Cross Validation*

Dapat disimpulkan bahwa dari tabel perbandingan diatas pengukuran nilai akurasi pada implementasi *K Nearest Neighbor* terbaik diperoleh pada *Nearest Neighbor* tanpa *Smote*  $K = 9$  dengan nilai rata-rata akurasi sebesar 71.956%. dan nilai rata-rata akurasi yang menggunakan *smote* diperoleh *Nearest Neighbor*  $K=3$  yaitu 80.144%. sedangkan nilai tertinggi dari rata-rata akurasi tertinggi terdapat pada yang menggunakan *Smote* sebesar 80.144% yang ada pada  $K=3$ .



#### 4. Confusion matrix

Berikut akan disajikan hasil pengujian dari proses klasifikasi tanpa menggunakan *SMOTE* dan proses klasifikasi menggunakan *SMOTE* dengan 20% data di uji dari 80% data latih menggunakan *confusion matrix*.

Pengukuran	Tanpa Smote				Menggunakan Smote			
	K3	K5	K7	K9	K3	K5	K7	K9
TP	8	6	4	5	8	6	4	5
TN	35	35	35	34	35	35	35	34
FP	2	2	2	3	2	2	2	3
FN	15	17	19	18	15	17	19	18
Akurasi	71.67	68.33	65	65	71.67	68.33	65	65
Presisi	80	75	66.67	62.5	80	75	66.67	62.5

Tabel 2. Perbandingan *Confusion matrix*

Dari tabel 2 dapat dijelaskan bahwa dari proses pengukuran pada implementasi *K Nearest Neighbor* menggunakan *SMOTE* maupun tanpa proses *SMOTE* sebelumnya menghasilkan nilai akurasi tertinggi sebesar 71.67% yang diperoleh pada *Nearest Neighbor*  $k = 3$  dan nilai presisi tertinggi sebesar 80% yang diperoleh pada *Nearest Neighbor*  $k = 3$ . Sehingga dapat disimpulkan bahwa hasil pengujian dari *confusion matrix* tanpa *SMOTE* maupun menggunakan *SMOTE* tidak ada perbedaan.

#### PENUTUP

##### 1. Kesimpulan

Dari serangkaian proses yang dilalui dalam penelitian ini yaitu Analisis Metode *K Nearest Neighbor* Terhadap Klasifikasi Data Pasien Penderita Gagal Jantung dengan data 209 dengan nilai *Nearest Neighbor*  $k = 3, 5, 7$  dan 9 serta validasi data menggunakan *Cross Fold Validation* dengan nilai  $k$  yaitu 10 *fold* memperoleh hasil yaitu:

- Dari proses pertama yaitu klasifikasi *K Nearest Neighbor* terhadap data pasien gagal jantung tanpa *SMOTE* diperoleh hasil validasi terbaik terdapat pada *Nearest Neighbor*  $k = 9$  dengan rata-rata akurasi sebesar 71.59% dan akurasi tertinggi yaitu sebesar 83.33% pada *fold* kedua. Sedangkan pada data uji akurasi tertinggi diperoleh pada *Nearest Neighbor*  $k = 3$  dengan nilai akurasi 71.66%.
- Dari proses kedua yaitu klasifikasi *K Nearest Neighbor* terhadap data pasien gagal jantung dengan menggunakan *SMOTE* diperoleh hasil validasi terbaik terdapat pada *Nearest Neighbor*  $k = 3$  dengan rata-rata akurasi sebesar 87.88% dengan akurasi tertinggi yaitu sebesar 80.14%

pada *fold* keenam dan delapan. Sedangkan pada data uji akurasi tertinggi diperoleh pada *Nearest Neighbor*  $k = 3$  dengan nilai akurasi 63.33%.

##### 2. Saran

Penulis selaku peneliti menyadari bahwa penelitian ini jauh dari kesempurnaan. Oleh karena itu penulis berkeinginan jika pembaca ingin melakukan pengembangan pada penelitian ini. beberapa aspek yang dapat dikembangkan dalam penelitian ini untuk memperoleh hasil yang lebih baik antara lain pengembang dapat membandingkan dengan metode klasifikasi lain, menambahkan teknik *preprocessing* yang lain, atau menggunakan *metric distance* yang lain dalam klasifikasi *K Nearest Neighbor* ini.

#### Daftar Pustaka

A. Somasundaram and U. S. Reddy. (2016). "Data Imbalance: Effects and Solutions for Classification of Large and Highly Imbalanced data," Proc. 1st Int. Conf. Res. Eng. Comput. Technol. (ICRECT 2016), no. November, pp. 28-34.

Alfisahrin, N. N., & Mantoro, T. (2014). Datamining techniques for optimization of liver disease classification. pp. 379-384). Presented at the Proceedings - 2013 International Conference on Advanced Computer Science Applications and Technology (ACSAAT 2013).

Agung, A. (2020). Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes. *Indonesian Journal of Data and Science*, 1(2), 29-33.

Alfaridhi, Purnawansyah P., Fattah, F., & Putri, I. P. (2020). Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung. *ILKOM Jurnal Ilmiah*, 12(2), 81-86.

B. Santoso, H. Wijayanto, K. A. Notodiputro, and B. Sartono. (2017). "Synthetic oversampling Methods for Handling Class Imbalanced Problems : A Review," in IOP Conference Series: Earth and Environmental Science, vol. 58, no. 1.

Black, J.M., & Hawks, J.K. (2005). *Medical Surgical Nursing : Clinical Management for Positive Outcomes, Volume II, 7th Edition*, Elsevier's Health Sciences Right Departement : Philadelphia.

Bustami, B. (2013). Penerapan Algoritma Naïve Bayes Untuk Mengklasifikasi Data Nasabah Asuransi. *TECHSI-Jurnal Teknik Informatika*, 5(2).

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). *SMOTE*: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357.

Chicco, D., & Jurman, G. (2020). *Machine Learning* can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC medical informatics and decision making*, 20(1), 16.

Emory Health Care. (2020). Heart Failure Statistics. [www.emoryhealthcare.org](http://www.emoryhealthcare.org). (Diakses pada tanggal 15 September 2020).

Fatoni, C. S., & Noviandha, F. D. (2018). Case Based Reasoning Diagnosis Penyakit Difteri dengan Algoritma *K-Nearest Neighbor*. *Creative Information Technology Journal*, 4(3), 220-232.

Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996). Knowledge Discovery and Data Mining: Towards a Unifying Framework. In *KDD* (Vol. 96, pp. 82-88).

Firdaus, D. (2017). Penggunaan Data Mining dalam Kegiatan Sistem Pembelajaran Berbantuan Komputer. *Format*, 6(2), 91-97.

Gustiani. (2012). "Algoritma *Nearest Neighbor* Untuk Menentukan Area Pemasaran Produk Batik Di Kota Pekalongan," *Jurnal Ilmiah UTeach Vol. 8*. STMIK Widya Pratama.

Han, J., & Kamber, M. (2006). Classification and prediction. *Data mining: Concepts and techniques*, 347-350.

Han, J., Pei, J., & Kamber, M. (2011). *Data mining concepts and techniques*. Elsevier.

Hidayah, Nikmatul. (2014). "Klasifikasi Perjurusan Sekolah Menengah Atas dengan Algoritma Naive Bayes Classifier pada SMAN 1 Sulah," *Universitas San Nuswantoro : Semarang*.

Ihdaniyati, A. I., & Arifah, S. (2009). Hubungan tingkat kecemasan dengan mekanisme coping pada pasien gagal jantung kongestif di RSUD Pandan Arang Boyolali.

Indrayanti, I., Sugianti, D., & Al Karomi, A. (2017). Optimasi Parameter K Pada Algoritma *K-Nearest Neighbour* Untuk Klasifikasi Penyakit Diabetes Mellitus. *Prosiding SNATIF*, 823-829.

Jananto, Arief. (2010). Perbandingan Performansi Algoritma *Nearest Neighbor* dan SLIQ untuk Prediksi Kinerja Akademik Mahasiswa Baru ( Studi Kasis Data Akademik Mahasiswa Fakultas Teknologi Informasi UNISBANK). *Jurnal Teknologi Informasi DINAMIK*, Vol : XV.

Karegowda, A. G., Jayaram, M. A., & Manjunath, A. S. (2012). Cascading k-means clustering and *k-Nearest Neighbor classifier* for categorization of diabetic patients. *International Journal of Engineering and Advanced Technology*, 1(3), 147-151.

Kasanah, A. N., Muladi, M., & Pujiyanto, U. (2019). Penerapan Teknik *SMOTE* untuk Mengatasi Imbalance *Class* dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 3(2), 196-201.

Leidiana, H. (2013). Penerapan algoritma *k-Nearest Neighbor* untuk penentuan resiko kredit kepemilikan kendaraan bermotor. *PIKSEL: Penelitian Ilmu Komputer Sistem Embedded and Logic*, 1(1), 65-76.

Lestari, M. (2015). Penerapan algoritma Klasifikasi *Nearest Neighbor* (K-NN) untuk mendeteksi penyakit jantung. *Faktor Exacta*, 7(4), 366-371.

Maulida, P. C., & Jadmiko, A. W. (2018). Gambaran Tipe Kepribadian pada Pasien Gagal Jantung Kongestif di RSUD Dr. Moewardi Surakarta (Doctoral dissertation, Universitas Muhammadiyah Surakarta).

Mozaffarian, D., Benjamin, E. J., Go, A. S., Arnett, D. K., Blaha, M. J., Cushman, M., ... & Howard, V. J. (2017). Executive summary: heart disease and stroke statistics—2016 update: a report from the American Heart Association. *Circulation*, 133(4), 447-454.

Mustafa, (2015). Perancangan Aplikasi Prediksi Kelulusan Tepat Waktu Bagi Mahasiswa Baru Dengan Teknik Data Mining (Studi Kasus : Data Akademik Mahasiswa STMIK Diponegara Makassar). *Citec Jurnal Ilmiah*, 1. STMIK Diponegara.

Workalemichael, T. G. (2007). Management of chronic heart failure: national clinical guideline. Edinburgh: Scottish Intercollegiate Guidelines Network.

POKATIN. (2013). *Situasi Kesehatan Jantung*. Jakarta: kementkes.

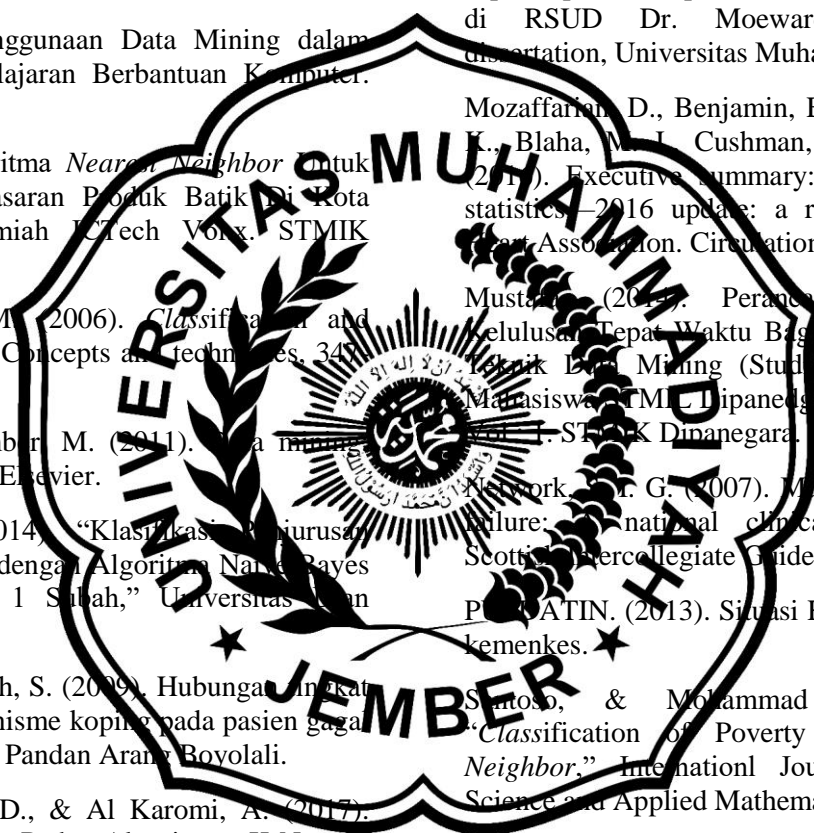
Santoso, & Mohammad Isa Irawan. (2012). "Classification of Poverty Levels using *KNearest Neighbor*." *International Journal of Computing and Science and Applied Mathematics*, vol. 2, No 1 March

Sari, P. P. M. K., Ernawati, E., & Pranowo, P. (2015). Kombinasi Metode *K-Nearest Neighbor* dan Naive Bayes Untuk Klasifikasi Data. *J. SEMNASTEKNIMEDIA*, 3, 37-41.

Sudoyo, Aru. W,dkk. (2009). *Buku Ajar Penyakit Dalam Jilid 2 Edisi 5*. Jakarta : Internal Publishing. Halaman 59.

Sulastri, H., & Gufroni, A. I. (2017). Penerapan data mining dalam pengelompokan penderita thalassaemia. *Jurnal Nasional Teknologi dan Sistem Informasi*, 3(2), 299-305.

Turban, E., Rainer, R. K., & Potter, R. E. (2005). *Introduction to information technology* (Vol. 2, pp. 51-62). John Wiley & Sons.



UC Irvine *Machine Learning* Repository. (2020). <https://archive.ics.uci.edu/ml/datasets/Heart+failure+clinical+records#>. (Diakses pada 15 September 2020).

Xhemali, D., J HINDE, C., & G STONE, R. (2009). Naïve bayes vs. decision trees vs. neural networks in the classification of *training* web pages. D. XHEMALI, CJ HINDE and Roger G. STONE," Naive Bayes vs. Decision Trees vs. Neural Networks in the *Classification of Training Web Pages*", *International Journal of Computer Science Issues*, IJCSI, Volume 4, Issue 1, pp16-23, September 2009, 4(1).

