

PERBANDINGAN ALGORITMA *FUZZY C-MEANS* DAN ALGORITMA *NAÏVE BAYES* DALAM MENENTUKAN KELUARGA PENERIMA MANFAAT (KPM) BERDASARKAN STATUS SOSIAL EKONOMI (SSE) TERENDAH

Putu Satya Saputra, Gede Rasben Dantes, I Gede Aris Gunadi

Program Studi Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha
Singaraja, Bali, Indonesia

E-mail: satput77@gmail.com, rasben.dantes@undiksha.ac.id,
igedearisgunadi@undiksha.ac.id

Abstrak

Penelitian ini bertujuan untuk mengetahui : hasil *clustering* algoritma *Fuzzy C-means*, hasil klasifikasi *Naïve Bayes*, dan tingkat akurasi penerapan *Fuzzy C-means* dan *Naïve Bayes* dalam penentuan penerima bantuan sosial di Kabupaten Buleleng berdasarkan Basis Data Terpadu (BDT). Data yang digunakan yaitu Basis Data Terpadu sesuai Kepmensos No. 71/huk/2018 yang didapat di Bidang Perlindungan dan Jaminan Sosial Dinas Sosial Kabupaten Buleleng. Data tersebut dikelompokkan menjadi 3 kelompok yaitu penerima bantuan Program Keluarga Harapan (PKH), Bantuan Sosial Pangan (BSP) dan Penerima Bantuan Iuran (PBI) Jaminan Kesehatan. Metode perhitungan akurasi data mining menggunakan *confusion matrix*. Hasil implementasi algoritma dengan 1350 data keluarga menunjukkan tingkat akurasi yang diperoleh algoritma *Naïve Bayes* lebih baik dari pada *Fuzzy C-means*. Nilai akurasi *Naïve bayes* sebesar 74% dan akurasi *Fuzzy C-means* sebesar 67%. Dari pengujian yang telah dilakukan dengan menggunakan perhitungan *confusion matrix* didapatkan hasil algoritma yang efektif digunakan dalam menentukan keluarga penerima manfaat tersebut adalah algoritma *Naïve Bayes*.

Kata kunci: data mining, naïve bayes, fuzzy c-means, confusion matrix, BDT

Abstract

This study aims to determine: the results of the Fuzzy C-means clustering algorithm, the results of the Naïve Bayes classification, and the accuracy of the application of Fuzzy C-means and Naïve Bayes in the determination of recipients of social assistance in Buleleng Regency based on the Integrated Database (BDT). The data used is Integrated Database according to Kepmensos No. 71 / huk / 2018 obtained in the Social Protection and Social Security Sector of the Regency of Buleleng. The data is grouped into 3 groups: recipients of the Family Hope Program (PKH), Social Food Assistance (BSP) and Health Insurance Benefit Recipients (PBI). The method of calculating the accuracy of data mining uses confusion matrix. The results of the implementation of the algorithm with 1350 family data indicate the level of accuracy obtained by the Naïve Bayes algorithm is better than Fuzzy C-means. Naïve Bayes accuracy value is 74% and Fuzzy C-means accuracy is 67%. From the tests that have been carried out using the confusion matrix calculation, the results of the effective algorithm used in determining the family of beneficiaries are the Naïve Bayes algorithm.

Keywords: data mining, naïve bayes, fuzzy c-means, confusion matrix, BDT

PENDAHULUAN

Indonesia merupakan salah satu Negara berkembang di Asia khususnya Asia Tenggara. Salah satu masalah yang sering dihadapi oleh Negara berkembang adalah kemiskinan. Badan Pusat Statistik (BPS) mencatat angka kemiskinan per September 2018 mencapai 9,66% atau jika dilihat angkanya maka jumlah rumah tangga miskin di Indonesia menjadi 25,67 juta orang (BPS, 2019). Basis Data Terpadu (BDT) digunakan sebagai sumber data untuk program pengentasan kemiskinan. Sehingga program apapun dalam rangka penanganan fakir miskin harus berpedoman kepada BDT. Basis Data Terpadu adalah sistem data elektronik berisi data nama dan alamat yang memuat informasi sosial, ekonomi, dan demografi dari individu dengan status kesejahteraan terendah di Indonesia (Permensos 10 Tahun 2016 Pasal 1).

Sumber utama BDT adalah hasil kegiatan Pendataan Program Perlindungan Sosial yang dilaksanakan oleh BPS pada bulan Juli sampai dengan Desember 2011 (Widianto, 2013) dan diolah lebih lanjut oleh Tim Nasional Percepatan Penanggulangan Kemiskinan (TNP2K, 2015). Dalam Survey Sosial Ekonomi Nasional (Susenas), pengelompokan tingkat kesejahteraan masyarakat terbagi menjadi Sangat Miskin, Miskin dan Hampir Miskin. Basis Data Terpadu digunakan untuk memperbaiki kualitas penetapan sasaran program - program perlindungan sosial. BDT membantu perencanaan program, memperbaiki penggunaan anggaran dan sumber daya program perlindungan sosial. Dengan menggunakan data dari BDT, jumlah dan sasaran penerima manfaat program dapat dianalisis sejak awal perencanaan program. Hal ini akan membantu mengurangi kesalahan dalam penetapan sasaran program perlindungan sosial. Data rumah tangga dalam BDT diurutkan menurut peringkat kesejahteraannya dengan metode *Proxy-Means Testing* (PMT). PMT digunakan untuk memperkirakan kondisi sosial-ekonomi setiap rumah tangga dengan menggunakan data karakteristik rumah tangga seperti jumlah anggota keluarga,

status pendidikan, kondisi rumah, kepemilikan aset dan lain-lain (TNP2K, 2015).

Permensos No. 28 Tahun 2017 menyebutkan modul pada aplikasi SIKS telah digunakan oleh lebih dari 256 (dua ratus lima puluh enam) daerah Kabupaten/Kota yang secara aktif melaksanakan verifikasi dan validasi (verivali) Data Terpadu Penanganan Fakir Miskin dan Orang Tidak Mampu. Melalui SIKS-NG, kesalahan karena *inclusion error* (menetapkan orang yang tidak memenuhi syarat sebagai penerima manfaat) dan *exclusion error* (tidak menetapkan orang yang memenuhi syarat sebagai penerima manfaat) dapat diminimalisasi (Permensos No. 28, 2017).

Data BDT pada aplikasi SIKS di Dinas Sosial Kabupaten Buleleng berjumlah 243.210 jiwa (40% keluarga dengan SSE terendah) yang terdiri dari 3 kelompok data KPM yaitu 256.448 jiwa penerima PIS, 37.215 keluarga penerima BSP, 61.005 jiwa penerima PKH (Widianto, 2013). Menurut salah satu Staff Pengolahan Data dan Informasi Kesejahteraan Sosial yang bertugas sebagai supervisor SIKS Dinas Sosial Kabupaten Buleleng, masih banyak rumah tangga miskin yang komplain mengenai bantuan sosial yang diterima. Masih banyak pula yang mengeluh membandingkan bantuan sosial tersebut malah diperoleh rumah tangga miskin yang lebih mampu dari mereka. Misalnya keluarga penerima BSP yang menerima bantuan subsidi beras (Bantuan Sosial Pangan) mengeluh dan merasa layak mendapat bantuan PKH yang menerima bantuan tunai melalui rekening empat kali dalam setahun. Perlu dilakukan evaluasi terhadap tingkat akurasi data yang ada di aplikasi SIKS dan adanya suatu sistem pembandingan dengan akurasi data yang lebih baik. Untuk mempertahankan kualitas data penerima bantuan sosial, Pemerintah Kabupaten Buleleng melalui Dinas Sosial telah melakukan upaya dengan melakukan validasi ulang kondisi sosial ekonomi masyarakat setiap setahun sekali. Namun upaya yang telah dilakukan tersebut masih memiliki beberapa kekurangan antara lain membutuhkan

waktu yang cukup lama dalam melakukan validasi dari rumah ke rumah, melibatkan berbagai pihak sebagai validator, serta biaya yang tinggi.

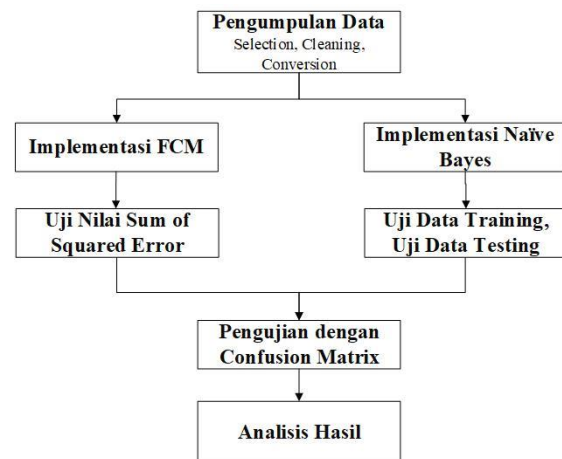
Terdapat beberapa metode pengelompokan data seperti *Analytic Hierarchy Process* (AHP), *Simple Additive Weighting* (SAW), *Fuzzy C-Means* (FCM), *Naïve Bayes* dan lain sebagainya. Adapun penelitian yang pernah dilakukan terkait pengelompokan data kemiskinan adalah sebagai berikut. 1) Penelitian dengan judul *Poverty Level Grouping Using SAW Method* (Rini, 2015) melakukan pemeringkatan tingkat kemiskinan 9 Kecamatan di Kabupaten Pringsewu dengan metode *Simple Additive Weighting* (SAW). Parameter yang digunakan yaitu fasilitas jalan, sumber penerangan, kriteria air bersih dan bahan bakar untuk memasak. 2) Penelitian dengan judul *Poverty Classification Using Analytic Hierarchy Process and K-means Clustering* (Diarsa, 2017) membagi data kemiskinan dalam beberapa kategori. Pada penelitian ini menyajikan kolaborasi metode untuk memproses data survei kemiskinan ke dalam kategori menggunakan *Analytic Hierarchy Process* (AHP) dan metode *k-means clustering* digunakan untuk mencari nilai batas untuk setiap kategori. Namun penelitian ini masih berupa *prototype* dan belum ada evaluasi keakuratan metode klasifikasi kemiskinan tersebut.

Fuzzy C-Mean (FCM) adalah suatu teknik pengelompokan data yang mana keberadaan tiap-tiap data dalam suatu *cluster* ditentukan oleh nilai keanggotaan (Pholsena and Pan, 2018). *Naive Bayes* adalah sebuah teknik klasifikasi probabilistik yang berdasarkan teorema bayes yang menggunakan asumsi keindependenan atribut (tidak ada kaitan antar atribut) dalam proses pengklasifikasiannya (Annur, 2018). Dengan memper-timbangkan hal tersebut pada penelitian ini akan dilakukan perbandingan akurasi Fuzzy C-Means dan Naïve Bayes dalam menentukan Keluarga Penerima Manfaat (KPM). Metode tersebut dibandingkan berdasarkan nilai akurasi menggunakan perhitungan *Confusion Matrix*. Penelitian ini diharapkan

dapat membantu pemerintah setempat agar bantuan sosial yang diberikan tepat sasaran.

METODE

Penelitian ini dilakukan mulai bulan Januari 2019 sampai dengan bulan Januari 2020. Pada penelitian ini, yang menjadi objek penelitian adalah penentuan penerima bantuan sosial berdasarkan Basis Data Terpadu. Tahapan penelitian yang dilakukan adalah sebagai berikut.



Gambar 1. Tahapan Penelitian

Data sosial ekonomi yang digunakan adalah data sosial ekonomi terendah masyarakat Buleleng tahun 2018 sesuai dengan Kepmensos No. 71/huk/2018. Data tersebut diperoleh dari database Sistem Informasi Kesejahteraan Sosial (SIKS) yang terdapat di Dinas Sosial Kabupaten Buleleng.

a. Pemilihan Data

Data sosial ekonomi terendah yang digunakan dalam penelitian ini meliputi data Basis Data Terpadu (BDT), Penerima Bantuan Iuran (PBI), Bantuan Sosial Pangan (BSP) dan Program Keluarga Harapan (PKH) Se-Kabupaten Buleleng.

b. Cleaning Data

Proses *cleaning data* merupakan proses analisa kualitas dari data BDT dengan cara menghapus data yang tidak lengkap atau memiliki format yang salah dalam basis data guna menghasilkan data berkualitas tinggi.

Tabel 1. Proses *Cleaning Data*

ID_BDT	Nama	Jumlah ART	Status Bangunan	Lantai	Peserta
5108020011000269	Ni Made Rusmini	3	Milik Sendiri	Bata Merah	PBI
5108020011000274	Ida Bagus Ngurah	4	Milik Sendiri	Keramik	PKH
5108020011000114	Made Yasa	2	Milik Sendiri	Keramik	
5108020011000121	I Made Parwata			Keramik	

Tabel 1. merupakan contoh data BDT yang tidak lengkap sehingga proses *cleaning data* dilakukan dengan cara

menghapus beberapa data yang tidak lengkap.

Tabel 2. Proses *Cleaning Data* (Lanjutan)

ID_BDT	Nama	Jumlah ART	Status Bangunan	Lantai	Peserta
5108020011000269	Ni Made Rusmini	3	Milik Sendiri	Bata Merah	PBI
5108020011000274	Ida Bagus Ngurah	4	Milik Sendiri	Keramik	PKH

Tabel 2. merupakan hasil *cleaning data* dari tabel 1. Terdapat 4 data keluarga yang dihapus karena data kepesertaan kosong atau tidak lengkap.

dalam bentuk yang lebih mudah diolah dengan alat bantu data mining yang ada. Data pada setiap indikator dikonversi ke dalam bentuk skor sesuai dengan ketentuan Pusat Data dan Informasi Kementerian Sosial melalui aplikasi Sistem Informasi Kesejahteraan Sosial (SIKS). Ketentuan skor dapat dilihat pada Tabel 3. berikut.

c. Konversi Data

Untuk mempermudah dalam proses mining selanjutnya maka data yang ada akan dilakukan proses konversi data ke

Tabel 3. Ketentuan Skor Tiap Indikator

No.	Indikator	Kriteria	Skor
1.	Jumlah Anggota Rumah Tangga (ART)	1	6
		2	5
		3	4
		4	3
		5	2
		> 5	1
2.	Status Bangunan Tinggal	1. Milik Sendiri	5
		2. Kontrak/Sewa	4
		3. Bekas Sewa	3
		4. Dinas	2
3.	Jenis Lantai	1. Marmer/Granit	10
		2. Keramik	9
		3. Permadani	8
		4. Ubin	7
		5. Papan	6
		6. Semen	5
		7. Bambu	4
		8. Kayu	3
		9. Tanah	2
		10. Lainnya	1

A. Pengujian Sum of Squared Error
 Pengujian Sum of Squared Error dilakukan pada algoritma Fuzzy C-means untuk menguji data training pada algoritma FCM. Contoh perhitungan Sum of Squared Error dilakukan pada kasus penentuan bantuan sosial sebagai berikut.

Tabel 4. Perubahan Iterasi Algoritma FCM

Iterasi	PBI	BSP	PKH
1	93,0	93,9	82,9
2	91,7	84,5	80,2
3	85,3	88,0	95,7

Tabel 4. merupakan hasil rata – rata setiap iterasi pada algoritma Fuzzy C-means. Data tersebut digunakan untuk menguji kinerja algoritma Fuzzy C-means dengan Sum of Squared Error. Selanjutnya menentukan nilai deviasi untuk setiap iterasi yang ditampilkan pada tabel 5.

Tabel 5. merupakan proses perhitungan untuk menentukan nilai deviasi setiap iterasi. Setelah mendapatkan nilai deviasi selanjutnya menjumlahkan nilai deviasi tersebut untuk mendapatkan nilai SSE pada masing – masing iterasi. Perubahan nilai SSE algoritma FCM ditampilkan pada tabel 6.

Tabel 5. Nilai Deviasi

Iterasi 1	Deviation	Deviation ²
93,051	3,085	9,523
93,906	3,941	15,531
82,938	-7,027	49,379
Rata - rata = 89,965		Jumlah = 74,434
Iterasi 2	Deviation	Deviation ²
91,785	6,250	39,068
84,576	-0,957	0,917
80,241	-5,292	28,013
Rata - rata = 85,534		Jumlah = 67,999
Iterasi 3	Deviation	Deviation ²
85,385	-4,321	18,677
88,020	-1,686	2,844
95,715	6,008	36,098
Rata - rata = 89,707		Jumlah = 57,619

Tabel 6. Perubahan Nilai Sum of Squared Error Algoritma FCM

Iterasi	PBI	BSP	PKH	SSE
1	93,05	93,90	82,93	74,43
2	91,78	84,57	80,24	67,99
3	85,38	88,02	95,71	57,61

Tabel 6. menunjukkan perubahan nilai Sum of Squared Error. Dari iterasi 1 sampai iterasi 3 terlihat semakin besar iterasi semakin kecil nilai SSE sehingga

kinerja algoritma FCM yang digunakan dapat dikatakan baik.

B. Pengujian dengan Confusion Matrix
 Penerapan *confusion matrix* pada kasus penerima bantuan sosial keluarga yang menerima Bantuan Sosial Pangan (BSP), Program Keluarga Harapan (PKH) atau Penerima Bantuan Iuran (PBI) dengan algoritma Naïve Bayes. Setelah algoritma tersebut dijalankan, hasil

prediksi dan data kepesertaan awal pada Tabel 7.

Tabel 7. Hasil Prediksi Algoritma Naïve Bayes

IDBDT	Kepesertaan Awal	Hasil Prediksi Naïve Bayes
IBBDT1	PBI	PBI
IBBDT2	PBI	BSP
IBBDT3	PKH	PKH
IBBDT4	BSP	BSP
IBBDT5	BSP	BSP
IBBDT6	PKH	PKH

Tabel 7. menampilkan data kepesertaan awal penerima bantuan sosial dengan menerapkan algoritma Naive Bayes. Selanjutnya ditampilkan dalam bentuk confusion tabel pada Tabel 8. untuk memudahkan mencari nilai akurasi, *error*, presisi dan *recall*.

Tabel 8. Confusion Matrix Naïve Bayes

Actual	Predicted		
	PBI	BSP	PKH
PBI	1	0	0
BSP	1	2	0
PKH	0	0	2

Dari hasil prediksi algoritma Naïve Bayes pada tabel 8. dapat dihitung akurasi, *error*, presisi dan *recall*.

$$\begin{aligned} \text{Akurasi} &= \frac{(TP + TN)}{(TP + TN + FP + FN)} * 100\% \\ &= \frac{5}{6} * 100\% = 83\% \\ \text{Error} &= \frac{(FP + FN)}{(TP + TN + FP + FN)} * 100\% \\ &= \frac{1}{6} * 100\% = 17\% \\ \text{Presisi} &= \frac{(TP)}{(TP + FP)} * 100\% \\ \text{Presisi}_{PBI} &= \frac{1}{2} * 100\% = 50\% \\ \text{Presisi}_{BSP} &= \frac{2}{2} * 100\% = 100\% \\ \text{Presisi}_{PKH} &= \frac{2}{2} * 100\% = 100\% \\ \text{Recall} &= \frac{(TP)}{(TP + FN)} * 100\% \\ \text{Recall}_{PBI} &= \frac{1}{1} * 100\% = 100\% \end{aligned}$$

$$\text{Recall}_{BSP} = \frac{2}{3} * 100\% = 67\%$$

$$\text{Recall}_{PKH} = \frac{2}{2} * 100\% = 100\%$$

HASIL DAN PEMBAHASAN

Penelitian ini menggunakan 2 perhitungan algoritma yang berbeda. Perhitungan yang pertama menggunakan 9 data keluarga dan perhitungan kedua menggunakan 1350 data keluarga yang diimplementasikan menggunakan bahasa pemrograman PHP. Data sosial ekonomi yang digunakan adalah data sosial ekonomi terendah masyarakat Buleleng tahun 2018 sesuai dengan Kepmensos No. 71/huk/2018. Data tersebut diperoleh dari database Sistem Informasi Kesejahteraan Sosial (SIKS) yang terdapat di Dinas Sosial Kabupaten Buleleng. Akurasi algoritma *Fuzzy C-means* maupun algoritma *Naïve Bayes* dihitung menggunakan *Confusion Table*. Hasil perhitungan manual metode Naïve Bayes dengan 9 data training dan 6 data testing menghasilkan kelompok Penerima Bantuan luran (PBI) berjumlah 4 keluarga, kelompok Bantuan Sosial Pangan (BSP) berjumlah 2 keluarga dan kelompok Program Keluarga Harapan (PKH) berjumlah 2 keluarga. Hasil perhitungan manual metode FCM dengan 6 data keluarga yang dipilih secara random menghasilkan 3 kelompok keluarga penerima manfaat. Kelompok Penerima Bantuan luran (PBI) berjumlah 4 keluarga, kelompok Bantuan Sosial Pangan (BSP) berjumlah 2 keluarga dan kelompok Program Keluarga Harapan (PKH) berjumlah 2 keluarga.

Sedangkan perbandingan antara algoritma *Fuzzy C-means* dengan algoritma *Naïve Bayes* dalam menentukan keluarga penerima manfaat setelah diimplementasikan dengan Bahasa Pemrograman PHP ditampilkan pada Tabel 9.

Tabel 9. Hasil Perbandingan Akurasi Algoritma Fuzzy C-means dengan Naïve Bayes

Jumlah Data	Naïve Bayes	Fuzzy C-means
1350	74,8%	67,1%

Perhitungan metode Naïve Bayes dengan 90% data training dari 1350 data keluarga menghasilkan akurasi sebesar 74%. Hasil implementasi metode FCM dengan 1350 data keluarga menggunakan Bahasa pemrograman PHP menghasilkan 3 kelompok keluarga penerima manfaat. Kelompok Penerima Bantuan Iuran (PBI) berjumlah 4 keluarga, kelompok Bantuan Sosial Pangan (BSP) berjumlah 2 keluarga dan kelompok Program Keluarga Harapan (PKH) berjumlah 2 keluarga. Perhitungan manual metode FCM menghasilkan akurasi sebesar 67%.

Kinerja algoritma Fuzzy C-means diuji menggunakan *Sum of Squared Error* (SSE). SSE merupakan salah satu metode pengukuran performansi pada *clustering*. Semakin kecil *Sum of Squared Error* pada *cluster* maka semakin baik *cluster* yang dihasilkan. Pada pengujian ini SSE yang di hitung merupakan SSE terbaik dari hasil uji tiap *cluster*. Berikut adalah SSE dari setiap *cluster* yang diuji ditampilkan pada Tabel 10.

Tabel 10. Nilai Sum of Squared Error Algoritma Fuzzy C-means

Iterasi	PBI	BSP	PKH	SSE
7	0,29	0,39	0,30	0,005
9	0,38	0,31	0,29	0,004
10	0,32	0,34	0,33	0,0001
11	0,33	0,33	0,33	0,000009

Tabel 10. menampilkan hasil Sum of Squared Error dari *clustering* algoritma Fuzzy C-means. SSE yang dihitung

merupakan hasil cluster derajat keanggotaan pada iterasi terakhir. Iterasi yang diuji dengan SSE adalah iterasi 7, iterasi 9, iterasi 10 dan iterasi 11. Hasil perhitungan SSE pada iterasi 7 yaitu 0,0053142886387441, pada iterasi 9 yaitu 0,0047903560973356, iterasi 10 yaitu 0,00015900038163128 dan iterasi 11 yaitu 0,000009007878. Dari hasil perhitungan SSE yang dilakukan pada iterasi yang berbeda terlihat perubahan nilai SSE tidak fluktuatif dan semakin besar iterasi semakin kecil nilai SSE yang dihasilkan.

SIMPULAN

Berdasarkan hasil dari penelitian yang telah dilakukan maka dapat ditarik kesimpulan algoritma Naïve Bayes maupun algoritma Fuzzy C-means dapat membantu pengelompokan data kemiskinan dengan akurasi algoritma Naïve Bayes sebesar 74% sedangkan akurasi algoritma Fuzzy C-means sebesar 67%.

DAFTAR PUSTAKA

- BPS (2019). Persentase Penduduk Miskin Pada September 2018 Sebesar 9.66 Persen. *Retrieved from Badan Pusat Statistik website: <https://www.bps.go.id>*.
- Widiyanto, B. (2013). Unifikasi Sistem Penetapan Sasaran Nasional. *Kementerian Sekretaris Negara, Sekretariat Wakil Presiden, Jakarta*.
- TNP2K (2015). Pengembangan, Pengelolaan, dan Pemanfaatan Basis Data Terpadu (BDT). *Tim Nasional Percepatan Penanggulangan Kemiskinan, Jakarta*.
- Rini, D. C. (2015). Klasifikasi Sinyal EEG Menggunakan metode fuzzy C-means clustering (FCM) dan adaptive neighborhood modified Backpropagation (ANMBP). *Jurnal Matematika" MANTIK", 1(1), 31-36*.
- Diarsa, I W.B. (2017). Pengembangan Sistem Pendukung Keputusan Dengan Kombinasi Metode Naïve Bayes Dan Dempster-Shafer Untuk Memberikan Rekomendasi Objek

Wisata Di Bangli. *Ejournal Universitas Pendidikan Ganesha*.

- Pholsena, K., & Pan, L. (2018, June). Traffic status evaluation based on possibilistic fuzzy C-means clustering algorithm. In *2018 IEEE Third International Conference on Data Science in Cyberspace (DSC)* (pp. 175-180). IEEE.
- Annur, H. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes. *ILKOM Jurnal Ilmiah*, *10*(2), 160-165.