

Penerapan Data Mining pada Penentuan Varian Rasa yang Paling Diminati di Roti Kacang Hj Eliya di Tebing Tinggi Menggunakan Algoritma K-Neighbors

Wirna Rizka Auliani¹, Abdul Halim Hasugian²

Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Jl. Lap. Golf No.12, Kp. Tengah, Kec. Pancur Batu, Kab. Deli Serdang, Sumatera Utara, 20353

e-mail: ¹wirnaaulia1@gmail.com ²abdulhalimhasugian@uinsu.ac.id

Submitted Date: January 23rd, 2024

Revised Date: January 27th, 2024

Reviewed Date: January 26th, 2024

Accepted Date: January 29th, 2024

Abstract

Growth and development in the business world is growing rapidly as it is now requiring entrepreneurs to be able to compete more fiercely for consumers' attention. Entrepreneurs must use various methods to attract consumer interest about a product with various businesses that are starting to emerge. Therefore, every entrepreneur is required to have the ability to do something that is considered better than competitors' businesses in order to face this competition. Umega Nut Bread Hj. Eliya Lubis in promoting her peanut bread products. The competitive environment that exists in the era of globalization will increasingly lead to economic systems and market mechanisms that position marketers to always balance and gain market share (market share). The KNN algorithm has several advantages, namely resilience to training data that has a lot of noise and is effective when the training data is large. Meanwhile, the weakness of KNN is that KNN needs to determine the value of the K parameter (the number of nearest neighbors), training based on distance is not clear about what type of distance should be used and which attributes should be used to get the best results, and the computational cost is quite high because it requires calculations. distance from each query instance to the entire training sample. The result of this research is a classification to determine scholarship recipients using the k-nearest neighbor research methodology to obtain the best accuracy results of 100% from a total of 46 training data, 15 testing data and a K value of 1. Of the 10 flavors of Hj Eliya Tebing Tinggi Peanut Bread, it was found that the best flavor variants based on sales were the flavor variants that had GOOD characteristics, namely the chocolate, chocolate cheese, green bean and black/meirah flavor variants.

Keywords : Data Mining; Variance; Nut Bread; K-Neighbors Neighbour

Abstrak

Pertumbuhan dan perkembangan di dalam dunia usaha bertumbuh dengan pesat seperti sekarang mewajibkan para pengusaha untuk bisa bersaing lebih ketat dalam memperebutkan perhatian konsumen. Pengusaha harus menggunakan berbagai macam cara untuk menarik minat konsumen tentang suatu produk dengan berbagai usaha yang mulai bermunculan. Oleh karena itu, setiap pengusaha diwajibkan untuk memiliki kemampuan terhadap sesuatu yang dianggap lebih baik dari usaha pesaing demi menghadapi persaingan tersebut. Umega Roti Kacang Hj. Eliya Lubis dalam mempromosikan produk roti kacang yang dimiliki. Fenomena persaingan yang ada dalam era globalisasi, akan semakin mengarahkan sistem perekonomian ke mekanisme pasar yang memposisikan pemasar untuk selalu mengembangkan dan merebut market share (pangsa pasar). Algoritma KNN memiliki beberapa kelebihan yaitu ketangguhan terhadap training data yang memiliki banyak noise dan efektif apabila training datanya besar. Sedangkan, kelemahan KNN adalah KNN perlu menentukan nilai dari parameter K (jumlah dari tetangga terdekat), training berdasarkan jarak tidak jelas mengenai jenis jarak apa yang harus digunakan dan atribut mana yang harus digunakan untuk mendapatkan hasil terbaik, dan biaya komputasi cukup tinggi karena diperlukan perhitungan jarak dari tiap query instance pada keseluruhan training sample. Hasil dari penelitian ini adalah klasifikasi untuk menentukan penerima beasiswa menggunakan metode k-nearest neighbor pengujian dengan



diperoleh hasil akurasi terbaik sebesar 100% dari jumlah data training 46, data testing 15 dan nilai K adalah 1. Dari 10 varian rasa Roti Kacang Hj Eliya Tebing Tinggi, diperoleh bahwa varian rasa terbaik berdasarkan penjualan adalah varian rasa yang memiliki keterangan BAIK, yaitu varian rasa coklat, coklat keju, kacang hijau dan kacang hitam/merah.

Kata kunci: Data Mining; Varians; Roti Kacang; K-Nears Neighbour

1 Pendahuluan

Perkembangan dunia usaha yang pesat seperti saat ini mengharuskan para pelaku usaha menghadapi persaingan yang sangat ketat dalam memperebutkan perhatian konsumen. Mereka akan melakukan berbagai macam cara untuk menarik minat konsumen terhadap suatu produk, khususnya mencari cara agar bisa menghadapi persaingan usaha sejenis yang mulai bermunculan.

Oleh karena itu, setiap pelaku usaha diharuskan untuk memiliki kemampuan-kemampuan akan sesuatu yang dianggap lebih baik dari usaha pesaing demi menghadapi persaingan tersebut. Salah satu jenis usaha yang sedang berkembang dengan pesat saat ini adalah bisnis pada bidang makanan. Roti kacang merupakan salah satu oleh-oleh produk makanan khas dari kota Tebing Tinggi. Produk makanan ini mulai muncul sekitar tahun 2005. Secara garis besar, roti kacang khas Tebing Tinggi ini memiliki banyak varian rasa, tidak hanya sebatas pada rasa kacang hijau ataupun kacang hitam (Morgan 2019).

Selain itu, karena pemasarannya yang juga sudah mulai meluas, tidak hanya masyarakat Tebing Tinggi saja yang membeli, tetapi banyak orang-orang dari luar kota Tebing Tinggi yang sekedar singgah ke kota ini tertarik untuk mencicipi bagaimana rasa dari roti kacang tersebut. Hal ini yang membuat beberapa produsen roti kacang di Tebing Tinggi mulai mengembangkan bisnis makanan ini. Kendati pasarnya cukup besar, persaingan bisnis roti kacang ini juga bisa di bilang ketat. Ada 3 merek roti kacang terkenal di kota Tebing Tinggi, antara lain, Roti Kacang Hj. Eliya Lubis, Roti Kacang Cap Rajawali.

UD. Umega Roti Kacang Hj. Eliya Lubis sebagai salah satu produsen roti kacang di Tebing Tinggi mencoba ikut ambil bagian dalam bersaing memperebutkan pangsa pasar roti kacang yang ada di kota Tebing Tinggi. Tak hanya rasa kacang hitam atau kacang hijau, UD. Umega Roti Kacang Hj. Eliya Lubis mulai membuat variasi berbagai

rasa roti kacang agar para konsumen tidak merasa bosan. Ada 5 variasi rasa pada awalnya, yakni kacang hijau (original), kacang hitam, nenas, strawberry, dan cappucino. Setelah beberapa tahun kemudian, barulah usaha ini mulai mengembangkan varian rasa baru yakni, coklat, coklat keju, keju, strawberry coklat dan durian. Hal ini merupakan strategi UD. Umega Roti Kacang Hj. Eliya Lubis dalam memasarkan produk roti kacang mereka mengingat pangsa pasarnya yang cukup luas, memiliki konsumen yang loyal dan adanya asosiasi merek terhadap kualitas yang dimiliki produk ini. Selain itu, tingkat awareness yang tinggi diharapkan akan lebih memudahkan UD. Umega Roti Kacang Hj. Eliya Lubis dalam mempromosikan produk roti kacang yang dimiliki. Fenomena persaingan yang ada dalam era globalisasi, akan semakin mengarahkan sistem perekonomian ke mekanisme pasar yang memposisikan pemasar untuk selalu mengembangkan dan merebut market share (pangsa pasar) (Dewi et al. 2022).

Algoritma KNN memiliki beberapa kelebihan yaitu ketangguhan terhadap training data yang memiliki banyak noise dan efektif apabila training data-nya besar. Sedangkan, kelemahan KNN adalah KNN perlu menentukan nilai dari parameter K (jumlah dari tetangga terdekat), training berdasarkan jarak tidak jelas mengenai jenis jarak apa yang harus digunakan dan atribut mana yang harus digunakan untuk mendapatkan hasil terbaik, dan biaya komputasi cukup tinggi karena diperlukan perhitungan jarak dari tiap query instance pada keseluruhan training sample (Hidayat dan Darussalam 2022).

K-Nearest Neighbour (KNN) adalah suatu metode yang menggunakan algoritma supervised dimana hasil dari query instance yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada KNN (Furqon et al, 2020). Tujuan dari algoritma ini adalah mengklasifikasikan obyek baru berdasarkan atribut dan training

sample. Classifier tidak menggunakan model apapun untuk dicocokkan dan hanya berdasarkan pada memori. Diberikan titik query, akan ditemukan sejumlah K obyek atau (titik training) yang paling dekat dengan titik query (Agustina et al, 2018). Klasifikasi menggunakan voting terbanyak di antara klasifikasi dari K obyek. Algoritma KNN menggunakan klasifikasi ketetangaan sebagai nilai prediksi dari query instance yang baru (Augusto 2019). Sebagai ilustrasi dari penerapan algoritma KNN adalah misalnya terdapat data hasil survei dengan kuesioner, untuk meminta pendapat (Kusrini 2019).

2 Tinjauan Pustaka

2.1 Metode Data Mining

Pada umumnya metode *data mining* dapat dikelompokkan ke dalam dua kategori, yaitu deskriptif dan prediktif. Metode deskriptif untuk mencari pola yang dapat dimengerti oleh manusia yang menjelaskan karakteristik dari data. Metode *clustering* menggunakan ciri-ciri tertentu dari data untuk melakukan pengelompokan data (Kusrini 2019).

2.2 Metode k-Nearest Neighbour

Algoritma *k-Nearest Neighbor* merupakan sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan jumlah k yang telah ditentukan dan mengklasifikasikan ke dalam kelas baru (Furqon et al, 2020). K -NN termasuk *algoritma supervised learning* yang hasil dari query instance yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada k -NN (Pail and Shiu, 2018). Kelas yang paling banyak muncul itu yang akan menjadi kelas hasil klasifikasi. Tujuan dari algoritma ini adalah mengklasifikasikan objek baru berdasarkan atribut dan *training sample* (Lu dan He 2017).

Algoritma *k-Nearest Neighbor* menggunakan klasifikasi ketetangaan (*neighbor*) sebagai nilai prediksi dari *query instance* yang baru. Algoritma ini sederhana, bekerja berdasarkan jarak terpendek dari query instance ke *training sample* untuk menentukan ketetanggaannya (Kusumantara, Alfian, and Yodistina 2019). Berikut beberapa formula perhitungan jarak yang digunakan dalam algoritma k -NN. Teknik pencarian tetangga

terdekat disesuaikan dengan dimensi data, proyeksi, dan kemudahan implementasi oleh pengguna. Teknik pencarian tetangga terdekat pada umumnya dilakukan dengan menggunakan formula jarak Euclidean (Utama 2017).

2.3 Konsep K-Nearest Neighbor

K-Nearest Neighbor (KNN) menjadi salah satu metode dalam top 10 metode data mining yang paling populer (Wu dan Kumar 2009). Metode KNN murni termasuk dalam klasifikasi yang *lazy learner* karena menunda proses pelatihan (atau bahkan tidak melakukan pelatihan sama sekali) sampai ada data uji yang ingin diketahui label kelasnya, maka metode baru akan menjalankan algoritmanya (Wu dan Kumair 2019). Algoritma KNN melakukan klasifikasi berdasarkan kemiripan suatu data dengan data yang lain (Wiley et al. 2019)

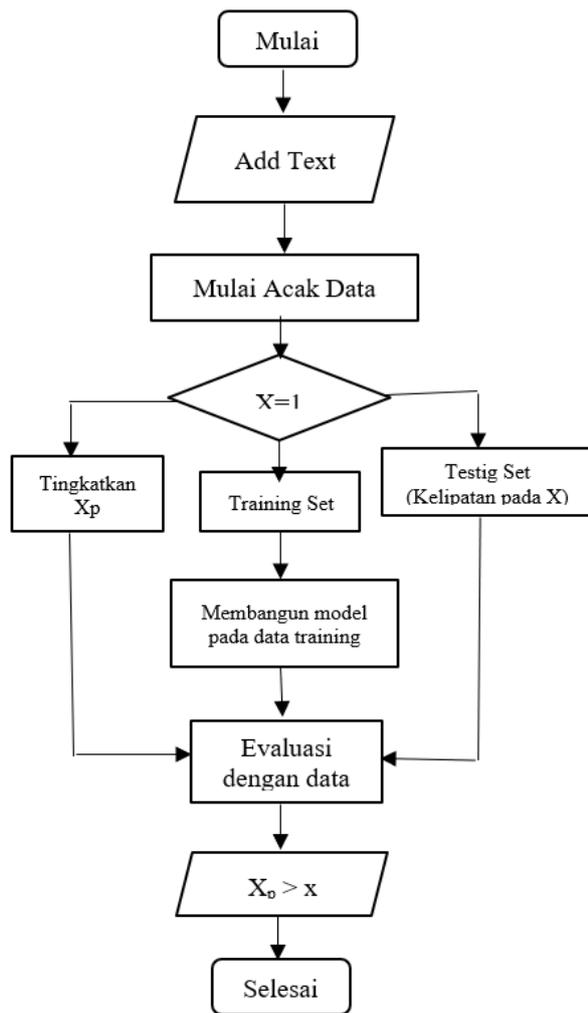
2.4 Rapid Miner

Rapid Miner merupakan perangkat lunak yang bersifat terbuka (*Open Source*), sebuah salah satu solusi untuk melakukan analisis terhadap Data Mining. *Rapid Miner* adalah *platform* perangkat lunak ilmu data yang dikembangkan oleh perusahaan bernama sama dengan yang menyediakan lingkungan terintegrasi untuk persiapan data, pembelajaran mesin, pembelajaran dalam, penambangan teks, dan analisis prediktif. Hal ini digunakan untuk bisnis dan komersial, juga untuk penelitian, pendidikan, pelatihan, *rapid prototyping*, dan pengembangan aplikasi serta mendukung semua langkah dalam proses pembelajaran mesin termasuk persiapan data, hasil visualisasi, validasi model dan optimasi (Aisyah and Putra 2022).

3 Metodologi Penelitian

Metode kuantitatif berbasis adalah metode penelitian yang didasarkan pada filosofi positivisme, digunakan dalam populasi atau subsampel. Pengumpulan data dengan instrumen penelitian kuantitatif/statistik Metode kuantitatif harus menguji hipotesis yang telah ditetapkan (Sugiyono 2016).

Untuk mendukung penelitian ini, sumber data yang digunakan yaitu data primer. Data primer diperoleh dari hasil wawancara dan dokumentasi dengan Roti Kacang Hj. Eliya.



Gambar 1. Flowchart Perhitungan Algoritma K-Nearest Neighbor

Prosedur dalam penelitian ini adalah sebagai berikut:

- Mengumpulkan data pustaka
Langkah-langkahnya adalah Mengumpulkan data penjualan Roti kacang Hj.Eliya
- Melaksanakan observasi di Roti kacang Hj.Eliya

Langkah-langkahnya adalah dengan mewawancarai KTU tentang keadaan penjualan dan produksi Roti kacang Hj.Eliya:

- Menganalisis deskriptif data
- Perhitungan varians rasa paling diminati dengan Algoritma KNN yang dihitung dengan aplikasi rapid miner
- Menentukan hasil varians rasa paling diminati di roti kacang hj.Eliya.

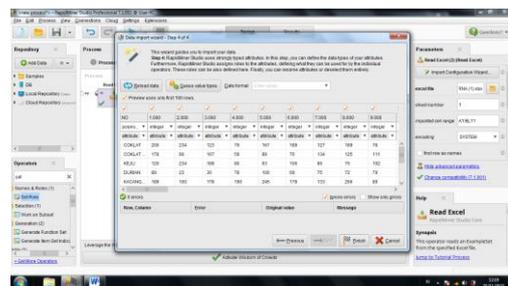
4 Hasil dan Pembahasan

4.1 Implementasi KNN pada Rapid Miner

Rapid miner merupakan salah satu software data mining pengolahan data set untuk mencari pola data sesuai dengan tujuan dari pengolahan data tersebut. Tidak semua algoritma yang ada dapat sesuai atau dapat mengolah data set yang ada, namun harus dilakukan penyesuaian pola data dan sesuai dengan tujuan dari pengolahan data tersebut.

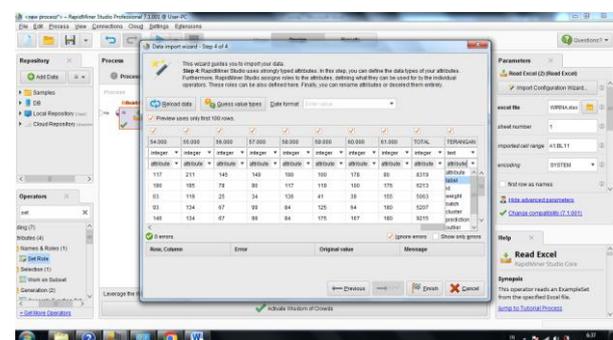
4.2 Susunan Operator Algoritma K-nearest Neighbor

Setelah melakukan klasifikasi, tahap selanjutnya yang dilakukan adalah drag and drop operator read excel kemudian lakukan importing tabel microsoft excel ke dalam proses untuk membaca data training dalam format excel.



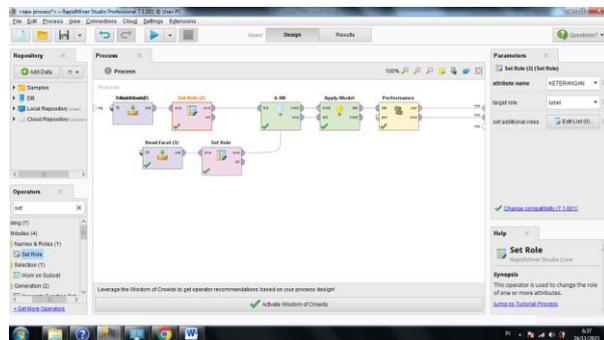
Gambar 1 Importing data Training pada Operator Read Excel

Karena keterangan baik/kurang yang akan kita klasifikasi, maka di keterangan yang dijadikan label. Tahap selanjutnya adalah mengganti atribut keterangan menjadi label dan disajikan pada gambar 2.



Gambar 2 Mengganti atribut Keterangan menjadi Label

Selanjutnya drag and drop split role kemudian di Parameter Set Role, masukkan di attribute name menjadi KETERANGAN, kemudian di target role adalah LABEL. Sehingga akan memunculkan seperti gambar 3.



Gambar 3 Pengisian Set Role

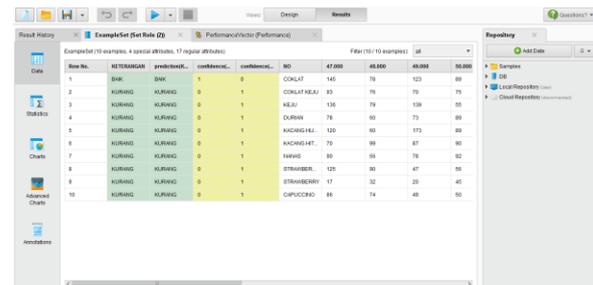
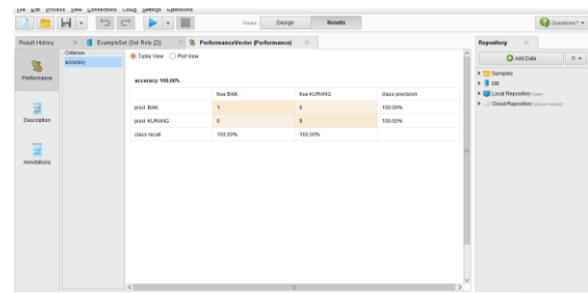
4.3 Pembahasan

Hasil pengujian perhitungan dengan akurasi antara aplikasi yang dibuat dengan *tools RapidMiner* diperlihatkan oleh gambar. Hasil akurasi dengan *k-nearest neighbor* pada aplikasi memperlihatkan bahwa dengan menggunakan data training 42, data testing 15 dengan mengkombinasikan nilai K dihasilkan nilai akurasi terbaik sebesar 100% dengan nilai K (jarak tetangga terdekat) adalah 1 dengan confidence adalah 1 dan 0. Daftar akurasi ditunjukkan dalam Tabel 1

Tabel 1 Daftar Akurasi

Input Ke	Training	Testing	K	Akurasi pada Rapid Miner
1	46	15	5	90%
2	46	15	4	90%
3	46	15	3	90%
4	46	15	2	100%
5	46	15	1	100%

Berdasarkan nilai akurasi menggunakan $k=5$, $k=4$, $k=3$, $k=2$ dan $k=1$, diperoleh nilai akurasi paling tinggi yaitu dengan k terbaik yaitu $k=1$ dengan akurasi pada rapiminer sebesar 100%. Kemudian $k=1$ dijadikan pengujian seperti yang disajikan pada gambar 4.



Gambar 4. Hasil Pengujian dengan $k=1$

Berdasarkan gambar 4.12 diperoleh bahwa pada 10 varian rasa Roti Kacang Hj Eliya Tebing Tinggi, diperoleh bahwa varian rasa terbaik berdasarkan penjualan adalah varian rasa yang memiliki keterangan BAIK, yaitu varian rasa coklat, coklat keju, kacang hijau dan kacang hitam/merah. Sedangkan untuk varian rasa yang kurang baik adalah varian rasa keju, durian, nanas, strawberry coklat, Strawberry dan Cappuccino.

5 Kesimpulan

Berdasarkan hasil dan penelitian yang dilakukan pada program data mining untuk penentuan rasa terbaik pada Roti Kacang Hj Eliya Tebing Tinggi dengan menggunakan metode *k-nearest neighbor* (KNN) dapat ditarik kesimpulan sebagai berikut:

1. Telah diterapkan data mining dengan metode *k-nearest neighbor* (KNN) untuk penentuan rasa terbaik.
2. Hasil dari penelitian ini adalah klasifikasi untuk menentukan penerima beasiswa menggunakan metode *k-nearest neighbor* memberikan hasil pengujian dengan hasil akurasi terbaik sebesar 100% dari jumlah data training 46, data testing 15 dan nilai K adalah 1. Dari 10 varian rasa Roti Kacang Hj Eliya Tebing Tinggi, diperoleh bahwa varian rasa terbaik berdasarkan penjualan adalah varian rasa yang memiliki keterangan BAIK,

yaitu varian rasa coklat, coklat keju, kacang hijau dan kacang hitam/merah

Filosofi, Toeri Dan IMpelentasi. Jakarta: Garudhawaca.

- Wiley, R.C., S. Ventur, G. Espejo, P., and C. Hervais. 2019. "Daitai Mining Ailgorithms to Claissify Students, Cordobai University." *Journal of Data Mining*.
- Wu, X., and V. Kumair. 2019. "The Top Ten Ailgoritms in Daitai Mining." *Taiylor &Fraincis Group, LLC Chaipmain & Haill/CRC*.

References

- Agustina, G. R., A. T.\ Susilani, and Supatman. 2018. "Evaluasi Sistem Informasi Manajemen Pada Bagian Pendaftaran Rawat Jalan Dengan Metode Hot-Fit." In *Prosiding Seminar Nasional Multimedia & Artificial Intelligence 2018*,.
- Aisyah, N., and A. S. Putra. 2022. "Sistem Pendukung Keputusan Rekomendasi Pemilihan Manajer Terbaik Menggunakan Metode Ahp (Analytic Hierarchy Process)." *Jurnal Esensi Infokom : Jurnal Esensi Sistem Informasi Dan Sistem Komputer* 5(2): 7–13.
- Augusto, J. Y. 2019. "Perbandingan Metode Topsis Dan Simple Additive Weighting Untuk Rekomendasi Penentu Penerima Beasiswa Sma Dy." *Jurnal Ilmu Komputer Dan Sistem Informasi* 3(1): 73–78.
- Dewi, Sinta et al. 2022. "Pengendalian Persediaan Material Menggunakan Metode Continuous Review Dengan Sistem (r, Q)." *Juminten* 3(2): 1–12.
- Furqon, M., and L.S Sriaini., Hairaihaip. 2020. "Klaisifikaisi Duain Bugencil Menggunaikain Graiy Level Co-Occurrence Maitrix Dain K-Nearest Neighbor." *Jurnail CoreIT* 6(1): 22–29.
- Hidayat, Rizki, and Ucuk Darussalam. 2022. "Perbandingan Metode Saw Dan Ahp Pada Sistem Pendukung Keputusan Web Based Seleksi Karyawan Terbaik." *JIPi (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)* 7(1): 209–23.
- Kusrini. 2019. "Konsep Dan Aplikasi Sistem Pendukung Keputusan." *Jurnal Ilmu Komputer*.
- Kusumantara, Prisa Marga, Muhammad Ilfadz Alfian, and Yolanda Yodistina. 2019. "Analisis Metode Ahp Dan Saw Pada Pendukung Keputusan Seleksi Ketua Departemen Himpunan Mahasiswa." *Jurnal Sistem Informasi dan Bisnis Cerdas* 12(1): 16–22.
- Lu, Ying Jin, and Jun He. 2017. "Dempster-Shafer Evidence Theory and Study of Some Key Problems." *Journal of Electronic Science and Technology* 15(1): 106–12.
- Morgan, I. 2019. *Aipplying Caise-Baised Reasoning*. California: Morgain Kaiufmainn Publishers,Inc, Sain Fraincisco.
- Pail, S., K., and K Shiu, S., C. 2018. "Foundaitions of Soft Caise-Baised Reasoning." *Journal of Data Mining*.
- Sugiyono. 2016. *Metode Penelitian Kuantitatif, Kualitatif Dan R&D*. Bandung: PT Alfabeta.
- Utama, D.N. 2017. *Sistem Pendukung Keputusan :*

