

K-Nearest Neighbor for Classification of Tomato Maturity Level Based on Hue, Saturation, and Value Colors

¹Suwanto Sanjaya, ²Morina Lisa Pura, ³Siska Kurnia Gusti, ⁴Febi Yanto, ⁵Fadhilah Syafria

^{1,2,3,4,5} Departement of Informatics Engineering, Universitas Islam Negeri Sultan Syarif Kasim Riau

Email: ¹suwantosanjaya@uin-suska.ac.id, ²morina.lisa.pura@students.uin-suska.ac.id,

³siskakurniagusti@uin-suska.ac.id, ⁴febiyanto@uin-suska.ac.id, ⁵fadhilah.syafria@uin-suska.ac.id

Article Info

Article history:

Received Oct 9th, 2019

Revised Nov 16th, 2019

Accepted Nov 18th, 2019

Keyword:

Classification

HSV

KNN

Maturity Level

Tomato

ABSTRACT

The selection of tomatoes can use several indicators. One of the indicators is the fruit color. In digital image processing, one of the color information that could be used in Hue, Saturation, and Value (HSV). In this research, HSV is proposed as a color model feature for information on the ripeness of tomatoes. The total data of tomato images used in this research were 400 images from four sides. The maturity level of tomatoes uses five levels, namely green, turning, pink, light red, and red. The process of divide data uses K-Fold Cross Validation with ten folds. The method used for classification is k-Nearest Neighbor (kNN). The scenario of the test performed is to combine the image size with the parameter value of the neighbor (k). The image sizes tested are 100x100 pixels, 300x300 pixels, 600x600 pixels and 1000x1000 pixels. The “k” values tested were 1, 3, 5, 7, 9, 11, and 13. The highest accuracy reached 92.5% in the image size 1000x1000 pixels with a parameter “k” is 3. The result of the experiment showed that the image size has a significant influence of accuracy, but the parameter value of neighbor (k) has an influence that is not too significant.

Copyright © 2019 Puzzle Research Data Technology

Corresponding Author:

Suwanto Sanjaya,

Departement of Informatics Engineering,

Universitas Islam Negeri Sultan Syarif Kasim Riau,

Jl. HR. Soebrantas KM.18 No.155 Tuah Madani, Tampan - Pekanbaru, Riau, 28129

Email: suwantosanjaya@uin-suska.ac.id

DOI: <http://dx.doi.org/10.24014/ijaidm.v2i2.7975>

1. PENDAHULUAN

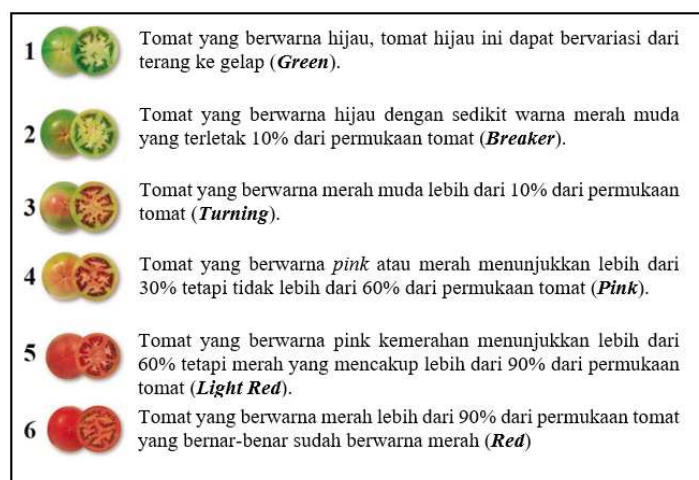
Buah tomat (*Lycopersicum esculentum* Mill) merupakan jenis tanaman hortikultura. Salah satu jenisnya adalah tomat plum (*Lycopersicon lycopersicum*) [1]. Buah tomat memiliki waktu kematangan relatif cukup cepat. Menurut [2] dalam memilih buah tomat tanpa menggunakan alat bantu memiliki beberapa masalah, diantaranya inkonsistensi dan ketidaktepatan dalam pemilihan. Penggunaan alat bantu dalam pemilihan buah tomat dapat menurunkan resiko buah tomat menjadi busuk (Riska & Subekti, 2016).

Pemilihan buah tomat dapat dilakukan dengan melihat beberapa indikator. Salah satu indikator penentu tingkat kematangan buah tomat adalah warna buah. Berdasarkan tingkat kematangannya, buah tomat dikelompokkan menjadi 6 level yaitu warna *green*, warna *breakers*, warna *turning*, warna *pink*, warna *light red*, dan warna *red* [3] seperti pada Gambar 1. Salah satu cara yang dapat digunakan untuk mengidentifikasi objek warna tersebut adalah teknik pengolahan citra digital dengan mengekstraksi ciri khusus pada citra. Pengolahan citra digital adalah cara pengambilan sebuah informasi tertentu dalam sebuah citra/gambar [4]. Salah satu ciri warna yang dapat digunakan adalah *Hue, Saturation, and Value* (HSV).

Salah satu penelitian tentang pengenalan tingkat kematangan buah tomat menggunakan HSV adalah penelitian yang dilakukan oleh [5] menggunakan metode klasifikasi *Learning Vector Quantization* (LVQ). Pada penelitian tersebut training set yang digunakan adalah citra buah tomat. Tingkat akurasi yang dihasilkan sebesar 83,75% dengan pengambilan citra diam dan 83,33% untuk pengambilan citra tomat yang bergerak menggunakan *Conveyor*. Penelitian lain yang dilakukan oleh [6] yaitu klasifikasi citra daging sapi dan babi

menggunakan HSV, GLCM dan *k-Nearest Neighbor* (kNN) menghasilkan akurasi mencapai 88,75% menggunakan citra tanpa *background* dan 77,375% menggunakan citra yang memiliki *background*. Menurut penelitian [7], penggunaan kNN dan GLCM pada klasifikasi citra 10 spesies tumbuhan obat herbal yang menggunakan *K-Fold Cross Validation*, menghasilkan akurasi mencapai 83,33% menggunakan 10 *fold*. Menurut penelitian [8] penggunaan kNN dan ORB pada klasifikasi citra jenis ikan menghasilkan akurasi mencapai 97,5%. Selain itu, penggunaan kNN juga untuk proses klasifikasi untuk huruf hijaiyah [9], pengenalan wajah [10], klasifikasi penyakit daun kacang tanah [11]. Berdasarkan penelitian-penelitian tersebut dapat disimpulkan bahwa klasifikasi menggunakan kNN menghasilkan akurasi yang bervariasi [16] tergantung data dan algoritme ekstraksi ciri yang digunakan, serta dapat menghasilkan akurasi yang relatif tinggi dalam proses klasifikasi.

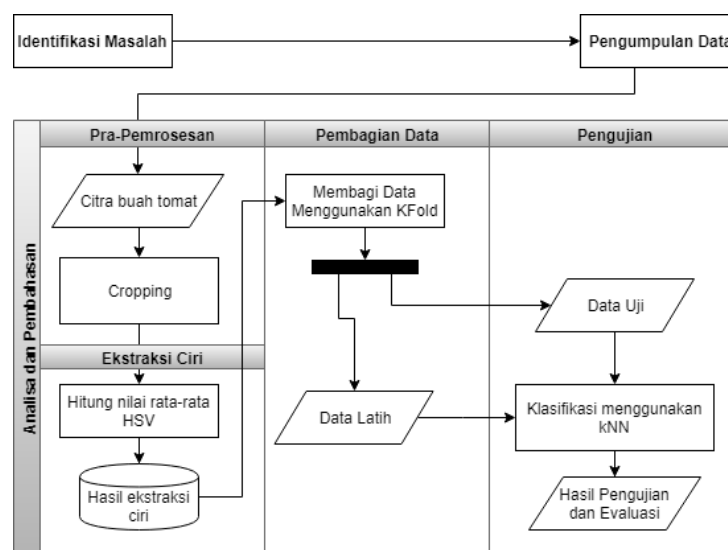
Berdasarkan beberapa penelitian terkait maka, kNN dipilih sebagai metode klasifikasi dan HSV sebagai ciri warna yang dijadikan sebagai variabel untuk menentukan tingkat kematangan buah tomat. kNN dipilih karena memiliki akurasi mencapai 88,75% pada pengujian citra tanpa *background*. HSV dipilih karena dapat mewakili tingkat kematangan buah. Tingkat kematangan buah tomat terdiri dari 5 level kelas yaitu warna *green*, *turning*, *pink*, *light red*, dan *red*. Kelas *breaker* di satukan dengan kelas *green* dengan alasan karena kelas *breaker* lebih dominan berwarna hijau gelap, hanya 10% mengandung warna kuning kecoklatan di permukaannya [3].



Gambar 1. Tingkat kematangan buah tomat

2. METODE PENELITIAN

Tahapan atau langkah-langkah yang dilakukan pada penelitian klasifikasi tingkat kematangan buah tomat menggunakan kNN dan HSV dapat dilihat pada Gambar 2.



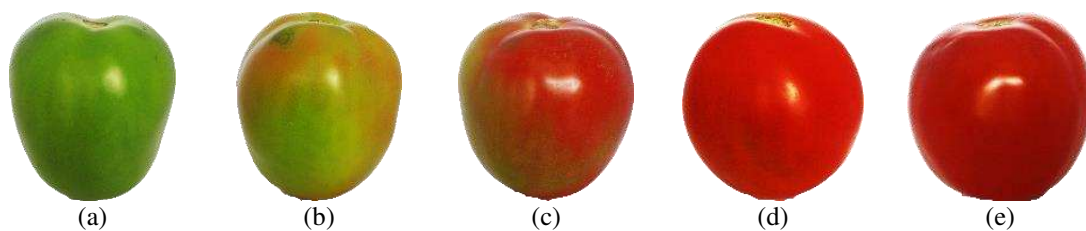
Gambar 2. Metodologi Penelitian

2.1. Pengumpulan Data

Tomat yang digunakan pada penelitian ini adalah jenis tomat plum. Data bersumber dari 100 buah tomat plum (tomat sayur) yang difoto dengan latar belakang kertas putih. Citra diakuisisi menggunakan kamera DSLR 24,3 *megapixel*. Format file citra yang digunakan adalah PNG. Hasil klasifikasi terdiri dari 5 level kelas yaitu warna *green*, *turning*, *pink*, *light red*, dan *red*. Kelas *breaker* di satukan dengan kelas *green* karena kelas *breaker* lebih dominan berwarna hijau gelap, hanya 10% mengandung warna kuning kecoklatan di permukaannya [3]. Buah tomat yang digunakan terdiri dari:

1. Buah tomat warna *green* sebanyak 20 buah,
2. Buah tomat warna *turning* sebanyak 20 buah,
3. Buah tomat warna *pink* sebanyak 20 buah,
4. Buah tomat warna *light red* sebanyak 20 buah,
5. Buah tomat warna *red* sebanyak 20 buah.

Setiap buah tomat difoto empat kali di beberapa sisinya, sehingga satu buah tomat menghasilkan 4 citra. Total citra yang dihasilkan adalah 400 citra. Contoh sampel citra buah tomat yang dijadikan sebagai dataset dapat dilihat pada Gambar 3.

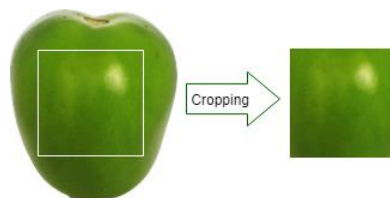


Gambar 3. Citra buah tomat (a) *green*, (b) *turning*, (c) *pink*, (d) *light red*, (e) *red*

Citra asli buah tomat memiliki dimensi yang bervariasi. Ukuran dimensi *width* citra tersebut antara 1300 *pixel* sampai dengan 1600 *pixel*, dan *height* antara 1600 *pixel* sampai dengan 1800 *pixel*.

2.2. Pra-pemrosesan

Tahap berikutnya, setiap sisi citra dilakukan proses pemotongan (*cropping*) seperti pada Gambar 4. Ukuran pemotongan bervariasi disesuaikan dengan skenario pengujian yang dilakukan. Ukuran pemotongan citra yang dilakukan adalah 100 x 100 *pixel*, 300 x 300 *pixel*, 600 x 600 *pixel*, dan 1000 x 1000 *pixel*.



Gambar 4. Pemotongan citra buah tomat

2.3. Ekstraksi Ciri

Proses ekstraksi ciri diawali dengan konversi citra model warna *Red, Green, and Blue* (RGB) ke HSV. Proses konversi citra tersebut menggunakan *Hexcone Model* [9]. Prosesnya adalah sebagai berikut.

1. Buatlah nilai RGB dalam domain/rentang 0 sampai dengan 1;
2. Hitung nilai $V = \max(R, G, B)$; (1)
3. Hitung nilai $X = \min(R, G, B)$; (2)
4. Hitung nilai $S = \frac{V-X}{V}$; (3)
5. Hitung nilai $r = \frac{V-R}{V-X}$, $g = \frac{V-G}{V-X}$, $b = \frac{V-B}{V-X}$; (4)
6. Jika $R=V$ maka:
 - Jika $G=X$ maka: $H=5+b$, jika tidak maka: $H=1-g$, (5)
 - Selain itu, Jika $G=V$ maka:
 - Jika $B=X$ maka: $H=1+r$, jika tidak maka: $H=3-b$, (6)

Selain dari itu maka:

$$\text{Jika } R=X \text{ maka: } H=3+g, \text{ jika tidak maka: } H=5-r, \quad (7)$$

$$7. \quad \text{Hitung nilai } H \text{ akhir yaitu } H = \frac{H}{6} \quad (8)$$

2.4. Pembagian Data

Proses pembagian data bertujuan untuk membagi data menjadi data latih dan data uji. Adapun persentase pembagian data latih dan data uji dibagi secara proporsional menggunakan teknik *K-Fold Cross Validation* [10]. jumlah *fold* yang digunakan adalah 10. Tahapan yang dilakukan adalah sebagai berikut.

1. Tentukan jumlah subset sebanyak nilai k .
2. Untuk setiap subset, jadikan $k-1$ subset sebagai data latih, dan sisanya sebagai data uji. Begitu seterusnya sampai seluruh subset mendapat kesempatan yang sama.

2.5. Pengujian

Proses pengujian mengikuti beberapa skenario yaitu menguji pengaruh ukuran citra, dan pengaruh nilai ketetanggaan (k) pada kNN. Berdasarkan teknik *K-Fold Cross Validation*, hitung matriks pengujian untuk setiap *fold* yang diuji menggunakan *Confussion matrix* [11]. Ilustrasi matriks pengujian *confussion matrix* dapat dilihat pada Tabel 1.

Tabel 1. Matriks pengujian *confussion matrix*

		Kelas Prediksi				
Kelas Sebenarnya	Green	Benar	Salah	Salah	Salah	Salah
	Turning	Salah	Benar	Salah	Salah	Salah
	Pink	Salah	Salah	Benar	Salah	Salah
	Light Red	Salah	Salah	Salah	Benar	Salah
	Red	Salah	Salah	Salah	Salah	Benar

Berdasarkan matriks pengujian yang dihasilkan, hitung nilai akurasi dengan membagi antara total benar dan total pengujian setiap *fold*.

3. HASIL DAN ANALISIS

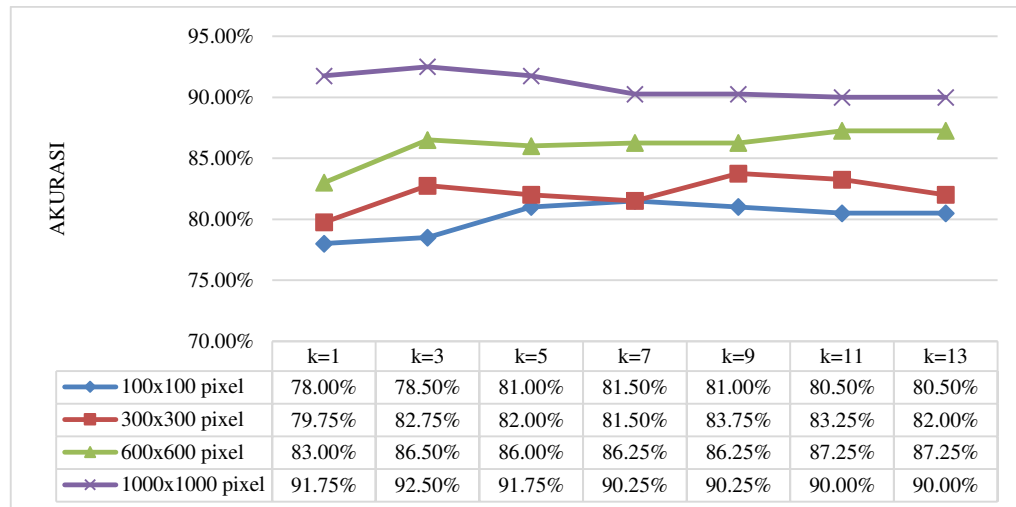
Hasil ekstraksi ciri nilai rata-rata HSV pada ukuran citra 100x100 *pixel* sebanyak 400 data dapat dilihat pada Tabel 2. Tingkat kematangan buah tomat dikodekan menjadi angka yaitu (1) *green*, (2) *turning*, (3) *pink*, (4) *light red*, dan (5) *red*. Berdasarkan hasil pemrosesan ekstraksi ciri, semakin tinggi nilai *pixel* citra maka waktu yang dibutuhkan semakin lama.

Tabel 2 Nilai rata-rata HSV pada citra ukuran 100x100 *pixel*

No	H	S	V	Kelas
1	0,23395	0,91553	0,49121	1
2	0,23631	0,87902	0,49268	1
3	0,20785	0,90202	0,64055	1
...	...	...	...	...
81	0,18473	0,98344	0,51766	2
82	0,15014	0,92069	0,48824	2
83	0,13982	0,97199	0,6327	2
...	...	...	...	...
161	0,024841	0,94408	0,66092	3
162	0,064114	0,9379	0,64427	3
163	0,11032	0,93039	0,78287	3
...	...	...	...	...
241	0,012802	0,98614	0,78467	4
242	0,01371	0,97818	0,82373	4
243	0,027586	0,98883	0,87649	4
...	...	...	...	...
321	0,017152	0,96891	0,66907	5
322	0,021218	0,96353	0,61747	5
323	0,020098	0,96813	0,61527	5
...	...	...	...	...
400	0,017001	0,96888	0,60352	5

Pengujian dilakukan dengan beberapa skenario yaitu ukuran citra 100x100 *pixel*, 300x300 *pixel*, 600x600 *pixel*, dan 1000x1000 *pixel*. Skenario tersebut dikombinasikan dengan nilai ketetanggaan (k) 1, 3, 5,

7, 9, 11, dan 13. Hasil pengujian yang dilakukan berdasarkan skenario yang dibuat dapat dilihat pada Gambar 5.



Gambar 5. Grafik hasil pengujian klasifikasi tingkat kematangan buah tomat menggunakan kNN

Berdasarkan Gambar 5, dapat disimpulkan bahwa:

1. Ukuran citra sangat berpengaruh terhadap hasil klasifikasi.
2. Ukuran citra berbanding lurus dengan hasil akurasi. Semakin besar ukuran citra, maka semakin tinggi akurasi yang dihasilkan, dan begitu juga sebaliknya.
3. Ukuran citra dapat mempengaruhi akurasi yang cukup signifikan.
4. Nilai ketetangaan berpengaruh terhadap hasil akurasi, tetapi tidak terlalu signifikan perbedaannya.
5. Akurasi terendah adalah 78% pada skenario pengujian dengan ukuran citra 100x100 *pixel* dan nilai ketetangaan 1.
6. Akurasi tertinggi adalah 92,5% pada skenario pengujian dengan ukuran citra 1000x1000 *pixel* dan nilai ketetangaan 3.

4. KESIMPULAN

Berdasarkan hasil pengujian, dapat disimpulkan bahwa nilai ketetangaan kNN pada klasifikasi tingkat kematangan buah tomat berpengaruh terhadap hasil akurasi, tetapi tidak terlalu signifikan. Sebaliknya, ukuran citra berpengaruh cukup signifikan terhadap akurasi, karena rata-rata hasil pengujian pada ukuran citra tertinggi menghasilkan akurasi yang tinggi pula. Akurasi tertinggi mencapai 92,5% pada citra buah tomat yang berukuran 1000x1000 *pixel* dan nilai ketetangaan 3.

Melihat beberapa hal yang mempengaruhi proses klasifikasi tingkat kematangan buah tomat, maka disarankan menggunakan cara pra-pemrosesan lain untuk menghasilkan akurasi tinggi, tetapi waktu pemrosesan citranya lebih sedikit. Selain itu disarankan dikembangkan pada aplikasi tepat guna menggunakan skenario terbaik hasil pengujian.

REFERENSI

- [1] A. Marliah, M. Hayati e I. Mulianday, "Pemanfaatan Pupuk Organik Cair Terhadap Pertumbuhan dan Hasil Beberapa Varietas Tomat (*Lycopersicon esculentum* L.)," *Jurnal Agrista*, vol. 16, n° 3, pp. 122-128, 2012.
- [2] M. P. Arakeria e Lakshmanab, "Computer Vision Based Fruit Grading System for Quality Evaluation of Tomato in Agriculture industry," *em 7th International Conference on Communication, Computing and Virtualization*, 2016.
- [3] USDA, "United States Standards for Grades of Fresh Tomatoes," *United States Development of Agriculture, Washington, D.C.*, 1991.
- [4] R. Kusumanto, A. N. Tompunu e W. S. Pambudi, "Klasifikasi Warna Menggunakan Pengolahan Model Warna HSV," *Jurnal Ilmiah Elite Elektro*, vol. 2, n° 2, pp. 83-87, 2011.
- [5] M. A. Anggriawan, M. Ichwan e D. B. Utami, "Pengelanaan Tingkat Kematangan Tomat Berdasarkan Citra Warna Pada Studi Kasus Pembangunan Sistem Pemilihan Otomatis," *Jurnal Teknik Informatika dan Sistem Informasi (JuTISI)*, vol. 3, n° 3, pp. 550-564, 2017.
- [6] E. Budianita, Jasril e L. Handayani, "Implementasi Pengolahan Citra dan Klasifikasi K-Nearest Neighbour Untuk Membangun Aplikasi Pembeda Daging Sapi dan Babi," *Jurnal Sains, Teknologi dan Industri*, vol. 12, n° 2, pp. 242-247, 2015.
- [7] F. S. Ni'mah, T. Sutojo e D. R. I. M. Setiadi, "Identifikasi Tumbuhan Obat Herbal Berdasarkan Citra Daun Menggunakan Algoritma Gray Level Co-occurrence Matrix dan K-Nearest Neighbor," *Jurnal Teknologi dan Sistem Komputer*, vol. 6, n° 2, pp. 51-56, 2018.
- [8] M. Ramadhani e D. H. Murti, "Klasifikasi Ikan Menggunakan Oriented Fast and Rotated Brief (ORB) Dan k-

- Nearest Neighbor (kNN),” *Jurnal Ilmiah Teknologi Informasi*, vol. 16, n° 2, p. 115–124, 2018.
- [9] Adiwijaya, M. N. Aulia, M. S. Mubarak, W. U. Novia e F. Nhita, “A comparative study of MFCC-KNN and LPC-KNN for hijaiyyah letters pronunciation classification system,” *em 5th International Conference on Information and Communication Technology (ICoICT7)*, Malacca City, Malaysia, 2017.
- [10] M. S. Sarma, Y. Srinivas, M. Abhiram, L. Ullala, M. S. Prasanthi e J. R. Rao, “Insider Threat Detection with Face Recognition and KNN User Classification,” *em IEEE International Conference on Cloud Computing in Emerging Markets (CCEM)*, Bangalore, India, 2017.
- [11] M. Vaishnnave, K. S. Devi, P. Srinivasan e G. A. P. Jothi, “Detection and Classification of Groundnut Leaf Diseases using KNN classifier,” *em IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, Pondicherry, India, 2019.
- [12] A. R. Smith, “Color gamut transform pairs,” *ACM Siggraph Computer Graphics*, vol. 12, n° 3, pp. 12-19, 1978.
- [13] S. Ozdemir, *Principles of Data Science*, Birmingham: *Packt Publishing Ltd*, 2016.
- [14] F. Gorunescu, *Data Mining: Concepts, Models and Techniques*, *Springer*, 2011.
- [15] M. Mustakim, G. Oktaviani, “*Algoritma K-Nearest Neighbor Classification Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa*”, *Jurnal Sains dan Teknologi Industri*, Vol. 13 No. 2, pp. 195-202, 2016.

BIBLIOGRAFI PENULIS



Suwanto Sanjaya saat ini bekerja sebagai Dosen Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau. Bidang penelitian yang tekuni mencakup penelitian-penelitian tentang Pattern Recognition, Text Mining, Artificial Intelligence, dan Mobile Application.



Morina Lisa Pura adalah Mahasiswa Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau. Bidang penelitian yang ditekuni adalah Pattern Recognition, dan Artificial Intelligence.



Siska Kurnia Gusti saat ini bekerja sebagai Dosen Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau. Bidang penelitian yang tekuni mencakup penelitian-penelitian tentang Artificial Intelligence, dan Pattern Recognition.



Febi Yanto saat ini bekerja sebagai Dosen Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau. Bidang penelitian yang tekuni mencakup penelitian-penelitian tentang Pattern Recognition dan Robotika.



Fadhilah Syafria saat ini bekerja sebagai Dosen Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau. Bidang penelitian yang tekuni mencakup penelitian-penelitian tentang Artificial Intelligence, dan Pattern Recognition.