



KLASIFIKASI TINGKAT PRODUKTIVITAS TANAMAN PADI DI KABUPATEN KARAWANG MENGUNAKAN METODE NAÏVE BAYES DAN K-FOLD CROSS VALIDATION

Asep Muhidin, Ahmad Burhan

Program Studi Teknik Informatika Sekolah Tinggi Teknologi Pelita Bangsa
asep.muhidin@pelitabangsa.ac.id

Abstraksi

Penelitian ini dilakukan dengan tiga tahap, yaitu tahap pemrosesan data, tahap pengujian dan tahap mencari akurasi model. Tahap pemrosesan data menggunakan *software* IBM SPSS Statistic 22 untuk pencarian data *missing* dan mengganti nilai *missing* dengan rata-rata/*mean*. Tahap pengujian menggunakan *software* RapidMiner Studio untuk mencari nilai akurasi dengan skema *k-fold validation* dan melakukan pengujiaannya sebanyak sepuluh kali. Hasil dari tiap pengujian yang didapatkan berupa *Confusion matrix* dan nilai yang diambil adalah *accuracy* yang akan dijadikan sebagai nilai pencarian dari akurasi model dalam klasifikasi tingkat produktivitas padi. Dari pengujian yang dilakukan sebanyak sepuluh kali didapatkan nilai akurasi dari model dalam mengklasifikasikan tingkat produktivitas dengan skema *k-fold validation* sebesar 87,22% ini membuktikan bahwa metode *naïve bayes* mempunyai klasifikasi yang baik. Hasil dari klasifikasi bisa menjadi acuan pihak pemerintah dalam menentukan kecamatan mana yang akan dijadikan perhatian khusus dalam budidaya padi di Kabupaten Karawang.

Kata kunci: *Naïve Bayes, K-Fold Validation, Rapid Miner, IBM Statistic 22, Produktivitas padi, Klasifikasi.*

Abstract

This research was conducted in three stages, namely the data processing stage, the testing stage and the stage of searching for model accuracy. The data processing stage uses IBM SPSS Statistics 22 software to find missing data and replace missing values with mean. The testing phase uses the RapidMiner Studio software to find the accuracy value with the k-fold validation scheme and to test it ten times. The results of each test obtained in the form of Confusion matrix and the value taken is accuracy which will be used as a search value of the accuracy of the model in the classification of rice productivity levels. From the tests carried out ten times, the accuracy value of the model in classifying productivity levels with the k-fold validation scheme of 87.22% proves that the naïve bayes method has a good classification. The results of the classification can be a reference for the government in determining which sub-districts will be of particular concern in rice cultivation in Karawang Regency.

Keywords: *Naïve Bayes, K-Fold Validation, Rapid Miner, IBM Statistic 22, Rice Productivity, Classification.*

1. Pendahuluan

Penelitian ini dilakukan pada lingkungan Dinas Kehutanan dan Pertanian Kabupaten Karawang dengan subyek penelitian adalah produktivitas tanaman padi. Karawang adalah daerah penghasil beras terbesar kedua di propinsi Jawa Barat.

Kesalahan dalam penentuan produktivitas tanaman padi akan dapat mengakibatkan antara lain : produktivitas menurun, gagal panen dan kerugian. Untuk membantu peningkatan produktivitas beras perlu adanya suatu model data yang mampu mengklasifikasikan data atau informasi yang masuk.

Adapun model data yang akan digunakan dalam proses pengklasifikasian data yang ada pada rentang waktu tertentu untuk dilakukan analisa lanjutan yaitu dengan menggunakan algoritma *Naïve Bayes* adapun untuk validasi menggunakan

K-FOLD CROSS VALIDATION. Adapun *tools* yang digunakan adalah *Rapidminer* dan *SPSS*.

2. Tinjauan Studi

2.1. Pengertian Data Mining

Menurut Chapman (2014), “Data mining membantu menemukan struktur yang mendasari dalam data, untuk mengubah data menjadi informasi, dan informasi menjadi pengetahuan. Penggalian data adalah ekstraksi informasi yang implisit, sebelumnya tidak diketahui, dan berpotensi berguna dari data. Model yang diekstraksi secara otomatis memberikan wawasan tentang perilaku pelanggan dan dalam proses menghasilkan data, tetapi juga dapat diterapkan, misalnya, secara otomatis mengklasifikasikan objek atau dokumen atau gambar ke dalam kategori tertentu, untuk memperkirakan variabel target numerik, untuk memprediksi nilai masa depan dari seri waktu yang diamati, dan banyak tugas lain di mana data membantu untuk membuat keputusan yang lebih baik atau bahkan untuk mengotomatisasi keputusan dan proses.”

Sedangkan menurut Julian (2015), “Data mining berbicara tentang memecahkan masalah dengan menganalisis data yang ada dalam database. Saat ini, hal itu memenuhi syarat sebagai ilmu pengetahuan dan teknologi untuk menjelajahi data untuk menemukan pola yang sudah ada yang tidak diketahui.”

2.2. Proses Knowledge Discovery in Database

Menurut Julian (2015), “Banyak orang membedakan DM sebagai sinonim dari proses *Knowledge Discovery in Databases* (KDD), sementara yang lain melihat DM sebagai langkah utama KDD. Ada berbagai definisi KDD. Misalnya, ada yang mendefinisikannya sebagai proses trivial mengidentifikasi pola yang valid, baru, berpotensi berguna, dan pada akhirnya dapat dipahami dalam data”

Pada KDD ini mengadopsi *hybridization* (hibridisasi) yang banyak digunakan dalam beberapa tahun terakhir yang mengkategorikan tahapan-tahapan ini menjadi enam langkah (Julian, 2015) :

1. *Spesifikasi Masalah/Problem Specification*
Merancang dan mengatur domain aplikasi, pengetahuan awal yang relevan yang diperoleh oleh para ahli dan tujuan akhir yang dikejar oleh pengguna akhir.
2. *Pemahaman masalah/Problem Understanding*
Pemahaman data yang dipilih untuk pendekatan dan pengetahuan ahli yang terkait untuk mencapai tingkat kehandalan yang tinggi.
3. *Pemrosesan Data/Data Preprocessing*
Tahap ini mencakup operasi untuk *Data*

Cleaning (seperti menangani noise dan data yang tidak konsisten), *Data Integration* (di mana beberapa sumber data dapat digabungkan menjadi satu), *Data Transformation* (di mana data diubah dan dikonsolidasikan ke dalam bentuk yang sesuai untuk spesifik Tugas DM) dan *Data reduction*, termasuk pemilihan dan ekstraksi fitur dan contoh dalam *database*. Fase ini akan menjadi fokus studi di seluruh buku.

4. *Penambangan Data/Datamining*
Ini adalah proses penting di mana metode yang digunakan untuk mengekstrak pola data yang valid. Langkah ini termasuk pilihan tugas DM yang paling sesuai, (seperti klasifikasi, regresi, *clustering* dan asosiasi)
5. *Evaluasi/Evaluation*
Memperkirakan dan menafsirkan pola-pola yang ditambah berdasarkan ukuran-ukuran yang menarik
6. *Hasil Eksploitasi/Result Exploitation*
Tahap terakhir mungkin melibatkan menggunakan pengetahuan secara langsung, menggabungkan pengetahuan ke dalam sistem lain untuk proses lebih lanjut atau hanya melaporkan pengetahuan yang ditemukan melalui alat visualisasi.

2.3. Data Preprocessing

Langkah selanjutnya adalah mempertanyakan data yang akan digunakan. Data masukan harus disediakan dalam jumlah, struktur, dan format yang sesuai dengan setiap tugas DM dengan sempurna. Sayangnya, *database* dunia nyata sangat dipengaruhi oleh faktor-faktor negatif seperti kehadiran *noise*, *MV*, data yang tidak konsisten dan berlebihan dan ukuran besar baik dalam dimensi, contoh dan fitur. Dengan demikian, data berkualitas rendah akan mengarah pada kinerja DM berkualitas rendah (Julian, 2015).

2.4. Supervised Learning

Dalam komunitas data mining, metode prediksi sering disebut sebagai *supervised learning*. Metode *supervised* dianggap mencoba menemukan hubungan antara atribut input (kadang-kadang disebut variabel atau fitur) dan atribut target (kadang-kadang disebut kelas). Hubungan yang dicari diwakili dalam struktur yang disebut model. Umumnya, model menggambarkan dan menjelaskan pengalaman, yang tersembunyi dalam data, dan yang dapat digunakan dalam prediksi nilai atribut target, ketika nilai atribut *input* diketahui. *Supervised learning* hadir di banyak domain aplikasi, seperti keuangan, obat-obatan dan teknik. Dalam skenario *supervised learning*, satu set pelatihan diberikan dan tujuannya adalah untuk membentuk

deskripsi yang dapat digunakan untuk memprediksi contoh yang tidak terlihat (Julian, 2015). Klasifikasi, regresi dan *time series* masuk dalam kategori *supervised learning*.

2.5. Unsupervised Learning

Dalam *Unsupervised learning* tidak ada guru (label dalam data) hanya data input yang tersedia. Dengan demikian tujuan dari *unsupervised learning* adalah menemukan keteraturan, ketidakberesan, hubungan, persamaan dan asosiasi dalam masukan. Dengan *unsupervised learning*, adalah mungkin untuk belajar model yang lebih besar dan lebih kompleks dari pada dengan *supervised learning*. *Clustering* masuk dalam kategori *unsupervised learning* (Julian, 2015).

2.6. Algoritma Naïve Bayes

Dalam pembelajaran machine learning, klasifikasi dianggap sebagai tuntunan dari metode supervised learning, yaitu menyimpulkan fungsi dari data pelatihan berlabel. Data pelatihan terdiri dari satu set contoh pelatihan, di mana setiap contoh adalah pasangan yang terdiri dari objek input (biasanya vektor dan nilai output yang diinginkan (biasanya label kelas). Dengan demikian, tugas dari algoritma klasifikasi adalah untuk menganalisis data pelatihan dan menghasilkan fungsi yang disimpulkan, yang dapat digunakan untuk mengklasifikasikan contoh baru. Sub-kelas umum klasifikasi adalah klasifikasi probabilistik, Algoritma klasifikasi probabilistik menggunakan inferensi statistik untuk menemukan kelas terbaik untuk contoh yang diberikan (Charu, 2015).

Salah satu algoritma klasifikasi probabilistik adalah naive bayes classifier, algoritma naive bayes merupakan penggolongan probabilistik sederhana berdasarkan penerapan teorema bayes dengan asumsi independensi yang kuat, dengan kata lain algoritma naive bayes mengasumsikan bahwa keberadaan nilai tertentu dari suatu atribut tidak terkait dengan kehadiran nilai atribut lainnya (Markus, 2014). Klasifikasi naive bayes sangat cocok ketika dimensi input tinggi, dan memiliki kinerja yang sebanding dengan beberapa metode klasifikasi lainnya seperti decision tree dan neural network classifier (Charu, 2015).

Berikut adalah persamaan dari teorema bayes:

$$p(H|D) = \frac{p(H)p(D|H)}{p(D)}$$

- Keterangan:
- P: Probabilitas
- D: Data dengan class yang belum diketahui.
- H: Hipotesis data merupakan class spesifik.
- p(H|D): Probabilitas hipotesis H berdasarkan kondisi D (posterior probability).
- p(H): Probabilitas hipotesis H (prior probability).
- p(D|H): Probabilitas D berdasarkan kondisi pada hipotesis H.
- p(D): Probabilitas D

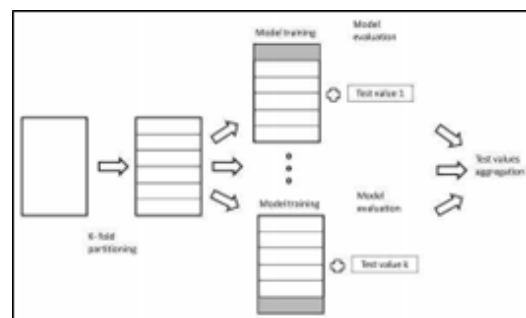
Langkah penyelesaian Naïve Bayes :

Dalam menyelesaikan permasalahan yang ada algoritma naive bayes dalam perhitungannya terdiri dari beberapa langkah perhitungan, yaitu :

1. Mulai
2. Baca data training a) Mencari nilai p(H) untuk setiap kelas. b) Mencari nilai p(D|H) untuk setiap kriteria dari setiap kelas. c) Mencari nilai probabilitas paling besar dari kelas (hasil p(H) x hasil p(D|H))
3. Menampilkan hasil probabilitas (probabilitas nilai kelas terbesar dijadikan acuan).
4. Selesai

2.7. K-Fold Cross Validation

Dalam *k-fold cross validation*, data asli dipartisi kedalam subset/bagian. Model ini kemudian dibangun menggunakan data dari subset K-1 (2,3,4,5, dst), dan terdapat bagian lain didalam subset yang digunakan untuk set tes. Bagian subset (*dataset*) harus lebih banyak dari pada set tes, dilakukan secara iteratif sampai memiliki model yang berbeda. Hasil dari masing-masing model K (akurasi) kemudian digabungkan menggunakan rata-rata untuk mendapatkan hasil akurasi dari keseluruhan data. *K-fold cross-validation* yang sering digunakan adalah *10-fold cross validation*. Manfaat menggunakan *k-fold cross-validation* adalah setiap record muncul dalam set tes tepat satu kali, kekurangannya adalah bahwa tugas validasi yang diperlukan dibuat lebih sulit (Daniel, 2014). Berikut adalah skema dari *K-Fold cross validation* :



Gambar 1. Proses K-Fold Cross Validation

2.8. Pengertian Produktivitas

Nurmala, dkk (2012), Produktivitas adalah kemampuan tanah untuk menghasilkan produksi tanaman tertentu dalam keadaan pengolahan tanah tertentu. Produktivitas merupakan perwujudan dari keseluruhan faktor-faktor (tanah dan non tanah) yang berpengaruh terhadap hasil tanaman yang lebih berdasarkan pada

pertimbangan ekonomi.

2.9. Faktor Yang Mempengaruhi Produktivitas Padi

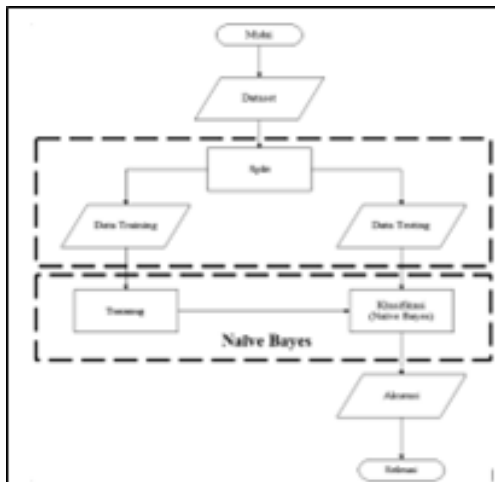
Berdasarkan data yang diperoleh dari Dinas Pertanian dan Kehutanan Kabupaten Karawang (2010

– 2015) bahwa yang mempengaruhi produktivitas padi di kabupaten karawang meliputi beberapa hal, antara lain : luas sawah, luas baku sawah, luas tanam, luas panen, produksi, curah hujan, hari hujan, *organisme* pengganggu tanaman.

3. Metode Penelitian

3.1. Proses Pengujian

Proses pengujian yang dimaksud pada tahap ini adalah proses pengujian kinerja metode klasifikasi *naïve bayes*. Untuk proses pengujian ini menggunakan *k-fold cross validation*. Metode evaluasi yang umum digunakan adalah *10-fold cross validation*, *10-fold cross validation* akan mengulang pengujian sebanyak 10 kali dan hasil pengukuran adalah nilai rata-rata dari 10 kali pengujian.



Gambar 2. Proses pengujian dengan K-Fold Cross Validation

3.2. Akurasi

Pada tahap ini dilakukan perhitungan rata-rata akurasi berdasarkan keseluruhan pengujian. Ini dilakukan untuk mengetahui seberapa besar akurasi dari keseluruhan data, sebagai acuan seberapa baik metode *naïve bayes* dalam memprediksi tingkat produktivitas tanaman padi. Tingkat akurasi menunjukkan tingkat kebenaran klasifikasi terhadap kelas. Semakin rendah akurasi berpengaruh terhadap tingkat kesalahannya sebaliknya untuk tingkat akurasi yang tinggi semakin rendah tingkat kesalahannya. Tingkat akurasi yang baik adalah tingkat presentase yang mendekati angka 100%. Untuk perhitungan akurasinya dapat dihitung sebagai berikut :

$$\text{Rata-rata Akurasi} = \frac{\sum \text{Hasil Akurasi Setiap Pengujian}}{\sum \text{Jumlah Total Pengujian}}$$

4. Hasil dan Pembahasan

a. Dataset

TAKHIN	KECAMAT	RATA-RATA	RATA-RATA	LUAS TAN	PRODUKSI	LUAS PAN	LUAS BAKU	LUAS SAW	PENGERG	TIKUS	WERENG	ESIPUT	MUJI	BAKTERI	HAMA	PU	BLASIT	PRODUKT
2010 Karawang	1029	1117	4542	32327	4533	2243	2243	36	30	24	0	0	0	0	0	0	0	0
2010 Karawang	1029	1117	3844	23207	3845	1847	1847	46	35	12	0	0	0	0	0	0	0	0
2010 Majalaya	8467	4867	4496	31165	4446	2233	2233	0	76	99	0	0	0	0	0	0	0	0
2010 Klati	2007	1317	5434	38495	5434	2392	2392	133	175	342	70	0	0	0	0	0	0	0
2010 Telukjambe	222	11	4256	27403	4196	2308	2308	0	0	119	0	0	0	0	0	0	0	0
2010 Telukjambe	222	11	2771	13238	2064	935	935	0	0	535	0	0	0	0	0	0	0	0
2010 Ciampel	159	3034	2136	10797	1507	852	852	101	83	377	29	0	0	0	0	0	0	0
2010 Pangkajene	1502	30	4889	34425	4606	2341	2341	103	74	244	70	0	0	0	0	0	0	0
2010 Tegayuhan	3502	30	4066	28638	3837	1912	1912	147	52	152	77	0	0	0	0	0	0	0
2010 Rangasde	1404	5	4016	28848	4817	2026	2026	77	44	37	0	0	0	0	0	0	0	0
2010 Jayawanta	1834	0	7242	48839	7242	3571	3571	40	145	0	0	0	0	0	0	0	0	0
2010 Kutawaluyo	1404	5	3744	63079	3744	4172	4172	100	78	74	0	0	0	0	0	0	0	0
2010 Butuyaya	1842	0	5862	68418	5862	4931	4931	553	499	441	37	0	0	0	0	0	0	0
2010 Tertuyaya	1759	0	11316	76736	11316	5658	5658	222	247	36	0	0	0	0	0	0	0	0
2010 Paksiyaya	0	0	6382	40820	5980	3166	3166	87	211	103	41	0	0	0	0	0	0	0
2010 Ploret	1842	7884	30312	71886	30312	5156	5156	251	179	180	74	0	0	0	0	0	0	0
2010 Cilebar	1404	5	20820	72346	20820	5427	5427	132	36	54	69	0	0	0	0	0	0	0
2010 Cibuyaya	1339	5334	7892	54926	7892	3948	3948	25	143	544	0	0	0	0	0	0	0	0
2010 Cikampek	3369	17	1282	8474	1282	641	641	6	39	74	10	0	0	0	0	0	0	0
2010 Purwasari	2960	17	3222	19037	3222	1811	1811	45	31	377	0	0	0	0	0	0	0	0
2010 Terasmy	1881	3029	5942	38769	5942	2521	2521	109	278	644	4	0	0	0	0	0	0	0
2010 Jatitani	221	1367	8208	52203	8208	4324	4324	39	181	1212	0	0	0	0	0	0	0	0

Gambar 3. Dataset

b. Preprocessing Data / Missing Value

	N	Missing	
		Count	Percent
RATARATACURAHUJAN	172	8	4.4
RATARATAHARIHUJAN	167	13	7.2
PENGGEREKBATANG	175	5	2.8
TIKUS	170	10	5.6
WERENGBATANGCOKEL	116	64	35.6
AT			
SIPUTMURBAI	78	102	56.7
BAKTERIHAWARDAUN	48	132	73.3
HAMAPUTIHVALSU	36	144	80.0
BLASIT	2	178	98.9
LUASTANAM	180	0	.0
PRODUKSI	180	0	.0
LUASPANEN	180	0	.0
LUASBAKUSAWAH	180	0	.0
LUASSAWAH	180	0	.0

Gambar 4. Missing Value

c. Hasil Transformasi Data

Untuk melakukan transformasi data, penulis menentukan nilai batas interval yang akan menjadi acuan untuk merubah data kedalam bentuk kategorikal yaitu dengan menentukan nilai tinggi, sedang dan rendah.

Gambar 5. Bentuk Data Set setelah transformasi Data

d. Proses Pengujian

Dari proses penghitungan *Naïve Bayes* di dapat hasil dengan produktivitas tinggi pada kelas (P)Produktivitas tinggi) yaitu sebesar 9,43131E-05. Setelah proses penghitungan maka akan

dilakukan proses pengujian dilakukan tahapan yang telah direncanakan yaitu dengan metode naïve bayes dengan menggunakan skema 10-fold *cross validation*, dengan artian dataset akan dibagi menjadi N bagian secara acak. Fold ke-1 adalah ketika bagian ke- 1 menjadi data *testing* dan sisanya menjadi data *training*, demikian seterusnya hingga sampai fold 10 bagian ke-10. Pengujian pada penelitian ini menggunakan tool rapid miner untuk mengolah dataset.

e. Analisa Hasil Klasifikasi Produktivitas Tanaman Padi

Dari beberapa pengujian di atas, data yang diambil untuk menjadi acuan dalam menilai performa naïve bayes dalam memprediksi produktivitas padi adalah nilai *accuracy*. Berikut ini adalah ringkasan dari tingkat akurasi tiap pengujian yang memiliki nilai berbeda-beda.

Tabel 1. Tingkat Akurasi Pengujian

No	Pengujian	Accuracy
1	Pengujian ke 1	83,33%
2	Pengujian ke 2	89,89%
3	Pengujian ke 3	88,89%
4	Pengujian ke 4	88,89%
5	Pengujian ke 5	88,89%
6	Pengujian ke 6	88,89%
7	Pengujian ke 7	77,78%
8	Pengujian ke 8	88,89%
9	Pengujian ke 9	94,44%
10	Pengujian ke 10	83,33%

5. Kesimpulan

Berdasarkan hasil penelitian ini maka didapat beberapa kesimpulan sebagai berikut:

1. Penelitian ini dilakukan dengan tiga tahap, yaitu tahap pemrosesan data, tahap pengujian dan tahap mencari akurasi model. Tahap pemrosesan data menggunakan software IBM SPSS Statistic 22 untuk pencarian data missing dan mengganti nilai missing dengan rata-rata mean. Tahap pengujian menggunakan software RapidMiner Studio untuk mencari nilai akurasinya dengan skema *k-fold validation* dan melakukan pengujiannya sebanyak sepuluh kali. Hasil dari tiap pengujian yang didapatkan berupa *Confusion matrix* dan nilai yang diambil adalah *accuracy* yang akan dijadikan sebagai nilai pencarian dari akurasi model dalam mengklasifikasi tingkat produktivitas padi.
2. Dari pengujian yang dilakukan sebanyak sepuluh kali didapatkan nilai akurasi dari model dalam memprediksi tingkat produktivitas dengan skema *k-fold validation* sebesar 87,22% ini membuktikan bahwa metode naïve bayes mempunyai klasifikasi yang baik.
3. Hasil dari klasifikasi bisa menjadi acuan

pihak pemerintah dalam menentukan kecamatan mana yang akan dijadikan perhatian khusus dalam budi daya padi di Kabupaten Karawang.

Daftar Pustaka

- [1] A.S Rosa., M, Shalahudin. 2014. *Rekayasa Perangkat Lunak Struktur Dan Berorientasi Objek*. Bandung : Informatika.
- [2] Aggarwal, Charu C. 2014. *Data Classification Algorithms And Applications*. Chapman And Hall/CRC Data Mining and Knowledge Discovery Series. United States : CRC Press.
- [3] B. Irawan, “Dinamika Produktivitas dan Kualitas Budi Daya Padi Sawah”, 2016, pp. 179–199.
- [4] BMKG. Pengertian Hari Hujan, Curah Hujan tersedia di <https://bmgk.go.id/> diakses pada tanggal 30 Agustus 2018.
- [5] BPS Nasional. Pengertian Luas Sawah, Luas Tanam, Luas Panen, Produksi tersedia di <https://bps.go.id/> diakses pada 30 Agustus 2018.
- [6] BPS Provinsi Jawa Barat. Data Produksi Gabah Kering. 2015 tersedia di <https://jabar.bps.go.id/> diakses pada tanggal 30 Agustus 2018.
- [7] Davis, Barbara., Radford, Daren. 2014. *Going Beyond The Waterfall : Managing Scope Effectively Across The Project Life Cycle*. U.S.A : J.Ross Publishing.
- [8] Dennis, Alan., Wixom, Barbara Haley., Tegarden, David. 2015. *Systems Analysis And Design : An Object-Oriented Approach with UML 5th Edition*. United States : John Wiley And Sons Ins.
- [9] Departemen Pertanian. Pengertian Luas Baku Sawah tersedia di <https://pertanian.go.id/> diakses pada tanggal 30 Agustus 2018.
- [10] Farizal, A. 2009. *Kajian Produktivitas Lahan Sawah Terhadap Kesejahteraan Petani Di Bulupayung Kecamatan Patimunan Kabupaten Cilacap*. Purwokerto:FKIP UMP.
- [11] Goldfrank, Lewis R., Flomenbaum, N.E., Nelson, L.S., Lewin, N.A., Howland, M.A., Hoffman, R.S. 2011. *Toxicologic Emergencies Ninth Edition*. The McGraw-Hill Companies, Inc.
- [12] Gorunescu, F. (2011). *Data Mining: Concepts, Models and Techniques*. Berlin: Springer-Verlag.
- [13] Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques (3rd ed.)*. San Francisco: Morgan Kaufmann Publishers Inc.
- [14] Hofmann, Markus., Klinkenberg, Ralf. 2014. *Rapid Miner Data Mining Use Cases And*

- Business Analytics Applications*. Chapman And Hall/CRC Data Mining and Knowledge Discovery Series. United States : CRC Press.
- [15] Indrajani. 2011. *Perancangan Basis Data dalam All in 1*. Jakarta : PT. Elex Media Komputindo.
- [16] Jananto, A. (2013). Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa. *Teknologi Informasi*, 18(1), 9–16.
- [17] Larose, Daniel.T., Larose, Chantal.D. 2014. *Discovering Knowledge In Data : An Introduction to Data Mining Second Edition*. Canada : John Wiley & Sons, Inc.
- [18] Laroussi, Hesham Mohammed Ma Al. 2015. *Implementasi Algoritma Naive Bayes Sebagai Proses Seleksi Penerima Beasiswa Libyan Embassy Berbasis Web*. Skripsi. Fakultas Sains Dan Teknologi. Universitas Islam Negeri Maulana Malik Ibrahim. Malang.