

Penerapan *Machine Learning* dengan Aplikasi Orange Data Mining Untuk Menentukan Jenis Buah Mangga

Heri Suroyo

Program Studi Teknik Informatika Universitas Bina Darma, Palembang, Indonesia

Email : herisuroyo@binadarma.ac.id

Abstrak

Penggunaan penambangan data kini semakin banyak diterapkan pada berbagai bidang. Pertanian termasuk bidang yang sangat cocok menerapkan teknik data mining untuk mengolah maupun menganalisis data-data pertanian. Beberapa teknik data mining seperti K-Means, Clustering cukup populer dan bisa dimanfaatkan. Tujuan dari riset ini adalah menerapkan teknik image analisis dengan menggunakan tools aplikasi Orange Data Mining untuk membantu petani mengetahui jenis buah mangga. Pada riset ini teknik Image clustering digunakan untuk menentukan jenis buah mangga dengan input berupa file image buah mangga. Saat ini setidaknya ada 5-7 jenis buah mangga yang bisa dibudidayakan di Indonesia. Aplikasi orange data mining adalah sebuah tools yang bisa digunakan untuk membantu dalam membedakan jenis buah mangga diperlukan agar petani bisa memilih dengan tepat jenis buah mangga yang akan dibudidayakan. Data mining dengan teknik image clustering bisa dimanfaatkan dalam hal ini. Penggunaan Orange Data Mining karena tools ini berbasis GUI sehingga bisa dimanfaatkan oleh orang awam sekalipun. Riset ini menghasilkan gambaran yang sederhana tentang proses penggunaan image analitik yang bisa dipergunakan pada bidang pertanian.

Kata Kunci: Machine Learning, Data Mining, Image Analytic, Metode clustering, Aplikasi Orange.

1. PENDAHULUAN

Indonesia merupakan negara kepulauan sekaligus negara maritim beriklim tropis yang dilalui jalur pegunungan api serta memiliki flora dan fauna yang beranekaragam sehingga menjadikan Indonesia memiliki sumber daya alam berlimpah ditambah penduduknya yang mencapai 258,7 juta jiwa pada tahun 2016. Itu semua merupakan modal produksi untuk semua sektor, khususnya sektor pertanian yang sangat cocok dengan keadaan alam Indonesia^[1].

Saat ini 82% dari total jumlah desa di Indonesia masih mengandalkan sektor pertanian^[2]. Meningkatkan literasi petani dalam menggunakan teknik data mining untuk menganalisis data perlu dipertimbangkan. Sebab dengan semakin banyak tersedianya teknologi informasi yang bisa diterapkan pada bidang pertanian diharapkan bisa berpengaruh pada produktifitas petani. Sistem analisis data baik dengan cara tradisional maupun non tradisional di bidang pertanian adalah murni semua menggunakan statistik. Penambangan data adalah teknik analisis data modern. Ini memiliki jangkauan luas aplikasi di bidang pertanian. Aplikasi teknik penambangan data dalam area pertanian dan area terkaitnya dipelajari. Teknik penambangan data yang berbeda telah digunakan di lapangan ini. Padahal, ada banyak teknik yang tersedia dalam penambangan data, beberapa metodologi seperti pendekatan K-means, tetangga K-terdekat, Bi-clustering yang populer saat ini tergantung pada sifat data^[4].

Dalam upaya budidaya tanaman buah di Indonesia salah satu masalah yang sering dihadapi adalah bagaimana mengidentifikasi jenis buah yang akan dibudidayakan. Hal ini disebabkan oleh belum adanya cara yang bisa membantu proses pemilihan buah secara cepat dan tepat. Oleh karena itu diperlukan metode yang tepat untuk mengelompokkan buah secara otomatis dengan tingkat ketepatan yang tinggi dan waktu yang singkat.

Dari uraian diatas, rumusan masalah dalam penelitian ini adalah Bagaimana Penerapan *Machine Learning* untuk menentukan jenis buah mangga menggunakan aplikasi Orange. Sementara tujuan penelitian adalah sebagai berikut ; 1) Menerapkan metode *image clustering* untuk menentukan jenis buah mangga, 2) Melakukan Analisis data mining pada image-image buah mangga dengan menggunakan aplikasi *Orange*.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan metode tindakan (*action research*) yaitu penelitian baik kualitatif maupun kuantitatif. Penelitian tindakan adalah cara melakukan masalah pada saat yang bersamaan. Penelitian tindakan ini merupakan metode yang didasarkan pada tindakan terhadap objek riset yang seringkali diselenggarakan pada suatu latar yang luas, seperti di rumah sakit, pabrik, sekolah, dan lain sebagainya^[3].

Jaganathan (2014) menulis tentang Studi Teknik Penambangan Data untuk Pertanian. Dalam studi ini disajikan beberapa teknik penambangan data yang umum digunakan di bidang pertanian. Beberapa teknik-teknik ini, seperti k-means, tetangga terdekat k dan bi-clustering dibahas dan aplikasi di pertanian untuk masing-masing teknik ini disajikan^[4]. Riska (2015) telah meneliti tentang Klasifikasi Jenis Tanaman Mangga Gadung dan Mangga Madu Berdasarkan Tulang Daun. Dalam penelitian tersebut Fitur tulang daun mangga dijadikan acuan untuk membedakan kedua jenis tanaman tersebut, karena setiap jenis tanaman memiliki sudut tulang daun yang berbeda. Adanya keberagaman tekstur tulang daun, penelitian ini mengusulkan dua metode pada proses segmentasi. Salah satu metodenya dengan dipilih untuk mendapatkan tekstur tulang daun yang terbaik. Sehingga, dapat dipisahkan bagian tulang daun sebagai *foreground* dan daging daun sebagai *background*. Sementara metode lainnya digunakan pada proses ekstraksi fitur untuk menentukan pola tulang daun. Fitur yang dihasilkan dihitung dengan *entropy* yang kemudian diklasifikasikan menggunakan SVM. Pengujian menggunakan *10-fold cross validation* dengan menunjukkan akurasi 78,5 %^[8].

Sujatha, (2017) meneliti tentang *Smart Farming* menggunakan K-means Clustering dan SVM Classifier dalam Pengolahan Gambar. Dalam makalah ini, tujuan utamanya adalah untuk membangun sistem yang cepat, andal, efektif, dan bebas kesalahan untuk mendeteksi bagian buah yang terkena dampak penyakit dan membantu pertanian untuk meningkatkan kualitas dan hasil yang lebih baik. Hal ini merupakan upaya menghemat waktu dan meningkatkan efisiensi industri berbasis buah di mana buah-buahan dimanfaatkan dalam jumlah besar untuk membuat berbagai produk. Pendekatan ini dapat digunakan dalam industri impor dan ekspor^[9].

Penambangan data adalah proses penemuan yang sebelumnya tidak diketahui dan berpotensi pola menarik dalam dataset besar. Informasi yang ditambang digunakan untuk mewakili sebagai model untuk prediksi atau klasifikasi. Kumpulan data dari domain pertanian tampak jauh lebih banyak kompleks dari dataset yang digunakan secara tradisional dalam pembelajaran mesin. Data mining terutama dikategorikan sebagai penambangan data deskriptif dan prediktif. Namun di bidang pertanian, terutama penambangan data prediktif bekas. Ada dua teknik utama yaitu klasifikasi dan pengelompokan. Teknik penambangan data telah mengedepankan pedoman untuk membuat rekomendasi untuk manajemen tanaman endemik lokal^[4].

Data mining adalah ekstraksi informasi atau pola yang penting atau menarik dari data yang ada di database. Dalam jurnal ilmiah, data mining juga dikenal dengan nama *Knowledge Discovery in Database (KDD)* ^[7]. Penjelasan dari setiap tahapannya adalah sebagai berikut:

- 1) *Data cleaning* (untuk menghilangkan noise data yang tidak konsisten).
- 2) *Data integration* (sumber data yang terpecah dapat disatukan).
- 3) *Data selection* (data yang relevan dengan tugas analisis dikembalikan ke dalam database).
- 4) *Data transformation* (data berubah atau bersatu menjadi bentuk yang tepat untuk menambang dengan ringkasan performa atau operasi agresif)
- 5) *Data mining* (proses esensial dimana metode tepat digunakan untuk mengekstrak pola data).
- 6) *Pattern evolution* (untuk mengidentifikasi pola yang benar-benar menarik yang mewakili pengetahuan berdasarkan atas beberapa tindakan yang menarik).
- 7) *Knowledge presentation* (gambaran teknik visualisasi dan pengetahuan digunakan untuk memberikan pengetahuan yang telah ditambang kepada *user*).

Analisis gambar (*image analytic*) adalah ekstraksi informasi yang bermakna dari gambar; terutama dari gambar digital dengan menggunakan teknik pemrosesan gambar digital. Tugas-tugas analisis gambar dapat sesederhana seperti membaca kode bar atau secanggih mengidentifikasi seseorang dari wajahnya. Komputer sangat diperlukan untuk analisis data dalam jumlah besar, untuk tugas-tugas yang memerlukan perhitungan yang rumit, atau untuk ekstraksi informasi kuantitatif.

Analisis gambar merupakan alat yang berharga untuk menggambarkan unit manajemen untuk aplikasi pertanian spesifik misalnya SiteSpecific. Permukaan tanah seringkali sangat bervariasi dalam hal hubungannya dengan potensi produksi pertanian. Karena sifat kompleks bentang alam ini, hanya peta skala besar yang dapat memberikan tingkat detail yang diperlukan untuk pertanian khusus^[5]. Teknik analisis citra terdiri dari 5 komponen dasar: (1) pengambilan gambar, (2) georeferensi, (3) peningkatan, (4) analisis, dan (5) klasifikasi. Pengolahan gambar dapat digunakan dalam aplikasi pertanian untuk tujuan berikut^[6]: 1) Untuk mendeteksi daun yang sakit, batang, buah, 2) Untuk mengukur area yang terkena penyakit, 3) Untuk menemukan bentuk area yang terkena dampak, 4) Untuk menentukan warna area yang terkena, 5) Untuk menentukan ukuran & bentuk buah.

Clustering adalah proses pengelompokan data ke beberapa kelompok sehingga data dalam satu cluster memiliki kemiripan yang minimum. *Clustering* merupakan metode segmentasi data yang telah diimplementasikan dalam berbagai bidang, seperti prediksi dan analisis masalah bisnis segmentasi pasar, marketing, zonasi wilayah hingga identifikasi objek dalam bidang komputer visual dan pengolahan citra. Ada 4 (empat) metode *clustering* yaitu ^[10]: 1) *Hierarchical Agglomerative Clustering*, 2) *Fuzzy Clustering*, 3) *Hierarchical Divisive Clustering*, 4) *Kohonen Self-Organizing Feature Maps*,

Algoritma K-Means adalah salah satu metode *clustering* non hirarki dengan cara mempartisi data yang ada ke dalam satu atau lebih *cluster*. Metode ini mempartisi data ke dalam *cluster*, sehingga data yang memiliki karakteristik yang sama akan berada dalam 1 cluster. Sedangkan Segmentasi gambar berarti mempartisi gambar ke dalam berbagai kelompok atau kelompok dengan fitur yang sama atau memiliki beberapa kesamaan. K-means merupakan metode pengelompokan partisi gambar kedalam kelompok sedemikian rupa sehingga setidaknya satu bagian dari *cluster* mengandung gambar dengan area utama dari bagian yang digunakan.

3. ANALISA DAN PEMBAHASAN

Untuk melakukan analisa citra dengan menggunakan Orange Data mining kita harus memasang atau menambahkan *add-ons Image Analytics*. Berikut koleksi *image* berbagai jenis mangga yang tumbuh di Indonesia.



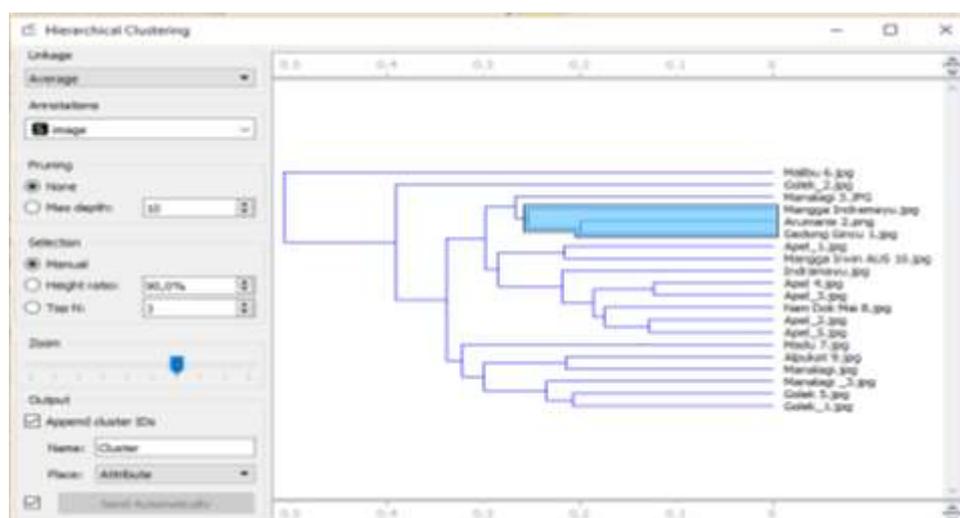
Gambar 1. Koleksi citra buah mangga sebagai data *trainer machine learning*.

Setelah kita menyiapkan koleksi citra yang akan digunakan sebagai *trainer data mining* pada *machine learning*. Selanjutnya dengan *image deskriptor* kita bisa mendapatkan informasi tentang nama file, ukuran file, serta panjang dan lebar citra yang diperlukan untuk *machine learning*. Kemudian dengan *widget Image Embedding* kita bisa mendapatkan fitur ruang multidimensiaonal dari setiap citra. *Image embedding* bekerja dengan mengirimkan data citra ke *server* untuk menghitung profil citra yang sebagai data awal untuk *machine learning*. Data *image deskriptor* dibutuhkan untuk mendiskripsikan data citra. Data citra mentah selanjutnya di transformasi ke bentuk vektor menggunakan *deep neural network*. Berikut tabel data *image deskriptor*.

	image name	image # 2017-2018/Prop image	size	width	height	n0	n1	n2	n3	n4	n5	n6	n7	n8	n9	n10	n11
1	Alpukat 9	Alpukat 9.jpg	50209	700	465	0.405	0.816	0.248	0.234	0.269	0.248	0.218	0.135	0.013	0.456	0.432	0.047
2	Apel 4	Apel 4.jpg	213865	1024	768	0.148	0.303	0.604	0.096	0.153	0.288	0.151	0.134	0.062	0.742	0.065	0.436
3	Apel_1	Apel_1.jpg	4890	225	225	0.171	0.069	0.122	0.055	0.102	0.507	0.166	0.032	0.053	1.195	0.555	0.631
4	Apel_2	Apel_2.jpg	6734	259	194	0.128	0.489	0.182	0.044	0.323	0.509	0.177	0.084	0.099	0.377	0.076	0.293
5	Apel_3	Apel_3.jpg	11022	239	211	0.081	0.454	0.214	0.317	0.212	0.333	0.197	0.094	0.150	1.059	0.103	0.573
6	Apel_5	Apel_5.jpg	11804	225	225	0.167	0.810	0.082	0.016	0.143	0.648	0.060	0.008	0.043	0.301	0.192	0.309
7	Arumanis 2	Arumanis 2.png	618475	872	612	0.137	0.041	0.146	0.017	0.006	0.165	0.118	0.024	0.049	0.841	0.179	0.388
8	Gedong Gincu 1	Gedong Gincu ...	222662	1270	942	0.212	0.482	0.122	0.116	0.012	0.327	0.224	0.149	0.212	0.787	0.301	0.049
9	Golek 5	Golek 5.jpg	209822	1000	801	0.107	0.196	0.194	0.138	0.006	0.184	0.271	0.208	0.005	0.543	0.486	0.089
10	Golek_1	Golek_1.jpg	9558	194	259	0.188	0.068	0.445	0.152	0.238	0.961	0.101	0.417	0.023	0.028	0.235	0.196
11	Golek_2	Golek_2.jpg	11428	195	258	0.384	0.480	0.086	0.050	0.260	0.309	0.092	0.037	0.004	0.198	0.299	0.105
12	Indramayu	Indramayu.jpg	79776	640	426	0.234	0.228	0.187	0.206	0.154	0.516	0.061	0.001	0.148	0.824	0.061	0.290
13	Madu 7	Madu 7.jpg	317497	1280	800	0.209	0.098	0.412	0.271	0.059	0.328	0.157	0.097	0.198	0.480	0.457	0.011
14	Malibu 6	Malibu 6.jpg	245100	946	523	0.600	0.071	0.534	0.139	0.226	0.605	0.098	0.749	0.019	0.804	0.299	0.289
15	Manalagi 3	Manalagi 3.JPG	1362401	2848	2136	0.138	0.267	0.117	0.109	0.172	0.002	0.002	0.089	0.042	1.074	0.106	0.202
16	Manalagi_3	Manalagi_3.jpg	22214	400	300	0.149	0.583	0.077	0.014	0.001	0.151	0.068	0.078	0.001	0.723	0.305	0.089
17	Manalagi	Manalagi.jpg	9409	256	197	0.344	0.276	0.067	0.025	0.309	0.158	0.045	0.342	0.019	0.257	0.274	0.069
18	Mangga Indra...	Mangga Indra...	17798	320	276	0.271	0.153	0.242	0.176	0.046	0.288	0.103	0.090	0.006	0.738	0.400	0.295
19	Mangga Irwin ...	Mangga Irwin ...	115084	400	300	0.240	0.194	0.425	0.167	0.186	0.320	0.082	0.054	0.026	1.607	0.140	0.588
20	Nam Dok Mai 8	Nam Dok Mai 8...	69262	572	850	0.462	0.312	0.056	0.175	0.180	0.318	0.122	0.174	0.054	0.278	0.094	0.321

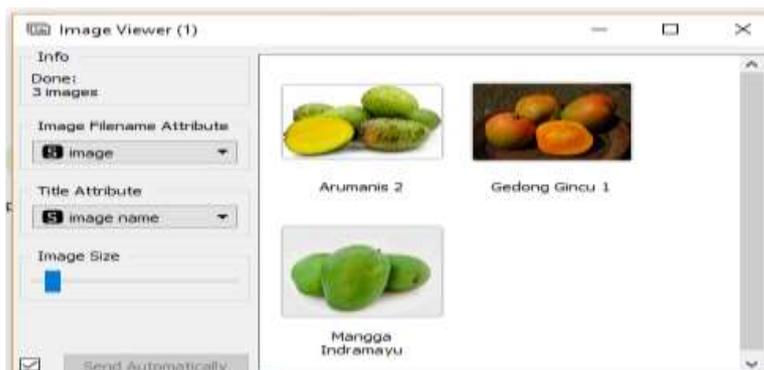
Gambar 2. Image Deskriptor berisi profil citra untuk *Mechine Learning*.

Selanjutnya dari *Image embedding* kita bisa menghitung simialiritas dengan menggunakan metode *cosine*. Selanjutnya dengan *Hierarchical Clustering* kita bisa mendapatkan pengelompokan image jenis buah mangga. Berikut gambar *Hierarchical Clustering*.



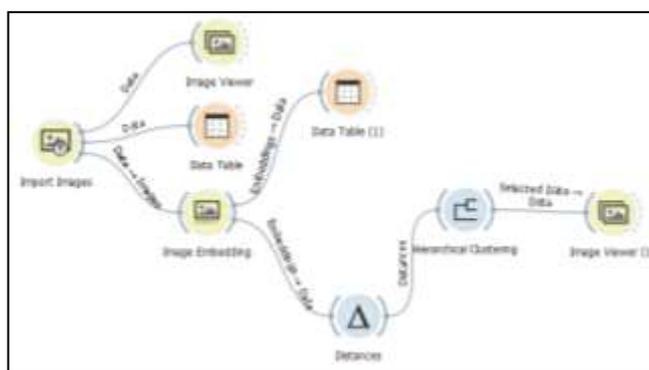
Gambar 3. Hierarchical Clustering Citra jenis Mangga

Selanjutnya dengan untuk melihat hasil pengelompokan citra dengan menggunakan Image Viewer. Berikut contoh hasil pengelompokan citra mangga dengan teknik image clustering.



Gambar 4. Contoh Pengelompokan citra mangga dengan teknik image clustering.

Berikut Gambar design widget Orange Data Mining untuk Penerapan Image Analytic untuk menentukan jenis buah mangga.



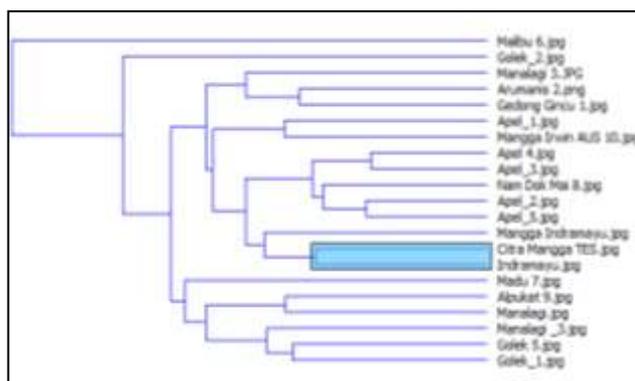
Gambar 5. Widget Orange data mining menentukan jenis buah mangga

Untuk melakukan pengujian maka dibutuhkan sampel file citra jenis mangga yang bisa diperoleh dengan memotret beberapa buah mangga di sekitar kita. Pada contoh ini akan digunakan citra gambar mangga sebagai berikut :



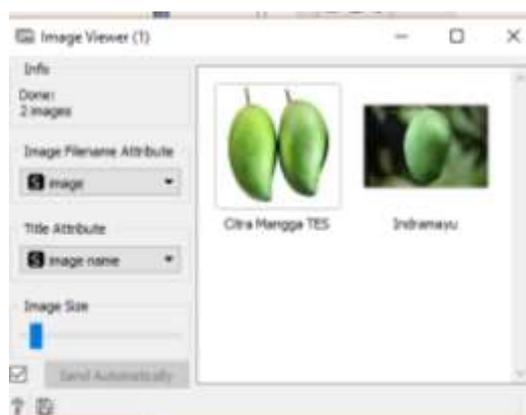
Gambar 6. Citra mangga untuk pengujian clustering

Setelah citra tes diatas dimasukan pada folder koleksi citra buah mangga kemudian melalui widget Import Image dilakukan reload untuk mengubah koleksi image pada machine learning. Selanjutnya melakukan proses menghitung ulang image deskriptor dengan widget Image Embedding dan hasilnya nampak pada Hierarchical Clustering yang berubah seperti pada gambar berikut ini.



Gambar 7. Hierarchical Clustering Hasil pengujian image Citra Buah Mangga Tes.

Dari pengujian diatas didapatkan Citra Mangga TES termasuk kelompok Mangga Indramayu dan jika di klik tanda blok pada Hierarchical Clustering nampak *Citra Mangga Tes.Jpg* berada satu kelompok dengan citra *Indramayu.jpg* seperti nampak pada gambar berikut.



Gambar 8. Hasil Pengelompokan Citra Mangga Tes.Jpg

4. KESIMPULAN

Riset ini menghasilkan gambaran yang sederhana tentang proses penggunaan *image analitic* untuk menentukan jenis buah mangga. Kesimpulan yang bisa disajikan pada pentian ini adalah sebagai berikut :

1. Teknik *image analitic* bisa digunakan untuk membantu dalam proses mengidentifikasi jenis buah berdasar citra dengan *machine learning*.
2. Aplikasi Orange Data Mining memiliki fitur yang cukup dan mudah dipahami untuk melakukan proses *image analitic* dengan metode *image clustering*.

REFERENCES

- [1]. Fajar, Muhammad, 2018, *Kinerja Sektor Pertanian Indonesia Periode 2012-2016*, <https://www.researchgate.net/publication/322298271> Working Paper - January 2018, diakses tanggal 04/07/2019 jam 09.20.
- [2]. Fajar dan Alfiyani, 2018, *Kinerja Sektor Pertanian Indonesia Periode 2012 – 2016*, : <https://www.researchgate.net/publication/322298271>, Working Paper - January 2018, diakses tanggal 04/07/2019 jam 09.20.
- [3]. Hasibuan, Zainal (2007), *Metodologi Penelitian Pada Bidang Ilmu Komputer Dan Teknologi Informasi Konsep, Teknik, Dan Aplikasi*, Fakultas ILKOM Universitas Indonesia.
- [4]. Jaganathan, et.al ,(2014) *A Study of Data Mining Techniques to Agriculture*, *IJRIT International Journal of Research in Information Technology*, Volume 2, Issue 4, April 2014, Pg: 306- 313.
- [5]. McCann, 1996, The Development of Management Units for Site-Specific Farming, http://www.usask.ca/soilsncrops/conference-proceedings/previous_years/Files/96/1996docs/365.PDF, diakses : 19/07/2018
- [6]. Priyanka, Sony et.al, 2015, Studi Pengolahan Citra di Pertanian untuk Mendeteksi Penyakit Tanaman, *International Journal of Computer Science and Mobile Computing*, IJCSMC, Vol. 4, Issue. 7, July 2015, pg.581 – 587
- [7]. Rerung , 2018, *Penerapan Data Mining dengan Memanfaatkan Metode Association Rule untuk Promosi Produk* , Jurnal Teknologi Rekayasa, Politeknik Sukabumi, JTERA, Vol. 3, No. 1, Juni 2018
- [8]. Riska , et.al, 2015, *Klasifikasi Jenis Tanaman Mangga Gadung dan Mangga Madu Berdasarkan Tulang Daun*, Jurnal Buana Informatika, Volume 6, Nomor 1, Januari 2015: 41-50.
- [9]. Sujatha (2017), Smart Farming using K-means Clustering and SVM Classifier in Image Processing, *International Journal of Science, Engineering and Technology Research (IJSETR)* Volume 6, Issue 11, November 2017, ISSN: 2278 -7798.
- [10]. Tiwari, 2011, *Application of Cluster Analysis in Agriculture– A Review Article*, *International Journal of Computer Applications* (0975 – 8887) Volume 36– No.4, December 2011.