

ISSN: 2460-8092, E-ISSN: 2548-6551



SYSTEMIC

INFORMATION SYSTEM AND INFORMATICS JOURNAL

Diterbitkan oleh :
Program Studi Sistem Informasi
Fakultas Sains dan Teknologi
Universitas Islam Negeri (UIN) Sunan Ampel
Surabaya

Volume 6 Nomor 2 - Desember 2020

SYSTEMIC

Information System and Informatics Journal

Volume 6 Nomor 2, Desember 2020

Pimpinan Redaksi

Indri Sudanawati Rozas (Universitas Islam Negeri Sunan Ampel Surabaya)

Mitra Bestari

Achmad Solichin (Universitas Budi Luhur)
Aeri Rachmad (Universitas Trunojoyo Madura)
Agus Hermanto (Universitas 17 Agustus 1945 Surabaya)
Ahmad Afif Supianto (Universitas Brawijaya Malang)
Ahmad Habib (Universitas 17 Agustus 1945 Surabaya)
Anik Vega Vitianingsih (Universitas Dr. Soetomo Surabaya)
Devi Karolita (Universitas Palangkaraya)
Dina Fitria Murad (Universitas Bina Nusantara Jakarta)
Dwi Puspitasari (Politeknik Negeri Malang)
Elly Antika (Politeknik Negeri Jember)
Hanung Prasetyo (Universitas Telkom)
Hartarto Junaedi (Institut Sains dan Teknologi Terpadu Surabaya)
Himawan (Sekolah Tinggi Teknologi Informasi NIIT)
Irwan Alnarus Kautsar (Universitas Muhammadiyah Sidoarjo)
Moh Noor Al-Azam (Universitas Narotama Surabaya)
Novi Prastiti (Universitas Trunojoyo)
Ridha Sefina Samosir (Institut Teknologi dan Bisnis Kalbis)
Robbi Rahim (Sekolah Tinggi Ilmu Manajemen Sukma)
Sucipto (Universitas Nusantara PGRI Kediri)
Sukirman (Universitas Muhammadiyah Surakarta)
Titin Agustin Nengsih (UIN Sulthan Thaha Saifuddin Jambi)
Titus Kristanto (Institut Teknologi Telkom Surabaya)
Uky Yudatama (Universitas Muhammadiyah Magelang)

Editor

Khalid (Universitas Islam Negeri Sunan Ampel Surabaya)
Dwi Rolliawati (Universitas Islam Negeri Sunan Ampel Surabaya)
Noor Wahyudi (Universitas Islam Negeri Sunan Ampel Surabaya)
M. Andik Izzudin (Universitas Islam Negeri Sunan Ampel Surabaya)
Mujib Ridwan (Universitas Islam Negeri Sunan Ampel Surabaya)
Nita Yalina (Universitas Islam Negeri Sunan Ampel Surabaya)
Ahmad Yusuf (Universitas Islam Negeri Sunan Ampel Surabaya)
Andhy Permadi (Universitas Islam Negeri Sunan Ampel Surabaya)

SYSTEMIC

Information System and Informatics Journal

Penerbit

Program Studi Sistem informasi
Fakultas Sains dan Teknologi
Universitas Islam Negeri Sunan Ampel Surabaya

Alamat Redaksi

Kampus Fakultas Sains dan teknologi
UIN Sunan Ampel Surabaya
Jl. Jend. A. Yani 117 Surabaya, 60237
E-Mail. systemic@uinsby.ac.id
Telp. (031) 8410298, Fax (031) 8413300

SYSTEMIC merupakan jurnal yang diterbitkan oleh Program Studi Sistem Informasi Fakultas Sains dan Teknologi, UIN Sunan Ampel Surabaya. **SYSTEMIC** diterbitkan 2 (dua) kali dalam satu tahun pada bulan Agustus dan Desember. Artikel yang dimuat di **SYSTEMIC** berisi pokok bahasan yang terkait dengan aspek pengembangan, kerangka teoritis, implementasi dan pengembangan sistem secara keseluruhan.

SYSTEMIC

Information System and Informatics Journal

Daftar Isi

1. Analisis Usabilitas Pada Website iLab Universitas Gunadarma Menggunakan Metode Heuristic Evaluation
Ahmad Apandi 1 - 7
2. Implementasi Perbandingan Algoritma Apriori Dan FP-Growth Untuk Mengetahui Pola Pembelian Konsumen Pada Produk Panel Di PT Surya Multi Perkasa Movinko
Diego Armando P Putra, Tresna M Fahrudin, Natalia Damastuti 8 - 13
3. Klasifikasi Multi Output pada Harga Smartphone Menggunakan Learning Vector Quantization (LVQ) dan Backpropagation (BP)
Dinita Rahmalia, Mohammad Syaiful Pradana, Teguh Herlambang 14-19
4. Pengenalan Karakter Huruf Braille dengan Metode Convolutional Neural Network
Muhammad Fahmi Herlambang, Asep Nana Hermana, Kurnia Ramadhan Putra 20-26
5. Rancang Bangun Kontrol Keasaman pH Tanaman dalam Sistem Hidroponik menggunakan Kontrol PID Berbasis Android
Ika Indriani Retna Wardani, Kunto Eko Susilo 27 - 32
6. Pemanfaatan Algoritma FP-Growth Untuk Menentukan Strategi Penjualan Pada Kedai Kopi Teras Garden
Adrian Marvel Ugrasena, Achmad Zakki Falani 38 – 43
7. Pemanfaatan Image Mining Untuk Klasifikasi Dan Prediksi Kematangan Tomat Menggunakan Metode Jaringan Saraf Tiruan Backpropagation
Firdaus, Nori Sahrin 38 – 46
8. Klasifikasi Text Judul Buku Perpustakaan Untuk Menentukan Kategori Buku Menggunakan K-Nearest Neighbor
Muhamad Kadafi 47 - 53

SYSTEMIC: Information System and Informatics Journal

ISSN: 2460-8092, 2548-6551 (e)

Vol 6 No 2 – Desember 2020

Analisis Usabilitas Pada Website iLab Universitas Gunadarma Menggunakan Metode Heuristic Evaluation**Ahmad Apandi**

Universitas Gunadarma

apandiahmad007@gmail.com**Kata Kunci***Heuristic Evaluation, iLab, User Interface, Usabilitas, Website***Abstrak**

Laboratorium mandiri (iLab) Universitas Gunadarma merupakan salah satu lab yang memberikan kemudahan dalam memberikan informasi melalui website. Kemudahan dalam mengakses website tidak lepas dari usabilitas website tersebut. Usabilitas yang baik sangat berpengaruh bagi pengguna, semakin baik usabilitasnya semakin mudah pengguna menggunakan website tersebut. Agar suatu website dapat dilihat oleh pengguna dengan waktu yang cukup lama, maka usabilitas adalah syarat penting yang harus dipenuhi. Dalam penelitian ini metode yang digunakan adalah heuristic evaluation. Heuristic adalah metode evaluasi yang digunakan untuk menemukan kesalahan pada desain antar muka. Berdasarkan pengukuran website iLab menggunakan metode evaluasi heuristic oleh Nielsen didapatkan bahwa 10 aspek yang diteliti mendapatkan nilai satu dan dua, yang berarti website memiliki beberapa kekurangan yang tidak menjadikan sebuah masalah atau dengan kata lain tidak dipermasalahkan dan tidak mengganggu pengguna saat mengakses website. Nilai severity rating tertinggi yaitu sebesar 2.00 terdapat pada aspek membantu pengguna untuk mengenali, mendiagnosa, dan memulihkan dari kesalahan.

Keywords*Heuristic Evaluation, iLab, User Interface, Usability, Website***Abstract**

Gunadarma University's independent laboratory (iLab) is one of the laboratories that provides convenience in providing information through the website. The ease of accessing the website cannot be separated from the website's usability. Good usability is very influential for website users and must be met. In this study, the method used is heuristic evaluation. Heuristic is an evaluation method used to find errors in interface design. Based on the measurement of the iLab website using the heuristic evaluation method by Nielsen, it was found that ten aspects studied got one and two values, which means that the iLab website has several deficiencies that do not cause a problem or in other words are not a problem and do not interfere with users when accessing the iLab website. The highest severity rating of 2.00 is in the aspect of helping user recognize, diagnose, and recover from errors.

1. Pendahuluan

Perguruan tinggi memiliki layanan administrasi dan layanan akademik yang meliputi layanan perkuliahan dan praktikum, kegiatan praktikum biasanya dilaksanakan di laboratorium. Sebagai salah satu layanan kegiatan akademik, laboratorium memiliki tanggung jawab memberikan layanan akademik yang baik. Salah satu cara dalam memberikan layanan akademik yang baik pada laboratorium adalah dengan memiliki sebuah *website* yang mampu membantu para pengguna dalam menemukan informasi terkait dengan kegiatan yang diselenggarakan oleh

laboratorium, seperti kegiatan praktikum dan penerimaan asisten baru. Salah satu laboratorium yang memanfaatkan *website* dalam kegiatannya adalah laboratorium mandiri (iLab). iLab merupakan lembaga di Universitas Gunadarma yang memiliki kewenangan terhadap penyelenggaraan praktikum, pemberian modul praktikum, serta penyampaian segala informasi yang berkaitan dengan kegiatan laboratorium.

Keberadaan iLab sangat penting bagi kegiatan perkuliahan, karena dari sinilah terjadi pemahaman mahasiswa terhadap teori yang telah diberikan oleh dosen di kelas. Selama ini keseluruhan kegiatan yang dilakukan sudah

banyak dijalankan melalui *website* seperti pembagian materi praktikum, jadwal praktikum, pemberian informasi-informasi terkait kegiatan praktikum, dan sebagainya.

Kemudahan dalam mengakses *website* tidak lepas dari usabilitas *website* tersebut. Usabilitas yang baik sangat berpengaruh bagi pengguna *website*. Semakin baik usabilitasnya semakin mudah pengguna menggunakan *website* tersebut. Agar suatu *website* dapat dilihat oleh pengguna dengan waktu yang cukup lama, maka usabilitas adalah syarat penting yang harus dipenuhi. Seiring dengan berkembangnya, teknologi usabilitas dari sebuah *website* perlu dilakukan karena pengguna menuntut adanya kemudahan dan kenyamanan ketika berinteraksi dalam sebuah *website*. Situs yang memiliki usabilitas tinggi memiliki peluang yang sangat besar untuk sering dikunjungi oleh para pengguna [1].

Usabilitas sebagai ukuran kualitas *user experience* atau pengalaman pengguna ketika berinteraksi dengan antar muka sebuah sistem, baik *website*, teknologi bergerak, aplikasi perangkat lunak, maupun peralatan lain yang dioperasikan oleh pengguna. Usabilitas mengacu pada teknik atau metode yang dipakai untuk meningkatkan kemudahan pengguna. Usabilitas yang kurang baik dapat disebabkan oleh beberapa faktor dan permasalahan yang sering dialami oleh para pengguna. Jika sebuah situs atau *website* gagal dalam menunjukkan secara jelas apa yang dapat dilakukan oleh situs tersebut, pengguna cenderung akan langsung meninggalkan situs [2] [3]. Penting bagi sebuah *website* akademik memiliki usabilitas yang baik supaya pengguna merasa mudah dan nyaman ketika menggunakan *website* dalam mencari dan menemukan informasi yang diinginkan. Hal ini juga menjadi landasan yang bermanfaat bagi akademik untuk mengungkap bahwa usabilitas menjadi salah satu faktor penting bagi pengguna untuk dapat menyelesaikan tugasnya atau tidak, serta ada tidaknya kepuasan dari pengguna ketika menyelesaikan tugasnya melalui situs atau *website* yang dikunjungi [4].

Untuk melakukan evaluasi terhadap antarmuka suatu *website*, ada beberapa metode atau cara yang dapat digunakan, yaitu [5]: *formal usability*, *pluralistic walkthrough*, *standards inspection*, *consistency inspection*, *feature inspection*, dan *heuristic method*. Untuk penelitian ini metode yang digunakan adalah *heuristic evaluation*. *Heuristic* merupakan salah satu metode evaluasi yang digunakan untuk menemukan kesalahan pada desain antar muka [1] [3] [4]. Metode *heuristic* adalah metode yang tepat digunakan untuk menganalisis usabilitas suatu desain antarmuka pengguna atau *User Interface Design* (UID). Metode *heuristic* memiliki beberapa keunggulan dibanding dengan metode lain seperti proses evaluasi yang cepat, dan penggunaan biaya yang rendah. Namun metode *heuristic* ini memiliki kelemahan yaitu hanya digunakan untuk

menganalisis tampilan antarmuka pengguna [6].

Beberapa penelitian terkait mengenai *usability* dan *heuristic evaluation* diantaranya:

Yemima Monica Gasela, dkk, dalam penelitiannya, "Analisis User Interface terhadap Website berbasis E-Learning dengan Metode Heuristic Evaluation" dari penelitian ini didapat desain *interface* dan kemudahan penggunaan suatu *website* yang memiliki basis *e-learning* ditemukan cukup baik, namun masih terdapat beberapa hal yang perlu ditingkatkan seperti karakteristik dalam kemudahan pengguna, metode evaluasi *heuristic* yang terdiri dari sepuluh aturan yang berprinsip pada penilaian suatu desain antar muka [7].

Mulia Sulistiyono dalam penelitiannya, "Evaluasi Heuristic Sistem Informasi Pelaporan Kerusakan Laboratorium Universitas AMIKOM Yogyakarta" dari penelitian ini didapat bahwa metode *heuristic* dapat digunakan untuk mengevaluasi sistem informasi berbasis website dengan melibatkan *evaluator* untuk memberikan masukan kemudian dikategorikan dalam prinsip-prinsip *heuristic*, dari hasil pengukuran menunjukkan nilai akhir untuk *severity rating* sebesar 3.50 yang diperoleh dari rata-rata nilai secara keseluruhan [8].

Bella Aulia Mustikaningtyas, dkk, dalam penelitiannya, "Analisis Usability Pada Website Universitas Brawijaya dengan Heuristic Evaluation", penelitian ini menganalisis aspek *usability*, aspek ini dibutuhkan untuk permasalahan yang ada pada sistem dan supaya mencapai tujuan tertentu secara efektif, efisien, dan mencapai kepuasan *user*. Uji preferensi *user* juga dilakukan untuk mengetahui respon *user* umum terhadap website UB, dan didapatkan hasil berupa perbandingan hasil analisis antara *heuristic evaluation* dengan uji preferensi *user* [9].

Untuk menguji usabilitas menggunakan metode *heuristic* perlu adanya *evaluator*; metode ini pertama kali ditemukan oleh Jakob Nielsen dan Rolf Molich pada tahun 1990. Beberapa keunggulan dari metode *heuristic* ini diantaranya: Proses dalam melakukan evaluasi yang cepat, mudah dalam melakukan suatu proses evaluasi, dan hemat dalam penggunaan biaya atau *cost* [10] [11] [12]. Pengujian menggunakan metode *heuristic* ini didasarkan kepada tingkat keparahan (*severity rating*) terhadap usabilitas suatu antar muka. *Severity rating* merupakan nilai yang diberikan berdasarkan tingkat keparahan yang ditemukan dari suatu permasalahan.

Ada sepuluh komponen yang perlu dianalisis dalam menggunakan metode *heuristic* yaitu: 1. *Visibility of system status* atau visibilitas status sistem, 2. *Match between system and the real world* atau kecocokan antara sistem dengan dunia nyata, 3. *User control and freedom* atau kontrol dan kebebasan pengguna, 4. *Consistency and standards* atau konsistensi dan standarisasi, 5. *Error prevention* atau pencegahan kesalahan, 6.

Recognition rather than recall atau pemahaman daripada ingatan, 7. *Flexibility* atau fleksibilitas, 8. *Aesthetic and minimalist design* atau estetika dan desain minimalis, 9. *Help user recognize, diagnose, and recover from errors* atau membantu pengguna mengenali, mendiagnosa, dan memulihkan kembali dari kesalahan, 10. *Help and documentation* atau bantuan dan dokumentasi.

Berdasarkan latar belakang tersebut, rumusan masalah pada penelitian ini adalah bagaimana melakukan evaluasi desain antar muka *website* iLab menggunakan metode *heuristic*? Tujuan dari penelitian ini adalah untuk mengetahui apakah antarmuka *website* iLab sudah sesuai dengan yang dibutuhkan oleh pengguna.

2. Metode Penelitian

Metode yang digunakan dalam penelitian ini yaitu dengan melakukan studi literatur tentang Interaksi Manusia dan Komputer serta teori mengenai *heuristic evaluation* Jakob Nielsen. Metode tersebut digunakan sebagai panduan untuk melakukan evaluasi usability desain antar muka *website* iLab. Langkah selanjutnya adalah dengan melakukan penyebaran kuisisioner ke responden untuk menilai desain antarmuka pengguna melalui pertanyaan yang didasarkan dari metode *heuristic evaluation* Jakob Nielsen. Kemudian hasil jawaban responden tersebut dikumpulkan dan diproses untuk ditarik kesimpulan dan saran, dimana hasil tersebut merupakan salah satu bentuk upaya untuk pengembangan *website* iLab.

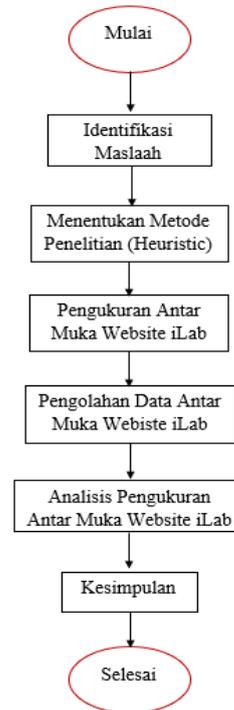


Gambar 1. Screenshoot *website* iLab [diambil tanggal 15 maret 2020]

Pada Gambar 1, dapat dilihat bahwa *website* iLab Universitas Gunadarma memiliki menu bar berupa Home, iLab, Aktivasi, Download, dan Open Recruitment. Pada halaman Home terdapat ILAB NEWS yang berisi berita terbaru seputar kegiatan lab, pada bagian ILAB PROCEDURE berisi mengenai aktivasi mahasiswa baru, tata tertib, tahapan praktikum, dan praktikum pengganti, pada bagian ILAB FACILITIES berisi mengenai cek jadwal praktikum, download modul, e-learning ilab, dan wifi kampus, pada bagian ILAB MEDIA berisi video-video mengenai layanan praktikum.

Berbagai layanan yang disajikan pada *website* tersebut untuk memudahkan mahasiswa untuk mendapatkan informasi seputar kegiatan iLab perlu dilakukan evaluasi usability untuk mengetahui apakah *website* tersebut sudah memenuhi kriteria atau belum berdasarkan metode yang digunakan.

Metode yang digunakan dalam penelitian ini lebih jelasnya dapat dilihat pada gambar 2.



Gambar 2. Tahapan metode penelitian

Berdasarkan Gambar 2, penelitian dimulai dari mengidentifikasi masalah, masalah yang muncul pada layanan *website* adalah kenyamanan dalam mengakses *website*, sehingga perlu dilakukan evaluasi apakah *website* sudah memberikan kenyamanan kepada mayoritas pengguna atau belum.

Kemudian menentukan metode penelitian berdasarkan referensi dari sumber pustaka, metode yang digunakan adalah *heuristic evaluation*. Metode ini merupakan panduan, aturan atau prinsip umum dalam menentukan keputusan yang digunakan. Pada pendekatan ini sekumpulan kriteria *usability* atau *heuristic* dapat diidentifikasi [13].

Tahap selanjutnya adalah pengukuran antar muka *website* dengan menggunakan 10 kriteria yang mengacu pada metode *heuristic* Nielsen. Selanjutnya dilakukan pengolahan data untuk mendapatkan nilai *severity rating* usabilitas pada *website*.

Setelah melakukan pengolahan data, tahap berikutnya melakukan analisis terhadap objek sesuai dengan prinsip metode yang digunakan yaitu *heuristic evaluation*, tujuannya adalah untuk mengevaluasi dan memperbaiki perancangan secara efektif. Evaluasi *heuristic* sangat baik digunakan sebagai teknik evaluasi desain [14, 15].

Untuk melakukan perhitungan menggunakan *heuristic evaluation* Jakob Nielsen dapat digunakan rumus (1) berikut [6]:

$$\Sigma A = (0 \cdot X) + (1 \cdot X) + (2 \cdot X) + (3 \cdot X) + (4 \cdot X) \quad (1)$$

Keterangan :

ΣA : merupakan jumlah dari skor *rating* dari sub aspek usabilitas dalam setiap aspek usabilitas (A1, A2, ..., A10)

Nilai 0-4: merupakan nilai *severity rating* atau peringkat keparahan.

X: merupakan poin usabilitas yang memiliki nilai 1/0 (1: Ya, 0: Tidak)

Selanjutnya, untuk mendapatkan nilai *severity rating* dari setiap aspek usabilitas dapat menggunakan rumus (2) berikut:

$$S = (\Sigma A) / n \quad (2)$$

Keterangan:

S: merupakan hasil nilai dari *severity rating* pada satu aspek usabilitas.

ΣA : merupakan jumlah dari skor *rating* dari sub aspek usabilitas dalam setiap aspek usabilitas (A1, A2, ..., A10).

n: merupakan banyaknya dari sub aspek usabilitas pada setiap aspek usabilitas.

Berikut merupakan penjelasan dari setiap aspek usabilitas yang dievaluasi:

- 1) Visibilitas status sistem
Aspek ini berisi mengenai navigasi *website*, keunikan tombol, dan notifikasi.
- 2) Kecocokan antara sistem dengan dunia nyata
Aspek ini berisi mengenai kejelasan dan keunikan dari simbol atau ikon, opsi menu, tata bahasa, penggunaan istilah yang sesuai pada *website*.
- 3) Kontrol dan kebebasan pengguna
Aspek ini berisi mengenai dialog konfirmasi ketika akan melakukan perintah dan kontrol sederhana yang terdapat pada *website*.
- 4) Konsistensi dan standarisasi
Aspek ini berisi mengenai konsistensi bahasa dan istilah yang digunakan, ikon, nama-nama menu, dan kontrol pada *website*.
- 5) Pencegahan kesalahan

Aspek ini berisi mengenai dialog atau notifikasi ketika terjadi kesalahan, *website* dapat mencegah kesalahan yang dilakukan oleh pengguna.

- 6) Pemahaman daripada ingatan
Aspek ini berisi mengenai penempatan tombol menu, peletakan elemen grafis yang akan muncul ketika dibutuhkan oleh pengguna.
- 7) Fleksibilitas
Aspek ini berisi mengenai kemudahan dalam mengakses informasi yang diinginkan, ketersediaan opsi lanjutan, dan pintasan ketika akan menjalankan perintah.
- 8) Estetika dan desain minimalis
Aspek ini berisi mengenai ikon yang mewakili perintah saling terkait, kejelasan informasi dari sebuah perintah, dan desain *website* yang tidak rumit.
- 9) Membantu pengguna untuk mengenali, mendiagnosa, dan memulihkan dari kesalahan
Aspek ini berisi mengenai ada tidaknya informasi ketika terjadi kesalahan, pengguna dapat memulihkan ketika terjadi kesalahan, serta terdapat saran ketika terjadi kesalahan.
- 10) Bantuan dan dokumentasi
Aspek ini berisi mengenai keakuratan bantuan dan dokumentasi, dokumentasi memungkinkan pengguna untuk memahami dan dapat melanjutkan langkah perintahnya dengan benar.

3. Hasil Dan Pembahasan

Pada penelitian ini, *severity rating* yang dipakai untuk menilai menggunakan skala angka 0 sampai 4. Angka tersebut menunjukkan tingkat keparahan yang terdapat pada *website*, semakin kecil angkanya semakin setuju fitur yang sudah ada dan tidak perlu menambahkan fitur baru. Sebaliknya, semakin tinggi angkanya, maka pentingnya fitur tersebut dan dibutuhkan oleh pengguna serta harus ada di dalam *website*.

Tabel 1. Klasifikasi *Severity Rating*

<i>Severity Rating</i>	Keterangan
0	Tidak Masalah
1	Masalah tidak penting, tidak apa-apa
2	Sedikit masalah, prioritas rendah
3	Masalah, prioritas sedang
4	Masalah sekali, prioritas tinggi

Penjelasan lebih lanjut mengenai skala *severity rating* Jakob Nielsen dapat dilihat pada Tabel 1, nilai *severity rating* memiliki skala 0 sampai 4, dimana nilai tersebut menunjukkan keparahan yang terdapat pada *website*, nilai 0 artinya tidak ada masalah usabilitas yang terdapat pada *website*, nilai 1 artinya masalah tidak penting atau tidak perlu diperbaiki kecuali ada waktu dalam pengerjaan proyek, nilai 2 artinya *website*

memiliki sedikit masalah dengan prioritas rendah, nilai 3 artinya *website* memiliki masalah dengan prioritas sedang, dan nilai 4 artinya *website* memiliki masalah sekali dengan prioritas tinggi.

Tabel 2 adalah aspek usabilitas dan sub usabilitas yang digunakan dalam menentukan *severiry rating* atau tingkat keparahan yang terdapat pada *website*.

Tabel 2. Aspek Usabilitas dan Sub Aspek Usabilitas

Aspek Usabilitas	Sub Aspek Usabilitas
Visibilitas status sistem	<ol style="list-style-type: none"> 1. Setiap halaman memiliki judul yang menggambarkan isi halaman 2. Terdapat respon yang membedakan secara visual ketika tindakan dari objek tertentu ditekan atau dipilih 3. Nama pada setiap menu sudah sesuai dengan konten 4. Tampilan menu pada navigasi sudah bisa menunjukkan perbedaan antara menu yang saat ini dipilih atau tidak
Kecocokan antara sistem dan dunia nyata	<ol style="list-style-type: none"> 1. Ikon atau simbol yang digunakan secara umum sudah biasa dikenal oleh pengguna 2. Nama menu sudah ditulis secara logis dan dapat dimengerti oleh pengguna 3. Penggunaan istilah yang sudah sesuai
Kontrol dan kebebasan pengguna	<ol style="list-style-type: none"> 1. Ada tombol bantuan ketika sistem tidak merespon apa pun 2. Pengguna memiliki keleluasaan dalam melakukan pencarian 3. Jika sistem memiliki menu halaman, pengguna dapat dengan mudah Kembali ke menu atau halaman sebelumnya
Konsistensi dan standarisasi	<ol style="list-style-type: none"> 1. Bahasa atau istilah yang digunakan pada sistem sudah konsisten 2. Penggunaan nama pada menu sudah sesuai 3. Ikon yang ditampilkan sudah sesuai 4. Tampilan web pada setiap halaman memiliki bentuk yang sama dan konsisten
Pencegah kesalahan	<ol style="list-style-type: none"> 1. Terdapat dialog atau notifikasi ketika terjadi kesalahan 2. Teks pada instruksi jelas dan tidak menyebabkan ambigu 3. Panduan navigasi di setiap halaman sudah baik
Pemahaman daripada ingatan	<ol style="list-style-type: none"> 1. Penempatan tombol menu dan elemen grafis yang dibutuhkan pengguna sudah sesuai 2. Terdapat pesan kesalahan yang bersifat teknik ketika web gagal mengakses halaman
Fleksibilitas	<ol style="list-style-type: none"> 1. Terdapat pintasan atau opsi

- lanjutan yang dibutuhkan oleh pengguna
2. Navigasi sudah ditempatkan dengan benar
3. Pengguna dapat memilih menu secara bebas

Estetika dan desain minimalis

1. Ikon yang mewakili perintah saling terkait
2. Pencarian menu mudah dikenali dan digunakan terutama untuk pemula
3. Tata letak menu yang biasa dan mudah diakses oleh pengguna
4. Pemilihan warna dan jenis huruf sudah sesuai

Membantu pengguna untuk mengenali, mendiagnosa, dan memulihkan dari kesalahan

1. Terdapat informasi yang muncul ketika terjadi kesalahan pada sistem
2. Penggunaan dapat memulihkan ketika terjadi kesalahan
3. Terdapat saran ketika terjadi kesalahan pada sistem

Bantuan dan dokumentasi

1. Penggunaan menu bantuan untuk pengguna sudah tepat
2. Fasilitas yang dapat menghubungi korespondensi atau pemilik situs web sudah sesuai

Setelah menentukan aspek dan sub aspek usabilitas seperti pada Tabel 2, maka dilakukan perhitungan nilai hasil evaluasi *heuristic* menggunakan Rumus 1 dan 2. Table 3 merupakan contoh perhitungan aspek usabilitas dengan satu responden. Pada Tabel 3, daftar nilai *severity rating* pada kolom C, D, E, F, G adalah nilai dari evaluasi *heuristic* yang dilakukan, kolom H merupakan kolom yang berisi jumlah *severity rating* dari sub aspek sesuai dengan Rumus 1 dan 2.

Table 3. Contoh Perhitungan Aspek Usabilitas

Aspek Usability	Sub Aspek Usability	Severity Rating (SR)					Jumlah SR Sub Aspek
		0	1	2	3	4	
A	B	C	D	E	F	G	H
1	1	0	1	0	0	0	1
	2	0	0	1	0	0	2
	3	0	1	0	0	0	1
	4	0	1	0	0	0	1
Jumlah SR Aspek (ΣA)							6
Hasil Nilai Severity Rating ($\Sigma A/n$)							1.25

Untuk melakukan evaluasi antar muka menggunakan *heuristic evaluation* menggunakan *severity rating*, maka perlu diadakan kuisioner. Kuisioner antarmuka *website* iLab ini melibatkan 50 orang responden, dengan rincian 32 laki-laki dan 18 perempuan, responden merupakan

mahasiswa aktif semester 6 dan 8 di Universitas Gunadarma yang mengakses *website* iLab.

Tabel 4. Hasil Pengolahan Nilai *Severity Rating* Responden Laki-laki

No	Aspek Usability	Severity Rating	
		A	B
1	Visibilitas status sistem	1.33	1
2	Kecocokan antara sistem dan dunia nyata	0.60	1
3	Kontrol dan kebebasan pengguna	1.89	2
4	Konsistensi dan standarisasi	1.26	1
5	Pencegah kesalahan	1.59	2
6	Pemahaman daripada ingatan	1.41	1
7	Fleksibilitas	0.88	1
8	Estetika dan desain minimalis	0.93	1
9	Membantu pengguna untuk mengenali, mendiagnosa, dan memulihkan dari kesalahan	1.93	2
10	Bantuan dan dokumentasi	1.81	2

Berdasarkan hasil kuesioner yang telah diikuti oleh responden tersebut selanjutnya dilakukan perhitungan nilai *severity rating* untuk masing-masing aspek yang ada. Dari hasil kuesioner dinyatakan bahwa 10 aspek dari metode *heuristic* oleh Nielsen pada responden laki-laki didapatkan nilai *severity rating* 1 (satu) dan 2 (dua). Nilai satu menandakan bahwa kesalahan masalah tidak penting, sedangkan nilai dua menandakan sedikit masalah namun prioritas rendah, berarti bahwa kesalahan atau kekurangan dapat ditolerir oleh pengguna. Nilai *severity rating* tertinggi yaitu terdapat pada aspek “membantu pengguna untuk mengenali, mendiagnosa, dan memulihkan dari kesalahan” dengan nilai *severity rating* sebesar 1.93. Hasil perhitungan untuk responden laki-laki dapat dilihat pada Tabel 4.

Tabel 5. Hasil Pengolahan Nilai *Severity Rating* Responden Perempuan

No	Aspek Usability	Severity Rating	
		A	B
1	Visibilitas status sistem	1.19	1
2	Kecocokan antara sistem dan dunia nyata	0.65	1
3	Kontrol dan kebebasan pengguna	1.85	2
4	Konsistensi dan standarisasi	1.32	1
5	Pencegah kesalahan	1.78	2
6	Pemahaman daripada ingatan	1.06	1
7	Fleksibilitas	1.00	1
8	Estetika dan desain minimalis	1.13	1
9	Membantu pengguna untuk mengenali, mendiagnosa, dan memulihkan dari kesalahan	2.13	2
10	Bantuan dan dokumentasi	1.42	2

Untuk responden perempuan didapatkan nilai *severity rating* 1 (satu) dan 2 (dua). Nilai *severity rating* tertinggi yaitu terdapat pada aspek “membantu pengguna untuk mengenali, mendiagnosa, dan memulihkan dari kesalahan” dengan nilai *severity rating* sebesar 2.13. Hasil perhitungan untuk responden perempuan dapat dilihat pada Tabel 5.

Untuk keseluruhan dari 50 responden tersebut didapatkan nilai *severity rating* 1 (satu) dan 2 (dua), berarti bahwa kesalahan atau kekurangan dapat ditolerir oleh pengguna. Oleh karena itu, masalah usabilitas yang terdapat pada *website* iLab tidak dipermasalahkan. Nilai *severity rating* tertinggi yaitu terdapat pada aspek “membantu pengguna untuk mengenali, mendiagnosa, dan memulihkan dari kesalahan” dengan nilai *severity rating* sebesar 2.00. Lebih lengkapnya dapat dilihat pada Tabel 6.

Tabel 6. Hasil Pengolahan Nilai *Severity Rating* Keseluruhan (50 Responden)

No	Aspek Usability	Severity Rating	
		A	B
1	Visibilitas status sistem	1.28	1
2	Kecocokan antara sistem dan dunia nyata	0.62	1
3	Kontrol dan kebebasan pengguna	1.87	2
4	Konsistensi dan standarisasi	1.28	1
5	Pencegah kesalahan	1.67	2
6	Pemahaman daripada ingatan	1.28	1
7	Fleksibilitas	0.92	1
8	Estetika dan desain minimalis	1.00	1
9	Membantu pengguna untuk mengenali, mendiagnosa, dan memulihkan dari kesalahan	2.00	2
10	Bantuan dan dokumentasi	1.67	2

Tabel 6 adalah hasil keseluruhan, di mana nilai *severity rating* pada kolom A merupakan hasil nilai mentah dengan 2 angka dibelakang koma, sedangkan untuk kolom B merupakan hasil nilai A yang dibulatkan ke bilangan bulat.

4. Kesimpulan

Berdasarkan pengukuran *website* iLab menggunakan metode evaluasi *heuristic* oleh Nielsen didapatkan bahwa 10 aspek yang diteliti bila dilihat berdasarkan jenis kelamin maka untuk responden laki-laki dan responden perempuan mendapatkan nilai satu dan dua. Nilai *severity rating* tertinggi pada responden laki-laki dan perempuan yaitu terdapat pada aspek “membantu pengguna untuk mengenali, mendiagnosa, dan memulihkan dari kesalahan” dengan nilai *severity rating* pada responden laki-laki sebesar 1.93 dan responden perempuan sebesar 2.13. Nilai tertinggi secara keseluruhan dari 50 responden tersebut mendapatkan nilai satu dan dua yang

berarti *website* iLab memiliki beberapa kekurangan yang tidak menjadikan sebuah masalah atau dengan kata lain tidak dipermasalahkan dan tidak mengganggu pengguna saat mengakses *website* iLab. Nilai *severity rating* tertinggi terdapat pada aspek membantu pengguna untuk mengenali, mendiagnosa, dan memulihkan dari kesalahan sebesar 2.00.

Daftar Pustaka

- [1] P. Krisnayani, I. K. Arthana, I. K., & I. G. Darmawiguna, "Analisa Usability pada Website Undiksha dengan Menggunakan Metode Heuristic Evaluation", *Karmapati*, vol. 5, no. 2, pp. 1-10, 2016.
- [2] L. D. Farida, "Pengukuran User Experience dengan Pendekatan Usability (Studi Kasus: Website Pariwisata di Asia Tenggara)", *Seminar Nasional Teknologi Informasi dan Multimedia*, Yogyakarta 6-7 Februari 2016.
- [3] P. Savitri, & M. Ispani, "Review Desain Interface Aplikasi Soppo Menggunakan Evaluasi Heuristik", *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, no. 6, vol. 1, pp. 95-100, 2015.
- [4] D. Caesaron, "Evaluasi Heuristic Desain Antar Muka (Interface) Portal Mahasiswa (Studi Kasus Portal Mahasiswa Universitas X)", *Jurnal Metris*, vol. 16, no. 1, pp. 9-14, 2015.
- [5] A. Folstad, "Users' Design Feedback in Usability Evaluation: a Literature Review", *HCCIS*, vol. 7, no.19, 2017.
- [6] R. F. Aziza & Y. T. Hidayat, "Analisis Usability Desain User Interface Pada Website Tokopedia Menggunakan Metode Heuristic Evaluation", *Jurnal Teknokompak*, vol. 13, no. 1, pp. 7-11, 2019.
- [7] Y. M. Gasela, P. Ranting, & J. F. Andry, "Analisis User Interface terhadap Website Berbasis E-Learning dengan Metode Heuristic Evaluation", *Jurnal Informatika*, vol. 5, no. 2, pp. 270-277, 2018.
- [8] M. Sulistiyono, "Evaluasi Heuristic Sistem Informasi Pelaporan Kerusakan Laboratorium Universitas Amikom Yogyakarta", *Jurnal Ilmiah DASI*, vol. 18, no. 1, pp. 37-43, 2017.
- [9] B. A. Mustikaningtyas, M. C. Saputra, & A. Pinandito, "Analisis Usability pada Website Universitas Brawijaya dengan Heuristic Evaluation", *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 3, pp. 188-193, 2016.
- [10] I. Khairina, Suprpto, & N. H. Wardani, "Analisis Usability pada Website Jawa Timur Park Group dengan Heuristic Evaluation" *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, no. 8, pp. 641-648, 2017.
- [11] Yadi, "Analisa Usability Pada Website Traveloka" *Jurnal Ilmiah Betrik*, vol. 09, no. 02, 2018.
- [12] F. Thewes, T. Herrmann, A. Kluge, "Validating Heuristic Evaluation Method An Application Test", *MuC'19: Proceedings of Mensch und Computer*, September 2019.
- [13] I. Almarashdeh & M. Alsmadi, "Heuristic Evaluation of Mobile Government Portal Services: An Experts Review". In *2016 11th International Conference for Internet Technology and Secured Transactions, ICITST*, 2016.
- [14] T. D. Susanto, A. I. Parsetyo, H. M. Astuti., "Web Usability Evaluation on BloobIS Website by Using Hallway Usability Testing Method and ISO 9241:11", *International Conference on Mathematics: Pure, Applied and Computation*, Surabaya 20 Oktober 2018.
- [15] N. I. Zulkifli, "Usability Testing Sistem Informasi Pendonor Darah (Studi Kasus di Unit Pelayanan Transfusi Darah RSUP Dr. Sardjito Yogyakarta)", *Jurnal for Information System*, vol. 1, no. 1, 2016.

SYSTEMIC: Information System and Informatics Journal

ISSN: 2460-8092, 2548-6551 (e)

Vol 6 No 2 – Desember 2020

Implementasi Perbandingan Algoritma Apriori Dan FP-Growth Untuk Mengetahui Pola Pembelian Konsumen Pada Produk Panel Di PT Surya Multi Perkasa MovinkoDiego Armando Pratama Putra¹, Tresna Maulana Fahrudin², Natalia Damastuti³

1,2,3) Universitas Narotama, Surabaya

armandodiego435@gmail.com¹, tresna.maulana@narotama.ac.id², natalia.damastuti@narotama.ac.id³**Kata Kunci***Apriori, FP-Growth, Konsumen, Pola pembelian, Produk Panel***Abstrak**

Beberapa perusahaan belum banyak memanfaatkan data transaksi pembelian konsumen sebagai salah satu strategi penjualannya, data transaksi ini meliputi barang apa saja yang sering dibeli oleh konsumen dalam satu transaksi pembelian pada struk dan waktu yang berbeda. Jika data transaksi tersebut dianalisis dan digali lebih mendalam, maka perusahaan mendapatkan suatu insight berupa analisis pola pembelian konsumen dan menguntungkan bagi perusahaan. Pada penelitian ini dilakukan analisis data transaksi pembelian konsumen menggunakan perbandingan algoritma Apriori dan FP-Growth, dimana keduanya merupakan kelompok Metode Association Rule yang bertujuan untuk mengetahui pola pembelian konsumen. Data yang digunakan pada penelitian ini diperoleh dari data transaksi pembelian produk panel pada PT Surya Multi Perkasa Movinko. Data transaksi tersebut terdiri dari 23 jenis item produk dan 492 transaksi. Hasil eksperimen dari penelitian ini menunjukkan bahwa kinerja terbaik algoritma Apriori dengan support factor sebesar 0.0054 dan confidence factor sebesar 0.30 menghasilkan 12 aturan asosiasi, sedangkan kinerja terbaik algoritma FP-Growth dengan support factor sebesar 2 dan confidence factor sebesar 0.7 menghasilkan 9 aturan asosiasi.

Keywords*Apriori, Consumer, FP-Growth, Purchase Pattern, Panel Products***Abstract**

Some companies have not used much consumer purchase transaction data as one of their sales strategies, this transaction data contains what items are often bought by consumers in one purchase transaction at a different time and structure. If the transaction data is analyzed and explored in more depth, the company will gain insight into consumer purchase patterns analysis and be profitable for the company. In this research, an analysis of consumer purchase transaction data was carried out using Apriori algorithm and FP-Growth, both of which are association rule method group that aims to determine consumer purchasing patterns. The data used in this study were obtained from panel product purchase transaction data at PT Surya Multi Perkasa Movinko. The transaction data consist of 23 types of product items and 492 transactions. The experimental results of this study showed that the best performance of Apriori algorithm with a support factor of 0.0054 and a confidence factor of 0.30 generating 12 association rules, while the best performance of FP-Growth algorithm with a supporting factor of 2 and a confidence factor of 0.7 generating 9 association rules.

1. Pendahuluan

PT Surya Multi Perkasa Movinko merupakan salah satu perusahaan yang memproduksi panel listrik. Kontraktor mekanikal elektrikal dan juga sebagai distributor resmi produk-produk elektrik schneider dan siemens. Awal berdirinya perusahaan adalah pada tahun 1993. PT Surya

Multi Perkasa Movinko terus berkembang dan menjadi salah satu perusahaan manufaktur lokal dalam menghasilkan produk perlengkapan kelistrikan. Berawal dari tahun 2010 hingga saat ini PT Surya Multi Perkasa Movinko sudah meraih pasar yang lebih luas dan mampu menyukseskan banyak proyek industri dan proyek komersial di Indonesia. PT Surya Multi Perkasa Movinko

memberikan layanan *electrical* panel, trafo dan juga *cable tray*.

Dalam dunia perbisnisan yang dimana semakin berkembang dan modern seperti saat ini. Sudah pasti terdapat banyak pesaing dari perusahaan-perusahaan lain dalam meningkatkan omset pada perusahaan, sehingga membuat para pebisnis ini harus menyusun strategi untuk menghadapi persaingan tersebut. Pada beberapa perusahaan masih banyak yang belum mengetahui bagaimana mengelola data transaksi pembelian konsumen, dimana data transaksi tersebut sebenarnya dapat menghasilkan pola pembelian meliputi barang apa saja yang sering dibeli oleh konsumen secara bersamaan dalam satu transaksi pembelian. Hal ini menunjukkan data transaksi penjualan masih belum dimanfaatkan dengan baik dan hanya dijadikan arsip saja.

Tinjauan penelitian pertama adalah karya Jordy Lasmana Putra, Mugi Raharjo, Tommi Alfian Armawan Sandi, Ridwan, Rizal Prasetyo (2019) dengan judul Implementasi Algoritma Apriori Terhadap Data Penjualan Pada Perusahaan Retail[1]. Pada penelitian diperoleh pola pembelian barang. Dimana para consumer lebih sering membeli barang *Cream Cupid Heart Coat Hanger*.

Tinjauan penelitian yang kedua adalah karya tulis dari Heroe Santoso, I Putu Hariyadi, Prayitno (2016), dengan judul Data Mining Analisa Pola Pembelian Produk Dengan Menggunakan Metode Algoritma Apriori[2]. dapat membantu untuk mengetahui produk barang apa saja yang jarang dibeli dan dapat membantu untuk mengetahui produk barang apa saja yang sering dibeli oleh konsumen, sehingga dapat menghasilkan alat alternative dalam meningkatkan strategi pemasaran dengan cara membuat diskon barang tertentu yang jarang di beli untuk menarik minat beli konsumen.

Tinjauan penelitian yang terakhir adalah karya tulis dari Luki Henando (2019), dengan judul Algoritma Apriori Dan FP-Growth Untuk Analisa Perbandingan Data Penjualan Laptop Berdasarkan Merk Yang Diminati Konsumen (Studi Kasus : INDOCOMPUTER Payakumbuh)[3]. Penelitian ini menggunakan dua perbandingan metode yang dapat ditarik kesimpulan, Dengan menggunakan Algoritma Apriori Dan FP-Growth dengan bantuan software WEKA 3.7.4 dapat menghasilkan suatu aturan (rules) yang merupakan kumpulan frequent itemset dengan nilai confidence yang tinggi. Dengan didapatkannya rules ini maka pimpinan dapat menggunakan aturan (rules) dalam membuat suatu strategi yang dapat meningkatkan jumlah penjualan laptop setiap harinya.

Pada penelitian ini, kami bertujuan untuk mengimplementasikan Metode Association Rule untuk mengetahui pola pembelian konsumen terhadap produk panel menggunakan perbandingan kinerja dari algoritma Apriori dan

FP-Growth. Untuk mengetahui pola pembelian konsumen yaitu dapat dilakukan dengan cara mengolah data transaksi penjualan pada PT Surya Multi Perkasa Movinko.

2. Metode Penelitian

Pada penelitian ini terdapat dua metode yang digunakan, metode tersebut adalah algoritma Apriori dan Algoritma FP-Growth, kedua metode tersebut merupakan kelompok dari Metode Association Rule. Pada bagian ini akan dibahas terkait teori keduanya.

2.1 Data Mining

Data Mining merupakan salah satu proses pencarian secara otomatis informasi yang berguna dalam tempat penyimpanan data berukuran besar. Dan untuk mencari pola atau informasi menarik dengan menggunakan teknik atau metode tertentu[[4]]. Dapat mengerjakan satu atau lebih dari teknik pembelajaran komputer (*machine learning*) dalam menganalisis dan mengekstraksi pengetahuan (*knowlegde*) secara otomatis [5].

Pengelompokan Data Mining dibagi menjadi beberapa kelompok [6], yaitu :

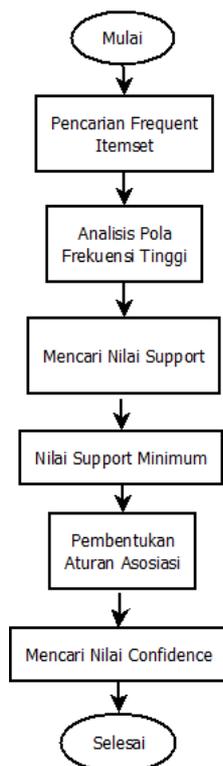
- a. Deskripsi
Deskripsi merupakan cara untuk mengidentifikasi pola yang muncul secara berulang pada suatu data dan mengubah pola tersebut menjadi aturan dan kriteria yang dapat dipahami dengan mudah
- b. Estimasi
Estimasi hampir memiliki kesamaan dengan klasifikasi, yang dimana pada variable target estimasi lebih ke arah numerik dari pada ke arah kategori.
- c. Prediksi
Prediksi merupakan sebuah nilai yang masih belum diketahui dan juga memperkirakan suatu nilai pada masa yang akan mendatang.
- d. Klasifikasi
Klasifikasi merupakan proses menemukan sebuah model atau fungsi yang mendeskripsikan dan membedakan data ke dalam kelas.
- e. Pengklasteran
Merupakan pengelompokan dari suatu *record*, pengamatan, atau memperhatikan dan membentuk kelas objek yang memiliki beberapa kemiripan
- f. Asosiasi
Asosiasi bertugas untuk menemukan atribut yang muncul dalam suatu waktu. Dalam dunia bisnis lebih umum disebut dengan analisis keranjang belanja (*market basket analysis*).

2.2 Association Rule

Association rule merupakan salah satu metode yang memiliki tujuan untuk menemukan

pola pembelian pada suatu transaksi, metode ini mendukung sistem rekomendasi melalui penemuan pola antar item dalam suatu transaksi yang terjadi [2]. Tugas dari association rule adalah untuk mencari aturan yang tidak mengcover untuk mengukur hubungan antara dua atau lebih atribut[7]. Langkah pertama yang perlu diperhatikan dalam Association Rule adalah dengan mengetahui seberapa sering kombinasi item yang muncul secara bersamaan (*frequent pattern*)[8]. Association rule memiliki bentuk LHS (*Left Hand Shake*) -> RHS (*Right Hand Shake*), dengan interpretasi jika setiap item di dalam LHS maka kemungkinan item didalam RHS juga akan dibeli[9][10].

2.3 Algoritma Apriori



Gambar 1. Flowchart Algoritma Apriori

Algoritma Apriori merupakan salah satu teknik data mining yang digunakan untuk menemukan aturan asosiasi pada kombinasi item[11][12][13]. Algoritma Apriori dibagi menjadi beberapa tahap atau yang biasa disebut iterasi[14]. Tahap pertama yang harus dilakukan adalah dengan melakukan pencarian *frequent itemset* (himpunan item-item yang memenuhi *minimum support*) terlebih dahulu.

Frequent itemset adalah sekumpulan item yang sering muncul secara bersamaan pada suatu dataset. Suatu asosiasi dapat diketahui dengan tolak ukur, yaitu dengan nilai *support* dan *confidence* [15].

Metodologi dasar analisis asosiasi terbagi menjadi dua tahap[16]:

a. Analisis pola frekuensi tinggi

Tahap ini digunakan untuk mencari kombinasi dari setiap item yang sudah memenuhi syarat minimum dari nilai *support*. Nilai *support* dapat dicari dengan menggunakan persamaan 1 seperti berikut:

$$Support(A) = \frac{Jumlah\ transaksi\ mengandung\ A}{Total\ transaksi} \times 100\% \quad (1)$$

Sedangkan nilai *support* dari 2 kombinasi item dapat diperoleh dengan persamaan 2 sebagai berikut:

$$Support(A \cap B) = \frac{Jumlah\ transaksi\ mengandung\ A\ dan\ B}{Total\ transaksi} \times 100\% \quad (2)$$

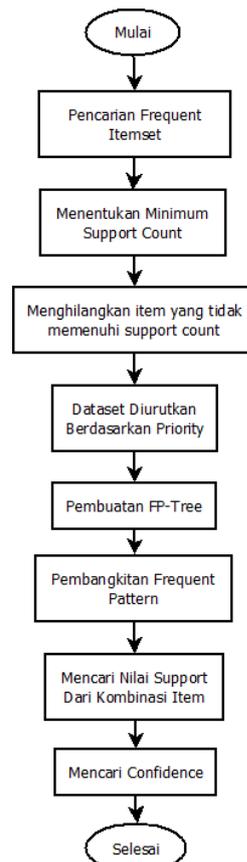
b. Pembentukan aturan asosiatif

Setelah semua pola frekuensi tinggi ditemukan, selanjutnya mencari aturan *asosiatif* yang sudah memenuhi syarat minimum untuk *confidence*, dengan cara menghitung *confidence* aturan *asosiatif* A_B

Nilai *confidence* aturan A_B dapat diperoleh dengan persamaan 3 sebagai berikut:

$$Confidence = P(B|A) = \frac{\sum Transaksi\ mengandung\ A\ dan\ B}{\sum Transaksi\ mengandung\ A} \times 100\% \quad (3)$$

2.4 Algoritma FP-Growth



Gambar 2. Flowchart Algoritma FP-Growth

Algoritma FP-Growth (*Frequent Pattern Growth*) merupakan salah satu algoritma yang dapat digunakan untuk menentukan himpunan data yang sering muncul (*frequent itemset*). Algoritma FP-Growth adalah pengembangan dari algoritma Apriori. Sehingga kekurangan yang terdapat pada algoritma Apriori dapat diperbaiki oleh algoritma FP-Growth. Algoritma Apriori memiliki kekurangan yang sangat besar, yaitu harus mendapatkan *frequent sets* yang diproduksi *candidate frequent sets* tanpa henti.

FP-Growth menggunakan konsep pembangunan *tree* dalam pencarian *frequent itemset*. Dengan menggunakan FP-Tree, algoritma FP-Growth dapat langsung mengekstrak *frequent itemset* dari FP-Tree. Hal tersebut yang menyebabkan algoritma FP-Growth lebih cepat dari algoritma Apriori[6].

Algoritma FP-Growth memiliki tiga tahapan utama[17], yaitu:

- Tahap pembangkitan *conditional pattern base*
Conditional pattern base merupakan *sub-database* yang berisi *prefix path* dan *suffix pattern* (pola akhiran).
- Tahap pembangkitan conditional FP-Tree
Pada tahap ini *support count* dari setiap item pada setiap *conditional pattern base* dijumlahkan.
- Tahap pencarian *frequent itemset*
Tahap ini merupakan pencarian *frequent itemset* yang memiliki lintasan tunggal (*single path*), kemudian didapatkan *frequent itemset* dengan melakukan kombinasi item untuk conditional FP-Tree.

3. Hasil Dan Pembahasan

Pada bagian ini akan dibahas hasil dari penelitian dengan mengimplementasikan algoritma Apriori dan FP-Growth untuk mengetahui bagaimana pola pembelian pada konsumen yang dihasilkan.

3.1 Hasil dan pembahasan algoritma Apriori

Pada Gambar 3 adalah langkah pertama dalam menganalisa algoritma Apriori pada *jupyterlab*, yaitu dengan cara meng-*install library* terlebih dahulu.

```
[1]: pip install apyori
Requirement already satisfied: apyori in c:\users\hp\anaconda3\lib\site-packages (1.1.2)
Note: you may need to restart the kernel to use updated packages.
```

Gambar 3. Install Apyori Library

Pada Gambar 4 adalah langkah untuk meng-*import* dan *load dataset*.

```
[2]: #importing the required dataset
import numpy as np
import pandas as pd
from apyori import apriori

[3]: #Loading the dataset
datapanel = pd.read_excel("C:/Users/hp/Documents/Diego/P.xlsx")

[4]: datapanel.head()
```

	Barang 1	Barang 2	Barang 3	Barang 4	Barang 5	Barang 6	Barang 7	Barang 8
0	Panel KWH	Panel Capacitor Bank	Panel ATS	Panel AMF	Panel COS	Panel Water Level Control	Panel LVMDP	Panel LVSDP
1	Panel KWH	NaN	NaN	NaN	NaN	NaN	Panel LVMDP	NaN
2	NaN	NaN	Panel ATS	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	Panel ATS	NaN	NaN	NaN	NaN	Panel LVSDP

Gambar 4. Import dan Load Dataset

Pada Gambar 5 adalah langkah untuk melihat garis dan kolom dari *dataset*, pada penelitian ini terdapat 492 transaksi dan 23 jenis barang.

```
[5]: datapanel.shape

[5]: (492, 23)

[6]: #converting the pandas dataframe into a list of lists
records = []
for i in range(0, 492):
    records.append([str(datapanel.values[i, j]) for j in range(0,23)])
```

Gambar 5. Jumlah Kolom dan Baris Dataset

Pada Gambar 6 adalah langkah untuk membangun apriori model dengan menentukan nilai *minimum support*, *minimum confidence*, *minimum lift* dan *minimum length*.

```
[7]: #Building the first apriori model
association_rules = apriori(records,
                             min_support=0.0054,
                             min_confidence=0.30,
                             min_lift=3,
                             min_length=2)

[8]: association_results = list(association_rules)

[9]: #Getting the number of rules
print(len(association_results))

12

[10]: #Glancing at the First rule
print(association_results[0])

RelationRecord(items=frozenset({'Panel GGCP', 'Box Panel',
atistics=[OrderedStatistic(items_base=frozenset({'Box Panel
confidence=1.0, lift=12.0), OrderedStatistic(items_base=fro
nel Aluminium'}), confidence=0.6000000000000001, lift=7.569
P', 'Box Panel Aluminium'}), items_add=frozenset({'Box Pane
```

Gambar 6. Membangun Model Apriori

Dalam penelitian ini hasil yang didapatkan menggunakan algoritma Apriori dengan nilai *minimum support* = 0.0054, nilai *minimum confidence* = 0.30, nilai *minimum lift* = 3, dan nilai *minimum length* = 2 menghasilkan 12 rule.

	title1	title2	support	confidence	lift
0	Box Panel	Panel GGCP	0.00609	1.0	12.0
1	Box Panel	Box Panel Outdoor	0.00609	0.60000	7.76842
2	Box Panel	Panel MVMDP	0.00609	0.60000	7.20000
3	Box Panel	Panel MVMDP	0.00813	0.66666	8.00000
4	Box Panel	Panel MDP	0.00609	0.60000	7.20000
5	Panel MVMDP	Box Panel Outdoor	0.00609	0.75	9.0
6	Panel MVMDP	Panel GGCP	0.00609	0.75	9.0
7	Panel MCC	Cable Tray	0.00609	0.75	8.20000
8	Panel Dol	Panel MDP	0.00609	0.75	9.0
9	Panel MVMDP	Panel PMCC	0.00609	1.0	12.0
10	Box Panel	Panel MVMDP	0.00609	0.60000	59.0400
11	Box Panel	Panel MVMDP	0.00609	0.5	6.15

Gambar 7. Hasil Analisis Pola Pembelian Konsumen Menggunakan Algoritma Apriori

Pada gambar 7 menunjukkan hasil dari analisis pola pembelian konsumen yang menggunakan algoritma Apriori yang dimana customer membeli box panel maka customer juga akan membeli panel MVMDP dengan tingkat support factor 0.00609 dan confidence 0.5

3.2 Hasil dan pembahasan algoritma FP-Growth

Pada Gambar 8 adalah langkah awal dalam menganalisis algoritma FP-Growth pada jupyterlab, dengan cara meng-install library terlebih dahulu.

```

+ 🔍 📄 ▶ ■ ⌂ Code ▼
[1]: pip install pyfpgrowth
Requirement already satisfied: pyfpgrowth in c:\users\hp\anaconda3\lib\site-packages (1.0)
Note: you may need to restart the kernel to use updated packages.
    
```

Gambar 8. Install pyfpgrowth Library

Pada Gambar 9 adalah langkah untuk membaca dataset.

```

[2]: with open('FIX2.csv', 'r') as read_obj:
      # pass the file object to reader() to get the reader object
      csv_reader = reader(read_obj)
      # Pass reader object to list() to get a list of lists
      list_of_rows = list(csv_reader)

[3]: transactions = list_of_rows
      transactions

[4]: [['Transactions'],
      ['Panel KWH', 'Panel LVMDP'],
      ['Panel ATS'],
      ['Panel ATS', 'Panel LVSDP', 'Box Panel Aluminium', 'Cable Tray'],
      ['Panel SDP', 'Box Panel Indoor'],
      ['Cable Tray'],
      ['Panel COS', 'Panel MDP', 'Panel Dol'],
      ['Panel ANF', 'Box Panel Outdoor'],
      ['Panel NGP'],
      ['Panel ATS', 'Box Panel Outdoor'],
      ['Panel ATS', 'Panel GGCP', 'Panel MVMDP', 'Cable Tray'],
    ]
    
```

Gambar 9. Import dan Load Dataset

Pada gambar 10 adalah langkah untuk mencari frequent pattern dengan nilai minimum support = 2.

```

[5]: patterns = pyfpgrowth.find_frequent_patterns(transactions, 2)
      patterns

[5]: (('Panel KWH', 'Panel LVMDP'): 2,
      ('Box Panel Indoor', 'Panel LVMDP'): 2,
      ('Panel GGCP', 'Panel LVMDP'): 2,
      ('Panel Capacitor Bank', 'Panel LVMDP'): 2,
      ('Panel LVMDP', 'Panel MDP'): 2,
      ('Cable Tray', 'Panel LVMDP'): 3,
      ('Cable Tray', 'Panel LVMDP', 'Panel MCC'): 2,
      ('Panel LVMDP', 'Panel MCC'): 4,
      ('Panel LVMDP', 'Panel VSD'): 6,
      ('Box Panel Aluminium', 'Panel Water Level Control'): 2,
      ('Panel NGP', 'Panel Water Level Control'): 2,
      ('Panel MVMDP', 'Panel Water Level Control'): 2,
      ('Panel Capacitor Bank', 'Panel Water Level Control'): 2,
    ]
    
```

Gambar 10. Mencari Frequent Pattern pada algoritma FP-Growth

Pada Gambar 11 adalah langkah untuk mencari rule dengan menggunakan nilai minimum confidence = 0.7.

```

[6]: rules = pyfpgrowth.generate_association_rules(patterns, 0.7)
      rules

[6]: (('Cable Tray', 'Panel MCC'): (('Panel LVMDP',), 1.0),
      ('Box Panel Indoor', 'Panel MVMDP'): (('Box Panel', 'Box Panel Outdoor'),
      1.0),
      ('Box Panel', 'Box Panel Indoor', 'Box Panel Outdoor'): (('Panel MVMDP',),
      1.0),
      ('Box Panel', 'Box Panel Indoor', 'Panel MVMDP'): (('Box Panel Outdoor',),
      1.0),
      ('Box Panel Indoor', 'Box Panel Outdoor', 'Panel MVMDP'): (('Box Panel',),
      1.0),
      ('Panel Dol', 'Panel MDP'): (('Panel COS',), 1.0),
      ('Panel Capacitor Bank', 'Panel MVMDP'): (('Panel PMCC',), 1.0),
      ('Box Panel Outdoor', 'Panel MVMDP'): (('Box Panel',), 0.75),
      ('Box Panel', 'Box Panel Aluminium'): (('Panel GGCP',), 1.0)}
    
```

Gambar 11. Generate Aturan Asosiasi Menggunakan algoritma FP-Growth

Dalam penelitian ini hasil yang didapatkan dengan menggunakan algoritma FP-Growth dengan nilai minimum support = 2 dan nilai minimum confidence = 0.7 menghasilkan 9 rule, berdasarkan pada Gambar 12 menunjukkan hasil dari output dengan algoritma FP-Growth. Terdapat percobaan menggunakan nilai minimum support 1 dan nilai minimum confidence 0.7 menghasilkan terlalu banyak rule, ketika menggunakan nilai minimum support 3 dan nilai minimum confidence 0.7 hanya dapat menghasilkan 1 rule, dan untuk nilai minimum support 4 dan 5 dengan nilai minimum confidence 0.7 tidak menghasilkan rule (0).

```

 (('Cable Tray', 'Panel MCC'): (('Panel LVMDP',), 1.0),
 ('Box Panel Indoor', 'Panel MVMDP'): (('Box Panel', 'Box Panel Outdoor'),
 1.0),
 ('Box Panel', 'Box Panel Indoor', 'Box Panel Outdoor'): (('Panel MVMDP',),
 1.0),
 ('Box Panel', 'Box Panel Indoor', 'Panel MVMDP'): (('Box Panel Outdoor',),
 1.0),
 ('Box Panel Indoor', 'Box Panel Outdoor', 'Panel MVMDP'): (('Box Panel',),
 1.0),
 ('Panel Dol', 'Panel MDP'): (('Panel COS',), 1.0),
 ('Panel Capacitor Bank', 'Panel MVMDP'): (('Panel PMCC',), 1.0),
 ('Box Panel Outdoor', 'Panel MVMDP'): (('Box Panel',), 0.75),
 ('Box Panel', 'Box Panel Aluminium'): (('Panel GGCP',), 1.0)}
    
```

Gambar 12. Hasil Analisis Pola Pembelian Konsumen Menggunakan Algoritma FP-Growth

Gambar 12 menunjukkan hasil dari analisis pola pembelian konsumen yang menggunakan algoritma FP-Growth yang dimana customer membeli box panel, box panel indoor dan box panel outdoor akan membeli panel MVMDP dengan nilai confidence 1.0

4. Kesimpulan

Algoritma Apriori dan FP-Growth merupakan kelompok dari Metode Association Rule. Keduanya cocok digunakan untuk mengetahui pola pembelian pada konsumen, dari uji coba perbandingan kedua metode tersebut, algoritma FP-Growth memiliki nilai confidence yang lebih tinggi dari pada algoritma Apriori, dan hasil rule yang diperoleh dari algoritma FP-Growth sudah mendekati dengan kasus yang ada di lapangan.

Daftar Pustaka

- [1] J. L. Putra, M. Raharjo, T. A. A. Sandi, R. Ridwan, and R. Prasetyo, "Implementasi Algoritma Apriori Terhadap Data Penjualan Pada Perusahaan Retail," *J. Pilar Nusa Mandiri*, vol. 15, no. 1, pp. 85–90, 2019, doi: 10.33480/pilar.v15i1.113.
- [2] H. Santoso, I. P. Hariyadi, and Prayitno, "Data Mining Analisa Pola Pembelian Produk," *Tek. Inform.*, no. 1, pp. 19–24, 2016, [Online]. Available: <http://ojs.amikom.ac.id/index.php/semnasteknomedia/article/download/1267/1200>.
- [3] K. Dan *et al.*, "ALGORITMA APRIORI DAN FP-GROWTH UNTUK ANALISA PERBANDINGAN DATA PENJUALAN LEPTOP BERDASARKAN MERK YANG DIMINATI KONSUMEN (STUDI KASUS : INDOCOMPUTER PAYAKUMBUH)," *J. Sist. Inf. Dan Manaj. Inform.*, vol. 6, no. 2, pp. 201–207, 2019.
- [4] K. N. Angraini, Indwiarti, and F. Nhita, "Implementasi Algoritma Fuzzy c-Covering untuk Mengetahui Pola Pembelian pada Data Transaksi Swalayan," vol. 5, no. 3, pp. 8198–8205, 2018.
- [5] J. Eska, "Penerapan Data Mining Untuk Prekdiksi Penjualan Wallpaper Menggunakan Algoritma C4.5 STMIK Royal Ksieran," *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 2, pp. 9–13, 2016.
- [6] A. Maulana and A. A. Fajrin, "Penerapan Data Mining Untuk Analisis Pola Pembelian Konsumen Dengan Algoritma Fp-Growth Pada Data Transaksi Penjualan Spare Part Motor," *Klik - Kumpul. J. Ilmu Komput.*, vol. 5, no. 1, p. 27, 2018, doi: 10.20527/klik.v5i1.100.
- [7] N. A. Hasibuan *et al.*, "Implementasi Data Mining Untuk Pengaturan Layout," vol. 4, no. 4, pp. 6–11, 2017.
- [8] W. Aprianti, K. A. Hafizd, and M. R. Rizani, "Implementasi Association Rules dengan Algoritma Apriori pada Dataset Kemiskinan," *Limits J. Math. Its Appl.*, vol. 14, no. 2, p. 57, 2017, doi: 10.12962/limits.v14i2.2933.
- [9] Dio Prima Mulya, "Analisa Dan Implementasi Association Rule Dengan Algoritma Fp-Growth Dalam Seleksi Pembelian Tanah Liat," *Teknol. dan Sist. Inf. Bisnis*, vol. 1, no. 1, pp. 47–57, 2019.
- [10] A. Setiawan and I. G. Anugrah, "Penentuan Pola Pembelian Konsumen pada Indomaret GKB Gresik dengan Metode FP-Growth," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 2, no. 2, p. 115, 2019, doi: 10.32672/jnkti.v2i2.1564.
- [11] T. M. Fahrudin, I. Syarif, and A. R. Barakbah, "Discovering patterns of NED-breast cancer based on association rules using apriori and FP-growth," *Int. Electron. Symp. Knowl. Creat. Intell. Comput.*, pp. 132–139, 2017, doi: 10.1109/KCIC.2017.8228576.
- [12] N. Rahmawati, Y. N. Nasution, and F. D. T. Amijaya, "Aplikasi Data Mining Market Basket Analysis untuk Menemukan Pola Pembelian di Toko Metro Utama Balikpapan," *J. EKSPONENSIAL*, vol. 8, no. 1, pp. 1–8, 2017, [Online]. Available: <http://jurnal.fmipa.unmul.ac.id/index.php/expoensial/article/view/69>.
- [13] M. Badrul, "Algoritma Asosiasi Dengan Algoritma Apriori Untuk Analisa Data Penjualan," *None*, vol. 12, no. 2, pp. 121–129, 2016.
- [14] E. T. L. Christin Nandari Dengen, Kusriani, "Penentuan Association Rule Pada Kelulusan Mahasiswa Menggunakan Algoritma Apriori," *Jurti*, vol. 3, no. 1, pp. 20–29, 2019.
- [15] A. Ikhwan, "A Novelty of Data Mining for FP-Growth Algorithm," *Int. J. Civ. Eng. Technol.*, vol. 9, no. 7, pp. 1660–1669, 2018.
- [16] I. P. Astuti, "Algoritma Apriori Untuk Menemukan Hubungan Antara Jurusan Sekolah Dengan Tingkat Kelulusan Mahasiswa," *J. Tek. Inform.*, vol. 12, no. 1, pp. 69–78, 2019, doi: 10.15408/jti.v12i1.10525.
- [17] R. Fitria, W. Nengsih, and D. H. Qudsi, "Implementasi Algoritma FP-Growth Dalam Penentuan Pola Hubungan Kecelakaan Lalu Lintas," *J. Sist. Inf.*, vol. 13, no. 2, p. 118, 2017, doi: 10.21609/jsi.v13i2.551.

SYSTEMIC: Information System and Informatics Journal

ISSN: 2460-8092, 2548-6551 (e)

Vol 6 No 2 – Desember 2020

Klasifikasi Multi Output pada Harga Smartphone Menggunakan Learning Vector Quantization (LVQ) dan Backpropagation (BP)Dinita Rahmalia¹, M. Syaiful Pradana², Teguh Herlambang³^{1,2} Universitas Islam Darul Ulum Lamongan³ Universitas Nahdlatul Ulama Surabayadinitarahmalia@gmail.com¹, syaifulp@unisda.ac.id², teguh@unusa.ac.id³**Kata Kunci***Klasifikasi, Neural Network, Learning Vector Quantization, Backpropagation, Pengenalan Pola***Abstrak**

Terdapat banyak smartphone dengan harga yang bermacam-macam yang terjual di pasaran. Harga smartphone dipengaruhi oleh beberapa komponen seperti berat, penyimpanan internal, memory (RAM), kamera belakang, kamera depan dan merek. Terdapat dua metode untuk mengklasifikasikan kelas harga smartphone di pasaran seperti Learning Vector Quantization (LVQ) dan Backpropagation (BP). Dari klasifikasi kelas harga smartphone di pasaran menggunakan LVQ and BP, terdapat perbedaan pada keduanya. LVQ mengklasifikasikan kelas harga smartphone dengan jarak euclidean dari bobot dan data pada iterasinya. BP mengklasifikasikan kelas harga smartphone dengan penurunan gradient dari target dan output pada iterasinya. Pada klasifikasi multi output, satu obyek bisa memiliki lebih dari satu output. Berdasarkan hasil simulasi, BP memberikan akurasi dan error yang lebih baik pada data training dan data testing daripada LVQ.

Keywords*Classification, Neural Network, Learning Vector Quantization, Backpropagation, Pattern Recognizing***Abstract**

There are many smartphones with various price sold in market. The price of smartphone is affected by some components such as weight, internal storage, memory (RAM), rear camera, front camera and brands. There are two methods for classifying price class of smartphone in market such as Learning Vector Quantization (LVQ) and Backpropagation (BP). From classifying price class of smartphone in market using LVQ and BP, there are the differences on the both of them. LVQ classifies price range of smartphone by euclidean distance of weight and data on its iteration. BP classifies price range of smartphone by gradient descent of target and output on its iteration. In multi output classification, one object may have multi output. Based on simulation results, BP gives the better accuracy and error rate in training data and testing data than LVQ.

1. Pendahuluan

Pada saat ini perkembangan teknologi informasi dan komunikasi bekerja sangat cepat, luas dan akurat. Salah satu produk yang paling utama dan banyak digunakan adalah smartphone. Smartphone adalah peralatan komunikasi (gadget) yang sangat penting untuk setiap orang sehingga proses komunikasi dapat berjalan sangat cepat. Karena itu, terdapat sangat banyak perusahaan smartphone bersaing memproduksi bermacam-macam smartphone dengan spesifikasi dan harga yang berbeda.

Di pasaran, terdapat banyak smartphone yang dijual dengan harga yang bermacam-macam. Berdasarkan data penjualan smartphone, harga smartphone tergantung pada beberapa komponen

seperti berat, penyimpanan internal, memory (RAM), kamera belakang, kamera depan dan merek, sehingga pada penelitian ini akan dilakukan klasifikasi kelas harga menggunakan metode Neural Network yaitu Learning Vector Quantization (LVQ) dan Backpropagation (BP).

Pada penelitian sebelumnya, aplikasi Neural Network telah diterapkan pada peramalan menggunakan Backpropagation (BP) pada peramalan cuaca [3], peramalan penyebaran penyakit [7], peramalan suhu udara [6] dan Adaptive Neuro Fuzzy Inference System (ANFIS) pada peramalan kelembaban udara [9], peramalan intensitas matahari [11]. Metode peramalan juga diterapkan menggunakan metode Kalman Filter pada peramalan harga saham [10], peramalan

persediaan darah [5]. Selain peramalan, juga terdapat klustering atau pengelompokan menggunakan Fuzzy Clustering Means pada pengelompokan bandara udara [4] dan pengelompokan produksi pertanian [8].

Pada penelitian ini, terdapat dua metode untuk mengklasifikasi harga smartphone di pasaran, yaitu Learning Vector Quantization (LVQ) dan Backpropagation (BP). LVQ dan BP merupakan jenis dari Neural Network yang sering digunakan pada klasifikasi data [1]. Pada proses klasifikasi menggunakan LVQ dan BP, terdapat perbedaan pada kedua metode tersebut. LVQ mengklasifikasi harga smartphone menggunakan jarak euclidean antara bobot (weight) dengan data di setiap iterasi. BP mengklasifikasi harga smartphone menggunakan penurunan gradien (gradient descent) dari target dan output di setiap iterasi. LVQ dan BP bekerja dengan membagi data ke dalam data training dan data testing dengan proporsi yang ditentukan dan mengupdate matriks bobot. Pada LVQ, matriks bobot berperan sebagai vektor referensi sedangkan pada BP, matriks bobot menghubungkan beberapa hidden layer

Pada klasifikasi multi output, satu obyek bisa memiliki lebih dari satu output. Berdasarkan hasil simulasi, BP memberikan accuracy yang lebih baik pada data training maupun data testing daripada LVQ.

2. Metode Penelitian

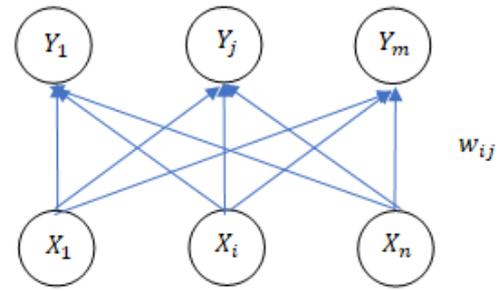
Metode yang digunakan pada penelitian adalah aplikasi Neural Network yang digunakan pada masalah klasifikasi yaitu Learning Vector Quantization (LVQ) dan Backpropagation (BP).

2.1 Learning Vector Quantization

Learning Vector Quantization (LVQ) ditemukan oleh Kohonen pada tahun 1990. Berdasarkan modifikasi pada bobot, terdapat dua jenis pelatihan (training) yaitu supervised dan unsupervised training. Pada supervised training, terdapat pasangan input dan target yang digunakan untuk melatih jaringan sampai matriks bobot optimal diperoleh. Untuk setiap training, inputs diberikan pada jaringan. Jaringan akan memproses output. Error antara output dan target dipropagasikan untuk mengupdate matriks bobot [1].

Pada unsupervised training, terdapat vektor bobot. Vektor bobot untuk suatu unit output pada jaringan berperan sebagai vektor untuk pola input. Selama training, suatu jaringan menentukan unit output yang terbaik untuk vektor input. Vektor bobot untuk pemenang disesuaikan menurut algoritma pada jaringan [1].

Model LVQ adalah model supervised training yang dapat dilihat pada Gambar 1.



Gambar 1. Model Jaringan Learning Vector Quantization (LVQ)

Pada LVQ, terdapat beberapa parameter, yaitu :

- N : jumlah data
- M : jumlah klasifikasi
- x_i : data ke- i
- c_i : klasifikasi pada data ke- i
- w_j : bobot dari klasifikasi ke- j

Algoritma LVQ adalah sebagai berikut :

Inisialisasi matriks bobot W sebagai vektor referensi.

$e = 1$

while($e \leq \max_epoch$ & & $dist \geq \min_dist$)

for($i = 1 : N$)

1. Hitung jarak euclidean $dist(j)$

$$dist(j) = \|w_j - x_i\| \quad j = 1, 2, \dots, M \quad (1)$$

2. Cari index J sehingga $dist(j)$ minimum.

3. Update $w_{j(\min)}$ dengan :

If $C_i = j(\min)$

$$w_{j(\min)}^{new} = w_{j(\min)}^{old} + \alpha(x_i - w_{j(\min)}^{old}) \quad (2)$$

Else

$$w_{j(\min)}^{new} = w_{j(\min)}^{old} - \alpha(x_i - w_{j(\min)}^{old}) \quad (3)$$

End

end

Hitung jarak euclidean pada persamaan (4)

$$dist = \sum_{d=1}^{datasize} \|x_d - w^*\| \quad (4)$$

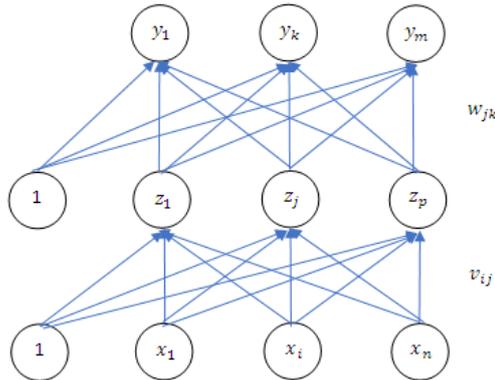
Turunkan learning rate $\alpha = \delta\alpha$ dengan δ adalah bilangan antara 0-1.

$e = e + 1$

end

2.2 Backpropagation

Backpropagation (BP) adalah jenis Neural Network yang digunakan pada proses estimasi. Model BP dapat dilihat pada Gambar 2.



Gambar 2. Model Jaringan Backpropagation (BP)

BP terdiri dari beberapa input x_1, x_2, \dots, x_n , beberapa hidden layer z_1, z_2, \dots, z_p , dan beberapa output y_1, y_2, \dots, y_m . Pada input dan hidden layer, terdapat bias yang memiliki nilai 1. Bobot v_{ij} menghubungkan input x_i ke hidden layer z_j . Bobot w_{jk} menghubungkan hidden layer z_j ke output y_k . Pada BP, terdapat tiga fase perhitungan yaitu forward propagation, backward propagation, dan update matriks bobot [1].

Pada Backpropagation (BP), terdapat fungsi aktivasi yang berfungsi menentukan output dari neuron. Argument dari fungsi aktivasi adalah kombinasi linear dari input dan bobot.

$$\begin{aligned} net &= \sum_i x_i w_i \\ f(net) &= f\left(\sum_i x_i w_i\right) \end{aligned} \quad (5)$$

Sifat dari fungsi aktivasi adalah kontinu, differentiable, dan bukan fungsi turun [1]. Pada penelitian ini, fungsi aktivasi yang dipakai adalah binary sigmoid dengan range (0-1) seperti pada persamaan (6).

$$\begin{aligned} f(net) &= \frac{1}{1 + e^{-net}} \\ f'(net) &= f(net)(1 - f(net)) \end{aligned} \quad (6)$$

Algoritma BP adalah sebagai berikut :

Inisialisasi matriks bobot V and W dengan bilangan random antara -0.5 sampai 0.5.

$e = 1$

while($e \leq \max_epoch$ & & $MSE \geq \min_MSE$)

for($d = 1 : \text{datasize}$)

1. Masing-masing input menerima sinyal dan melanjutkan ke setiap hidden layer melalui

forward propagation pada persamaan (7) sampai persamaan (10) dan backward propagation pada persamaan (11) sampai persamaan (15).

Forward Propagation

2. Hitung semua output $z_j, j = 1, 2, \dots, p$ pada hidden layer

$$z_net_j = v_{oj} + \sum_{i=1}^n x_i v_{ij} \quad (7)$$

$$z_j = f(z_net_j) = \frac{1}{1 + e^{-z_net_j}} \quad (8)$$

3. Hitung semua output $y_k, k = 1, 2, \dots, m$ pada output layer

$$y_net_k = w_{ok} + \sum_{j=1}^p z_j w_{jk} \quad (9)$$

$$y_k = f(y_net_k) = \frac{1}{1 + e^{-y_net_k}} \quad (10)$$

Backward Propagation

4. Hitung factor δ output berdasarkan error pada setiap output layer $y_k, k = 1, 2, \dots, m$

$$\delta_k = (t_k - y_k) f'(y_net_k) \quad (11)$$

5. Hitung perubahan bobot

$$\begin{aligned} \Delta w_{jk} &= \alpha \delta_k z_j, \\ k &= 1, 2, \dots, m \quad j = 0, 1, 2, \dots, p \end{aligned} \quad (12)$$

6. Hitung factor δ hidden layer berdasarkan error pada setiap hidden layer $z_j, j = 1, 2, \dots, p$

$$\delta_net_j = \sum_{k=1}^m \delta_k w_{jk} \quad (13)$$

$$\delta_j = \delta_net_j f'(z_net_j) \quad (14)$$

7. Hitung perubahan bobot

$$\begin{aligned} \Delta v_{ij} &= \alpha \delta_j x_i, \\ j &= 1, 2, \dots, p \quad i = 0, 1, 2, \dots, n \end{aligned} \quad (15)$$

Update Matriks Bobot

8. Update matriks bobot yang baru

$$w_{jk} = w_{jk} + \Delta w_{jk} \quad (16)$$

$$v_{ij} = v_{ij} + \Delta v_{ij} \quad (17)$$

end

De-normalisasi output untuk membandingkan dengan target asli

Hitung Root of Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{datasize} \sum_{d=1}^{datasize} \frac{1}{m} \sum_{k=1}^m (T_{dk} - Y_{dk})^2} \quad (18)$$

with T_{dk} is target value and Y_{dk} is outputs.

$e = e + 1$

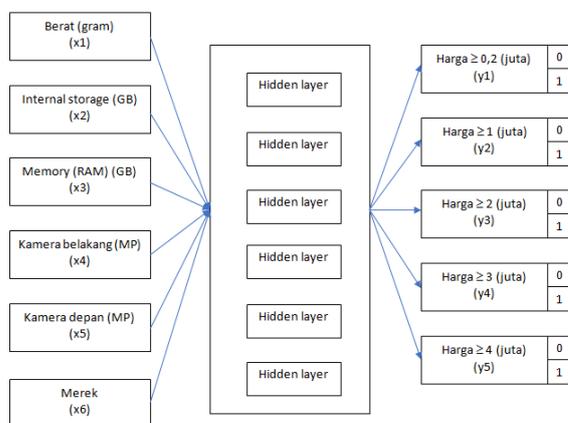
end

2.3 Konstruksi pada Data

Pada klasifikasi harga smartphone di pasaran, terdapat dataset dari smartphone, dimana atribut yang digunakan sebagai input adalah : berat (gram), penyimpanan internal (GB), memory (RAM) (GB), kamera belakang (MP), kamera depan (MP) dan merek seperti pada Gambar 3. Atribut tersebut akan digunakan sebagai input pada algoritma BP untuk menghasilkan output, yaitu kelas harga smartphone. Sedangkan pada LVQ, akan diambil sampel tiap kelas sebagai vektor reference. Kelas harga smartphone yang digunakan adalah :

1. Kelas 1 (harga smartphone di atas Rp. 200.000)
2. Kelas 2 (harga smartphone di atas Rp. 1.000.000)
3. Kelas 3 (harga smartphone di atas Rp. 2.000.000)
4. Kelas 4 (harga smartphone di atas Rp. 3.000.000)
5. Kelas 5 (harga smartphone di atas Rp. 4.000.000)

Pada klasifikasi multi output, pada jaringan BP terdapat layer output lebih dari satu seperti pada Gambar 3, yaitu harga di atas Rp. 200.000, harga di atas Rp. 1.000.000, harga di atas Rp. 2.000.000, harga di atas Rp. 3.000.000, dan harga di atas Rp. 5.000.000.



Gambar 3. Konstruksi Data pada Jaringan Backpropagation (BP)

Satu obyek bisa memiliki lebih dari satu output. Contoh smarphone harga Rp. 2.500.000, masuk ke dalam Kelas 3, yaitu di atas Rp.

2.000.000. Namun juga pasti masuk dalam Kelas 1 (di atas Rp. 200.000) dan Kelas 2 (di atas Rp. 1.000.000), tapi tidak untuk Kelas 4 (di atas Rp. 3.000.000) dan Kelas 5 (di atas Rp. 4.000.000), sehingga output dari smarphone tersebut adalah [11100].

Pada klasifikasi, performace yang digunakan adalah accuracy, error rate, recall, specificity, dan precision dengan rumus sebagai berikut [2] :

$$accuracy = \frac{TP + TN}{P + N} \times 100\% \quad (19)$$

$$error\ rate = \frac{FP + FN}{P + N} \times 100\% \quad (20)$$

$$recall = \frac{TP}{P} \times 100\% \quad (21)$$

$$specificity = \frac{TN}{N} \times 100\% \quad (22)$$

$$precision = \frac{TP}{TP + FP} \times 100\% \quad (23)$$

Dengan penjelasan :

TP : Jumlah tuple positif yang diprediksi benar sebagai positif oleh classifier

TN : Jumlah tuple negatif yang diprediksi benar sebagai negatif oleh classifier

FP : Jumlah tuple negatif yang diprediksi salah sebagai positif oleh classifier

FN : Jumlah tuple positif yang diprediksi salah sebagai negatif oleh classifier

P : Jumlah tuple positif pada data target

N : Jumlah tuple negatif pada data target

3. Hasil Dan Pembahasan

Data smartphone yang digunakan diperoleh dari daftar smartphone yang dijual di pasaran pada Februari - Maret 2020 dengan 80 jenis smartphone. Kemudian klasifikasi ke dalam Kelas 1 (harga smartphone di atas Rp. 200.000), Kelas 2 (harga smartphone di atas Rp. 1.000.000), Kelas 3 (harga smartphone di atas Rp. 2.000.000), Kelas 4 (harga smartphone di atas Rp. 3.000.000), dan Kelas 5 (harga smartphone di atas Rp. 4.000.000)

Sebelum menerapkan estimasi dan proses klasifikasi, bagi data ke dalam data training dengan proporsi 80% dari semua data dan data testing dengan proporsi 20% sisanya. Kemudian normalisir data sehingga berada di antara 0-1 [2].

3.1 Klasifikasi pada Learning Vector Quantization

Parameter yang digunakan pada simulasi LVQ adalah :

Learning rate α : 0.6

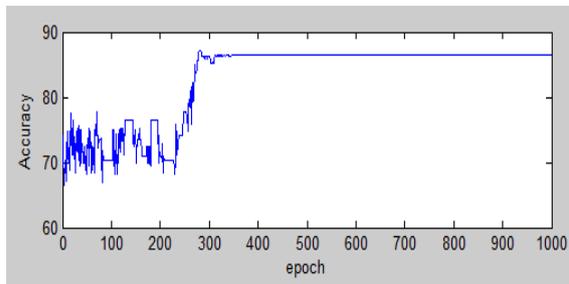
Learning rate δ : 0.9

Maximum epoch : 1000

Simulasi LVQ pada klasifikasi kelas harga smartphone dapat dilihat pada Gambar 4. Pertama-tama, inialisasi matriks bobot dari lima kelas :

1. Kelas 1 (harga smartphone di atas Rp. 200.000)
2. Kelas 2 (harga smartphone di atas Rp. 1.000.000)
3. Kelas 3 (harga smartphone di atas Rp. 2.000.000)
4. Kelas 4 (harga smartphone di atas Rp. 3.000.000)
5. Kelas 5 (harga smartphone di atas Rp. 4.000.000)

Gunakan pada data training menggunakan LVQ sampai maksimum epoch dan proses konvergensi dapat dilihat pada Gambar 4. Terlihat pada awal epoch, accuracy pada persamaan (19) yang dihasilkan cukup kecil. Pada proses optimisasi, accuracy meningkat dan konvergen dengan accuracy 86.5625 % pada data training. Kemudian matriks bobot optimal pada data training digunakan pada data testing dan menghasilkan accuracy 71.25 % pada data testing.



Gambar 4. Proses Optimisasi Accuracy Algoritma LVQ pada Data Training

Sehingga performance dari algoritma LVQ pada data training dan data testing seperti pada Tabel 1.

Tabel 1. Performance Klasifikasi pada Algoritma LVQ

Training Data (64 data)	
Accuration	86,5625 %
Error rate	13,4375 %
Recall	93,4884 %
Specificity	72,3810 %
Precision	87,3913 %
Testing Data (16 data)	
Accuration	71,25 %
Error rate	28,75 %
Recall	68,5185 %
Specificity	76,9231 %
Precision	86,0465 %

3.2 Klasifikasi pada Backpropagation

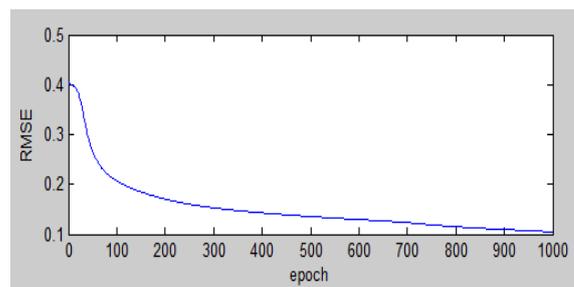
Parameter yang digunakan pada simulasi BP adalah :

Learning rate α : 0.2

Jumlah hidden layer : 2
Maximum epoch : 1000

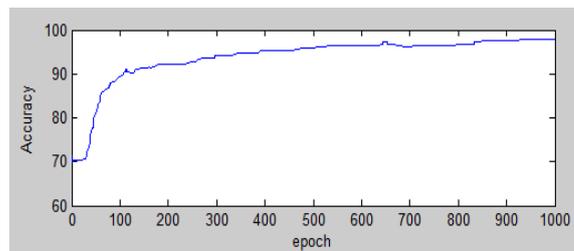
Simulasi BP pada klasifikasi kelas harga smartphome dapat dilihat pada Gambar 5 pada optimisasi nilai RMSE dan Gambar 6 pada optimisasi nilai accuracy.

Pada optimisasi nilai RMSE, inialisasi matriks bobot dan gunakan pada training data menggunakan BP sampai maksimum epoch dan proses konvergensi dapat dilihat pada Gambar 5. Terlihat pada awal epoch, RMSE pada persamaan (18) yang dihasilkan cukup besar. Pada proses optimisasi, RMSE menurun dan konvergen dengan nilai RMSE adalah 0,1048 pada data training. Kemudian matriks bobot optimal pada data training digunakan pada data testing dan menghasilkan RMSE 0,2293 pada data testing.



Gambar 5. Proses Optimisasi RMSE Algoritma BP pada Data Training

Pada optimisasi nilai accuracy, inialisasi matriks bobot dan gunakan pada training data menggunakan BP sampai maksimum epoch dan proses konvergensi dapat dilihat pada Gambar 6. Terlihat pada awal epoch, accuracy pada persamaan (19) yang dihasilkan cukup kecil. Pada proses optimisasi, accuracy meningkat dan konvergen dengan nilai accuracy adalah 97,8125 % pada data training. Kemudian matriks bobot optimal pada data training digunakan pada data testing dan menghasilkan accuracy 90 % pada data testing.



Gambar 6. Proses Optimisasi Accuracy Algoritma BP pada Data Training

Sehingga performance dari algoritma BP pada data training dan data testing seperti pada Tabel 2.

Tabel 2. Performance Klasifikasi pada Algoritma BP

Training Data (64 data)	
Accuration	97,8125 %
Error rate	2,1875 %
Recall	96,7442 %
Specificity	100 %
Precision	100 %
Testing Data (16 data)	
Accuration	90 %
Error rate	10 %
Recall	88,8889 %
Specificity	92,3077 %
Precision	96 %

4. Kesimpulan

Terdapat dua metode yang digunakan pada klasifikasi kelas harga smartphone di pasaran yaitu Learning Vector Quantization (LVQ) dan Backpropagation (BP). Pada proses klasifikasi menggunakan LVQ dan BP, terdapat perbedaan pada kedua metode tersebut. LVQ mengklasifikasi harga smartphone menggunakan jarak euclidean antara bobot (weight) dengan data di setiap iterasi. BP mengklasifikasi harga smartphone menggunakan penurunan gradien (gradient descent) dari target dan output di setiap iterasi. Berdasarkan hasil simulasi, BP memberikan accuracy yang lebih baik pada data training maupun data testing daripada LVQ. Pengembangan dari penelitian ini adalah teknik klasifikasi dengan big data menggunakan machine learning process.

Daftar Pustaka

- [1] L. Fausett, *Fundamental of Neural Networks*, Prentice Hall, New York, 1994
- [2] J. Han, M. Kamber, J. Pei, *Data Mining Concepts and Techniques*, Elsevier, New York, 2012
- [3] D. Rahmalia, T. Herlambang, "Prediksi Cuaca Menggunakan Algoritma Particle Swarm Optimization-Neural Network (PSO-NN)" *In Prosiding Seminar Nasional Matematika dan Aplikasinya*, pp. 41-48, 2017
- [4] D. Rahmalia, T. Herlambang, "Application Kohonen Network and Fuzzy C Means for Clustering Airports Based on Frequency of Flight," *Kinetik : Game Technology, Information System, Computer Network, Computing* ; vol. 3, no. 3, pp. 229-236. 2018.
- [5] A. Muhith, T. Herlambang, Irhamah, D. Rahmalia, "Estimation of Thrombocyte Concentrate (TC) and Whole Blood (WB) Using Unscented Kalman Filter," *International Journal of Advanced Science and Technology* ; vol. 9, no. 8. 2020.
- [6] D. Rahmalia, N. Aini, "Pengaruh Korelasi Data pada Peramalan Suhu Udara Menggunakan Backpropagation Neural Network," *Zeta Math Journal* ; vol. 4, no. 1, pp.1-6. 2018.
- [7] D. Rahmalia, M.S. Pradana, "Backpropagation Neural Network pada Data yang Tak Stationer," *Jurnal Riset dan Aplikasi Matematika (JRAM)* ; vol. 3, no. 1, pp. 32-42. 2019.
- [8] A. Rohmatullah, D. Rahmalia, M.S. Pradana, "Klasterisasi Data Pertanian di Kabupaten Lamongan Menggunakan Algoritma K-Means dan Fuzzy C Means," *Jurnal Ilmiah Teknosains* ; vol. 5, no. 2, pp. 86-93. 2020.
- [9] D. Rahmalia, A. Rohmatullah, "Pengaruh Korelasi Data pada Peramalan Kelembaban Udara Menggunakan Adaptive Neuro Fuzzy Inference System (ANFIS)," *Applied Technology and Computing Science Journal* ; vol. 2, no. 1, pp. 10-24. 2019.
- [10] D.F. Karya, P. Katias, T. Herlambang, D. Rahmalia, "Development of Unscented Kalman Filter Algorithm for Stock Price Estimation" *In Journal of Physics : Conference Series*, 2019
- [11] D. Rahmalia et al, "Comparison Between Neural Network (NN) and Adaptive Neuro Fuzzy Inference System (ANFIS) on Sunlight Intensity Prediction Based on Air Temperature and Humidity" *In Journal of Physics : Conference Series*, 2019

SYSTEMIC: Information System and Informatics Journal

ISSN: 2460-8092, 2548-6551 (e)

Vol 6 No 2 – Desember 2020

Pengenalan Karakter Huruf Braille dengan Metode *Convolutional Neural Network*Muhammad Fahmi Herlambang¹, Asep Nana Hermana², Kurnia Ramadhan Putra³^{1,2,3} Institut Teknologi Nasional Bandungmfahmiherlambang@gmail.com¹, asep_nana@itenas.ac.id², kurniaramadhan@itenas.ac.id³**Kata Kunci***Pengolahan Citra, Pengenalan Karakter Braille, Convolutional Neural Network.***Abstrak**

Karakter huruf Braille terdiri dari 6 titik yang dirancang sedemikian rupa sehingga menjadi sebuah sistem penulisan yang dapat digunakan untuk membantu tunanetra. Akan tetapi, membaca huruf Braille tidaklah mudah karena selain harus memahami huruf Braille tersebut juga dibutuhkan sensitivitas jari yang cukup agar dapat membaca huruf Braille. Adapun penelitian tentang pengenalan huruf Braille menggunakan teknologi kecerdasan buatan, salah satunya *deep learning*. Metode *deep learning* yang digunakan adalah *Convolutional Neural Network (CNN)*. Metode *CNN* dapat digunakan dalam pengenalan wajah, analisis dokumen, klasifikasi gambar, dan sebagainya. Pada penelitian ini, metode *CNN* digunakan untuk pengenalan karakter huruf Braille. Sistem melakukan proses pengenalan karakter huruf braille per karakter dengan model yang sudah dilatih dengan dataset dengan 26 karakter. Hasil yang didapat mencapai tingkat akurasi 81.54% untuk citra karakter Braille yang diakuisisi dengan smartphone dengan kemiringan antara 0 hingga 4 derajat dan jarak 30cm dengan model training dengan *learning rate 0.0001* dan *optimizer Adam*.

Keywords*Image processing, Braille Character Recognition, Convolutional Neural Network.***Abstract**

Braille characters consists of 6 dots that are designed in such way to be a writing system to help blind people. However, learning or reading Braille characters isn't an easy thing to do, because fingers sensitivity and understanding the writing system are needed to be able to read Braille. Therefore, there are some researches on Braille characters recognition with different methods and technologies, such as *deep learning*. The *Convolutional Neural Network (CNN)* is used. *CNN* method has been used in various recognition researches, such as face recognition, document analysis, image classification, etc. In this research, the *CNN* method is used to perform Braille characters recognition. The system performs the Braille character recognition process per character based on a model that has been trained using a dataset with the 26 Braille characters. The result of 81.54% accuracy is achieved for Braille character image acquisition with a smartphone with 0 to 4 degrees tilting and 30cm distance with training model using *learning rate of 0.0001* and *Adam optimizer*.

1. Pendahuluan

Manusia pada umumnya dapat memperoleh informasi melalui dua cara, yaitu lisan dan tulisan. Informasi lisan biasanya didapatkan lewat televisi, video berita, orang lain, radio, dan lain-lain. Sementara, informasi tulisan dapat diperoleh lewat artikel, koran, surat, tulisan, dan lain-lain. Akan tetapi, bagi orang yang memiliki disabilitas seperti tunanetra memiliki kesulitan dalam memperoleh informasi. Oleh karena itu, dikembangkan jenis huruf yang dikhususkan untuk tunanetra, yaitu huruf Braille.

Huruf Braille terdiri dari 6 titik yang dirancang sedemikian rupa dan menjadi sebuah sistem penulisan yang dapat digunakan untuk membantu tunanetra. Akan tetapi, membaca huruf Braille tidaklah mudah karena selain harus memahami huruf Braille tersebut juga dibutuhkan sensitivitas jari yang cukup agar dapat membaca huruf Braille. Maka dari itu, dengan teknologi yang ada, terdapat penelitian tentang pengenalan huruf Braille menggunakan teknologi kecerdasan buatan, salah satunya *deep learning*.

Deep learning adalah salah satu kelas dari *Machine Learning*, yang merupakan salah satu

kelas dari *Artificial Intelligence*. *Deep Learning* terdiri dari banyak metode, beberapa yang paling sering digunakan adalah *Artificial Neural Network* (ANN), *Convolutional Neural Network* (CNN), *Multilayer Perceptrons* (MLP), dan *Recurrent Neural Network* (RNN). *Convolutional Neural Network* (CNN) merupakan salah satu metode deep learning yang mampu melakukan pengenalan gambar dan klasifikasi gambar. Metode CNN sudah banyak digunakan untuk klasifikasi wajah, plat nomor kendaraan, identifikasi jenis sampah, dan lain-lain.

Ada beberapa penelitian yang telah dilakukan mengenai karakter Braille, seperti yang dilakukan oleh (Joko Subur, dkk., 2015) [10] tentang pengenalan karakter Braille menggunakan metode ANN (*Artificial Neural Network*) menggunakan webcam untuk menangkap citra dengan hasil akurasi hingga 99%, akan tetapi kemiringan gambar mempengaruhi keakurasian, dimana akuisisi citra dengan kemiringan gambar yang terbatas pada 1 derajat. Oleh karena itu, dilakukan penelitian pengenalan karakter Braille dengan metode CNN (*Convolutional Neural Network*) dengan tujuan menguji akurasi pengenalan karakter jika akuisisi citra menggunakan *smartphone* dengan kemiringan 0 hingga 4 derajat. Untuk mempermudah pengguna, dirancang aplikasi berbasis *python*. Adapun, fitur konversi huruf Braille ke huruf latin dan ucapan. Diharapkan dengan menggunakan metode CNN, keakurasian pengenalan karakter Braille mencapai angka diatas 80% dengan kemiringan 0 hingga 4 derajat.

2. Metode Penelitian

2.1 Huruf Braille

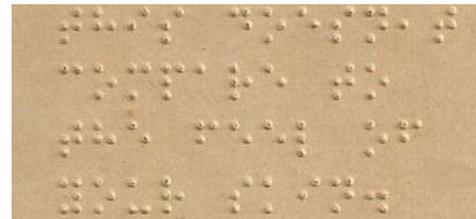
Huruf Braille merupakan suatu sistem untuk membaca dan menulis lewat sentuhan. Braille menggunakan karakter yang dibentuk lewat kombinasi enam titik timbul yang diatur dalam sel Braille dalam dua kolom vertikal dengan tiga titik pada kedua kolom dan dirancang sedemikian rupa dan menjadi sebuah sistem penulisan yang dapat digunakan untuk membantu orang tunanetra. Karakter Braille yang sederhana dibentuk oleh satu atau lebih titik dan menempati seluruh sel Braille. Untuk kemudahan, titik-titik pada sel Braille dirujuk dengan angka dan menyesuaikan dengan tombol pada Braillewriter.

Ada 3 metode untuk mentranskripsi Braille, yaitu Braillewriter, program pada komputer, dan Braille slate dan stylus. Braillewriter dan komputer bekerja dengan cara yang sama yaitu dengan menggunakan 6 tombol yang sesuai dengan sel Braille. Sementara, jika menggunakan Braille slate titik akan timbul pada sisi kertas yang satunya. Sehingga, penulisannya dilakukan dari kanan ke kiri, jadi saat kertasnya dibalik dapat dibaca dari kiri ke kanan. Maka dari itu, titik 1, 2,

dan 3 yang seharusnya ada di sebelah kiri, menjadi di sebelah kanan. Sementara, titik 4, 5, dan 6 yang seharusnya ada di sebelah kanan, menjadi di sebelah kiri. Sehingga, ketika kertasnya dibalik titik 1, 2, dan 3 akan ada di sebelah kiri dan sebaliknya untuk titik 4, 5, dan 6. Karakter huruf braille ditampilkan pada gambar 1. Contoh penulisan Braille ditampilkan pada gambar 2.

⠁	⠃	⠉	⠇	⠑	⠋	⠏	⠗	⠎	⠚
a	b	c	d	e	f	g	h	i	j
⠅	⠊	⠍	⠉	⠏	⠗	⠞	⠠	⠡	⠢
k	l	m	n	o	p	q	r	s	t
⠠	⠡	⠢	⠣	⠤	⠥	⠦			
u	v	w	x	y	z				

Gambar 1. Karakter Huruf Braille [11]



Gambar 2. Contoh Penulisan Karakter Huruf Braille [12]

2.2 Deep Learning

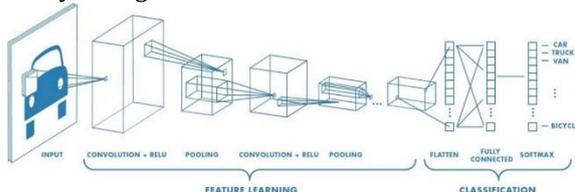
Deep Learning adalah salah satu fungsi kecerdasan buatan yang meniru cara kerja otak manusia dalam memproses data dan menciptakan pola untuk digunakan dalam pengambilan keputusan. *Deep Learning* merupakan bagian dari *Machine Learning*. *Deep Learning* terdiri dari banyak kelas atau jenis, beberapa yang paling sering digunakan adalah *Convolutional Neural Network* (CNN), *Multilayer Perceptrons* (MLP), dan *Recurrent Neural Network* (RNN).

Keunggulan *deep learning* adalah memiliki performa terbaik dalam menyelesaikan masalah kompleks, mengurangi kebutuhan untuk rekayasa ciri, dan memiliki arsitektur yang mampu beradaptasi terhadap permasalahan baru dengan mudah. Sementara, kekurangan dari *deep learning* adalah membutuhkan banyak data, proses *training* yang memakan waktu, dan *overfitting*.

2.3 Convolutional Neural Network

Convolutional Neural Network (CNN) merupakan salah satu kelas dari deep learning yang mampu melakukan pengenalan gambar dan klasifikasi gambar. Metode CNN merupakan suatu kelas pada neural network yang berspesialisasi dalam memproses data yang memiliki topologi seperti grid, misalnya gambar. Metode CNN dapat digunakan dalam pengenalan wajah, analisis dokumen, klasifikasi gambar, klasifikasi video, dsb.

Metode CNN mengklasifikasi gambar dengan memproses gambar yang diinput, kemudian mengklasifikasikannya pada kategori tertentu misalnya pada citra manusia, terdapat wajah, mata, bibir, hidung, tangan, dll. Gambar akan dibuat menjadi array berisi nilai pada setiap *pixel* dengan resolusi tinggi*panjang*dimensi yang disebut *channel*. Dimana, channel ini biasanya terdiri dari 3 buah yang berarti citra merupakan gambar RGB dengan masing-masing lapisan (*channel*) merepresentasikan *Red-Green-Blue* atau 1 lapisan jika gambar grayscale. Akan tetapi, jumlah lapisan juga bisa melebihi 3, bahkan hingga ratusan yang merepresentasikan berbagai warna lainnya dengan arsitektur RGB.



Gambar 1 Ilustrasi Arsitektur Convolutional Neural Network [13]

Seperti yang ditampilkan pada gambar 3, Arsitektur CNN dibagi menjadi dua bagian besar, yaitu *Feature Extraction Layer* dan *Fully Connected Layer*. Pada *Feature Extraction Layer* terjadi “*encoding*” dari sebuah citra menjadi *feature map* yang berupa angka-angka yang merepresentasikan citra tersebut (*Feature Extraction*). *Feature Extraction Layer* terdiri dari *convolution layer* dan *pooling layer*. *Convolution layer* merupakan bagian utama pada metode CNN yang menjadi pembeda dengan *neural network* lainnya. *Convolution layer* adalah lapisan pertama yang mengekstraksi ciri dari citra yang dimasukkan. *Convolution* menjaga hubungan antara *pixel* dengan mempelajari ciri citra menggunakan operasi matematis antara matriks citra dengan filter atau *kernel*. *Kernel* adalah sebuah operator yang diterapkan ke seluruh citra untuk mendapatkan nilai *array* dari sebuah citra. *Kernel* adalah matriks biasanya berukuran 3*3 atau 5*5 yang berisi nilai acak antara -1 dan 1. Hasil konvolusi matriks citra dengan filter (*kernel*) tersebut disebut *feature map*. Adapun, ReLU atau *Rectified Linear Unit* yang berfungsi untuk mengubah nilai negatif pada *feature map* menjadi positif. *Pooling layer* merupakan lapisan yang berfungsi untuk mengurangi ukuran spasial untuk mengurangi jumlah parameter dan perhitungan ketika ukuran citra terlalu besar. *Pooling layer* beroperasi pada setiap *feature map* secara independen.

Fully Connected Layer merupakan bagian dimana hasil *feature map* dari *feature extraction layer* yang berbentuk *array* multi-dimensi diubah menjadi vektor agar dapat dibentuk menjadi *fully connected layer* seperti sebuah jaringan syaraf (*neural network*). Kemudian, semua ciri yang sudah terbentuk menjadi jaringan syaraf

dikombinasikan untuk membuat suatu model. Lalu, dengan fungsi aktivasi seperti *softmax* atau *sigmoid* digunakan untuk mengklasifikasikan output misalnya, mobil, wajah, kucing, macan, dll. Singkatnya, cara kerja metode CNN yaitu, CNN akan melatih dan menguji setiap gambar melalui serangkaian proses. Dimulai dari pemecahan gambar menjadi gambar yang lebih kecil yang tumpang tindih, kemudian memasukkan setiap gambar yang lebih kecil ke *neural network* yang lebih kecil, menyimpan hasil dari masing-masing gambar kecil ke dalam *array* baru, *downsampling* atau mengurangi ukuran spasial untuk mengurangi jumlah parameter dan perhitungan ketika ukuran citra terlalu besar, dan membuat prediksi.

Kelebihan dari metode CNN adalah dapat secara otomatis mengekstraksi ciri penting dari setiap citra tanpa bantuan manusia, selain itu metode CNN juga lebih efisien dibandingkan metode *neural network* lainnya terutama untuk memori dan kompleksitas. Sementara, kekurangan dari metode CNN adalah membutuhkan banyak data latih, proses pelatihan (*training*) yang memakan waktu, dan *overfitting* atau karena terlalu banyak data latih maka algoritma kehilangan kemampuan untuk menggeneralisasi.

2.4 Metode Penelitian

Dalam penelitian ini penulis melakukan beberapa metodologi untuk memperoleh data, informasi dan melakukan penelitian. Berikut ini adalah metodologi yang dilakukan.

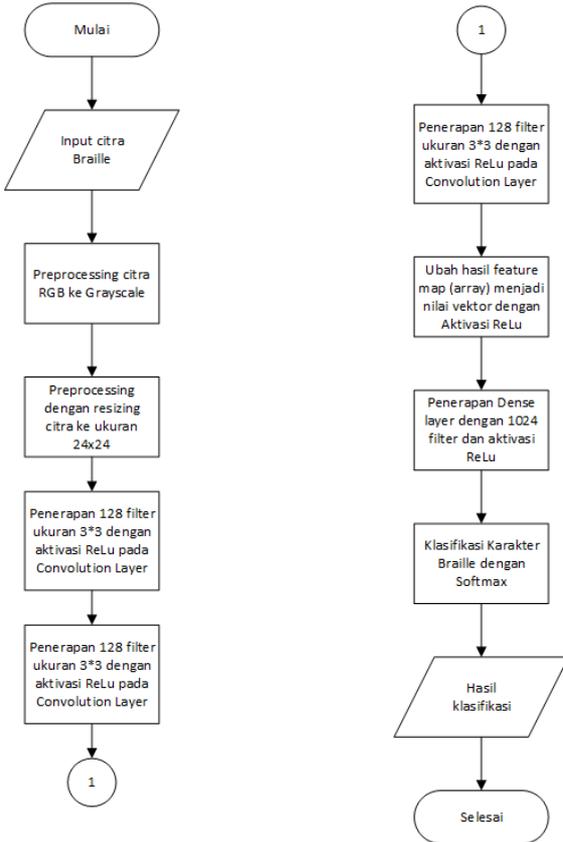
- Observasi: Penulis melakukan observasi atau pengamatan secara langsung untuk mempelajari, mengamati, dan mengumpulkan data serta informasi yang digunakan dalam penelitian ini.
- Studi Pustaka: Penulis melakukan penelitian kepustakaan untuk memperoleh aspek-aspek teoritis dalam pengumpulan data yang berhubungan dengan masalah yang ditinjau dalam penyusunan laporan penelitian ini.
- Studi Literatur: Proses pengumpulan data melalui berbagai jurnal dan penelitian yang terkait dengan penelitian ini.
- Eksperimen: Hasil dari ketiga tahap sebelumnya dan percobaan untuk membuat dan menguji model *Convolutional Neural Network* untuk pengenalan karakter huruf Braille.

2.5 Analisis Kebutuhan

Dalam proses perancangan sistem, diperlukan analisis untuk mengetahui kebutuhan sistem. Kebutuhan secara keseluruhan dalam melaksanakan perancangan sistem terbagi menjadi kebutuhan perangkat keras (*hardware*) dan kebutuhan (*software*).

2.6 Perancangan Sistem

Dalam proses perancangan sistem, dibuat *flowchart* sistem yang digambarkan pada gambar 4.



Gambar 4 Flowchart Sistem

Sistem melakukan pembuatan model yang diawali dengan *preprocessing* citra RGB menjadi *grayscale* untuk memperkecil ukuran *layer*, kemudian *resizing* ke ukuran 24x24. Lalu, penerapan arsitektur CNN mulai dari penerapan 128 filter ukuran 3x3 dengan aktivasi ReLu pada *Convolution Layer* sebanyak 3 kali, kemudian hasil *feature map* berupa array diubah menjadi nilai vektor dengan aktivasi ReLu, kemudian penerapan *Dense layer* dengan 1024 filter dan aktivasi ReLu. Kemudian, proses klasifikasi karakter menggunakan fungsi aktivasi softmax. Terakhir, hasil klasifikasi dikeluarkan oleh sistem.

2.7 Proses Pengumpulan Data Latih

Data latih diperoleh dari *website* github milik *user* HelenGezahegn[14], yang kemudian dirancang ulang agar sesuai dengan penelitian yang dilakukan. Data latih berjumlah total 20771, dengan rata-rata 798 untuk masing-masing 26 karakter dari A sampai Z.

2.8 Preprocessing

Pada proses pelatihan dan pengujian terdapat proses *preprocessing* yang digambarkan pada gambar 5.



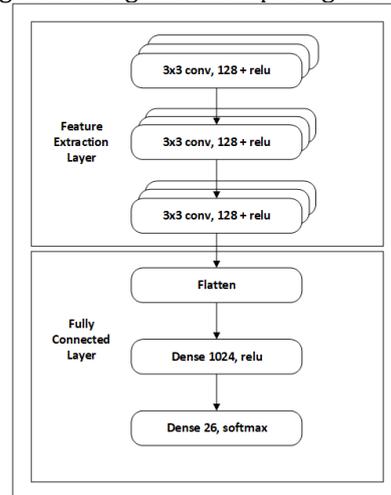
Gambar 5 Flowchart Sistem

Proses *preprocessing* terdiri dari:

1. Pendeteksian Citra RGB.
2. Proses *grayscale*, untuk memperkecil ukuran *layer* dan mempermudah komputasi.
3. Proses *resizing* atau perubahan ukuran citra menjadi 24x24

2.9 Proses Convolutional Neural Network

Setelah dilakukan *preprocessing*, lalu dilakukan penerapan metode CNN. Arsitektur CNN yang digunakan digambarkan pada gambar 6.



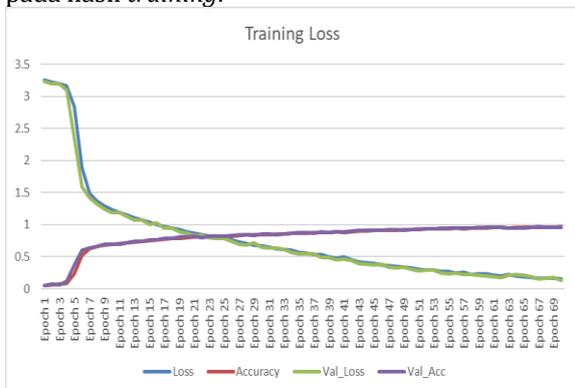
Gambar 6 Arsitektur Convolutional Neural Network

Pada arsitektur CNN yang digunakan untuk melakukan penelitian ini terdiri dari 3 buah *convolutional layer* dengan filter 3x3 dan fungsi aktivasi relu untuk mengubah semua nilai negative dalam array menjadi nol pada *Feature Extraction Layer*, dan pada *Fully Connected Layer* terdapat Flatten untuk mengubah array multi-dimensi menjadi vektor, dan 2 *dense layer* beserta fungsi aktivasi softmax untuk mengklasifikasi kategori citra yang sesuai.

3. Hasil Dan Pembahasan

3.1 Pengujian Training

Proses *training* dilakukan sebanyak 70 epoch dengan nilai *learning rate* 0.0001, *optimizer* Adam, dan fungsi aktivasi Softmax. Proses *training* dilakukan dalam kurun waktu sekitar 530s per *epoch*. Dari hasil *training* diperoleh nilai *loss* dan *accuracy*. Nilai *loss* adalah nilai yang digunakan untuk mengukur kinerja model saat proses *training* dan mengestimasi *error*. Pada gambar 7 adalah kurva perbandingan nilai *loss* dan akurasi pada hasil *training*.



Gambar 7 Kurva Perbandingan Nilai Loss dan Akurasi

Berdasarkan gambar 7, maka dapat dilihat bahwa nilai akurasi yang dicapai adalah 97.6%, nilai *loss* 1.3%, nilai *validation accuracy* 97.56%, dan nilai *validation loss* 1.3% dengan waktu rata-rata 27ms per langkahnya.

3.2 Pengujian Kinerja Sistem

Untuk mengukur kinerja sistem, dilakukan pengujian dengan menghitung nilai *confusion matrix*, yang terdiri dari nilai akurasi, presisi, *recall*, dan *F1 score*. Sistem dapat mengenali karakter braille yang didapatkan dengan menggunakan persamaan-persamaan berikut:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots (1)$$

$$Presisi = \frac{TP}{TP+FP} \times 100\% \dots\dots\dots (2)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \dots\dots\dots (3)$$

$$F1\ Score = 2 \times \frac{(Presisi \times Recall)}{(Presisi+Recall)} \dots\dots\dots (4)$$

Dimana:
 TP (*True Positive*), yaitu jumlah data positif yang terklasifikasi dengan benar oleh sistem.
 TN (*True Negative*), yaitu jumlah data negatif yang terklasifikasi dengan benar oleh sistem.
 FN (*False Negative*), yaitu jumlah data negatif namun terklasifikasi salah oleh sistem.
 FP (*False Positive*), yaitu jumlah data positif namun terklasifikasi salah oleh sistem.

Pengujian dilakukan dengan 5 tingkat kemiringan akuisisi citra, yaitu 0 derajat, 1 derajat, 2 derajat, 3 derajat, dan 4 derajat terhadap setiap karakter

dari A sampai Z, dengan total 130 pengujian. Pengujian yang dilakukan menggunakan kertas putih bertuliskan karakter huruf Braille berwarna hitam dengan jarak akuisisi citra 30cm. Berdasarkan hasil keseluruhan pengujian sistem maka dapat dihitung.

$$Akurasi = \frac{106 + 0}{106 + 0 + 24 + 0} = 81.54\%$$

$$Presisi = \frac{106}{106 + 24} \times 100\% = 81.54\%$$

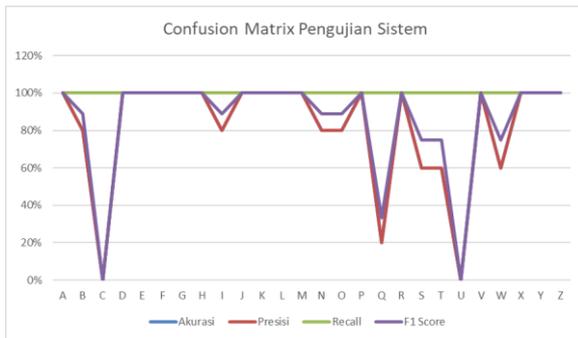
$$Recall = \frac{106}{106 + 0} \times 100\% = 100\%$$

$$F1\ Score = 2 \times \frac{(81.54\% \times 100\%)}{(81.54\%+100\%)} = 89.83\%$$

Hasil pengujian per karakter ditampilkan pada tabel 1 dan kurva perbandingan hasil pengujian ditampilkan pada gambar 8.

Tabel 1 Hasil Pengujian

No	Huruf	Akurasi	Presisi	Recall	F1 Score
1	A	100%	100%	100%	100%
2	B	80%	80%	100%	88.89%
3	C	0%	0%	100%	0%
4	D	100%	100%	100%	100%
5	E	100%	100%	100%	100%
6	F	100%	100%	100%	100%
7	G	100%	100%	100%	100%
8	H	100%	100%	100%	100%
9	I	80%	80%	100%	88.89%
10	J	100%	100%	100%	100%
11	K	100%	100%	100%	100%
12	L	100%	100%	100%	100%
13	M	100%	100%	100%	100%
14	N	80%	80%	100%	88.89%
15	O	80%	80%	100%	88.89%
16	P	100%	100%	100%	100%
17	Q	20%	20%	100%	33%
18	R	100%	100%	100%	100%
19	S	60%	60%	100%	75%
20	T	60%	60%	100%	75%
21	U	0%	0%	100%	0%
22	V	100%	100%	100%	100%
23	W	60%	60%	100%	75%
24	X	100%	100%	100%	100%
25	Y	100%	100%	100%	100%
26	Z	100%	100%	100%	100%



Gambar 2 Kurva Perbandingan Confusion Matrix Sistem

Berdasarkan hasil perhitungan secara keseluruhan sistem, didapatkan nilai akurasi sebesar 81.54%, nilai presisi 81.54%, nilai *recall* 100%, dan nilai *F1 score* 89.83% untuk semua 26 karakter. Sementara, berdasarkan tabel 1, dapat disimpulkan bahwa nilai confusion matrix dari masing-masing 26 karakter bervariasi, ada yang sudah mencapai tingkat keakurasian 100%, adapula yang mencapai keakurasian 0%. Kesalahan prediksi total terjadi pada huruf C dan U, dimana untuk kedua karakter tersebut tingkat keakurasian yaitu 0%, kemudian pada karakter huruf Q dicapai tingkat keakurasian sebesar 20% dan *F1 score* 33%. Selanjutnya, karakter huruf S, T, dan W mencapai tingkat akurasi 60% dan *F1 score* 75%. Lalu, untuk karakter huruf B, I, N, dan O mencapai tingkat akurasi 80% dengan *F1 score* 88.89%. Sementara untuk 16 karakter lainnya, mencapai tingkat akurasi, nilai presisi, recall, dan *F1 score* sempurna yaitu 100%. Berdasarkan hasil tersebut, maka dapat disimpulkan bahwa model ini dapat mengenali karakter huruf braille dengan baik untuk studi kasus karakter braille dengan akuisisi citra menggunakan *smartphone* dan tingkat kemiringan antara 0 hingga 4 derajat. Sementara pada beberapa karakter, yaitu B, C, I, N, O, Q, S, T, U, dan W masih belum optimal, hal ini dipengaruhi oleh beberapa hal seperti tingkat kemiringan, proses pemotongan citra yang tidak rapi sehingga citra tidak dapat diprediksi dengan benar. Selain itu, data latih yang digunakan juga mempengaruhi hasil prediksi, semakin banyak dan semakin variatif data latihnya maka akan semakin baik akurasinya.

4. Kesimpulan

Setelah dilakukan penelitian, maka dapat disimpulkan bahwa pengenalan karakter braille dengan metode *Convolutional Neural Network* memiliki tingkat akurasi sebesar 81.54%, nilai presisi 81.54%, nilai *recall* 100%, dan nilai *F1 score* 89.83% untuk studi kasus karakter braille dengan akuisisi citra menggunakan *smartphone* dan tingkat kemiringan antara 0 hingga 4 derajat dengan jarak 30cm menggunakan model *training* dengan *learning rate* 0.0001 dan *optimizer* adam.

Daftar Pustaka

- [1] Brownlee, Jason. (2019). What Is Deep Learning. Retrieved December 2019 from 'https://machinelearningmastery.com/what-is-deep-learning/'
- [2] Dertat, Arden. (2017). Applied Deep Learning - Part 4: Convolutional Neural Networks. Retrieved December 2019 from 'https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2'
- [3] Guo, G., & Zhang, N. (2019). A survey on deep learning based face recognition. *Computer Vision and Image Understanding*. 102805
- [4] Lee, S., Jung, S., & Song, H. (2018). CNN-Based drug recognition and braille embosser system for the blind. *Journal of Computing Science and Engineering*, 12(4), 149–156.
- [5] Mousa, A., Hiary, H., Alomari, R., & Alnemer, L. (2013). Smart braille system recognizer. *IJCSI International Journal of Computer Science Issues*, 10(6), 52–60.
- [6] Pokharna, Harsh. (2016). The Best Explanation of Convolutional Neural Networks on the Internet. Retrieved December 2019 from 'https://medium.com/technologymadeeasy/the-best-explanation-of-convolutional-neural-networks-on-the-internet-fbb8b1ad5df8'
- [7] Prabhu. (2018). Understanding of Convolutional Neural Network (CNN) – Deep Learning. Retrieved December 2019 from 'https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148'
- [8] Qolbiyatul, Lina. (2019). Apa Itu Convolutional Neural Network?, [What is Convolutional Neural Network?]. Retrieved February 25, 2020 from 'https://medium.com/@16611110/apa-itu-convolutional-neural-network-836f70b193a4'
- [9] Sathe, P. (2019). Waste Segregation using Convolutional Neural Network. *International Journal for Research in Applied Science and Engineering Technology*, 7(4), 932–937.
- [10] Subur, J., Sardjono, T. A., & Mardiyanto, R. (2016). Braille Character Recognition Using Artificial Neural Network. *IPTEK Journal of Proceedings Series*, 2(1), 83–84.
- [11] Anonim. Retrieved July 3, 2020 from 'https://sites.google.com/site/sharingvision/louis-braile/braille-script'
- [12] Róžańska, Natt. Braille Poem Night Writing. Retrieved January 2020 from 'https://www.etsy.com/listing/111235779/braille-poem-night-writing'

- [13] Saha, Sumit. (2018). A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way. Retrieved December 2019 from 'https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53'
- [14] Gezahegn, Helen (2018). Optical Braille Recognition. Retrieved December 2019 from 'https://github.com/HelenGezahegn/ae-ye-alliance'

SYSTEMIC: Information System and Informatics Journal

ISSN: 2460-8092, 2548-6551 (e)

Vol 6 No 2 – Desember 2020

Rancang Bangun Kontrol Keasaman pH Tanaman dalam Sistem Hidroponik menggunakan Kontrol PID Berbasis AndroidIka Indriani Retna Wardani ¹, Kunto Eko Susilo²^{1,2}) Universitas Narotama, SurabayaIkaindrianiiretnawardani.17@fasilkom.narotama.ac.id¹, kunto.eko.susilo@narotama.ac.id²**Kata Kunci***Hidroponik, IoT, MQTT, PID, Android***Abstrak**

Semakin padatnya keadaan bumi dengan manusia, tentunya berpengaruh terhadap tersedianya lahan di permukaan bumi. Tentu saja ini berpengaruh pada kestabilan bercocok tanam apalagi di daerah perkotaan. Hidroponik adalah salah satu solusi pintar yang bisa diambil untuk bercocok tanam di lahan sempit atau bahkan tanpa lahan tanah. hidroponik sangat membutuhkan lingkungan yang terkontrol untuk menghindari penurunan kualitas tanaman hingga layunya tanaman. Salah satu parameter yang penting adalah derajat keasaman (pH) larutan nutrisi, dimana pH sangat berpengaruh terhadap kelarutan dan ketersediaan unsur hara dalam larutan nutrisi. Perkembangan IoT (Internet of Things) saat ini sudah mencakup berbagai aspek salah satunya dalam dunia pertanian, dengan protokol MQTT Broker pemantauan dan pengukuran pH tanaman pada sistem hidroponik dapat dilakukan secara jarak jauh. Metode yang dilakukan adalah menggunakan PID kontrol dimana pH tanaman akan mencapai set point yang ditentukan dengan stabil. Hasil yang diharapkan adalah dapat membantu untuk melakukan kontrol pH tanaman secara jarak jauh melalui android untuk mencegah penurunan kualitas tanaman.

Keywords*Hydroponic, IoT, MQTT, PID, Android***Abstract**

The more dense condition of the earth with humans, of course, affect the availability of land on the surface of the earth. Of course this affects the stability of farming especially in urban areas. Hydroponics is one of the smart solutions that can be taken to grow crops in narrow areas or even without land. hydroponics really need a controlled environment to avoid decreasing the quality of plants until the plants wither. One important parameter is the degree of acidity (pH) of the nutrient solution, where pH is very influential on the solubility and availability of nutrients in the nutrient solution. The development of the IoT (Internet of Things) currently covers various aspects, one of which is in agriculture, with the MQTT Broker protocol monitoring and measuring plant pH on a hydroponic system can be done remotely. The method used is to use PID control where the pH of the plant will reach a set point that is determined stably. The expected result is that it can help to control plant pH remotely through android to prevent deterioration in plant quality.

1. Pendahuluan

Hidroponik adalah salah satu cara untuk menanam tanaman dalam skala besar tanpa memerlukan lahan yang sangat cocok untuk dibudidayakan didaerah perkotaan.dimana semakin sempitnya lahan diperkotaan, hidroponik menjadi salah satu solusi. Karena hidroponik menganut sistem bercocok tanam di atas air atau tanpa tanah. Dengan munculnya solusi seperti ini, tidak akan ada lagi halangan untuk bercocok tanam dan malas untuk mengkonsumsi sayur Sistem

Hidroponik tidak memerlukan lahan yang begitu luas, penggunaan lahan yang efisien,

tanaman lebih bersih, penggunaan pupuk dan air lebih efisien. Namun disisi lain, sistem hidroponik memerlukan ketelatenan, ketelitian dan pemantauan terus-menerus. Salah satu parameter yang penting adalah derajat keasaman (pH) larutan nutrisi, dimana pH sangat berpengaruh terhadap kelarutan dan ketersediaan unsur hara dalam larutan nutrisi. Penggunaan alat ukur manual yang sebenarnya menyita waktu jika si pemilik sedang sibuk dengan pekerjaan kantornya. Pemantauan dan pengelolaan pertanian secara manual dapat menciptakan banyak masalah selama musim hujan musim seperti irigasi dan pemupukan dan kekurangan tenaga kerja. Petani akan sulit tidur malam karena mereka harus pergi

DOI : [10.29080/systemic.v6i2.970](https://doi.org/10.29080/systemic.v6i2.970)This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

ke pertanian di malam hari dan memantau tanaman mereka[1].

Dengan masalah ini, sangat penting untuk merancang pemantauan dan kontrol cerdas secara real-time sistem. Kontrol real-time dari perangkat dunia nyata telah menunjukkan nilai komersial dan memudahkan konsumen. Informasi lingkungan real-time dapat dikumpulkan dari jarak jauh dan ditransfer ke tempat itu dan dapat diproses untuk menemukan masalah, menyimpan data, atau mengambil tindakan yang diperlukan [2]. Maka dari itu diperlukan solusi untuk memantau kondisi tanaman dan kontrol secara otomatis jika kondisi tersebut tidak sesuai dengan kondisi yang diharapkan.

Pertanian terintegrasi meningkatkan hasil panen, daur ulang nutrisi mengarah pada penguatan ketahanan lingkungan. Sistem pertanian terintegrasi terdiri dari serangkaian praktik penghematan sumber daya yang bertujuan untuk mencapai laba tinggi yang dapat diterima dan tingkat produksi yang berkelanjutan, sambil meminimalkan efek negatif dari pertanian intensif dan melestarikan lingkungan[3].

Melalui manipulasi dan pengelolaan tanaman secara hati-hati yang tumbuh pada lingkungan meliputi jumlah air, tingkat pH dan kombinasi nutrisi spesifik, tanaman dapat didorong untuk tumbuh lebih cepat[4].

Ukuran pH pada umumnya berkisar antara 0 hingga 14. Cairan akan sama jika kurang dari 7 dan akan basa jika lebih dari 7. Nilai pH 7 artinya cairan tersebut netral. Jika air terdapat banyak senyawa asam maupun basa dapat mengakibatkan korosif dan bisa mengganggu kesehatan tanaman. Rentan pH air yang diperbolehkan untuk larutan nutrisi hidroponik antara 5,5 hingga 7,5 [5] [6].

Dengan memanfaatkan kemajuan ilmu masa kini yang telah memudahkan pekerjaan manusia dilakukan pengembangan pada sistem hidroponik yaitu dengan mengontrol pH tanaman sayur media tanam hidroponik menggunakan Arduino sebagai mikrokontroler. Untuk mengatasi pada sisi kelemahan hidroponik maka dibuat rancang bangun kontrol keasamaan pH menggunakan kontrol PID berbasis android. Nutrisi akan di pantau melalui sistem yang bernama PID Control, dimana pH tanaman akan di kontrol agar mencapai kestabilan derajat keasamaan pH tanaman[7].

Pengontrol PID (Proportional Integral Derivative) terdiri dari tiga komponen, yaitu komponen Proporsional (P), komponen Integral (I) dan komponen Derivatif (D). Pengontrol PID akan menghasilkan aksi kontrol dengan membandingkan kesalahan atau error yang merupakan selisih dari process variable dan set point sebagai masukan [8][9].

Melakukan kontrol terhadap pH tanaman dengan protokol MQTT dimana MQTT merupakan jembatan antara server dengan pengguna. MQTT adalah arsitektur Client-Server dengan topologi

hub-and-spoke dan sensor tujuan bertindak sebagai client yang membuat koneksi TCP ke server disebut Broker, yang bertanggung jawab untuk mengirimkan data antara pesan yang merupakan Publisher dan Subscriber. Keunikan utama dari kegunaan dari MQTT ini adalah dapat dilakukan integrasi dengan baik ke dalam perangkat berbasis Internet of Things ,karena tidak memerlukan modifikasi pada MQTT brokers atau clients [10] [11].

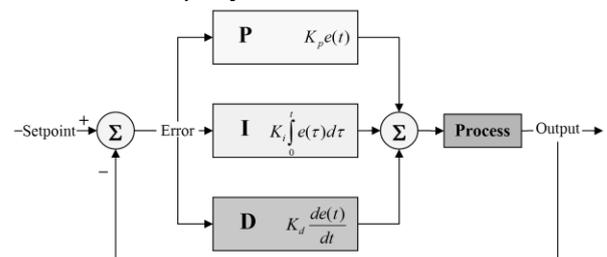
Selanjutnya data yang diterima akan ditampilkan pada sebuah aplikasi pada MQTT Dashboard. Ketika pH dimasukkan pada set point, otomatis mikrokontroler akan bekerja melakukan setting pH tanaman seperti set point yang diberikan.

2. Metode Penelitian

2.1 Teori PID

PID (Proportional, Integral dan Derivative) merupakan pengontrol konvensional yang merupakan bentuk sistematis yang baik karena dapat mengubah error menjadi nol, yaitu dengan menyamakan proses variabel sama dengan set point sehingga kestabilan pengontrolan dapat tercapai.

Setiap kekurangan dan kelebihan dari masing-masing kontroler P, I dan D dapat saling menutupi dengan menggabungkan ketiganya secara paralel menjadi kontroler proporsional plus integral plus differensial (kontroler PID). Nilai P dihasilkan dari nilai error yang masuk dalam sensor, kemudian nilai I merupakan akumulasi dari error sebelumnya, kemudian D akan memprediksi kesalahan selanjutnya.



Gambar 1. Blok Diagram Kontrol PID

1. Kendali Proporsional

Pengendali proporsional atau Gain bertindak sebagai penguat yang mampu mengubah output dari sistem secara proporsional tanpa memberikan efek dinamik pada kinerja pengendali tersebut. Kendali tersebut dinyatakan pada persamaan berikut :

$$u(t) = K_p e(t)$$

2. Kendali Integral

Pengendali integral atau Reset merupakan pengendali yang berfungsi untuk

memperbaiki respon steady state dari sistem sehingga kendali ini mampu memperbaiki error sistem. Kendali tersebut dinyatakan pada persamaan berikut:

$$u(t) = K_I \int_0^t e(t) . dt$$

3. Kendali Derivative

Pengendali derivative atau rate merupakan suatu kendali yang berfungsi untuk memperbaiki respontransien dari sistem. Kendali tersebut dinyatakan pada persamaan berikut :

$$u(t) = K_P \frac{d}{dt} e(t)$$

Pada pengontrolan PID, terdapat kontrol loop dengan cara kerja yaitu pada proses awal yaitu perhitungan dari proses variabel, sensor atau transmitter akan memperhitungkan beberapa variabel pada proses yang akan terjadi seperti temptertatur, suhu, tinggi, tekanan dll. Kemudian mengubah sinyal sensor kedalam transmisi ke sistem kontrol. Tahap kedua yaitu algoritma kontrol, algoritma matematika pada sistem kontrol akan mengeksekusi beberapa periode (setiap detik atau lebih) sinyal keluar untuk dilakukan transmisi ke elemen kontrol akhir. Tahap ketiga adalah elemen kontrol akhir, perangkat yang menerima sinyal dari kontroler dan memanipulasi proses, mengubah flow rate beberapa material. Tahap terakhir yaitu pada proses, proses merespon untuk mengubah variable manipulasi dengan perubahan hasil dan perhitungan variabel

2.2 Metode Perancangan Sistem

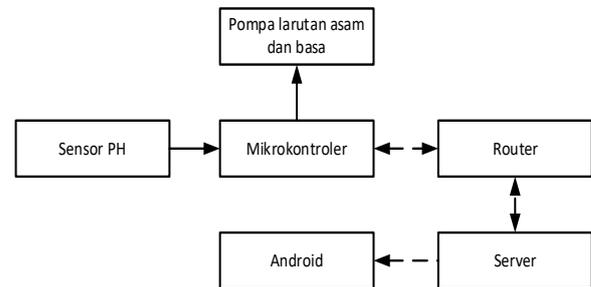
Pada perancangan penelitian ini dibutuhkan beberapa komponen seperti mikrokontroler yaitu Wemos D1 Mini, Sensor pH meter, pompa larutan asam dan basa, pengaduk, router dan android sebagai pemberi perintah.

A. Diagram Blok

Perancangan dan pembuatan alat pada penelitian ini dimudahkan dengan adanya Diagram Blok seperti pada Gambar 2. Pada diagram flow sistem ini di mana mikrokontroler melakukan perintah untuk mengirim data kepada server dan juga pompa larutan asam dan basa. Di mana pH tanaman yang belum mencapai set point sesuai masukan akan dilakukan penambahan larutan asam atau basa sesuai kebutuhan. Server dengan mikrokontroler disambung dengan supply internet dari router. Server yang digunakan disini adalah VPS di mana VPS sendiri dapat dengan mudah digunakan karena pengguna dapat mengaksesnya dimanapun dan kapanpun data pada mikrokontroler dikirimkan kepada server dengan protokol MQTT itulah mengapa pengguna atau user dapat dengan mudah mengaksesnya.

Pada penelitian ini hal yang paling utama terletak pada sensor pH dimana sensor akan melakukan pengecekan pH tanaman. Ph tanaman

harus dilakukan kontrol secara berkala karena hal ini mempunyai fungsi penting untuk melakukan cek apakah tanaman dalam keadaan baik atau tidak. Wemos D1 Mini dapat membaca sensor analog yang dikirimkan oleh sensor pH dan merubahnya kedalam digital kemudian dilakukan kontrol terhadap pompa larutan asam dan basa.



Gambar 2. Diagram Blok

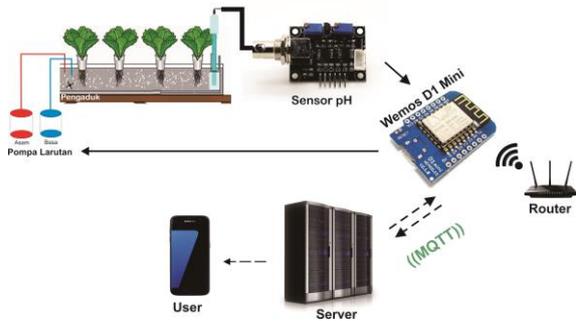
B. Perancangan Software dan Hardware Analisa Permasalahan

Dalam perancangan sistem ini ada beberapa hal yang perlu diperhatikan dalam hal program aplikasi hasil dari monitoring akan ditampilkan pada MQTT dashboard atau aplikasi di mana data yang dikirim dari mikrokontroler bersifat real time.

Perancangan software disini adalah merancang sebuah program yang dimasukkan dalam mikrokontroler. Pemrograman dibuat menggunakan bahasa pemrograman dengan program Arduino IDE. Alur jalannya program pada sistem ini adalah dengan melakukan pengukuran pada larutan cairan asam dan basa pada tanaman hidroponik, hasil pengukuran akan digunakan sebagai masukan pada mikrokontroler untuk menghidupkan pompa larutan asam atau basa. Sistem akan berlangsung secara terus menerus sampai nilai pH larutan sesuai dengan setpoint yang ditentukan

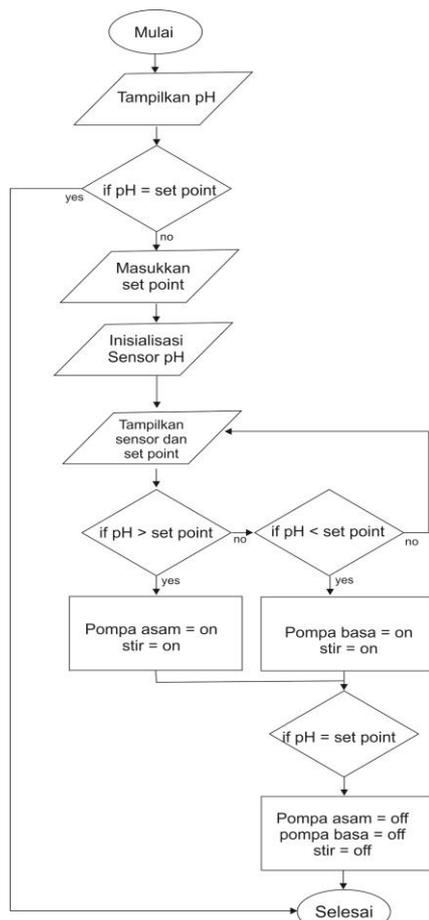
Pembuatan Hardware

Pada rancang bangun ini pada awalnya pengguna akan melakukan monitoring pH tanaman dengan menggunakan protokol MQTT di mana aplikasi telah terinstal pada smartphone ketika pH tanaman tidak pada set point yang dimasukkan pengguna melakukan set point kemudian dikirimkan untuk diterima oleh mikrokontroler menggunakan protokol MQTT kemudian akan dilakukan penambahan cairan larutan tergantung oleh set point pada pH tanaman setelah cairan diberikan pada tanaman akan dilakukan pengadukan. Kemudian pH meter akan melakukan pengecekan kembali untuk memastikan pH benar pada set point. Setelah selesai maka pompa larutan asam atau basa dan pengaduk otomatis akan mati.



Gambar 3. Desain Perancangan Alat

Berdasarkan Gambar 3 di atas menunjukkan Wemos D1 Mini mempunyai peran penting dalam melakukan kontrol dan monitoring tanaman.



Gambar 4. Diagram Alir Proram

Wemos D1 Mini akan menerima data analog dari pH meter dimana akan diterjemahkan ke dalam data digital oleh Wemos D1 Mini, kemudian diteruskan kepada server menuju user atau pengguna. Jembatan antara server dan Wemos D1 Mini menggunakan protokol MQTT digunakan untuk mensuplai internet untuk Wemos D1 Mini. Protokol MQTT menyusun konsep publish atau subscribe yang membutuhkan bandwidth relatif kecil setelah ditangkap oleh server maka akan diteruskan ditampilkan pada aplikasi. Pompa larutan asam dan basa akan menyala ketika pH

tanaman lebih set point, maka pompa larutan asam akan menyala, sebaliknya jika pH tanaman kurang dari set point, artinya pH tanaman sedang asam sehingga pompa larutan basa akan menyala. Pengaduk akan bekerja setelah penambahan cairan, setelah pH tanaman sesuai dengan set point, maka sistem akan selesai.

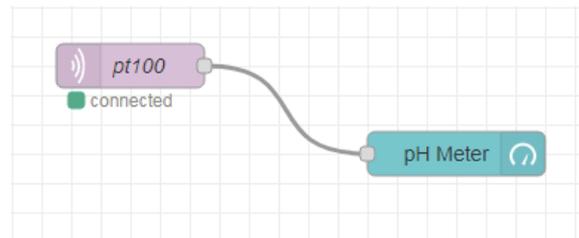
Sedangkan diagram alir pada Gambar 4 menjelaskan bahwa ketika pH tanaman tidak mencapai set point yang telah ditentukan, sistem akan terus bekerja secara sistematis sehingga mencapai set point yang dibutuhkan. Namun jika pH tanaman sudah mencapai set point yang ditentukan, sistem otomatis akan mati.

3. Hasil dan Pembahasan

Pada bagian ini, jelaskan hasil dari penelitian dan berikan penjelasan secara komperhensif dari hasil penelitian anda. Hasil dapat disajikan dalam bentuk grafik, tabel, gambar sehingga para pembaca dapat dengan mudah memahami hasil penelitian.

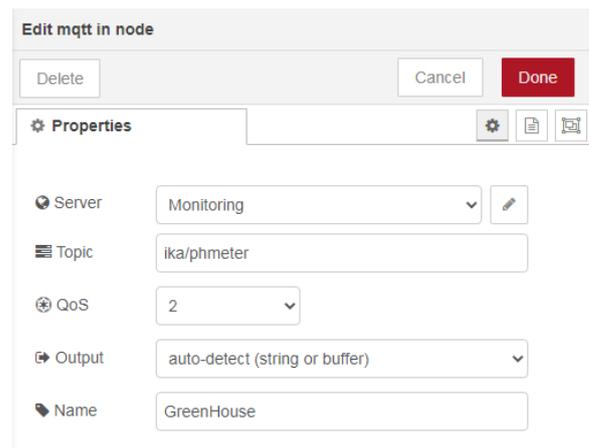
A. Konfigurasi pada MQTT

Pengujian ini memiliki tujuan memastikan bahwa sudah terpasangnya protokol MQTT agar berjalan sesuai dengan fungsinya saat nanti dijalankan.



Gambar 5. Node-red

Pada bagian ini akan ditunjukkan untuk konfigurasi sensor pH meter pada dashboard node-red. Gambar 5. merupakan gambar untuk konfigurasi dashboard ph meter. Dimana untuk sensor ph memiliki inisialisasi topik yaitu ika/phmeter.



Gambar 6. Edit node-red

B. Pengujian Sensor

Pada pengujian sensor ini dilakukan dengan tujuan untuk meastikan tingkat akurasi sensor ketika di jankan secara manual dan otomatis.

Pengujian ini dilakukan dengan dimasukkan pH buffer dengan nilai ph 5.0 perbedaan terjadi pada pengujian melalui dua ph-meter yaitu dengan Analog dan digital

Table 1. Hasil Kalibrasi Sensor pH

Sample	pH Meter Analog	pH Meter Digital
pH 5,0	5,0	5,0
pH 7,1	7,0	7,0

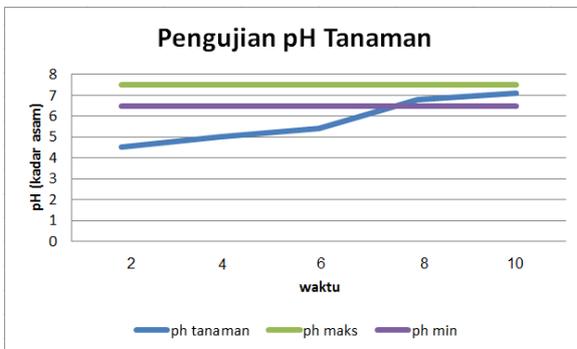
Pada pengujian selanjutnya, dimasukkan lebih banyak cairan buffer secara berkala sehingga didapat data sebagai berikut

Table 2. Hasil pengujian Sensor pH

Ph Meter Analog	pH Meter Digital	Selisih
5,0	5,0	0
5,3	5,4	0,1
5,5	5,4	0,1
6,0	6,0	0
6,5	6,2	0,3

Selisih paling tinggi yang dihasilkan pengujian pada tabel 2 sensor ph dalam percobaan adalah 0,3 sedangkan nilai tertentu adalah nol dapat disimpulkan berdasarkan tabel sensor yang digunakan dalam keadaan baik.

Pengujian Kadar Asam (pH) Tanaman

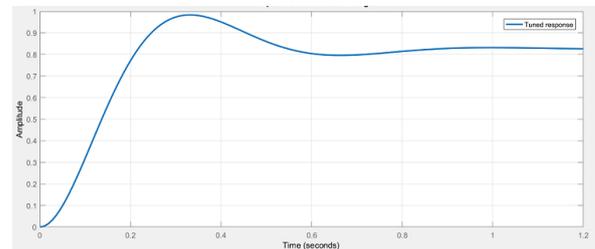


Gambar 4. Diagram Alir Proram

Dalam melakukan pengontrolan terhadap kadar pH air, saat cairan asam dan basa ditambahkan, sistem pengontrolan ini perlu menunggu beberapa saat agar cairan asam dan basa ercampur merata didalam air agar dapat diketahui perubahan kadar pH yang terjadi setelah ditambahkan cairan asam dan basa.

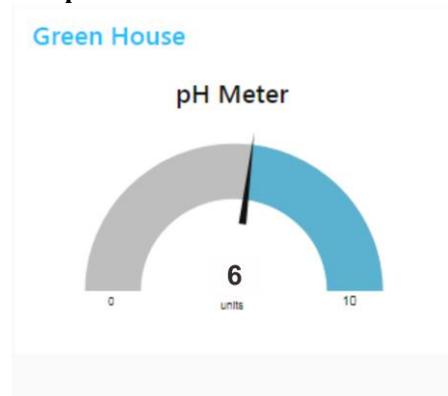
Agar pH tetap pada steady state nya inilah peran PID kontrol untuk melakukan sesuai fungsinya.

Ketika nilai sensor pH semakin mendekati nilai set point, maka respon nilai PID menjadi semakin kecil dan mendekali nilai titik nol, semakin besar nilai seslisih error antara nilai set point maka nilai sensor akan semakin menjauhi titik nol. Hasil penjumlahan antara arameter P+I+D berfungsi sebagai parameter pembanding terhadap parameter proses.



Gambar 4. PID Kontrol

C. Tampilan Akhir Sistem



Gambar 4. Tampilan Dashboard

Dengan tampilan user interface yang mudah untuk dibaca, dapat memudahkan untuk dilakukan kontrol.

4. Kesimpulan

Dari hasil pengujian yang dilakukan sebelumnya dapat disimpulkan bahwa rancang bangun yang dibuat mendapat error kecil antara 0-0,3 dari beberapa waktu yang di ambil. dengan pengujian pada penelitian ini diharapkan dapat memberi sebuah solusi untuk para petani konvensional yang masih menggunakan metode manual untuk bercocok tanam secara hidroponik.

Daftar Pustaka

[1] S. Rosminah, C. Zalina, and N. Ismail, "Increasing Productivity in Agriculture Through Integrated Smart Architecture of Irrigation Systems with LORA Technology," vol. 9, no. 11, pp. 264–274, 2019.

- [2] N. Che, S. Rosminah, and N. Omar, "IOT-Based Intelligent Green Houses (IGH) using Lo-Ra Technology," vol. 9, no. 11, pp. 274–283, 2019.
- [3] M. Suhana, A. Aziz, N. Che, and N. Ismail, "Lo-Ra Based Framework for Smart Green House Monitoring Systems," vol. 9, no. 11, pp. 251–263, 2019.
- [4] A. Ullah, S. Aktar, N. Sutar, R. Kabir, and A. Hossain, "Cost Effective Smart Hydroponic Monitoring and Controlling System Using IoT," pp. 142–154, 2019.
- [5] M. Fakhruzzaini and H. Aprilianto, "Sistem Otomatisasi Pengontrolan Volume Dan PH Air Pada Hidroponik," vol. D, pp. 1335–1344.
- [6] P. Studi, S. Teknik, F. Teknik, U. N. Surabaya, and M. S. Zuhrie, "RANCANG BANGUN SISTEM MONITORING DAN PENGONTROLAN PH NUTRISI PADA HIDROPONIK SITEM NUTRIENT FILM TECHNIQUE (NFT) MENGGUNAKAN PENGENDALI PID BERBASIS ARDUINO UNO Ahmad Nur Fuad."
- [7] Z. Buana, O. Candra, J. T. Elektro, F. Teknik, and U. N. Padang, "SISTEM PEMANTAUAN TANAMAN SAYUR DENGAN MEDIA," vol. V, no. 1, 2019.
- [8] F. Logic, "PERBANDINGAN SISTEM PENGONTROLAN PID KONVENSIONAL DENGAN PENGONTROLAN CMAC , FUZZY LOGIC DAN ANN PADA WATER LEVEL," vol. 17, no. 3, pp. 129–141, 2013.
- [9] J. A. Shaw, "The PID Control Algorithm How it works , how to tune it , and how to use it," 2003.
- [10] K. Kularbphetong, U. Ampant, and N. Kongrodj, "An Automated Hydroponics System Based on Mobile Application," vol. 9, no. 8, pp. 8–12, 2019.
- [11] P. Colombo and E. Ferrari, "Access Control Enforcement within MQTT-based Internet of Things Ecosystems," pp. 223–234, 2018.

SYSTEMIC: Information System and Informatics Journal

ISSN: 2460-8092, 2548-6551 (e)

Vol 6 No 2 – Desember 2020

Pemanfaatan Algoritma FP-Growth Untuk Menentukan Strategi Penjualan Pada Kedai Kopi Teras Garden

Adrian Marvel Ugrasena¹, Achmad Zakki Falani²^{1,2}) Universitas Narotama Surabayaadrian.marvel@mhs.fasilkom.narotama.ac.id¹, achmad.zakki@nanrotama.ac.id²

Kata Kunci

Association rule, fp-growth, pola pembelian.

Abstrak

Data mining merupakan proses pengambilan informasi dari sekumpulan data. Pada penelitian ini akan mencoba menerapkan data mining pada data transaksi di Kedai Kopi Teras Garden untuk strategi penjualan dengan cara membuat rekomendasi item mana yang cocok dan sesuai untuk dijual secara sistem paket atau dijual bersamaan. Penelitian ini menggunakan metode asosiasi rule dengan algoritma fp-growth untuk menemukan pola pembelian customer pada Kedai Kopi Teras Garden. Penelitian ini nantinya akan menghasilkan beberapa rule untuk digunakan sebagai rekomendasi penjualan sesuai dengan data yang sudah di kumpulkan dan diproses dengan metode association rule. Data akan dibagi menjadi 2 yaitu data pada musim kemarau dan musim hujan karena ada perubahan pola pembelian customer sesuai musim yang sedang berlangsung.

Keywords

Association rule, fp-growth, purchase pattern.

Abstract

Data mining is a process extracting data from the dataset. In this paper will try to apply data mining on data transaction in Kedai Kopi Teras Garden for sales strategy by creating recommendation item that suit to be sold by package system or the other words by to be sold simultaneously. This paper use association rule method with fp-growth algorithm to find customer purchase pattern on Kedai Kopi Teras Garden. The output of this paper will create some rules for recommendation item that can be sold as a package according to data that already collected and processed with association rule method. Data will be divided by 2, dry season data and rainy season data because there's customer pattern change accordingly with the season.

1. Pendahuluan

Perumahan *The Quality Garden* mempunyai sebuah kedai kopi yang berada didepan pintu masuk yang bernama Kedai Kopi Teras Garden. Kedai ini berlokasi di Gamping Wetan, Gamping, Kecamatan Krian, Kabupaten Sidoarjo, Jawa Timur. Banyak muda mudi yang menghabiskan waktu bersama dikedai ini sambil menikmati makanan dan minuman yang dijual dikedai ini. Kedai ini juga menyediakan *live music* setiap hari sabtu malam untuk meramaikan suasana Kedai Kopi Teras Garden. Pihak manajemen dari Kedai Kopi Teras Garden ingin menjual makanan dan minuman dengan sistem paket yang efektif atau dijual secara bersamaan dengan tujuan meningkatkan penjualan. Tetapi pihak manajemen belum memahami bagaimana cara menemukan kombinasi item mana yang cocok atau sesuai untuk dijual secara bersamaan. Tetapi pihak manajemen belum memahami bagaimana cara menemukan kombinasi item mana yang cocok atau sesuai untuk dijual

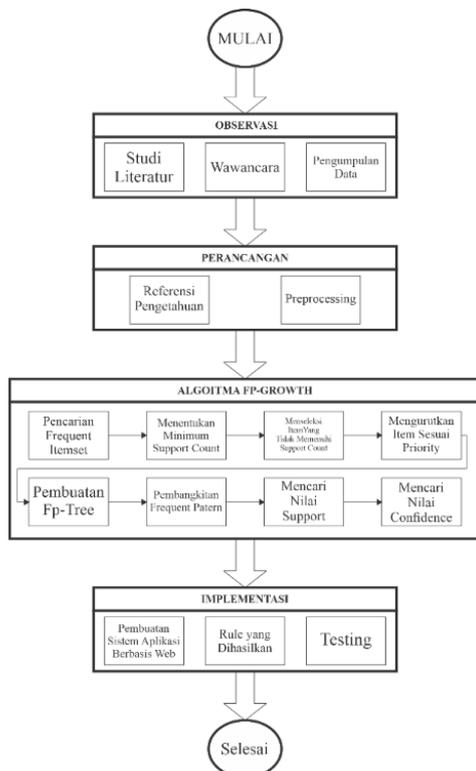
secara bersamaan. Pihak Manajemen perlu untuk merubah sistem yang tradisional ke sistem yang lebih professional[1].

Data mining biasanya digunakan untuk memproses data yang berukuran sangat besar untuk mengambil informasi dan pengetahuan dari data tersebut [2]. Data mining bisa juga digunakan untuk menemukan pola pembelian dari *customer* dalam membeli beberapa kombinasi barang dengan menggunakan metode tertentu. Data mining sering disebut juga *knowledge discovery in database* (KDD) adalah sebuah proses yang meliputi pengumpulan data, penggunaan data untuk menentukan pola atau hubungan dalam sebuah data yang besar [3]. Metode yang digunakan untuk menemukan pola pembelian dari *customer* yaitu *association rule mining*. Ada beberapa algoritma yang biasanya digunakan dalam metode ini, salah satunya adalah *fp-growth* (*frequent pattern growth*).

Dalam beberapa tahun yang lalu, *frequent pattern mining* digunakan secara signifikan untuk menemukan pengetahuan yang menarik yang tersembunyi dari sekumpulan data. Banyak algoritma yang telah dikembangkan untuk mempercepat proses pengambilan informasi [4]. Algoritma *fp-growth* adalah pengembangan dari algoritma apriori dalam menemukan kombinasi item yang tepat menurut data yang sudah di kumpulkan sebelumnya. Algoritma apriori yang diciptakan oleh Agrawal et al pada 1990 membutuhkan untuk membaca database beberapa kali dan menghasilkan jumlah kandidat kombinasi yang besar, Han et al menciptakan *fp-growth* yang mempunyai struktur data berupa *fp-tree* untuk memperbaiki kinerja algoritma apriori [5]. *Fp-growth* akan memakan banyak waktu dalam pembuatan *fp-tree* dan membangkitkan *frequent pattern*. Pada *association rule* dapat ditentukan dengan 2 nilai parameter yaitu *support* dan nilai *confidence* [6].

Data yang digunakan didapat dari pihak manajemen Kedai Kopi Teras Garden. Karena biasanya pola pembelian *customer* berubah sesuai dengan cuaca yang sedang berlangsung, maka data akan dibagi menjadi 2 yaitu data musim kemarau dan musim hujan. Sebagai contoh di musim kemarau, *customer* membeli lebih banyak minuman dingin dari pada minuman hangat. Begitu juga sebaliknya pada musim dingin *customer* membeli lebih banyak minuman hangat dari pada minuman dingin.

2. Metode penelitian



Gambar 1 Flowchart

Penelitian ini menggunakan metode *association rule* dengan algoritma *frequent pattern growth* untuk mencari kombinasi dari makanan dan minuman yang di jual di Kedai Kopi Teras Garden. Penelitian ini memilih tahapan atau proses yang bisa dilihat pada Gambar 1 untuk menggambarkan bagaimana jalannya penelitian ini.

2.1 Association Rule

Penelitian ini bertujuan untuk menemukan pola pembelian dari *customer* yang nanti digunakan untuk menemukan kombinasi item yang cocok atau sesuai dijual secara bersamaan di Kedai Kopi Teras Garden.

1. Observasi

Pada tahap ini, peneliti mengamati strategi penjualan apa yang sudah ada pada Kedai Kopi Teras Garden. Studi literature dilakukan untuk mengumpulkan jurnal yang digunakan sebagai referensi atau rujukan untuk penelitian ini. Wawancara dilakukan kepada pihak manajemen tentang bagaimana penjualan yang sedang berjalan, menu apa yang paling sering terjual dan menu mana yang biasanya dibeli *customer* secara bersamaan. Pengumpulan data juga dilakukan untuk menganalisa pola pembelian *customer*. Pengumpulan data dilakukan dengan cara meminta langsung kepada pihak manajemen, yang diperlukan untuk penelitian ini hanya data penjualan.

2. Perancangan

Dibutuhkan perancangan terlebih dahulu sebelum masuk ke pembuatan aplikasi untuk menganalisa pola pembelian *customer*. Dilakukan pengumpulan referensi pengetahuan berupa pengetahuan tentang bahasa pemrograman python beserta library *fp-growth* yang digunakan dalam penelitian ini. Selanjutnya tahap *preprocessing*, pada tahap ini data di olah terlebih dahulu sebelum dimasukan ke dalam proses analisa untuk hasil yang lebih optimal. Tahap *preprocessing* yang dilakukan pada penelitian ini adalah menghapus item yang sama pada satu pesanan atau transaksi untuk menghindari hasil rekomendasi item yang sama seperti jika membeli *ice tea* maka akan membeli *ice tea*. Hasil rekomendasi ini akan kurang relevan untuk dijadikan hasil penelitian.

3. Algoritma FP-Growth

Data yang telah melalui tahap *preprocessing*, akan dilakukan analisa pada tahap ini. Adapun beberapa langkah dari algoritma ini sebagai berikut.

a. Pencarian *frequent itemset*

Pada tahap pencarian *frequent itemset* ini, data yang sudah melalui tahap *preprocessing* sebelumnya akan dimasukan kedalam tabel yang mempunyai atribut nomer transaksi dan pesanan seperti pada Tabel 1.

Tabel 1 Daftar pesanan

No	Pesanan
1	Ice Tea,IceCapuccino,French Fries
2	Ice Tea,Lemon Tea,Kebab,French Fries
3	Ice Capuccino,French Fries,Donat
4	Ice Tea,Spagethi,French Fries
5	Ice Choco Oreo,Ice Choco Delight,French Fries
6	Hot Vietnam Drip,Pisang Coklat Keju
7	Lemon Tea,Hot Vietnam Drip,Jamur Crispy,Onion Ring
8	Hot Capuccino,Ice Coffe Latte,Donat
9	Ice Tea,Ice Choco Delight,Ice Choco Oreo,Onion Ring
10	Lemon Tea,ice Coffe Latte,French Fries,Pisang Coklat Keju

b. Menentukan *Minimum Support Count*

Perlu menentukan *minimum support count* untuk menyeleksi item yang kurang laku atau jarang dibeli untuk memudahkan tahap berikutnya. Pada tahap ini peneliti akan menentukan *minimum support count* sebesar 20%. Jadi item yang memiliki support count dibawah 20% akan di buang atau tidak akan digunakan untuk perhitungan kedepannya. Untuk menentukan support count digunakan persamaan(1), jumlah transaksi yang mengandung item A dibagi dengan total transaksi. Untuk mencari nilai support dari 2 item digunakan persamaan (2), Jumlah transaksi yang mengandung item A dan B dibagi dengan total transaksi.

$$Support(A) = \frac{Jumlah\ transaksi\ yang\ mengandung\ item\ A}{Total\ transaksi} (1)$$

$$Support(A \cap B) = \frac{Jumlah\ transaksi\ yang\ mengandung\ item\ A\ dan\ B}{Total\ transaksi} (2)$$

c. Menyeleksi Item Yang Tidak Memenuhi *Support Count*

Asumsi minimum support adalah 20% dari total transaksi pada penelitian ini, jadi minimum *support count* adalah $(20\% \times 10) = 2$. Karena minimum support adalah 2, maka item yang memiliki nilai *support* kurang dari 2 akan dihilangkan.

d. Mengurutkan Item Sesuai *Priority*

Pada tahap ini item akan di urutkan sesuai *priority* atau mengurutkan item dari frekuensi yang tertinggi ke frekuensi yang terendah. Seperti pada tabel 3 *french fries* mempunyai frekuensi tertinggi sebesar 6, makan *french fries* akan diletakan pada nomor 1.

Kedua adalah *ice tea* yang mempunyai frekuensi sebesar 4 maka diletakan nomor 2. Tahap ini akan dilakukan pada semua item yang memenuhi support count. Dan akan dikembalikan ke tabel data transaksi dengan sesuai urutan dari frekuensi yang terbesar ke frekuensi yang terkecil seperti pada tabel 4.

Tabel 2 *Support count*

Item	Frekuensi
Ice Tea	4
French Fries	6
Lemon Tea	3
Kebab	1
Ice Capuccino	2
Ice Choco Delight	2
Donat	2
Hot Capuccino	1
Spagethi	1
Ice Choco Oreo	1
Hot Vietnam Drip	2
Pisang Coklat Keju	1
Jamur Crispy	1
Ice Coffe Latte	1
Onion Ring	2
Ice Coffe Latte	1

Tabel 3 Yang sudah diseleksi dan di urutkan

No	Pesanan
1	French Fries,Ice Tea,IceCapuccino
2	French Fries,Ice Tea,Lemon Tea
3	French Fries,Ice Capuccino,Donat
4	French Fries,Ice Tea
5	French Fries,Ice Choco Delight
6	Hot Vietnam Drip
7	Lemon Tea,Hot Vietnam Drip,Onion Ring
8	Hot Capuccino,Donat
9	Ice Tea,Ice Choco Delight,Onion Ring
10	French Fries,Lemon Tea

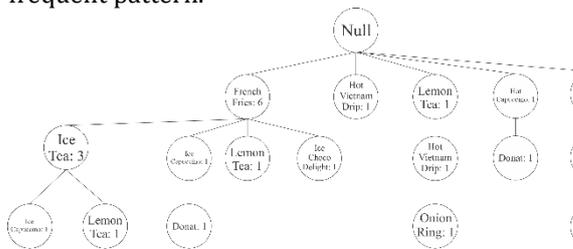
Tabel 4 Pesanan yang sudah di urutkan

Item	Frekuensi
French Fries	6
Ice Tea	4
Lemon Tea	3
Ice Capuccino	2
Ice Choco Delight	2
Donat	2
Hot Vietnam Drip	2
Onion Ring	2

e. Pembuatan *FP-Tree*

Setelah diurutkan sesuai *priority*, tahap berikutnya adalah membuat *fp-tree*. Sebuah *fp-tree* akan selalu dimulai dengan node null. Di dalam *fp tree* juga terdapat support count yang menandakan item melewati transaksi yang sama. Dan di isi mulai dari transaksi nomor 1. French Fries, Ice Tea dan Ice Capuccino diletakan dalam satu jalur kebawah. Transaksi nomor 2 French Fries, Ice Tea dan Lemon Tea. Tidak perlu membuat jalur baru untuk French Fries karena sudah ada sebelumnya, tinggal menambahkan support count pada French Fries sebelumnya. Ice Tea juga tidak membentuk jalur baru setelah French Fries, tinggal menambahkan support count pada Ice Tea sebelumnya. Lalu akan membuat jalur baru

setelah Ice tea karena Belum ada Lemon Tea setelah French Fries dan Ice Tea. Akan selalu membuat jalur baru jika belum ada yang melewati item tersebut dan jika sudah ada, tinggal menambahkan pada support count seperti pada gambar 2. Semua pesanan pada data transaksi yang sudah diurutkan semuanya akan ditulis pada fp tree ini untuk diproses ke tahap selanjutnya. Setelah semua pesanan pada data transaksi ditulis kedalam fp tree, maka akan dilanjutkan pada tahap pembangkitan frequent pattern.



f. Pembangkitan Frequent Pattern

Pada langkah ini semua *suffix* atau akhir dari semua jalur yang terdapat pada fp tree akan di tulis pada sebuah tabel beserta jalur yang telah di lalui.

Tabel 5 frequent pattern

Suffix	Frequent Itemset
Ice Capuccino	{Ice Capuccino, Ice Tea, French Fries}
Lemon Tea	{Lemont Tea, Ice Tea, French Fries}, {Lemon Tea, French Fries}
Hot Capuccino	{Hot Capuccino, Ice Tea, French Fries}
Ice Tea	{Ice Tea, French Fries}
Donat	{Donat, Ice Capuccino, French Fries}, {Donat, Hot Capuccino}
Ice Choco Delig	{Ice Choco Delight, French Fries}
Hot Vietnam Dr	{Hot Vietnam Drip}
Onion Ring	{Onion Ring, Hot Vietnam Drip, Lemon Tea}, {Onion Ring, Ice Choco Delight, Ice Tea}

Dari tabel ini didapatkan beberapa pattern, selanjutnya akan dicari nilai support dan nilai confidence untuk setiap pattern. Untuk mencari nilai confidence digunakan persamaan (3), Jumlah transaksi yang mengandung item A dan B dibagi dengan jumlah transaksi yang mengandung item A.

$$Confidence P(A|B) = \frac{\text{Jumlah transaksi yang mengandung A dan B}}{\text{Jumlah transaksi yang mengandung A}} \quad (3)$$

Jika semua sudah tertulis pada tabel maka akan memunculkan pattern seperti berikut.

- Pattern 1: {Ice Capuccino, Ice Tea, French Fries} support = 0.1 & confidence = 1
- Pattern 2: {Lemon Tea, Ice Tea, French Fries} support = 0,1 & confidence = 1
- Pattern 3: {Lemon Tea, French Fries} support = 0,2 confidence = 0,66
- Pattern 4: {Hot Capuccino, Ice Tea, French Fries} support = 0,1 & confidence = 1
- Pattern 5: {Ice Tea, French Fries} support = 0,3 & confidence = 0,75

- Pattern 6: {Donat, Ice Capuccino, French Fries} support = 0,1 & confidence = 1
- Pattern 7: {Donat, Hot Capuccino} support = 0,1 & confidence = 1
- Pattern 8: {Ice Choco Delight, French Fries} support = 0,1 & confidence = 1
- Pattern 9: {Hot Vietnam Drip} support = 0,1 & confidence = 1
- Pattern 10: {Onion Ring, Hot Vietnam Drip, Lemon Tea} support = 0,1 & confidence = 1
- Pattern 11: {Onion Ring, Ice Choco Delight, Ice Tea} support = 0,1 & confidence = 1

Dari perhitungan nilai support dan nilai confidence yang sudah dilakukan, didapatkan hasil 2 rule dengan nilai terbesar, yaitu.

- a. Rule 1, jika membeli Ice Tea maka akan membeli French Fries dengan nilai support mencapai 30% dan nilai confidence mencapai 75%.
- b. Rule 2, Jika Membeli Lemon Tea maka akan membeli French Fries dengan nilai support mencapai 20% dan nilai confidence mencapai 66%.

Dua rule ini akan dijadikan rekomendasi kombinasi item yang akan dijual secara bersamaan atau sistem paket karena 2 rule mempunyai kombinasi item yang cukup kuat.

3. Hasil dan Diskusi

Pada bagian ini akan menjelaskan hasil dan evaluasi dari metode yang digunakan dalam penelitian ini. Penelitian ini menggunakan jupyter lab untuk menganalisa data. Penelitian ini juga menggunakan Bahasa pemrograman python dan menggunakan library fp growth dari python.

3.1 Dataset

Data yang digunakan dalam metode ini adalah data yang berformat csv yang telah melalui tahap *preprocessing* sebelumnya untuk mendapatkan hasil yang optimal. Data didapat dari pihak manajemen Kedai Kopi Teras Garden langsung. Data akan dibagi menjadi 2 yaitu data musim kemarau dan musim hujan untuk menghasilkan hasil rekomendasi yang sesuai untuk musim kemarau dan musim hujan. Data pada musim kemarau sebanyak 5727 transaksi, sedangkan data pada musim hujan sebanyak 4371. Data musim kemarau diambil dalam periode 6 bulan mulai dari bulan Maret 2019 sampai dengan bulan Agustus 2019. Data musim hujan diambil dalam periode 6 bulan juga dari bulan September 2019 sampai dengan bulan Februari 2020. Data pada musim hujan lebih sedikit dikarenakan ada penurunan jumlah *customer* pada musim hujan. Kedai Kopi Teras Garden mempunyai pengunjung yang lebih sedikit pada musim hujan daripada musim kemarau. Perbedaan musin juga mempengaruhi hasil yang didapat karena pada musim kemarau pengunjung Kedai Kopi Teras

Garden cenderung untuk membeli minuman dingin. Begitu juga sebaliknya, pada musim hujan pengunjung lebih banyak memesan minuman hangat.

3.2 Testing

Pengujian pada penelitian ini menggunakan jupyter lab untuk analisa data yang didapat dari manajemen Kedai Kopi Teras Garden. Untuk hasil yang maksimal, maka diperlukan tahap *preprocessing* terlebih dahulu. Tahap pertama untuk *preprocessing* adalah menghilangkan item yang sama pada setiap transaksi. Hal ini perlu dilakukan untuk menghindari hasil output dengan item yang sama contoh jika memesan ice tea maka akan membeli ice tea. Hasil ini kurang relevan untuk dijadikan rekomendasi penjualan. Tahap kedua adalah menghapus *cell* kosong pada file csv untuk bisa diproses lebih lanjut. Tahap berikutnya adalah mengubah file csv ke dalam struktur data berbentuk list dalam pemrograman python.

Data pada musim kemarau dan musim hujan akan ditentukan *minimum support count* sebesar 20% dan *minimum confidence* sebesar 0,5. Berdasarkan data yang di dapat dari 6 bulan musim kemarau dan musim hujan didapatkan 1 rule setiap musim yang sesuai dengan *minimum support* dan *minimum confidence* yang sudah ditentukan pada penelitian ini. Pada musim kemarau didapatkan rule French Fries dan Ice Tea dengan tingkat *confidence* sebesar 0,61. Dapat diartikan jika Membeli French Fries maka akan membeli Ice Tea.

```
[11]: {'French Fries',): ('Ice Tea',), 0.6151639344262295}
```

Gambar 2 hasil dari data musim kemarau

```
[28]: {'Donat',): ('Hot Cappuccino',), 0.5837696335078534}
```

Gambar 3 hasil dari data musim hujan

Sedangkan pada musim hujan didapatkan rule Donat dan Hot Capuccino dengan tingkat *confidence* sebesar 0, 58. Dapat di artikan Jika membeli Donat maka akan membeli Hot Capuccino. Hasil kedua analisa tersebut dilakukan pada Jupyter lab dengan menggunakan Bahasa pemrograman python dan menggunakan library fp growth.

4. Kesimpulan

Berdasarkan hasil uji coba yang dilakukan pada penelitian ini didapatkan 2 rule yang terdiri dari 1 rule pada musim kemarau dan 1 rule pada musim hujan yang memenuhi syarat *minimum support count* sebesar 20% dan *minimum confidence* sebesar 0,5 untuk digunakan sebagai rekomendasi

penjualan dengan sistem paket atau dijual secara bersamaan. Rule yang dihasilkan sebagai berikut. Pada musim kemarau, jika *customer* membeli French Fries maka *customer* juga akan membeli Ice Tea dengan tingkat *confidence* sebesar 0,61. Sedangkan pada musim Hujan, jika *customer* membeli Donat maka *customer* juga akan membeli Hot Capuccino dengan tingkat *confidence* sebesar 0,58. Data yang di dapat dari Kedai Kopi Teras Garden pada musim hujan lebih sedikit karena terdapat penurunan jumlah pengunjung saat musim hujan.

5. DAFTAR PUSTAKA

- [1] B. A. Arasy Alimudin, Soebandi, Ahmad Zakki Falani, Eman Setiawan, "The Use of Fuzzy Logic in Developing Competitiveness Strategy Toward Small and Medium Enterprises (Groceries)," *Adv. Sci. Lett.*, vol. 23, no. 12, 2017.
- [2] M. Narvekar and S. Fatma, "An optimized algorithm for association rule mining using FP tree," *Procedia - Procedia Comput. Sci.*, vol. 45, pp. 101-110, 2015, doi: 10.1016/j.procs.2015.03.097.
- [3] A. Ikhwan, "A NOVELTY OF DATA MINING FOR FP-GROWTH ALGORITHM," vol. 9, no. 7, pp. 1660-1669, 2018.
- [4] K. Gadia and K. Bhowmick, "Parallel text mining in multicore systems using FP-Tree algorithm," *Procedia - Procedia Comput. Sci.*, vol. 45, pp. 111-117, 2015, doi: 10.1016/j.procs.2015.03.100.
- [5] Y. Li and S. Yin, "Mining Algorithm for Weighted FP-Growth Frequent Item Sets based on Ordered FP-Tree," vol. 5, no. October, pp. 154-158, 2019.
- [6] E. Regar and W. Silfianti, "Visit Patterns Analysis of Foreign Tourist in Indonesian Territory Using Frequent Pattern Growth (FP - Growth) Algorithm," vol. 7, no. 9, pp. 212-217, 2018.

SYSTEMIC: Information System and Informatics Journal

ISSN: 2460-8092, 2548-6551 (e)

Vol 6 No 2 – Desember 2020

Pemanfaatan Image Mining Untuk Klasifikasi Dan Prediksi Kematangan Tomat Menggunakan Metode Jaringan Saraf Tiruan BackpropagationFirdaus¹, Nori Sahrin²¹ AMIK Boekittinggi² Sekolah Tinggi Ilmu Ekonomi Riaufirdaus6ta@gmail.com¹, norisahrin84@gmail.com²**Kata Kunci***Image Mining, Jaringan Saraf Tiruan, Backpropagation, Tomat, RGB***Abstrak**

Klasifikasi dan prediksi terhadap tingkat kematangan tomat secara otomatisasi menggunakan metode Jaringan Saraf Tiruan Backpropagation. Penentuan kematangan bidang pertanian masih diterapkan secara manual. Dengan adanya perkembangan teknologi di bidang image mining, penentuan kematangan tomat dapat dilakukan secara otomatisasi. Metode yang dipakai dalam pembuatan sistem ini ialah Jaringan Saraf Tiruan. Algoritma yang digunakan ialah backpropagation. Output kematangan tomat terdiri dari tiga kategori yaitu belum matang, setengah matang dan matang. Data training dan data testing sebanyak 60 buah yang digunakan. Arsitektur backpropagation pada penelitian ini berupa 3 input layer, 4 hidden layer, dan 1 output layer. Fungsi aktivasi yang digunakan dari input ke hidden layer ialah sigmoid biner, sedangkan dari hidden layer ke output ialah fungsi identitas (purelin). Ekstraksi citra dalam bentuk nilai minimum RGB berguna sebagai input. Diproses dan menghasilkan output tingkat kematangan dan prediksi kematangan. Hasil pengujian sistem data training memperoleh nilai keakuratan sebanyak 96,67% dan data testing senilai 90%.

Keywords*Image Mining, Artificial Neural Networks, Backpropagation, Tomato, RGB***Abstract**

Classification and prediction of tomato maturity level are automated using the Backpropagation Neural Network method. The determination of maturity in agriculture is still applied manually. With the development of technology in the field of image mining, determining the maturity of tomatoes can be done automatically. The method used in making this system is an Artificial Neural Network. The algorithm used is backpropagation. The output of tomato ripeness consists of three categories, namely immature, half ripe and ripe. 60 training data and testing data were used. The backpropagation architecture in this study consists of 3 input layers, 4 hidden layers, and 1 output layer. The activation function used from input to hidden layer is binary sigmoid, while from hidden layer to output is the identity function (purelin). Image extraction in the form of RGB minimum value is useful as input. Processed and produces output maturity level and maturity prediction. The results of testing the training data system obtained an accuracy value of 96.67% and testing data of 90%.

1. Pendahuluan

Indonesia merupakan negara agraris dengan kekayaan sumber daya alam yang melimpah dan posisi Indonesia yang strategis. Curah hujan yang tinggi, iklim yang tropis, serta tanah yang subur menjadi keuntungan yang dapat memudahkan tumbuhnya berbagai produksi pertanian. Salah satu hasil sektor pertanian ialah tomat. Tomat atau *Lycopersicon esculentum* menjadi salah satu produk pertanian yang bermanfaat untuk tubuh manusia^[1]. Berdasarkan data dari Badan Pusat Statistik dan

Direktoral Jenderal Hortikultura, diketahui tingkat produksi tomat di provinsi Sumatera Barat tahun 2017 mencapai 101,292 ton dan menduduki posisi kedua penghasil terbesar di Indonesia. Hal ini tentu merupakan peluang yang besar bagi kelompok tani yang ada di Sumatera Barat untuk menghasilkan tomat dengan kualitas yang baik dan menjaga mutu tomat agar tetap baik saat berada di tangan konsumen.

Namun penentuan tingkat kematangan tomat masih dilakukan secara manual melalui

pengamatan langsung visual manusia. Identifikasi dengan cara manual memiliki kelemahan yaitu membutuhkan waktu yang lama dan bersifat subjektif. Permasalahan utamanya ialah waktu perjalanan dari produsen hingga ke tangan konsumen. Selama waktu tersebut berlalu, proses pematangan tomat terjadi. Jika waktu yang digunakan tidak efektif, maka kemungkinan tomat untuk busuk sebelum dikonsumsi juga dapat terjadi sehingga dapat merugikan petani maupun penjual.

Teknologi informasi memungkinkan penentuan tingkat kematangan tomat secara otomatisasi yaitu menggunakan *data mining* dengan metode Jaringan Syaraf Tiruan (JST). JST adalah sistem komputasi dimana arsitekturnya meniru konsep dan cara kerja dari sel saraf biologis di dalam otak. Tujuannya untuk memberikan informasi kepada masyarakat yang membutuhkan dalam pengklasifikasian tingkat kematangan tomat dengan hasil yang lebih objektif, waktu yang cepat, dan hasil yang akurat. Serta diperlukan prediksi kematangan tomat yang efektif hingga sampai ke tangan konsumen. Jaringan Saraf Tiruan dapat diguahkan untuk klasifikasi buah karena menghasilkan tingkat akurasi yang lebih tinggi dibandingkan dengan metode lainnya^[2].

Beberapa tahun terakhir, teknik *image processing* telah banyak ditemukan semakin berguna dalam industri buah^[3]. Penelitian terkait *image processing* dengan metode jaringan saraf tiruan (*artificial neural network*) untuk bidang pertanian telah dilakukan oleh beberapa peneliti, seperti K. Sandhiya, dkk. dengan judul "Smart Fruit Classification using Neural Networks", Sella Kusumaningtyas dan Rosa Andrie A. dengan judul jurnal "Identifikasi Kematangan Buah Tomat Berdasarkan Warna Menggunakan Metode Jaringan Syaraf Tiruan", M. Angga Anggriawan, dkk. pada jurnalnya dengan judul "Pengenalan Tingkat Kematangan Tomat Berdasarkan Citra Warna Pada Studi Kasus Pembangunan Sistem Pemilihan Otomatis", Dila Deswari, dkk. dengan judul jurnal "Identifikasi Kematangan Buah Tomat Menggunakan Metoda Backpropagation", jurnal "Identifikasi Tingkat Ketuaan dan Kematangan Pepaya (*Carica papaya L.*) IPB 1 dengan Pengolahan Cira Digital dan Jaringan Syaraf Tiruan" oleh Enrico Syaefullah, Hadi K. Purwadaria, Sutrisno, dan Suroso.

Maka dari itu, penelitian menggunakan metode klasifikasi jaringan saraf tiruan untuk mengklasifikasikan tingkat kematangan tomat pada kelompok tani penghasil tomat berdasarkan warna menggunakan teknik pengolahan citra serta algoritma *backpropagation*. Salah satu keunggulan metode jaringan saraf tiruan ialah kemampuan yang dimiliki untuk melakukan prediksi. Algoritma *backpropagation* sangat bermanfaat, cukup handal serta cukup mudah dipahami. Objek yang diamati ialah tomat pada tingkat kematangan yang memiliki citra warna yang kontras dari warna buah

muda hingga matang. Citra diambil menggunakan kamera dengan kualitas pencahayaan yang sama serta sudut pengambilan yang sama.

2. Tinjauan Pustaka

2.1 Data Mining

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakit dari berbagai *database* besar^[4].

2.2 Image mining

Image mining ialah memahami gambar dengan pola tertentu agar mendapatkan informasi. Saat ini, *image mining* secara cepat mendapat perhatian diantara para peneliti di bidang *data mining*, *information retrieval*, dan *multimedia databases* karena potensinya yang berguna dalam menemukan pola gambar. Sistem *image mining* dapat secara otomatis mendapatkan informasi yang bernilai pengetahuan dari data gambar yang ada. *Image mining* bertujuan untuk penentuan pola gambar yang signifikan dalam koleksi gambar yang diberikan dan data alfanumerik terkait^[5].

2.3 Citra RGB

Citra RGB merupakan citra yang terdiri dari 3 kanal dengan komponen warna merah/*red* (R), hijau/*green* (G), biru/*blue* (B). Rentang nilai pada masing-masing kanal yaitu dari 0 – 255. Gambar 2.4 merupakan contoh citra RGB. Karena citra ini memiliki 3 kanal, maka untuk satu piksel pada citra ini memerlukan $3 \times 8\text{-bit} = 24\text{-bit}$. Citra berwarna disebut juga dengan citra 24-bit. Model warna RGB bertujuan untuk penginderaan dan presentasi gambar yang dilihat dalam bentuk visual pada berbagai alat elektronik seperti televisi dan fotografi^[6].

2.4 Pengolahan citra

Pengolan citra merupakan proses untuk mengubah kualitas citra atau gambar menjadi lebih baik, khususnya menggunakan perangkat komputer sebagai *hardware* (Munir, 2004). Seringkali citra yang asli mengalami penurunan kualitas akibat beberapa pengaruh seperti citra mengandung cacat atau derau (*noise*), warna citra terlalu kontras, kurang tajam (*blur*), dan lain sebagainya yang dapat memengaruhi informasi yang disampaikan. Citra tersebut tentu perlu diperbaiki agar kualitasnya menjadi lebih baik sehingga informasi yang ditangkap dapat akurat dan jelas. *Input* dari program pengolahan citra adalah citra dan *output* adalah citra pula. Namun *output* citra memiliki kualitas yang lebih bagus dibandingkan dengan *input* citra.

2.5 Tomat

Dalam bahasa latin disebut dengan

Lycopersicon esculentum adalah tumbuhan dari keluarga *Solanaceae*. Tomat merupakan salah satu jenis produk hortikultura yang dapat berguna untuk kesehatan tubuh, serta mempunyai peluang pasar yang cukup menjanjikan. Komposisi tomat mengandung gizi yang cukup lengkap. Salah satu perubahan yang menonjol dalam proses pematangan buah ialah perubahan warna (Winarno, 1981).

2.6 Jaringan Saraf Tiruan

Penelitian ini menggunakan salah satu dari metode dalam *data mining* yaitu Jaringan Saraf Tiruan. Jaringan Saraf Tiruan atau *Artificial Neural Network* ialah metode yang menerapkan sistem jaringan saraf hidup makhluk hidup terutama manusia dengan mengadopsi cara kerja neuron secara biologi fokus pada cara kerja otak. Jaringan saraf tiruan terdiri dari kumpulan *node* (neuron) dan relasi. Ada tiga tipe *node*, yaitu *input*, *hidden*, dan *output*.

2.7 Metode *backpropagation*

Metode ini pertama kali diperkenalkan oleh Paul Werbos pada 1974, kemudian dikemukakan kembali oleh David Parker pada 1982 selanjutnya dipopulerkan oleh Rumelhart dan McClelland pada 1986. Aturan pelatihan jaringan *backpropagation* terdiri dari 2 tahapan yaitu *feedforward* dan *backward propagation*. Pada jaringan diberikan sekumpulan contoh pelatihan yang disebut sebagai set pelatihan. Set pelatihan ini terdiri dari vektor input dan vektor output yang akan menjadi target pelatihannya. Keluaran dari jaringan berupa sebuah vektor output aktual.

Siklus setiap perubahan bobot (*epoch*) dilakukan pada setiap set pelatihan hingga kondisi berhenti tercapai. Algoritma *backpropagation* terdiri dari 3 tahapan (Puspitaningrum, 2006) yaitu:

1. Tahapan umpan maju (*feedforward*)
2. Tahapan umpan mundur (*backward propagation*)
3. Tahap meng-*update* bobot dan bias

3. Metode Penelitian

3.1. Analisa Kebutuhan Fungsional dan Non-fungsional Sistem

Kebutuhan fungsional ialah kebutuhan akan proses yang dapat dilakukan oleh sistem. Berdasarkan penelitian ini, kebutuhan fungsional sistem tersebut antara lain sebagai berikut:

1. Sistem dapat melakukan input citra tomat melalui direktori dan melalui pengambilan langsung menggunakan *webcam* yang dihubungkan ke PC.
2. Sistem dapat menampilkan nilai histogram RGB dari citra tomat.
3. Sistem dapat menentukan nilai minimum RGB tomat berdasarkan pengolahan citra.

4. Sistem dapat melakukan pemotongan atau *cropping* pada citra asli sesuai dengan daerah pemotongan yang diinginkan.
5. Sistem dapat menghasilkan klasifikasi dan prediksi kematangan tomat berdasarkan data *training* dengan perhitungan metode *backpropagation*.

Kebutuhan non-fungsional ialah kebutuhan yang berisi proses-proses yang akan mempengaruhi sistem. Berdasarkan penelitian ini, proses-proses tersebut antara lain sebagai berikut:

1. Faktor cahaya yang diberikan.
2. Faktor latar pengambilan citra tomat.

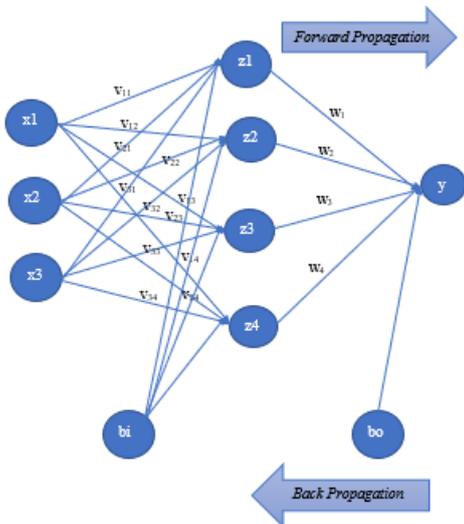
3.2 Tahapan Penelitian



Gambar 3. 1 Tahapan Penelitian

3.3 Rancangan Arsitektur *Backpropagation*

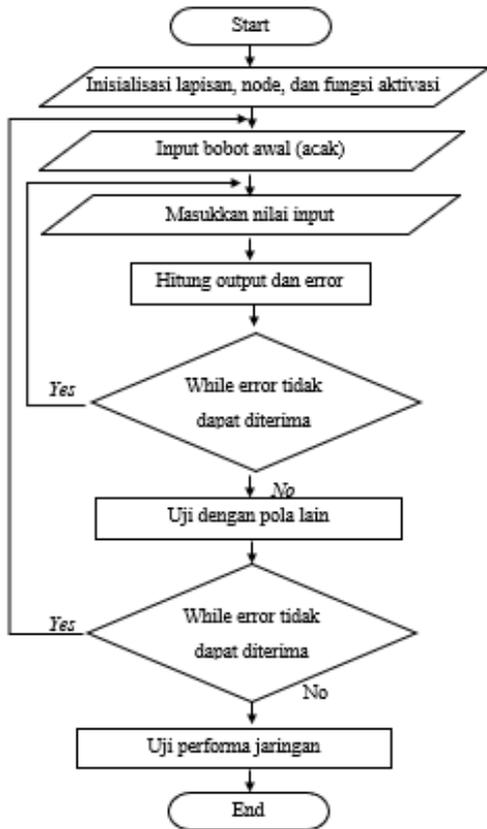
Langkah pertama yang dilakukan dalam perancangan penelitian ini dimulai dari perancangan arsitektur *backpropagation*, seperti pada gambar 3.2. Proses tersebut terdiri dari 3 unit *input layer*, 4 node *hidden layer*, dan 3 unit *output layer*. Pada bagian *input layer* (x) berisi nilai RGB dari tomat yang telah didapatkan dari hasil pengolahan citra.



Gambar 3. 2 Arsitektur Backpropagation

3.4 Rancangan Proses Jaringan Saraf Tiruan

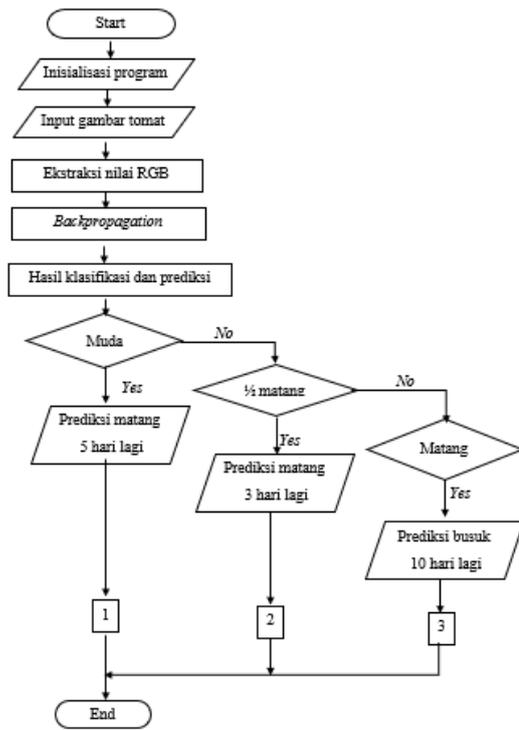
Flowchart jaringan saraf tiruan ditunjukkan pada gambar 3.4 berikut.



Gambar 3. 3 Flowchart Jaringan Saraf Tiruan

3.5 Flowchart Pengujian

Apabila pelatihan jaringan saraf tiruan telah didapatkan dan sistem yang dibangun telah selesai, maka tahap selanjutnya ialah melakukan pengujian terhadap data *training* dan data *testing*. Semua data dihitung nilai kebenaran dan kesalahan. Berdasarkan nilai tersebut didapatkan hasil tingkat akurasi. Flowchart untuk tahapan pengujian ditunjukkan pada gambar 3.4 berikut



Gambar 3. 4 Flowchart Pengujian

4. Implementasi

Pada penelitian ini dibuat sistem untuk mengetahui klasifikasi kematangan tomat dan prediksi dengan metode Jaringan Saraf Tiruan. Objek tomat diambil dari kelompok tani penghasil tomat di Alahan Panjang dapat dilihat pada gambar 4.1.



Gambar 4. 1 Kebun Tomat di Alahan Panjang

Data-data citra tomat yang sudah dipetik dengan menggunakan kamera *smartphone* dengan latar belakang solid berwarna putih menggunakan alat berupa mini *photo studio* seperti pada gambar 4.2. Proses pengambilan gambar tomat disesuaikan dengan pencahayaan yang sama.



Gambar 4. 2 Pengambilan Citra Tomat

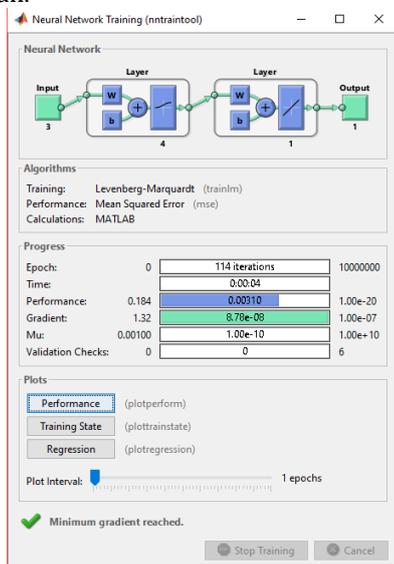
Total data yang didapatkan sebanyak 60 buah citra tomat. Pada data *training* terdapat total 30 citra tomat. Masing-masing kategori terdiri dari 10 citra. Sedangkan pada data *testing* terdapat 30 citra tomat.

4.1 Implementasi Pelatihan Data Latih

Data latih dikumpulkan dan dilatih menggunakan aplikasi MATLAB dengan metode jaringan saraf tiruan agar mendapatkan nilai bobot dan bias dari data tersebut. Gambar tomat sebagai data latih dicari terlebih dahulu nilai minimum RGB pada masing-masing citra.

Inisialisasi jaringan adalah hal pertama yang perlu dilakukan untuk menghitung *backpropagation* menggunakan Matlab. Perintah yang dipakai untuk membentuk jaringan adalah **newff**. Fungsi aktivasi yang digunakan diinput dari unit input ke layer tersembunyi adalah fungsi aktivasi sigmoid biner (logsig), sedangkan fungsi aktivasi dari layer tersembunyi ke output ialah fungsi identitas (purelin).

Setelah program tersebut dijalankan, maka bentuk jaringan pelatihan akan ditampilkan beserta atribut dan hasil dari *training* yang dilakukan.



Gambar 4. 3 Bentuk Jaringan Pelatihan

Hasil perhitungan nilai bobot dan bias dengan menggunakan MATLAB disimpan ke *excel*.

Nilai bobot *Input Weight (IW)* dan *Layer Weight (LW)* dan *Input Bias* dan *Layer Bias (IB dan LB)* dari pelatihan ini adalah:

	v1	v2	v3
IW =	0.267801589	0.120305327	0.280376744
	7.309172349	-7.012592424	1.510508462
	2.438326094	-4.397011898	1.942057488
	-	1.419873875	1.022442117
	1.697743454		

				65
LW =	1.579075076	-	0.278303465	-
		1.690360577		0.3893544414

IB =	0.001888
------	----------

	0.407338
	0.027037
	0.74368

LB =	0.666315622
------	-------------

Setelah memperoleh hasil pelatihan data latih dengan menggunakan MATLAB, selanjutnya menguji hasil tersebut dengan menggunakan rumus *backpropagation*. Nilai output yang didapat berdasarkan metode *backpropagation*. Prosedur pengujian aplikasinya adalah sebagai berikut:

Langkah 0 : Inisialisasi bobot (algoritma pelatihan).

Langkah 1 : Untuk tiap vektor masukan, lakukan langkah 2-4.

Langkah 2 : Hitung nilai di neuron *hidden layer* serta fungsi aktivasi sigmoid biner. Misalnya kita menghitung nilai output untuk nilai input minimum RGB yaitu 87, 97, dan 34.

$$z_{in_j} = \sum_{i=1}^n x_i v_{ji} + b_i$$

$$j=1, \quad z_{in_1} = (x_1 v_{11}) + (x_2 v_{21}) + (x_3 v_{31}) + b_1$$

$$z_{in_1} = (87 * 0.26780158) + (97 * 0.12030532) + (34 * 0.28037674) + 0.00188848$$

$$z_{in_1} = 44.50116429 + 0.00188848$$

$$z_{in_1} = 44.50305277$$

$$j=2, \quad z_{in_2} = (x_1 v_{12}) + (x_2 v_{22}) + (x_3 v_{32}) + b_2$$

$$z_{in_2} = (87 * 7.30917234) + (97 * -7.012592424) + (34 * 1.51050846) + 0.407337765$$

$$z_{in_2} = 7.033816925 + 0.407337765$$

$$z_{in_2} = 7.441154691$$

$$j=3, \quad z_{in_3} = (x_1 v_{13}) + (x_2 v_{23}) + (x_3 v_{33}) + b_3$$

$$z_{in_3} = (87 * 2.43832609) + (97 * -4.39701189) + (34 * 1.94205748) + 0.0270365070$$

$$z_{in_3} = -148.3458293 + 0.0270365070$$

$$z_{in_3} = -148.3187928$$

$$j=4, \quad z_{in_4} = (x_1 v_{14}) + (x_2 v_{24}) + (x_3 v_{34}) + b_4$$

$$z_{in_4} = (87 * 2.43832609) + (97 * 1.4198738752) + (34 * 1.0224421165) + 0.7436802508$$

$$z_{in_4} = 24.78711737 + 0.0270365070$$

$$z_{in_4} = 25.53079762$$

Fungsi Aktivasi

$$z_j = f(z_{in_j}) = \frac{1}{1 + e^{-z_{in_j}}}$$

$$j=1, \quad z_1 = f(z_{in_1}) = \frac{1}{1 + e^{-44.50305277}} = 1$$

$$j=2, \quad z_2 = f(z_{in_2}) = \frac{1}{1 + e^{-7.441154691}} = 0.999414$$

$$j=3, \quad z_3 = f(z_{in_3}) = \frac{1}{1 + e^{-(-148.3187928)}} = 3.85e^{-}$$

$$z_4 = f(z_{in_4}) = \frac{1}{1 + e^{-25.53079762}} = 1$$

Langkah 5 : Masing-masing unit output (y_k , $k=1,2,3,\dots,n$)

dikalikan dengan bobot dan dijumlahkan serta ditambahkan dengan indeks bias outputnya.

$$y_{in_k} = \sum_{j=1}^p z_j w_j + b_o$$

$$y_{in_1} = (z_1 w_1) + (z_2 w_2) + (z_3 w_3) + (z_4 w_4) + 0.666315622$$

$$y_{in_1} = (1 * 1.579075076) + (0.999414 * -1.690360577) + (3.85e - 65 * 0.2783034652) + (1 * -0.3893544413) + 0.666315622$$

$$y_{in_1} = 0.16667$$

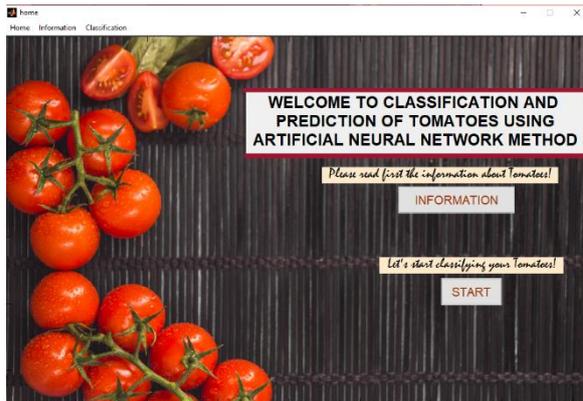
$$y_k = f(y_{in_k}) = y_{in_k} = 0.16667$$

Langkah 6 : Sesuaikan nilai output pada masing-masing kategori

Pada pemrograman sistem ini, nilai keluaran berupa y_k yang telah diperoleh akan menjadi parameter untuk kondisi penentuan tingkat kematangan dari tomat. Berdasarkan nilai y_k tersebut didapatkan tingkat kematangan tomat kategori belum matang bernilai rentang 0.166667, setengah matang bernilai 0.5000, dan matang bernilai 0.8333.

4.2 Implementasi User Interface

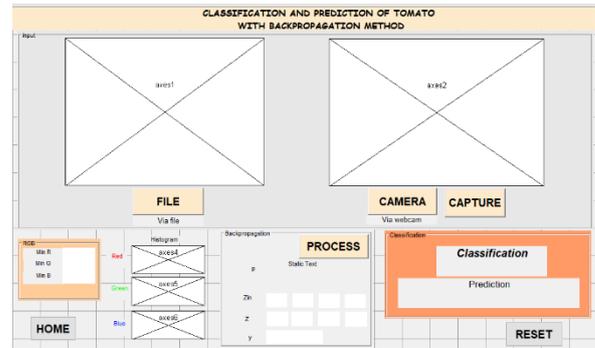
Aplikasi ini ditampilkan dalam bentuk *Graphical User Interface* (GUI). Sistem ini menggunakan tiga halaman, yaitu halaman *home* seperti pada gambar 4.4, *information* pada gambar 4.5, dan *classification* pada gambar 4.6.



Gambar 4. 4 Tampilan Halaman Home

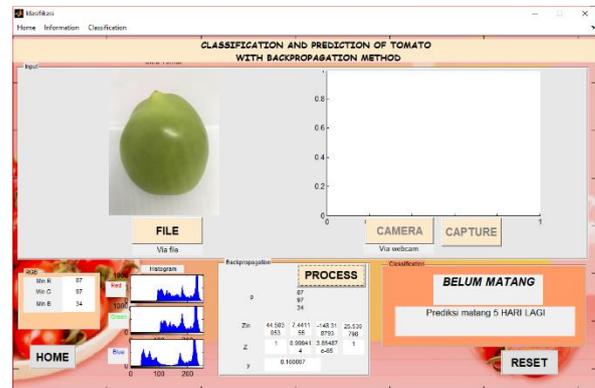


Gambar 4. 5 Tampilan Halaman Information

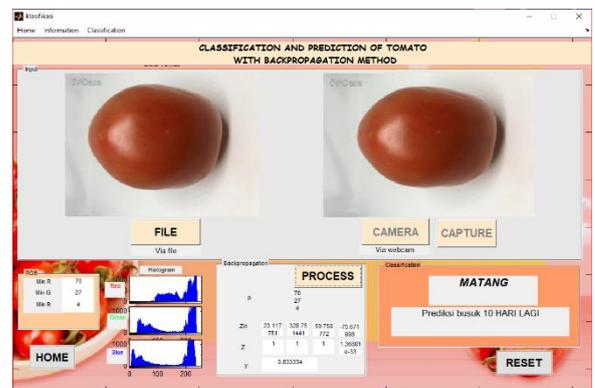


Gambar 4. 6 Tampilan Halaman Classification

Apabila aplikasi tersebut dijalankan, maka akan menampilkan hasil klasifikasi dan prediksi citra seperti pada gambar 4.7 jika melalui direktori citra tomat yang telah ada. Proses aplikasi ini dimulai dari pengambilan gambar tomat dapat melalui direktori yang telah ada maupun melalui *capture* langsung dengan menggunakan *webcam* seperti pada gambar 4.8. Apabila melalui direktori, maka *user* dapat menekan tombol *file* dan memilih citra tomat yang akan diklasifikasikan sehingga citra tersebut dapat ditampilkan. Kemudian tombol *process* akan menampilkan hasil dari ekstraksi citra berupa nilai minimum RGB dan perhitungan metode *backpropagation*.



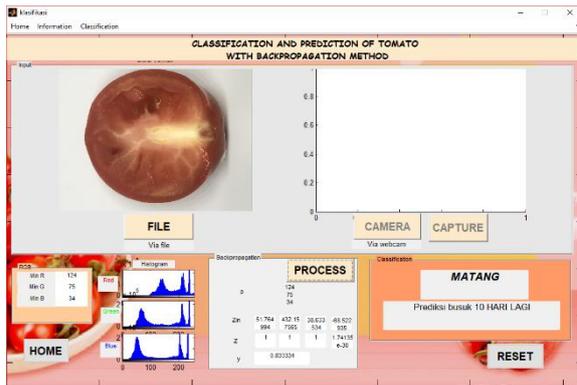
Gambar 4. 7 Input Citra melalui Direktori



Gambar 4. 8 Input Citra melalui Webcam

Input citra tomat juga diuji dengan menggunakan citra penampang tomat yang dipotong melintang. Hal ini untuk mengetahui kondisi kematangan pada tomat bagian dalam. Contoh klasifikasi kematangan tomat dengan

penampang melintang dapat dilihat pada gambar 4.9.



Gambar 4. 9 Pengujian Penampang Tomat

4.3 Pengujian

Pengujian dilakukan dengan cara menguji masing-masing data *training* dan data *testing* dengan menggunakan sistem yang telah dibuat. Pengujian sistem ini menerapkan *black box test*, yaitu pengujian hanya mengamati hasil eksekusi program melalui data uji pada sistem dan memeriksa fungsi-fungsi yang ada pada sistem atau aplikasi tersebut.

1. Pengujian Data Training

Tabel 4.1 menunjukkan data hasil pengujian terhadap data *training*.

Tabel 4. 1 Pengujian Data Training

No	Gambar	Nilai Min RGB			Hasil Hitung	Klasi fikasi	Kebe naran
		R	G	B			
1		87	97	34	0.166	Belum Matang	Benar
2		59	63	19	0.165	Belum Matang	Benar
3		61	64	27	0.165	Belum Matang	Benar
4		82	83	32	0.165	Belum Matang	Benar
5		76	77	28	0.165	Belum Matang	Benar
6		88	90	33	0.165	Belum Matang	Benar
7		73	74	22	0.17	Belum Matang	Benar
8		68	72	38	0.165	Belum Matang	Benar
9		57	60	15	0.169	Belum Matang	Benar
10		68	71	36	0.165	Belum Matang	Benar
11		75	75	18	0.494	Setengah Matang	Benar
12		67	74	20	0.500	Setengah Matang	Benar
13		89	62	17	0.555	Setengah Matang	Benar

14		92	78	28	0.555	Setengah Matang	Benar
15		84	80	26	0.494	Setengah Matang	Benar
16		73	65	16	0.555	Setengah Matang	Benar
17		81	79	22	0.512	Setengah Matang	Benar
18		93	85	21	0.555	Setengah Matang	Benar
19		72	67	24	0.500	Setengah Matang	Benar
20		81	61	23	0.555	Setengah Matang	Benar
21		59	28	9	0.833	Matang	Benar
22		87	54	19	0.833	Matang	Benar
23		65	33	9	0.833	Matang	Benar
24		74	37	14	0.833	Matang	Benar
25		68	59	19	0.555	Setengah Matang	Salah
26		61	28	12	0.833	Matang	Benar
27		59	30	14	0.833	Matang	Benar
28		66	31	14	0.833	Matang	Benar
29		61	31	10	0.833	Matang	Benar
30		57	27	10	0.833	Matang	Benar

$$Akurasi = \frac{\text{jumlah data benar}}{\text{jumlah seluruh data}} \times 100\%$$

$$Akurasi = \frac{29}{30} \times 100\% = 96,67\%$$

2. Pengujian Data Testing

Pengujian Data Testing dapat dilihat pada tabel 4.2.

Tabel 4. 2 Pengujian Data Testing

No	Gambar	Nilai Min RGB			Hasil hitung	Klasi fikasi	Kebe naran
		R	G	B			
1		60	64	17	0.162	Belum Matang	Benar
2		56	58	13	0.248	Belum Matang	Benar
3		60	68	16	1.856	Matang	Salah
4		14	15	17	0.380	Setengah Matang	Salah
5		43	46	11	0.171	Belum Matang	Benar
6		38	40	11	0.171	Belum Matang	Benar
7		59	61	13	0.314	Belum Matang	Benar
8		53	55	24	0.165	Belum Matang	Benar

9		57	59	22	0.165	Belum Matang	Benar
10		45	47	25	0.165	Belum Matang	Benar
11		38	26	10	0.582	Setengah Matang	Benar
12		67	52	8	0.555	Setengah Matang	Benar
13		67	63	11	0.555	Setengah Matang	Benar
14		42	38	12	0.550	Setengah Matang	Benar
15		64	53	9	0.555	Setengah Matang	Benar
16		67	59	13	0.555	Setengah Matang	Benar
17		66	58	14	0.555	Setengah Matang	Benar
18		39	25	4	0.555	Setengah Matang	Benar
19		60	41	18	0.758	Matang	Salah
20		71	68	20	0.532	Setengah Matang	Benar
21		59	25	5	0.833	Matang	Benar
22		58	26	7	0.833	Matang	Benar
23		47	22	2	0.833	Matang	Benar
24		55	23	8	0.833	Matang	Benar
25		63	24	8	0.833	Matang	Benar
26		46	21	5	0.833	Matang	Benar
27		53	29	11	0.833	Matang	Benar
28		51	26	5	0.833	Matang	Benar
29		51	25	8	0.833	Matang	Benar
30		46	13	0	0.833	Matang	Benar

$$Akurasi = \frac{27}{30} \times 100\% = 90\%$$

Berdasarkan hasil pengujian dari klasifikasi tingkat kematangan tomat menggunakan metode Jaringan Saraf Tiruan *Backpropagation*, didapatkan hasil akurasi yang bagus. Nilai akurasi pengujian terhadap data *training* ialah sebanyak 96,7%, nilai akurasi pengujian terhadap data *testing* ialah sebanyak 90%, dan nilai akurasi terhadap pengujian data penampang melintang tomat ialah sebanyak 80%. Sehingga metode yang digunakan sesuai dengan hasil yang diharapkan. Pada beberapa data, terdapat kesalahan dalam menghasilkan klasifikasi tingkat kematangan. Hal ini terjadi karena

kesalahan sistem dalam membaca nilai RGB yang diinputkan. Pada pengujian data *training* terdapat kesalahan satu data. Sedangkan pada pengujian data *testing*, terdapat kesalahan klasifikasi sebanyak 3 buah data. Pada data tomat kategori matang menghasilkan klasifikasi tingkat kematangan matang. Hal ini disebabkan karena terdapat kemiripan nilai RGB pada data matang dengan setengah matang.

5. Kesimpulan dan Saran

Dari hasil penelitian tentang pemanfaatan *image mining* untuk klasifikasi dan prediksi tingkat kematangan tomat menggunakan metode Jaringan Saraf Tiruan, dapat diambil kesimpulan sebagai berikut:

1. Berdasarkan penelitian yang dilakukan terhadap klasifikasi dan prediksi kematangan tomat, metode jaringan saraf tiruan algoritma *backpropagation* dengan struktur 3 *neuron input layer*, 4 *neuron hidden layer* dan 1 *neuron output layer* mampu melakukan pengklasifikasian berdasarkan unsur warna RGB dengan memanfaatkan *image mining*.

2. Berdasarkan penelitian yang dilakukan terhadap klasifikasi dan prediksi kematangan tomat menghasilkan tingkat akurasi pada pengujian menggunakan data *training* diperoleh hasil sebesar 96,7% sedangkan pengujian menggunakan data *testing* diperoleh hasil sebesar 90%. Sedangkan pada pengujian terhadap data penampang tomat diperoleh hasil sebesar 80%. Hasil keakuratan tergantung pada kualitas gambar yang dihasilkan.

3. Perangkat lunak yang dibangun mampu mengelompokkan kematangan tomat yaitu belum matang, setengah matang, dan matang.

Saran

Berdasarkan kesimpulan hasil penelitian, disarankan untuk metode Jaringan Saraf Tiruan yang diterapkan menggunakan lebih banyak *hidden layer* agar mendapatkan keakuratan yang lebih baik. Selain itu, aplikasi dapat dikembangkan berupa aplikasi *android* atau *web* sebagai aplikasi yang dipakai langsung oleh masyarakat. Kualitas citra yang dihasilkan juga dapat ditingkatkan dengan cara segmentasi citra dan metode lainnya.

DAFTAR PUSTAKA

- [1] Prakash, A. (2001). Antioxidant Activity. *Medallion Laboratories Analytical*, vol. 19, No.2.
- [2] Sandhiya, K., Vidhya, M., Shivaranjani, M., & Saranya, S. (2017). Smart Fruit Classification using Neural Networks. *International Journal of Trend in Scientific Research and Development (IJTSRD)*, Vol.2 Page: 1303.

- [3] Prabha D., S., & SateheeshKumar, J. (2012). A Study on Image Processing Methods for Fruit Classification. *Research Gate*, 403.
- [4] Turban, E. (2005). *Decision Support Systems and Intelligent Systems*. Yogyakarta: Andi.
- [5] Zhang, J., Hsu, W., & Lee, M. (2002). Image Mining: Trends and Developments. *Journal of Intelligent Information Systems*, 7-23
- [6] Silvana, M., & Kurnia, R. (2015). Skin and Clothes Matching Seeded by Color System Selection. *TELKOMNIKA Indonesian Journal of Electrical Engineering*, vol.14 hal. 509.

SYSTEMIC: Information System and Informatics Journal

ISSN: 2460-8092, 2548-6551 (e)

Vol 6 No 2 – Desember 2020

Klasifikasi Text Judul Buku Perpustakaan Untuk Menentukan Kategori Buku Menggunakan K-Nearest Neighbor**Muhamad Kadafi**

Universitas Islam Negeri (UIN) Raden Fatah Palembang

kadafi_uin@radenfatah.ac.id**Kata Kunci***Nearest Neighbor Classifier, Data mining, Perpustakaan.***Abstrak**

Kebutuhan terhadap informasi dalam bentuk buku ataupun artikel ilmiah pada Perpustakaan UIN Raden Fatah Palembang semakin terus meningkat, untuk mempermudah dalam pencarian informasi buku salah satunya adalah dengan mengelompokkan buku berdasarkan jenis kategorinya. Dalam mengelompokkan data buku perpustakaan Metode Nearest Neighbor Classifier pada data mining dapat di kombinasikan dengan teknik ekstraksi data text untuk melakukan klasifikasi data text judul buku perpustakaan. Tujuan dari penelitian ini adalah untuk mengklasifikasikan text judul buku perpustakaan dengan menggunakan Nearest Neighbor Classifier untuk menentukan jenis kategori buku. Metode penelitian ini menggunakan teknik klasifikasi data mining Nearest Neighbor Classifier. Hasil dari penelitian ini adalah nilai akurasi tertinggi terdapat pada K=12 yaitu sebesar 72.50%, dan model yang terbentuk dapat digunakan untuk mengklasifikasikan buku dengan label 2x0, 150, 2x2, 400, 020, 2x1, 657, 500, 375, 302.2, 800 dan tidak dapat di gunakan untuk klasifikasi buku dengan label kelas 070, 370, 330, 300, 600, 340, 700.

Keywords*Nearest Neighbor Classifier, Data mining, Library***Abstract**

The need for information in the form of books or scientific articles at the Library of UIN Raden Fatah Palembang continues to increase. To make it easier to find book information, one of which is by classifying books based on the type of category. In classifying library book data, the Nearest Neighbor Classifier method in data mining can be combined with text data extraction techniques to classify library book title text data. The purpose of this study was to classify the text title of library books using the Nearest Neighbor Classifier to determine the type of book category. This research method uses the Nearest Neighbor Classifier data mining classification technique. The results of this study are that the highest accuracy value is found at K = 12, which is 72.50%, and the model formed can be used to classify books with labels 2x0, 150, 2x2, 400, 020, 2x1, 657, 500, 375, 302.2, 800. and cannot be used for classifying books with class labels 070, 370, 330, 300, 600, 340, 700.

1. Pendahuluan

Universitas Islam Negeri (UIN) Raden Fatah Palembang adalah salah satu perguruan tinggi negeri di kota Palembang Sumatera Selatan. UIN Raden Fatah Palembang dilengkapi dengan berbagai fasilitas untuk menunjang kegiatan akademik, dimana salah satunya adalah fasilitas berupa perpustakaan. Perpustakaan adalah fasilitas atau tempat menyediakan sarana bahan bacaan. Tujuan dari perpustakaan khususnya perguruan tinggi adalah memberikan layanan informasi untuk kegiatan belajar, penelitian, dan pengabdian masyarakat dalam rangka

melaksanakan Tri Dharma Perguruan Tinggi [1].

Perpustakaan UIN Raden Fatah Palembang memiliki koleksi buku yang cukup banyak dari berbagai cabang ilmu pengetahuan. Koleksi buku yang tersedia dapat digunakan oleh dosen, mahasiswa dan staf untuk mencari referensi buku atau sebagai bahan bacaan. Pada Perpustakaan UIN Raden Fatah Palembang kebutuhan terhadap informasi dalam bentuk buku ataupun artikel ilmiah semakin meningkat, hal tersebut dapat dilihat dari banyaknya jumlah buku yang terdapat pada perpustakaan. Maka untuk membantu mempermudah dalam pencarian informasi buku

salah satunya adalah dengan mengelompokkan buku berdasarkan jenisnya. Pengelompokan buku ini dibutuhkan untuk mempermudah pencarian informasi yaitu berupa topik yang menggambarkan pokok pembahasan secara umum [2].

Dalam mengelompokkan data buku perpustakaan para pengambil keputusan bisa memanfaatkan gudang data yang sudah dimiliki untuk melakukan analisa dalam mengambil keputusan, salah satunya yaitu dengan cara penggalian informasi atau pola yang penting dan menarik dari data jumlah besar, yang disebut dengan data mining. Data mining diartikan sebagai menambang data atau upaya untuk menggali informasi yang berharga dan berguna pada database yang sangat besar.

Salah satu teknik yang digunakan untuk menentukan pola tersebut yaitu dengan menggunakan teknik klasifikasi pada data mining dengan metode Nearest Neighbor Classifier. Nearest Neighbor Classifier adalah suatu metode yang menggunakan algoritma supervised [3], [4], [5]. Supervised learning bertujuan untuk menemukan pola baru dalam data dengan menghubungkan pola data yang sudah ada dengan data baru. Tujuan dari Nearest Neighbor Classifier adalah untuk mengklasifikasi objek baru berdasarkan atribut dan training samples [4], [5]. Dimana hasil dari uji sampel yang baru di klasifikasikan berdasarkan mayoritas dari kategori pada ketetanggaan terdekat. Nearest Neighbor Classifier adalah metode yang menentukan nilai jarak pada pengujian data testing dengan data training berdasarkan nilai terkecil dari ketetanggaan terdekat [6].

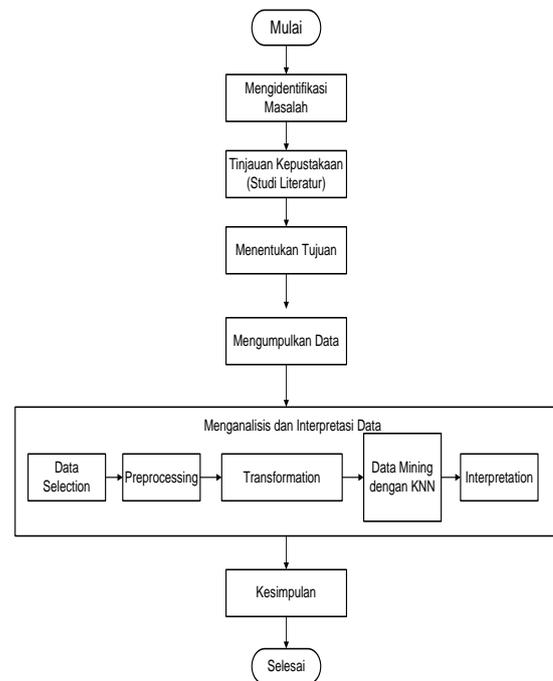
Untuk mengklasifikasikan buku perpustakaan, data buku yang telah tersedia sebelumnya dapat digunakan sebagai data training yaitu berupa data teks judul buku perpustakaan beserta label klasifikasinya. Nearest Neighbor Classifier pada data mining dapat di kombinasikan dengan teknik ekstraksi data text untuk melakukan klasifikasi dokumen, dalam hal ini adalah text judul buku perpustakaan yang disebut dengan text mining. Text mining merupakan proses menambang data yang berupa text dimana sumber data biasanya di dapat dari dokumen dan tujuannya adalah untuk mencari kata – kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen [7]. Tujuan dari text mining adalah mengekstrak informasi yang berguna dari sumber data, sumber data yang digunakan pada text mining adalah sekumpulan dokumen yang memiliki format yang tidak terstruktur melalui indentifikasi dan eksplorasi pola yang menarik. Adapun tugas khusus dari text mining antara lain, pengkategorian text (text categorization) dan Pengelompokan Text (text clustering) [3].

2. Metode Penelitian

Metode penelitian yang digunakan adalah pendekatan kuantitatif karena proses pengolahan data yang akan digunakan pada penelitian ini bersifat kuantitatif, digunakan untuk menghitung dan mengukur pada saat analisis dan interpretasi data. Dan hasilnya berupa data prediksi klasifikasi kategori buku jenis penelitian yang digunakan adalah penelitian eksperimen.

2.1 Tahapan Penelitian

Adapun Tahapan Penelitian adalah seperti pada gambar 1:



Gambar 1. Tahapan Penelitian

Tahapan yang akan ditempuh yaitu:

1. Mengidentifikasi Masalah
Masalah yang diidentifikasi dalam penelitian ini adalah bagaimana mengkategorikan buku berdasarkan pada text judul buku perpustakaan dengan menggunakan Nearest Neighbor Classifier?
2. Tinjauan Kepustakaan (Studi Literatur)
Literatur – literatur yang di pakai sebagai bahan refensi dalam penelitian ini adalah dari jurnal – jurnal ilmiah, modul pembelajaran, dan buku tentang data mining. literatur – literatur ini akan menjadi pedoman untuk melakukan penelitian agar memudahkan proses penelitian.
3. Menentukan Tujuan
Tujuan pada penelitian ini adalah mengklasifikasikan text judul buku perpustakaan dengan menggunakan Nearest Neighbor Classifier untuk menentukan kategori buku.
4. Mengumpulkan Data
Metode pengumpulan data dilakukan dengan cara melakukan pengamatan langsung di

Perpustakaan UIN Raden Fatah Palembang. Teknik yang digunakan adalah teknik analisis dokumen yaitu data transaksi peminjaman buku perpustakaan tahun 2015, 2016, 2017.

5. Menganalisis dan Interpretasi Data

Tahapan ini merupakan pengolahan data mining dengan mengikuti tahapan KDD (Knowledge Discovery In Database), dan metode yang digunakan untuk proses pengolahan data mining menggunakan klasifikasi Nearest Neighbor.

6. Kesimpulan

Membuat kesimpulan dari hasil penelitian dan memberikan saran untuk pihak perpustakaan agar dapat menjadi lebih baik lagi.

2.2 Nearest Neighbor Classifier

Metode Nearest Neighbor Classifier pertama kali di jelaskan pada awal tahun 1950-an. Nearest Neighbor classifier di dasarkan pada pembelajaran dengan analogi, yaitu dengan membandingkan tuple uji yang diberikan dengan tuple pelatihan yang mirip dengannya. Tuple pelatihan di deskripsikan dengan n atribut. Setiap tuple merepresentasikan suatu titik dalam ruang berdimensi n. Dengan cara ini, semua tuple pelatihan disimpan dalam ruang pola berdimensi n. Ketika diberi tuple yang tidak diketahui K-NN Classifier mencari ruang pola untuk K tuple pelatihan yang paling dekat dengan tuple yang tidak diketahui. Tuple pelatihan K ini adalah K “Ketetanggan Terdekat” dari tuple yang tidak diketahui.

“Kedekatan” di definisikan dalam metrik jarak, misalnya seperti jarak Euclidean. Jarak Euclidean antara dua titik atau tuple, misal $X_1 = (x_{11}, x_{12}, \dots, x_{1n})$ dan $X_2 = (x_{21}, x_{22}, \dots, x_{2n})$, adalah

$$dist(X_1, X_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2} \dots (1)$$

Dengan kata lain untuk setiap atribut numerik, kita mengambil selisih antara nilai yang y. nilai cosinus 0 berarti kedua vector berada pada 90 derajat satu sama lain (orthogonal) dan tidak memiliki kecocokan. Semakin dekat nilai cosinus ke 1, semakin kecil sudutnya dan semakin besar kecocokan antar vector. Ukuran kesamaan kosinus disebut sebagai ukuran non metrik.

2.4 TF - IDF

TF-IDF adalah singkatan dari istilah term frequency - inverse document frequency. Term frequency adalah berapa kali term tersebut muncul dalam dokumen. Document frequency adalah istilah untuk jumlah dokumen yang berisi term tertentu. Persamaan untuk nilai - nilai ini ditunjukkan pada gambar persamaan berikut [8] :

$$tf.idf(t, d) = tf(t, d).idf(t)$$

sesuai dari atribut pada tuple X_1 dan di tuple X_2 , mengkuadratkan dan mengakumulasinya. Akar kuadrat jumlah total jarak yang terakumulasi. Biasanya, kita menormalkan setiap atribut sebelum menggunakan persamaan (1). Ini membantu mencegah atribut dengan rentang awalnya besar melebihi atribut dengan rentang awalnya lebih kecil. Normalisasi Min - maks, misalnya, dapat digunakan untuk mengubah nilai v dari atribut numerik A menjadi v' dalam rentang [0,1] dengan menghitung.

$$v' = \frac{v - min_A}{max_A - min_A}, \dots (2)$$

Dimana min_A dan max_A adalah nilai minimum dan maksimum dari atribut A.

Untuk klasifikasi Nearest Neighbor, tuple yang tidak diketahui diberi kelas yang paling umum di antara k ketetanggan terdekat nya. Ketika $k = 1$, tuple yang tidak diketahui diberikan kelas dari tuple pelatihan yang paling dekat dengannya dalam ruang pola [5].

2.3 Cossine Similarity

Cossine similarity adalah ukuran kemiripan yang dapat digunakan untuk membandingkan dokumen atau memberikan peringkat dokumen sehubungan dengan vector dari query kata yang diberikan. Misalkan x dan y adalah dua vector untuk perbandingan. Menggunakan ukuran kosinus sebagai fungsi kesamaan.

$$sim(x, y) = \frac{x.y}{||x|| ||y||}, \dots (3)$$

Dimana $||x||$ adalah norma Euclidean dari vector $x = (x_1, x_2, \dots, x_p)$ di definisikan sebagai $\sqrt{X_1^2 + X_2^2 + \dots + X_p^2}$ secara konseptual, ini adalah Panjang vector. Demikian pula $||y||$ adalah norma Euclidean dari vektor y. ukuran tersebut menghitung cosinus dari sudut antara vector x dan

$$tf(t, d) = \sum_{i \in d}^{|d|} 1 \{ d_1 = t \}, \dots (4)$$

$$idf(t) = \text{Log} \left(\frac{|D|}{\sum_{d \in D} 1 \{ t \in d \}} \right)$$

2.5 Evaluasi Model Klasifikasi

Evaluasi terhadap suatu klasifier umumnya di lakukan menggunakan sebuah himpunan data uji, yang tidak digunakan dalam pelatihan classifier tersebut, dengan suatu ukuran tertentu. Terdapat sejumlah ukuran yang dapat digunakan untuk menilai atau mengevaluasi model klasifikasi, di antaranya adalah : accuracy atau tingkat pengenalan, error rate atau tingkat kesalahan atau kekeliruan klasifikasi, recall atau sensitivity atau true positive rate, specificity atau true negative

rate, precision, F-measure atau F_1 atau F-Score atau rata-rata harmonic dari precision dan recall, dan F_β (J. Han et al.2012), yang secara ringkas ilustrasikan pada Tabel 1.

Tabel 1. Ukuran Evaluasi Klasifikasi

No	Ukuran	Rumus
1.	Accuracy atau tingkat pengenalan	$\frac{TP + TN}{P + N}$
2.	Error rate atau tingkat kesalahan atau kekeliruan klasifikasi	$\frac{FP + FN}{P + N}$
3.	Recall atau sensitivitas atau true positive rate	$\frac{TP}{P}$
4.	Specificity atau true negative rate	$\frac{TN}{N}$
5.	Precision	$\frac{TP}{TP + FP}$
6.	F atau F_1 atau F-Score atau rata-rata harmonic dari precision dan recall	$\frac{2 \times precision \times recall}{precision + recall}$
7.	F_β , dimana β adalah sebuah bilangan riil nonnegatif	$\frac{1 + \beta^2 \times precision \times recall}{\beta^2 \times precision + recall}$

empat istilah sangat penting untuk memahami semua ukuran evaluasi dalam tabel tersebut adalah sebagai berikut :

1. TP atau True Positives adalah jumlah tuple positif yang dilabeli benar oleh classifier. Yang dimaksud tuple positif adalah tuple actual yang berlabel positif, seperti tuple dengan label Bonus = 'Ya'.
2. TN atau True Negative jumlah tuple negative yang dilabeli dengan benar oleh classifier. Yang dimaksud tuple negative adalah tuple

actual yang berlabel negative, seperti tuple dengan label Bonus = 'Tidak'.

3. FP atau False Positives adalah jumlah tuple negative yang salah dilabeli oleh classifier. Misalnya, sebuah tuple pelanggan yang berlabel Bonus = 'Tidak' tetapi dilabeli Bonus = 'Ya'
4. FN atau False Negatives adalah jumlah tuple positif yang salah dilabeli oleh classifier. Misalnya, sebuah tuple pelanggan yang berlabel Bonus = 'Ya' tetapi oleh classifier di labeli dengan Bonus = 'Tidak'.

3. Hasil dan Pembahasan

Adapun tahapan yang digunakan dalam pengolahan data mining pada penelitian ini, yaitu mengikuti tahapan *Knowledge Discovery in Database* (KDD). Adalah sebagai berikut :

3.1 Data Selection

Data yang digunakan adalah data transaksi peminjaman buku tahun 2015, 2016, 2017. Transaksi Data peminjaman buku pada tahun 2015,2016, 2017 ada sebanyak 3239 Buah transaksi. Dibawah ini merupakan contoh sebagian dari data transaksi peminjaman buku.

Pada tahapan ini penulis hanya menggunakan 2 atribut yang nanti akan digunakan untuk proses pengolahan data mining. Yaitu :

1. Judul Buku.
Berisi informasi judul buku.
2. Klasifikasi
Klasifikasi adalah atribut yang berisi informasi kode jenis buku, penggolongan buku berdasarkan jenis buku.

3.2 Preprocessing

Pada tahapan ini *preprocessing* ini akan dilakukan proses integrasi data untuk menghubungkan tabel data peminjaman, selanjutnya dilakukan data *cleaning* untuk menghasilkan *dataset* yang bersih sehingga dapat digunakan dalam tahap berikutnya yaitu mining. Berikut merupakan penjelasan dari kedua proses tersebut yaitu:

1. Integrasi Data
Tahap ini adalah proses penggabungan data dari berbagai *database* yang berbeda, sehingga data tersebut saling berintegrasi. Data integrasi dilakukan pada atribut - atribut yang mengidentifikasikan entitas-entitas yang unik. Pada tahapan ini tidak ada penggabungan data dikarenakan data yang diambil berasal dari satu *database*.
2. Data *Cleaning*
Pada tahapan ini data yang tidak *relevan*, *missing value*, dan *redundant* harus dibersihkan. Hal ini dikarenakan data yang *relevan*, tidak *missing value*, dan tidak *redundant* merupakan syarat awal dalam

melakukan data mining. Suatu data dikatakan *missing value* jika terdapat atribut dalam *dataset* yang tidak berisi nilai atau kosong, sedangkan data dikatakan *redundant* jika dalam satu *dataset* lebih dari satu *record* yang berisi nilai yang sama.

3.3 Transformation

Tahap Transformation merupakan tahapan merubah data kedalam bentuk yang sesuai untuk selanjutnya di proses dalam pengolahan data mining. Berikut pada gambar 2 adalah sebagian transformasi data transaksi buku 2015, 2016, 2017.

	A	B
1	text	label
2	fiqh manajemen zakat di Indonesia	2x0
3	jurnalistik hukum komunikasi massa jangkau era cyber communication milenium tiga	070
4	toward easy learning kiat sukses ajar di guru tinggi	570
5	10 jurus larang kok masih mau bisnis cara biasa	530
6	110 soal iman yang sehat akal	2x0
7	15 masalah fikih yang hangat kontroversial	2x0
8	20 salah dalam didik anak	570
9	22 nasihat abadi halus budi	300
10	25 pengaruh jiwa dan akal anak	150
11	254 hadits qudsi	2x2
12	40 hadis nabi saw telah imam khomeini atas hadis mistis dan akhlak	2x2
13	5 jam ajar olah data dengan spss17	600
14	530 hadits sahih bukhari - muslim	2x2
15	65 manusia langit jalan hidup sahabat rasulullah saw	2x0
16	7 langkah awal tuju karier idam	300
17	81 putus hukum rasulullah saw	2x0
18	9 presentasi kreatif dengan powerpoint 2007	600
19	95 strategi ajar multiple intelligences	600
20	a concise introduction to linguistics	400
21	a practical english grammar	400
22	a successful foreign language guide	400

Gambar 2. Transformasi Data Peminjaman Buku

Berdasarkan gambar 2 di atas hasil dari transformasi data setelah melewati tahapan *data selection*, dan *preprocessing* menghasilkan Data Buku sebanyak 2.742 Judul buku serta klasifikasinya, yang terdiri 2 atribut yaitu:

1. Text : Atribut yang berisi informasi judul buku.
2. Label : Atribut yang berisi informasi klasifikasi buku.

Jumlah buku yang terklasifikasi sebanyak 18 kelas. Berikut daftar buku yang klasifikasi. Pada Tabel 1 adalah daftar distribusi kelas buku.

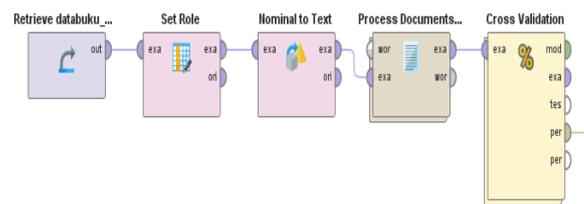
Tabel 2. Daftar Distribusi Buku

No.	kelas	label kelas	jumlah
1	2x0	Agama Islam	755
2	070	Pers, Jurnalisme, Penerbitan, Persuratkabaran	18
3	370	Pendidikan	271
4	330	Ilmu Ekonomi	139
5	300	Ilmu Sosial	559
6	150	Psikologi	113
7	2x2	Hadist dan yang berkaitan	71
8	600	Teknologi (Ilmu Terapan)	198
9	400	Bahasa	86
10	020	Ilmu Perpustakaan dan Informasi	37

11	2x1	Alquran dan Ilmu yang berkaitan	143
12	657	Akuntansi	23
13	500	Sains dan Matematika	47
14	340	Ilmu Hukum	183
15	375	Kurikulum	15
16	302.2	Komunikasi	55
17	800	Kesusastraan	22
18	700	Kesenian dan Rekreasi	7
Total			2742

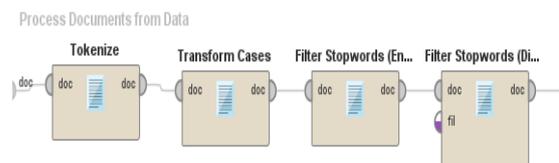
3.4 Data Mining

Untuk memprosesan data mining peneliti menggunakan software Rapidminer Studio versi 9.7.



Gambar 3. Pengolahan Data dengan Rapid Miner

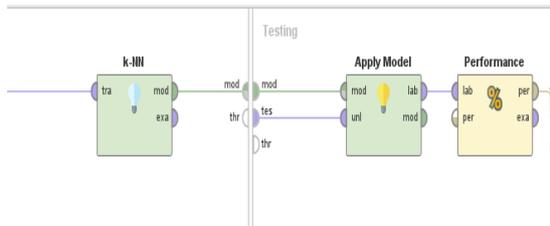
Sebelum pengolahan data ada 5 buah operator yang dibutuhkan yaitu, retrieve data buku, Set Role, Nominal to Text, Proses Document, Cross Validation. Operator retrieve data buku digunakan untuk membaca data buku yang akan di proses, operator Set Role digunakan untuk menentukan field atau bagian yang akan digunakan sebagai label. Pada penelitian ini atribut yang digunakan adalah atribut text dan atribut label. Set role juga digunakan untuk mengubah atribut role (misal, regular, special, label, id, dll). Operator nominal to text digunakan untuk mengubah jenis atribut nominal yang dipilih untuk teks. Serta memetakan semua nilai atribut ke dalam nilai string yang sesuai. Operator Process Document From Data digunakan untuk membersihkan data agar menjadi vector, vector yang digunakan adalah TF-IDF, yang dapat digunakan untuk perhitungan algoritma diantaranya adalah seperti subproses dari process dokumen form data gambar di bawah berikut :



Gambar 4. Sub proses dari Operator Proses Document From Data

Di dalam operator proses document terdapat operator Tokenize, Transform Cases, Filter

Stopword English, Filter Stopword Dictionary. Setelah dokumen menjadi vector yang dapat dihitung karena nilai teks tersebut menjadi nominal maka selanjutnya adalah proses validasi dengan cross validation, k-fold yang digunakan adalah K Fold=10. Di dalam operator cross validation terdapat subproses yang berisi operator validasi menggunakan algoritma K-NN dan ukuran kedekatan yang digunakan adalah cosine similarty, dan operator apply model yang digunakan untuk membuat model dari proses validasi, serta operator performance yang digunakan untuk mengukur kinerja dari model yang terbentuk.



Gambar 5. Sub proses dari Operator Cross Validation

Di dalam operator Cross validation terdapat operator KNN, Apply Model, Performance.

Berikut ini adalah hasil pengujian akurasi dengan menggunakan nilai k yang berbeda :

k	1	2	3	4	5	6	7	8	9
Accuracy (%)	60.83	60.79	66.59	68.53	69.18	69.62	70.28	71.23	72.07
k	10	11	12	13	14	15	16	17	18
Accuracy (%)	72.17	72.28	72.50	72.39	72.39	72.43	72.39	72.32	72.43

Gambar 6. Hasil Pengujian K=1 sampai K=18

Dari hasil pengujian seperti pada gambar 6 di atas dapat dilihat bahwa nilai akurasi semakin meningkat dengan bertambahnya nilai k hal ini disebabkan karena dimensi klasifikasi pada label dalam penelitian ini ada sebanyak 18 klasifikasi. Pada kasus penelitian ini dengan menambahkan nilai k maka semakin banyak pula rujukan yang dapat di gunakan untuk menentukan sebuah dokumen termasuk ke dalam kelompok kelas yang mana dalam suatu label atribut klasifikasi. Pada Tabel di atas dapat dilihat bahwa nilai akurasi tertinggi terdapat di k = 12 yaitu sebesar 72.50%.

3.5 Interpretation

Selain nilai akurasi pada penelitian ini juga dilihat nilai ukuran lain seperti precision dan recall, di karenakan distribusi data class yang digunakan tidak merata (imbalace) (seperti pada table 1). ukuran akurasi bekerja paling baik ketika kelas data di distribusikan secara merata. Pengukuran lain seperti sensitivity (atau recall), specificity, precision, F dan F_{β} lebih cocok untuk masalah keseimbangan kelas [5].

Class precision : dapat dianggap sebagai ukuran ketepatan (yaitu berapa persentase tuple yang diberi label positif sebenarnya).

Misal seperti pada tabel 2 Untuk kelas 2x0 class precision adalah 76.65% artinya sebanyak 805 tuple terprediksi ke dalam label class 2x0, tetapi hanya sebanyak 617 tuple yang masuk ke dalam label class 2x0 yang sebenarnya.

Dan class recall : adalah ukuran kelengkapan (berapa persentase tuple positif yang diberi label seperti itu)

Misal seperti pada tabel untuk kelas 2x0 class recallnya adalah 81.72% artinya dari 755 jumlah tuple 2x0, yang terprediksi benar adalah sebanyak 617 tuple.

Nilai dari precision dan recall biasanya digunakan bersama, dimana nilai precision dibandingkan dengan nilai recall, atau sebaliknya. Cara alternatifnya adalah dengan mengkombinasikan nilai precision dan recall menjadi satu ukuran dengan F measure (atau di kenal dengan F1 Score). F1 Score adalah harmonic mean dari nilai precision dan recall. Berikut adalah hasil dari F1 Score.

Tabel 3. Nilai F1 Score k=12

class	precision	recall	F1 Score
2x0	76.65%	81.72%	79.10%
070	100.00%	44.44%	61.53%
370	59.01%	70.11%	64.08%
330	72.73%	63.31%	67.69%
300	61.11%	72.81%	66.45%
150	87.50%	68.14%	76.62%
2x2	76.12%	71.83%	73.91%
600	85.37%	53.03%	65.42%
400	92.00%	80.23%	85.71%
020	90.00%	72.97%	80.60%
2x1	88.98%	73.43%	80.46%
657	100.00%	86.96%	93.03%
500	93.75%	63.83%	75.95%
340	67.20%	68.31%	67.75%
375	100.00%	60.00%	75.00%
302.2	86.96%	72.73%	79.21%
800	84.21%	72.73%	78.05%
700	57.14%	57.14%	57.14%

Dari table 3 di atas dapat dilihat bahwa dari nilai F1 Score, class yang mempunyai nilai > 72.50% akurasi, adalah 2x0, 150, 2x2, 400, 020, 2x1, 657, 500, 375, 302.2 800. Dan class yang mempunyai F1 Score < 72.50% akurasi, adalah 070, 370, 330, 300, 600, 340, 700.

4. Kesimpulan

Berdasarkan hasil analisis data dan pembahasan, dapat diuraikan kesimpulan yang di dapat dari hasil penelitian ini adalah :

1. Untuk memberikan tingkat kesalahan minimum nilai k pada Nearest Neighbor Classifier dapat dipilih dengan cara melakukan serangkaian set pengujian pada dataset.
2. Semakin besar dimensi klasifikasi pada label kelas maka nilai k akan semakin bertambah, dengan menambahkan nilai k maka semakin banyak pula rujukan yang dapat digunakan untuk menentukan sebuah dokumen masuk ke dalam suatu kelas.
3. Untuk dataset yang memiliki data kelas yang berdistribusi tidak merata (imbalance) maka evaluasi model klasifikasi tidak hanya dilihat dari nilai akurasi, nilai recall atau sensitivity, nilai precision, dan F measure.
4. Nilai k = 12 memiliki akurasi sebesar 72.50%, jika dilihat dari nilai F1 Score, maka kelas terklasifikasi yang nilainya lebih dari 70% adalah 2x0, 150, 2x2, 400, 020, 2x1, 657, 500, 375, 302.2 800, dan dapat digunakan untuk mengklasifikasikan buku dengan label klasifikasi tersebut. Sebaliknya untuk label kelas 070, 370, 330, 300, 600, 340, 700, tidak dapat di jadikan rujukan untuk klasifikasi buku, karena memiliki nilai F1 score dibawah 70%.

- [8] G. Miner, J. Elder IV, A. Fast, T. Hill, R. Nisbet, dan D. Delen, *Practical text mining and statistical analysis for non-structured text data applications*. Academic Press, 2012.

Daftar Pustaka

- [1] F. A. Wiranto, Supriyanto, dan Suryaningsih, "Perpustakaan Menjawab Tantangan Jaman," 1997.
- [2] M. R. Herga, "IMPLEMENTASI TEXT MINING SISTEM KLASIFIKASI DAN PENCARIAN KONTEN BUKU PERPUSTAKAAN MENGGUNAKAN ALGORITMA," hlm. 6.
- [3] E. K. Putri dan T. Setiadi, "PENERAPAN TEXT MINING PADA SISTEM KLASIFIKASI," vol. 2, hlm. 11, 2014.
- [4] D. T. Larose, *Discovering Knowledge in Data : an Introduction to Data Mining*. Canada: John Wiley & Sons, Inc., 2005.
- [5] J. Han, J. Pei, dan M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.
- [6] G. Gan, C. Ma, dan J. Wu, *Data clustering: theory, algorithms, and applications*. SIAM, 2007.
- [7] R. Feldman dan J. Sanger, *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge university press, 2007.



SYSTEMIC
INFORMATION SYSTEM AND INFORMATICS JOURNAL



9 772460 809004