



INSYST

Journal of Intelligent System and Computation

p-ISSN: 2621-9220

e-ISSN: 2722-1962

Volume 2 Nomor 1, April 2020



Published By **Lembaga Penelitian dan Pengabdian Masyarakat (LPPM)**
Institut Sains dan Teknologi Terpadu Surabaya (ISTTS)
formerly **Sekolah Tinggi Teknik Surabaya (STTS)**



Managed By
Departement of Informatics
Institut Sains dan Teknologi Terpadu Surabaya (ISTTS)

INSYST

Journal of Intelligent System and Computation

Volume 02 Nomor 01 April 2020

Editor in Chief:

Dr. Yosi Kristian, S.Kom, M.Kom.
Institut Sains dan Teknologi Terpadu Surabaya, Indonesia

Managing Editor:

Dr. Esther Irawati Setiawan, S.Kom., M.Kom.
Institut Sains dan Teknologi Terpadu Surabaya, Indonesia

Reddy Alexandro H., S.Kom., M.Kom.
Institut Sains dan Teknologi Terpadu Surabaya, Indonesia

Editorial Board:

Dr. Ir. Endang Setyati, M.T.
Institut Sains dan Teknologi Terpadu Surabaya, Indonesia

Ir. Edwin Pramana, M.App.Sc, Ph.D
Institut Sains dan Teknologi Terpadu Surabaya, Indonesia

Prof. Dr. Ir. Mauridhi Hery Purnomo, M.T.
Institut Sepuluh November, Indonesia

Hindriyanto Dwi Purnomo, Ph.D.
Universitas Kristen Satya Wacana, Salatiga, Indonesia

Hendrawan Armanto, S.Kom., M.Kom.
Institut Sains dan Teknologi Terpadu Surabaya, Indonesia

Dr. Lukman Zaman PCSW, M.Kom.
Institut Sains dan Teknologi Terpadu Surabaya, Indonesia

Dr. Diana Purwitasari, S.Kom., M.Sc.
Institut Sepuluh November, Indonesia

Dr. Joan Santoso, S.Kom., M.Kom.
Institut Sains dan Teknologi Terpadu Surabaya, Indonesia

INSYST

Journal of Intelligent System and Computation

Volume 02 Nomor 01 April 2020

Reviewer:

Teguh Wahyono, S.Kom., M.Cs.

Universitas Kristen Satya Wacana, Salatiga, Indonesia

Dr. Anang Kukuh Adisusilo, ST, MT.

Universitas Wijaya Kusuma, Surabaya, Indonesia

Dr. I Ketut Eddy Purnama, ST., MT.

Institut Sepuluh November, Indonesia

Prof. Dr. Benny Tjahjono, M.Sc.

Coventry University, United Kingdom

Dr. Ir. Gunawan, M.Kom.

Institut Sains dan Teknologi Terpadu Surabaya, Indonesia

Dr. Umi Laili Yuhana S.Kom., M.Sc.

Institut Sepuluh November, Indonesia

Dr. Tita Karlita, S.Kom., M.Kom.

Politeknik Elektronika Negeri Surabaya, Indonesia

Dr. Ir. Rika Rokhana, M.T.

Politeknik Elektronika Negeri Surabaya, Indonesia

Dr. I Made Gede Sunarya, S.Kom., M.Cs.

Universitas Pendidikan Ganesha, Indonesia

Dr. Yuni Yamasari, S.Kom., M.Kom.

Universitas Negeri Surabaya, Indonesia

Dr. Adri Gabriel Sooai, S.T., M.T.

Universitas Katolik Widya Mandira, Indonesia

INSYST

Journal of Intelligent System and Computation

Volume 02 Nomor 01 April 2020

Daftar Isi

Pengaruh Penggunaan Augmented Reality pada Pembelajaran Sistem Saluran Pernapasan dan Sistem Saluran Pencernaan pada tubuh manusia Rico Fiyah Hady, Hendrawan Armanto	1
Klasifikasi Tekstur Serat Kayu pada Citra Mikroskopik Veneer Memanfaatkan Deep Convolutional Neural Network Suriani Alamgunawan, Yosi Kristian	6
Analisis Trending Topik untuk Percakapan Media Sosial dengan Menggunakan Topic Modelling Berbasis Algoritme LDA Ahmad Syaifuddin, Reddy Alexandro Harianto, Joan Santoso	12
Hyper Sudoku Solver dengan Menggunakan Harris Hawks Optimization Algorithm Eric Dinata, Herman Budiarto, Hendrawan Armanto	20
Information Extraction Berbasis Rule Untuk Soal Ujian Stefanus Nico Soenardjo, Gunawan	28
Market Basket Analysis untuk Swalayan KSU Sumber Makmur dengan Algoritma FP Growth Ramadhan, Esther Irawati Setiawan	34
Pemanfaatan Behavior Tree dan Fuzzy Waypoint Tactic pada Game Strategi "War of Zombies" Michael Patria Christie, Andreas	40

Pengaruh Penggunaan *Augmented Reality* pada Pembelajaran Sistem Saluran Pernapasan dan Sistem Saluran Pencernaan pada tubuh manusia

Rico Fiyan Hady, *Teknologi Informasi Institut Sains dan Teknologi Terpadu Surabaya (iSTTS)*,
Hendrawan Armanto, *Teknik Informatika Institut Sains dan Teknologi Terpadu Surabaya (iSTTS)*.

Abstrak—Abstrak dituliskan *Augmented Reality (AR)* merupakan peluang baru untuk mempermudah kegiatan manusia dalam mengilustrasikan sesuatu. Karena pada dasarnya konsep *AR* adalah menampilkan objek grafis menjadi tampak nyata seolah ada dihadapan kita. Dalam segi pendidikan hal ini akan sangat berpengaruh mengingat konsep dari *AR* itu sendiri. Aplikasi ini akan mempermudah interaksi dua arah mengenai materi yang sedang disampaikan khususnya dalam jenjang sekolah dasar. Peserta didik seolah akan melihat objek yang sedang diajarkan berada didepan mereka. Hal ini membuat peserta didik dapat mengerti secara kompleks dan terstruktur mengenai sistem pencernaan dan sistem pernapasan pada tubuh manusia dengan hanya menggunakan *smartphone* dan sebuah gambar sebagai marker. Hasil penelitian disalah satu sekolah mengungkap hasil belajar siswa dengan menggunakan metode pembelajaran *AR* akan meningkatkan pemahaman peserta didik dalam mengenal organ sesuai dengan kuisioner yang sudah diberikan sebesar 36%. Aplikasi ini berhasil membantu belajar peserta didik untuk mempelajari sistem pernapasan dan sistem pencernaan pada tubuh manusia dengan menggunakan teknologi *Augmented Reality (AR)*.

Kata Kunci—*Augmentend Reality, Annona, Sistem Pencernaan, Sistem Pernapasan, Pembelajaran*

I. PENDAHULUAN

Media pendidikan sangat diperlukan sebagai perantara penyampai pesan, guna meminimalkan kegagalan selama proses komunikasi berlangsung. Bethay (2014) mengemukakan bahwa proses belajar adalah proses penyampaian pesan/materi dari pemberi pesan (guru) ke penerima pesan (peserta didik). Media pembelajaran yang menarik juga sangat diperlukan bagi siswa SD, dikarenakan sistem pembelajaran yang diberikan pada siswa SD adalah belajar sambil bermain. Media pembelajaran sangat diperlukan guru untuk membantu menyampaikan materi dalam sebuah proses pembelajaran. Proses pembelajaran yang baik haruslah memuat aspek interaktif, menyenangkan, menantang, memotivasi dan memberikan ruang lebih bagi siswa untuk dapat mengembangkan kreativitas dan kemandirian, sesuai dengan bakat dan minat siswa [1]. Meskipun guru hanya sebagai fasilitator dalam sebuah pembelajaran, dan siswa yang dituntut untuk lebih aktif, guru harus mampu membuat suasana pembelajaran yang menyenangkan untuk merangsang siswa lebih aktif dalam belajar.

Rico Fiyan Hady, Departemen Teknologi Informasi, Institut Sains dan Teknologi Terpadu Surabaya, Surabaya, Jawa Timur, Indonesia (e-mail: rico.fiyan.hady@gmail.com)

Hendrawan Armanto, Departemen Teknik Informatika, Institut Sains dan Teknologi Terpadu Surabaya, Surabaya, Jawa Timur, Indonesia (email:hendrawan@stts.edu)

Dalam pembelajaran jenjang sekolah dasar atau SD salah satu mata pelajaran yang diampu adalah IPA yang dimana bagian pada pembahasannya adalah saluran pencernaan dan saluran pernapasan pada manusia. Mengingat pembelajaran pada jenjang SD rata-rata siswa masih berumur 9-10 tahun. Dalam mengajarkan siswa jenjang SD untuk materi saluran pencernaan dan saluran pernapasan pada tubuh manusia tentunya siswa pada umumnya akan diberikan gambaran sistem, fungsi, dan objek dari saluran pernapasan dan juga saluran pencernaan pada tubuh manusia. Dengan demikian siswa tersebut hanyalah berimajinasi tentang bentuk dari setiap organ yang terkait dalam sistem tersebut. Secara teknis guru menjelaskan materi tersebut secara lisan dan arsitektur dari materi tersebut digambarkan dengan menggunakan spidol di papan tulis, kemudian diberikan anak panah untuk menunjukkan tahapan atau setiap alurnya. Dengan media tersebut siswa akan menjadi paham, tetapi setelah beberapa lama siswa akan merasa bosan karena gambar terlihat sederhana dan hanya terlihat satu dimensi. Gambar tersebut kurang memotivasi siswa dalam belajar, sehingga proses pembelajaran dengan materi saluran pencernaan dan saluran pernapasan pada tubuh manusia ini kurang menarik dan membuat siswa merasa jenuh dan merasa kesulitan untuk menyelesaikan soal tentang materi tersebut. Hal ini mengakibatkan hasil belajar siswa rendah.

Augmented Reality adalah teknologi yang menggabungkan benda maya dua dimensi 2D dan ataupun tiga dimensi 3D ke dalam sebuah lingkungan nyata tiga dimensi lalu memproyeksikan benda-benda maya tersebut dalam waktu nyata. *Augmented Reality* memiliki banyak peluang untuk terus dikembangkan, tidak ketinggal dalam pendidikan. Dari berapa hasil penelitian dalam bidang pendidikan, *augmented reality* menunjukkan hasil yang positif dan membuktikan bahwa *augmented reality* dapat diterapkan dengan baik dalam dunia pendidikan, khususnya sebagai sebuah media pembelajaran.

Teknologi ini sangat tepat digunakan dalam pembelajaran IPA khususnya pada sistem saluran pernapasan dan saluran pencernaan pada tubuh manusia. Dengan teknologi ini berbagai bentuk dari setiap organ yang berkaitan pada sistem saluran pernapasan dan saluran pencernaan manusia akan ditampilkan dalam bentuk gambar tiga dimensi yang dapat bergerak, sehingga gambar terkesan hidup dan tidak membosankan. Animasi tiga dimensi dalam *augmented reality* ini dapat membantu siswa dalam memahami materi yang diberikan oleh pendidik, seperti proses awal sampai akhir seperti pada sistem saluran pernapasan mulai dari oksigen masuk kehidung sampai dihirup kembali dari hidung menjadi karbondioksida begitu juga pada saluran pencernaan manusia mulai dari makanan yang masuk dari mulut sampai menjadi kotoran dan dikeluarkan dari dubur akan lebih jelas jika ditampilkan

dengan teknologi augmented reality karena alur dan tahapannya dibuat bergerak. Selain itu semua organ yang berkesinambungan dalam sistem tersebut dapat akan terlihat nyata dihadapan kita. Berdasarkan uraian diatas penelitian ini dilakukan dengan tujuan yang ingin dicapai, yaitu untuk mengetahui pengaruh media pembelajaran berbasis augmented reality terhadap peningkatan hasil belajar siswa tingkat SD pada kelas VI khususnya untuk mata pelajaran IPA dengan materi sistem saluran pencernaan dan saluran pernapasan pada tubuh manusia.

II. PENELITIAN TERKAIT

Penelitian yang dilakukan oleh Cheng-Ping Chen dan Chang-Wha Wa (2015) yang berjudul *The Effect of learning style on Mobile Augmented-Reality Facilitated English Vocabulary Learning*. [2] Dalam tulisannya mereka menjelaskan bahwa mereka merancang sebuah model pembelajaran baru dalam belajar berbahasa Inggris. Dengan model pembelajaran seperti ini akan membuat siswa untuk dapat benar-benar melihat, menyentuh, dan berinteraksi dengan *vocabulary* yang mereka pelajari. Dengan bermodalkan internet dan *smartphone* yang berkamera siswa akan mengetahui sama benda dalam bahasa inggris beserta dengan ejaan inggrisnya apabila kamera pada *smartphone* tersebut disorotkan pada suatu benda di sekitar kita. Dalam implementasinya penulis membuat suatu permainan dalam belajar berbahasa inggris. Pertama yang harus dilakukan adalah membagi menjadi beberapa tim. Lalu tahapan dalam permainan ini yaitu, 1) mereka membuat petunjuk untuk item pertama. 2) mereka akan mendapat informasi nama dan cara mengeja item tersebut dalam bahasa inggris jika menemukan item yang sesuai dengan petunjuk. 3) lalu mereka harus melakukan tangkap layar dan menirukan pengucapan item tersebut dalam bahasa dengan baik dan benar untuk mendapatkan item selanjutnya.

Penelitian yang dilakukan oleh Malek El Kouzi, Abdihakim Mao, dan Diego Zambrano (2019) yang berjudul *An Educational Augmented Reality Application for Elementary School Students Focusing on the Human Skeletal System*. Adalah sebuah penelitian yang dilakukan pada jenjang Sekolah Dasar yang berfokus pada sistem gerak pada tulang manusia [3]. Penulis memaparkan bahwa partisipan pada pembelajarannya adalah siswa yang berumur 10 dan 11 pada tingkatan kelas 5 di Sekolah Dasar. Total siswa pada kelas tersebut adalah 30 siswa dengan 1 orang guru. Secara garis besar konsep AR yang mereka buat mengikuti buku pelajaran yang sedang siswa pelajari yang dimana gambar sekletal 2D yang disebut dengan marker nantinya disorot dengan menggunakan tablet dan akan muncul diatas gambar sekletal 2D tadi tetapi dalam bentuk 3D. Peserta didik dapat merotasi gambar 3D tersebut dengan cara merotasikan buku mereka sehingga akan nampak bagian sisi lain dari gambar 3D skeletal itu sendiri. Pada jurnalnya penulis mengatakan bahwa mereka mempunyai 5 contoh gambar tulang yang dapat di interaksikan dengan siswa. Lima gambar tulang yang dimaksud meliputi, tulang kepala manusia, tulang selangka, tulang rusuk, tulang paha, dan tulang pinggang.

Penelitian yang dilakukan oleh Bambang Cahyono, Muhammad Bambang Firdaus, dan Edy Budiman (2018) yang berjudul *Augmented Reality Applied to Geometry Education*. Penulis memaparkan bahwa pelajaran

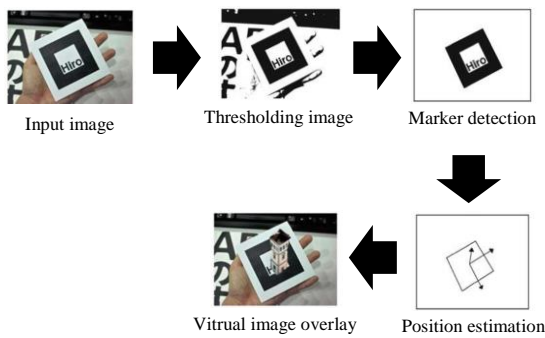
matematika adalah mata pelajaran yang rata-rata tidak disukai oleh anak kecil maka dari itu perlu untuk membuat cara belajara dengan model lain yaitu dengan bentuk 3D sehingga siswa dapat membangun apa yang mereka inginkan dari pengalaman yang mereka miliki serta dapat berinteraksi [4]. Tidak hanya berimajinasi atas bangunan geometri yang sudah di gambarkan oleh gurunya. Dalam implementasinya penulis menggunakan strategi *Multimedia Development Life Cycle* (MDLC) yang tahapan ini dibagi menjadi 5 yaitu, *Concept, Material Collecting, Assembly, Testing, Distribution*

Penelitian yang dilakukan oleh Mohammad Fahim Hossain, Sudipta Barman, AKM Baharul Haque (2019) adalah penelitian yang dilakukan di Bangladesh yang dimana pada setiap hari siswa diharuskan belajar dua atau lebih bahasa yaitu Bengali dan juga Bahasa Inggris [5]. Mereka harus mengenal 11 huruf vokal dan juga 39 konsonan atau huruf mati pada bahasa Bengali dan pada Bahasa Inggris terdapat 26 huruf dan simbol – simbol yang berbeda pada setiap bahasa. Jadi total yang harus dihafal adalah 76 simbol yang berbeda. Hal ini tentunya cukup menyulitkan mereka untuk menghafal satu persatu setiap simbol yang ada sehingga akan timbul rasa bosan. Dan dari sini seorang guru mengerti bahwa proses pembelajaran haruslah kreatif dan interaktif. Kita semua tahu bahwa anak kecil jaman sekarang mereka terlahir dengan *smartphone* yang selalu ada di tangan mereka yang artinya generasi muda mayoritas lebih paham mengenai cara menggunakan *smartphone* dan mereka menggunakan ini sebagai salah satu sumber belajar selain buku. Dengan begitu penulis memberikan solusi sebuah *Augmented Reality* (AR) yang berada pada genggamannya. Sebuah buku dengan teknologi *Augmented Reality* (AR) akan mereka butuhkan dalam belajar. Buku ini akan penuh dengan visualisasi dan suara yang tampak nyata dengan model-model 3D buah dan juga semua huruf dari kedua bahasa. Pada setiap model 3D juga dapat di putar untuk dapat mengetahui setiap sisinya sehingga hal ini akan menarik.

Penelitian yang dilakukan oleh Carolina Mateo dan Heriot Watt (2018) mereka menceritakan bahwa untuk mendapatkan level pemahaman yang tinggi dalam memahami sebuah konsep abstrak maka dibutuhkan visualisasi dalam bentuk objek 3D dan manipulasi yang kompleks secara teori dari segi matematika sebagai contoh adalah tentang *electromagnetic* (EM) [6]. Sebuah proyek contoh dalam pengembangan tentang *elctromagnetic* tentunya membutuhkan itu demi tercapainya proyek yang baik dengan finishing sesuai sasaran yang sudah direncanakan. Siswa yang belajar dalam bidang *electromagnetic* (EM) tentunya membutuhkan fokus yang tinggi khususnya bagi siswa pada tahun ke-3 dimana mereka membutuhkan level pemahaman yang tinggi. Untuk mendapatkan itu maka dibuat sebuah simulasi 3D dengan teknologi *Augmented Reality* (AR) sehingga dapat memvisualisasikan dan meningkatkan pemahaman siswa mengenai bidang yang dimaksud.

III. AUGMENTED REALITY

Augmented Reality (AR) dapat diartikan sebagai teknologi yang menggabungkan atau mencitrakan antara benda maya dua dimensi atau tiga dimensi kedalam sebuah lingkungan nyata tiga dimensi lalu memproyeksikan benda maya tersebut secara real-time. Untuk mencitrakan gambar tiga dimensi tersebut, sistem Augmented Reality terlebih dahulu harus melakukan penglihatan atau vision terhadap lingkungan yang akan dicitrakan objek virtual, kemudian dilakukanlah proses tracking terhadap image target dan menentukan letak citra objek virtual tersebut. Kemudian objek tersebut akan dikenali, atau dianalisis. Setelah dikenali dan di analisis posisi dan orientasinya, maka komputer akan melakukan proses pencitraan objek tersebut, dan ditampilkan pada perlengkapan display [7]. Proses tracking pada AR merupakan gambaran singkat tentang proses kerja AR. Proses dimulai dari pengambilan gambar marker dengan kamera. Kemudian masuk ke dalam proses tracking (pelacakan atau registrasi marker) yang meliputi : thresholding, marker detection dan proses render menggabungkan gambar asli dan komponen virtual menggunakan perhitungan pose dan orientasi marker. Hasil keluaran pelacakan marker ditampilkan pada layar komputer atau smartphone.



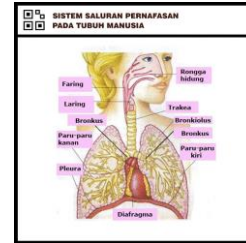
Gambar 1. Proses tracking AR (Nanang, 2018)

IV. MARKER

Merupakan sebuah penanda khusus yang memiliki pola tertentu yang akan dikenali oleh kamera, sehingga saat kamera mendeteksi marker maka objek 3 dimensi dapat ditampilkan. Augmented Reality dapat dikategorikan menjadi dua, berdasarkan ada atau tidaknya marker yaitu menggunakan marker based tracking dan tanpa marker (markless). Marker based tracking dimana komputer akan mengenali posisi dan orientasi marker dan menciptakan dunia virtual atau 3 dimendis yaitu titik (0,0,0) dan 3 sumbu yaitu X, Y, Z [8]. Marker based tracking dapat merupakan berupa foto sebuah objek nyata atau gambar buatan dengan pola unik. Markless merupakan sebuah metode pelacaan (tracking) dimana AR menggunakan objek di dunia nyata sebagai marker atau tanpa menggunakan marker buatan.

Marker atau juga bisa disebut sebagai image target memiliki kriteria untuk dapat dilacak oleh sistem AR yang meliputi :

- fitur gambarnya kaya (polanya unik) misal gambar pemandangan, gambar sekumpulan orang, kolase dan lain sebagainya.
- Kontrasnya bagus, gelap dan terangnya jelas
- Tidak ada pengulangan pola, misal lapangan rumput, kotak-kotak
- Grafik warnanya 8 atau 24 bit format PNG atau JPG, ukurannya kurang dari 2 MB JPGs haruslah RGB atau grayscale (bukan CMYK)



Gambar 2. Contoh marker

V. METODE PENELITIAN

Metode pengujian digunakan untuk mengukur tingkat perkembangan dan pengenalan serta mengingat siswa sebelum dan setelah pembelajaran dengan menggunakan model AR ini. Jenis pengujian berupa *pre-test*, *post-test*, dan kuisioner.

Proses pengujian dengan *pre-test* dan *post-test* dilakukan untuk mengetahui perbedaan dari sebelum dan sesudah pembelajaran dengan menggunakan model AR ini. Jenis soal yang diberikan pada *pre-test* dan *post-test* berupa soal uji pada umumnya

A. Soal Pre-Test

Tujuan dari *pre-test* untuk mengetahui tingkat kemampuan pemahaman siswa terhadap materi sistem saluran pernapasan dan sistem saluran pencernaan pada tubuh manusia sebelum diberikan model pembelajaran berbasis *Augmented Reality*. Dan hasil jawaban dari *pre-test* nantinya akan dijadikan hasil perbandingan dari hasil jawaban *post-test*

B. Soal Post-Test

Merupakan soal yang diberikan kepada peserta didik ketika sudah di bagi menjadi dua kelompok yaitu kelompok eksperimen dan kelompok konvensional yang dimana pembelajaran antar kelompok tersebut adalah berbeda. Pada kelompok eksperimen menggunakan pembelajaran berbasis *Augmented Reality*. Pada kelompok konvensional menggunakan pembelajaran seperti biasa yang sudah dilakukan sehari-hari pada kelas tersebut. Tujuan dari memberikan soal *post-test* ini adalah untuk mengetahui pengukuran tingkat kemampuan dan pemahaman siswa terhadap perbedaan materi yang telah ditentukan. Dan hasil dari soal *post-test* ini akan disandingkan dengan hasil dari soal *pre-test*.

C. Metode Pembelajaran

Secara teknis pada metode pembelajaran ini seluruh siswa dengan jumlah 50 siswa akan dibagi menjadi dua kelompok yaitu kelompok eksperimen dan kelompok konvensional yang dimana setiap kelompok terdiri dari 25 orang.

Kelompok eksperimen akan menerima pembelajaran dengan menggunakan media pembelajaran *Augmented Reality* (AR) sedangkan Kelompok Konvensional akan menerima pembelajaran secara konvensional sebagaimana guru di sekolah tersebut memberikan materi terhadap siswanya.

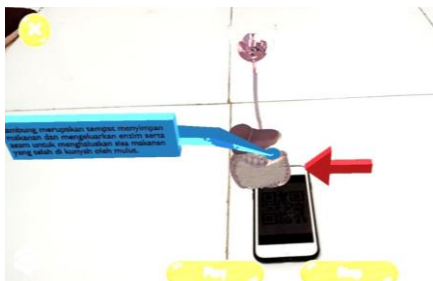
D. Quiz

Pada kelompok eksperimen akan terdapat quiz yang dimana pada quiz tersebut terdapat tiga tingkatan yaitu mudah, sedang, sulit. Pada implementasinya quiz ini terdapat menu tersendiri yang dimana setelah pembelajaran mengenai alur dari setiap sistem dan juga mengenal bentuk serta nama organ maka user akan dihadapkan pada menu quiz untuk melatih daya ingat dari pembelajaran sebelumnya. Adapun ruang lingkup quiz pada masing-masing tingkatannya adalah sebagai berikut:

- a. **Tingkatan mudah**
 Pada tingkatan ini user atau siswa akan dihadapkan dengan pertanyaan-pertanyaan mengenai nama-nama organ yang berperan pada setiap alurnya.
- b. **Tingkatan sedang**
 Untuk pertanyaan quiz pada tingkatan sedang ini user atau siswa akan dihadapkan dengan pertanyaan-pertanyaan dalam urutan organ pada setiap sistemnya.
- c. **Tingkatan sulit**
 Pada tingkatan sulit ini pertanyaan-pertanyaan yang akan muncul adalah tentang fungsi dari setiap organ sesuai dengan penjelasan dalam animasi



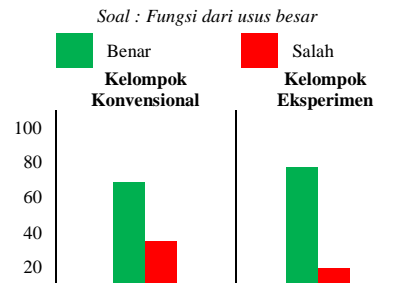
Gambar 3. Diatas marker ada model organ tubuh manusia



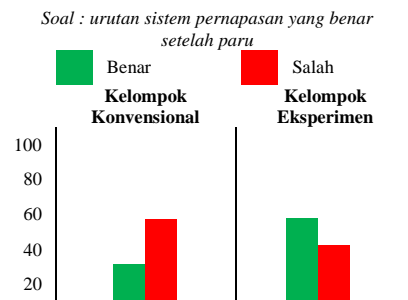
Gambar 4. Animasi alur sistem pencernaan Pada tubuh manusia

VI. HASIL

Dengan metode yang sudah dijelaskan dalam melakukan penelitian ini maka hasil yang dicapai adalah berupa nilai siswa ketika *pre-test* dan *post-test* yang dimana hasil tersebut disandingkan untuk mengetahui signifikansi pengaruh media pembelajaran dengan menggunakan *Augmented Reality*(AR) di jenjang sekolah dasar khususnya pada materi sistem saluran pernapasan dan sistem saluran pencernaan pada tubuh manusia. Pengelolaan data analisis statistik untuk mengukur signifikansi antara hasil *pre-test* dan *post-test* menggunakan ANOVA (*Analysis Of Variance*). Hasil hitung dari Anova ini yang dimana sebagai acuan adalah *F* hitung (*F*) dan *F* Tabel (*F crit*) apabila *F* hitung lebih kecil dari pada *F* tabel maka media pembelajaran ini tidak berpengaruh terhadap hasil belajar siswa, sebaliknya apabila *F* hitung lebih besar dari pada *F* tabel maka media pembelajaran ini berpengaruh terhadap hasil belajar siswa. dapat dilihat pada Tabel 3.



Gambar 5. Diagram jawaban soal Tentang fungsi dari usus besar



Gambar 6. Diagram jawaban soal Tentang urutan dari sistem pernapasan

Berdasarkan hasil kuisioner tentang aplikasi yang sudah digunakan kepada peserta didik yang untuk mengetahui kelayakannya, menganalisa hasil kuisioner tersebut dengan acuan persentase dan didapatkan data sebagai berikut :

Tabel 1. Konversi hasil kegunaan sistem

No.	Persentase	Interpretasi
1	0% - 20%	Sangat Tidak Layak
2	21% - 40%	Tidak Layak
3	41% - 60%	Cukup Layak
4	61% - 80%	Layak
5	81% - 100%	Sangat Layak

Tabel 2.
Hasil kuisioner uji kelayakan aplikasi

Pertanyaan	Persentase	Uraian
Aplikasi mudah digunakan	80	Layak
Suka bermain dengan AR	76	Layak
Membantu dalam pembelajaran	64	Layak
Tertarik AR kedepannya	76	Layak
Membantu dalam visualisasi	60	Cukup layak

Tabel 3. Hitung hasil signifikansi menggunakan ANNOVA

SUMMARY				
Groups	Count	Sum	Average	Variance
pre-test	25	1795	71,8	42,66667
post-test	25	1950	78	16,75

ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	480,5	1	480,5	16,17391	0,000204	4,042652
Within Groups	1426	48	29,70833			
Total	1906,5	49				

VII. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan dengan melakukan implementasi teknologi *Augmented Reality* (AR) terhadap pembelajaran pada jenjang sekolah dasar khususnya materi sistem saluran pernapasan dan sistem saluran pencernaan pada tubuh manusia dapat mempengaruhi motivasi dan hasil belajar peserta didik.

Penelitian masa mendatang akan dikembangkan lebih jauh pada materi pembelajaran yang lain atau pada jenjang sekolah yang lebih tinggi misalnya pada jenjang sekolah menengah pertama atau jenjang sekolah menengah atas yang mencakup materi lebih luas dan mendalam

DAFTAR PUSTAKA

- [1] I. Mustaqim, "Pengembangan media pembelajaran berbasis augmented reality," *J. Edukasi Elektro*, vol. 1, no. 1, 2017.
- [2] C. Shao, G. L. Ciampaglia, O. Varol, K. C. Yang, A. Flammini, and F. Menczer, "The spread of low-credibility content by social bots," *Nat. Commun.*, vol. 9, no. 1, 2018, doi: 10.1038/s41467-018-06930-7.
- [3] M. El Kouzi, A. Mao, and D. Zambrano, "An educational augmented reality application for elementary school students focusing on the human skeletal system," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2019, pp. 1594–1599.
- [4] B. Cahyono, M. B. Firdaus, E. Budiman, and M. Wati, "Augmented reality applied to geometry education," in *2018 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT)*, 2018, pp. 299–303.
- [5] M. F. Hossain, S. Barman, and A. K. M. B. Haque, "Augmented Reality for Education; AR Children's Book," in *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*, 2019, pp. 2568–2571.
- [6] C. Mateo-Segura, "A new Augmented-Reality platform for Electromagnetic Education," in *2018 International Conference on Electromagnetics in Advanced Applications (ICEAA)*, 2018, pp. 174–177.
- [7] C. Juan, F. Beatrice, and J. Cano, "An augmented reality system for learning the interior of the human body," in *2008 Eighth IEEE*

- [8] *International Conference on Advanced Learning Technologies*, 2008, pp. 186–188.
L. Chien-Yu, J.-T. Chao, and H.-S. Wei, "Augmented reality-based assistive technology for handicapped children," in *2010 International Symposium on Computer, Communication, Control and Automation (3CA)*, 2010, vol. 1, pp. 61–64.

Klasifikasi Tekstur Serat Kayu pada Citra Mikroskopik Veneer Memanfaatkan Deep Convolutional Neural Network

Suriani Alamgunawan, *Teknologi Informasi Institut Sains dan Teknologi Terpadu Surabaya (iSTTS)*,
Dr. Yosi Kristian, *Teknik Informatika Institut Sains dan Teknologi Terpadu Surabaya (iSTTS)*.

Abstrak—Convolutional Neural Network sebagai salah satu metode Deep Learning yang paling sering digunakan dalam klasifikasi, khususnya pada citra. Terkenal dengan kedalaman dan kemampuan dalam menentukan parameter sendiri, yang memungkinkan CNN mampu mengeksplor citra tanpa batas. Tujuan penelitian ini adalah untuk meneliti klasifikasi tekstur serat kayu pada citra mikroskopik veneer dengan CNN. Model CNN akan dibangun menggunakan MConv dan arsitektur lapisan akan didesain menggunakan EfficientNet. Diharapkan dapat tercapai tingkat akurasi yang tinggi dengan penggunaan jumlah parameter yang sedikit. Dalam penelitian ini akan mendesain empat model arsitektur CNN, yaitu model RGB tanpa contrast stretching, RGB dengan contrast stretching, Grayscale tanpa contrast stretching dan Grayscale dengan contrast stretching. Proses ujicoba akan mencakup proses pelatihan, validasi dan uji pada masing-masing input citra pada setiap model arsitektur. Dengan menggunakan penghitungan softmax sebagai penentu kelas klasifikasi. SGD optimizer digunakan sebagai optimization dengan learning rate 1e-1. Hasil penelitian akan dievaluasi dengan menghitung akurasi dan error dengan menggunakan metode F1-score. Penggunaan channel RGB tanpa contrast stretching sebagai citra input menunjukkan hasil uji coba yang terbaik.

Kata Kunci—Contrast Stretching, Convolutional Neural Network, EfficientNet, Grayleveling.

I. PENDAHULUAN

Klasifikasi tekstur serat kayu sangat menarik untuk digunakan sebagai bahan penelitian. Veneer sebagai salah satu hasil dari olahan kayu yang banyak digunakan sebagai bahan untuk mempercantik dan memperindah peralatan rumah tangga dan desain interior. Dengan keindahan dan keunikan yang timbul dari serat dan warna kayu asalnya, hal ini membuat veneer mempunyai nilai lebih sebagai komoditas dagang. Seluruh kayu yang ada dapat diolah menjadi veneer, hal ini memberikan nilai tambah sekaligus tingkat kesulitan dalam mengenali jenis kayu. Dikarenakan jenis ragam, warna, tekstur dan arah serat kayu sangat berpengaruh dalam pengenalan veneer. Kemampuan dan pengalaman sangat diperlukan dalam mengenali jenis kayu asal veneer.

Dengan pemanfaatan digital mikroskopik sebagai salah satu media penunjang dalam pencitraan serat kayu pada veneer. Dimana digital mikroskopik digunakan sebagai salah satu alat bantu yang berfungsi untuk mengambil citra serat dengan

pembesaran 1000x. Misalnya seperti penelitian yang dilakukan oleh Luiz G. Hafemann [1] dengan menggunakan digital mikroskopik dan penerapan model deep neural network dapat mencapai tingkat akurasi yang tinggi dalam mengklasifikasikan spesies hutan Brasil. Dan digunakan pada penelitian yang dilakukan oleh Jinmei Yang [2] untuk mengklasifikasi, mendeteksi dan mengidentifikasi 25 jenis kayu langka.

Diharapkan dengan penerapan metode klasifikasi berbasis Convolutional Neural Networks dalam perancangan sistem pengklasifikasian jenis kayu pada tekstur serat kayu dengan citra mikroskopis dapat memberikan tingkat akurasi pengklasifikasian di atas 80%.

II. LANDASAN TEORI

Gunakan dokumen ini sebagai template agar format paper anda sesuai dengan format paper yang kami perlukan dan ketikan teks anda di dalamnya. Akan tetapi apabila anda ingin melakukan layouting sendiri berikut adalah panduannya.

A. Convolutional Neural Network

Struktur jaringan CNN [3] pada dasarnya dapat dikelompokkan menjadi dua tahapan yaitu Feature Learning dan Classification. Dimana dalam tahapan feature learning meliputi Convolutional Layer, Rectified Linear Unit (ReLU) dan Pooling Layer. Sedangkan pada tahapan Classification mengerucut pada proses pengklasifikasian neuron untuk dapat menjadi satu kelas terpilih.

Lapisan feature learning berguna untuk mentransasikan input menjadi sebuah fitur berdasarkan ciri input dalam bentuk angka vector. Kemampuan pembelajaran dari CNN dimulai dengan pembelajaran dalam pengenalan sebuah objek pada beberapa regional, dengan menggunakan sliding window. Proses window ini bergerak dari satu regional ke regional yang lain dengan tujuan untuk mencari aspek yang paling informatif, yang dikenal dengan sebutan filter. Filter yang sama akan bergerak pada beberapa regional yang berbeda.

Setiap pergerakan filter dan input yang bergeser berkali-kali menghasilkan suatu fitur baru sebagai output atau biasa disebut dengan feature map atau activation map. Penghitungan output berdasarkan neuron yang terhubung ke dalam input, dimana masing-masing akan menghitung produk titik (dot function) antara bobot dan wilayah yang

terhubung dengan volume input. Nilai pixel input citra akan dikalikan dengan nilai pada filter.

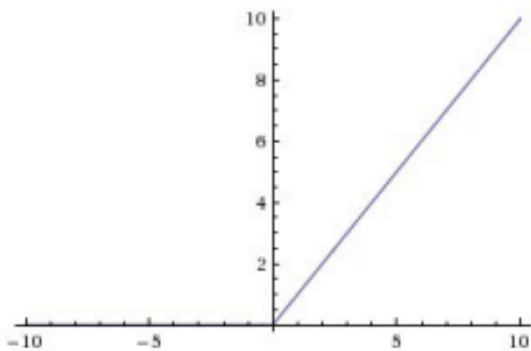
Dalam proses konvolusi terdapat beberapa pengaturan yang dapat dilakukan, meliputi ukuran filter, jumlah stride dan penggunaan zero padding. Filter size ditentukan tergantung pada berapa banyak aspek informative yang ingin ditangkap saat proses pengolahan sebuah regional.

B. Fungsi Aktivasi

Fungsi aktivasi merupakan bagian penting dalam sebuah jaringan neural network yang tidak dapat kita lewatkan. Dimana fungsi aktivasi berperan penting untuk membuat sebuah jaringan neural menjadi non-linear. Terdapat banyak fungsi aktivasi yang dapat digunakan dalam model, setiap fungsi aktivasi memiliki kekurangan dan kelebihan masing-masing. Fungsi aktivasi yang akan dibahas adalah ReLU (Rectified Linear Unit), Sigmoid dan ReLU6.

ReLU (Rectified Linear Unit). ReLU akan menghilangkan vanishing gradient dengan cara penerapan fungsi aktivasi elemen. Aktivasi elemen akan dilakukan dengan membuat pembatas pada bilangan nol, artinya apabila $x \leq 0$ maka $x = 0$ dan apabila $x > 0$ maka $x = x$.

$$f(x) = \max(0, x) \tag{1}$$

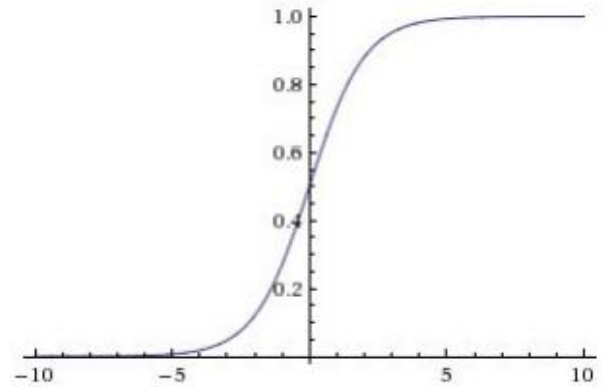


Gambar 1 . Rectified Linear Unit (ReLU)

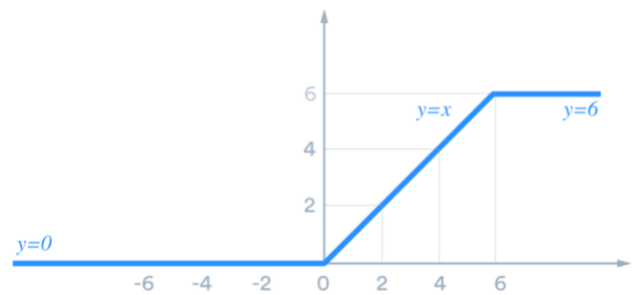
Sigmoid adalah fungsi aktivasi yang akan mengubah angka menjadi non-linear dengan rentang nilai dari 0 sampai 1. Sigmoid mulai jarang digunakan dikarenakan sigmoid akan mematikan gradient, di saat aktivasi neuron mengeluarkan nilai yang berada pada akhir 0 atau 1, maka gradient di rentangan ini bernilai hampir 0. Dikarenakan gradient memiliki nilai sangat kecil, sigmoid akan mematikan gradient. Selain itu output sigmoid tidak zero-centered, yang berdampak pada ketidakdinamisan gradient descent. Jika data selalu positif maka gradient pada weights selama backpropagation menjadi semua positif atau semua negatif.

ReLU6 [4] adalah pengembangan dari ReLU, dimana ReLU mempunyai rentangan $(0, x)$, yang artinya apabila $x \leq 0$ maka $x = 0$ dan apabila $x > 0$ maka $x = x$. Hal ini berarti nilai x bisa tidak terbatas, yang berakibat kepada membengkaknya nilai aktivasi. Solusinya adalah dengan menetapkan nilai x dengan angka tertentu. Angka yang digunakan adalah 6, maka fungsi aktivasi menjadi:

$$y = \min(\max(x, 0), 6) \tag{2}$$



Gambar 2. Sigmoid



Gambar 3. ReLU6

C. Pooling Layer

Lapisan pooling merupakan lapisan yang berfungsi sebagai untuk memperkecil dimensi dari citra. Dikarenakan tahapan ini merupakan downsampling sehingga memperkecil parameter, maka kecepatan komputasi dapat lebih meningkat. Metode Pooling yang paling sering digunakan adalah Max Pooling dan Average Pooling. Max Pooling menentukan nilai maksimum, sedangkan Average Pooling menentukan nilai rata-rata. Max Pooling lebih banyak digunakan karena dengan max pooling filter dapat mengeksplorasi nilai tertinggi pada citra. Sedangkan Average Pooling mencari nilai rata-rata pada tiap dimensi.

D. Evaluasi Hasil

Evaluasi hasil uji coba dilakukan dengan mengukur nilai precision, recall dan F-1 Score. Precision (P) adalah perbandingan antara nilai True Possitive (TP) dengan total data yang diprediksi positif. Recall (R) adalah perbandingan antara nilai True Possitive (TP) dengan total data yang sebenarnya positif. F-1 Score adalah metode yang menyeimbangkan antara Precision dan Recall, dimana nilai terbaik adalah 1 dan nilai terburuk adalah 0. Nilai F-1 Score yang baik menunjukkan bahwa model klasifikasi memiliki nilai precision dan recall yang baik.

$$P = \frac{TP}{TP+FP} \tag{3}$$

$$R = \frac{TP}{TP+FN} \tag{4}$$

$$F1 = 2 \frac{PR}{P+R} \tag{5}$$

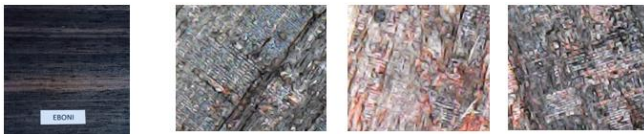
III. METODE PENELITIAN

Metode penelitian yang digunakan dalam proses pengklasifikasian tekstur serat kayu veneer pada citra mikroskopis, meliputi dataset yang akan digunakan, metode klasifikasi, dan arsitektur CNN yang akan disusun.

A. Dataset

Dataset yang digunakan berupa image mikroskopis atas veneer, yang diperoleh dengan mengumpulkan fisik veneer sesuai dengan jenis kelas yang akan menjadi output. Setelah fisik veneer terkumpul maka veneer akan diambil citranya menggunakan digital mikroskopis. Jumlah dataset yang digunakan untuk penelitian total sebanyak 9.144 citra. Adapun pembagian penggunaan dataset dalam penelitian ini adalah 60% untuk proses training, 20% proses validasi dan 20% proses testing.

Citra yang berhasil dikumpulkan per kelas terdiri dari kelas Sungkai sebanyak 804 citra, kelas Mahoni sebanyak 792 citra, kelas Eboni sebanyak 828 citra, kelas Mindi sebanyak 684 citra, kelas Sengin sebanyak 492 citra, kelas Mangga sebanyak 840 citra, kelas Ceri sebanyak 1.500 citra, kelas Walnut sebanyak 1.056 citra, kelas Keruing sebanyak 828 citra, dan kelas Jati Kampung sebanyak 1.320.



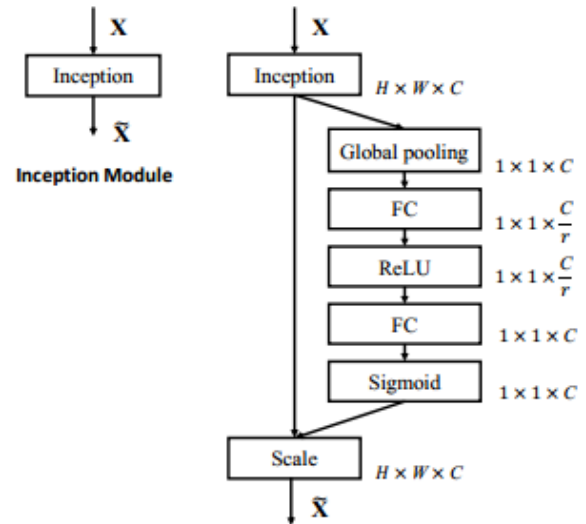
Gambar 4. Contoh fisik veneer dan citra mikroskopik

B. Metode Klasifikasi

Metode klasifikasi yang digunakan adalah convolutional neural network dengan model arsitektur yang dibangun menggunakan pendekatan Mobile Inverted Bottleneck Conv (MBCConv) Block[5] dan dibuat menyerupai EfficientNet[6][7]

Alasan penggunaan layer dan block untuk meningkatkan performa efektivitas dan efisiensi dari sebuah arsitektur jaringan neural. Dengan mampu meningkatkan pencapaian tingkat akurasi lebih tinggi dengan jumlah parameter yang relatif lebih sedikit, ringan untuk dapat diimplementasikan dan secara otomatis meningkatkan kecepatan. Tingkat akurasi merupakan alasan yang terpenting dan paling utama dalam penyusunan neural network.

Penerapan Squeeze and Excitation (SE) Block[8][9] dan Mobile Inverted Bottleneck Conv (MBCConv) Block[10][10] telah terbukti dapat mempengaruhi tingkat akurasi dengan jumlah parameter yang minimal, yang memungkinkan pengaplikasian di desktop, laptop dan mobile devices.



Gambar 5. SE – Inception Module

C. Arsitektur CNN

Model struktur arsitektur CNN pada penelitian ini terlihat pada tabel I, yang menampilkan jumlah lapisan CNN yang akan dijalankan sebelum akhirnya menuju pada proses penentuan klasifikasi dari masing-masing kelas dengan metode softmax.

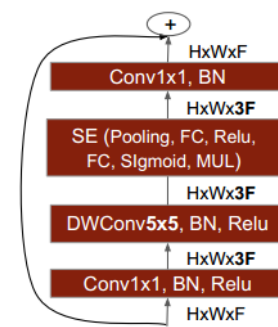
TABEL I
MODEL STRUKTUR ARSITEKTUR CNN

Layer	C (Total Filter)	Stride	R	Padding	Input	Output
Conv2D 3x3	32	2	-	Same	128x128x3	64x64x32
MBCCConv3 (k5x5)	16	1	2	-	64x64x32	64x64x16
MBCCConv6 (k3x3)	32	1	-	-	64x64x16	64x64x32
MBCCConv3 (k5x5)	48	2	4	-	64x64x32	32x32x48
MBCCConv3 (k3x3)	64	1	4	-	32x32x48	32x32x64
MBCCConv3 (k3x3)	80	2	4	-	32x32x64	16x16x80
MBCCConv3 (k5x5)	96	1	6	-	16x16x80	16x16x96
MBCCConv3 (k5x5)	112	2	6	-	16x16x96	8x8x112
MBCCConv3 (k3x3)	128	1	6	-	8x8x112	8x8x128
MBCCConv6 (k5x5)	140	2	-	-	8x8x128	4x4x140
Conv2D 1x1	256	1	-	Valid	4x4x140	4x4x256
Swish	-	-	-	-	4x4x256	4x4x256
Global Average Pooling	-	-	-	-	4x4x256	256
Fully Connected	256	-	-	-	256	256
ReLU6	-	-	-	-	256	256
Global Average Pooling	100	-	-	-	256	100
Softmax	10	-	-	-	100	10

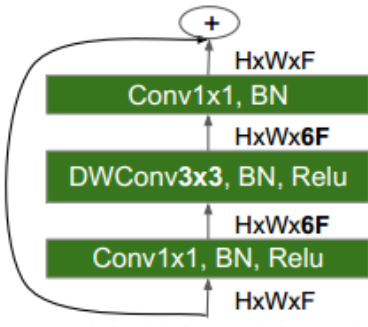
Terdapat 11 lapisan pada convolutional layer dan 2 fully connected, sebelum data dimasukkan ke dalam proses penentuan kelas.



Gambar 6. EfficientNet yang disusun



Gambar 7. MBCConv3 (k5x5)[11]

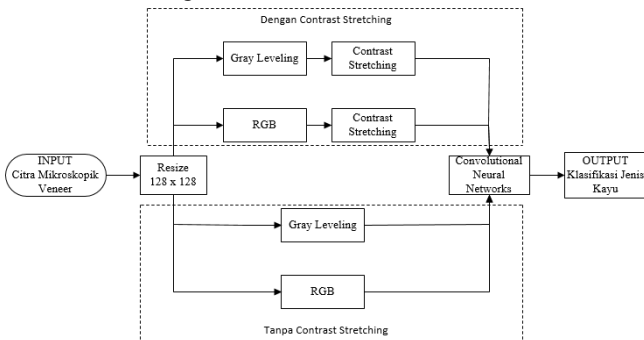


Gambar 8. MBConv6 (k3x3)[11]

D. Alur Penelitian

Alur penelitian dalam penelitian ini dapat dilihat pada gambar 9, dimana terdapat empat proses uji coba sesuai dengan model penelitian yang sudah ditetapkan.

Penelitian RGB tanpa contrast stretching, dalam penelitian ini menggunakan input citra mikroskopis veneer 3 channel (RGB) tanpa melalui proses contrast stretching sebelum masuk dalam model CNN. Penelitian RGB dengan contrast stretching, Penelitian ini menggunakan input citra mikroskopis veneer 3 channel (RGB) dan melalui proses contrast stretching sebelum masuk dalam model CNN.



Gambar 9. Skema penelitian

Penelitian grayscale tanpa contrast stretching, dalam penelitian ini menggunakan input citra mikroskopis veneer satu channel (Grayscale) tanpa contrast stretching sebelum masuk dalam model CNN. Penelitian grayscale dengan contrast stretching, Penelitian ini menggunakan input citra mikroskopis veneer satu channel (Grayscale) dan melalui proses contrast stretching.

IV. HASIL UJI COBA

Hasil uji coba yang dilakukan dalam mengevaluasi semua model yang telah disusun dalam penelitian ini dalam mengklasifikasikan tekstur serat kayu pada citra mikroskopis veneer. Hasil uji coba akan dijelaskan secara runtut hasil training yang diperoleh mulai dari model RGB tanpa contrast stretching, RGB dengan contrast stretching, Grayscale tanpa contrast stretching dan Grayscale dengan contrast stretching.

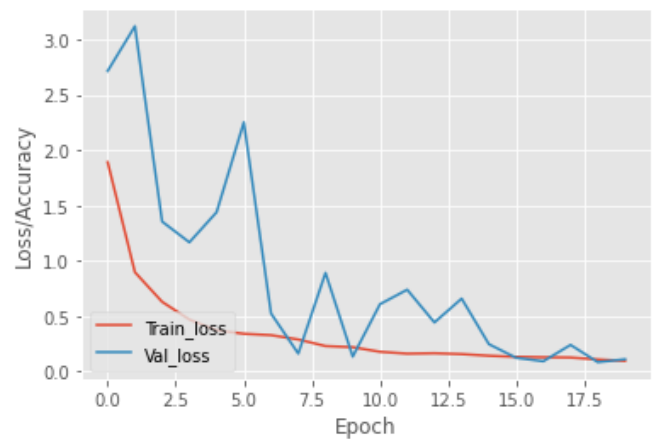
A. RGB tanpa Contrast Stretching

Hasil uji coba klasifikasi tekstur serat kayu pada citra mikroskopis pada veneer dengan menggunakan model arsitektur CNN atas input dataset RGB tanpa contrast

stretching, menunjukkan nilai F-1 score sebesar 0.97 dengan Precision sebesar 0.96 dan Recall sebesar 0.97.

TABEL II
HASIL TRAINING RGB TANPA CONTRAST STRETCHING

KELAS	GRAYSCALE - STRETCH			
	Precision	Recall	F1-Score	Support
Sungkai	0.71	0.75	0.73	161
Mahoni	0.74	0.89	0.81	158
Eboni	0.88	0.97	0.92	166
Mindi	0.69	0.58	0.63	137
Sengon	0.66	0.93	0.77	98
Mangga	0.92	0.65	0.77	168
Ceri	0.80	0.78	0.79	300
Walnut	0.62	0.83	0.71	211
Keruing	0.72	0.38	0.50	166
Jati Kampung	0.74	0.70	0.72	264
				1,829



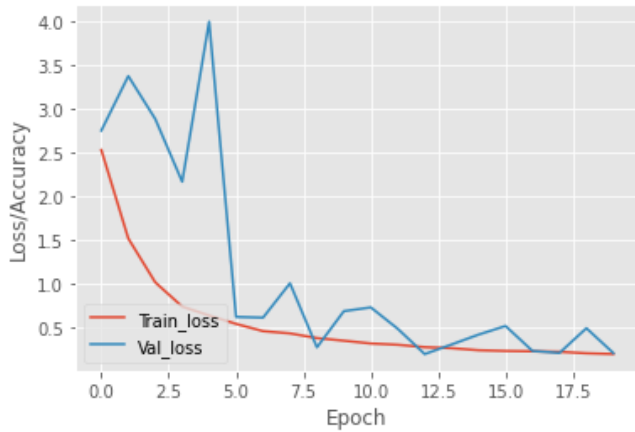
Gambar 10. Grafik loss/accuracy RGB tanpa contrast stretching

B. RGB dengan Contrast Stretching

Hasil uji coba klasifikasi tekstur serat kayu pada citra mikroskopis pada veneer dengan menggunakan model arsitektur CNN atas input dataset RGB dengan contrast stretching, menunjukkan nilai F-1 score sebesar 0.94 dengan Precision sebesar 0.93 dan Recall sebesar 0.93.

TABEL III
HASIL TRAINING RGB DENGAN CONTRAST STRETCHING

KELAS	RGB - STRETCH			
	Precision	Recall	F1-Score	Support
Sungkai	0.93	0.99	0.96	161
Mahoni	0.99	0.96	0.97	158
Eboni	0.93	0.98	0.95	166
Mindi	0.94	0.77	0.84	137
Sengon	1.00	0.95	0.97	98
Mangga	0.94	0.96	0.95	168
Ceri	0.92	0.90	0.91	300
Walnut	0.97	0.93	0.95	211
Keruing	0.81	0.96	0.88	166
Jati Kampung	0.95	0.94	0.95	264
				1,829



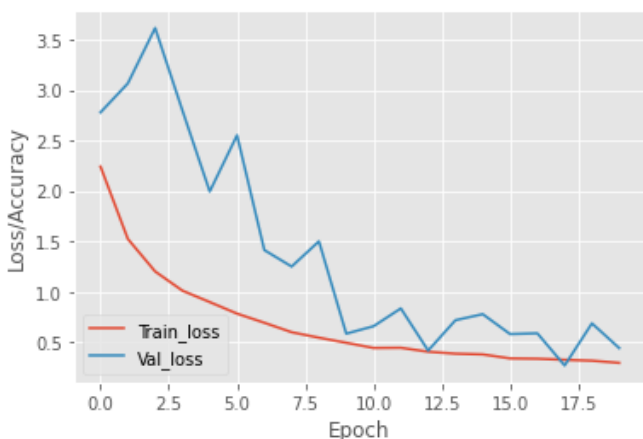
Gambar 11. Grafik loss/accuracy RGB dengan contrast stretching

C. Grayscale tanpa Contrast Stretching

Hasil uji coba klasifikasi tekstur serat kayu pada citra mikroskopis pada veneer dengan menggunakan model arsitektur CNN atas input dataset grayscale tanpa contrast stretching, menunjukkan nilai F-1 score sebesar 0.91 dengan Precision sebesar 0.91 dan Recall sebesar 0.90.

TABEL IV
HASIL TRAINING GRAYSCALE TANPA CONTRAST STRETCHING

KELAS	GRAYSCALE - NO STRETCH			
	Precision	Recall	F1-Score	Support
Sungkai	0.82	0.99	0.94	161
Mahoni	0.89	0.99	0.94	158
Eboni	0.86	1.00	0.93	166
Mindi	0.76	0.95	0.84	137
Sengon	1.00	0.84	0.91	98
Mangga	1.00	0.80	0.89	168
Ceri	0.94	0.91	0.93	300
Walnut	0.94	0.89	0.91	211
Keruing	0.92	0.92	0.92	166
Jati Kampung	0.95	0.80	0.87	264
				1,829



Gambar 12. Grafik loss/accuracy grayscale tanpa contrast stretching

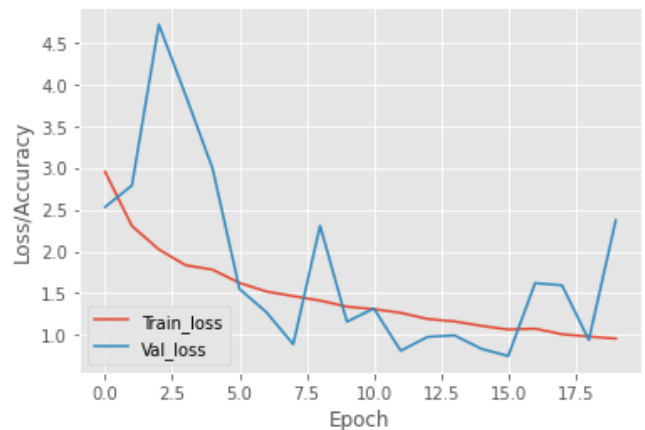
D. Grayscale dengan Contrast Stretching

Hasil uji coba klasifikasi tekstur serat kayu pada citra mikroskopis pada veneer dengan menggunakan model arsitektur CNN atas input dataset grayscale dengan contrast stretching, menunjukkan nilai F-1 score sebesar 0.75 dengan

Precision sebesar 0.75 dan Recall sebesar 0.74.

TABEL V
HASIL TRAINING GRAYSCALE DENGAN CONTRAST STRETCHING

KELAS	GRAYSCALE - STRETCH			
	Precision	Recall	F1-Score	Support
Sungkai	0.71	0.75	0.73	161
Mahoni	0.74	0.89	0.81	158
Eboni	0.88	0.97	0.92	166
Mindi	0.69	0.58	0.63	137
Sengon	0.66	0.93	0.77	98
Mangga	0.92	0.65	0.77	168
Ceri	0.80	0.78	0.79	300
Walnut	0.62	0.83	0.71	211
Keruing	0.72	0.38	0.50	166
Jati Kampung	0.74	0.70	0.72	264
				1,829



Gambar 13. Grafik loss/accuracy grayscale dengan contrast stretching

V. KESIMPULAN

Penelitian ini membuktikan bahwa kemampuan penerapan Convolutional Neural Network untuk memecahkan permasalahan dalam bidang citra. Model yang disusun dalam penelitian ini mampu menghasilkan tingkat akurasi dalam mengklasifikasikan tekstur serat kayu dalam citra mikroskopis pada veneer sebesar $\geq 80\%$.

Dari hasil uji coba atas 4 model arsitektur yang disusun, yaitu model RGB tanpa contrast stretching, RGB dengan contrast stretching, Grayscale tanpa contrast stretching dan Grayscale dengan contrast stretching. Model RGB tanpa contrast stretching mampu memberikan tingkat akurasi paling tinggi yaitu sebesar 0.97. Disusul berikutnya oleh model RGB dengan contrast stretching sebesar 0.93, model Grayscale tanpa contrast stretching sebesar 0.90 dan terakhir model Grayscale dengan contrast stretching sebesar 0.74.

Dari hasil uji coba menunjukkan bahwa data input dalam bentuk warna (RGB) mempunyai peran penting terhadap pencapaian hasil training dan hasil testing yang lebih bagus, baik atas pencapaian nilai precision, recall maupun tingkat akurasi.

Penerapan contrast stretching tidak cocok digunakan dalam klasifikasi tekstur serat kayu pada citra mikroskopis veneer. Penggunaan contrast stretching menurunkan kinerja model dalam mencapai akurasi. Dimana pada setiap model

dalam penelitian yang menggunakan contrast stretching justru mengalami penurunan tingkat akurasi. Hal ini terbukti dengan menurunnya tingkat akurasi sebesar 0.04 untuk RGB dan 0.16 untuk grayscale.

Meskipun data input dalam bentuk RGB mampu memberikan hasil yang lebih baik, di lain sisi data input dalam bentuk RGB memerlukan waktu training yang lebih lama yaitu sebanyak rata-rata 2 detik per epoch.

DAFTAR PUSTAKA

- [1] L. G. Hafemann, L. S. Oliveira, and P. Cavalin, "Forest species recognition using deep convolutional neural networks," 2014, doi: 10.1109/ICPR.2014.199.
- [2] J. Yang, P. Huang, F. Dai, Y. Sun, L. Wang, and H. Bi, "Application of Deep Learning in Wood Classification," 2019 *IEEE Int. Conf. Comput. Sci. Educ. Informatiz. CSEI 2019*, pp. 124–129, 2019, doi: 10.1109/CSEI47661.2019.8938960.
- [3] I. MathWorks, "Convolutional Neural Network - MATLAB & Simulink." p. 1.
- [4] C. Doshi, "Why Relu? Tips for using Relu. Comparison between Relu, Leaky Relu, and Relu-6.," pp. 2–7, 2019.
- [5] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 4510–4520, 2018, doi: 10.1109/CVPR.2018.00474.
- [6] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *36th Int. Conf. Mach. Learn. ICML 2019*, vol. 2019-June, pp. 10691–10700, 2019.
- [7] M. Tan, "Blog AutoML and Model Scaling," pp. 1–4, 2019.
- [8] J. Hu, "Squeeze-and-Excitation_Networks_CVPR_2018_paper.pdf," *Cvpr*, pp. 7132–7141, 2018.
- [9] K. Ta and P. Follow, "Squeeze and Excitation Networks (Hu et al .," pp. 3–6, 2020.
- [10] A. Howard *et al.*, "Searching for mobileNetV3," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2019-October, pp. 1314–1324, 2019, doi: 10.1109/ICCV.2019.00140.
- [11] M. Tan *et al.*, "Mnasnet: Platform-aware neural architecture search for mobile," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 2815–2823, 2019, doi: 10.1109/CVPR.2019.00293.

Analisis Trending Topik untuk Percakapan Media Sosial dengan Menggunakan Topic Modelling Berbasis Algoritme LDA

Ahmad Syaifuddin, *Teknologi Informasi iSTTS*, Reddy Alexandro Harianto, *Teknik Informatika iSTTS*, dan Joan Santoso, *Teknik Informatika iSTTS*

Abstrak— Aplikasi *WhatsApp* merupakan salah satu aplikasi *chatting* yang sangat populer terutama di Indonesia. *WhatsApp* mempunyai data unik karena memiliki pola pesan dan topik yang beragam dan sangat cepat berubah, sehingga untuk mengidentifikasi suatu topik dari kumpulan pesan tersebut sangat sulit dan menghabiskan banyak waktu jika dilakukan secara manual. Salah satu cara untuk mendapatkan informasi tersirat dari media sosial tersebut yaitu dengan melakukan pemodelan topik. Penelitian ini dilakukan untuk menganalisis penerapan metode LDA (*Latent Dirichlet Allocation*) dalam mengidentifikasi topik apa saja yang sedang dibahas pada grup *WhatsApp* di Universitas Islam Majapahit serta melakukan eksperimen pemodelan topik dengan menambahkan atribut waktu dalam penyusunan dokumen. Penelitian ini menghasilkan model topik dan nilai evaluasi *f-measure* dari model topik berdasarkan uji coba yang dilakukan. Metode LDA dipilih untuk melakukan pemodelan topik dengan memanfaatkan library LDA pada python serta menerapkan standar *text-preprocessing* dan menambahkan *slang words removal* untuk menangani kata tidak baku dan singkatan pada *chat logs*. Pengujian model topik dilakukan dengan uji *human in the loop* menggunakan *word intrusion task* kepada pakar Bahasa Indonesia. Hasil evaluasi LDA didapatkan hasil percobaan terbaik dengan mengubah dokumen menjadi 10 menit dan menggabungkan dengan *reply chat* pada percakapan grup *WhatsApp* merupakan salah satu cara dalam meningkatkan hasil pemodelan topik menggunakan algoritma *Latent Dirichlet Allocation* (LDA), didapatkan nilai *precision* sebesar 0.9294, nilai *recall* sebesar 0.7900 dan nilai *f-measure* sebesar 0.8541.

Kata Kunci—Topic Modelling, Latent Dirichlet Allocation, LDA, Media Sosial.

I. PENDAHULUAN

WhatsApp merupakan alat komunikasi yang bisa menimbulkan tantangan unik dalam pengelolaannya. Misalnya pada sebuah grup, banyak anggota yang cenderung mengabaikan kiriman atau tidak merespon pesan [1].

Ahmad Syaifuddin, Teknologi Informasi Institut Sains dan Teknologi Terpadu Surabaya, Surabaya, Jawa Timur, Indonesia (e-mail: syaifuddin.skm@gmail.com)

Reddy Alexandro Harianto, Teknologi Informasi Institut Sains dan Teknologi Terpadu Surabaya, Surabaya, Jawa Timur, Indonesia (e-mail: reddy@stts.edu)

Joan Santoso, Teknologi Informasi Institut Sains dan Teknologi Terpadu Surabaya, Surabaya, Jawa Timur, Indonesia (e-mail: joan@stts.edu)

atau membalas postingan [2]. Karena sifat datanya yang cepat sekali berubah dan pengguna tidak sepenuhnya *online* atau memataui grup *WhatsApp* membuat pengguna kesulitan dalam mendapatkan informasi topik apa yang sedang dibahas pada grup *WhatsApp*. Padahal, kumpulan pesan tersebut merupakan sumber data yang sangat berpotensi untuk memberikan informasi yang penting bagi anggota lain.

Penelitian tentang pemodelan topik telah banyak dilakukan, banyak diantara penelitian menggunakan data dari *Twitter*, *Facebook*, *Online News*, *Blog*, jurnal dan lain sebagainya, namun untuk aplikasi *mobile chatting* khususnya *WhatsApp* belum banyak ditemukan topik penelitian di bidang *topic modelling* [3]. Padahal manfaat yang didapatkan dari data tersebut sangatlah banyak, misalnya untuk grup di Perguruan Tinggi. Salah satu fungsi utama grup di Perguruan Tinggi adalah untuk memudahkan komunikasi, penyampaian informasi yang penting dari pimpinan, keluhan, dan hal lain yang bersifat urgen.

Berdasarkan latar belakang tersebut, penelitian ini akan menawarkan solusi dalam melakukan analisis pemodelan topik menggunakan metode *Latent Dirichlet Allocation* pada grup *WhatsApp*. Analisis pemodelan topik digunakan untuk mengetahui tren topik yang muncul pada obrolan grup, sehingga memudahkan dalam mengetahui informasi apa yang sedang dibahas pada grup tersebut. Sebagai bagian dari *natural language programming*, maka langkah yang akan diimplementasikan mengikuti tahapan penelitian dalam ranah *text mining* secara umum, dengan menambahkan atribut *reply chat* yang dimiliki oleh *WhatsApp* dalam melakukan *pre-processing*.

Kontribusi paper ini adalah:

1. Analisis trending topik dengan LDA pada time frame tertentu yang dilakukan di Bahasa Indonesia.
2. Visualisasi trending topik dengan LDA untuk menggunakan atribut waktu dan *reply* bertingkat.

II. PENELITIAN PENJUNJANG

Guixian Xu [4] melakukan penelitian dalam melakukan ekstraksi topik dari teks berita menggunakan model *Latent Dirichlet Allocation* (LDA) dan metode *Gibbs sampling* untuk melakukan spekulasi parameter. Kemudian membandingkan dengan metode *K-Mean* dalam pendeteksian topik. Jumlah data yang digunakan berasal dari berita berlabel yang diterbitkan oleh *Sogou Lab* sebanyak 3.000 laporan berita. Hasilnya metode *LDA* lebih andal

daripada model *k-means clustering* untuk penentuan topik yang potensial.

YaJun Du [5] melakukan penelitian dalam melakukan mengestrak dan melacak topik pada *micro-blog* menggunakan improfikasi dari metode *Latent Dirichlet Allocation* (LDA). Penelitian ini menggunakan lima fitur unik blog mikro untuk mendorong distribusi probabilitas gabungan dari semua kata dan topik, dan meningkatkan LDA ke dalam model ekstraksi topik bernama MF-LDA (*Micro-blog features Latent Dirichlet Allocation*). Data yang digunakan dalam eksperimen model MF-LDA yaitu sekitar 500.000 data posting pada blog Mikro Sina yang dikumpulkan dari Mei 2016 hingga Desember 2016. Sedangkan pada pengujian hot topic mengambil sekitar 300.000 posting mikro-blog yang terkait dengan hampir 100 topik hangat dari platform Sina Micro-blog. Hasil percobaan menunjukkan bahwa model MF-LDA yang diusulkan berkinerja lebih baik dengan *perplexity* yang lebih rendah dan CR yang lebih tinggi daripada model LDA.

Dongjin Yu [6] melakukan penelitian tentang pemodelan topik hirarki data twitter untuk pemrosesan analitik online dengan mengusulkan model topik yang disebut twitter hierarchical latent Dirichlet alokasi (thLDA), dengan membandingkan dengan metode LDA biasa dan metode hLDA. Eksperimen dilakukan pada sejumlah besar data Twitter yang dikumpulkan melalui REST API Twitter. diperoleh 10.160.317 percakapan dari 6.907 tweet. Hasil penelitian didapatkan skor PMI pada model thLDA sedikit lebih tinggi daripada dua metode lainnya.

Ajai Gaur [7] melakukan penelitian tentang pemanfaatan LDA dalam *topic modelling* untuk identifikasi artikel yang diterbitkan oleh *Organization Research Methods* (ORM) sejak pertama kali terbit yaitu identifikasi kemiripan setiap topik pada seluruh artikel yang berasumsi bahwa topik yang muncul akan bervariasi tergantung pada latar belakang disiplin penulis. Hasil penelitiannya didapatkan sebanyak 15 topik yang sangat mirip dari 421 artikel ORM.

Terkait pelacakan topik, penelitian Jui-Feng Yeh [8] tentang pendeteksian dan pelacakan topik dalam percakapan lisan model *Conceptual Dynamic Latent Dirichlet Allocation* (CDLDA). Dataset berasal dari Chia Yi Chinese Conversation Dialogue Corpus (CYCCDC) yang diseleksi secara random didapatkan 769 dialog. Hasilnya didapatkan bahwa model CDLDA dapat secara dinamis mengekstraksi dan melacak topik dibandingkan dengan LDA tradisional. Basri [9] dalam penelitiannya yaitu klasterisasi dengan algoritma K-Means untuk identifikasi topik informasi publik media sosial di Kota Surabaya.

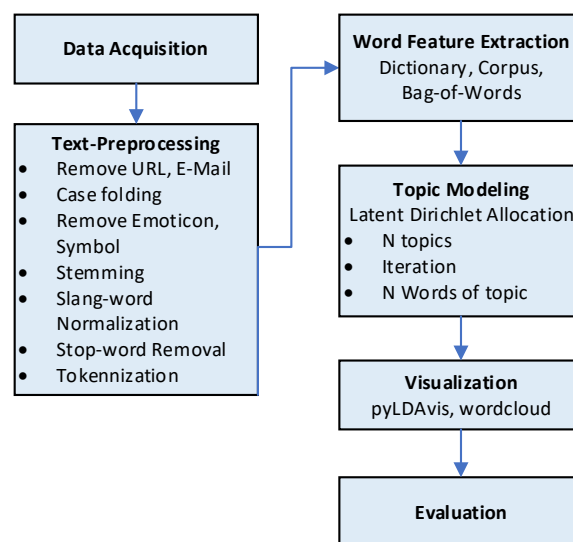
Khusus untuk grup WhatsApp, Cendana [10] dengan penelitiannya tentang pra-pemrosesan teks pada grup WhatsApp untuk pemodelan topik, pra-proses yang dilakukan yaitu tokenisasi, *filtering* dan *stemming*. Hasil penelitian didapatkan pada tahap pra-proses ditemukan kendala seperti, penggunaan bahasa tidak baku dalam chat, tidak dapat melakukan pemrosesan terhadap data yang besar. Penelitian ini tidak dilakukan normalisasi kata tidak baku pada tahap pra-proses. Penelitian Rosenfeld [11] tentang *WhatsApp Usage Patterns and Prediction Models*. Sumber datanya yaitu 4 juta pesan dari 100 pengguna dengan melihat

kebiasaan dan perilaku pengiriman pesan berdasarkan *gender* dan usia. Model prediksi yang digunakan adalah algoritma C.45 dan Bayesian networks.

Penelitian Premalatha [12] dengan melakukan analisis sentimen dengan metode *Analytical Sandbox* terhadap pesan yang terdapat dalam grup WA dan visualisasinya sehingga didapatkan hasil opini positif dan negatif yang diilustrasikan dalam bentuk emotikon (*anger, fear, disgust, anticipation, joy, sadness, surprise, trust* dan *negative*). Sanchita Patil [13] melakukan implementasi bahasa pemrograman R untuk prediksi level kecanduan dari pengguna grup WA berdasarkan umur dan *gender*.

III. METODOLOGI

Alur sistem penelitian menggunakan langkah standar pada analisis *text mining* yaitu standar *pre-processing* ditambah dengan *remove URL-email, normalization slang word, stemming, remove stopword* dan *processing* menggunakan LDA, ditunjukkan pada gambar 1.



GAMBAR 1.
ALUR SISTEM

Pada langkah pertama, akuisisi data yaitu dengan mengunduh *chat logs* pada *WhatsApp Web*, kemudian langkah kedua dilakukan *text-preprocessing*, selanjutnya melakukan ekstraksi fitur dengan membuat *dictionary, corpus* dan *bag-of-word*. Selanjutnya data diolah dengan algoritma LDA dan divisualisasi menggunakan library *pyLDAvis* dan *wordcloud*. Langkah terakhir yaitu pengujian menggunakan uji koherensi topik dan evaluasi *f-measure*.

A. Pengumpulan Data

Data didapatkan dari hasil *download chat* menggunakan fitur *WhatsApp Web* menggunakan aplikasi *Google Chrome*, kemudian ditambahkan *extension* atau plugin *Backup WhatsApp Chats*. Ekstensi *Backup WhatsApp Chats* dapat diunduh secara gratis melalui *Chrome Web Store*. Penulis menggunakan Ekstensi *Backup WhatsApp Chats* dikarenakan entitas yang dihasilkan lebih banyak bila dibandingkan dengan ekspor *chats* menggunakan fitur *export* yang terdapat pada *WhatsApp mobile*. Berikut merupakan tampilan dari percakapan yang ada pada grup *WhatsApp*. Sumber data yang dianalisa adalah data pesan yang dikirim oleh pengguna grup

WhatsApp Dosen Universitas Islam Majapahit pada bulan Desember tahun 2019. Data grup tersebut dipilih karena dianggap mempunyai kemiripan pada topik ataupun konten yang disampaikan, serta sudah mewakili civitas akademika di Universitas Islam Majapahit. Data yang digunakan adalah data teks saja, tidak termasuk gambar, video, emoticon dan file lainnya.

Input LDA berupa corpus yang terdiri dari beberapa dokumen, pada penelitian ini dokumen diambil berdasarkan pesan yang dikirim oleh user (1 pesan = 1 dokumen). Entitas *MessageBody* merupakan entitas utama dalam pembentukan dokumen, yaitu 1 isian pesan pada kolom *MessageBody* mewakili 1 dokumen, dalam hal ini *UserName* tidak dilibatkan dalam pembentukan dokumen. Pada eksperimen lebih lanjut, peneliti mengubah dokumen normal (1 pesan = 1 dokumen) berdasarkan reply chat dan kumpulan pesan dalam interval 10 menit.

B. Preprocessing

Langkah pra-proses data mencakup beberapa langkah utama pengerjaan yakni membersihkan data, *stemming* dan *stoword removal*.

1. Pembersihan data dilakukan untuk menghilangkan penulisan huruf besar menjadi huruf kecil, menghilangkan adanya karakter yang tidak diperlukan, menghapus token yang hanya terdiri dari angka. Selain itu, karena data yang diproses berasal dari pesan sosial media, maka perlu penanganan khusus seperti menghapus URL dan alamat email, menghapus *emoticon*, menghapus media gambar, video dan menghapus bilangan angka.
2. Normalisasi kata tidak baku, pesan pada whataspp sering kali diketik menggunakan kata slang atau singkatan, sehingga perlu melakukan pengumpulan kata tidak baku dan baku untuk digunakan sebagai *base dictionary*.
3. *Stemming* dilakukan untuk mengurangi kata-kata infleksi dalam Bahasa Indonesia menjadi bentuk dasarnya. Tahap *stemming* menggunakan library *sastrawi* dengan fungsi *stem*.
4. *Stopword removal* mengacu pada susunan *stopword* yang memanfaatkan *stopword removal indonesian* dari library *nltk*, dan *slang stopwords* yang didapatkan peneliti dari temuan hasil ekspor data.
5. Menghapus dokumen yang mempunyai kata kurang dari dua. Dokumen yang hanya mempunyai satu kata penyusun dianggap tidak mempunyai arti dalam dokumen, sehingga data tersebut harus dihapus dan tidak dipakai sebagai data proses.

Pra-proses untuk mendapatkan dokumen reply chat, dilakukan dengan mengolah file csv hasil download dari WhatsApp Web menggunakan aplikasi *Libre Office* untuk menggabungkan isi pesan yang termasuk replay chat dan pesan yang dikirim dalam durasi 10 menit menjadi 1 buah dokumen. Urutan aturan dalam menggabungkan pesan menjadi dokumen sebagai berikut:

1. Pertama, pesan yang merupakan balasan dari pesan sebelumnya digabungkan terlebih dahulu
2. Kedua, pesan yang dikirim dalam interval waktu 10 menit digabungkan, dimulai dari menit ke 0-10, kemudian menit ke 11-20 dan seterusnya.

3. Ketiga, pesan yang sudah digabungkan pada tahap pertama dapat digabungkan lagi pada tahap kedua, jika pesan tersebut masuk dalam interval waktu tertentu.
4. Menyimpan hasil penggabungan pesan pada kolom bernama "MergedMessage".

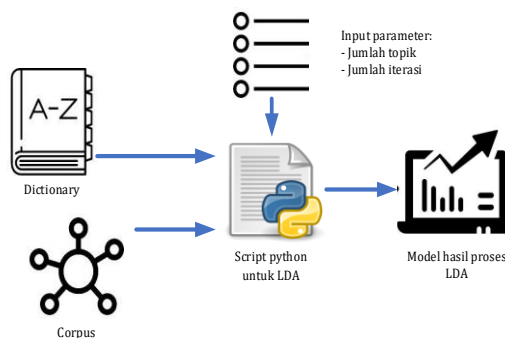


GAMBAR 2.
ALUR MEMPERSIAPKAN DATA

C. Latent Dirichlet Allocation (LDA)

LDA (*Latent Dirichlet Allocation*) adalah model probabilistik dalam *topic modelling* pada data teks untuk mendapatkan informasi berupa model topik. Model topik LDA dapat mewujudkan dimensi reduksi representasi teks dalam ruang semantik, dan memodelkan teks dengan probabilitas kosa kata, yang meringankan masalah data sampai batas tertentu [4].

Proses LDA diawali dengan konversi dokumen ke dalam bentuk *dictionary* dan mengkonversi *dictionary* ke dalam bentuk matriks dokumen atau *corpus*, kemudian membentuk model topik dengan menggunakan algoritma LDA. Proses algoritma LDA membutuhkan parameter input yang dalam hal penelitian ini yang digunakan adalah jumlah topik sebanyak 10, jumlah iterasi 50 dan jumlah kata dalam satu topik yaitu 10.



GAMBAR 3.
ALUR TOPIC MODELLING DENGAN LDA

Algoritma LDA terdiri dari proses inialisasi, proses iterasi dan pengambilan sampel, proses membaca parameter akhir [14].

1. Tahap inialisasi merupakan proses untuk menentukan frekuensi kemunculan kata dari setiap kata pada setiap file teks. Proses ini dilakukan pada teks hasil *pre-processing* data. Proses inialisasi dilakukan dengan langkah:
 - a. Menentukan index dari setiap kata pada dokumen
 - b. Menghitung frekuensi kemunculan setiap kata pada setiap dokumen menggunakan *Bag-of-Words (BoW)*
 - c. Menentukan topik setiap kata dengan random berdasarkan nilai frekuensi kemunculan kata (z_0)

LDA membutuhkan nilai topik yang ditentukan terlebih dahulu. Selain itu, dalam algoritma LDA, tidak ada nilai awal yang diberikan untuk setiap kata dalam dokumen, sehingga menghasilkan setiap kata memiliki nilai ketidakpastian. Dalam penelitian ini, nilai (z_0) diambil secara acak yang menjadi nilai awal untuk setiap kata dalam dokumen.

- d. Menentukan matriks kata-topik dan dokumen-topik
- e. Menghitung jumlah total dari distribusi kata-topik dan dokumen-topik, dan menyimpan hasil matriks. Distribusi kata, W_i , pada tiap topik, Z_i , NW dilihat pada persamaan berikut.

$$NW = \begin{Bmatrix} NW_{w1,z1} & NW_{w1,z2} & \dots & NW_{w1,zk} \\ NW_{w2,z1} & NW_{w2,z2} & \dots & NW_{w2,z1} \\ \dots & \dots & \dots & \dots \\ NW_{wn,z1} & NW_{wn,z2} & \dots & NW_{wn,zk} \end{Bmatrix} \quad (1)$$

Persamaan (1), W_n adalah *term* ke- n pada vocab dan Z_k adalah Topik ke- k sedangkan $NW_{wn,zk}$ banyak *term* ke- n yang berlabel topik ke- k

Kemudian membuat matriks distribusi topik, Z_i , pada tiap dokumen, d_i , ND

$$ND = \begin{Bmatrix} ND_{d1,z1} & ND_{d1,z2} & \dots & ND_{d1,zk} \\ ND_{d2,z1} & ND_{d2,z2} & \dots & ND_{d2,z1} \\ \dots & \dots & \dots & \dots \\ ND_{dn,z1} & ND_{dn,z2} & \dots & ND_{dn,zk} \end{Bmatrix} \quad (2)$$

Persamaan (2), dimana d_m adalah dokumen ke- m , Z_k adalah topik ke- k dan $ND_{dn,zk}$ adalah banyak label topik ke- k pada Dokumen ke- n

Pada akhir proses inialisasi, dilakukan proses perhitungan jumlah setiap topik dalam dokumen (nd) dan jumlah setiap kata dalam topik (nw) yang akan digunakan dalam proses iterasi dan pengambilan sampel topik. Nilai total semua nd juga dihitung sebagai $sumnd$ dan nilai total semua nw sebagai $sumnw$. Nilai dari $sumnkm$ dan $sumnw$ digunakan untuk mengurangi dan menambahkan nilai nd dan nw di sembarang perubahan topik yang terjadi pada setiap kata. Jumlah distribusi NW dan ND sebagai berikut:

$$NWSum_{zi} = \sum_{j=1}^m NW_{wj,zi} \dots \dots \quad (3)$$

$$NDSum_{di} = \sum_{j=1}^k ND_{di,zj} \dots \dots \quad (4)$$

Persamaan (3), $NWSum_{zi}$ adalah jumlah seluruh kata dalam setiap topik, $NDSum_{di}$ adalah jumlah seluruh topik dalam setiap dokumen, $NW_{wj,zi}$ adalah jumlah setiap kata dalam topik dan $ND_{di,zj}$ adalah setiap topik dalam dokumen

2. Tahap sampling topik merupakan proses untuk menentukan topik baru dari setiap kata pada setiap dokumen. Proses ini dilakukan pada teks hasil *pre-processing*. Proses sampling topik dilakukan dengan langkah:

- a. Menghitung probabilitas kata pada topik

$$\phi_{ij} = \frac{C_{ij}^{WT} + \eta}{\sum_{k=1}^W C_{kj}^{WT} + W\eta} \quad (5)$$

Persamaan (5), ϕ_{ij} adalah probabilitas dari kata i untuk topik j , C_{ij}^{WT} adalah Jumlah kata i pada topik j , WT adalah kata-topik, η adalah nilai beta (merupakan parameter dirichlet), $\sum_{k=1}^W C_{kj}^{WT}$ adalah jumlah seluruh kata pada topik j , k adalah indek topik dan W dalam Jumlah seluruh kata pada dokumen.

Menghitung probabilitas dokumen pada topik

$$\theta_{dj} = \frac{C_{dj}^{DT} + \alpha}{\sum_{k=1}^T C_{dk}^{DT} + T\alpha} \quad (6)$$

Persamaan (6), θ_{dj} adalah proporsi topik j dalam dokumen d , C_{dj}^{DT} adalah jumlah topik j pada dokumen d , DT adalah dokumen-topik, α adalah nilai alpha, $\sum_{k=1}^T C_{dk}^{DT}$ adalah jumlah seluruh topik pada dokumen d T adalah jumlah seluruh topik yang sudah ditentukan

- b. Menentukan topik baru dari setiap kata dengan distribusi multinomial (posterior) berdasarkan nilai probabilitas kata tertinggi.

Parameter distribusi posterior dimana bobot tersebut didapatkan dari nilai distribusi probabilitas kata-topik dikali distribusi probabilitas topik-dokumen.

$$P(z_i = j | z_i, w_i, d_i) = \frac{C_{ij}^{WT} + \eta}{\sum_{k=1}^W C_{kj}^{WT} + W\eta} \times \frac{C_{dj}^{DT} + \alpha}{\sum_{k=1}^T C_{dk}^{DT} + T\alpha} \quad (7)$$

Persamaan (7), $P(z_i = j)$ adalah probabilitas posterior untuk kata i yang diberikan topik j , z_i adalah merupakan penugasan topik dari semua kata w_i adalah indeks kata ke i dan d_i adalah dokumen yang berisi kata

- c. Menyimpan hasil distribusi posterior
- d. Langkah-langkah ini dilakukan sebanyak n iterasi/pengulangan sampai mencapai kondisi konvergen.
3. Tahap perhitungan parameter final merupakan proses untuk menghitung jumlah dokumen untuk setiap topik dan jumlah kata untuk setiap topik berdasarkan matriks kata-topik dan dokumen-topik yang telah konvergen.

D. Evaluasi LDA

Evaluasi hasil model topik dilakukan dengan menggunakan perhitungan *precision*, *recall* dan *f-measure* yang ditampilkan dalam matrik evaluasi [15]. *Precision* (P) merupakan nilai pembagian dari jumlah item relevan yang diperoleh terhadap jumlah seluruh item yang diperoleh, rumus mencari *precision* dapat dilihat dalam persamaan (1).

$$P(\text{relevant} | \text{retrieved}) = \frac{\text{relevant items retrieved}}{\text{retrieved items}} \quad (8)$$

Recall (R) merupakan nilai pembagian dari jumlah item relevan yang diperoleh terhadap jumlah seluruh item yang relevan, rumus mencari recall dapat dilihat dalam persamaan (2)

$$R(\text{retrieved}|\text{relevant}) = \frac{\text{relevant items retrieved}}{\text{relevant items}} \quad (9)$$

F-Measure (F) merupakan bobot rata-rata dari nilai P dan R, rumus mencari recall dapat dilihat dalam persamaan (3).

$$F = \frac{2PR}{P+R} \quad (10)$$

IV. UJI COBA SISTEM

Pembentukan model LDA dilakukan dengan mengimplementasikan 2 kali uji coba, yaitu model dengan dokumen sesuai hasil unduh data *WhatsApp* (uji coba tahap 1), model dengan dokumen hasil unduh data *WhatsApp* dalam durasi 10 menit serta hasil gabungan reply chat (uji coba tahap 2). Uji coba LDA dilakukan dengan bahasa python menggunakan library gensim dan *word cloud* untuk visualisasi model topik.

A. Uji Coba pada Dokumen Normal

Jumlah data mentah yang dimuat sebagai sumber data masukan dapat dilihat pada tabel berikut. Data mentah yang dimuat selanjutnya dijadikan sumber data input. Tahap pra-proses cukup efektif dalam menyeleksi data, terlihat sebanyak 443 data yang terhapus setelah dilakukan pra-proses. Data yang telah melalui tahap pra-proses dapat dilihat pada tabel sebagai berikut:

TABEL 1.
DATA UJI COBA PADA DOKUMEN NORMAL

	Jumlah dokumen	Jumlah token
Sebelum pra-proses	773	5702
Setelah pra-proses	330	1614

Hasil uji coba tahap pertama yang dilakukan dengan menggunakan dokumen sesuai hasil ekspor data *WhatsApp* dapat ditampilkan bawa tabel dibawah, yaitu berupa topik, kata penyusun topik dan nilai probabilitas dari masing-masing kata, berikut cuplikan model topik percobaan pertama dapat dilihat pada tabel 2.

TABEL 2.
HASIL MODEL TOPIK PADA DOKUMEN NORMAL

Topik	Model Topik
T1	pmb, 0.0419 + tim, 0.0196 + thp, 0.0189 + personil, 0.0113 + prodi, 0.0113 + sistem, 0.0113 + unim, 0.0086 + promosi, 0.0086 + bijak, 0.0086 + review, 0.0058
T2	data, 0.0242 + prodi, 0.0212 + dosen, 0.0152 + mahasiswa, 0.0122 + bimbingan, 0.0122 + guru, 0.0122 + didik, 0.0092 + teliti, 0.0092 + seminar, 0.0062 + riset, 0.0062
T3	wisata, 0.0451 + mojosuro, 0.0307 + air, 0.0162 + panas, 0.0130 + kuasa, 0.0114 + padusan, 0.0098 + pipa, 0.0098 + banjir, 0.0098 + sungai, 0.0095 + cinta, 0.0081
T4	kirim, 0.0343 + mahasiswa, 0.0311 + pkm, 0.0186 + tulis, 0.0186 + proposal, 0.0160 + email, 0.0160 + terima, 0.0148 + dosen, 0.0106 + kampus, 0.0081 + judul, 0.0081
T5	ajar, 0.0258 + didik, 0.0221 + mendikbud, 0.0203 + bijak, 0.0148 + sekolah, 0.0148 + uji, 0.0129 + guru, 0.0129 + kualitas, 0.0111 + data, 0.0093 + presiden, 0.0093
T6	ipk, 0.0136 + indonesia, 0.0119 + kejar, 0.0119 + kerja, 0.0117 + didik, 0.0102 + riset, 0.0102 + bangsa, 0.0102 + sekolah, 0.0102 + peringkat, 0.0102 + urut, 0.0102
T7	allah, 0.1415 + dosa, 0.0864 + hati, 0.0589 + maha, 0.0425 + rezeki, 0.0320 + sembuh, 0.0209 + guru, 0.0209 + terima, 0.0150 + keluarga, 0.0149 + urus, 0.0142
T8	khotimah, 0.0296 + husnul, 0.0259 + rojiun, 0.0182 + lillahi, 0.0163 + amal, 0.0130 + almarhum, 0.0112 + kanker, 0.0112 + sembuh, 0.0094 + jantung, 0.0094 + allah, 0.0094
T9	listrik, 0.0174 + pln, 0.0161 + didik, 0.0134 + mati, 0.0134 + dosen, 0.0107 + mobil, 0.0094 + parkir, 0.0094 + lebaran, 0.0094 + usaha, 0.0094 + sepeda, 0.0081
T10	dosen, 0.0229 + insulin, 0.0216 + glukosa, 0.0173 + unim, 0.0155 + gula, 0.0145 + mahasiswa, 0.0141 + diabetes, 0.0116 + parkir, 0.0111 + manusia, 0.0101 + karbo, 0.0101

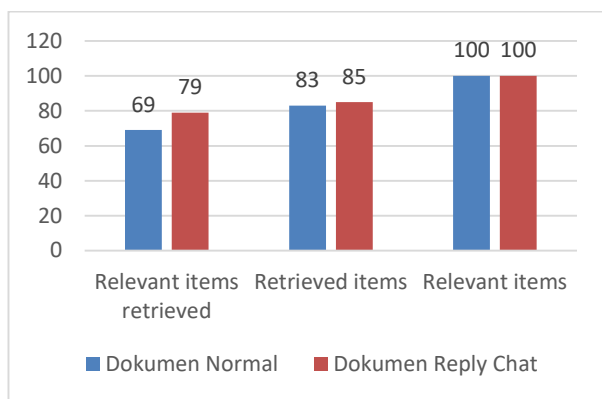
B. Uji Coba Pada Dokumen Reply Chat

Distribusi probabilitas kata dalam 10 topik yang tersusun dari 10 kata dalam masing-masing topik dari model dokumen sesuai hasil ekspor data *WhatsApp* serta gabungan dari reply chat, dapat dilihat pada tabel hasil pembentukan model LDA dengan dokumen hasil ekspor data *WhatsApp* dalam durasi 10 menit dan gabungan reply chat. Jumlah data mentah yang dimuat sebagai sumber data masukan dapat dilihat pada tabel berikut. Data mentah yang dimuat selanjutnya dijadikan sumber data input berikut cuplikan data yang digunakan pada uji coba kedua, dapat dilihat pada tabel 3.

TABEL 3.
DATA UJI COBA PADA DOKUMEN REPLY CHAT

	Jumlah dokumen	Jumlah token
Sebelum pra-proses	340	5702
Setelah pra-proses	205	1664

Hasil uji coba tahap kedua yang dilakukan dengan menggunakan dokumen hasil ekspor data *WhatsApp* berdasarkan interval waktu 10 menit dan reply chat dapat ditampilkan bawa tabel dibawah, yaitu berupa topik, kata penyusun topik dan nilai probabilitas dari masing-masing kata, berikut cuplikan model topik percobaan pertama dapat dilihat pada tabel 4.



GAMBAR 6.
HASIL MODEL TOPIK PADA DOKUMEN REPLY CHAT

Evaluasi dengan matriks pengukuran *precision*, *recall* dan *f-measure*. Berikut matrik hasil evaluasi yang dilakukan.

TABEL 5.
MARIK HASIL EVALUASI

Uji Coba	Dokumen normal	Dokumen Reply Chat
Precision	0.8313	0.9294
Recall	0.6900	0.7900
F-Measure	0.7541	0.8541

Berdasarkan hasil evaluasi diatas, didapatkan bahwa nilai *precision*, *recall* dan *f-measure* pada uji coba kedua (dengan mengubah dokumen menjadi 10 menit dan gabungan *reply chat*) memiliki nilai yang lebih tinggi bila dibandingkan dengan uji coba pertama yang telah dilakukan. Didapatkan nilai *precision* sebesar 0.9294, nilai *recall* sebesar 0.7900 dan nilai *f-measure* sebesar 0.8541. Hal ini menunjukkan bahwa mengubah dokumen menjadi 10 menit dan menggabungkan *reply chat* pada percakapan grup *WhatsApp* merupakan salah satu cara dalam meningkatkan hasil pemodelan topik menggunakan algoritma *Latent Dirichlet Allocation* (LDA). Hal ini karena anggota grup pada *WhatsApp* sebagian besar merespon pesan yang dikirimkan oleh anggota lain dalam jangka waktu yang cukup lama, sehingga meningkatkan dokumen dengan interval 10 menit dapat mengumpulkan percakapan dengan topik diskusi yang sama selain itu hasil *reply chat* juga sangat berpengaruh besar dalam pembentukan dokumen. Penelitian topik modelling pada grup *WhatsApp* menggunakan *Latent Dirichlet Allocation* (LDA) dapat dengan baik mengidentifikasi topik-topik yang sering dibicarakan anggota grup *WhatsApp* dengan kondisi mengubah dokumen menjadi 10 menit dan menggabungkan *reply chat*.

Grup *WhatsApp* cukup efektif guna mendukung kinerja dari civitas akademika di perguruan tinggi. Sesuai penelitian Andjani [16] penggunaan media komunikasi *WhatsApp* di instansi sangat baik dan dapat membantu dalam peningkatan efektivitas kinerja karyawan. Penelitian Rahartri [17] juga menjelaskan *WhatsApp* merupakan pengganti SMS yang praktis dan tepat waktu dalam mengirimkan pesan karena lebih simpel dan mudah digunakan pada layanan jasa informasi ilmiah di kawasan Puspptek.

Uji coba pertama didapatkan hasil evaluasi *precision*, *recall* dan *f-measure* paling kecil dengan nilai *precision* sebesar 0.8313, nilai *recall* sebesar 0.6900 dan nilai *f-*

measure sebesar 0.7541. Hal ini disebabkan karena respon anggota dalam membalas pesan dari anggota lain dengan durasi yang lama, sehingga topik yang dibicarakan tidak terkumpul menjadi satu dokumen. Selain itu, karena karakter dari pesan *WhatsApp* yang kebanyakan ditulis dengan kalimat yang pendek, sehingga pada saat tahap *pre-processing* kata penyusun dokumen ternormalisasi menjadi sedikit atau mungkin hilang, karena penulis tidak membuang dokumen yang hanya berisi satu kata.

V. KESIMPULAN

Latent Dirichlet Allocation (LDA) dapat dengan baik mengidentifikasi topik-topik yang sering dibicarakan anggota grup *WhatsApp*, dibuktikan dengan topik yang dihimpun dari pakar, menyatakan lebih dari 80% topik yang didapatkan dari sistem sesuai dengan model/kata penyusun topik. Eksperimen dengan menambahkan atribut waktu dalam penyusunan dokumen, didapatkan bahwa mengubah dokumen menjadi 10 menit dan menggabungkan *reply chat* didapatkan hasil yang lebih baik bila dibandingkan dengan dokumen normal yaitu sebanyak 79 kata relevan yang diterima oleh pakar. Hasil uji *f-measure* didapatkan bahwa dokumen dengan interval 10 menit dan menggabungkan *reply chat* mempunyai nilai yang tinggi bila dibandingkan dengan dokumen normal yaitu sebesar 0.8541.

Berdasarkan pengamatan penulis masih ada beberapa kata yang tidak bisa dilakukan *stemming* menggunakan *sastrawi python* yakni sebesar 4 dari 100 random kata, bagi peneliti selanjutnya dapat meningkatkan performa dalam *pre-processing* pada proses *stemming*, sehingga dapat meningkatkan nilai probabilitas kata dan efisiensi iterasi pada perhitungan LDA. Selain itu penggunaan LDA pada library gensim dalam pembentukan topik kurang konsisten, maka untuk penelitian lebih lanjut dapat menambahkan langkah khusus dalam penentuan z_0 atau membuat dictionary dan corpus yang dapat meningkatkan stabilitas dalam membuat model topik, misalnya menggunakan n-gram atau lainnya.

DAFTAR PUSTAKA

- [1] B. T. Shambhu, "WhatsApp: Unlocking The Goldmine." New Delhi, India: Educreation Publishing, 2016.
- [2] A. O. Afolaranmi, "Towards the Possibility of Internet Ministry as an Alternative Pastoral Ministry in Nigeria during the COVID-19 Pandemic," *Int. J. Inf. Technol. Lang. Stud.*, vol. 4, no. 2, 2020.
- [3] Z. Tong and H. Zhang, "A text mining research based on LDA topic modelling," in *International Conference on Computer Science, Engineering and Information Technology*, 2016, pp. 201–210.
- [4] G. Xu, Y. Meng, Z. Chen, X. Qiu, C. Wang, and H. Yao, "Research on topic detection and tracking for online news texts," *IEEE access*, vol. 7, pp. 58407–58418, 2019.
- [5] Y. Du, Y. Yi, X. Li, X. Chen, Y. Fan, and F. Su, "Extracting and tracking hot topics of micro-blogs based on improved Latent Dirichlet Allocation," *Eng. Appl. Artif. Intell.*, vol. 87, p. 103279, 2020.
- [6] D. Yu, D. Xu, D. Wang, and Z. Ni, "Hierarchical topic modeling of Twitter data for online analytical processing," *IEEE Access*, vol. 7, pp. 12373–12385, 2019.
- [7] A. Piepenbrink and A. S. Gaur, "Topic models as a novel approach to identify themes in content analysis," in *Academy of Management Proceedings*, 2017, vol. 2017, no. 1, p. 11335.
- [8] J.-F. Yeh, Y.-S. Tan, and C.-H. Lee, "Topic detection and tracking for conversational content by using conceptual dynamic latent Dirichlet allocation," *Neurocomputing*, vol. 216, pp. 310–

- 318, 2016.
- [9] M. H. Basri, "Identifikasi Topik Informasi Publik Media Sosial Di Kota Surabaya Berdasarkan Klasterisasi Teks Pada Twitter Dengan Menggunakan Algoritma K-means," Institut Teknologi Sepuluh Nopember, 2015.
- [10] M. Cendana and S. D. H. Permana, "Pra-Pemrosesan Teks Pada Grup Whatsapp Untuk Pemodelan Topik," *J. Mantik Penusa*, vol. 3, no. 3, 2019.
- [11] A. Rosenfeld, S. Sina, D. Same, O. Avidov, and S. Kraus, "WhatsApp usage patterns and prediction models," 2016.
- [12] C. Premalatha and S. J. Rani, "SENTIMENTAL ANALYSIS OF WHATSAPP DATA USING DATA ANALYTICS TECHNIQUES," *J. Data Min. Manag. (e-ISSN 2456-9437)*, vol. 2, 2017, Accessed: Jul. 05, 2021. [Online]. Available: <http://matjournals.in/index.php/JoDMM/article/view/1906>.
- [13] S. Patil, "WhatsApp Group Data Analysis with R," *Int. J. Comput. Appl.*, vol. 154, no. 4, 2016, doi: 10.5120/ijca2016912116.
- [14] P. M. Prihatini, I. Putra, I. A. D. Giriantari, and M. Sudarma, "Fuzzy-gibbs latent dirichlet allocation model for feature extraction on Indonesian documents," *Contemp. Eng. Sci.*, vol. 10, pp. 403–421, 2017.
- [15] P. M. Prihatini, I. K. Suryawan, and I. N. Mandia, "Feature extraction for document text using Latent Dirichlet Allocation," in *Journal of Physics: Conference Series*, 2018, vol. 953, no. 1, p. 12047.
- [16] A. Anjani, I. A. Ratnamulyani, and A. A. Kusumadinata, "Penggunaan Media Komunikasi Whatsapp terhadap Efektivitas Kinerja Karyawan," *J. Komun.*, vol. 4, no. 1, 2018.
- [17] L. Rahartri, "' WHATSAPP' MEDIA KOMUNIKASI EFEKTIF MASA KINI (STUDI KASUS PADA LAYANAN JASA INFORMASI ILMIAH DI KAWASAN PUSPIPTEK)," *VISI PUSTAKA Bul. Jar. Inf. Antar Perpust.*, vol. 21, no. 2, pp. 147–156, 2019.

Hyper Sudoku Solver dengan Menggunakan Harris Hawks Optimization Algorithm

Eric Dinata, *Teknologi Informasi Institut Sains dan Teknologi Terpadu Surabaya*, Herman Budiarto, *Teknik Informatika Institut Sains dan Teknologi Terpadu Surabaya*, dan Hendrawan Armanto, *Teknik Informatika Institut Sains dan Teknologi Terpadu Surabaya*

Abstrak—Sudoku merupakan salah satu permainan klasik yang digemari banyak orang. Sebagai salah satu permainan papan, Sudoku mempunyai banyak varian, salah satunya Hyper Sudoku. Hyper Sudoku mempunyai tingkat kesulitan yang lebih tinggi daripada Sudoku biasa. Tingkat kompleksitas yang tinggi membuat permainan ini menjadi *brain teaser* yang baik dan sangat cocok diambil sebagai media untuk menguji algoritma metaheuristik. Algoritma yang populer pada dekade terakhir ini adalah algoritma metaheuristik berbasis populasi, yang mengadaptasi perilaku binatang dalam memecahkan permasalahan optimasi, salah satunya adalah Harris Hawks Optimization (HHO). Seperti kebanyakan metode *swarm intelligence* (SI) lainnya, algoritma ini mengandalkan proses *diversification* dan *intensification*. Selain itu, HHO mempunyai empat strategi khusus untuk mencari solusi dengan kondisi yang berbeda. HHO mampu mencakup solusi multi dimensi, sehingga sangat cocok diimplementasikan pada persoalan Hyper Sudoku. Pada proses pengujian, digunakan dua *setting parameter* yang berbeda, tiga macam persoalan Hyper Sudoku, dan tiga puluh *independent run* untuk mencapai hasil yang diinginkan. Berdasarkan hasil pengujian, dapat disimpulkan bahwa tingkat keberhasilan untuk mencari solusi pada persoalan Hyper Sudoku dengan menggunakan HHO berkisar antara 86 hingga 88%, dilihat dari *fitness value*-nya.

Kata Kunci—HHO, Hyper Sudoku, metaheuristik, optimasi.

I. PENDAHULUAN

Permasalahan optimasi yang ada pada kehidupan sehari-hari, sering kali tidak dapat diselesaikan dengan persamaan matematika biasa. Hal ini membuat banyak orang mencoba melakukan pendekatan dengan algoritma metaheuristik. Algoritma metaheuristik secara umum dapat dibagi menjadi tiga kategori, yaitu *physical-based*, *evolutionary-based*, dan *swarm-based*. Beberapa metode yang populer meliputi Genetic Algorithms (GA), Particle Swarm Optimization (PSO), dan Ant Colony Optimization (ACO).

Dalam dekade terakhir, muncul banyak metode *swarm-based* dengan menggunakan nama-nama hewan, karena

Eric Dinata, Teknologi Informasi Institut Sains dan Teknologi Terpadu Surabaya, Surabaya, Jawa Timur, Indonesia (e-mail: eric.dinata@gmail.com)

Herman Budiarto, Teknik Informatika Institut Sains dan Teknologi Terpadu Surabaya, Surabaya, Jawa Timur, Indonesia (e-mail: herman.budiarto@gmail.com)

Hendrawan Armanto, Teknik Informatika Institut Sains dan Teknologi Terpadu Surabaya, Surabaya, Jawa Timur, Indonesia (e-mail: hendrawan@stts.edu)

memang metode ini mengimplementasikan perilaku natural dari jenis-jenis hewan tertentu pada algoritmanya, yang dianggap efektif dalam mencari solusi dalam masalah optimasi. Penelitian ini membahas sebuah metode baru yang dipublikasikan pada tahun 2019, yaitu Harris Hawks Optimization. Algoritma ini menggunakan pendekatan dari sekumpulan elang yang memantau dan mengepung mangsanya. Algoritma ini juga menggunakan dua fase, yakni *exploration (diversification)* dan *exploitation (intensification)*. Kedua fase ini umum dipakai pada metode-metode yang menggunakan *Swarm Intelligence* (SI).

Untuk permasalahan optimasi yang dipakai, penelitian ini menggunakan permainan Hyper Sudoku sebagai pokok masalah. Di mana permainan ini merupakan varian dari permainan Sudoku yang populer sejak tahun 1979, baik secara kertas maupun *online*. Hyper Sudoku juga sering dikenal sebagai Windoku, Four-square Sudoku, dan NRC Sudoku (pertama kali muncul pada koran Belanda NRC Handelsblad). Seperti yang kita ketahui, Sudoku standar memakai grid 9x9 dan memiliki aturan pokok, yaitu tidak boleh ada angka yang sama pada setiap baris, setiap kolom, dan setiap area yang ditentukan, dalam hal ini, sembilan area 3x3. Sedangkan Hyper Sudoku memiliki aturan tambahan, yaitu adanya area tambahan sebanyak empat kotak 3x3 (Four-square Sudoku) yang menyerupai jendela (Windoku).

II. HARRIS HAWKS OPTIMIZATION

Harris Hawks Optimization [1] merupakan *population-based metaheuristic algorithm*, dengan memimik pergerakan natural dari sekumpulan burung elang jenis Harris dalam mencari mangsa.



Gambar. 1. Elang Harris

Kunci utama dari HHO adalah perilaku kooperatif dan gaya pengejaran para elang yang disebut sebagai *surprise*

pounce (terkaman kejutan). Strategi ini dilakukan bersama-sama dengan cara mengepung sang mangsa dari berbagai arah. Ada beberapa pola pengejaran yang bisa dilakukan, bergantung pada berbagai skenario dan pola sang mangsa untuk melarikan diri. Dari berbagai pola tersebut, maka dibentuklah algoritma optimasinya.

Seperti kebanyakan metode *swarm intelligence* (SI) lainnya, HHO dibagi menjadi dua fase, yaitu fase *exploration* (eksplorasi) dan fase *exploitation* (exploitasi). Yang unik di HHO adalah adanya *perching strategies* (strategi hinggap para elang) pada fase *exploration* dan *escaping behavior* (perilaku melarikan diri mangsa) pada fase *exploitation*.

Ada beberapa pengembangan algoritma HHO, seperti *Harris Hawks optimization with information exchange* [2](IEHHO), yang menambahkan adanya pertukaran informasi antar elang, dan *Horizontal and vertical crossover of Harris hawk optimizer with Nelder-Mead simplex for parameter estimation of photovoltaic models* [3](CCNMHHO), yang menambahkan *crisscross optimizer* dan *Nelder-Mead simplex algorithm*, yang dipakai dalam perhitungan *solar power system* (sistem tenaga surya).

A. Fase pada HHO

Variabel E adalah energi sang mangsa untuk melarikan diri. Jika nilai $|E| \geq 1$, berarti energi sang mangsa masih sangat banyak, bisa berpindah tempat dengan cepat, sehingga pada saat ini para elang memasuki fase pencarian, atau fase *exploration*. Sedangkan jika nilai $|E| < 1$, maka dimulailah fase *exploitation*, dengan catatan apabila $|E| \geq 0.5$, maka akan memasuki fase *soft besiege* dan apabila $|E| < 0.5$, maka akan memasuki fase *hard besiege*.

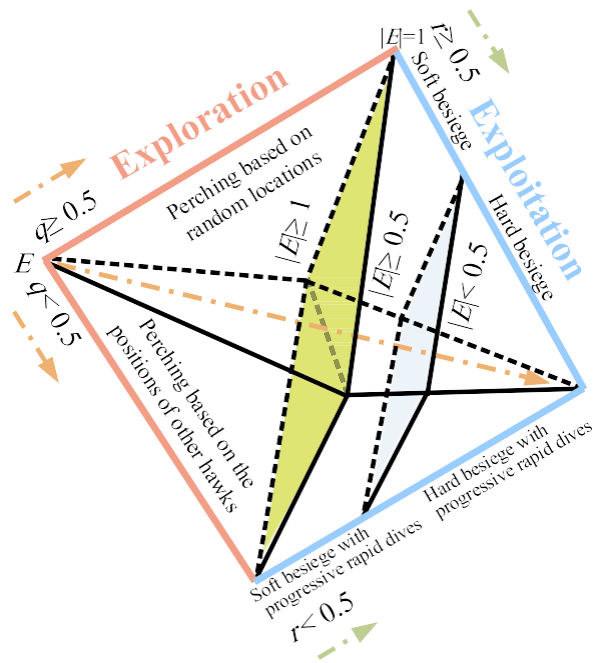
Variabel q adalah strategi hinggap, jika nilai $q < 0.5$, maka elang akan hinggap pada suatu posisi, dengan memperhitungkan posisi elang-elang lainnya. Jika nilai $q \geq 0.5$, maka elang akan hinggap pada pohon random atau bisa dibidang posisi random.

Variabel r adalah kemungkinan sang mangsa untuk melarikan diri. Jika nilai $r < 0.5$, maka kemungkinan sang mangsa untuk melarikan diri adalah besar. Sebaliknya, jika nilai $r \geq 0.5$, maka kemungkinan mangsa untuk melarikan diri adalah kecil.

B. Fase exploration

Pada fase ini, para elang akan melacak dan mendeteksi sang mangsa dengan penglihatan tajam mereka. Namun, hal ini tidak selalu mudah, dikarenakan sang mangsa sering kali tidak begitu saja muncul di area penglihatan para elang. Oleh karena itu, para elang akan menunggu, mengamati, dan memantau untuk mencari mangsa, bahkan sampai berjam-jam.

Pada HHO, para elang ibarat kandidat solusi dan kandidat terbaik pada setiap iterasi adalah yang mempunyai posisi yang sama dengan sang mangsa, atau yang paling mendekati.



Gambar. 2. Fase pada algoritma Harris Hawks Optimization

Pseudocode dari HHO

Input: jumlah populasi (N) dan maksimum iterasi (T)

Output: posisi kelinci dan fitness valuenya

Inisialisasi populasi random $X_i (i = 1, 2, \dots, N)$

while (kondisi berhenti belum tercapai) **do**

 Hitung fitness value elang

 Set X_{rabbit} sebagai posisi kelinci (best location)

for (setiap elang (X_i)) **do**

 Update energi awal (E_0) dan kekuatan lompat (J)

$E_0 = 2 \text{rand}() - 1$

$J = 2(1 - \text{rand}())$

 Update energi dengan persamaan (3)

 Fase eksplorasi

if ($|E| \geq 1$) **then**

 Update lokasi dengan persamaan (1)

 Fase eksploitasi

if ($|E| < 1$) **then**

 Soft besiege

if ($r \geq 0.5$ and $|E| \geq 0.5$) **then**

 Update lokasi dengan persamaan (4)

 Hard besiege

else if ($r \geq 0.5$ and $|E| < 0.5$) **then**

 Update lokasi dengan persamaan (6)

 Soft besiege with progressive rapid dives

else if ($r < 0.5$ and $|E| \geq 0.5$) **then**

 Update lokasi dengan persamaan (10)

 Hard besiege with progressive rapid dives

else if ($r < 0.5$ and $|E| < 0.5$) **then**

 Update lokasi dengan persamaan (11)

Return X_{rabbit}

Pada fase ini terdapat dua strategi, yang pertama, seekor elang akan hinggap pada suatu posisi dimana posisi ini dihitung atau ditentukan dari posisi kawan-kawannya dan juga posisi mangsa (supaya mereka berada pada posisi yang cukup dekat untuk menerkam mangsa), ini digambarkan pada persamaan 1, di mana kondisi $q < 0.5$, dan yang kedua, seekor elang akan hinggap pada posisi random, di mana

pada dunia nyata posisi ini adalah letak pepohonan yang dihindangi oleh para elang tersebut, ini digambarkan pada persamaan 1, dengan kondisi $q \geq 0.5$.

$$X(t + 1) = \begin{cases} X_{rand}(t) - r_1 |X_{rand}(t) - 2r_2 X(t)| & q \geq 0.5 \\ (X_{rabbit}(t) - X_m(t)) - r_3(LB + r_4(UB - LB)) & q < 0.5 \end{cases} \quad (1)$$

Variabel q merupakan *perching strategies* (strategi hinggap), di mana $X(t + 1)$ adalah posisi *vector* dari para elang pada iterasi selanjutnya, $X_{rabbit}(t)$ adalah posisi dari sang mangsa, kelinci, kemudian $X(t)$ adalah posisi *vector* para elang yang sekarang. Variabel r_1, r_2, r_3, r_4 , dan q adalah angka random di antara 0 dan 1, yang akan berganti terus pada setiap iterasi, kemudian LB dan UB adalah batas bawah (*lower bound*) dan atas (*upper bound*) untuk setiap variabel. $X_{rand}(t)$ adalah elang yang dipilih secara acak dari populasi saat ini, dan X_m adalah rata-rata posisi elang dari populasi saat ini.

Untuk menghitung rata-rata posisi elang dari populasi saat ini, digunakan persamaan berikut:

$$X_m(t) = \frac{1}{N} \sum_{i=1}^N X_i(t) \quad (2)$$

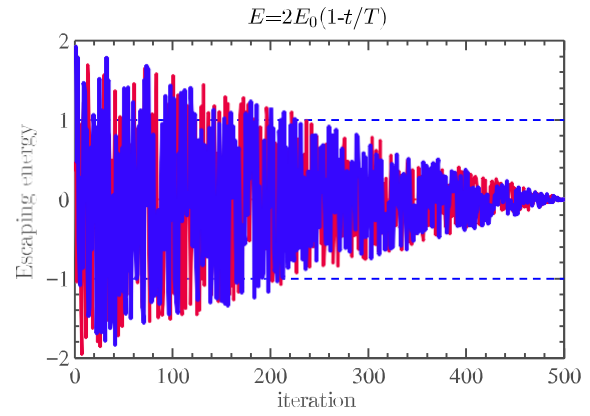
Di mana $X_i(t)$ merupakan lokasi dari setiap elang pada iterasi t dan N merupakan jumlah total elang.

C. Transisi dari *exploration* ke *exploitation*

Algoritma HHO akan mengalami transisi dari fase *exploration* ke fase *exploitation*. Kemudian, akan terjadi perubahan *exploitative behavior* (perilaku eksploitasi), yang bergantung pada *escaping energy* (tenaga untuk melarikan diri) dari sang mangsa dan juga *chance of prey escaping* (kemungkinan sang mangsa untuk berhasil melarikan diri). Untuk memodelkan kondisi ini, maka digunakan persamaan sebagai berikut:

$$E = 2E_0(1 - \frac{t}{T}) \quad (3)$$

Di mana E merupakan tenaga untuk melarikan diri sang mangsa, T adalah jumlah maksimum iterasi, dan E_0 adalah energi awal sang mangsa. Pada HHO, E_0 akan berganti secara acak pada setiap iterasi, dengan interval antara -1 dan 1. Jika nilai dari E_0 menurun dari 0 ke -1, maka itu berarti si kelinci melemah secara fisik. Sedangkan, sebaliknya, jika nilai E_0 naik dari 0 ke 1, maka itu berarti si kelinci mengalami penguatan secara fisik. Dinamika energi untuk melarikan diri ini cenderung mengalami penurunan pada proses iterasi. Pada saat $|E| \geq 1$, para elang akan mencari pada area yang berbeda untuk menjelajahi lokasi sang mangsa, yaitu dengan dilakukannya fase *exploration*. Sedangkan pada saat $|E| < 1$, algoritma ini akan memanfaatkan solusi-solusi yang berdekatan (*neighborhood*) pada fase *exploitation*. Singkatnya, fase *exploration* terjadi saat $|E| \geq 1$, sedangkan fase *exploitation* akan terjadi setelahnya, pada saat $|E| < 1$.



Gambar. 3. Hasil simulasi E

Simulasi E ini dilakukan dua kali, dengan jumlah iterasi masing-masing 500. Terdapat garis berwarna biru dan merah, yang mewakili simulasi pertama dan ke-dua. Seperti yang terlihat di gambar 3, *escaping energy* atau tenaga yang dimiliki sang mangsa untuk melarikan diri semakin lama akan semakin berkurang, atau mendekati 0.

D. Fase *exploitation*

Pada fase ini, para elang Harris akan melakukan *surprise pounce* (terkaman kejutan), atau disebut juga sebagai *seven kills*, dengan cara menyerang mangsa yang telah dideteksi pada fase sebelumnya. Akan tetapi, secara natural, sang mangsa akan sering kali mencoba untuk melarikan diri. Oleh karena itu, diperlukan strategi pengejaran yang berbeda pada kondisi sebenarnya. Berdasarkan dari perilaku melarikan diri dari sang mangsa dan strategi pengejaran para elang, maka ada empat macam strategi yang bisa dilakukan, yaitu *soft besiege*, *hard besiege*, *soft besiege with progressive rapid dives*, dan *hard besiege with progressive rapid dives*.

Sang mangsa akan selalu mencoba untuk melarikan diri dari situasi yang mengancam dirinya. Variabel r adalah besarnya peluang dari sang mangsa untuk berhasil melarikan diri (jika $r < 0.5$) dan gagal untuk melarikan diri (jika $r \geq 0.5$), sebelum *surprise pounce*. Apapun yang dilakukan sang mangsa, para elang akan melakukan *hard besiege* atau *soft besiege* untuk menangkap mangsa tersebut. Itu berarti para elang akan mengitari sang mangsa dari arah yang berbeda, secara *hard* atau *soft*, tergantung dari energi yang tersisa dari sang mangsa. Pada situasi sebenarnya, para elang akan mendekat perlahan-lahan menuju sang mangsa untuk memperbesar kesempatan mereka dalam bekerja sama membunuh sang mangsa dengan melakukan *surprise pounce*. Setelah beberapa menit, sang mangsa akan kehilangan banyak tenaga. Kemudian, para elang akan memperkuat proses pengepungan untuk menangkap sang mangsa yang sudah kehilangan banyak tenaga tersebut. Untuk memodelkan strategi ini dan menentukan apakah HHO akan menggunakan strategi *soft besiege* atau *hard besiege*, maka parameter E akan digunakan.

Dalam hal ini, jika $|E| \geq 0.5$, maka strategi *soft besiege* akan digunakan, dan jika $|E| < 0.5$, maka strategi *hard besiege* yang akan digunakan.

E. Soft besiege

Pada saat sang mangsa masih mempunyai tenaga yang cukup untuk melarikan diri dan kemudian mencoba untuk melompat ke sana kemari, para elang akan mengelilinginya secara perlahan, dengan tujuan untuk membuat sang mangsa kehabisan tenaga dan pada akhirnya akan melakukan *surprise pounce* (terkaman kejutan). Strategi ini terjadi jika $r \geq 0.5$ dan $|E| \geq 0.5$. Perilaku ini digambarkan dengan persamaan berikut:

$$X(t + 1) = \Delta X(t) - E|JX_{rabbit}(t) - X(t)| \quad (4)$$

$$\Delta X(t) = X_{rabbit}(t) - X(t) \quad (5)$$

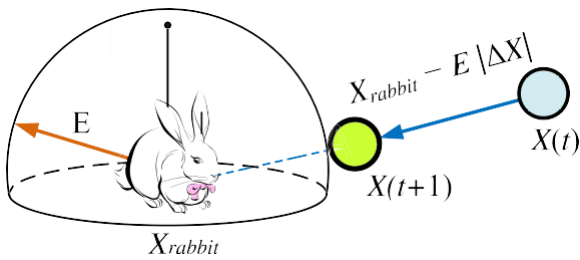
Di mana $\Delta X(t)$ adalah selisih antara posisi vector sang mangsa dan lokasi saat ini, pada iterasi t . Variabel r_5 merupakan angka random di antara 0 dan 1, yang akan digunakan untuk menghitung $J = 2(1 - r_5)$, yang merupakan kekuatan melompat kelinci sepanjang proses melarikan diri. Nilai J akan berganti dengan nilai acak pada setiap iterasi, untuk men-simulasi pergerakan natural kelinci.

F. Hard besiege

Pada saat sang mangsa sudah kehabisan tenaga dan kemungkinan untuk melarikan diri semakin kecil, maka para elang tidak lagi mengelilinginya secara perlahan, namun sudah bersiap untuk melakukan *surprise pounce*. Strategi ini terjadi jika $r \geq 0.5$ dan $|E| < 0.5$. Pada situasi ini, posisi elang pada iterasi selanjutnya akan dihitung dengan menggunakan persamaan berikut:

$$X(t + 1) = X_{rabbit}(t) - E|\Delta X(t)| \quad (6)$$

Model sederhana untuk menggambarkan langkah ini, dapat dilihat dari gambar berikut, dengan menggunakan satu ekor elang sebagai contoh:



Gambar. 4. Contoh vector pada hard besiege

G. Soft besiege with progressive rapid dives

Pada saat sang mangsa masih mempunyai tenaga yang cukup untuk melarikan diri, para elang tetap akan melakukan soft besiege sebelum melakukan *surprise pounce*. Strategi ini terjadi jika $|E| \geq 0.5$, namun $r < 0.5$.

Untuk memodelkan pola pergerakan melarikan diri dari sang mangsa dan juga *leapfrog movement* (pergerakan melompat), maka digunakanlah fungsi *levy flight* (LF). Fungsi LF ini digunakan untuk meniru pergerakan *zigzag* (berkelok-kelok) dari sang mangsa pada saat melarikan diri,

serta pergerakan tidak teratur dari para elang pada saat melakukan *rapid dives* (pergerakan menekuk secara cepat) untuk menangkap sang mangsa.

Untuk melakukan *soft besiege*, para elang akan mengevaluasi untuk menentukan pergerakan mereka selanjutnya melalui persamaan berikut:

$$Y = X_{rabbit}(t) - E|JX_{rabbit}(t) - X(t)| \quad (7)$$

Kemudian, mereka akan membandingkan pergerakan ini dengan pergerakan sebelumnya untuk menentukan apakah itu merupakan sebuah pergerakan yang bagus atau tidak. Jika tidak, maka mereka akan mulai melakukan pergerakan yang tidak teratur dan mendadak untuk mendekati sang mangsa. Pergerakan ini akan digambarkan melalui persamaan berikut:

$$Z = Y + S \times LF(D) \quad (8)$$

Di mana D adalah dimensi dari masalah, S adalah vector random dengan ukuran $1 \times D$, dan LF adalah fungsi *levy flight*, yang dapat dihitung dengan menggunakan persamaan berikut:

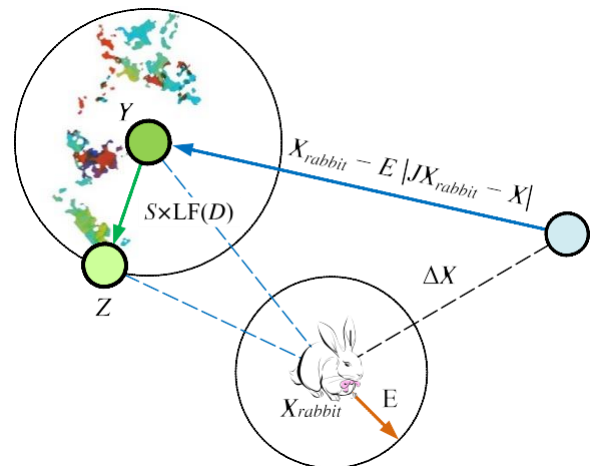
$$LF(x) = 0.01 \times \frac{u \times \sigma}{|v|^{\frac{1}{\beta}}}, \sigma = \left(\frac{\Gamma(1+\beta) \times \sin(\frac{\pi\beta}{2})}{\Gamma(\frac{1+\beta}{2}) \times \beta \times 2^{\frac{\beta-1}{2}}} \right)^{\frac{1}{\beta}} \quad (9)$$

Di mana u dan v merupakan bilangan random di antara 0 dan 1, kemudian β adalah konstanta yang ditetapkan pada nilai 1.5.

Pada akhirnya, untuk meng-*update* posisi para elang pada fase *soft besiege* ini, digunakanlah persamaan sebagai berikut:

$$X(t + 1) = \begin{cases} Y & \text{if } F(Y) < F(X(t)) \\ Z & \text{if } F(Z) < F(X(t)) \end{cases} \quad (10)$$

Di mana nilai Y dan Z didapatkan dari persamaan (7) dan (8). Ilustrasi sederhana langkah ini dapat dilihat pada gambar berikut, dengan menggunakan seekor elang sebagai contoh:



Gambar. 5. Contoh vector pada soft besiege with progressive rapid dives

Pada gambar 5 ini, terdapat titik-titik berwarna-warni, yang merupakan jejak lokasi dari pola perhitungan LF sebelumnya, yang kemudian mencapai posisi Z . Pada setiap langkah, hanya posisi Y atau Z yang lebih baik saja yang akan dipilih sebagai posisi selanjutnya. Strategi ini akan diterapkan pada setiap elang yang ada pada populasi.

H. Hard besiege with progressive rapid dives

Pada saat sang mangsa sudah tidak mempunyai tenaga yang cukup untuk melarikan diri, maka para elang akan melakukan *hard besiege* sebelum melakukan *surprise pounce*. Situasi pada langkah ini mirip dengan pada langkah *soft besiege*, hanya saja, pada langkah ini para elang mencoba memperkecil jarak rata-rata mereka dengan sang mangsa. Strategi ini terjadi jika $|E| < 0.5$ dan $r < 0.5$. Oleh karena itu, digunakanlah persamaan berikut:

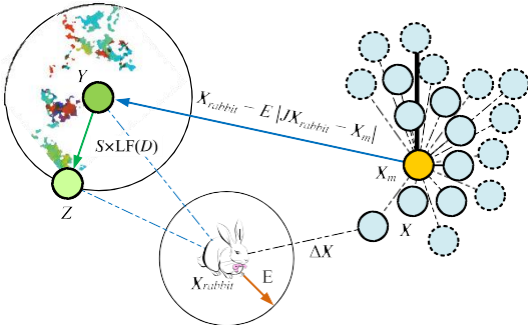
$$X(t + 1) = \begin{cases} Y & \text{if } F(Y) < F(X(t)) \\ Z & \text{if } F(Z) < F(X(t)) \end{cases} \quad (11)$$

Di mana nilai Y dan Z didapatkan dari persamaan berikut:

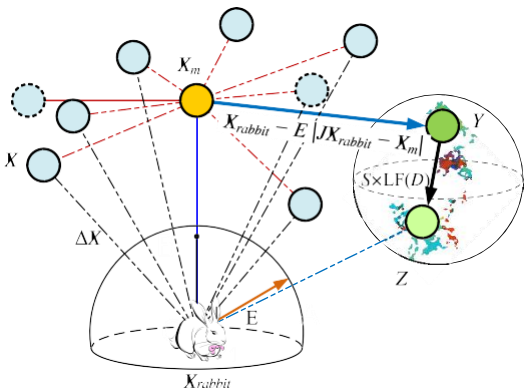
$$Y = X_{rabbit}(t) - E \lfloor |X_{rabbit}(t) - X_m(t) \rfloor \quad (12)$$

$$Z = Y + S \times LF(D) \quad (13)$$

Variabel $X_m(t)$ dapat dihitung dengan menggunakan persamaan (2.2). Ilustrasi sederhana dari langkah ini dapat dilihat dari dua gambar berikut:



Gambar. 6. Contoh vector pada hard besiege with progressive rapid dives (2D)



Gambar. 6. Contoh vector pada hard besiege with progressive rapid dives (3D)

Seperti halnya pada *soft besiege with progressive rapid dives*, titik-titik berwarna-warni merupakan jejak lokasi dari pola perhitungan LF sebelumnya dan hanya nilai Y atau Z saja yang akan dipakai sebagai posisi baru pada iterasi selanjutnya.

III. HYPER SUDOKU

Hyper Sudoku[4], [5] merupakan varian dari Sudoku standar, yang memiliki besar *grid* yang sama (9x9), yang terdiri dari 9 *region* 3x3. Yang membedakan Hyper Sudoku dari Sudoku standar adalah adanya 4 tambahan *region*^{[6][7]}, yang mengarah kepada adanya *constraint* tambahan, tidak boleh ada angka kembar pada *region* tambahan ini.

Angka-angka awal yang diberi dinamakan sebagai *givens*[6], [7]. Jumlah *givens* tidak mempengaruhi tingkat kesulitan. Untuk mencapai Sudoku dengan solusi unik, jumlah *givens* minimal untuk Sudoku standar adalah 17, sedangkan untuk Hyper Sudoku belum diketahui pasti, namun telah ditemukan persoalan Hyper Sudoku dengan 11 *givens*. Sudoku dengan solusi unik dinamakan sebagai *proper*^[7] Sudoku.

								1
		2					3	4
				5	1			
					6	5		
	7		3					8
		3						
				8				
5	8						9	
6	9							

Gambar. 7. Hyper Sudoku

Constraint pada Sudoku [8]:

- 1) Angka harus berada pada *range* 1 sampai dengan 9
- 2) Tidak ada angka yang sama pada setiap baris
- 3) Tidak ada angka yang sama pada setiap kolom
- 4) Tidak ada angka yang sama pada setiap *region*

Constraint pada Hyper Sudoku:

- 5) Tidak ada angka yang sama pada setiap *region* tambahan

Setiap *constraint* pada Hyper Sudoku ini akan dipakai untuk menghitung *fitness value* dalam satu solusi.

IV. IMPLEMENTASI

Implementasi algoritma HHO diawali dengan inisialisasi populasi elang, dengan angka random dalam batasan 0 (batas bawah) sampai dengan 1 (batas atas). Seekor elang mewakili satu solusi, di mana satu solusi mempunyai 81 dimensi (81 angka) yang di-*sort* secara per baris dan diberi indeks. Sembilan angka pertama mewakili baris pertama, sembilan angka ke-dua mewakili baris ke-dua, dan seterusnya. Dengan diberi indeks dan pengurutan per baris, maka dijamin tidak ada angka yang sama pada setiap baris.

TABEL I
INISIALISASI SOLUSI HYPER SUDOKU DENGAN HHO

X	unsorted	u-index	sorted	s-index
1	0.0185268	1	0.0185268	1
2	0.0751726	2	0.0751726	2
3	0.1938518	5	0.1418404	3
4	0.2032932	3	0.1938518	4
5	0.1418404	4	0.2032932	5
6	0.8858696	9	0.2326893	6
7	0.44012	7	0.44012	7
8	0.6946542	8	0.6946542	8
9	0.2326893	6	0.8858696	9

Angka yang ditampilkan mewakili baris pertama saja (sembilan angka pertama).

Angka-angka awal (*unsorted*) akan diurutkan (*sorted*), kemudian diberi indeks (*s-index*), setelah itu akan dilakukan *reverse indexing* untuk mengetahui indeks pada angka-angka awal (*u-index*).

Setelah itu, *fitness value* akan dihitung untuk setiap solusi, dengan cara menghitung banyaknya pelanggaran *constraint* terlebih dahulu:

- 1) Menghitung jumlah kotak kosong (a)
- 2) Menghitung jumlah kotak yang melanggar *constraint* baris (b)
- 3) Menghitung jumlah kotak yang melanggar *constraint* kolom (c)
- 4) Menghitung jumlah kotak yang melanggar *constraint region* awal pada Sudoku (d)
- 5) Menghitung jumlah kotak yang melanggar *constraint region* tambahan pada Hyper Sudoku (e)

$$Error = \frac{1}{5} \left(\frac{(a+b+c+d)}{81} + \frac{e}{36} \right) \quad (14)$$

Variabel a, b, c, dan d dibagi 81 karena angka maksimum pada *constraint-constraint* tersebut bernilai 81 (dari kotak 9x9). Sedangkan untuk variabel e dibagi 36 dikarenakan angka maksimum untuk *constraint* ini bernilai 36 (dari 4 kotak 3x3).

Berikutnya, karena *Error* adalah besarnya pelanggaran *constraint* dalam bentuk persentase (antara 0 dan 1), maka untuk menghitung *fitness value* (*FV*), digunakan persamaan sebagai berikut:

$$FV = 1 - Error \quad (15)$$

Nilai *fitness value* berkisar antara 0 dan 1, di mana nilai 0 merupakan yang paling jelek dan nilai 1 merupakan yang paling baik. Jika *fitness value* bernilai 1, berarti solusi untuk soal Hyper Sudoku telah berhasil ditemukan.

Setelah proses inisialisasi, solusi dengan *fitness value* terbaik akan dipakai sebagai patokan awal, atau pada HHO dipakai istilah kelinci (X_{rabbit}) dan pada setiap iterasi dan fase yang dilalui setiap elang, elang terbaik akan dibandingkan dengan solusi terbaik tersebut. Apabila elang baru mempunyai *fitness value* yang lebih baik, maka X_{rabbit} akan di-*update*, apabila tidak, maka X_{rabbit} yang lama-lah yang akan dipakai untuk iterasi berikutnya.

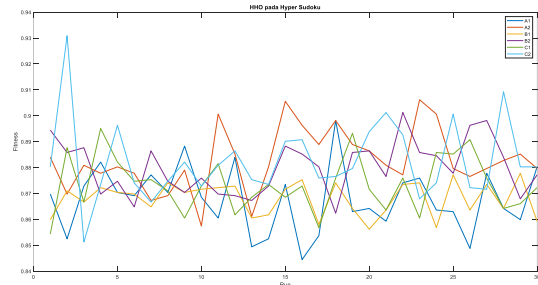
V. UJI COBA

Uji coba yang dilakukan adalah uji coba parameter, yaitu jumlah populasi, jumlah maksimum iterasi, dan juga jumlah *independent run*. Ada dua *setting parameter* yang diuji-cobakan, yang pertama adalah jumlah populasi 20 dengan jumlah maksimum iterasi 100. Kemudian yang ke-dua adalah jumlah populasi 30 dengan jumlah maksimum iterasi 500. Ke-dua *setting parameter* ini diuji-cobakan pada tiga soal Hyper Sudoku yang memiliki jumlah *givens* yang berbeda, masing-masing dilakukan 30 *independent run*.

TABEL II
UJI COBA HHO PADA HYPER SUDOKU DENGAN 30 RUN

HHO	A1	A2	B1	B2	C1	C2
Givens	15	15	18	18	19	19
Populasi	20	30	20	30	20	30
Iterasi	100	500	100	500	100	500
Run	30	30	30	30	30	30
Avg Time (s)	3.95	22.14	4.09	27.56	4.03	27.88
Avg Fitness	0.868	0.883	0.868	0.88	0.873	0.883
Run-1	0.87	0.884	0.86	0.894	0.854	0.88
Run-2	0.852	0.87	0.871	0.886	0.888	0.931
Run-3	0.873	0.881	0.867	0.888	0.867	0.851
Run-4	0.882	0.878	0.872	0.87	0.895	0.873
Run-5	0.87	0.88	0.87	0.875	0.882	0.896
Run-6	0.869	0.878	0.87	0.865	0.875	0.874
Run-7	0.877	0.867	0.865	0.886	0.875	0.867
Run-8	0.87	0.869	0.874	0.875	0.871	0.875
Run-9	0.888	0.879	0.87	0.87	0.86	0.882
Run-10	0.869	0.857	0.872	0.876	0.873	0.873
Run-11	0.86	0.901	0.872	0.87	0.881	0.881
Run-12	0.884	0.885	0.873	0.869	0.862	0.886
Run-13	0.849	0.861	0.86	0.867	0.869	0.875
Run-14	0.852	0.88	0.862	0.873	0.873	0.873
Run-15	0.873	0.906	0.872	0.888	0.869	0.89
Run-16	0.844	0.896	0.875	0.885	0.873	0.891
Run-17	0.854	0.889	0.858	0.88	0.857	0.876
Run-18	0.898	0.898	0.874	0.862	0.876	0.877
Run-19	0.863	0.889	0.864	0.886	0.893	0.88
Run-20	0.864	0.886	0.856	0.886	0.872	0.894
Run-21	0.859	0.881	0.864	0.877	0.864	0.901
Run-22	0.874	0.877	0.873	0.901	0.876	0.893
Run-23	0.876	0.906	0.874	0.886	0.86	0.868
Run-24	0.864	0.901	0.857	0.885	0.886	0.874

Huruf A mewakili soal A, begitu juga dengan huruf B dan C. Sedangkan angka pada A1 merupakan *setting parameter* 1 dan angka pada A2 merupakan *setting parameter* 2.



Gambar. 8. Grafik uji coba HHO pada Hyper Sudoku dengan 3 jenis persoalan, 2 macam *setting parameter*, dan 30 run.

Ada tiga jenis soal Hyper Sudoku dengan jumlah *givens* yang berbeda-beda yang digunakan pada uji coba ini, yakni soal A dengan jumlah *givens* sebanyak 15, soal B dengan 18, dan soal C dengan 19.

6			3			8		
				1	2			
		7						
				3		9		
			4	9				
		9	7					
5								
		4				7		
	8							

Gambar. 9. Soal Hyper Sudoku A

	5	4		3				
					9			
	3							4
		1			4			
	7		8	9		3		
	8							
			1			9		6
				7		5		

Gambar. 10. Soal Hyper Sudoku B

3	4			2			9	6
						3		
5			1			7		
						4	3	
2			7	9		1		
		6			3			
		5						
4								

Gambar. 11. Soal Hyper Sudoku C

Ada dua jenis setting parameter yang digunakan pada uji coba ini. Berikut perbandingan hasil kedua setting parameter tersebut pada masing-masing persoalan.

TABEL III

UJI COBA HHO PADA HYPER SUDOKU DENGAN SETTING PARAMETER 1

HHO	Soal A1	Soal B1	Soal C1
Givens	15	18	19
Populasi	20	20	20
Iterasi	100	100	100
Run	30	30	30
Avg Time (s)	3.952196	4.0927832	4.0303001
Avg Fitness	0.8676955	0.8680247	0.8734774

Uji coba dengan setting parameter 1 menggunakan jumlah populasi 20 dan jumlah maksimum iterasi 100.

TABEL IV

UJI COBA HHO PADA HYPER SUDOKU DENGAN SETTING PARAMETER 2

HHO	Soal A2	Soal B2	Soal C2
Givens	15	18	19
Populasi	30	30	30
Iterasi	500	500	500
Run	30	30	30
Avg Time (s)	22.141024	27.555999	27.884046
Avg Fitness	0.8827572	0.8800412	0.8825103

Uji coba dengan setting parameter 1 menggunakan jumlah populasi 30 dan jumlah maksimum iterasi 500.

Tabel III dan IV digunakan untuk membandingkan hasil antar persoalan Hyper Sudoku, dengan jumlah givens yang berbeda. Pada tabel III, terlihat peningkatan average fitness, namun tidak cukup signifikan. Pada tabel IV agak sedikit berbeda, terlihat average fitness pada soal A2 lebih tinggi dari average fitness pada soal yang lain, meskipun tidak banyak.

TABEL V

UJI COBA HHO PADA HYPER SUDOKU SOAL A

HHO	Soal A1	Soal A2
Givens	15	15
Populasi	20	30
Iterasi	100	500
Run	30	30
Avg Time (s)	3.952196	22.1410238
Avg Fitness	0.867695473	0.882757202

Uji coba dengan setting parameter 1 dan 2 pada Hyper Sudoku soal A.

TABEL VI

UJI COBA HHO PADA HYPER SUDOKU SOAL B

HHO	Soal B1	Soal B2
Givens	18	18
Populasi	20	30
Iterasi	100	500
Run	30	30
Avg Time (s)	4.0927832	27.5559993
Avg Fitness	0.868024691	0.880041152

Uji coba dengan setting parameter 1 dan 2 pada Hyper Sudoku soal B.

TABEL VII

UJI COBA HHO PADA HYPER SUDOKU SOAL C

HHO	Soal C1	Soal C2
Givens	19	19
Populasi	20	30
Iterasi	100	500
Run	30	30
Avg Time (s)	4.0303001	27.8840455
Avg Fitness	0.873477366	0.882510288

Uji coba dengan setting parameter 1 dan 2 pada Hyper Sudoku soal C.

Tabel V, VI, dan VII membandingkan hasil antar soal dengan setting parameter yang berbeda, apakah setting parameter 1 lebih baik atau setting parameter 2. Dari ketiga tabel, dapat dilihat bahwa average fitness dengan parameter ke-2 selalu lebih tinggi dari average fitness dengan parameter yang pertama.

Dari tabel II, dapat dilihat bahwa nilai *fitness* tertinggi ada pada *run* ke-2 soal C, dengan menggunakan *setting parameter* 2, yaitu bernilai sebesar 0.930864198.

3	4	1	5	2	7	8	9	6
7	6	5	1	8	9	3	4	2
4	8	9	2	3	5	1	6	7
5	3	4	1	6	8	7	2	9
1	7	8	5	9	2	4	3	6
2	3	5	7	9	4	1	8	6
4	8	6	9	5	3	2	7	1
9	1	5	4	2	3	6	7	8
4	8	2	1	7	6	9	3	5

Gambar. 12. Angka *fitness* tertinggi yang dicapai

Pada gambar 12, angka berwarna hitam berarti angka soal, angka berwarna hijau berarti angka yang benar, dan angka yang merah berarti angka yang salah.

VI. KESIMPULAN

Dari sini hasil uji coba, dapat ditarik beberapa kesimpulan, yang pertama adalah, *jumlah givens* yang lebih banyak tidak menjamin *fitness value* yang lebih tinggi. Yang kedua, jumlah populasi dan jumlah maksimum iterasi yang lebih banyak akan meningkatkan *fitness value*. Yang ketiga, tingkat keberhasilan untuk menemukan solusi pada semua parameter dan persoalan berkisar antara 86-88%.

Selain itu, dapat disimpulkan bahwa waktu komputasi tidak bergantung pada banyaknya jumlah *givens* pada persoalan. Namun, untuk perbedaan parameter, jumlah populasi dan jumlah maksimum iterasi, jelas mempengaruhi waktu komputasi secara signifikan.

Daftar Pustaka

- [1] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, "Harris hawks optimization: Algorithm and applications," *Futur. Gener. Comput. Syst.*, vol. 97, 2019, doi: 10.1016/j.future.2019.02.028.
- [2] C. Qu, W. He, X. Peng, and X. Peng, "Harris hawks optimization with information exchange," *Appl. Math. Model.*, vol. 84, pp. 52–75, 2020.
- [3] Y. Liu *et al.*, "Horizontal and vertical crossover of Harris hawk optimizer with Nelder-Mead simplex for parameter estimation of photovoltaic models," *Energy Convers. Manag.*, vol. 223, 2020, doi: 10.1016/j.enconman.2020.113211.
- [4] J. M. Weiss, "Genetic algorithms and sudoku," in *Midwest Instruction and Computing Symposium (MICS 2009)*, 2009, pp. 1–9.
- [5] B. Michel, "Mathematics of NRC-Sudoku," 2007.
- [6] M. Weller, "Counting, generating, and solving Sudoku," *Ph. D. Diss.*, 2008.
- [7] A. Moraglio and J. Togelius, "Geometric particle swarm optimization for the sudoku puzzle," in *Proceedings of the 9th annual conference on Genetic and evolutionary computation*, 2007, pp. 118–125.
- [8] S. McGerty, "Solving Sudoku puzzles with particle swarm optimisation," *Final Report, Macquarie Univ.*, 2009.

INFORMATION EXTRACTION BERBASIS RULE UNTUK SOAL UJIAN

Stefanus Nico Soenardjo, *Teknologi Informasi Institut Sains dan Teknologi Terpadu Surabaya,*
Gunawan, *Teknik Informatika Institut Sains dan Teknologi Terpadu Surabaya*

Abstrak—Proses information extraction dapat dilakukan pada beberapa macam media, seperti artikel berita, tanya jawab dan sebagainya. Penelitian ini mencoba untuk melakukan information extraction pada media soal ujian yang dilengkapi dengan jawaban.

Pendekatan pengolahan informasi yang dibahas dalam penelitian ini adalah information extraction berbasis rule. Informasi yang hendak digali adalah informasi data soal ujian beserta jawabannya. Inputan dalam penelitian ini pasangan file soal dan jawaban milik Cambridge. Ada beberapa mata pelajaran yang digunakan, yaitu Biologi, Matematika dan Ekonomi. Jenis soal yang digunakan juga ada beberapa macam, yaitu pilihan ganda dan esai. Hasil penelitian ini diharapkan bisa menjadi media pembelajaran.

Penelitian dilakukan dengan menggunakan sebanyak 100 pasang data soal dan ujian. Sistem akan menerima 2 inputan file dengan format PDF. Kedua file ini merupakan pasangan soal dan jawaban. Proses yang dilakukan adalah file akan dirubah menjadi 2, yaitu file HTML dan file PNG. File HTML mengandung semua teks soal dan file PNG mengandung semua gambar dari soal. Sistem akan mengambil teks dan gambar dari masing-masing soal dan jawaban berdasar rule yang sudah ditentukan. Penentuan rule dilakukan secara manual dengan mempelajari pola-pola data yang terdapat dalam tag HTML. Setelah proses ekstraksi, soal dan jawaban ini dipasangkan sesuai dengan nomor urutnya masing-masing. Pasangan soal dan jawaban ini kemudian akan disimpan ke dalam database. Dari hasil penelitian, tingkat akurasi yang didapatkan adalah sekitar 46%. Kendala utama yang dihadapi adalah format soal dan jawaban yang tidak standar sehingga menimbulkan kesulitan dalam proses ekstraksi informasi.

Kata Kunci—Information Extraction, berbasis rule, soal ujian, PDF.

I. PENDAHULUAN

Pandemi virus corona (Covid-19) pada tahun 2020 sangat mengganggu aktivitas banyak orang. Banyak orang yang mengurangi aktivitas di luar rumah bila memungkinkan. Kantor-kantor ada yang menerapkan pola kerja WFH (Work From Home) untuk bagian/divisi yang tidak memerlukan kontak fisik langsung untuk melaksanakan pekerjaannya, kadang juga ada yang menerapkan masuk kantor bergantian untuk menghindari penyebaran virus.

Stefanus Nico Soenardjo, Teknologi Informasi Institut Sains dan Teknologi Terpadu Surabaya, Surabaya, Jawa Timur, Indonesia (e-mail: stefanusnicoid@gmail.com)

Gunawan, Teknik Informatika Institut Sains dan Teknologi Terpadu Surabaya, Surabaya, Jawa Timur, Indonesia (e-mail: gunawan@stts.edu)

Salah satu bagian kehidupan yang sangat terdampak oleh pandemi corona dalam hal ini adalah sektor pendidikan. Dalam keadaan normal, kegiatan pendidikan dilakukan dengan bertatap muka antara pengajar (guru) dengan para murid dalam ruangan kelas di sekolah. Dalam satu kelas ada beberapa murid, jumlah murid beragam, tergantung kebijakan dari masing-masing sekolah. Ditengah pandemi corona ini, pertemuan fisik antara guru dan murid ini tidak memungkinkan, karena berisiko besar untuk menyebarkan virus corona. Oleh karena itu kegiatan belajar dan mengajar dilakukan melalui jalan online. Beberapa aplikasi yang biasa digunakan seperti: Zoom, Micosoft Team dan Google Class.

Dengan pembelajaran online maka guru cenderung akan lebih banyak untuk memberikan tugas lebih banyak bila dibandingkan pembelajaran konvensional. Pemberian penjelasan mengenai suatu materi secara online mempunyai kekurangan bila dibandingkan dengan secara tatap muka langsung. Supaya murid lebih bisa menangkap materi maka diberi tugas yang porsinya lebih banyak, dengan harapan bila murid lebih banyak praktek maka akan lebih mengerti materi yang diberikan.

Bagian yang diharapkan bisa dibantu dalam penelitian ini adalah mengenai pemberian latihan atau ujian online. Di internet ada banyak soal ujian yang disediakan oleh beberapa institusi pendidikan. Dataset yang digunakan dalam penelitian adalah milik soal ujian milik Cambridge International Examinations. Dataset ini terdiri dari pasangan file PDF soal dan jawaban. Penelitian ini berusaha untuk melakukan pendekatan berbasis rule untuk mengekstrak soal serta jawaban kemudian berusaha untuk memasangkan antara soal dengan jawabannya. Pasangan antara soal dan jawaban ini kemudian disimpan dalam suatu bentuk database dan kemudian bisa dimanfaatkan lebih lanjut untuk kegiatan pembelajaran.

Lebih lanjut, pada penelitian ini penekanan dilakukan pada pengambilan atau ekstraksi informasi dari soal-soal ujian dalam bentuk pdf. Data soal yang digunakan milik Cambridge International Examinations, yang berupa pasangan antara file Question Paper (QP) dan file Mark Scheme (MS). File QP berisi soal-soal ujian, sedangkan file MS berisi jawaban. Di dalam file QP terdapat beberapa macam jenis soal, seperti esai, pilihan ganda, ada yang memakai gambar dan sebagainya. Pendekatan yang akan dilakukan untuk melakukan Information Extraction (IE) dalam penelitian ini adalah dengan menggunakan sistem berbasis rule. Penentuan rule akan melihat kondisi dari struktur dokumen yang terdapat pada file soal dan jawaban. Hasil ekstraksi yang berupa pasangan antara soal dan

jawaban ini diharapkan bisa dijadikan sebagai bahan pembelajaran.

II. TINJAUAN PUSTAKA

Menurut Wikipedia, Information Extraction (IE) merupakan proses pengambilan informasi dari dokumen yang tidak terstruktur atau semi terstruktur menjadi informasi yang terstruktur. IE [1]–[3] sendiri merupakan salah 1 bentuk dari Natural Language Processing (NLP). Media yang menjadi sumber dari IE bisa berasal dari berbagai macam sumber, seperti: gambar, video, audio, maupun teks.

IE berbasis rule [4], [5] menggunakan satu set atau beberapa aturan dalam proses pengambilan informasi. Proses penentuan aturan ini dengan memperhatikan bagaimana kondisi dalam dokumen, terutama ciri-ciri yang terdapat dalam teks yang menjadi tujuan utama dalam pengambilan informasi. Metode pendekatan yang dilakukan untuk proses IE dibagi menjadi 2 dimensi, yaitu:

1. Hand-coded atau Learning-based

Pada sistem berbasis Hand-coded, diperlukan manusia untuk mendefinisikan rule atau regular expression untuk melakukan proses IE. Untuk bisa menghasilkan rule yang baik, maka diperlukan orang yang mengerti mengenai domain linguistic dan pemrograman untuk bisa menghasilkan rule yang baik. Pada sisi lain, untuk Learning-based diperlukan data yang sudah dilabeli secara manual sebagai bahan pelatihan. Tentu data yang digunakan untuk proses pelatihan harus valid, karena mempengaruhi hasil IE yang akan dilakukan oleh sistem. Pemilihan antara kedua pendekatan ini dipengaruhi oleh kondisi informasi yang akan digunakan, baik dari segi struktur maupun noise-noise yang terdapat dalam data.

2. Rule-based atau Statistical

Rule based menggunakan suatu set aturan yang sudah di prediksi untuk melakukan IE, sedangkan statistical menggunakan suatu system pembobotan di mana bobot yang didapat menjadi dasar acuan untuk melakukan IE. Rule-based lebih mudah untuk diinterpretasikan dan dikembangkan, sedangkan untuk statistical lebih baik dalam menangani noise dalam data yang tidak tersruktur. Penerapan Rule-based lebih baik untuk data yang terbatas dalam domain tertentu, sedangkan statistical lebih baik untuk domain umum seperti pencarian fakta dari transkrip pidato.

Rule yang digunakan bisa ditentukan sendiri sesuai dengan keperluan ataupun bisa dihasilkan dengan melakukan pelatihan oleh program dengan menggunakan metode-metode tertentu. Tentu saja kondisi data sangat berperan penting dalam pembentukan rule, ada kalanya data memerlukan semacam preprocessing terlebih dahulu. Pada penelitian ini proses pembentukan rule ditentukan sendiri dengan melihat kondisi soal ujian.

Dalam pembuatan penelitian ini digunakan beberapa tinjauan pustaka untuk membantu dalam proses penelitian [6], [7]. Tinjauan pustaka yang digunakan tidak sama persis dengan penelitian ini, tetapi memiliki prinsip pendekatan yang mirip. Tinjauan Pustaka ini membantu dalam proses untuk menetapkan pendekatan yang mau digunakan dalam melakukan ekstraksi terhadap soal-soal ujian.

III. INFORMATION EXTRACTION BERBASIS RULE UNTUK SOAL UJIAN

A. Analisa Permasalahan

Masalah utama dalam penelitian ini adalah melakukan information extraction terhadap soal-soal ujian beserta dengan jawabannya. Inputan berupa pasangan file soal ujian dan jawaban dengan format PDF. Pendekatan yang dipilih dalam penelitian ini adalah information extraction berbasis rule. Pada Tabel 1 dapat dilihat keuntungan dan kekurangan dari penggunaan information extraction yang berbasis rule dengan yang berbasis machine learning.

TABEL I
PERBANDINGAN RULE BASED DAN MACHINE LEARNING BASED

Rule Based	Machine Learning Based
<p>Keuntungan</p> <ol style="list-style-type: none"> 1. Deklaratif 2. Mudah dipahami 3. Mudah dimaintain 4. Mudah untuk memasukkan suatu domain pengetahuan 5. Mudah untuk melacak penyebab kesalahan dan memperbaikinya 	<p>Keuntungan</p> <ol style="list-style-type: none"> 1. Dapat dilatih 2. Beradaptasi 3. Mengurangi kerja manual untuk mengadaptasi suatu domain pengetahuan
<p>Kekurangan</p> <ol style="list-style-type: none"> 1. Heuristik 2. Membutuhkan kerja manual yang banyak 	<p>Kekurangan</p> <ol style="list-style-type: none"> 1. Membutuhkan data berlabel 2. Membutuhkan pelatihan ulang 3. Membutuhkan keahlian machine learning untuk menggunakan atau maintain

Proses ekstraksi akan di lakukan pertama kali pada file soal. Bentuk soal yang terdapat dalam file pdf ini ada beberapa macam, seperti soal esai, pilihan ganda dan sebagainya. Pada masing-masing jenis soal mempunyai tingkat kesulitan sendiri-sendiri. Hal yang paling menyulitkan adalah tidak adanya standarisasi mengenai format soal. Walaupun dalam satu mata pelajaran yang sama dan jenis soal yang sama, belum tentu format penulisan pada file soal dan jawabannya sama. Beberapa pendekatan yang dilakukan untuk mengatasi masalah ini:

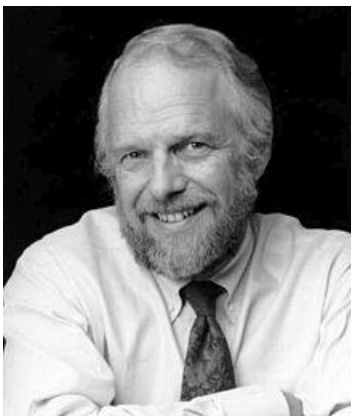
1. Membatasi jumlah mata pelajaran. Mata pelajaran yang digunakan dalam penelitian ini adalah Biologi, Matematika dan Ekonomi. Dengan pembatasan ini, maka setidaknya akan mengurangi banyaknya variasi pola soal dan jawaban yang harus dilakukan secara manual.
2. Mencari ciri format penulisan sebanyak mungkin, sehingga sistem dapat mengidentifikasi soal dan jawaban sebanyak mungkin. Hal ini penting karena mempengaruhi dari tingkat keberhasilan dalam melakukan information extraction.
3. Rule dibentuk hardcode dan statis dalam program, karena melihat kerumitan dalam parsing dokumen HTML. Bentuk rule akan mengikuti pola dari tag HTML hasil ekstraksi.

Untuk proses mengakses file PDF akan digunakan beberapa library tambahan. Pada fase ini file PDF akan diubah menjadi HTML sehingga lebih mudah untuk diproses. Dari file HTML akan dimulai proses untuk

mengidentifikasi soal dan jawaban. Pada proses ini, sistem akan berusaha untuk mencari pola dalam tag HTML untuk mencari soal dan jawaban. Soal dan jawaban mempunyai pola tag HTML sendiri-sendiri. Pencarian pola dalam hal ini mempunyai peranan dalam pembentukan rule, sebagai dasar dalam information extraction. Setelah berhasil mendapatkan soal dan jawaban maka langkah selanjutnya adalah memasang soal ke jawaban yang sesuai. Setelah mendapatkan pasangan soal dan jawaban, maka sistem akan menyimpan file tersebut ke dalam database sesuai dengan format yang telah ditentukan.

B. File PDF

PDF adalah kependekan untuk Portable Document Format, PDF merupakan format file yang dikembangkan oleh Adobe. Pada tahun 1991, salah satu pendiri Adobe Dr. John Warnock meluncurkan revolusi kertas-ke-digital dengan ide yang disebutnya, Camelot Project. Tujuannya adalah agar siapa pun dapat mengubah dokumen ke dalam bentuk digital, sehingga kemudian dapat mengirim nya ke mana pun dan mencetak kembali dokumen tersebut. Pada tahun 1992, Camelot telah berkembang menjadi PDF. Saat ini, format tersebut dipercaya oleh bisnis di seluruh dunia.



Gambar. 1. Dr. John Warnock

PDF mampu menampung berbagai format seperti teks, gambar, tabel dan sebagainya. PDF bisa dikirim secara digital seperti melalui email atau media komunikasi yang lain. Pihak penerima dapat membuka file tersebut dan mendapat tampilan yang sesuai dengan kondisi sama dengan yang dibuat oleh pihak pengirim (mempertahankan jenis huruf (font), gambar, grafik serta tata letak yang tepat dari berkas aslinya). PDF juga menyediakan enkripsi dan tanda tangan digital.

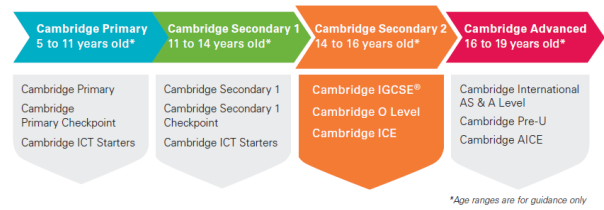
C. Input dan Output Sistem

Dataset yang digunakan dalam penelitian adalah milik soal ujian milik Cambridge International Examinations, yang terdiri dari:

1. Ordinary Level (O)

O Level merupakan program sertifikasi yang dipelopori oleh University of Cambridge. Soal ujian ini untuk anak usia 14-16 tahun. Setara dengan IGCSE. Cambridge O Level menerima 620.000 pendaftaran setahun di lebih dari 50

negara di seluruh dunia.



Gambar. 2. Cambridge International Examinations

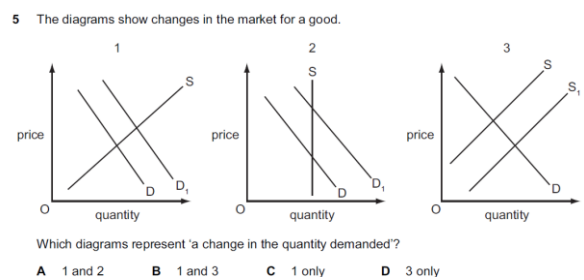
2. International General Certificate of Secondary Examinations (IGCSE)

IGCSE merupakan program sertifikasi yang dipelopori oleh University of Cambridge. Soal ujian ini untuk anak usia 14-16 tahun. Urutan jenjang program bisa dilihat pada Gambar3.2. IGCSE juga merupakan dasar yang ideal untuk program tingkat yang lebih tinggi seperti A Level. Setara dengan O Level. Daftar grup mata pelajaran IGCSE adalah Languages, Humanities and Social Sciences, Sciences, Mathematics, Creative and Professional.

3. Advance Level (A)

A Level merupakan program sertifikasi yang dipelopori oleh University of Cambridge, ditujukan untuk siswa berusia 16–19 tahun sebelum masuk ke jenjang universitas, Cambridge International AS & A Levels diikuti oleh lebih dari 175.000 siswa di lebih dari 125 negara setiap tahun. Daftar grup mata pelajaran A Level adalah English, Mathematics, Science, Languages, Humanities, Technology, Social Sciences, The Arts, General Studies.

Untuk setiap level mempunyai ujian dalam berbagai mata pelajaran, daftar bisa dilihat di Lampiran A. Data yang digunakan berupa sepasang file pdf yang terdiri dari Question Paper (QP) dan Mark Scheme (MS). Contoh pasangan antara QP dan MS dapat dilihat pada Gambar 3.1.



Gambar. 3. Contoh Question Paper

Pada Gambar 3 dan 4 bisa dilihat contoh Question Paper dan Mark Scheme dengan mata pelajaran ekonomi. Jenis soal pada gambar ini adalah soal pilihan ganda.

RULE CHECK ANCHOR

1. IF setelah cleanup RETURN angka THEN (cleanup = memecah br, dan ambil posisi masing-masing br ada yang isinya berupa angka)
2. IF angkanya urut THEN titik tersebut dikembalikan sebagai anchor.

Anchor yang dimaksud dalam hal ini adalah titik awal untuk pengambilan soal, yaitu nomor soal. Hal ini dilakukan dengan membaca isi didalam tag HTML jika ditemukan angka numerik di awal tag, maka anggap sebagai nomor hingga sampai ketemu nomor selanjutnya yang urut.

RULE PENGAMBILAN SOAL BERDASARKAN:

1. CHECK dari <p> ke <p> selanjutnya: IF batas kiri sama dengan anchor THEN jika YA ke point 2, ELSE ke point 4
2. IF batas atas/kiri masih limit dan angka yang ditentukan (menghindari pengambilan nomer halaman) THEN jika iya ke point 3
3. Masukkan komponen html dan text ke dalam array
4. IF content dalam batas kiri ada yang mengandung angka, dan sesuai urutan (jika sebelumnya 1, maka yang dicari 2) THEN jika YA ke point 5
5. Masukkan semua data array html ke dalam class tersendiri, Selesai pengecekan soal didapat jumlah soal

2. Simpan semua jumlah content yang mengandung angka di masing-masing batas kiri
3. IF dari masing-masing jumlah data di batas kiri ada yang sama dengan soal dan isi angkanya urut THEN batas kiri tersebut diambil sebagai patokan soal, ELSE loncat ke point 4
4. IF ada 2 kelompok batas kiri yang jumlahnya sama dengan jumlah soal, THEN 2 batas kiri tersebut akan dimerge (mengatasi terjadinya kunci jawaban yang batas kirinya variatif). jika tidak loncat ke point 5
5. IF angka semua angka pada batas kiri dalam range tertentu, AND dimerge jika jumlahnya sesuai dengan jumlah soal THEN batas kiri tersebut dijadikan anchor, RLSE ke point 6
6. Dari halaman kedua ambil semua jawaban dengan pemisahan spasi cocokkan dengan jumlah soal (hal ini karena ada beberapa soal yang 2kolom) untuk point 1-5 pengambilan jawaban sama dengan pengambilan soal dari anchor ke anchor lainnya

IV. UJICOBA

Dari total pengujians sebanyak 100 pasangan soal dan jawaban, yang berhasil di ekstrak adalah sebanyak 46 soal. Maka bisa dilihat tingkat keberhasilan pengambilan soal adalah sebanyak 46%. Angka ini masih dibawah angka harapan pada hipotesa yaitu sebanyak 65%.

V. KESIMPULAN

Berdasarkan hasil penelitian ini, penulis mengambil kesimpulan dari penelitian information extraction berbasis rule untuk soal ujian, yaitu

1. Hal yang paling mempengaruhi dalam proses information extraction adalah format dari soal dan jawaban yang menjadi inputan. Tidak ada standar baku dalam penulisan soal dan jawaban. Dalam satu mata pelajaran yang sama pun, format yang digunakan bisa berbeda-beda sehingga sangat mengganggu proses information extraction.
2. Untuk tetap bisa mempertahankan bentuk soal, beberapa tag HTML yang berasal dari hasil proses preprocessing dimana inputan awal PDF diubah menjadi HTML, ikut disimpan ke dalam database, hal ini akan membantu dalam proses penampilan ulang. Bila hal ini tidak dilakukan maka akan sulit untuk menampilkan kembali bentuk awal dari soal, karena tidak ditemukan elemen lain penanda lokasi dan bentuk soal.
3. Program sudah berhasil untuk melakukan proses ekstraksi pada data berbentuk teks, gambar dan formula matematika. Namun pada waktu penyajian data masih sering terjadi pergeseran layout, terutama pada data yang mempunyai tabel. Hal ini merupakan salah satu kendala utama, yang dipengaruhi tidak adanya format yang baku dalam penulisan soal dan jawaban.
4. Pembentukan rule dilakukan secara manual dengan meneliti pola dari tag HTML hasil proses



Gambar. 8. Contoh Isi HTML

Pada Gambar 8 akan menunjukkan ilustrasi bagaimana mengambil soal. Bisa dilihat kalau bagian A, B dan C merupakan awal dari soal. Maka dalam hal ini digunakan Rule Check Anchor dan Rule Pengambilan soal. Pada bagian A ditemukan data numerik yaitu angka 11, kemudian pembacaan perbaris diteruskan hingga sampai pada bagian B. Ternyata pada awal bagian B ditemukan angka 12 maka sudah memenuhi kriteria rule, maka awal bagian A sampai ke awal bagian B dianggap sebagai 1 soal tersendiri. Koordinat Top dan Left pada tag HTML yaitu (94,74) hingga (254,74) juga disimpan untuk melakukan crop pada image background.

RULE tentukan anchor jawaban :

1. Cari batas kiri masing-masing jawaban dengan cara yang sama dengan soal

preprocessing. Informasi yang ingin diambil berada didalam tag HTML ini. Pada penelitian ini rule yang terbentuk tidak dalam bentuk yang ideal, yaitu hardcode dan statis pada program. Ada 3 rule utama yaitu mencari anchor poin, yaitu titik start untuk pengambilan data (dengan mengidentifikasi nomor soal atau nomor jawaban), rule untuk pengambilan soal dan rule untuk pengambilan jawaban.

5. Hasil ekstraksi data pasangan soal dan jawaban bisa digunakan sebagai bahan pembelajaran. Pemanfaatan bisa disesuaikan sesuai dengan pihak yang menggunakan, baik pengajar maupun murid. Untuk pengajar bisa digunakan untuk menyusun materi dari ujian, untuk murid kumpulan soal dan jawaban bisa digunakan sebagai untuk latihan.
6. Hipotesis kurang terpenuhi yaitu hanya 46% dari angka harapan semula sekitar 65%. Tidak tercapainya angka harapan ini terutama dipengaruhi oleh format soal yang tidak standar.

DAFTAR PUSTAKA

- [1] J. Tang, M. Hong, D. Zhang, B. Liang, and J. Li, "Information extraction: Methodologies and applications," in *Emerging Technologies of Text Mining: Techniques and Applications*, 2007.
- [2] S. Sarawagi, *Information Extraction*. Now Publishers, 2008.
- [3] R. Gaizauskas and Y. Wilks, "Information extraction: Beyond document retrieval," *J. Doc.*, vol. 54, no. 1, 1998, doi: 10.1108/EUM0000000007162.
- [4] "A Rule-based Information Extraction System," *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 9, 2019, doi: 10.35940/ijitee.i8156.078919.
- [5] L. Chiticariu, Y. Li, and F. R. Reiss, "Rule-based information extraction is dead! Long live rule-based information extraction systems!," 2013.
- [6] Y. Lin, Z. Jun, M. Hongyan, Z. Zhongwei, and F. Zhanfang, "A method of extracting the semi-structured data implication rules," in *Procedia Computer Science*, 2018, vol. 131, doi: 10.1016/j.procs.2018.04.315.
- [7] N. Bhutani, Y. Suhara, W. C. Tan, A. Halevy, and H. V. Jagadish, "Open information extraction from question-answer pairs," in *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 2019, vol. 1, doi: 10.18653/v1/n19-1239.

Market Basket Analysis untuk Swalayan KSU Sumber Makmur dengan Algoritma FP Growth

Ramadhan, *Teknologi Informasi Institut Sains dan Teknologi Terpadu Surabaya*,
Esther Irawati Setiawan, *Teknik Informatika Institut Sains dan Teknologi Terpadu Surabaya*.

Abstrak—Salah satu teknik data mining yang populer digunakan adalah association data mining atau yang biasa disebut dengan istilah market basket analysis. Market basket didefinisikan sebagai suatu itemset yang dibeli secara bersamaan oleh pelanggan dalam suatu transaksi. Market basket analysis adalah suatu sarana untuk meningkatkan penjualan. Metode ini dimulai dengan mencari sejumlah frequent itemset dan dilanjutkan dengan pembentukan aturan-aturan asosiasi. Algoritma Apriori dan frequent pattern growth adalah dua algoritma yang sangat populer untuk menemukan sejumlah frequent itemset dari data-data transaksi yang tersimpan dalam basis data. Dalam penelitian ini algoritma frequent pattern growth (FP Growth) digunakan untuk menemukan sejumlah aturan asosiasi dari basis data transaksi penjualan di Swalayan KSU Sumber Makmur (Trenggalek). Dari hasil pengolahan data didapatkan pola pembelian paling kuat berupa jika membeli pasta gigi maka dimungkinkan juga akan membeli sabun dan jika membeli shampo juga akan membeli sabun dengan tingkat keyakinan (confidence) 63% dan 62 %.

Kata Kunci—data mining, association rules, market basket analysis, Apriori, FP-growth.

I. PENDAHULUAN

Sistem komputerisasi dapat membantu pencatatan dan pengolahan data transaksi pembelian dan penjualan barang di toko swalayan atau supermarket. Penggunaan sistem komputerisasi tersebut umumnya terbatas untuk pengelolaan barang. Swalayan KSU Sumber Makmur – Trenggalek telah menggunakan sistem komputerisasi untuk mencatat transaksi penjualan tetapi belum dimanfaatkan untuk mendapatkan *knowledge* untuk meningkatkan transaksi penjualan.

Market Basket Analysis merupakan sebuah analisis terhadap kebiasaan customer berbelanja pada supermarket dengan cara menemukan asosiasi dan korelasi di antara berbagai macam item yang dimasukkan customer di dalam keranjang belanjanya. Market Basket Analysis paling banyak digunakan untuk menganalisa data dalam rangka keperluan strategi pemasaran, desain katalog, dan proses pembuatan keputusan bisnis.

Ramadhan, Departemen Teknologi Informasi, Institut Sains dan Teknologi Terpadu Surabaya, Surabaya, Jawa Timur, Indonesia (e-mail: radokdr@gmail.com)

Esther Irawati Setiawan, Departemen Teknologi Informasi, Institut Sains dan Teknologi Terpadu Surabaya, Surabaya, Jawa Timur, Indonesia (e-mail: esther@stts.edu)

Produk-produk yang sering dibeli secara bersamaan dapat ditempatkan secara berdekatan sehingga konsumen dapat dengan mudah menemukan apa yang dicari. Dengan demikian para konsumen akan merasa puas dan penjualan juga akan meningkat. *Market Basket Analysis* bertujuan untuk mengetahui *item* apa saja yang sering dibeli bersamaan oleh customer. *Item* di sini diartikan sebagai berbagai macam produk atau barang pada supermarket.

Banyak algoritma yang bisa digunakan untuk melakukan *mining frequent itemset* untuk mengetahui hubungan antar *item* dalam transaksi penjualan. Apriori adalah algoritma pertama dan menjadi dasar dalam pencarian *frequent itemset* yang diusulkan oleh R. Agrawal dan R. Srikant di tahun 1994 [1]. FP Growth merupakan algoritma yang saat ini dikenal cepat dan umum digunakan tanpa pembangkitan *candidate generation*. FP Growth meringkas informasi transaksi ke dalam bentuk pohon.

II. TINJAUAN PUSTAKA

A. Market Basket Analysis

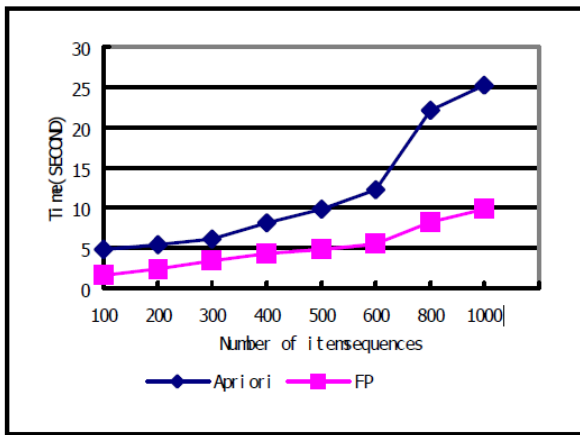
Pada paper [3] disebutkan bahwa *market basket analysis* atau analisis keranjang belanja memiliki tujuan untuk mengidentifikasi produk, atau kelompok produk, yang cenderung terjadi bersamaan (terkait) dalam transaksi pembelian (keranjang). Pengetahuan yang diperoleh dari analisis keranjang belanja bisa sangat berharga; misalnya, dapat digunakan oleh supermarket untuk mengatur ulang tata letaknya, mengambil produk yang sering dijual bersama dan menempatkannya berdekatan. Dapat juga digunakan untuk meningkatkan efisiensi promosi: produk yang terkait seharusnya tidak dipromosikan pada saat yang sama. Dengan mempromosikan secara tepat salah satu produk terkait memungkinkan meningkatkan penjualan produk itu dan mendapatkan penjualan yang menyertainya peningkatan untuk produk terkait. Basis data Untuk analisis keranjang belanja terdiri dari semua transaksi yang dilakukan dalam periode penjualan tertentu (misalnya satu tahun) dan di lokasi penjualan tertentu (misalnya di suatu lokasi supermarket). Konsumen bisa tampil lebih dari sekali di database. Bahkan, konsumen akan muncul didatabase setiap kali mereka melakukan transaksi di lokasi penjualan. Tujuan dari analisis ini adalah untuk menemukan yang terbaik kombinasi produk yang sering dibeli oleh pelanggan.

Penulis melakukan penelitian pada implementasi metode data mining untuk Six Sigma metodologi. Kemudian dipilih

market basket analysis untuk meningkatkan fase, dengan itu untuk memprediksi kebiasaan pelanggan.

B. Analisis Asosiasi

Pada paper [2] penulis melakukan penelitian tentang seberapa efektif metode FP Growth. Ekperimen dilakukan dengan 2 jenis data yang memiliki 10 ribu dan 100 ribu transaksi. Eksperimen dilakukan menggunakan algoritma Apriori dan FP Growth. Perbandingan waktu proses ditunjukkan oleh Gambar. 1.



Gambar. 1. Perbandingan waktu proses Apriori dan FP Growth

Analisis asosiasi atau *association rules mining* adalah teknik data mining untuk menemukan aturan asosiasi antara suatu kombinasi *item*. *Interestingness measure* yang dapat digunakan dalam data mining adalah :

- a. Support, adalah suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu *item* atau *itemset* dari keseluruhan transaksi. Nilai support sebuah *item* diperoleh dengan menggunakan rumus berikut:

$$Support(A) = \frac{Jumlah\ transaksi\ mengandung\ A}{Total\ Transaksi}$$

Nilai support dari 2 *item* diperoleh dengan menggunakan rumus :

$$Support(A, B) = \frac{\sum Transaksi\ mengandung\ A\ dan\ B}{\sum Transaksi}$$

- b. Confidence, adalah suatu ukuran yang menunjukkan hubungan antar dua *item* secara conditional (berdasarkan suatu kondisi tertentu). Confidence dihitung dengan rumus :

$$P(B|A) = \frac{\sum Transaksi\ mengandung\ A\ dan\ B}{\sum Transaksi\ Mengandung\ A}$$

C. Algoritma Apriori

Pada paper [1] penulis memperkenalkan algoritma yang diberi nama Apriori. Langkah pertama dari algoritma Apriori menghitung kemunculan *item* untuk menentukan suatu *itemsets* besar. Langkah berikutnya adalah mencari *k*, yang terdiri dari dua fase. Pertama, *itemsets* besar *L_{k-1}* ditemukan di jalur (*k-1*) digunakan untuk menghasilkan kandidat *itemsets* *c_k*, menggunakan fungsi *apriorigen*. Selanjutnya, database dipindai dan dukungan kandidat di *c_k* dihitung. Agar lebih cepat perlu langkah efisien untuk menentukan kandidat di *c_k* yang terdapat di transaksi yang diberikan *t*.

Bentuk algoritma Apriori dapat dituliskan seperti nampak di Gambar. 2.

```

L1 = {frequent itemset with one element}
for (k=2; Lk-1 ≠ ∅; k++)
{
  Ck = apriori-gen(Lk-1); //pembuatan
                             kandidat
                             //baru
  for all transactions t
  {
    C't = subset(Ck, t); //kandidat yang
                          //tampil pada t
    for all candidates c ∈ C't, do
      c.count ++;
  }
  Lk = {c ∈ C't | c.count ≥ minsup}
}
return ∪k Lk;

```

Gambar. 2. Algoritma Apriori

dimana :

- L : himpunan frequent itemset
- minsup : minimum support
- C : himpunan kandidat itemset
- c : kandidat itemset
- t : transaksi

D. Algoritma FP Growth

Pada paper [4] penulis melakukan perbandingan kinerja algoritma Apriori dengan FP Growth. Algoritma FP Growth membantu menyelesaikan 2 masalah utama pada algoritma apriori, yaitu : menghasilkan sejumlah besar kumpulan kandidat dan berulang kali memindai database. FP Growth diusulkan untuk untuk memberi solusi atas permasalahan di algoritma Apriori dengan menggunakan pendekatan yang berbeda dari paradigma yang digunakan pada algoritma Apriori, yaitu : strategi *divide-and-conquer*. Kelebihan dari algoritma FP Growth adalah hanya melakukan 2 kali pemindaian ke database dan tidak membentuk kumpulan kandidat yang besar. Pertama kali algoritma FP Growth membentuk freuent

pattern tree atau disebut juga FP-Tree. Proses penyusunan FP-Tree dari mulai representasi awal transaksi untuk menentukan frequent itemset, setelah didapatkan frequent itemset kemudian dilanjutkan ke perhitungan *association rule*, menghitung *rule* yang dihasilkan dari penerapan metode FP Growth untuk mendapatkan pola terbaik dari pola-pola yang ada. Hasil akhir pola dari perhitungan *association rule* merupakan hasil rekomendasi sebagai bahan acuan terbaik dalam menentukan item atau produk yang paling sering dibeli oleh konsumen.

Algoritma FP Growth ditunjukkan oleh Gambar. 3.

Algorithm FP-growth

Input: $\mathcal{D}, \sigma, I \subseteq \mathcal{I}$

Output: $\mathcal{F}[I](\mathcal{D}, \sigma)$

```

 $\mathcal{F}[I] := \{\}$ 
for all  $i \in \mathcal{I}$  occurring in  $\mathcal{D}$  do
   $\mathcal{F}[I] := \mathcal{F}[I] \cup \{I \cup \{i\}\}$ 
   $H := \{\}$ ;  $\mathcal{D}^i := \{\}$ 
  for all  $j \in \mathcal{I}$  occurring in  $\mathcal{D}$  such
  that  $j > i$  do
    if  $supp_{\mathcal{D}}(I \cup \{i, j\}) \geq \sigma$  then
       $H := H \cup \{j\}$ 
  for all  $(tid, X) \in \mathcal{D}$  with  $i \in X$  do
     $\mathcal{D}^i := \mathcal{D}^i \cup \{(tid, X \cap H)\}$ 
  Compute  $\mathcal{F}[I \cup \{i\}](\mathcal{D}^i, \sigma)$ 
   $\mathcal{F}[I] := \mathcal{F}[I] \cup \mathcal{F}[I \cup \{i\}]$ 

```

Gambar. 3. Algoritma FP Growth

Metodologi algoritma FP-Growth bisa diperinci sebagai berikut :

1. Menentukan nilai *minimum support*, misalnya ditentukan 40%. Dengan asumsi jumlah transaksi 9 maka nilai *minimum support count* adalah $40/100 \cdot 9 = 3,6$ (dibulatkan ke atas jadi 4).
2. Berikutnya adalah menghitung jumlah kemunculan masing-masing item (*support count*).
3. Selanjutnya cari *item* yang memenuhi *minimum support count*, urutkan *support count* secara *descending* (besar ke kecil) dan *itemset* secara *ascending* (kecil ke besar).
4. Daftar *itemset* dan *support count* itu bisa disebut juga sebagai *header table*, selanjutnya dibuat database transaksi baru yang mana dilakukan pengurutan *list of item* sesuai urutan *support count* sekaligus membuang *item* yang tidak ada di dalam *header table* tersebut.
5. Dengan menggunakan database transaksi baru dilakukan pembuatan FP-Tree, dimulai dari transaksi ke-1, 2, dan seterusnya.
6. Dengan menggunakan FP-Tree yang telah dibangun di tahap sebelumnya, dilakukan pembangkitan *conditional pattern base*.
7. Hitung *support count* dari setiap *item* pada setiap *conditional pattern base*, untuk masing-masing *item* yang memiliki jumlah *support count* lebih besar atau sama dengan *minimum support count* akan dibangkitkan menjadi *conditional FP-Tree*.
8. Berikutnya adalah pencarian *Frequent Itemset*,

apabila *conditional FP-Tree* merupakan lintasan tunggal (*single path*), maka didapatkan *frequent itemset* dengan melakukan kombinasi *item* untuk setiap *conditional FP-Tree*. Jika bukan lintasan tunggal, maka dilakukan pembangkitan FP-Growth secara *rekursif* (proses memanggil dirinya sendiri).

9. Tahap selanjutnya adalah pencarian *strong association rule*. *Strong association rule* adalah *rule* yang memenuhi syarat *minimum support* dan *minimum confidence*. Tahap pencarian *strong association rule* dilakukan dengan langkah :

- a. Untuk setiap *frequent itemset* l dari , dibangkitkan semua *non-empty subset* dari l .
- b. Untuk setiap *non-empty subset* s dari l , menghasilkan *rule* :

$$s \Rightarrow (l-s) \text{ jika } \frac{\text{support_count}(l)}{\text{support_count}(s)} \geq \text{min_conf}$$

dimana *min_conf* adalah *minimum confidence*.

Untuk analisis digunakan dataset *french retail dan bakery shop*. Minimum support ditentukan sebesar 1% dan minimum confidence adalah 50%. Komputer yang digunakan untuk ujicoba memiliki spesifikasi : processor Inter® Core™ i5-4210U, memori utama 4GB, dan operating system Microsoft Windows 10.

Kesimpulan dari penelitian pada paper tersebut menyatakan bahwa FP Growth memerlukan waktu lebih cepat pada segala pengujian, baik saat produk dikurangi maupun tidak.

E. Hybrid-Dimension Association Rules

Pada paper [5] penulis menerapkan konsep multi-dimensional yang mana digunakan 2 faktor atau variable untuk keputusan ini, yaitu : barang yang dibeli bersamaan dan waktu transaksi. Terdapat beberapa langkah untuk melakukan data mining, dimulai dari pre-prosesing, menentukan *frequent itemsets*, menentukan *association rules*.

Percobaan dilakukan menggunakan data transaksi dari minimarket X yang ingin melakukan analisis kebiasaan berbelanja pelanggan.

Hasil dari percobaan menunjukkan suatu *rules* bahwa pelanggan berbelanja di hari Minggu untuk membeli *Aqua Isi Ulang* dan *Club Isi Ulang*, yang mana hal ini memenuhi kriteria *Hybrid-Dimension Association Rules*.

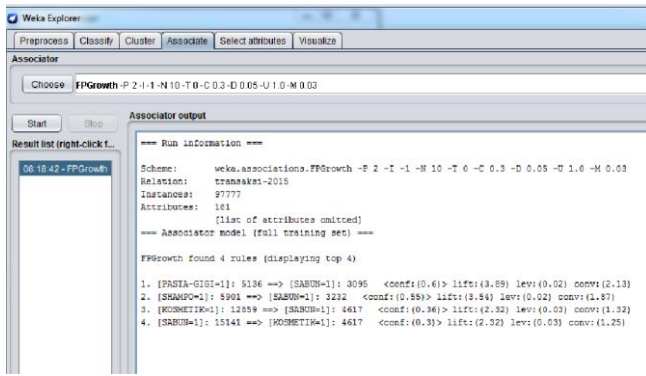
F. Perbedaan Algoritma Apriori dan FP-Growth

Secara umum bisa digambarkan bahwa ada beberapa hal yang membedakan antara algoritma Apriori dan algoritma FP-Growth, berikut ini adalah ringkasannya :

Kekurangan algoritma Apriori :

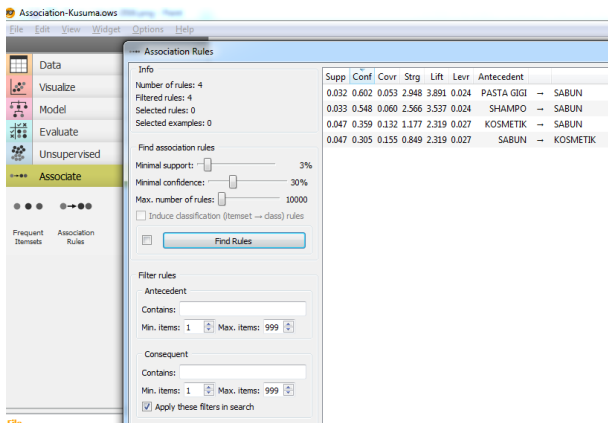
- Membutuhkan ruang penyimpanan besar
- Membutuhkan waktu proses lebih lama
- Menggunakan memori lebih banyak
- Kurang akurat
- Melakukan pembacaan ke database berulang kali

confidence dengan menggunakan data tahun 2015-2019 yang memiliki 218 *attributs* dan 612.439 *instances* ditunjukkan oleh Tabel 1.



Gambar. 5. Contoh *Associator output* aplikasi Weka

Pada aplikasi orange bisa dipilih nilai minimum support dan minimum confidence yang diinginkan di bagian *slider* minimal support dan minimal confidence. Dari pengolahan data dengan aplikasi Orange didapatkan tampilan seperti pada Gambar. 6.



Gambar. 6. Contoh tampilan *Association rules* aplikasi Orange

Agar mendapatkan nilai minimum support yang tepat maka dilakukan percobaan dengan nilai tertentu misalnya 30%. Dengan nilai minimum support 30% ternyata belum menghasilkan *rules* karena variasi barang yang diolah cukup banyak yakni mencapai 218 jenis barang untuk seluruh data antara tahun 2015-2019 dan jumlah *instance/* transaksi adalah 612.349. Dengan dataset semacam itu maka nilai minimum support yang bisa menghasilkan *rules* dan tidak terlalu banyak *rules* yang dihasilkan adalah 0.03 atau 3%.

Nilai minimum confidence juga dilakukan percobaan dengan menggunakan nilai yang cukup rendah dulu, misalnya : 50% atau 0.5 di aplikasi Weka. Dengan kombinasi nilai minimum support 30% dan nilai minimum confidence 50% ternyata hanya didapatkan 2 *rules* sehingga dicoba lagi dengan menurunkan nilai minimum confidence ke 20%, hasilnya didapatkan 12 *rules*. Karena *rules* yang muncul sudah terlalu banyak dan yang terendah bernilai confidence 0.21 atau 21% maka ditentukan menggunakan nilai confidence 30% yang menghasilkan 4 *rules*.

Hasil uji coba berbagai macam nilai *minimum support* dan

TABEL 1
HASIL UJI COBA BERBAGAI MACAM NILAI *MINIMUM SUPPORT* DAN *CONFIDENCE*

Minimum Support (%)	MINIMUM CONFIDENCE (%)	JUMLAH ASSOCIATI ON RULE	NILAI CONFIDENCE CE TERTINGGI
10	90	0	0
5	90	0	0
1	80	0	0
1	70	2	80
1	60	7	80
1	50	8	80
1	40	20	80
1	30	43	80
1	20	102	80
1	10	166	80
2	10	42	62
3	10	12	63
4	10	4	62
5	10	2	37
6	10	0	0
5	20	2	37
4	30	3	62
3	30	4	63

Perbandingan hasil association rule dari kedua aplikasi tersebut nampak seperti pada Tabel 2. Dari Tabel 2. bisa diketahui jika *rule* terbaik atau menarik adalah :

$$\text{beli PASTA GIGI} \Rightarrow \text{beli SABUN} [\text{support} = 3,7\%, \text{confidence} = 63\%].$$

Itu berarti pelanggan yang membeli pasta gigi dan sabun secara bersamaan terjadi pada 3,7% dari keseluruhan transaksi dan tingkat keyakinan untuk dibeli bersama-sama sebesar 63%. Kenapa dikatakan menarik karena nilai support dan confidence cukup tinggi.

V. KESIMPULAN

Nilai minimum support yang tepat pada penelitian dengan menggunakan data transaksi penjualan di Swalayan KSU Sumber Makmur adalah 0.03 atau 3%. Sedangkan nilai minimum confidence di angka 0.3 atau 30%.

Dari hasil pengolahan data didapatkan 2 *strong association* dengan nilai confidence paling tinggi, yaitu: jika membeli pasta gigi maka ada kemungkinan sebesar 63% akan membeli sabun dan jika membeli shampo maka ada kemungkinan 62% akan membeli sabun.

Pencarian *strong association* menggunakan aplikasi Weka versi 3.8.4 dan Orange versi 3.24.1 menghasilkan *output* yang sama meskipun ada perbedaan cara menampilkan nilai *confidence* yang mana Orange dengan 1 desimal sedangkan Weka tanpa desimal.

TABEL 2

PERBANDINGAN ASSOCIATION RULES HASIL DARI WEKA DAN ORANGE

TAHUN	ANTECEDENT	CONSEQUENT	WEKA		ORANGE
			SUPPORT	CONFIDENCE	CONFIDENCE
2015	PASTA GIGI	SABUN	3,2	60,2	60
	SHAMPO	SABUN	3,3	54,8	55
	KOSMETIK	SABUN	4,7	35,9	36
	SABUN	KOSMETIK	4,7	30,5	30
2016	PASTA GIGI	SABUN	3,1	63,1	63
	SHAMPO	SABUN	3,3	55,9	56
	KOSMETIK	SABUN	4,6	36,5	37
2017	PASTA GIGI	SABUN	4,0	64,0	64
	SHAMPO	SABUN	4,4	62,5	62
	BABY CARE	SABUN	3,2	40,7	41
	KOSMETIK	SABUN	5,1	38,4	38
2018	SHAMPO	SABUN	4,8	66,1	66
	PASTA GIGI	SABUN	4,0	64,5	65
	KOSMETIK	SABUN	5,3	38,5	38
	SABUN	KOSMETIK	5,3	30,1	30
2019	SHAMPO	SABUN	4,7	64,7	65
	PASTA GIGI	SABUN	3,8	62,4	62
	KOSMETIK	SABUN	5,5	36,7	37
	SABUN	KOSMETIK	5,5	32,3	32
	SUSU	SNACK	4,2	30,4	30
2015-2019	PASTA GIGI	SABUN	3,7	63,0	63
	SHAMPO	SABUN	4,2	61,9	62
	KOSMETIK	SABUN	5,1	37,3	37
	SABUN	KOSMETIK	5,1	30,2	30

DAFTAR PUSTAKA

- [1] R. Agrawal, R. Srikant, and others, "Fast algorithms for mining association rules," in *Proc. 20th int. conf. very large data bases, VLDB*, 1994, vol. 1215, pp. 487-499.
- [2] W. Zhang, H. Liao, and N. Zhao, "Research on the FP growth algorithm about association rule mining," in *2008 International Seminar on Business and Information Management*, 2008, vol. 1, pp. 315-318.
- [3] A. Trnka, "Market basket analysis with data mining methods," in *2010 International Conference on Networking and Information Technology*, 2010, pp. 446-450.
- [4] M. Hossain, A. H. M. S. Sattar, and M. K. Paul, "Market Basket Analysis Using Apriori and FP Growth Algorithm," in *2019 22nd International Conference on Computer and Information Technology (ICCIT)*, 2019, pp. 1-6.
- [5] D. H. Setiabudi, G. S. Budhi, I. W. J. Purnama, and A. Noertjahyana, "Data mining market basket analysis' using hybrid-dimension association rules, case study in Minimarket X," in *2011 International Conference on Uncertainty Reasoning and Knowledge Engineering*, 2011, vol. 1, pp. 196-199.
- [6] J. Han, M. Kamber, and J. Pei, "Data mining concepts and techniques third edition," *Morgan Kaufmann Ser. Data Manag. Syst.*, vol. 5, no. 4, pp. 83-124, 2011.

Pemanfaatan Behavior Tree dan Fuzzy Waypoint Tactic pada Game Strategi "War of Zombies"

Michael Patria Christie, *Departemen Informatika, Institut Sains dan Teknologi Terpadu Surabaya,*
Andreas, *Fakultas Ekonomi dan Bisnis, Universitas Pelita Harapan*

Abstrak— Dalam video game jaman sekarang, hampir semua permainan memakai kecerdasan buatan sebagai cara agar non-player karakter agar dapat berinteraksi dengan game. Karena itu pada penelitian ini bertujuan untuk mencoba menerapkan kecerdasan buatan behavior tree dan juga Fuzzy waypoint tactic. Project game "War of Zombies" ini merupakan permainan strategi dimana pemain yang berperan sebagai zombie mencoba melenyapkan manusia sebagai musuhnya. Pada permainan ini selain waypoint tactic dan behavior tree terdapat beberapa metode kecerdasan buatan yang diterapkan seperti A* pathfinding, djikstra dan juga fuzzy logic. Behavior tree merupakan metode panduan berisi langkah-langkah bagi karakter untuk menyelesaikan suatu tugas. Sedangkan fuzzy waypoint tactic merupakan cara bagi developer untuk memberitahu lokasi-lokasi strategis yang dapat membantu karakter dalam menyelesaikan tugas, seperti misalnya bersembunyi atau menembak. Pada game bertema strategi seperti ini, fuzzy logic sangat membantu untuk membuat suatu keputusan bagi karakter. Selain itu, fuzzy logic juga merupakan metode yang baik dalam waypoint tactic untuk menentukan lokasi strategis yang paling menguntungkan diantara pilihan-pilihan lokasi yang telah disediakan. Behavior tree juga merupakan metode yang baik dalam mengatur perilaku setiap karakter yang ada pada game strategi.

Kata Kunci— Artificial Intelligent, Behavior Tree, Djikstra, War of Zombies, Waypoint Tactics.

I. PENDAHULUAN

PERKEMBANGAN teknologi yang sangat pesat akhir-akhir ini membawa banyak perubahan dalam perkembangan perangkat lunak, salah satu jenis perangkat lunak yang sangat populer dewasa ini adalah game, game merupakan perangkat lunak yang digunakan sebagai sarana hiburan oleh siapapun tanpa memandang status, usia maupun gender. Salah satu fitur yang umum dan dapat ditemui pada hampir semua game yang ada pada zaman sekarang adalah Artificial Intelligence atau AI, AI adalah kepintaran buatan yang dipasang pada sistem komputer sehingga membuat player dapat bermain melawan komputer seolah-olah sedang melawan player lain.

Michael Patria Christie, Department of Informatics, Institut Sains dan Terapan teknologi Terpadu Surabaya, Surabaya, Jawa Timur, Indonesia (e-mail: mpatriac@gmail.com)

Andreas, Department of Management, Faculty of Economy and Business, Universitas Pelita Harapan, Surabaya, Jawa Timur, Indonesia (e-mail: andreas.jodhinata@gmail.com)

AI merupakan system yang menentukan bagaimana komputer tersebut berpikir sehingga hal ini sangat menarik jika diterapkan pada karakter musuh maupun karakter pembantu agar karakter tersebut dapat memilih dan menyesuaikan tindakan apa yang tepat pada situasi yang terjadi dalam game dan tidak terlihat berperilaku secara asal-asalan.

Semisalnya saja, kapan karakter musuh mulai menyerang dan juga bagaimana cara menyerang pemain? Apakah ketika karakter musuh berada di belakang pemain atau hanya ketika pemain tersebut menyerang pemain lain? Kondisi semacam itulah yang wajib ditentukan oleh developer dalam menentukan sebuah AI yang baik dan bagus. Apabila kualitas AI pada karakter kurang baik maka kualitas pada game tersebut akan turun. Karena itulah AI pada game juga dapat menentukan kualitas pada game.

Penelitian ini merupakan penerapan dua jenis Artificial Intelligence, yaitu Behavior Tree dan Fuzzy Waypoint Tactics, pada karakter musuh dan karakter pembantu milik musuh dalam sebuah game bernama War of Zombies. Sehingga musuh dapat memilih dan menyesuaikan tindakan-tindakan yang diambil berdasarkan kondisi yang ada. Mulai dari menentukan jenis serangan musuh terhadap karakter player berdasarkan lingkup jarak karakter, dimana posisi yang baik untuk bersembunyi, sampai pada menentukan kapan dan bagaimana karakter pembantu melakukan tugasnya atau kabur.

II. TEORI PENUNJANG

A. Artificial Intelligence

Artificial Intelligence atau Kecerdasan Buatan [1] adalah kecerdasan yang dimiliki oleh mesin yang dapat bertindak seperti layaknya manusia. Pada game, AI bertindak dalam mengendalikan karakter musuh dan karakter pembantu dari musuh agar dapat bermain melawan player.

B. Behavior Tree

Behavior tree [2]–[5] adalah suatu model matematis berbentuk tree yang dibuat untuk merepresentasikan perilaku dari suatu entitas yang diprogram. Behavior tree adalah sesuatu yang sangat fundamental dalam pengembangan game. Behavior tree mirip dengan finite state machine, hanya saja tidak seperti finite state machine yang berisi keadaan-keadaan, behavior tree berisi kumpulan instruksi dan perintah yang harus dijalankan oleh developer karena

behavior tree mudah dimengerti oleh manusia sehingga tingkat kesalahan yang dibuat menjadi minimum.

C. Waypoint Tactics

Pada pathfinding, waypoint [6] biasanya lebih dikenal dengan sebutan representative point atau node. Developer akan menaruh waypoint sebagai cara untuk memberitahu karakter area yang bisa dilewati. Lalu untuk mengetahui rute terdekat maka tiap koneksi antar waypoint akan diberi bobot seperti jarak atau waktu, dan developer akan menggunakan algoritma seperti djikstra atau A* untuk mencari rute terdekat [7].

D. Panda BT

Panda BT merupakan scripting framework behavior tree untuk unity. Framework ini memungkinkan untuk menentukan logika yang kompleks, skalabel dan dapat digunakan kembali dalam pembuatan game. Panda behavior diaplikasikan sebagai komponen pada gameobject. Komponen ini dapat dipasang pada segala gameobject untuk membuat behavior yang terdiri dari kombinasi C# dan BT scripts yang dapat ditulis dengan menggunakan bahasa bawaan yang minimalist.

III. DESAIN GAME

A. Informasi Umum

Game war of zombies ini merupakan game dengan genre strategi seperti age of empire atau total war three kingdoms. Game ini adalah permainan single player dimana player akan melawan karakter dengan kecerdasan buatan. Game ini berbasis platform desktop windows. Konsep desain pada game ini adalah menggunakan konsep gambar dua dimensi. Game akan dibuat dengan menggunakan unity sebagai physic game engine dan juga visual studio sebagai code editor. Target usia untuk game ini adalah remaja berumur 13 tahun sampai 20 tahun.

Pada game ini terdapat 2 jenis karakter yaitu karakter zombie dan manusia. Setiap jenis pada karakter zombie dan manusia memiliki tipe yang berbeda beda. Yang akan dimainkan oleh player adalah karakter zombie. Dan yang akan menjadi musuh adalah karakter manusia. Berikut penjelasan setiap karakter yang dipakai pada game war of zombies.

- Karakter utama
Player adalah zombie pertama yang memiliki kesadaran dan kecerdasan. Player merupakan zombie yang berhasil menularkan kesadaran dan kecerdasan tersebut ke zombie yang lain. Karena itu, player diberikan peringkat teratas oleh zombie-zombie lain sebagai zombie leader. Sebagai zombie leader, player dapat membuat strategi dan menerapkannya dalam game. User dapat membuat unit dan menentukan zombie yang berada pada unit tersebut dan memberikan perintah melalui ghost general yang berfungsi sebagai pemimpin dalam setiap unit.

- Ghost General
Ghost general merupakan pemimpin sebuah unit. Player dapat memberikan perintah ke setiap unit melalui masing-masing ghost general pada setiap unit. Setiap unit akan terdapat 1 ghost general. Ketika user menambahkan unit, ghost general akan otomatis ditambahkan dalam unit.
Ghost general merupakan sarana bagi player untuk dapat memberikan perintah kepada setiap team. Ghost general hanya dapat berjalan dan dapat menyerang. Namun walaupun tidak dapat menyerang, ghost general tidak dapat diserang oleh musuh menggunakan apapun. Untuk mengalahkan ghost general, seluruh zombie pada team yang dipimpin oleh ghost general tersebut harus dikalahkan semua. Barulah ghost general akan mati.
- Iron Zombie
Iron zombie adalah zombie yang memiliki badan seperti slime karena karakteristik badannya. Iron zombie memiliki karakteristik seperti magnet. Sehingga badannya yang tertutupi oleh pasir besi, sehingga membuat bentuknya seperti slime. Karena badannya yang tertutupi oleh besi, iron zombie memiliki darah dan pertahanan yang cukup besar sehingga sangat cocok untuk melindungi zombie-zombie lain. Iron zombie juga memiliki kemampuan untuk mengubah bentuk pasir-pasir besi yang mengelilingi tubuhnya sesuai yang dia inginkan.
Iron slime tidak memiliki kemampuan untuk menyerang musuh. Namun sebagai gantinya iron slime memiliki darah yang tebal. Ketika perintah defend diberikan, maka iron zombie akan berubah menjadi seperti dinding. Dengan begitu akan memberikan perlindungan kepada karakter-karakter zombie yang lain dengan lebih efektif.
- Suicide zombie
Suicide Zombie adalah zombie yang berasal dari mayat teroris. Kegemarannya dalam meledekan bomb terbawa hingga mati. Badannya dapat mengeluarkan sebuah cairan enzim. Cairan tersebut merupakan gabungan dari zat-zat kimia yang cukup mudah meledak. Tubuhnya yang dapat dengan mudah terbakar dapat membuat ledakan cukup besar di area tersebut, dan memberikan damage yang cukup besar.
Apabila ghost colonel memberikan perintah untuk menyerang, maka suicide zombie akan pergi ke area yang ditentukan. Ketika sampai di tempat atau target yang ditentukan, maka akan muncul ledakan dengan jarak yang luas dan damage yang besar. Karena karakteristiknya, suicide zombie hanya dapat melakukan serangan hanya satu kali. Setelah melakukan serangan, suicide zombie akan langsung meninggal karena satu-satunya tipe serangan yang bisa dilakukan oleh suicide zombie adalah bom bunuh diri.
- Zombie gunner
Zombie gunner adalah zombie yang terlahir dari prajurit yang sudah meninggal dan dipersenjatai dengan pistol. Senjata yang biasa digunakan adalah senjata riddle buatan Russia yaitu AK-47. Karena ketika masih hidup mereka adalah prajurit, maka mereka memiliki kemampuan yang setara dengan prajurit manusia.

Mereka memiliki kemampuan menembak yang sangat baik sehingga dapat menjadi ancaman bagi manusia.

Ketika ghost general memberikan perintah untuk menyerang, maka zombie gunner akan mengeluarkan riffle AK-47 dan mulai menembak target yang telah ditentukan. Ketika musuh sudah meninggal, maka zombie gunner akan berdiam diri sampai ada perintah selanjutnya. Namun apabila ghost general memberikan perintah massacre kepada zombie gunner, maka zombie gunner akan pergi menuju mendekati target lalu mulai menyerang hingga target meninggal dan akan mencari target lainnya secara otomatis hingga perintah untuk regroup diberikan.

- Blacksmith

Blacksmith adalah karakter manusia yang berfungsi untuk membetulkan sesuatu. Seperti misalnya sandbag, atau tembok yang berfungsi untuk melindungi manusia lain dari serangan zombie. Sayangnya, karena mereka tidak ahli dalam menembak, mereka tidak diberikan senjata sama sekali. Namun dengan tidak diberikan senjata, mereka dapat membawa muatan bahan baku lebih banyak untuk membetulkan tembok atau pintu.

Dalam game, blacksmith akan menuju ke tempat dimana tembok atau sandbag yang terkena damage. Ketika sudah di lokasi, blacksmith akan membetulkan tembok dan sandbag yang terkena serangan. Sandbag dan tembok yang hancur tidak dapat dibetulkan sama sekali. Ketika tidak ada lagi yang dibetulkan, maka blacksmith akan pergi ke tempat bersembunyi untuk menghindari serangan musuh. Ketika markas sudah ditembus oleh zombie, blacksmith akan lebih focus ke melarikan diri daripada melakukan tugasnya untuk membetulkan dinding atau pintu markas.

- Doctor

Doctor memiliki tugas untuk menyembuhkan orang yang terluka. Doctor akan pergi kesana kemari untuk menyembuhkan setiap karakter yang terluka. Karena lebih ahli dalam bidang kesehatan, doctor tidak memiliki pengetahuan mengenai cara menembak dan lain-lain. Karena itu doctor tidak memiliki kemampuan untuk menyerang.

Setiap area biasanya mempunyai 1-2 dokter. Dalam game, kemampuan doctor mirip dengan blacksmith. Doctor akan pergi menuju lokasi soldier yang terluka dan mulai menyembuhkan pasien. Setelah pasien sembuh, doctor akan pindah lokasi dan menyembuhkan pasien lainnya. Dokter tidak bisa menyembuhkan doctor lain, yang difokuskan untuk disembuhkan oleh doctor adalah soldier karena soldier memiliki tugas untuk menyerang zombie. Doctor tidak memiliki kemampuan untuk menyerang zombie. Ketika markas berhasil ditembus zombie, maka doctor akan pergi melarikan diri dari kejaran zombie. Ketika tidak sibuk menyembuhkan soldier, doctor akan pergi untuk bersembunyi.

- Soldier

Soldier merupakan prajurit yang telah dipersenjatai. Soldier memiliki kemampuan yang mirip dengan zombie gunner. Hanya saja soldier juga memiliki senjata lain yaitu pisau. Sehingga membuat karakter ini tidak hanya baik dalam serangan dekat, namun juga serangan jauh. Dalam game, soldier akan pergi ke lokasi yang dirasa cocok untuk menembak dan juga bertahan.

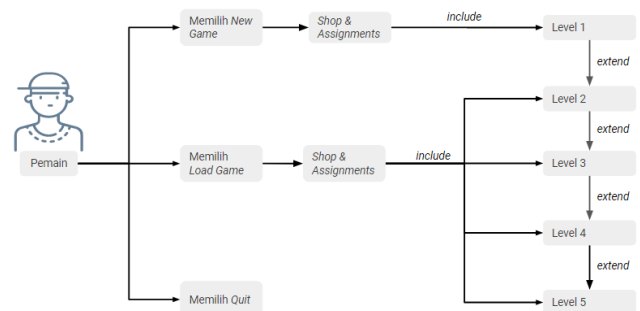
Soldier akan menuju lokasi yang menguntungkan dimana terdapat cover dan penglihatan yang bagus untuk menembak. Ketika terdapat zombie yang mendekat, maka soldier akan mengeluarkan pisanya untuk menyerang zombie secara dekat.

Pada game ini, soldier terbagi menjadi dua team yaitu, defend team dan mobile team. Defend team memiliki tugas untuk melindungi area yang ditugaskan. Sedangkan mobile team merupakan team yang berfungsi untuk membantu area lain menyerang zombie. Hampir di setiap stage, akan terdapat soldier yang ditugaskan sebagai mobile team. Tugas dari mobile team adalah sebagai bala bantuan untuk team-team lain yang kesusahan dalam menangkis serangan zombie. Tidak seperti mobile team, defend team tidak dapat pergi ke lokasi yang lain selain yang ditentukan.

B. Gameplay

Seperti yang diberitahukan di informasi umum, game war of zombies merupakan game dua dimensi dan memiliki genre strategi. Game ini memaki top up view dimana player melihat situasi lapangan dari atas. Konsep ini membantu agar player dapat melihat kondisi lapangan secara mudah. Berikut akan dijelaskan mengenai konsep pada game ini.

Ketika pemain memulai game ini, yang pertama kali ditampilkan adalah menu utama. Pada menu utama terdapat New Game, Load, Setting, Achievement, Exit. Ketika player memilih New Game atau Load, maka yang ditampilkan di layar pertama kali adalah Shop and Assignment. Pada tampilan tersebut, player akan diarahkan untuk mengatur timnya terlebih dahulu. Setelah tim sudah diatur, maka tampilan akan berubah sesuai dengan menu yang dipilih sebelumnya. Apabila, player memilih New Game, maka tampilan akan dilanjutkan ke level 1. Namun, jika player memilih Load, maka tampilan akan dilanjutkan ke level terakhir yang dimainkan oleh player.



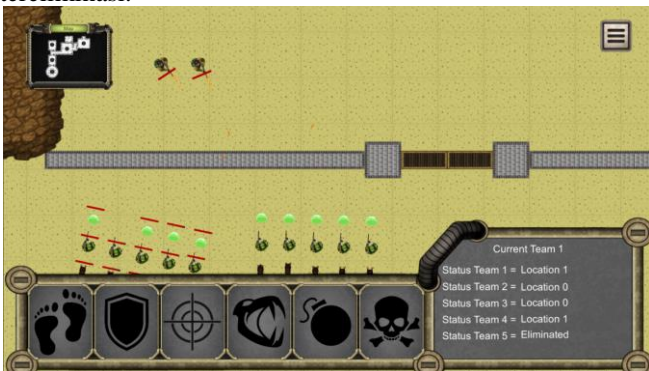
Gambar 1. Use Case Diagram Game

Pada Shop and Assignment terdapat 4 menu yaitu Assignment, Missions, Shop, dan Start. Pada menu assignment, player dapat mengatur timnya. Pada menu mission, player dapat melihat misi di stage tersebut. Player juga dapat mengetahui jumlah zombie dalam satu tim untuk tiap jenisnya, lalu jumlah timnya, dan juga jumlah maksimum zombie yang dipakai dalam keseluruhan tim. Untuk menyelesaikan stage, player wajib menyelesaikan misi yang diberikan tersebut. Pada menu shop, player dapat membeli zombie dan juga items. Player dapat memilih start untuk memulai game.



Gambar 2. Design Assignment

Pada saat memulai game, pada tampilan terdapat map, log, dan commands. Button commands adalah fitur yang dapat digunakan player untuk memberikan perintah kepada tim zombie. Sedangkan log menampilkan status lokasi dari setiap tim yang digunakan. Ketika salah satu tim gagal dan musnah, maka log akan menampilkan status bahwa tim tersebut telah tereliminasi.



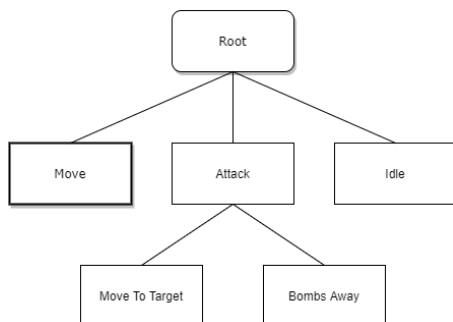
Gambar 3. Design Gameplay

Untuk memberikan perintah pada tim, pertama-tama player memilih ghost general dari tim yang dipilih. Lalu player dapat menginputkan command yang ingin digunakan, yaitu move, shoot, bite, bombs away, massacre, regroup, dan defend.

IV. BEHAVIOR TREE

A. Behavior Tree pada Player

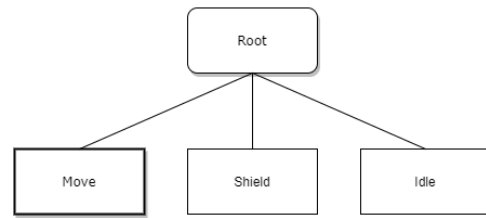
1. Suicide Zombie



Gambar 4. Behavior Tree Suicide Zombie

Node root merupakan composite tipe selector dimana bergantung input dari player. Root akan menuju ke move, attack, atau idle.

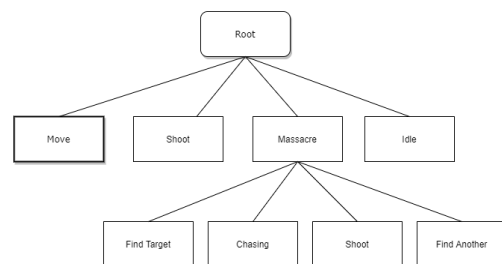
2. Iron Zombie



Gambar 5. Behavior Tree Iron Zombie

Karena karakteristiknya yang cukup sederhana, maka iron zombie hanya memiliki empat node, yaitu Root, Move, Shield, dan Idle.

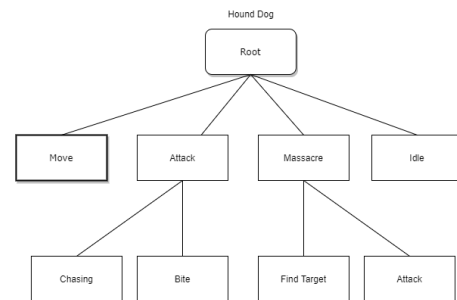
3. Tree Gunner



Gambar 6. Behavior Tree Gunner

Pada gunner, terdapat beberapa node dan node root merupakan tipe selector. Dari node root akan dilanjutkan ke node Move, Shoot, Massacre dan Idle.

4. Hound Dog

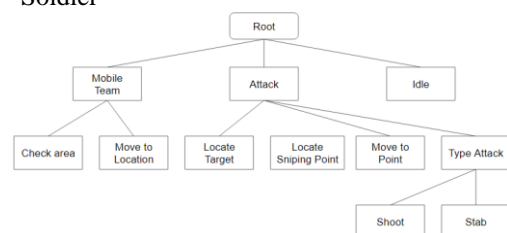


Gambar 7. Behavior Tree Hound Dog

Pada Hound Dog, Node Root bersifat composite tipe selector. Jika tidak ada perintah dari player, maka hound dog akan masuk dalam node idle. Selain node idle, juga ada node move, attack, dan massacre.

B. Behavior Tree pada Enemy

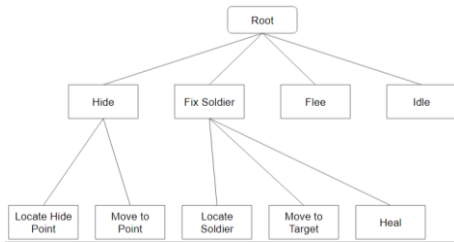
1. Soldier



Gambar 8. Behavior Tree Soldier

Node root dari Behavior Tree Soldier ini merupakan composite tipe selector. Dari node root yang akan diperiksa pertama kali adalah node yang paling kiri, yaitu node mobile team. Jika node mobile attack menghasilkan nilai false, maka akan berlanjut ke node attack. Ketika node mobile team dan node attack mengembalikan value berupa false, maka node idle akan aktif dan selalu mengembalikan nilai true.

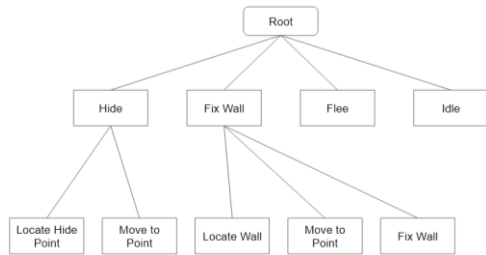
2. Doctor



Gambar 9. Behavior Tree Doctor

Node Root dari Behavior Tree untuk doctor ini bersifat composite tipe Selector.

3. Blacksmith



Gambar 10. Behavior Tree Blacksmith

Node Root dari Behavior Tree untuk Blacksmith ini bersifat composite tipe Selector. Dari node root ke node anaknya, selama markas belum ditembus oleh zombie, maka system akan memeriksa dua hal yaitu, apakah ada tembok atau pintu yang rusak dan apakah ada zombie di sekitar situ. Jika tidak ada tembok atau pintu yang rusak dan tidak ada zombie, maka node idle akan menghasilkan value true sehingga proses idle berlangsung.

V. FUZZY WAYPOINT TACTICS

Waypoint merupakan suatu tanda lokasi yang diinputkan secara manual maupun otomatis oleh developer. Pada pathfinding, waypoint adalah sebuah posisi yang digunakan oleh algoritma pathfinding untuk menghubungkan dua area yang berbeda. Sehingga karakter tidak perlu berjalan melalui tembok atau mengambil rute yang aneh bagi pemain.

Pada waypoint tactics, waypoint digunakan sebagai sarana bagi developer untuk memberitahu posisi-posisi yang strategis bagi karakter untuk dapat menyelesaikan sebuah tugas. Secara teori, waypoint akan diberikan informasi tambahan agar system dapat membuat keputusan yang lebih kompleks dan informative.

Pada game ini terdapat dua jenis waypoint yang akan dipergunakan. Yang pertama adalah hide waypoint yaitu waypoint yang digunakan untuk berlindung dari serangan musuh dan juga sniping waypoint yaitu waypoint untuk

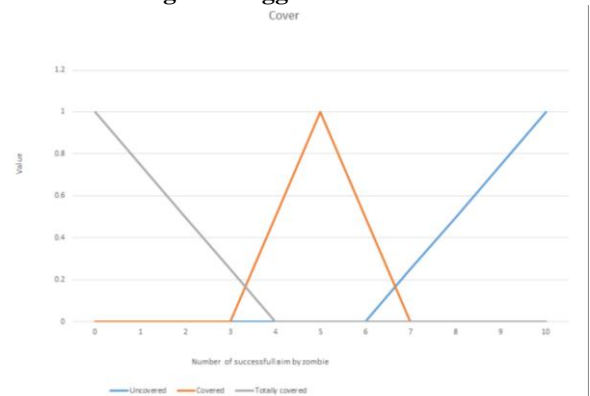
menembak. Untuk mendapatkan hide waypoint, hal yang harus diperhatikan adalah apakah bidikan musuh dapat mengenai karakter. Dan sniping waypoint adalah perlindungan dari bidikan serangan musuh dan juga apakah bidikan sniping waypoint dapat mengenai musuh. Untuk menentukan hasil nilai akhir sniping point, maka system akan menggunakan algoritma fuzzy logic sebagai bantuan pembuat keputusan mengenai sniping waypoint terbaik. Fuzzy Waypoint nantinya akan ditaruh di tempat tempat seperti di balik tembok, pintu atau tempat-tempat lain yang sekiranya dapat memberikan perlindungan yang baik.

Terdapat empat aturan yang akan diterapkan pada fuzzy logic yang akan digunakan yaitu:

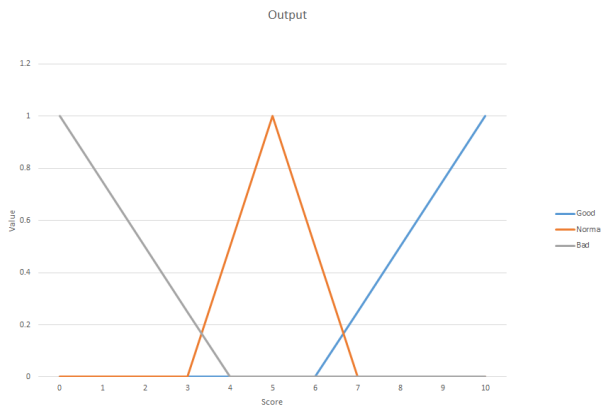
- Jika cover (uncovered) atau aim target (can't get target) maka output bad
- Jika cover (covered atau totally covered) dan aim target (can get target) maka output normal
- Jika cover (covered) dan aim target (target spotted) maka output normal
- Jika cover (totally covered) dan aim target (target spotted) maka output good



Gambar 11. Fungsi Keanggotaan Nilai Kualitas Bidik



Gambar 12. Fungsi Keanggotaan Nilai Kualitas Cover



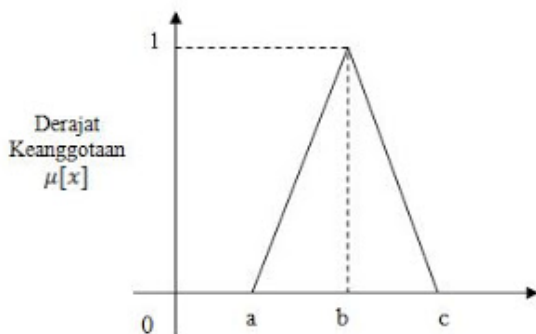
Gambar 13. Grafik Fungsi Keanggotaan Output

$$\mu[x, a, b] = \begin{cases} 0; & x \leq a \\ (x - a) / (b - a); & a \leq x \leq b \\ 1; & x \geq b \end{cases} \quad (1)$$

Formula 1 digunakan apabila bentuk kategorinya berbentuk seperti kategori target spotted pada grafik fungsi keanggotaan aim target. Variable a merupakan nilai kualitas aim target terendah pada kategori target spotted dan variable b adalah nilai kualitas aim tertinggi pada kategori target spotted. Apabila nilai kualitas cover berada diantara a atau b maka menggunakan rumus $(x-a)/(b-a)$ dimana x adalah nilai kualitas cover. Nilai cover akan bernilai 0 apabila sama dengan a dan nilai 1 apabila sama dengan b.

$$\mu[x, a, b] = \begin{cases} (b-x) / (b-a); & a \leq x \leq b \\ 0; & x \geq b \end{cases} \quad (2)$$

Rumus di atas digunakan apabila bentuk kategorinya berbentuk seperti kategori can't get target pada grafik fungsi keanggotaan aim target. b merupakan nilai kualitas aim target terendah pada kategori can't get target dan a adalah nilai kualitas aim tertinggi pada kategori can't get target. Apabila nilai kualitas cover berada pada a atau b maka menggunakan rumus $(b-x)/(b-a)$ dimana x adalah nilai kualitas cover. Nilai cover akan bernilai apabila sama dengan b.



Gambar 16. Representasi Kurva Segitiga

$$triangle(x : a, b, c) = \begin{cases} 0 & x < a \\ (x - a) / (b - a) & a \leq x \leq b \\ (c - x) / (c - b) & b \leq x \leq c \\ 0 & x > c \end{cases} \quad (3)$$

Rumus di atas memiliki bentuk segitiga seperti bentuk garis kategori pada output. Dapat dilihat di gambar 5.4, titik a adalah titik awal kategori dengan nilai 0, lalu b merupakan titik tertinggi dengan nilai 1 dan c adalah titik akhir dengan nilai 0. Apabila nilai kualitas berada pada b dan a maka menggunakan rumus $(x-a)/(b-a)$. Apabila berada di posisi $(c-x)/(c-b)$. X adalah nilai kualitas waypoint. Dan apabila x berada di posisi a atau pada posisi c maka nilai x adalah 0.

VI. KESIMPULAN

Pada penelitian ini telah dikembangkan sebuah game yang memanfaatkan dua buah teknik AI for Games: Fuzzy Way Point Tactics dan Behavior Tree. Kedua hal ini dapat membantu AI pada game War of Zombies sehingga berperilaku lebih cerdas. Beberapa kesimpulan teknis yang didapat selama penelitian antara lain:

1. Pada game bertema strategi seperti war of zombies ini, fuzzy logic sangat effective dalam membantu AI dalam mengambil keputusan bagi setiap karakter
2. Behavior Tree merupakan metode yang sangat baik dalam mengatur langkah yang harus diselesaikan oleh karakter atau perilaku karakter.
3. Untuk mencari tempat yang baik dalam melakukan suatu aksi seperti bersembunyi ataupun menembak, waypoint tactic adalah cara yang efektif dan mudah diterapkan bagi developer game.

Beberapa saran untuk penelitian lebih lanjut antara lain adalah:

1. Jenis desain diagram behavior tree akan lebih baik jika dibuat symbol dan keterangan yang lebih lengkap untuk memperjelas alur dan fungsi dari tiap node.
2. Library pendukung untuk mengimplementasikan behavior tree masih tergolong baru sehingga banyak fitur yang dirasa masih bisa ditingkatkan. Penulis berharap dalam waktu kedepan, setelah banyak library mendukung behavior tree, dapat lebih banyak penelitian tentang AI pada Game yang memanfaatkan behavior tree.

DAFTAR PUSTAKA

- [1] I. Millington and J. Funge, *Artificial intelligence for games*. CRC Press, 2018.
- [2] M. Colledanchise and P. Ögren, *Behavior trees in robotics and AI: An introduction*. CRC Press, 2018.
- [3] M. Nicolau, D. Perez-Liebana, M. O'Neill, and A. Brabazon, "Evolutionary behavior tree approaches for navigating platform games," *IEEE Trans. Comput. Intell. AI Games*, vol. 9, no. 3, pp. 227–238, 2016.
- [4] P. A. Lindsay, S. Kromodimoeljo, P. A. Strooper, and M. Almorisy, "Automation of test case generation from behavior tree

- requirements models,” in *2015 24th Australasian Software Engineering Conference*, 2015, pp. 118–127.
- [5] P. A. Lindsay, “Behavior trees: from systems engineering to software engineering,” in *2010 8th IEEE International Conference on Software Engineering and Formal Methods*, 2010, pp. 21–30.
- [6] L. Lidén and others, “Strategic and tactical reasoning with waypoints,” 2002.
- [7] A. C. Prasetyo, M. P. Arnandi, H. S. Hudnanto, and B. Setiaji, “Perbandingan Algoritma Astar dan Dijkstra Dalam Menentukan Rute Terdekat,” *Sisfotenika*, vol. 9, no. 1, pp. 36–46, 2019.

INSYST

Journal of Intelligent System and Computation

Volume 02 Nomor 01 April 2020

Author Guidelines

- Manuscript should be written in Indonesia and be submitted online via journal website. Online Submission will be charged at no Cost
- Manuscript should not exceed 15 pages including embedded figures and tables, without any appendix, and the file should be in Microsoft Office (.doc/.docx). [download template](#)
- Title, Abstract and Keywords must be written in bilingual
- Title should be less than 15 words
- Abstracts consists of no more than 200 words, contains the essence of the article and includes a brief background, objectives, methods and results or findings of the study. Abstract is written in one paragraph.
- Keywords are written in Indonesia and English three to five words/phrases, separated with coma and consist of important words/phrases from the article.
- Author's name, affiliation, affiliation address and email. State clearly and include country's name on your affiliation address.
- The main text of the writing should be consists of: Introduction, Method, Result and Discussion, and Conclusion; followed by Acknowledgment and Reference
- Introduction State adequate background, issues and objectives, avoiding a detailed literature survey or a summary of the results. Explain how you addressed the problem and clearly state the aims of your study.
- Used method is the scientific in the form of study of literature, observation, surveys, interviews, Focus Group Discussion, system testing or simulation and other techniques commonly used in the world of research. It is also recommended to describe analysis techniques used briefly and clearly, so that the reader can easily understand.
- Results should be clear, concise and not in the form of raw data. Discussion should explore the significance of the results of the work, not repeat them. Avoid extensive citations and discussion of published literature. INSYST will do the final formatting of your paper.
- Conclusion should lead the reader to important matter of the paper. Authors are allowed to include suggestion or recommendation in this section. Write conclusion, suggestion and/or recommendation in narrative form (avoid of using bulleting and numbering)
- Acknowledgments. It is highly recommended to acknowledge a person and/or organizations helping author(s) in many ways. Sponsor and financial support acknowledgments should be included in this section. Should you have lots of parties

to be acknowledged, state your acknowledgments only in one paragraph. Avoid of using bulleting and numbering in this section

- The number of references are not less than 10 with at least 8 primary references. Primary references are include journal, thesis, disertasion and all kinds of research reports. All refferences must come from source published in last 7 years.
- Figure and table should be in black and white, and if it is made in color, it should be readable when it is later printed in black and white.
- Figure and table should be clearly readable and in a proportional measure to the overall page.

Tim Redaksi

Journal of Intelligent System and Computation

Departement of Informatics

Institut Sains dan Teknologi Terpadu Surabaya

Jl. Ngagel Jaya Tengah 73-77 Surabaya

Email: insyst@istts.ac.id

Website: <https://jurnal.stts.edu/index.php/INSYST/index>