

ALAT VISUALISASI ANALITIK DATA UNTUK PETUA SEKUTUAN

Jasmine Tan She Ming

Assoc. Prof. Dr. Zulaiha Ali Othman

Fakulti Teknologi dan Sains Maklumat, Universiti Kebangsaan Malaysia

ABSTRAK

Petua sekutuan merupakan satu kaedah yang dapat membantu pengguna untuk mendapat maklumat daripada data yang banyak. Tambahan pula, petua sekutuan yang diperolehi selepas kaedah algoritma apriori adalah maklumat yang berguna untuk pengguna tentang data tersebut. Kini, hasil daripada algoritma apriori yang akan didapati di dalam bentuk petua sekutuan yang terlalu banyak telah menyebabkan pengguna untuk menghadapi masalah untuk menganalisis dan memahami datanya. Kajian ini akan menghasilkan satu alat visualisasi yang dapat mengubah petua sekutuan yang terdapat di dalam satu teks yang panjang ke satu bentuk visualisasi yang dapat memudahkan pengguna untuk memahami petua sekutuan yang telah dijana selepas apriori algoritma. Data yang akan digunakan untuk kajian ini adalah data yang terdiri daripada 9 atribut dan setiap atribut akan mempunyai keadaan seperti tinggi, sederhana dan rendah untuk setiap keadaan yang berlaku. Data akan diproses dan disimpan di dalam jadual dengan melihat jumlah atribut untuk setiap data. Selepas itu, data akan diproses dan dikeluarkan di dalam satu bentuk visualisasi yang sesuai. Kajian ini akan mengeluarkan beberapa bentuk visualisasi yang berbeza yang akan bergantung kepada jumlah atribut yang terlibat.

1 PENGENALAN

Petua sekutuan merupakan satu kaedah dalam perlombongan data. Tetapi, petua sekutuan boleh didapati di dalam hasil yang banyak selepas dianalisis dengan menggunakan algoritma apriori. Ini menyebabkan pengguna perlu menganalisis data satu per satu kerana keputusan daripada algoritma apriori yang terlalu banyak untuk alat visualisasi untuk menggambarkan semua. Proses ini sangat mengambil masa dan tenaga pengguna yang membuat tindakan tersebut.

Algoritma apriori juga merupakan satu algoritma untuk perlombongan data. Hasil daripada algoritma apriori dapat digunakan untuk menganalisis data dengan mengklasifikasikan data dengan berdasarkan keadaan. Algoritma apriori merupakan satu kaedah untuk mencari item yang berulang di dalam satu dataset. Kaedah ini menggunakan pengetahuan sebelumnya untuk mencari item yang telah berulang. Apriori algoritma merupakan salah satu kaedah yang digunakan untuk data yang banyak dan hasil daripada algoritma apriori boleh didapati di dalam beberapa keadaan dengan berdasarkan item yang berulang di dalam keadaan tersebut. Algoritma ini dapat membantu pengguna untuk meningkatkan hasil jualan dengan mengenal pasti barang-barang yang biasa dibeli atau barang yang disukai oleh pembeli.

Hasil daripada algoritma apriori didapati di dalam teks yang panjang dan ini merupakan masalah kepada pengguna untuk mendapat maklumat daripadanya. Untuk memudahkan pengguna, alat visualisasi telah digunakan untuk menunjukkan maklumat yang terdapat di dalam data yang banyak tersebut. Alat visualisasi dapat memudahkan pengguna untuk memahami tentang maklumat tersebut.

Alat visualisasi merupakan satu kaedah untuk menunjukkan maklumat dengan menggunakan grafik. Alat visualisasi dapat memudahkan pengguna untuk lebih memahami maklumat yang terdapat berdasarkan data tersebut. Alat visualisasi yang terdapat di pasaran sekarang tidak dapat menggambarkan semua keadaan untuk satu kumpulan data yang banyak. Keadaan ini berlaku dan akan menyebabkan pengguna untuk menghadapi masalah untuk memahami dan mendapat maklumat tentang data tersebut.

Alat visualisasi yang terdapat di pasaran adalah terhad untuk pengguna kerana bukan semua alat visualisasi dapat menggambarkan semua maklumat daripada petua sekutuan. Sebagai contoh, Sisense yang terlalu susah untuk pengguna yang tiada kemahiran dalam menggunakan komputer untuk menggunakannya kerana alat ini memerlukan teknikal untuk menggunakannya. Tableau juga merupakan salah satu alat visualisasi yang terhad kerana graf boleh digunakan untuk maklumat yang terhad sahaja. Maklumat yang terlalu banyak tidak boleh menggunakan Tableau. Selain itu, qlikview juga merupakan alat yang terhad kerana ingatan capaian rawak (RAM) yang terhad dan aplikasi yang tidak sering dibangunkan ke versi yang baru. Petua sekutuan yang mudah dapat digambarkan dengan menggunakan alat visualisasi yang terdapat di pasaran tetapi jikalau petua sekutuan yang susah dan banyak, ini akan menjadi satu masalah kepada pengguna kerana tiada alat visualisasi yang terdapat di pasaran yang dapat menukarkan kepada bentuk grafik.

2 PENYATAAN MASALAH

Alat visualisasi yang terdapat di pasaran adalah terhad untuk menunjukkan maklumat atau informasi untuk data yang banyak. Terdapat kaedah yang dapat mengekstrak maklumat daripada data seperti apriori algoritma yang akan dapat membawa hasil keluaran maklumat di dalam satu teks yang panjang dan boleh terdapat di pasaran tetapi tiada alat visualisasi yang terdapat di pasaran yang dapat membantu pengguna untuk mengubah maklumat yang terdapat di dalam bentuk satu teks yang panjang ke dalam satu bentuk grafik. Masalah ini akan mengakibatkan pengguna untuk menghadapi masalah untuk mengetahui semua maklumat yang berkaitan dengan data dan mungkin akan hilang maklumat yang berkaitan dengan data kerana data yang terlalu banyak untuk pengguna menganalisis satu per satu. Alat visualisasi yang terdapat di pasaran yang terhad tidak dapat membantu pengguna untuk menukar petua sekutuan ke dalam bentuk grafik. Contoh alat visualisasi yang terdapat seperti Tableau, Qlikview dan Sisense yang tidak dapat digunakan jikalau petua sekutuan yang terlalu banyak untuk diproses.

3 OBJEKTIF KAJIAN

Objektif kajian ini adalah untuk mengeluarkan satu alat yang dapat membantu pengguna untuk memahami maklumat yang terdapat di dalam data yang banyak melalui alat visualisasi. Objektif kajian khususnya untuk mencapai objektif utama tersebut:

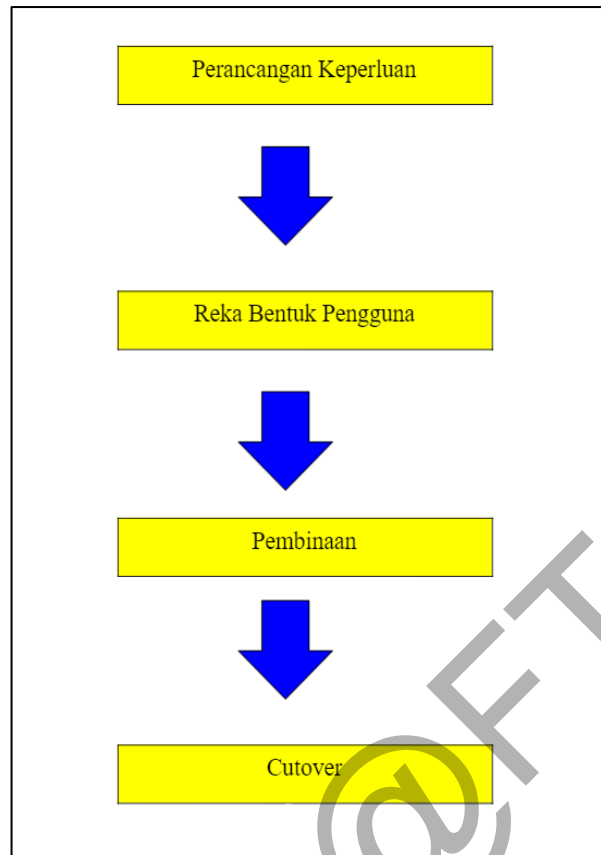
- Membangunkan satu alat visualisasi untuk analisis petua sekutuan yang banyak

4 METOD KAJIAN

Kaedah yang akan digunakan untuk kajian ini adalah kaedah pembangunan aplikasi pesat (RAD). Kaedah pembangunan aplikasi pesat merupakan salah satu kaedah di bawah kaedah pembangunan tangkas yang mementingkan fungsi perisian dan maklum balas daripada pengguna tentang penambah baik untuk sistem.

Kaedah pembangunan tangkas merupakan kaedah yang mementingkan interaksi di antara pengguna. Selain itu, kaedah ini juga mementingkan masa dengan membuat perisian lebih daripada membuat dokumentasi. Tambahan pula, kaedah ini juga akan memberi tumpuan kepada komen pengguna dan akan sering mengubah suai perisian tersebut. Dengan ini, kaedah ini dapat sentiasa meningkatkan perisian ke tahap yang lebih tinggi. Kaedah pembangunan tangkas merangkumi pelbagai pendekatan untuk pembangunan perisian di mana keperluan dan penyelesaian dipenuhi. Pembangunan perisian tangkas menganjurkan perancangan, pembangunan dan peningkatan terhadap perisian yang sudah ada. Kaedah ini dipilih kerana kaedah ini merupakan kaedah yang mudah untuk menggabungkan keperluan pengguna dalam proses menjalankan kajian ini.

Terdapat beberapa fasa semasa menjalankan kaedah pembangunan aplikasi pesat (RAD) :



Rajah 4 Fasa-fasa kaedah pembangunan aplikasi pesat (RAD)

4.1 Fasa Perancangan

Fasa ini adalah fasa perancangan keperluan. Fasa ini merupakan fasa yang menggabungkan element sistem perancangan dan fasa analisis sistem bagi Kitaran Pembangunan Sistem (SDLC). Fasa ini akan menyelesaikan keperluan pengguna kerana fasa ini akan melibatkan proses pengenalpastian masalah dan objektif. Data yang digunakan untuk kajian ini akan dikumpulkan dan diagihkan kepada pelbagai jadual dengan berdasarkan jumlah atribut yang terdapat di dalam setiap petua sekutuan. Tafsiran maklumat akan dikumpul dan dianalisis di dalam fasa ini.

4.2 Fasa Reka Bentuk

Fasa ini adalah fasa reka bentuk pengguna. Untuk fasa ini, model dan prototaip yang mewakili sistem akan dibangunkan. Sistem yang direka akan bergantung dengan kesukaan reka bentuk pengguna. Dalam fasa ini, reka bentuk untuk setiap keadaan akan direka dengan berdasarkan kesesuaian untuk memaparkan data tersebut dalam bentuk grafik. Bentuk grafik yang direka akan menggunakan pelbagai warna untuk menandakan setiap atribut dengan keadaannya. Fasa ini akan melibatkan beberapa proses yang penting, iaitu reka bentuk untuk antara muka alat visualisasi dan juga carta bar untuk setiap peringkat.

4.3 Fasa Pembinaan

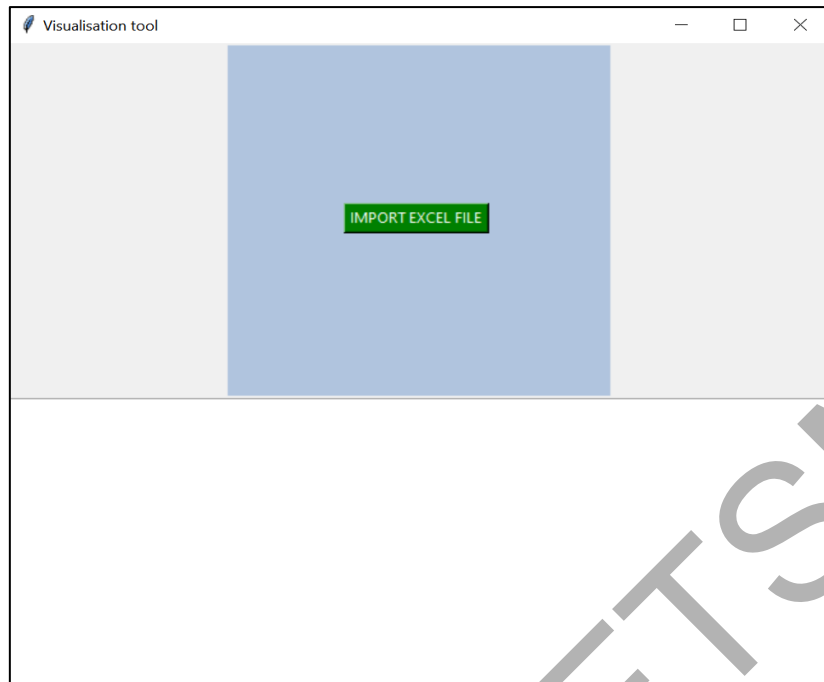
Fasa ini adalah fasa pembinaan. Fasa ini menumpukan kepada tugas untuk membangunkan program. Fasa ini adalah membolehkan pengguna untuk memahami, mengubah dan meluluskan model yang memenuhi keperluan pengguna yang menggunakan sistem ini. Dalam fasa ini, pengguna masih boleh menukar atau menyatakan pendapatnya untuk menambah baik. Dalam fasa ini, model akan diuji untuk ketepatan maklumat dan membuat pengaturcaraan terhadap program ini.

4.4 Fasa Cantas (*Cutover*)

Fasa ini adalah fasa cantas. Fasa ini merupakan fasa yang terakhir. Aktiviti yang akan dilakukan pada fasa ini adalah pengujian sistem yang telah dibangunkan. Sistem yang dibangunkan mestilah dapat memenuhi keperluan pengguna. Sekiranya gagal untuk mencapai objektif kajian untuk projek ini, penyelarasan perlu dijalankan.

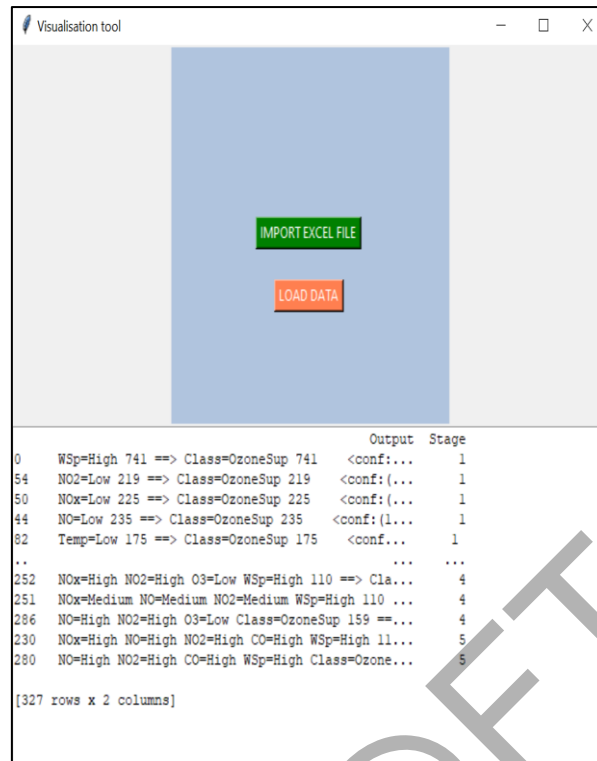
5 HASIL KAJIAN

Bahagian ini akan membincangkan hasil daripada proses pembangunan alat visualisasi untuk petua sekutuan. Alat visualisasi akan dibangunkan di dalam Python dengan menggunakan IDE Visual Studio Code. Alat ini yang dibangunkan akan memerlukan pengguna untuk memuat naik fail excel ke dalam program.



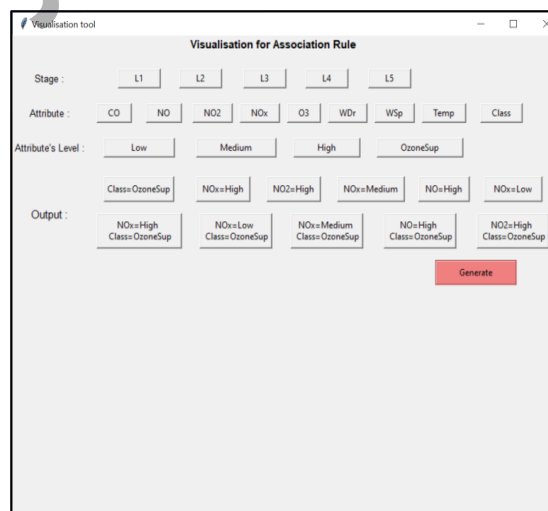
Rajah 2 Reka bentuk muka depan

Selepas memuat naik fail excel ke dalam alat visualisasi, program akan membahagikan petua sekutuan kepada peringkat masing-masing dengan bergantung jumlah atribut yang terdapat di dalam petua sekutuan. Sebagai contohnya, semua petua sekutuan yang hanya melibatkan 1 atribut akan disimpan di dalam bahagian L1, semua petua sekutuan yang melibatkan 2 atribut akan disimpan di dalam bahagian L2 dan sebagainya. Seterusnya, program akan mengambil atribut, keadaan atribut dan output daripada petua sekutuan tersebut dan simpan di dalam senarai yang baru. Selepas dianalisis, muka depan akan memaparkan data di dalam fail tersebut. Selepas itu, pengguna boleh melihat data tersebut dan menekan butang "Load Data" selepas memastikan bahawa data tersebut akan digunakan untuk alat visualisasi tersebut.



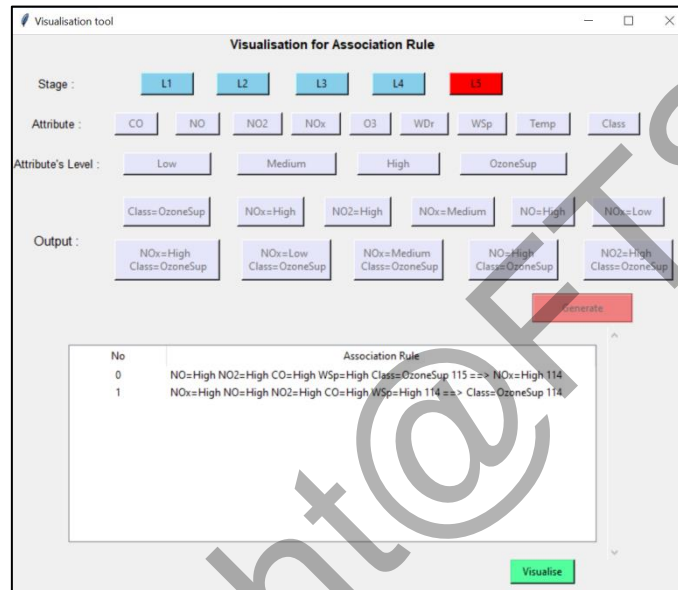
Rajah 3 Antara muka selepas memuat naik fail excel

Selepas itu, alat visualisasi akan menggunakan data tersebut untuk memaparkan data berdasarkan peringkat di dalam bentuk jadual dan juga di dalam bentuk grafik. Jadi, selepas pengguna menekan butang “Load Data”, antara muka untuk pengguna memilih input iaitu peringkat, atribut, keadaan atribut dan output dengan berdasarkan peringkat yang pengguna memilih akan dipaparkan.



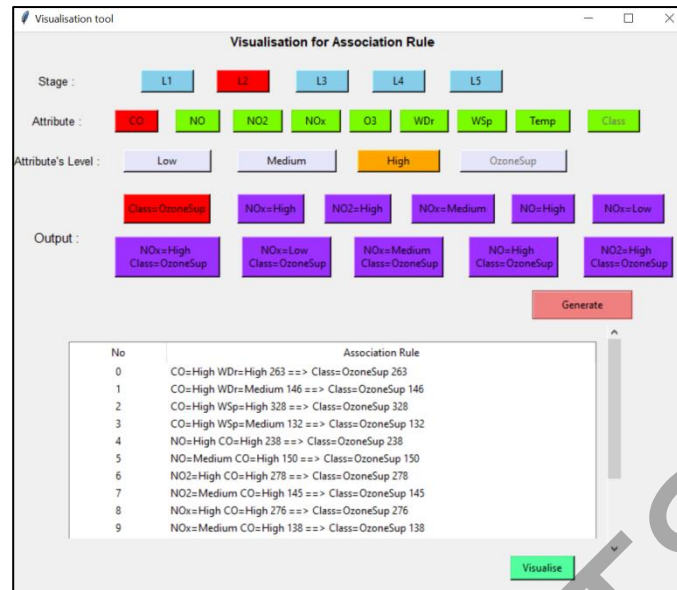
Rajah 4 Antara muka untuk pengguna memilih input

Selepas itu, pengguna boleh memilih peringkat dengan berdasarkan pilihan pengguna. Untuk peringkat L1 dan L5, pengguna tidak perlu memilih input atribut, keadaan atribut dan output kerana data yang tidak banyak di dalamnya jadi alat visualisasi akan menyebabkan butang yang berkaitan kepada tidak aktif supaya pengguna tidak akan memilihnya.



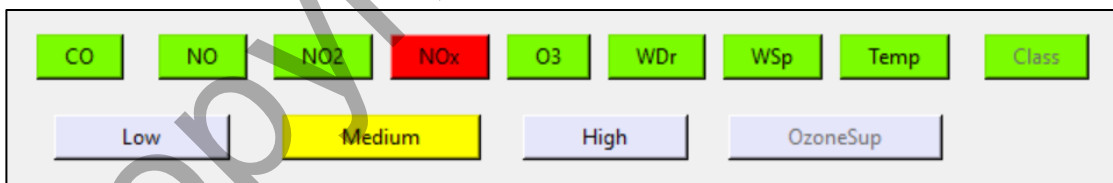
Rajah 5 Contoh antara muka yang mempunyai butang yang tidak boleh dipilih oleh pengguna

Selepas itu, untuk peringkat L2 hingga peringkat L4, pengguna boleh memilih atribut, keadaan atribut dan output sebagai input.

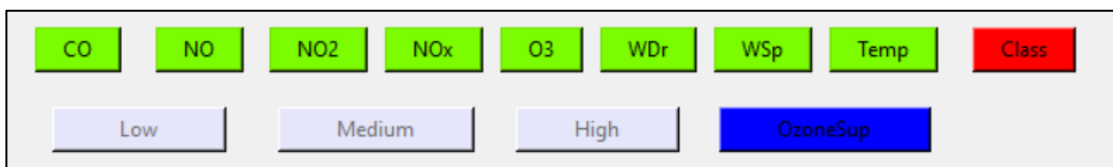


Rajah 6 Contoh antara muka untuk memaparkan senarai yang berdasarkan pilihan pengguna

Untuk atribut 'CO', 'NO', 'NO2', 'NOx', 'O3', 'WDr', 'WSp' dan 'Temp', butang 'Low', 'Medium' dan 'High' boleh dipilih oleh pengguna manakala untuk atribut 'Class', butang 'OzoneSup' sahaja boleh dipilih. Untuk butang 'Low', warnanya akan ditukar ke warna merah, butang 'Medium' akan ditukar ke warna kuning dan butang 'High' akan ditukar ke warna oren jika pengguna memilihnya. Manakala, butang 'OzoneSup' akan ditukar ke warna biru jika dipilih oleh pengguna di dalam sistem.



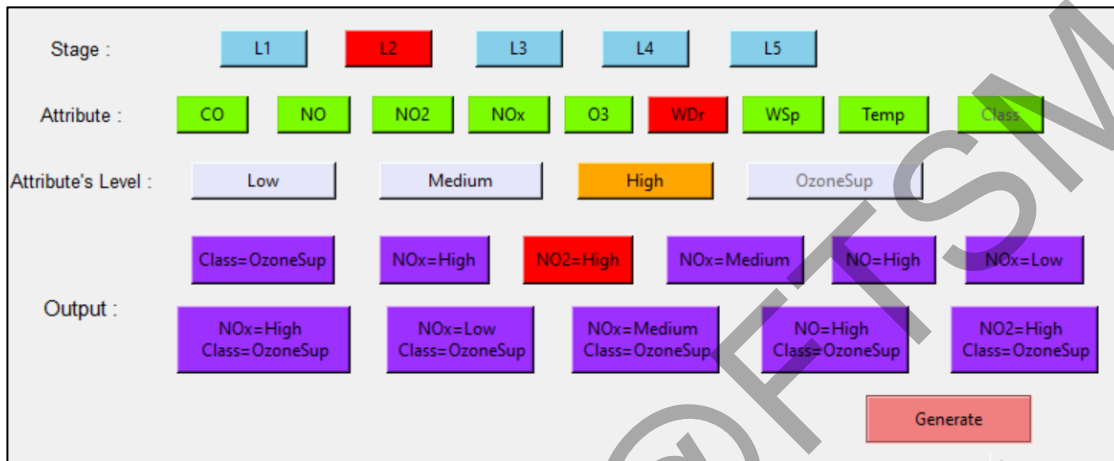
Rajah 7 Atribut dengan butang 'Low', 'Medium' dan 'High'



Rajah 8 Atribut dengan butang 'OzoneSup'

Setiap baris mempunyai warna sendiri di mana baris pertama merupakan peringat

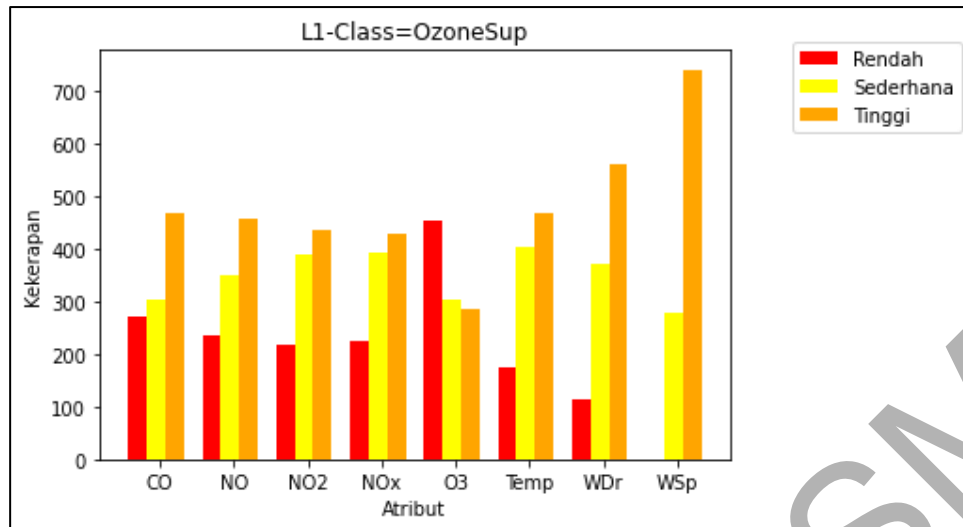
dengan warnanya biru muda, baris kedua adalah atribut dengan warnanya hijau, baris ketiga adalah keadaan untuk atribut yang telah dipilih dan warnanya adalah berdasarkan keadaan sendiri yang boleh terdapat di atas dan baris yang terakhir adalah output dengan warnanya ungu. Selepas pengguna memilih butang untuk setiap baris, butang tersebut akan ditukar ke warna merah.



Rajah 9 Contoh warna untuk butang di setiap baris

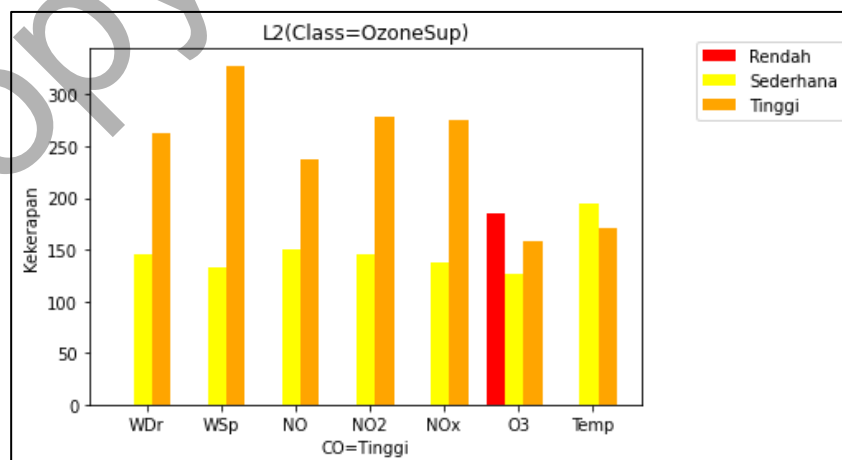
Untuk memaparkan carta bar peringkat L1 dan L5, pengguna hanya perlu menekan butang hijau di mana butang boleh didapati di bawah jadual. Manakala untuk peringkat L2 hingga L4, pengguna hanya perlu menekan butang merah di mana butang merah tersebut boleh didapati di atas jadual di antara muka alat visualisasi.

Carta bar untuk L(1) yang terdapat di contoh di bawah adalah semua data adalah mempunyai output "Class=OzoneSup". Setiap keadaan mempunyai warna yang berbeza. Warna merah mewakili rendah, kuning mewakili sederhana dan oren pula mewakili tinggi. Paksi x adalah mewakili atribut dan paksi y adalah mewakili kekerapan untuk setiap atribut. Semua atribut mempunyai 3 keadaan kecuali atribut WSp mempunyai keadaan sederhana dan tinggi sahaja dengan melihat rajah yang terdapat di bawah.



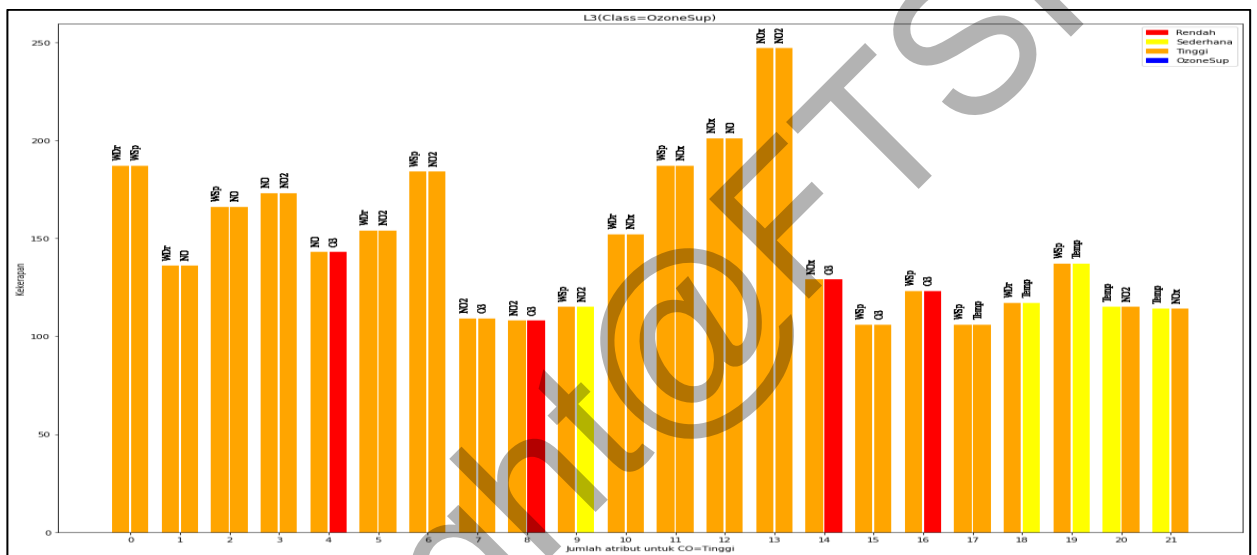
Rajah 10 Carta bar untuk L(1)

Carta bar untuk L(2) pula adalah seperti di bawah. Paksi x merupakan atribut CO=Tinggi seperti rajah di bawah. Atribut untuk paksi x merupakan input yang dipilih oleh pengguna. Paksi y merupakan kekerapan untuk setiap atribut yang bersama dengan CO=Tinggi. Atribut yang terdapat di dalam carta bar adalah L(2) jadi atribut yang terdapat di dalam carta bar adalah bersama dengan atribut yang terdapat di paksi x kerana peringkat L2 mempunyai 2 atribut untuk petua sekutuan tersebut untuk berlaku. Sebagai contoh, rajah di bawah merupakan atribut CO dengan atribut lain yang terdapat di dalam carta bar. Tambahan pula, atribut yang dipilih oleh pengguna adalah atribut yang dipilih dengan keadaannya bersama.



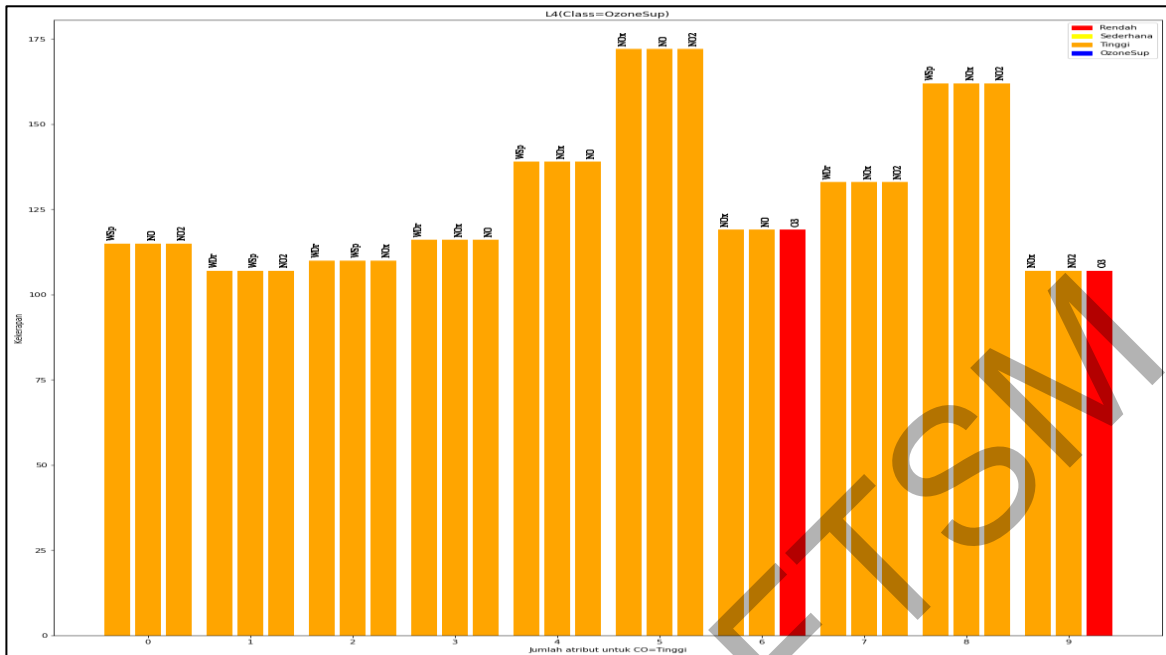
Rajah 11 Carta bar untuk L(2)

Carta di bawah merupakan contoh untuk L(3). Paksi x adalah atribut CO dengan keadaannya tinggi dan paksi y adalah kekerapan. L(3) merupakan 3 item yang bersama dengan menghasilkan output. Sebagai contoh, atribut CO dengan 2 lagi atribut dalam carta bar telah bersama menghasilkan output “Class=OzoneSup”. Setiap atribut mempunyai keadaan yang berlainan di atas carta bar tersebut yang mendorong output untuk berlaku. Setiap keadaan dan atribut mempunyai warna yang berlainan untuk carta bar di atas supaya pengguna dapat mengetahui atribut apa dan dengan keadaan apa yang mendorong output itu berlaku.



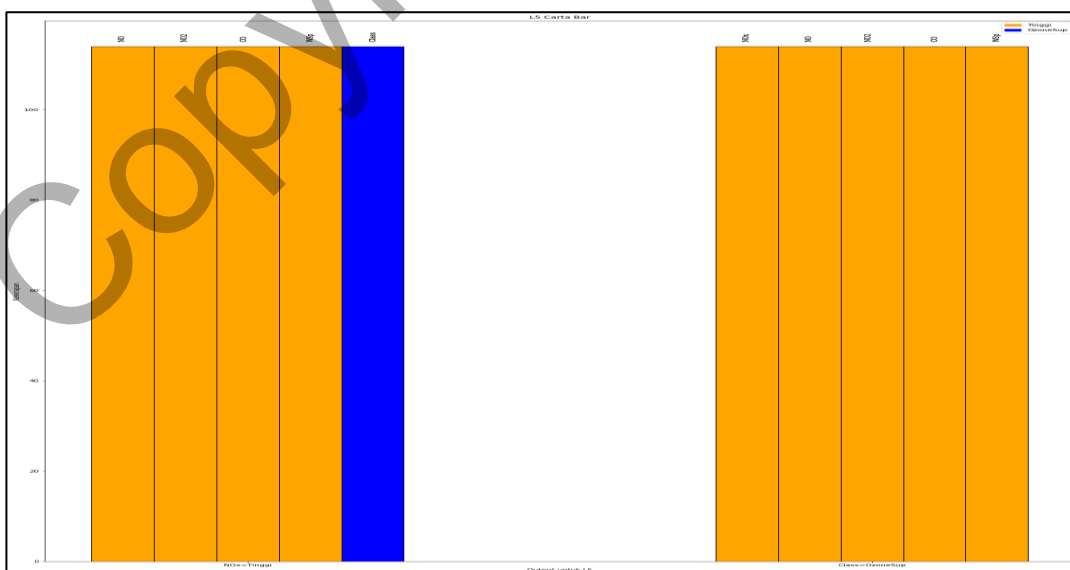
Rajah 12 Carta bar untuk L(3)

Carta bar di bawah adalah carta bar untuk L(4). Paksi x dan paksi y adalah sama dengan carta bar untuk L(3) tetapi perbezaannya adalah L(4) mempunyai 4 item untuk menghasilkan output dan L(3) mempunyai 3 item untuk menghasilkan output. 4 item adalah paksi x dengan atribut yang lain di carta bar. Setiap bar mempunyai 3 warna yang juga bermaksud 3 atribut dengan keadaannya yang mendorong output untuk berlaku. 3 atribut dengan paksi x yang merupakan CO telah menghasilkan 4 atribut yang mendorong output “Class=OzoneSup” untuk berlaku.



Rajah 13 Carta bar untuk L(4)

Carta bar untuk L(5) adalah berbeza dengan L(1) hingga L(4). Paksi x untuk L(5) adalah output dan paksi y adalah kekerapan. Paksi x merupakan output kerana untuk L(5) yang kajian ini melakukan mempunyai 2 output sahaja. Outputnya adalah “Class=OzoneSup” dan “NOx=Tinggi”. Setiap bar untuk output tersebut mempunyai 5 warna yang mewakili 5 atribut dengan keadaannya.



Rajah 14 Carta bar untuk L(5)

6 KESIMPULAN

Kesimpulannya, alat visualisasi untuk petua sekutuan boleh membantu pengguna untuk memahami dan mengambil maklumat daripada algoritma apriori data. Alat ini boleh membantu pengguna untuk menganalisis dan menukarkan data di dalam teks yang panjang kepada bentuk grafik dan di dalam bentuk jadual selepas dianalisis. Dengan menggunakan alat visualisasi ini, pengguna boleh mengambil dan memahami data dengan mudah. Alat ini juga memudahkan pengguna kerana pengguna tidak perlu menganalisis petua sekutuan satu per satu. Pengguna hanya perlu memuat naik fail ke dalam alat visualisasi sahaja.

Sistem ini hanya akan menganalisis petua sekutuan yang dapat daripada algoritma apriori. Kedua, alat ini hanya memberi pengguna untuk memuat naik fail excel sahaja. Selain itu, sistem yang dibangunkan adalah berdasarkan dataset yang diberi, jadi jika dataset yang berlainan yang diguna di dalam sistem, reka bentuk untuk antara muka perlu menukar ke atribut yang telah digunakan.

7 RUJUKAN

- D.Magdalene Delighta Angeline. 2013. Association Rule Generation for Student Performance Analysis using Apriori Algorithm.
- Rakesh Agrawal, Ramakrishnan Srikant. 1994. Fast Algorithms for Mining Association Rules.
- Hsin-Kai Wu, Joseph S. Krajcik, Elliot Soloway. 2001. Promoting understanding of chemical representations: Students' use of a visualization tool in the classroom.
- Pak Chung Wong, P. Whitney, J. Thomas. 1999. Visualizing association rules for text mining.
- Kelsie Anderson. 2019. What is Rapid Application Development(RAD).
- Abdullah Saad Al-Malaise. 2013. Implementation of Apriori Algorithm to Analyze Organization Data: Building Decision Support System.
- Santhosh Kumar, K.V.Rukmani. 2010. Implementatuin of Web Usage Mining Using APRIORI and FP Growth Algorithms.
- Teemu Rajala, Mikko-Jussi Laakso, Erkki Kaila, Tapio Salakoski. 2008. Effectiveness of

Program Visualization: A Case Study with the ViLLE Tool.

Data visualization beginner's guide: a definition, examples, and learning resources.

[https://www.tableau.com/learn/articles/data-](https://www.tableau.com/learn/articles/data-visualization#:~:targetText=Data%20visualization%20is%20the%20graphical,outliers%2C%20and%20patterns%20in%20data.)

[visualization#:~:targetText=Data%20visualization%20is%20the%20graphical,outliers%2C%20and%20patterns%20in%20data.](https://www.tableau.com/learn/articles/data-visualization#:~:targetText=Data%20visualization%20is%20the%20graphical,outliers%2C%20and%20patterns%20in%20data.)

Copyright@FTSM