

Penerapan Graf pada *Neural Network* dalam Aplikasi *Song Recognition* Shazam

Muhammad Fikri Naufal - 13591518¹
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia
¹13519158@std.stei.itb.ac.id

Abstract—*Song recognition* merupakan sebuah sistem yang dapat mengenali lagu dengan input berbentuk audio. Di antara beberapa nama-nama besar lain, Shazam merupakan salah satu aplikasi paling terkenal yang menggunakan sistem *song recognition*. Sistem *song recognition* didasari oleh sistem pengenalan suara atau *voice recognition*. *Voice recognition* mengenali suatu suara dengan membandingkan pola karakteristiknya dengan sinyal suara yang menjadi referensi atau acuan yang telah disimpan pada database. Pada prosesnya, *song recognition* menggunakan penerapan teori graf pada *neural network* dalam *encoder* dan *decoder* yang dipakai.

Keywords—Graf, *neural network*, Shazam, *song recognition*, *voice recognition*.

I. PENDAHULUAN

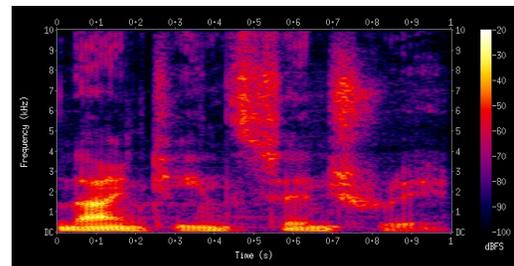
Shazam adalah sebuah aplikasi sosluler yang dapat mengenali musik. Pada tahun 1999, Shazam Entertainment Limited didirikan oleh Chris Barton, Philip Inghelbrecht, Avery Wang, dan Dhiraj Mukherjee. Di tahun 2002, cara Shazam dioperasikan yaitu memanggil “2580” pada ponsel, menahan panggilan mengikuti musik, lalu akan dikirimkan judul lagu dan artis dalam bentuk SMS. Pada tahun 2008, Shazam menjadi salah satu aplikasi pertama di Apple App Store. Di tahun 2011, bukan hanya musik, Shazam dapat mendeteksi program TV dan iklan.



Gambar 1.1. Logo Shazam
Sumber: [3]

Pada saat ini, aplikasi Shazam sudah dapat diunduh di berbagai platform, seperti Android dan iOS. Cara kerja Shazam sangat mudah dan unik. Misal anda datang ke sebuah toko dan menyukai musik yang diputar. Jalankan aplikasi dan tekan tombol berlogo Shazam. Aplikasi Shazam akan membuat digital fingerprint dari audio yang terdengar dan setelah beberapa detik,

ditemukan kecocokan audio dengan database yang berisi jutaan lagu dan acara televisi. Kemudian anda akan diberi informasi seputar lagu yang terdengar tersebut, mulai dari nama artis, judul lagu, lirik, video, tiket konser, dan lain-lain.



Gambar 1.2. Contoh Spektrogram
Sumber: [7]

Salah satu hal menarik dari cara kerja Shazam yaitu *song recognition*, di mana setiap audio atau lagu memiliki representasi digital fingerprint yang unik untuk dicocokkan dengan database. Audio digital fingerprint dibuat dengan salah satu teknik untuk membuat grafik frekuensi waktu yang disebut spektrogram. Algoritma yang digunakan Shazam yaitu memilih titik-titik puncak energi dalam spektrogram yang dibentuk sebagai frekuensi yang mewakili audio untuk mengabaikan kebisingan. Dari audio digital fingerprint ini, musik dapat dibedakan berdasarkan genre nya. Misal musik metal memiliki kecenderungan frekuensi di bagian riff gitar yang garang, vokal berenergi tinggi dan dentuman drum.

Pada proses pembentukan digital fingerprint ini, digunakan teori graf dalam *neural network* yang dipakai pada *encoder* dan *decoder* yang akan dibahas pada bagian-bagian selanjutnya.

II. DASAR TEORI

A. Graf

1. Definisi Graf

Graf adalah sebuah struktur yang digunakan untuk menggambarkan objek-objek diskrit dan hubungan di antara objek-objek tersebut.

Secara detail, graf didefinisikan sebagai $G = (V, E)$ dengan V yaitu himpunan tidak kosong dari simpul-simpul (vertices) yang menggambarkan objek-objek diskrit, misal $\{v_1, v_2, v_3,$

..., v_n }, dan E yaitu himpunan dari sisi (edge) yang menghubungkan objek atau simpul, misal $\{e_1, e_2, e_3, \dots, e_n\}$. Sisi e yang menghubungkan simpul u dengan simpul v dinotasikan dengan: $e = (u, v)$.

2. Jenis-Jenis Graf

Graf digolongkan dalam dua jenis berdasarkan ada atau tidaknya sisi ganda dan gelang, yaitu graf sederhana dan graf tak-sederhana.

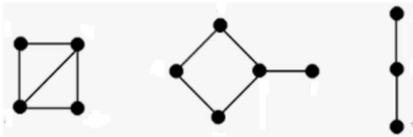
a) Graf Sederhana

Graf sederhana adalah graf yang tidak mengandung gelang maupun sisi ganda. Pada graf sederhana, sisi merupakan struktur berisi pasangan tidak terurut sehingga:

$$(u, v) = (v, u)$$

Sisi pada graf ini menghubungkan dua simpul yang berbeda, dengan notasi:

$$e = (u, v) \\ u \neq v$$



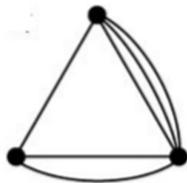
Gambar 2.1. Graf Sederhana
Sumber: [1]

b) Graf Tak-Sederhana

Graf tak-sederhana (unsimple graph) adalah graf yang mengandung sisi ganda atau gelang dinamakan. Graf tak-sederhana digolongkan lagi berdasarkan ada atau tidaknya gelang pada graf yaitu graf ganda dan graf semu.

1. Graf Ganda (Multi-Graph)

Graf ganda adalah graf yang mengandung sisi ganda. Sisi ganda adalah dua buah sisi atau lebih yang menghubungkan dua buah simpul yang sama.

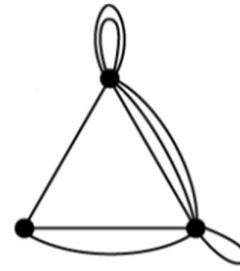


Gambar 2.2. Graf Ganda
Sumber: [1]

2. Graf Semu (Pseudo-Graph)

Graf semu adalah graf yang mengandung sisi gelang. Gelang merupakan sisi yang menghubungkan simpul dengan simpul itu sendiri, dengan notasi:

$$e = (u, v) \\ u = v$$



Gambar 2.3. Graf Semu
Sumber: [1]

Graf juga dapat digolongkan berdasarkan orientasi arah pada sisi, yaitu graf tidak berarah dan graf berarah.

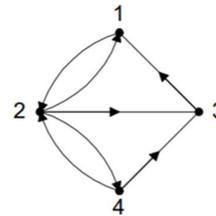
a) Graf Tidak Berarah

Graf tidak berarah adalah graf yang sisinya tidak berarah sehingga urutan pasangan simpul yang dihubungkan sisi tidak perlu diperhatikan, yaitu:

$$(u, v) = (v, u)$$

b) Graf Berarah

Graf berarah adalah graf yang pada sisinya mengandung arah. Pada gambar, orientasi arah pada graf biasanya direpresentasikan sebagai tanda panah. Urutan pasangan simpul pada sisi diperhatikan sehingga $(u, v) \neq (v, u)$. Pada sisi (u, v) , u adalah simpul asal dan v adalah simpul tujuan.



Gambar 2.4. Graf Berarah
Sumber: [1]

3. Terminologi pada Graf

Graf memiliki beberapa istilah dasar yang perlu diketahui, yaitu ketetanggaan, beririsan, simpul terpencil, graf kosong, derajat, lintasan, sirkuit, terhubung, upagraf, komplemen, upagraf rentang, *cut-set*, dan graf berbobot.

a) Ketetanggaan (*Adjacency*)

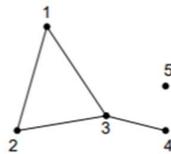
Ketetanggaan pada graf diartikan sebagai dua buah simpul pada graf tak berarah G yang dihubungkan langsung oleh sebuah sisi. Hal ini dilambangkan dengan (u, v) , sama dengan lambang sisi, yang memiliki arti yaitu simpul u bertetangga dengan simpul v .

b) Beririsan (*Incidency*)

Misal jika notasi sisi $e = (u, v)$, menandakan sisi e beririsan dengan simpul u dan simpul v , yaitu sisi e menghubungkan langsung simpul u dan v .

c) Simpul Terpencil (*Isolated Vertex*)

Simpul terpencil pada graf diartikan sebagai simpul yang tidak mempunyai sisi yang berurusan dengannya atau simpul yang tidak memiliki tetangga. Tinjau gambar 2.5., pada graf, simpul 5 tidak memiliki simpul tetangga sehingga simpul 5 adalah simpul terpencil.



Gambar 2.5. Graf yang memiliki simpul terpencil
Sumber: [1]

d) Graf Kosong (*Null Graph*)

Graf Kosong adalah graf yang himpunan sisinya merupakan himpunan kosong. Semua simpul pada graf kosong adalah simpul terpencil.

e) Derajat (*Degree*)

Derajat pada simpul graf adalah jumlah sisi graf yang berurusan dengan simpul tersebut. Derajat pada sebuah simpul v dinotasikan sebagai $d(v)$. Pada gambar 2.5., derajat-derajat simpulnya yaitu:

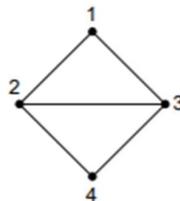
$$\begin{aligned} d(1) &= d(2) = 2 \\ d(3) &= 3 \\ d(4) &= 1 \\ d(5) &= 0 \end{aligned}$$

f) Lintasan (*Path*)

Lintasan yang panjangnya n dari simpul awal v_0 ke simpul tujuan v_n pada graf G adalah barisan berselangseling simpul, dengan notasi $v_0, e_0, v_1, e_1, v_2, e_2, \dots, v_{n-1}, e_{n-1}, v_n, e_n$ dengan sisi $e_1 = (v_0, v_1), e_2 = (v_1, v_2) \dots, e_n = (v_{n-1}, v_n)$ himpunan sisi $E = \{e_1, e_2, \dots, e_n\}$. Lintasan juga dapat disebut jalur. Panjang lintasan dapat dihitung dengan menghitung jumlah sisi dalam lintasan tersebut. Lintasan sederhana (simple path) adalah lintasan dengan semua simpulnya berbeda.

g) Sirkuit (*Circuit*)

Siklus atau sirkuit adalah lintasan yang berawal dan berakhir pada simpul yang sama. Sirkuit termasuk lintasan tertutup. Panjang sirkuit dapat dihitung dengan menghitung jumlah sisi pada sirkuit tersebut. Pada gambar 2.6., 1,2,4,3 adalah sebuah sirkuit



Gambar 2.6. Contoh Sirkuit Graf
Sumber: [1]

h) Terhubung (*Connected*)

Dua buah simpul pada graf dapat dikatakan terhubung jika ada lintasan dari suatu simpul asal ke simpul lain. Pada graf tidak berarah, graf dapat disebut sebagai graf terhubung (connected graf) jika untuk setiap pasang simpul u dan v pada himpunan V pada graf, terdapat lintasan dari u ke v , sedangkan graf dapat disebut sebagai graf tidak terhubung (disconnected graf) jika untuk setiap pasang simpul u dan v pada himpunan V pada graf, tidak terdapat lintasan dari u ke v .

Pada graf berarah, graf dapat disebut sebagai graf terhubung jika graf tidak berarahnya terhubung. Pada graf berarah, graf terhubung dibedakan menjadi dua, yaitu graf terhubung kuat (strongly connected) dan graf terhubung lemah (weakly connected). Graf terhubung kuat adalah graf yang memiliki lintasan dari simpul u ke v dan juga lintasan dari simpul v ke u sedangkan jika u dan v tidak terhubung kuat tetapi terhubung pada graf tidak berarahnya, maka u dan v dikatakan terhubung lemah (weakly connected).

i) Upagraf dan Komplemen Upagraf

Sebuah graf $G_1 = (V_1, E_1)$ dikatakan sebagai upagraf dari graf $G = (V, E)$ jika $V_1 \subseteq V$ dan $E_1 \subseteq E$. Sebuah graf $G_2 = (V_2, E_2)$ dikatakan komplemen dari upagraf G terhadap G_1 jika $E_2 = E - E_1$ dan V_2 merupakan himpunan simpul yang anggotanya bersisian dengannya. Komponen graf (connected component) adalah jumlah maksimum upagraf terhubung dalam graf tersebut. Pada graf berarah, komponen terhubung kuat (strongly connected component) adalah jumlah maksimum upagraf yang terhubung kuat.

j) Upagraf rentang

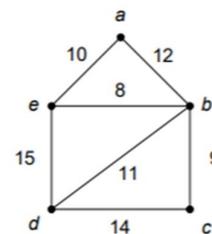
Upagraf $G_1 = (V_1, E_1)$ dari $G = (V, E)$ dikatakan upagraf rentang jika $V_1 = V$ (yaitu G_1 mengandung semua simpul dari G).

k) Cut-Set

Cut-set dari graf terhubung G adalah himpunan sisi yang bila dibuang dari G menyebabkan G tidak terhubung. Jadi, cut-set selalu menghasilkan dua buah komponen.

l) Graf Berbobot (*Weighted Graph*)

Graf berbobot adalah graf yang setiap sisinya diberi sebuah harga (bobot).



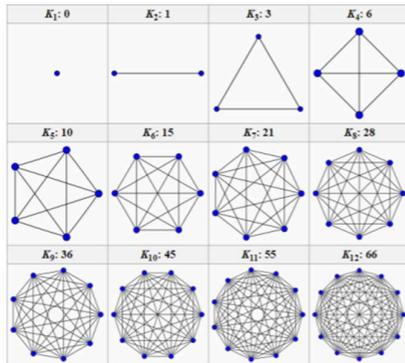
Gambar 2.7. Contoh Graf Berbobot
Sumber: [1]

4. Graf Khusus

Terdapat beberapa graf khusus yang memiliki pola atau bentuk tertentu yaitu:

a) Graf Lengkap

Graf lengkap adalah graf sederhana yang setiap simpulnya mempunyai sisi ke semua simpul lainnya. Semua simpul pada graf lengkap saling bertetangga. Graf lengkap dengan n buah simpul dilambangkan dengan K_n . Jumlah sisi pada graf lengkap yang terdiri dari n buah simpul adalah $n(n - 1)/2$.



Gambar 2.8. Graf Lengkap
Sumber: [1]

b) Graf Lingkaran

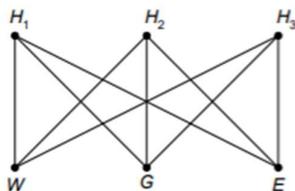
Graf lingkaran adalah graf sederhana yang setiap simpulnya berderajat dua. Graf lingkaran dengan n simpul dilambangkan dengan C_n . Pada gambar 2.8., K_3 adalah graf lingkaran.

c) Graf Teratur

Graf yang setiap simpulnya mempunyai derajat yang sama disebut graf teratur. Apabila derajat setiap simpul adalah r, maka graf tersebut disebut sebagai graf teratur derajat r. Jumlah sisi pada graf teratur adalah $nr/2$.

d) Graf Bipartit

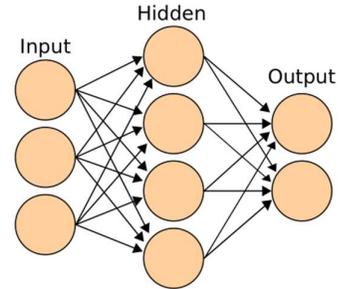
Graf G yang himpunan simpulnya dapat dipisah menjadi dua himpunan bagian V_1 dan V_2 , sedemikian sehingga setiap sisi pada G menghubungkan sebuah simpul di V_1 ke sebuah simpul di V_2 disebut graf bipartit dan dinyatakan sebagai $G(V_1, V_2)$. Pada gambar 2.9., graf $K_{3,3}$ adalah contoh graf bipartit.



Gambar 2.9. Graf $K_{3,3}$
Sumber: [1]

B. Artificial Neural Network

Artificial neural network atau jaringan saraf tiruan (JST) adalah jaringan dari sekelompok unit pemroses kecil yang dimodelkan berdasarkan sistem saraf manusia. JST merupakan sistem adaptif yang dapat mengubah strukturnya untuk memecahkan masalah berdasarkan informasi eksternal maupun internal yang mengalir melalui jaringan tersebut. Oleh karena sifatnya yang adaptif, JST juga sering disebut dengan jaringan adaptif.



Gambar 2.10. Gambaran Lapisan JST
Sumber: [6]

JST dimodelkan dari fungsi matematika $f(X) = Y$. JST menggunakan istilah “jaringan” yaitu merujuk kepada koneksi dari neuron-neuron yang diletakkan di lapisan berbeda. Secara umum, lapisan JST yaitu:

1. Lapis masukan (input layer) terdiri dari neuron yang menerima data masukan dari variabel X. Semua neuron pada lapis ini dapat terhubung ke neuron pada lapisan tersembunyi atau langsung ke lapisan luaran jika jaringan tidak menggunakan lapisan tersembunyi.
2. Lapisan tersembunyi (hidden layer) terdiri dari neuron yang menerima data dari lapisan masukan.
3. Lapisan luaran (output layer) terdiri dari neuron yang menerima data dari lapisan tersembunyi atau langsung dari lapisan masukan yang nilai luarannya melambangkan hasil kalkulasi dari X menjadi nilai Y.

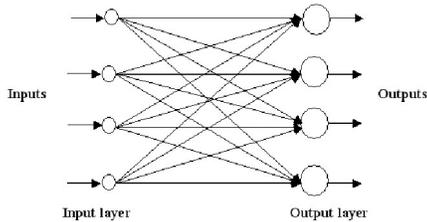
Secara matematis, neuron merupakan sebuah fungsi yang menerima masukan dari lapisan sebelumnya $g_i(x)$ (lapisan ke-i). Fungsi ini pada umumnya mengolah sebuah vektor untuk kemudian diubah ke nilai skalar melalui komposisi nonlinear weighted sum.

III. PENERAPAN GRAF PADA ARTIFICIAL NEURAL NETWORK

Pada artificial neural network, graf digunakan untuk menggambarkan arsitektur jaringan, yaitu penyusunan neuron pada lapisan-lapisan dan pola koneksinya dalam dan antarlapisan. Perlu diperhatikan bahwa unit masukan tidak terhitung sebagai lapisan karena tidak melakukan komputasi. Jadi, jumlah lapisan pada jaringan ditentukan berdasarkan lapisan yang berisikan bobot antar koneksi dari neuron-neuron. Beberapa contoh dari arsitektur jaringan saraf tiruan, yaitu:

1. Single Layer Network

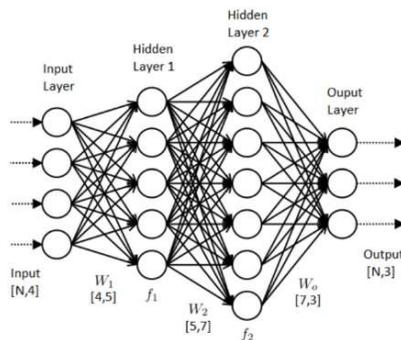
Jaringan dengan lapisan tunggal hanya memiliki satu lapisan dengan bobot-bobot terhubung. Jaringan ini hanya menerima masukan kemudian secara langsung akan mengolahnya menjadi keluaran tanpa harus melalui lapisan tersembunyi. Neuron-neuron pada lapisan masukan terhubung langsung dengan neuron-neuron pada lapisan keluaran.



Gambar 3.1. Ilustrasi single layer network
Sumber: [6]

2. Multi Layer Network

Jaringan dengan banyak lapisan memiliki satu atau lebih lapisan yang terletak di antara lapisan masukan dan lapisan keluaran. Lapisan tersebut disebut lapisan tersembunyi (hidden layer). Jaringan dengan banyak lapisan ini dapat menyelesaikan permasalahan yang lebih sulit daripada lapisan dengan lapisan tunggal, tentu saja dengan pembelajaran yang lebih rumit. Namun demikian, pada banyak kasus, pembelajaran pada jaringan dengan banyak lapisan ini lebih sukses dalam menyelesaikan masalah.



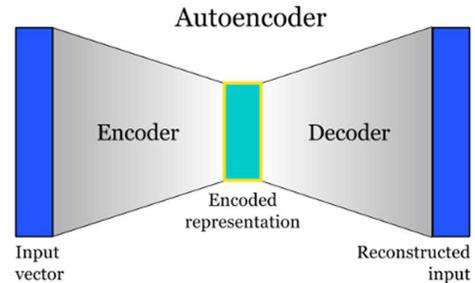
Gambar 3.1. Ilustrasi single layer network
Sumber: [6]

IV. PENERAPAN ARTIFICIAL NEURAL NETWORK PADA SISTEM SONG RECOGNITION DALAM APLIKASI SHAZAM

Telah dijelaskan pada bagian pendahuluan bahwa telah dibentuk audio digital fingerprint yang dikenal sebagai signature. Signature akan dipakai sebagai input dari sistem jaringan saraf tiruan yang akan dilatih.

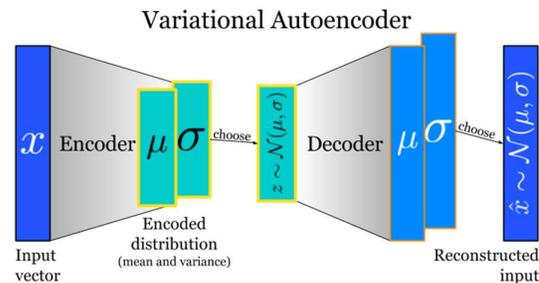
Arsitektur jaringan yang dipilih Shazam adalah model autoencoder variasional atau VAE. AE akan dilatih untuk merekonstruksi input signature yang diberikan sedekat mungkin, yaitu representasi yang dipadatkan di tengah (condensed).

Selama pelatihan, kerugian (loss) dihitung dari selisih mean squared dari input dan output yang telah direkonstruksi. Parameter dari encoder dan decoder lalu diperbarui dan disesuaikan dengan kontribusinya terhadap nilai kerugian (proses backprop).



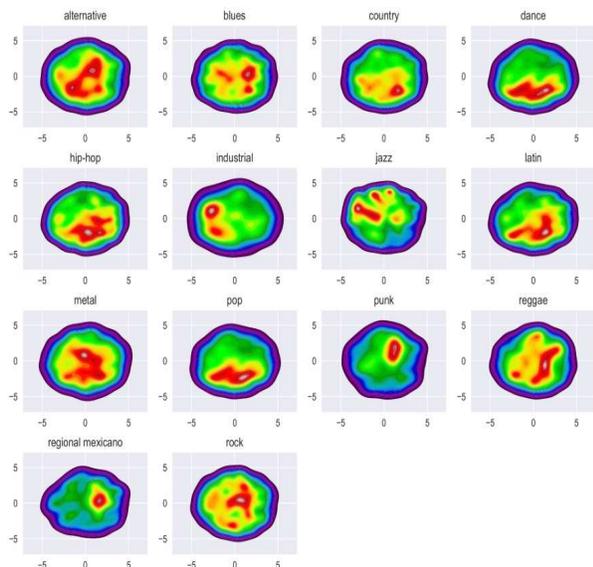
Gambar 4.1. Model Autoencoder
Sumber: [2]

Model pada gambar 4.1. lalu dimodifikasi dengan menambahkan metode Bayesian variasional pada encoder dan decodernya. Model ini belajar untuk memperkirakan distribusi probabilitas dari representasi yang telah di-encode. Lalu model ini mengambil vector dari distribusi tersebut sebagai input dari decoder. Setelah itu dapat diperoleh rekonstruksi dengan memilihnya dari distribusi output. Fungsi kerugian (loss) dari VAE menjadi sedikit lebih rumit karena berhubungan dengan distribusi probabilitas, tetapi lebih menguntungkan karena membentuk distribusi yang lebih bermakna



Gambar 4.2. Model Modifikasi Variasional Autoencoder
Sumber: [2]

Hasil dari pelatihan model dapat divisualisasikan dengan model t-SNE untuk mengubah beberapa hasil yang berdimensi tinggi menjadi dua dimensi untuk setiap genre musik (diwakili sumbu x dan y pada grafik). Untuk menunjukkan distribusi dari banyak trek lagu, digunakan warna ungu dan biru untuk menunjukkan sedikit trek yang berdekatan dan warna jingga, merah, dan putih sebagai area yang menunjukkan banyaknya trek yang berdekatan.



Gambar 4.3. Visualisasi Hasil Distribusi Trek Tiap Genre
Sumber: [2]

V. KESIMPULAN

Kesimpulan yang dapat diambil dari makalah ini yaitu banyaknya aplikasi dari ilmu Matematika Diskrit di bidang teknologi informasi. Salah satunya adalah struktur graf yang dapat merepresentasikan arsitektur jaringan dari Artificial Neural Network atau jaringan saraf tiruan. Jaringan saraf tiruan ini memiliki banyak sekali aplikasi yang salah satunya adalah pembuatan sistem song recognition untuk mengenali lagu dengan input berbentuk audio.

VI. UCAPAN TERIMA KASIH

Puji syukur penulis panjatkan kepada Allah SWT atas berkat, rahmat, dan hidayah-Nya, dapat terselesaikannya tugas makalah ini dengan baik. Pada kesempatan ini, penulis ingin mengucapkan terima kasih kepada Ibu Dra. Harlili S., M.Sc. selaku Dosen Mata Kuliah IF2120 Matematika Diskrit Kelas 02 atas bimbingan dan ilmu yang telah diberikan. Penulis juga ingin menyampaikan terima kasih kepada kedua orang tua atas dukungan yang selalu diberikan kepada penulis. Penulis juga mengucapkan terima kasih kepada teman-teman yang senantiasa memberikan bantuan dalam penyusunan makalah ini.

REFERENCES

- [1] Munir, Rinaldi. 2016. *Matematika Diskrit*, Edisi Revisi Keenam. Bandung: Informatika Bandung
- [2] O'Brien, Tim. 2017. Learning to Understand Music from Shazam <https://medium.com/@tsob/learning-to-understand-music-from-shazam-56a60788b62f>. Diakses 8 November pada pukul 23.30 WIB
- [3] Shazam. 2020. <https://www.shazam.com/company>. Diakses 8 November pada pukul 22.15 WIB
- [4] Hertz, J., Palmer, R.G., Krogh. A.S. 1990. Introduction to the theory of neural computation, Perseus Books.
- [5] Dwivedi, Priya. 2018. Using CNNs and RNNs for Music Genre Recognition. <https://towardsdatascience.com/using-cnns-and-rnns-for-music-genre-recognition-2435fb2ed6af>. Diakses 9 November pada pukul 21.55 WIB

- [6] https://www.academia.edu/34779977/MAKALAH_JARINGAN_SYAR_AF_TIRUAN.pdf. Diakses pada 8 Desember pukul 20.13 WIB
- [7] http://audio.rightmark.org/lukin/pub/aes_adapt/popup.html Diakses pada 8 Desember pukul 20.00 WIB

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Jakarta, 11 Desember 2020

Muhammad Fikri Naufal 13519158