

Aplikasi Pohon Keputusan dalam Klasifikasi Lagu Populer pada Spotify

Ramadhana Bhanuharya Wishnumurti 13519203¹

Program Studi Teknik Informatika

Sekolah Teknik Elektro dan Informatika

Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia

¹13519203@std.stei.itb.ac.id

Abstrak—Spotify adalah sebuah perusahaan teknologi yang menyediakan layanan streaming musik dan media yang berasal dari Swedia. Model bisnis Spotify menawarkan dua tingkatan bagi penggunaannya, yaitu Spotify gratis dan Spotify Premium dengan fitur-fitur seperti kualitas audio yang tinggi, musik *offline*, dan tidak ada iklan. Sampai saat ini, pengguna Spotify telah mencapai 286 juta orang dengan 130 juta orang yang menggunakan Spotify Premium. Spotify menyediakan Web API yang bisa diakses oleh para pengembang dan menyediakan metadata tentang lagu, artis, popularitas, dan hal lainnya yang berasal langsung dari aktivitas jutaan penggunanya. Penulis melakukan klasifikasi dengan pohon keputusan untuk mencari tahu karakteristik lagu yang memiliki nilai popularitas tinggi berdasarkan data yang disediakan oleh Spotify Web API.

Keywords—Pohon keputusan, Spotify, musik, *classification*

I. PENDAHULUAN

Musik adalah susunan suara yang memiliki elemen seperti nada, ritme, dinamika, kualitas sonik timbre, dan tekstur yang dihasilkan dari alat-alat yang menghasilkan bunyian tersebut. Musik berkembang sedemikian rupa dari zaman dahulu hingga saat ini dimana kita hampir tidak bisa terlepas olehnya. Industri musik juga menjadi salah satu bagian terpenting dari dunia hiburan dimana hampir setiap harinya ada sebuah lagu yang dirilis oleh musisi dari seluruh dunia.

Saat ini kebanyakan orang mengakses musik melalui platform digital, salah satu yang saat ini mulai dingemari adalah platform streaming audio dalam aplikasi seperti Spotify, Apple Music, Youtube Music, dan lainnya. Spotify adalah salah satu dari platform tersebut yang digemari karena kemudahan akses karena tersedianya aplikasi *mobile* dan katalog musik yang tersedia sangat banyak. Perusahaan asal Swedia ini memiliki 50 juta katalog lagu, dan sekitar 700 ribu podcast yang bisa dinikmati oleh 286 juta pengguna setianya.



Gambar 1 Logo Spotify (Sumber :

https://id.m.wikipedia.org/wiki/Berkas:Spotify_logo_with_text.svg
diakses 10 Desember 2020 22:45 WIB)

Spotify menyediakan dua tipe akun pengguna yaitu Spotify gratis dan Spotify Premium. Spotify Premium membolehkan pengguna untuk membuka Batasan dari fitur Spotify gratis seperti tanpa iklan, *download* lagu supaya dapat diakses *offline*, dan lainnya. Spotify juga menawarkan fitur seperti rekomendasi musik, tampilan lirik, dan Wrapped yang meringkas momen musik dari pengguna pada akhir tahun.

Spotify juga menyediakan API Web yang mengambil konten Spotify seperti data lagu, karakteristik dari lagu, dan lainnya. Banyak pengembang memanfaatkan API Web Spotify ini untuk membuat aplikasi web yang berhubungan dengan Spotify seperti website untuk melihat musik favorit pengguna, menyortir music pengguna, dan untuk merekomendasi album berdasarkan data dari pengguna. Contoh dari data yang tersedia adalah seperti karakteristik lagu seperti tempo, irama, sampai ke tingkat popularitas.

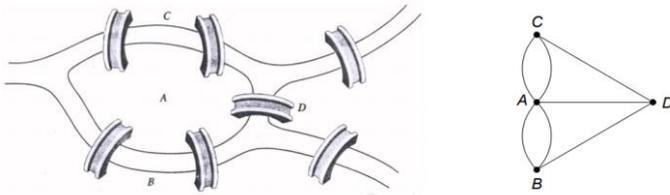
Oleh karena itu, penulis memikirkan sebuah ide untuk menggunakan pohon keputusan untuk melakukan klasifikasi terhadap data lagu spotify dan mencari karakteristik dari lagu yang memiliki nilai popularitas tinggi. Hal ini bisa digunakan oleh para musisi, produser, dan label musik supaya mengetahui karakteristik musik apa yang bisa mendapatkan tingkat popularitas tinggi pada aplikasi Spotify, salah satu platform streaming audio terbesar untuk menghasilkan profit dan menjangkau lebih banyak pendengar.

II. DASAR TEORI

A. Graf

Graf adalah bentuk susunan diagram yang menunjukkan relasi dan hubungan antara objek diskrit dalam sebuah pasangan himpunan simpul dan sisi yang saling terhubung. Graf didefinisikan sebagai G yang memiliki komponen himpunan tidak-kosong dari simpul-simpul (V) dan himpunan sisi yang menghubungkan sepasang simpul (E) dan dapat dituliskan sebagai notasi $G = (V, E)$.

Pemakaian pertama teori graf dapat dilacak pada tahun 1735 saat Leonhard Euler (1707-83) memberikan solusi untuk masalah Jembatan Königsberg dengan mempresentasikan jembatan tersebut dengan sisi dan daratan sebagai simpul sehingga setiap jembatan dapat dilalui orang tepat sekali dan kembali ke tempat semula.

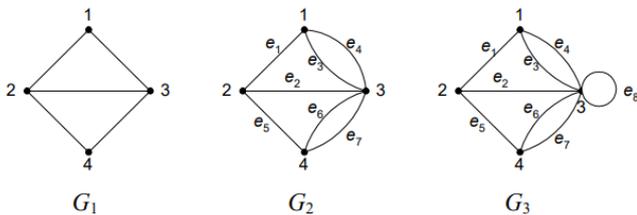


Gambar 2 Permasalahan Jembatan Königsberg

(Sumber : Powerpoint Matematika Diskrit IF2120/Rinaldi Munir diakses 10 Desember 2020 22:50 WIB)

Berdasarkan ada atau tidak adanya gelang/sisi ganda pada sebuah graf, maka graf dibedakan menjadi:

1. Graf Sederhana, yaitu graf yang tidak mengandung gelang dan sisi ganda.
2. Graf Tak-Sederhana, yaitu graf yang mengandung sisi ganda, sisi gelang.



Gambar 3 Graf Sederhana, Graf Ganda, dan Graf Semu

(Sumber : Powerpoint Matematika Diskrit IF2120/Rinaldi Munir diakses 10 Desember 2020 22:53 WIB)

Berdasarkan orientasi arah pada sisi, graf dibedakan menjadi 2 jenis yaitu:

1. Graf Tak-Berarah, yaitu graf yang tidak memiliki orientasi arah pada sisinya.
2. Graf Berarah, yaitu graf yang memiliki orientasi arah pada sisinya.

Penerapan graf pada ilmu pengetahuan sangat banyak seperti pada rangkaian listrik, isomer senyawa kimia karbon, jejaring makanan, pengujian program, dan penerapan bermanfaat lainnya.

B. Pohon

Pohon adalah bentuk tak berarah dari graf di mana dua simpul terhubung pada satu jalur dan tidak mengandung sirkuit. Setiap simpul pada pohon terhubung dalam satu jalur dan saling terhubung dengan simpul lainnya yang terus bercabang sampai ujung dari pohon tersebut.

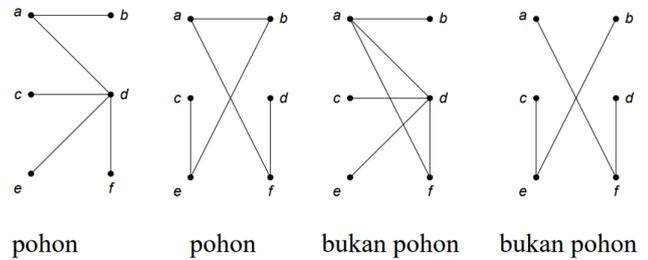
Pada gambar diatas, sebuah graf dapat dideskripsikan menjadi pohon ketika tidak berbentuk sirkuit, berbeda dengan contoh bukan pohon bukan pohon yang bentuk grafnya menjadi sebuah sirkuit yang terhubung. Istilah pohon diciptakan oleh Arthue Cayley, seorang matematikawan Inggris pada tahun 1857.

C. Properti Pohon

Teorema pohon sendiri adalah misal $G = (V,E)$ adalah graf tak-berarah dengan jumlah simpul berupa n , maka pernyataan dibawah akan ekuivalen:

1. G adalah pohon.

2. Setiap pasang simpul di dalam G akan terhubung dengan lintasan tunggal.
3. G terhubung dan memiliki sisi sebanyak $m = n - 1$.
4. G tidak terhubung sirkuit dan memiliki sisi sebanyak $m = n - 1$.
5. G tidak memiliki sirkuit dan penambahan satu sisi pada graf tersebut akan membuatnya menjadi satu sirkuit.
6. G terhubung dan semua sisinya adalah jembatan.

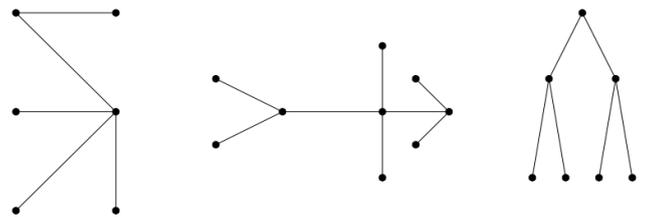


Gambar 4 Contoh Pohon dan Bukan Pohon

(Sumber : Powerpoint Matematika Diskrit IF2120/Rinaldi Munir diakses 10 Desember 2020 22:55 WIB)

D. Hutan

Hutan adalah sekumpulan pohon yang memiliki hubungan saling lepas atau graf tidak terhubung yang tidak mengandung sirkuit. Berikut ilustrasi dari hutan:

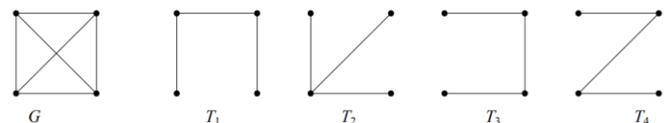


Gambar 5 Ilustrasi dari Hutan

(Sumber : Powerpoint Matematika Diskrit IF2120/Rinaldi Munir diakses 10 Desember 2020 22:55 WIB)

E. Pohon Merentang

Pohon merentang adalah graf terhubung atau upgraf yang berupa pohon. Pohon merentang terbentuk dengan memutus sirkuit pada graf. Pohon merentang memiliki $n-1$ tepi, di mana n adalah banyaknya simpul (*nodes*). Dari mengurangi maksimum $e - n + 2$ tepi dari graf, kita bisa membangun sebuah pohon merentang dan sebuah graf lengkap dapat memiliki $n^2 - n$ pohon merentang.



Gambar 6 Ilustrasi dari Pohon Merentang

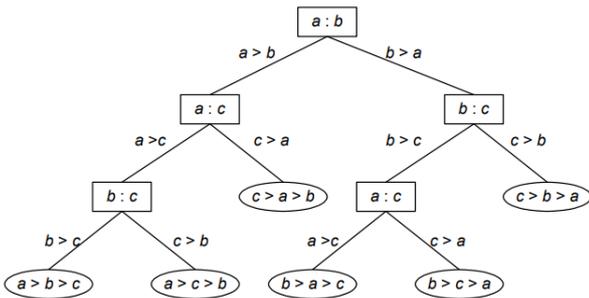
(Sumber : Powerpoint Matematika Diskrit IF2120/Rinaldi Munir diakses 10 Desember 2020 22:56 WIB)

F. Pohon Berakar

Pohon berakar adalah pohon yang satu buah simpulnya berlaku sebagai akar dan pada sisinya diberikan arah sehingga menjadi sebuah graf berarah. Bentuk pohon berakar sangat

mirip dengan bentuk pohon alami.

G. Pohon Keputusan



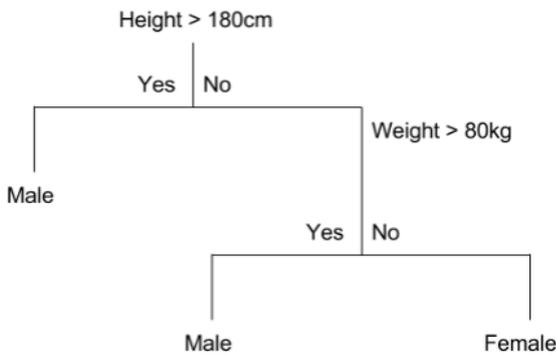
Gambar 7 Ilustrasi dari Pohon Keputusan untuk mengurutkan 3 buah elemen

(Sumber : Powerpoint Matematika Diskrit IF2120/Rinaldi Munir diakses 10 Desember 2020 22:57 WIB)

Pohon keputusan adalah aplikasi dari konsep pohon untuk mengambil sebuah keputusan dengan struktur data pada sebuah pohon dan melihat sebuah item pengamatan (cabang) dengan melihat kondisi setelahnya (daun). Pohon keputusan digunakan sebagai representasi keputusan dan pengambilan keputusan sehingga dapat dilihat secara visual dan eksplisit. Pohon keputusan sering diimplementasikan dalam memprediksi dalam pembelajaran mesin dengan model yang melakukan kalkulasi menggunakan model kemungkinan kondisional.

H. Klasifikasi

Klasifikasi adalah salah satu jenis dari pohon keputusan yang melihat hasil dari variabel menjadi sebuah kategori diskrit (ya/tidak). Pohon keputusan untuk klasifikasi dibangun dengan proses yang bernama partisi rekursif biner, yaitu proses berulang untuk memisahkan data yang tersedia menjadi banyak partisi lalu membaginya lagi pada cabang selanjutnya.



Gambar 8 Ilustrasi dari Klasifikasi

(Sumber : <https://medium.com/swlh/decision-tree-classification-de64fc4d5aac> diakses 11 Desember 2020 20:57 WIB)

III. APLIKASI POHON KEPUTUSAN

Untuk aplikasi pohon keputusan pada klasifikasi lagu populer di Spotify, penulis menggunakan set data Spotify pada Kaggle yang berasal dari API Web Spotify. Set data ini memiliki 160 ribu koleksi lagu yang berasal dari tahun 1921 sampai dengan

2020. Digunakan program R Studio untuk mengolah data ini menjadi model pohon keputusan.



Gambar 9 Set Data Spotify

(Sumber : https://www.kaggle.com/yamaerenay/spotify-dataset-19212020-160k-tracks?select=data_w_genres.csv diakses 11 Desember 2020 21:30 WIB)

Data (data.csv) ini terdiri dari *valence*, *year*, *acousticness*, *artist*, *danceability*, *duration_ms*, *energy*, *explicit*, *id*, *instrumentalness*, *key*, *liveness*, *loudness*, *mode*, *name*, *popularity*, *release_date*, *speechiness*, dan *tempo*. Penulis berencana hanya menggunakan beberapa data yang memberi informasi karakteristik dari lagu dan juga data popularitas, oleh karena itu dibuatlah sebuah data baru bernama *df* yang memiliki struktur seperti ini (*artist*, *name*, *year*, *tempo*, *speechiness*, *acousticness*, *danceability*, *duration_ms*, *energy*, *explicit*, *instrumentalness*, *key*, *liveness*, *loudness*, dan *popularity*).

```
str(data)
df <- data[,c("artists", "name", "year", "tempo", "speechiness", "acousticness", "danceability", "duration_ms", "energy", "explicit", "instrumentalness", "key", "liveness", "loudness", "popularity")]
df
```

Gambar 10 Pembentukan tabel data "df" pada R Studio

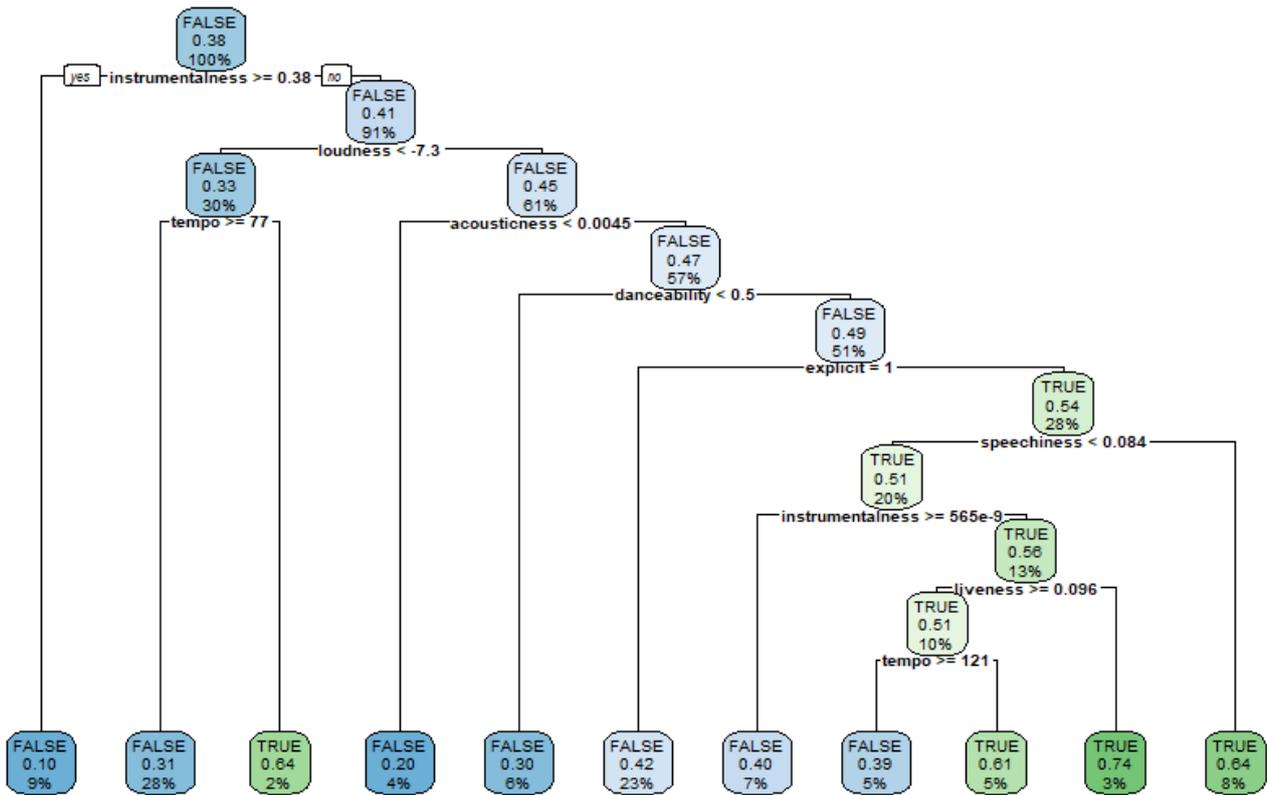
Lalu untuk level popularitas yang dijadikan standar minimum adalah pada angka 70 karena pada Spotify Top 50 Song 2019, nilai terkecil pada lagu yang masuk pada penghargaan itu adalah dengan nilai popularitas 70. Lalu dilakukan sortir untuk lagu dari tahun 2019 dan 2020 untuk data yang lebih relevan dengan trend pengguna Spotify.

```
data2020 <- df[df$year == 2020,]
above <- data2020[popularity >= 70,]
data2020$rank <- c(above)
```

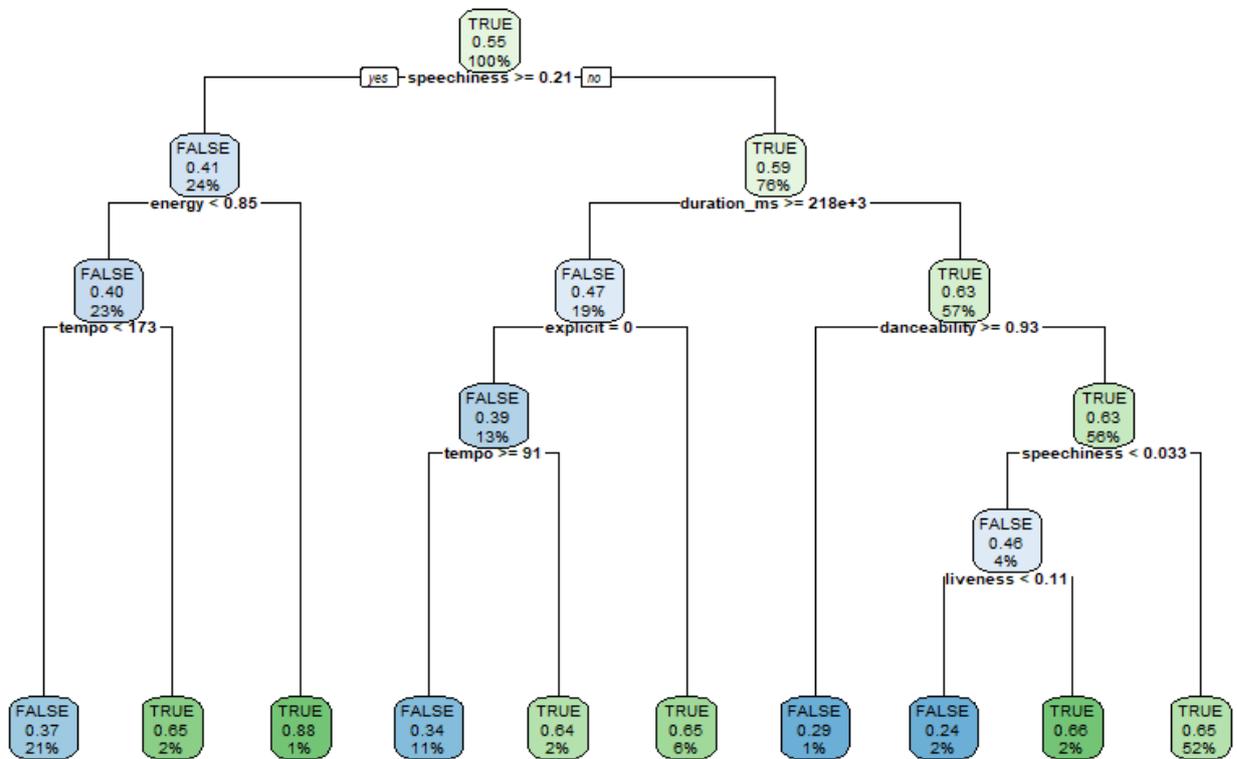
```
library(rpart)
library(rpart.plot)
set.seed(1)
classifier <- rpart(rank ~ tempo + speechiness + acousticness + danceability + duration_ms + energy + explicit + instrumentalness + key + liveness + loudness, data=data2020, method="class")
rpart.plot(classifier, cex=0.6)
```

Gambar 11 Kode Klasifikasi pada bahasa Pemrograman R

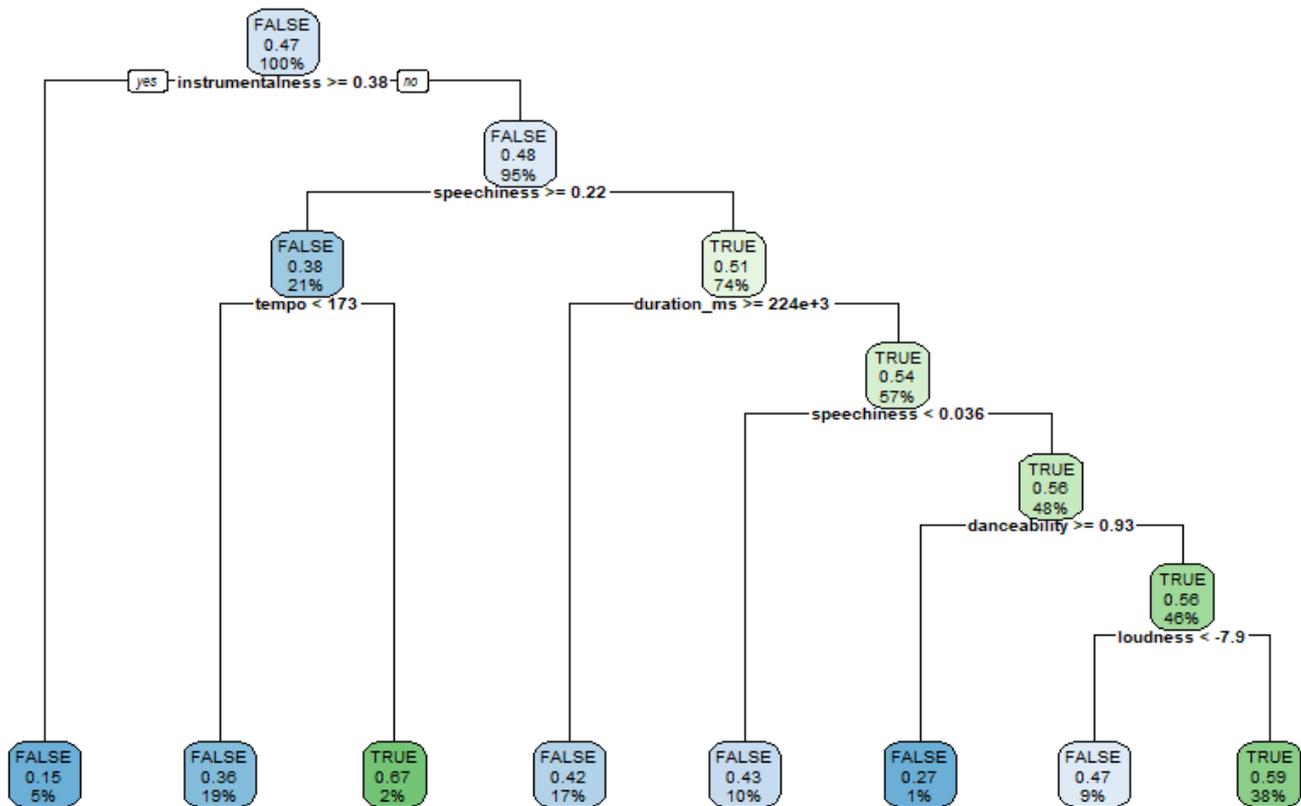
Pada kode diatas, dilakukan pengambilan data yaitu pada tahun 2020 dengan nilai popularitas lagu diatas 70. Pada kode diatas juga dilakukan pengkategorian data yaitu bernilai TRUE untuk lagu yang memiliki nilai popularitas diatas 70 dan bernilai FALSE untuk data yang memiliki nilai popularitas dibawah 70. Lalu menggunakan library *rpart* kita melakukan klasifikasi dengan fungsi *classifier* dengan menurangi rank dengan variabel karakteristik dari lagu. Hal ini dilakukan 3 kali yaitu data pada tahun 2019, tahun 2020, dan tahun 2019-2020 untuk melihat perbedaan dari tren dari pengguna dan juga jenis lagu yang populer pada setiap tahun.



Gambar 12 Pohon Keputusan untuk Popularitas lagu pada tahun 2019



Gambar 13 Pohon Keputusan untuk Popularitas lagu pada tahun 2020



Gambar 14 Pohon Keputusan untuk Popularitas lagu pada tahun 2019-2020

Algoritma classifier pada bahasa pemrograman R membuat pohon keputusan dihasilkan dengan membandingkan karakteristik dari setiap lagu untuk mencari variabel apa yang berpengaruh sehingga lagu bisa mendapat nilai popularitas diatas 70. Pada setiap pohon keputusan faktor yang mempengaruhi juga berbeda dan kita bisa melihat perbedaan bentuk pohon keputusan pada tahun 2019, 2010, dan 2019-2020.

IV. KESIMPULAN

Dari pohon keputusan diatas (Gambar 12,13,14) kita bisa melihat karakteristik lagu populer yang memiliki nilai *popularity* diatas 70 (TRUE/bewarna hijau). Dari percobaan ini penulis berharap bahwa para musisi, produser, dan label musik bisa menggunakan konsep pohon keputusan untuk menentukan jenis dan karakteristik music secara nilai kuantitatif dari setiap komponennya. Sebuah pola pasti bisa dicari untuk membuat sebuah lagu hits yang bisa menduduki peringkat tinggi pada nilai popularitas di Spotify.

Dari model klasifikasi pohon keputusan ini masih banyak hal yang bisa dikembangkan dan diperbaiki supaya hasilnya lebih akurat. Oleh karena itu, penulis akan sebisa mungkin untuk melanjutkan projek ini dilain waktu sehingga kita bisa mengetahui dengan lebih baik dan detail untuk karakteristik lagu yang bisa mendapat nilai popularitas Spotify yang tinggi. Lalu untuk menentukan standar minimum nilai popularitas masih bisa dikembangkan dengan mencari tahu data-data statistik sehingga standar minimum nilai popularitas bisa lebih jelas dan

spesifik untuk setiap lagu yang berada pada Spotify. Semoga hasil dari penelitian ini bisa menginspirasi para musisi yang ingin mencetak lagu hits sehingga bisa mencapai nilai popularitas di Spotify dan juga menginspirasi para pengembang untuk menciptakan model prediksi yang lebih baik lagi.

V. PENUTUP

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Tuhan yang Maha Esa karena dengan bantuannya penulis bisa menyelesaikan makalah ini dengan tepat waktu. Penulis juga ingin mengucapkan terima kasih sebesar-besarnya kepada Ibu Farizka Zakhralativa S.T.,M.T. dan dosen IF 2120 Matematika Diskrit lainnya yang telah mengajarkan konsep-konsep matematika diskrit sehingga penulis bisa membuat makalah berjudul “Aplikasi Pohon Keputusan dalam Klasifikasi Lagu Populer di Spotify”.

Penulis juga ingin mengucapkan banyak terima kasih kepada Ibu dari penulis yang sudah mendukung penulis selama berkuliah daring dari rumah. Semoga semua ilmu yang didapatkan penulis selama belajar mata kuliah IF 2120 dan selama menulis tugas ini bisa menjadi bekal yang baik untuk penerapan ilmu dari Teknik informatika pada dunia nyata.

REFERENSI

[1] <http://informatika.stei.itb.ac.id/~rinaldi.munir/Matdis/2020-2021/Graf-2020-Bagian1.pdf> diakses pada 10 Desember 2020

- [2] <http://informatika.stei.itb.ac.id/~rinaldi.munir/Matdis/2020-2021/Graf-2020-Bagian2.pdf> diakses pada 10 Desember 2020
- [3] <http://informatika.stei.itb.ac.id/~rinaldi.munir/Matdis/2020-2021/Graf-2020-Bagian3.pdf> diakses pada 10 Desember 2020
- [4] <https://towardsdatascience.com/decision-trees-in-machine-learning-641b9c4e8052> diakses pada 10 Desember 2020
- [5] https://www.kaggle.com/yamaerenay/spotify-dataset-19212020-160k-tracks?select=data_w_genres.csv diakses pada 10 Desember 2020
- [6] https://www.researchgate.net/publication/317731072_Popular_Decision_Tree_Algorithms_of_Data_Mining_Techniques_A_Review diakses pada 11 Desember 2020

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.

Bandung, 11 Desember 2020



Ramadhana B. Wishnumurti 13519203