

TUGAS AKHIR

ANALISIS DATA LAGU TOP 200 PADA SPOTIFY TAHUN 2020-2021



Viola Kinanti Putri Pramono

NPM: 6181901027

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2024

FINAL PROJECT

DATA ANALYSIS OF TOP 200 SONGS ON SPOTIFY 2020-2021



Viola Kinanti Putri Pramono

NPM: 6181901027

**DEPARTMENT OF INFORMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2024**

LEMBAR PENGESAHAN

**ANALISIS DATA LAGU TOP 200 PADA SPOTIFY TAHUN
2020-2021**

Viola Kinanti Putri Pramono

NPM: 6181901027

Bandung, 12 Januari 2024

Menyetujui,

Pembimbing

**Digitally signed
by Vania Natali**

Vania Natali, M.T.

Ketua Tim Penguji

**Digitally signed
by Pascal**

Alfadian Nugroho

Pascal Alfadian, Nugroho, M.Comp.

Anggota Tim Penguji

**Digitally signed
by Keenan**

Adiwijaya Leman

Keenan Adiwijaya Leman, M.T.

Mengetahui,

Ketua Program Studi

**Digitally signed
by Lionov**

Lionov, Ph.D.

PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa tugas akhir dengan judul:

ANALISIS DATA LAGU TOP 200 PADA SPOTIFY TAHUN 2020-2021

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,
Tanggal 12 Januari 2024



Viola Kinanti Putri Pramono
NPM: 6181901027

ABSTRAK

Lagu menjadi salah satu hiburan yang banyak digemari oleh masyarakat dan setiap orang memiliki selera musiknya masing-masing. Begitu juga dengan para musisi yang memiliki genre musik dan karakternya masing-masing. Oleh karena itu ada beragam lagu yang dapat dinikmati oleh masyarakat di seluruh dunia, salah satu media yang banyak digunakan sebagai layanan untuk mendengarkan lagu maupun sebagai media untuk para musisi mempublikasikan lagunya adalah Spotify. Spotify sendiri merupakan layanan musik digital yang memberikan akses ke berbagai lagu dari musisi di seluruh dunia. Spotify memiliki beberapa fitur, salah satunya adalah tangga lagu seperti *Top 50 Global*, *Top 50 Indonesia*, *Top 200 Global* dan lainnya. Bagi para musisi, artis dan pelaku bidang industri musik lainnya jika lagu yang mereka buat ataupun lagu yang mereka bawakan bisa masuk ke dalam tangga lagu Spotify tentunya akan memberikan pengaruh baik, seperti meningkatnya jumlah pendengar dan lagu yang lebih dikenal masyarakat. Namun tidak semua lagu dapat masuk ke dalam tangga lagu Spotify, oleh karena itu penelitian ini dibuat untuk mencari faktor-faktor apa saja yang dapat membuat sebuah lagu masuk dan bertahan dalam tangga lagu *Top 200 Spotify* dengan menganalisis data data lagu *Top 200 Spotify* tahun 2020-2021 dengan menggunakan teknik *data mining*.

Data mining merupakan proses untuk menemukan pola-pola menarik dan informasi yang menarik dari data dengan jumlah yang banyak. Kumpulan data yang didapatkan akan diproses sehingga dapat menghasilkan pengetahuan baru yang dapat digunakan. Sebelum melakukan pengolahan data untuk menemukan pola-pola perlu dilakukan tahapan *data science* yaitu pengumpulan data dengan mencari data-data yang diperlukan dalam penelitian ini. Penyiapan dan eksplorasi data dengan melakukan pembersihan dan mempelajari data yang akan digunakan. Analisis data dengan mengolah data dengan teknik *data mining* untuk menemukan pola-pola dan pembuatan model klasifikasi serta model *clustering*.

Pada penelitian ini dilakukan pengumpulan data utama yaitu data lagu *Top 200 Spotify* tahun 2020-2021 Global dan data tambahan yaitu data lagu *Top 200 Spotify* tahun 2020-2021 Indonesia. Kemudian dilakukan pembersihan dan penyiapan data untuk tahapan eksplorasi. Dilakukanlah eksplorasi univariat data untuk mempelajari sifat data dan eksplorasi bivariat untuk analisis lanjutan dan pemilihan atribut yang akan digunakan dalam pembuatan model klasifikasi dan model *clustering*. Penelitian ini menghasilkan informasi dari hasil analisis data yang disampaikan dalam visualisasi data dan juga model klasifikasi untuk memprediksi jumlah kemunculan sebuah lagu dalam tangga lagu *Top 200 Spotify* dengan algoritma K-Nearest Neighbor (KNN). Dibuat juga model *clustering* menggunakan algoritma K-Means untuk melihat pola-pola dalam dataset yang digunakan. Sebagai kesimpulan hasil penelitian ini mengungkapkan bahwa artis dan genre dari sebuah lagu berkorelasi dengan peringkat tertinggi dari sebuah lagu dalam tangga lagu, namun tidak berkorelasi dengan jumlah kemunculan sebuah lagu dalam tangga lagu. Kemudian dapat dilakukan prediksi jumlah kemunculan sebuah lagu dalam tangga lagu berdasarkan nilai popularitas lagu dan peringkat tertinggi yang dicapai lagu tersebut.

Kata-kata kunci: Penambangan Data, Analisis Data, Klasifikasi, *Clustering*

ABSTRACT

Songs are a form of entertainment that is popular with many people and everyone has their own taste in music. Likewise with musicians who have their own musical genres and characters. Therefore, there are various songs that can be enjoyed by people all over the world. One of the media that is widely used as a service for listening to songs and as a medium for musicians to publish their songs is Spotify. Spotify itself is a digital music service that provides access to various songs from musicians around the world. Spotify has several features, one of which is song charts such as Top 50 Global, Top 50 Indonesia, Top 200 Global and others. For musicians, artists and other players in the music industry, if the songs they compose or the songs they perform can enter the Spotify charts, it will certainly have a good impact, such as increasing the number of listeners and songs that are better known to the public. However, not all songs can enter the Spotify charts, therefore this research was created to find out what factors can make a song enter and remain in the Spotify Top 200 charts by analyzing data from the 2020 Spotify Top 200 songs. 2021 using data mining techniques.

Data mining is the process of finding interesting patterns and interesting information from large amounts of data. The data collection obtained will be processed so that it can produce new knowledge that can be used. Before carrying out data processing to find patterns, it is necessary to carry out the *data science* stage, namely data collection by searching for the data needed in this research. Data preparation and exploration by cleaning and studying the data to be used. Data analysis by processing data using *data mining* techniques to find patterns and create classification models and *clustering* models.

In this research, the main data was collected, namely the *Top 200* Spotify song data for 2020-2021 Global and additional data, namely the *Top 200* Spotify song data for 2020-2021 Indonesia. Then the data is cleaned and prepared for the exploration stage. Univariate data exploration was carried out to study the nature of the data and bivariate exploration for further analysis and selection of attributes that will be used in creating classification models and *clustering* models. This research produces information from the results of data analysis presented in data visualization and also a classification model to predict the number of times charted of a song in the *Top 200* Spotify charts using the K-Nearest Neighbor (KNN) algorithm. A *clustering* model was also created using the K-Means algorithm to see patterns in the dataset used. In conclusion, the results of this research reveal that the artist and genre of a song correlate with the highest ranking of a song in the charts, but do not correlate with the number of appearances of a song in the charts. Then you can predict the number of appearances of a song in the charts based on the song's popularity value and the highest ranking achieved by the song.

Keywords: Data Mining, Data Analysis, Classification, Clustering

*Dipersembahkan untuk keluarga dan teman-teman saya yang selalu
memberikan saya dukungan dalam berbagai keputusan berharga
dalam hidup saya...*

KATA PENGANTAR

Puji dan syukur penulis panjatkan pada Tuhan Yang Maha Esa karena atas berkat dan rahmatNya penulis berhasil menyelesaikan penyusunan tugas akhir ini yang berjudul "Analisis Data Lagu Top 200 pada Spotify Tahun 2020-2021". Tugas akhir ini dibuat sebagai salah satu syarat untuk menyelesaikan pendidikan di Program Studi Teknik Informatika, Fakultas Teknologi Informasi dan Sains, Universitas Katolik Parahyangan, Bandung. Penulis menyadari bahwa penyusunan tugas akhir ini tidak akan berhasil tanpa dukungan dan doa dari berbagai pihak, oleh karena itu penulis ingin mengucapkan terima kasih kepada:

1. Orang tua penulis yang selalu mendukung, mendoakan dan percaya pada penulis.
2. Ibu Vania Natali, M.T. yang sudah membimbing dan memberikan arahan selama proses penyusunan tugas akhir.
3. Bapak Pascal Alfadian, Nugroho, M.Comp. dan Bapak Keenan Adiwijaya Leman, M.T. yang telah memberikan saran dan masukan untuk tugas akhir ini.
4. Keluarga yang mendukung dan mendoakan penulis.
5. Teman seperjuangan selama kuliah yaitu Alma, Caca, Ira, Levi, Axel dan Nanda yang sudah menemani dan mengisi masa-masa perkuliahan penulis.
6. Teman terdekat penulis Angel, Nisrina dan Faszya selalu menyemangati penulis.
7. Seluruh pihak lain yang tidak dapat disebutkan satu per satu.

Akhir kata penulis berharap agar tugas akhir ini dapat berguna bagi seluruh pihak yang memerlukannya. Tidak lupa penulis juga memohon maaf apa bila terdapat kesalahan dan kekurangan dalam hasil penyusunan tugas akhir ini. . .

Bandung, Januari 2024

Penulis

DAFTAR ISI

KATA PENGANTAR	xv
DAFTAR ISI	xvii
DAFTAR GAMBAR	xix
DAFTAR TABEL	xxiii
DAFTAR KODE PROGRAM	xxv
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	4
1.4 Batasan Masalah	4
1.5 Metodologi	4
1.6 Sistematika Pembahasan	4
2 LANDASAN TEORI	7
2.1 <i>Data Science</i>	7
2.2 Pengukuran <i>Central Tendency</i>	9
2.2.1 <i>Mean</i>	9
2.2.2 <i>Median</i>	9
2.2.3 <i>Modus</i>	11
2.3 <i>Data Mining</i>	11
2.3.1 <i>Machine Learning</i>	12
2.3.2 Klasifikasi	13
2.3.3 <i>Clustering</i>	16
2.4 Visualisasi Data	18
2.5 Korelasi	22
2.5.1 Chi-Square	22
2.5.2 Korelasi Pearson	23
2.5.3 Korelasi Spearman	25
2.6 Python <i>Library</i>	26
2.7 Elemen dalam Musik	27
3 ANALISIS PENYELESAIAN MASALAH	29
3.1 Analisis Masalah	29
3.2 Eksperimen Penyiapan Data	29
3.3 Eksperimen Eksplorasi Data	32
3.3.1 Eksperimen Penerapan Pengukuran <i>Central Tendency</i>	32
3.3.2 Eksperimen Penerapan Visualisasi Data	33
3.4 Eksperimen Analisis Data	37

3.4.1	Eksperimen Pembuatan Model Klasifikasi	37
3.4.2	Eksperimen Pembuatan Model <i>Clustering</i>	39
4	PENAMBANGAN DATA	43
4.1	Pengumpulan Data	43
4.2	Pembersihan dan Penyiapan Data	44
4.3	Eksplorasi Data	45
4.4	Eksplorasi dan Analisis Data untuk Atribut <i>Highest Charting Position</i>	68
4.5	Pembangunan dan Evaluasi Model Klasifikasi Atribut <i>Highest Charting Position</i>	77
4.6	Eksplorasi dan Analisis Data untuk Atribut <i>Number of Times Charted</i>	82
4.7	Pembangunan dan Evaluasi Model Klasifikasi Atribut <i>Number of Times Charted</i>	92
4.8	Pembangunan Model <i>Clustering</i>	98
4.8.1	Model <i>Clustering</i> dengan Empat Atribut	99
4.8.2	Model <i>Clustering</i> dengan Atribut Elemen Musik	101
4.9	Pengumpulan Data Tambahan	105
4.10	Pembersihan dan Penyiapan Data Tambahan	106
4.11	Eksplorasi dan Analisis Data Tambahan	107
5	PELUNCURAN MODEL DAN PENGUJIAN	111
5.1	Deskripsi Perangkat Lunak	111
5.2	Rancangan Tampilan Antarmuka	111
5.3	Implementasi Tampilan Antarmuka	113
5.4	Pengujian Model pada Perangkat Lunak	116
6	KESIMPULAN DAN SARAN	119
6.1	Kesimpulan	119
6.2	Saran	121
	DAFTAR REFERENSI	123
	A KODE PROGRAM	125

DAFTAR GAMBAR

1.1	Tangga lagu Spotify	1
1.2	Spotify <i>Top 200 Global</i>	2
1.3	Spotify (a) <i>Top 200 Indonesia</i> , (b) <i>Top 200 South Korea</i> , (c) <i>Top 200 Netherlands</i>	3
2.1	Tahapan <i>data science</i>	7
2.2	Gambar a adalah data dengan distribusi <i>negatively skewed</i> , gambar b adalah data dengan distribusi normal, dan gambar c adalah data dengan distribusi <i>positively skewed</i>	10
2.3	<i>Modus</i> dalam <i>bar chart</i>	11
2.4	<i>Data mining</i> mengadopsi berbagai teknik dari berbagai domain.	12
2.5	Klasifikasi data – proses mempelajari data di mana model klasifikasi dibangun.	13
2.6	Klasifikasi data – proses klasifikasi di mana model digunakan untuk memprediksi.	13
2.7	Decision Tree	14
2.8	K-Nearest Neighbor	15
2.9	Agglomerative	17
2.10	K-Means	17
2.11	Elbow method	17
2.12	<i>Line chart</i>	18
2.13	<i>Bar chart</i> vertikal dan <i>bar chart</i> horizontal	19
2.14	histogram	19
2.15	<i>Pie chart</i>	20
2.16	<i>Scatter plot</i> dengan korelasi positif (a) dan korelasi negatif (b)	20
2.17	Contoh <i>scatter plot</i> tanpa korelasi	20
2.18	<i>Boxplot</i>	21
2.19	<i>Choropleth map</i>	21
2.20	<i>Critical values of the chi-square distribution</i>	23
3.1	Histogram atribut <i>Streams</i>	35
3.2	<i>Boxplot</i> atribut <i>Highest Charting Position</i>	35
3.3	<i>Bar chart</i> atribut <i>Artist</i>	36
3.4	<i>Scatter plot</i> atribut <i>Highest Charting Position</i> dan atribut <i>Streams</i>	36
3.5	<i>Elbow method</i> untuk model clustering	40
4.1	Contoh data dengan <i>missing value</i>	45
4.2	Top 10 nilai <i>Highest Charting Position</i> terbanyak	46
4.3	Bottom 10 nilai <i>Highest Charting Position</i> terbanyak	46
4.4	Top 10 nilai <i>Number of Times Charted</i> terbanyak	47
4.5	Bottom 10 nilai <i>Number of Times Charted</i> terbanyak	47
4.6	Boxplot untuk atribut <i>Number of Times Charted</i>	47
4.7	Histogram untuk atribut <i>Streams</i>	48
4.8	Boxplot untuk atribut <i>Streams</i>	48
4.9	Histogram untuk atribut <i>Artist Followers</i>	49
4.10	Boxplot untuk atribut <i>Artist Followers</i>	49
4.11	Boxplot untuk atribut <i>Popularity</i>	50

4.12	<i>Top 10</i> nilai <i>Popularity</i> terbanyak	51
4.13	<i>Bottom 10</i> nilai <i>Popularity</i> terbanyak	51
4.14	Lagu-lagu dengan nilai <i>popularity</i> di bawah 43	51
4.15	Lagu-lagu dengan nilai <i>popularity</i> di atas 43	52
4.16	Histogram untuk atribut <i>Danceability</i>	52
4.17	Boxplot untuk atribut <i>Danceability</i>	53
4.18	Lagu-lagu dengan nilai <i>danceability</i> di bawah 0,31	53
4.19	Lagu-lagu dengan nilai <i>danceability</i> di atas 0,8	53
4.20	Histogram untuk atribut <i>Energy</i>	54
4.21	Boxplot untuk atribut <i>Energy</i>	54
4.22	Lagu-lagu dengan nilai <i>energy</i> di bawah 0,2	55
4.23	Lagu-lagu dengan nilai <i>energy</i> di atas 0,8	55
4.24	Histogram untuk atribut <i>Loudness</i>	56
4.25	Boxplot untuk atribut <i>Loudness</i>	56
4.26	Histogram untuk atribut <i>Speechiness</i>	57
4.27	Boxplot untuk atribut <i>Speechiness</i>	57
4.28	Lagu-lagu dengan nilai <i>speechiness</i> di atas 0,5	58
4.29	Lagu-lagu dengan nilai <i>speechiness</i> di bawah 0,3	58
4.30	Histogram untuk atribut <i>Acousticness</i>	59
4.31	Boxplot untuk atribut <i>Acousticness</i>	59
4.32	Lagu-lagu dengan nilai <i>acousticness</i> di atas 0,9	60
4.33	Lagu-lagu dengan nilai <i>acoustic</i> di bawah 0,1	60
4.34	Histogram untuk atribut <i>Liveness</i>	61
4.35	Boxplot untuk atribut <i>Liveness</i>	61
4.36	Histogram untuk atribut <i>Tempo</i>	62
4.37	Boxplot untuk atribut <i>Tempo</i>	62
4.38	Histogram untuk atribut <i>Duration</i>	63
4.39	Boxplot untuk atribut <i>Duration</i>	63
4.40	Histogram untuk atribut <i>Valence</i>	64
4.41	Boxplot untuk atribut <i>Valence</i>	64
4.42	Lagu-lagu dengan nilai <i>valence</i> di bawah 0,1	65
4.43	Lagu-lagu dengan nilai <i>valence</i> di atas 0,9	65
4.44	10 artis dengan kemunculan paling banyak	66
4.45	10 artis dengan kemunculan paling sedikit	66
4.46	10 <i>genre</i> dengan kemunculan paling banyak	67
4.47	10 <i>genre</i> dengan kemunculan paling sedikit	67
4.48	<i>Scatter plot</i> untuk atribut <i>Highest Charting Position</i> dan <i>Number of Times Charted</i>	68
4.49	<i>Scatter plot</i> untuk atribut <i>Highest Charting Position</i> dan <i>Streams</i>	69
4.50	<i>Scatter plot</i> untuk atribut <i>Highest Charting Position</i> dan <i>Artist Followers</i>	69
4.51	<i>Scatter plot</i> untuk atribut <i>Highest Charting Position</i> dan <i>Popularity</i>	70
4.52	<i>Scatter plot</i> untuk atribut <i>Highest Charting Position</i> dan <i>Duration</i>	70
4.53	<i>Scatter plot</i> untuk atribut <i>Highest Charting Position</i> dan <i>Tempo</i>	71
4.54	<i>Scatter plot</i> untuk atribut <i>Highest Charting Position</i> dan <i>Energy</i>	71
4.55	<i>Scatter plot</i> untuk atribut <i>Highest Charting Position</i> dan <i>Danceability</i>	72
4.56	<i>Scatter plot</i> untuk atribut <i>Highest Charting Position</i> dan <i>top 10 Genre</i>	73
4.57	<i>Stacked bar plot</i> untuk atribut <i>Highest Charting Position</i> dan <i>top 10 Genre</i>	73
4.58	10 <i>genre</i> dengan lagu pada peringkat 1-50 terbanyak	74
4.59	10 <i>genre</i> dengan rata-rata popularitas lagu tertinggi	74
4.60	<i>Scatter plot</i> untuk atribut <i>Highest Charting Position</i> dan <i>top 15 Artist</i>	75
4.61	<i>Stacked bar plot</i> untuk atribut <i>Highest Charting Position</i> dan <i>top 10 Artist</i>	76
4.62	10 artis dengan lagu pada peringkat 1-50 terbanyak	76

4.63	10 artis dengan rata-rata popularitas lagu tertinggi	77
4.64	<i>Line chart</i> untuk akurasi model klasifikasi menggunakan algoritma KNN dengan $k = 1$ hingga $k = 100$	79
4.65	<i>Stacked bar plot</i> untuk melihat proporsi hasil prediksi dengan menggunakan model klasifikasi yang dihasilkan	80
4.66	<i>Scatter plot</i> untuk atribut <i>Number of Times Charted</i> dan <i>Highest Charting Position</i>	83
4.67	<i>Scatter plot</i> untuk atribut <i>Number of Times Charted</i> dan <i>Streams</i>	83
4.68	<i>Scatter plot</i> untuk atribut <i>Number of Times Charted</i> dan <i>Artist Followers</i>	84
4.69	<i>Scatter plot</i> untuk atribut <i>Number of Times Charted</i> dan <i>Popularity</i>	84
4.70	<i>Scatter plot</i> untuk atribut <i>Number of Times Charted</i> dan <i>Duration</i>	85
4.71	<i>Scatter plot</i> untuk atribut <i>Number of Times Charted</i> dan <i>Tempo</i>	85
4.72	<i>Scatter plot</i> untuk atribut <i>Number of Times Charted</i> dan <i>Energy</i>	86
4.73	<i>Scatter plot</i> untuk atribut <i>Number of Times Charted</i> dan <i>Danceability</i>	86
4.74	<i>Scatter plot</i> untuk atribut <i>Number of Times Charted</i> dan <i>top 10 Genre</i>	87
4.75	<i>Stacked bar plot</i> untuk atribut <i>Number of Times Charted</i> dan <i>top 10 Genre</i>	88
4.76	<i>Bar plot top 10 genre</i> dengan nilai rata-rata <i>Number of Times Charted</i> tertinggi	88
4.77	<i>Bar plot</i> jumlah lagu dari <i>top 10 genre</i> dengan nilai rata-rata <i>Number of Times Charted</i> tertinggi	89
4.78	<i>Scatter plot</i> untuk atribut <i>Number of Times Charted</i> dan <i>top 10 Artist</i>	90
4.79	<i>Stacked bar plot</i> untuk atribut <i>Number of Times Charted</i> dan <i>top 10 Artist</i>	90
4.80	<i>Bar plot top 10 artis</i> dengan nilai rata-rata <i>Number of Times Charted</i> tertinggi	91
4.81	<i>Bar plot</i> jumlah lagu dari <i>top 10 artis</i> dengan nilai rata-rata <i>Number of Times Charted</i> tertinggi	91
4.82	<i>Bar plot</i> jumlah data untuk setiap label.	93
4.83	<i>Bar plot</i> jumlah data untuk setiap label.	94
4.84	<i>Bar plot</i> jumlah data untuk setiap label.	94
4.85	<i>Line chart</i> untuk akurasi model klasifikasi menggunakan algoritma KNN dengan $k = 1$ hingga $k = 100$	96
4.86	<i>Stacked bar plot</i> untuk melihat proporsi hasil prediksi dengan menggunakan model klasifikasi yang dihasilkan	97
4.87	<i>Elbow method</i> untuk model <i>cluster</i> dengan empat atribut	99
4.88	Jumlah anggota untuk setiap <i>cluster</i>	100
4.89	Lagu-lagu dalam <i>cluster 2</i>	101
4.90	Lagu-lagu dalam <i>cluster 4</i>	101
4.91	<i>Elbow method</i> untuk model <i>cluster</i> dengan atribut-atribut elemen musik	103
4.92	Jumlah anggota untuk setiap <i>cluster</i>	104
4.93	Lagu-lagu dalam <i>cluster 3</i>	105
4.94	10 artis dengan kemunculan paling banyak di tangga lagu Indonesia	107
4.95	10 artis dengan kemunculan paling banyak di tangga lagu Global	108
4.96	Proporsi <i>highest charting postion</i> untuk <i>Top 10 artis</i> pada tangga lagu Indonesia	109
4.97	Proporsi <i>highest charting postion</i> untuk <i>Top 10 artis</i> pada tangga lagu Global	110
5.1	Rancangan halaman <i>dashboard</i>	112
5.2	Rancangan halaman peluncuran model	112
5.3	Rancangan <i>bar menu</i>	113
5.4	Halaman <i>dashboard</i> data lagu <i>Top 200 Spotify</i> tahun 2020-2021	114
5.5	Halaman <i>dashboard</i> data lagu <i>Top 200 Spotify</i> tahun 2020-2021	114
5.6	<i>Bar menu</i> untuk berpindah ke halaman lain	115
5.7	Halaman prediksi jumlah kemunculan sebuah lagu pada tangga lagu <i>Top 200 Spotify</i>	116
5.8	Contoh hasil prediksi jumlah kemunculan sebuah lagu dalam tangga lagu <i>Top 200 Spotify</i>	116

DAFTAR TABEL

2.1	<i>Confusion matrix</i>	15
2.2	Contoh <i>confusion matrix</i> untuk perhitungan akurasi, <i>precision</i> dan <i>recall</i>	16
2.3	Contoh tabel kontingensi 2x2 untuk korelasi gender dengan genre buku yang disukai.	22
2.4	Kekuatan korelasi berdasarkan nilai korelasi (r)	24
2.5	Contoh tabel data untuk perhitungan korelasi Pearson	24
2.6	Contoh tabel data untuk perhitungan korelasi Spearman	25
2.7	Hasil perhitungan d^2 untuk setiap baris	25
3.1	Contoh dataset yang masih kotor	30
3.2	Contoh dataset yang sudah bersih	30
3.3	Contoh dataset sebelum ditransformasi	31
3.4	Contoh dataset sesudah ditransformasi	32
3.5	Dataset sederhana untuk melakukan eksplorasi data dengan pengukuran <i>central tendency</i>	33
3.6	Dataset sederhana untuk melakukan eksplorasi data dengan visualisasi data	34
3.7	Potongan dataset untuk melakukan analisis data	37
3.8	Hasil perhitungan evaluasi model klasifikasi	37
3.9	Cofusion matrix untuk model klasifikasi	38
3.10	Metrik evaluasi untuk model klasifikasi	38
3.11	Hasil <i>clustering</i>	40
4.1	Tabel atribut dataset lagu <i>Top 200</i> pada Spotify tahun 2020-2021.	43
4.2	Pengukuran <i>central tendency</i> atribut <i>Highest Charting Position</i>	45
4.3	Pengukuran <i>central tendency</i> atribut <i>Number of Times Charted</i>	46
4.4	Pengukuran <i>central tendency</i> atribut <i>Number of Times Charted</i> setelah dilakukan penyesuaian nilai data	46
4.5	Pengukuran <i>central tendency</i> atribut <i>Streams</i>	48
4.6	Pengukuran <i>central tendency</i> atribut <i>Artist Followers</i>	50
4.7	Pengukuran <i>central tendency</i> atribut <i>Popularity</i>	50
4.8	Pengukuran <i>central tendency</i> atribut <i>Danceability</i>	53
4.9	Pengukuran <i>central tendency</i> atribut <i>Energy</i>	55
4.10	Pengukuran <i>central tendency</i> atribut <i>Loudness</i>	56
4.11	Pengukuran <i>central tendency</i> atribut <i>Speechiness</i>	58
4.12	Pengukuran <i>central tendency</i> atribut <i>Acousticness</i>	59
4.13	Pengukuran <i>central tendency</i> atribut <i>Liveness</i>	60
4.14	Pengukuran <i>central tendency</i> atribut <i>Tempo</i>	62
4.15	Pengukuran <i>central tendency</i> atribut <i>Duration</i>	63
4.16	Pengukuran <i>central tendency</i> atribut <i>Valence</i>	65
4.17	Tabel sebelum dilakukan <i>split</i> pada atribut <i>Artis</i>	65
4.18	Tabel setelah dilakukan <i>split</i> pada atribut <i>Artis</i>	66
4.19	Tabel sebelum dilakukan <i>split</i> pada atribut <i>Genre</i>	67
4.20	Tabel setelah dilakukan <i>split</i> pada atribut <i>Genre</i>	68
4.21	Hasil perhitungan evaluasi model klasifikasi dengan satu atribut prediktor	77

4.22 Hasil perhitungan evaluasi model klasifikasi dengan satu atribut prediktor setelah dilakukan pelabelan baru	78
4.23 Hasil perhitungan evaluasi model klasifikasi dengan dua atribut prediktor setelah dilakukan pelabelan baru	78
4.24 Hasil perhitungan evaluasi model klasifikasi dengan tiga atribut prediktor setelah dilakukan pelabelan baru	79
4.25 <i>Confusion matrix</i> untuk model klasifikasi menggunakan algoritma KNN dengan $k = 30$	80
4.26 Metrik evaluasi untuk model klasifikasi menggunakan algoritma KNN dengan $k = 14$	81
4.27 Hasil perhitungan evaluasi model klasifikasi dengan atribut <i>Artist</i> sebagai atribut prediktor	82
4.28 Contoh peringkat tertinggi lagu-lagu dari seorang artis yang tidak selalu sama . .	82
4.29 Hasil perhitungan evaluasi model klasifikasi dengan satu atribut prediktor	92
4.30 Hasil perhitungan evaluasi model klasifikasi dengan satu atribut prediktor setelah dilakukan pelabelan baru (2 label)	92
4.31 Hasil perhitungan evaluasi model klasifikasi dengan satu atribut prediktor setelah dilakukan pelabelan baru (3 label)	93
4.32 Hasil perhitungan evaluasi model klasifikasi dengan satu atribut prediktor setelah dilakukan pelabelan baru (4 label)	94
4.33 Hasil perhitungan evaluasi model klasifikasi dengan dua atribut prediktor setelah dilakukan pelabelan baru (4 label)	95
4.34 Hasil perhitungan evaluasi model klasifikasi dengan tiga atribut prediktor setelah dilakukan pelabelan baru (4 label)	96
4.35 <i>Confusion matrix</i> untuk model klasifikasi menggunakan algoritma KNN dengan $k = 33$	97
4.36 Metrik evaluasi untuk model klasifikasi menggunakan algoritma KNN dengan $k = 33$	97
4.37 Hasil <i>cluster</i> dengan atribut <i>Highest Charting Position</i> , <i>Number of Times Charted</i> , <i>Popularity</i> , dan <i>Streams</i> berupa <i>centroid</i> dan jumlah anggota tiap kelompok	99
4.38 Atribut-atribut yang dapat menggambarkan elemen musik.	102
4.39 Hasil <i>cluster</i> dengan atribut <i>Highest Charting Position</i> , <i>Number of Times Charted</i> , <i>Popularity</i> , dan <i>Streams</i> berupa <i>centroid</i> dan jumlah anggota tiap kelompok	103
4.40 Nilai atribut <i>Highest Charting Position</i> terkecil dan terbesar dalam masing-masing <i>cluster</i>	104
4.41 Tabel atribut dataset Spotify Chart	106
4.42 Contoh dataset tambahan sebelum dilakukan pembersihan	106
4.43 Contoh dataset tambahan setelah dilakukan pembersihan, tersisa data lagu dari tangga lagu <i>Top 200</i> tahun 2020-2021	107
4.44 Contoh dataset tambahan setelah dilakukan pemilihan data lagu yang masuk ke dalam tangga lagu <i>Top 200</i> tahun 2020-2021 di Indonesia	107
4.45 Contoh dataset tambahan sebelum dilakukan pemilihan nilai atribut <i>rank</i> tertinggi	108
4.46 Contoh dataset tambahan setelah dilakukan pemilihan nilai atribut <i>rank</i> tertinggi	108

DAFTAR KODE PROGRAM

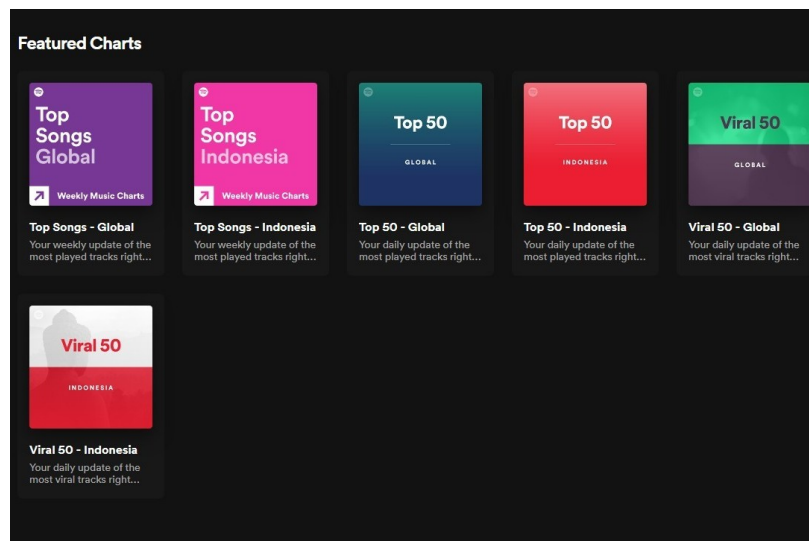
3.1	Kode untuk menghapus baris dengan nilai NaN	30
3.2	Kode untuk memecah nilai pada atribut Genre ke dalam beberapa baris	31
3.3	Kode untuk membuat histogram dari atribut Streams	34
3.4	Kode untuk membuat boxplot dari atribut Highest Charting Position	34
3.5	Kode untuk membuat boxplot dari atribut Highest Charting Position	35
3.6	Kode untuk membuat scatter plot dari atribut Highest Charting Position dan atribut Streams	36
3.7	Kode untuk membuat model klasifikasi dan akurasainya	38
3.8	Kode untuk membuat confusion matrix serta mencari nilai precision dan recall untuk model klasifikasi	39
3.9	Kode untuk membuat model clustering dan mencari nilai k terbaik	40
3.10	Kode untuk membuat model clustering dan mencari nilai centroid serta jumlah setiap cluster	41
A.1	Kode pembersihan dan penyiapan serta eksplorasi data utama	125
A.2	Kode pembuatan model klasifikasi	133
A.3	Kode pembuatan model <i>clustering</i>	140
A.4	Kode pembersihan dan penyiapan serta eksplorasi data tambahan	142
A.5	Kode pembuatan halaman <i>dashboard</i> pada perangkat lunak	143
A.6	Kode pembuatan halaman prediksi pada perangkat lunak	144
A.7	Kode untuk <i>styling</i> perangkat lunak halaman <i>dashboard</i>	145
A.8	Kode untuk <i>styling</i> perangkat lunak halaman prediksi	146

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Lagu merupakan ragam suara yang berirama atau ragam menyanyi seperti dalam membaca, bernyanyi dan musik.¹ Lagu menjadi salah satu hiburan yang banyak digemari oleh masyarakat dan setiap orang memiliki selera musiknya masing-masing. Begitu juga dengan para musisi yang memiliki genre musik dan karakternya masing-masing. Oleh karena itu ada beragam lagu yang dapat dinikmati oleh masyarakat di seluruh dunia, salah satu media yang banyak digunakan sebagai layanan untuk mendengarkan lagu maupun sebagai media untuk para musisi mempublikasikan lagunya adalah Spotify.



Gambar 1.1: Tangga lagu Spotify

Spotify adalah layanan musik digital, *podcast*, dan video dari Swedia yang memberikan akses ke berbagai lagu dan konten lain dari kreator di seluruh dunia, dengan fungsi utama memutar musik dan memberikan rekomendasi sesuai dengan selera pendengar.² Dengan perkembangan jumlah pendengar setiap tahunnya, Spotify menjadi salah satu layanan *streaming* audio yang banyak digunakan.

Spotify memiliki beberapa fitur, salah satunya adalah tangga lagu Spotify harian dan mingguan seperti *Top 50 Global*, *Top 50 Indonesia*, *Top 200 Global* dan lainnya yang dapat dilihat pada Gambar 1.1³. Tangga lagu Spotify akan menampilkan lagu-lagu yang berhasil masuk ke dalam tangga lagu mingguan Spotify. Pada tahun 2022 Spotify memiliki sekitar 82.000.000 (delapan

¹Badan Pengembangan dan Pembinaan Bahasa Kementerian Pendidikan, Budaya dan Teknologi Republik Indonesia. Lagu. <https://kbbi.kemdikbud.go.id/entri/lagu>. 23 Juni 2022.

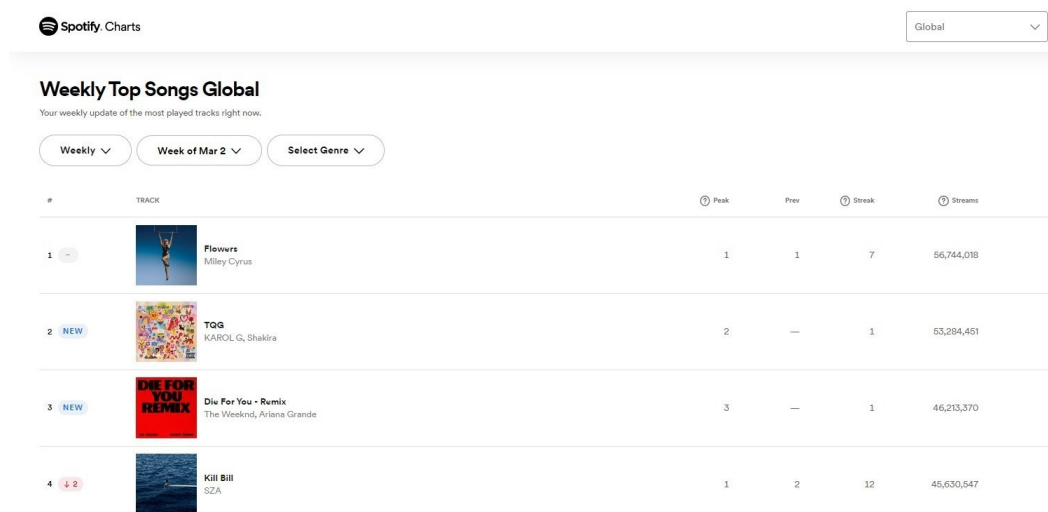
²Spotify AB. Apa itu Spotify?. <https://support.spotify.com/id/article/what-is-spotify/>. 23 Juni 2022.

³Gambar diambil dari <https://open.spotify.com/genre/browse-charts-tab> (2022)

puluh dua juta) lagu dengan 11.000.000 (sebelas juta) artis dan kreator. Dengan beragam lagu, *genre*, artis dan banyaknya pendengar pada Spotify, setiap minggunya peringkat pada tangga lagu Spotify pun dapat berubah-ubah menampilkan berbagai lagu populer seperti pada Gambar 1.2⁴ yang menampilkan beberapa lagu dari tangga lagu mingguan *Top 200 Global* pada tanggal 24 Februari 2023 hingga 2 Maret 2023. Dari tangga lagu pada minggu tersebut didapati 4 lagu pada peringkat teratas yaitu:

- Miley Cyrus - *Flowers*
- Karol G, Shakira - *TQG*
- The Weeknd, Ariana Grande - *Die For You - Remix*
- SZA - *Kill Bill*

Selain tangga lagu Global, Spotify juga memiliki tangga lagu untuk suatu negara seperti tangga lagu mingguan *Top 200 Indonesia*, *Top 200 South Korea*, dan *Top 200 Netherlands* yang menampilkan lagu-lagu yang banyak didengarkan dinegara tersebut seperti pada Gambar 1.3⁵



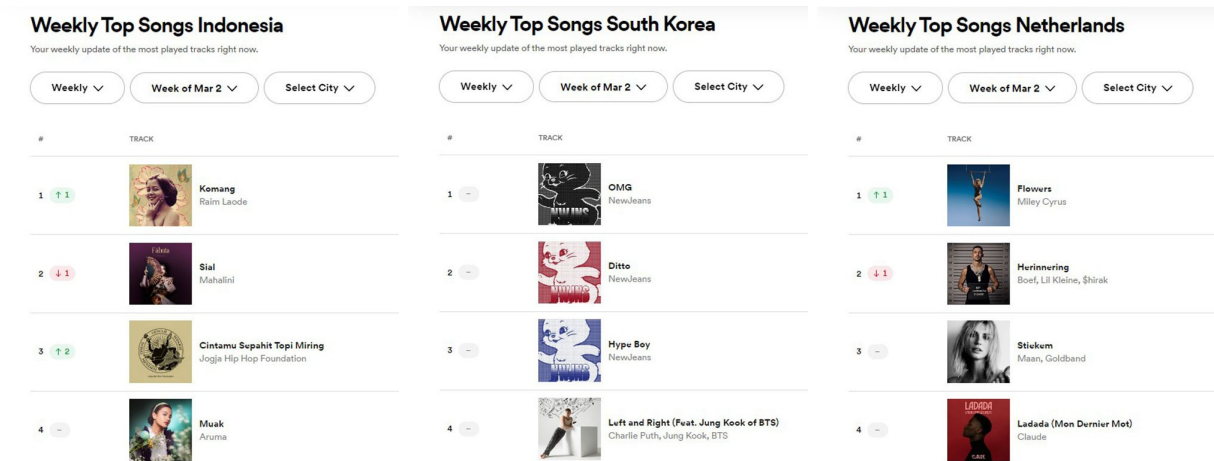
#	TRACK	Peak	Prev	Streak	Streams
1	<i>Flowers</i> Miley Cyrus	1	1	7	56,744,018
2	<i>TQG</i> KAROL G, Shakira	2	—	1	53,284,451
3	<i>Die For You - Remix</i> The Weeknd, Ariana Grande	3	—	1	46,213,370
4	<i>Kill Bill</i> SZA	1	2	12	45,630,547

Gambar 1.2: Spotify *Top 200 Global*

Ada berbagai hal yang dapat menjadi faktor untuk membuat sebuah lagu bisa masuk ke dalam tangga lagu Spotify, seperti popularitas artis, jumlah pengikut artis, jumlah pendengar lagu, genre lagu, dan lainnya. Bagi para musisi, artis, dan pelaku bidang industri musik lainnya jika lagu yang mereka buat ataupun lagu yang mereka bawakan bisa masuk ke dalam tangga lagu Spotify, tentunya akan memberikan pengaruh baik. Lagu-lagu pada tangga lagu akan mendapatkan jumlah pendengar yang lebih banyak, sehingga baik lagu, artis dan musisi akan lebih dikenali oleh pendengar. Lalu dengan terus bertambahnya jumlah pengguna Spotify, masuk ke dalam tangga lagu Spotify tentu menjadi salah satu cara untuk meningkatkan jumlah pendengar. Karena tangga lagu Spotify menjadi salah satu cara untuk para pengguna Spotify mencari dan mengikuti perkembangan lagu-lagu terbaru dan terbaik.

⁴Gambar diambil dari <https://charts.spotify.com/charts/view/regional-global-weekly/2023-03-02> (2023)

⁵Gambar (a) di ambil dari <https://charts.spotify.com/charts/view/regional-id-weekly/2023-03-02>, Gambar (b) diambil dari <https://charts.spotify.com/charts/view/regional-kr-weekly/2023-03-02>, Gambar (c) diambil dari <https://charts.spotify.com/charts/view/regional-nl-weekly/2023-03-02> (2023)



Gambar 1.3: Spotify (a) *Top 200 Indonesia*, (b) *Top 200 South Korea*, (c) *Top 200 Netherlands*

Namun tidak semua lagu dapat masuk ke dalam tangga lagu Spotify, oleh karena itu penelitian ini dibuat untuk mencari faktor-faktor apa saja yang dapat membuat sebuah lagu masuk dan bertahan dalam tangga lagu *Top 200* Spotify, dengan menganalisis data lagu-lagu *Top 200* Spotify tahun 2020-2021 menggunakan *data mining*. Penelitian ini akan menggunakan data lagu *Top 200* Spotify 2020-2021 yang didapatkan dari Kaggle yang diunggah oleh user Sashank Pillai. Kaggle sendiri merupakan sebuah web yang memberikan akses kepada penggunanya untuk mencari berbagai dataset yang dapat digunakan dalam kegiatan *data science*.⁶ Dataset yang digunakan pada penelitian ini memiliki 23 atribut seperti judul lagu, nama artis, jumlah pengikut artis, *genre*, posisi tertinggi pada tangga lagu, dan jumlah suatu lagu masuk ke dalam tangga lagu. Data yang digunakan akan diolah dengan menggunakan teknik *data mining* seperti *clustering* dan regresi, untuk dapat menemukan pola-pola atau *trend* dari data lagu *Top 200* Spotify tahun 2020-2021. Melalui data yang diolah dan pola atau *trend* yang ditemukan, bisa didapatkan informasi-informasi seperti ciri-ciri lagu yang berhasil masuk ke dalam *Top 200* Spotify, fitur yang berpengaruh untuk meningkatkan kemungkinan sebuah lagu masuk ke dalam *Top 200* Spotify, artis yang lagunya paling banyak masuk ke dalam *Top 200* Spotify, dan negara-negara yang banyak menyumbangkan lagu dalam *Top 200* Spotify.

Pada penelitian ini akan dibuat model untuk menganalisis data lagu *Top 200* Spotify tahun 2020-2021 untuk melihat pola atau *trend* dari data yang digunakan. Kemudian untuk mempermudah pengguna dalam menangkap informasi, hasil dari analisis data yang dilakukan akan ditampilkan dengan menggunakan visualisasi data.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, maka rumusan masalah dari penelitian ini adalah sebagai berikut:

1. Bagaimana menyiapkan data berupa data lagu *Top 200* pada Spotify tahun 2020-2021?
2. Pola atau *trend* apa saja yang dapat ditemukan pada data lagu *Top 200* pada Spotify tahun 2020-2021?
3. Bagaimana mengevaluasi pola atau *trend* lagu *Top 200* pada Spotify tahun 2020-2021 yang ditemukan?
4. Bagaimana menampilkan pola atau *trend* lagu *Top 200* pada Spotify tahun 2020-2021 ke dalam perangkat lunak?

⁶Mahmoud, H. M. (2022) *What is a Kaggle?*. <https://www.kaggle.com/general/328265>. 11 Maret 2023

1.3 Tujuan

Tujuan dari penelitian ini adalah sebagai berikut:

1. Menyiapkan data berupa data lagu *Top 200* pada Spotify tahun 2020-2021.
2. Mencari pola atau *trend* lagu *Top 200* pada Spotify tahun 2020-2021 dengan menggunakan teknik *data mining*.
3. Mengevaluasi pola atau *trend* lagu *Top 200* pada Spotify tahun 2020-2021.
4. Menampilkan pola atau *trend* lagu *Top 200* pada Spotify tahun 2020-2021 ke dalam perangkat lunak.

1.4 Batasan Masalah

Batasan-batasan masalah yang diambil dalam melakukan penelitian ini adalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini diambil dari situs penyedia dataset Kaggle.

1.5 Metodologi

Metodologi yang dilakukan untuk melakukan penelitian ini adalah sebagai berikut:

1. Melakukan studi literatur
Studi literatur dilakukan untuk memahami teori-teori yang digunakan dalam penelitian yaitu, teknik *data mining* seperti *clustering* dan regresi untuk menemukan pola atau *trend* dari data.
2. Menganalisis metode dan data
Analisis dilakukan untuk mengetahui metode apa yang paling tepat untuk digunakan dalam tahapan-tahapan analisis data, mulai dari mempelajari data, pembersihan dan penyiapan data, integrasi data, analisis data, hingga evaluasi hasil analisis data. Analisis juga dilakukan pada data-data yang akan digunakan untuk mengenali sifat dari data, agar data dapat diolah dan digunakan dengan tepat dalam penelitian.
3. Mencari pola atau *trend*
Pencarian pola atau *trend* pada data dilakukan sebagai bentuk analisis data yang dilakukan terhadap data yang digunakan. Untuk mendapatkan pola atau *trend* dari data, akan dipilih fitur-fitur yang sesuai untuk digunakan dalam pembuatan model.
4. Menguji model dan analisis hasil
Pengujian dilakukan agar model yang dibuat dapat menghasilkan hasil analisis data berupa pola atau *trend* yang terbaik. Model yang dibuat akan diuji, dievaluasi dan dimodifikasi, dimana proses tersebut bisa dilakukan berulang kali untuk bisa mendapat hasil terbaik.
5. Menulis dokumen
Penulisan dokumen tugas akhir yang terdiri dari bab-bab utama tugas akhir ini.

1.6 Sistematika Pembahasan

Sistematika penulisan pada penelitian ini adalah sebagai berikut:

1. Bab Pendahuluan:
Pada bab ini dibahas latar belakang dibuatnya tugas akhir ini, rumusan masalah, tujuan, batasan masalah, metodologi dan sistematika pembahasan.
2. Bab Landasan Teori:
Pada bab ini dibahas berbagai teori yang dibutuhkan sebagai dasar dari pengerjaan tugas akhir ini, yaitu mengenai *data science*, pengukuran *central tendency*, *data mining*, visualisasi data, korelasi, Python *library* yang digunakan dalam penelitian ini, dan elemen dalam musik.
3. Bab Analisis Penyelesaian Masalah:
Pada bab ini dibahas mengenai eksperimen pengerjaan tahapan *data science* dengan berbagai Python *library* menggunakan dataset berukuran kecil.

4. Bab Penambangan Data:

Pada bab ini dibahas proses pengolahan data mulai dari pengumpulan data, pembersihan dan penyiapan data, eksplorasi data, analisis data, hingga pembuatan model klasifikasi dan model *clustering*.

5. Bab Peluncuran Model dan Pengujian:

Pada bab ini dibahas penampilan hasil analisis dan peluncuran model pada perangkat lunak.

6. Bab Kesimpulan dan Saran:

Pada bab ini dibahas kesimpulan dari seluruh proses penelitian yang dilakukan serta saran untuk pengembangan penelitian ini.