## 3.1 Dataset

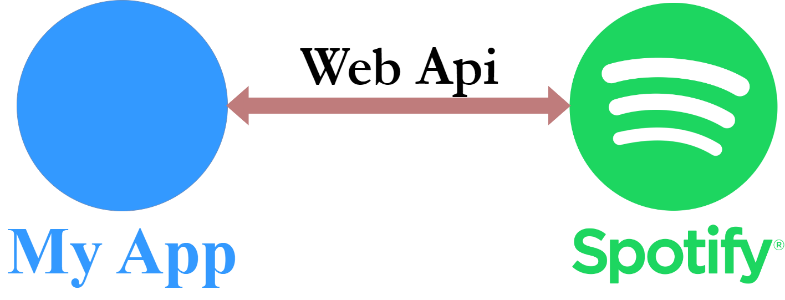
Veri setinin hazırlığı ihtiyacınız olan verileri toplama, okunabilir ve anlaşılır formata getirme ve kontrol etmek büyük önem taşımaktadır. Doğru bir veri seti hem zaman ve emek tasarrufuna hem de daha doğru ve verimli algoritmaların gelişiminde büyük rol oynamaktadır.

Projemizde kullanacağımız veri seti farklı çalma listelerinden oluşan 3783 şarkıdan oluşmaktadır. Veri setinin oluşturulması için Spotify Web API’nin bize sağladığı geliştirici platformu kullanılmaktadır.

API bir uygulama için programlama ara yüzü olarak tanımlanmaktadır. Şirketlerin program verilerini ve işlevlerini, dışarıdan üçüncü taraf geliştiricilere, çalışma ortaklarına ve şirketlerindeki dahili departmanlara paylaşılmasını sağlamaktadır. Yazılımların bir diğeri ile iletişim kurmasına ve birbirlerinin verilerinden ve işlevlerinden yararlanmasına olanak tanımaktadır. Programcıların yazılım geliştirmek için harcadıkları çaba ve zaman azalmaktadır [9].

Basit REST ilkelerine dayanan Spotify Web API uç noktaları, doğrudan spotify veri kataloğundan müzik sanatçıları, albümler ve parçalar hakkında json meta verilerini döndürür [10].

Spotify'ın Web API'sinin piyasaya sürülmesinden bu yana, birçok yetenekli içerik oluşturucunun geliştirici platformunu kullanarak birçok içerik üretmiştir. Spotify tarafından hesap hizmeti aracılığı ile spotify geliştirici hesabı sahibi olunması gerekmektedir. Spotify kullanıcılara şarkı önerirken, şarkılara bazı özellikler atar. Python programlama dili yardımı ile web api sayesinde veri seti oluşturulmuştur.



**Şekil 3.1.1 Api Bağlantı Şeması**

Oluşturduğumuz veri seti 18 sütundan oluşmaktadır. Bunlar, şarkının tanımlayıcısı, şarkı url, şarkı adı, sanatçı adı, key, mode, danceability, energy, loudness, acousticness, speechiness, instrumentalness, liveness, tempo, valance, duration (ms), time signature, popularity. Bu özellikler Tablo 3.1.1 de gösterilmiştir [11].

**Tablo** 3.1.**1**

|  |  |  |
| --- | --- | --- |
| Özellik | Veri Türü | Tanım |
| key | integer | Parçanın içinde bulunduğu anahtar. Tamsayılar, standart Pitch Class (Perde Sınıfı) gösterimini kullanarak perdelere eşlenir. Örneğin. 0 = C, 1 = C♯/D♭, 2 = D, vb. Bir anahtar algılanmadıysa, değer -1'dir. |
| mode | integer | Mod, bir parçanın kipliğini (majör veya minör), melodik içeriğinin türetildiği gam türünü belirtir. Majör 1 ile temsil edilir ve minör 0'dır. |
| danceability | float | Tempo, bir parçanın dans edilme potansiyelini tanımlayan bir ölçümdür. Bu ölçüm, ritm istikrarı, vuruş gücü ve genel düzenlilik gibi müzik öğelerinin bir kombinasyonuna dayanır. Ölçüm aralığı 0 ila 1 arasındadır, 0 değeri en az dans edilebilir olarak ve 1 değeri ise en yüksek dans edilebilirlik potansiyeli olarak kabul edilir. |
| energy | float | Enerji, bir parçanın algısal yoğunluk ve aktivite seviyesini ölçen bir parametredir. Bu ölçüm 0 ile 1 arasında bir skala üzerinde yapılır. Yüksek enerjili parçalar, hızlı, yüksek sesli ve gürültülü hissettirirken, düşük enerjili parçalar daha yavaş, sessiz ve sakin hissettirir. Örneğin, bir heavy metal parçası yüksek enerjiye sahipken, bir Bach başlangıcı düşük enerjiye sahip olacaktır. Enerji özelliğinin belirlenmesinde kullanılan algısal özellikler arasında dinamik aralık, algılanan ses yüksekliği, tını, başlangıç hızı ve genel entropi yer alır. |
| loudness | float | Ses yüksekliği, bir parçanın genel ses seviyesini ölçen bir parametredir. Bu ölçüm desibel (dB) cinsinden yapılır. Bir parçanın ses yüksekliği değerleri, parçanın tüm uzunluğu boyunca ortalaması alınarak hesaplanır ve diğer parçalarla karşılaştırmak için kullanılır. Ses yüksekliği, bir sesin fiziksel gücünün (genlik) psikolojik karşılığıdır. Değerler tipik olarak -60 ile 0 dB arasında değişir. |
| acousticness | float | Akustiklik, bir parçanın akustik olup olmadığının ölçüldüğü bir parametredir. Bu ölçüm 0 ile 1 arasında bir skala üzerinde yapılır. 1.0 değeri parçanın tamamen akustik olduğuna dair yüksek bir güveni ifade eder. |
| speechiness |  | Konuşmanın varlığı, bir parçada algılanan bir özniteliktir. Bu öznitelik, parçada konuşulan kelimelerin varlığını ölçer. Öznitelik değeri, kayıtların konuşma içeriğine göre değişebilir. Örneğin, talk show, sesli kitap veya şiir gibi kayıtlar için, öznitelik değeri 1.0'a yakın olacaktır. 0,66 ve üzeri değerler, muhtemelen tamamen konuşulan sözcüklerden oluşan parçaları tanımlar. 0,33 ile 0,66 arasındaki değerler, rap müzik gibi durumlar da dahil olmak üzere, bölümler halinde veya katmanlı olarak hem müzik hem de konuşma içerebilen parçaları tanımlar. 0,33 ve altındaki değerler ise büyük olasılıkla sadece müzik içeren parçaları temsil eder. Bu öznitelik, parçaların konuşma içeriği, türü veya amacı gibi faktörlere göre değişebilir. |
| instrumentalness | float | Vokallik, bir parçada vokal içerip içermediğinin tahmin edildiği bir özniteliktir. Bu öznitelik, parçada seslendirilen kelime veya sesleri değerlendirir. Örneğin, 'ooh' ve 'aah' gibi sesler, parçanın enstrümantal olduğuna dair yüksek bir güveni temsil ederken, rap veya konuşulan kelime parçaları açıkça 'vokal' olarak değerlendirilir. Enstrümentallik değeri 1.0'a ne kadar yakınsa, parçanın vokal içerik içermeme olasılığı o kadar yüksektir. 0,5'in üzerindeki değerlerin enstrümantal izleri temsil etmesi amaçlanmıştır, ancak bu değer 1.0'a yaklaştıkça güven daha yüksektir. |
| liveness | float | Canlılık, bir parçanın canlı olarak kaydedilip kaydedilmediğini algıladığı bir özniteliktir. Canlılık değeri, parçanın kayıt ortamına göre değişebilir. Örneğin, bir konser kaydı genellikle daha yüksek bir canlılık değerine sahip olacaktır. Daha yüksek canlılık değerleri, parçanın canlı olarak gerçekleştirilme olasılığının artmasını temsil eder. 0,8'in üzerindeki bir değer, parçanın canlı olarak kaydedildiğine dair güçlü bir olasılık sağlar. |
| tempo | float | Tempo, bir parçanın dakikada kaç vuruş yaptığını ölçen bir özelliktir. Tempo, müzik terminolojisinde parçanın hızı olarak tanımlanır ve dakikada gerçekleşen vuruş sayısı (BPM) cinsinden ölçülür. |
| valence | float | Bir parçanın müzikal olarak ifade ettiği pozitifliği ölçen 0.0 ile 1.0 arasında bir özellik. Yüksek değerli parçalar daha olumlu ve pozitif duyguları yansıtırken (örneğin mutlu, neşeli, öforik) düşük değerli izler daha olumsuz ve negatif duyguları yansıtmaktadır (örneğin üzgün, depresif, kızgın). |
| duration\_ms | integer | Bir parçanın süresini milisaniye cinsinde belirtir. |
| time\_signature | integer | Bir parçanın ritmik öğelerini tanımlayan tahmini bir zaman imzası. Zaman imzası, parçanın kaç çubukta (veya ölçüde) kaç vuruş olduğunu belirtir ve genellikle "3/4" veya "4/4" gibi notasyonlarla gösterilir. |

Kaynaklar

1. Ofoeda, J., Boateng, R., & Effah, J. (2019). Application Programming Interface (API) Research. *International Journal of Enterprise Information Systems*, *15*(3), 76–95. https://doi.org/10.4018/IJEIS.2019070105
2. Schettino, V. J., Braga, R., David, J. M. N., & Araújo, M. A. P. (2017). Spotify characterization as a software ecosystem. *Proceedings of the 11th Brazilian Symposium on Software Components, Architectures, and Reuse - SBCARS ’17*, 1–10. https://doi.org/10.1145/3132498.3133836
3. Al-Beitawi, Z., Salehan, M., & Zhang, S. (2020). What makes a song trend? Cluster analysis of musical attributes for Spotify top trending songs. Journal of Marketing Development and Competitiveness, 14(3), 79-91.

Terimler sözlüğüne = integer, float, url, ms, api