

MESTRADO
GESTÃO COMERCIAL

SPREADING THE WORD EM REDE: O CASO DA PROPAGAÇÃO DE HITS NO SPOTIFY

Inês Maria Moreira Terrinha

M

2024



SPREADING THE WORD EM REDE:

O CASO DA PROPAGAÇÃO DE HITS NO SPOTIFY

Inês Maria Moreira Terrinha

Dissertação

Mestrado em Gestão Comercial

Orientado por
Professor Pedro Campos

2024

Resumo

Dada a robusta economia que envolve a Indústria da Música e a transformação digital por que esta tem passado na última década, é crucial que todos os *stakeholders* compreendam as novas dinâmicas de mercado e os novos comportamentos dos consumidores. Assim, há a necessidade de músicos e *Record Labels* saberem como desenvolver uma estratégia eficaz nesta era dominada por grandes *players* do *streaming*, como o Spotify.

Tal como outros produtos, a música espalha-se dentro das redes sociais por meio da difusão e do contágio. Estes conceitos ganham ainda mais relevância, à medida que as interações sociais são cada vez mais facilitadas graças à *social media* e a fenómenos como o *Electronic Word of Mouth* (EWOM).

Assim, no decorrer desta tese, foi analisada a influência das redes sociais no consumo de música, tanto ao nível das ações dos utilizadores, como ao nível das músicas que ouvem, nomeadamente no caso dos *hits*. Para além disso, foram analisados os contextos em que os indivíduos ouvem música, e ainda o respetivo impacto no comportamento do consumidor e nos seus hábitos de audição.

Para atingir os objetivos mencionados, várias técnicas de análise de dados foram desenvolvidas: *k-means clustering* para categorizar as músicas, cálculo da *weighted closeness* para analisar redes, regressões para analisar a interação entre a centralidade dos indivíduos nas redes e os hábitos de audição de *hits*, e, por fim, foram apurados três modelos para avaliar o que influencia os utilizadores a tomarem determinadas decisões: *Fast Large Margin*, Árvores de Decisão e *Neural Networks*. Foram utilizadas as plataformas Excel, RapidMiner e R para as análises descritas, tendo por base uma amostra do maior *dataset* já disponibilizado pelo Spotify.

Com este trabalho, foi possível entender melhor o consumo de música no mundo do *streaming*, tendo inclusive em conta diferentes estilos musicais e, assim, desenvolver observações importantes que podem ser benéficas para artistas e *Record Labels* no que diz concerne à promoção, *marketing*, distribuição e estratégia comercial.

Palavras-chave: Indústria da Música – Spotify – Redes Sociais – Difusão – Estratégia Comercial – Comportamento do Consumidor – Análise de Dados

Abstract

Given the strong economy involving the Music Industry and the digital transformation that it has been undergoing for the past decade, it is crucial that all stakeholders understand the new market dynamics and consumer behaviours. Thus, there is an urgent need for musicians and record labels to know how to strategize effectively in an era dominated by key streaming players like Spotify.

Similarly to other products, music spreads within social networks through diffusion and contagion. These concepts are even more relevant today, as social interactions have become easier due to social media and phenomena like Electronic Word of Mouth (EWOM).

Therefore, in this thesis, the influence of social networks on music consumption was analyzed, both from the perspective of users' actions and the music they listen to, particularly in the case of hits. Additionally, the contexts in which users listen to music and their impact on consumer behavior and listening habits were also explored.

To achieve the mentioned objectives, various data analytics techniques were applied: k-means clustering to categorize music tracks, weighted closeness calculation to analyze networks, regressions to study the interaction between network centrality and the habit of listening to hits, and three models to assess what influences users to take specific actions: Fast Large Margin, Decision Trees, and Neural Networks. The platforms Excel, RapidMiner, and R were used for these analyzes conducted on a sample of the largest dataset ever published by Spotify.

With this work, it was possible to better understand music consumption in the Spotify environment across different musical genres and, ultimately, provide managerial recommendations beneficial for artists and music labels concerning promotion, marketing, distribution, and commercial strategy.

Keywords: Music Industry – Spotify – Social Networks – Diffusion – Commercial Strategy – Consumer Behaviour – Data Analytics

Índice

1.	Introdução.....	1
2.	Enquadramento Teórico	3
2.1.	Música: indivíduo e sociedade	3
2.2.	Difusão e contágio em redes.....	4
2.3.	A indústria da música na era digital	5
2.4.	Spotify: funcionamento, <i>features</i> e papel social	7
2.5.	Os novos hábitos de consumo e dinâmicas de mercado	10
2.6.	Os novos processos de tomada de decisão na era do <i>streaming</i>	12
3.	Questões de Investigação, Dados e Metodologia.....	13
3.1.	Questões de Investigação	13
3.1.1.	De que forma o consumo de música varia consoante o contexto?.....	13
3.1.2.	Qual a relação entre redes sociais de indivíduos e o surgimento de <i>hits</i> ?	13
3.1.3.	A centralidade dos utilizadores influencia-os na escolha entre ter uma experiência guiada pelo Spotify ou mais autónoma?	14
3.2.	Bases de Dados.....	14
3.3.	Metodologia.....	18
3.3.1.	<i>Clustering</i> das faixas musicais	18
3.3.2.	Redes de <i>users</i> no Spotify.....	19
3.3.3.	Algoritmos para classificação das experiências de consumo no Spotify	20
4.	Resultados	22
4.1.	Resultados Obtidos	22
4.1.1.	<i>Clustering</i> das Faixas Musicais.....	22
4.1.2.	Hábitos de consumo nos diferentes contextos Spotify.....	25
4.1.3.	Audição de <i>Hits</i>	29
4.1.4.	Influência da Centralidade dos <i>Users</i> na audição de <i>Hits</i>	30
4.1.5.	Determinantes da experiência de consumo	32
4.2.	Interpretação dos Resultados.....	36
5.	Considerações Finais	39
5.1.	Limitações.....	39
5.2.	Conclusão	39
6.	Referências Bibliográficas	42
7.	Anexo.....	49

7.1.	Processos informáticos (RapidMiner e R)	49
7.2.	Distribuição dos dados	51
7.3.	Modelos desenvolvidos para classificação das experiências de consumo	51

Índice de Figuras

Figura 1 – Receita Total da Indústria Fonográfica Mundial (fonte autora: adaptado de (IFPI, 2024))	5
Figura 2 – Serviços prestados pelas Record Labels (fonte autora: adaptado de (IFPI, 2024))	6
Figura 3 – Market Shares do mercado de Global Music Streaming no terceiro trimestre de 2023 (fonte autora: adaptado de MIDiA.....	7
Figura 4 – Distribuição da Acousticness nos 10 Clusters.....	22
Figura 5 – Distribuição da Danceability nos 10 Clusters.....	22
Figura 6 – Distribuição da Energy nos 10 Clusters.....	22
Figura 7 – Distribuição da Instrumentalness nos 10 Clusters	22
Figura 8 – Distribuição da Liveness nos 10 Clusters.....	23
Figura 9 – Distribuição da Loudness nos 10 Clusters.....	23
Figura 10 – Distribuição da Speechiness nos 10 Clusters	23
Figura 11 – Distribuição do Tempo nos 10 Clusters	23
Figura 12 – Distribuição da Valence nos 10 Clusters.....	23
Figura 13 – Distribuição da US_popularity_estimate nos 10 Clusters	23
Figura 14 – Distribuição da Key nos 10 Clusters	23
Figura 15 – Distribuição da Time_signature nos 10 Clusters	24
Figura 16 – Distribuição do Mode nos 10 Clusters.....	24
Figura 17 – Distribuição das audições por cluster nos 6 contextos.....	29
Figura 18 – Distribuição da US_popularity_estimate nos 6 contextos	29
Figura 19 – Gráfico de dispersão (centralidade dos users e hits que ouvem) do user_fav_cluster_1.....	30
Figura 20 – Gráfico de dispersão (centralidade dos users e hits que ouvem) do user_fav_cluster_2.....	30
Figura 21 – Gráfico de dispersão (centralidade dos users e hits que ouvem) do user_fav_cluster_3.....	30
Figura 22 – Gráfico de dispersão (centralidade dos users e hits que ouvem) do user_fav_cluster_4.....	30
Figura 23 – Gráfico de dispersão (centralidade dos users e hits que ouvem) do user_fav_cluster_5.....	31

Figura 24 – Gráfico de dispersão (centralidade dos users e hits que ouvem) do user_fav_cluster_6.....	31
Figura 25 – Gráfico de dispersão (centralidade dos users e hits que ouvem) do user_fav_cluster_7.....	31
Figura 26 – Gráfico de dispersão (centralidade dos users e hits que ouvem) do user_fav_cluster_8.....	31
Figura 27 – Gráfico de dispersão (centralidade dos users e hits que ouvem) do user_fav_cluster_9.....	31
Figura 28 – Gráfico de dispersão (centralidade dos users e hits que ouvem) do user_fav_cluster_10.....	31
Figura 29 – Classificação das faixas em estilos musicais (KNN): processo em RapidMiner	49
Figura 30 – Clustering das faixas (K-Means Clustering): processo em RapidMiner	49
Figura 31 – Árvore de Decisão do user_fav_cluster_1.....	51
Figura 32 – Árvore de Decisão do user_fav_cluster_3.....	53
Figura 33 – Árvore de Decisão do user_fav_cluster_4.....	54
Figura 34 – Árvore de Decisão do user_fav_cluster_5.....	55
Figura 35 – Árvore de Decisão do user_fav_cluster_6.....	57
Figura 36 – Árvore de Decisão do user_fav_cluster_7.....	59
Figura 37 – Árvore de Decisão do user_fav_cluster_8.....	61
Figura 38 – Árvore de Decisão do user_fav_cluster_9.....	62
Figura 39 – Árvore de Decisão do user_fav_cluster_10	63

Índice de Tabelas

Tabela 1 – Distribuição de users e logs por user_fav_cluster	20
Tabela 2 – Distribuição das Faixas por Clusters (número e respetiva percentagem de faixas face ao total)	24
Tabela 3 – Distribuição de todos os logs por cluster e is_fav_cluster (número de logs e respetiva percentagem face ao total)	26
Tabela 4 – Distribuição dos Logs em catalog por cluster e is_fav_cluster (número de logs e respetiva percentagem face ao total)	26
Tabela 5 – Distribuição dos logs em charts por cluster e is_fav_cluster (número de logs e respetiva percentagem face ao total)	26
Tabela 6 – Distribuição dos logs em editorial_playlist por cluster e is_fav_cluster (número de logs e percentagem face ao total)	27
Tabela 7 – Distribuição dos logs em personalized_playlist por cluster e is_fav_cluster (número de logs e respetiva percentagem face ao total)	27
Tabela 8 – Distribuição dos logs em radio por cluster e is_fav_cluster (número de logs e respetiva percentagem face ao total)	28
Tabela 9 – Distribuição dos logs em user_collection por cluster e is_fav_cluster (número de logs e respetiva percentagem face ao total)	28
Tabela 10 – Métricas de desempenho dos modelos desenvolvidos para o user_fav_cluster_1	32
Tabela 11 – Métricas de desempenho dos modelos desenvolvidos para o user_fav_cluster_3	32
Tabela 12 – Métricas de desempenho dos modelos desenvolvidos para o user_fav_cluster_4	33
Tabela 13 – Métricas de desempenho dos modelos desenvolvidos para o user_fav_cluster_5	33
Tabela 14 – Métricas de desempenho dos modelos desenvolvidos para o user_fav_cluster_6	34
Tabela 15 – Métricas de desempenho dos modelos desenvolvidos para o user_fav_cluster_7	34
Tabela 16 – Métricas de desempenho dos modelos desenvolvidos para o user_fav_cluster_8	34
Tabela 17 – Métricas de desempenho dos modelos desenvolvidos para o	

user_fav_cluster_9.....	35
Tabela 18 – Métricas de desempenho dos modelos desenvolvidos para o	
user_fav_cluster_10.....	35
Tabela 19 – Código R para obter a weighted_closeness dos indivíduos a partir da matriz da	
distribuição dos logs dos users por cluster	50
Tabela 20 – Distribuição de todos os logs por context_type e not_skipped.....	51
Tabela 21 – Neural Network do user_fav_cluster_1.....	51
Tabela 22 – Neural Network do user_fav_cluster_3.....	53
Tabela 23 – Neural Network do user_fav_cluster_4.....	54
Tabela 24 – Neural Network do user_fav_cluster_5.....	55
Tabela 25 – Neural Network do user_fav_cluster_6.....	57
Tabela 26 – Neural Network do user_fav_cluster_7.....	59
Tabela 27 – Neural Network do user_fav_cluster_8.....	61
Tabela 28 – Neural Network do user_fav_cluster_9.....	62
Tabela 29 – Neural Network do user_fav_cluster_10.....	63

Índice de Equações

Equação 1 – Fast Large Margin do user_fav_cluster_1.....	51
Equação 2 – Fast Large Margin do user_fav_cluster_3.....	53
Equação 3 – Fast Large Margin do user_fav_cluster_4.....	54
Equação 4 – Fast Large Margin do user_fav_cluster_5.....	55
Equação 5 – Fast Large Margin do user_fav_cluster_6.....	57
Equação 6 – Fast Large Margin do user_fav_cluster_7.....	59
Equação 7 – Fast Large Margin do user_fav_cluster_8.....	61
Equação 8 – Fast Large Margin do user_fav_cluster_9.....	62
Equação 9 – Fast Large Margin do user_fav_cluster_10.....	63

1. Introdução

Nos últimos 20 anos, tem-se assistido a uma transformação digital do consumo de *media* e cultura, com uma crescente tendência de adesão a serviços de *streaming* como o Spotify na música, a Netflix no entretenimento ou o Kindle na literatura (Colbjørnsen, 2021). Com a mudança, deram-se profundas alterações nestas cadeias de distribuição, nas condições de acesso e nos hábitos de consumo (Colbjørnsen, 2021). A indústria da música foi a primeira a fazer esta transformação e é hoje aquela onde os novos modelos de negócio, dinâmicas de mercado e plataformas *online* de acesso estão mais consolidadas e estabelecidas, sendo assim um bom caso de estudo capaz de elucidar o futuro de outros setores, onde estas mudanças não atingiram ainda o seu expoente máximo (Maasø, 2018; Prey et al., 2022).

A música ultrapassa já largamente o campo da arte, envolvendo uma importante e lucrativa indústria que, em 2018, impactou o PIB norte-americano em mais de 170 mil milhões de dólares (RIAA, 2020) e o PIB da União Europeia em mais de 80 mil milhões de euros (Economics, 2020). Com tal dinamismo, é importante assegurar que os diversos *players* e *stakeholders* do setor se sabem adaptar e conseguem suceder num mundo dominado pelo *streaming*, especialmente pelo Spotify, plataforma líder de mercado (Maasø & Hagen, 2020).

No Spotify, o consumo é marcado pela existência de um algoritmo capaz de oferecer recomendações, cruzando experiências dos utilizadores e de outros indivíduos semelhantes e também pelo domínio das *playlists* (sejam estas criadas pelo Spotify ou por utilizadores) (Aguilar & Waldfogel, 2021). Como em outros meios culturais, também no seio da música as redes de utilizadores, a difusão e a definição da identidade são fatores marcantes e exponeciados pela *internet* e redes sociais (Hagen & Lüders, 2017; Park, Lee, et al., 2022; Ren & Kauffman, 2017).

Nesta dissertação, será estudado o consumo de músicas em plataformas de *streaming*, com o objetivo de entender o que influencia os consumidores a tomar determinadas decisões no seu consumo, quer ao nível das ações tomadas, quer ao nível das músicas ouvidas.

Para tal, será explorada a maior e mais completa base de dados já disponibilizada pelo Spotify: The Music Streaming Sessions Dataset, MSSD (Brost et al., 2019), por forma a entender como varia o consumo dos utilizadores consoante as condições e contextos do consumo, bem como o impacto das redes e centralidades dos indivíduos no consumo de músicas, especialmente no caso de *hits*. Por fim, serão analisadas as ações dos utilizadores, numa perspetiva de perceber se os seus posicionamentos nas redes os influenciam a deixarem-se ser guiados pelo Spotify na sua experiência e consumos ou, pelo contrário, têm uma experiência

mais autónoma e proativa, procurando eles próprios por faixas e artistas ou até construindo novas *playlists*. Com este estudo, será possível responder às seguintes questões de investigação: “De que forma o consumo de música varia consoante o contexto?”, “Qual a relação entre redes sociais de indivíduos e o surgimento de hits?” e “A centralidade dos utilizadores influencia-os na escolha entre ter uma experiência guiada pelo Spotify ou mais autónoma?”. Para desenvolver os pontos anteriormente apresentados, serão, entre outros, calculadas métricas relativas às redes e aplicados os métodos *Fast Large Margin*, Árvores de Decisão e *Neural Networks*. Para o desenvolvimento destas metodologias serão utilizadas as plataformas R, Excel e RapidMiner.

Todas as conclusões apuradas nesta dissertação são de particular interesse para artistas e *Record Labels*, na medida em que quão melhor entenderem os seus públicos e hábitos de procura e consumo, mais capazes estarão de tomar decisões estratégicas e de atingir o sucesso, nomeadamente ao nível do lançamento de novas músicas, um momento crucial para atingir o sucesso e fazer da música um *hit*.

Esta dissertação é constituída por 5 capítulos. No Capítulo 2, será desenvolvido o enquadramento teórico baseado na revisão da literatura existente e no Capítulo 3, serão apresentadas as questões de investigação, a metodologia e os dados utilizados. No Capítulo 4 serão apurados e discutidos os resultados e, por fim, o Capítulo 5 constituirá a conclusão.

2. Enquadramento Teórico

Neste capítulo é feito um enquadramento teórico do tema dividido nas seguintes secções: Música: Indivíduo e Sociedade; Difusão e Contágio em Redes; A Indústria da Música na era digital; Spotify: funcionamento, *features* e papel social; Os novos hábitos de consumo e dinâmicas de mercado e Os novos processos de tomada de decisão na era do *Streaming*.

2.1. Música: indivíduo e sociedade

O papel e relevância da cultura nas sociedades é imenso: Macdonald (1953) explica que, no decorrer do século XIX, com o surgimento de democracias e com uma maior escolarização da sociedade, a cultura deixou de se destinar apenas à elite, passando a ser “única e exclusivamente para o consumo de massas, tal como a pastilha elástica” (Macdonald, 1953). Surgiram assim fenómenos de sucesso à escala global e celebridades seguidas e idolatradas pelas massas, abrindo a porta ao estabelecimento de novos modelos de negócio em torno do *mainstream* ou *kitsch* (termo alemão para definir cultura de massas), sustentadas na produção em série de livros, discos, ... e numa sociedade cada vez mais exigente nos seus consumos culturais (Macdonald, 1953). Assim, esta transformação culminou numa homogeneização cultural, sem nunca alterar o papel preponderante da cultura na identidade dos povos e sociedades (Mcrobbie, 1999; Williams, 1958).

Também na música esta massificação assumiu importantes contornos através da *Pop Culture*, intensificando o seu papel enquanto meio de expressão individual e de comunidades (Hesmondhalgh, 2013; Mcrobbie, 1999; Williams, 1958). Na verdade, hoje a relevância da música ultrapassa o campo da arte e o seio da intimidade, assumindo contornos políticos (são exemplo as inúmeras músicas de denúncia e intervenção), sociais (movimentos nacionalistas como o *Cool Britannia* inspiraram-se largamente na música *pop*), económicos (em 2018, a indústria musical da União Europeia representava dois milhões de postos de trabalho e contribuía com 81,9 mil milhões de euros para o PIB dos países constituintes da UE), etc. (Economics, 2020; Hesmondhalgh, 2013; Mcrobbie, 1999).

É relevante considerar o papel social da música, enquanto facilitador para a expressão de ideias, emoções e afetos, motor para a formação de comunidades, instigador ao desenvolvimento de relações interpessoais, etc. (De Smet & Dhaenens, 2022; Hagen & Lüders, 2017; Hosey et al., 2019; Leong & Wright, 2013; Maasø, 2018; Park, Lee, et al., 2022; Park, Redmond, et al., 2022; Siles et al., 2019). No fundo, a música serve de mote à construção social de uma identidade coletiva na qual os indivíduos se pretendem integrar, levando-os a

criar a sua própria identidade pessoal não apenas numa perspetiva de expressão da imagem que têm de si próprios, mas também numa tentativa de gestão das impressões que pretendem causar e do *status*, posicionamento e popularidade que pretendem atingir (Hagen & Lüders, 2017; Leong & Wright, 2013; Park, Lee, et al., 2022; Park, Redmond, et al., 2022). Quer isto dizer que a música é um importante ativo para os relacionamentos e interações sociais, podendo não só ser um tópico de conversa ou elemento de aproximação entre indivíduos, especialmente quando na forma de partilha de recomendações e/ou descobertas, mas constituindo também o mote para grandes eventos culturais e sociais como concertos ou festivais (Lee et al., 2019; Park & Kaneshiro, 2021; Trocchia et al., 2002). Assim, interesses partilhados por um determinado artista, banda ou género musical servem frequentemente de base para a formação e a afiliação a comunidades (sejam estas pequenos e coesos grupos de amigos ou grandes movimentos sociais como os *punks* ou o *hip-hop*) (Brooks & Conroy, 2010; Clark & Lonsdale, 2022; Lee et al., 2019; Park, Lee, et al., 2022; Prey, 2018; Trocchia et al., 2002).

2.2. Difusão e contágio em redes

A forma como um produto, imagem, música, ... se espalha entre os indivíduos, pode ser explicada à luz dos modelos epidemiológicos: partindo de um primeiro indivíduo “infetado” por uma ideia ou prática, esta espalha-se por entre complexas redes de pessoas integradas numa determinada comunidade ou envolvendo diferentes comunidades interligadas (Kumar & Sinha, 2021; Valente & Vega Yon, 2020). Este processo é conhecido como Difusão de Informação ou Difusão de Inovações e é altamente dependente dos *opinion leaders*, os primeiros “infetados”: indivíduos altamente influenciadores sobre a sua vasta rede de conhecimentos e conexões, mas também bastante inovadores e atentos a novidades, por isso *Early Adopters*, e que partilham ativamente estas descobertas com outros (Litterio et al., 2017; Tepper & Hargittai, 2009).

Com a Internet, as redes foram exponenciadas e a difusão tornou-se bem mais eficaz, com a informação a fluir mais rapidamente e a tornar-se acessível a um público cada vez maior, que já não precisa sequer de se conhecer entre si (Huete-Alcocer, 2017; Kumar & Sinha, 2021). Com isto, as próprias marcas têm vindo a investir cada vez mais no *marketing* digital, tirando usufruto das dinâmicas online, nomeadamente o EWOM, *Electronic Word of Mouth* (partilha pública e *online* de opiniões ou informações sobre um determinado assunto através de, por exemplo, publicações, partilhas ou comentários) (Gulati & Eirinaki, 2019; Huete-Alcocer, 2017; Tepper & Hargittai, 2009; Trocchia et al., 2002; Verma & Yadav, 2021; Zhang & Negus,

2021; Zhao et al., 2020).

Assim, a *internet*, o EWOM e a difusão alteraram as dinâmicas de procura e consumo, na medida em que trouxeram poder aos consumidores: as decisões de consumo são cada vez mais ponderadas, informadas e inteligentes, havendo inclusivamente uma procura ativa dos indivíduos por recomendações e *online reviews* (Litterio et al., 2017; Verma & Yadav, 2021).

2.3. A indústria da música na era digital

Atualmente, a música envolve toda uma importante indústria que, em 2023, gerou mais de 26.2 mil milhões de dólares em receitas, valor que tem vindo a aumentar constantemente nos últimos nove anos (IFPI, 2024), como se pode constatar na Figura 1.

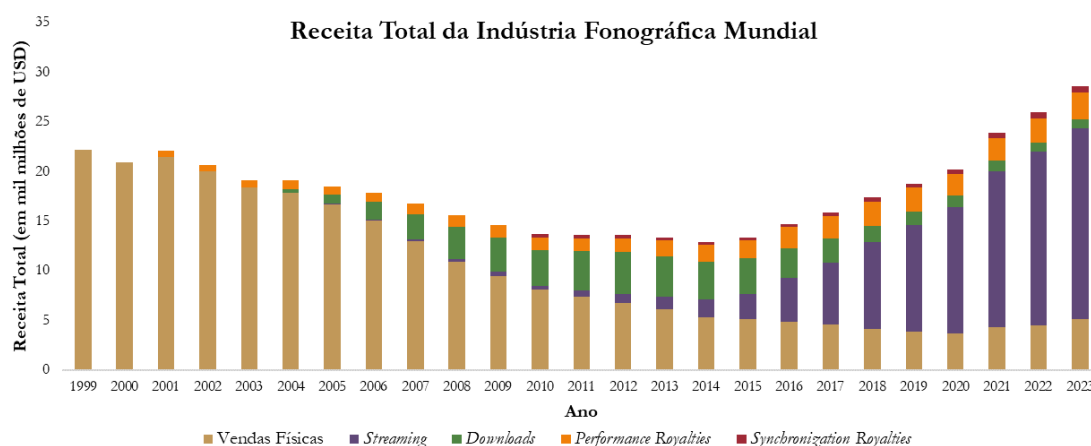


Figura 1 – Receita Total da Indústria Fonográfica Mundial (fonte autora: adaptado de (IFPI, 2024))

É possível verificar que as receitas se distribuem entre vendas físicas de álbuns, vendas *online* de música (*downloads*), recebimentos provenientes das plataformas de *streaming* (conforme as audiências), *Performance Royalties* (pagos pelo direito da reprodução da música publicamente, por exemplo, numa rádio) e *Synchronization Royalties* (pagos pelo licenciamento para sincronização com conteúdo visual em produções como filmes, programas de televisão ou anúncios) (IFPI, 2024).

Esta indústria é dominada pelas *Record Labels* ou Agências Discográficas: empresas *experts* na gestão das carreiras dos artistas, impulsionando as suas vendas e resultados, determinando as melhores estratégias comerciais, desenvolvendo planos de *marketing* e comunicação, etc. (IFPI, 2024). Esta é, hoje, uma indústria altamente competitiva, onde o sucesso comercial de um artista envolve avultados investimentos (anualmente, são investidos 3.2 mil milhões de dólares em ações de *marketing* para promoção dos artistas e respetivo catálogo e 3.9 mil milhões de dólares em A&R, *Artists and Repertoire*, a atividade responsável pela pesquisa,

angariação e desenvolvimento dos artistas) (Harrison, 2021; IFPI, 2024; Negus, 2019). A atividade destas empresas envolve uma complexa e variada rede de serviços prestados, que se encontram estruturados na Figura 2, que segue a linha temporal normal: depois de uma *label* estabelecer parceria com o artista, passa ao desenvolvimento criativo e, após o lançamento das músicas, fomenta o sucesso comercial (Harrison, 2021; IFPI, 2024; Negus, 2019; Towse, 2020).

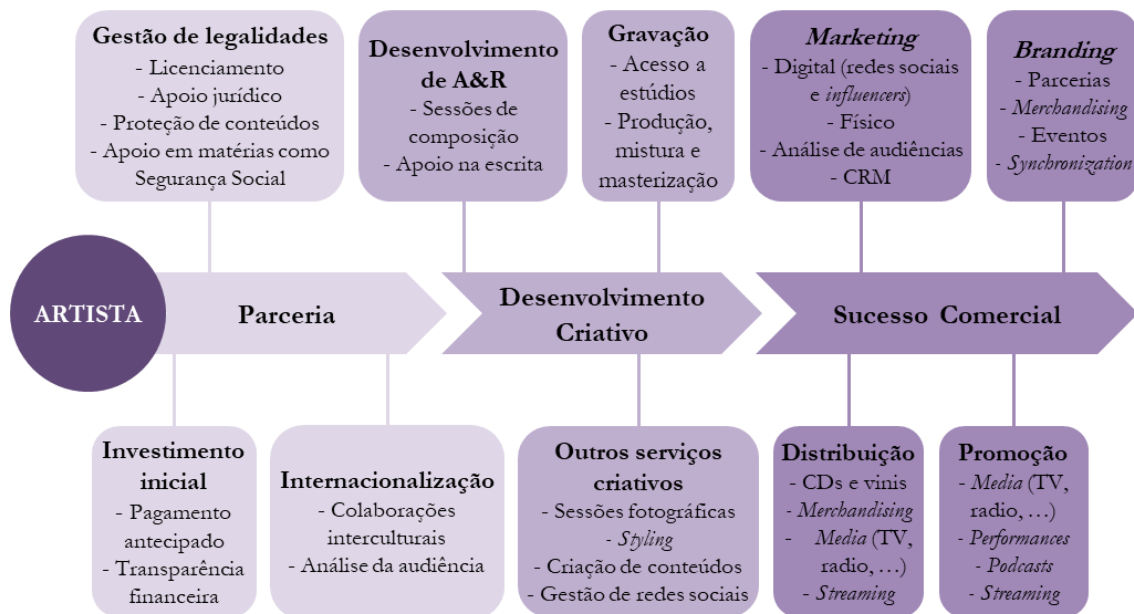


Figura 2 – Serviços prestados pelas Record Labels (fonte autora: adaptado de (IFPI, 2024))

As *Record Labels* podem ser de dois tipos: *Major Labels*, que integram os três grandes e históricos grupos multinacionais líderes de mercado, *the Big Three* (Universal Music Group, Sony Music Entertainment e Warner Music Group) e são geralmente associadas ao meio *mainstream*, ou *Independent Labels*, que são as restantes empresas: organizações mais pequenas, que gerem menos artistas e que são frequentemente especializadas em determinados estilos ou até nichos, apostando na inovação e disrupção (Mall, 2018; Prey et al., 2022).

Na Figura 2, pode-se verificar que, em meados da década de 2010, depois de um declínio de mais de dez anos, as receitas da indústria musical recuperaram fruto do aumento das receitas em plataformas de *streaming* (contrastando com a contínua diminuição das vendas físicas) (Harrison, 2021; IFPI, 2024; Kaimann et al., 2021). As **Error! Reference source not found.** plataformas de *streaming* de música surgiram ao longo das décadas de 2000 e 2010 e transformaram disruptivamente o panorama do mercado musical: hoje os seus utilizadores têm acesso a um consumo *on demand* de uma enorme variedade e diversidade de opções, resultando numa maior democratização e facilidade no consumo (Freeman et al., 2022;

Maasø & Hagen, 2020; Tepper & Hargittai, 2009; Towse, 2020). Com um modelo de negócio *freemium* (de utilização gratuita, mas com determinadas funcionalidades a serem apenas acessíveis mediante o pagamento de uma *fee* (Freeman et al., 2022; Hujran et al., 2020; Maasø & Hagen, 2020; Towse, 2020)), estas plataformas contam com mais de 700 milhões de utilizadores¹ e constituem a maior fonte de receitas da indústria fonográfica, assumindo-se assim como a “nova rádio”: influenciam descobertas, determinam padrões de consumo e originam sucessos (Harrison, 2021; Morgan, 2020; Prey et al., 2022). Desta forma, o *streaming* é hoje o mais importante eixo estratégico desta indústria, pelo que é importante que seja profundamente entendido, monitorizado e capitalizado pelos vários *players* do setor (Harrison, 2021; Hujran et al., 2020; Kaimann et al., 2021).

2.4. Spotify: funcionamento, *features* e papel social

Atualmente, a plataforma líder em *global streaming* é o Spotify, com mais de 626 milhões de utilizadores, dos quais 246 são *premium*, e mais de 100 milhões de músicas².

Com um vasto espólio de opções musicais à sua disposição, os utilizadores veem as suas decisões de consumo serem dificultadas, requerendo, assim, alguma orientação (*online* ou *offline*), nomeadamente no momento da descoberta (Kaimann et al., 2021). Por este motivo, o Spotify apresenta um sistema de recomendação automatizado e personalizado, e que sugere músicas com base nos hábitos de audição dos consumidores, comparando-os com ou-

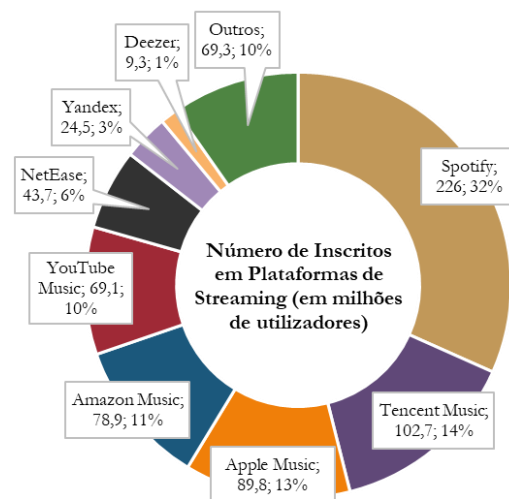


Figura 3 – Market Shares do mercado de Global Music Streaming no terceiro trimestre de 2023 (fonte autora: adaptado de MIDiA¹)

tros de hábitos semelhantes (De Assunção & Zaina, 2022; De Smet & Dhaenens, 2022; Freeman et al., 2022; Hosey et al., 2019; Hujran et al., 2020; Kaimann et al., 2021; Lyu, 2021; Morgan, 2020; Prey, 2018; Snickars, 2017). O Spotify mantém o funcionamento do seu algoritmo envolvido em bastante secretismo e confidencialidade, mas, ainda assim, sabemos que

¹ <https://www.midiaresearch.com/blog/music-subscriber-market-shares-2023-new-momentum>

² <https://newsroom.spotify.com/company-info/>

através de técnicas de *Machine Learning* e Inteligência Artificial, este algoritmo recolhe dados do utilizador para lhe apresentar sugestões alinhadas com os seus interesses e preferências. Ele é particularmente interessante, na medida que, em vez de utilizar dados como idade ou género para caracterizar os utilizadores, recorre ao *Collaborative Filtering* para identificar grupos de utilizadores com consumos e gostos semelhantes, baseando-se em dados que vão desde os hábitos de consumo até às fotografias tiradas por um indivíduo ou à sua localização (Hujran et al., 2020). Por outro lado, o algoritmo tira igualmente partido das tecnologias de *Music Information Retrieval*, por forma a caracterizar as diferentes músicas, com base em metadados como *audio features*, e, assim, identificar também músicas semelhantes entre si (De Assunção & Zaina, 2022; Petridis et al., 2022; Prey, 2018).

No Spotify, é possível criar *playlists* compostas por diferentes faixas, sendo tanto possível um indivíduo construir e ouvir as suas próprias *playlists*, como ouvir as criadas por outros utilizadores (De Smet & Dhaenens, 2022; Kaimann et al., 2021; Park, Redmond, et al., 2022; Siles et al., 2019). Existem também *playlists* criadas pelo próprio Spotify: as *playlists* editoriais, criadas por editores especialistas em curadoria musical, e as *playlists* personalizadas, criadas pelo algoritmo e customizadas especialmente para o utilizador (De Smet & Dhaenens, 2022; Eriksson, 2020; Freeman et al., 2022; Hosey et al., 2019; Morgan, 2020; Park & Kaneshiro, 2021; Prey, 2018; Razlogova, 2020; Snickars, 2017; Villermet et al., 2021).

Para além disso, existem na plataforma os *Charts*: os famosos *TOPs* de popularidade, onde figuram os maiores *hits* do momento, podendo ser organizados por estilo, país de origem, período de referência, etc. (Aguiar & Waldfogel, 2021; Kaimann et al., 2021; Maasø & Hagen, 2020). A presença nos *charts* beneficia os artistas, trazendo os ouvintes dos *hits* para o seu restante repertório, dado todo o *buzz* que é gerado (Kaimann et al., 2021; Maasø & Hagen, 2020), e atraindo novos públicos, levando a uma maior penetração no mercado (Kaimann et al., 2021; Maasø & Hagen, 2020). Percebe-se, por isso, que figurar num *chart* não é meramente um barómetro de *performance* comercial, mas também um veículo de promoção, dado o destaque e a acrescida visibilidade que lhes são dados (Aguiar & Waldfogel, 2021; Kaimann et al., 2021).

Os utilizadores podem ainda conceber o seu próprio repositório, guardando artistas, músicas, faixas ou álbuns, ouvir em “modo livre”, aventurando-se pelo vasto catálogo disponível ou deixar-se guiar totalmente pelo algoritmo, ouvindo o Radio (Lüders, 2021; Snickars, 2017; Villermet et al., 2021).

Assim, o Spotify surge como um espaço onde os seus utilizadores conseguem construir a sua

tão desejada identidade musical, sendo que a plataforma se assemelha até a uma rede social: apesar de os conteúdos serem publicados e detidos pelos artistas, os utilizadores conseguem levar a cabo uma autêntica curadoria musical, através da construção da sua própria biblioteca, constituída pelas *playlists* guardadas e repositório formado (De Smet & Dhaenens, 2022; Lüders, 2021). Para além disso, há ainda lugar a uma certa interação social na própria plataforma: através dos seguidores (função que permite saber o que outros utilizadores estão a ouvir naquele momento), dos perfis (onde se pode aceder aos artistas e músicas mais ouvidos por um utilizador, bem como às suas *playlists*) e das *playlists* colaborativas (*playlists* criadas em conjunto por vários utilizadores) (De Smet & Dhaenens, 2022; Hagen & Lüders, 2017; Hujran et al., 2020; Park & Kaneshiro, 2021; Park, Redmond, et al., 2022; Petridis et al., 2022; Prey, 2018; Siles et al., 2019).

São também frequentes as partilhas de conteúdos Spotify noutras plataformas, destacando-se os exemplos claros que atestam a importância dada pelos utilizadores à identidade musical construída: *dating apps* como o Tinder ou o Bubble dão um destaque proeminente no perfil do utilizador aos hábitos de consumo no Spotify (De Smet & Dhaenens, 2022); desde 2016, o Spotify sumariza o ano decorrido de cada utilizador, indicando-lhe as músicas, géneros e artistas mais ouvidos, gerando anualmente uma campanha viral nas redes sociais: só em 2021, 120 milhões de utilizadores acederam ao seu *Spotify Wrapped* e, destes, 60 milhões partilharam os resultados nas redes sociais sendo que, só no TikTok, vídeos com a *hashtag* #SpotifyWrapped acumulam já 66.5 mil milhões de visualizações³ (Habeeb, 2024). Assim, mais uma vez, entende-se que a importância da identidade musical ultrapassa o indivíduo: existe a necessidade de partilhar os hábitos, gostos e descobertas musicais com outras pessoas (De Smet & Dhaenens, 2022; Lee et al., 2019; Li & Liu, 2020; Park & Kaneshiro, 2021; Park, Lee, et al., 2022; Villermet et al., 2021), prática exponenciada pelas redes sociais, onde o EWOM tanto pode ser uma partilha com um único amigo como com uma toda rede de seguidores (De Smet & Dhaenens, 2022; Hagen & Lüders, 2017; Li & Liu, 2020; Park & Kaneshiro, 2021; Park, Lee, et al., 2022; Park, Redmond, et al., 2022; Ren & Kauffman, 2017).

Conclui-se, então, que a jornada pelo Spotify pode ser assim de dois tipos: fruto de uma orientação do próprio Spotify (*radio*, *charts*, *playlists* editoriais e *playlists* personalizadas)

³ <https://www.forbes.com/sites/conormurray/2023/11/28/spotify-wrapped-2023-comes-soon-heres-how-it-became-a-viral-and-widely-copied-marketing-tactic/>

(Anderson et al., 2020; Hosey et al., 2019; Kaimann et al., 2021) ou numa experiência mais autónoma, guiada pelo próprio utilizador, baseando-se nos seus gostos ou deixando-se influenciar por fatores externos: a sua rede, filmes, séries, eventos, redes sociais, rádio, etc. (Anderson et al., 2020; Lüders, 2021; Maasø, 2018), tarefa facilitada por recursos como o Shazam ou o Google Sound Search, que permitem identificar instantaneamente uma música através de um simples clique (Hujran et al., 2020).

2.5. Os novos hábitos de consumo e dinâmicas de mercado

Apesar da qualidade do algoritmo e de toda a diversidade de opções de escolha que figura no Spotify, *charts*, *playlists* personalizadas, *playlists* editoriais e radio tendem a concentrar músicas populares ou que o algoritmo prevê que se tornarão populares, dado possuírem características alinhadas com as preferências gerais dos utilizadores, o que cria dificuldades para os consumidores que desejam descobrir músicas de nicho ou fora do *mainstream*, além de limitar a criatividade e o trabalho das *Record Labels* (Aguiar & Waldfogel, 2021; Freeman et al., 2022; Hujran et al., 2020; Kaimann et al., 2021; Maasø & Hagen, 2020; Prey, 2018; Villermet et al., 2021). Por outro lado, o algoritmo é bastante fiel aos hábitos dos utilizadores, fazendo, por isso, sugestões que os mantêm na sua “bolha”, em detrimento de lhes trazer diversidade, aspeto desejado por maioria dos utilizadores (Freeman et al., 2022; Kaimann et al., 2021; Mok et al., 2022; Petridis et al., 2022; Prey, 2018). Assim, os indivíduos acabam por alternar entre ambas as experiências, tirando o máximo partido do Spotify e “educando” também o seu algoritmo (Anderson et al., 2020; Prey et al., 2022; Villermet et al., 2021). No entanto, a experiência guiada pela plataforma mantém uma significativa preponderância, o que confere ao Spotify poder sobre a procura e as decisões de consumo dos utilizadores e torna os artistas e *Record Labels* de certa forma dependentes do Spotify (Aguiar & Waldfogel, 2021; Eriksson, 2020; Morgan, 2020; Prey et al., 2022; Razlogova, 2020; Villermet et al., 2021).

Os utilizadores recorrem assim frequentemente às suas redes para descobrir novas músicas, artistas, géneros, etc., valorizando estas recomendações, na medida em conhecem a fonte e sentem alguma identificação com ela: trate-se de amigos capazes de dar conselhos pessoais ou até *influencers* com quem partilhem preferências (Hagen & Lüders, 2017; Hosey et al., 2019; Lüders, 2021; Park & Kaneshiro, 2021; Park, Lee, et al., 2022; Towse, 2020). Autores como Park, Lee, et al. (2022) ou Park and Kaneshiro (2021) aferiram inclusive já que as recomendações de pessoas acabam mesmo por gerar mais atenção e dedicação comparativamente às

do algoritmo, aumentando a probabilidade em serem bem-sucedidas (Jun et al., 2015; Lee et al., 2019; Park, Lee, et al., 2022). Neste âmbito, os *opinion leaders* e *innovators* destacam-se claramente, sendo capazes de sugerir músicas de nicho ou menos populares, de apresentarem diversidade e até de desafiar as preferências dos seus pares com sugestões mais “fora da caixa” (Lüders, 2021; Park & Kaneshiro, 2021; Park, Lee, et al., 2022; Petridis et al., 2022; Tepper & Hargittai, 2009).

Um interessante exemplo do efeito de contágio entre pares no seio da indústria musical foi investigado por Ternovski and Yasseri (2020) que aferiram que um concerto não só aumenta as audições do próprio artista nos indivíduos que assistiram ao evento, como também gera um impacto adicional de mais 5% de audições deste artista em cada indivíduo pertencente às redes da assistência (Ternovski & Yasseri, 2020).

Neste seguimento, as *playlists* de utilizadores são importantes veículos para a descoberta e difusão de músicas, sabendo que os indivíduos podem ouvir e guardar as *playlists* da sua rede, com destaque para as dos referidos *opinion leaders* e *innovators* (De Smet & Dhaenens, 2022; Hagen & Lüders, 2017; Park & Kaneshiro, 2021; Prey, 2018; Prey et al., 2022; Tepper & Hargittai, 2009).

Para além disso, as *playlists* constituem hoje um importante meio de expressão de ideias, preferências e emoções com um vasto alcance: não é sequer necessário um indivíduo conhecer o criador da *playlist* para a ouvir: basta que quando se deparar com ela (pesquisando uma determinada palavra, por exemplo), se sinta identificado e representado (De Smet & Dhaenens, 2022; Hagen & Lüders, 2017; Lüders, 2021; Negus, 2019; Park & Kaneshiro, 2021; Park, Redmond, et al., 2022; Petridis et al., 2022; Prey, 2018; Siles et al., 2019)

Conclui-se assim que a transição para o *streaming* trouxe também uma democratização ao mercado, pois deixou de existir um custo financeiro associado à descoberta e partilha de música. Os utilizadores não têm já necessidade de comprar o álbum em questão, o que obviamente facilita e promove novas experiências, dando maior poder de escolha e decisão ao consumidor. Para além disso, releva também o papel do “orientador” de descobertas e consumos, pois o único custo associado é mesmo o fator tempo: os indivíduos ambicionam receber as sugestões mais eficazes, que não os façam perder tempo com músicas que não irão gostar (De Smet & Dhaenens, 2022; Hujran et al., 2020; Kaimann et al., 2021; Li & Liu, 2020; Lyu, 2021; Maasø, 2018; Tepper & Hargittai, 2009; Way et al., 2020). Esta alteração das condições de acesso trouxe também volatilidade ao mercado: geralmente, o pico de audiência de uma música é relativamente próximo do seu lançamento, marcado por um consume *on*

repeat (Kaimann et al., 2021; Lüders, 2021). Porém, este sucesso pode ser, por vezes, extremamente curto, dado que existe elevada rotatividade e o tal consumo *on repeat* traz frequentemente alguma saturação aos consumidores (Kaimann et al., 2021; Lüders, 2021).

2.6. Os novos processos de tomada de decisão na era do *streaming*

Com a explicada transformação digital e social da indústria musical, alteraram-se também as condições comerciais de distribuição e oferta do catálogo musical: as barreiras à entrada são cada menos preponderantes, dado que o custo de distribuição é próximo de zero e houve uma notória redução dos custos de produção e promoção (Hujran et al., 2020; Kaimann et al., 2021; Prey et al., 2022; Tepper & Hargittai, 2009; Towse, 2020). Tudo isto não só trouxe um enfraquecimento da vantagem competitiva das *Major Labels*, perturbando a sua hegemonia e abrindo espaço a que *Independent Labels* ganhem força e destaque, mas provocou também um aumento da oferta musical, acirrando a concorrência entre artistas e alterando os padrões de difusão das músicas (Kaimann et al., 2021). Assim, tanto para as *Major Labels* recuperarem o seu posicionamento, como para as *Independent Labels* assumirem definitivamente uma posição de destaque, será determinante estas conhecerem e dominarem as dinâmicas de atração e retenção de ouvintes nas novas condições e circunstâncias de mercado (Hujran et al., 2020; Kaimann et al., 2021; Maasø & Hagen, 2020; Negus, 2019).

Concluindo, se artistas e *Record Labels* precisam dominar o *streaming* para executar as suas estratégias comerciais, seja penetrar em novos mercados, lançar músicas, manter *shares*, ... devem então investir os seus esforços onde realmente conseguem interferir: se os conteúdos do Spotify são da responsabilidade dos curadores e do algoritmo, então importa aos *players* referidos dominar estratégias de difusão em redes, chegar às *playlists* mais difundidas, entender as suas audiências, direcionar estratégias de *marketing*, etc. (Aguiar & Waldfogel, 2021; Kaimann et al., 2021; Maasø & Hagen, 2020; Morgan, 2020; Negus, 2019; Prey et al., 2022). Para o atingimento destes objetivos, é fundamental que as *Record Labels* continuem a investir em análise de dados e *Business Intelligence*, tirando partido de todos os dados que recolhem e/ou têm acesso, destacando-se os disponibilizados pelo Spotify, onde é possível aceder a *insights* sobre a *performance* de faixas e os perfis dos ouvintes (Aguiar & Waldfogel, 2021; Hujran et al., 2020; Maasø & Hagen, 2020; Negus, 2019).

3. Questões de Investigação, Dados e Metodologia

Neste capítulo, serão enquadradas e explicadas as três questões de investigação desta dissertação. Depois, serão apresentados os dados utilizados, bem como a metodologia aplicada.

3.1. Questões de Investigação

3.1.1. De que forma o consumo de música varia consoante o contexto?

Como já explorado anteriormente, os utilizadores do Spotify podem ouvir músicas em diferentes contextos (*radio*, *charts*, diferentes tipos de *playlists* ou experiências mais autónomas), o que influencia as dinâmicas de consumo dos indivíduos (Hosey et al., 2019; Prey et al., 2022). Vários autores abordaram já este tema, principalmente através de entrevistas, concluindo que os utilizadores tendem a ter consumos mais alinhados com as suas preferências, quando guiados pelo Spotify (Aguar & Waldfogel, 2021; Freeman et al., 2022; Hujran et al., 2020; Kaimann et al., 2021; Maasø & Hagen, 2020; Prey, 2018; Villermet et al., 2021). Ou seja, concluiu-se que experiências mais autónomas como *user collection* ou *catalog* abrem espaço a uma maior diversidade no consumo.

Depois de este tema já ter sido então bastante discutido na literatura, importa agora explorar tópicos como a popularidade, as preferências do utilizador ou a diversidade do consumo, partindo do de uma amostra do maior e mais completo *dataset* alguma vez disponibilizado pelo Spotify e, desta forma, complementando as conclusões de outros autores com análises de dados reais.

3.1.2. Qual a relação entre redes sociais de indivíduos e o surgimento de *hits*?

Sabendo a relevância que a partilha de músicas tem nas interações sociais (Tepper & Hargittai, 2009) e do real impacto deste fenómeno nos hábitos de consumo dos utilizadores (Huete-Alcocer, 2017; Sanjeev Verma and Neha, 2021), existe uma vasta literatura sobre o tema.

Assim, depois de vários autores terem já aferido que redes de indivíduos determinam padrões de difusão e consumo de músicas (De Smet & Dhaenens, 2022; Hagen & Lüders, 2017; Li & Liu, 2020; Park & Kaneshiro, 2021; Park, Lee, et al., 2022; Park, Redmond, et al., 2022; Ren & Kauffman, 2017), importa agora aplicar igualmente esta metodologia ao campo da HSS, *Hit Song Science*, por forma a perceber se a dinâmica das redes influencia particularmente o consumo de *hits*. Para este efeito, ir-se-á verificar se utilizadores mais centrais (que influenciam mais e são mais influenciados) ouvem mais *hits*.

3.1.3. A centralidade dos utilizadores influencia-os na escolha entre ter uma experiência guiada pelo Spotify ou mais autónoma?

Conforme visto anteriormente, os utilizadores podem optar por uma experiência guiada pelo próprio Spotify (*radio*, *charts*, *playlists* editoriais e *playlists* personalizadas) (Anderson et al., 2020; Hosey et al., 2019; Kaimann et al., 2021) ou por uma experiência mais autónoma, guiada pelo próprio utilizador, baseando-se nos seus gostos ou deixando-se influenciar por fatores externos: a sua rede, filmes, séries, eventos, redes sociais, rádio, etc. (Anderson et al., 2020; Hujran et al., 2020; Lüders, 2021; Maasø, 2018).

Vários autores investigaram já o que leva os utilizadores a optarem por experiências guiadas pelo algoritmo (De Assunção & Zaina, 2022; De Smet & Dhaenens, 2022; Freeman et al., 2022; Hosey et al., 2019; Hujran et al., 2020; Kaimann et al., 2021; Lyu, 2021; Morgan, 2020; Prey, 2018; Snickars, 2017). No entanto, é interessante também incluir aqui os restantes contextos dominados pela plataforma, pelo que se tentará entender o que leva os consumidores a tomar esta decisão de consumo e, principalmente, se esta escolha é influenciada pelas centralidades e redes dos utilizadores e se está ligada às popularidades das músicas.

3.2. Bases de Dados

Para responder a estas questões, será utilizada a base de dados The Music Streaming Sessions Dataset, MSSD (Brost et al., 2019), a maior e mais completa base de dados já disponibilizada pelo Spotify. Consiste em 160 milhões de sessões de *streaming*, com detalhes das interações do utilizador com a plataforma, características das músicas e outros metadados. Cada sessão é um período de audição de até 20 faixas (20 *logs*) com um limite de inatividade entre faixas consecutivas de 60 segundos. As sessões incluídas no *dataset* são selecionadas aleatoriamente a partir de todas as sessões elegíveis ao longo de um período de 8 semanas. Foram excluídas sessões que incluíssem músicas que não atendessem aos critérios mínimos de popularidade (Brost et al., 2019). Existe uma versão “mini” da base de dados completa, que contém apenas 167.880 *logs*. Dados os limites de capacidade de processamento dos computadores e programas utilizados, foi usada a versão “mini” da base de dados. De agora em diante, sempre que for mencionada a base de dados ou o *dataset*, será uma referência à versão “mini” dos 167.880 *logs*, correspondendo cada *log* a uma audição.

A base de dados dos *logs* está organizada da seguinte forma, correspondendo cada linha a um *log* e estando também identificadas a sessão, a música e outros atributos (Brost et al., 2019):

- **session_id**: ID da sessão em que o *log* se integra;
- **session_position**: Número inteiro de 1 a 20 que identifica a posição do *log* na sessão;
- **session_length**: Número inteiro de 1 a 20 que aponta a duração da sessão;
- **track_id_clean**: ID da faixa ouvida neste *log*;
- **skip_1**: Variável booleana (*True/False*) que indica se a faixa foi apenas ouvida "very briefly";
- **skip_2**: Variável booleana (*True/False*) que indica se a faixa foi apenas ouvida "briefly";
- **skip_3**: Variável booleana (*True/False*) que indica se quase toda a faixa foi ouvida;
- **not_skipped**: Variável booleana (*True/False*) que indica se a faixa foi ouvida até ao fim;
- **context_switch**: Variável booleana (*True/False*) que indica se o utilizador mudou de contexto do *log* anterior para o *log* corrente;
- **no_pause_before_play**: Variável booleana (*True/False*) que indica se houve uma paragem entre o *log* anterior e o *log* corrente;
- **short_pause_before_play**: Variável booleana (*True/False*) que indica se houve uma paragem curta entre o *log* anterior e o *log* corrente;
- **long_pause_before_play**: Variável booleana (*True/False*) que indica se houve uma paragem longa entre o *log* anterior e o *log* corrente;
- **hist_user_behavior_n_seekfwd**: Número de vezes que o utilizador avançou uma parte da faixa durante o *log*;
- **hist_user_behavior_n_seekback**: Número de vezes que o utilizador retrocedeu uma parte da faixa durante o *log*;
- **hist_user_behavior_is_shuffle**: Variável booleana (*True/False*) que indica se o utilizador estava a ouvir no modo "shuffle" (aleatório);
- **hour_of_day**: Hora do *log*;
- **date**: Data do *log*;
- **premium**: Variável booleana (*True/False*) que indica se o utilizador tinha uma conta *premium*;
- **context_type**: Contexto do *log*;
- **hist_user_behavior_reason_start**: Ação do utilizador que originou o *log* corrente;

- **hist_user_behavior_reason_end:** Ação do utilizador que terminou o *log* corrente.

Na base de dados das características e atributos das músicas, cada linha corresponde a uma faixa (*track_id*, detalhado na tabela abaixo) que corresponde ao atributo *track_id_clean* da anteriormente referida base de dados dos *logs*. De seguida, são explicados os campos que serão relevantes nesta investigação (Barone et al., 2017; Heggli et al., 2021; Lyu, 2021; Nijkamp, 2018; Sandag & Manueke, 2020):

- **track_id:** ID da faixa (cruza com o *track_id_clean* da base de dados de *logs*);
- **duration:** Duração da faixa em segundos;
- **release_year:** Ano de lançamento da faixa;
- **us_popularity_estimate:** Popularidade da faixa, em percentil, nos EUA a 12 de outubro de 2018;
- **acousticness:** Nível de Confiança de 0 a 1 que indica se a música foi produzida com instrumentos acústicos, incluindo voz (uma faixa com valores mais próximos de 1 é produzida maioritariamente sem recurso a instrumentos eletrónicos ou elétricos);
- **danceability:** Nível de Confiança de 0 a 1 que representa a adequabilidade da música para dançar (uma faixa com valores mais próximos de 1 é mais dançável);
- **energy:** Nível de Confiança de 0 a 1 que mede a intensidade da faixa (uma faixa com valores mais próximos de 1 é mais energética, ou seja, mais rápida, intensa e frenética; um prelúdio de Bach terá baixa energia enquanto uma faixa de *Death Metal* terá valores elevados);
- **instrumentalness:** Nível de Confiança de 0 a 1 que avalia se a faixa contém vozes (a cantar ou a falar), em que se prevê que acima de 0,5 a faixa seja instrumental;
- **key:** Número inteiro entre -1 e 11 que identifica a tonalidade da faixa (quando é -1, não se identificou a tonalidade) (é utilizada a *Pitch Class Notation*);
- **liveness:** Nível de Confiança de 0 a 1, em que 1 significa que a faixa tem um público presente na gravação (prevê-se que acima de 0,8 a faixa já inclua público);
- **loudness:** Medida da intensidade média da faixa em decibéis;
- **mode:** Indica se a faixa está em modo maior (1) ou menor (0) (notar que o modo maior é percecionado como sendo mais animado e feliz que o modo menor);
- **speechiness:** Nível de Confiança de 0 a 1, que identifica a presença de palavras faladas na faixa (abaixo de 0,33 deverá conter apenas música; entre 0,33 e 0,66 deverá conter música e partes faladas (ex.: *rap*); entre 0,66 e 1 deverá conter apenas partes

faladas (ex.: *talk-show*, *podcast*));

- **tempo**: Medida do andamento da faixa em BPM, *Beats Per Minute*;
- **time_signature**: Número inteiro entre 3 e 7 que indica o esquema rítmico/unidade de compasso da faixa;
- **valence**: Nível de Confiança de 0 a 1 que avalia o efeito psicológico da música: valores de próximos de 0 estão associados a tristeza ou depressão, e mais próximos de 1 estão associados a alegria ou euforia.

Todas estas *audio features* estão disponíveis no website Spotify for Developers⁴ e são calculadas pelo Spotify, que não divulga os algoritmos e métodos de obtenção destes dados (Heggli et al., 2021). A base de dados relativa às características das faixas tem ainda outros atributos que não serão explorados (*dynrangementmean*, *organism*, *bounciness*, *beat_strength*, *flatness*, *mechanism*, *acoustic_vector_0*, *acoustic_vector_1*, *acoustic_vector_2*, *acoustic_vector_3*, *acoustic_vector_4*, *acoustic_vector_5*, *acoustic_vector_6*, *acoustic_vector_7*), dado que o seu significado não é explicado em Spotify for Developers³, não sendo inclusive incluídos em várias publicações baseadas nesta base de dados (Ignatius Moses Setiadi et al., 2020; Lyu, 2021; Nijkamp, 2018; Sandag & Manueke, 2020).

Este trabalho é antecedido por várias publicações que, partindo da mesma base de dados ou outras semelhantes, exploraram as áreas de *Hit Song Science* e *Music Information Retrieval*, com o objetivo de prever que *audio features* determinariam maiores probabilidades de a música ser bem-sucedida, assumindo que os *hits* partilham certas características entre si (Barone et al., 2017; Nijkamp, 2018; Sandag & Manueke, 2020). Uma outra área que tem suscitado bastante atenção é o funcionamento do Spotify, nomeadamente ao nível do algoritmo e das *playlists* editoriais (Anderson et al., 2020; Eriksson, 2020; Freeman et al., 2022; Prey, 2018; Razlogova, 2020). Existem também várias publicações que exploram as temáticas de difusão e redes, pretendendo explicar o comportamento dos indivíduos na era do *streaming* e as consequentes alterações da dinâmica da Indústria Musical e respetivos *stakeholders* (Gulati & Eirinaki, 2019; Kaimann et al., 2021; Lee et al., 2019; Lyu, 2021).

Desta forma, surge a necessidade de agregar todas estas áreas, relacionando estilos musicais e *audio features* com o comportamento dos indivíduos, desde a adesão a *hits* à forma como consomem música no Spotify: se guiados pelo Spotify ou seguindo uma experiência mais

⁴ <https://developer.spotify.com/documentation/web-api/reference/get-audio-features>

personalizada, absorvendo todas as influências das suas respectivas redes (Nijkamp, 2018; Villermet et al., 2021).

3.3. Metodologia

3.3.1. *Clustering* das faixas musicais

Para investigar as questões propostas, é necessário começar por integrar as diferentes faixas musicais nas suas respectivas categorias: os estilos musicais, classificação dominante e que é entendida e utilizada pelo mercado, tanto por consumidores, como por *labels* e artistas (Anderson et al., 2020). No entanto, esta classificação não consta no *dataset* e não é sequer incluída em Spotify for Developers³, pelo que surge então a necessidade de classificar as faixas.

Neste seguimento, tentou-se criar um classificador K-NN (usando a aplicação RapidMiner (Mierswa et al., 2006)) tendo por base diferentes *datasets* disponíveis em Kaggle.com (Henrique, 2019; Noah, 2021; Singh, 2019) que incluem os atributos já apresentados, mas também o estilo musical. O classificador K-Nearest Neighbors (KNN) é um algoritmo de *machine learning* supervisionado que classifica os dados com base nas respectivas distâncias aos vizinhos mais próximos, tendo já sido aplicado de forma bem-sucedida a esta mesma base de dados em outras publicações (Gulmatico et al., 2022; Ignatius Moses Setiadi et al., 2020). Verificou-se que os resultados não foram satisfatórios devido à disparidade de universos de estilos musicais utilizados para treinar o classificador (mais ou menos abrangentes, mais ou menos limitados): por exemplo, enquanto uma das bases de dados continha apenas 6 estilos musicais (entre eles o *rock*), numa outra existiam já 114 estilos musicais diferentes (comparando ao outro *dataset*, aqui já tínhamos 28 estilos distintos de *rock*).

Assim, tentou-se criar um outro classificador K-NN, mas agora tendo por base apenas o *dataset* com o universo mais reduzido de estilos musicais (6). No entanto, os resultados desta classificação também não foram satisfatórios, pois geraram uma distribuição muito concentrada das 10.000 faixas num determinado estilo.

Como não era possível verificar se esta concentração estaria correta, abandonou-se a tentativa de classificar diretamente as faixas em estilos musicais e, então, foi construído um processo de *clustering* que originou 10 *clusters*, através do processo K-Means Clustering, normalizando previamente as diferentes *features* para assegurar que todas tivessem influências equivalentes.

3.3.2. Redes de *users* no Spotify

Para construir então as redes de difusão a partir desta base de dados, assumiu-se que cada sessão (*session_id*) corresponde a um utilizador diferente e que as redes são bipartidas, dado que os nós podem ser de dois tipos: indivíduos ou músicas. Os indivíduos só se podem ligar a músicas, dado que não existem ligações diretas entre indivíduos, nem entre músicas.

Recorrendo à Análise de Redes Sociais (DNA, *Social Network Analysis*), é possível aplicar a estrutura das redes às interações sociais e relacionais, bem como às decorrentes difusões de informação (Kumar & Sinha, 2021; Litterio et al., 2017), podendo as redes assumir diferentes modelos de acordo com, por exemplo, a rapidez a que flui a informação, a proximidade entre indivíduos, a estrutura das redes ou até mesmo o meio de comunicação (Kumar & Sinha, 2021; Valente & Vega Yon, 2020). Neste caso, em vez de ser usada a rede bipartida (músicas-utilizadores), será utilizada uma projeção desta rede, contendo apenas utilizadores. Com base nestas redes, será calculada a centralidade dos indivíduos, que revela a importância de um indivíduo na rede: quanto maior for a sua centralidade, mais exposto estará ele a novas informações, maior alcance terá (chegando a mais indivíduos) e mais suscetível a ser “infetado” estará também (Kumar & Sinha, 2021; Opsahl et al., 2010). Neste trabalho, será usada como medida de centralidade a *closeness*, que corresponde ao inverso da média das distâncias de um nó a cada um dos restantes nós (Opsahl et al., 2010; Valente & Vega Yon, 2020).

Sabendo que cada indivíduo distribui o seu consumo pelos 10 *clusters*, foi desenvolvida uma matriz onde cada linha corresponde a um *user*, existindo 10 colunas (uma por cada cluster) com as percentagens de audições em cada um destes *clusters* (somando os valores de cada linha, obtém-se sempre 100%).

Depois, para calcular a *closeness*, foi utilizado o software R (R Core Team, 2021), recorrendo ao *package* *igraph* (Csardi & Nepusz, 2006), para, a partir da referida matriz, ser obtida a *weighted_closeness* de cada indivíduo, utilizando o código presente no anexo. Não foi possível obter a *weighted closeness* de 3222 *users* (68% dos *users*, 72% dos *logs*), o que compromete qualquer análise feita a este atributo.

Foi também identificado o *cluster* preferido de cada *user* (*user_fav_cluster*), e partir daí constituídas redes, recuperando assim o conceito explorado anteriormente das comunidades (os indivíduos tendem a organizar-se em grupos com interesses semelhantes e é principalmente nestes ambientes de preferências partilhadas que se partilham mais recomendações (Brooks & Conroy, 2010; Clark & Lonsdale, 2022; Lee et al., 2019; Park, Lee, et al., 2022; Prey, 2018; Trocchia et al., 2002)). Assim, foram formadas 10 redes, representando 10 comunidades

unidas por gostos em comum: têm o mesmo `user_fav_cluster`. Foi, depois, aplicado o código R (R Core Team, 2021) anteriormente referido às 10 redes, agora para obter a `fav_cluster_user_weighted_closeness`. De notar que neste cálculo já foi possível obter dados para todos os *users*. Assim, no decorrer deste trabalho será considerada a `fav_cluster_user_weighted_closeness`. Na seguinte tabela, é possível observar a constituição das 10 redes, detalhando o número de *users* que as integram e as respetivas quantidades de *logs*, bem como as percentagens associadas face ao total de *users/logs*.

Tabela 1 – Distribuição de *users* e *logs* por `user_fav_cluster`

	<code>fav_cluster_1</code>	<code>fav_cluster_2</code>	<code>fav_cluster_3</code>	<code>fav_cluster_4</code>	<code>fav_cluster_5</code>
no.	132	3	82	705	1618
users	1,32%	0,03%	0,82%	7,05%	16,18%
no.	2164	45	1425	11618	26782
log	1,29%	0,03%	0,85%	6,92%	15,95%
	<code>fav_cluster_6</code>	<code>fav_cluster_7</code>	<code>fav_cluster_8</code>	<code>fav_cluster_9</code>	<code>fav_cluster_10</code>
no.	4971	413	314	666	1096
users	49,71%	4,13%	3,14%	6,66%	10,96%
no.	84600	6691	5220	11088	18247
logs	50,39%	3,99%	3,11%	6,60%	10,87%

3.3.3. Algoritmos para classificação das experiências de consumo no Spotify

Como já foi abordado anteriormente, importa distinguir as experiências dos utilizadores: se são guiados pelo Spotify nos seus consumos e descobertas ou se são mais autónomos, consumindo conteúdos de forma mais independente, guiando-se por preferências próprias e influências dos pares. A questão ganha especial relevância se considerarmos as redes e os fenómenos associados de EWOM, difusão e *opinion leaders*: será que utilizadores mais centrais, movidos pelas influências das redes, irão ter experiências mais autónomas e depender menos do Spotify? E, por outro lado, será que a audição de faixas mais populares esta associada a experiências guiadas pelo próprio Spotify?

Assim, foram definidas duas experiências diferentes (`experience_type`):

- `guided_by_spotify`, correspondendo aos contextos *charts*, *editorial_playlist*, *personalized_playlist* e *radio*;
- `guided_by_user`, correspondendo aos contextos *catalog* e *user_collection*.

Para entender o que determina o tipo de consumo numa determinada experiência de um utilizador, foi criado um modelo preditivo em que `experience_type` é a variável dependente. Com recurso ao RapidMiner (Mierswa et al., 2006), foram então testados diferentes tipos de modelos e diferentes combinações das várias variáveis disponíveis, até encontrar os modelos

que melhor se ajustavam. Assim, serão apresentados três modelos por cada comunidade, cada `user_fav_cluster`: Árvore de Decisão, *Fast Large Margin* e *Neural Networks*:

- Árvores de decisão são um modelo de *machine learning* destinado a resolver problemas de classificação e regressão (Rahman et al., 2023). Este algoritmo é representado na forma de uma estrutura hierárquica invertida (uma árvore invertida), onde cada nó representa uma decisão relativa a uma condição de uma determinada variável e as folhas indicam as classes ou previsões finais (Rahman et al., 2023).
- O *Fast Large Margin* foi utilizado no ambiente AutoModel do RapidMiner (Mierswa et al., 2006), um *automated machine learning* (AutoML) capaz de selecionar algoritmos, ajustar parâmetros, automatizar processos, ... (Rahman et al., 2023) Assim, o AutoModel utiliza mecanismos de otimização computacional para selecionar automaticamente as melhores características e algoritmos, reduzindo a complexidade e aumentando a eficiência (Rahman et al., 2023). O *Fast Large Margin* (FLM) é um algoritmo de *machine learning* de classificação e previsão baseado nos modelos de *Support Vector Machines* (SVM), mas com a capacidade de processar *datasets* com milhões de dados e atributos⁵. Tal como o SVM, o FLM maximiza as distâncias entre as diferentes classes e determina uma equação semelhante às regressões logísticas⁴.
- As Redes Neurais (*Neural Networks*) são modelos de *machine learning* inspirados no funcionamento do cérebro humano e utilizados para classificação, regressão e reconhecimento de padrões (Kovács & Ghous, 2020; Kriegeskorte & Golan, 2019; Wu & Feng, 2018). São compostas por nós (neurónios artificiais) interligados e organizados em *layers*: o *input*, uma ou mais *hidden layers* e o *output* (Kovács & Ghous, 2020; Kriegeskorte & Golan, 2019; Wu & Feng, 2018). A partir do *input* (as variáveis do *dataset* em questão), os dados são processados nas *hidden layers* e, consoante os pesos e *biases* associados a cada variável e nó, é obtido o *output* (Kovács & Ghous, 2020; Kriegeskorte & Golan, 2019; Wu & Feng, 2018).

⁵ https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/support_vector_machines/fast_large_margin.html

4. Resultados

Na sequência da metodologia anteriormente apresentada, neste capítulo, serão expostos os resultados obtidos, incluindo o processo de *clustering* das faixas musicais, o estudo dos hábitos de consumo dos utilizadores nos diferentes contextos do Spotify, a análise das audições de *bits*, a análise da influência das redes nas audições de *bits* e, ainda, as observações tiradas a partir dos modelos desenvolvidos para a classificação das experiências de consumo no Spotify. Por fim, serão discutidos os resultados obtidos

4.1. Resultados Obtidos

4.1.1. *Clustering* das Faixas Musicais

Os 10 *clusters* obtidos através do *K-Means Clustering* refletem distribuições semelhantes dos diferentes atributos das faixas musicais, pelo que se pode fazer a sua caracterização e até um certo paralelismo com determinados estilos, apesar de não terem sido constituídos com base nesta classificação.

Para este efeito, importa então analisar os seguintes *boxplots* que evidenciam a distribuição dos diferentes metadados nas 50.704 músicas organizadas em 10 *clusters*.

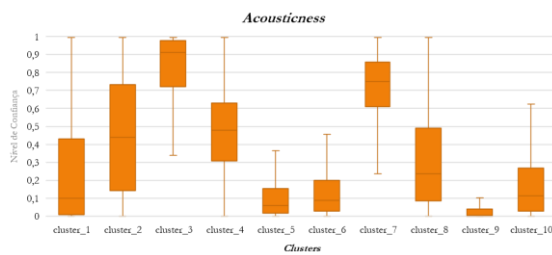


Figura 4 – Distribuição da *Acousticness* nos 10 *Clusters*

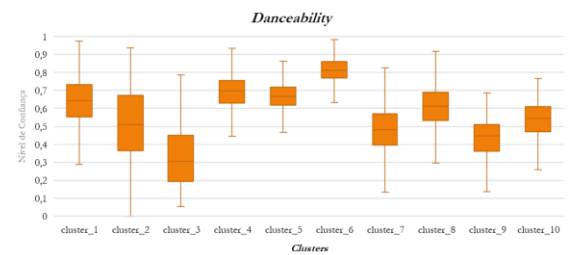


Figura 5 – Distribuição da *Danceability* nos 10 *Clusters*

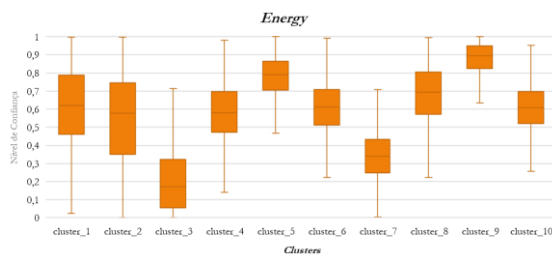


Figura 6 – Distribuição da *Energy* nos 10 *Clusters*

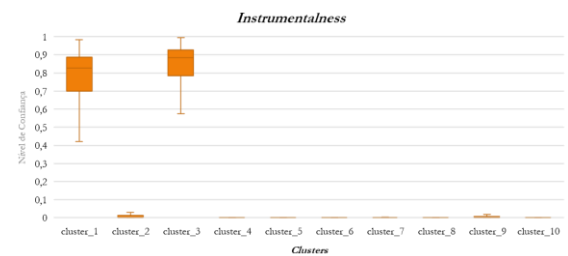


Figura 7 – Distribuição da *Instrumentalness* nos 10 *Clusters*

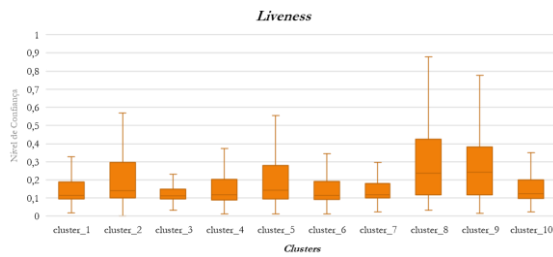


Figura 8 – Distribuição da Liveness nos 10 Clusters

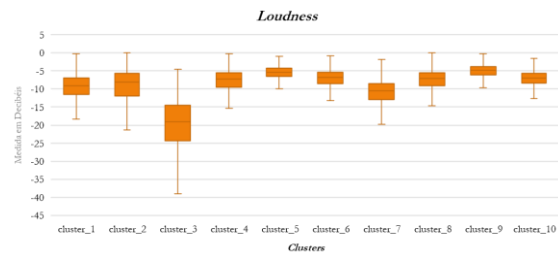


Figura 9 – Distribuição da Loudness nos 10 Clusters

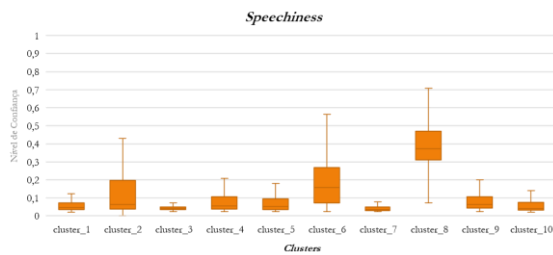


Figura 10 – Distribuição da Speechiness nos 10 Clusters

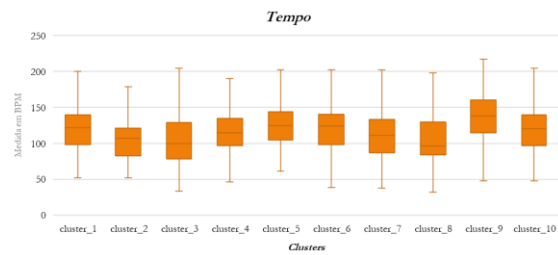


Figura 11 – Distribuição do Tempo nos 10 Clusters

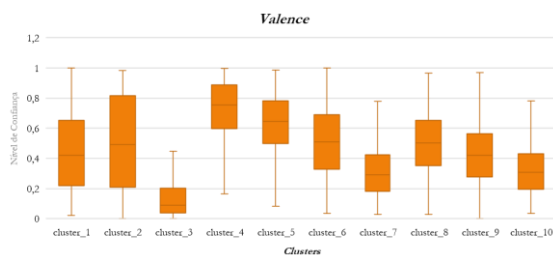


Figura 12 – Distribuição da Valence nos 10 Clusters



Figura 13 – Distribuição da US_popularity_estimate nos 10 Clusters

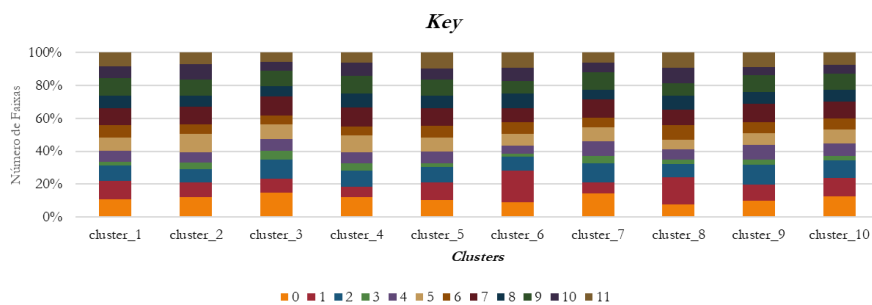


Figura 14 – Distribuição da Key nos 10 Clusters

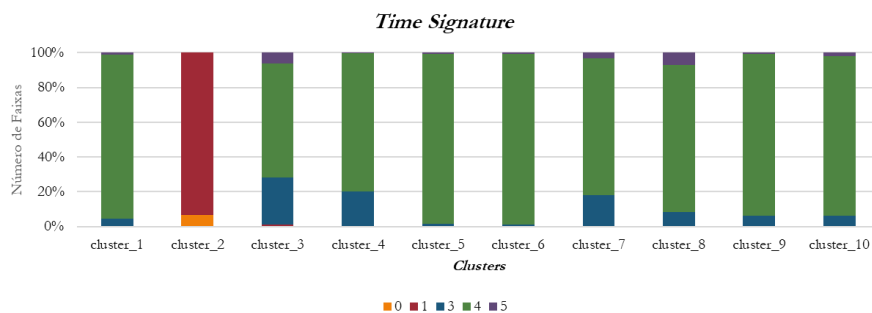


Figura 15 – Distribuição da *Time_signature* nos 10 Clusters



Figura 16 – Distribuição do *Mode* nos 10 Clusters

Tabela 2 – Distribuição das Faixas por Clusters (número e respetiva percentagem de faixas face ao total)

	clus- ter_1	clus- ter_2	clus- ter_3	clus- ter_4	clus- ter_5	clus- ter_6	clus- ter_7	clus- ter_8	clus- ter_9	clus- ter_10
faixas	1906	426	1256	5972	8761	9308	4490	3885	6639	8061
	4%	1%	2%	12%	17%	18%	9%	8%	13%	16%

Aferiu-se então que:

- O cluster 1 tem um elevada *Instrumentalness* e não tão alta *Acousticness*, o que revela que apesar de provavelmente ser música maioritariamente produzida sem recurso a vozes, inclui instrumentos elétricos e/ou eletrónicos. Tem ainda elevada *Loudness* e *Danceability*, pelo que se assumirá que se trata de EDM, *Electronic Dance Music*.
- O cluster 3 tem um elevado valor para a *Acousticness* e para a *Instrumentalness*, revelando ser constituído por música instrumental (acústica e sem recurso a vozes). Os baixos valores de *Valence*, *Energy* e *Loudness* evidenciam que será música suave e delicada. Assim, supõe-se que se tratará de música clássica/instrumental calma, como explicado anteriormente (recuperando inclusive a indicação do Spotify de que um prelúdio de Bach teria baixa energia³).
- Os clusters 2 e 4 são constituídos por músicas com presença de vozes, dada a reduzida *instrumentalness*, e têm *acousticness*, *energy*, *loudness* e *danceability* médias a elevadas. Para facilitar o entendimento deste trabalho, assumiremos tratar-se de *funk*, dadas as

características partilhadas.

- Os clusters 5, 6, 9 e 10 são bastante semelhantes: música cantada (baixa *instrumentalness*), com forte presença de instrumentos eletrónicos ou elétricos (baixa *acousticness*) e também com *danceability* e *energy* algo elevadas. Para facilitar o entendimento deste trabalho, assumiremos tratar-se de *pop-rock*, dadas as características partilhadas.
- O cluster 7 tem elevada *acousticness* e *valence*, baixa *danceability* e *energy* e muito reduzida *Instrumentalness*, pelo que percebemos que se tratará de música cantada, acústica e tranquila, mas que, ainda assim, transmite emoções positivas. Para facilitar o entendimento deste trabalho, assumiremos tratar-se de *jazz*, dadas as características partilhadas.
- O cluster 8 destaca-se dos demais por ter um número considerável de faixas com *speechiness* entre 0,33 e 0,66, que, como explicado anteriormente³, revela tratar-se de música *rap*.

4.1.2. Hábitos de consumo nos diferentes contextos Spotify

O atributo `context_type` identifica o contexto no qual o *log* se insere e pode constituir os seguintes tipos, todos já previamente aprofundados:

- **editorial_playlist:** *playlists* editoriais.
- **user_collection:** a biblioteca do utilizador, incluindo artistas, álbuns, faixas e *playlists*.
- **radio:** recomendações do algoritmo associadas a uma determinada música, álbum ou artista.
- **personalized_playlist:** *playlists* desenvolvidas pelo algoritmo especificamente para cada utilizador.
- **catalog:** todo o catálogo de faixas disponível para os utilizadores.
- **charts:** *rankings* onde constam as 50 músicas ou álbuns mais ouvidos globalmente ou numa determinada categoria, região, período de tempo, etc.

Nesta sequência, importa então analisar o comportamento dos *users* nos diferentes contextos. Para tal, será analisado o consumo de músicas por *cluster* e se corresponde às preferências do utilizador e a respetiva variação consoante o contexto em questão.

Assim, a seguinte tabela ilustra a distribuição dos *logs* por *cluster* (e se corresponde ou não ao *cluster* preferido), bem como a respetiva percentagem.

Tabela 3 – Distribuição de todos os logs por cluster e is_fav_cluster (número de logs e respetiva percentagem face ao total)

cluster	cluster_1		cluster_2		cluster_3		cluster_4		cluster_5	
	2748		1038		1640		15110		27595	
	2%		1%		1%		9%		16%	
is_fav_cluster	True	False	True	False	True	False	True	False	True	False
	1007	1741	16	1022	923	717	5891	9219	11682	15913
	37%	63%	2%	98%	56%	44%	39%	61%	42%	58%

cluster	cluster_6		cluster_7		cluster_8		cluster_9		cluster_10		Total	
	58185		10269		12320		12384		26591		167880	
	35%		6%		7%		7%		16%		100%	
is_fav_cluster	True	False	True	False	True	False	True	False	True	False	True	False
	49328	8857	3049	7220	2501	9819	6301	6083	8010	18581	88708	79172
	85%	15%	30%	70%	20%	80%	51%	49%	30%	70%	53%	47%

A seguinte tabela evidencia a distribuição dos logs por cluster (e se corresponde ou não ao cluster preferido), bem como a respetiva percentagem, no contexto catalog.

Tabela 4 – Distribuição dos Logs em catalog por cluster e is_fav_cluster (número de logs e respetiva percentagem face ao total)

cluster	cluster_1		cluster_2		cluster_3		cluster_4		cluster_5	
	733		360		636		3630		5667	
	2%		1%		2%		9%		15%	
is_fav_cluster	True	False	True	False	True	False	True	False	True	False
	284	449	12	348	399	237	1844	1786	2622	3045
	39%	61%	3%	97%	63%	37%	51%	49%	46%	54%

cluster	cluster_6		cluster_7		cluster_8		cluster_9		cluster_10		Total catalog	
	11223		3204		3020		3704		6086		38263	
	29%		8%		8%		10%		16%		100%	
is_fav_cluster	True	False	True	False	True	False	True	False	True	False	True	False
	9446	1777	1208	1996	837	2183	2433	1271	2304	3782	21389	16874
	84%	16%	38%	62%	28%	72%	66%	34%	38%	62%	56%	44%

A seguinte tabela evidencia a distribuição dos logs por cluster (e se corresponde ou não ao cluster preferido), bem como a respetiva percentagem, no contexto charts.

Tabela 5 – Distribuição dos logs em charts por cluster e is_fav_cluster (número de logs e respetiva percentagem face ao total)

cluster	cluster_1		cluster_2		cluster_3		cluster_4		cluster_5	
	0		0		0		150		360	
	0%		0%		0%		6%		15%	
is_fav_cluster	True	False	True	False	True	False	True	False	True	False
	0	0	0	0	0	0	0	150	49	311
	0%	0%	0%	0%	0%	0%	0%	100%	14%	86%

cluster	cluster_6		cluster_7		cluster_8		cluster_9		cluster_10		Total charts	
	1433		14		102		4		378		2441	
	59%		1%		4%		0%		15%		100%	
is_fav_cluster	True	False	True	False	True	False	True	False	True	False	True	False
	1368	65	0	14	0	102	0	4	6	372	1423	1018
	95%		5%		0%		100%		0%		100%	

A seguinte tabela evidencia a distribuição dos *logs* por cluster (e se corresponde ou não ao cluster preferido), bem como a respetiva percentagem, no contexto editorial_playlist.

Tabela 6 – Distribuição dos logs em editorial_playlist por cluster e is_fav_cluster (número de logs e percentagem face ao total)

cluster	cluster_1		cluster_2		cluster_3		cluster_4		cluster_5	
	416		164		363		3072		6346	
	1%		0%		1%		9%		18%	
is_fav_cluster	True	False	True	False	True	False	True	False	True	False
	223	193	0	164	261	102	1034	2038	2824	3522
	54%		46%		0%		100%		72%	

cluster	cluster_6		cluster_7		cluster_8		cluster_9		cluster_10		Total editorial_playlist	
	14776		1644		1875		1794		5213		35663	
	41%		5%		5%		5%		15%		100%	
is_fav_cluster	True	False	True	False	True	False	True	False	True	False	True	False
	13100	1676	500	1144	120	1755	748	1046	1107	4106	19917	15746
	89%		11%		30%		70%		6%		94%	

A seguinte tabela evidencia a distribuição dos *logs* por cluster (e se corresponde ou não ao cluster preferido), bem como a respetiva percentagem, no contexto personalized_playlist.

Tabela 7 – Distribuição dos logs em personalized_playlist por cluster e is_fav_cluster (número de logs e respetiva percentagem face ao total)

cluster	cluster_1		cluster_2		cluster_3		cluster_4		cluster_5	
	155		19		31		331		591	
	5%		1%		1%		11%		20%	
is_fav_cluster	True	False	True	False	True	False	True	False	True	False
	70	85	0	19	0	31	122	209	269	322
	45%		55%		0%		100%		0%	

cluster	cluster_6		cluster_7		cluster_8		cluster_9		cluster_10		Total personalized_playlist	
	547		232		216		319		535		2976	
	18%		8%		7%		11%		18%		100%	
is_fav_cluster	True	False	True	False	True	False	True	False	True	False	True	False
	318	229	71	161	34	182	98	221	153	382	1135	1841
	58%		42%		31%		69%		16%		84%	

A seguinte tabela evidencia a distribuição dos *logs* por cluster (e se corresponde ou não ao cluster preferido), bem como a respetiva percentagem, no contexto radio.

Tabela 8 – Distribuição dos logs em radio por cluster e is_fav_cluster (número de logs e respetiva percentagem face ao total)

cluster	cluster_1		cluster_2		cluster_3		cluster_4		cluster_5	
	284		114		119		2420		3755	
	1%		1%		1%		11%		18%	
is_fav_cluster	True	False	True	False	True	False	True	False	True	False
	97	187	0	114	68	51	1244	1176	1612	2143
	34%	66%	0%	100%	57%	43%	51%	49%	43%	57%

cluster	cluster_6		cluster_7		cluster_8		cluster_9		cluster_10		Total radio	
	7408		1079		1226		1370		3332		21107	
	35%		5%		6%		6%		16%		100%	
is_fav_cluster	True	False	True	False	True	False	True	False	True	False	True	False
	6426	982	366	713	370	856	682	688	1222	2110	12087	9020
	87%	13%	34%	66%	30%	70%	50%	50%	37%	63%	57%	43%

A seguinte tabela evidencia a distribuição dos logs por cluster (e se corresponde ou não ao cluster preferido), bem como a respetiva percentagem, no contexto user_collection.

Tabela 9 – Distribuição dos logs em user_collection por cluster e is_fav_cluster (número de logs e respetiva percentagem face ao total)

cluster	cluster_1		cluster_2		cluster_3		cluster_4		cluster_5	
	1160		381		491		5507		10876	
	2%		1%		1%		8%		16%	
is_fav_cluster	True	False	True	False	True	False	True	False	True	False
	333	827	4	377	195	296	1647	3860	4306	6570
	29%	71%	1%	99%	40%	60%	30%	70%	40%	60%

cluster	cluster_6		cluster_7		cluster_8		cluster_9		cluster_10		Total user_collection	
	22798		4096		5881		5193		11047		67430	
	34%		6%		9%		8%		16%		100%	
is_fav_cluster	True	False	True	False	True	False	True	False	True	False	True	False
	18670	4128	904	3192	1140	4741	2340	2853	3218	7829	32757	34673
	82%	18%	22%	78%	19%	81%	45%	55%	29%	71%	49%	51%

Com a análise das tabelas, pode-se concluir que é nas personalized_playlists que os utilizadores mais ouvem músicas diferentes das suas preferências, contrariamente ao que seria espetável. Também no user_collection, os utilizadores ouvem mais músicas diferentes das suas preferências, apesar de ser uma diferença de apenas 2%, bem menos expressiva do que no contexto anteriormente referido.

No seguinte gráfico, verifica-se ainda que é nas personalized_playlists e no catalog que existe uma distribuição mais homogénea das audições por cluster.

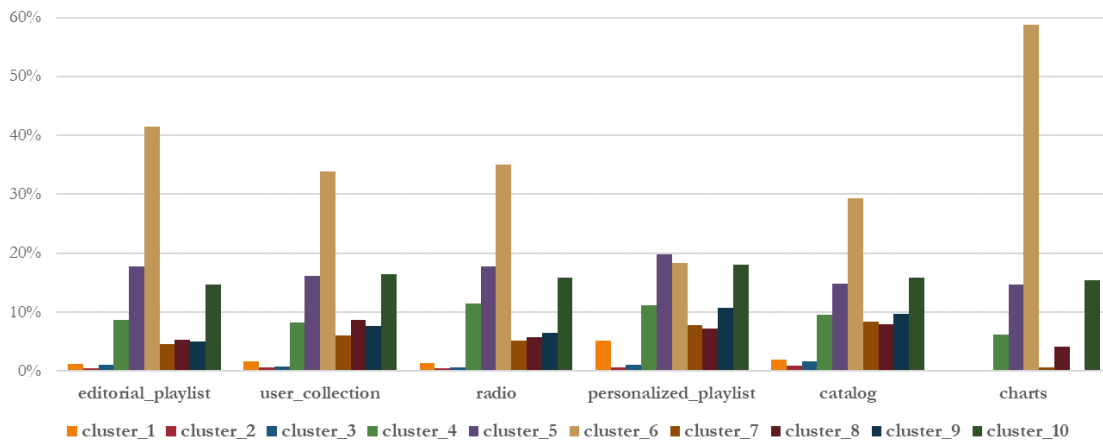


Figura 17 – Distribuição das audições por cluster nos 6 contextos

4.1.3. Audição de *Hits*

Como explicado anteriormente, um dos objetivos deste estudo é o de entender se efetivamente *users* mais centrais, ou seja, que são mais influenciados e influentes nas redes, são mais propensos a ouvir *hits*.

Para tal, iremos considerar a variável `us_popularity_estimate` de cada faixa, considerando a distribuição da popularidade por contexto, como evidenciado na seguinte *boxplot*.

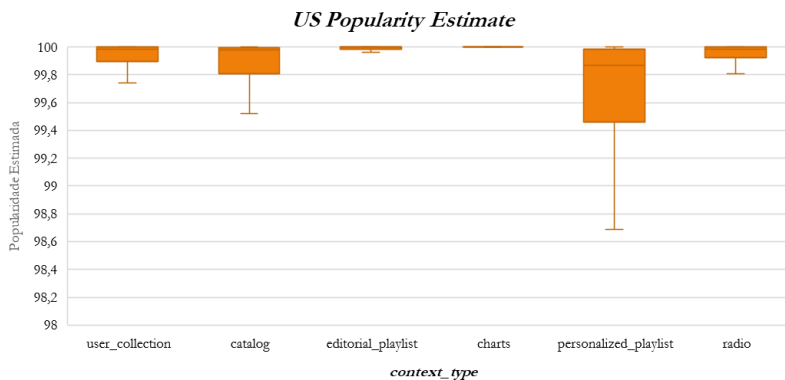


Figura 18 – Distribuição da `US_popularity_estimate` nos 6 contextos

É interessante perceber que, tal como esperado, os *charts* concentram as músicas mais populares (com valores muito próximos de 100), mas também a *user collection*, a *editorial_playlist* e o *radio* concentram popula-

ridades elevadas. Onde encontramos a maior dispersão de popularidade é mesmo nas *personalized_playlist*.

Demonstramos assim algo já apontado pela literatura: o Spotify aposta principalmente em músicas com sucesso comprovado, seja por intermédio do seu algoritmo, *radio*, ou por meio das *playlists* criadas pelos seus curadores, *editorial_playlist*. No entanto, é surpreendente o facto de *personalized_playlists* terem a maior dispersão de popularidades. Possivelmente, as músicas menos populares serão músicas que algoritmo e curadores preveem vir a atingir elevada popularidade, por partilhem características similares com as músicas mais populares.

4.1.4. Influência da Centralidade dos *Users* na audição de *Hits*

Para aprofundar a análise às músicas mais populares, foram identificadas as 50 músicas com maior popularidade em cada comunidade de indivíduos que partilham o mesmo *cluster* mais ouvido e, para perceber qual a influência das redes no surgimento destes *hits*, recorreu-se a gráficos de dispersão para perceber a relação entre as seguintes variáveis:

- fav_cluster_weighted_closeness (a centralidade do *user* na rede da sua comunidade: os indivíduos que partilham o mesmo *cluster* preferido)
- % de músicas ouvidas que integram TOP50 da comunidade (o rácio entre o número de músicas ouvidas pelo *user* que integram as 50 mais populares desta comunidade e o número total de músicas ouvidas pelo *user*)

Note-se que, por exemplo, as músicas da comunidade do *fav_cluster_1* não pertencem necessariamente ao *cluster* 1: são, sim, ouvidas por *users* cujo *cluster* mais ouvido é o 1.

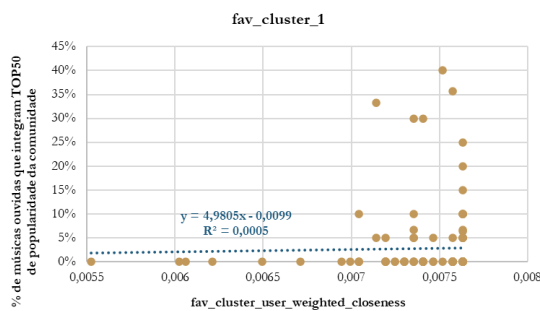


Figura 19 – Gráfico de dispersão (centralidade dos *users* e *hits* que ouvem) do *user_fav_cluster_1*

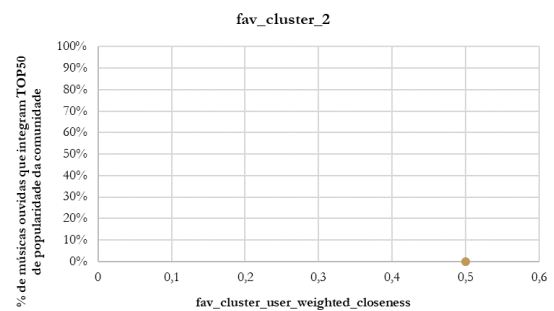


Figura 20 – Gráfico de dispersão (centralidade dos *users* e *hits* que ouvem) do *user_fav_cluster_2*

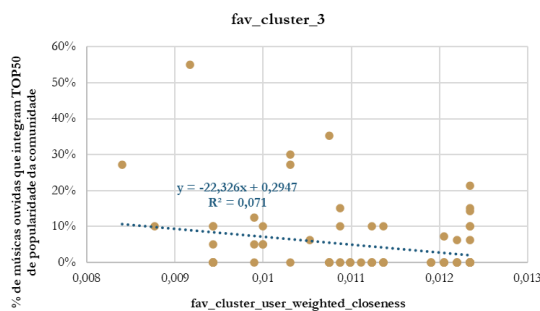


Figura 21 – Gráfico de dispersão (centralidade dos *users* e *hits* que ouvem) do *user_fav_cluster_3*

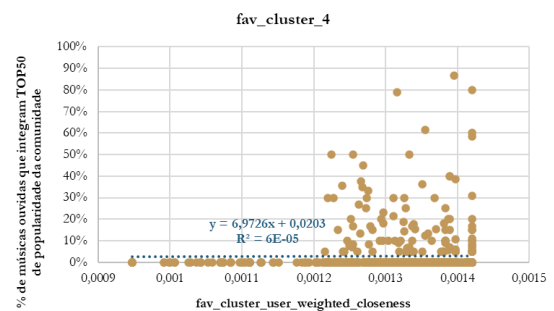


Figura 22 – Gráfico de dispersão (centralidade dos *users* e *hits* que ouvem) do *user_fav_cluster_4*

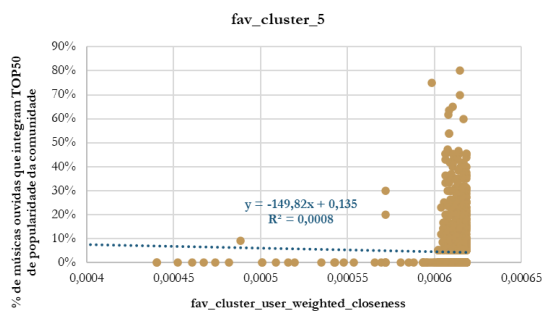


Figura 23 – Gráfico de dispersão (centralidade dos users e hits que ouvem) do user_fav_cluster_5

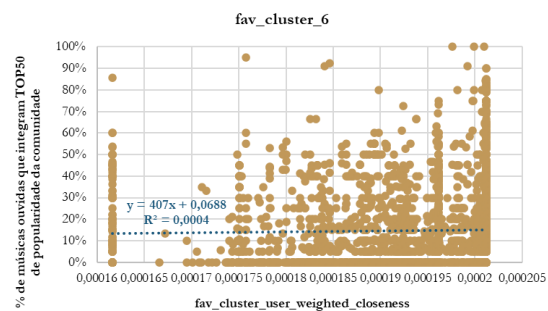


Figura 24 – Gráfico de dispersão (centralidade dos users e hits que ouvem) do user_fav_cluster_6

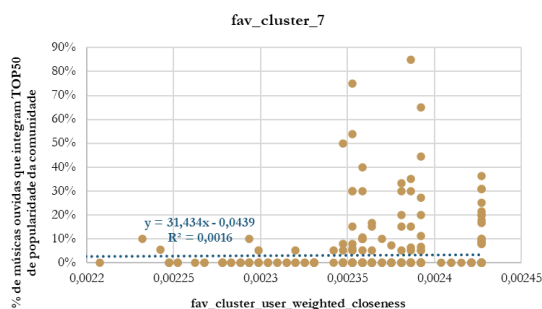


Figura 25 – Gráfico de dispersão (centralidade dos users e hits que ouvem) do user_fav_cluster_7

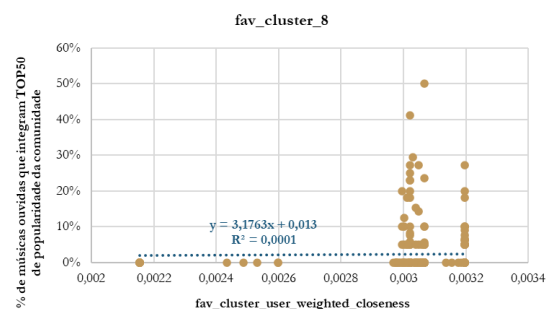


Figura 26 – Gráfico de dispersão (centralidade dos users e hits que ouvem) do user_fav_cluster_8

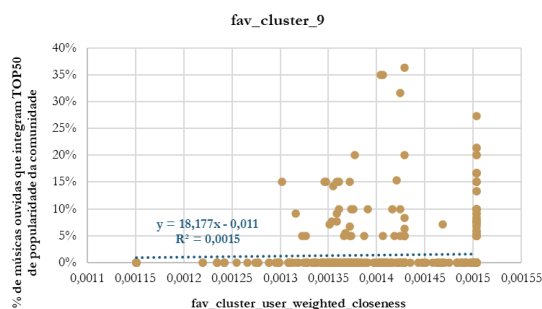


Figura 27 – Gráfico de dispersão (centralidade dos users e hits que ouvem) do user_fav_cluster_9

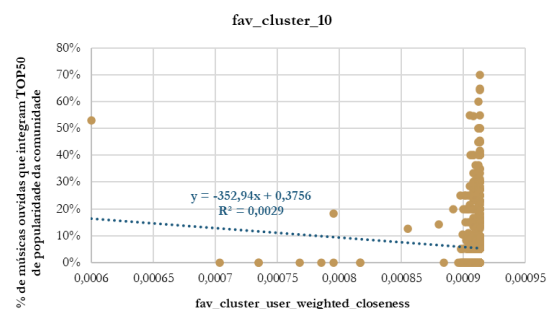


Figura 28 – Gráfico de dispersão (centralidade dos users e hits que ouvem) do user_fav_cluster_10

Nos gráficos representados, percebe-se uma tendência para os *users* que mais ouvem músicas do TOP50 da comunidade serem exatamente aqueles com maior *weighted closeness*, apesar de não ser possível constituir uma regressão (pois os dados são bastantes dispersos, originando um R^2 muito baixo). Tal permite-nos aferir que os indivíduos que mais *hits* ouvem são mais centrais nas suas respetivas redes, sendo mais influenciados e influentes. Tal sugere um papel preponderante das interações sociais no fenómeno da difusão das tendências musicais ao nível das comunidades.

Dada a reduzida amostra do user_fav_cluster_3, não foi possível obter um gráfico de dispersão e regressão associada neste cluster.

4.1.5. Determinantes da experiência de consumo

Conforme explicado no ponto 3.3.3, foram desenvolvidos três modelos (*Árvores de Decisão*, *Fast Large Margin* e *Neural Networks*) em cada comunidade (cada *user_fav_cluster*), para entender o que determina o tipo de consumo numa determinada experiência de um utilizador, foi criado um modelo preditivo em que *experience_type* é a variável dependente. Estes modelos encontram-se descritos no anexo e, para cada *user_fav_cluster*, foi selecionado o melhor entre os três: o que tivesse a melhor *accuracy*.

De seguida, são apresentadas as métricas de desempenho dos modelos desenvolvidos para cada *user_fav_cluster*, bem como as observações tiradas. É de notar que os dados obtidos levantam vários pontos interessantes, no entanto, dado que o foco deste trabalho são os *bits* (popularidade) e as redes (*closeness*), a análise focar-se-á principalmente nesses atributos.

user_fav_cluster = cluster_1

Tabela 10 – Métricas de desempenho dos modelos desenvolvidos para o *user_fav_cluster_1*

	Accuracy	Precision		Recall		AUC
		guided_by_ user	guided_by_ spotify	guided_by_ user	guided_by_ spotify	
Fast Large Margin	58%	68%	45%	63%	50%	61%
Decision Tree	69%	73%	61%	78%	56%	71%
Neural Net	68%	72%	60%	81%	47%	69%

Assim, foi selecionada a *Decision Tree*, com uma *accuracy* de 69%.

Neste modelo, verifica-se que faixas com popularidade mais elevadas estão ligadas a experiências guiadas pelo Spotify e que utilizadores com centralidades mais elevadas estão ligados a experiências pelo próprio utilizador. Também a hora do dia e o facto de ser ou não *premium* têm forte influência.

user_fav_cluster = cluster_2

Não foi possível aplicar os modelos, dado o reduzido número de *logs* da amostra.

user_fav_cluster = cluster_3

Tabela 11 – Métricas de desempenho dos modelos desenvolvidos para o *user_fav_cluster_3*

	Accuracy	Precision		Recall		AUC
		guided_by_ user	guided_by_ spotify	guided_by_ user	guided_by_ spotify	
Fast Large Margin	76%	75%	88%	98%	33%	68%
Decision Tree	83%	86%	76%	89%	71%	84%
Neural Net	87%	93%	76%	86%	87%	93%

Assim, foi selecionada a *Neural Net*, com uma *accuracy* de 87%.

Os resultados obtidos mostram que a variável dependente (a experiência) é principalmente influenciada pela *weighted closeness* e a popularidade, sendo que as músicas mais populares são mais ouvidas em experiências guiadas pelo Spotify. Quanto à centralidade, existem tanto nós que ligam a centralidade à experiência guiada pelo utilizador como à guiada pelo Spotify, no entanto, aferimos que os nós onde o peso associado à centralidade é maior são exatamente aqueles em que uma maior *weighted closeness* implica experiências guiadas pelo utilizador.

user_fav_cluster = cluster_4

Tabela 12 – Métricas de desempenho dos modelos desenvolvidos para o user_fav_cluster_4

	Accuracy	Precision		Recall		AUC
		guided_by_ user	guided_by_ spotify	guided_by_ user	guided_by_ spotify	
Fast Large Margin	61%	61%	0%	100%	0%	57%
Decision Tree	66%	71%	56%	74%	53%	67%
Neural Net	60%	71%	49%	59%	62%	65%

Assim, foi selecionada a *Decision Tree*, com uma *accuracy* de 66%.

Com este modelo, é possível verificar que uma menor centralidade está associada a uma experiência guiada pelo Spotify e que faixas menos populares estão ligadas a experiências guiadas pelo utilizador. Neste cluster, também têm forte influência a hora do dia do consumo e o facto de ser ou não uma audição em modo *shuffle*.

user_fav_cluster = cluster_5

Tabela 13 – Métricas de desempenho dos modelos desenvolvidos para o user_fav_cluster_5

	Accuracy	Precision		Recall		AUC
		guided_by_ user	guided_by_ spotify	guided_by_ user	guided_by_ spotify	
Fast Large Margin	60%	60%	42%	97%	4%	56%
Decision Tree	63%	66%	54%	78%	39%	62%
Neural Net	59%	68%	49%	62%	56%	61%

Assim, foi selecionada a *Decision Tree*, com uma *accuracy* de 63%.

Neste modelo, denota-se claramente que popularidades mais baixas estão associadas a experiências guiadas pelo utilizador e não é observável qualquer influência da centralidade. Destaca-se aqui também a influência do cluster da faixa: em popularidades mais baixas, a experiência tende a ser guiada pelo utilizador, quando não se trata do cluster preferido dos utilizadores.

user_fav_cluster = cluster_6

Tabela 14 – Métricas de desempenho dos modelos desenvolvidos para o user_fav_cluster_6

	Accuracy	Precision		Recall		AUC
		guided_by_ user	guided_by_ spotify	guided_by_ user	guided_by_ spotify	
Fast Large Margin	61%	69%	51%	62%	59%	64%
Decision Tree	60%	60%	85%	100%	1%	56%
Neural Net	64%	69%	56%	74%	50%	67%

Assim, foi selecionada a *Neural Net*, com uma *accuracy* de 64%.

Com a análise dos resultados obtidos, vê-se uma distribuição variada dos efeitos da centralidade, tendo tanto nós em que o peso desta variável é expressivo e negativo como outros em que o peso é também acentuado, mas positivo. Por outro lado, o efeito da popularidade é notório: elevadas popularidades estão claramente associadas a experiências guiadas pelo Spotify.

user_fav_cluster = cluster_7

Tabela 15 – Métricas de desempenho dos modelos desenvolvidos para o user_fav_cluster_7

	Accuracy	Precision		Recall		AUC
		guided_by_ user	guided_by_ spotify	guided_by_ user	guided_by_ spotify	
Fast Large Margin	66%	73%	42%	80%	32%	60%
Decision Tree	66%	69%	45%	92%	14%	57%
Neural Net	67%	78%	47%	74%	52%	66%

Assim, foi selecionada a *Neural Net*, com uma *accuracy* de 67%.

Verifica-se que se destacam os efeitos da centralidade dos utilizadores e da popularidade das faixas: indivíduos mais centrais ouvem mais em contextos guiados pelo Spotify e uma maior popularidade da música em questão potencia também que seja ouvida em experiência guiada pelo Spotify.

user_fav_cluster = cluster_8

Tabela 16 – Métricas de desempenho dos modelos desenvolvidos para o user_fav_cluster_8

	Accuracy	Precision		Recall		AUC
		guided_by_ user	guided_by_ spotify	guided_by_ user	guided_by_ spotify	
Fast Large Margin	81%	81%	0%	100%	0%	57%
Decision Tree	83%	85%	59%	95%	30%	73%
Neural Net	82%	83%	58%	97%	19%	69%

Assim, foi selecionada a *Decision Tree*, com uma *accuracy* de 83%.

Com este modelo, verifica-se que o efeito mais evidente é o da centralidade: indivíduos mais

centrais ouvem mais música numa experiência guiada pelo utilizador. Maiores popularidades, de outro modo, são marcadas por experiências guiadas pelo Spotify.

user_fav_cluster = cluster_9

Tabela 17 – Métricas de desempenho dos modelos desenvolvidos para o user_fav_cluster_9

	Accuracy	Precision		Recall		AUC
		guided_by_user	guided_by_spotify	guided_by_user	guided_by_spotify	
Fast Large Margin	74%	74%	0%	100%	0%	54%
Decision Tree	72%	76%	51%	87%	33%	68%
Neural Net	74%	74%	0%	100%	0%	65%

Assim, foi selecionada a *Decision Tree*, com uma *accuracy* de 72%. Neste caso, não foi selecionado o modelo com maior *accuracy* dado que tanto o *Fast Large Margin* como a *Neural Net* concentram todos os resultados numa só classificação (guided_by_user).

No modelo apurado, é possível observar a forte influência da popularidade das faixas que leva a que menores popularidades estejam associadas a experiências guiadas pelo próprio utilizador. Verifica-se ainda que existe mais dispersão quanto à associação entre centralidade e experiência, na medida em que existem tanto casos em maiores centralidades potenciam experiências guiadas pelo Spotify como casos que evidenciam o oposto.

user_fav_cluster = cluster_10

Tabela 18 – Métricas de desempenho dos modelos desenvolvidos para o user_fav_cluster_10

	Accuracy	Precision		Recall		AUC
		guided_by_user	guided_by_spotify	guided_by_user	guided_by_spotify	
Fast Large Margin	69%	69%	0%	100%	0%	62%
Decision Tree	69%	71%	51%	92%	18%	62%
Neural Net	69%	69%	48%	99%	1%	60%

Assim, foi selecionada a *Decision Tree*, com uma *accuracy* de 69%.

Nesta *Decision Tree*, vê-se que popularidades baixas estão principalmente associadas a experiências guiadas pelo utilizador e ainda que a influência do *timing* do consumo e do modo *shuffle* é também notória.

4.2. Interpretação dos Resultados

Recuperando a análise do ponto 3.3.1, denominar-se-á o cluster 1 como EDM, o cluster 2 como *funk* 1, o cluster 3 como música clássica, o cluster 4 como *funk* 2, o cluster 5 como *pop-rock* 1, o cluster 6 como *pop-rock* 2, o cluster 7 como *jazz*, o cluster 8 como *rap*, o cluster 9 como *pop-rock* 3 e o cluster 10 como *pop-rock* 4. Todos os estilos identificados são meras *assumptions*, pelo que estas análises devem ter mais em conta as características marcantes de cada *cluster*, do que propriamente o nome atribuído. Na verdade, com a identificação dos clusters com estilos musicais, pretende-se acima de tudo facilitar a discussão e compreensão dos resultados.

Com a construção das 10 comunidades dos estilos musicais preferidos, foi possível observar que as comunidades com mais expressão são as do *pop-rock* 1, *pop-rock* 2 e *pop-rock* 4. A distribuição dos utilizadores pelas comunidades dos estilos preferidos segue aproximadamente a distribuição das audições por estilo.

É nas *personalized_playlists* e no *user_collection* que os utilizadores mais ouvem músicas diferentes das suas preferências (ou seja, de estilos diferentes do preferido), e é nas *personalized_playlists* e no *catalog* que os utilizadores consomem música de forma mais dispersa, dado que, nestes contextos, existe maior dispersão quanto à popularidade das músicas ouvidas e maior dispersão quanto ao estilo ouvido.

Assim, confirmamos que o consumo guiado pelo utilizador é mais diferenciado face ao guiado pelo algoritmo, com exceção das *personalized playlists*, as *playlists* criadas por algoritmo e curadores, onde figura a maior diversidade musical a nível de género e popularidade. Deste modo, conclusões de estudos anteriores são contrariadas no que toca às *playlists* personalizadas (Freeman et al., 2022; Kaimann et al., 2021; Mok et al., 2022; Petridis et al., 2022; Prey, 2018).

Na análise de redes, pretendeu-se aferir o papel dos indivíduos mais centrais, ou seja, que mais influenciam e que mais são influenciados, na dispersão dos *bits*. Percebemos que, apesar de não ter sido possível encontrar regressões com R^2 satisfatório, existe efetivamente um padrão entre os consumidores de *bits*: aqueles que mais ouvem as músicas mais populares são efetivamente os indivíduos com maior centralidade nas redes das suas comunidades. Ou seja, toda a dinâmica previamente explorada das redes, difusão, EWOM e *opinion leaders* tem também um papel relevante na indústria fonográfica ao nível do consumo de *bits*.

Estas conclusões tiradas são menos evidentes nos gráficos da música clássica (cluster 3) e do

pop-rock 2 (cluster 6). Tal poderá ser explicado possivelmente por características das próprias comunidades: sendo a música clássica um consumo de nicho e, muitas vezes, de músicas mais antigas, é mais difícil encontrar *hits* propriamente ditos, por outro lado, o *pop-rock 2* é o estilo mais ouvido, com os *hits* mais bem-sucedidos, possivelmente com dinâmicas de viralização que já ultrapassam as próprias redes.

Como visto anteriormente, com a última questão de investigação pretendeu-se aferir qual a influência da centralidade dos utilizadores nas suas experiências no Spotify. Sabendo que a popularidade das faixas ouvidas está também ligada às redes, faz obviamente sentido analisar igualmente este efeito. Com o desenvolvimento dos três modelos previamente explorados, foi possível então aferir estas relações. Desde logo, todos os estilos corroboraram a hipótese relativa à popularidade, com exceção do *pop-rock 1*. Já no que concerne à centralidade, não foi possível verificar a relação proposta nos estilos *pop-rock 1*, *pop-rock 2*, *jazz*, *pop-rock 3* e *pop-rock 4*.

Assim, verificamos que em contextos guiados pelo Spotify, os utilizadores acabam a ouvir músicas bem mais populares que nos seus consumos mais pessoais. No entanto, possivelmente esta diferença seria ainda mais notória se retirássemos da experiência guiada pelo Spotify o contexto das *personalized playlists* que, como vimos anteriormente, contém uma vasta dispersão de estilos e popularidades. Por outro lado, não foi possível verificar qualquer relação entre a centralidade dos utilizadores e a experiência no Spotify nos estilos mais genéricos, os *pop-rock*, o que claramente poderá ser uma evidência de que no *mainstream* (são efetivamente os quatro estilos mais ouvidos) a influência da centralidade é menor, dado que o acesso à música e à descoberta é muito mais facilitado tanto pelo Spotify como por outros meios e é nos estilos e comunidades de nicho que as redes e a difusão assumem o maior destaque.

Confirmando então que a centralidade dos *users* impacta o consumo de música e está diretamente relacionada com as audições de *hits*, importa lembrar a importância do EWOM e dos *opinion leaders*. A verdade é que a exposição dos *opinion leaders* às músicas pode bem ser o gatilho para que estas faixas se espalhem dentro das redes e atinjam o sucesso comercial. Sabendo que os picos de consumo de músicas tendem a ser temporalmente próximos do seu lançamento, entende-se que, deste modo, é importante investir logo nos *opinion leaders* (Valente & Vega Yon, 2020). Tal poderá concretizado através de *release parties* (oportunidade chave para convidar indivíduos estratégicos e expô-los à nova música), por parcerias com

criadores de conteúdo nas redes sociais (os vídeos curtos com música de *background* dominam em plataformas como TikTok ou Instagram e são importantes veículos de transmissão das faixas, criando autênticos fenômenos virais), por parcerias com curadores de *playlists* influentes (já existem hoje serviços como o soundcamp.com que fazem esta ligação entre músicos e curadores) e, claro, por uma eficaz estratégia de promoção antes e após o lançamento (por forma a, antes do lançamento, criar a ânsia de conhecer a nova música e, após o lançamento, gerar *buzz* em torno da novidade, curiosidade em ouvir e vontade de partilhar)⁶. Também *playlists* de marcas e/ou *influencers* constituem uma parte importante da difusão e são um ponto de partida sólido para as faixas atingirem novos públicos.

⁶ <https://grin.co/blog/digital-marketing-for-musicians/>

5. Considerações Finais

Neste último capítulo, serão apresentadas as considerações finais, incluindo as limitações inerentes a esta tese e, por fim, a conclusão.

5.1. Limitações

A presente dissertação envolve limitações decorrentes das *assumptions* feitas e dados utilizados.

Primeiramente, o facto de ter sido utilizada a versão mini do MSSD (Brost et al., 2019) limita os resultados obtidos, pelo que se sugere que a mesma análise feita nesta dissertação deva ser aplicada também à base de dados inteira, por forma a verificar se são apuradas as mesmas conclusões.

Por outro lado, o *clustering* das faixas musicais pode ter envolvido vários erros, no entanto, não tendo informações relativas ao nome da música ou identidade do artista, não é possível sequer testar uma amostra dos dados de modo a verificar se a classificação foi corretamente efetuada.

Uma outra limitação que surgiu no decorrer do presente trabalho está relacionada com a ausência de informação quanto ao utilizador em questão: sabe-se apenas o *ID* da sessão e foi mesmo assumido que cada sessão corresponderia a um utilizador diferente.

No seguimento dos dois anteriores pontos, surge uma quarta lacuna nesta investigação: não poderá ser feita uma análise de congruência/consistência dos hábitos de consumo dos utilizadores ao nível de estilos e/ou características musicais, devido à ausência do identificador de utilizador. Seria interessante perceber, por exemplo, se quando um utilizador ouve pela primeira vez uma nova música, tem mais ou menos tendência a voltar a ouvi-la consoante o alinhamento dos seus gostos e hábitos pessoais com o estilo musical da faixa em questão.

5.2. Conclusão

Com os números que a indústria fonográfica tem vindo a atingir no decorrer dos últimos anos, surgiu um especial interesse em entender as dinâmicas gestionárias por trás dos artistas e seus repertórios (Economics, 2020; Hesmondhalgh, 2013; McRobbie, 1999). Esta relevância acentuou-se ainda mais na última década com o surgimento e sucesso do *streaming* que veio transformar profundamente todas as dinâmicas de mercado, desde a oferta à procura (Hujran et al., 2020; Kaimann et al., 2021; Prey et al., 2022; Tepper & Hargittai, 2009; Towse, 2020). Assim, é especialmente importante para os artistas e empresas do setor entender o novo

consumo dos indivíduos.

O Spotify, plataforma líder de mercado, tem hoje um poder ímpar na indústria, sendo inclusive capaz de determinar sucessos, e tem uma série de contextos de audição para os utilizadores totalmente dominados por ele (Aguilar & Waldfogel, 2021; Eriksson, 2020; Morgan, 2020; Prey et al., 2022; Razlogova, 2020; Villermet et al., 2021). Se em *charts* e *playlists* editoriais ainda sabemos os mecanismos que lhes dão origem, em *personalized playlists* e *radio*, estes são completamente desconhecidos, pelo que importa a artistas e *labels* saber dominar outras dinâmicas de mercado como a difusão, as redes ou o EWOM (De Smet & Dhaenens, 2022; Lee et al., 2019; Li & Liu, 2020; Park & Kaneshiro, 2021; Park, Lee, et al., 2022; Villermet et al., 2021). A verdade é que músicas menos populares num determinado momento podem efetivamente tornar-se faixas com elevada popularidade mais tarde (depois da tal difusão) e figurar então numa posição de destaque no Spotify e nas suas dependências como *charts* ou *playlists* editoriais.

Assim, no decorrer deste trabalho, foi desenvolvida uma problemática frequentemente explorada na literatura (De Smet & Dhaenens, 2022; Lee et al., 2019; Li & Liu, 2020; Park & Kaneshiro, 2021; Park, Lee, et al., 2022; Villermet et al., 2021), mas, desta vez, tendo por base de trabalho um *dataset* concreto, real e disponibilizado pelo próprio Spotify: The Music Streaming Sessions Dataset, MSSD (Brost et al., 2019)

No decorrer do trabalho, foram aplicados diversos métodos de análise de dados para entender ligações entre contextos de audição, popularidade, redes e *bits*. Percebeu-se que existe efetivamente uma relação entre a variedade de músicas ouvidas e o contexto de audição: em experiências guiadas pelo utilizador, ouvem-se músicas de diferentes popularidades e diferentes estilos, mas é nas *personalized playlists* que encontramos a maior dispersão de popularidades e estilos, contrariamente ao que é defendido em várias publicações (Aguilar & Waldfogel, 2021; Freeman et al., 2022; Hujran et al., 2020; Kaimann et al., 2021; Maasø & Hagen, 2020; Prey, 2018; Villermet et al., 2021).

Foi também ainda possível aferir que utilizadores mais centrais (influenciam mais e são mais influenciados) têm um maior consumo de *bits* e também se distinguem por ter experiências mais guiadas por si próprios: ou seja, procuram e guardam ativamente músicas, artistas, álbuns, etc. É interessante denotar que esta relação não foi observada nos estilos do *mainstream*, possivelmente, por aí as dinâmicas serem diferentes, na medida em que é o género mais publicitado pelo Spotify e também por outros meios de comunicação, retirando relevância às redes. Por fim, foi possível observar que músicas mais populares estão ligadas a experiências

guiadas pelo próprio Spotify.

O presente trabalho focou-se mais na distinção entre a experiência guiada pelo utilizador e guiada pelo Spotify, mas, como foi concluído, as *personalized playlists* podem também ser interessantes veículos de dispersão de uma determinada música na medida em que são o contexto com a maior variedade de músicas a nível de popularidade e estilo, bem como o contexto onde os utilizadores mais ouvem faixas de estilos diferentes do seu “preferido”. Assim, seria interessante também futuramente entender qual o papel destas *playlists* na difusão e o que leva os utilizadores a optarem por este contexto e, claro, como podem então *record labels* e artistas figurar neste contexto, capitalizando-o para seu benefício.

Importa, por fim, lembrar que a indústria fonográfica foi a primeira indústria cultural a ultrapassar a digitalização e a primeira a ser totalmente dominada pelo *streaming*, pelo que qualquer resultado apurado neste âmbito não é apenas uma conclusão aplicável à indústria da música, mas, sim, uma possível ilustração daquilo que será o futuro de outras indústrias onde a transformação estará ainda a iniciar-se (Maasø, 2018; Prey et al., 2022).

6. Referências Bibliográficas

- Aguiar, L., & Waldfogel, J. (2021). Platforms, Power, and Promotion: Evidence from Spotify Playlists. *The Journal of Industrial Economics*, 69(3), 653-691. <https://doi.org/10.1111/joie.12263>
- Anderson, A., Maystre, L., Anderson, I., Mehrotra, R., & Lalmas, M. (2020). *Algorithmic Effects on the Diversity of Consumption on Spotify* Proceedings of The Web Conference 2020, Taipei, Taiwan.
- Barone, M. D., Bansal, J., & Woolhouse, M. H. (2017). Acoustic Features Influence Musical Choices Across Multiple Genres. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.00931>
- Brooks, S., & Conroy, T. (2010). Hip-Hop Culture in a Global Context: Interdisciplinary and Cross-Categorical Investigation. *American Behavioral Scientist*, 55(1), 3-8. <https://doi.org/10.1177/0002764210381723>
- Brost, B., Mehrotra, R., & Jehan, T. (2019). The music streaming sessions dataset. In *The World Wide Web Conference*, 2594-2600.
- Clark, A. B., & Lonsdale, A. J. (2022). Music preference, social identity, and collective self-esteem. *Psychology of Music*, 51(4), 1119-1131. <https://doi.org/10.1177/03057356221126202>
- Colbjørnsen, T. (2021). The streaming network: Conceptualizing distribution economy, technology, and power in streaming media services. *Convergence: The International Journal of Research into New Media Technologies*, 27(5), 1264-1287. <https://doi.org/10.1177/1354856520966911>
- Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695. <https://doi.org/10.5281/zenodo.7682609>
- De Assunção, W. G., & Zaina, L. A. M. (2022). Evaluating user experience in music discovery on deezer and spotify. Proceedings of the 21st Brazilian Symposium on Human Factors in Computing Systems, Diamantina, Brazil.
- De Smet, B., & Dhaenens, F. (2022). “Music Makes the People Come Together”. In *Identities and Intimacies on Social Media* (pp. 66-81). Routledge. <https://doi.org/10.4324/9781003250982-6>
- Economics, O. (2020). *The Economic Impact of Music in Europe*. https://www.ifpi.org/wp-content/uploads/2020/11/IFPI_music_in_Europe_WEB_spreads.pdf
- Eriksson, M. (2020). The editorial playlist as container technology: on Spotify and the

- logistical role of digital music packages. *Journal of Cultural Economy*, 13(4), 415-427.
<https://doi.org/10.1080/17530350.2019.1708780>
- Freeman, S., Gibbs, M., & Nansen, B. (2022). 'Don't mess with my algorithm': Exploring the relationship between listeners and automated curation and recommendation on music streaming services. *First Monday*. <https://doi.org/10.5210/fm.v27i1.11783>
- Gulati, A., & Eirinaki, M. (2019). With a Little Help from My Friends (and Their Friends): Influence Neighborhoods for Social Recommendations. The World Wide Web Conference, San Francisco, CA, USA.
- Gulmatico, J., Susa, J., Malbog, M. A., Acoba, A., Nipas, M., & Nicolas-Mindoro, J. (2022). *SpotiPred: A Machine Learning Approach Prediction of Spotify Music Popularity by Audio Features* 2022 Second International Conference on Power, Control and Computing Technologies (ICPC2T), Raipur, India.
- Habeeb, H. (2024). *Wrapped Insights: A Data-Driven Approach to Personalizing User Experiences in a Digital Tipping Platform* CUNY Academic Works.
https://academicworks.cuny.edu/gc_etds/5907
- Hagen, A. N., & Lüders, M. (2017). Social streaming? Navigating music as personal and social. *Convergence: The International Journal of Research into New Media Technologies*, 23(6), 643-659. <https://doi.org/10.1177/1354856516673298>
- Harrison, A. (2021). *Music: The Business (8th edition)*. Ebury Publishing.
- Heggli, O. A., Stupacher, J., & Vuust, P. (2021). Diurnal fluctuations in musical preference. *Royal Society Open Science*, 8(11). <https://doi.org/10.1098/rsos.210885>
- Henrique, L. (2019). *Top Spotify songs from 2010-2019 - BY YEAR* (Version 1).
<https://www.kaggle.com/datasets/leonardopena/top-spotify-songs-from-20102019-by-year>
- Hesmondhalgh, D. (2013). *Why Music Matters*. Wiley.
- Hosey, C., Vujović, L., St. Thomas, B., Garcia-Gathright, J., & Thom, J. (2019). Just Give Me What I Want. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland Uk.
- Huete-Alcocer, N. (2017). A Literature Review of Word of Mouth and Electronic Word of Mouth: Implications for Consumer Behavior [Mini Review]. *Frontiers in Psychology*, 8.
<https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2017.01256>
- Hujran, O., Alikaj, A., Durrani, U. K., & Al-Dmour, N. (2020). Big Data and its Effect on

- the Music Industry. Proceedings of the 3rd International Conference on Software Engineering and Information Management, Sydney, NSW, Australia.
- IFPI (International Federation of the Phonographic Industry). (2024). *GLOBAL MUSIC REPORT 2024 — STATE OF THE INDUSTRY*. <https://globalmusicreport.ifpi.org/>
- Ignatius Moses Setiadi, D. R., Satriya Rahardwika, D., Rachmawanto, E. H., Atika Sari, C., Irawan, C., Kusumaningrum, D. P., Nuri, & Trusthi, S. L. (2020). Comparison of SVM, KNN, and NB Classifier for Genre Music Classification based on Metadata. 2020 International Seminar on Application for Technology of Information and Communication (iSemantic), Semarang, Indonesia.
- Jun, S., Kim, D., Jeon, M., Rho, S., & Hwang, E. (2015). Social mix: automatic music recommendation and mixing scheme based on social network analysis. *The Journal of Supercomputing*, 71(6), 1933-1954. <https://doi.org/10.1007/s11227-014-1182-1>
- Kaimann, D., Tanneberg, I., & Cox, J. (2021). “I will survive”: Online streaming and the chart survival of music tracks. *Managerial and Decision Economics*, 42(1), 3-20. <https://doi.org/10.1002/mde.3226>
- Kovács, L., & Ghous, H. (2020). Efficiency comparison of Python and RapidMiner. *Multidiszciplináris Tudományok*, 10(3), 212-220. <https://doi.org/10.35925/j.multi.2020.3.26>
- Kriegeskorte, N., & Golan, T. (2019). Neural network models and deep learning. *Current Biology*, 29(7), R231-R236. <https://doi.org/10.1016/j.cub.2019.02.034>
- Kumar, P., & Sinha, A. (2021). Information diffusion modeling and analysis for socially interacting networks. *Social Network Analysis and Mining*, 11(1). <https://doi.org/10.1007/s13278-020-00719-7>
- Lee, J. H., Pritchard, L., & Hubbles, C. (2019). Can We Listen To It Together?: Factors Influencing Reception of Music Recommendations and Post-Recommendation Behavior. 20th ISMIR Conference, Delft, The Netherlands.
- Leong, T. W., & Wright, P. C. (2013). Revisiting social practices surrounding music. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Paris, France.
- Li, G., & Liu, R. (2020). Original Music Album Diffusion Sustainability in Social Network-Based Communities: A Network Embedded Perspective. *IEEE Access*, 8, 53107-53115. <https://doi.org/10.1109/access.2020.2977947>

- Litterio, A. M., Nantes, E. A., Larrosa, J. M., & Gómez, L. J. (2017). Marketing and social networks: a criterion for detecting opinion leaders. *European Journal of Management and Business Economics*, 26(3), 347-366. <https://doi.org/10.1108/ejmbe-10-2017-020>
- Lüders, M. (2021). Pushing music: People's continued will to archive versus Spotify's will to make them explore. *European Journal of Cultural Studies*, 24(4), 952-969. <https://doi.org/10.1177/1367549419862943>
- Lyu, X. (2021). Role of underlying diffusion mechanism and information digitization in a global network of streaming music. 2021 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS), Shenyang, China.
- Maasø, A. (2018). Music Streaming, Festivals, and the Eventization of Music. *Popular Music and Society*, 41(2), 154-175. <https://doi.org/10.1080/03007766.2016.1231001>
- Maasø, A., & Hagen, A. N. (2020). Metrics and decision-making in music streaming. *Popular Communication*, 18(1), 18-31. <https://doi.org/10.1080/15405702.2019.1701675>
- Macdonald, D. (1953). A Theory of Mass Culture. *Diogenes*, 1(3), 1-17. <https://doi.org/10.1177/039219215300100301>
- Mall, A. (2018). Concentration, diversity, and consequences: Privileging independent over major record labels. *Popular Music*, 37(3), 444-465. <https://doi.org/10.1017/S0261143018000375>
- Mcrobbie, A. (1999). *In the Culture Society*. Routledge. <https://doi.org/10.4324/9781315004808>
- Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., & Euler, T. (2006). YALE: Rapid Prototyping for Complex Data Mining Tasks. Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, Philadelphia, PA, USA.
- Mok, L., Way, S. F., Maystre, L., & Anderson, A. (2022). The Dynamics of Exploration on Spotify. *Proceedings of the International AAAI Conference on Web and Social Media*, 16, 663-674. <https://doi.org/10.1609/icwsm.v16i1.19324>
- Morgan, B. A. (2020). Revenue, access, and engagement via the in-house curated Spotify playlist in Australia. *Popular Communication*, 18(1), 32-47. <https://doi.org/10.1080/15405702.2019.1649678>
- Negus, K. (2019). From creator to data: the post-record music industry and the digital conglomerates. *Media, Culture & Society*, 41(3), 367-384. <https://doi.org/10.1177/0163443718799395>

- Nijkamp, R. (2018). *Prediction of product success: explaining song popularity by audio features from Spotify data* University of Twente]. <http://essay.utwente.nl/75422/>
- Noah. (2021). *Spotify Genre Audio features* (Version 1). <https://www.kaggle.com/datasets/naoh1092/spotify-genre-audio-features>
- Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node Centrality in Weighted Networks: Generalizing Degree and Shortest Paths. *Social Networks - SOC NETWORKS*, 32, 245-251. <https://doi.org/10.1016/j.socnet.2010.03.006>
- Park, S. Y., & Kaneshiro, B. (2021). Social Music Curation That Works. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), 1-27. <https://doi.org/10.1145/3449191>
- Park, S. Y., Lee, K. Y., & Lee, J. H. (2022). Cross-Cultural Exploration of Music Sharing. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2), 1-28. <https://doi.org/10.1145/3555108>
- Park, S. Y., Redmond, E., Berger, J., & Kaneshiro, B. (2022). Hitting Pause: How User Perceptions of Collaborative Playlists Evolved in the United States During the COVID-19 Pandemic. CHI Conference on Human Factors in Computing Systems, New Orleans, LA, USA.
- Petridis, S., Daskalova, N., Mennicken, S., Way, S. F., Lamere, P., & Thom, J. (2022). TastePaths: Enabling Deeper Exploration and Understanding of Personal Preferences in Recommender Systems. 27th International Conference on Intelligent User Interfaces, Helsinki, Finland.
- Prey, R. (2018). Nothing personal: algorithmic individuation on music streaming platforms. *Media, Culture & Society*, 40(7), 1086-1100. <https://doi.org/10.1177/0163443717745147>
- Prey, R., Esteve Del Valle, M., & Zwerwer, L. (2022). Platform pop: disentangling Spotify's intermediary role in the music industry. *Information, Communication & Society*, 25(1), 74-92. <https://doi.org/10.1080/1369118x.2020.1761859>
- Rahman, R. A., Masrom, S., Mohamad, M., Sari, E. N., Saragih, F., & Rahman, A. S. A. (2023). Comparisons of automated machine learning (AutoML) in predicting whistleblowing of academic dishonesty with demographic and theory of planned behavior. *MethodsX*, 11, 102364. <https://doi.org/10.1016/j.mex.2023.102364>
- Razlogova, E. (2020). Provincializing Spotify: Radio, algorithms and conviviality. *Radio Journal: International Studies in Broadcast & Audio Media*, 18(1), 29-42. https://doi.org/10.1386/rjao_00014_1

- R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. In R Foundation for Statistical Computing. <https://www.R-project.org/>
- Ren, J., & Kauffman, R. J. (2017). Understanding music track popularity in a social network. Proceedings of the 25th European Conference on Information Systems ECIS, Guimarães, Portugal.
- RIAA (Recording Industry Association of America). (2020). *The U.S. Music Industries: Jobs & Benefits*. <https://www.riaa.com/wp-content/uploads/2021/02/The-U.S.-Music-Industries-Jobs-Benefits-2020-Report-Executive-Summary.pdf>
- Sandag, G. A., & Manueke, A. M. (2020, 2020-10-27). Predictive Models for Popularity of Solo and Group Singers in Spotify Using Decision Tree. 2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS), Manado, Indonesia.
- Sanjeev Verma and Neha, Y. (2021). Past, Present, and Future of Electronic Word of Mouth (EWOM). *Journal of Interactive Marketing*, 53, 111-128. <https://doi.org/https://doi.org/10.1016/j.intmar.2020.07.001>
- Siles, I., Segura-Castillo, A., Sancho, M., & Solís-Quesada, R. (2019). Genres as Social Affect: Cultivating Moods and Emotions through Playlists on Spotify. *Social Media + Society*, 5(2), 205630511984751. <https://doi.org/10.1177/2056305119847514>
- Singh, S. (2019). *Spotify - All Time Top 2000s Mega Dataset* (<https://www.kaggle.com/datasets/iamsumat/spotify-top-2000s-mega-dataset>)
- Snickars, P. (2017). More of the Same – On Spotify Radio. *Culture Unbound*, 9(2), 184-211. <https://doi.org/10.3384/cu.2000.1525.1792184>
- Tepper, S. J., & Hargittai, E. (2009). Pathways to music exploration in a digital age. *Poetics*, 37(3), 227-249. <https://doi.org/10.1016/j.poetic.2009.03.003>
- Ternovski, J., & Yasseri, T. (2020). Social complex contagion in music listenership: A natural experiment with 1.3 million participants. *Social Networks*, 61, 144-152. <https://doi.org/10.1016/j.socnet.2019.10.005>
- Towse, R. (2020). Dealing with digital: the economic organisation of streamed music. *Media, Culture & Society*, 42(7-8), 1461-1478. <https://doi.org/10.1177/0163443720919376>
- Trocchia, P. J., Apps, M. M., & McNish, S. E. (2002). A typology of music consumption motivations. *International Business & Economics Research Journal (IBER)*, 1(9).
- Valente, T. W., & Vega Yon, G. G. (2020). Diffusion/Contagion Processes on Social Networks. *Health Education & Behavior*, 47(2), 235-248. <https://doi.org/10.1177/1090198120901497>

- Verma, S., & Yadav, N. (2021). Past, Present, and Future of Electronic Word of Mouth (EWOM). *Journal of Interactive Marketing*, 53(1), 111-128. <https://doi.org/10.1016/j.intmar.2020.07.001>
- Villermet, Q., Poiroux, J., Moussallam, M., Louail, T., & Roth, C. (2021). Follow the guides: disentangling human and algorithmic curation in online music consumption. *Proceedings of the 15th ACM Conference on Recommender Systems*, 380-389. <https://doi.org/10.1145/3460231.3474269>
- Way, S. F., Garcia-Gathright, J., & Cramer, H. (2020). Local Trends in Global Music Streaming. *Proceedings of the International AAAI Conference on Web and Social Media*, 14, 705-714. <https://doi.org/10.1609/icwsm.v14i1.7336>
- Williams, R. (1958). *Culture and society*. Chatto & Windus. <https://doi.org/0-7012-0792-2>
- Wu, Y.-C., & Feng, J.-W. (2018). Development and Application of Artificial Neural Network. *Wireless Personal Communications*, 102(2), 1645-1656. <https://doi.org/10.1007/s11277-017-5224-x>
- Zhang, Q., & Negus, K. (2021). Stages, Platforms, Streams: The Economies and Industries of Live Music after Digitalization. *Popular Music and Society*, 44(5), 539-557. <https://doi.org/10.1080/03007766.2021.1921909>
- Zhao, Y., Wang, L., Tang, H., & Zhang, Y. (2020). Electronic word-of-mouth and consumer purchase intentions in social e-commerce. *Electronic Commerce Research and Applications*, 41, 100980. <https://doi.org/https://doi.org/10.1016/j.elerap.2020.100980>

7. Anexo

7.1. Processos informáticos (RapidMiner e R)

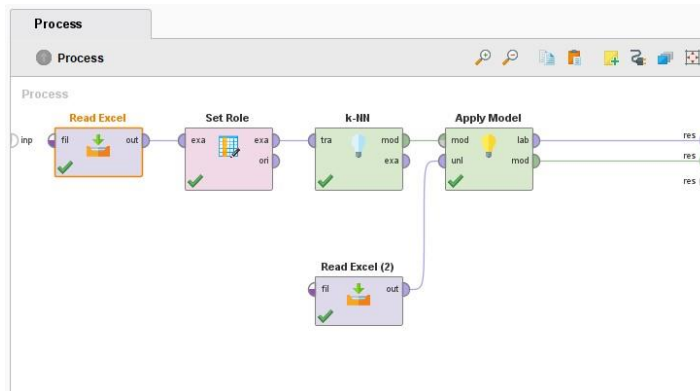


Figura 29 – Classificação das faixas em estilos musicais (KNN): processo em RapidMiner

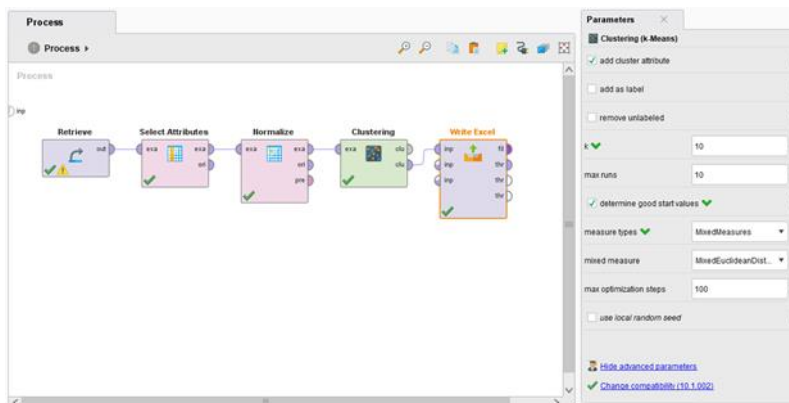


Figura 30 – Clustering das faixas (K-Means Clustering): processo em RapidMiner

Tabela 19 – Código R para obter a *weighted_closeness* dos indivíduos a partir da matriz da distribuição dos logs dos users por cluster

```
#1. Chamar a livreria igraph
library(igraph)

#2. Calcular matriz de distâncias
#cada linha
d<-dist(log_database)
d<-as.matrix(1/d)
d<-d>0.015
diag(d) <- 0

#3. Ler e converter os dados para serem trabalhados no packa-ge igraph
adj_matrix<-d
non_zero_elements <- which(adj_matrix != 0, arr.ind = TRUE)
#4. Encontrar os elementos diferentes de zero na matriz de adjacência
non_zero_elements <- which(adj_matrix != 0, arr.ind = TRUE)

#5. Obter os nomes dos nós
nodes <- rownames(adj_matrix)

#6. Criar a listagem de arestas com pesos
edge_list <- data.frame(
  from = nodes[non_zero_elements[, 1]],
  to = nodes[non_zero_elements[, 2]],
  weight = adj_matrix[non_zero_elements]
)

#7. Criar o grafo
graph <- graph_from_data_frame(edge_list, directed = FALSE)

#8. Calcular métricas
weighted_closeness <- closeness(graph, weights = E(graph)$weight)

#9. Exportar csv
write.csv2(weighted_closeness, "weighted_closenesses.csv",
row.names=TRUE)
```

7.2. Distribuição dos dados

Tabela 20 – Distribuição de todos os logs por context_type e not_skipped

context_type	editorial_playlist		user_collection		catalog		charts		radio		personalized_playlist		Total	
	35663		67430		38263		2441		21107		2976		167880	
	21%		40%		23%		1%		13%		2%		100%	
not_skipped	True	False	True	False	True	False	True	False	True	False	True	False	True	False
	12191	23472	21314	46116	14198	24065	733	1708	6405	14702	1043	1933	55884	111996
	34%	66%	32%	68%	37%	63%	30%	70%	30%	70%	35%	65%	33%	67%

7.3. Modelos desenvolvidos para classificação das experiências de consumo

Equação 1 – Fast Large Margin do user_fav_cluster_1

$$\begin{aligned}
 \text{experience_type} = & \\
 & - 0,582 * \text{hist_user_behavior_reason_start} + 0,023 * \text{premium} \\
 & - 0,163 * \text{hist_user_behavior_is_shuffle} + 0,010 * \text{song_US_popularity} \\
 & + 0,083 * \text{track_cluster_is_fav_cluster} + 0,003 * \text{hour_of_day} \\
 & - 0,003 * \text{fav_cluster_user_weighted_closeness} + 0,213
 \end{aligned}$$

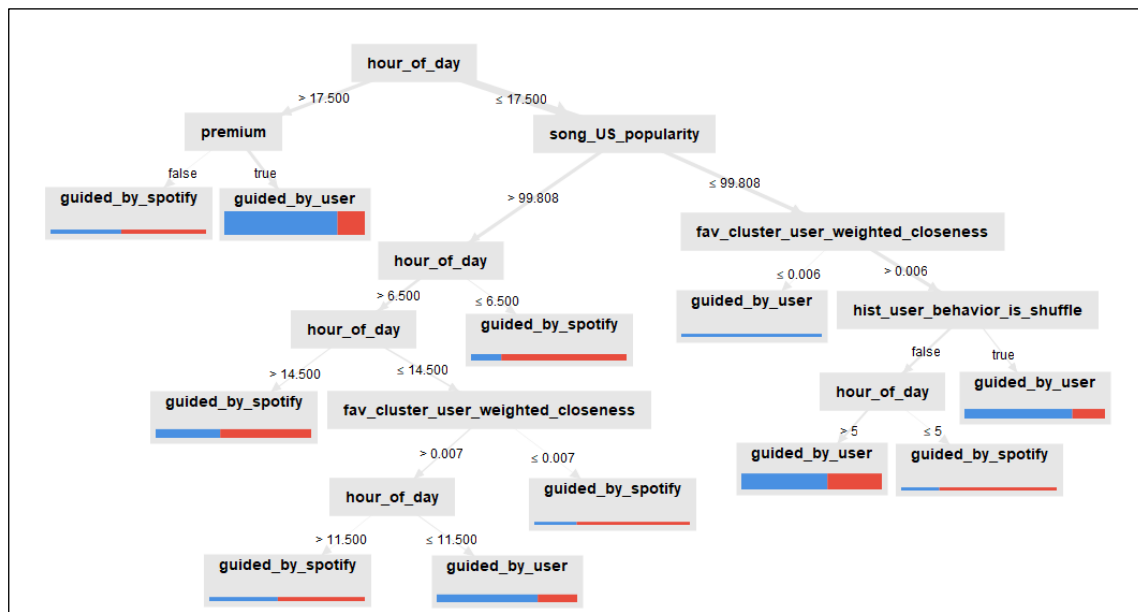


Figura 31 – Árvore de Decisão do user_fav_cluster_1

Tabela 21 – Neural Network do user_fav_cluster_1

Hidden Layer 1							
Node		1	2	3	4	5	6
track_cluster_is_fav_cluster	= true	4,3	0,1	-0,1	-0,4	-3,5	0,1
	= false	-4,4	-0,1	0,1	0,4	3,5	0,0
hist_user_behavior_is_shuffle	= true	5,2	4,7	-2,5	2,0	2,1	-0,2
	= false	-5,2	-4,7	2,4	-1,9	-2,0	0,2

premium	= true	3,0	-2,8	-0,3	1,1	3,4	-0,7
	= false	-3,0	2,8	0,3	-1,1	-3,4	0,7
song_US_popularity		-10,1	9,8	-3,5	-7,8	0,6	-2,1
fav_cluster_user_weighted_closeness		-15,1	-6,8	-7,4	-3,5	-15,4	6,3
Bias		1,6	-3,6	1,3	-0,7	-0,7	-6,8
Output							
Class (sigmoid)		U	S	U	S	U	S
		3,8	-3,8	-3,6	3,6	-3,2	3,2
		2,6	-2,6	3,0	-3,0	2,3	-2,3

U = guided_by_user; S = guided_by_spotify

Equação 2 – Fast Large Margin do user_fav_cluster_3

$$\begin{aligned}
 \text{experience_type} = & \\
 & - 0,033 * \text{not_skipped} + 0,009 * \text{premium} \\
 & - 0,551 * \text{hist_user_behavior_is_shuffle} + 0,004 * \text{song_US_popularity} \\
 & - 0,062 * \text{track_cluster_is_fav_cluster} + 0,008 * \text{hour_of_day} \\
 & + 0,002 * \text{fav_cluster_user_weighted_closeness} + 0,042
 \end{aligned}$$

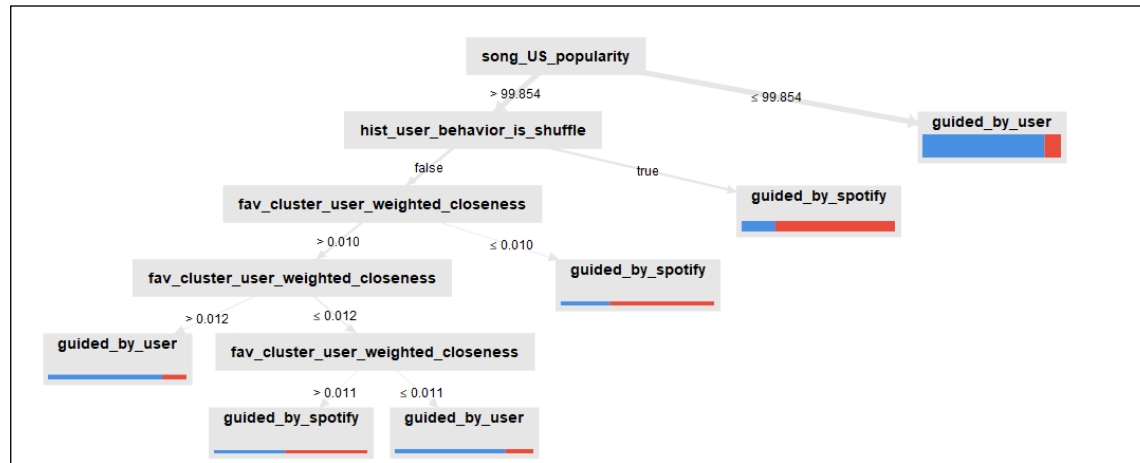


Figura 32 – Árvore de Decisão do user_fav_cluster_3

Tabela 22 – Neural Network do user_fav_cluster_3

Hidden Layer 16							
Node		1	2	3	4	5	6
track_cluster_is_fav_cluster	= true	1,2	-7,3	0,4	0,6	-1,8	-4,1
	= false	-1,1	7,3	-0,3	-0,6	1,8	4,1
hist_user_behavior_is_shuffle	= true	-4,0	-5,5	2,2	-5,1	-5,5	-3,3
	= false	4,0	5,5	-2,2	5,0	5,6	3,4
premium	= true	-3,0	6,7	18,5	1,2	-6,2	-1,3
	= false	3,1	-6,8	-18,5	-1,2	6,2	1,3
song_US_popularity		-9,7	-10,2	-57,6	-10,0	-19,3	-1,7
fav_cluster_user_weighted_closeness		-2,6	-18,4	-5,8	35,8	33,3	-28,8
Bias		0,0	8,1	17,9	0,1	-4,4	-1,2
Output							
Class (sigmoid)		U	S	U	S	U	S
		6,9	-6,9	1,9	-1,9	3,2	-3,2
		3,4	-3,4	3,9	-3,9	1,8	-1,8

U = guided_by_user; S = guided_by_spotify

Equação 3 – Fast Large Margin do *user_fav_cluster_4*

$$\begin{aligned}
 \text{experience_type} = & \\
 & - 0,003 * \text{hour_of_day} + 0,094 * \text{premium} \\
 & - 1,452 * \text{hist_user_behavior_is_shuffle} - 0,015 * \text{song_US_popularity} \\
 & + 0,088 * \text{track_cluster_is_fav_cluster} + 1,030 \\
 & + 0,002 * \text{fav_cluster_user_weighted_closeness}
 \end{aligned}$$

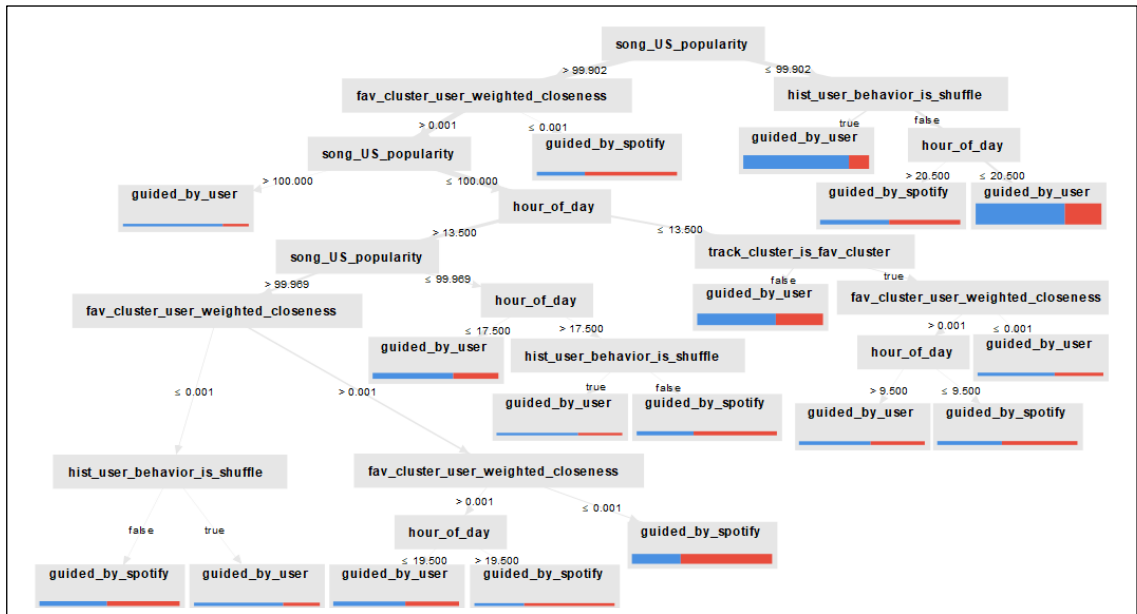


Figura 33 – Árvore de Decisão do *user_fav_cluster_4*

Tabela 23 – Neural Network do *user_fav_cluster_4*

Hidden Layer 1																	
Node		1		2		3		4		5		6					
track_cluster_is_fav_cluster	= true	-4,6		-11,3		0,1		-0,5		-0,4		0,2					
	= false	4,6		11,3		0,0		0,5		0,4		-0,2					
hist_user_behavior_is_shuffle	= true	5,9		-3,7		14,7		0,3		8,3		-0,1					
	= false	-5,9		3,7		-14,7		-0,4		-8,2		0,1					
premium	= true	-0,7		-1,6		1,4		2,2		-7,1		-0,5					
	= false	0,7		1,5		-1,3		-2,3		7,1		0,4					
song_US_popularity		-4,2		2,9		-21,5		-40,6		-10,7		-41,9					
fav_cluster_user_weighted_closeness		-20,1		-20,5		-35,5		36,8		-15,8		-0,1					
Bias		-10,4		-14,0		11,2		1,5		-3,1		38,3					
Output																	
Class (sigmoid)		U		S		U		S		U		S		U		S	
		0,8	-0,8	1,2	-1,2	0,6	-0,6	0,7	-0,7	1,1	-1,1	1,4	-1,4				

U = guided_by_user; S = guided_by_spotify

Equação 4 – Fast Large Margin do *user_fav_cluster_5*

$$\begin{aligned}
 \text{experience_type} = & \\
 & - 2,176 * \text{hist_user_behavior_reason_start} & - 0,005 * \text{premium} \\
 & - 1,562 * \text{hist_user_behavior_is_shuffle} & + 0,021 * \text{song_US_popularity} \\
 & + 0,042 * \text{track_cluster_is_fav_cluster} & + 0,002 * \text{hour_of_day} \\
 & + 0,000 * \text{fav_cluster_user_weighted_closeness} & + 0,389
 \end{aligned}$$

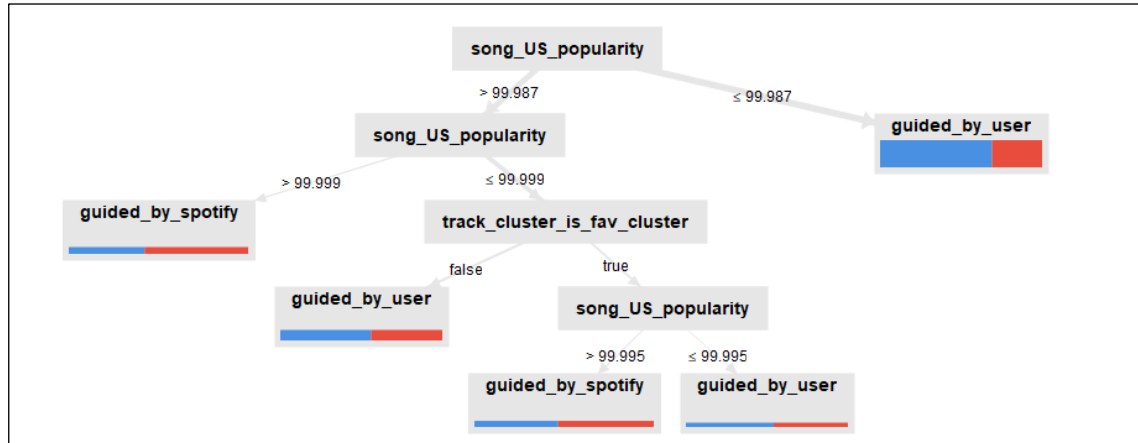


Figura 34 – Árvore de Decisão do *user_fav_cluster_5*

Tabela 24 – Neural Network do *user_fav_cluster_5*

Hidden Layer 1												
Node		1	2	3	4	5	6	7	8	9	10	11
track_cluster_is_fav_cluster	= true	-4,8	-1,2	-2,2	-3,4	-0,3	0,7	-0,2	4,6	0,1	0,8	0,2
	= false	4,7	1,2	2,2	3,5	0,3	-0,7	0,2	-4,5	-0,1	-0,7	-0,1
hist_user_behavior_is_shuffle	= true	10,4	-1,1	-5,5	2,0	-1,3	1,7	3,6	8,3	4,1	-5,5	-3,0
	= false	-10,3	1,1	5,4	-2,0	1,3	-1,7	-3,7	-8,4	-4,1	5,6	3,0
premium	= true	2,1	-0,3	-1,6	2,6	-4,4	5,0	-6,1	2,9	-1,6	5,5	-3,5
	= false	-2,2	0,3	1,6	-2,6	4,4	-5,0	6,1	-2,9	1,6	-5,5	3,5
hist_user_behavior_reason_start	= fwdbtn	10,6	-0,5	-3,6	0,4	-0,6	-1,7	6,4	9,4	2,9	-8,6	1,3
	= clickrow	3,4	2,0	-1,6	0,0	-1,3	12,0	-7,6	5,1	4,0	13,6	5,9
	= appload	-0,4	0,2	-2,3	0,9	-0,5	4,0	2,4	0,2	0,2	-0,3	0,5
	= trackdone	2,4	2,9	0,5	2,5	0,5	-0,6	6,4	-0,7	5,8	5,7	-0,3
	= backbtn	-0,8	1,3	10,2	9,3	6,0	-0,1	7,7	10,1	2,0	7,6	0,0
	= playbtn	1,9	0,5	-1,9	1,7	-0,1	2,2	2,6	-0,2	-0,1	-1,1	0,5
	= remote	2,8	1,5	3,3	1,6	1,1	3,0	3,2	2,6	1,3	2,5	1,9
	= trackerror	-0,2	1,5	6,3	-1,1	1,0	3,2	-1,9	4,0	0,9	3,0	2,0
	= endplay	3,3	1,7	2,1	2,5	1,1	3,8	3,3	5,1	2,9	7,0	2,0
song_US_popularity		-7,4	0,3	1,1	0,0	-11,9	8,6	-9,9	2,3	-5,2	2,2	0,4
fav_cluster_user_weighted_closeness		-20,7	-1,6	-8,3	-9,6	-3,7	-8,8	4,5	-10,4	-8,2	8,4	-4,1
hour_of_day		13,7	0,6	2,4	-12,6	4,5	9,8	-35,9	15,4	31,6	-12,9	-8,1
Bias		-3,2	-1,6	-1,9	-2,5	-1,0	-3,6	-3,3	-5,1	-2,8	-4,2	-2,0

		Output																	
Class (sigmoid)		U	S	P	S	P	S	P	S	P	S	P	S	P	S	P	S	P	S
		0,7	-0,7	0,5	-0,5	1,0	-1,0	0,8	-0,8	0,9	-0,9	0,9	-0,9	1,3	-1,3	0,8	-0,8	0,9	-0,9

U = guided_by_user; S = guided_by_spotify

Equação 5 – Fast Large Margin do *user_fav_cluster_6*

$$\begin{aligned}
 \text{experience_type} = & \\
 & - 2,261 * \text{hist_user_behavior_reason_start} & - 2,250 * \text{premium} \\
 & - 2,444 * \text{hist_user_behavior_is_shuffle} & + 0,006 * \text{song_US_popularity} \\
 & - 1,797 * \text{track_cluster_is_fav_cluster} & - 0,004 * \text{hour_of_day} \\
 & + 0,002 * \text{fav_cluster_user_weighted_closeness} & + 2,243
 \end{aligned}$$

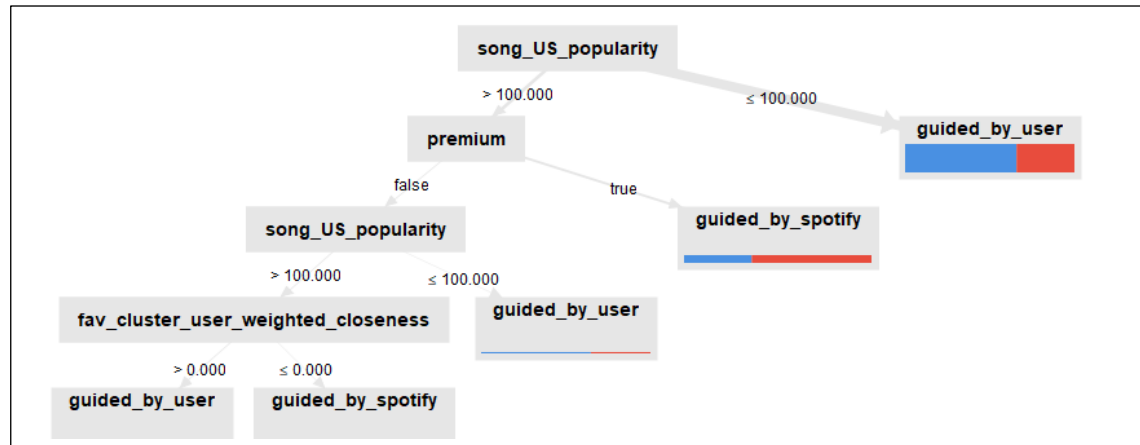


Figura 35 – Árvore de Decisão do *user_fav_cluster_6*

Tabela 25 – Neural Network do *user_fav_cluster_6*

Hidden Layer 1												
Node		1	2	3	4	5	6	7	8	9	10	11
track_cluster_is_fav_cluster	= true	-16,0	0,4	-0,4	13,1	-3,9	-13,3	-0,4	-3,3	-0,6	5,3	0,8
	= false	16,0	-0,4	0,5	-13,0	3,9	13,3	0,4	3,4	0,6	-5,3	-0,8
hist_user_behavior_is_shuffle	= true	-9,7	13,8	5,2	0,2	3,4	8,1	-2,0	-3,0	9,8	-7,3	-8,1
	= false	9,8	-13,7	-5,2	-0,2	-3,5	-8,1	2,0	2,9	-9,8	7,4	8,1
premium	= true	-3,2	-7,0	-1,3	0,2	0,3	2,4	-13,3	-3,7	0,3	-9,4	5,8
	= false	3,1	7,0	1,3	-0,2	-0,4	-2,4	13,3	3,7	-0,3	9,4	-5,9
hist_user_behavior_reason_start	= clickrow	10,4	21,0	4,0	-8,5	2,5	1,5	-9,8	2,1	17,2	5,5	4,1
	= trackdone	6,5	-11,8	0,1	-6,9	6,4	4,0	2,0	1,2	-3,6	17,1	-3,8
	= fwdbtn	-8,2	-10,2	2,5	-8,1	11,2	5,0	1,6	7,4	-0,4	2,2	-9,8
	= apload	4,5	-27,7	1,7	-7,5	1,5	2,5	-0,7	0,9	-2,9	-3,3	1,4
	= backbtn	-3,7	16,8	11,0	-8,6	6,3	-1,0	24,5	2,1	4,5	3,5	11,2
	= endplay	5,6	-7,3	4,0	-5,6	7,1	3,7	4,4	4,3	3,8	7,6	4,0
	= playbtn	5,9	-6,3	-0,3	-12,2	2,6	2,6	1,2	2,4	-4,2	0,7	1,7
	= remote	1,2	-8,4	-1,3	-3,5	2,8	4,6	4,4	3,2	6,2	8,5	12,2
	= trackerror	4,9	-6,7	5,7	-3,8	6,3	2,8	2,6	3,1	6,0	10,1	6,2
song_US_popularity		9,6	-31,8	-11,8	-92,9	-7,4	-9,0	-7,4	-6,5	-6,9	-9,2	-5,0
fav_cluster_user_weighted_closeness		-26,8	29,7	17,6	4,1	26,7	-25,0	-18,6	-0,6	6,1	3,7	-12,7
hour_of_day		-4,1	-11,8	14,4	0,8	15,0	3,4	-23,2	11,1	-19,8	23,2	-7,6
Bias		-3,9	5,7	-3,9	9,3	-6,7	-3,6	-4,3	-3,8	-3,8	-7,4	-3,9

Output																				
Class (sigmoid)	U	S	U	S	U	S	U	S	U	S	U	S	U	S	U	S	U	S	U	S
	0,3	-0,3	0,4	-0,4	0,6	-0,6	1,3	-1,3	0,2	-0,2	1,4	-1,4	1,2	-1,2	0,0	0,0	0,9	-0,9	0,6	-0,6
U = guided_by_user; S = guided_by_spotify																				

U = guided_by_user; S = guided_by_spotify

Equação 6 – Fast Large Margin do *user_fav_cluster_7*

$$\begin{aligned}
 \text{experience_type} = & \\
 & - 1,613 * \text{hist_user_behavior_reason_start} \quad - 1,220 * \text{premium} \\
 & - 1,035 * \text{hist_user_behavior_is_shuffle} \quad + 0,002 * \text{hour_of_day} \\
 & + 0,447 * \text{track_cluster_is_fav_cluster} \quad + 1,403 \\
 & + 0,004 * \text{fav_cluster_user_weighted_closeness}
 \end{aligned}$$

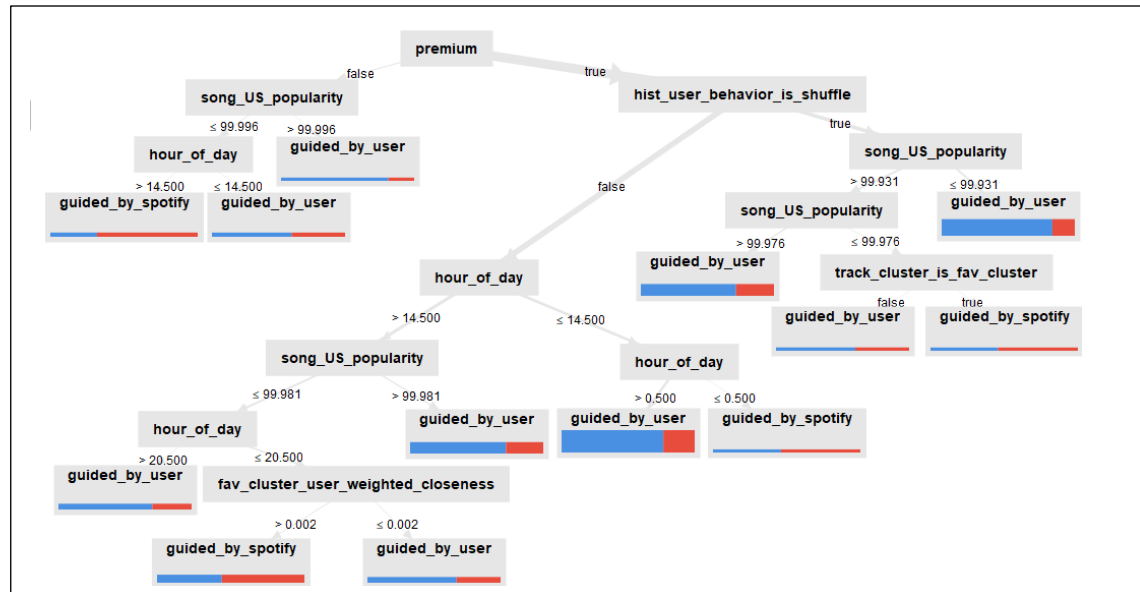


Figura 36 – Árvore de Decisão do *user_fav_cluster_7*

Tabela 26 – Neural Network do *user_fav_cluster_7*

Hidden Layer 1												
Node		1	2	3	4	5	6	7	8	9	10	11
track_cluster_is_fav_cluster	= true	9,2	0,2	2,4	1,8	-0,5	1,3	-0,5	-10,8	-2,8	1,4	2,1
	= false	-9,2	-0,2	-2,4	-1,8	0,5	-1,3	0,5	10,8	2,8	-1,4	-2,0
hist_user_behavior_is_shuffle	= true	3,2	-2,8	8,1	7,4	9,5	5,8	11,6	8,9	5,1	-1,1	7,8
	= false	-3,1	2,9	-8,2	-7,4	-9,6	-5,8	-11,6	-9,0	-5,1	1,2	-7,8
premium	= true	0,3	8,8	3,9	4,8	-8,1	7,3	9,5	9,2	2,0	0,7	4,3
	= false	-0,3	-8,8	-3,9	-4,9	8,1	-7,2	-9,4	-9,2	-2,0	-0,7	-4,3
hist_user_behavior_reason_start	= clickrow	7,1	24,9	3,3	1,0	-4,2	-2,9	5,3	2,8	4,5	0,8	0,5
	= trackdone	7,3	-5,4	-2,0	-1,3	-2,0	19,1	-5,3	3,2	3,1	3,0	-0,5
	= fwdbtn	-5,8	-8,7	5,8	4,0	3,7	-7,7	6,6	12,7	-0,1	-0,9	6,6
	= backbtn	-2,9	17,1	1,8	4,4	3,3	13,9	1,1	0,4	-0,2	1,3	3,5
	= appload	4,3	11,1	4,4	2,9	3,3	-2,2	4,7	6,8	-2,7	1,8	3,2
	= playbtn	3,0	2,0	3,6	3,0	5,8	3,9	3,6	8,2	1,2	2,0	3,9
	= remote	2,1	-5,3	3,2	2,5	1,9	1,6	3,2	7,8	0,5	1,7	3,5
	= endplay	2,7	3,5	3,9	3,1	2,2	3,3	3,6	8,2	1,2	3,0	3,9
	= trackerror	2,4	5,8	1,6	1,8	1,6	-2,3	2,8	7,2	1,0	1,4	2,6
song_US_popularity		-23,6	1,8	-13,0	-18,2	0,7	4,5	-27,1	-4,4	-13,5	-1,3	-13,7

fav_cluster_user_weighted_closeness	0,0	12,9	2,3	-1,3	-16,2	-1,4	-0,6	-18,0	-0,7	1,8	1,5										
hour_of_day	-1,5	-23,4	10,7	3,4	-21,0	-8,3	0,5	26,9	0,2	-3,1	10,6										
Bias	-2,8	-6,5	-3,7	-3,0	-2,2	-3,8	-3,7	-8,1	-1,2	-2,0	-3,9										
Output																					
Class (sigmoid)	U	S	U	S	U	S	U	S	U	S	U	S	U	S	U	S					
	2,7	-2,7	2,1	-2,1	0,9	-0,9	1,6	-1,6	1,9	-1,9	1,3	-1,3	1,2	-1,2	1,6	-1,6	0,6	-0,6	0,2	-0,2	0,8

U = guided_by_user; S = guided_by_spotify

Equação 7 – Fast Large Margin do *user_fav_cluster_8*

$$\begin{aligned}
 \text{experience_type} = & \\
 & - 0,009 * \text{hist_user_behavior_reason_start} & - 0,020 * \text{premium} \\
 & - 0,048 * \text{hist_user_behavior_is_shuffle} & + 0,005 * \text{song_US_popularity} \\
 & - 0,001 * \text{track_cluster_is_fav_cluster} & + 0,008 * \text{hour_of_day} \\
 & + 0,000 * \text{fav_cluster_user_weighted_closeness} & + 0,001
 \end{aligned}$$

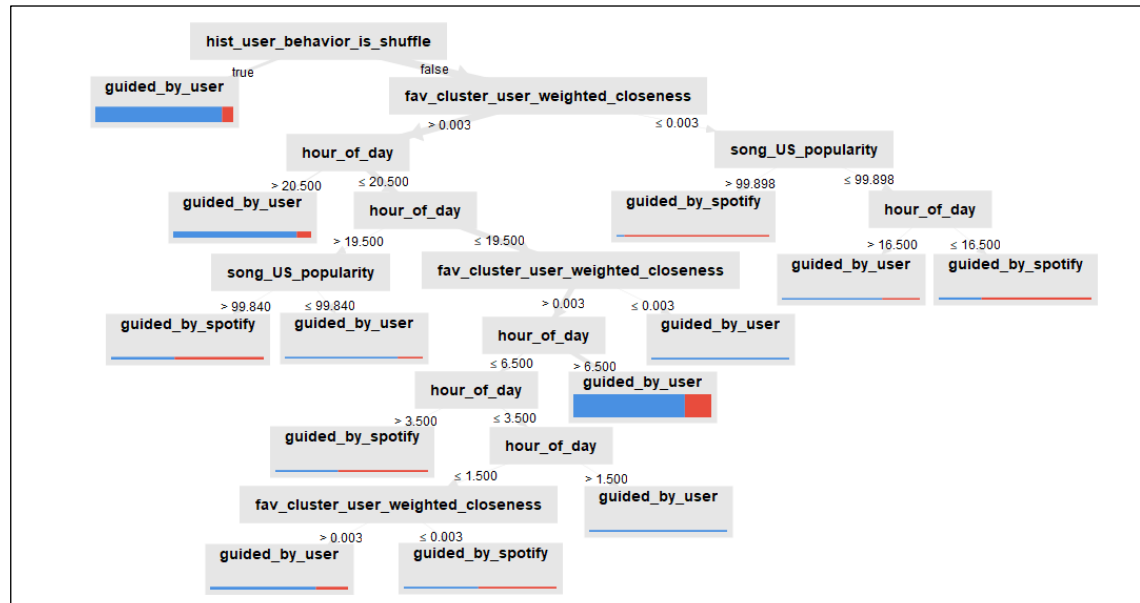


Figura 37 – Árvore de Decisão do *user_fav_cluster_8*

Tabela 27 – Neural Network do *user_fav_cluster_8*

Hidden Layer 1							
Node		1	2	3	4	5	6
track_cluster_is_fav_cluster	= true	4,3	0,1	-0,1	-0,4	-3,5	0,1
	= false	-4,4	-0,1	0,1	0,4	3,5	0,0
hist_user_behavior_is_shuffle	= true	5,2	4,7	-2,5	2,0	2,1	-0,2
	= false	-5,2	-4,7	2,4	-1,9	-2,0	0,2
premium	= true	3,0	-2,8	-0,3	1,1	3,4	-0,7
	= false	-3,0	2,8	0,3	-1,1	-3,4	0,7
song_US_popularity		-10,1	9,8	-3,5	-7,8	0,6	-2,1
fav_cluster_user_weighted_closeness		-15,1	-6,8	-7,4	-3,5	-15,4	6,3
Bias		1,6	-3,6	1,3	-0,7	-0,7	-6,8
Output							
Class (sigmoid)		U	S	U	S	U	S
		5,3	-5,3	-1,8	1,8	1,0	-1,0
		3,9	-3,9	-1,8	1,8	0,6	-0,6

U = guided_by_user; S = guided_by_spotify

Equação 8 – Fast Large Margin do user_fav_cluster_9

$$\begin{aligned}
 \text{experience_type} = & \\
 & - 1,279 * \text{hist_user_behavior_reason_start} \quad - 0,220 * \text{premium} \\
 & - 0,4277 * \text{hist_user_behavior_is_shuffle} \quad + 0,010 * \text{song_US_popularity} \\
 & - 0,329 * \text{track_cluster_is_fav_cluster} \quad - 0,000 * \text{hour_of_day} \\
 & - 0,000 * \text{fav_cluster_user_weighted_closeness} \quad + 0,501
 \end{aligned}$$

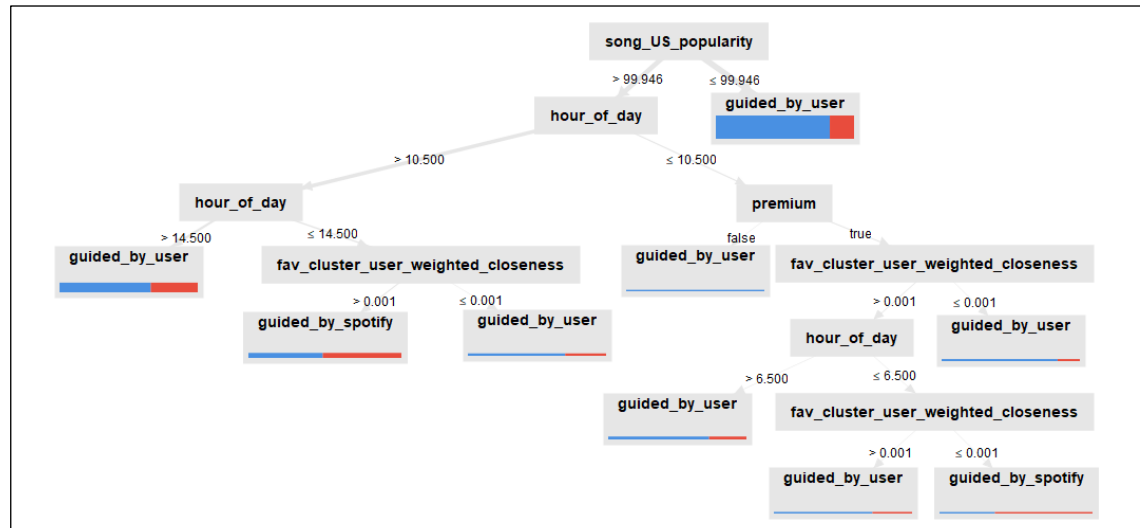


Figura 38 – Árvore de Decisão do user_fav_cluster_9

Tabela 28 – Neural Network do user_fav_cluster_9

Hidden Layer 1											
Node		1	2	3	4	5	6				
track_cluster_is_fav_cluster	= true	-5,1	0,3	1,4	0,2	0,1	-1,2				
	= false	5,1	-0,3	-1,3	-0,2	-0,1	1,3				
hist_user_behavior_is_shuffle	= true	3,3	-1,9	3,4	-0,3	0,1	-0,7				
	= false	-3,3	1,9	-3,3	0,3	-0,1	0,8				
premium	= true	-1,5	-8,5	-3,5	15,1	10,4	1,6				
	= false	1,5	8,5	3,5	-15,1	-10,4	-1,6				
song_US_popularity		-1,6	-17,6	-5,5	-45,5	-33,6	-11,9				
fav_cluster_user_weighted_closeness		-22,4	-16,4	-3,1	0,7	6,5	-3,3				
Bias		-1,5	10,8	-3,6	11,4	3,1	1,9				
Output											
Class (sigmoid)		U	S	U	S	U	S	U	S		
		0,9	-0,9	1,5	-1,5	1,7	-1,7	1,2	-1,2	1,2	-1,2

U = guided_by_user; S = guided_by_spotify

Equação 9 – Fast Large Margin do user_fav_cluster_10

$$\begin{aligned}
 \text{experience_type} = & \\
 & - 2,206 * \text{hist_user_behavior_reason_start} + 0,033 * \text{premium} \\
 & - 2,156 * \text{hist_user_behavior_is_shuffle} + 0,004 * \text{song_US_popularity} \\
 & + 0,028 * \text{track_cluster_is_fav_cluster} - 0,004 * \text{hour_of_day} \\
 & + 0,000 * \text{fav_cluster_user_weighted_closeness} + 0,124
 \end{aligned}$$

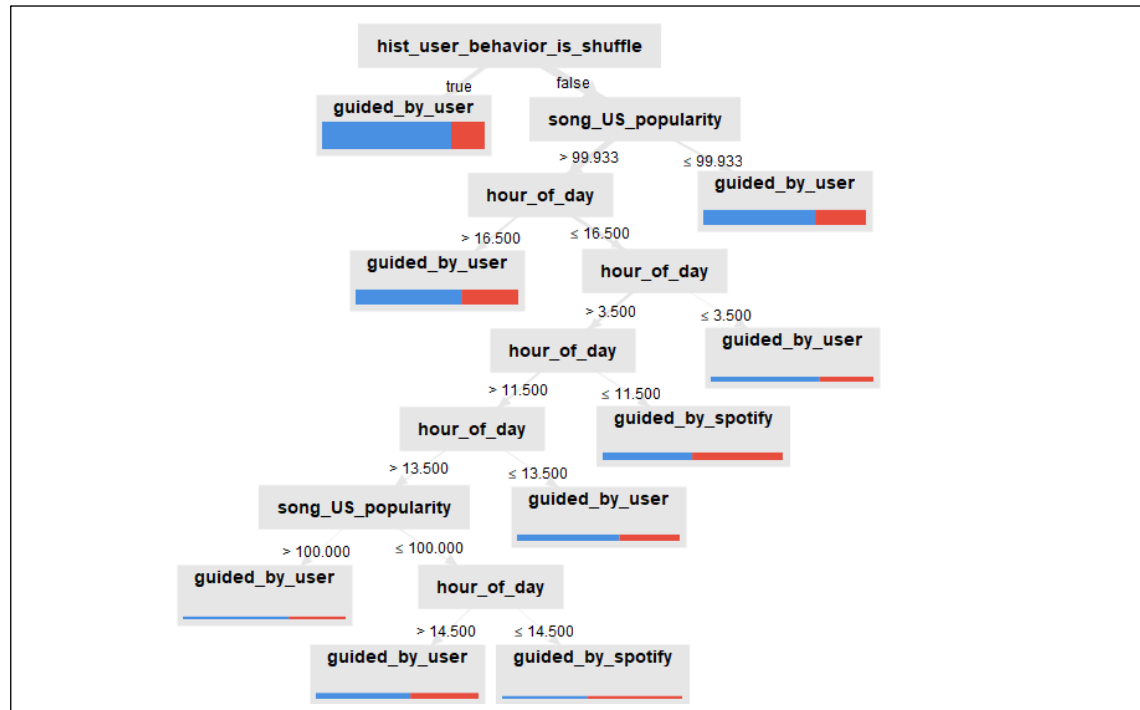


Figura 39 – Árvore de Decisão do user_fav_cluster_10

Tabela 29 – Neural Network do user_fav_cluster_10

Hidden Layer 1							
Node		1	2	3	4	5	6
track_cluster_is_fav_cluster	= true	0,0	0,1	0,0	0,0	0,1	-0,3
	= false	-0,1	-0,1	0,1	0,0	-0,1	0,4
hist_user_behavior_is_shuffle	= true	0,5	0,7	0,9	0,8	-1,8	0,5
	= false	-0,5	-0,8	-1,0	-0,8	1,8	-0,5
premium	= true	2,1	-2,7	0,1	0,1	2,5	0,4
	= false	-2,1	2,6	-0,1	0,0	-2,6	-0,5
song_US_popularity		-11,7	0,4	0,8	1,1	-0,7	-6,0
fav_cluster_user_weighted_closeness		6,4	-5,4	-6,4	-5,4	-10,6	-0,9
Bias		0,6	-2,6	2,1	1,2	-0,6	-0,7
Output							
Class (sigmoid)		U	S	U	S	U	S
		0,8	-0,8	2,2	-2,2	2,0	-2,0
		1,4	-1,4	-3,2	3,2	-0,9	0,9

U = guided_by_user; S = guided_by_spotify

