

Título provisional: Análisis del sistema de recomendación y marketing Basado en Datos Demográficos y del Comportamiento del Usuario de Spotify

Justificación

Spotify es una de las plataformas de streaming más populares del mundo, utilizada por más de 500 millones de usuarios activos mensuales, según sus datos de 2023. Una de sus características distintivas es su capacidad para ofrecer recomendaciones musicales personalizadas a través de algoritmos avanzados de machine learning y big data. La lista de reproducción semanal "Descubrimiento Semanal" y el reporte anual "Spotify Wrapped" son dos de las principales formas en las que Spotify personaliza la experiencia de usuario.

El marketing personalizado, que busca adaptar los productos y servicios a las preferencias individuales de los consumidores, es fundamental en la actualidad. Las recomendaciones de Spotify son un claro ejemplo de cómo los datos del usuario, como sus preferencias musicales e historial de escucha, se utilizan para crear una experiencia altamente personalizada.

Esta investigación busca profundizar en cómo las características demográficas, obtenidas a través de encuestas (Google Forms), junto con los datos recopilados en el "Spotify Wrapped", influyen en la personalización de las recomendaciones. Este estudio es relevante no solo para entender a grandes rasgos el funcionamiento interno del algoritmo de Spotify, sino también para extraer aprendizajes aplicables a otras industrias que utilicen marketing personalizado.

Objetivos

1. **Analizar cómo las características demográficas de los usuarios** (edad, género, ubicación) **y sus hábitos de escucha afectan las recomendaciones que Spotify proporciona.**
2. **Comparar patrones de recomendación entre distintos grupos demográficos,** identificando posibles diferencias o similitudes en la personalización de Spotify.
3. **Proporcionar recomendaciones basadas en los resultados obtenidos,** con el fin de aplicar mejoras en estrategias de marketing personalizado en otras industrias.
4. **(En duda) Desarrollar una API que permita crear perfiles que utilicen Spotify de forma autónoma siguiendo unas instrucciones previamente establecidas.**

Metodología

Fuentes de datos:

- **Información demográfica:** Se obtendrá a través de un cuestionario de Google Forms distribuido a usuarios reales de Spotify, recopilando datos como edad, género, ubicación geográfica, y hábitos generales de consumo musical.
- **Datos del 'Spotify Wrapped':** Este informe, proporcionado por Spotify a finales de cada año, ofrece un resumen detallado del comportamiento musical de los usuarios, incluyendo el número de horas de escucha, los géneros más escuchados, las canciones y artistas más reproducidos, entre otros datos clave.

Variables a estudiar

- **Variables demográficas:**
 - Edad: ¿Cómo varían las recomendaciones según el rango de edad?
 - Género: ¿Existen diferencias notables en las recomendaciones según el género del usuario?
 - Ubicación geográfica: ¿Spotify personaliza sus recomendaciones según la región o país?
- **Variables de comportamiento musical** (extraídas del 'Spotify Wrapped'):
 - Géneros más escuchados: Relación entre los géneros más escuchados del usuario y las recomendaciones futuras.
 - Artistas más reproducidos: Cómo el algoritmo de recomendación introduce artistas similares o expande las recomendaciones más allá de los favoritos del usuario.
 - Frecuencia de uso: Usuarios que interactúan más con la plataforma (mayor tiempo de escucha) pueden recibir recomendaciones más personalizadas o variaciones en la calidad de la personalización.
 - Número de canciones nuevas descubiertas: Relación entre la apertura a nuevas canciones y la adaptación del algoritmo a las preferencias.

Interacción entre las variables:

Se analizará la relación entre las características demográficas y los comportamientos extraídos del 'Spotify Wrapped' para determinar qué combinación de variables influye más en la personalización de las recomendaciones.

Método de análisis:

1. Recolección de datos: Se obtendrán datos de los cuestionarios de Google Forms y de los informes del 'Spotify Wrapped' a principios de diciembre.
2. Análisis comparativo: Se utilizarán técnicas de machine learning para comparar las recomendaciones semanales del 'Descubrimiento Semanal' con los datos demográficos y los patrones de comportamiento musical. Se emplearán métricas (tales como Jaccard

- Similarity) para medir la similitud entre las playlists recomendadas y las canciones más escuchadas del usuario.
3. Estadísticas descriptivas y multivariadas: Se realizará un análisis estadístico descriptivo para entender las tendencias generales, seguido de modelos multivariados que permitan identificar cómo las variables influyen en la personalización.
 4. Análisis temporal: Se evaluará cómo cambian las recomendaciones a lo largo del tiempo y si los patrones de escucha repetidos influyen en las nuevas recomendaciones.

Bibliografía

- Anderson, C. (2006). *The Long Tail: Why the Future of Business Is Selling Less of More*. Hyperion. <https://www.longtail.com>
- Chen, J., & Luo, J. (2019). Recommender Systems: Algorithms, Challenges, Evaluation, and Future Trends. *ACM Computing Surveys*, 52(5), 1-38.
<https://dl.acm.org/doi/10.1145/3298689>
- Davidson, J., Livingston, B., & Sandler, B. (2010). The YouTube video recommendation system. *Proceedings of the Fourth ACM Conference on Recommender Systems*, 293-296.
<https://dl.acm.org/doi/10.1145/1864708.1864770>
- Gómez-Uribe, C., & Hunt, N. (2015). The Netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems*, 6(4), 1-19.
<https://dl.acm.org/doi/10.1145/2843948>
- Herlocker, J. L., Konstan, J. A., & Riedl, J. (2002). An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms. *Information Retrieval*, 5(4), 287-310.
<https://link.springer.com/article/10.1023/A:1020443909834>
- Ibañez Jara, E. E., & Soto Gaete, J. M. (2020). Las playlist de Spotify: Una forma de curatoría musical contemporánea. Universidad Academia de Humanismo Cristiano.
<http://www.uahc.cl/musica>
- Koren, Y. (2009). Collaborative filtering with temporal dynamics. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 447-456.
<https://dl.acm.org/doi/10.1145/1557019.1557072>
- Linden, G., Smith, B., & York, J. (2003). Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1), 76-80.
<https://ieeexplore.ieee.org/document/1167344>
- Ricci, F., Rokach, L., & Shapira, B. (2015). Recommender systems: Introduction and challenges. En F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender Systems Handbook* (pp. 1-34). Springer. https://link.springer.com/chapter/10.1007/978-1-4899-7637-6_1

Aitana Herrero Castro
5ºE2 Analytics

Salakhutdinov, R., & Mnih, A. (2007). Probabilistic matrix factorization. *Proceedings of the 21st International Conference on Neural Information Processing Systems*, 1257-1264.
<https://papers.nips.cc/paper/2007/file/0198d8f59e57bede2edaefc2a8d333e-Paper.pdf>

Schäfer, H., & Thompson, B. (2018). Recommender Systems: AI for Personalized Music Discovery. *Journal of Computer Science and Technology*, 33(2), 223-237.
<https://www.springer.com/journal/11390>

Sharma, M., & Mahajan, A. (2020). Deep Learning Based Recommender System for Online Retail Platforms. *Journal of Data Science and Analytics*, 5(4), 335-350.
<https://link.springer.com/article/10.1007/s41060-020-00221-4>

Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*, 1(1), 1-20. <https://www.hindawi.com/journals/aai/2009/421425/>

van den Oord, A., Dieleman, S., & Schrauwen, B. (2013). Deep content-based music recommendation. *Proceedings of the 27th International Conference on Neural Information Processing Systems*, 2643-2651.
<https://proceedings.neurips.cc/paper/2013/file/b0d8e60b5b18f67fb6c0d9f37063806c-Paper.pdf>

Zhang, Y., & Yang, Q. (2019). A survey on multi-task learning. *IEEE Transactions on Knowledge and Data Engineering*, 31(10), 1865-1883.
<https://ieeexplore.ieee.org/document/8620073>

Cronograma

Octubre-Noviembre 2024:

- Creación y distribución del cuestionario de Google Forms para recopilar datos demográficos.
- Análisis inicial de datos preliminares obtenidos de Spotify.

Diciembre 2024:

- Recolección de datos del 'Spotify Wrapped'.
- Primer análisis comparativo entre datos demográficos y comportamiento en Spotify.

Enero - Febrero 2025:

- Análisis detallado de las recomendaciones personalizadas y evaluación de la similitud entre playlists.
- Elaboración de conclusiones preliminares.

Marzo 2025:

- Redacción y presentación del informe final.

Resultados esperados

Se espera obtener una mejor comprensión de cómo Spotify personaliza las recomendaciones basándose en datos demográficos y patrones de escucha específicos. El análisis permitirá evaluar si los algoritmos de recomendación varían significativamente según características demográficas o patrones de comportamiento musical, lo que podría proporcionar información valiosa para la mejora de estrategias de marketing personalizado en otras industrias.

Limitaciones

1. Acceso limitado a los detalles internos del algoritmo de Spotify: El estudio se basará en observaciones externas de las recomendaciones proporcionadas por Spotify, sin acceso directo a los factores internos exactos que el algoritmo utiliza.
2. Validez de los datos demográficos: La calidad del análisis dependerá de la precisión y representatividad de los datos demográficos obtenidos a través del cuestionario.
3. Variabilidad en los patrones de escucha: Los hábitos musicales de los usuarios pueden variar de manera significativa, lo que podría afectar las conclusiones sobre la personalización.
4. Limitaciones de tiempo: El estudio se llevará a cabo durante un periodo limitado, lo que podría restringir la capacidad para observar cambios a largo plazo en las recomendaciones personalizadas.
5. Dificultades técnicas en la API: La implementación de una API que interactúe con los datos de Spotify puede representar un desafío técnico, limitando el acceso a la frecuencia de las recomendaciones o al volumen de datos obtenidos.