ANÁLISIS Y CONVERSIÓN DE VARI5ABLE DE CONTEO DE USO EN VARIABLE CALIFICACIÓN PARA LA ELABORACIÓN DE UN MODELO DE RECOMENDACIÓN DE CANCIONES.

ANALYSIS AND CONVERSION OF A USE COUNTING VARIABLE INTO A QUALIFICATION VARIABLE FOR THE PREPARATION OF A SONG RECOMMENDATION MODEL.

Cristian Jose Castro Sanchez, Magister en Supply Chains Management y logística, Especialista en Gestión de redes de valor e Ingeniero Industrial. Fundación Universitaria Los Libertadores. Bogotá, Colombia. Contacto: cjcastros01@libertadores.edu.co.

Jhon Gonzalez Veloza, Fundación Universitaria Los Libertadores. Bogotá, Colombia. Contacto: jjgonzalezv02@libertadores.edu.co

RESUMEN

Este artículo presenta un análisis de datos que permite múltiples formas para la conversión de la variable de conteo en una variable de calificación, para así generar un modelo de recomendación de canciones basado en un dataset de rock clásico a través del filtro colaborativo de la herramienta Turicreate de Machine Learning con la ayuda del software Python de Ubuntu. Como resultado, se encontró que la base de datos posee un alto sesgo el cual implicó analizar la data en profundidad y crear un modelo que contemple con mayor peso los datos escuchados más de una vez por los usuarios. El modelo de recomendación posee una precisión 0.075 y 0.20 recuperación mayor que el modelo que le antecede en valores comparativos de rendimiento, esto demuestra que el modelo brinda una recomendación satisfactoria ya sea filtrando por usuario, título de canción o artista. Se evidenció que el modelo provee una metodología ágil y confiable para el entendimiento lógico del análisis y creación de un sistema de recomendación. Como conclusión general se observa que, entre distintas formas de mitigar el sesgo.

Palabras clave: Sistema de recomendación, Turicreate, Machine Learning.

ABSTRACT

This article presents a data analysis that allows multiple ways to convert the count variable into a rating variable, in order to generate a song recommendation model based on a classic rock dataset through the collaborative filter of the Turicreate tool. Machine Learning with the help of Ubuntu Python software. As a result, it was found that the database has a high bias, which involved analyzing the data in depth and creating a model that considers with greater weight the data heard more than once by users. The recommendation model has an accuracy of 0.075 and 0.20 recovery higher than the model that precedes it in comparative performance values, this shows that the model provides a satisfactory recommendation whether filtering by user, song title or artist. It was evidenced that the model provides an agile and reliable methodology for the logical understanding of the analysis and creation of a recommendation system. As a general conclusion, it is observed that, among different ways of mitigating the bias.

Keywords: Recommendation system, Turicreate, Machine Learning.

INTRODUCCIÓN

Según Laguna (2013) el creciente número de datos que genera la humanidad en el uso de sistemas de información y acceso a tecnologías, debido al internet de las cosas, así como el acceso a un mar de opciones de contenido, genera interés por la mejor forma de realizar contenido a usuarios, como películas, libros, canciones, podcast, entre otros. En cada interacción con los usuarios se obtiene un dato de uso, dicho conteo es de vital importancia en especial en los casos que no tiene una calificación de gusto o preferencia directa del usuario. El presente estudio busca convertir la variable de conteo del número de veces que se escucha una canción en una calificación de gusto permite generar recomendaciones asertivas al disminuir el tiempo de elección del usuario,

así como aumentar su nivel de confianza y autonomía con la herramienta o sistema digital. De esta manera estudiantes, investigadores y desarrolladores encontrarán la metodología para comprender, analizar y apropiar una forma eficaz de convertir un conteo en la calificación o puntuación, esto al no tener una calificación directa del usuario para el ítem de contenido (libros, canciones, entre otros elementos).

Los beneficios principales de un sistema de recomendación es el de incrementar la satisfacción del usuario, esto es posible gracias a que se puede mejorar la experiencia del usuario con el sitio o la aplicación, a través de la facilidad para el usuario de encontrar las recomendaciones interesantes y relevantes optimizando el tiempo y uso de la plataforma, otra ventaja es la de Incrementar la fidelidad del usuario, por medio de la generación de una experiencia memorable al reconocer a un antiguo cliente y lo trata como un visitante valioso; esta es una norma mal característica de un RS (sistema de recomendación), ya que muchos RS calculan recomendaciones, aprovechando la información adquirida del usuario en interacciones anteriores, por ejemplo, sus calificaciones de elementos, en el mismo sentido el comprender mejor lo que quiere el usuario es la razón fundamental de desarrollar sistema de recomendación, dado que puede el conocer y aprovechar el conocimiento de las preferencias y gustos es valioso en muchas aplicaciones.

Según Lasse R. (2018) define la IA como la habilidad de los ordenadores para hacer actividades que normalmente requieren inteligencia humana.

Una parte del gran campo llamado IA es el Machine Learning (ML), el cual se describe como aprendizaje autónomo; Según Luca M. (2019) Machine Learning puede ser clasificado de acuerdo con la cantidad y tipo de supervisión que tienen durante su entrenamiento, existen 4 tipos de categorías principales, Aprendizaje Supervisado (Superviced Learning), Aprendizaje No

Supervisado (Unsupervised Learning), Aprendizaje Semi Supervisado (Semi Supervised Learning) y Aprendizaje De Reforzamiento (Reinforcement Learning).

Machine Learning identifica dos tipos de aprendizaje principales, el supervisado depende de la información que previamente está etiquetada, podría ser el que un ordenador logre diferenciar imágenes de autos, de barcos, etc. luego, normalmente el modelo pone los rótulos de acuerdo a las etiquetas predefinidas, el resultado se espera en términos de calidad y efectividad de la información; contrario al aprendizaje supervisado, el aprendizaje no supervisado no cuenta con ninguna información previa, ya que al algoritmo se le entrega poca responsabilidad de alguna etiqueta, del cual no cuenta con ninguna indicación previa, pero se le entrega una gran cantidad de datos con los rasgos propios de un objeto como partes o aspectos que forman parte de una motocicleta, carro o un avión, donde este modelo es ajustado a las observaciones para que pueda verificar y crear sus propias etiquetas a partir de la información compilada; algunos ejemplos serían como hallar y clasificar información, morfología en oraciones, identificar objetos u elementos en imágenes varias, etc.

Se toma como base él data set para la construcción del modelo, este se precisa por Romero, Aceituno, Pedro D. (2019) como la materia prima esencia a la hora de construir sistemas predictivos, por tanto, es el primer paso en el proceso de construcción del modelo es la de recolectar, conocer y entender los datos necesarios para construirlo, entrenarlo y validarlo. Igualmente, reconocen la importancia de realizar un análisis exploratorio de los datos: Una vez extraídos los datos de la base de datos, prosiguió el realizar un análisis exploratorio de los mismos con el fin de ganar conocimiento y comprender mejor las características de los datos.

Según Calvo M (2017), los sistemas de recomendación surgieron como un área de investigación independiente a mediados de la década de 1990, con el auge de internet de las cosas

y la disrupción tecnológica los sistemas de recomendación gozan en la actualidad de gran atención para su afinamiento, una metodología de sistema de recomendación es el basado en el filtro de recomendaciones de artículos que les hayan gustado a usuarios similares (aquellos con gustos similares). Este enfoque es denominado filtrado colaborativo y su razón de ser es que, si el usuario activo acordó en el pasado con algunos usuarios, luego las otras recomendaciones provenientes de estos similares los usuarios también deben ser relevantes y de interés para el usuario activo.

Un conjunto de herramientas para crear modelos en Machine Learning es Turicreate, la función: turicreate.item_similarity_recommender utiliza en una primera etapa de construcción de modelos al encontrar la similitud entre todos los pares de elementos. Esta función de similitud puede tomar muchas formas, como la correlación entre las calificaciones o el coseno de esos vectores de calificación. Al igual que en los sistemas de usuario-usuario, las funciones de similitud pueden usar calificaciones normalizadas.

Este modelo primero calcula la similitud entre elementos empleando las observaciones de los usuarios que han interactuado con ambos elementos. Hay tres opciones de métricas de similitud para usar: 'jaccard', 'coseno' y 'pearson'.

La similitud de Jaccard se utiliza para medir la similitud entre dos conjuntos de elementos. En el contexto de la recomendación, la similitud de Jaccard entre dos elementos dado por el conjunto de usuarios que calificación el artículo. Jaccard es una buena opción cuando uno solo tiene retroalimentación implícita de los elementos (por ejemplo, las personas los calificaron o no), o cuando a uno no le importa cuántas estrellas recibieron los artículos. Si es necesario comparar las calificaciones de los elementos, se recomienda la similitud de Coseno y Pearson.

La similitud de coseno entre dos elementos dad por el conjunto de usuarios que calificaron el artículo versus el conjunto de usuarios que calificaron ambos elementos. Un problema con la

similitud del coseno es que no considera las diferencias en la media y la varianza de las calificaciones hechas a los elementos.

Otra medida popular que compara las calificaciones donde se han eliminado los efectos de las medias y la varianza es la similitud de la correlación de Pearson, la cual utiliza los elementos más similares a los elementos ya calificados de un usuario para generar una lista de recomendaciones. Por lo general, este cálculo es una suma ponderada o una regresión lineal. Esta forma de recomendación es análoga a "las personas que valoran muy bien el artículo X, como usted, también tienden a valorar muy bien el artículo Y, y usted todavía no ha puntuado el artículo Y, por lo que debería intentarlo".

Según Neter, J. et al. (1996) las transformaciones de Box y Cox son una familia de transformaciones potenciales usadas en estadística para corregir sesgos en la distribución de errores, para corregir varianzas desiguales (para diferentes valores de la variable predictora) y principalmente para corregir la no linealidad en la relación (mejorar correlación entre las variables). Esta transformación recibe el nombre de los estadísticos George E. P. Box y David Cox.

Una vez creado el modelo, se procede a su validación se validan al comparar cualitativamente entre ellos el resultado, se poseen los siguientes indicadores Precisión – Recall, Según Calderon O, JOSE D; Morales T, Lennin F; Roncal M, María E; Solórzano R, William G. (2021), describen la primera representa la relación entre elementos recuperados relevantes y el número de elementos relevantes (Elemento generado por modelo no perteneciente al conjunto original del modelo o segmente buscado). Así mismo, ellos clarifican que el F1 Score Tiene en cuenta la precisión y la recuperación de la prueba para calcular su puntuación, es el promedio

armónico de la precisión y recuperación. Alcanza un mejor valor en 1 (precision y recall) y peor en 0.

En términos de recuperación de información, la precisión representa la relación entre elementos recuperados relevantes y el número de elementos relevantes. La recuperación representa la relación entre elementos recuperados relevantes y el número de elementos originales del modelo.

Recuperación- Recall

La proporción de elementos que le gustan a un usuario en realidad fueron recomendados, está dada por:

$$Recall = \frac{tp}{tp + fn}$$

Aquí tp representa la cantidad de elementos recomendados a un usuario que le gustan y tp + fn representa el total de elementos que le gustan a un usuario

(Si a un usuario le gustan 5 artículos y el motor de recomendación decidió mostrar 3 de ellos, recall será de 0.6)

Precisión- Precision

De todos los elementos recomendados, la cantidad de etiquetas que le gustó realmente al usuario, está dada por:

$$Precision = \frac{tp}{tp + fn}$$

Aquí tp representa la cantidad de elementos recomendados para un usuario que le gustan y tp + fp representa el total de elementos recomendados para un usuario

(Si se recomendaron 5 elementos al usuario, de los cuales le gustaron 4, entonces la precisión será de 0.8)

- A mayor precisión, mejores recomendaciones.

METODOLOGÍA

Esta investigación tiene un alcance descriptivo, al tener como intensión la de encontrar diferentes métodos para convertir la variable de conteo del número de veces que se escucha la canción en una calificación de gusto, esto permitirá crear un modelo de recomendación de canciones el cual facilitará la toma de decisiones de usuarios creando una experiencia más ágil y rápida, así mismo, apoyará la creación de contenidos de acuerdo con las tendencias predichas. Desde el alcance se considera un enfoque cualitativo al contar con una base de datos con elementos de listas por cada usuario, para su descripción y análisis se utilizan diversas herramientas estadísticas.

La investigación se basa en la herramienta Turicreate de Machine Learning, para el análisis de la base de datos y creación de varios modelos de recomendación. La técnica es el filtro colaborativo la cual se basa en encontrar preferencias entre usuarios, concatenándolos por gustos miliares previamente; para el análisis estadístico se utiliza la plataforma Python. Para el desarrollo se planean las siguientes fases: análisis descriptivo de la variable objetivo y creación de métodos de elaborar modelos de recomendación.

1. **Análisis descriptivo de la variable objetivo**

Incluye dos procesos en el primero se analizará de manera descriptiva la base de datos usada y en el segundo se describirá el nivel de sesgo de la variable objetivo.

Se identificaron los principales elementos de la base de datos los cuales son user_id, song_id, listen_cound, title, artis, song y id_song_user. El número total de filas del dataset es 1.116.609, la cantidad de usuarios en la base de datos es de 66.346. La base de datos contiene múltiples usuarios de los cuales cada uno consta de n número de canciones que ha escuchado, además, se tiene en cuenta el número de veces que el usuario a escuchado dicha canción.

La estructura genérica del dataset se muestra a continuación. (Veas la tabla 1)

Tabla 1. Estructura dataset

user_id	song_id	listen_count	title	artist
b80344d063b5ccb3212f76538 f3d9e43d87dca9e ...	SOAKIMP12A8C130995	1	The Cove	Jack Johnson
b80344d063b5ccb3212f76538 f3d9e43d87dca9e ...	SOBBMDR12A8C13253B	2	Entre Dos Aguas	Paco De Lucia
b80344d063b5ccb3212f76538 f3d9e43d87dca9e ...	SOBXHDL12A81C204C0	1	Stronger	Kanye West
b80344d063b5ccb3212f76538 f3d9e43d87dca9e ...	SOBYHAJ12A6701BF1D	1	Constellations	Jack Johnson
b80344d063b5ccb3212f76538 f3d9e43d87dca9e ...	SODACBL12A8C13C273	1	Learn To Fly	Foo Fighters
b80344d063b5ccb3212f76538 f3d9e43d87dca9e ...	SODDNQT12A6D4F5F7E	5	Apuesta Por El Rock 'N' Roll ...	Héroes del Silencio
b80344d063b5ccb3212f76538 f3d9e43d87dca9e ...	SODXRTY12AB0180F3B	1	Paper Gangsta	Lady GaGa
b80344d063b5ccb3212f76538 f3d9e43d87dca9e ...	SOFGUAY12AB017B0A8	1	Stacked Actors	Foo Fighters
b80344d063b5ccb3212f76538 f3d9e43d87dca9e ...	SOFRQTD12A81C233C0	1	Sehr kosmisch	Harmonia
b80344d063b5ccb3212f76538 f3d9e43d87dca9e ...	SOHQWYZ12A6D4FA701	1	Heaven's gonna burn your eyes ...	Thievery Corporation feat. Emiliana Torrini ...

song
The Cove - Jack Johnson
Entre Dos Aguas - Paco De Lucia ...
Stronger - Kanye West
Constellations - Jack Johnson ...
Learn To Fly - Foo Fighters ...
Apuesta Por El Rock 'N' Roll - Héroes del ...
Paper Gangsta - Lady GaGa
Stacked Actors - Foo Fighters ...
Sehr kosmisch - Harmonia
Heaven's gonna burn your eyes - Thievery ...

Posteriormente, se validó que cada usuario tuviera un único dato para cada canción, esto dado por agrupando por la columna id_song_user y hacemos un filtro de mayor a menor en la cuenta y observamos que la cuenta máxima es 1, por lo tanto, confirma que la concatenación de user_id con song_id es único

Se validó la canción y artista más escuchado en el dataset de manera general y específica por usuario.

En el análisis de variable objetivo se observó un sesgo alto al evaluar el peso de la variable listen_count, es decir, se evaluó cual es el número de veces que se escucha una canción de manera general, identificando que la mayoría de las canciones fueron escuchadas una única vez, esto equivale a 623.998 veces (55,8% del total de canciones del dataset), así mismo, se han escuchado dos veces la misma canción un total de 169.875 veces (15,2% de total canciones del dataset). Lo anterior evidencia un sesgo alto, lo cual es una característica para tener en cuenta en la elaboración del modelo de recomendación (Véase la tabla No.2).

Tabla 2. Segmentación de listen_count

listen_count	count	listen_count	count
1	623998	307	1
2	169875	225	1
3	78816	220	1
5	56102	453	1
4	46401	270	1
6	26421	218	1
7	18042	192	1
8	13359	183	1
10	11755	243	1
9	10129	286	1

Con el fin de crear alternativas para gestionar el sesgo de la variable objetivo se aplicó a la variable liste_count un logaritmo en base 10, con el fin de identificar y segmentar en la base de datos las canciones que se escucharon una única vez y las de más (canciones escuchadas dos o más veces), lo anterior se realizó utilizando escalamiento entre cero y uno, esto dio como resultado una nueva columna de datos llamada gusto_escalado.

2. Creación de modelos de acuerdo con los métodos de recomendación Turicreate.

En la librería Turicreate encontramos múltiples formas de crear un modelo de recomendación, de los cuales se crearon los siguientes modelos:

2.1 Modelo de recomendación simple.

Este modelo de recomendación emplea la fórmula popularity_model, la cual de manera lógica sugiere canciones de acuerdo con el número de veces que se escuchó una canción, este no usa ningún tipo de filtro colaborativo entre usuarios. El modelo planteado se realizó con los datos originales de la base de datos, es decir, no se realizó ningún tratamiento a estos para eliminar el sesgo o conversión de la variable de conteo en una calificación de gusto.

Dato más alto de Precisión: 0.0025

Dato más alto de Recuperación: 0.003

Equilibrio precisión (P)- recuperación (R): P-R (0.0020-0.002)

2.2 Modelo de recomendación de canciones personalizado.

Este modelo de recomendación utiliza la fórmula personalized_model, la cual de manera lógica utiliza un filtro colaborativo entre usuarios de acuerdo con una canción común entre estos. La recomendación que el modelo da como resultado consta de 10 canciones y puede proponer seleccionado un usuario o seleccionando una canción, en cada variación según la selección el

modelo sugiere de manera distinta canciones recomendadas. Al igual que el modelo 1 este modelo se desarrolló sin tratamiento alguno del sesgo o conversión del conteo en una calificación de gusto.

Dato más alto de Precisión: 0.025

Dato más alto de Recuperación: 0.03

Equilibrio precisión (P)- recuperación (R): P-R (0.020-0.02)

2.3 Modelo de recomendación de canciones personalizado empleando la variable gusto_escalado.

Igual que en el modelo 2 se usa la formula personalized_model de la herramienta Turicreate, sin embargo, se realiza un tratamiento a los datos de la variable de conteo (número de veces que se escucha una canción) con el fin de convertir este conteo en una calificación del gusto o preferencia, lo anterior es de vital importancia dado el alto sesgo de los datos; se realiza el siguiente tratamiento:

- Obtenemos la columna listen count y le aplicamos logaritmo en base 10.
- Adicionamos dicha columna al dataframe
- La columna del log la dividimos entre el máximo menos el mínimo
- Adicionamos la información al dataframe

Una vez aplicado el logaritmo en base 10 a la variable objetivo se observó que la mayoría de las canciones (623.998) se escucharon una única vez por lo cual en la nueva columna tiene un valor de cero.

Ahora bien, identificadas que la mayoría de las canciones se escucharon una única vez se desarrolló un modelo que tuviera en cuenta todos los datos menos los que tuvieran en la columna

gusto_escalado el valor de cero, es decir, no se toma en cuenta las canciones que se escucharon una única vez.

Dato más alto de Precisión: 0.027

Dato más alto de Recuperación: 0.13

Equilibrio precisión (P)- recuperación (R): P-R (0.025-0.05)

2.4 Modelo de recomendación de canciones personalizado utilizando media y desviación estándar.

Con el fin de convertir la variable de conteo del número de veces de canciones que se escuchó una canción y con el objetivo de gestionar el sesgo de la variable objetivo, se llevaron a cabo las siguientes actividades:

- Cálculo de la media a la variable objetivo (listen_count).

  La media de listenc_count es: 3.207687471621676

  La desviación estándar de liste_count es: 7.20253035130171

- Calculo desviación estándar, límite inferior y límite superior.

Se tomó una desviación estándar para crear un modelo para no tomar los datos atípicos, los cuales serían entre otras las canciones que solo se escucharon una única vez o que se escucharon muchas veces por un único usuario y se calculó el límite inferior y superior.

El total de los datos del modelo propuesto es de 1054898.

- Establecer un modelo que tome en cuenta el límite inferior y superior.

Dato más alto de Precisión: 0.175

Dato más alto de Recuperación: 0.40

Equilibrio precisión (P)- recuperación (R): P-R (0.075-0.20)

Los anteriores resultados reflejan que el modelo 4 genera un 7,5 % más de datos con referente al modelo 3 que le gustaron al usuario y 20% más de datos sugeridos o recomendados adicionales con referencia al modelo 3.

2.5 Modelo de recomendación de canciones personalizado utilizando Box Cox

Debido a que los datos de la variable objetivo no siguen una curva normal, se procedió a utilizar la metodología Box Cox (forma de transformar variables dependientes no normales en una curva normal) con el objetivo de mitigar el sesgo de los datos, con dicha metodología se logo mover el pico de la izquierda hacia el centro.

Se realizo la siguiente metodología:

- Transformar los datos de la variable objetivo mediante Box Cox.
- Incluir los resultados en el dataframe.
- Crear el modelo de recomendación de canciones de acuerdo con los nuevos datos transformados.

Dato más alto de Precisión: 0.175

Dato más alto de Recuperación: 0.40

Equilibrio precisión (P)- recuperación (R): P-R (0.075-0.20)

Los anteriores resultados reflejan que el modelo 5 genera un 7,5 % más de datos con referente al modelo 3 que le gustaron al usuario y 20% más de datos sugeridos o recomendados adicionales con referencia al modelo 3. Estos datos son muy similares al modelo 4.

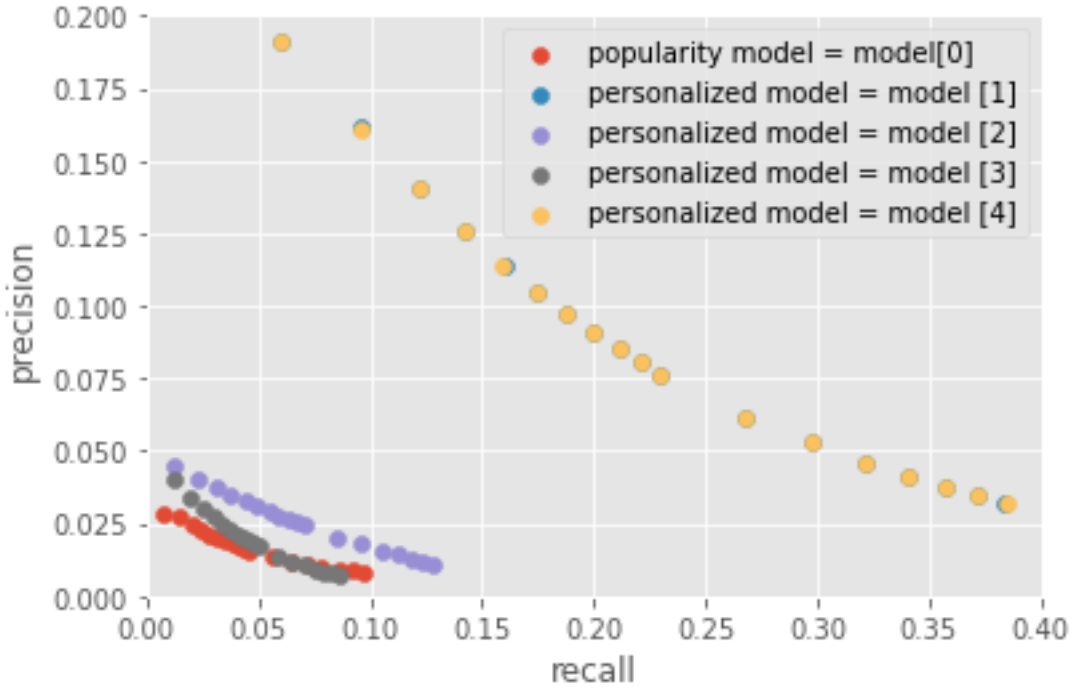
RESULTADOS

La conversión de la variable de conteo a una etiqueta de gusto o preferencia por las canciones se cumplió satisfactoriamente al poder reconocer el verdadero gusto que posee cada usuario, esto

según los datos observados de equilibrio precisión (P)- recuperación (R): P-R (0.075-0.20) del modelo 4 y 5.

El sesgo se de los datos de la variable objetivo se logro mitigar con el uso de la metodología propuesta en el modelo 4 y 5, es así como se cumple con el objetivo de proponer una metodología de análisis y tratamiento del sesgo reflejado en un modelo de recomendación de canciones con alta precision y recuperación frente a los modelos simples sin ningún tratamiento.

De manera general se observan 10 recomendaciones de canciones de acuerdo con un usuario o canción seleccionado, así mismo frente a la comparación entre modelos se utilizó la comparación de la precisión y recuperación entre modelos, la cual permite comparar formalmente la popularidad entre los modelos de acuerdo con las variables de recuperación -recall y precisión-precisión.



Se observa que el modelo 4 y 5 posee una mejor precisión y recuperación en comparativa con los modelos 1,2 y 3. Esto evidencia mejoría en términos de recuperación (arroja muchos datos) y precisión (la mayoría de las etiquetas predichas son correctas en comparación con las etiquetas de entrenamiento).

Este modelo 4 fue el que se obtuvo luego de realizar un tratamiento al sesgo y convertir el conteo del número de veces en una calificación para realizar un modelo de recomendación óptimo en términos de precisión y recuperación.

CONCLUSIONES

Las implicaciones de los resultados fue la estructuración de una manera fácil y ágil de modelar un sistema de recomendación, obteniendo resultados favorables en la capacidad de predecir etiquetas de recomendación.

El modelo de recomendación con más alta recuperación y precisión frente a los demás modelos una mejor y más eficientes predicciones, esto una vez se ajustó el modelo según el sesgo de la variable objetivo.

El objetivo propuesto es entender la lógica y la manera óptima de análisis de la base de datos y modelamiento del recomendador, en pro de obtener beneficios de los datos captados en múltiples plataformas el modelo facilitara la toma de decisiones de usuarios creando una experiencia más ágil y rápida, así mismo, apoyara la creación de contenidos de acuerdo a las tendencias predichas, por lo cual el resultado cumple con el objetivo propuesto dado que se creó una metodología y modelo que facilitara el procesamiento de otras bases de datos como libros, artículos, audios, canciones y demás, generando así un modelo de recomendación con datos de fácil y ágil de utilizar.

Se obtiene como resultado un modelo que en la comparativa cualitativa entre los modelos obtiene un punto de equilibrio precisión (P)- recuperación (R): P-R (0.075-0.20).

Se observa que una variable para tener en cuenta es el sesgo de los datos debido a que pueden afectar en gran manera la predicción de etiquetas.

Este modelo permitió validar la metodología planteada y la potencia de la herramienta Turicreate.

REFERENCIAS BIBLIOGRÁFICAS

Alba, T. (2018) Cómo y por qué aplicar la Escucha Activa en Redes Sociales. Recuperado el 25 de marzo de 2020 de https://blog.mailrelay.com/es/2018/02/22/escucha-activa

Mendenhall, Robert J. Beaver, Barbara M, Beaver.(2016) Introducción A La Probabilidad Y Estadística. Cencage Learning

Janns A. (2020) Modelo predictivo para el reconocimiento de actividades humanas basado en técnicas de Machine Learning y de selección de características. Cencage Learning

Gustavo B, Jorge D, Oscar L, José R. Quevedo, Jaime A, Antonio B y Juan J. (2020) Modelo predictivo para el reconocimiento de actividades humanas basado en técnicas de Machine Learning y de selección de características. Centro de Inteligencia Artificial. Universidad de Oviedo en Gijón Campus de Viesques

Gustavo F. (2020) Selección de Atributos en Métodos no Lineales de Aprendizaje de Preferencias para el análisis de Datos Sensoriales. Centro de Inteligencia Artificial- Universidad de Oviedo

Alonso, M., Gómez C. y Vilares, J. (2015) Seguimiento y análisis automático de contenidos en redes sociales. En III Congreso Nacional de I+D en Defensa y Seguridad. Universidad de Coruña España

Altamirano, M, Cevallos, L, González, I., Echeverria B, Pedro F. (2009) Implementación de un sistema basado en minería de datos para la obtención de las preferencias estudiantiles de nivel superíor para la planificación de materias. Facultad de Ingeniería en Electricidad y Computación Escuela Superior Politécnica del Litoral

Marín V, Darwin G, Pineda T, Ian A. (2019) Modelo predictivo machine learning aplicado a análisis de datos hidrometeorológicos para un SAT en represas. Universidad tecnológica del Peru

Calvo M (2017) Text Analytics para Procesado Semántico. Universidad de Vigo

Romero, Aceituno, Pedro D. (2019) Desarrollo y evaluación de un modelo predictivo basado en machine learning para estudiar y predecir el comportamiento del absentismo en prestaciones sanitarias. Universidad de Jaén- Escuela Politécnica Superior de Linares

Miguel R. (2019) Desarrollo y evaluación de un modelo predictivo basado en machine learning para estudiar y predecir el comportamiento del absentismo en prestaciones sanitarias. Universidad nacional del Altiplano.

Jay L. Devore. (2019) Probabilidad y Estadística para Ingeniería y Ciencias. Cengage Learning.

Juan J. Padial. (2019) Técnicas de programación «deep learning». Universidad de Málaga.

Luca M. (2019). Machine Learning For Dummies. Universidad de Málaga.

Lasse R. (2018) Inteligencia Artificial. Alienta.

JOSÉ C. (2018) Análisis De Sentimiento En Twitter. Espacios.

Calderon O, JOSE D; Morales T, Lennin F; Roncal M, María E; Solórzano R, William G. (2021) Uso De Algoritmos De Machine Learning Para El Diagnóstico De Melanomas. Universidad De Piura.

Neter, J. et al. (1996) Applied Linear Statistical Models. McGrawHill

Herrán Gascón, A., Cortina Selva, M. y González Sánchez, I. (2006). La muerte y su didáctica: manual para educación infantil, primaria y secundaria. Madrid: Universitas.

Maria Laguna (2013). Estudio sobre el uso de internet y sus aplicaciones en el alumnado de último año de carrera. Alicante: Universidad de Alicante.