25 CONGRESO INTERNACIONAL DE CIENCIAS ADMINISTRATIVAS
6, 7 Y 8 DE OCTUBRE 2021

LA SATISFACCIÓN DEL TURISTA CON EL DESTINO VISITADO: COMPARACIÓN DE MODELOS PREDICTIVOS USANDO FACTORES MOTIVACIONALES

Área de investigación: Mercadotecnia

Guberney Muñeton Santa
Instituto de Estudios Regionales
Universidad de Antioquia
Colombia
guberney.muneton@udea.edu.co

Juan Gabriel Vanegas López
Facultad de Ciencias Económicas y Administrativas
Visión de las Américas I. U.
Colombia
juan.vanegas@uam.edu.co

LA SATISFACCIÓN DEL TURISTA CON EL DESTINO VISITADO: COMPARACIÓN DE MODELOS PREDICTIVOS USANDO FACTORES MOTIVACIONALES

Resumen

El nivel de satisfacción de un turista con el destino visitado se asume ampliamente dependiente de las motivaciones previas que tenía de éste. Para observar esta perspectiva relacional se utilizó un conjunto de datos de 386 turistas que visitaron la ciudad de Medellín (Colombia). Se emplearon cuatro modelos de aprendizaje estadístico: regresión logística, árboles aleatorios, máquinas de soporte vectorial y el incremento extremo del gradiente. Los hallazgos muestran que el mejor algoritmo son los árboles aleatorios, seguido de las máquinas de soporte vectorial con precisión de 0,63 y 0,62, respectivamente; con el algoritmo de bosques aleatorios se predice con más precisión los turistas que califican como agradable el destino turístico en comparación con otras ciudades visitadas. Por su parte, para identificar los turistas que califican de forma más negativa el destino, las máquinas de soporte vectorial mostraron mejor precisión. El desempeño del modelo para generalizar el resultado fue modesto en la base de prueba. Se abre un área de investigación en la evaluación de los destinos turísticos con las percepciones de los turistas usando modelos de aprendizaje estadístico.

Palabras clave: satisfacción del turista, motivaciones, modelos de evaluación, aprendizaje estadístico.

Introducción

Las motivaciones del turista ocupan un papel preponderante en los modelos de relacionamiento con la imagen, satisfacción y lealtad con el destino visitado (Yoon & Uysal, 2005). No obstante, la evidencia empírica que corrobora la importancia relativa de las motivaciones provienen estrictamente de modelos relacionales basados en estructuras de ecuaciones estructurales (Do Valle *et al.*, 2006; Prebensen *et al.*, 2010; Albayrak & Caber, 2018; Luna-Cortés, 2020) y modelos logísticos (Jang & Cai, 2002; Huang *et al.*, 2018; Lam-González *et al.*, 2019). En este

sentido, el objetivo central de este trabajo se centra en predecir la satisfacción del visitante mediado por las variables de motivación. De esta manera, si el conjunto de factores motivacionales explica o predice el nivel de satisfacción del turista, es necesario implementar modelos basados en aprendizaje estadístico para demostrar tanto la estimación como la predicción de estos factores sobre los cambios en la satisfacción.

En el presente trabajo se aborda esta brecha investigativa desde la perspectiva metodológica. Así, se utilizan tres de los principales algoritmos por su desempeño y rendimiento en las competencias de clasificación tipo *Kaggle*: máquinas de soporte vectorial, árboles aleatorios y el incremento extremo del gradiente (*XGBoost*). Además, se utiliza la regresión logística como algoritmo de base para la comparación de los rendimientos, por ser de los más usados en el área de investigación de motivaciones del turista. Los algoritmos que implementamos se usan para predecir la satisfacción de los turistas con base en conjunto de variables motivacionales (*push* y *pull*).

Revisión de la literatura: motivaciones y satisfacción

Diversos tipos de factores determinan la elección de un destino por parte de un turista. La literatura académica ha abordado los determinantes críticos de las motivaciones, la satisfacción y la imagen que se forja un visitante al momento de disfrutar de un destino (Yoon & Uysal, 2005; do Valle *et al.*, 2006; Chi & Qu, 2008; Lee, 2009; Olague de la Cruz, 2015). Vale la pena reseñar algunos estudios que soportan este trabajo. La mayor parte de éstos, optan por metodología multivariadas como forma de explicar las variables independientes y su relacionamiento con la variable dependiente. Así, Yoon & Uysal (2005) en su estudio estudian un modelo de ecuaciones estructurales para comprender mejor las motivaciones que tienen los turistas para viajar a un destino, además de aportar al tema de la relación que tienen las motivaciones "*pull* y *push*", satisfacción y lealtad del destino. Usando esta misma técnica, do Valle *et al.* (2006) indagan por la relación que existe entre la satisfacción del viaje e intención de lealtad al destino. Chi & Qu (2008) ofrecen un enfoque sobre la lealtad del destino mediante la examinación de la relación entre la imagen del destino, los atributos turísticos y la satisfacción. Lee (2009) examina el comportamiento de los turistas con cinco factores, imagen del destino, satisfacción, motivación, actitud y comportamiento futuro, en el turismo de humedales. Autores

como Villamediana-Pedrosa *et al.* (2020) y Dean & Suhartanto (2019) analizan las motivaciones usando modelos de regresión con escalamiento óptimo en el primer caso, mientras que en el segundo se usó con mínimos cuadrados parciales; la estrategia de usar el escalonamiento óptimo permite convertir datos de tipo tabulación a una forma continua, la cual permite un amplio uso de estadística paramétrica. Para el caso latinoamericano, más concretamente en Monterrey (México) Olague de la Cruz (2015) propone un modelo capaz de medir y explicar la satisfacción y lealtad de los turistas.

En la literatura, también se encuentran estudios comparados a nivel de varios países. Tal es el caso del trabajo desarrollado por Kozak (2002) que determina la existencia de diferencias motivacionales de turistas alemanes y británicos al momento de visitar España y Turquía; al final el autor encuentra que los motivos difieren por nacionalidades y el lugar visitado. Otros ejercicios han mostrado relaciones comparativas en asistencia a eventos (Nicholson & Pearce, 2001); estrategias de mercado en viajes de placer hacia Corea (Kim & Prideaux, 2005); casos comparados de huéspedes de hoteles tailandeses y australianos (O'Mahony, Sophonsiri & Turner, 2013); o hasta el caso de diferencias intraculturales entre turistas chinos y taiwaneses (Kim, Penny Wan & Pan (2015), y la cultura musical en las motivaciones de turistas chinos y rusos (Whang, Yong & Ko 2016). Por su parte, Dos Santos (2010) compara los problemas que pueden existir en los destinos y la imagen que tienen los turistas de estos en Londres, París y Roma, encontrando que se debe diseñar estratégicamente la imagen del destino, con el fin de brindar servicios de calidad que generen motivación y satisfacción.

Muy pocos ejercicios usan técnicas alternativas en el caso concreto de la satisfacción. Uno de éstos expuso una máquina de soporte vectoriales para la evaluación del grado de satisfacción del turista (Li *et al.*, 2009). Otro de los trabajos es desarrollado por Ghaderi *et al.* (2018) que propone un modelo de redes neuronales en la evaluación de la gestión de la cadena de suministro del turismo que influye en la satisfacción de los turistas. En otros temas específicos de turismo, las herramientas que aplican el aprendizaje estadístico van ganando terreno; aunque es de destacar la creciente utilización de este tipo de ayudas en los ejercicios de modelación que van desde redes neuronales artificiales, árboles de decisión, métodos probabilísticos y bayesianos, aprendizaje basado en

instancias, conjuntos (*ensembles*), agrupamiento y minería de reglas de asociación (Guerra-Montenegro *et al.*, 2021).

Para el análisis de segmentación de turistas se han usado técnicas de redes neuronales con algoritmos autoorganizados y de retropropagación (Bloom, 2004). También, para el análisis de segmentación de mercado en hoteles y predicción de elecciones de viaje, se han combinado técnicas de agrupación y reducción de dimensiones (mapas auto-organizados y descomposición de valores singulares de orden superior) con técnicas de predicción usando árboles de decisión (Ahani *et al.*, 2019); para el sentido del lugar en un atractivo turístico se han usado modelos de regresión logística y modelos de análisis no supervisado como la asignación de Dirichlet latente (Song *et al.*, 2021); y para determinar los factores subyacentes en el valor del cliente en restaurantes se han empleado redes neuronales artificiales para el modelamiento de tópicos usando como variables de entrada la incrustación de palabras (*word embeddings*, por su sigla en inglés) (Kwon *et al.*, 2020). Sin embargo, dada la cantidad de algoritmos disponibles con demostrados rendimientos en las competencias computacionales tipo *Kaggle*, se abre la discusión sobre el mejor modelo los problemas de clasificación o identificación de patrones en el tema de satisfacción de los turistas con el destino turístico.

Materiales y métodos

Datos

La información fue obtenida a través de un cuestionario dirigido a 404 turistas que visitaron la ciudad de Medellín en diciembre de 2019. La recolección de información se realizó en diferentes hoteles y lugares turísticos de la ciudad. Luego del proceso de depuración quedaron 386 registros válidos con información completa. El cuestionario aplicado estaba compuesto por características sociodemográficas, motivaciones inherentes al viajero y motivaciones inherentes al destino, y la satisfacción con el destino visitado. Las variables de motivaciones son usadas para analizar la percepción del destino con relación a otras experiencias de los viajeros; en la literatura estas variables se denominan como endógenas (*push*) y exógenas (*pull*) al viajero. Éstas se midieron en una escala de 5 puntos, 1 nada importante, 2 poco importante, 3 neutral, 4 importante, 5 muy importante. Dichas variables serán la predictoras

del modelo y se observan en los siguientes cuadros; el primer cuadro muestra las motivaciones de empuje y el segundo las motivaciones de atracción.

Tabla 1

Motivaciones específicas según categorías

	Motivaciones de empuje		**Motivaciones de atracción**
*push*1.1	Estar activo físicamente	*pull*1.1	Es una ciudad moderna
*push*1.2	Reunirse o conocer otras personas	*pull*1.2	Tiene una atmósfera exótica
*push*1.3	Encontrar algo nuevo y excitante	*pull*1.3	Museos y galerías de arte
*push*1.4	Experimentar como viven otras personas	*pull*1.4	Restaurantes de alta calidad
*push*2.1	Experimentar nuevos o diferentes estilos de vida	*pull*2.1	Variedad de actividades para realizar
*push*2.2	Probar nuevos alimentos o comidas	*pull*2.2	Alojamiento económico
*push*2.3	Visitar lugares históricos	*pull*2.3	Destino de bajo costo
*push*2.4	Conocer gente nueva	*pull*2.4	Restaurantes económicos
*push*2.5	Ser libre para actuar como quiera	*pull*3.1	Se puede recorrer fácilmente
*push*3.1	Cambiar de ambiente por tanto trabajo	*pull*3.2	Tiene un muy buen clima
*push*3.2	Ir a lugares que mis amigos no han visitado	*pull*3.3	Es segura
*push*3.3	Hablar sobre experiencias de viaje en el futuro	*pull*6.0	Paisaje excepcional
*push*3.4	Redescubrir experiencias pasadas	*pull*6.1	Variedad de patrimonio cultural diferente al mío
*push*3.5	Escapar del estrés diario	*pull*6.2	Gente local interesante y amable
*push*5.1	Visitar lugares que mi familia ha estado	*pull*6.3	Diferentes culturas
*push*5.2	Visitar amigos o parientes-Estar junto a familiares	*pull*6.4	Pueblos antiguos e históricos
*push*6.1	Alejarme de las demandas del hogar	*pull*7.1	Limpieza de sus calles
*push*6.2	Vivir un estilo de vida más simple	*pull*7.2	Realizar compras
*push*7.1	Estar entretenido y pasarla bien	*pull*8.1	Vida nocturna y entretenimiento
*push*7.2	Viajar a bajos precios	*pull*8.2	Cocina local

Motivaciones de empuje		Motivaciones de atracción	
*push*8.1	Sentirme como en casa lejos de ella	*pull*10.1	Actividades económicas
*push*8.2	Conocer tanto como sea posible	*pull*10.2	Actividades deportivas
		*pull*10.3	Turismo sexual
		*pull*10.4	Turismo médico

La variable de interés es la satisfacción con el destino en comparación con otras ciudades visitadas, la pregunta que se realizó fue: "¿Cómo calificaría Usted a Medellín como un destino para vacaciones comparado con otras ciudades similares que ha visitado?". La variable de interese se midió en una escala de 5 valores desde 1=mucho peor hasta 5=mucho mejor). El valor de 1 no se presentó en las respuestas, los porcentajes de distribución de las demás respuestas fueron: 2 el 4,92%, 3 el 28,49%, 4 el 38,86% y la categoría 5 el 27,72%. Para la clasificación se colapsaron los valores 1, 2 y 3 en la clase "Aceptable" y los valores 4 y 5 en la clase "Agradable". La clase "Agradable" es la principal y se recodifica como 1 y la clase "Aceptable" como 0. La base de datos no presenta valores perdidos, todas las variables tienen datos completos para cada observación.

Modelos

El problema es de clasificación, entonces, los algoritmos de clasificación estiman funciones para asignar las observaciones en una de las clases dadas en el problema. Para el presente trabajo implementamos una regresión logística (*Rl*), *Support Vector Machines* o máquinas de soporte vectorial (SVM), *Random Forest* o árboles aleatorios (*RF*) y *eXtreme Gradient Boosting* o incremento extremo del gradiente (*XGBoost*). La línea base de comparación será el resultado de la regresión logística que se considera un método de estimación paramétrica, es decir, parte del conocimiento a priori de la distribución de los datos. La comparación se hará contra los demás algoritmos que son de estimación no paramétrica: SVM, RF y *XGBoost*, este último considerado el "*gold standard*" para los tipos de datos de tabulación, es decir, cuando las variables predictoras son nominales u ordinales.

Regresión logística

La regresión logística permite predecir una respuesta binaria del problema de clasificación, así: $p(X) = Pr(Y=1 \mid X)$, donde la probabilidad de X sigue la siguiente función logística:

$$p(X) = \frac{e^{\beta_0+\beta_1 X}}{1 + e^{\beta_0+\beta_1 X}}$$

El ajuste se hace por el método de máxima verosimilitud. La regresión logística produce una curva en función de S, entonces, de acuerdo al valor de X se obtiene la predicción. La regresión logística es un método de clasificación usado como línea base de comparación, debido a que su desempeño ha demostrado ser menor que los algoritmos más avanzado como los árboles aleatorios o las máquinas de soporte vectorial.

Máquinas de soporte vectorial

Las máquinas de soporte vectorial son una generalización de los clasificadores de margen máxima, pero con capacidad para acomodarse a problemas donde las clases a separar tienen fronteras no lineales (James *et al.*, 2013). Para manejar el problema en el cual la frontera de separación de clases es no lineal, se extienden las llamadas máquinas de soporte vectorial de margen dura. Entonces, el espacio original de medición de las características se cambia aplicando una serie de funciones sobre las variables predictoras y resolviendo un problema de optimizar el hiperplano separador. Al usar una función no lineal sobre las variables originales, el algoritmo mapea el vector de variables originales en un espacio de dimensión superior llamado espacio de características. Sin embargo, una ventaja del algoritmo es la posibilidad de usar el truco *Kernel* para evitar el manejo explícito del espacio de características, aunque se mapean las variables originales en el espacio de características (Abe, 2005). Por consiguiente, el *Kernel* hace parte de la elección que debe realizarse al implementar el algoritmo, puesto que existen funciones *Kernel*s lineales, polinomiales, de base radial, basados en distancia de Mahalanobis, además de funciones *Kernel* desarrolladas para problemas específicos de procesamiento de imágenes, clasificación de texto o reconocimiento de voz; sin embargo, el *Kernel* de base radial ha sido el más usado (Abe, 2005).

Árboles aleatorios (Random Forest)

Es uno de los algoritmos más usados para las tareas de clasificación por sus altos desempeños y la identificación de las variables explicativas. El algoritmo es un método de ensamblaje que usa diferentes árboles de decisión construidos sobre muestras de entrenamiento *boostrapeadas*; entonces, cada vez que se construye un árbol se hace un muestreo aleatorio para elegir m variables predictoras que se usan para la partición de cada nodo del árbol. La elección típica de la cantidad de predictoras $m \approx \sqrt{p}$ donde p es el número total de variables predictoras. La idea de usar diferentes variables predictoras para la construcción de los árboles genera la aleatoriedad y reduce la varianza, comparado con los métodos clásicos basados en árboles (James *et al.*, 2013). El algoritmo es robusto a *outliers* y el ruido, también es más rápido que el *Adaboost* y el *Bagging*, podemos obtener una estimación interna del error, la correlación y las variables de importancia (Breiman, 2001).

XGBoost

Al igual que los árboles aleatorios, el *XGBoost* está basado en ensamblaje de árboles. La idea del ensamblaje es reducir el error de predicción o error de generalización, en el caso del *XGBoost* se enfoca en reducir el sesgo, mejorando la velocidad de cálculo y la precisión. A diferencia del RF, el *XGBoost* usa el algoritmo de gradiente descendente para la optimización (*Gradient Boosting*), la idea es ir corrigiendo los errores del modelo y aprender en las siguientes iteraciones para mejorar el rendimiento hasta que no se logren mejoras.

Experimentos

Para estimar los cuatro algoritmos de aprendizaje estadístico, se realizó una malla de búsqueda con 20 combinaciones diferentes de los hiperparámetros de los modelos ajustados. El entrenamiento se realizó con el paquete *tydimodels* de *R-project* perteneciente al ecosistema de *tidyverse*. Para realizar una eficiente comparación entre los modelos se usó un flujo de trabajo común entre ellos, es decir, se usa la misma validación cruzada de 10 particiones con 5 repeticiones en cada una; igualmente, se usan las mismas métricas de desempeño, a saber: la precisión (*accuracy*), la curva *roc*, la sensibilidad y especificidad. En total, se entrenaron 3031 modelos. Una vez realizado el preprocesamiento a

los datos, se balancearon las clases usando un método de elección por muestreo aleatorio de las observaciones con la clase predominante, así se obtuvo una base de datos balanceada para el entrenamiento; así mismo, por los procesos de optimización de los algoritmos usamos una matriz de *one-hot* dicotomizando las variables para correr los algoritmos de SVM y *XGBoost*, para el caso de la regresión logística y los árboles aleatorios no se aplicó la transformación de las variables. Los parámetros de los modelos que se ajustaron en la malla de búsqueda y la validación cruzada, fueron los siguientes:

- En el algoritmo de árboles aleatorios: el número de variables que se usa en cada partición del nodo (*mtry*), la cantidad de árboles se fijaron en 1000, el mínimo de observaciones para activar un nodo dentro de cada árbol *(min_n*).

- En las máquinas de soporte vectorial: el costo de la violación de la restricción, es decir, la constante del término de regularización en la formulación del lagrangiano (*cost*); el valor del sigma del *Kernel* gaussiano (*rbf_sigma*).

- En el *XGBoost*: el número de variables que se usa en cada partición del nodo (*mtry*), el número de árboles del ensamblaje, el mínimo de observaciones para activar un nodo dentro de cada árbol (*min_n*), la profundidad máxima de cada árbol (*tree_depth*), la tasa de aprendizaje (*learn_rate*), la reducción en la función de pérdida (*loss_reduction*) para crear una nueva partición en el árbol, el número de observaciones para el ajuste en una estimación (*sample_size*).

Resultados

Análisis descriptivo

Las figuras 1 y 2 muestran las respuestas de los turistas evaluando los factores individuales que inciden en la elección del destino (*push*) y los factores o características del destino mismo (*pull*); la calificación tiene una escala de 1 a 5, donde 1 es nada importante, 2 es importante, 3 es neutral, 4 importante y 5 muy importante.

El resultado descriptivo muestra que las variables "*push* 5.1" y "*push* 5.2" son aquellas donde más del 50% de los turistas responde que son aspectos nada importantes, relacionados con visita a familiares o recomendaciones de ellos (Fig 1). Igualmente, se observa un porcentaje alto de respuestas (alrededor del 45%) con valoraciones de nada importante y poco importante en los aspectos "*push* 3.2" y "*push* 3.4", que están relacionados con ir a los lugares que los amigos no han visitado y redescubrir experiencias pasadas. En cuanto a las variables con calificaciones altas de importancia, se observan el grupo 1, desde la 1.2 asociado con la emoción y la pasión por viajar, el grupo 2 hasta la 2.4 que se relaciona con el conocimiento y el aprendizaje; la variable 7.1 relacionado con pasarla bien y la 8.2 con conocer tanto como sea posible.

Para el caso de las motivaciones de atracción (Fig. 2), aquellas relacionadas con un tipo de turismo particular (sexual o médico) aparecieron como las menos importantes para los visitantes, así como el tema gastronómico de calidad (1.4) y realización de compras (7.2). Las motivaciones con una calificación de mayor importancia para los viajeros aquellas en el grupo de la cultura y escenario natural de la región (6.0 a 6.4) y el tema climatológico (3.2) y la variedad actividades a realizar (2.1).

Figura 1

Distribución de frecuencias en las respuestas de motivaciones push

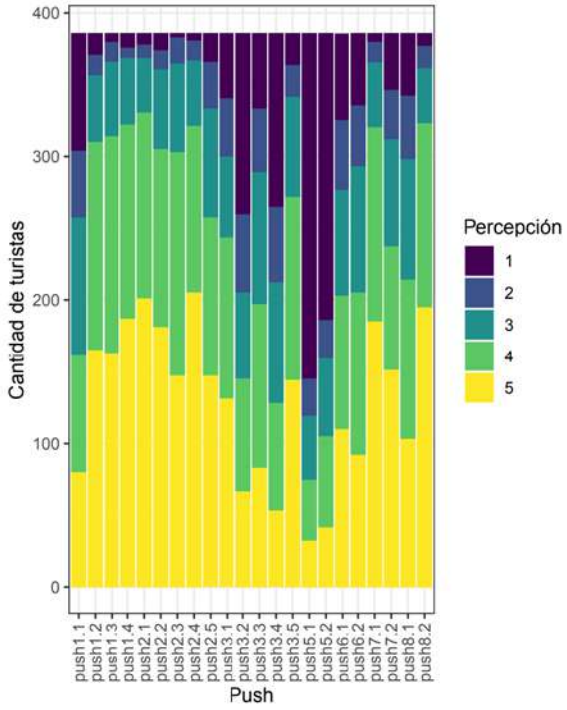
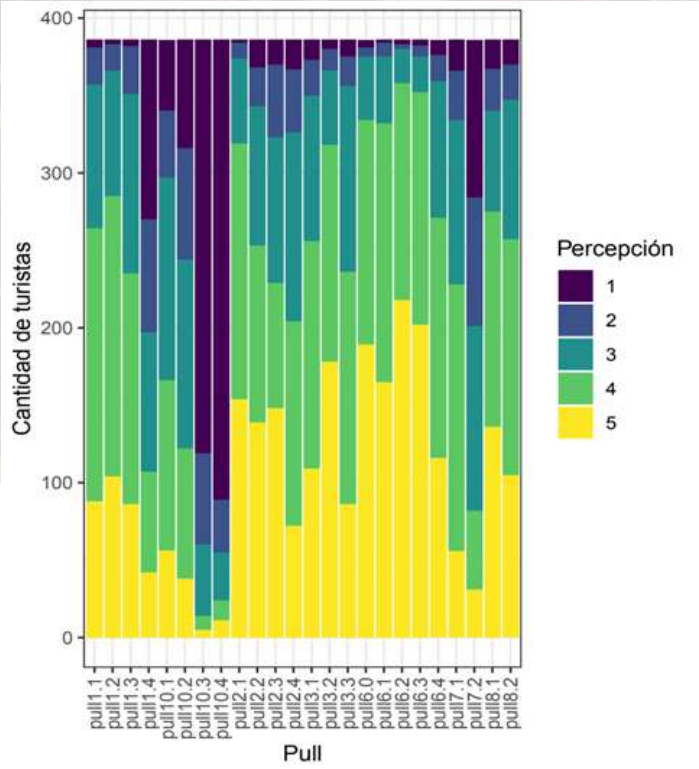


Figura 2

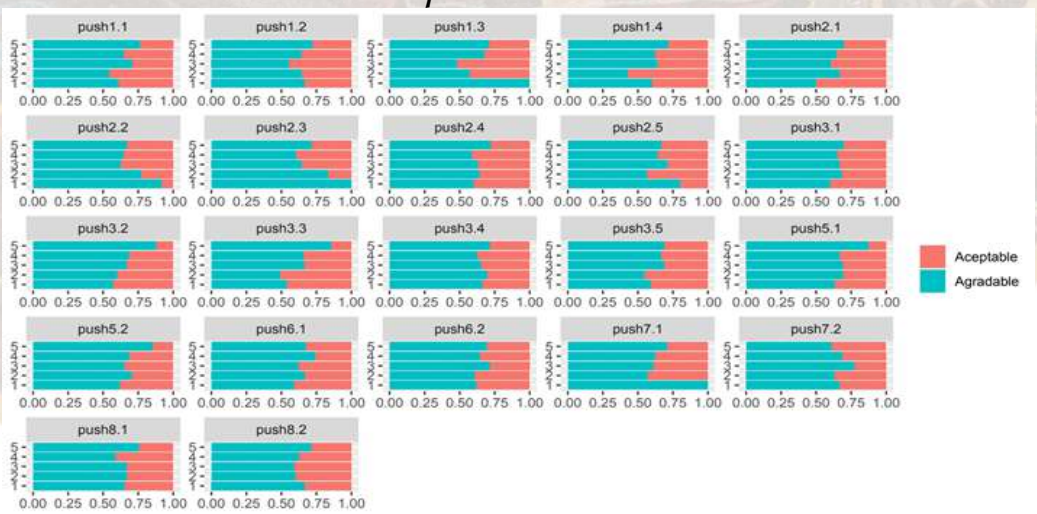
Distribución de frecuencias en las respuestas de motivaciones pull



Ahora bien, si se cruza cada nivel de la respuesta ordenada de los factores motivacionales con los dos niveles de la variable explicativa codificada en aceptable y agradable, se pueden observar algunos rasgos distintivos. Las figuras 3 y 4 ofrecen este panorama. Así, se puede esperar que motivaciones del grupo de la relajación (3.1 a 3.4), del grupo de lazos (5.1 y 5.2), así como del grupo de comodidad (8.1 a 8.2) tendieron a asignar un nivel de satisfacción más alto, lo cual está relacionado con la valoración agradable de la ciudad.

Figura 3

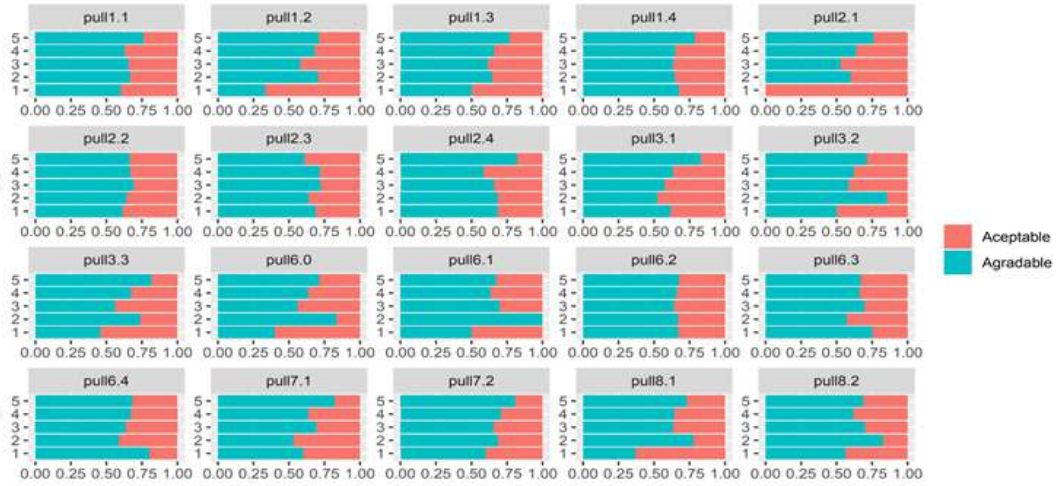
Distribución de respuestas en la variable salida según motivaciones push



Entretanto, las motivaciones inherentes al destino es más complejo observar una tendencia de respuesta en la relación de una separación lineal clara entre las variables respuesta de los niveles superiores con el nivel de satisfacción de la ciudad. De esta forma, para el tema de los algoritmos propuestos es más retador evidenciar esta situación.

Figura 4

Distribución de respuestas en la variable salida según motivaciones pull



Comparación de modelos y desempeño

La Figura 5 muestra la comparación de los modelos entrenados, los árboles aleatorios resultados ser el mejor modelo para la clasificación de la comparación de ciudades. Los mejores hiperparámetros ajustados en el malla de búsqueda fueron: *mtry* = 4 y *min_n* = 30 con 1.000 árboles establecido por defecto. La precisión del modelo es modesta, logrando un 0,63 (*accuracy*) y un 0,66 en la precisión del área bajo la curva *roc*. El resultado modesto se puede deber a la limitación en la cantidad de los datos para realizar el proceso de entrenamiento, igualmente, a la complejidad del problema evidenciado en los resultados descriptivos, donde no se observaba una separación de las clases intuitiva. Es importante señalar que el algoritmo de árboles aleatorios tiene mayor capacidad para predecir la clase positiva (sensibilidad de 0,72), es decir, es capaz de identificar la clase "Agradable" con más precisión que la clase "Aceptable", pues la especificidad fue de 0,58. En este sentido, el modelo logra capturar si la ciudad es agradable en comparación con las otras que ha visitado el turista.

Figura 5

Comparación de modelos

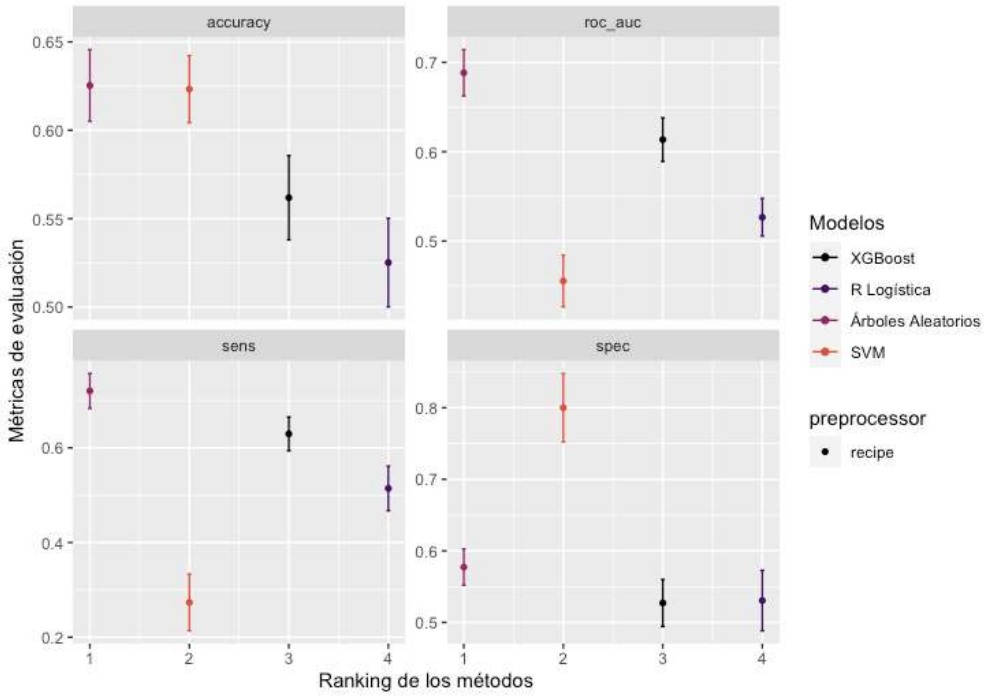


Tabla 2

Jerarquía de los mejores modelos y sus métricas asociadas dentro cada algoritmo

Modelo	Métrica	Promedio	Ranking
rand_forest	accuracy	0,63	1
rand_forest	roc_auc	0,69	1
rand_forest	sens	0,72	1
rand_forest	spec	0,58	1
svm_rbf	accuracy	0,62	2
svm_rbf	roc_auc	0,46	2
svm_rbf	sens	0,27	2
svm_rbf	spec	0,80	2
boost_tree	accuracy	0,56	3
boost_tree	roc_auc	0,61	3
boost_tree	sens	0,63	3
boost_tree	spec	0,53	3
logistic_reg	accuracy	0,53	4
logistic_reg	roc_auc	0,53	4
logistic_reg	sens	0,51	4
logistic_reg	spec	0,53	4

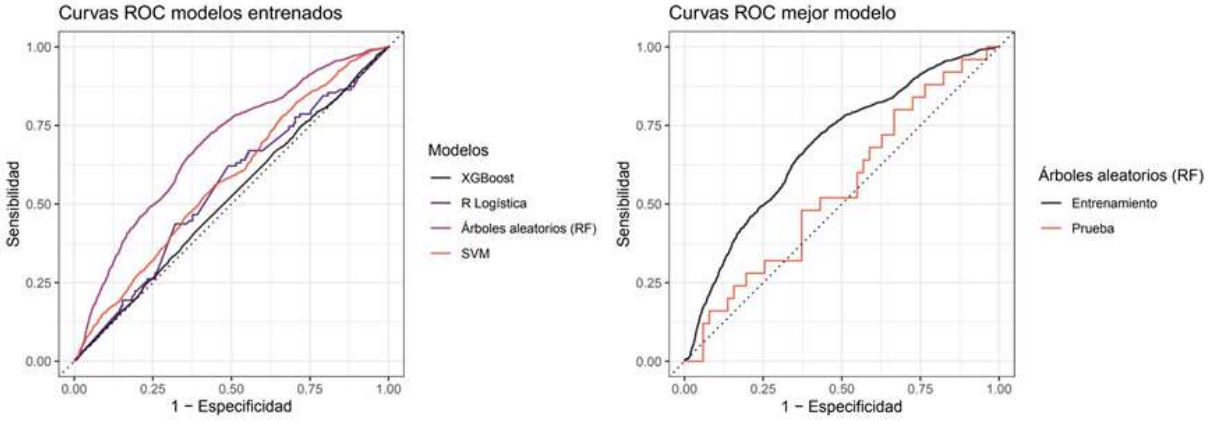
Con un rendimiento similar se observa el algoritmo de las máquinas de soporte vectorial con un *Kernel Gaussiano* (Fig. 5). Se obtuvo una precisión de 0,62, solo un punto por debajo de los árboles aleatorios. Los interesante con el resultado se observa en el valor de la especificidad que llega a 0,80, es decir, con el modelo se puede lograr predecir la clase "Aceptable" con buen desempeño. En este sentido, las máquinas de soporte vectorial permiten clasificar los turistas que evalúan la ciudad destino, en función de otras visitadas, de manera más negativa. Si se quiere clasificar los turistas con valoraciones positiva la mejor opción son los árboles aleatorios, en caso contrario, si se quiere clasificar a los turistas con valores negativos del destino, la opción más recomendada es usar las máquinas de soporte vectorial. Los demás modelos presentan desempeños pobres para problema (Tabla 2).

El desempeño de los modelos se aprecia en la figura 6, donde se presenta la curva *roc* de cada modelo entrenado y la curva solo para un mejor modelo, pero en el conjunto de datos de prueba. Se esperaba que el modelo tuviera un modesto desempeño en el conjunto de datos de

prueba, puesto que el desempeño en entrenamiento no fue el mejor. La precisión de clasificación en el conjunto de prueba llego a 0,53; en cuanto a las medidas de desempeño de especificidad y sensibilidad, el resultado es poco alentador puesto que ambas mediciones ubican en 0,53. Es decir, el modelo no logró generalizar la capacidad para predecir más una clase que otra.

Figura 6

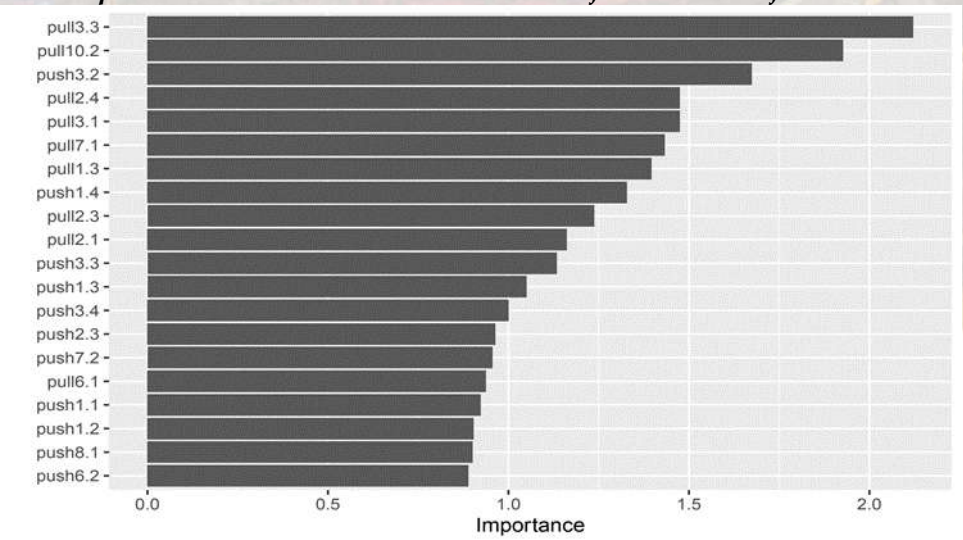
El primer gráfico son las curvas ROC de los cuatro modelos entrenados. El segundo gráfico las curvas roc del mejor modelo (RF) en el conjunto de datos de entrenamiento y prueba.



Una ventaja de los métodos basados en árboles es la posibilidad de identificar la importancia de las variables predictoras en el modelo y así obtener explicación de la clasificación. En la figura 7 se muestran las 20 variables más importantes del modelo de árboles aleatorios. Destacan en su orden: sensación de seguridad, la posibilidad de practicar actividades deportivas, ir a lugares que los amigos no han visitado, conocer gente nueva y se puede recorrer fácilmente.

Figura 7

Importancia de las variables en el mejor modelo ajustado (RF)



Conclusiones

En este estudio se probaron algunos de los principales modelos de clasificación encontrados en la literatura relacionada con el aprendizaje estadístico. Se comparan cuatro modelos de clasificación para clasificar los turistas que evalúan el destino turístico como "Agradable" o "Aceptable". La clasificación se realiza tomando con referencia un conjunto de factores motivacionales tanto internos, denominados "*push*" y externos, denominados "*pull*". Los algoritmos probados fueron la regresión logística, los árboles aleatorios, las máquinas de soporte vectorial y el *XGBoos*t; los resultados en el conjunto de entrenamiento evidencian que el mejor modelo fue los árboles aleatorios logrando un 0,63 de precisión, seguido por las máquinas de soporte vectorial con *Kernel Gaussiano* que alcanzaron un 0,62 de precisión.

Al estimar los modelos en los datos de prueba, los desempeños fueron modesto y solo se logran valores de precisión del 0,53. Cabe anotar que, en nuestro estudio evidenciamos mejor precisión de clasificación en los algoritmos de árboles aleatorios y máquinas de soporte vectorial que en el algoritmo de *XGBoost*, aunque este último ha sido uno de los principales para la precisión de las competencias de aprendizaje estadístico.

La ventaja de los árboles aleatorios permite identificar las variables que más aportaron a la composición del modelo. Las motivaciones de sensación de seguridad, la posibilidad de practicar actividades deportivas e ir a lugares que los amigos no han visitado, explican la satisfacción del viajero, principalmente en la categoría aceptable. Del análisis desprende que esos factores elevan la ponderación de la estadía comparado con otros viajes realizados por el turista.

Los resultados se consideran como línea base para le investigación de modelos de aprendizaje estadístico y la motivación de los turistas a visitar un destino, un área de estudio prometedora por los constantes avances en la modelación y los diferentes algoritmos que pueden ayudar a la toma de decisiones informada. La estimación de diversos modelos sirve como herramienta de planificación para desarrollar propuestas de atracción turística a través de estrategias de marketing o de planificación de recursos del destino. Puesto que los modelos presentan desempeños diferentes, por ejemplo, los árboles aleatorios tienen mejor desempeño para predecir los turistas que califican con agradable del destino, mientras las máquinas de soporte vectorial tienen mejor desempeño prediciendo la clase contraria. Ambos en el conjunto de datos de entrenamiento, puesto que las pruebas de generalización del modelo en los datos de prueba no logran resultados satisfactorios.

El estudio tiene limitación en datos, con pocos datos el proceso de entrenamiento es limitado porque el algoritmo no logra aprender la variabilidad de los datos para lograr mejores resultados. Igualmente, las variables asociadas como relaciones causales de la elección de destino turístico requieren más discusión teórica y empírica para lograr precisar las variables predictoras. En trabajos futuros se espera probar con redes neuronales para datos categóricos, aunque la gran limitación son los pocos datos con los que contamos para resolver el problema. Igualmente, se plantea estrategias de aumentación de datos o corregir el desbalance de las clases aumentando muestras en la clase con menor representación en la base original, usando métodos de *upsample*.

Referencias

Ahani, A., Nilashi, M., Ibrahim, O., Sanzogni, L., & Weaven, S. (2019). Market segmentation and travel choice prediction in Spa hotels through TripAdvisor's online reviews. *International Journal of Hospitality Management*, 80, 52–77.

Albayrak, T., & Caber, M. (2018). Examining the relationship between tourist motivation and satisfaction by two competing methods. *Tourism Management*, 69, 201–213.

Almeida, P. (2010). La Imagen de un Destino Turístico como Antecedente de la Decisión de Visita: análisis comparativo entre los destinos.

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5–32. Abe, S. (2005). *Support vector machines for pattern classification*. (Vol. 2) Springer.

Chen, N., Xie, W., Welsch, R., Larson, K., & Xie, J. (2017). Comprehensive predictions of tourists' next visit location based on call detail records using machine learning and deep learning methods. In 2017 *IEEE International Congress on Big Data (BigData Congress)* (pp. 1–6).

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (*pp. 785–794).*

Chi, C., & Qu, H. (2008). Examining the structural relationships of destination image, tourist satisfaction and destination loyalty: An integrated approach. *Tourism Management, 29(4), 624–636.*

Dean, D., & Suhartanto, D. (2019). The formation of visitor behavioral intention to creative tourism: the role of push–Pull motivation. *Asia Pacific Journal of Tourism Research, 24(5), 393–403.*

Do Valle, P., Silva, J., Mendes, J., & Guerreiro, M. (2006). Tourist satisfaction and destination loyalty intention: A structural and categorical analysis. *International Journal of Business Science & Applied Management (IJBSAM), 1(1), 25–44.*

Dos Santos Almeida, P. (2010). La Imagen de un Destino Turístico como Antecedente de la Decisión de Visita: análisis comparativo entre los destinos Londres, París y Roma. (Tesis Doctoral, Universidad de Extremadura).

Fernández, M., & Picos, A. (2005). Predicciones en el nivel de satisfacción percibida por los turistas a partir de variables

motivacionales y de valoración de la visita. Información Comercial Española, 821, 241.

Ghaderi, Z., Hatamifar, P., & Khalilzadeh, J. (2018). Analysis of tourist satisfaction in tourism supply chain management. *Anatolia, 29(3), 433–444*.

Guerra-Montenegro, J., Sanchez-Medina, J., La\~na, I., Sanchez-Rodriguez, D., Alonso-Gonzalez, I., & Del Ser, J. (2021). Computational Intelligence in the hospitality industry: A systematic literature review and a prospect of challenges. *Applied Soft Computing*, 107082.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. (Vol. 112) Springer.

Jang, S., & Cai, L. (2002). Travel motivations and destination choice: A study of British outbound market. *Journal of Travel & Tourism Marketing, 13(3),111–133.*

Bloom, J. (2004). Tourist market segmentation with linear and non-linear techniques. *Tourism Management, 25(6), 723–733*.

Kim, S., & Prideaux, B. (2005). Marketing implications arising from a comparative study of international pleasure tourist motivations and other travel-related characteristics of visitors to Korea. Tourism Management, 26(3), 347-357.

Kim, S. S., Penny Wan, Y. K., & Pan, S. (2015). Differences in tourist attitude and behavior between Mainland Chinese and Taiwanese tourists. Journal of Travel & Tourism Marketing, 32(1-2), 100-119.

Kozak, M. (2002). Comparative analysis of tourist motivations by nationality and destinations. Tourism management, 23(3), 221-232.

Kwon, W., Lee, M., & Back, K.J. (2020). Exploring the underlying factors of customer value in restaurants: A machine learning approach. *International Journal of Hospitality Management, 91, 102643*.

Lam-González, Y., León, C., & León, J. (2019). Coopetition in maritime tourism: assessing the effect of previous islands' choice and experience in tourist satisfaction. *Sustainability, 11(22), 6334.*

Lee, H., & Kang, Y. (2021). Mining tourists' destinations and preferences through LSTM-based text classification and spatial clustering using Flickr data. *Spatial Information Research*, 1–15.

Lee, T. (2009). A structural model to examine how destination image, attitude, and motivation affect the future behavior of tourists. *Leisure Sciences, 31(3), 215–236*.

Lee, Y., & Chang, Y. (2019). Uncovering Los Angeles Tourists' Patterns Using Geospatial Analysis and Supervised Machine Learning with Random Forest Predictors. In 2019 *International Conference on Computational Science and Computational Intelligence (CSCI*) (pp. 1275–1280).

Luna-Cortés, G. (2020). Análisis de la percepción de los estadounidenses que visitan Colombia. Un modelo de ecuaciones estructurales. *Estudios y perspectivas en turismo, 29(1), 51–71.*

Mirzaalian, F., & Halpenny, E. (2021). Exploring destination loyalty: Application of social media analytics in a nature-based tourism setting. *Journal of Destination Marketing & Management, 20, 100598.*

Nicholson, R., & Pearce, D. (2001). Why do people attend events: A comparative analysis of visitor motivations at four South Island events. *Journal of Travel Research*, 39(4), 449-460.

O'Mahony, G. B., Sophonsiri, S., & Turner, L. W. (2013). The impact of the antecedents of relationship development on Thai and Australian resort hotels guests. International *Journal of Hospitality Management*, 34, 214-226.

Olague de la Cruz, J. (2015). La imagen del destino y la motivación de viaje como determinantes de la satisfacción y lealtad del turismo urbano de ocio en Monterrey, México (Tesis Doctoral, Universidad Autónoma de Nuevo León).

Paolanti, M., Mancini, A., Frontoni, E., Felicetti, A., Marinelli, L., Marcheggiani, E., & Pierdicca, R. (2021). Tourism destination management using sentiment analysis and geo-location information: a deep learning approach. *Information Technology & Tourism*, 1–24.

Prebensen, N., Skallerud, K., & Chen, J. (2010). Tourist motivation with sun and sand destinations: satisfaction and the wom-effect. *Journal of Travel & Tourism Marketing, 27(8), 858–873.*

Pulido-Fernández, J., Cárdenas-García, P., & Durán-Román, J. (2017). Socio-economic profile of tourism expenditure in emerging urban-cultural destinations. *Tourism Planning & Development, 14(3), 318–336*.

Song, Y., Wang, R., Fernandez, J., & Li, D. (2021). Investigating sense of place of the Las Vegas Strip using online reviews and machine learning approaches. *Landscape and Urban Planning, 205, 103956.*

Valls, A., Gibert, K., Orellana, A., & Antón-Clavé, S. (2018). Using ontology-based clustering to understand the push and pull factors for British tourists visiting a Mediterranean coastal destination. *Information & Management, 55(2), 145–159.*

Villamediana-Pedrosa, J., Vila-Lopez, N., & Küster-Boluda, I. (2020). Predictors of tourist engagement: Travel motives and tourism destination profiles. *Journal of Destination Marketing & Management, 16, 100412.*

Whang, H., Yong, S., & Ko, E. (2016). Pop culture, destination images, and visit intentions: Theory and research on travel motivations of Chinese and Russian tourists. *Journal of Business Research*, 69(2), 631-641.

Yoon, Y., & Uysal, M. (2005). An examination of the effects of motivation and satisfaction on destination loyalty: a structural model. *Tourism Management, 26(1), 45–56.*