



Universidad de Buenos Aires
Facultad de Ciencias Económicas
Escuela de Estudios de Posgrado



Universidad de Buenos Aires Facultad de Ciencias Económicas Escuela de Estudios de Posgrado

**CARRERA DE ESPECIALIZACIÓN EN
MÉTODOS CUANTITATIVOS PARA LA GESTIÓN Y ANÁLISIS DE
DATOS EN ORGANIZACIONES**

TRABAJO FINAL DE ESPECIALIZACIÓN

***Predicción de la demanda de pasajes aéreos a través de
técnicas de aprendizaje automático:
Aplicación al estudio de caso de una agencia de viajes***

AUTORA: CALÁ, MARÍA BELÉN

DICIEMBRE, 2019



Resumen

En el presente trabajo se utilizaron modelos de minería de datos para predecir la cantidad de pasajes aéreos vendidos a las regiones de Europa y Caribe y a su vez, determinar las variables que influyen en su venta. La elección de destinos pertenecientes a regiones distintas como caso de estudio se hizo con el objetivo de poder comparar los resultados obtenidos en destinos que no son similares en relación con sus características vacacionales.

Para ello, se entrenó el modelo con cuatro algoritmos de clasificación (*Decision Tree*, *K-NN*, *Random Forest* y *Gradient Boosting Trees*) sobre una base de 83.332 búsquedas diarias desde agosto 2018 a septiembre 2019; para evaluar la calidad de los modelos probados, se utilizó la métrica del error cuadrático medio (RMSE).

Palabras clave: *Aprendizaje automático, Predicción, Gestión, Demanda, Base de datos, Turismo, Pasajes aéreos.*



Estructura

Resumen	2
Introducción	4
Capítulo 1: Desafíos de la gestión de la demanda en la industria de viajes	7
1.1 El impacto de la tecnología en la industria de viajes	7
1.2 El manejo de grandes volúmenes de datos en las agencias de viajes online	8
1.3 La gestión de la demanda en la industria de viajes y su impacto en los negocios	10
Capítulo 2: Metodología y datos utilizados para el desarrollo del modelo de predicción de la demanda de pasajes aéreos	15
2.1 Selección de los datos	15
2.2 Preprocesamiento de los datos	16
2.3 Técnicas de minería de datos y métricas de evaluación	19
Capítulo 3: Implementación del modelo predictivo de la demanda de pasajes aéreos	21
3.1 Aplicación del modelo a través de técnicas de aprendizaje automático	21
3.2 Interpretación y evaluación de resultados	22
Conclusión.....	25
Futuros estudios.....	27
Referencias bibliográficas	28



Introducción

El sector al que una empresa pertenece es el factor clave para formular su estrategia competitiva. Dado que, el mismo determina el comportamiento que la organización debe seguir para lograr el éxito (Porter, 2008). En el sector turístico aparecen múltiples factores que configuran un nuevo paradigma y nuevas condiciones de la demanda. Entre estos factores cabe destacar el cambio tecnológico, la globalización de la actividad, la desregulación de los mercados y la fragmentación de la demanda (Rodríguez-Zulaica, 2017).

Asimismo, en este nuevo entorno en el que se encuentra la industria de viajes se destacan dos aspectos importantes. Por un lado, surgieron las agencias de viaje online u OTAS- *Online Travels Agencies*- que se dedican a vender servicios turísticos en internet y no dependen de sucursales u oficinas (Abad, 2006). Y por el otro lado, la complejidad de la industria les exige a las empresas que venden pasajes aéreos a resolver los problemas propios de maximización de ingresos mediante datos, análisis e inteligencia de negocio (Amadeus, 2018).

El esquema de ingresos de las agencias de viajes consta de tres fuentes. Dos de estas fuentes, los incentivos y las comisiones, son ingresos provenientes de las líneas aéreas por la venta de sus pasajes aéreos. Mientras que la otra variable proviene de los cargos que se les cobra a los clientes por el servicio prestado. Esto es decir que la manera que tienen las agencias de viajes de inducir a los clientes a inclinarse por la compra de determinada línea aérea es a través de los cargos (Abad, 2006).

En efecto, con el fin de realizar una conclusión integral que se adecue al contexto competitivo y exigente de la industria de viajes, el presente trabajo tiene como objetivo predecir la demanda de pasajes aéreos a las regiones de Europa y Caribe, y a su vez, analizar cuáles son las variables que influyen en la decisión de compra dichos destinos de viaje. Cabe aclarar que, de acuerdo con el destino, el pasajero tiene un comportamiento distinto, es decir que, las variables que definen la decisión de compra podrían variar también.

Asimismo, el presente trabajo es aplicado en un caso práctico de una agencia de viaje online denominada con el nombre ficticio de “Vuela Vuela”. Cabe aclarar que se hace referencia a la plataforma de venta online con un nombre de fantasía con el fin de cuidar la confidencialidad de la información brindada por la empresa y así, preservar su identidad. La empresa “Vuela Vuela” pertenece al sector retail; la misma vende pasajes aéreos, hoteles, departamentos en



alquiler, paquetes, cruceros, alquiler de autos y asistencia al viajero. Sin embargo, en el presente trabajo se hace foco en la predicción de la demanda del producto de aéreos.

Por consiguiente, la pregunta de investigación que guía el presente trabajo es la siguiente:

¿Cuáles son los métodos de aprendizaje automático que predicen con mayor precisión la demanda de pasajes aéreos de la agencia de viajes “Vuela Vuela” a las regiones de Caribe y Europa? ¿Qué acciones comerciales se pueden recomendar para maximizar los ingresos de la agencia de viajes en dichas regiones utilizando los modelos encontrados?

Con el fin de realizar una conclusión integral que responda a la pregunta formulada, el presente trabajo tiene como **objetivo general** desarrollar un modelo que prediga la demanda de pasajes aéreos de una agencia de viajes a las regiones de Europa y Caribe. Para cumplir con dicho objetivo general, se plantean los siguientes **objetivos específicos**:

1. Recopilar y preparar los datos para el armado del modelo.
2. Desarrollar diferentes modelos.
3. Analizar cuáles son las características que explican los cambios en la demanda a las regiones de Europa y Caribe.
4. Analizar las características del modelo que predice con mayor exactitud la demanda de pasajes aéreos a las regiones estudiadas.
5. Proponer estrategias de precios diferenciados que maximicen los ingresos de los aéreos de la agencia en los destinos de Europa y Caribe.

Para lograr un mejor entendimiento del análisis realizado, en un **primer apartado** se presentan los cambios que provocó la tecnología en los datos y la gestión de la demanda en la industria del turismo. Se profundiza en el impacto de la tecnología tanto en la industria de viajes como también en la gestión de la demanda de dicho sector. A su vez, se expone el problema que enfrenta la agencia de viajes estudiada frente a los grandes volúmenes de datos que genera la industria del turismo.

Luego, en un **segundo apartado**, se hace foco en la selección y procesamiento de los datos que serán utilizado para crear el set de datos que será entrenado con los algoritmos de clasificación que son explicados en el **tercer apartado**. En este capítulo, se explica sobre el modelo predictivo de la demanda de pasajes aéreos, haciendo foco en las técnicas de minería de datos y las métricas de evaluación, y luego, la aplicación, interpretación y evaluación de los algoritmos de clasificación utilizados para entrenar el modelo.



Universidad de Buenos Aires
Facultad de Ciencias Económicas
Escuela de Estudios de Posgrado



El cuarto apartado, contiene una **conclusión** integradora de todo el trabajo, sintetizando las principales resoluciones obtenidas como resultados del estudio realizado. Y, por último, se encuentra el apartado de **futuros estudios** en donde se exponen las oportunidades propuestas para crear futuras líneas de investigación.



Capítulo 1: Desafíos de la gestión de la demanda en la industria de viajes

En la actualidad, las firmas se encuentran inmersas en un contexto altamente competitivo por el que se enfrentan a posibles riesgos y contingencias futuras. A los efectos de mitigar estos factores, contar con una lectura exhaustiva de las principales variables del mercado pertinente comprende una ventaja competitiva para los agentes económicos. Ante este escenario, junto con los empleados, la información representa un activo de singular importancia en la operación de la firma (Mcafee et al., 2012).

Por lo cual, al tener presente que, en la actualidad, la comprensión y el manejo eficaz de la información conlleva uno de los principales desafíos que enfrentan las compañías, se presenta el siguiente capítulo. En el mismo se brinda conocimiento sobre el contexto altamente competitivo en el cual comercializan las organizaciones pertenecientes a la industria de viajes. A su vez, se propone la comprensión del impacto que tiene la tecnología en la agencia de viajes estudiada y en la gestión de la demanda del sector turismo.

1.1 El impacto de la tecnología en la industria de viajes

La evolución en la industria de viajes se puede ver afectado por los eventos geopolíticos, el crecimiento económico, las limitaciones de capacidad y los problemas de sustentabilidad. Primero, los eventos geopolíticos y de seguridad están afectando de manera negativa al turismo. Debido a que las amenazas de terrorismo conducen a restricciones en viajes, mayor seguridad y control fronterizo que se traducen en una caída en la demanda. Segundo, los niveles de crecimiento económico tienen un impacto significativo en el crecimiento de la industria de viajes. Por ejemplo, la tasa de crecimiento promedio de pasajeros que volaron es alrededor de 1,5 veces el PIB global, aunque multiplicador varía sustancialmente de una región a otra. Por lo tanto, las futuras recesiones o aumentos económicos tendrán un fuerte impacto en los volúmenes de viaje (Amadeus, 2018).

Tercero, el crecimiento de la industria se ve limitado por la capacidad, ya sea, la infraestructura de viajes como también del aeropuerto. Por ejemplo, en el verano de 2018, 204 aeropuertos fueron designadas instalaciones coordinadas de nivel 3, lo que significa que no tienen la capacidad de pista, rampa o puerta para manejar todo el volumen de los vuelos que a las aerolíneas les gustaría operar. Por último, el sector de viajes y turismo se ve desafiado por cuestiones de sustentabilidad tales como destinos superpoblados, desigualdades de ingresos y



variabilidad climática. Estos temas y otros relacionados con la sustentabilidad, podrían tener un marco legal mas restricto que afecte de manera negativa a la industria de viajes (Amadeus, 2018).

En este contexto, los autores afirman que la tecnología es importante para mejorar una amplia variedad de operaciones centrales para los proveedores de viajes. Primero, la tecnología permite realizar cambios en el momento en caso de interrupción del viaje o cambio de planes. Segundo, el uso del aprendizaje automático hace posible ajustar resultados de búsqueda basados en las preferencias del viajero. Cabe aclarar que, conocer dichas preferencias es útil para crear ofertas altamente personalizadas y contextuales para el próximo vuelo, habitación de hotel o alquiler del auto del viajero (Amadeus, 2018).

Tercero, la tecnología facilita las iniciativas emergentes destinadas a proporcionar una identidad digital estrecha vinculadas a la identidad física del usuario están permitiendo un procesamiento perfecto y personalizado del viajero. Por último, las nuevas tecnologías permiten una nueva forma para compartir, almacenar y usar datos, lo que lleva a una mayor eficiencia y procesos transparentes de intercambio de información durante el viaje (Amadeus, 2018).

Por lo cual, la evolución y el crecimiento de la industria de viajes son altamente dependientes en tecnología. Debido a que la tecnología permite conectar a los viajeros con los diferentes proveedores, permitiendo a los viajeros buscar y reservar lo óptimo para su viaje, que consta quizás de varios servicios de viaje (por ejemplo, avión, hotel y auto) (Amadeus, 2018).

1.2 El manejo de grandes volúmenes de datos en las agencias de viajes online

La evolución de las tecnologías móviles ha desencadenado una explotación de los datos generados diariamente a través de los diferentes dispositivos móviles y electrónicos que son utilizados en la actualidad (Mérida et al., 2017). Asimismo, se proyecta que se generen aproximadamente 40% más cantidad de datos anualmente que en el periodo anterior (Beyer, 2011). A su vez, el autor destaca que existe un 85% de cantidad de datos generados cada año de nuevas fuentes de información. Los autores Hilbert & López (2011) agregan que el 94% de los datos son digitales mientras que el 6% restante es analógico. Tal es así que, las compañías de todos los sectores de la industria almacenan por lo menos 100 terabytes de información en sus sistemas principales de cómputo y muchas de ella están llegando a 1 petabyte de información (Hilbert & López, 2011).



Dichos autores afirman que dicha expansión de los datos digitales desencadenó una necesidad en el mercado de explotar la incansable cantidad de datos. Sin embargo, los autores Schroeck et al. (2012) exponen que para algunas compañías el uso del Big Data no es novedad. Como ejemplo, una empresa de extracción de petróleo analiza terabytes de información geológica y las bolsas de valores procesan millones de transacciones por minuto. No obstante, dichos autores explican dos tendencias importantes que hacen de esta era de Big Data algo completamente diferente.

La primera tendencia es la digitalización de prácticamente “todo” da lugar a nuevos tipos de grandes datos en tiempo real en un amplio abanico de sectores. Muchos de ellos son datos no normalizados: por ejemplo, datos en streaming, geoespaciales o generados por sensores que no encajan bien en los warehouses relacionales, tradicionales y estructurados. Luego, la segunda tendencia son las tecnologías y técnicas de análisis avanzado de hoy en día que permiten a las empresas extraer conocimientos de los datos con un nivel de sofisticación, velocidad y precisión que no han sido vistos anteriormente (Schroeck et al., 2012).

Por lo cual, esto hace que la información, en la actualidad, sea considerada como uno de los principales recursos que poseen las compañías. La misma es tan importante para el éxito empresarial como las funciones de contabilidad, finanzas, administración de operaciones, marketing y administración de recursos humanos (Schroeck et al, 2012). Los autores McAfee et al. (2012), agregan que aquellas empresas que basan sus decisiones en los datos – proclamadas *data-driven* – logran alcanzar sus objetivos tanto financieros como operacionales. Schroeck et al. (2012) explican que esto se logra debido a que las empresas que utilizan Big Data identifican eficazmente los requisitos del negocio y, a continuación, adaptan la infraestructura, las fuentes de datos y la analítica a fin de respaldar la oportunidad de negocio. En particular, el Big Data permite el procesamiento de grandes **volúmenes** a alta **velocidad** con una amplia **variedad** de datos, permitiéndole a los gerentes la toma de decisiones basada en la evidencia en lugar de su intuición (McAfee et al., 2012).

Dicha expansión en los datos digitales tiene un impacto directo en aquellas organizaciones que operan en entornos comerciales inciertos. Debido a que, sin una visión adecuada de cómo se vería su perspectiva a corto y largo plazo, estas organizaciones podrían tomar decisiones inadecuadas sobre la gestión del negocio (Bodea & Ferguson, 2014). De acuerdo con la capacidad de análisis que poseen las organizaciones, se puede identificar tres tipos de empresas: *aspiracionales*, *experimentadas* y *transformadas*. El tipo de capacidad que posea la compañía



para analizar los datos, determinan la eficiencia en la implementación y gestión de los datos como fuente de información (LaValle et al., 2011).

Las empresas aspiracionales son aquellas que no cuentan con procesos, herramientas o personas capacitadas en recolectar, entender, incorporar y accionar en las ideas analizadas. Luego, las empresas experimentadas cuentan con experiencia en el análisis de datos, y esto les permite concentrar sus esfuerzos en desarrollar mejores maneras de recolectar e incorporar información, para luego actuar de manera eficiente en su análisis. Por último, las empresas transformadas, a diferencia de las empresas aspiracionales, cuentan con procesos, herramientas o personas capacitadas y eficientes y a su vez, no destinan todos sus esfuerzos a reducir costos, dado que, posiblemente sus procesos ya se encuentren automatizados (LaValle et al., 2011).

La unidad de análisis de la presente investigación se trata de una organización del tipo aspiracional, debido a que, no puede gestionar los grandes volúmenes de datos provenientes de su operación diaria y esto impacta directamente en la toma de decisiones a nivel estratégico de la empresa. En efecto, existe un gran desafío en pos de lograr que las decisiones se tomen, no sólo, en base a la intuición de quienes la gestionan, sino también que sea complementada con la evidencia proveniente de los datos. A su vez, el análisis y gestión de los datos no es sólo un desafío sino también una gran oportunidad para la compañía, debido a su impacto directo en las decisiones de carácter productivas, económicas y competitivas.

1.3 La gestión de la demanda en la industria de viajes y su impacto en los negocios

La gestión de la demanda- o *revenue management*- hace referencia a la colección de estrategias y tácticas que usan las empresas para gestionar científicamente la demanda de sus productos y servicios (Talluri & Van Ryzin, 2006). La práctica de la gestión de la demanda comenzó en la industria de las aerolíneas estrictamente como precios diferenciados, y luego se expandió a otras industrias, inicialmente en los campos de viajes y turismo, con características económicas similares a las de las aerolíneas. Finalmente, tanto las técnicas empleadas como la gama de industrias se extendieron, hasta el punto donde en la actualidad es aplicada en una amplia variedad de organizaciones (Huefner, 2015).

En los primeros tiempos, era posible identificar las industrias u organizaciones que tenían la capacidad de aplicar las técnicas de gestión de ingresos a través de las siguientes características. En primer lugar, aquellas industrias que son incapaces de cambiar fácilmente la



capacidad total en respuesta a variaciones en la demanda. Es el caso de las aerolíneas que tienen un determinado número de aviones y vuelos programados, por lo que los vuelos con asientos vacíos no pueden ser utilizados en un futuro. Esta característica generalmente se aplica a las industrias de servicios; las industrias de productos generalmente tienen la oportunidad de inventariar productos no vendidos para su venta futura (Huefner, 2015).

En segundo lugar, la industria debe caracterizarse por tener costos fijos altos y costos variables bajos. Esto es decir que los costos incrementales al vender una unidad más son mínimos, por ejemplo, en un vuelo de una línea aérea, los costos del avión, la tripulación y los servicios de apoyo no son afectados al agregar otro pasajero. En consecuencia, al tomar una decisión en relación con los ingresos, no tiene un impacto directo en los costos (Huefner, 2015).

En tercer lugar, se caracteriza por patrones de demanda inciertos o que varían de acuerdo con el tiempo. El autor cita el ejemplo de una aerolínea que puede encontrar que sus vuelos de madrugada y tarde están bastante llenos de viajeros que viajan por negocios en el mismo día, pero su mediodía y quizás sus vuelos de fin de semana están subutilizados. Por lo cual, las técnicas de gestión de ingresos son muy efectivas para atraer clientes en los tiempos subutilizados. Dado que, es posible asumir que algunos clientes son sensibles al precio y pueden ser inducidos a cambiar sus tiempos si el precio es atractivo. Por último, con el fin de atraer clientes en tiempos de servicio subutilizados, es importante que la empresa tenga la capacidad de pronosticar la demanda, para identificar los momentos en que probablemente exista un exceso de capacidad de servicio (Huefner, 2015).

A pesar de que estas cuatro condiciones eran comunes en las primeras aplicaciones de la gestión de ingresos, no son restrictivas. Estas condiciones no son inherentemente necesarias, pero sí hacen el análisis de la aplicación más fácil. A medida que la administración de ingresos ha crecido en su aplicabilidad, el concepto de gestión de ingresos se extendió a otras industrias con características menos restrictivas. La gestión de los ingresos ahora tiene aplicaciones en un amplio espectro de industrias: en conciertos, en hospitales, en telecomunicaciones, en los canales venta al por menor -o retail-, entre otros (Huefner, 2015).

A su vez, los autores Talluri & Van Ryzin (2006) resumen en que no es necesario tener condiciones similares a las de las aerolíneas para aplicar la gestión de la demanda en las organizaciones. Sin embargo, exponen las siguientes condiciones que favorecen la aplicación de las técnicas de la gestión de la demanda. Primero, la **heterogeneidad del cliente**, es decir, a mayor heterogeneidad entre los clientes, existe un mayor potencial para aplicar técnicas de la



gestión de la demanda de manera estratégica y táctica para mejorar los ingresos. Debido a que, si todos los clientes percibieran el producto de manera idéntica, existe un menor potencial para explotar variaciones en la disposición a pagar.

Segundo, la **variabilidad de la demanda e incertidumbre**, es decir, cuanto más varía la demanda con el tiempo (debido a las estacionalidades, los choques, etc.) y cuanta más incertidumbre se tenga sobre la demanda futura, más difíciles son las decisiones de gestión de la demanda. Por lo tanto, aumenta el potencial para tomar malas decisiones, y se vuelve importante contar con herramientas sofisticadas para evaluar el resultado compensaciones complejas (Talluri & Van Ryzin, 2006).

Tercero, la **inflexibilidad de la producción**. La misma conduce a una mayor interacción en la gestión de la demanda en diferentes momentos, entre diferentes segmentos de clientes, a través de diferentes productos de una línea de productos, y a través de diferentes canales de distribución. La complejidad aumenta y las consecuencias de las malas decisiones se vuelven más agudas (Talluri & Van Ryzin, 2006).

Cuarto, la **infraestructura de datos y sistemas de información** para aplicar la gestión de la demanda en las organizaciones se requiere de datos para caracterizar con precisión y modelar la demanda. También requiere de sistemas para recopilar y almacenar los datos y para implementar y monitorear las decisiones resultantes en tiempo real. Por último, **la cultura de la organización**: la cultura de la empresa puede no ser receptiva a la innovación o puede valorar enfoques más intuitivos para la resolución de problemas (Talluri & Van Ryzin, 2006).

En la industria de viajes, las aerolíneas enfrentan una serie de desafíos a medida que las tarifas continúan siendo sujeto a la presión de la baja de los precios, lo que acelera la necesidad de desarrollar estrategias minoristas efectivas para asegurar el crecimiento y la rentabilidad (Amadeus, 2018). En efecto, las prácticas de gestión de ingresos y análisis de precios son muy útiles para el contexto actual en el que se encuentra la industria. Debido a que las prácticas de gestión de ingresos se basan en los datos históricos de ventas para estimar analíticamente los pronósticos de demanda que luego son utilizados en los modelos de optimización para establecer y actualizar la capacidad o los precios a clientes específicamente segmentados para maximizar el beneficio (Bodea & Ferguson, 2014).

Con el objetivo de enfrentar este desafío, la tecnología se adhiere a los estándares de la industria al tiempo que les permite a las aerolíneas a diferenciarse a través de la predicción de las necesidades del viajero y así, adaptar las ofertas de las líneas aéreas (Amadeus, 2018). Los



autores Talluri & Van Ryzin (2006) exponen que la verdadera innovación no son las decisiones basadas en la gestión de la demanda, sino que es el método utilizado para tomar dichas decisiones. Este método es definido por los autores como un enfoque tecnológicamente sofisticado, detallado e intensamente operativo para tomar decisiones de gestión de la demanda.

Este método tecnológico utilizado para tomar decisiones sobre la gestión de la demanda es impulsado por dos fuerzas complementarias. Primero, los avances científicos en la investigación económica, estadística y operativa permiten modelar la demanda y las condiciones económicas, cuantificar las incertidumbres que enfrentan quienes toman las decisiones, predecir la respuesta del mercado y a su vez, calcular las soluciones óptimas para problemas complejos. Segundo, los avances en la tecnología de la información proporcionan la capacidad de automatizar transacciones, capturar y almacenar grandes cantidades de datos, ejecutar rápidamente algoritmos complejos y luego, implementar y gestionar decisiones altamente detalladas de gestión de la demanda (Talluri & Van Ryzin, 2006).

En otras palabras, dichos autores explican que ahora es posible gestionar la demanda en una escala y complejidad que sería impensable a través de medios manuales. Con el objetivo de dimensionar la cantidad de datos, se expone el ejemplo de una aerolínea que tiene miles de vuelos al día y brinda servicio entre cientos de miles de pares de origen-destino, cada uno de los cuales se vende a docenas de precios; por lo que sería muy complejo gestionar su demanda de manera manual.

A su vez, en la industria de viajes, dicho método en la gestión de la demanda impacta de manera positiva en la calidad de las decisiones. Dado que, el enfoque avanzado de ofertas personalizadas apoyado en las decisiones basadas en los datos les permite a los proveedores tomar conocimiento sobre sus clientes y así, detectar de manera proactiva las oportunidades o riesgos operativos (Amadeus, 2018).

Los autores Talluri & Van Ryzin (2006) agregan que la interacción hombre-máquina ofrece a la empresa lo mejor de la toma de decisiones tanto humana como automatizada. Debido a que, los encargados de tomar las decisiones no sólo se apoyan en su experiencia, sino que también lo hacen en los resultados que proveen los modelos y sistemas que manejan las decisiones rutinarias de gestión de la demanda de forma automatizada. De acuerdo con los autores McAfee et al. (2012) las tomas de decisiones basadas en los datos son más efectivas que aquellas basadas en la intuición. Dado que el Big Data permite gestionar, analizar y predecir y así, alcanzar mejores intervenciones que aquellas alcanzadas con analytics.



Universidad de Buenos Aires
Facultad de Ciencias Económicas
Escuela de Estudios de Posgrado



En efecto, se ha comprobado que el impacto económico de la gestión de la demanda es significativo, habiéndose demostrado aumentos en los ingresos del 5% o más reportados en las industrias que lo aplicaron (Talluri & Van Ryzin, 2006). Debido a que las técnicas de la gestión de la demanda proporcionan una guía para tomar decisiones eficientes con respecto a los ingresos que le permite a la organización planear sus estrategias por adelantado (Huefner, 2015).



Capítulo 2: Metodología y datos utilizados para el desarrollo del modelo de predicción de la demanda de pasajes aéreos

El presente trabajo utiliza como guía el proceso Knowledge Discovery in Databases (KDD) para su elaboración. El mismo es un proceso de análisis de bases de datos que busca encontrar relaciones inesperadas que son de interés o valor para el poseedor de dicha base de datos. Este proceso iterativo consiste en cinco etapas: *Integración o Selección, Preprocesamiento, Transformación, Minería de datos e Interpretación y evaluación* (Hand, 1998).

En este capítulo, se presentan las etapas de la selección, preprocesamiento de los datos y minería de datos. En la etapa de **Integración o Selección**, se seleccionan las variables y las fuentes a considerar en el proceso completo, por lo que se refiere a la creación del conjunto de datos como la base de datos de estudio en el proceso. Luego, en la segunda etapa llamada **Preprocesamiento**, se hace foco en el análisis y la limpieza de los datos, es decir, se produce el tratamiento de los valores tanto ausentes (missing) como de aquellos que se encuentran fuera de rango (outliers). La tercera etapa es la **Transformación**. En la misma, se generan nuevas variables a partir del estudio de la naturaleza de las variables originales (Hand, 1998).

La cuarta etapa es La **Minería de datos**. La misma consiste en la aplicación de análisis de datos para descubrir un algoritmo que produzca una particular enumeración de patrones a partir de los datos. Por ende, se selecciona el modelo y algoritmo a utilizar, bajo los supuestos que mantienen los objetivos primarios del estudio (Hand, 1998).

2.1 Selección de los datos

El presente análisis se realiza en base a los datos de la venta de pasajes aéreos de la agencia de viajes “Vuela Vuela”. Con el objetivo de diseñar y generar un set de datos que pueda tener un impacto significativo en dicha empresa, se han utilizado los siguientes conjuntos de datos:

- Datos de búsqueda: incluye la cantidad de búsquedas diarias realizadas por destino en la página web.
- Base de datos de transacciones: contiene la cantidad de transacciones vendidas por día con el detalle de la fecha de compra y el destino y la fecha de viaje.
- Base de datos de la competitividad de la agencia: son los datos históricos de los precios de la agencia “Vuela Vuela” y de sus dos (2) competidores directos.

En base a estos datos, se seleccionaron diez (10) atributos que se consideraron relevantes para la realización del modelo. En primer lugar, se encuentra el campo de “Fecha de búsqueda”,



el cual contiene las fechas diarias en las que se realizó la búsqueda. Luego, el campo de “Salida de viaje” hace referencia al inicio del viaje y el mismo dura una semana en todos los casos. En tercer lugar, la base tiene un campo llamado Destino en donde se encuentran “Londres”, “Madrid”, “Barcelona”, “Paris” y “Roma”, los cuales hacen referencia al análisis perteneciente a la región de Europa. Mientras que los destinos de “Cancún”, “La Habana” y “Punta Cana” se refieren a la región de Caribe.

En relación con los precios, existen tres campos que muestran el precio de venta de distintas agencias de viajes. Asimismo, existen dos campos que se denominan “Variación” y muestran la variación porcentual en los precios entre cada agencia en la misma fecha y destino de búsqueda. Luego, el campo denominado “Búsquedas únicas” expone la cantidad de búsquedas en el Sitio Web al destino analizado. Por último, se encuentra el campo de “Transacciones” que muestra la cantidad de transacciones vendidas por día a cada destino. Este atributo es el que se busca predecir. En efecto, los expuestos anteriormente son las variables que podrían influir en la venta de pasajes aéreos. Estos se encuentran resumidos en la Tabla I.

Tabla I: Resumen de los atributos de la base datos

Atributo	Tipo	Descripción
Fecha de búsqueda	Fecha Corta	Fecha de la búsqueda
Salida de Viaje	Fecha Corta	Inicio del viaje
Destino	Polinomial	Destino al que se quiere viajar
Precio de venta de Agencia “Vuela Vuela”	Entero	El precio de la agencia al destino y en el periodo de viaje buscado de la Agencia “Vuela Vuela”
Precio de venta de Agencia 2	Entero	El precio de la agencia al destino y en el periodo de viaje buscado de la competencia de la Agencia “Vuela Vuela”
Precio de venta de Agencia 3	Entero	El precio de la agencia al destino y en el periodo de viaje buscado de la competencia de la Agencia “Vuela Vuela”
Diferencia Agencia “Vuela Vuela” versus Agencia 2	Real	Variación entre los precios de la Agencia “Vuela Vuela” versus Agencia 2
Diferencia Agencia “Vuela Vuela” versus Agencia 3	Real	Variación entre los precios de la Agencia “Vuela Vuela” versus Agencia 3
Búsquedas únicas	Entero	Cantidad de búsquedas diarias por destino
Transacciones	Entero	Cantidad de pasajes aéreos vendidos por día y por destino en la fecha de búsqueda

Fuente: Elaboración propia en base a los datos de Destinos de Europa y Caribe.

2.2 Preprocesamiento de los datos

Luego de seleccionar las variables y las fuentes a considerar para crear la base de datos de estudio en el proceso, se procedió a analizar los datos con el fin de conocer en mayor



profundidad la base con la que se está trabajando. Por ello, se ha realizado un análisis de estadística descriptiva a los precios de cada agencia, las diferencias porcentuales entre los precios, las búsquedas únicas y la cantidad de transacciones. El análisis para ambas regiones se encuentra expuesto en las Tablas II y III.

Tabla II: Estadística descriptiva de los destinos de Europa

EUROPA	Precio Agencia "Vuela Vuela"	Precio Agencia 2	Precio Agencia 3	Diferencia Agencia "Vuela Vuela" vs Agencia 2	Diferencia Agencia "Vuela Vuela" vs Agencia 3	Búsquedas únicas Agencia "Vuela Vuela"	Transacciones Agencia "Vuela Vuela"
Máximo	\$ 98.313,00	\$ 72.364,00	\$ 81.752,00	94,47%	138,09%	10438,00	110,00
Mínimo	\$ 13.573,00	\$ 13.599,00	\$ 13.533,00	-55,63%	-42,11%	12,00	0,00
Promedio	\$ 32.518,51	\$ 30.264,20	\$ 30.706,46	7,43%	6,04%	318,28	1,35
Desvío Estándar	\$ 8.965,49	\$ 7.741,18	\$ 8.108,43	11,27%	10,92%	526,47	5,29
Moda	\$ 30.848,00	\$ 29.777,00	\$ 28.706,00	0,00%	0,00%	52,00	0,00
Mediana (Percentil 50)	\$ 31.674,00	\$ 29.624,00	\$ 29.903,00	4,74%	3,98%	197,00	0,00
Coefficiente de Asimetría	\$ 0,78	\$ 0,75	\$ 0,88	141,00%	154,77%	11,60	14,67
Coefficiente de Variación	\$ 0,28	\$ 0,26	\$ 0,26	151,74%	180,75%	1,65	3,91

Fuente: Elaboración propia en base a los datos de Destinos de Europa.

Tabla III: Estadística descriptiva de los destinos del Caribe

CARIBE	Precio Agencia "Vuela Vuela"	Precio Agencia 2	Precio Agencia 3	Diferencia Agencia "Vuela Vuela" vs Agencia 2	Diferencia Agencia "Vuela Vuela" vs Agencia 3	Búsquedas únicas Agencia "Vuela Vuela"	Transacciones Agencia "Vuela Vuela"
Máximo	\$ 104.622,00	\$ 101.351,00	\$ 97.897,00	125,18%	136,74%	5337,00	60,00
Mínimo	\$ 13.343,00	\$ 13.484,00	\$ 12.954,00	-58,29%	-53,94%	22,00	0,00
Promedio	\$ 22.834,08	\$ 22.638,78	\$ 22.776,59	0,98%	0,39%	303,12	1,40
Desvío Estándar	\$ 5.453,35	\$ 5.250,39	\$ 5.267,05	6,92%	7,53%	413,11	4,20
Moda	\$ 19.368,00	\$ 18.858,00	\$ 17.448,00	0,00%	0,00%	120,00	0,00
Mediana (Percentil 50)	\$ 21.499,84	\$ 21.136,00	\$ 21.393,00	0,75%	0,32%	183,00	0,00
Coefficiente de Asimetría	\$ 1,68	\$ 1,60	\$ 1,66	97,45%	129,26%	6,50	8,76
Coefficiente de Variación	\$ 0,24	\$ 0,23	\$ 0,23	708,88%	1948,62%	1,36	3,00

Fuente: Elaboración propia en base a los datos de Destinos del Caribe.

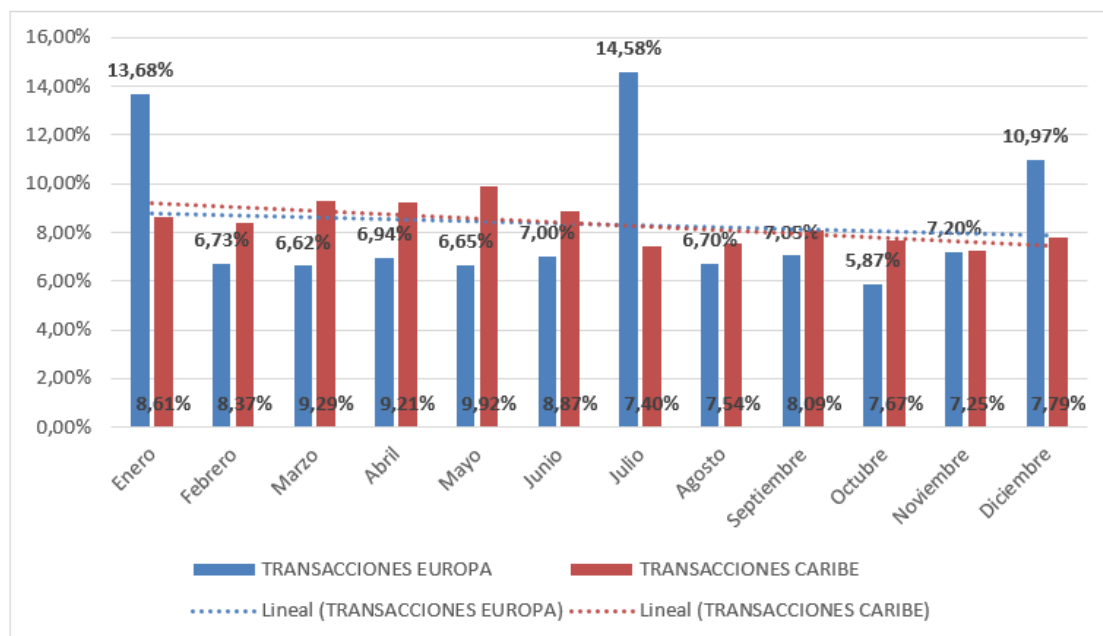


En efecto, es posible observar que las transacciones promedio de Europa son 1,35 por día mientras que en los destinos del Caribe es 1,4 transacciones diarias. Luego, los precios promedio de la Agencia “Vuela Vuela” para ambas regiones son más caros versus la competencia. En efecto, en el caso de Europa, la diferencia promedio versus la Agencia 2 es de 7,43% mientras que versus la Agencia 3 es 6,04%. En el caso de la región Caribe, sucede que la diferencia promedio versus la Agencia 2 es de 0,98% y versus la Agencia 3 es 0,39%. Por lo cual, existe una menor diferencia en los precios en esta última región que en la región de Europa versus la competencia.

A través del coeficiente del desvío estándar, se deduce que la presencia de valores extremos incrementa el rango de variabilidad en cada serie (valores mínimos y máximos muy dispares). Por tal razón, se procedió a realizar un tratamiento de outliers con el fin de balancear la base.

Por último, en el Gráfico I se muestran los pasajes vendidos a cada destino distribuidos de acuerdo con el mes en el que el pasajero desea volar. De esta manera, es posible detectar si existe una preferencia en el mes en el que se desea viajar a cada una de las regiones estudiadas.

Gráfico I: Distribución de la venta de acuerdo con la fecha de viaje



Fuente: Elaboración propia en base a los datos de Destinos de Europa y Caribe.

En efecto, en el gráfico I se demuestra que, en el caso de Europa, el 30% de las transacciones vendidas, realizan su viaje en enero y julio y el tercer mes en el que los clientes prefieren viajar



a dicho destino es en diciembre. Por su parte, en el caso del Caribe, no presenta picos de venta abruptos en un mes en particular, es decir, que no se detecta un periodo de viaje en el que los clientes prefieren viajar a esta región.

2.3 Técnicas de minería de datos y métricas de evaluación

En la etapa de minería de datos, se exponen los algoritmos de clasificación con los que el modelo será entrenado para la selección de la técnica más adecuada: *Decision Tree*, *KNN*, *Random Forest* y *Gradient Boosting Trees*. Los mismos son explicados a continuación.

Primero, **Decision Tree** es un algoritmo de clasificación que se utiliza para separar un conjunto de datos en clases que pertenecen a la variable de respuesta. Por lo general, la variable de respuesta tiene dos clases: Sí o No (1 o 0). Si la variable de respuesta tiene más de dos categorías, entonces se han desarrollado variantes del algoritmo del árbol de decisión que pueden aplicarse. En cualquier caso, los árboles de decisión se usan cuando la respuesta o la variable objetivo es de naturaleza categórica (Quinlan, 1986).

Segundo, **KNN** es un método no paramétrico, no asume distribuciones subyacentes en los datos, sólo asume que inputs similares tendrán outputs similares. La k en el algoritmo k -NN indica el número de registros de entrenamiento cercanos que deben considerarse al hacer la predicción para un registro de prueba sin etiqueta. Como la clase del registro objetivo se evalúa mediante votación, a k generalmente se le asigna un número impar para un problema de dos clases (Kotu & Deshpande, 2014). Los autores destacan que el algoritmo maneja de manera natural los problemas multiclase.

Tercero, **Random Forest** es un concepto que fue presentado por primera vez por Leo Breiman y Adele Cutler. El mismo es un tipo de ensamble que utiliza una técnica similar al Bagging. Al decidir dividir cada nodo en un árbol de decisión, a diferencia de Bagging, Random Forest solo considera un subconjunto aleatorio de todos los atributos en el conjunto de entrenamiento. Para reducir el error de generalización, el algoritmo se aleatoriza en dos niveles, selección de registros de entrenamiento y selección de atributos, en el funcionamiento interno de cada clasificador base (Kotu & Deshpande, 2014).

Cuarto, **Gradient Boosting Trees** es un tipo de algoritmo que produce diferentes modelos individuales (árboles de decisión, por ejemplo) cuyos resultados se van agregando de modo que el resultado final *-clasificador de ensamble-* está formado por un modelo que es una combinación de los anteriores *-clasificadores débiles-*, pero con una capacidad de predicción



muy superior a la de los modelos individuales en los que se basa (Chambers y Dinsmore, 2015). En las sucesivas iteraciones, Gradient Boosting Trees aprende y minimiza los errores de los modelos anteriores y ajusta los árboles de decisión a los residuos o errores con el fin de ir actualizando y minimizando los mismos. Este algoritmo, además de aprender de los errores de los múltiples modelos a medida que los va generando (como lo hacen los otros algoritmos basados en boosting), Gradient Boosting Trees permite la existencia de valores extremos, correlaciones altas entre las variables, relaciones no lineales, la presencia de valores perdidos y admite el uso de variables categóricas como independientes (Campillo et al., 2018)

Luego, cada uno de dichos modelos, se le evaluará la calidad de sus resultados utilizando el criterio la métrica del error cuadrático medio (RMSE). El mismo mide la cantidad de error que hay entre dos conjuntos de datos. En otras palabras, compara un valor predicho y un valor observado o conocido (Kotu & Deshpande, 2014).



Capítulo 3: Implementación del modelo predictivo de la demanda de pasajes aéreos

En el presente capítulo, se presentan la última etapa - *Interpretación y evaluación*- del proceso Knowledge Discovery in Databases (KDD).

En la etapa de **Interpretación y Evaluación**, se involucran las medidas de evaluación y la trasposición de resultados técnicos a niveles comerciales, de tal manera, que la aplicación del procedimiento converja a acciones correctivas en el negocio, que solucionen el fenómeno estudiado (Hand, 1998).

3.1 Aplicación del modelo a través de técnicas de aprendizaje automático

En el presente trabajo, las técnicas de minería de datos tienen como objetivo predecir la cantidad de transacciones vendidas por región -*Europa y Caribe*-. De acuerdo con los autores Kotu & Deshpande (2014), la variable que el modelo busca predecir se denomina variable dependiente y la misma se predice en función de un conjunto de variables predictoras (variables independientes). Las variables independientes en el presente trabajo son: “Salida de viaje”, “Destino”, “Precio de venta de la Agencia “Vuela Vuela”, “Precio de venta la Agencia 2”, “Precio de venta la Agencia 3”, “Diferencia agencia “Vuela Vuela” vs Agencia 2”, “Diferencia agencia “Vuela Vuela” vs Agencia 3” y “Búsquedas únicas”. Cabe aclarar que el modelo no se entrena sobre el atributo “Fecha de búsqueda”, dado que, fue clasificado como “ID”.

Luego, el set de datos descrito en el apartado 2.1 es ingresado en el programa RapidMiner Studio con el objetivo de poner a prueba los algoritmos de clasificación utilizados en el presente trabajo. Todos los algoritmos reciben como entrada la base de datos con la que aprenderán y validarán el modelo generado. Cabe aclarar que, a cada algoritmo se le configuraron los parámetros correspondientes con el objetivo de lograr el mejor funcionamiento del modelo.

Una vez ingresado el set de datos, se incorpora el operador “Split Data”, con el cual con el cual se divide la base de datos en un 80% para entrenar el algoritmo y con el 20% restante se probará la calidad de los resultados. Luego, se probaron sobre una base de 83.332 búsquedas diarias desde agosto 2018 a septiembre 2019 los algoritmos de clasificación explicados en el apartado 3.1: *Decision Tree, KNN, Random Forest y Gradient Boosting Trees*. El hecho de probar distintos operadores permitió realizar una comparación entre las distintas predicciones obtenidas por cada uno de ellos y así, poder acercarse al mejor resultado. Por último, se evaluó

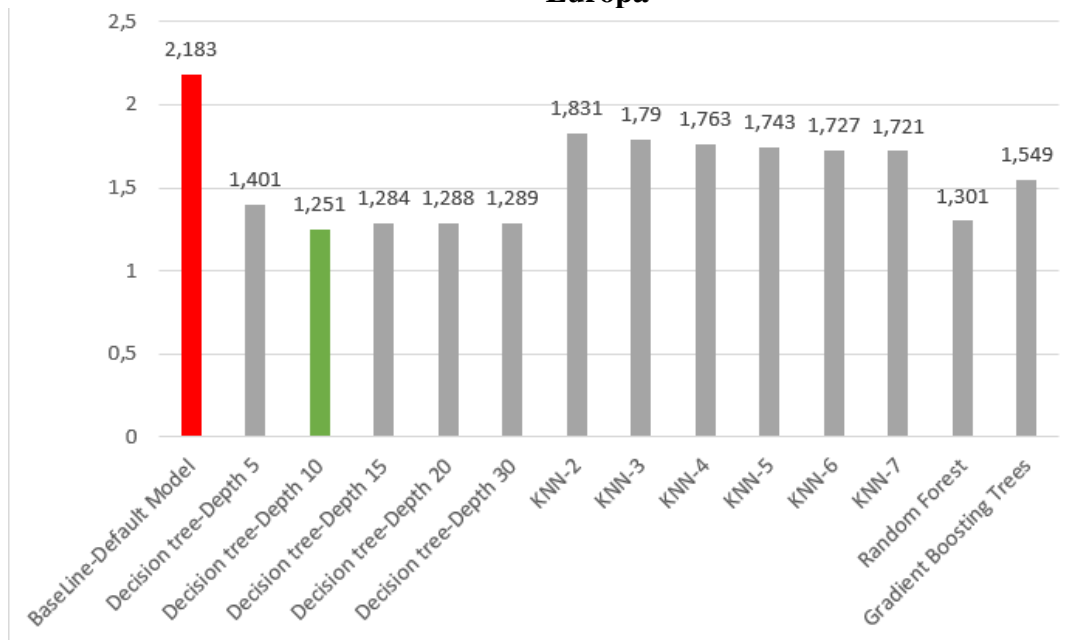


la calidad de dichos modelos a través de la métrica del error cuadrático medio (RMSE) para evaluar la calidad de dichos modelos aplicados para encontrar aquel que prediga con mayor exactitud la demanda de pasajes aéreos.

3.2 Interpretación y evaluación de resultados

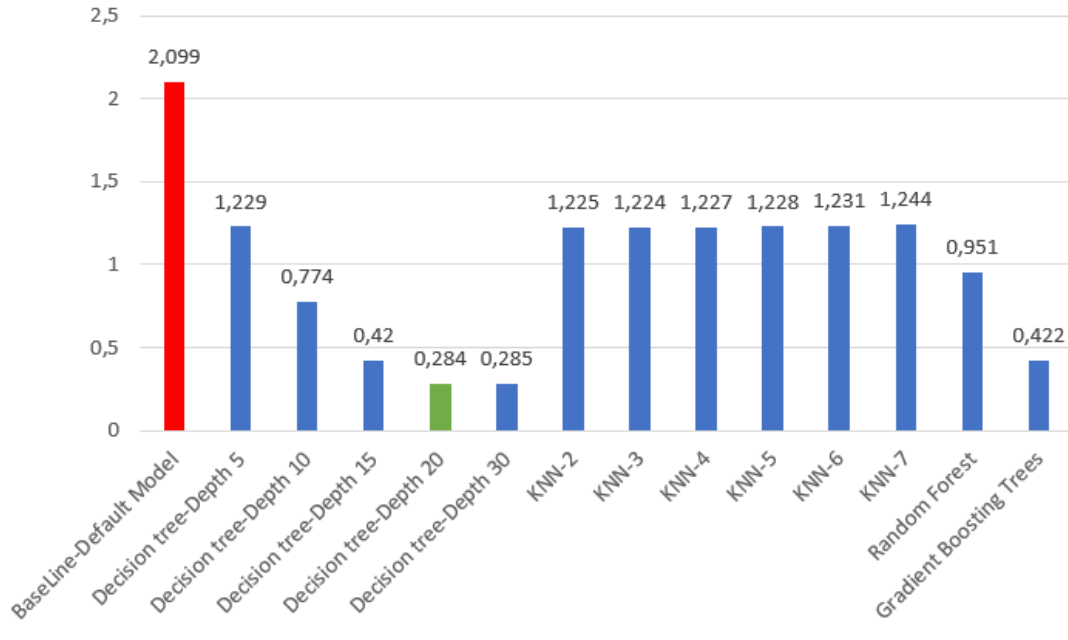
Los resultados de los distintos algoritmos entrenados para cada región fueron comparados entre sí con el fin de encontrar aquel que predecía con mayor precisión la demanda de pasajes aéreos. Dicha información se expone en los gráficos II y III.

Gráfico II: Performance de los distintos modelos de acuerdo con RMSE en Europa



Fuente: Elaboración propia en base a los resultados obtenidos en RapidMiner

Gráfico III: Performance de los distintos modelos de acuerdo con RMSE en el Caribe



Fuente: Elaboración propia en base a los resultados obtenidos en RapidMiner

En efecto, a partir de las pruebas realizadas con los distintos modelos, es posible inferir que el mejor algoritmo para predecir la demanda de pasajes aéreos es **Decision Tree**. En el caso de la región de Europa, Decision Tree con una profundidad de 10 (RMSE: 1,251) mejora la predicción de transacciones vendidas en 0,932 respecto al error cuadrático medio inicial (2,183). En el caso de la región del Caribe, Decision Tree con profundidad de 20 (RMSE: 0,284) mejora el modelo en 1,815 respecto al error cuadrático medio del punto de partida (2,099).

Por otro lado, con el fin de recomendar acciones de gestión de la demanda, es importante conocer las variables que tienen mayor influencia sobre la variación de la cantidad de transacciones vendidas. En efecto, esta información se expone en la Tabla IV.



Tabla IV: Peso de las variables por destino

VARIABLES	PESO EN EUROPA	PESO EN CARIBE
Salida de Viaje	28,90%	21,40%
Destino	3,40%	0,30%
Precio de venta de la Agencia "Vuela Vuela"	20,70%	20,10%
Precio de venta de Agencia 2	13,30%	5,70%
Precio de venta de Agencia 3	9,50%	4,20%
Diferencia Agencia la Agencia "Vuela Vuela" versus Agencia 2	8,60%	12,30%
Diferencia Agencia la Agencia "Vuela Vuela" versus Agencia 3	6,20%	5,10%
Búsquedas únicas	9,40%	30,90%

Fuente: Elaboración propia en base a los resultados obtenidos en RapidMiner

Por lo cual, las variables que influyen en la demanda de pasajes aéreos varían de acuerdo con la región que se esté estudiando. En el caso de Europa, el periodo de viaje es la variable que mayor influencia tiene sobre los pasajes aéreos vendidos (su peso es 28,9%). Por otro lado, en lo que respecta a la región del Caribe, la variable que mayor influencia tiene sobre las transacciones vendidas es la cantidad de búsquedas únicas en el Sitio Web pesando un 30,9%.



Conclusión

El objetivo del presente trabajo era responder a la siguiente pregunta: ¿Cuáles son los métodos de aprendizaje automático que predicen con mayor precisión la demanda de pasajes aéreos de la agencia de viajes “Vuela Vuela” a las regiones de Caribe y Europa? ¿Qué acciones comerciales se pueden recomendar para maximizar los ingresos de la agencia de viajes en dichas regiones utilizando los modelos encontrados?

En efecto, durante el trabajo realizado se buscó alcanzar una conclusión que responda a la pregunta planteada y a su vez, que le fuese de utilidad a la agencia de viajes para contar con las herramientas necesarias para tomar decisiones en el mercado competitivo y exigente en el cual se encuentra.

Para ello, se entrenó el modelo con cuatro (4) algoritmos de clasificación (Decision Tree, K-NN, Random Forest y Gradient Boosting Trees) sobre una base de 83.332 búsquedas; para evaluar la calidad de los modelos probados, se utilizó la métrica del error cuadrático medio (RMSE). Se ha comprobado **Decision Tree** es el modelo que mejor predice la demanda de pasajes aéreos de la agencia “Vuela Vuela”.

En el caso de los destinos del Caribe, el algoritmo que mejor resultados arrojó fue Decision Tree con una profundidad de 20, debido a que el error cuadrático medio fue de 0,284 transacciones. Asimismo, en el caso de los destinos de Europa, Decision Tree también fue el algoritmo que mostró mejores resultados, en este caso con una profundidad de 10, de debido a que el error cuadrático medio fue 1,251 transacciones.

Asimismo, en relación con las variables que influyen en la cantidad de transacciones vendidas, es posible afirmar que las mismas cambian de acuerdo con el destino. En el caso de Europa, la variable que más influye en la decisión de compra es el periodo de viaje, siendo los meses más importantes enero y julio. Esto es decir que, los usuarios eligen su pasaje aéreo por la fecha en la que desean volar, siendo esto más importante que otras variables como la cantidad de búsquedas únicas y la competitividad en precios de la agencia en donde están comprando el pasaje aéreo. Por otro lado, en el caso de la región del Caribe, la variable que mayor influencia tiene en las transacciones vendidas es la cantidad de búsquedas en la Página Web de la agencia de viajes.

Por lo cual, en el caso de la región de Europa, se le recomienda la agencia de viajes “Vuela Vuela”, realizar una estrategia de precios diferenciada por fecha de vuelo. Esto es decir que, en



la temporada alta de viaje- *enero y julio*- los precios sean más caros, dado que, la demanda es alta. Mientras que, se recomienda incentivar la venta los meses de viaje en donde la demanda es baja a través de acciones que el cliente perciba un descuento en el precio. Estas acciones pueden orientarse a vender el pasaje sin cargos extra por el servicio otorgado, otorgar cuotas sin interés u ofrecer cupones de descuento a dichos destinos.

Luego, en el caso de la región del Caribe, en donde la variable que mayor influencia tiene sobre la demanda de pasajes aéreos es la cantidad de búsquedas, se recomiendan acciones relacionadas con el Marketing online. En otras palabras, se le recomienda a la agencia “Vuela Vuela” que en sus Páginas Webs muestre destinos como Cancún, Punta Cana o La Habana, también que programe los Web Push y los mails enviados a su base de clientes con ofertas a dichos destinos. De esta manera, la agencia estaría incentivando la venta a estos destinos a través de persuadir al usuario a realizar búsquedas a estos destinos.

En efecto, la organización utilizada como caso de estudio, podrá conocer con exactitud su demanda y así, tener las herramientas necesarias para desarrollar sistemas de gestión de la demanda que permitan maximizar los ingresos de la compañía. De esta manera, la agencia de viajes “Vuela Vuela” basará sus decisiones no sólo en la experiencia de quien gestiona, sino también en la información que le aporte el análisis de los datos realizado.



Futuros estudios

El modelo propuesto abre oportunidades para crear futuras líneas de investigación. Primero, se recomienda incluir al set de datos variables provenientes de fuente de datos externas como, por ejemplo, variables económicas, estaciones del año, fechas en las que hay Campañas importantes de Marketing en el mercado online como Hot Sale o Cyber Monday, entre otras variables externas. De esta manera, se podría alcanzar un mayor entendimiento de las variables que afectan la demanda de pasajes aéreos.

Segundo, existe la posibilidad de utilizar el modelo en otros destinos y productos (hoteles, autos, cruceros, entre otros) diferentes a los analizados en el presente trabajo. De este modo, la agencia de viajes estudiada tendría un conocimiento más preciso sobre la demanda total de su organización.

Tercero, la implementación del modelo en la agencia de viajes es una oportunidad para generar valor a la compañía. A su vez, en cuarto lugar, existe la oportunidad de implementar informáticamente el modelo con datos en tiempo real. Ambas implementaciones le permitirían a la empresa mejorar la calidad de las decisiones, al basarse en los datos y a su vez, realizarlo en el momento adecuado.

Por último, al considerar realizar las mejoras propuestas anteriormente, el modelo podría ser comercializado en la industria de turismo como una herramienta que permite mejorar la toma de decisiones de las organizaciones, lo cual garantiza también una mejora en los ingresos.



Referencias bibliográficas

- ABAD, C. E. J. (2006). *“Producción y venta de servicios turísticos en agencias de viajes”*. Editorial Paraninfo.
- AMADEUS, IT. GROUP SA (2018). *“Informe Global: Análisis de la actividad empresarial, financiera y de sostenibilidad”*.
- BARRIENTOS, F., & RÍOS, S. A. (2013). *“Aplicación de Minería de Datos para Predecir Fuga de Clientes en la Industria de las Telecomunicaciones”*. Revista ingeniería de Sistemas, 27, 73-107.
- BEYER, M. (2011). *“Information Management in the 21st Century”*. New York: Gartner.
- BODEA, T., & FERGUSON, M. (2014). *“Segmentation, revenue management and pricing analytics”*. Editorial Routledge.
- BOWEN, R., & SMITH, A. R. (2014). *“Developing an enterprisewide data strategy”*. Healthcare Financial Management, 68(4), 86-90.
- BRIEMAN, L. F. (1984). *“Classification and Regression Trees”*. Chapman and Hall.
- CAMPILLO, J. P., VARGAS, J. M., & IBÁÑEZ, P. C. (2018). *“Análisis de la utilidad del algoritmo Gradient Boosting Machine (GBM) en la predicción del fracaso empresarial”*. Spanish Journal of Finance and Accounting/Revista Española de Financiación y Contabilidad, 47(4), 507-532.
- CHAMBERS, M. & DINSMORE T. W. (2015). *“Advanced Analytics Methodologies: Driving Business Value with Analytics”*. Pearson Education, New Jersey.
- DESAI, S.K. (2018). *“Data Analytics for the Insurance Industry: A Gold Mine”*. Critical Essays on Indian Writing in English. Dharwar: Karnatak University; Madras: distributors, Macmillan Company of India.
- FAYYAD, U., PIATETSKY-SHAPIRO, G., & SMYTH, P. (1996). *“From data mining to knowledge discovery in databases”*. AI magazine, 17(3), 37-37.
- HAND, D. J. (1998). *“Data mining: statistics and more?”*. The American Statistician, 52(2), 112-118.
- HILBERT, M., & LÓPEZ, P. (2011). *“The world’s technological capacity to store, communicate, and compute information. Science”*, 332(6025), 60-65.
- HUEFNER, R. (2015). *“Revenue Management: A path to increased profits”*. Editorial Business Expert Press.
- KOTU, V., & DESHPANDE, B. (2014). *“Predictive analytics and data mining: concepts and practice with rapidminer”*. Morgan Kaufmann.
- LAVALLE, LESSER, SHOCKLEY, HOPKINS Y KRUSCHWITZ (2011). *Big Data, Analytics and the Path from Insights to Value*. MitSloan Management Review.



- MÉRIDA, C., RÍOS, R., KOBAYASHI, A., & RAYMUNDO, C. (2017). “*Modelo arquitectónico de información para una plataforma de Big Data para el Sector Turístico*”. In Proceedings of the Decima Sexta Conferencia Iberoamericana en Sistemas, Cibernética e Informática (CISCI 2017), Décimo Cuarto Simposium Iberoamericano en Educación, Cibernética e Informática, SIECI 2017-Memorias (pp. 321-328).
- MCAFEE, A., BRYNJOLFSSON, E., DAVENPORT, T. H., PATIL, D. J., & BARTON, D. (2012). “*Big data: the management revolution*”. Harvard business review, 90(10), 60-68.
- PORTER, M.E. (2008). “*Estrategia Competitiva: Técnicas para el análisis de los sectores industriales y de la competencia*”, trigésima octava reimpresión. México D.F., México: Grupo Editorial Patria.
- QUINLAN, J. R. (1986). “*Induction of decision trees*”. Machine learning, 1(1), 81-106.
- RODRÍGUEZ-ZULAICA, A. (2017). “*¿Cómo gestionar una agencia de viajes en internet?*”. Editorial UOC.
- SCHROECK, M., SHOCKLEY, R., SMART, J., ROMERO, D., & TUFANO, P. (2012). “*Analytics: el uso de Big Data en el mundo real*”. IBM Institute for Business Value, Oxford, Informe ejecutivo.
- TALLURI, K. T., & VAN RYZIN, G. J. (2006). “*The theory and practice of revenue management (Vol. 68)*”. Editorial Springer Science & Business Media.