

## *Abstract*

Technological advances in recent years have enabled the emergence of new data sources and with it, the storage of large amounts of data or 'Big Data' has become increasingly important. More and more scientific studies are using these Big Data sources to try to improve understanding in various scientific fields. In tourism economics, many of these sources have already been used to predict the behavior of real variables. In tourism, the usefulness of these new data sources lies in the fact that they can help to understand the behavior of tourists, from their spatial-temporal patterns to which attractions and activities are the most popular in the destination, and therefore, they can help in the decision making of economic agents.

Therefore, this thesis tries to better understand which Big Data sources are the most useful when dealing with tourism variables and also to propose methodological improvements so that these sources can be applied to the field of rural tourism, and more specifically, to the prediction of tourists.

In this thesis several advances in this aspect are presented: First, a classification of data sources that every tourist generates during his tourist process and that compose his digital footprint. Then, with respect to this classification, Google Trends is chosen as the most appropriate source to help predict tourist demand, but accuracy problems are found, which are demonstrated and exemplified. Further on, it is demonstrated how this accuracy error is generated through the GT sampling process and solutions are proposed to alleviate this error, namely by obtaining more extractions and using their mean. Finally, this method is tested for the prediction of monthly overnight stays in rural tourism accommodations in Spain.

In summary, the contribution that this thesis aims to make is to provide a better understanding of Big Data sources and help to generate good practices in the use of them so that they can be applied to the prediction of real variables in rural tourism, in a way that streamlines and improves the decision making of economic agents.

## *Resumen*

Los avances tecnológicos de los últimos años han permitido la aparición de nuevas fuentes de datos y con ello, el almacenamiento de grandes cantidades de datos o 'Big Data' se ha cobrado cada vez mayor importancia. Cada vez más y más estudios científicos utilizan estas fuentes de 'Big Data' para tratar de mejorar el entendimiento en diversos campos científicos. En la economía del turismo ya se han utilizado muchas de estas fuentes para predecir el comportamiento de variables reales. En turismo, la utilidad de estas nuevas fuentes de datos reside en que pueden ayudar a entender el comportamiento de los turistas, desde sus patrones espaciotemporales hasta qué atracciones y actividades son las más populares en el destino, y por tanto, pueden ayudar en la toma de decisiones de los agentes económicos.

Por tanto, esta tesis intenta entender mejor cuáles son las fuentes de Big Data que resultan más útiles a la hora de lidiar con variables turísticas y además proponer mejoras metodológicas para que dichas fuentes se puedan aplicar al campo del turismo rural, y más concretamente, a la predicción de turistas.

En esta tesis se presentan varios avances en este aspecto: Primero, una clasificación de fuentes de datos que genera todo turista durante su proceso turístico y que componen su huella digital. Después, respecto a esta clasificación, se escoge Google Trends como la fuente más adecuada para ayudar a predecir la demanda turística, pero se encuentran problemas de precisión, que son demostrados y ejemplificados. Más adelante, se demuestra cómo se genera este error de precisión a través del proceso de muestreo de GT y se proponen soluciones para aliviar este error, a saber, obteniendo más extracciones y utilizando su media. Finalmente, este método se pone a prueba para la predicción de pernoctaciones mensuales en alojamientos de turismo rural en España.

En resumen, la contribución que esta tesis pretende hacer, es aportar una mayor comprensión de las fuentes de Big Data y ayudar a generar buenas prácticas en el uso de las mismas para que se puedan aplicar a la predicción de variables reales en el turismo rural, de forma que agilice y mejore la toma de decisiones de los agentes económicos.

## *Resum*

Els avanços tecnològics dels últims anys han permès l'aparició de noves fonts de dades i amb això, l'emmagatzematge de grans quantitats de dades o 'Big Data' s'ha cobrat cada vegada major importància. Cada vegada més i més estudis científics utilitzen aquestes fonts de 'Big Data' per a tractar de millorar l'enteniment en diversos camps científics. En l'economia del turisme ja s'han utilitzat moltes d'aquestes fonts per a predir el comportament de variables reals. En turisme, la utilitat d'aquestes noves fonts de dades resideix en què poden ajudar a entendre el comportament dels turistes, des dels seus patrons espaciotemporals fins a quines atraccions i activitats són les més populars en el destí, i per tant, poden ajudar en la presa de decisions dels agents econòmics.

Per tant, aquesta tesi intenta entendre millor quines són les fonts de Big Data que resulten més útils a l'hora de bregar amb variables turístiques i a més proposar millores metodològiques perquè aquestes fonts es puguin aplicar al camp del turisme rural, i més concretament, a la predicció de turistes.

En aquesta tesi es presenten diversos avanços en aquest aspecte: Primer, una classificació de fonts de dades que genera tot turista durant el seu procés turístic i que componen la seua empremta digital. Després, respecte a aquesta classificació, es tria Google Trends com la font més adequada per a ajudar a predir la demanda turística, però es troben problemes de precisió, que són demostrats i exemplificats. Més endavant, es demostra com es genera aquest error de precisió a través del procés de mostreig de GT i es proposen solucions per a alleujar aquest error, a saber, obtenint més extraccions i utilitzant la seua mitjana. Finalment, aquest mètode es posa a prova per a la predicció de pernотacions mensuals en allotjaments de turisme rural a Espanya.

En resum, la contribució que aquesta tesi pretén fer, és aportar una major comprensió de les fonts de Big Data i ajudar a generar bones pràctiques en l'ús de les mateixes perquè es puguin aplicar a la predicció de variables reals en el turisme rural, de manera que agilitze i millore la presa de decisions dels agents econòmics.