## ABSTRACT

The rapid growth of the Indian domestic tourism sector has opened up an immense opportunity for all its stakeholders. Unfortunately, the Indian government has not been able to fully capitalize on this, especially in terms of its marketing strategy. In fact, it's not only the government, but also different stakeholders, who are primarily responsible for implementing destination marketing. Till date, extant research on the tourism sector, has been focused on stakeholder involvement vis a vis their complex relationships in destination marketing. Thus, in this research, we look to fully utilize the concept of Stakeholder theory, in order to address some of the problems of market segmentation and classification, especially in the Indian context. The research approach followed includes both qualitative and quantitative methods. The variables employed have been identified from two sources, i.e. extant literature along with semi-structured interviews with stakeholders. Extensive literature review was done to identify a comprehensive list of unique variables. Furthermore, twelve semi-structured interviews were conducted with three sets of stakeholders (i.e. four experts from tour operators, five experts from accommodation service providers, and three experts from Government tourism officials). Herein, the aim was to understand their market segmentation process, based on which we develop the relevant market segmentation variables. A total of 346 unique variables were identified from literature review (i.e. 122 variables for accommodation service providers, 134 variables for tour operators, and 90 variables for tourism policy makers). These were shared with the stakeholders that were interviewed. Post our discussions and deliberations with them, finally a total of 103 variables (out of 346 initially identified) were chosen for the study. The breakup of these variables are as follows: 37 variables related to accommodation service providers, 34 variables associated with tour operators, and 32 variables linked to tourism policy makers. Further, based on these 103 variables, three sets of final questionnaires were prepared. Each questionnaire had two parts; the first was about the respondents' socio-demographic profile, while the second aimed at incorporating the respondents' opinions based on the significance of variables converted into statements. A 5-point Likert-type scale with an interval scale ranging from 1 (Not at all important) to 5 (Extremely important) was used. 1309 complete responses (i.e. 412 responses for accommodation service providers, 438 responses for tour operators, and 459 responses for tourism policy makers) were collected from Indian domestic tourists. Factor analysis method based on principal component analysis (PCA) with varimax rotation was done in order to explore the dimensions from the data collated from the

domestic tourists. Post the factor analysis, the factor score of each of the respondents was saved for further analysis. The next step of the study included segmentation, using K-means clustering techniques, applied to the heterogeneous population. These were divided into a homogeneous subgroup of similar behavior tourists. The results obtained suggested three, five and six cluster solutions that were possible for accommodation service providers, tour operators, and tourism policy makers respectively. Further, a classification model was implemented for Indian domestic tourists for each of the stakeholders. The inputs received to the classification model were the variables obtained from the stakeholders' responses. Notably, the classification models were developed using respective clustered groups, as being the desired group. Three classifiers were developed, namely, C5.0 decision tree classifier, K-Nearest neighbor classifier, and support vector machine (SVM) classifier. Test data set was prepared by collecting responses from customers whose contact details were shared by tour operators and accommodation service providers (i.e. 72 customer responses data from tour operators and 31 customer data from accommodation service providers) based on our initial set of questionnaire survey. The cluster belongingness of test data was shared by stakeholders. Importantly, we could not get any customer data from the policy makers. Our empirical results (based on performance metrics) suggested that given sufficient training data, SVMs tend to be the best classifiers on test data; in fact, this was true for both of the stakeholders. Thus, our results hold substantial implications to the stakeholders' perspective of successful destination marketing, which would positively contribute to formulating useful policies, planning and strategy, primarily at the government's level.

*Keywords:* Indian tourism, Segmentation, Domestic market, Factor analysis, Stakeholders, Classification, Tourism policy, K-means clustering, C5.0 decision tree, K-Nearest neighbor, Support vector machine, Planning and strategy