Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Leveraging AI for advanced analytics to forecast altered tourism industry parameters: A COVID-19 motivated study

Ankur Kumar\textsuperscript{a}, Subhas Chandra Misra\textsuperscript{a,}\textsuperscript{*}, Felix T.S. Chan\textsuperscript{b,}\textsuperscript{*}

\textsuperscript{a} Industrial Management Engineering, Indian Institute of Technology, Kanpur, Kanpur, Uttar Pradesh, India
\textsuperscript{b} Department of Decision Sciences, Macau University of Science and Technology, Taipa, Macao

\section*{A R T I C L E   I N F O}

\textbf{Keywords:}
COVID-19
Tourism industry
Artificial Intelligence
Foreign Tourist Arrivals
Random forest model

\section*{A B S T R A C T}

COVID-19 pandemic has given a sudden shock to economy indices worldwide and especially to the tourism sector, which is already very sensitive to such crises as natural calamities, terrorist activities, virus outbreaks and unwanted conditions. The economic implications for a reduction in tourism demand, and the need to analyse post-COVID-19 tourism motivates our research.

This study aims to forecast the future trends for foreign tourist arrivals and foreign exchange earnings for India and to formulate a model to predict the future trends based on the COVID-19 parameters, vaccinations and stringency index (Government travelling guidelines). In the study, we have developed artificial intelligence models (random forest, linear regression) using the stacked based ensemble learning method for the development of base models and meta models for the study of COVID-19 and its effect on the tourism industry. The architecture of a stacking model consists of two or more base models, often referred to as level-0 models, and a meta-model that combines the predictions of the base models, and is referred to as a level-1 model (\textit{Smyth & Wolpert, 1999}). The results show that the projected losses require quick action on developing new practices to sustain and complement the resilience of tourism per se.

\section*{1. Introduction}

The COVID-19 virus can survive on surfaces for several hours, and the first case can be traced back to 17 November 2019. The zoonotic virus quickly spread across the globe, thanks to the extremely liberal movement of goods and people. The virus made its way to Europe where Italy, UK, France and Spain were the most affected countries. The virus then reached the USA and severely hit the country with more than two and a half million fatalities up to the 1st of January 2021. By February 2020, most of the countries around the world were infected by the disease. Since the first case in China, the world has been under the influence of the virus outbreak, which was declared as a pandemic by the World Health Organization (WHO) on 11 March 2020. Following the development of COVID-19 around the globe and the precautionary measures taken to curb the spread, the overall world economy has been under severe strain, especially the tourism industry. (\textit{Altan & Karasu, 2020}).

As of the 20th Nov 2020, the total number of cases globally has crossed the 58 million mark, with total deaths around 1.4 million. Although the situation is ameliorating in Europe with the number of active cases consistently declining, the virus is still creating havoc in major countries like the USA, India, Brazil and Russia.

Confirmed COVID-19 cases in India have reached almost 8 million and the total number of fatalities crossed the psychologically painful mark of 1.2 million by 21st October 2020. The data has been taken from the Indian Government website (Indian economy gets a tourism boost; a tenth of GDP and crores of jobs - The Financial Express, n.d.).

Certainly, the modern world has never faced such a situation for at least a century. During recent times, travel and tourism had reached a pinnacle and experienced ever-increasing numbers due to greater international mobility and connectivity. Travel in India is restricted to domestic options like most countries. Officials are considering when to continue international travel in the pre-pandemic levels and we may get a verdict on foreign travel soon. By taking precautions to limit the effect

Abbreviations: R\textsuperscript{M}, Input dataset of M number, where M is a finite real number; N, Dataset of Labels; D, Training Data Set; h\textsubscript{0}, base regressor of t number of training data points; H(x\textsubscript{1}), Level 1 regressor based on the level 0 regressor h\textsubscript{t}; x\textsubscript{1}, x\textsubscript{2}, x\textsubscript{3}.... where x\textsubscript{i} is an input dataset; y\textsubscript{1}, y\textsubscript{2}, y\textsubscript{3}.... where y\textsubscript{i} is an output dataset label; i,j, finite real numbers.

* Corresponding authors.
E-mail addresses: sahajankur@gmail.com (A. Kumar), subhasm@iitk.ac.in (S.C. Misra), tschan@must.edu.mo (F.T.S. Chan).

https://doi.org/10.1016/j.eswa.2022.118628

Received 29 March 2021; Received in revised form 16 August 2022; Accepted 16 August 2022
Available online 22 August 2022
0957-4174/© 2022 Elsevier Ltd. All rights reserved.
of the virus, international connectivity has come to a halt like never before, grounding the tourism industry. Before performing any further probing on the tourism industry or the effect of Covid-19 on the tourism industry, it is essential to address the fundamental questions to understand the current and past tourism status and its parameters, and the methods utilized to perform the forecasting analysis.

Prior to the pandemic, the tourism industry was prosperous. Especially in the last decade, the tourism industry grew like never before. All the various sectors of this industry, viz. tourist attractions, hotels & hospitality, transportation modes, etc were all on the rise in their respective domains and these aspects complemented each other and helped the tourism industry.

According to The World Travel and Tourism Council (WTTC), tourism generated 194 US billion dollars or 6.8% of India’s GDP in 2019 and supported 39.80 million jobs which are 8% of the total employment. The sector was predicted to grow at an annual rate of 6.9% to 460 billion dollars by 2028 which is 9.9% of GDP based on pre-covid research. But these numbers were massively disturbed due to the sudden arrival of the Covid-19 pandemic, whose arrival changed the entire scenario. Travel was completely shut off during the lockdown. This did not just affect India, but also all those nations which implemented complete lockdown. Even in the absence of lockdown, tourists themselves were hesitant in making unnecessary travel.

However, a year after the pandemic hit, the travel industry has again started to recover. Countries like UAE are amongst the first to open the option to travel to the country to fully vaccinated tourists with all pre-cautions. Even large economies like the USA have also lifted the travel option to travel to the country to fully vaccinated tourists with all pre-cautions. Even large economies like the USA have also lifted the travel ban from across almost all nations. India, too, has followed suit and local tourism has almost come back to its original shape, although, the inbound foreign tourists are yet to arrive in full numbers to India.

Several past studies have shown how the tourism industry was shaped by external effects, be it a pandemic or other known or unknown parameters. Most existing studies followed a conservative approach of estimating the trend, seasonality or performing time series analysis. Several incidents that have happened worldwide affected tourism, including Ebola (EVD, which originated in the Democratic Republic of Congo and Sudan in 1976. It was transmitted from human to human via contact by bodily fluids or the blood of an infected person. The mortality rate in Ebola was higher than COVID-19. The WHO publicly announced an outbreak of Ebola on the 23rd of March 2014 (The health issues of the disease are not discussed in this paper). Sierra Leone, Liberia and Guinea were the highly affected countries by Ebola. The growth in all three countries started to decline after the outbreak of this epidemic. The scenario was even worse in tourism-dependent countries due to the sudden drop in the number of tourists, although the outbreak was controlled by taking effective measures worldwide. Still, the economies of these countries were severely hit. At this point, forecasting tourist demand with subjective or hypothetical models would have high importance in planning in terms of making policies to cope with these crises.

Globally, all countries have put immense efforts to stop the unbridled pandemic by imposing strict partial and complete lockdowns, prohibiting the movement of citizens and products. Although the imposed lockdowns in different countries have helped in flattening the curve to some extent, they had an extremely negative effect on the world economy. This directly hit the national economies of the country and by taking all these measures to stop the pandemic, national and international tourism industry came to a halt. A total PTAs in 2020. With restrictions on movement across the world, the tourism industry has been inevitably hard-hit. The closing of international borders, widely imposed travel restrictions and the persisting fear of the spread of coronavirus infections have drastically reduced the demand for international as well as domestic tourism. (Gössling et al., 2020).

India has been known for its hospitality to guests, and the tourism and hospitality industry is now a significant economic sector. With improving connectivity across the globe and reducing distances, this sector has immense potential to contribute even more than its present numbers in terms of revenue and employment. The tourism and hospitality industry can be seen as a group with various other sectors which include: 1) Food and beverage; 2) Travel and tourism; 3) Lodging; 4) Entertainment and recreation. In 2016, the tourism and hospitality industry accounted for 10.2% of the global GDP which is equal to the US $7.6 trillion and 292 million jobs. Its contribution to GDP growth was 3.1% (+"UNWTO World Tourism Barometer and Statistical Annex, 2020). India, as of 2019, generated nearly 10 per cent of the GDP from the tourism and hospitality sector (World Travel & Tourism Economic Impact Report | Destinations International, n.d.). In addition, it accounted for more than 8 per cent of total employment in the country in 2019, i.e., around 42 million jobs (Indian economy gets a tourism boost; a tenth of GDP and crores of jobs - The Financial Express, n.d.). With so much uncertainty in the situation due to the pandemic, the people associated with the tourism and hospitality sector will arguably be the worst hit.

In terms of numbers, as of June 18th, 2021, more than 178 million people in the world and 29.5 Million people in India have been affected by this pandemic. Zooming in, these are just the recorded ones by the official sources. The actual numbers are way beyond these, and tourists are very much aware of this.

The tourism industry is also witnessing unprecedented trends similar to the nature of the pandemic. This study includes how the patterns of arrival, stay, and expenditure has changed because of the COVID-19 pandemic. Moreover, the hospitality industry is experiencing a plethora of required changes at the grass-root level which used to be rather unnoticed and accepted by the travellers in the past, viz., the norms of social distancing, the highest levels of hygiene and protection measures, and the ability to create secure bio-bubbles, all come with their band of underlying costs and additional layers of increased competition within.

Countries like India, which are known for their dense population base and more per unit footfalls of travellers are more susceptible to the high-end tourists avoiding their visits to these countries. They are now preferring to move to more open and less-dense locations, and are seeking more facilitated and secure services. The economic implications for a decline in tourism demand, and the need to understand the quantitative and qualitative driving parameters of tourism intentions in the COVID-19 and post COVID-19 era motivates our research.

According to the latest research, which incorporated the Covid-19 effect on this industry, the following are some statistics about Indian tourism.

The FTAs in April 2019 were 7,71,353 as compared to 7,45,051 in April 2018 registering a growth of 3.5%. Further to this, FTAs during the period January-April 2019 were 39,35,293 as compared to 38,60,871 in January- April 2018 registering a growth of 1.9%. The number of Foreign Tourist Arrivals (FTAs) in India during 2019 increased to 10.93 million as compared to 10.56 million in 2018. The growth rate in FTAs during 2019 over 2018 was 3.5% as compared to 5.20% during 2018 over 2017 (Indian Tourism And Hospitality Industry Analysis Presentation | IBEF, n.d.).

The share of India in international tourist arrivals in 2019 was 1.23%. India accounted for 4.97% of international tourist arrivals in Asia Pacific Region in 2019, with the rank of 7th. About 77.4% of the FTAs entered India through air routes followed by 21.7% by land routes and 0.9% by sea routes. Delhi and Mumbai airports accounted for about 41.7% of total PTAs in India. The top 15 source marked for FTAs in India in 2019 were Bangladesh followed by the United States, United Kingdom, Australia, Canada, China, Malaysia, Sri Lanka, Germany, Russian Federation, France, Japan, Singapore, Thailand and Nepal. The top 15 countries accounted for about 76.3% of total FTAs in India in 2019.

Tourism continues to play an important role as a foreign exchange earner for the country. In 2019, foreign exchange earnings (FEE) from tourism were the 30.06 billion USD as compared to 28.59 billion USD in

A. Kumar et al.
2018, registering a growth of 5.1%. The number of domestic tourist visits in India during 2019 was 2321 million as compared to 1853 million in 2018, with a growth rate of 25.3 %. The number of Indian Nationals’ Departures from India during 2019 was 26.91 million as compared to 26.29 million in 2018, registering a growth rate of 2.4%. About 98.2% of Indian Nationals’ Departures through air routes followed by land routes 1.6% and 0.3% by sea routes. Delhi and Mumbai airports accounted for about 41.7% of the total Indian Nationals’ Departures. The Top 10 source countries for INDs in 2019 were UAE followed by Saudi Arabia, USA, Thailand, Singapore, Qatar, Kuwait, Oman, UK and Malaysia. The Top 10 countries accounted for about 71.51% of the total INDs in 2019(Indian Tourism And Hospitality Industry Analysis Presentation | IBEF, n.d).

Another key policy issue that affects the tourism industry is the presence of seasonality in tourist arrivals which further creates instability in demand and revenue. The seasonal component in a time series distorts the prediction attempts for policy-making. In this context, it is essential to suggest an accurate method of producing a reliable forecast of foreign tourist arrivals. The data on India’s inbound tourism from January 2001 to June 2018 were used for preparing the forecast for the period July-2018 to June-2020. Based on Mean Absolute Error, Mean Absolute Percentage Error and Mean Square Error, the findings infer the relative efficiency of our models in forecasting the foreign tourist arrivals in India.

Thus, to reduce the perceived negative impacts of seasonality in Indian inbound tourism and to ensure foreign tourist visits round the year, niche products best suited for Indian climatic and socio-cultural-institutional conditions need to be introduced and promoted on a large scale at the national and global levels. This paper first explores the pattern of seasonality in foreign tourist arrivals in India, and then examine the forecasting performance of two multiplicative univariate time-series models, Holt-Winters’ model and the Seasonal ARIMA model, for producing a better forecast of foreign tourist arrivals in India. The time series plot of the monthly data on foreign tourist arrivals in the country not only reveals the rising trend in such arrivals but also unveils peak and lean seasons in Indian inbound tourism. These findings are significant for the policy circle working on the sustainable development of tourism in India. The finding of a rising trend in foreign tourist arrivals signals the government as well as private stakeholders remaining prepared to welcome an increasing number of international tourists in the years ahead. The finding of seasonality in Indian inbound tourism indicates December as the peak month and May as the lean month of every year.

Any major world event sways the economic performance of a country’s industries. A global pandemic of such a degree as that of COVID-19, being no different, has distorted the economic indicators worldwide. At a time when even going out of the house is not advisable, one of the many sectors which bears the upshot is tourism and hospitality. It is difficult to analyse what is the current impact and expected future repercussions which could shape the output from the industry. This paper aims to analyse the impact of the pandemic on the tourism sector using Artificial Intelligence (AI) modelling (random forest & linear regression modelling). Furthermore, it also aims to model forecasting of the performance of the sector based on various scenarios, such as government strategies, so as to tackle the pandemic.

With distances reducing and incomes rising, the tourism and hospitality sectors have become significant contributors to the global economy. They have predicted a 1.25 trillion rise (1685 million USD) loss in the tourism industry alone (Indian Tourism And Hospitality Industry Analysis Presentation | IBEF, n.d). Under normal circumstances, tourism demand forecasts can be predicted by qualitative and/or quantitative methods, but the coronavirus pandemic has changed everything and caused a sudden shock to the economic indices worldwide (Chakhroarty et al., 2020).

This study aims to predict the impact of COVID-19 on foreign exchange earnings and foreign tourist visits for India under different scenarios, and can be seen as an early-stage analysis at a time when the pandemic is creating a havoc among the various industries, particularly for the tourism industry. The results show that the projected losses require quick action in developing new practices which are sustainable and complement the resilience of the tourism sector per se. The data has been taken from the University of Oxford website (Coronavirus Pandemic (COVID-19) – the data · Our World in Data, n.d). Based on our understanding of effects of Covid-19 on the tourism and hospitality industry, past and current trends of tourism, influential parameters, and overview of past studies in this field, our paper aims to advance the research in this field by leveraging Artificial Intelligence for advanced analytics to forecast altered tourism industry parameters. The further sections in this paper provide holistic study of the same, starting with the review of the existing literature in this field and addressing the gap unfulfilled by these studies.

2. Literature review

A great number of news articles, journals, expert opinions and research papers pertinent to the topic was referred to for this study, analysing the situation both qualitatively and quantitatively, specific to certain countries and regions. Karim et al.,2020) analysed the situation qualitatively and tried to shed light on the losses that the virus caused to various sectors including aviation, hotel and tourism (Ceylan, 2020; Karim & Haque, 2020). The results of the study by Bakar and Rosbi (2020) in “Effect of Coronavirus disease (COVID-19) to the tourism industry” showed that the pandemic has created fear that has resulted in low numbers of tourists. The pandemic is spreading even after the lockdowns in many countries, an important factor when considering the travel decisions of the public. Therefore, in balancing the supply and demand, the prices in the tourism sector have significantly come down with the decrease in the number of foreign tourist arrivals (Bakar & Rosbi, 2020. The suggested framework in the research paper “Panemics, tourism and global change: a rapid assessment of COVID-19” by the Gößling et al.,2020) analysed the various data available and presented various scenarios as and how it is presented by various organizations like WATA (World Association of Travel Agencies) & WTTO (World Travel & Tourism Online). The methodology included a thorough analysis of the data already available and does not necessarily forecast the performance but presents sustainable arguments (Gößling et al., 2020; Toğaçar et al., 2020).

Thorough research by Maphanga & Henama (2019), based on the literature review of currently available data, suggested ways to build effective strategies to reduce future negative impacts. It emphasizes better management of outbreaks in the future in “The Tourism Impact of Ebola in Africa: Lessons on Crisis Management” (Maphanga & Henama, 2019). “No Ebola…still doomed: – The Ebola-induced tourism crisis” presented an analysis of the shocks and shortcomings in planning to cope with outbreaks by Marina Novelli (2020) (Goebel et al., 2016; Novelli et al., 2020; Sifolo & Sifolo, 2015). Kapiki (2011) in the paper “The Impact of Economic Crisis on Tourism and Hospitality: Results from a Study in Greece” identified the key indicators of sector performance like basic data and performance indicators (Kapiki, 2011).

In another research paper titled, “Seasonality in Tourism and Forecasting Foreign Tourist Arrivals in India”, Mishra et al.2018) explored the pattern of seasonality in foreign tourist arrivals in India, and then examined the forecasting performance of two multiplicative univariate time-series models, the Holt-Winters’ model and the Seasonal ARIMA model, for producing a better forecast of foreign tourist arrivals in India. The time series plot of the monthly data on foreign tourist arrivals in the country not only revealed the rising trend in such arrivals, but also identified the peak and lean seasons in Indian inbound tourism (Mishra et al., 2018). Further, the comparison of MAPE, MAE and MSE, concluded on the better forecasting ability of Holt-Winters’ multiplicative model in the context of foreign tourist arrivals in India. These findings are significant for the policy circle working on the sustainable
development of tourism in India. The finding of a rising trend in foreign tourist arrivals signals the government as well as private stakeholders to be prepared to welcome an increasing number of international tourists in the years ahead. The finding of seasonality in the Indian inbound tourism indicates December as the peak month and May as the lean month each year.

In another research paper, Chenguang Wu et al. (2020) have explored scenario forecasting for global tourism. This study provides innovative forecasts of the probabilities of certain scenarios of tourism demand. The scenarios of interest are constructed in relation to increasing tourism and economic growth. The likelihood forecasts based on these scenarios provide valuable information for destination policy makers. The time-varying parameter panel vector autoregressive (TVP-PVAR) model is adopted for scenario forecasting. Both the accuracy rate and the Brier score are used to evaluate the forecasting performance. A global set of 25 tourism destinations was empirically examined, and the results confirmed that the TVP-PVAR model with a time-varying error covariance matrix is generally a promising tool for forecasting. Their study contributes to tourism forecasting literature in advocating the use of scenario forecasting to facilitate industry decision making in situations wherein forecasts are defined by two or more dimensions simultaneously. In addition, it is the first study to introduce the TVP-PVAR model to tourism demand forecasting (Chenguang Wu et al., 2020).

A research study titled “Forecasting tourism recovery amid COVID-19” by Zhang et al. (2021) implied the significant impact of the coronavirus 2019 (COVID-19) pandemic on global tourism activity methods to predict the recovery of tourism from the devastating effects of COVID-19. In the study, econometric and judgmental methods were combined to forecast the possible paths to tourism recovery in Hong Kong. The autoregressive distributed lag-error correction model was used to generate baseline forecasts, and Delphi adjustments based on different recovery scenarios were performed to reflect different levels of severity in terms of the pandemic’s influence. These forecasts were also used to evaluate the economic effects of the COVID-19 pandemic on the tourism industry in Hong Kong (Zhang et al., 2021). Forecasting tourism demand drew interest to the more advanced forecasting techniques, together with the requirement for more accurate forecasts of tourism demand at the destination level due to the constant growth of world tourism, this has led us to evaluate the forecasting performance of neural modelling relative to that of time series methods at a regional level (Claveria & Torra, 2014). Seasonality and volatility are important features of tourism data, which makes it a particularly favourable context in which to compare the forecasting performance of linear models to that of nonlinear alternative approaches. Pre-processed official statistical data of overnight stays and tourist arrivals from all the different countries to Catalonia from 2001 to 2009 was used in the study. When comparing the forecasting accuracy of the different techniques for different time horizons, autoregressive integrated moving average models outperform self-exciting threshold autoregressions and artificial neural network models, especially for shorter horizons. These results suggest that the there is a trade-off between the degree of pre-processing and the accuracy of the forecasts obtained with neural networks, which are more suitable in the presence of nonlinearity in the data. In spite of the significant differences between countries, which can be explained by different patterns of consumer behaviour, it was found that forecasts of tourist arrivals were more accurate than forecasts of overnight visits.

In other research, authors researched Forecasting tourism demand with a composite search index (Li et al., 2017). They proposed a new forecasting framework with search trend data which were applied to the prediction of Beijing tourist volumes. First, they introduced a GDFM that used the common components of search trend data to construct a more comprehensive index. Second, they compare this new index with a traditional time series model and the PCA-based index model commonly used in past studies. They evaluated the performance of the econometric models with different indices by using static and dynamic tests. The empirical study indicated that their framework has a more favourable performance than other econometric models. First, a significant co-integration relationship exists between the index and Beijing tourist volumes. Second, the Granger causality tests suggested that search trend data lead the actual tourist volumes. Third, they demonstrated that the econometric model with the new index has the best forecasting accuracy in the one-week and four-week forecasts. They also conducted rolling window forecasts to check the robustness. The empirical results validate our framework, which offers a suitable solution for better manipulating large-scale search trend data.

An empirical analysis by Tiwari et al. (2018) on “Foreign tourist arrivals in India from major source countries” stated that the impact on the tourist arrivals from eight out of the top 17 countries to India is temporary, although the other nine countries could bear a permanent shock. Hence, the government should give high importance to attracting tourists from these nine countries during developing a recovery plan and make long-term plans for these nine countries to promote the tourism industry (Tiwari et al., 2018).

Understanding tourism resilience: Adapting to social, political, and economic change” by Joseph Cheer states the parameters and success factors in the tourism industry, and presents a graph depicting the resilience parameters with respect to change rate and tourism scale (Joseph M. Cheer, 2018). The finding in the paper “The future of the global tourism system post COVID-19” by Khalilzadeh (2020) suggested that countries hit with SARS-such as Singapore and Hong Kong have done better in containing COVID-19, which could be instrumental in the possibility of containing another viral outbreak in future and how it will be handled, eventually relieving the impact on the industry. The measures taken could provide better insights into sustainable practices which makes the industry better equipped in the next crisis (Khalilzadeh, 2020).

A Scenario Analysis “Assessing the Short-term Impacts of COVID-19 Pandemic on Foreign Visitors Demand for Turkey”, started with giving a timeline of the virus. Further, it quantitatively forecasted tourism in Turkey, taking various possibilities on the length of the lockdown (Al-Awadh et al., 2020; Topaça et al., 2020). The paper “The Great Patient of Coronavirus- COVID-19” studies the impact on tourism industries in various countries like China, Australia and Italy. It also discusses the opportunities in the future that the virus has created (Folini & Theodore, 2020). An article that was published by Cambridge Econometrics predicted the trends in the economy in the upcoming years by looking at different sectors through different ways, like the number of working hours lost, change in consumption and many more (Ardalakani et al., 2020). An analysis by the UNWTO (United Nations World Tourism Organization) also predicted the loss to tourism and allied industries such as air travel, restaurants and hospitality (Bakar & Rosbi, 2020; Ceylan, 2020; Folinas & Theodore, 2020).

Another study “Crisis Management During the SARS Threat” presented the case study of a hotel in Hong Kong. This paper identifies the impact of SARS on hotels in Hong Kong, and the relevant human resources strategies used to tackle the crisis. The findings show that unpaid leave and involuntary separation were the most common immediate solutions adopted by the hotels (Leung & Lam, 2004).

One of the studies done for understanding the tourist arrival recovery pattern was the analysis based on catastrophe theory by taking the cusp catastrophe model as its foundation, the study proposed a well-grounded approach to understanding the nature of the recovery processes and in explaining the difference between the recovery patterns displayed by arrivals from Japan and those from Hong Kong and USA. Implications regarding tourism promotion policies were drawn from the analysis (Mao et al., 2009).

In another work, the authors have tried to incorporate an empirical approach to study the effects of SARS on the Hong Kong tourism industry. The analysis found that data series of 24 affected countries contain unit roots and hence any form of exogenous shocks, like the SARS epidemic, can have permanent impact on the number of tourist
Lastly, the aftereffects of the pandemic are modelled with the scenario COVID-19, the effect has been modelled to forecast the expected loss. The Simple model of auto-regression has been used to analyse the personality index models of the expected growth of Foreign Tourist Arrivals (FTAs) in 2020 as per the original trend (i.e. if the pandemic in India are noted. The Simple model of auto-regression has been used to manage the negative effect of SARS (Au et al., 2005).

The studies in the past have induced several quantitative measures into their models, which are for-sure very much required. However, this pandemic, which was undoubtedly unforeseen, requires a more detailed study which not just covers both quantitative and qualitative parameters which are affecting the patterns of the tourism industry as a whole, but also the level of granularity required to assess the behaviour of the traveller. This study has been motivated from the fact that the most basic elements from the behavioural changes originating out of the consequences of this pandemic and henceforth their effects need to be analysed in a manner which holistically covers all the underlying aspects.

3. Objectives of the research

In our study, a scenario formulation approach is applied to analyse the range of losses in the tourism and hospitality industry due to the COVID-19 pandemic. Prior to the same modelling, the trend and seasonality index models of the expected growth of Foreign Tourist Arrivals in India are noted. The Simple model of auto-regression has been used to predict the FTA in 2020 as per the original trend (i.e. if the pandemic had not happened). Further, with the help of the critical factors of COVID-19, the effect has been modelled to forecast the expected loss. Lastly, the aftereffects of the pandemic are modelled with the scenario formulation approach. After taking all the parameters into account, We would like to validate our objective based on the Hypotheses given in the below Table 1:

| Table 1 | Hypotheses for the Objective function. |
|---------|---------------------------------------|
| H1:     | There is no relationship between the number of deaths and the number of vaccinations. |
| H2:     | There is no relationship between the number of new cases identified and the number of vaccinations. |
| H3:     | There is no relationship between the number of deaths and the stringency Index Population |
| H4:     | There is no relationship between the number of deaths and new test per thousand Population |
| H5:     | There is no relationship between the number of new cases identified and new test per thousand Population |
| H6:     | There is no difference among the MAE Score obtained in all the five models |
| H7:     | There is no recovery in FTAs of India as per the predictions of various models |

arrivals. Included in this category are Japan, Taiwan, the US and the UK, which are the main source of tourists for Hong Kong. The paper recommends that authorities take source-country-specific measures to manage the negative effect of SARS (Au et al., 2005).

Fig. 1. Year on Year Real Quarterly growth of Foreign Tourist Arrivals in India compared to the previous year.

4. Scope of the study

We aim to capture the foreign tourist demand and foreign exchange earnings through tourism with the appropriate tools. Data has been collected from the Ministry of Tourism website, Monthly Press releases. This being so, it was extremely challenging to include all the other factors like contactless hospitality, tourism perception, and interpersonal factors, etc. and substantially include these in the paper as sufficient data are not available for these factors. To address and include these factors, the currently available data points alone are not sufficient and it would require another survey-questionnaire based approach to holistically capture the above mentioned parameters and the authors feel that this approach can lead to a significant research paper. Henceforth, this calls for another study in future to incorporate these factors.

Fig. 1 depicts the Year-on-Year quarterly growth of Foreign Tourist Arrivals (FTAs) in India. The terrorist attacks of 9/11 at the World Trade Centre affected tourist arrivals to India by more than 20% in 2001 Q4, and the shock took at least 4 more quarters to recover from the losses in tourist arrivals since the previous year. Similarly, during the global financial crisis of 2009, the loss of FTAs can be seen from 2008 Q4 which fell to around 13% by 2009 Q1, the same took 4 quarters for recovery. Much can be said about the resilience of the tourism industry through this trend. Usually, it has taken around four quarters to recover from YoY losses in FTA. This trend can be expected to be followed for this uncertain pandemic situation where we still don’t know when the travel bans will be lifted and when life will come back to normal. The further direction of research can be hypothesized in relevant cases. However, this situation is an unparalleled crisis where most countries have banned international travel, a move which has never been implemented before. This is bound to largely impact on the resilience of the industry. Thus, the study aims to estimate FTAs and relate the same with Foreign Exchange Earnings (FEE). The Y-axis shows the percentage of Year on Year quarterly growth of Foreign Tourist arrivals compared to previous year.

As per the recovery trends, every time the FTAs fell, they rose significantly after recovery. A similar trend may be seen but it is still dependent on a lot of unaccounted for and unknown variables. Any
Further development is focused on how the pandemic spreads in the country. In accordance with recent trends in 2018, as depicted in Fig. 2, the top 3 countries (United States, United Kingdom & Bangladesh) alone accounted for around 45% of FTAs in India. The top 15 countries accounted for around 70% of the same which tells us that the situation in these countries largely affect the arrivals in the nation.

Fig. 3 shows a heatmap of the various states in India with respect to tourist visits by foreign nationals in 2018. It can be seen that Tamil Nadu accounts for the greatest number of foreign tourist visits at 21%, closely followed by Maharashtra, Uttar Pradesh and Delhi with 18%, 14% and 10% respectively. More than 80% of the visits can be attributed to the above listed four states, along with Punjab, Rajasthan, Telangana and West Bengal. These states being the most popular traditionally for foreign tourists do attract the foreign tourists. Suppose, these states are severely hit by the pandemic, inbound visits may decrease thus impacting on the overall tourism which is very low as of now. The data has been taken from the Indian government website (Annual Reports Archive | Ministry Of Tourism | Government of India, n.d).

Fig. 4 shows the contribution of the mode of transport facilitating foreign tourist visitors to India. Moreover, air travel has always been the favourable source mode for tourist arrivals. It had been on a decline since the year 2013 as the land route share started increasing rapidly, but still, around 80% of FTAs can be attributed to air travel. In 2018, 79.6% FTAs were attributed to air transport and 17.6% to land transport, the sea route has never been a favourable option for tourists in India although it too has been on the rise since 2015, staying at around 0.8% in 2018 (Annual Reports Archive | Ministry Of Tourism | Government of India, n.d). It is another observation which may impact on the arrivals with further developments. Given that the airlines may decide to run aeroplanes at half capacity, there may be a direct 50% loss in FTAs in addition to the indirect decline due to the pandemic. Till now, all the above data are considered as the major source concerning FTAs, major ports of entry, most favourable state for tourist visitors, etc may highly impact on the effect of the pandemic on the industry. The Y-axis shows the percentage share of mode of transportation for FTA.

Fig. 5 shows the percentage of foreign tourist arrivals at different ports of entry to India. For example, in Fig. 5, Delhi and Mumbai ports lead with almost 45% of the share of FTAs. The situation as of now is relatively worse for both these places due to the less demand of tourists. Similarly, the severity of the infection in source countries may affect demand and likewise, for destination states too, the infection spread is bound to affect the tourist demand.

Fig. 6 shows the monthly FTAs in the Ebola-affected country Sierra Leone, somewhere before and after the epidemic from 2013 to 19. Initially, there was a sudden drop in FTAs. Even after the four years of the epidemic, the FTAs in Sierra Leone could not reach the baseline before the epidemic. One model has been developed for the prediction of FTAs in India by training the model using this data. The Y-axis shows the number of foreign tourist arrivals in Sierra Leone. The data has been taken from the Indian Government website (Sierra Leone Tourist Arrivals | 2009-2019 Data | 2020-2022 Forecast | Historical, n.d.).

Fig. 7 shows the density plot of FTAs with critical factors during and before the pandemic. The data has been taken from the Indian Government website (Annual Reports Archive | Ministry Of Tourism | Government of India, n.d) for all the pre-processing and training.
purposes and data has been plotted using Matplotlib library (Barrett et al., 2005).

The following critical factors are considered in the modelling:

1. COVID-19 Counts: Total number of COVID-19 Cases on that specific day has been considered for modelling
2. Deaths: Total number of COVID-19 Deaths on that specific day has been considered for modelling
3. Stringency Index
4. COVID-19 Test: Total number of COVID-19 Tests on that specific day has been considered for modelling

FTAs were affected by the sudden shock initially with increasing numbers of COVID-19 Counts, Deaths Stringency Index & COVID-19 Tests. The reason for including these factors in the model is that FTAs are highly affected by this. The fatality rate is a derived factor from the COVID-19 counts & deaths by the virus. Fig. 8 shows the Pair plot of FTAs in India with critical factors, during and before the pandemic. FTAs decrease with increase in all other four parameters which shows strong negative correlation and can be observed with the given pair plot.

Fig. 9 shows the correlation plot of FTAs in India with critical factors during and before the pandemic up to 21st October 2020. FTAs have a strongly negative correlation with all the other factors and mostly with the stringency index which shows that the restriction of movement by the Government plays an important role in determining the number of FTAs in any country.

Fig. 10 shows the correlation matrix of FTAs in India with critical
factors during and before the pandemic up to 21 October, 2020. FTAs have strongly negative correlations of $-0.915$, $-0.55$, $-0.51$ and $-0.49$ with the stringency index, deaths, tests & COVID-19 counts respectively.

Fig. 11 shows a comparison between the real and modelled forecasts of tourist arrivals from the year 2001 to 2020 Quarter 1. The data has been taken from the Indian Government website (Annual Reports Archive | Ministry Of Tourism | Government of India, n.d.).

Trend and Seasonality analysis showed that foreign tourist arrivals were high in the first and last quarters and low during the second and third quarters annually over the last two decades at least. The forecasted values were predicted by a devised model which developed a seasonality index for the data already available. Furthermore, the annual footfall of 2020, as estimated by the autoregressive model, is modelled into the data and it showed a forecasted growth of around 6.18% during the first quarter of 2020 itself. However, due to the pandemic, the real figures dropped drastically and there was a steep loss of more than 22% in the initial quarter when compared to the same quarter in 2019 in regard to FTAs (Annual Reports Archive | Ministry Of Tourism | Government of India, n.d.).

5. Methodology

In this research, the stacked based ensemble learning method has been used for the development of the base models and the meta model. The architecture of a stacking model consists two or more base-models, often referred to as level-0 models, and a meta model that combines the predictions of base-models, referred to as a level-1 model (Smyth & Wolpert, 1999) (Cui et al., 2021).

Prediction of a meta-model is based on the predictions of the base models on sample data, along with the input and output of the training dataset.

There are two major approaches to predict and forecast industry outputs statistically: quantitative and qualitative methods of forecasting. Quantitative approaches try to predict what will happen in the future by calculating the past trends and the relationship between variables affecting demand (Crouch, 1994). Any prediction based on a quantitative method is a forecast of previous data i.e. historical performance of the indices. Any quantitative forecast makes a classic ceteris paribus assumption. However, during the COVID-19 pandemic, there is huge uncertainty as to what will happen in the near future and what conditions can be modelled to a particular extent if critical success factors can be considered during the modelling. Any history-less event cannot be predicted by quantitative methods (Schnaars, 1987) Since COVID-19 is a one of a kind global pandemic which has caused 100% of the tourist destinations to be affected, it is impossible to accurately model all the irregularities in a statistical approach unless we relate it to a similar event and normalize the parameters respectively. In such cases, subjective methods, called qualitative or judgmental, can be used to forecast the future due to the advantage of not requiring historical data (Song, 2003). Such methods, classified as qualitative approaches, are appropriate where historical data are insufficient or inappropriate in forecasting the future.
Scenario analysis is one of the techniques used in the economics, financial or other fields to predict the future (Darvish et al., 2018). Strictly based on the information available, this research aims to analyse the resilience of the tourism industry with several projections for the number of foreign tourist arrivals to India after the international travel ban is lifted. The above context suggests that various scenarios can be generated where the rate of change of demand can be used to project the expected arrivals and compare with the trends along with the previous year i.e., 2019. The research attempts to track every possible situation with the time of border closure and estimated opening. These Government regulations can be defined as a stringency index, calculated based on the restrictions policy factors for travelling in a particular country (COVID-19: Government Response Stringency Index, Dec 19, 2020, n. d.).

Since Machine Learning and Artificial Intelligence techniques provide the optimal support to analyse the data, develop heuristics and predict the output with better accuracy, these techniques have been used in forecasting the trends, considering various scenarios using pandas, NumPy libraries. COVID-19 being one of a kind, data is convoluted, and this calls for the use of machine learning algorithms that have the ability to model such complex data.

A working methodology is proposed in this paper and as the data increases in future, we can switch from traditional machine learning algorithms to complex neural network algorithms.

Machine learning models become more accurate due to their ability to adapt according to the training data, providing flexible models which modify itself in the small change in the training data (Goldani et al., 2021).

5.1. Framework

The idea is to relate FTAs with social distancing norms, stringency index, COVID-19 counts and fatality rate or deaths using multi-level modelling. (Rao et al., 2020; Wu et al., 2019) Each country’s FTAs have been calculated based on the previous month’s data. In one model, FTAs have been predicted comparing the Ebola-affected countries tourism patterns before and after the epidemic.

The top 12 countries accounted for around 70% of the FTAs arrivals in India in the last 7 years, which tells us that the situation in these countries may largely affect the arrivals in India. Hence, these 12
country parameters have been taken individually and the remaining countries’ data have been taken as other countries’ data, due to the unavailability of data and to keep the model as accurate as possible.

Mathematical Model:
\[ X = \{x_i \in \mathbb{R}^M\} \text{ and a set of labels } Y = \{y_i \in \mathbb{N}\} \text{ and a Training Set } D = \{(x_i, y_i)\} \]

1 Step 1: Learn first level regressor.

For \( t \leftarrow 1 \) to \( T \) do,

- Learn a base regressor \( h_t \) based on \( D \).

2 Step 2: Construct new data set from \( D \).

For \( t \leftarrow 1 \) to \( m \) do,

- Construct a new data set that contains \( \{x_i^\text{new}, y_i\} \), where \( x_i^\text{new} = \{h_j(x_i) \text{ for } j = 1 \text{ to } T\} \).

- output \( = y_i \).

3 Step 3: Construct new data set by adding \( y_j \) to \( D \).

For \( m \leftarrow 1 \) to \( M \) do,

- Construct a new data set that contains \( \{x_i^\text{new}, y_i\} \), where \( x_i^\text{new} = \{h_j(x_i) \text{ for } j = 1 \text{ to } M\} \).

4 Step 4: Learn a second-level regressor.

Learn a new regressor \( h_\text{new} \) based on the newly constructed data set.

Return \( H(x) = h_\text{new}(h_1(x), h_2(x), \ldots, h_T(x)) \).

Models have been developed by using the stacked based ensemble

Fig. 9. Correlation factor plot of FTAs in India with critical factors during and before the pandemic.

Fig. 10. Correlation matrix of FTAs in India with critical factors during and before the pandemic.

Fig. 11. The graph shows an analysis of how FTAs vary with respect to the quarters of the same year.
learning method. The architecture and mathematical model represents the stacked based ensemble model consisting two or more base models, often referred to as level-0 models, and a meta-model that combines the predictions of the base models, and is referred to as a level-1 models. Level-0 models have been developed based on the raw dataset although Level-1 models have been developed based on Level-0 models and input dataset.

Model:

$$FTAs = f(x_1, x_2)$$

Abbreviations:
- $$x_1$$: natural growth
- $$x_2$$: loss due to pandemic (risk).

Functional form:

$$FTAs = f(y_1, y_2, y_3, y_4, y_5)$$

Abbreviations:
- FTAs: monthly number of foreign tourist arrivals.
- $$y_1$$: monthly positive COVID-19 cases.
- $$y_2$$: monthly deaths due to COVID-19.
- $$y_3$$: average stringency index monthly.
- $$y_4$$: monthly conducted COVID-19 cases.
- $$y_5$$: natural trend before COVID-19.

COVID-19 daily data has been used during the modelling to increase the accuracy of the model.

Models: linear regression model, random forest model.

Approach 1: In level-0 modelling, base models of COVID-19 counts, deaths, stringency index & tests have been constructed for each country. A second base model has been constructed for the FTA’s of the Top 12 countries and other countries based on each country’s parameters in the next step. In level-1 modelling, a meta model for India’s FTAs has been predicted (constructed) based on the number of FTA’s predicted (base models) in Level 1 modelling (Rao et al., 2020).

Approach 2: In level-0 modelling, base models of the COVID-19 counts, deaths, stringency index & tests have been predicted (constructed) for each country. In level-1 modelling, India’s number of FTAs have been predicted based on the parameters of all countries predicted in level-0 modelling (base models).

Approach 3: In this approach, the base model has been trained for the Ebola-affected countries and results have been predicted for India. Parameters have been normalized to compare both scenarios and constant factors have been taken into consideration to predict the outputs. Ensemble methods & logistical regression (Baniamerian et al., 2019; Pradeepkumar & Ravi, 2017) have been used to predict the number of FTAs. Since studies showed that running at half the capacity would be ideal for this situation, henceforth, the demand may go down to half during the first month of the opening of borders. Furthermore, the scenarios where the demand declines follow a proportional pattern in the tourism sector. The data used for the study were obtained from the Tourism Ministry database of the Government of India up to July 2020. Tourism is currently a severely affected sector from the pandemic and UNWTO has revised its 2020 forecast for international arrivals for a second time, and it also emphasizes that any predictions are likely to be further revised (UNWTO, 2020).

Scenarios: The discussed scenarios aim to include all possible situations with given parameters and provide a range for the estimated loss in Foreign Tourist Arrivals (FTAs).

The UNWTO announced the expectation in claiming that international tourist arrivals will be down 60% to 80% for 2020 when compared to 2019 because of travel restrictions (UNWTO, 2020). On the other hand, tourism professionals are hopeful for an after-COVID-19 recovery. Based on such estimations of the various stakeholders, the study estimated the projections for foreign tourist arrivals in the Indian context without considering the impact of COVID-19. Monthly international tourist numbers were calculated using Eq. (1) as shown below (Günay et al., 2020).

$$FNE \ [2020] = FNE \ [2019] - 1 * (1 \mp r)$$  \hspace{1cm} (1)

$$FEE \ [2020] = FEE \ [2019] - 1 * (1 \mp r)$$  \hspace{1cm} (2)

where r is the decline rate. This calculation has been done for each forecasted month. The number of yearly arriving foreigners has been calculated as the sum of the monthly data for 2020. Lastly, five models have been developed to forecast the monthly data for the next five years using Linear Regression and Random Forest methods.

The data concerning the second quarter of 2020, i.e., April – June has been given as zero in the ministry database. Since a strict implementation of the travel ban was observed in these months and international travel was not allowed, the footfall of foreign tourists during these months is taken to be zero.

5.2. Autoregression

Autoregression has been used in the annual FTA observations since 2000, to predict the annual FTA in the year 2020. R squared and the adjusted R square value shows that the model is accurate. With the same
model, a forecasted value of 11.6 million foreign tourist arrivals was predicted accounting for 6.14% growth in annual foreign tourist arrivals prior to the pandemic, although the situation has changed all the prior calculations.

The given p-value for the Lag is also less than the 0.05 level of significance, implying that the lag accurately models the fluctuations in arrivals. An important thing to note about autoregression is that it assumes that a single time shock affects the series infinitely in future and can be said to be true in the case of tourist arrivals.

6. Experimental analysis

1. Predictions using Ebola Affected countries: As discussed earlier in the methodology, the model has been trained for the Ebola-affected countries and the results have been predicted for India after normalizing the pandemic and epidemic parameters for accurately modelling the effects of all possible declines in the arrivals and foreign exchange. Parameters have been normalized to compare both scenarios and constant factors have been taken into consideration to predict the outputs. Sierra Leone was severely affected with Ebola and their tourism sector has still not recovered from the shock even six years after the epidemic. FTAs were normalized during training the models so that the effect of the epidemic can be modelled in an pandemic. In regard to the fatality rate, Ebola cases were used as an independent variable to train the number of FTAs during the modelling.

Fig. 12 shows the FTAs prediction for India using linear regression which predicts that even after four years, it will recover only around 80% as compared to 2019 with a mean absolute error of 860.18 for FTAs, 231540.9 for COVID-19 count and & 0.37 for fatality rate (monthly) respectively during training.

The model values are quite practical as it is expected to take a longer time for the industry to recover from the effect of COVID-19. The Y-axis shows the number of FTAs prediction for India using linear regression.

2. Prediction using Linear Regression with FTAs: In level-0 modelling, base models of COVID-19 counts, deaths, stringency index & tests have been constructed for each country using Linear Regression. A second base model has been constructed for FTA’s of the Top 12 countries and other countries based on each country’s parameters. In level-1 modelling, a meta model for India’s FTAs has been predicted (constructed) based on the number of FTA’s predicted (base models) in Level 1 modelling. A stacked based ensemble model has been used for accurately modelling the effects of all possible declines in the arrivals using Linear Regression (Wesseling, 1976).

In step 1 modelling, a base model of COVID-19 counts, deaths, stringency index & tests have been modelled for each country for the Top 12 countries and other countries accounted for FTAs in India with an overall mean absolute error of 7409.

In step 2 modelling, the base models of the FTA’s for the Top 12 countries and other countries have been predicted as the dependent variable based on each country’s parameters (COVID-19 counts, deaths, stringency index & tests) with an overall mean absolute error of 320.

In step 3 modelling, a meta model of India’s number of FTAs has been predicted based on the number of FTA’s predicted in Level 2 modelling and India’s features with an overall mean absolute error of 0.046.

Fig. 13 shows the FTAs prediction which shows that the shock will keep affecting the Tourism industry. It might be one of the possibilities as predicted by the model based on the input data given to it, which can be validated in the future. The Y-axis shows the number of foreign tourist arrivals.

3. Prediction using Linear regression without FTAs: As discussed earlier in the methodology, approach-3 has been used for accurately modelling the effects of all possible declines in the arrivals using Linear Regression. There is only one difference in other countries, the FTAs are not predicted during modelling. However, the final prediction is done using the features (COVID-19 counts, deaths, stringency index & tests) of all countries.

In step 1 modelling, base models of COVID-19 counts, deaths, stringency index & tests have been developed for each country for the Top 12 countries which accounted for FTAs in India with an overall mean absolute error of 7409.

In step 2 modelling, a meta model of India’s number of FTAs has been predicted based on the number of FTA’s predicted in the base model of step 1 and India’s features with an overall mean absolute error of 3927. The error is high as compared to the previous method as other countries FTAs are not considered during the modelling, which indicates the significance of this particular factor.

Fig. 14 shows the FTAs prediction of India which predicts that after the four years, the recovery will be around 70% as compared to 2019. The predicted values, although not the same for the subsequent years, do not show much change. The Y-axis indicates the number of foreign
tourist arrivals.

4. Prediction using Random Forest regression with FTAs: As discussed earlier in the methodology, approach 1 has been used for accurately modelling the effects of all possible declines in the arrivals using Random Forest.

   In step-1 modelling, base models of COVID-19 counts, deaths, stringency index & tests have been constructed for each country using random forest for India with an overall mean absolute error of 1034.

   In step-2 modelling, a base model of FTA’s for the Top 12 countries and other countries has been predicted as the dependent variable based on each country’s parameters (COVID-19 counts, deaths, stringency index & tests) with an overall mean absolute error of 2.8.

   In step 3 modelling, a meta model of India’s number of FTAs has been predicted based on the number of FTA’s predicted in Level 2 modelling and India’s parameters with overall mean absolute error of 16.4.

   Fig. 15 shows the FTAs prediction for India which predicts that after the four years, it will recover around 20% as compared to 2019. The predicted values although not the same for the subsequent years, do not show much change.

5. Prediction using Random Forest regression without FTAs: As discussed earlier in methodology, approach 2 has been used for accurately modelling the effects of all possible declines in the arrivals and foreign exchange using Random Forest. There is only one difference, other countries FTAs are not being predicted. However, the final prediction is being done using the features (COVID-19 counts, deaths, stringency index & tests) of all countries.

   In step 1 modelling, base models of COVID-19 counts, deaths, stringency index & tests have been developed for each country for the Top 12 countries accounting for FTAs in India with an overall mean absolute error of 1107.

   In step 2 modelling, a meta model of India’s number of FTAs has been predicted based on the number of FTA’s predicted in the base model of step 1 and India’s features with an overall mean absolute error of 36.

   Fig. 16 shows the FTAs prediction in India which predicts the sudden rise and drop in the number of FTAs, considering several waves of the pandemic. The model predicts the future values mainly based on the 2020 data and sudden drop in FTAs.

   Fig. 17 shows the scatter plot of FTAs in India with critical factors after prediction. Predicted values similarly show a strong negative correlation in FTAs with all the other factors which can be seen in the given scatter plot.
Fig. 18 shows the pair plot of FTAs in India with critical factors after prediction. Predicted values similarly show a strong negative correlation in the FTA correlation with all the other factors which can be seen in the given pair plots.

Fig. 19 shows the correlation matrix of FTAs in India with critical factors after prediction. FTAs have strongly negatively correlations of –0.877, –0.527, –0.471 & –0.458 with the stringency index, deaths, COVID-19 counts & tests respectively.

Fig. 20 shows the heatmap of the correlation of Foreign Tourist Arrivals in India with four major critical factors considered during the modelling after prediction. All the factors are negatively correlated to these highly dense areas which show the more data points.

Fig. 21 shows the correlation plot of FTAs in India with critical factors after prediction. FTAs have a strong negative correlation with all the other factors and mostly with the stringency index of arrival and departure countries which states that the restriction of movement by the Government plays an important role in determining the number of FTAs in any country. (COVID-19: Government Response Stringency Index, Dec 19, 2020, n.d.).

Several organizations and agencies around the world are focusing on possibilities as to how to tackle it best, vaccines being one of the most effective ways. Vaccines save millions of lives each year and work by training and preparing the body’s immune system by recognizing and fighting off the virus. After the vaccination, if the virus enters the body again, the chance of getting infected is reduced, preventing illness.

There are several effective vaccines that prevent people from getting seriously ill or dying from COVID-19. The severity of the infection is reduced with vaccination. The chances that a person might die of COVID-19 are less when a person has been vaccinated. Approved COVID-19 vaccines provide a high degree of protection against the disease, like getting seriously ill or dying. While ensuring safety, there are some minor side effects that a person may or may not experience after getting vaccinated. The side effects work as signs indicating the person’s body is building protection to COVID-19. The common side effects include arm soreness, mild fever, tiredness, headaches, muscle or joint aches.

In regard to vaccines and some other parameters such as the healthcare services, tests done for covid-19, stringency during the pandemic can affect the number of positive cases and the number of deaths due to COVID-19 for a particular country. Since vaccines are considered to be the most effective means for tackling this COVID-19 pandemic, we need to analyse this parameter in our studies to formulate the model, since we have very limited tourism data during vaccination. At this stage, we can compare it with the other Covid-19 parameters, which affect the number of FTAs in a country, whether stringency, tests done for covid-19, the number of people fully vaccinated, the total number of people vaccinated at least with the first dose can be combined with the vaccination parameter to optimize the vaccines’ effect.

By analysing the plots and from our trained model, it can be inferred that a relationship exists between the number of deaths and number of vaccinations as shown in Fig. 22. In the multiple linear regression model, the p-values for the vaccination variable people vaccinated per hundred is greater than 0.05, implying a null hypothesis cannot be rejected with R squared value of 0.920 and mean absolute error of 0.3894. Therefore vaccination will have positive impact on the number of foreign tourist arrival in India, as the impact of Covid-19 will decrease with increase in the number of vaccinated people.

By modelling the drop in FEE in a similar way to FTA, from a 20 to 70% decline in annual earnings, & the FEE drop can be anywhere between 20% and 74%. The study has some important limitations, and the findings should be handled under the same conditions. The research is conducted by using alternative probabilistic scenarios to predict the foreign tourist arrivals in India. In this respect, the findings of the study should be considered under probable scenarios. On the other hand, providing information and shedding light on the stakeholders for planning and future decisions makes the study important. It is aimed at...
suggesting probable implications of various scenarios and further policy measures. The model aims to predict the foreign arrivals and relates the same with foreign exchange earnings from tourism as per the considered scenarios. Fig. 23 shows the extreme scenarios of the impact on FEE. The best-case scenario if international travel resumes from July 2020 and no demand loss is seen, a loss of US$ 10.32 billion is still inevitable. The magnitude of this amount emphasizes how important a contributor is tourism in the GDP of India.

Fig. 24 shows the mean of predicted range of Tourism recovery in India for all models in Percentage and indicates the continuous recovery after sudden drop. Result of two models represents the steep rise after the drop due to the COVID-19. It is expected that the industry will bounce back when the situation returns to near-normal and starts its revival positively.

7. Results

Tourism is a sector which is very sensitive to global shocks. The recent COVID-19 pandemic has been the biggest health crisis that the modern world has ever had to face. The global economy has been affected many times in the past and tourism has been hit too, but this is the very first time when all of the pressure has come down to the tourism sector globally and all of it has been halted. International travel bans have ensured the loss of the annual footfall in tourist destinations. As per UNWTO, during the global meltdown of 2009, foreign tourist arrivals were reduced by as much as 14% during the first quarter of the same year, courtesy the infamous 9/11 and Indian parliament terrorist attacks, and it fell by as much as 23% in the last quarter of 2001.

The Coronavirus spread rapidly across the globe and India closed its international borders along with a full lockdown from 25th March 2020. June 1 saw the nation ease its regulations regarding domestic travel but the normal volume of traffic and demand has not been reached yet. It isn’t expected to rise that much depending on the developments, with preventive measures against the pandemic extended from day to day. The model was aimed at correctly forecasting the foreign tourist arrivals in India following the pandemic. It is incorrect to say that accurate prediction is possible because it is dependent on a lot of factors which are unknown as variables in the community as of now, but definitely with more data, the model accuracy can be improved. A framework and model have been created to analyse and predict the situation, and with
more data the results will be improved. The model can be used to forecast the factors in similar scenarios.

According to the most recent data on arrivals, it is aimed at analysing and forecasting the change in foreign tourist arrivals in India due to COVID-19 under alternative hypothetical scenarios. The findings of the study show that demand may decline anything between 20 and 70%. The losses in annual FTAs for the year 2020 are projected to be 72% for a 90% demand decline and recovery is seen in all the models as shown in the various models shown in Fig. 25 with significant p values shown in the Table 2. A demand decline which is different for every model or an increasing trend of demand or a sudden increase and then a gradual drop, anything may be possible depending on further developments. This pandemic is set to cause unparalleled damage and loss to the tourism sector which may take years to recover. These effects will recover in the coming years, but one thing that must be noted is that keeping healthy is emphasized as the most important thing for human beings.

8. Discussion

The findings of the study show that demand may decline anything between 20 and 70%. The losses in annual FTAs for the year 2020 is projected to be 72% for 90% demand decline and recovery is seen in all the models. A demand decline which is different for every model or an increasing trend of demand or a sudden increase and then a gradual drop, anything may be possible depending on the further developments. More precisely, another surge or wave of the pandemic will push the industry downwards, but simultaneously increased and speedy vaccination should be able to pull the industry back to normalcy, which would take up to few years to get back to its original shape. This also
puts 100 + Million jobs at stake. As per the outcomes of the study, it can be inferred that not before 2024–2025, can the industry achieve its pre-crisis shape. The hospitality industry should prepare itself well enough not just to witness any of the above-mentioned implications, but also figure out robust methods to bounce themselves back from the after-effects.

By analysing the performance of the models and significant value, it can be inferred that:

a) There exists a relationship between the number of deaths and number of vaccinations.
b) There exists a relationship between the number of deaths and the number of new cases
c) There is a difference among the Mean Absolute Error Score obtained in all the five models.
d) There will be a recovery in FTAs of India as per the predictions of various models.

Trained model was used to make predictions with a high R squared value and a low value of mean squared error (MSE) as compared with the linear regression models which is providing higher MSE values. This provides evidence in support of Hypothesis H₁, H₂, H₆ & H₇.

In the multiple linear regression & random forest model, the p-values for these hypotheses are <0.05, implying that the null hypothesis can be rejected, which does provide evidence in support of the above-stated hypotheses.

| Hypothesis                                                                 | p-value | Result   |
|----------------------------------------------------------------------------|--------|----------|
| H₁: There is no relationship between the number of deaths and the number of vaccinations. | <0.001 | Rejected |
| H₂: There is no relationship between the number of new cases identified and the number of vaccinations. | <0.001 | Rejected |
| H₃: There is no relationship between the number of deaths and the stringency Index | 0.227  | Accepted |
| H₄: There is no relationship between the number of deaths and new test per thousand Population | 0.373  | Accepted |

(continued on next page)
Fig. 23. The expected range of shock to Foreign Exchange Earnings by Tourism in India in US$ billions.

Fig. 24. The mean of predicted range of Tourism recovery in India for all models in Percentage.

Fig. 25. Tourism recovery of predicted range in India for all models in Percentage.
The practical aspect of the study is that policies can be formed based on these projections for reviving the Tourism sector at the earliest. The outputs of all the models show the extremity of the drop in FTAs annually, using all the models in Fig. 26. This means, if we are to face the worst scenario, it will be one of the worst tourism crises that India has ever experienced. This result shows that COVID-19 will be far more destructive than other health borne outbreaks such as SARS (McAleer & Vella, 2010).

Health outbreaks have never impacted on Indian tourism before. The surge in the Nipah virus back in 2018, didn’t impact on the annual footfall negatively. So, it is worth noting that this is the first time in recent decades that a health crisis has impacted very severely on the tourism industry. Recent news about COVID-19 shows that the future of the pandemic is still unclear. On the other hand, we also carry optimistic views. To reduce the impact of the virus on Indian tourism, it is absolutely necessary to formulate recovery plans and implement them urgently. The loss in Foreign Exchange Earnings (FEE) has been around 17.5% in the first three months of 2020 as compared to the last year. The year started positively when January 2020 saw a growth of almost 11% over the previous year. Subsequently, as the cases rose, February 2020 showed a paltry growth of 1.6% and finally during the month when everything was locked down, March 2020 approximated losses of around 66% in the FEE.

It will be one of the most hit sectors for sure because everything in tourism is somewhat dependent on travelling, which was the most affected during this crisis. COVID-19 may turn out to be one such crisis or the start of a world where such threats to life occur now and then. That being said, a lot can be further dwelled upon to enhance the response of the sector to such situations in future. An important thing to

### Table 2

| Hypothesis | p-value | Result |
|------------|---------|--------|
| H5: There is no relationship between the number of new cases identified and new test per thousand Population | 0.416 | Accepted |
| H6: There is no difference among the MAE Score obtained in all the five models | <0.001 | Rejected |
| H7: There is no recovery in FTAs of India as per the predictions of various models | <0.001 | Rejected |

9. Conclusions

COVID-19 is a one of a kind phenomenon, therefore forecasting its impact on the Tourism sector can help Governments to make the appropriate policies for reviving the sector at the earliest time. Hence, the research has been conducted for forecasting the probable impact of COVID-19 on the sector.

The outputs of all the models show the extremity of the drop in FTAs annually, using all the models in Fig. 26. This means, if we are to face the worst scenario, it will be one of the worst tourism crises that India has ever experienced. This result shows that COVID-19 will be far more destructive than other health borne outbreaks such as SARS (McAleer & Vella, 2010).

Health outbreaks have never impacted on Indian tourism before. The surge in the Nipah virus back in 2018, didn’t impact on the annual footfall negatively. So, it is worth noting that this is the first time in recent decades that a health crisis has impacted very severely on the tourism industry. Recent news about COVID-19 shows that the future of the pandemic is still unclear. On the other hand, we also carry optimistic views. To reduce the impact of the virus on Indian tourism, it is absolutely necessary to formulate recovery plans and implement them urgently. The loss in Foreign Exchange Earnings (FEE) has been around 17.5% in the first three months of 2020 as compared to the last year. The year started positively when January 2020 saw a growth of almost 11% over the previous year. Subsequently, as the cases rose, February 2020 showed a paltry growth of 1.6% and finally during the month when everything was locked down, March 2020 approximated losses of around 66% in the FEE.

It will be one of the most hit sectors for sure because everything in tourism is somewhat dependent on travelling, which was the most affected during this crisis. COVID-19 may turn out to be one such crisis or the start of a world where such threats to life occur now and then. That being said, a lot can be further dwelled upon to enhance the response of the sector to such situations in future. An important thing to

![Fig. 26. The expected range of shock to Foreign tourist arrivals in India using all the models.](image-url)
These sustainable practices along with a greater emphasis on domestic tourism will help the sector recover efficiently and thus come back stronger in the near future. In addition, the resilience of the industry is being tested time and is required to be more sustainable and adopt newer practices which promote and safeguard the industry from any other further shocks like this one, or at least be better prepared (Chakraborty et al., 2020).

CRediT authorship contribution statement

Ankur Kumar: Software, Data curation, Formal analysis. Subhas Chandra Misra: Visualization, Investigation, Validation, Writing – review & editing. Felix T.S. Chan: Conceptualization, Methodology, Writing – original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Al-Awadhi, A. M., Alsafi, K., Al-Awadhi, A., & Alhammadi, S. (2020). Death and contagious infectious diseases: Impact of the COVID-19 virus on stock market returns. *Journal of Behavioral and Experimental Finance, 27*. https://doi.org/10.1016/j.jbef.2020.100326

Altan, A., & Karasu, S. (2020). Recognition of COVID-19 disease from X-ray images by hybrid model consisting of 2D curvelet transform, chaotic salp swarm algorithm and deep learning technique. *Chaos, Solitons and Fractals, 140*(July). https://doi.org/10.1016/j.chaos.2020.110071

Annual Reports Archive | Ministry Of Tourism | Government of India. (n.d.). Retrieved December 21, 2020, from https://tourism.gov.in/media/annual-reports/annual-reports-archive.

Ardakani, A. A., Kanafi, A. R., Acharya, U. R., Khadem, N., & Mohammadi, A. (2020). Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. *Computers in Biology and Medicine, 121*. https://doi.org/10.1016/j.compbiomed.2020.103795

Au, A. K. M., Ramazamy, B., & Yeung, M. C. H. (2005). The effects of SARS on the Hong Kong tourism industry: An empirical evaluation. *Asia Pacific Journal of Tourism Research, 10*(1), 85–95. https://doi.org/10.1080/1094166042000130236

Bakar, N. A., & Roshni, S. (2020). Effect of Coronavirus disease (COVID-19) to tourism industry. *International Journal of Advanced Engineering Research and Science, 7*(4), 189–193. https://doi.org/10.22101/ijers.74.23.

Baniamerian, A., Bashiri, M., & Tavakoli-Moghaddam, R. (2019). Modified variable neighborhood search and genetic algorithm for profitable heterogeneous vehicle routing problem with cross-docking. *Applied Soft Computing Journal, 75*, 441–460. https://doi.org/10.1016/j.asoc.2018.11.029

![Fig. 27. Domestic vs Foreign Tourist Visits in India since 2001 with the same year as a baseline.](image-url)
