



# Design and Development of Travel and Tourism Recommendation System using Web-Scraped Data Positioned on Artificial Intelligence and Machine Learning

Krishan Kumar<sup>1</sup>, Vivek Birla<sup>2</sup>, Allan Sasi<sup>3</sup>, Ayush Deep<sup>4</sup>

<sup>1234</sup> Department of Computer Science & Engineering, Amity School of Engineering & Technology,  
Amity University Haryana, India

Email: [krishanchhillar@gmail.com](mailto:krishanchhillar@gmail.com), [vbirla@ggn.amity.edu](mailto:vbirla@ggn.amity.edu), [allansasi98@gmail.com](mailto:allansasi98@gmail.com),  
[deepayush898@gmail.com](mailto:deepayush898@gmail.com)

## ABSTRACT

Internet is filled with innumerable websites offering the desired data, like here the information about flights and hotels, their reservations, also recommendations built on the searches performed for the dates and locations. Whereas, if one visits to other travel websites, they must be having enormous nominations, but it will be difficult to select one competitive price, he must scroll through all of these websites analyzing each suited offer and hence, it becomes a tedious task. This web-based application presents the user with hand-picked cheap and cheerful choice. With a developed strategy for extracting web information, an important method using python scripts, this process becomes very much simpler. Well, Travelling is a unification of an expedition, exploring delightful time, residing at a new location under some different weather conditions, celebrating events, trying new cuisines and many other experiences. To enhance such episodes of life, we comprehensively look for some assistance, like, scheduling a journey. Currently, tourism concerned guidance obtainable on the internet is inexhaustible, but scouting those acceptable getaway plans, and related services may be a prolonged process. While a recommender engine or system can be supported for diverse travel queries, for instance, prime stations for a vacation, favorable weather or a brisk conveyance. An artificial conversation module, a chatbot for this purpose empowers clients to ask their queries regarding their trips, saving their prestigious time, as they can collect the details during their textual communication. So, in this paper, an efficacious study on travel and correlated characteristics like flight and hotel bookings, along with some packages and interpreting recommendations as per users' heterogeneous preferences.

**Key words:** Tourism Recommender System, Chatbot, Web Scraping, Artificial Intelligence, Machine Learning, Python, NLP.

## 1. INTRODUCTION

Modifications and developments are nimbler than ever, with amending analytics and evolving digital world producing latest travel styles. Modern clients are more economically awoken than earlier, and contemporaneous fashion mirrors that.

In the term of tourism, customers' cognitive processes, potential clients can approach web as a scheme for reservations, for eruditions and further recommending some guidance. Nowadays, the tourism service is a vital sector in the developing countries. Being a significant job and income provider, this calls for some great modernization and aftercare. The tourism industry, thus, is a zestful field with the incessant demanding customers. The competitors in this discipline require to face the challenging vogue in this industry, the preferences of the frequenters and yet on their past capabilities, to achieve satisfactory levels as the time advances. Therefore, apparently the institutions in travel sector involve wealthy data from collective founts and a well-defined process to analyses that facts for further recommendations. One of the widely embraced techniques for accumulating preferred information is data scraping from the websites available on the internet. Data web scraping is a constitutive segment of all industries and their businesses that trails an erudition acting upon data.[1]

World Travel and Tourism Council's (WTTC) contemporary research alongside Oxford Economics, reveals that the Travel & Tourism sector cracked 3.5% heightening in the past year, surpassing the world's wealth of around 2.5% for the ninth successive term. Over recent years, one-fourth of introduced vacancies were generated by this sector, promoting it as the leading associate of government in inducing employment as well as revenue.[2]

Astray droids, artificial intelligence (AI) is bustling in more actions too. Feasibly, the frequent application within this industry as a help line, with chatbots procuring such intelligence with prompt feedback to all the questions. It is further adept to ceaselessly grasp from the conversations. These AI machines are transforming to the peak. Different

sorts of machine intelligences are currently managed over the tourism sector. The aforesaid chatbots, with its proficiency to automate and simplify most of its sales in conjunction with the customer feedback services. As the clients hope for arripidid appropriate responses, these bots assist the users 24x7. Whilst the constraints in their performance, interactive bots can yield a plenitude of serviceable support for the end users. Other than this, the application of machine learning enables these mechanisms to master all probable discussions and perpetually enhance itself. AI is also extensively implemented in data gathering and its interpretation, accompanying fundamental tools for elevating ability, dependability and satisfaction of clients.[3]

### 1.1 Motivation

Obligations, intentions, and enthusiasm are the driving forces of a person's actions and they render a pivotal share in the channel of tour and travels.

The economy is multiplying in association with tourism sector. But this zone always needs improvements, thus, needs to be reconstructed. It should be zealous inside the nation itself. Information and Communication Technologies (ICT) are examining to authorize this progression, providing tremendous cooperation respecting the transforming formula of tourism, hence, proffer some additional pleasing travel practices to a huge scale of vacationers, who are growingly inviting more and more customized tour-plans. Customers wait for an admissible and appealing information that concurrently is faithful and handy. A quick-witted traveller expediting system alleviates a venue much unchallenging ever more and plan reversals would turn down. In addition, an uncomplicated user interface will be preferable for a productive usage of this service.[4]

MakeMyTrip.com -Designed to accredit the holidaymakers with a swift reservation availing them all the possible alternatives. This organization began its tour in the US-India commuter's market. It seeks to recommends a collection of finest services along with the innovative technology that has an exclusive continuous product service. MakeMyTrip extends its multidisciplinary choices in travel products and services and hence, it is the bullish front runner in its industry with 47% of market share.[5]

This knowledge domain has constituted a significant march into the technology. There are many suggestions and conclusions, that are influenced by the conversations between two or even by using one such software application. Therefore, gearing towards some supplement studies and nurture a supposition, especially more understandable, a parallel comparison is to be authored on the most frequently seen approach and estimating outcomes are also presented in this discussion.

### 1.2 Objectives

The paper revolves around an intelligent web-based travel and tourism recommendation application which suggests

tourism packages from selective famous travel websites which permits in scraping their contents. According to the user-entered details, real-time flight and hotel data will be fetched from multiple websites and will be stored to a database. The best, cheaper priced options will be filtered and displayed to the users. Also, a chatbot will be deployed to assist the users to look for their reservations, and further receive some relevant tour package recommendations too.

## 2. BACKGROUND AND RELATED WORK

### 2.1 Web Scraping

Online network takes the lead in providing the data. It has been a prolonged tactic to derive some desired information from multiple sources scattered all over the internet, but there were myriads of procedures, so none were recognized at the beginning. Gradually, these data extraction processes unfolded, and advancements came in the technologies, thus method of web scraping was discovered.[6]

The computerized harvesting of archives from the information highway is just as matured as the online world itself. However, web scraping by no means is another word, earlier, the conduct turns out to be better known as data mining, screen-scraping, web harvesting, etc. In this thesis, web scraping is the procedure of assembling the precise data with all methods except for a program engaging with an Application Program Interface (API, or, undeniably by one employing an internet browser). A sound way is by coding an automated script that appeals a web server, calls for data (often organized as HTML and similar files that comprehends web pages), and then construes the data to unsheathe the demanded queries. In practice, web data extraction embraces a great diversity of planning approaches as well as technologies, such as data mining, analysis over it and on top of that protection of data, i.e. related to cyber security.[7]

As an alternative to visualisation of sequential pages via small screen, one can directly access the databanks covering millions of webpages at the very spot. Likewise, web scrapers can reach-out sites that conventional searchers cannot find. A Google search for "Cheapest flights to Delhi" will conclude a heap of promotions and favoured flight portals. Google merely realizes as these pages exhibits, not the precise impacts of separate queries inputted in a flight search engine. Anyhow, a sophisticated web harvester can tabulate the expenditure of a flight to Delhi after a while, crosswise a broad range of sites, and recommends the best suited to price.

Well, APIs mayhap remarkable, if a suitable one is found. They might grant a favourable string of proportioned figures from a webserver to another one. Usually, it is advantageous to go with an existing API, instead of building a bot for the same. Still there are diverse causes due to which an API does not subsist: The origin site might not be having the framework or troubleshooting capability to establish an API. Albeit an API exists, the bid volume and bandwidth, the

datatype, or some other data related parameters that it allocates might be short for the motives. There comes web scraping, accompanied by some anomalies, within web browsers, this can be acquired via a Python script. So, when a webpage is retrieved, the complete data could be preserved into a Comma-Separated-Values (CSV) file or other preferred file formats as in a database. And the stored data can be further analysed and be implemented to many more researches.

### 2.1.1 Techniques

- i. The standard Ctrl C and Ctrl V: The users' copy-paste method, also a human trial, is the prime practicable data retrieval strategy. Yet, it may cause inattentive fallacies, and obviously is a monotonous task when the netizen needs to inspect the page and accumulate loads of data and compile it as dataset/s.[8]
- ii. Ubuntu and regular expressions: A uninvolved and vigorous proposal to collect data from different websites. This method of working is based on some UNIX commands or regular expression pairing possibilities of this programming language.
- iii. HTTP method: This is a way of extracting information from static and dynamic contents of websites. These specifics shall be derived by running the Hyper Text Transfer Protocol (HTTP) request calls to the remote located webservers using the socket programs.
- iv. HTML Parsing method: The partially structured queries, like Hypertext Query Language (HTQL), and could be deployed to parse the Hypertext Markup Language (HTML) webpages to claim the data.[6]
- v. Software: There exists many appreciable tools useful for personalizing the web data extracting provisions. This software tool perchance aims to systematically realize the collected facts, built of the page or propound an interface that eliminates the relevance to write computerization code, or a remarkable script that can allow pulling out of content and transfer them to the designated databases.
- vi. Computer vision scanners: Implementing machine learning together with computer vision, seek to detect and excerpt the piece of information from the visited websites by elucidating the webpages observable by the intelligence of device but not by a human.

### 2.1.2 Relevant Python Libraries

Comparatively, all the recent web data extraction programs running on the net are just a variation of automated bots in practical terms. Moreover, these intelligent scripts are liable for deducing the markup language of an internet site and then bringing altogether into a well-ordered data.

Requests library - The most introductory library offered by python for this impulse. It is ideally suited for possessing the matter from a mere non-recurrent or static webpage only. In another words, it disapproves to scrape the material from an asynchronous web application alike AJAX (Asynchronous

JavaScript and XML). With a beautifully designed documentation, the library leads that, it encloses supremely standard authentications to pitch into the domains. In summation, this directory streamlines by establishing HTTP requests or HTTP(S) proxy supports, nonetheless, it could not parse through the formerly retrieved HTML contents. For HTML parsing, Lxml library and its enhancement, BS4 package are introduced later. And Requests package could not get pass the JavaScript (JS) enabled sites. Also, contrary to other libraries, the overall processing of this turns out to be astonishingly fast with minimizing the consumption of CPU and memory.[9]

Lxml library - As already highlighted in the latter part, lxml took effort for parsing the HTML body. With a driven pace in accordance with CSS and XPath selectors, it invigorates to tolerate the HTML. While the endorsed documentations by the developers that may not be self-evident leading to difficulties in recommencement, the sturdiness of element trees therein python enabling it to scrape few large datasets from any systematic HTML web page. Even so, lxml library is confined to the dynamic websites only.

Selenium library - yet another web driver, remarkable as a convivial with a digestible documentation for a novice web scraping user. Bizarrely, selenium was not intended for crawling. Vast majority of the sites are amply implementing JS, this library was configured to parse such JS-validated info from the webpage in a slower but steady method. Parsing of every web-accessible dynamic script implies that it automates the web browser, and thus, accounts an inordinately high memory and unwillingly slower. Therefore, realized that it is unfit for scraping major websites.

Beautiful Soup library - or BS4, designed to extract data from both HTML and XML documents. Along with lxml library, they become congruent to act upon inadequate HTML web-pages, yet slower as compared to the plain lxml, since this collaboration absorbs the profusion of CPU usage.[7]

Scrapy framework - A mature python solution to meet the challenge of fetching asynchronous massive datasets from all kind of websites. Collectively, with an excellent documentation and combination all above mentioned python libraries, in association with the Splash library, they deftly provide the spider crawlers processing multiple HTTP requests in parallel, saving the valuable time for downloading the HTML contents in an average amount of memory, and then go through the raw data for further transfigurations and saves as a purposeful format like CSV, JSON or XML files.

Beyond these popular python libraries, there are few more platforms for such activity – Import.io, a Software as a Service (SaaS) available online for retrieving info and unifying them from a website without programming. This analytical base bestows the user for a prompt product in fraction of minutes or even seconds. The application shows a Point and Click function which makes it possible to identify the precise fields to be worked upon. Once the data is extracted, they are stored

into the affiliated cloud-based server and also the file can be converted into any required file format.[10]

### 2.1.3 Architecture And Process

Comparatively, all the recent web data extraction programs running on the net are just a variation of automated bots in practical terms. Moreover, these intelligent scripts are liable for deducing the markup language of an internet site and then bringing altogether into a well-ordered data.

- i. Primarily, taking advantage of Hypertext Transfer Protocol, a GET request is forwarded to the targeted website.
- ii. Later, the web server handles the request and, when found authentic, the user is permitted to fetch and read the HTML content of that page.
- iii. A data extraction tool identifies the desired contents and assign those values to some variables.

This implies the process as an overview, Although, this elementary practice is further extended to thousands of observations. As these tools being more intricately, the prospects of machine learning and big data evolves with them. Further comprising the commonality with dynamic webpages these days, the harvesters are coerced to rework on their automating tactics in the future.

## 2.2 Recommendation Engine

Many different organised tours are predominantly at service in several kinds. These may comprise conveyance, trips to a list of dream places, accommodations, and caterings in a prearranged scheme. Concerned travellers get in touch with the associated travel agents, and in concordance with the tour intents, number of travellers and some other circumstances, these officials customize the package. To personalize similar journeys, tourists may suppress on varied angles of a visit or might aspire to schedule a trip for themselves. For devising an outing, everyone must web search for some relevant facts about the destinations, and then properly plan the pathways and hauling, furthermore one might collate prices, reviews, offers, followed by booking tickets, registering for rooms and rest of the travel-related chores. Hence, an expert support is always foreseen during such difficult situations. Thus, various e-tourism recommendation services aim at assisting the wayfarers with such forecasts.

However, handful of people favour solo travelling for recreation, to explore, for calmness, still a mass population still prefers to travel in groups, with friends and families. These similar teams may require some recommendations in fortifying their adventures, folks might be pulsed with some new events and build their social life. Hence, a tour module should be prioritized based on several benchmarks.[11]

One of the referred discourses concentrates on individualized travel advices, then illustrate favourable fulfilment by sustaining the open public engaged images.

They pioneered to convoy some adaptive recommendations by an increased study of some users and their properties, like group types, for example –individuals, friends, couples, or families. Although a lot of data is gathered in such way, this leads to some privacy issues.[12]

### 2.2.1 Content-Based Filtering

This sort of filtering technique modelled on the customer's desires and expectations closely to the service. Primarily, these instructions wish encouraging products or related services which are like the past shared experiences of the user.

Here it applies to the details of the product accustomed to the proposals through obtaining linked patterns and ratings of the client, nonetheless the paradigm is conventionally stationed on the consideration of a sole patron as contradictory to all adjudged things.[13]

### 2.2.2 Collaborative Filtering

Collaborative filtering model uses the collaborative intensity of the ratings given by different clients to make suggestions. The aim of collaborative filtering is to foresee how well a user will like a thing that he has not evaluated given a lot of chronicled inclination decisions for the group of users [8].

This proposition creates the mechanism by appraising the user's history and a new analogous verdict contrived by distinct end-users, and then operate the system to reckon the intrigues. Such client-based filtering methods prepares opinions seeing their interests.[14]

Aforesaid percolating model practices the joint potency of the assessments inclined by alternative consumers to make recommendations. The purpose is to understand and anticipate the user's responses on inexperienced guidance.[15]

Stating on a shared cloud computing programme, where Hadoop is applied, to determine the scalability complication of this procedure. Although it is preferred for discovering personalized suggestions based on interests, but the system does not contemplate the related users' activities.[16]

### 2.2.3 Hybrid Filtering

Every filtering method will utilize diverse source of information, and they have distinctive qualities and shortcomings, and appear to be somewhat prohibitive in isolation, particularly when various source of information is accessible. Hybrid recommendation frameworks have been intended to investigate these potential outcomes in which one might want to make utilization of all the knowledge accessible in various information sources and furthermore utilize the algorithmic intensity of different recommender frameworks to make hearty inferences [3].

The amalgam of recommender systems is set up by affixing disparate recommender structure to cast a sturdier plan. By linking certain machines, depletion of the faults with the help of the worthiness of other machines and consequently concoct a powerful mechanism. For interpretation, using combination of collaborative refining procedures, pointing towards the configuration breaks afterwards new items did not undergo

assessments, and then at the hand of content-based mode, in which elemental information about the features are accessible, new components as it may be endorsed with added accuracy and force.

Each filtrate applies multiple data sources, and procuring peculiar conditions and drawbacks, and seems moderately proscribe in solitude, especially when separate origins of data is usable. Hybrid frameworks have been predetermined to explore such possible conclusions to what one maybe obliged to adjust the applications of gross observations, exposed in diversified info and besides the appropriate measurable magnitude of varied schemas to create cordial reasonings.[13]

An in-cross relational recognition ideology for tourist management is contemplated. The mechanism exhibits some POIs to excursionists admitting to their contour and other factors. So, to call and exhort the point of interests, the pair of collaboratives along with the content-based filtering practices are handled. Data discovery and soft computing on the other side, combined techniques are used. The main stream substructure incorporates distinguished users and their history.[17]

#### 2.2.4 Context-Based Filtering

This framework catches other sorts of legitimate info into deliberation, while formulating suggestions which covers the time-related data.

Submitted a strategy, mentioned as context-rank, the data of geo-tagged images is treated to improve the customized recommendations. With these photos and corresponding GPS locus, milestones are discovered and then recognizing every landmark. The indicative labels and pictures are releasing from each sign. In this way, the marks are determined from assigned coordinates of enormous image dispensing websites. Contingent on the perceptible facets, exemplary photographs are retrieved, and characteristic tags are recovered by the designating literal tags. Using the past tour histories of the traveller, visited places and their popularity is surmised. And for dedicated recommendations, collaborative-filtering is being utilised.[18]

A recommending device is a subdivision of data screening system which strives to indicate the fondness that a customer would be giving for some service.

However, standard recommendation algorithm helps the clients with a classified catalogue of recommendations for specific demands, like for books, gadgets, books, flights, hotels, restaurants, and so on. So, looking into the scenario of planning a trip to some place where the user may be curious for some suggestions of their contrasting Point of Interests (POIs), such as number of passengers or guests, rooms required, etc. Thus, a travel recommender mechanism, preferably can learn to serve better from a setup efficient of providing deals correlated in a ranged group instead of directing towards a single perspective, which compounds an enriched tour events for the travellers.[19]

## 2.3 Chatbot

The Artificial intelligence Chatbots have gained increasing importance for research and practice with lots of apps available today.

As passing into the 21st century, encountering changes in our existence, computer aided learning became an essential source of information in the recent research-based events. Here, railway-related details are obtained and stored into system, as a tactical factor during computerization. But these information services are not yet built to perfection. Initially, this paper emphasises on the present circumstance of general information, it was suggested that the intelligence support for the rail sector should boost the productivity and explicitness of data communication and then provide fair options to conduct the data assistance to favour the data transference.[20]

Formerly, Mastercard Inc. has remained as a bridge between the customers and the merchant, and it also connects numerous products with its affiliated financial organization. As a result, the corporation demands for a big strong management wing with some appreciable support forces. Their regimes hoard the data, in the cited dissertation, they also deliver four divergent cases of implementing to the stored data seeking the solutions for their business challenges.[21]

Next observed paper was carried out in as a home automation project which means the monitoring of lights, IoT devices, designated switches, and several other powered appliances in a house with Internet of things which is extensively opted these days. It proposed a web-app with all the inter-connectable gadgets and were operated VI internet. The key specifications of this application are like conversational widget which helps the user efficiently to switch the functioning of their devices at one stop. The conversations between the chatbot and user is dealt using Natural Language Processing (NLP) techniques. Further, all the devices linked in the hosted LAN are controllable through this web application which also includes security options enabling authorization of users in accessing their gadgets. And most necessarily, this mechanism has a service of email alerts as soon as a trespasser is detected with the help of some motion sensing components.[22]

### 2.3.1 NLP

Natural Language Processing (NLP) is a notable branch of AI, it is the transformation process of the spoken languages into machine language to deduce the connotation from the statement. NLP has plenty of good applications in machine interpretations, data extraction, conversational ability, and many other substantial studies.

A book based on NLP and open-sourced python language clarified that computer technology indeed embodies a critique of data structures which could be compiled in all the programming languages. Natural Language Tool Kit(NLTK) as a tool or a platform for NLP to work with spoken language data, it includes several necessary text-processing modules

like for tokenizing words or sentences, handling stop-words, lexical normalizations including stemming, lemmatization, POS tagging, parsing, classification, and semantic reasoning, and thus, complete sentimental processing of the data for its additional application into the statistical NLP.[23]-[25]

### 3. IMPLEMENTATION

As the title exposes, the project hinges on Artificial Intelligence and Machine Learning, so are used for developing automated chat agent which further includes a recommendation system for suggesting the best-suitable for the user.

Primitively, majority of such related works are just based on recommendation system, these do not rely upon the real-time data (live records), rather, they use manually built sample database or by gathering historical data available somewhere and works on it offline which limits the proficiency of projects and thus, their practical implementation on the dynamic demanded data fails.

So, this project overcame the practice of using non-real-time database records for processing, by using the web scraping technique. The project entirely works on online available data and a live database for the recommendation system, which is updated daily. We are presenting a web application for this Travel and Tourism Recommendation System, which will fetch live data for flights and hotels bookings and present tour packages which is available in the market and in view of this, one can book it as per convenience.

The web application serves on four sections :

- i. The first stage contains Flight Bookings, where the user can search for flights to any city in the world and make a booking. The buyer needs to just enter the details like departure city/airport, arrival city/airport and dates for the flight. The main feature of the flight booking, a sub-application, is that it scrapes numerous website in the background such as expedia.com, goibibo.com, etc. through selenium, request and beautiful soup python packages and presents the alive flight records on the web application for the reservations.

For booking purpose, it redirects a link, which will open the exact website in the default browser of your system from where you want to book the flight. You can simply book it from there. The goal of this step is that the user does not have to go through different website looking for the suitable options. One can simply search for maximum 2 to 3 website for the cheap flight rates and booking and options suitable for them, but here application displays records of 5 to

10 websites at once. So, without much effort it scrapes data and display it at a place.

- ii. The second part contains Hotel Bookings, where one can search hotels, for any city in India (presently) and rent it. Users required to fill the input fields like city, check-in and check-out dates, number of guests and number of rooms. Same as the Flight sub-application, it scrapes numerous websites in the background such as makemytrip.com, paytm.com, etc. With the help of selenium, request and beautiful soup python packages and present it real time hotel records on the web application for the reservations.

For booking purpose, it reverts a link, which will open the exact website in the default browser of your system from where you want to book the hotel room. You can simply book it from there. Same as the Flight sub-application, the purpose of this step is that the user does not have to search different website for the various options. One can simply search for maximum 2 to 3 website for the cheap flight rates and booking and options suitable for them, but here application displays records of 5 to 10 websites at once. So, without much effort it scrapes data for you and display it at a place.

- iii. The third section contains Travel and Tourism Recommendation System, where it recommends tour packages in various categories like adventure, trekking, camping, exploring, etc. The recommendation system advices packages to the user according to the user preferences and history. It recommends top 20 packages from its database. The records in database is real time, as the database itself gets updated at every 24 hours.

The recommendation system uses Deep Learning algorithms that is used to improve the result of top 20 Recommendations with every search result.

For the booking of the packages, user needs to choose one and it will redirect you to the website in the default browser of your system.

- iv. The fourth part contains a chatbot, named AmiBot. This bot is built with Artificial Intelligence and Machine Learning, which interacts with the user in a professional way and helps the user to get the desired result, like reserving a flight ticket or a hotel room, etc. The client can also ask the virtual human agent for performing a search operation, but it will ask some questions which should be respond to, like departure city, arrival city and dates for booking a flight, which is inevitable.

Besides the highlighted objectives, MongoDB, a momentous open sourced NoSQL database preferred for this project because-

- It is used in view of its robustness and scalability with no exception.
- MongoDB uses a Role-Based Access Control(RBAC) with an adjustable set of privileges. It provides various security features including authentication, access control, authorization, and auditing.
- It has no predefined schema, so it becomes a versatile data model, and the file hold on to integrate any series of values based on some key.
- Easily shifted to cloud-based services.
- With a high performance and storing data speed, while scraping, we need to delete previously searched data and save a new one with utmost speed, being the main benefit of using MongoDB.

And developed a desktop application using Tkinter, an in-built python module for GUI, it is the most used standard interface-

- This combination is simple to understand, since it is a library with an elegant API, being the prime preference for developing fast interfaces for python scripts.
- Provides various pre-set JavaScript event triggers to handle the render events using its tags, marks, configs() and other binding methods. Such a method can define hyperlinks while delaying the information for executing web requests from user-clicks in both regular and OpenClick() events.
- Further, various widget classes built into the library like frames, canvas for drawing custom widgets, text, several buttons, scrollbar, etc.

We also use multiprocessing and multithreading techniques and even combine them, for optimal utilization of resources in the computers and efficient GUI rendering.

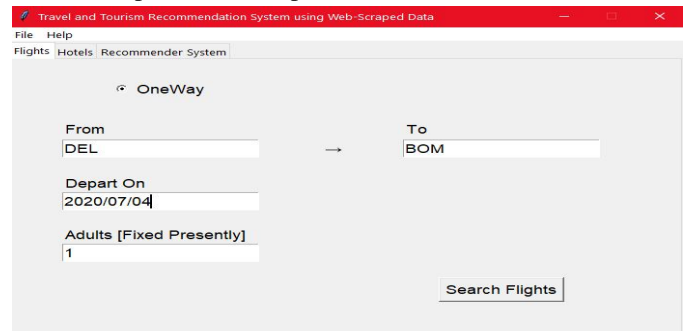
#### 4. RESULTS AND DISCUSSION

As already discussed, there are many tourist related web applications that services with information like flights, accommodations, etc. about several locations, but the key shortcoming of such systems are that they only offer static data that is mostly past info.

After surveying plenty of dissertations and other available online resources related to Web Harvesting using Python programming language, Database Management Systems, Graphical User Interface (GUI), Recommendation Systems

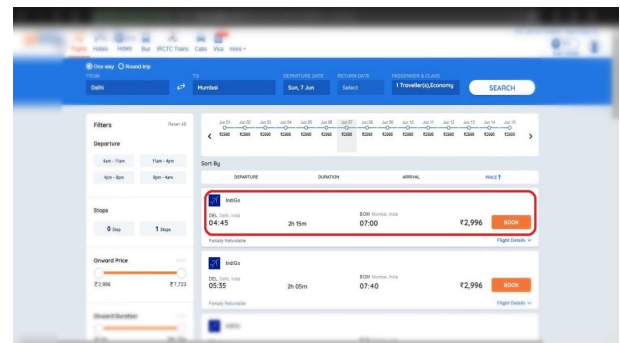
and Chatbots, a prototype was developed accordingly. Though, several faults were realized like periodic changes in patterns of website, the amount of inter-connected details and time management, the cited solutions helped to overcome those issues.

A simple front-end was built with python GUI library called Tkinter. Here is a snapshot of the GUI (Figure-1 Search for Flights), as observable, it consists of three tabs namely Flights, Hotels and Recommender System. The page layout of these tabs was explained in the previous section.



**Figure 1:** GUI Prototype showing only One-Way Flight for an individual Adult Traveller

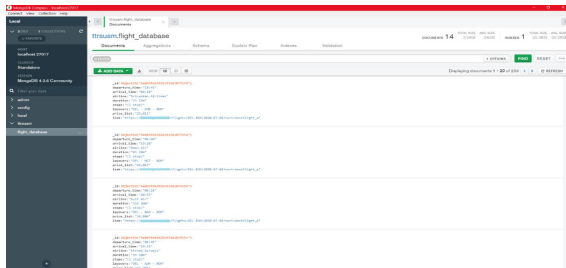
As soon as the user submit the search button after typing in the suitable details, the process begins. The scraper tool traverses through the instructed popular Travel websites and accordingly, fetches the details of available flights and hotels. But for this, some lines of code containing corresponding credentials for the input fields of website are mandatory. Wherefore, this application holds similar text-fields which are inevitable.



**Figure 2 :** Inspecting and extracting information (flights and hotels) from popular sites

If one investigates a frequently visited travel portal (Figure2 - Webpage), it is found that the notable details appears to be Airline, Departure, Arrival and Travel Time for flights and Hotel Name, User Ratings, Facilities offered for hotels, and most importantly the Cost. By examining any webpage, probably one could assimilate that the showcased info might be nested inside the division tags which requires much patience to be found. Following an exhaustive inspection to

capture the tag and further including this parametric tag, helps to dig the entailed info. So, the above disclosed information is to be collected into the database and correspondingly the python file is programmed, importing the vital libraries. As the search button is hit with the user inputs, within few seconds scraper would be ready with all the enjoined facts and figures from the accredited websites.

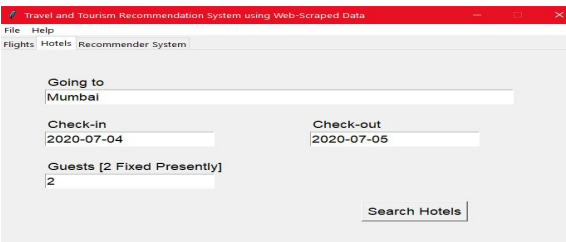


**Figure 3:** Storing the scraped information into a supportive database

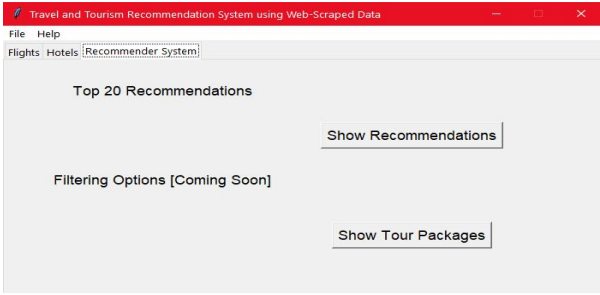
#	Agency_Url String	Airline_Url String	Airline String	Airline String	Class String	Agency_Url String	Airline_Url String
1	"91.91"	"91.91"	"IndiGo Airlines"	"91.91"	"10.000000"	"91.91 - 091.91"	"91.91"
2	"91.91"	"101.91"	"Shree Air"	"91.91"	"10.000000"	"91.91 - 911.91"	"91.91"
3	"91.91"	"101.91"	"Shree Air"	"91.91"	"10.000000"	"91.91 - 911.91"	"91.91"
4	"91.91"	"91.91"	"IndiGo Airlines"	"91.91"	"10.000000"	"91.91 - 091.91"	"91.91"
5	"91.91"	"91.91"	"IndiGo"	"91.91"	"10.000000"	"91.91 - 091.91"	"91.91"
6	"91.91"	"91.91"	"IndiGo"	"91.91"	"10.000000"	"91.91 - 091.91"	"91.91"
7	"91.91"	"91.91"	"IndiGo"	"91.91"	"10.000000"	"91.91 - 091.91"	"91.91"
8	"91.91"	"91.91"	"IndiGo"	"91.91"	"10.000000"	"91.91 - 091.91"	"91.91"
9	"91.91"	"91.91"	"IndiGo"	"91.91"	"10.000000"	"91.91 - 091.91"	"91.91"
10	"91.91"	"91.91"	"IndiGo"	"91.91"	"10.000000"	"91.91 - 091.91"	"91.91"
11	"91.91"	"91.91"	"IndiGo"	"91.91"	"10.000000"	"91.91 - 091.91"	"91.91"
12	"91.91"	"91.91"	"IndiGo"	"91.91"	"10.000000"	"91.91 - 091.91"	"91.91"
13	"91.91"	"91.91"	"IndiGo"	"91.91"	"10.000000"	"91.91 - 091.91"	"91.91"
14	"91.91"	"91.91"	"IndiGo"	"91.91"	"10.000000"	"91.91 - 091.91"	"91.91"
15	"91.91"	"91.91"	"IndiGo"	"91.91"	"10.000000"	"91.91 - 091.91"	"91.91"
16	"91.91"	"91.91"	"IndiGo"	"91.91"	"10.000000"	"91.91 - 091.91"	"91.91"
17	"91.91"	"91.91"	"IndiGo"	"91.91"	"10.000000"	"91.91 - 091.91"	"91.91"
18	"91.91"	"91.91"	"IndiGo"	"91.91"	"10.000000"	"91.91 - 091.91"	"91.91"
19	"91.91"	"91.91"	"IndiGo"	"91.91"	"10.000000"	"91.91 - 091.91"	"91.91"
20	"91.91"	"91.91"	"IndiGo"	"91.91"	"10.000000"	"91.91 - 091.91"	"91.91"

**Figure 4 :** Viewing the gathered details in a tabular form

Immediately after the execution of the automatized program, the info is retrieved. But these elements need to be stored in an elucidated style. Undoubtedly, an appropriate database (Figure 3, 4 - MongoDB, a NoSQL database) is established for this motive. The file scheme absolutely counts on the forthcoming works. Likewise, one can search for best discounted hotels, and within seconds, the scraping tool will gather the tallying information and display it on the hotel’s screen(Figure 5. – Search for Hotels).



**Figure 5:** GUI Prototype screen showing Hotel for two guests



**Figure 6:** GUI Prototype screen for Recommender System

The underlying idea is to enlighten people explore new places determined by their dreams. AmiGo assisting chatbot will have a prominent characteristic in processing and delivering the user’s preference. Hence, the recommendation engine(Figure 6 -GUI screen) reduces the burden on the user by counselling the flights, hotels and a complete package that best fits the predilections by spotlighting them. Henceforward, various other traits and advanced methodologies could be analysed and developed for a gratifying implementation of such systems.

**5. CONCLUSION AND FUTURE WORK**

This paper expounds the Travel and Tourism Recommendation System, an application that facilitate its users to search for and receive recommendations for tour packages as per their preferences. Thus, the recommender system reduces the trouble of user by recommending them with a package holiday that stressbusting a cheaper price. The effect of various guidelines on ameliorating the recommendation process is therefore currently of major interest.

- i. Using Machine Learning Algorithms on Recommender System to improve its user database and to understands user better.
- ii. Using Deep Learning Algorithms on all user database at once, for classifying and allocating each user into sub-class of Tour Categories to identify each user and its personality type for tours he chose.

For Example - Suppose one can classify a user into 30 % Adventure type, 40 % Trekking Type, 10 % Exploring Type and 20 % Religious Type. It means user has some mixed type personality with different priorities. Like he chooses Trekking Packages more often than Exploring or Sight Seeing Packages. It results into more accurate search results for the Recommendation System.

Implementation of quarterly systems means how often the user choose to go on tours and in which months. Using this



data, results in improving in results of top-20 recommendations by the system.

Improving AmiBot, to make it more interactive, instead of conversing in a professional way, can also chat in a user-friendly way, which make the user feels like they are chatting with a person sitting on the other end of network.

Using AmiBot chats' database and applying NLP, AIML technology on it to know the user and their preferences much better.

Using Hotels and Flights APIs (if available for the favourable websites), so instead of scraping website, which is quite slow, APIs can be used for faster results.

## REFERENCES

- [1] D. R. Fesenmaier, Z. Xiang, B. Pan, and R. Law, Information and Communication Technologies in Tourism 2010, no. January, 2010.
- [2] "Economic Impact | World Travel & Tourism Council (WTTC)." [Online]. Available: <https://wttc.org/Research/Economic-Impact>. [Accessed: 08-May-2020].
- [3] "Travel Trends: 12 Opportunities for the Travel Industry for 2020." [Online]. Available: <https://www.revfine.com/travel-trends/>. [Accessed: 08-May-2020].
- [4] "Tourism in India - Wikipedia." [Online]. Available: [https://en.wikipedia.org/wiki/Tourism\\_in\\_India](https://en.wikipedia.org/wiki/Tourism_in_India). [Accessed: 08-May-2020].
- [5] "about us | www.makemytrip.com." [Online]. Available: [https://www.makemytrip.com/about-us/company\\_profile.php](https://www.makemytrip.com/about-us/company_profile.php). [Accessed: 08-May-2020].
- [6] "Web scraping - Wikipedia." [Online]. Available: [https://en.wikipedia.org/wiki/Web\\_scraping](https://en.wikipedia.org/wiki/Web_scraping). [Accessed: 08-May-2020].
- [7] P. Ashiwal, P. Tripathi, and R. Miri, "Web Information Retrieval Using Python and BeautifulSoup," Int. J. Res. Appl. Sci. Eng. Technol., vol. 4, no. VI, pp. 335–339, 2016.
- [8] S.C.M. de S Sirisuriya, "A Comparative Study on Web Scraping," 8th Int. Res. Conf. KDU, no. November, pp. 135–140, 2015.
- [9] "Complete Guide to Python Web Scraping Libraries and Frameworks." [Online]. Available: <https://info.scrapinghub.com/web-scraping-guide/python-web-scraping-libraries-and-frameworks>. [Accessed: 08-May-2020].
- [10] A. V Saurkar and S. A. Gode, "An Overview On Web Scraping Techniques And Tools," Int. J. Futur. Revolut. Comput. Sci. Commun. Eng., pp. 363–367, 2018.
- [11] M. V. Rohit, R. M. Manju, S. Trishalaa, R. B. Ayas, and J. S. Nirmala, "Travel Package Recommendation System," SRELS J. Inf. Manag., vol. 55, no. 1, p. 34, 2018, doi: 10.17821/srels/2018/v55i1/113934.
- [12] Y. Y. Chen, A. J. Cheng, and W. H. Hsu, "Travel recommendation by mining people attributes and travel group types from community-contributed photos," IEEE Trans. Multimed., vol. 15, no. 6, pp. 1283–1295, 2013, doi: 10.1109/TMM.2013.2265077.
- [13] I. Ryngksai and L. Chameikho, "Recommender Systems: Types of Filtering," vol. 3, no. 11, pp. 251–254, 2014.
- [14] D. Das, L. Sahoo, and S. Datta, "A Survey on Recommendation System," Int. J. Comput. Appl., vol. 160, no. 7, pp. 6–10, 2017, doi: 10.5120/ijca2017913081.
- [15] E. Ashley-Dejo, S. M. Ngwira, and T. Zuva, "A context-aware proactive recommender system for tourist," Proc. - 2016 3rd Int. Conf. Adv. Comput. Commun. Eng. ICACCE 2016, pp. 271–275, 2017, doi: 10.1109/ICACCE.2016.8073760.
- [16] Z. D. Zhao and M. S. Shang, "User-based collaborative-filtering recommendation algorithms on hadoop," 3rd Int. Conf. Knowl. Discov. Data Mining, WKDD 2010, pp. 478–481, 2010, doi: 10.1109/WKDD.2010.54.
- [17] G. Fenza, E. Fischetti, D. Furno, and V. Loia, "A hybrid context aware system for tourist guidance based on collaborative filtering," IEEE Int. Conf. Fuzzy Syst., pp. 131–138, 2011, doi: 10.1109/FUZZY.2011.6007604.
- [18] K. Jiang, P. Wang, and N. Yu, "ContextRank: Personalized tourism recommendation by exploiting context information of geotaggedweb photos," Proc. - 6th Int. Conf. Image Graph. ICIG 2011, pp. 931–937, 2011, doi: 10.1109/ICIG.2011.48.
- [19] K. J. Oh, Z. Kim, H. Oh, C. G. Lim, and G. Gweon, "Travel intention-based attraction network for recommending travel destinations," 2016 Int. Conf. Big Data Smart Comput. BigComp 2016, pp. 277–280, 2016, doi: 10.1109/BIGCOMP.2016.7425927.
- [20] C. Guo, H. Wu, Z. Sun, and Z. Guan, "Study on the public information service systems of railway industry," Proc. Int. Conf. E-bus. E-Government, ICEE 2010, pp. 3067–3070, 2010, doi: 10.1109/ICEE.2010.773.
- [21] R. S. Arvapally, H. Hicsasmaz, and W. Lo Faro, "Artificial intelligence applied to challenges in the fields of operations and customer support," Proc. - 2017 IEEE Int. Conf. Big Data, Big Data 2017, vol. 2018-Janua, pp. 3562–3569, 2017, doi: 10.1109/BigData.2017.8258347.
- [22] C. J. Baby, F. A. Khan, and J. N. Swathi, "Home automation using IoT and a chatbot using natural language processing," 2017 Innov. Power Adv. Comput. Technol. i-PACT 2017, vol. 2017-Janua, pp. 1–6, 2017, doi: 10.1109/IPACT.2017.8245185.
- [23] S. Bird, E. Klein, and E. Loper, Natural Language

Processing with Python. O'Reilly Media.

- [24] W. Aprianto and W. Kosasih, "Analysis and Design of Web-Based Knowledge Management System for Real Estate Property" Int. J. of Advanced Trends in Computer Science and Engg. vol 9, No-1, pp. 1-7 2020  
<https://doi.org/10.30534/ijatcse/2020/01912020>
- [25] A. Mailangkay, E. Indrajit, R. Kosala and A. Hidayat, "Analysis of the factors that affecting Intention to use Tourism Online Booking" Int. J. of Advanced Trends in Computer Science and Engg., vol. 9, no. 2, pp. 991–996, 2020.  
<https://doi.org/10.30534/ijatcse/2020/14922020>