The Role of Data Technologies with Machine Learning Approaches in Makkah Religious Seasons

Waleed Al Shehri

Department of Computer Science, College of Computer in Al-Lith, Umm Al-Qura University, Makkah, Saudi Arabia

Abstract
Hajj is a fundamental pillar of Islam that all Muslims must perform at least once in their lives. However, Umrah can be performed several times yearly, depending on people's abilities. Every year, Muslims from all over the world travel to Saudi Arabia to perform Hajj. Hajj and Umrah pilgrims face multiple issues due to the large volume of people at the same time and place during the event. Therefore, a system is needed to facilitate the people's smooth execution of Hajj and Umrah procedures. Multiple devices are already installed in Makkah, but it would be better to suggest the data architectures with the help of machine learning approaches. The proposed system analyzes the services provided to the pilgrims regarding gender, location, and foreign pilgrims. The proposed system addressed the research problem of analyzing the Hajj pilgrim dataset most effectively. In addition, Visualizations of the proposed method showed the system's performance using data architectures. Machine learning algorithms classify whether male pilgrims are more significant than female pilgrims. Several algorithms were proposed to classify the data, including logistic regression, Naive Bayes, K-nearest neighbors, decision trees, random forests, and XGBoost. The decision tree accuracy value was 62.83%, whereas K-nearest Neighbors had 62.86%; other classifiers have lower accuracy than these. The open-source dataset was analyzed using different data architectures to store the data, and then machine learning approaches were used to classify the dataset.

Keywords:
Hajj Pilgrims, Data Analysis, Big Data Technologies, Data Architectures, Data Mesh, Data Fabric, Machine Learning.

1. Introduction

Makkah, located in the Kingdom of Saudi Arabia, has become a religious tourism destination for Muslim visitors worldwide to perform religious rituals in holy places there. In addition, almost all modern pilgrims are interested in visiting historical religious monuments and sites on their journey to perform the rites of Hajj and Umrah [1]. Despite the great efforts and the unlimited technical and logistical support provided by the government of the Kingdom of Saudi Arabia and the concerted efforts of all ministries to manage these vast crowds of Muslims coming to perform the rituals, activating the role of some methods and strategies of modern information technologies may contribute significantly to advancing this organizational process to broader horizons in terms of governance and management of big data for pilgrims and Umrah performers for a successful religious journey without any obstacles.
The Hajj procedure is a lengthy process with various steps. Pilgrims worldwide require Hajj services, which should be available online through a single distributed repository. Therefore, the following benefits are obligated:

- Hajj Regulations for All World Muslims.
- International and Domestic Pilgrim's Rights and Regulations.
- Hajj procedure and basic service amenities.

These services are required for every pilgrim who wishes to perform Hajj and must be obtained online and with high-speed access. The centralized data repository and distributed network can help pilgrims and administration get these services in a timely manner. However, because the data set is large enough to make the pilgrimage, big data techniques may play an important role in managing the massive data volume with the analysis of these proposed techniques that best fit this problem [2]. A data lake, data mesh, or data fabric can be helpful when the dataset is large. They should be low-cost and feature-rich. Furthermore, the Hajj procedure necessitates a secure, distributed repository and is quick. As a result, there is a need to create a system that will aid in the smooth execution of the Hajj process [3]. Multiple devices have already been installed in Makkah, but it would be advantageous to suggest data architectures using machine learning approaches with other cutting-edge technologies that may be used to collect and manage data. [4]. The remaining portion of this study is divided into the following sections: the second discusses existing approaches, and the third discusses the proposed method and methodology. Finally, Section 4 discusses the findings and conclusions.

Manuscript received August 5, 2023
Manuscript revised August 20, 2023
https://doi.org/10.22937/IJCSNS.2023.23.8.4

2. Literature Review

This section discussed the existing approaches related to the problem. Many techniques were already applied, but a few latest methods are discussed below in detail.

This article [5] focused on the control development of using a large crowd dataset, particularly those at Hajj and Umrah. Video surveillance and analysis have grown in significance as a means of enhancing the matters of security of pilgrims in Makkah, Saudi Arabia. Since so many people are attending Hajj, which is thought to be very unusual, it was not easy to analyze video data precisely. Where computer vision applications were used to analyze the pilgrims perfectly, the CNN approach was used in the proposed system to analyze the pilgrims. The study also suggested a method for measuring and determining crowd density. The technique was proposed for analyzing and estimating crowds based on the deep-learning approaches in the proposed system. CNN has many pros compared to the other failed models, and CNN could solve the problems. The research study provided a convolutional neural network-based paradigm for crowd counting (CNNs). In hitherto unexplored target situations, the aim is to assess crowd video maps. The following obstacles must be overcome to succeed: The complexity of widely used multiplicity analysis is insufficient to support the comparison. The model used a structure to identify each person in the crowd. With an astounding Mean Absolute Error of 200 and Mean Square Error (MSE) of 240, this method outperforms the industry norm (an average of 135.54 improvements). The conclusion was it has a density map and prediction results from various conventional methods in the proposed system's new dataset HAJJ-Crowd for assessment and testing.

Internet of Things (IoT) technology is associated with many industries, including the scientific, commercial, industrial, health, and transportation sectors [6]. In the current era, where we progressively rely on, some essential technological sectors include AI, Big Data, IoT, and many more. This study is considered extremely important and urgent since it pertains to around 20% of the population, specifically Muslims, who compete each year to be chosen to lead Islamic rites. Based on the study field, research instrument, statistical approach, data type, data collection techniques, and research citations. The paper recommends new research and practical directions based on a thorough analysis of recent technologies. The analytical survey is crucial to consider improving Hajj and Umrah services in light of current technology and discussing other outstanding issues that need further in-depth research. There are four components to the research methodology for this work. The first stage studies earlier studies on technology and its uses. The studies already classified in stage one were divided into six additional categories in stage two: research field, research instrument, statistical approach, kind of statistical data, data collection tools, and research citations. The results are then thoroughly presented and discussed separately during Stage 3. Finally, Stage 4 examines applications and modern technologies that might be used in the Hajj and Umrah systems, which are strongly advised and thoroughly explored, and other open issues that need more in-depth research. It is recommended that research be done on how IoT smart devices that employ NFC might help Hajj and Umrah decision-makers. Additionally, monitoring systems are being improved for disgraceful conduct. Several research projects may also be conducted to develop ground-breaking smart building strategies for human grouping.

This article [7] recommended boosting Hajj services by advancing exploratory data visualization technologies. The proposed system showed how the most recent approach to data visualization varies from conventional explanatory graphics and why it is necessary as a current, efficient exploratory decision-making approach from a large amount of data (big data) collected from all Hajj agencies. We may utilize data visualization to gather and communicate understanding by providing some real-world instances of examining global experiences. For example, to represent the results of a study on what pilgrims typically do while camping in Mina, we use the method to visualize some actual data from the Hajj season of the previous year. The research study focused on the technique to identify the everyday habits among these pilgrims based on their nationalities averaged over the preceding three years. We also offer a straightforward theoretical scenario to understand further the connection between the various pilgrim nations and their homes in Makah near the big mosque.

Another example is using a virtual map of Mina. The natural way to use big data with decision-makers,

particularly for Hajj services, is to use data visualization to study the data. Each service map has to have more in-depth information to emphasize the demands that help to enhance the service. This study established that familiar map figures and graphics lack the supporting data required to take the proper action in a timely way because too much data can prevent decision-makers from coming to the optimal conclusion that would enhance the services. To improve services for Hajj, it is hoped that this data visualization technique will bring together several entities that do not yet have interactive scientific references.

Many infrastructures, population dynamics, and service-related challenges continue to make it challenging to manage crowded events effectively [8] despite the frequent occurrence of disasters like fires, riots, and stampedes resulting in significant situations. There are no clear principles or laws for managing and controlling crowds. Every year, Muslims from all over the world assemble at Makkah to perform it. This essay explores crowd analytics by examining sizable pilgrimage groups participating in the yearly Hajj event. A technological answer for Hajj has been proposed: crowd analytics. These include, for instance, mobile crowd sensing and computing and unmanned aerial vehicles (UAVs). This research did explore trends and issues in crowd management and crowd monitoring during Hajj. Studies on urban analytics for Hajj have recently gained increasing popularity. We examined several Hajj-related tactics in this study. Computer vision-based techniques, such as CCTV security cameras that transmit audio and video, were used during Hajj. Additionally, many studies used FLIR techniques to determine crowd density and respond appropriately.

Along with the first, several location-based tactics were proposed, including inserting RFID chips in pilgrims' phones so their whereabouts could be followed up. There are particular crowd management tendencies in data and information gathering for the Hajj environment. Because Makkah is located in a valley surrounded by mountains, notably Arafat Mountain, which pilgrims must visit to complete their Hajj, recent developments in UAV technology suggest they may be considered remote monitoring of difficult circumstances. Although their images are not steady, UAVs can visit new places without installation.

There are particular crowd management tendencies in data and information gathering for the Hajj environment. Because Makkah is located in a valley surrounded by mountains, notably Arafat Mountain, which pilgrims must visit to complete their Hajj, recent developments in UAV technology suggest they may be considered remote monitoring of difficult circumstances. Although their images are not steady, UAVs can visit new places without installation. Research has not yet been done to identify whether MERS-CoV knowledge, attitudes, and practices of pilgrims on the annual Hajj pilgrimage need to be altered. In a cross-sectional survey research project, convenience sampling was employed. Participants have experienced Muslims who spoke Arabic and English. They were 35 (24.5e43.5) years old on average. The average knowledge and attitude ratings among pilgrims about MERS-CoV were 5 (IQR 4e7) and 6 (IQR 5e7), respectively. Although education level and job position were positively connected were significantly correlated with gender and age (P 0.05). Significant correlations between knowledge, attitude, and behavior were found. More excellent knowledge was found to predict positive practice. These findings will give intervention designers concepts about spreading specific messages to improve MERS-CoV practice, change attitudes, and raise understanding [9]. Muslims travel to Makkah for the pilgrimages of Hajj and Umrah [10]. Hajj is recognized as the most significant event in the world. Every adult Muslim in good health and has financial resources must complete Hajj and Umrah at least once. Traveling around 15 kilometers, Hajj pilgrims, also known as Hajjis, can accomplish so on foot, by bus, or by train. Hajj has performed by around 3 million Muslims annually, and the Kingdom of Saudi Arabia wants to see this number rise. Since 1989, professionals in various ICT sectors have tried to apply their expertise to address issues and challenges related to

Hajj and Umrah. Several papers have examined attempts to employ technology in Hajj and Umrah research, but none were exhaustive. In this study, I grouped research projects that have used ICTs to address issues with Hajj and Umrah during the preceding 33 years into ten categories depending on their relevance. Hajj has performed by around 3 million Muslims annually, and Saudi Arabia wants to see this number rise. Since 1989, professionals in various ICT sectors have tried to apply their expertise to address issues and challenges related to Hajj and Umrah. Several papers have examined attempts to employ technology in Hajj and Umrah research, but

none were exhaustive. In this study, I grouped research projects that have used ICTs to address issues with Hajj and Umrah during the preceding 33 years into ten categories depending on their relevance. These are fantastic guidelines for more research in this field. Researchers have attempted to overcome these problems with cutting-edge technology for decades in various ICT sectors. Although few publications have looked into technology use during Hajj and Umrah in the past, none of them was comprehensive. The papers under consideration tried to use various ICT solutions to tackle Hajj and Umrah difficulties. Based on their usefulness and necessity, I divided past research studies in the literature into ten categories for this study. Additionally, I divided the study projects into 33 categories depending on the technology employed to address Hajj and Umrah issues. Finally, I discussed open problems and new technologies that may be used to solve these problems to provide nine directions for future study in this field.

3. The Proposed Methodology

This stage of the proposal will discuss the architecture proposed for the solution of the above case study using the given technologies. We proposed a technique using real-time data based on a recent research study. The dataset will be extracted from the "Hajj" season records. We will use data acquisition and analytics and then apply Machine Learning techniques to get the optimum solution in real-time.

3.1 Input Data

This step represents how we can get the data; the data resources are shown in Figure 1. The data can be acquired in three ways: using IoT Devices or sensors, and these devices are specifically for getting the health records and security purpose data. We can collect data from external resources for the people using their complete visa process and Air Tickets records. Third, we can get the data using internal resources such as Hajj Process and restaurant or hotel reservations. The data which will be collected is entirely raw, and it may be in any form.

3.2 Data Lakes

We collected the raw data from the different resources in three forms: structured, unstructured, and semi-structured data. Data Lakes will be stored in emails, PDFs, social media tools, images, videos, etc. Data Lakes is a storage that can store data like Big Data storage, but it stores more data than Big Data at a low cost.

3.3 Data Analytics

After getting the data, the given data is not in the form of processing, so we need to pre-process the data. This data will be pre-processed and can be set in the form to pass to the model. Here we build the model using machine learning algorithms and deep learning algorithms. There is also a need to train the model using the previous records and then test the model using new data.

3.4 Key Business Insights the Data

At this stage, we will represent the results from the model, and the graphical representation shows the model building, training, and testing. This step gives the final insights from the data.

3.5 Data Mesh

Data Mesh can be used to take the operational data as input, and it will be performed using ETL (Extract, Transform, and Load) the data into the proposed model. In the end, after performing ETL, the data will be shown as analytics.

3.6 Data Fabric

Data perform essential reporting and services such as facilitating the pilgrims properly using their techniques. While data comes from the "Input Data," which has various types of data such as files, logs, emails, social media data, etc. After that, Data Lakes, Data Mesh, and Data Fabric technologies are used to store and manage the data. Finally, all the records go to Data Analytics, where the data analysis is performed, and the results are shown in the next phase. Machine learning algorithms are used to achieve the classification purpose. Logistics regression, Naïve Bayes, K-nearest Neighbors, Decision tree, Random Forest, and XGBoost algorithms were proposed for data classification.

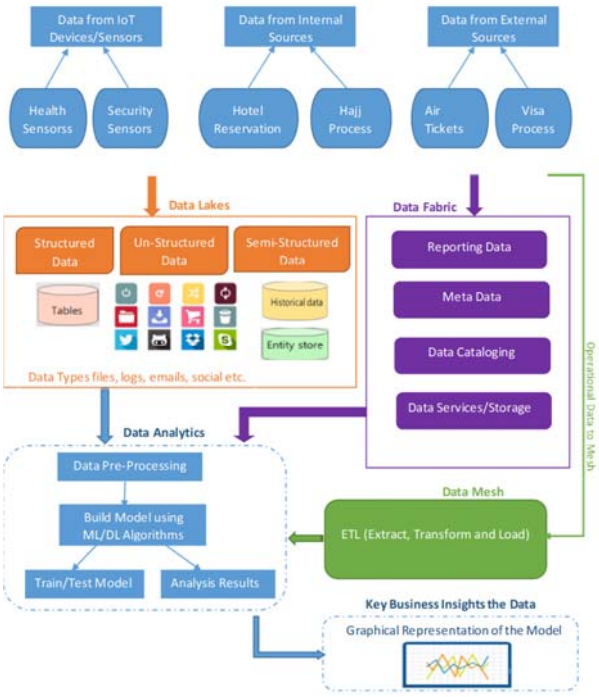

Figure 1: The Proposed methodology flow

4. Results and Discussion

The proposed system was evaluated using the open-source dataset. First, the Hajj pilgrim's dataset was used to analyze the services provided for the people on the Hajj occasion [11]. Next, the dataset was stored for training the model once, and then the unseen dataset was used to get the performance of the proposed system.

The first analysis was to see the number of pilgrims per year. It was started from 1995 to 2021. As can be seen, the number of pilgrims changes every year. The last year, 2021, was COVID-19, so the number of pilgrims was not the same as earlier. Nevertheless, it accounted for the total number of pilgrims in 2021 was 58745.

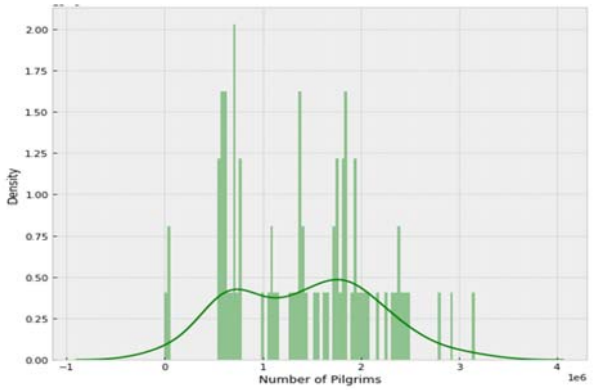

Figure 2: Number of pilgrims for the different years from 1995 to 2021

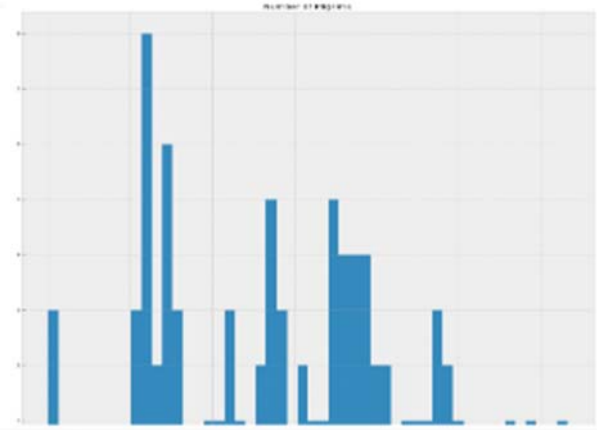

Figure 3: The range values for the number of pilgrims in the histogram

The ratio is seen in the second experiment, where the proposed system used domestic and foreign pilgrims. Multiple services were provided for the pilgrims, and the proposed system analyzed each service. Transport services, including vehicles equipped for the pilgrims.

An open-source dataset was used for the Umra and Hajj performers by gender classification, pilgrims' age group, and the pilgrim's nationality. In addition, exploratory data analysis was performed to analyze each column's contribution to the data. For example, the Gender column was portioned into Males with a value of 1 and Females with 0. Figure 3 shows the gender contribution in the data where 1 represents the male and 0 for the Female contributions.

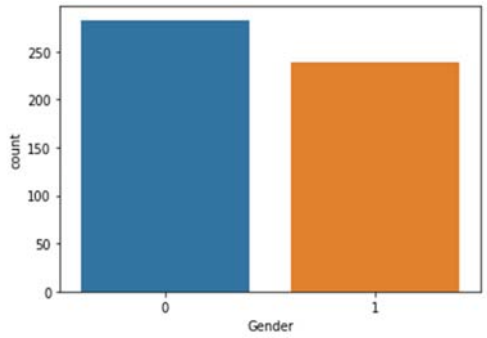

Figure 4: Gender variable contribution in the data (Male:1 and Femal:0)

The open-source dataset categorizes nationalities as Saudi, non-Saudi, and total. However, the total was ignored in this model because it was not required in the model's development. Figure 4 depicts the nationality column's category.

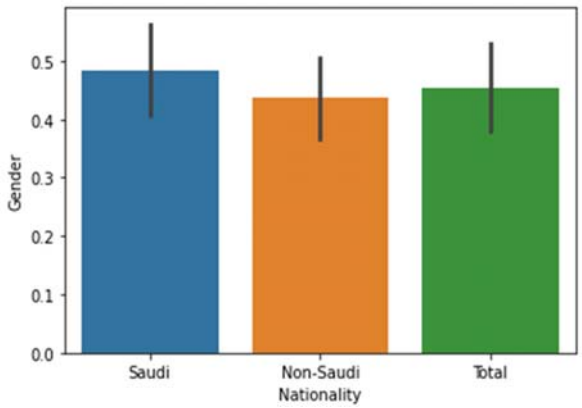

Figure 5: EDA for the nationality column in the dataset

Machine learning algorithms are used for model training and testing. Machine learning algorithms are used for the model training where the evaluated model has the values for the model. Logistic Regression (LR) was used to train the model, and an accuracy value was 60% obtained in the given dataset. Naïve Bayes (NB) algorithm didn't perform well, and the accuracy value was only 47.62%. K-Nearest Neighbors (K-NN) obtained the second-highest accuracy value against the open-source dataset, 62.83%. A Decision Tree (DT) was used to train the model, and the accuracy values against this algorithm were highest from other algorithms, which were 62.86%. Random Forest accuracy was 59.05%, and XGBoost accuracy was 58.1%. The complete comparison is shown in Table 1 for all the algorithms.

Table 1: Results obtained from different classifiers

No.	Algorithm Name	Accuracy (%)
1	Logistic Regression	60.00
2	Naïve Bayes	47.62
3	K-Nearest Neighbors	62.83
4	Decision Tree	62.86
5	XGBoost	58.1

Figure 5 shows the complete representation of each algorithm's results, where the Decision Tree outperforms other algorithms. At the same time, K-Nearest Neighbors was the second algorithm after the Decision Tree. However, the accuracy values were not good because there was not a sufficient available dataset. In the future, the dataset can be increased for the betterment.

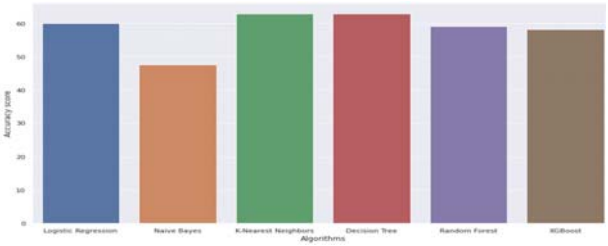

Figure 6: Machine learning algorithms' accuracy values during classification

The proposed system was evaluated with an accuracy measure on the given open-source dataset. The performance of the proposed method was compared with some different machine learning algorithms. Figure 5 shows the performance for the given machine learning algorithms where the decision tree outperformed this dataset's other algorithms.

5. Conclusion

It was concluded that Hajj was the central pillar of Islam; Muslims offer this occasion once a year. Muslim pilgrims are the most worshipers who travel in many countries. Every religion has its norms and rules. People wanted to visit Macca, located in Saudi Arabia. Every Muslim people wants to travel toward Macca for the love of Allah. Therefore, there is a need to develop a system that can facilitate the people for smooth execution of the Hajj process. Multiple devices are already installed in Makkah, but it was better to suggest the data architectures with the help of machine learning approaches. The proposed system

analysis the services provided to the pilgrims. The proposed system performed better in tackling the research problem. Visuals of the proposed method showed the system's performance using data architectures.

References

[1] Rustika, Rustika, Ratih Oemiati, Al Asyary, and Tety Rachmawati. "An evaluation of health policy implementation for Hajj pilgrims in Indonesia." Journal of Epidemiology and Global Health 10, no. 4 (2020): 263.

[2] Alfelali, Mohammad, Osamah Barasheed, Parvaiz Koul, Al-Mamoon Badahdah, Hamid Bokhary, Mohamed Tashani, Marwan Bakarman, Gulam Khandaker, Robert Booy, and Harunor Rashid. "Influenza vaccine effectiveness among Hajj pilgrims: a test-negative case-control analysis of data from different Hajj years." Expert review of vaccines 18, no. 10 (2019): 1103-1114.

[3] Alharthi, Nemshan, and Adnan Gutub. "Data visualization to explore improving decision-making within Hajj services." Scientific Modelling and Research 2, no. 1 (2017): 9-18.

[4] Goni, Mohammed Dauda, Nyi Nyi Naing, Habsah Hasan, Nadiah Wan-Arfah, Zakuan Zainy Deris, Wan Nor Arifin, Tengku Mohammad Ariff Raja Hussin, Abdulwali Sabo Abdulrahman, Aisha Abubakar Baaba, and Muhammad Rafie Arshad. "Development and validation of knowledge, attitude and practice questionnaire for prevention of respiratory tract infections among Malaysian Hajj pilgrims." BMC Public Health 20, no. 1 (2020): 1-10.

[5] Bhuiyan, Md Roman, Junaidi Abdullah, Noramiza Hashim, Fahmid Al Farid, Jia Uddin, Norra Abdullah, and Mohd Ali Samsudin. "Crowd density estimation using deep learning for hajj pilgrimage video analytics." F1000Research 10 (2021).

[6] Shambour, Mohd Khaled, and Adnan Gutub. "Progress of IoT research technologies and applications serving Hajj and Umrah." Arabian Journal for Science and Engineering 47, no. 2 (2022): 1253-1273.

[7] Alharthi, Nemshan, and Adnan Gutub. "Data visualization to explore improving decision-making within Hajj services." Scientific Modelling and Research 2, no. 1 (2017): 9-18.

[8] Alabdulkarim, Lamia, Wafa Alrajhi, and Ebtesam Aloboud. "Urban analytics in crowd management in the context of Hajj." In International Conference on Social Computing and Social Media, pp. 249-257. Springer, Cham, 2016.

[9] Alhomoud, Faten, and Farah Alhomoud. ""Your Health Essential for Your Hajj": Muslim pilgrims' knowledge, attitudes and practices regarding Middle East respiratory syndrome coronavirus (MERS-CoV) during Hajj season." Journal of Infection and Chemotherapy 23, no. 5 (2017): 286-292.

[10] Showail, Ahmad J. "Solving Hajj and Umrah Challenges Using Information and Communication Technology: A Survey." IEEE Access 10 (2022): 75404-75427.

[11] Alharthi, Online Available] "https://datasource.kapsarc.org/pages/home/"