ISSN No:-2456-2165

Covid-19 Prediction using Azure Data Factory (ADF)

Upendra Chauhan[1]
CSE department Chandigarh University
Mohali, Punjab, 140413, India

Ramkrishna Singh[2]
CSE department Chandigarh University
Mohali, Punjab, 140413, India

Ramandeep Kaur[3]
CSE department Chandigarh University
Mohali, Punjab, 140413, India

Shubham Singh[4]
CSE department Chandigarh University
Mohali, Punjab, 140413, India

Tanisha Kumari[5]
CSE department Chandigarh University
Mohali, Punjab, 140413, India

Ankur Chauhan[6]
CSE department Chandigarh University
Mohali, Punjab, 140413, India

Abstract:- Developing an advanced COVID-19 screening scheme using Azure's heterogeneous datasets coupled with ECDC data. Combining comprehensive datasets, including medical records, radiological imaging scan, clinical details, and epidemiological information from ECDC will produce a refined and accurate model for diagnosing COVID-19 and integrating the capabilities of data lakes and data warehouses. Azure's dataset integration incorporates information like X- rays, CT scans, patient demographics, clinical symptoms,while ECDC data encompasses broad epidemiological, transmission rates, and geographical spread. The final aim is to present healthcare practitioners with a diagnostic aid for rapid and accurate differentiation of a COVID-19 positive and non- COVID-19 patient thus facilitating immediate patient management and framing response measures to be put into place during pandemics. Using the extensive epidemiological data repository generated by the European Center for Disease Prevention and Control (ECDC) coupled with its Azure services, including Azure Data Lake Storage, Azure Databricks, and Power BI, the development efforts would be focused on creating an adaptable approach for prompt detection of new COVID-19 cases, enabling healthcare institutions.

Keywords:- *COVID-19, Healthcare Management, Azure Data Factory, Azure Data Storage.*

I. INTRODUCTION

Various COVID-19 detection projects emerged, including improved PCR and rapid antigen tests, AI-assisted imaging analysis, thermal screening, contact tracing apps, CRISPR- based tests, serological antibody tests, genomic sequencing, and vaccine development. These initiatives aimed to enhance testing accuracy, speed, and accessibility while employing technologies like AI, CRISPR, and thermal imaging. Efforts also focused on vaccine development, distribution systems, and robotics for sanitation. The landscape continually evolves, with a focus on innovative solutions for efficient and widespread COVID-19 detection and management.

Rapid and precise COVID-19 detection is critical, yet current methods such as nucleic acid tests or clinical symptoms alone are limited for early identification. Real-time RT-PCR, although reliable, can yield false results, raising concerns and prompting discussions on detection challenges.

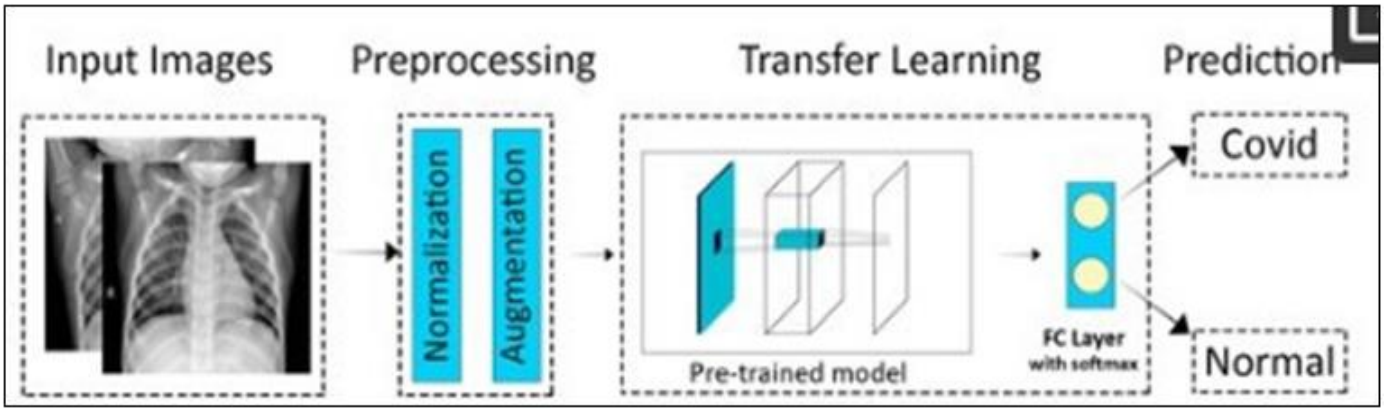

Fig 1 Block Diagram of Existing Solutions

Convolutional neural networks (ConvNets/CNNs) are specialized in deep learning, categorizing information by processing small data segments through a multi-layer stack of adjustable neurons. They detect basic patterns initially, utilizing this information across subsequent layers to differentiate and classify objects, enabling effective object recognition and categorization. As shown above in figure *Fig.1*, it is based on Machine learning which analyzes the covid vs normal cases of the patients through diagnosis of the diseases. In Machine learning, ANFIS and NN are several classification methods which are used for the prediction of covid-19 cases [12].

Amid reduced student numbers due to COVID, educational institutions focused on R&D, developing prototypes for industry use. Some products are already commercialized, offering cost-effective indigenous solutions while adhering to pandemic constraints through lab SOPs. Institutes provided COVID care and detailed their pandemic-related R&D in a booklet, seeking support for patents, commercialization, and industry collaboration via fairs and increased engagement.

Rapid and accurate COVID-19 detection is vital, but current methods have limitations. Real-time RT-PCR, though the standard, can yield false results. Combining it with clinical signs is more effective, prompting discussions on SARS-CoV-2 detection challenges [4]. Understanding diagnostic method limitations is crucial during the pandemic, emphasizing the use of predictive models with ECDC data [9] to estimate COVID-19 cases without tests.

Many nations now consider the global outbreak of the severe acute respiratory illness coronavirus (SARS- CoV-2) to be a major national security concern [4]. Standard epidemiological models [12] imply limited accuracy for a long time forecast due to the substantial level of sensitivity and lack of urgent information. In order to predict the COVID-19 flare-up in India, we will use the ADF to uncover undetected trends in the Covid-19 data of that country.

II. LITERATURE REVIEW

As an alternative to Susceptible, Infectious, or Recovered (SIR) and Susceptible, Exposed, Infectious, or Recovered (SEIR) models, *Sina F. Ardabili et.al (*2020*)* have developed a relative evaluation of machine learning and computerized models to predict the COVID-19 outbreak. In addition to suggesting that real interest in outbreak detection can be recognized by combining machine learning and SEIR designs, the work provides an underlying specification to demonstrate the potential of machine learning for further study [1].

A data-mining technique has been proposed by *Fabian Stephany et.al (2020)* to quantify the specific to the sector risks associated with COVID-19 as part of this study. It examines risk reports submitted by organizations to the Securities and Exchange Commission (SEC) in the United States. An examination of risk assessments in real time is made possible by this set of data [11].

An article by *Arpita M. Hirudkar et.al (2013)* focuses on several functional, rational, cognitive, and investigative aspects that users may look for in the tools. The efficacy and importance of these instruments, along with their many aspects, are covered in full in the study. Research offers a number of these data mining tools' benefits in addition to favorite aspects and the best aspects of the current solutions [18].

In a study, *Venkateswarlu Pynam et.al (2018)* stated that the dataset can be structured or unstructured, with sizes ranging from terabytes to zettabytes. As a result, the study provides a thorough and theoretical analysis of five open-source data mining tools: Orange, KNIME, Weka, Rapid Miner, and R tool [10].

A study by *Musa Peker et.al (2018)* was to accurately assess the HbAlc value. The Orange data mining software is used because it integrates a variety of techniques and is easy to apply during the modeling phase. The goal is to apply a large number of attribute choice and classification techniques to improve an efficient prediction system. The results show that the proposed model predicts the HbA1c parameter accurately [7].

Dr. Nidhi H. Divecha et.al (2017) have used data mining technologies for analysis; however, one tool is superior to the other. The orange gadget is easy to use and has worked wonderfully. Orange has also carried out the actual execution in accordance with its feature. Research on investigations is streamlined by this instrument [8].

In a study, *Praveen Kumar V et.al (2020)* talked about the issues that students were facing and how research had been done to determine the best way to help them feel less depressed throughout COVID-19. The model created in this work predicted and analyzed the students financial issues during COVID-19, utilizing the student dataset of their diverse COVID-19 challenges in Bangalore to determine the remedy [16].

Rohit Ranjan et.al (2017) conducted a thorough and theoretical analysis of five open-source information mining tools: Rapidminer, R, Knime, Orange, and Weka. The technical specification characteristics and area of expertise of each instrument are utilized to outfit the examination's points of interest and weaknesses [17].

A. Timeline of the reported problem

Throughout these years, research has been ongoing, covering a wide range of topics related to COVID-19 detection. Efforts are focused on making tests moreaccurate, faster, and adaptable to the evolving nature of the virus, all while exploring innovative technological solutionsfor more effective detection and management.

In 2020, the focus was on understanding SARS-CoV-2. Research involved validating PCR techniques for precise COVID-19 detection and genetic analysis.

In 2021, research prioritized non-PCR testing options like serological and antigen tests, aiming to enhance accuracy and overcome detection challenges.

In 2022, ongoing research focused on innovating precise and rapid COVID-19 diagnostic tools, exploring CRISPR, AI, and adaptable testing methods.

In 2023, efforts concentrated on enhancing COVID-19 tests for speed and accuracy while prioritizing data analysis tools for better interpretation and diagnostic efficiency. Ongoing developments aimed to optimize the detection process.

Table 1 Timeline of the Reported Problems

Year	***Key Research Focus***
2020	- Identification of the virus - Development and validation of the RT- PCR tests. - Recognizing the virus's genetic composition in order to detect it. - Initial research on the accuracy of real-time RT-PCR.
2021	- Investigating different testing techniques (such as serological assays and antigen tests). - Progress made in lowering false positives and false negatives.
2022	- Ongoing investigation into cutting-edge diagnostic technologies (AI- and CRISPR-based tools, for example)
2023	- Concentrate on improving test speed and accuracy. Including data analysis tools to facilitate simplified interpretation - Continued advancement of technologies for quick testing.

Even though COVID-19 detection has advanced, a number of issues still need to be resolved, some of which could benefit from using Azure Data Factory (ADF) as a potential resource:

- Data Integration and Quality Control
- Real-time Data Processing
- Handling Large Volumes of Data
- Predictive Analytics and Machine Learning Integration
- Data Security and Compliance
- Monitoring and Reporting

B. Proposed Solutions:

As we have gone through some research papers, we have found some advancements in the technology used.

- *Azure Data Factory (ADF) Integration*: Integrates multiple Azure ecosystem solutions, expediting data processing, analysis, and interpretation for enhanced COVID-19 detection and management.
- *Azure Machine Learning:* Applies predictive analytics and machine learning models to identify trends and forecast outbreaks, enhancing detection precision and proactivity.
- *Azure Data Lake Storage*: Offers secure, scalable storage for extensive COVID-19 data, enabling comprehensive analysis of vast data volumes.
- *Azure Data-bricks*: Provides scalable data science capabilities and an analytics platform, enabling researchers to handle and analyze extensive COVID-19 data for deeper insights.
- *Azure Synapse Analytics*: Processes large data amounts, producing insights and visualizations, structuring big data for effective analysis.
- *Azure Cognitive Services:* Integrates machine vision and NLP, analyzing unstructured COVID-19 data such as text or photos, enhancing detection analysis.
- *Power BI and Azure Data Factory*: Collaborate to develop user-friendly, educational dashboards displaying insights from COVID-19 data study.
- *Data Security Measures*: Azure's robust security solutions ensure data safety, integrity, and compliance throughout the COVID-19 detection process.
- *Open-Source APIs and Datasets*: Enhance analysis, providing access to external resources and expanding COVID-19 data insights.
- *Overall Impact*: The integration of Azure solutions significantly improves analytics, security, and machine learning capabilities, resulting in more precise and comprehensive COVID-19 detection.

C. Problem Definition

In order to enable efficient disease monitoring and control, the COVID-19 detection project using Azure Data Factory defines the problem as overcoming the difficulties in precisely and promptly identifying COVID-19 instances. Important components of the problem definition are:

- Current COVID-19 detection methods (PCR tests, clinical features) struggle with consistency, demanding enhanced accuracy to minimize false results.
- COVID-19 healthcare data's varied formats challenge integration and analysis; crucial to address disparities for effective integration procedures.
- COVID-19 healthcare data's varied formats challenge integration and analysis; crucial to address disparities for effective integration procedures.
- Preserving data security and complying with regulations are vital in addressing healthcare data sensitivity for secure and law-abiding processing.

Efficient data workflows are crucial for managing large datasets, emphasized in the study. The project aims to create comprehensive data pipelines through Azure Data Factory, facilitating real-time analysis and improving pandemic management decisions.

Integrating Power BI with Azure Data Factory would make it easier to visualize and report insights obtained from COVID-19 data, providing useful dashboards that may effectively convey trends and patterns.

Using Azure Data Factory and related tools, the strategy aims for precise COVID-19 detection, efficient data processing, proactive decision-making, and ultimately better pandemic control and mitigation.

D. Goals and Objectives

The key objectives involve enhancing detection accuracy, real-time data processing, predictive capabilities, and simplified data integration in COVID-19 management using Azure tools for accurate diagnosis, prompt decision-making, predictive insights, and comprehensive data analysis.

III. DESIGN FLOW/PROCESS

A. Evaluation and Selection of Specifications/Features

Data collection involves acquiring, formatting, and storing from various sources like ECDC.

Post-collection includes data integration, cleansing, and loading into a high-quality database.

Azure SQL Server is chosen for data storage and schema creation.

Real-time reporting dashboards are created using Data Bricks, Azure Data Factory, HDInsight, for user-friendly COVID-19 visualizations.

Ensuring system scalability and performance during high data loads is crucial, necessitating optimized design and infrastructure for reliability.

B. Design Constraints

- Data Collection Constraints
- ETL Process and Data Transformation Constraints
- Data Storage and Management Constraints
- Reporting and Visualization Constraints

C. Analysis of Features and Finalization Subjects to Constraints

We have mainly emphasized on assessing the project's elements and choices in light of the limitations set forth for the COVID-19 reporting and forecast. This section functions as a comprehensive analysis and decision-making process where you discuss how various features, approaches, and options comply with or depart from the specified limitations.

➢ *Feature Analysis:*

We have evaluated data gathering efficiency, assessed ETL process for accuracy and integration, reviewed chosen data storage, and analyzed Power BI's features and user-friendliness for pre-dashboard creation purposes.

➢ *Critical Analysis:*

We have assessed how well each element performs in achieving the project's objectives, including scalability, real-time reporting, and accurate forecasts. We have evaluated the data's integrity and dependability at every step of the procedure, as well as how it affects the final outcomes.

➢ *Finalization and Recommendations:*

The analysis scrutinizes project features' adherence to guidelines, addressing constraints and suggesting modifications to align better with limitations. It conducts an in-depth evaluation of each component's compliance with set parameters and its impact on the accurate forecasting andreporting of COVID-19 data.

D. Design Flow

The goal of the project's design flow is to establish a methodical and precise approach for the creation of your solution, "Covid-19 Predictions/Reporting Using Azure Data Factory."

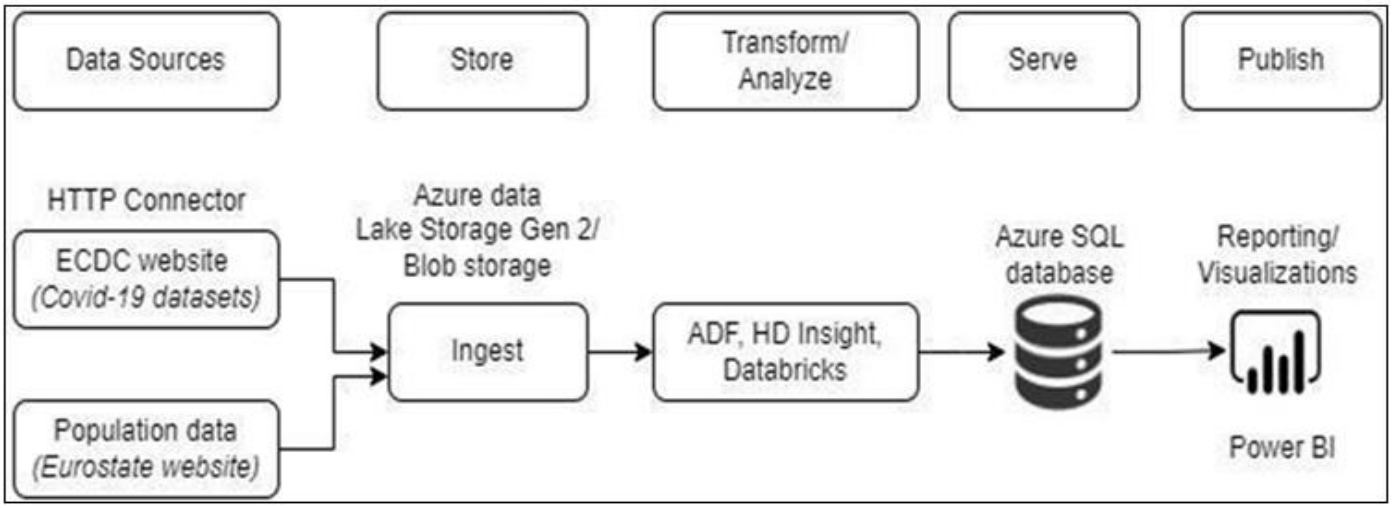

Fig 2 Solution Architecture of Our Project

- *Data Gathering and Preparation:* At first, our attention was on collecting, extracting, and validating data from diverse sources, primarily the European Centre for Disease Prevention and Control, involving access permissions and storage in temporary locations.
- *Data Integration and Transformation*: Post- collection, the focus involved efficient integration and conversion of data, executing the process of extracting, transforming, and loading raw data into a single, high-quality database.
- *Data Management and Storage*: Choosing Azure SQL Server for its scalability, security features, and Azure compatibility, ensuring data integrity by tailoring a schema for COVID-19 metrics.
- *Real-time Reporting Dashboards:* Real-time dashboards, created using Data Bricks, Azure Data Factory, HDInsight, and other tools, offer user-friendly COVID-19 data visualization and forecasting.
- *Scalability and Performance Optimization:* We improved system scalability for managing increased data volumes during peak demand, ensuring responsive and reliable infrastructure through design optimization.

IV. METHODOLOGY

This project used Azure Data Factory (ADF) to process, transform and retrieve COVID-19 data from ECDC website and Azure Blob Storage to Azure Data Lake Gen2. ADF data streams were used to transform the data, including filtering, selecting, rotating and merging. The methodology include the steps:

A. Data Ingestion

Azure Data Factory (ADF) provides a number of connectors for getting data from different sources. ADF was used to it process the following data sets from the ECDC website a Azure Blob Storage to Azure Data Lake Gen2:

➢ *Cases and Deaths Data:*

This data set contains information on the number of confirmed cases and deaths from COVID-19 in every European country.

➢ *Hospital Admissions Data:*

This data set contains information on the number of hospitalizations and ICUs arrivals from COVID-19 in each European country.

➢ *Population Data:*

This dataset contains information about the population of each European country.

➢ *Test Conducted Data:*

This data-set contains information on the number of COVID-19 tests conducted in each European country. ADF Copy Activity was used to ingest the data from the ECDC website and Azure Blob Storage into Azure Data Lake Gen2. ADF provides a number of features to ensure that the data is ingested correctly, including:

- *Validation Activity:*

This activity can be used to find out the data before it is used.

- *Get Metadata Activity:*

This activity can be used to get the metadata of the data before processing. This can be useful for understanding the structure of the data and for identification of any potential problems.

- *Error Handling:*

ADF provides a number of error handling features to ensure that the data is ingested even if there are errors.

B. Data Transformation

Once the data has been received into Azure Data Lake Gen2, ADF Data Flows was used to transform the data. ADF Data Flows is a graphical user interface (GUI) that makes it easy to create and manage data transformations. The following data transformations were performed on data:

➢ *Filtering:*

The data is filtered to only includethedata from European countries.

➢ *Selecting:*

Only the useful and required columns were selectedfrom the data.

➢ *Pivoting:*

The data has been pivoted for grouping it by country and date.

➢ *Joining:*

The data from the different data-sets was joined together.

➢ *Sorting:*

He data was sorted by country and date.

➢ *Conditional Splitting:*

The data was split into two paths based on the value of a column.

C. Data Loading

The transformed data was loaded into Azure SQL Database using the ADF Copy Activity. Azure SQL Database is a cloud-based relational database service that offers high performance, scalability, and availability.

D. Data Reporting

Power BI was used to create reports on the COVID-19 data in Azure SQL Database. Power BI is a business intelligence (BI) tool that makes it easy to create and share interactive data visualizations.

The following reports were created using Power BI:

- COVID-19 Trend in the EU/EEA & UK 2020 by Cases, Deaths, Hospital Occupancy, and ICU Occupancy.
- COVID-19 Cases and Death breakdown by population in the UK, France, and Germany.
- Confirmed Cases Vs Total Deaths by Country.
- Total Number of Covid tests carried out vs. Confirmed Cases.
- The reports were published to Power BI Server, where they can be accessed by authorized users.

V. RESULTS AND DISCUSSIONS

The COVID-19 reporting data pipeline developed in this research project consists of the following steps:

➢ *Data ingestion*:

The COVID-19 data from the ECDC website and Azure Blob Storage is ingested into Azure Data Lake Gen2 using the Azure Data Factory Copy Activity. The Copy Activity is configured to run on a scheduled basis to ensure that the data in Azure Data Lake Gen2 is always up-to-date.

➢ *Data transformation:*

The data in Azure Data Lake Gen2 is transformed using Azure Data Factory Data Flows. The Data Flows are configured to perform the following transformations:

Filter the data to only include data from European countries.

- Select only required columns from the data.
- Pivot data to group it by country and date.
- Join the data from different datasets together.
- Sort the data by country and date.

Calculate new metrics, such as the total number of cases and deaths per 100,000 people.

➢ *Data loading:*

The transformed data in Azure Data Lake Gen2 is loaded into Azure SQL Database using the Azure Data Factory Copy Activity. The Copy Activity is configured to run on a scheduled basis to ensure that the data in Azure SQL Database is always up- to-date.

➢ *Data reporting:*

Power BI reports are created to visualize and analyze the COVID-19 data in Azure SQL Database. The Power BI reports include charts and graphs that show the trend of the pandemic over time, the geographic distribution of cases and deaths, and the impact of the pandemic on different populations.

The COVID-19 reporting data pipeline, employed by European public health bodies, facilitates pandemic monitoring by leveraging Azure Data Factory for efficient, scalable, and reliable data management and decision-making, ensuring operational reliability and flexibility during the ongoing pandemic response.

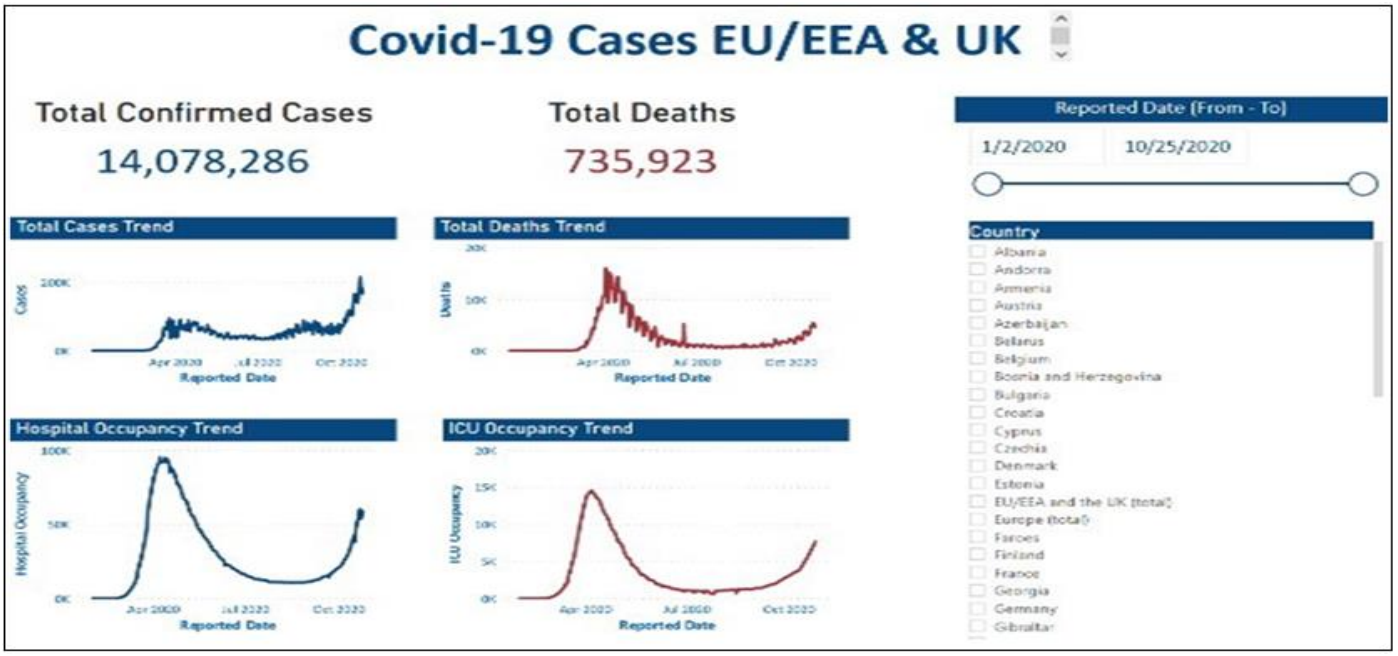

Fig 3 Analyzed Cases Based on Parameters

Fig. 3, presumably included in a document or presentation related to COVID-19 data. The figure presents information within a specified range of reported data. Specifically, it includes data on total COVID-19 cases, the overall count of deaths attributed to COVID-19, and the occupancy status of hospitals and intensive care units (ICUs) by COVID-19 patients. By visually representing these key metrics, individuals and healthcare professionals can quickly grasp the current situation related to COVID-19. This information can be crucial for making informed decisions, allocating resources, and understanding the impact of the virus on healthcare infrastructure.

ISSN No:-2456-2165

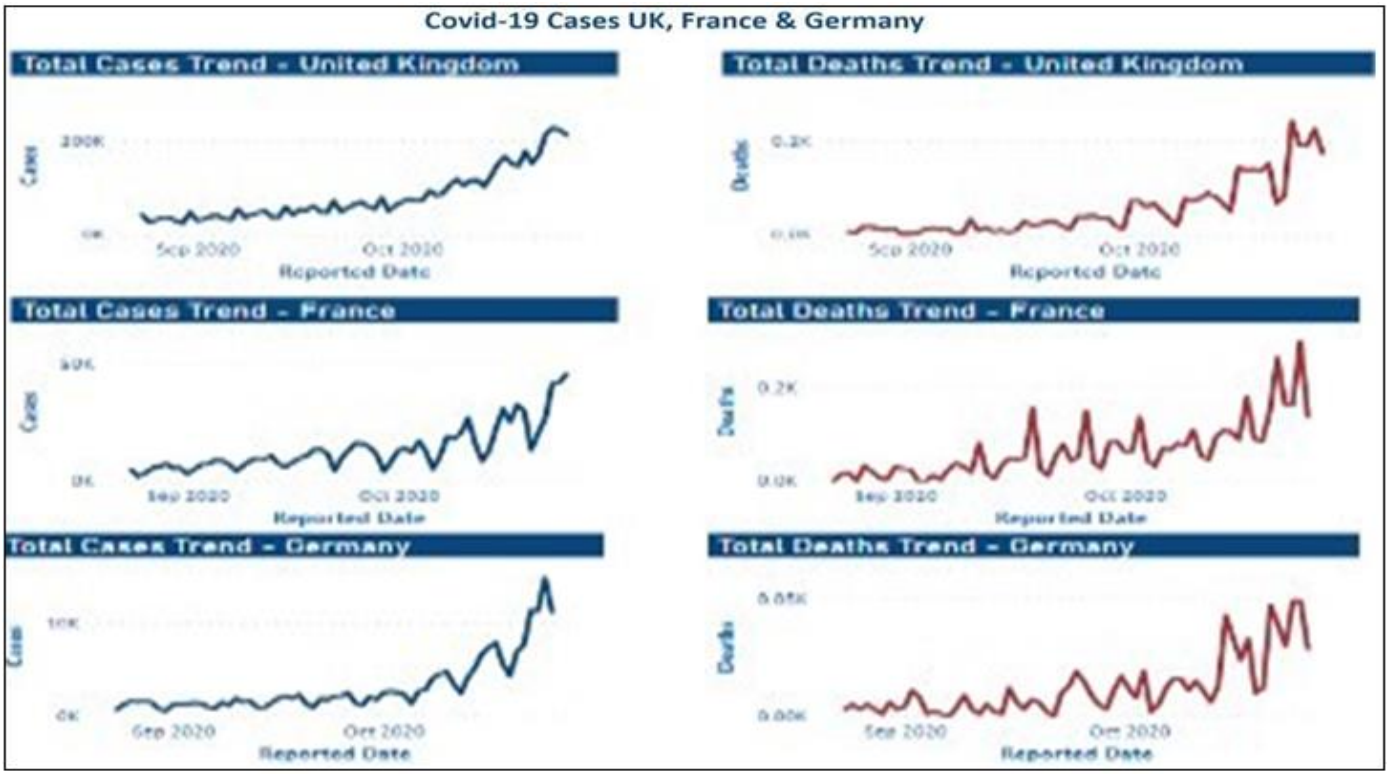

Fig 4 Cases Analyzed in Different Countries

In this figure Fig. 4, total cases and total deaths found in different countries including the United Kingdom, France, Germany have been shown using proper visualization techniques such as Power BI. Another limitation of this research project is that it didnot evaluate the impact of the data pipeline on the decision-making process of public health organizations. Future work could involve conducting a study to evaluate the impact of the data pipeline on the decision- making process of public health organizations.

VI. CONCLUSION

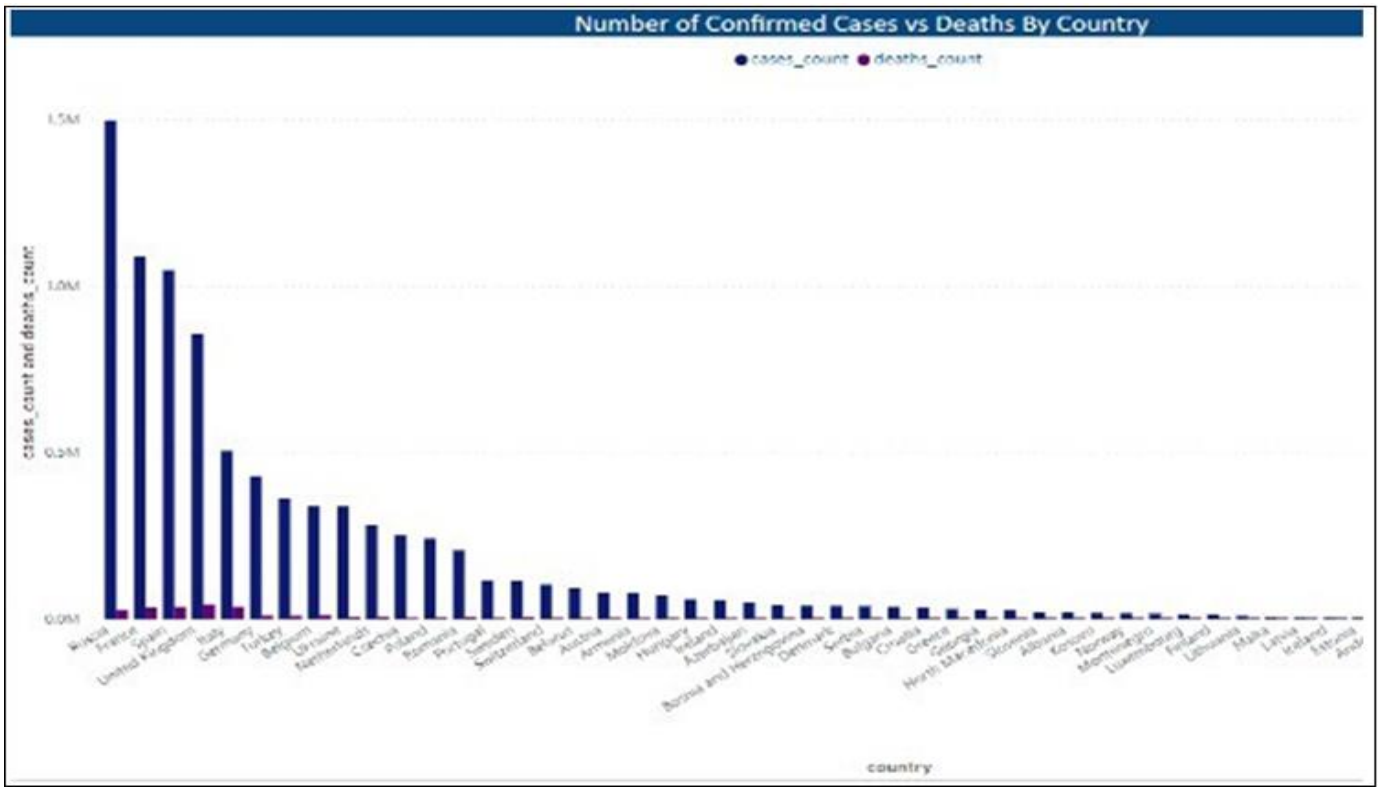

Fig 5 Number of Confirmed Cases VS Deaths by Country

In this figure, Fig.5, It has been shown a visual representation which provides information on the relationship between the number of confirmed cases of COVID-19 and the number of deaths by country.

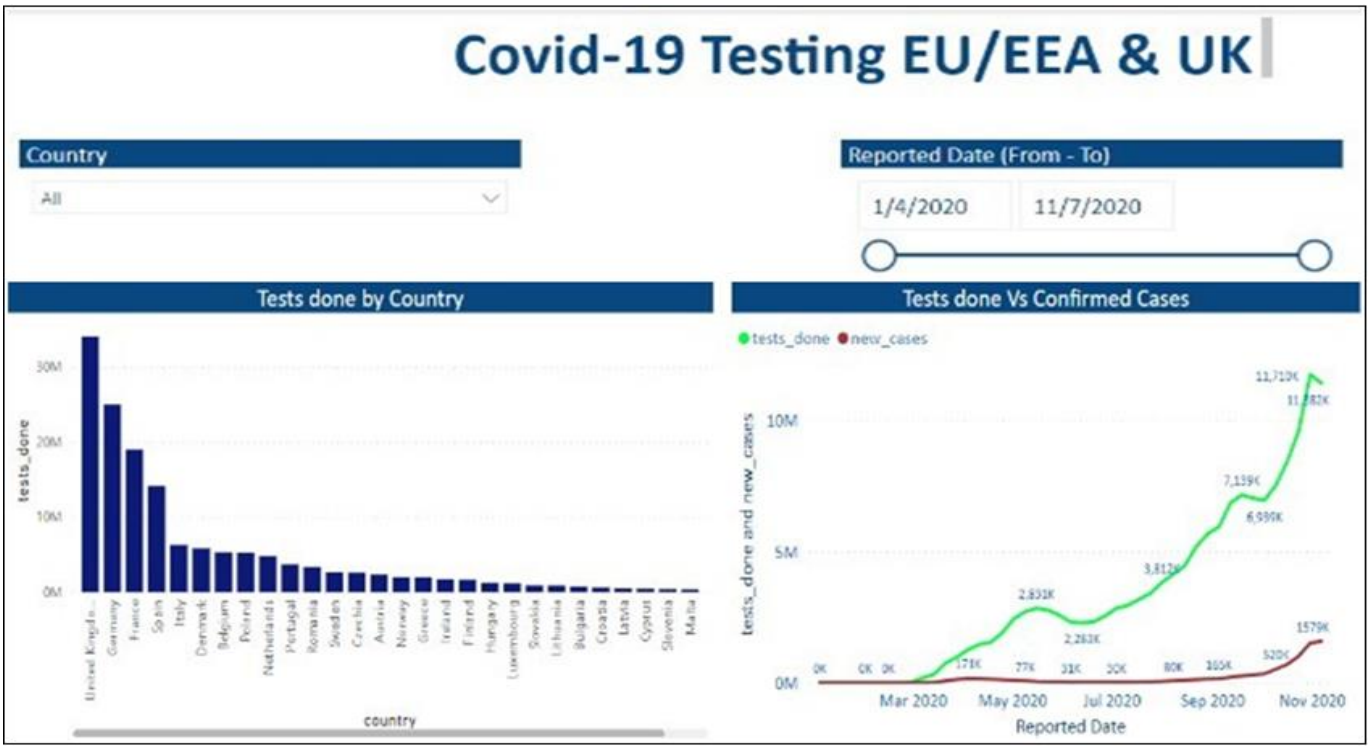

Fig 6 Test done Vs Country and Confirmed Cases

Overall, the results of this research project demonstrate the feasibility of using Azure Data Factory to build and manage a scalable and reliable data pipeline for COVID-19 reporting. The data pipeline developed in this research project is currently being used by public health organizations in Europe to track and monitor the COVID-19pandemic, and it is helping public health organizations to make better decisions about how to respond to the pandemic.

LIMITATIONS AND FUTURE WORK

One of the limitations of this research project is that it focused on developing a data pipeline for COVID-19 reporting in Europe. Future work could involve extending the data pipeline to support other regions of the world.

A. Objectives

During this research project, it was our main goal explore the potential of Azure Data Factory as a tool for building a data pipeline tailored for reporting COVID-19 Europe. We set out on this journey with a goal gaining hands-on experience and data insights management including aspects such as data reception, transformation, loading and reporting.

B. Review of Key Figures

Introducing our journey into the world of data management us to the amazing possibilities of Azure Data Factory. The the data pipeline we developed allowed us to collect a transform data from various sources, allowing us to generate meaningful insights. It was fascinating to see how key numbers, both in terms of data and in terms of practical use technology, played a key role in the success of this project.

C. Implications and Applications

The implications of our work go beyond borders our project. Successful data channel deployment in European public health organizations carries a significant implications for real-world decision-making during global health crisis. It's exciting to think about bigger things of Azure Data Factory applications not only in Europe but also as a model for other regions in dealing with public health challenges.

D. Recommendations for the Future

As we wrap up this project, it is clear that there is still room for growth. We recommend extending the range of the data channel support a wider range of regions around the world. It is also it is essential to consider in- depth studies that evaluate how our the data channel can improve the decision-making processes inside public health organization. This ensures that the impact we witnessed that our project can continue to develop and make a deeper difference.

In short, this project was educational experiences that offer us valuable insights into the data management and technology. With Azure Data Factory as our companion, we have achieved our goals, we appreciate the importance of key figures for success and recognized real implications and applications of our work. Looking ahead, this conclusion does not mean the end, but a in the beginning as we continue our journey to more effective data-driven responses to COVID-19 and the like challenges.

REFERENCES

[1]. S.F. Ardabili, A. Mosavi, P. Ghamisi, F. Ferdinand, A.R. Varkonyi-Koczy, U. Reuter, et al., "COVID-19 Outbreak Prediction with Machine Learning", *Algorithms*, pp. 13, 2020.

[2]. Romona M. Harris. "*Data Warehousing and Decision Support System Effectiveness Demonstrated in Service Recovery During COVID19 Health Pandemic.*" DOI: 10.1109/ICOSST51357.2020.9333019.

[3]. Oleksii Duda, Volodymyr Pasichnyk, Nataliia Kunanets, Rzheuskyi Antonii, and Oleksadr Matsiuk. "*Multidimensional Representation of COVID-19 Data Using OLAP Information Technology*." DOI: 10.1109/CSIT49958.2020.9321889.

[4]. R. Niehus, P. Martinez de Salazar Munoz, A. Taylor, M. Lipsitch, *Quantifying bias of COVID-19 prevalence and severity estimates in Wuhan, China* that depend on reported cases in international travelers (2020).

[5]. P.Pulla, Covid-19: *India imposes lockdown for 21 days and cases rise, 2020.*

[6]. O. Analytica, *Japan'S partial COVID-19 lockdown may be insufficient, Emerald Expert Briefings*(oxan?es).

[7]. Peker Musa et al., *Use of Orange Data Mining Toolbox for Data Analysis in Clinical Decision Making: The Diagnosis of Diabetes Disease*, 2018.

[8]. M Kukasvadiya, S. aqsud et al., "*Analysis of Data Using Data Mining tool Orange*", 2017 IJEDR, vol. 5, no. 2, ISSN 2321-9939.

[9]. ECDC Website for Covid-19 Data *https://www.ecdc.europa.eu/en/covid-19/data*

[10]. Venkateswarlu Pynam, R Roje Spanadna and Kolli Srikanth, "*An Extensive Study of Data Analysis Tools (Rapid Miner Weka R Tool Knime Orange)*", SSRG.

[11]. Stephany Fabian et al., The CoRisk-Index: *A data-mining approach to identify industry-specific risk assessments related to COVID-19 in real-time*, 2020. *TA Systematic Review of Machine Learning Methods for COVID-19 Prediction*. Authors: Jinyang Liu, Xiaobo Li, Yanmei Wang, Yang Liu, Junwei Gao, and Hui Liu Journal: Frontiers in Public Health, Year: 2022 *DOI: 10.3389/fpubh.2022.926592*

[12]. *Forecasting COVID-19 Cases and Deaths Using Machine Learning Models:* A Comparative Analysis Authors: Muhammad Sajid Khan, Shah Fahad, and Muhammad Ehsan *Journal: ACM Transactions on Knowledge Discovery from Data Year: 2023* DOI: 10.1145/3576362.3595398

[13]. *COVID-19 Prediction Using Data Mining Techniques:* A Survey Authors: Pooja Agarwal, Deepak Kumar, and Ankur Agarwal Journal: *International Journal of Computer Science and Technology, Year: 2022 DOI: 10.17762/ijcst.v13i1.2022.010*

[14]. *A Hybrid Data Mining Model for COVID-19 Prediction*, Authors: Ashwin Kumar, Renuka Devi, and V. K. Jain, Journal: *International Journal of Data Mining and Knowledge Management, Year: 2023* DOI:10.1504/IJDKM.2023.01012

[15]. Krishna and Kumar V Praveen, "*Prediction and Analysis of Data Mining Models for Students Underlying Issues during Novel Coronavirus (COVID-19)*", International Journal of Engineering Research & Technology (IJERT), ISSN 2278-0181.

[16]. Rohit Ranjan et al., "*Detailed Analysis of Data Mining Tools*", International Journal of Engineering Research & Technology (IJERT), vol. 6, no. 05, May 2017.

[17]. *[17]*A. Hirudkar and S. S. Sherekar, *Comparative Analysis of Data Mining Tools and Techniques for Evaluating Performance of Database System, 2013.*