REPORT ON COVID 19 DATA SET OF PAKISTAN

REPORT IS FOR THE PREDICTION OF EXPIRY STATUS SO THAT THE NECESSORY PRECAUTIONS HAS BEEN MADE .

REPORT IS SUBMITTED BY :   SEAT NO

                       ALAMGIR KHAN   19122013

                       IRUM NAZ   19122183

CLASS   :   BSCS (MORNING REGULAR)

SEMESTER   :   6TH

SECTION   :   A

SUUBMITTED TO:MAM UZMA   (FUUAST GULSHAN -E-IQBAL KARACHI)

SUBMISSION DATE   :   December 26, 2023

SUBMISSION DAY   :   TUESDAY

ABSTRACT

In this project we review methods of cluster and naïve bayes analysis in the context of classifying patients on the basis of expired rate type observations . The report is written as the necessity of prediction in medical industry have a huge importance regarding the precautions and necessary steps for patients health as well as to prevent other people from getting such a threatful conditions. As COVID-19 cases were reported from Islamabad and Karachi on February 26, 2020. Pakistan being one of the most densely populated countries in Asia, with a population of 204.65 million, and Karachi being the largest metropolitan city in Pakistan, has been greatly vulnerable to this outbreak. Owing to its present economic condition, health care resources, and the occurrence of previous outbreaks, the Centers for Disease Control and Prevention had already issued a level 3 warning for international travelers to Pakistan. The administration had a huge responsibility to constrain the spread through a timely response and the adoption of appropriate measures to avoid any major catastrophe. The disease was initially difficult to contain, especially because of noncompliance of the general population to the necessary measures and timely reporting of symptoms. Within 45 days, on April 10, 2020, Pakistan reported 4601 confirmed cases with a death toll approaching 66 individuals. The country's testing capacity was limited during the early months of the pandemic, and while high-income countries were conducting large-scale randomized tests to estimate the actual number of confirmed cases, Pakistan was forced to carry out priority-based testing and rely on the enforcement of strict quarantine and isolation strategies to contain the pandemic. Contact-tracing, however, was an effective strategy that not only helped limit the spread of the virus but also helped predict its route through different regions of the country and across different age groups.This report study contains data set of Covid 19 for predictions by useful techniques present in Data Mining so it can help in finding the numerus variations of virus wave in different regions so the medical staff could done their best in helping people to survive and become safe during pandamic.

TABLE OF CONTENTS

REPORT FOR THE PREDICTION OF EXPIRIES DURING COVID 19

TO ,   DATE 26DECEMBER,2023

Mam Uzma (FUUAST Gulshan -e-Iqbal campus)

From   :   Irum Naz and Alamgir Khan

Subject   :   Report on prediction of deaths during Covid 19 in Pakistan

Mam,

The spread of COVID-19 in Pakistan began with the first case being reported on February 26. Pakistan could be a potential hotspot for COVID-19 owing to its high population of 204.65 million and its struggling health care and economic systems. Pakistan was able to tackle the challenge with relatively mild repercussions. The present analysis has been conducted to highlight the situation of the disease in Pakistan in 2020 and the measures taken by various stakeholders coupled with support from the community to abate the risk of catastrophic spread of the virus. So data set of three months of Pakistan regarding covid 19 used for prediction. Data set include following attributes:-

01	Date
02	Region
03	Cumulative Test Positive
04	Cumulative Tests Performed
05	Tests Performed in last 24 hours
06	New (last 24 hrs)
07	Discharged
08	Still admitted
09	Home Quarantine
10	Expired
11	Expired status

Two operators of classification

1. NAÏVE BAYES
2. CLUSTER

was used in this project so that the prediction for the expiries has been made easy and provided a reliable information. Hence the predictive modeling in healthcare brings a variety of benefits to patients, providers and payers. The technology improves care outcomes by enabling hospitals to deliver proactive and more personalized treatment, reduce readmissions, and boost patient satisfaction rates.

In this project one of the two models known as "NAÏVE BAYES" used to develop a clinical Decision analysis tool which allows decision makers to apply evidence based medicine and make objective clinical decision when faced with complex situations. To develop a reliable and effective decision model which can provide high classification accuracy with simple and easy to understand representation of gathered knowledge. To reduce the number of false alerts in decision making systems and also for detecting artifacts in clinical imaging data. Naïve bayes classifier used for handling classification, prediction and accuracy index of data set. In this project collected data set of 413 rows split into two portions of 70% and 30% from which 1 portion is used as system and, so machine can learn from it and capable to implement the model on unseen new data, So it can give prediction of newly provided data to user. when operator NAÏVE BAYES applied on this data, it gave accuracy rate equal to 94.35% according to following table view:-

	True normal	True alert	True Dangerour	Class precision
Pred.Normal	60	0	0	100.00%
Pred.Alert	3	11	4	61.11%
Pred.Dangerour	0	0	46	100.00%
Class recall	95.24%	100.00%	92.00%	

While in another technique Cluster model applied which is a technique to group similar observations into a number of clusters based on the observed value of several variables for each individual. A frequently used applied statistical technique that

helps to reveal hidden structures and clusters found in large data set . Cluster analysis used to identify typical health care utilization patterns based on empirical data. In this project cluster analysis used to overcome the problem of knowing about how many regions have more death rates and how many regions are at low risk so it is applied on same data to divide them in n time of cluster. Cluster here used to separate the data of different group. Each group contains similar data. After applied cluster model the process results into following :-

CLUSTER MODEL

Cluster 0 : 318 items

Cluster 1 : 60 items

Cluster 2 : 34 items

Total number of items : 412

Hence it separated the data into three groups predicting that cluster 0 was at low risk of expiries, cluster 1 have high risk while cluster 2 also raises from high to minimum expiries. Although the given project is completed after applying these two models on the data set of covid 19 providing such a results which helps in knowing about the availability of the medical equipment, medicines, vaccines needed, beds needed, regions at risk and all safety precautions where to used more then normal.

CONCLUSION :-

The experimental results showed that the NAÏVE BAYES provided the best results although it is a simple and versatile classifier. Since the computations are cheap, the NAÏVE BAYES classifier works very efficiently for large datasets. In this project it proves to be an effective method for the expired and diseased sequence classification with better accuracy . On the other hand , clustering of covid 19 data to discover useful patterns by distributing data into three clusters group proved to be a helpful in understanding the needs in pandemic's spread .

RECOMMENDATIONS:-

The data mining techniques can be implemented in association with Naïve Bayes classifier algorithm and Clustering. Clearly the collaboration of these tools performs a prominent role in diagnosing the coronavirus infection or deaths encountered during covid 19 which is declared as pandemic by WHO. The suggested machanisms showed impressive results which may lead to further refinements to utilize data mining, machine learning, artificial intelligence and information technology for diagnosing the patients for coronavirus. There is a dire need to find out both the cases of coronavirus, which are with the symptoms or asymptotic , expirires or registered cases. It is only the way to break the chain and to control the spread of coronavirus. In future, the similar mechanism can be designed with more attributes and using other data mining tools.

REFERENCES:-

1. Pakistan's Response to COVID-19: Overcoming National and International Hypes to Fight the Pandemic - PMC (nih.gov)
2. A Nationally Representative Survey of COVID-19 in Pakistan, 2021–2022 - PMC (nih.gov)
3. Pakistan COVID-19 Dataset - dataset by dekhpakistan | data.world
4. MICRODATA (COVID-19) | Pakistan Bureau of Statistics (pbs.gov.pk)
5. Survey on COVID-19 | Pakistan Bureau of Statistics (pbs.gov.pk)
6. Medical Data Classification with Naive Bayes Approach (scialert.net)
7. (PDF) Data clustering application in medicine (researchgate.net)
8. Clustering of COVID-19 data for knowledge discovery using c-means and fuzzy c-means - PMC (nih.gov)