Crowd Prediction at Various Public Places for Covid-19 Spread Prevention Employing Linear Regression

S. Md. Riyaz Naik[1], Farooq Sunar Mahammad[1*], Mulla Suleman[2], S Danish Hussain[3], S Waseem Akram[4], Narasimha Reddy[5], N Naresh[6]

Department of Computer Science Engineering, Santhiram Engineering College, Nandyal
Email: riyaz.cse@srecnandyal.edu.in

ABSTRACT

Input from a simple sensor-based people counters module is taken as training data for a Linear Regression and a K-Means model for predicting crowds at various public places. The influencing factor can be either time or distance and the outcome factor be crowd strength at a place. By training a Linear Regression and a K-Means algorithm, an efficient predictive model for crowd prediction can be created.

Keywords: Linear Regression, K-Means, Prediction, Machine learning

1. INTRODUCTION

The covid-19 pandemic took a huge toll on human life and the economy. It was a wake-up call to the global community to take care of health as the highest priority. Managing hospitals and health departments during such a pandemic situation has become a big challenge. The Covid-19 pandemic has created a large number of acutely demanding situations for healthcare transport agencies, which include inadequate capability, delivery shortages, want for care redecorated, and monetary loss. Complex technological know-how perspectives health care transport agencies as complicated adaptive structures that perform in exceptionally complicated and unpredictable environments. The perspective assumes that plenty of organizational existence is unknowable, unsure, or unpredictable and as a result can not be standardized and controlled. This had created a call for an easy but powerful software program to manipulate covid-19 unfold prevention. This software program can be utilized by the authority's health departments.

2. LITERATURE SURVEY

Primarily based on the inference from our reference base paper we got here to recognize that "the system includes three cellular crowdsourcing applications that permit the general public, medical doctors and pharmacies to document illnesses and drugs sales in actual time. Records concerning illnesses for the year 2017 have been retrieved and the corresponding each day weather statistics specifically temperature, humidity and wind for that 12-month changed then extracted and delivered to this dataset. An artificial Neural network (ANN) was changed into then trained with this dataset which was used to expect the propagation of the illnesses which can be monitored. The prediction changed into accomplished primarily based on the number of suggested sicknesses in the modern-day day together with climate forecasts for the impending days and the effects were promising. The model has been evaluated ensuing in an accuracy of 90%. Eventually, we accept as true with that this type of system can be very useful to the ministry which could then make knowledgeable decisions to counteract the viable propagation of diseases."[1]

Based on the review of the literature we propose an efficient software model using linear regression and K-Means algorithms.
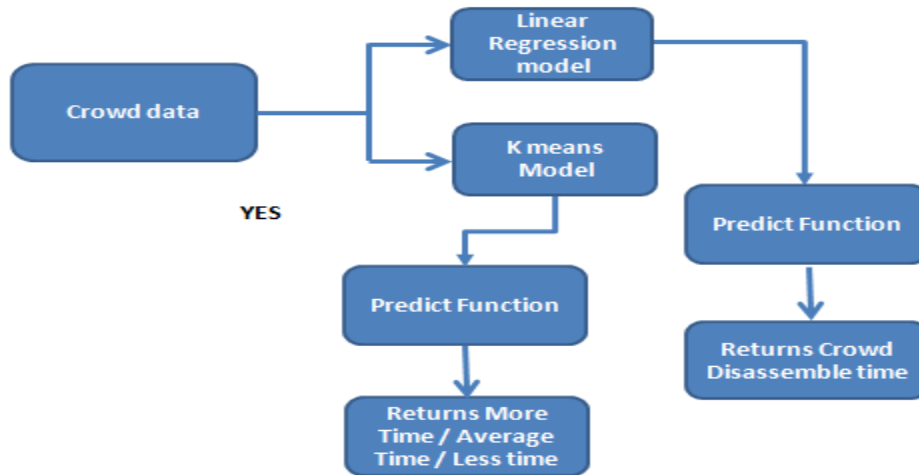
3. WORKING

Proposed system:

Figure 1 Block Diagram

In this project, the crowd strength at various public places with respect to the time is calculated by the linear regression model obtained by training it with a real-time dataset. Starting from the afternoon at 3 PM, as time goes by the crowd strength increases. Linear Regression is a regression machine learning algorithm. In this project, two-dimensional Linear Regression is employed. That is, one input parameter and one output parameter are taken and the given dataset is plotted in a 2D graph. The output model of Linear regression is the best fit line, which is the line with 'y = mx + c' equation passing closest to all the given points of the dataset. The model of this machine learning algorithm with just one input and one output parameter can be described with the value of the slope 'm' and the value of the y-intercept 'c'. This algorithm maps the proportionality, be it direct or inverse, between the input and the output parameter.

Linear Regression Model:

"This method showcases the courting among the impartial linear variable (X-axis) and the structured variable (Y-axis). It is applied to get the value based on the value of another number. If there's a solo enter number (x), it's uncomplicated regression. However, if there are a couple of entering digits, it is multiple regressions. This method version offers a sloped straight line describing the connection inside the numbers." [2]
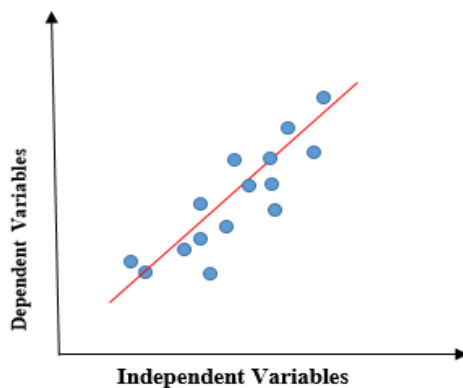

Figure 2 Linear Regression graph

"This method showcases the courting among the impartial linear variable (X-axis) and the structured variable (Y-axis). It is applied to get the value based on the value of another number. If there's a solo enter number (x), it's uncomplicated regression. However, if there are a couple of entering digits, it is multiple regressions. This method version offers a sloped straight line describing the connection inside the numbers." [2]

K-Means Model:
"K-means Clustering is an unmanaged mastering algorithm that is used to resolve the clustering problems in machine learning or records technology. In this subject matter, we are able to examine what the k-approach clustering algorithm is, and the way the algorithm works, in conjunction with the Python implementation of k-method clustering. It lets us cluster the information into distinct groups and is a handy manner to find out the kinds of groups within the unlabeled dataset on its own without the want for any schooling. It is a centroid-primarily based algorithm, in which each cluster is related to a centroid. The principal purpose of this set of rules is to decrease the sum of distances among the statistics factors and their corresponding clusters. The algorithm takes the unlabeled dataset as input, divides the dataset into k-quantity of clusters, and repeats the system until it no longer discovers the first-rate clusters. The value of the k needs to be predetermined on this algorithm. Determines the excellent value for k-center factors or centroids through an iterative technique. Assigns each information point to its closest k-center. the one's data points which might be close to the specific k-center create a cluster."[3]
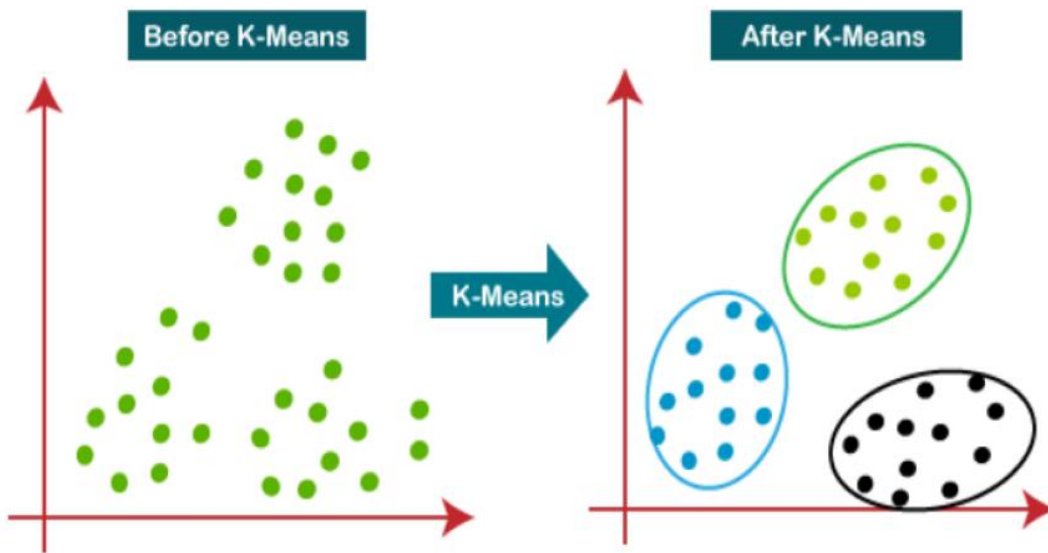

Figure 3 K-Means graph

How does the K-Means Work?
"The operation of the k-way set of rules is defined within the under steps.
Step-1: pick out the quantity k to determine the wide variety of clusters.
Step-2: pick random k factors or centroids. (It could be different from the input dataset).
Step-3: Assign every data factor to its closest centroid, if you want to form the predefined k clusters.
Step-4: Calculate the variance and area of a brand-new centroid in every cluster.
Step-5: Repeat the 1/3 steps, which means reassigning every record point to the brand-new closest centroid of every cluster.
Step-6: If any reassignment happens, then visit step-4 or else pass to finish.
Step-7: The version is prepared." [3]

4. ARTIFICIAL INTELLIGENCE

The algorithmic glide provided in this paper falls underneath the field of computing Intelligence and machine studying. supplied under is a definition of these technologies.

"Robotic or artificial intelligence (AI) is intelligence tested via machines, unlike the natural intelligence displayed with the aid of humans and animals, which entails cognizance and emotionality. The difference between the previous and the latter classes is frequently found via the acronym selected. Robust AI is commonly labeled as artificial general intelligence (AGI) whilst trying to emulate 'natural' intelligence was referred to as artificial biological intelligence (ABI). Main AI textbooks outline the sphere because of the study of sensible retailers: any device that perceives its surroundings and takes moves that maximize its risk of achieving its desires. Colloquially, the term artificial intelligence is regularly used to explain machines that mimic cognitive features that people partner with the human thoughts, which includes mastering and trouble fixing."[4]

"As technologies end up increasingly more successful, duties taken into consideration to require intelligence are frequently eliminated from the definition of AI, an occurrence called the AI impact. A quip in Tesler's rule state AI is anything that hasn't been executed yet. As an instance, optical individual recognition is often excluded from matters taken into consideration to be AI, having emerged as a habitual technology. Modern-day system talents typically categorized as AI encompass efficiently expertise in human speaking abilities, competing at the highest degree in strategic recreation structures (consisting of chess and move), and additionally imperfect-records video games like poker, self-using vehicles, smart routing in content material transport networks, and military simulations." [4]
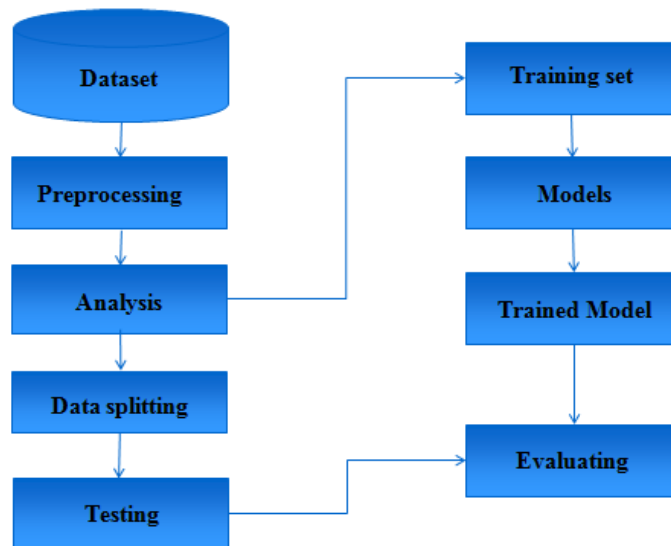

Figure 4 Flow Chart

"AI studies have been divided into subfields that frequently fail to talk with each different. those sub-fields are primarily based on technical issues, along with precise desires (e.g. robotics or system studying), the usage of unique equipment (good judgment or synthetic neural networks), or deep philosophical variations. Sub-fields have additionally been based totally on social elements (particular establishments or the work of specific researchers)." [4]

"Within the twenty-first century, AI techniques have skilled a resurgence following concurrent advances in pc strength, big quantities of records, and theoretical expertise; and AI strategies have to end up an essential part of the technology enterprise, assisting to clear up many difficult troubles in computer science, software engineering and operations studies." [4]

5. MACHINE LEARNING

"Machine learning (ML) involves the research of computing algorithms that improve themselves over the duration as a consequence of expertise plus information. It's considered an element of computer intelligence. Artificial learning algorithms create a framework relying on retraining information to generate forecasts or judgments despite having to get specifically configured to accomplish it. Machine intelligence algorithms were utilized across a vast range of applications, including healthcare, spam filtration, audio identification, and desktop sight, where developing algorithms to perform those required, jobs are complicated or impossible." [5]
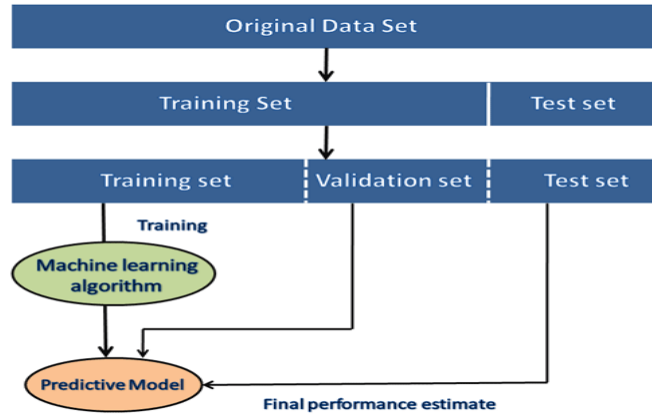

Figure 5 Machine learning dataset

"Machine intelligence and computing analytics are intimately linked, which concentrates upon utilizing computers to make forecasts; nevertheless, nearly most computer intelligence involves statistical analysis. The discipline of machine intelligence benefits from computational efficiency research since it provides tools, concepts, and applicable fields. Data extraction is a similar branch of research that focuses on uncontrolled learning for interactive data processing. Computer intelligence is also known as anticipatory analytics when intelligence is used to solve commercial challenges." [5]

5. INDUSTRY 4.0

"This same Fourth Economic Transformation (4IR aka Industrial 4.0) involves the continued digitization of old commercial and international operations, utilizing new intelligent systems, according to several websites. Regarding increasing robotics, massive device connectivity (M2M) plus the network interconnected objects (IoT) are combined enhanced personality and interaction, as well as the development of intelligent technologies which can assess and resolve problems without any requirement of living thing interaction." [6]

Automation under Industry 4.0 has a particular schema or pattern at its outset. Presented below is how automation in the mass production industry, as well as consumer-level products, are built-in in today's technological era.

The schema presented in Figure 6 has a lot of other components involved but the generic outline of it stands justifiable for all kinds of automation today.
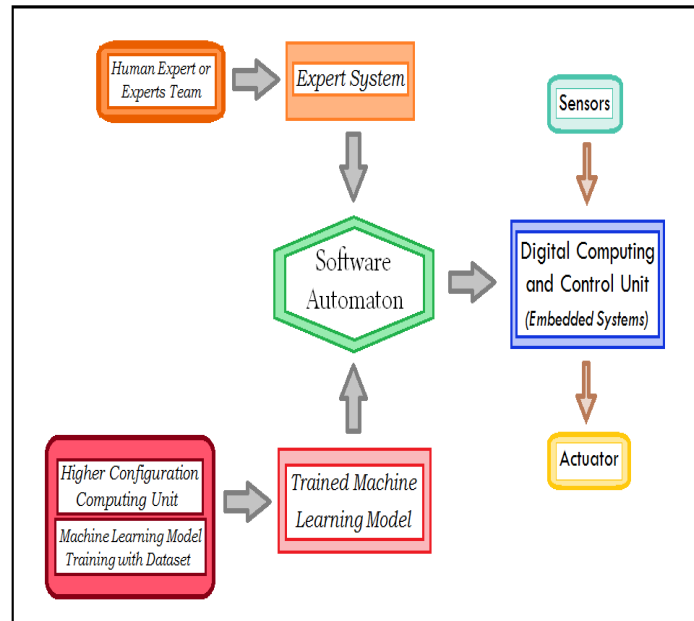
Figure 6 Schema of Automation

The software automation of the regular automation model, which is the status quo, was built by a human expert or a team of human experts till now. With the advent of machine learning technology, the software automaton was not fully directly designed by human experts. The human experts build the machine mastering software and give the real-world data set as training information. Within the formats of a numerical framework, a computer intelligence device finds statistical connections between the intake and outcome variables of the database. This mathematical model can be downloaded as a working software module to other electronic computing devices. This mathematical model is referred to as the 'trained machine intelligence module'. The software automaton of all the current digital implanted devices is a mathematical model that gives a numerical output for a numerical input based on arithmetic and logical conditions. This software automaton, as explained above can be either directly developed by a set of human experts by means of setting the boundary conditions themselves based on observation and requirement or can be downloaded as an executable module from machine mastering training systems that are trained with the relevant dataset. In whatever way the software automaton is developed, it can be loaded onto the relevant embedded computing module that can be used for either sensor-based closed-loop automation or open-loop automation.

The technological components of Industry 4.0 include IoT, augmented reality, virtual reality, cloud computing, 3D printing, big data analytics, networking, data security, human-machine interaction and others. IoT is a very effective way to collect real-world data. Sensors integrated with data acquisition and transmission systems can be placed anywhere and the collected data can be pre-processed if required and used as datasets to train machine learning models.

Cloud computing is employed for the optimized utilization of computing resources. There are many third-party vendors like Google and Amazon which are very reliable in terms of data security and speed of computation. These services offer companies and organisations a cheap and reliable way to leverage digital intelligence as well as machine learning to their advantage.

Analytics on big datasets is the set of technological components involved with collecting, collating and managing large quantities of data for analytics and decision making. When so much data is involved, especially with third-party service providers, data security plays an important role.

One of the paramount concerns about Industry 4.0 is the unemployment it can create due to powerful automation. The field of

human-machine interactions and co-working has been a very developing field now to mitigate the above-mentioned problem.

6. RESULTS AND DISCUSSION

Time starting from 3:00 PM	Crowd in Closest Food Court
3	47
3.76	43
3.9	56
4	70
4.1	100
4.7	123
4.86	135
5.4	200
5.55	254
6.75	297
6.84	284
7.43	302
7.86	350
8.2	396
8.33	405

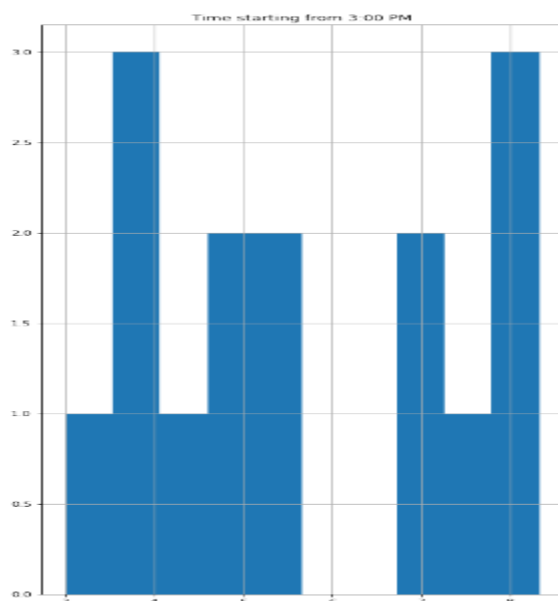**Figure 7 Input Sample Dataset**

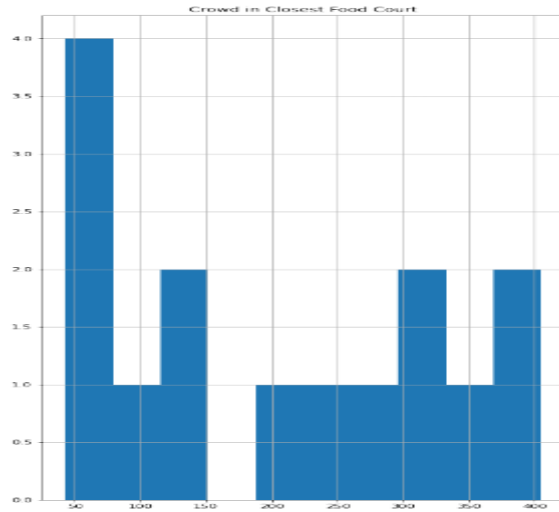**Figure 8 Crowd prediction starting from 3 PM**

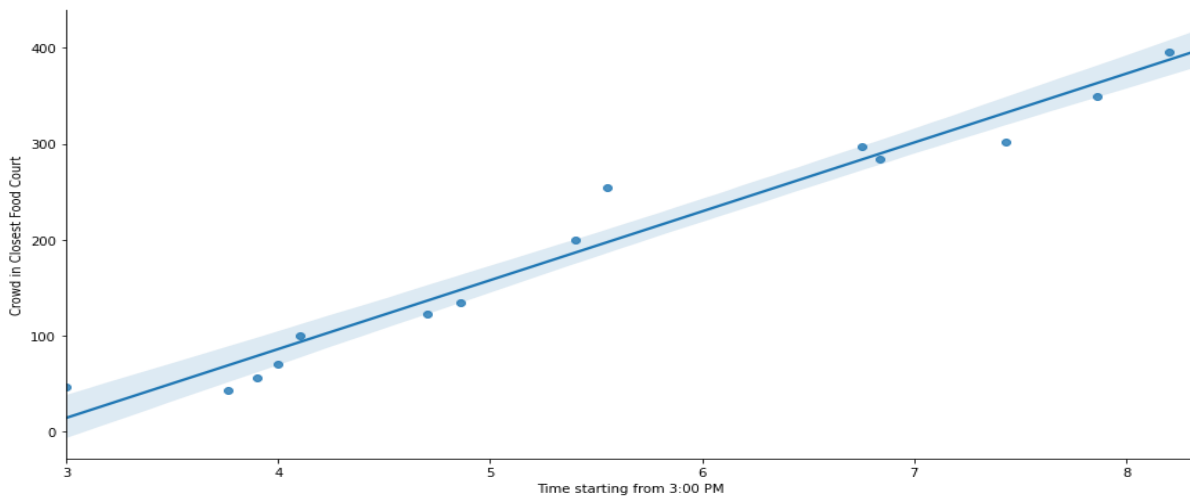**Figure 9 Crowd in the closest food court**

Figure 10 Output graph

This machine learning-based predictive model is implemented by employing Python programming language. The relevant library files have been included for the execution of the code. The dataset was given as input and the predictive model was obtained. The prediction model was tested and the results were satisfactory. The output graph of the machine learning model has been presented above. In this project, a Linear Regression algorithm is employed. The two parameters are the time starting from 3:00 PM and crowd strength in the closest food court from an office complex. The Linear Regression model maps the direct proportionality between the two above mentioned parameters.

7. CONCLUSION AND FUTURE WORK

This software predicts how much crowd will gather at a particular place at a given time in the evening based on a trained machine intelligence model. This software automaton can be used by government agencies like the health department and police department. A sufficient number of health workers can be dispatched to control the crowd gathering at such places. This software automaton can be made more real-world by adding more practical parameters and increasing the dataset size.

Automation under Industry 4.0 has a particular schema or pattern at its outset. The software automaton of the conventional automation model, which is the status quo, was built by a human expert or a team of human experts till now. With the advent of machine learning technology, the software automaton was not fully directly designed by human experts. The human experts build the machine learning software and give the real-world data set as training information. The machine-learning software program identifies the sample between the input and the output parameters of the dataset in the shape of a mathematical version. This mathematical model can be downloaded as a working software module to other electronic computing devices.

REFERENCES

[1] Pratima Kurtah, Yusrah Takun and Leckraj Nagowah, "Disease Propagation Prediction using Machine Learning for Crowd sourcing Mobile Applications", July 2019, Conference: 2019 7th International Conference on Information and Communication Technology (ICoICT)

[2] https://www.analyticsvidhya.com/blog/2021/06/linear-regression-in-machine-learning/

[3] https://www.javatpoint.com/k-means-clustering-algorithm-in-machine-learning

[4] https://en.wikipedia.org/wiki/Artificial_intelligence

[5] https://en.wikipedia.org/wiki/Machine_learning

[6] https://en.wikipedia.org/wiki/Fourth_Industrial_Revolution