JOIV : Int. J. Inform. Visualization, 8(1) - March 2024 107-114

INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

Dark Web Financial Fraud Identification Using Mathematical Models in Healthcare Domain

Anand Singh Rajawat [a], S.B. Goyal [b,*], Ram Kumar Solanki [a], Amit Gadekar [c], Dipak Patil [d]

[a] School of Computer Sciences and Engineering, Sandip University, Nashik, 422213, India
[b] Faculty of Information Technology, City University, Petaling Jaya, 46100, Malaysia
[c] Sandip Institute of Technology and Research Centre, Sandip University Nashik, 422213, India
[d] Sandip Institute of Engineering and Management, Sandip University Nashik, 422213, India
Corresponding author: *drsbgoyal@gmail.com

***Abstract*— The so-called "dark web" has emerged as the most trustworthy platform for thieves to launch their enterprises. The healthcare industry has become a haven for illegal activities such as the sale of medical gadgets, trafficking in human beings, and the purchase of organs. This is because the sector provides a high level of privacy, which makes it an ideal location for engaging in unlawful operations. In this field of research, linear regression is utilized to uncover previously unknown patterns in customer demand. A vector will be created using a time series of medical equipment purchases to do this. When we look at the data the case firm gave us, we notice that people tend to desire to purchase products in one of three ways. After that, we sort the hospitals into groups according to the course of the trend vector by employing a technique known as "hierarchical clustering," which we apply to the data. According to the research findings, the trend-based clustering method is an excellent way to partition hospitals into subgroups that share similar tendencies. According to our model evaluations, no one model can reliably produce the most accurate forecasts for each cluster when used by itself. Some models can be utilized to make accurate predictions, and these models apply to a wide variety of time series that exhibit various patterns.**

***Keywords*— Dark web; fraud identification; mathematical models; healthcare.**

Manuscript received 5 Dec. 2022; revised 29 Jul. 2023; accepted 29 Sep. 2023. Date of publication 31 Mar. 2024.

I. INTRODUCTION

The "dark web" is a part of the internet that can't be found using the usual ways to discover the web. Some of the illegal things that can be done with its help are selling drugs and guns and stealing money. Financial theft is a big problem in the medical field. In 2021, fraud in the financial sector is expected to cost healthcare businesses $10 billion. In a case like this, there are many ways to lie. Fraud in the healthcare business on the dark web can be found using mathematical techniques. These programs can sort through the information on the dark web to look for warning signs that point to scams. The logistic regression model [1], [2] is a mathematical model that can help with this task. Using this method, a person can understand how likely a specific deal will involve fraud. The support vector machine (SVM), which stands for "support vector machine," is another model that shows this. This model looks at some different factors to decide whether or not the transaction is fraudulent. In the healthcare business, financial fraud that takes place on the dark web can be found with the help of mathematical models [3]. But there are still some problems that need to be talked about and fixed. The dark web is a bit of a struggle because it changes all the time. This means that the mathematical models will need to be changed all the time so that they can keep up with the latest changes. Because the info from the dark web is noisy, it can't be trusted, which is another problem. Because of this, the math models used to predict what will happen need to account for some errors. Even with these problems, mathematical modelling is starting to look like it could be an excellent way to find financial fraud on the dark web in the healthcare business. By using mathematical models, doctors and hospitals can protect their patients from economic loss and make sure they are safe.

Here are three objectives for Dark Web Financial Fraud Identification Using Mathematical Models in the Healthcare Domain:

- To use mathematical models to find fraudulent activity in the healthcare field.

- To propose a mathematical model that can uncover illicit online financial scams.
- To use preexisting machine learning (Logistic regression and SVM) techniques for an impact analysis of healthcare fraud

II. Materials and Method

This study determined what kinds of financial services can be found on the Dark Web, what those services are, and how likely people are to be scammed. The results of this study were put together to teach law enforcement and the public about how criminal behavior on the Dark Web is constantly changing. To reach this goal, we will build a link between cybercrime studies and well-known texts in criminology.

Jung et al. [4] studied the different kinds of fake financial services, goods, and schemes on the Dark Web. The idea of "regular activity" in the financial market on the Dark Web is studied to learn more about the basic things that lead to fraud and scams. The data show how important it is to raise public awareness, programs to protect consumers, and law enforcement control of the financial sector in the fight against cybercrime.

Nguyen et al. [5] recommend several different strategies that can be used to detect fraudulent behavior more effectively. It focuses on enhancing detection skills by modifying existing characteristics and establishing new ones, as well as coping with datasets with significant variation. By improving the effectiveness of fraud detection algorithms, these proposals contribute to the ongoing battle against financial crime on the Dark Web.

Acin [6] shows how the Dark Web and sites like Dream Market have helped the online black market grow. Even though Dream Market is well-known, its closing is not expected to significantly affect the value of the illegal market as a whole, which is still doing well. This article shows how the criminal ecosystem on the Dark Web is growing. Hackers can buy tools on the dark web to attack people and businesses. By drawing attention to the above environment, this piece also shows how vital strong cybersecurity measures are.

Liu et al. [7] studied how important it is to ensure financial fraud detection tools are tailored to each business. The results show that government officials, investors, and auditing agencies can use customized models to find examples of unethical financial behavior in companies traded on the stock market. Because these models are so good at finding and stopping fraud in the business world, they are directly to blame for the improved financial stability from this success.

Wilson [8] investigated how important secret networks and protected data are on the Dark Web. Companies face big risks because private information can be bought on the Dark Web. This information includes details about clients, intellectual property, and financial operations. The results of this study show how important it is to keep sensitive information from getting into the wrong hands while it is being kept or sent online.

This research, when taken as a whole, contributes to our growing body of information regarding financial fraud on the dark web and its implications for economies worldwide. These studies assist law enforcement agencies, businesses, and the general public in gaining a better understanding of how to recognize illegal conduct on the Dark Web, how prevalent it is, and the reasons why it is critical to take preventative measures.

Fraudulent The Use of Mathematical Models in Inventory Control and Cybersecurity in the Age of Suspicious Financial Transactions and the Deep Web. Due to the rise of new technologies and the dark web, financial scams and hacks are more common and dangerous in the digital age we live in now. This is partly because current technologies work better than they used to. The Dark Web is a hidden part of the Internet and can only be reached through safe pathways. It has become a center for high-level criminal activity, which has made the types of money theft that used to be straightforward a lot harder. In this situation, mathematical models are beneficial because they use symbols, numbers, and pictures to show how complicated systems work and how they are connected. With the help of mathematical models, businesses can strengthen their defenses against possible threats by improving product management and digital security.

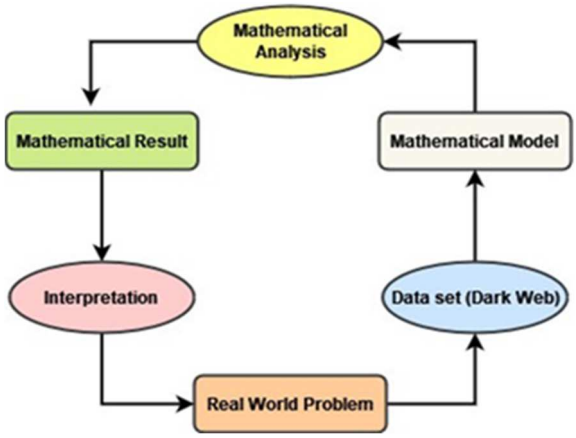

Fig.1 Process of Mathematical Modelling

Keeping the opportunity cost of holding on to items to a minimum requires good inventory management. Companies can use mathematical models to do "what-if" analyses, determine how much stock to keep on hand, and develop effective ways to keep costs low while making the most money. Stock theory gives managers and executives tools, like mathematical proofs and calculations, to help them run and handle their businesses well. This helps make things run more smoothly and effectively. The goal of illegal financial deals is almost always to improve one's financial situation. This is especially true when it comes to money scams. Cybercriminals sell stolen information or use it to get into bank accounts or steal identities. As the financial sector depends increasingly on digital technologies like online banking, virtual currencies, and fund transfers, institutions must take the necessary steps to protect themselves from these threats. Inventory, as defined by researchers in the field of inventory management, is the total number of things that can be sold, made, or kept for later use. The goal of inventory management is the same whether you call the stock "work in progress" or "stock in trade": to meet customer wants and help the business grow. Inventory management tools are very important to businesses when it comes to keeping track of stocks and making sure deliveries happen on time. Lastly, the Dark Web [9] [10] is always changing, and fake financial operations and cyberattacks are becoming more common.

This is a big problem for companies all over the world. Use mathematical models for inventory control and cyber defense. This is one of the most important things you can do to protect yourself from these dangers. Using mathematical analysis and reasoning tools, businesses can improve their inventory, boost security, and protect their financial assets. Ultimately, this makes it easier for the companies to deal with new problems [11], [12].

A. Dark Web Financial Fraud Identification Model Developments

Figure 1 shows the process of naming items during one round of the running cycle. At the first time 't' = 0, a preliminary replenishment of Q devices is done, S devices are shipped to fill back-orders, and a maximum of I devices are kept in the original offering. These things happen at the same time. As time goes from 't' = 0 to 't' = 1, zero-time gadgets slowly decrease due to use and wear and tear until none are left [13]. This state is when backlogged products are finally restocked through subsequent replenishment at Tt, where Tt is the length of the backlog and the resulting sales losses.

B. Pricing Mathematical Model

To assist the analysis of inventory management and related expenses, we have assumed the following assumptions and notations in this study:

- C1 stands for the average cost of holding a unit of inventory over time, which is the same for all units.
- C2: Indicates the cost associated with device shortages on a per-device, per-time basis.
- C3: This represents the average price paid for a single item of stock.
- The cost per unit of the device, denoted by C4, represents the expense incurred by the business.
- C5: Represents the revenue loss cost due to a device not being available or delayed delivery.
- The quantity of stock on hand at a specific point in time 't' is expressed by the variable I(t).
- Demand Price, or D(t), is the selling price of an item at a time 't' larger than zero.
- A cycle's length, represented by the symbol 'T,' is the time needed to perform a whole inventory management cycle.
- S: Stands for the consistent quantity used in inventory management.
- The lag time, abbreviated tT, is a valuable indicator of bottlenecks and delays in inventory restocking.
- Refers to the efficiency of instantaneous velocity as calculated by Weibull's three-parameter distribution (I1).
- When the inventory parameter is non-zero, the associated scaling parameter is considered.
- If the amount in the storage facility is zero, but the amount on hand is more significant than zero, we write the latter as 0.
- Time is a sign of deterioration since it represents the depletion of resources over time.
- During the time interval t1, there will be zero inventory.

Our research aims to help supply chain and logistics professionals make educated decisions by thoroughly examining inventory management, related expenses, and the impact of various factors on the inventory level, all as indicated by the notations and assumptions as follows:

Assumptions

- I The investigated inventory procedure is a one-unit operation.
- The Horizon Time, H, is unlimited, and there is a well-established protocol for preparing cycles of time T.Demand fee is characteristic of stock, i.e.,

$$D(t) = \begin{cases} \alpha + \beta I(t), I(t) > 0 \\ \alpha, \ I(t) \leq 0 \end{cases} \quad (1)$$

- Replenishment lead and instantaneous time is 0.
- Deteriorated devices are not updated/repaired during certain cycle.
- Holding price, buying price, shortage expense as well as device cost stay constant over time. Which $\alpha >$ zero stands for the original demand rate as well and $\beta >$ zero is the coefficient of inventory impact on product sales.
- Distinctive rate of decay is based on interval, $0< t < 1$; which seems sensible due to improvements in contemporary storage facility.
- Distribution of period to degeneration of the things follows three-parameter Weibull's Distribution,

Deficiencies in stock are allowed and accumulated to ensure that at inventory amount 0, a few arrived interests are permitted. We anticipate the backlogging fee depends upon the length of hanging time, which is small for the following replenishment. The scope of customers who would probably wish to accept delay buy in the current moment, 't' diminishes with the keeping up time, Tt prior to the following replenishment. Consequently, to determine the backlogging number as B('t') for Tt <1.

Following the above assumptions and notations, Weibull's distribution function is, i.e.,

$$f(t) = \alpha\beta/(t - \gamma)1\text{- }\beta e\ [\alpha+(t-\gamma)\ \beta], t > 0 \quad (2)$$

The instantaneous speed functionality is as follows:

$$\theta(t) = \alpha\ \beta/(t - \gamma)1\text{- }\beta. \quad (3)$$

The accounting process begins with Q devices of the item, and the quantity of listed inventory is impacted by the speed of product depletion from the device. During time intervals [6], [0, γ], inventory depletes due to demand. Just what about the interval, [γ, t1]? The listing depletes thanks primarily to demand and partially due to deterioration. With time t_1, the inventory amount is 0, and most of the need hereafter, i.e., (T − t1), is partly backlogged. The next following replenishment renews a whole number of back ordered goods; it is that assumed deterioration time for products follows three parameters of Weibull's distributions. The graphical representation of which is discussed [7] listing device is provided in Figure 1. For exactness, it is supposed that not merely opportunity cost per device C_3 is higher compared to device buy expense C_4, but additionally f(t) is log-concave and positive in (0, H], it is concluded that entire number of backorders throughout it cycles as

$$\int_{s_{i-1}}^{t_i} \frac{f(\mu)}{1+\propto(t_i-u)} du, \text{ where } s_{i-1} \leq t \leq t_i, 1 \leq i \leq n, \quad (4)$$

and order quantity of replenishment cycle as follows:

$$Q_i = \int_{s_i-1}^{t_i} \frac{f(t)}{1+\propto(t_i-t)}dt + \int_{t_i}^{s_i} e^{-\theta(t_i-t)}\ f(t)dt,\quad 1 \le i \le n, \quad (5)$$

C. Model with Zero I_m and Problem Description

Here, the merchant receives I_m devices at 't' = zero. Thus, the listing begins with I_m devices at beginning of each cycle, then steadily depletes to zero for $t = T_1$ because of the mixture impact of deterioration and need. A graphical representation [8], [14] of the deemed catalogue device is provided may be drawn to time interval [zero, T_d], the listing does not have any deterioration. Deterioration happens to time interval [T_d, T_1] in a continuous deterioration pace θ. Thus, the changes in the listing at any time 't' following differential equations.

$$\frac{d[I_1(t)]}{dt} = -(a + bI_1(t)),\ 0 \le t \le T_d \quad (6)$$

$$\frac{d[I_2(t)]}{dt} + \theta I_2(t) = -(a + bI_1(t)),\ T_d \le t \le T_1 \quad (7)$$

$$\frac{dI_3(t)}{dt} = \frac{-a}{1+\delta(T-t)},\ T_1 \le t \le T \quad (8)$$

With boundary values
$(0) = I_m$,
$(t_1) =$ zero
And
$(t) =$ zero.
The solution of these differential equations is as follows:

$$I_1(t) = \frac{a}{b}(e^{-bt} - 1) + I_m e^{-bt}, 0 \le t \le T_d \quad (9)$$

$$I_2(t) = \frac{a}{b+\theta}(e^{(\theta+b)(T_1-t)} - 1) + I_m e^{-bt}, T_d \le t \le T_1 \quad (10)$$

$$I_3(t) = \frac{-a}{\delta}[\log(1+\delta(T-T_1)) - \log(1+\delta(T-t))],\ T_1 \le t \le T_d \quad (11)$$

D. Fraud Requirements Detection and Modelling Formulation

Here, the stock framework developments as pursues m I devices of items contact base in the stock framework toward the beginning of every cycle. Of time interim [0, t_d], the stock amount is diminishing [11] simply inferable from the interest rate. The stock amount is dropping to 0 due to decay [12] as well as fascination throughout time interim [t_d, t_1]. At that time, the deficiency interim maintains as far as you possibly can of the existing request cycle. The whole treatment is rehashed. Within view of the above-mentioned portrayal, the differential state [13] talking on the stock status is provided as follows:

$$\frac{dI(t)}{dt} = \begin{cases} -(a+bt+ct^2) & 0 \le t \le t_d \\ -(a+bt+ct^2) - \theta I_2(t) & t_d \le t \le t_1 \\ -\left(\frac{1}{B}\right)\overline{[1+\delta(T-t)]}power(-1) & t_1 \le t \le T \end{cases} \quad (12)$$

With BC's I (0) = I_m, I (t_1) =0. The solution of Eq. (3.48) is as follows:

$$I(t) = \begin{cases} I_1(t) & \text{if } 0 \le t \le t_d \\ I_2(t) & \text{if } t_d \le t \le t_1 \\ I_3(t) & \text{if } t_1 \le t \le T \end{cases} \quad (13)$$

$$\frac{dI_1(t)}{dt} = -(a+bt+ct^2),\ 0 \le t \le t_d \quad (14)$$

$$\frac{dI_2(t)}{dt} + \theta I_2(t) = -(a+bt+ct^2), t_d \le t \le t_1 \quad (15)$$

$$\frac{dI_3(t)}{dt} = \frac{-B}{1+\delta(T-t)}, t_1 \le t \le T \quad (16)$$

Solution of (3.50), (3.51), (3.52) is

$$I_1(t) = -\left(at + \frac{bt^2}{2} + \frac{ct^2}{3}\right) + I_m, 0 \le t \le t_d \quad (17)$$

$$I_2(t) = \left(\frac{2c}{\theta^3} - \frac{b}{\theta^2} + \frac{a}{\theta}\right)(e^{\theta t_1} - e^{\theta t}) + \left(\frac{b}{\theta} + \frac{2c}{\theta^2}\right)(t_1 e^{\theta t_1} - te^{\theta t}) + \frac{c}{\theta}(t_1{}^2 e^{\theta t_1} - t^2 e^{\theta t}),$$

where $t_d \le t \le t_1$ (3.54)

$$I_3(t) = \left\{\frac{-B}{\delta}\begin{bmatrix}\log(1+\delta(T-t_1)) \\ -\log(1+\delta(T-t))\end{bmatrix}\right\},\ t_d \le t \le t_1 \quad (18)$$

The continuity of d/dt [I(t)] at t, gives

$$I_m = at_d + \frac{bt_d^2}{2} + \frac{ct_d^3}{3} + \left(\frac{2c}{\theta^3} - \frac{b}{\theta^2} + \frac{a}{\theta}\right)(e^{\theta t_1} - e^{\theta t_d}) + \left(\frac{b}{\theta} - \frac{2c}{\theta^2}\right)(t_1 e^{\theta t_d} - t_d e^{\theta t_d}) + \frac{c}{\theta}(t_1^2 e^{\theta t_d} - t_d^2 e^{\theta t_d})$$

E. Dark Web Financial Fraud Cycle Formulation

The inventory cycle begins for 't' = zero with the original inventory amount of I_{max} units of time interval (zero,) t_μ the inventory amount reduces just owing to demand. The need for the item is time reliant. When during period (t, t_μ) inventory is dropping to 0 because of demand speed as well as deterioration each and length of one cycle (t, T) shortage begins and on account [15] of partial backlogging certain sales are lost. The differential equations associated the change of device are

$$\frac{dI_1(t)}{dt} = -(a+bt),\ 0 \le t \le t_\mu \quad (19)$$

$$\frac{dI_2(t)}{dt} + \theta t I_2(t) = -(a+bt),\ t_\mu \le t \le t_1 \quad (20)$$

$$\frac{dI_3(t)}{dt} = -(a+bt)/e^{\delta(T-t)},\ t_1 \le t \le T \quad (20)$$

with BC's

$I_1(0) = I_{max.}$, and $I_2(t_1) = 0$,

F. Formulation Of Retailer's Cost Model

A mathematical model forever time product with immediate design request produced. The cycle begins with an important degree of inventory i.e., I_{max} devices, level of listing drops at a faster pace in basic period (0, 1) due to curiosity and crumbling, when it touches base value 0 quality at precious time $t = t_1$. Currently deficiencies happened for time (t_1, T) that is fairly delay bought with time ward backlogging fee. At the end of cycle, the inventory completed a most

elevated quantity of deficiency S and once that brand new request is placed to zero-in the excess [16] [17]. The adjustment in inventory degree I (t) as for point eventually is as given:

$$d/dt(I) = -(a + bt), 0 \leq t \leq \mu_0 \quad (22)$$

$$d/dt(I) + \theta(t)I(t) = -(a + bt), \quad \mu_0 \leq t \leq t_1 \quad (23)$$

In this section, the objective is to create probably the most excellent estimations of T as well as t_1 that limit Z (T, t_1). For a foreordained estimation of T, taking second and ist request subsidiaries of Z (T, t_1) regarding t_1,

G. Sensitivity Analysis with Respect to Parameters

An evaluation of how sensitive a system is to change can be carried out using a technique known as sensitivity analysis [15]. This technique involves changing just one parameter at a time while continuing to maintain track of the other variables in their individual states. The results of an investigation into numerous facets of influence are shown in the tables that may be seen below:

TABLE I
SENSITIVITY ANALYSIS FOR DEMAND PARAMETER B

% Change	T *	t_1*	Q*	Z *(T *, t_1*)
-40	17.7268	15	2130.06	1910.91
-20	17.7244	15	2133.86	1915.67
0	17.7219	15	2137.66	1920.43
+20	17.7195	15	2141.46	1925.19
+ 40	17.7171	15	2145.26	1929.95

TABLE II
SENSITIVITY ANALYSIS FOR LIFETIME PARAMETER 0

% Change	T *	t1*	Q*	Z*(T*, t1*)
-40	18.9940	15	2431.79	2616.33
-20	18.3755	15	2292.30	2291.17
0	17.7219	15	2137.66	1920.43
+20	17.0623	15	1978.95	1514.85
+ 40	16.4322	15	1827.27	1094.59

TABLE III
SENSITIVITY ANALYSIS FOR BACKLOGGING PARAMETER A

%Change	T *	t1*	Q*	Z *(T *, t1*)
-40	17.1836	15	2137.66	1947.73
-20	17.4456	15	2137.66	19
0	17.7219	15	2137.66	1920.43
+20	18.0100	15	2137.66	1906.54
+ 40	18.3084	15	2137.66	1892.67

TABLE IV
SENSITIVITY ANALYSIS FOR DETERIORATION PARAMETER Θ

%Change	T *	t1*	Q*	Z *(T *, t1*)
-40	16.7224	15	1887.10	1292.33
-20	17.2138	15	2012.38	1610.94
0	17.7219	15	2137.66	1920.43
+20	18.2478	15	2262.94	2221.03
+ 40	18.7924	15	2388.22	2512.94

III. RESULTS AND DISCUSSION

During this investigation, we collaborated with a company that specializes in the production and distribution of medical equipment and has developed extensive connections with hospitals located all over India. Our investigation was restricted to a few different medical devices, and the demand for data provided by the manufacturer served as the primary source for our findings. The primary purpose of this research was to develop a data-driven prediction model that could be used for supply chain planning, when applied to an environment that was comparable to the dark web. For reasons of privacy, each piece of information has been discretized within the range [0, 1].

To determine whether or not the suggested tactic was effective, we relied on two well accepted standards (Eq.5 and Eq.6) [18]. In these equations, 'y' represents the actual value of the dependent variable, 'f(x)' represents the fitted value of the forecasting variable, and 'n' represents the total number of data pairings that include both 'y' and 'f(x)'.

By applying Eq. 5, we were able to calculate the RMSE while taking into consideration both positive and negative errors. The accuracy of the predictive model can be evaluated in a manner that is objective and transparent using this method. In addition, in order to ensure that our findings were accurate, we utilized the MAPE method, which can be found in Equation 6, even though our sample size was relatively low.

By adhering to these stringent requirements, we were able to construct a data-driven predictive model for supply chain planning in the medical equipment industry. This model has enabled us to improve our decision-making and increase the efficiency with which we distribute products. Our research was carried out in a manner that respected people's rights to privacy. As a result, we were able to make moral judgements on the use of the data we gathered, which in turn inspired confidence among all involved parties.

$$RMSE = \sqrt{\left(\frac{1}{n}\right).\sum_{i=1}^{n} y - f(x))^2} \quad (24)$$

$$MAPE = \left(\frac{1}{n}\right).\sum_{i=1}^{n} |(y - f(x))/y| \quad (25)$$

All hypertext links and section bookmarks will be removed from papers during the processing of papers for publication. If you need to refer to an Internet email address or URL in your paper, you must type out the address or URL fully in Regular font.

A. Trend based Clustering.

For this study, information from hospitals H1 through H5 has been put together. It has been noticed that Hypotheses 2 and 5 have the most need for medical tools compared to Hypotheses 1, 3, and 4. As trends in medical equipment demand are always changing, the study plans to include hospital demand patterns from all times of the year [18]. A method called "clustering" is used to put medical facilities in order based on the needs they meet [15]. Hierarchical clustering is chosen as the main method for this reason. Hierarchical clustering will be used to rank hospitals based on how much demand they get. This will help us learn more about demand trends and possible relationships [19]. An important part of the study is to turn the raw time series data into useful demand trends. To do this, a linear regression method is used to change the time series data. This method is able to catch the underlying changes in demand over time. This study uses hierarchical clustering and linear regression to change time series data. The goal is to come up with a comprehensive [13], data-driven plan for understanding, analyzing, and adapting to the changing world of medical

equipment demand in a variety of healthcare settings. The hospital organization based on demand trends could help the medical equipment business a lot with supply chain planning, allocating resources, and making decisions in general [19].

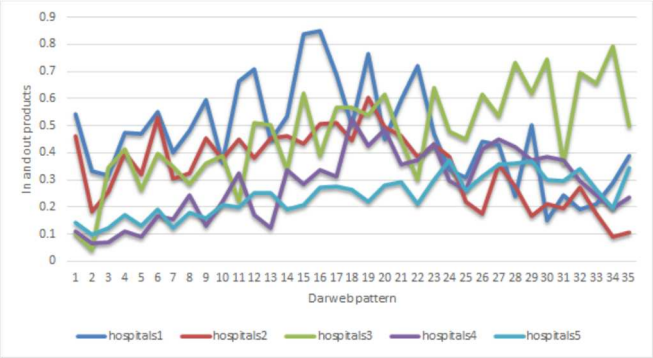

Fig. 2 Illustrates timelines.

TABLE V
THE USE OF A TREND-BASED CLUSTERING SYSTEM TO A COMPARISON OF APPROACHES TO PREDICTING THE DEMAND FOR MEDICAL DEVICES

Cluster	Number of hospitals	Coefficient 1 (average)	Coefficient 2 (average)
1	5	2.33	−2.37
2	5	3.67	0.48
3	8	1.54	2.04
4	100	0.06	0.08

Some hospitals may not have previous year's demand records; therefore data must be cleaned before grouping patients. By eliminating duplicate hospitals, each institution uses linear regression to forecast patient numbers. Linear regression provides two equations that show how institution demand evolves over time [20]. Linear regression data are used. We analyze Fig. 3's H1 data using linear regression. Two regression equations exhibit these tendencies. In the medical industry, annual sales trends are challenging to anticipate because companies change their business strategy and introduce new items based on market feedback. A linear regression coefficient over this time period [21] shows how the company's approach has changed. Next, hospitals will be grouped hierarchically [22]. First, second, and third clusters contain medical facilities in order (5, 6, 7, 9, 10, 11). Hospitals were grouped into the four groups shown. Table 5 displays cluster sizes and average coefficients, and Figure 7 depicts cluster growth.

Four groups of people get very different findings. Two hospitals in Cluster 1 noticed demand rise when the new product came available [23] [12]. The inventive product or service seems to have made the most money in cluster 1. Cluster 2 has two hospitals, and both saw a spike in patients before the new product came available. They became more interested afterward. The new product appears to have stolen market share from the older goods in cluster 2. All seven cluster 3 hospitals have grown in recent years. The new items didn't transform hospitals. The new items did not modify Cluster 4's trend. Figure 5 shows the four classes' three-year aggregate demand. Here's the link. Cluster 1's two hospitals account for 20% of 36-month demand. Clusters 2–4 account for 19%, 27%, and 36% of overall demand, respectively. Cluster 5 doesn't contribute to demand. To maintain their high quality of service, they must please their current consumers [24]. Medium-sized medical facilities may not be able to handle new product risks. Cluster 4 hospitals have enough supplies. Patients at smaller hospitals may be less likely to choose pricey operations for less critical diseases. This may be a contributing element [25].

B. Train, Adapt, Inform, and Monitor

Accessing threat data might not be a good way to find criminals on the dark web, where users can hide their identities. This could be the case if law enforcement and financial institutions don't work together. The best way to stop financial fraud [26] is to always know about the newest risks. This will let us take steps to stop attacks before they happen and act quickly if an attack attempt is found. Some things that medical institutions can do with the information they get from paste sites, dark web marketplaces, and dark web forums are listed below. Change your security rules based on the weaknesses that have been found and the hacking techniques that your enemies have used. For instance, medical institutions should add more security measures [27] [28]to stop or spot bank drops, fake document fraud, and as well as to protect weak technology (like cloud services) that is talked about on the dark web. All these steps are important to stop or find fraud. Tell the bosses, the workers, and the customers what's going on. One of the most effective ways to stop financial fraud is to follow security best practices in a strict and consistent way [29] [30]. By looking at the financial risks on the dark web, institutions can keep their executives, staff, and customers up to date on what possible attacks look like and how to deal with them. This information could also help keep the institution saft. Even if the dangers aren't directly aimed at the medical institutions brand, it can still be better prepared for the future if it actively looks for threats to the medical.

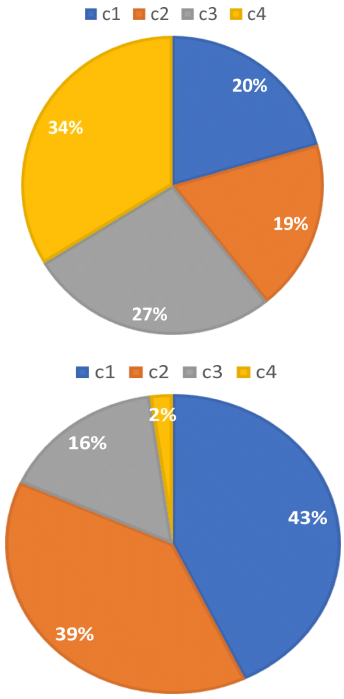

Fig. 3 Medical device demand forecasting

Medical device demand forecasting in Dark web environment. This study uses five different forecasting

methods, each of which is best for a different group of data, to make predictions about future demand. To predict overall demand, we must first figure out which models work best for each set of data and then add up the results of those models. Here, the trend-based method that was suggested is tested to see how well it can predict future demand.

The total demand is compared to models that try to predict direct demand to see how well they do. In Table 2, we can see how well each of the five forecast models worked with each of the four sets of data. With a great total score of 54.10 MAPE, BPNN stands out as the best of the five tried and tested models for cluster 1. But BPNN doesn't do as well as the best in groups 2 and 4. Table 2 and Figure 6 show a very interesting trend: the pattern of cluster 1 between months 18 and 19 changes a lot after month 19. These results show that BPNN's dynamic predictions can keep up with changes in demand. The score for Cluster 2 is the second worst. This shows that the suggested mathematical model works best with time series that follow a steady trend. So, in this case, it's possible that exponential smoothing would work better.

This study shows that the trend-based approach is a good way to predict future demand because it uses the best parts of different forecasting methods and looks at the results across different groups. This method is flexible and useful in the always-changing area of demand forecasting. This is shown by the fact that BPNN was found to be the best method for some groups while other methods were looked at for other types.

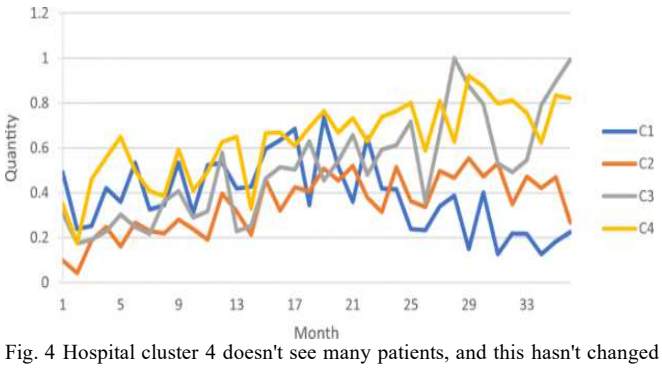

Fig. 4 Hospital cluster 4 doesn't see many patients, and this hasn't changed significantly over time.

A mathematical model should be used to estimate hospital demand with low demand. LR isn't the best model across all four categories. Medical device time series display linear and nonlinear trends. With hybrid models, we can handle linear and nonlinear patterns simultaneously, improving accuracy. The clandestine marketplace known as the "dark web" has emerged as a hub for various illicit activities, particularly within the healthcare industry. This concerning development highlights the industry's emphasis on maintaining high levels of privacy, inadvertently creating a shield for nefarious operations. Exploiting this vulnerability, activities such as organ trafficking and human trafficking thrive. Our research aimed to uncover concealed patterns that may indicate such transactions using the statistical tool of linear regression.

By converting a time series of medical equipment purchases into vectors, we could identify patterns in purchasing behaviors. Analyzing the provided data from the case company, we discovered three predominant purchasing patterns among customers. This finding prompted us to utilize the hierarchical clustering methodology, effectively classifying hospitals based on the direction of their trend vector. The results of this clustering process indicate a strong inclination for hospitals to follow distinct purchasing trajectories, revealing valuable insights for future analyses.

Our research's subsequent phase involved predicting future cluster utilization using five different models. The diversity within these models reflects the complex nature of the data and the challenges it presents. An important outcome of our study is the realization that no single model can consistently provide accurate predictions across all clusters. This emphasizes adopting a flexible approach when predicting future trends, as different time series with varying patterns may require diverse predictive models. In summary, while the clustering methodology based on trends has proven effective in categorizing hospitals with similar purchasing behaviors, searching for the most reliable predictive model remains ongoing. Our research has laid the foundation for further investigations into the intricacies of the "dark web" and its infiltration into the healthcare sector. To address these critical issues, stakeholders must employ a multifaceted approach combining predictive analytics with vigilant monitoring to suppress illicit activities.

IV. Conclusion

In this study, a two-part method was used: first, linear regression was used to look at raw temporal data and create a trend vector; second, hierarchical clustering was used to group hospitals by location; and third, the trend vector was used to make sure that the analysis was the same across clusters. This study looked at raw data about time with linear regression to find a trend vector. Hospitals are ranked based on how well they take care of their patients. The method is essential because it can quickly and correctly spot patterns and trends among healthcare providers. Fraud experts who work in cybersecurity are well aware of how dangerous hackers who target healthcare facilities like hospitals and clinics can be. You might be surprised by how much thought and planning these scammers put into their plans. If victims are going to be able to filter and protect critical information, they need easy access to specialized tools and independent threat researchers.

Because the deep and black web could contain dangers, financial institutions use advanced search engines that use machine learning technology. This is done to reduce the chances of these things happening. These technologies make it possible for financial security teams to look at vast amounts of risk data without having direct access to networks like I2P and Tor. This makes finding and analyzing information about online threats safer and easier. This study helps us learn more about how well healthcare institutions work and how vulnerable they are to cyber threats. It uses linear regression, hierarchical clustering, and cutting-edge search methods. This study's results help develop good cybersecurity strategies that protect private patient data and make healthcare systems less vulnerable to cyberattacks.

References

[1] Z. Ahmad, S. Rahim, M. Zubair, and J. Abdul-Ghafar, “Artificial intelligence (AI) in medicine, current applications and future role with special emphasis on its potential and promise in pathology: present and

future impact, obstacles including costs and acceptance among pathologists, practical and philosophical considerations. A comprehensive review," Diagnostic Pathology, vol. 16, no. 1, Mar. 2021, doi: 10.1186/s13000-021-01085-4.

[2] Karthikeyan Ramalingam, "Use Of Artificial Intelligence in Histopathological Interpretation - A Mini Review," International Journal of Histopathological Interpretation, vol. 12, no. 1, pp. 34–39, Jun. 2023, doi: 10.56501/intjhistopatholinterpret.v12i1.883.

[3] N. Anantrasirichai and D. Bull, "Artificial intelligence in the creative industries: a review," Artificial Intelligence Review, vol. 55, no. 1, pp. 589–656, Jul. 2021, doi: 10.1007/s10462-021-10039-7.

[4] B. R. Jung, K.-S. Choi, and C. S. Lee, "Dynamics of Dark Web Financial Marketplaces: An Exploratory Study of Underground Fraud and Scam Business," CrimRxiv, Sep. 2022, doi:10.21428/cb6ab371.dbbe560f.

[5] N. Nguyen et al., "A Proposed Model for Card Fraud Detection Based on CatBoost and Deep Neural Network," IEEE Access, vol. 10, pp. 96852–96861, 2022, doi: 10.1109/access.2022.3205416.

[6] V. Acin, "Making sense of the dark web," Computer Fraud & Security, vol. 2019, no. 7, pp. 17–19, Jan. 2019, doi: 10.1016/s1361-3723(19)30075-2.

[7] X. Liu and M. Fan, "Identification and Early Warning of Financial Fraud Risk Based on Bidirectional Long-Short Term Memory Model," Mathematical Problems in Engineering, vol. 2022, pp. 1–8, Jul. 2022, doi: 10.1155/2022/2342312.

[8] E. Wilson, "Disrupting dark web supply chains to protect precious data," Computer Fraud & Security, vol. 2019, no. 4, pp. 6–9, Apr. 2019, doi: 10.1016/s1361-3723(19)30039-9.

[9] L. Tkachenko, E. Andrey, G. Pozdeeva, and V. Romanyuk, "Modern approaches of detecting financial statement fraud," SHS Web of Conferences, vol. 80, p. 01024, 2020, doi:10.1051/shsconf/20208001024.

[10] V. Shpyrko and B. Koval, "Fraud detection models and payment transactions analysis using machine learning," SHS Web of Conferences, vol. 65, p. 02002, 2019, doi:10.1051/shsconf/20196502002.

[11] H. Wang, Z. Wang, B. Zhang, and J. Zhou, "Information collection for fraud detection in P2P financial market," MATEC Web of Conferences, vol. 189, p. 06006, 2018, doi:10.1051/matecconf/201818906006.

[12] M. Zarour et al., "Ensuring data integrity of healthcare information in the era of digital health," Healthcare Technology Letters, vol. 8, no. 3, pp. 66–77, Apr. 2021, doi: 10.1049/htl2.12008.

[13] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," SN Computer Science, vol. 2, no. 3, Mar. 2021, doi: 10.1007/s42979-021-00592-x.

[14] V. Shpyrko and B. Koval, "Fraud detection models and payment transactions analysis using machine learning," SHS Web of Conferences, vol. 65, p. 02002, 2019, doi:10.1051/shsconf/20196502002.

[15] A. Bermudez-Villalva and G. Stringhini, "The shady economy: Understanding the difference in trading activity from underground forums in different layers of the Web," 2021 APWG Symposium on Electronic Crime Research (eCrime), Dec. 2021, doi:10.1109/ecrime54498.2021.9738751.

[16] M. Herland, R. A. Bauder, and T. M. Khoshgoftaar, "The effects of class rarity on the evaluation of supervised healthcare fraud detection models," Journal of Big Data, vol. 6, no. 1, Feb. 2019, doi:10.1186/s40537-019-0181-8.

[17] K. S. Sangher, A. Singh, H. M. Pandey, and V. Kumar, "Towards Safe Cyber Practices: Developing a Proactive Cyber-Threat Intelligence System for Dark Web Forum Content by Identifying Cybercrimes," Information, vol. 14, no. 6, p. 349, Jun. 2023, doi:10.3390/info14060349.

[18] Yinhong Shi, "The Rightist Turn in Japanese Politics and Its Implications for China-Japan Relations," in *Research Series on the Chinese Dream and China's Development Path*, 1st ed., vol. 1, Li Yang and Li Peilin, Eds., New York: Springer Science and Business Media LLC, 2021, pp. 171–186.

[19] S. Xu, H. K. Chan, E. Ch'ng, and K. H. Tan, "A comparison of forecasting methods for medical device demand using trend-based clustering scheme," Journal of Data, Information and Management, vol. 2, no. 2, pp. 85–94, Mar. 2020, doi: 10.1007/s42488-020-00026-y.

[20] H. Shi, Y. Chen, and J.-Y. Hu, "Deep learning on information retrieval using agent flow e-mail reply system for IoT enterprise customer service," Journal of Ambient Intelligence and Humanized Computing, Mar. 2021, doi: 10.1007/s12652-021-02991-7.

[21] D. Kolevski, K. Michael, R. Abbas, and M. Freeman, "Cloud Computing Data Breaches in News Media: Disclosure of Personal and Sensitive Data," 2022 IEEE International Symposium on Technology and Society (ISTAS), Nov. 2022, doi:10.1109/istas55053.2022.10227100.

[22] S. Nazah, S. Huda, J. H. Abawajy, and M. M. Hassan, "An Unsupervised Model for Identifying and Characterizing Dark Web Forums," IEEE Access, vol. 9, pp. 112871–112892, 2021, doi:10.1109/access.2021.3103319.

[23] A. K. Pandey et al., "Key Issues in Healthcare Data Integrity: Analysis and Recommendations," IEEE Access, vol. 8, pp. 40612–40628, 2020, doi: 10.1109/access.2020.2976687.

[24] V. Jesus and H. J. Pandit, "Consent Receipts for a Usable and Auditable Web of Personal Data," IEEE Access, vol. 10, pp. 28545–28563, 2022, doi: 10.1109/access.2022.3157850.

[25] M. R. Arshad, M. Hussain, H. Tahir, S. Qadir, F. I. Ahmed Memon, and Y. Javed, "Forensic Analysis of Tor Browser on Windows 10 and Android 10 Operating Systems," IEEE Access, vol. 9, pp. 141273–141294, 2021, doi: 10.1109/access.2021.3119724.

[26] I. Matloob, S. Khan, H. ur Rahman, and F. Hussain, "Medical Health Benefit Management System for Real-Time Notification of Fraud Using Historical Medical Records," Applied Sciences, vol. 10, no. 15, p. 5144, Jul. 2020, doi: 10.3390/app10155144.

[27] S. Dalal, B. Seth, M. Radulescu, C. Secara, and C. Tolea, "Predicting Fraud in Financial Payment Services through Optimized Hyper-Parameter-Tuned XGBoost Model," Mathematics, vol. 10, no. 24, p. 4679, Dec. 2022, doi: 10.3390/math10244679.

[28] A. A. Ali, A. M. Khedr, M. El-Bannany, and S. Kanakkayil, "A Powerful Predicting Model for Financial Statement Fraud Based on Optimized XGBoost Ensemble Learning Technique," Applied Sciences, vol. 13, no. 4, p. 2272, Feb. 2023, doi:10.3390/app13042272.

[29] T. Ashfaq et al., "A Machine Learning and Blockchain Based Efficient Fraud Detection Mechanism," Sensors, vol. 22, no. 19, p. 7162, Sep. 2022, doi: 10.3390/s22197162.

[30] T. K. Shakir, R. Scharif, and M. M. Nasir, "A Proposed Blockchain based System for Secure Data Management of Computer Networks," Journal of Cybersecurity and Information Management, vol. 11, no. 2, pp. 36–46, 2023, doi: 10.54216/jcim.110204.