The International Journal of Engineering and Science (IJES)
|| Volume || 12 || Issue || 5|| May || Pages || PP 13-22 || 2023 ||
ISSN (e): 2319-1813 ISSN (p): 20-24-1805

# Efficient Model for Mining Emerging Patterns in Financial Transitions

Laila Abd- Ellatif[1]
*Department of Information Systems*
*Faculty of Computers and Artificial Intelligence*
*Helwan University Cairo, Egypt*

Noha kamel[3]
*Department of Information Systems*
*Faculty of Computers and Artificial Intelligence*
*Helwan University Cairo, Egypt*

Mahmoud AbdEllatif[3]
*Faculty of Computers & AI*
*Helwan University, Egypt*
*College of Business*
*University of Jeddah , Saudi Arabia*

***ABSTRACT***

Over the years, transactional databases accumulate enormous amounts of timestamped data on an organization's clients. Mining these timestamped data can assist business executives in making better judgements by analyzing their consumers' transactions over time. A time series is a collection of observations that are all recorded at the same time. The finding of significant differences between datasets is the most critical challenge in data mining. Dong and Li [1] are the first to propose the concept of emerging patterns (EPs). EPs are a group of itemsets whose support varies greatly from one dataset to the next. Emerging patterns are groups of itemsets. Or, in other terms, they are sets of sets." Emerging Pattern Mining (EPM) is a data mining task that attempts to uncover discriminative patterns that can describe emergent behavior in connection to a property of interest. Despite the fact that more and more studies focus on emerging patterns in relational databases, few studies have investigated mining emerging patterns in time series. The framework we propose in this paper is divided into two phases: Transformation of data: The time series data is transformed into a symbolic representation. After integrating the notion of the product life cycle, emerging pattern mining was used to uncover four emerging patterns. Experiments with financial data acquired from the Egyptian Stock Exchange were conducted to assess the efficacy of the suggested framework.

***Keywords***
*Emerging patterns; product life cycle, trend patterns.*

Date of Submission: 20-05-2023 Date of Acceptance: 03-06-2023

## I. INTRODUCTION

Machine learning methods in data mining include two primary methods. The first is supervised learning, which gives data and classification results. By training the machine to analyze the relationship between the given data and its predefined results, it can discover relationships that help build classification models and predict what will happen in New cases. The second method is unsupervised learning, where data is given to the machine whose results are not specified. Unsupervised learning refers to tasks such as summarizing or mining association rules. Unsupervised learning also deals with data analysis intending to formulate a description of the relationships between data elements without relying on attributes that explain these relationships.[37,38]

The analysis of systems behavior patterns and their components through the monitored behavior data is considered one of the essential applications in the field of knowledge detection techniques through data mining. Pattern analysis helps in diagnosing behavior and formulating prediction models. The knowledge gained from this data can be used to make smart decisions in various fields of applications. [33,39]

Contrast pattern mining (CPM)is a popular and important sub-field of data mining, as variance patterns describe significant differences between sets of data. There are many types of contrast patterns, including the following types, Fig 1 illustrated them:

1. Contrast Sub-graph
2. Converging Patterns
3. Emerging Patterns
4. Conditional Contrast Patterns
5. Minimal Distinguishing Subsequence
6. Contrast Sequential Patterns

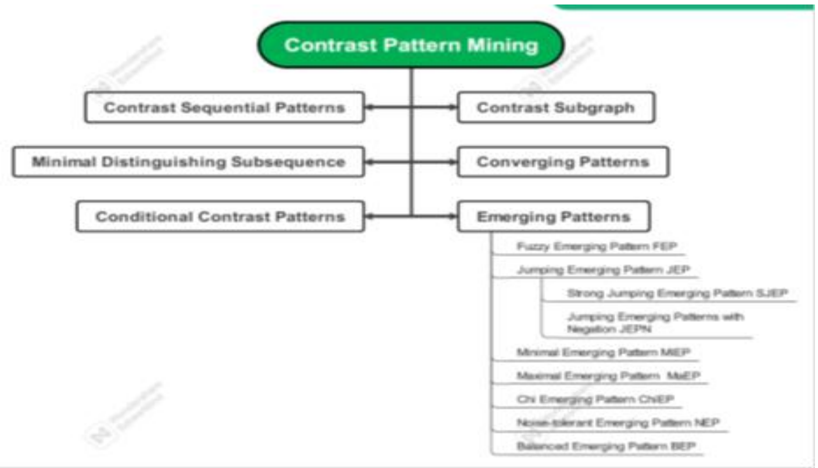



**Fig 1: Types of Contrast Pattern Mining (prepared by Authors)**

EPM is a subset of CPM. Emerging patterns have been suggested by Dong et al. in 1999 and are defined as item groups with significantly increasing support from one data set to another. Trends that appear in the timestamp database can be captured by EPs. EPs can also find useful contrasts between categories of data. Minimal discriminatory consequences of discrete successive patterns.[34,35]

One of the most crucial tasks for managers is to come up with new marketing strategies to deal with competitive and changing market conditions. Managers are bothered by this issue since it is challenging to identify shifting customer preferences. Emerging patterns can identify new patterns in timestamped databases or significant differences between different types of data.

Emerging patterns (EPs), the set of patterns that reflect increasing frequencies from one data set to another so they are vital in making different decisions. Whereas, in fixed data sets such as those with known classifications (males vs. females, treated (cured vs. untreated), the process of discovering emerging patterns contributes to identifying hidden patterns of variation between data sets, which contributes to decision support such as building classifiers , predicting disease probability, detecting patterns in gene expression data, etc. In sequencing datasets, emerging patterns are useful in decision making such as studying and understanding customer behavior, predicting future purchases, etc.[36]

In this article, we try to study and analyze behavior patterns in the product life cycle, as it is well known that the product life cycle includes four stages, namely[4]: (1) Growth stage patterns (2) Rapid-Rise stage patterns (3) Decline stage patterns (4) Rapid-Sink stage patterns to discover trend patterns in data of product service behavior . The conventional representation of product life cycle curves is a bell-shaped curve with its stages (see Fig. 2).

The product enters the market at the beginning of its life, sales climb gradually, which represents introduction stage so Profit is nonexistent due to the significant start-up expenditures. as a result of advertising and marketing to raise consumer awareness of the product and its benefits. in The growth stage: The gradual expansion of the market introduction and development stages has now given way to a rapid upturn as the product takes off and profits skyrocket. At maturity stage, sales growth slows since the product has earned acceptance from the bulk of possible customers. Profit may stabilize or drop owing to greater competition. If the company needs to increase customer number, it can gain new customers, create new categories, or entice clients from other companies. In the last stage which describes decline stage, many enterprises will begin to move onto alternative endeavors as market saturation means there is no longer any profit to be gained. Some enterprises may survive the decrease and continue to offer the product, but manufacturing is likely to be on a smaller scale and pricing and profit margins may become low. Sales slip downhill as profit destroys.


**Fig.2 A product life cycle**

The framework is divided into two stages: Data Transformation [5]: The transformation method eliminates certain erroneous and weak data points, which increases the efficiency of EPs mining. We use the symbolic aggregate approximation as one of the methods of transformation the data time series into a symbolic series, considering Perceptually Important Points (PIPs) approaches:-
1- Splitting time series data into consecutive subsequences.
2- Clustering these subsequences using an algorithm.
3- The symbolic time series is then created by replacing each subsequence with the name of its cluster.

Mining trend patterns: Exploiting developing patterns to find something fascinating in the initial time series The support variant of data sets is the sole criterion for discovering emerging patterns in traditional mining of emerging patterns. However, precisely distinguishing the varying trend of emerging patterns in business is very important. Managers' decision-making would be more accurate if the types of evolving trends could be accurately defined.
Millions of investors buying and selling assets on the Egyptian Stock Exchange interact dynamically, establishing the suggested framework to mine these four sorts of patterns from the financial time series data. For the purpose of assessing the effectiveness of the suggested framework, experiments were also carried out using financial data obtained from the Egyptian stock exchange.

## II. RELATED WORK

Numerical time series had traditionally been difficult to analyze and evaluate in the field of time series data mining because of their frequent high dimension and quantity of data. The majority of these methods were numerical, such as the discrete Fourier transform (DFT), principal component analysis (PCA), piecewise aggregate approximation (PAA).discrete wavelet transform (DWT), singular value decomposition (SVD), adaptive piecewise constant approximation (APCA). Many methods had been suggested for dimensionality reduction as well as to define and represent the original data.

Beyond their ease of use, readability, and efficiency for time series representation, symbolic approaches had also been widely used because they allowed for the application of algorithms from other fields, such as bioinformatics, text processing, and information retrieval. Each sequence would be rapidly and readily reduced from n dimensions to k dimensions using PAA. PAA Drawbacks: The mean value representation and equal length segmentation might overlook some significant features or patterns. Other methods, such as the following, had recently been suggested: (SAX) [9] referred to symbolic aggregate approximation. (2) PIPs referred to perceptually important points [8].

**Symbolic Aggregative approXimation (SAX)**

Lin and Keogh [9] proposed SAX, which relied on PAA. In SAX, each time series was turned into a PAA representation and then symbolized as a sequence of symbolic strings. SAX discretized PAA values using equal-sized regions under the Gaussian curve, resulting in breakpoints. Most of the cut-off values representing the coding zones are based on the Gaussian distribution for symbolize the time series.

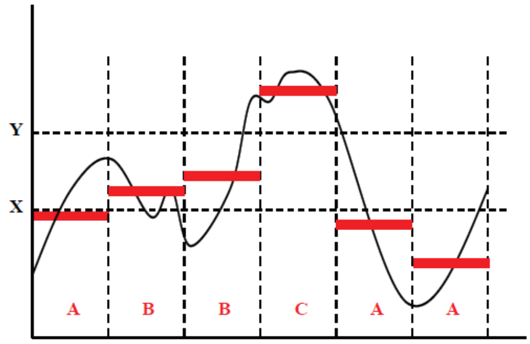



**Fig.3 An example of SAX transformation**

As shown in Figure 3, a time series was separated into six equal-length segments and represented as a string ABBCAA by cut values X and Y. SAX was the first symbolic representation of time series using an approximation distance function that had lower bounds using Euclidean distance [9]. It was simple and useful.

**Perceptually important points (PIPs) [10].**

Characterize time series data based on the assumption that regularly emerging patterns can usually be abstractly represented by a few essential points [11]. These points had perceptually important influences in relation to human eyesight. The n-dimensional time series data will be.

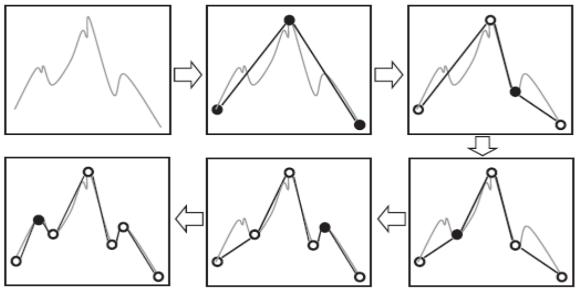


**Fig.4 An example of PIPs with seven points characterized**

A time series sequence was defined by seven points, as shown in Figure 4. (1) The first and second PIPs were defined as the sequence's first and end points. (2) The next PIP in the series was the one with the greatest distance between its two neighboring PIPs. (3) Pattern detection stops when it becomes the number of PIPs meets the user-specified parameter k . The distance between point PA(xA, yA) to its two adjacent PIPs P1(x1, y2) and P2(x2, y2) represents the vertical distance between points PA & PC, where PC(xC, yC) was the projection of PA on the line P1P2.

**Emerging Patterns Mining:**

The primary purpose of this research was to mine emerging patterns for noteworthy trends. EPs are groups of itemsets whose popularity varies greatly from one dataset to the next. EPs were proposed [12] as a type of pattern whose support varies dramatically from one data class to the next: "The larger the difference in pattern support, the more interesting the patterns." EPs were used to: predict the occurrence of disorders such as acute lymphoblastic leukaemia [13]. Investigate high-dimensional data sets, such as gene expression data [14]. Emerging trends in timestamped databases could be captured using timestamped databases. Capture useful contrasts between data classes. Emerging pattern classifiers: are valuable tools to solve real problems in fields like: Bioinformatics, Intruder detection , Rare event forecasting , Streaming data analysis , Human activity recognition , Privacy preserving data mining , Anomaly detection in network connection data .[33]

The two most common uses of EPs were in finance and bioinformatics. Using EP mining (EPM), Li and Wong [15] identified gene clusters and used them to analyse a dataset of colon cancers. In order to develop a method to identify changes in consumer segments between timestamped datasets, Huang et al. [16] mined the changes in medical practise for bronchial lung cancer clinical pathways, Kim et al. [17] applied the EP concepts. Tsai and Shieh [18] studied recently developing sequential patterns to look for trends in consumer behaviour. The main goal of Shie et al.'s [19] study was to apply mobile sequential pattern mining to mobile commerce environments in order to uncover user behaviour patterns. Deng and Zaane [20] proposed model to detect emerging sequences (ES) fitting various constraints; offered a gap constraint and maintain the uniqueness of items by assessing ESs with their occurrences. Ceci et al. [21] used a multi-relational strategy to deal with the degree of complexity involved in locating EPs from spatial data. In addition, as evidenced in works on EPs and leaping EPs, EP research has focused on the application of the recognised patterns for categorization purposes. Garca-Vico and others. [22] focused on the discriminative element of EP mining's descriptive element. They collected the most recent methods that have been offered in the literature and organised them into a taxonomy so that they might have a comprehensive comprehension of the task's overall vision. Terecki & Walczak [23] detected jumping emerging patterns with negation (JEPNs). Intriguing information can be found in JEPNs, which are also useful for classifying things. Chu et al. [24] proposed the EFI-Mine approach to efficiently and effectively mine temporal emerging common item sets from data streams. According to [25], the authors introduced a novel EPs weighting method. in order to mine the co-occurrence correlations of visual terms using EPs. Vimieiro and Moscato looked at the problem of mining disjunctive emergent patterns in high-dimensional biological datasets. [26]. But in [27] A special dynamic structure (DFP-SEPSF) was suggested in order to build the frequent pattern tree whenever new features are introduced and to mine emerging patterns online. Find newly developed hotel characteristics that matter to international travellers. Li et al. employ a pattern mining strategy that is still in development. [28]. While reviewing the paper[29], Khan et al. processed temporal data using dual support Apriori. Technique (DSAT); Through the application of DSAT using the sliding window methodology in analyzing time series data, it is possible to discover JEPs as one of the patterns of variation in the data. The DSAT technique exploits pre-mined timestamp data by utilizing the sliding window idea. It detects all JEPs, as in 'nave' approaches, but uses less memory and grows linearly with large datasets. In [10] TSEPs Miner, researcher a suggested algorithm, used time gap limitations to mine time series EPs. The proposed approach, TSEPs Miner, was the beginning of time series analyses. Mai Van Hoan's approach, transcribed by Miner, involves time series mining components: time series symbolization and looking for frequent patterns by employing sliding window techniques. A.M. Garc´ıa-Vico. [30] presented the algorithm EvAEP as a solution to a photovoltaic technology-related real-world issue. When applied to the issue of concentrating photovoltaic technology, which was focused on the generation of power while lowering associated costs, the algorithm was an evolutionary fuzzy system for emerging pattern mining. Weng Cheng-Hsiung [4] The concept of following changes in customer buying behaviors over several successive time periods was used to observe sales trends in dynamic markets through emerging patterns mining.

We detect that the majority of earlier research, some of which concentrated on converting numerical time series data into symbolic time series data, concentrated on mining only newly emerging patterns or frequently occurring patterns or through sliding windows, and some studies analyzed trend patterns through supermarket data in the form of transactions; consequently, there is no approach that focuses on mining trend patterns in symbolic time series.

The proposed methodology differs from the preceding work in that it makes use of the concept of EPs to uncover trends in symbolic time series data that span the product life cycle and help investors make informed decisions.

## III. PROPOSED FRAMEWORK

The recommended framework for mining trends and patterns from time series data, data transformation, and EPs mining algorithms is divided into two phases. Figure 5 depicts the proposed mining trend pattern framework

### 3.1 First Phase: The Symbolization

Representation of Time Series In this section, we will look at approaches for reducing the dimensionality of numerical time series. Transform an original numerical time series into a simpler, easier-to-analyze form. Time series characterization: Each value of a time series is recorded at regular time intervals in the domain of time series, such as seconds, hourly, daily, and so on. Numerical time series are challenging to analyze and handle due to their enormous amount of data and high dimension. By converting the numerical time series to a symbolic time series via:-

#### 3.1.1 Splitting:

Create subsequences out of numerical time series. By splitting the original time series T into k segments, we produce a set of subsequences.

where, $k = \frac{n}{p}$ n denoted the length of the time series T
p denoted the number of points in each segment

The type of time series affects the choice of k. For example: day, year, etc. A numeric time series T = {t1, t2, … , tn} of length n . A subsequence length p: number points in each segment.

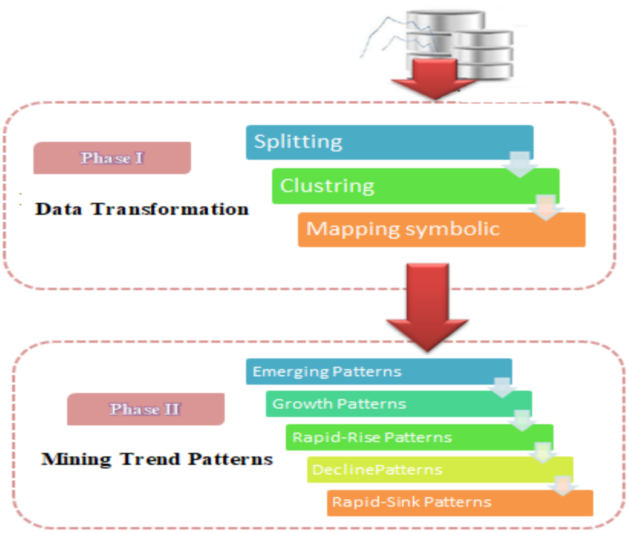



**Fig.5 Mining trend Patterns Framework**

**3.1.2 Clustering for Time Series**

Clustering is the process of categorizing input data into classes or clusters so that objects inside a cluster are notably different from those in other clusters while still sharing a high degree of similarity. There are two types of time series clustering: Clustering as a whole Taking into account a collection of unique time series data Sort the input data into categories so that similar time series can be grouped together. Clustering of Subsequences Given only one time series, to produce a new set of time series datasets.t, split the input time series.

k-means: The k-means algorithm divides a set of n input objects into k clusters (groups), where k is the predetermined number of clusters. The first k beginning clusters (centroids) of k objects are chosen from the dataset via the algorithm. Each object is then given its closest centroid. The new centroid is then calculated for each group. This process iterates until there are no more changes to the object assignments or a stable state can be established, or until we reach a maximum number of iterations.

**3.1.3 Mapping Symbolic:**

Each subsequence is given a cluster at the end of this process and can therefore be symbolized by the symbol for that cluster. Each subsequence is initially swapped out for the name of its cluster to form the symbolic time series.

A symbolic representation of a time series T: $\overline{T} = \{\overline{t1}, \ldots, \overline{tk}\}$ where $k = \frac{n}{p}$

**3.2 Second Phase: Mining trends patterns**

To discover four emerging patterns after integrating with the concept of product life cycle.

**3.2.1 Support:**

Support is the frequency with which an itemset appears in the input data. If an itemset's support exceeds a set of user-defined support thresholds, it is considered to be frequent. A transaction T may contain F if $F \subseteq T$. Where we might be able to deduce the association rule $F \Rightarrow G$, where $F \subset I$ , $G \subset I$, and $F \cap G = \emptyset$. The Association rule $F \Rightarrow G$ holds in the transaction set Di with support s, where s is the ratio of the transaction in Di that contains $F \cup G$. The following is the formal expression of sup($F \cup G$).

Sup ($F \cup G$, $D_i$) = $\frac{|F \cup G|}{|Di|}$ [31] , where $|D_i|$ being the number of transactions in $D_i$ and $|F \cup Y|$ being the number of transactions containing $F \cup Y$ in $D_i$.

**3.2.2 Growth Rate:**

An itemset F is called an Emerging Pattern if the supp(F) $\geq \sigma$ and GR(F) $\geq \delta$ where $\sigma$ and $\delta$ is user specified support and growth rate thresholds respectively. Given two dataset D1, D2, an itemset F, and the support of F in D1 and D2 then the Growth Rate of F from D1 to D2 GrowthRate(F) is:

$$\text{GrowthRate (F)} = \begin{cases} 0 & \text{if}(\text{supp}(F, D1) = 0 \text{ and } \text{supp}(F, D2) = 0), \\ \infty & \text{if}(\text{supp}(F, D1) = 0 \text{ and } \text{supp}(F, D2) \neq 0), \\ \frac{\text{supp}(F, D2)}{\text{supp}(F, D1)} & \text{otherwise.} \end{cases} \quad [32]$$

**3.2.3 Growth Pattern:**

Growth patterns may be one of two types: exponential growth, which occurs in an ideal environment, and logistical growth, where environmental pressures hinder growth rates [35].

Since GR(F,$D_{ta}$ , $D_{tb}$ ) $\geq \sigma_{grG}$ and GR(F,$D_{tb}$ , $D_{tc}$) $\geq \sigma_{grG}$, and $\sigma_{grG} \geq 1$ where $D_{ta}$ , $D_{tb}$, ; $D_{tc}$ are three transaction sets, ta,tb, and tc are the user's defined timestamps for ta < tb < tc, and $\sigma_{grG} \geq 1$ is the user's defined growth rate threshold.

We have GR (F,$D_{ta}$ , $D_{tc}$)

$$= \frac{\sup(F, D_{tc})}{\sup(F, D_{ta})} = \frac{\sup(F, D_{tb})}{\sup(F, D_{ta})} \times \frac{\sup(F, D_{tc})}{\sup(F, D_{tb})} = GR(F, D_{ta}, D_{tb}) \times GR(F, D_{tb}, D_{tc}) \geq \sigma_{grG} \times \sigma_{grG} = (\sigma_{grG})^2. \quad [4]$$

### 3.2.3 Rapid-Rise Pattern:

when the data disclosure on rapid rise pattern we find $GR(F,D_{ta}, D_{tb}) \geq \sigma_{grG}$ and $GR(F,D_{tb}, D_{tc}) \geq GR(F,D_{ta}, D_{tb}) \times (\sigma_{grG})^{\alpha}$, and $\sigma_{grG}$ ($\geq 1$) where $D_{ta}$, $D_{tb}$, and $D_{tc}$ are three transaction sets ta, tb, and tc are the user's defined timestamps for ta < tb < tc, $\sigma_{grG}$ ($\geq 1$) is the user's defined growth rate threshold, and $\alpha$ ($\geq 1$) is the user's defined Rapid-Rise parameter.

We have $GR(F,D_{ta}, D_{tc})$

$= \frac{sup(F,D_{tc})}{sup(F,D_{ta})} = \frac{sup(F,D_{tb})}{sup(F,D_{ta})} \times \frac{sup(F,D_{tc})}{sup(F,D_{tb})} = GR(F,D_{ta},D_{tb}) \times GR(F,D_{tb},D_{tc}) \geq GR(F,D_{ta},D_{tb}) \times GR(F,D_{ta},D_{tb}) \times (\sigma_{grG})^{\alpha} = (\sigma_{grG})^2 \times (\sigma_{grG})^{\alpha} = (\sigma_{grG})^{2+\alpha}$. [4]

### 3.2.4 Decline Pattern]:

the data display decline pattern when $GR(F,D_{ta}, D_{tb}) \leq \sigma_{grD}$ and $GR(F,D_{ta}, D_{tb}) \leq \sigma_{grD}$, $\sigma_{grD}$ ($\leq 1$) where $D_{ta}$, $D_{tb}$, and $D_{tc}$ are three transaction sets and ta, tb and tc are the user's defined timestamps for ta < tb < tc, and $\sigma_{grD}$ ($\leq 1$) is the user's defined decline rate threshold.

We have $GR(F,D_{ta}, D_{tc})$

$= \frac{sup(F,D_{tc})}{sup(F,D_{ta})} = \frac{sup(F,D_{tb})}{sup(F,D_{ta})} \times \frac{sup(F,D_{tc})}{sup(F,D_{tb})} = GR(F,D_{ta},D_{tb}) \times GR(F,D_{tb},D_{tc}) \leq \sigma_{grD} \times \sigma_{grD} = (\sigma_{grD})^2$. [4]

### 3.2.6 Rapid-Sink Pattern:

To discover this kind of patterns which is called a rapid-sink pattern. It must be $GR(F,D_{ta}, D_{tb}) \leq \sigma_{grD}$ and $GR(F,D_{tb}, D_{tc}) \leq GR(F,D_{ta}, D_{tb}) \times (\sigma_{grD})^{\beta}$, and $\sigma_{grD}$ ($\leq 1$), where $D_{ta}$, $D_{tb}$, and $D_{tc}$ are three transaction sets. ta, tb, and tc are the user's defined timestamps for ta < tb < tc, $\sigma_{grD}$ ($\leq 1$) is the user's defined decline rate threshold, and $\beta$ ($\geq 1$) is the user's defined Rapid-Sink parameter.

We have GR(F,Dta ,Dtc)

$= \frac{sup(F,D_{tc})}{sup(F,D_{ta})} = \frac{sup(F,D_{tb})}{sup(F,D_{ta})} \times \frac{sup(F,D_{tc})}{sup(F,D_{tb})} = GR(F,D_{ta},D_{tb}) \times GR(F,D_{tb},D_{tc})$
$\leq GR(F,D_{ta},D_{tb}) \times GR(F,D_{ta},D_{tb}) \times (\sigma_{grD})^{\beta} = \sigma_{grD} \times (\sigma_{grD})^{1+\beta} = (\sigma_{grD})^{2+\beta}$. [4]

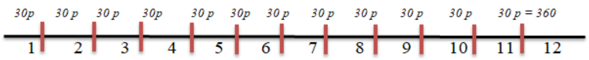
## IV. CONDUCTED RESULT

### 4.1 Splitting

n = 12(month) X 30(day) X 1(once a day) = 360
T (original ts) = {t1, t2, …….t360}

P (no. of points in each segment)
Suppose one month, P= 30 X 1 = 30

$$k = \frac{n}{p}$$ K (no. of segments)




### 4.2 Clustering:

We use a clustering algorithm, such as k - means, to cluster the collection of subsequences. A number of clusters C.
Suppose: C = 5
A set of symbols S of size $|S| = c$.
S = { A, B, C, D, E }
We have K= 12 segment or sequence, By using k-mean algorithm we distribute each sequence to cluster.

**4.3 Mapping Symbolic:**

At the end of this step, each subsequence is assigned into one cluster, and can thus be represented by the symbol associated with this cluster.

The symbolic time series is created by first replacing each subsequence with the name of its cluster. A symbolic representation of a time series T: T = { t1, … , tk} where $k = \frac{n}{p}$

Suppose

C1 – A    C2 – B    C3 – C    C4 – D    C5 - E

| 1 , 7 | 5 . 11 | 9 , 12 | 2 , 4 , 6 | 3, 8, 10 |
|---|---|---|---|---|

1 – 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9 - 10 - 11 - 12

{A - D - E - D - B - D - A - E - C - E - B - C}

Items = { A, B, C, D, E}

Let us set:

Support Threshold $\sigma_{SUP} = 0.20$

Growth Rate Threshold $\sigma_{grG} = 1.20$

Decline Rate Threshold $\sigma_{grD} = 0.80$

Rapid-Rise rate $\alpha = 1$

Rapid-Sink rate $\beta = 1$

**Table 1: Transactions Items datasets**

| | $D_1$ January – April | | | | | $D_2$ May – August | | | | | $D_3$ September – December | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | A | B | C | D | E | A | B | C | D | E |
| 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 |
| 2 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 3 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 |
| 4 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 |
| 5 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |

**Table 2: Discovering frequent item sets from D1, D2 and D3 transaction sets**

| Generated items | Sup ($D_1$) | FP $\sigma_{sup}=0.20$ | Sup ($D_2$) | FP $\sigma_{sup}=0.20$ | Sup ($D_1$) | FP $\sigma_{sup}=0.20$ |
|---|---|---|---|---|---|---|
| A | 4/5=0.8 | ✓ | 2/5=0.4 | ✓ | 1/5=0.2 | ✓ |
| B | 4/5=0.8 | ✓ | 5/5=1 | ✓ | 1/5=0.2 | ✓ |
| C | 2/5=0.4 | ✓ | 2/5=0.4 | ✓ | 2/5=0.4 | ✓ |
| D | 2/5=0.4 | ✓ | 3/5=0.6 | ✓ | 5/5=1 | ✓ |
| E | 4/5=0.8 | ✓ | 2/5=0.4 | ✓ | 3/5=0.6 | ✓ |
| AB | 4/5=0.8 | ✓ | 2/5=0.4 | ✓ | 1/5=0.2 | ✓ |
| AC | 1/5=0.2 | ✓ | 2/5=0.4 | ✓ | 1/5=0.2 | ✓ |
| AD | 1/5=0.2 | ✓ | 0/5=0.0 | ✓ | 1/5=0.2 | ✓ |
| AE | 3/5=0.6 | ✓ | 0/5=0.0 | ✓ | 0/5=0.0 | ✓ |
| BC | 1/5=0.2 | ✓ | 2/5=0.4 | ✓ | 1/5=0.2 | ✓ |
| BD | 1/5=0.2 | ✓ | 3/5=0.6 | ✓ | 1/5=0.2 | ✓ |
| BE | 3/5=0.6 | ✓ | 2/5=0.4 | ✓ | 0/5=0.0 | ✓ |
| CD | 2/5=0.4 | ✓ | 0/5=0.0 | ✓ | 2/5=0.4 | ✓ |
| CE | 1/5=0.2 | ✓ | 0/5=0.0 | ✓ | 0/5=0.0 | ✓ |
| DE | 1/5=0.2 | ✓ | 2/5=0.4 | ✓ | 3/5=0.6 | ✓ |

**Table 3: Calculation the growth rates of frequent item sets.**

| Generated items | Sup ($D_1$) | Sup ($D_2$) | Sup ($D_1$) | GR(D1,D2) | GR(D2,D3) | FP |
|---|---|---|---|---|---|---|
| A | 4/5=0.8 | 2/5=0.4 | 1/5=0.2 | 2/4=0.5 | 1/2=0.5 | X |
| B | 4/5=0.8 | 5/5=1 | 1/5=0.2 | 5/4=1.2 | 1/5=0.2 | (D1, D2) |
| C | 2/5=0.4 | 2/5=0.4 | 2/5=0.4 | 2/2=1 | 2/2=1 | X |
| D | 2/5=0.4 | 3/5=0.6 | 5/5=1 | 3/2=1.5 | 6/3=2 | (D1, D2) (D1, D2) |
| E | 4/5=0.8 | 2/5=0.4 | 3/5=0.6 | 2/4=0.5 | 3/2=1.5 | (D1, D2) |
| AB | 4/5=0.8 | 2/5=0.4 | 1/5=0.2 | 2/4=0.5 | 1/2=0.5 | X |
| AC | 1/5=0.2 | 2/5=0.4 | 1/5=0.2 | 2/1=2 | 1/2=0.5 | (D1, D2) |
| BC | 1/5=0.2 | 2/5=0.4 | 1/5=0.2 | 2/1=2 | 1/2=0.5 | (D1, D2) |
| BD | 1/5=0.2 | 3/5=0.6 | 1/5=0.2 | 3/1=3 | 1/3=0.33 | (D1, D2) |
| DE | 1/5=0.2 | 2/5=0.4 | 3/5=0.6 | 2/1=2 | 3/2=1.5 | (D1, D2) |

**Table 4: Identify trend patterns.**

| Generated items | GR(D1,D2) | GR(D2,D3) | PATTERNS(D1,D3) | | | | FP |
|---|---|---|---|---|---|---|---|
| | | | GROWTH | RAPID RISE | DECLINE | RAPID SINK | |
| A | 2/4=0.5 | 1/2=0.5 | 0.25 | 0.25 | 0.25 | 0.25 | X |
| B | 5/4=1.2 | 1/5=0.2 | 0.24 | 0.24 | 0.24 | 0.24 | (D1,D2) |
| C | 2/2=1 | 2/2=1 | 1 | 1 | 1 | 1 | X |
| D | 3/2=1.5 | 6/3=2 | 3 | 3 | 3 | 3 | (D1,D2) (D1,D2) |
| E | 2/4=0.5 | 3/2=1.5 | 0.75 | 0.75 | 0.75 | 0.75 | (D1,D2) |
| AB | 2/4=0.5 | 1/2=0.5 | 0.25 | 0.25 | 0.25 | 0.25 | X |
| AC | 2/1=2 | 1/2=0.5 | 1 | 1 | 1 | 1 | (D1,D2) |
| BC | 2/1=2 | 1/2=0.5 | 1 | 1 | 1 | 1 | (D1,D2) |
| BD | 3/1=3 | 1/3=0.33 | 0.99 | 0.99 | 0.99 | 0.99 | (D1,D2) |
| DE | 2/1=2 | 3/2=1.5 | 3 | 3 | 3 | 3 | (D1,D2) |

**Evaluate the results.**

Growth patterns were examined through this study, along with quick rising patterns, through in-depth analysis of the many and varied patterns discussed in this research. In addition, the study examined how to analyze Decline patterns, mainly how to do it quickly using data sets with a temporal component. The proposed model can therefore give managers knowledge support to aid in formulating and developing their financial and administrative strategies based on these results.

To verify the accuracy of the results, we use the standard evaluation model for evaluating the accuracy of classification models. The following table (Table 5) and equations measure the performance levels of the model.[37]

**Table 5: Evaluation Measures Matrix & Equations**

| Actual class | Predicted class | | | |
|---|---|---|---|---|
| | | Yes | No | Total |
| | Yes | TP | FN | P |
| | No | FP | TN | N |
| | Total | P' | N' | P+N |

| Equations Of Measures | |
|---|---|
| Accuracy | $(TP+TN)/(P+N)$ |
| Precision | $TP/(TP+FP)$ |

**In this Case study:**

We extract the following knowledge bases, see table 6:

**Table 6: Extracted Association rules.**

| Pattern | Rule |
|---|---|
| **Growth** | If $GR(F,D_{ta}, D_{b}) \geq \sigma_{grG}$ AND $GR(F,D_{tb}, D_{tc}) \geq \sigma_{grG}$, AND $\sigma_{grG} \geq 1$ Then $GR(F,D_{ta}, D_{tc}) = GR(F,D_{ta},D_{tb}) \times GR(F,D_{tb},D_{tc}) \geq (\sigma_{grG})^{2}$ |
| **Rapid – Rise** | If $GR(F,D_{ta}, D_{tb}) \geq \sigma_{grG}$ AND $GR(F,D_{tb}, D_{tc}) \geq GR(F,D_{ta}, D_{tb}) \times (\sigma_{grG})^{\alpha}$, AND $\sigma_{grG}$ ($\geq 1$) Then $GR(F,D_{ta}, D_{tc}) = GR(F,D_{ta},D_{tb}) \times GR(F,D_{tb},D_{tc}) \geq (\sigma_{grG})^{2+\alpha}$ |
| **Decline** | If $GR(F,D_{ta}, D_{b}) \geq \sigma_{grG}$ AND $GR(X,D_{tb}, D_{tc}) \geq \sigma_{grG}$, AND $\sigma_{grG} \geq 1$ Then $GR(F,D_{ta}, D_{tc}) = GR(F,D_{ta},D_{tb}) \times GR(F,D_{tb},D_{tc}) \geq (\sigma_{grG})^{2}$ |
| **Rapid - Sink** | If $GR(F,D_{ta}, D_{tb}) \leq \sigma_{grD}$ AND $GR(F,D_{tb}, D_{tc}) \leq GR(F,D_{ta}, D_{tb}) \times (\sigma_{grD})^{\beta}$, AND $\sigma_{grD}$ ($\leq 1$) Then $GR(F,D_{ta},D_{tc}) = GR(F,D_{ta},D_{tb}) \times GR(F,D_{tb},D_{tc}) \leq (\sigma_{grD})^{2+\beta}$ |

**In addition, we can add the following notes:**

1. Clusters D, DE is a Growth patterns
2. A rapid increase in sales growth Cluster D is Rapid-Rise pattern; managers should employ a number of strategies to maintain this rapid market expansion, such as enhancing product quality and design and including new features.
3. Clusters D, DE is a Growth patterns
4. Cluster A is a Decline pattern.

## V. CONCLUSION

Analyzing the data to explore the Contrast in the patterns of behavior of the variables under study helps in monitoring the relationships between these variables. Then it contributes to discovering the hidden knowledge within the data. The proposed approach differs from the earlier work in that it finds trend patterns throughout the product life cycle by applying EP ideas to symbolic time series data. Data transformation and trend pattern mining are the two sections of the framework. The framework provides four useful and intriguing patterns—the growth pattern, rapid-rise pattern, decline pattern, and rapid-sink pattern—with the idea of EPs in symbolic time series data. Emerging patterns that reveal trends in particular industries or businesses can be found by analysing stock market data, which can assist investors in making wise selections. Four trend patterns are also identified by this framework in addition to EPs. Results from the evaluation of this framework have been quite positive.

## VI. FUTURE WORK

In the future, implement the suggested framework to handle a variety of emerging patterns, such as Jumping Emerging Patterns (JEPs), Minimal Emerging Patterns (MinEPs), Maximal Emerging Patterns (MaxEPs), Essential Jumping Emerging Patterns (EJEPs), Noise-Tolerant Emerging Patterns (NEPs), constrained emerging patterns (CEPs), and JEPs with Negation (JEPNs). Chi EPs are developing patterns in Chi.

## REFERENCES

[1]. G. Dong and J. Li (1999) Efficient mining of emerging patterns: Discovering trends and differences, Proc. of the 5th ACM SIGMOD Int. Conf. on Knowledge Discovery and Data Mining, pp.43-52.

[2]. A.M. García-Vico,1 C.J. Carmona,2,3 D. Martín,4 M. García-Borroto4 and M.J. del Jesus1* (2017) An overview of emerging pattern mining in supervised descriptive rule discovery: taxonomy, empirical study, trends, and prospects, ©The Authors. WIREs Data Mining and Knowledge Discovery published by Wiley Periodicals, Inc.

[3]. Hamad Alhammady (2007) Mining Streaming Emerging Patterns from Streaming Data, 1-4244-1031-2/07/$25.00© IEEE

[4]. Cheng-Hsiung Weng1,2 • Tony, Cheng-Kui Huang3 (2018) Observation of sales trends by mining emerging patterns in dynamic Markets, © Springer Science+Business Media, LLC, part of Springer Nature

[5]. Mai Van Hoan, Matthieu Exbrayat (2013) Time Series Symbolization and Search for Frequent Patterns, ACM 978-1-4503-2454-0/13/12$15.00.

[6]. R. Agrawal, C. Faloutsos and A. Swami (1993) Efficient similarity search in sequence databases, Proc. Of the 4th Conf. on Foundations of Data Organization and Algorithms, pp.69-84.

[7]. K. Chan and W. Fu (1999) Efficient time series matching by wavelets, Proc. of the 15th IEEE Int. Conf. on Data Engineering, pp.126-133.

[8]. E. Keogh, K. Chakrabarti, M. Pazzani and S. Mehrotra (2001) Dimensionality reduction for fast similarity search in large time series databases, Knowledge and Information Systems, vol.3, no.3, pp.263-286.

[9]. J. Lin, E. Keogh, S. Lonardi and B. Chiu (2003) A symbolic representation of time series, with implications for streaming algorithms, Proc. of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, pp.2-11.

[10]. Hsieh-Hui Yu1, Chun-Hao Chen2 and Vincent S. Tseng1 (2011) MINING EMERGING PATTERNS FROM TIME SERIES DATA WITH TIME GAP CONSTRAINT, ICIC International© ISSN 1349-4198

[11]. F. L. Chung, T. C. Fu, V. T. Y. Ng and R. W. P. Luk (2004) An evolutionary approach to pattern-based time series segmentation, IEEE Trans. Evolutionary Computation, vol.8, no.5, pp.471-489.

[12]. Dong, G., Li, J. (1999) Efficient mining of emerging patterns: Discovering trends and differences. In: International Conference on Knowledge Discovery and Data Mining, pp. 43 – 52. ACM Press, New York.

[13]. Li, J., Liu, H., Ng, S.-K., Wong, L. (2003) Discovery of significant rules for classifying cancer diagnosis data. In: European Conference on Computational Biology, Supplement of Bioinformatics, pp. 93 – 102.

[14]. Li, J., Liu, H., Downing, J. (2003) Simple rules underlying gene expression profiles of more than six subtypes of acute lymphoblastic leukemia. Bioinformatics 19(1), 71 – 78.

[15]. Li J,Wong L ( 2002) Identifying good diagnostic gene groups from gene expression profiles using the concept of emerging patterns. Bioinformatics 18(5):725–734

[16]. Huang Z, Gan C, Lu X, Huan H (2013) mining the changes of medical behaviors for clinical pathways. Stud Health Technol Inf 192(1–2):117–121

[17]. Kim JK, Song HS, Kim TS, Kim HK (2005) Detecting the change of customer behavior based on decision tree analysis. Expert Syst 22(4):193–205
[18]. Tsai CY, Shieh YC (2009) A change detection method for sequential patterns. Decis Support Syst 46(2):501–511
[19]. Shie BE, Philip SY, Tseng VS (2013) Mining interesting user behavior patterns in mobile commerce environments. Appl Intell 38(3):418–435
[20]. Deng K, Za¨ıane OR (2010) An occurrence based approach to mine emerging sequences. Springer, Berlin, pp 275–284
[21]. Ceci M, Appice A, Malerba D (2007) Discovering emerging patterns in spatial databases: a multi-relational approach. In: European conference on principles of data mining and knowledge discovery. Springer, Berlin, pp 390–397
[22]. Garc´ıa-Vico AM, Carmona CJ, Mart´ın D, Garc´ıa-Borroto M, del Jesus MJ (2018) An overview of emerging pattern mining in supervised descriptive rule discovery: taxonomy, empirical study, trends, and prospects, vol 8, no 1. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery
[23]. Terlecki P, Walczak K (2007) Jumping emerging patterns with negation in transaction databases–classification and discovery. Inf Sci 177(24):5675–5690
[24]. Chu CJ, Tseng VS, Liang T (2009) RETRACTED: efficient mining of temporal emerging itemsets from data streams. Expert Syst Appl 36(1):885–893
[25]. Ding G, Wang J, Qin K (2010) A visual word weighting scheme based on emerging itemsets for video annotation. Inf Process Lett 110(16):692–696
[26]. Vimieiro R, Moscato P (2014) A new method for mining disjunctive emerging patterns in high-dimensional datasets using hypergraphs. Inf Syst 40:1–10
[27]. Alavi F, Hashemi S (2015) DFP-SEPSF: a dynamic frequent pattern tree to mine strong emerging patterns in streamwise features. Eng Appl Artif Intell 37:54–70
[28]. Li G, Law R, Vu HQ, Rong J, Zhao XR (2015) Identifying emerging hotel preferences using emerging pattern mining technique. Tourism Manag 46:311–321
[29]. Khan MS, Coenen F, Reid D, Patel R, Archer L (2010) A sliding windows based dual support framework for discovering emerging trends from temporal data. Knowl-Based Syst 23(4):316–322
[30]. A.M. Garc´ıa-Vico1, J. Montes2, J. Aguilera2, C.J. Carmona3(B) (2017) Analysing Concentrating Photovoltaics Technology Through the Use of Emerging Pattern Mining, Springer International Publishing AG M. Gra˜na et al. (eds.), International Joint Conference SOCO'16-CISIS'16-ICEUTE'16, Advances in Intelligent Systems and Computing 527, DOI 10.1007/978-3-319-47364-2 32
[31]. R. Agrawal and R. Srikant (1994) Fast algorithms for mining association rules, Proc. of the 20th VLDB Conference, pp.478-499.
[32]. J. M. Ale and G. H. Rossi (2000) An approach to discovering temporal association rules, Proc. of ACM Symposium on Applied Computing, pp.294-300.
[33]. Sarker, I.H. Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective. SN COMPUT. SCI. 2, 377 (2021). https://doi.org/10.1007/s42979-021-00765-8
[34]. CHEN, Y. et al. Contrast Pattern Mining: A Survey. [s. l.], 2022. Disponível em: https://search-ebscohost-com.sdl.idm.oclc.org/login.aspx?direct=true&db=edsarx&AN=edsarx.2209.13556&site=eds-live. Acesso em: 24 jun. 2023.
[35]. Bornmann, L., Haunschild, R. & Mutz, R. Growth rates of modern science: a latent piecewise growth curve approach to model publication numbers from established and new literature databases. *Humanit Soc Sci Commun* **8**, 224 (2021). https://doi.org/10.1057/s41599-021-00903-w
[36]. Nofong, V.M., 2015, Mining Productive Emerging Patterns and Their Application in Trend Prediction, Proceedings of the 13-th Australasian Data Mining Conference (AusDM 2015), Sydney, Australia, https://crpit.scem.westernsydney.edu.au/confpapers/CRPITV168Nofong.pdf
[37]. Han, Jiawei.et. al ,2022, Data mining : concepts and techniques, 4th edition, Morgan Kaufmann, Cambridge, MA
[38]. Li, Q., Zhao, L., Lee, Y.-C., Sassan, A., & Lin, J. (2020). CPM: A General Feature Dependency Pattern Mining Framework for Contrast Multivariate Time Series. Pattern Recognition, 107711. doi:10.1016/j.patcog.2020.107711
[39]. Wu, R., Li, Q., & Chen, X. (2019). *Mining contrast sequential pattern based on subsequence time distribution variation with discreteness constraints. Applied Intelligence.* doi:10.1007/s10489-019-01492-7