International Journal on Emerging Technologies **10**(3): 38-44(2019)

ISSN No. (Print) : 0975-8364
ISSN No. (Online) : 2249-3255

Predictive Mining Model for Transactional Data Pattern using Probabilistic Based Decision Tree Model

M.I. Thariq Hussan[1] and H.S. Saini[2]
[1]Professor & Head, Department of Information Technology,
Guru Nanak Institutions Technical Campus, Hyderabad, Telangana, India.
[2]Managing Director, Guru Nanak Institutions, Hyderabad, Telangana, India.

(Corresponding author: M.I. Thariq Hussani)
(Received 31 May 2019, Revised 17 August 2019 Accepted 28 August 2019)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: Given the huge amount of collected transactional data that comprises a complex event, a predictive data mining extracts the most relevant data based on user behavior system. The lack of an efficient extraction pattern that can simultaneously handle both the simple and complex event function in multivariate temporal patterns makes existing optimization search procedure inefficient in improving the rate of prediction. Also, prediction using the tree-based model involving a feature and threshold for each event increases the prediction rate, but higher complexity involved in balancing tree features makes the system to be of less efficient. In this paper, a Predictive Data Pattern mining using Probabilistic-based Decision Tree (PDP-PDT) is proposed. This model takes the transactional data pattern structure and performs the primary characterization of decision tree sublevels. The probabilities of complex events outcomes are also measured using the Ewens's sampling formula which describes the probabilities associated with identifying decision points. PDP-PDT model reduces the processing time and balances the decision tree structure. It is simple and efficient in terms of true positive prediction rate, decision tree accuracy level, and event prediction rate.

Keywords: Ewens's Sampling Formula, Decision Tree Model, Transactional Data Pattern, Predictive Data Mining.

I. INTRODUCTION

Predictive data mining is done for the purpose of using business intelligence or other data to forecast or predict trends. This type of data mining can help business leaders make better decisions and can add value to the efforts of the analytics team.

Stock market prediction has been an area of intense interest due to the potential of obtaining a very high return on the invested money in a very short time. However, according to the efficient market hypothesis, all such attempts at prediction are futile as all the information that could affect the behavior of stock price or the market index must have been already incorporated into the current market quotation. Technical analysis has been used since a very long time for predicting the future behavior of the stock price.

The requirements for preserving transactional data [6] [17] and the accompanying challenges facing the companies and institutions who aim to re-use these data for analysis or research have been addressed. It points to potential solution within current legal and ethical frameworks, but focus on positioning the problem of re-using these data from a preservation perspective.

Multivariate Reconstructed Phase Space [16] observed multivariate temporal patterns that are high of distinctive in nature and performed event predictive on dynamic data system. The MRPS model used a Bayesian approach to optimize the data being mined but complex event function was not predicted effectively.

Prediction with Tree-Based Models (PTBM) [11] involved a feature on each event prediction. Tree model-based prediction checked for complex events periodically but the processing time was of higher complexity rate on balancing the tree features.

Prediction model using genetic algorithm on mobile web structure [2] provides a technique called MAA increase the number of models related to the region and requested services are observed. It improves the performance in terms of precision, number of mobile models generated, execution time and increasing the prediction accuracy. The framework of optimized user behavior prediction model integrates the temporary and permanent register information and is stored immediately in the form of integrated logs which have higher precision and minimize the time for determining user behavior.

Though different types of distance or similarity functions were evaluated using decision tree models, satisfactory performances were not achieved. A satisfactory approach based on a similarity function called the Extended SubTree(EST) [1] was designed. The approach EST preserved the tree structure in an efficient manner by way of mapping sub-trees rather than nodes minimizing the runtime. But the computational complexity involved was high.

A novel symbiotic machine learning approach proposed to pattern detection [14] and developing predictive models for the onset of first-episode psychosis. It investigates how different patterns of cannabis use related to new cases of psychosis via association analysis and Bayesian techniques. However, the nearest neighbour for temporal sequences was discarded.

A research community of emerging field involves student learning experiences by predicting student's performance [5]. The classification technique applied on datasets to find out hidden information and pattern from database of educational environment. The performance is evaluated based on the parameters like accuracy, precision and recall. The prediction of user participation

[18] in online social networking sites examined based on association rules built with respect to user activeness of current posts. However, the model was not proved to be optimal.

An efficient forecasting system using decision tree model [8] was introduced through which efficient traffic flow prediction was obtained. The integrated model was proved to be efficient only in predicting the traffic flow.

Our contributions are in three folds. First, we present an efficient probabilistic based decision tree model for efficient prediction of the data pattern in a relatively lesser amount of time. Second, our method achieves highly efficient identification of decision points by applying Ewens's sampling formula by obtaining the probability of an outcome based on the constructed decision tree path. Finally, to optimize the complex event prediction on transactional data pattern using Ewens's sampling formula.

The rest of this paper is organized as follows. The next section introduces the related works on predictive mining model in different areas. Section 3 illustrates our proposed predictive data pattern mining using probabilistic based decision tree model which includes tree sublevel construction using transactional data decision tree algorithm, followed by construction of decision tree path using Ewens's sampling formula. Experimental results with dataset descriptions are demonstrated in Section 4. Finally, the conclusion is addressed in Section 5.

II. RELATED WORK

The classification of the cases was developed based on ID3 decision tree algorithm [12] give certain case an assessment of the comprehensive analysis. It can make the attribute selection become more reasonable and avoid compatibility with real attribute.

Medical data mining unearths the latent relationships among clinical attributes for finding interesting facts which helps experts in health care in decision making. Frequent patterns [4] in transactional medical databases that occur periodically are exploited to know the temporal aspects of various diseases. Periodical frequent patterns between the years and monthly patterns extracted using the modified K-means algorithm which improves quality of services with strategic and expert decision making.

A semantic biclustering [9] aims to detect interpretable rectangular patterns in binary data matrices. It requires the included elements can be jointly described in terms of semantic annotations pertaining to both rows and columns. Biclustering algorithm with the semantic descriptions detects sets of compact biclusters. However, the model was confined to a single area.

Hybrid Frequent Itemset Mining [15] utilizes the vertical layout of dataset to solve the problem of scanning the dataset in each iteration. Vertical dataset carries information to find support of each itemset and includes some enhancements to reduce number of candidate item set. It incorporates the concept of resilient distributed datasets and performs in-memory processing to optimize the execution time of operation. However, the nature of distribution remained unaddressed.

With the increasing use and amount of millions of records related to web-based scenarios, they are stored on several web servers, proxy servers, client machines, or organizational databases. Apriori and FP-growth algorithms [10] were applied for performing relational rule mining with respect to temporal attributes. With this, a number of hidden information was obtained at a relatively lesser amount of time.

A hybrid data mining [7] method developed to manage the limited bandwidth in a university network more effectively. It uses the clustering and classification techniques to detect, analyze and predict students' behavioural patterns in a university network and identify the main factors that affect their tendency in using internet.

In fractal time series classification, each class that was singled out unites model time series [13] with the same fractal properties. The self-similarity degree estimation performed and demonstrated a high probability of detection.

In order to overcome above limitations, PDP-PDT model is developed in this chapter for identifying the relations between the moving transactional data patterns with minimum processing time. PDP-PDT model used Transactional Data Decision Tree Algorithm for identifying the moving transactional data patterns efficiently and to reduce the processing time and improve the true positive prediction rate.

III. PREDICTIVE DATA PATTERN MINING USING PROBABILISTIC BASED DECISION TREE MODEL

This section describes the flow structure of generic prediction mining model and the functional representation of predictive data pattern mining using probabilistic based decision tree model. Then, the measurements of probabilistic based decision tree model and decision point identification is discussed by reviewing certain existing works proposed for efficient extraction of pattern mining.

The main objective of the proposed work is to identify the relations between the moving transactional data patterns for predicting dynamic object behavior in stock exchange. To be more specific, the main objective behind the design of PDP-PDT model is efficient prediction of data points relatively moving within the structural transactional pattern. The generic predictive mining structure is depicted in Fig. 1.

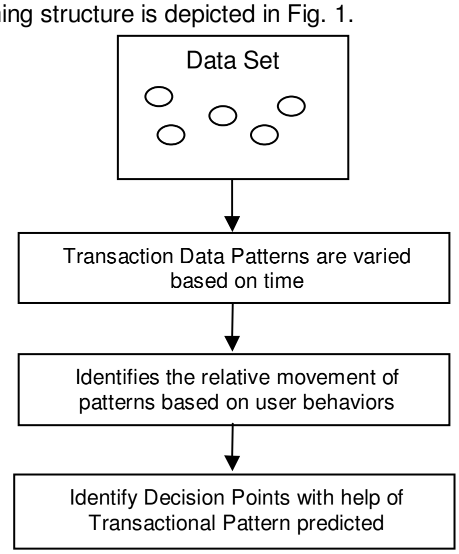

Fig. 1. Flow Structure of Generic Prediction Mining Model.

As shown in Fig. 1, PDP-PDT model initially takes the data set as input. The data set contains standard set of

data for the experimental evaluation where the transactional data pattern gets varied with respect to time. Then, PDP-PDT model identifies the relations between the moving transactional data patterns based on their user behaviors. Finally, PDP-PDT model efficiently identifies the decision points with the aid of predicted transactional data patterns.
The generic prediction mining procedure in PDP-PDT model helps to predict the relations between the moving transactional data pattern. As the values involved in stock exchange gets varied on daily basis, the PDP-PDT model predicts the data accurately with regular changes observed in the stock exchange results. The overall functional representation of PDP-PDT model is illustrated in Fig. 2.

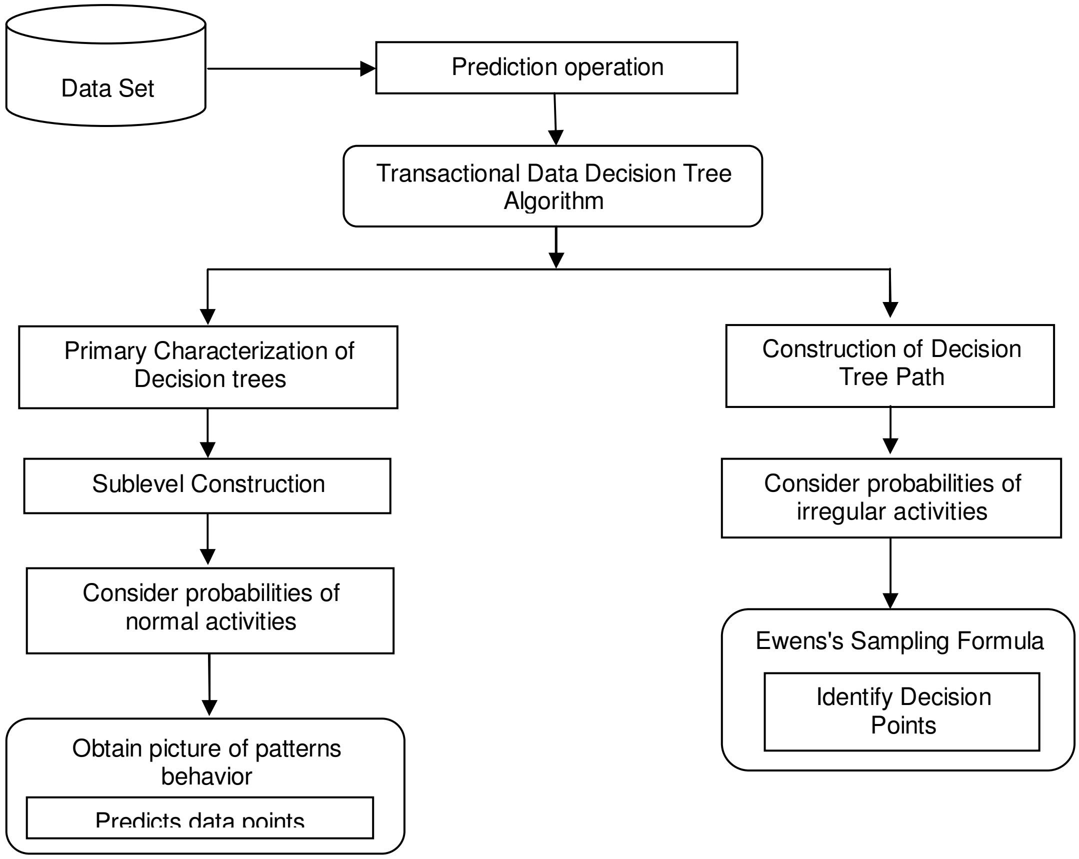

Fig. 2. Functional Representation of PDP-PDT Model.

As illustrated in Fig. 2, PDP-PDT model performs the effective prediction of transactional data with help of Transactional Data Decision Tree Algorithm. PDP-PDT model takes stock exchange data data set as input to predict the dynamic object behaviors. The prediction in PDP-PDT model attains accurate decision points in detecting the various changes observed with minimal processing time. From the Fig. 2, the operation of predictive mining is carried out in two different phases. The first phase is to characterize the decision tree primarily using the sublevels where the sublevel is constructed with the aid of Transactional Data Decision Tree algorithm. The transactional data decision tree algorithm considers the entire path in decision tree. The decision tree is used for efficient ordering of pattern that considers the probabilities of normal activities in order to obtain the picture of patterns, behavior and that accurately predicts the data points which results in improved true positive prediction rate than the existing MRPS model.
The second phase performs efficient construction of Decision Tree Path to identify the decision points where the transactional data patterns are aggregated to assemble detail profile revealing time occurrence of irregular activities, node involved in abnormalities with the help of probabilistic outcomes. With the help of identified irregular activities, Ewens's Sampling Formula identifies the probabilistic result rate with relatively lesser amount of time when compared to the existing tree based prediction models. The elaboration on primary characterization of decision trees with the construction of probabilistic based decision tree is briefed in the forthcoming subsections.

A. Primary Characterization of Decision Trees
The main objective behind the design of characterization of decision trees is to obtain the sublevels of the decision tree for accurate prediction using the Transactional Data Decision Tree algorithm, taken as an interesting variant to construct the sublevels. To start with, the PDP-PDT model is initially applied to characterize the patterns and then split the moving data points. Decision tree in PDP-PDT model is highly flexible in handling the complex data events.
Let us consider a scenario where complex events are represented as ‘E’ for predicting the transactional data pattern using the decision tree with ‘M’ and ‘N’ relatively moving data patterns that are to be considered as the input and output patterns for effective identification of the prediction rate. Using PDP-PDT model, let M(x) represents the moving

patterns on condition ‘x’ then the output predicted model is N(y). The basic idea behind the generic characterization of the decision tree using PDP-PDT model is to define the level of separation where each level wise separation for the entire data patterns is obtained as:

$$E(n) = \sum_{x \in M} P[M = x] \sum_{y \in N} P[N = y] \quad (1)$$

For the effective construction of a probabilistic-based decision tree, ‘n’ complex events are obtained using E(n) with ‘P’ representing the probabilistic count for moving patterns ‘M’ and ‘N’. The results of probabilistic test are obtained from the above equation and PDP-PDT model achieve higher result rate on characterization of decision tree. The transaction data decision tree is obtained using the pattern separation with the aid of the confidence value. With this, the specific tree path performs the ordering of patterns for easy prediction using the confidence value.

The results of prediction are shown using leaf nodes with the aid of a probabilistic-based decision tree in PDP-PDT model. This is because the prediction in the leaf nodes is easily performed by measuring and counting the sublevels where the first sublevel count in the decision tree is computed as

$$P[M1(x)] = P[M1(d1, d2, d3 \dots dn)] \quad (2)$$

Let us consider the input data pattern to be of ‘M’ is fetched through the root node and followed by it, processed to characterize using the data points d1,d2,d3...dn where M1 clearly specifies the first level of pattern separation for accurate prediction results. In a similar manner, the second sublevel count in the decision tree is computed as:

$$P[M1(x)] * P[M2(x)] = P[M1(d1, d2, d3 \dots dn)] * P[M2(d1, d2, d3 \dots dn)] \quad (3)$$

The second level computes the cumulative joint probabilities based prediction on relational moving data patterns. In a similar manner, the above equation is applied for ‘n’ sublevels of the moving data patterns. With this, the joint probabilities combine the transactional data patterns to construct an effective decision tree.

B. Transactional Data Decision Tree (TDDT) Algorithm

The algorithmic steps involved in the design of TDDT are given below that efficiently constructs the sublevels in the decision tree with higher confidence level where the sublevels are separated when the transaction moving data pattern varies from each group. Similar structures in PDP-PDT model are grouped together and also the similar data patterns are placed on the same side of the decision tree whereas the dissimilar structure is placed on the other part of the decision tree. The algorithmic description of TDDT is explained as given below:

//Transactional Data Decision Tree algorithm
Begin
Step 1: Decision tree with root node ‘R’ symbolize the start of moving patterns.
Step 2: Complex event node ‘E’ of ‘n’ types taken for prediction of moving relational data pattern.
Step 3: Decision tree represent node M(x) and N(y)
Step 3.1: M(x) - Input relational moving pattern with ‘x’ values for predicting.
Step 3.2: N(y) - Output relational moving pattern with ‘y’ value predicted.
Step 4: Sublevel count is computed through (2) and (3)
Step 4.1: All the sublevels are joined using cumulative joint probabilities based prediction.
Step 5: Joint probabilities result combine all the sublevels and perform effective decision tree construction.
End

The algorithmic step is explained clearly with very simple example set that takes into consideration the stock trade of different countries with varying index points used for experiment with which the algorithmic step predicts the data pattern in a relatively lesser amount of time. The stock trading points vary on the daily basis for different ‘n’ types of shares. Sometimes the complex event also takes place on the stock trade and therefore transactional data is used for predicting the result set. With the construction of joint probability, combine the entire sublevels for effective decision tree construction.

C. Probabilistic based Decision Tree Path Construction

Once the primary characterization of decision trees is formed the second phase in PDP-PDT constructs the decision tree with a probabilistic procedure. The decision tree path construction with a higher relative movement of transactional data easily processes the entire system and enhances the prediction accuracy.

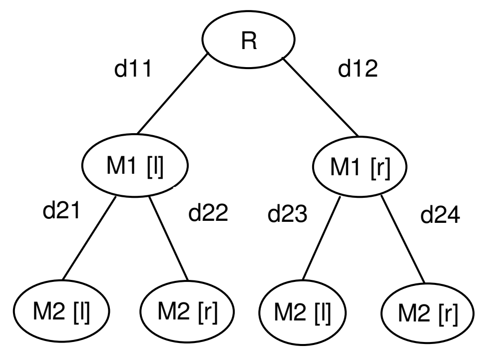

Fig. 3. Decision Tree Path Construction.

Fig. 3 depicts the decision tree path construction using the root node ‘R’ comprising of data points ‘d11’ and ‘d12’ respectively on sublevel 1 of the decision tree. The decision tree on the subsequent sublevels helps to easily predict the relatively moving data patterns.

D. Ewens's Sampling Formula

With the efficient probability of outcome generated from the construction of a decision tree path, decision points are identified using Ewens's sampling formula. Ewens’s sampling process in PDP-PDT model takes all the different types of the transactional data pattern to perform efficient prediction processing. The processing time taken for the relatively moving transactional data pattern is formularized as:

$$\mathrm{P}\big((\mathrm{d11}, \mathrm{d12}, \mathrm{d21}, \mathrm{d22}, \mathrm{d23}, \mathrm{d24}), \mathrm{time}\big) = \frac{\mathrm{n}!}{\mathrm{time}(\mathrm{time}+1)+\cdots(\mathrm{time}+\mathrm{n}-1)} \prod_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{time}\,\frac{\mathrm{M_i}}{\mathrm{i}^{\mathrm{M_i}}\mathrm{M_i}!} \quad (4)$$

The probabilistic result ‘P’ of prediction is received for different sublevel data point. In order to predict the processing time, the time for ‘n’ types of complex events are considered whereas the level of each relative moving data points is measured based on the M_i input data patterns.

IV. EXPERIMENTAL SETUP

A. The Dataset

The dataset used to evaluate the performance of PDP-PDT is based on the Istanbul Stock Exchange (ISE) [3] dataset from UCI repository and is experimented on JAVA platform. This study focuses on analyzing the transactional data pattern to perform the predictive pattern mining using the ISE extracted from UCI repository during the financial turmoil period from February 2015 to June 2017. The stock exchange returns with 100 indexes tested and the result compared among the existing MRPS and PTBM with PDP-PDT model.

B. Preprocessing

N*N sample are taken to define and uses a unique dataset compiled from 25-minute intraday values of ISE-100 stock market return index of Germany, UK, Japan, Brazil, Singapore and Turkey. The experiment is conducted on factors such as processing time, true positive prediction rate and decision tree accuracy level. The experiment runs on an Intel Core-2 Duo with 6 GB memory.

C. Metrics

In order to measure the accuracy of prediction pattern mining methods, we observe the average tree depth from 2 to 14 for experimental purpose. The following well-known metrics are used in the predictive mining area:

True positive prediction rate (TPPR) for PDP-PDT is the ratio of a number of similar data patterns Number_{SDP} observed to the number of dissimilar data patterns Number_{DDP} observed for transaction data extracted from ISE.

$$TPPR = \frac{Number_{SDP}}{Number_{SDP} + Number_{DDP}} \quad (5)$$

The PDP-PDT model measures the decision tree accuracy level (DTA) by obtaining the ratio of a number of correct predictions Number_{CP} made for complex events to the total number of predictions Total_P observed for a given size of tree Size_T.

$$DTA = \frac{Number_{CP}}{Total_P} * Size_T \quad (6)$$

The processing time observed for relatively moving transactional data pattern is obtained with different sublevel data points for 'n' complex events which are measured in terms of milliseconds (ms). Finally, the event prediction rate using PDP-PDT model is measured, where a similar structure is grouped together cumulative joint probabilities based prediction increasing the event prediction rate.

D. Experimental Results

D.1 True Positive Prediction Rate (TPPR) by PDP-PDT, MRPS and PTBM

Table 1 shows the average TPPR of PDP-PDT, MRPS, and PTBM with financial turmoil period from February 2015 to June 2017 using ISE dataset compiled from 25-minute intraday values. The values in the table show the performance difference between a number of similar data patterns and a number of dissimilar data patterns on the relational movement of transaction data. The results demonstrate that PDP-PDT model provide an improvement in TPPR comparing with MRPS and PTBM.

Table 1: Tabulation of TPPR.

Average Tree Depth	True Positive Prediction Rate (%)		
	PDP-PDT	MRPS	PTBM
2	0.45	0.25	0.20
4	0.75	0.55	0.50
6	1.25	1.05	1.00
8	1.45	1.25	1.20
10	1.85	1.65	1.60
12	1.95	1.75	1.70
14	2.50	2.20	2.15

However, using PDP-PDT model, the TPPR for handling complex data events significantly outperforms the other two methods, MRPS, and PTBM respectively. PDP-PDT model provides the highest rate of TPPR, which is 10-44% and 12-55% better than the existing. This is because, with the design characterization of decision trees, the obtained sublevels of decision tree results in accurate prediction using the TDDT algorithm. The transactional data decision tree highly flexible in handling complex data events improves the TPPR.

D.2. Decision Tree Accuracy Level (DTA) using Predictive Mining Models

To evaluate the classification performance of decision tree accuracy level using predictive mining models with respect to the size of the tree (i.e., using the number of indices), two well-known predictive mining models are designed and compared. The experimental setting considered for evaluating the DTA is to consider the stock trade of different countries with the size of the tree with varying index points consisting of 100 indices. For experimental purposes, the size of the tree in terms of indices used ranges from 10 to 80.

The effectiveness of the DTA using predictive pattern mining model by comparing the results with those of MRPS and PTBM is illustrated in Fig. 4. The improvement of the DTA is mainly due to the consideration of the probabilistic-based decision tree on the characterization of decision tree during prediction.

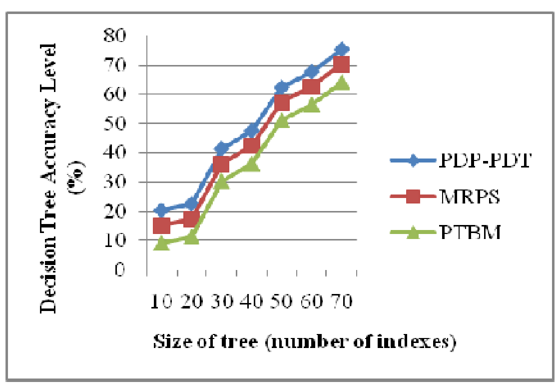

Fig. 4. Measure of Decision Tree Accuracy Level.

The probabilistic-based decision tree can retrieve the similar data patterns as well as dissimilar data patterns on two different sides of the decision trees with the aid of cumulative joint probabilities. In contrast, the conventional prediction based pattern mining model does not consider the complex events on transactional data pattern. In the case of ISE dataset, most of the key information of stock exchange is organized with regard to working days. Thus, the transaction data pattern observed on the basis of time and relative movement of patterns using the user behavior is sufficient to retrieve both simple and complex events thereby the performance using PDP-PDT is significantly improved. In our experiments, PDP-PDT achieves 7.44-24.93% improvement in DTA compared to MRPS and 14.77-55.06% improvement over PTBM.

D.3. Comparison Analysis of Processing Time

In order to examine the efficiency of the PDP-PDT model using TDDT algorithm, the processing time measured with an average tree depth of 14 using ISE dataset.

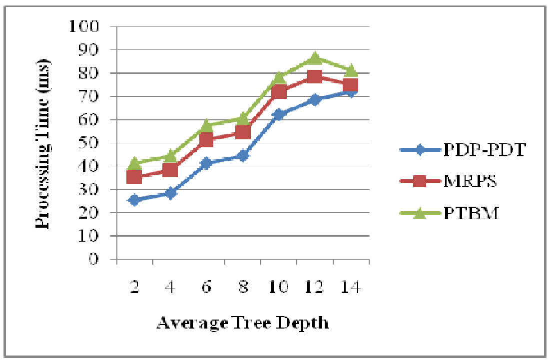

Fig. 5. Measure of Processing Time.

For most of the average tree depth, the results of processing time are performed within a reasonable time and improvement observed from 3-38%. The PDP-PDT model using TDDT algorithm directly accesses the data by constructing the sublevels in the decision tree with a higher confidence value, the processing time depends on the average tree depth.

To explore the influence of processing time on PDP-PDT model, the experiments were performed by applying seven different tree depth obtained from ISE data set as depicted in Fig. 5. The processing time observed is comparatively improved because the decision points identified using probabilistic outcomes with the aid of Ewens's sampling formula.

V. CONCLUSION AND FUTURE SCOPE

In this work, an efficient predictive mining model is analyzed and optimized the complex event prediction on transactional data pattern to increase the true positive prediction rate by using a probabilistic based decision tree model. This model predicts data pattern which is fairly relevant over relatively moving transaction data points by identifying the decision points to the relevance. The probabilistic based decision tree model identifies the relative importance of decision tree sublevels and determines the relevance of relative movement of patterns based on user behavior. Based on this measure, a new transactional data decision tree algorithm proposed which reflects the weights of similar and dissimilar data patterns relative to the confidence value. It has been derived the measure for complex events based on the probability of outcome and evaluation of root node. It is observed that the proposed model provides accurate search results with relatively lesser processing time and improved decision tree accuracy level compared to the existing predictive mining models. As a future scope, the performance of prediction can be improved by determining user behavior in complex event data transaction.

ACKNOWLEDGEMENT

This work was supported by the research department of Guru Nanak Institutions. We would like to thank the almighty and family members who have provided continuous support to guide me in improving the paper.

Conflict of Interest: Nil

REFERENCES

[1]. Ali Shahbazi, and James Miller, (2014). Extended Subtree: A new Similarity Function for Tree Structured Data. *IEEE Transactions on Knowledge and Data Engineering,* Vol. *26*(4): 864–877.

[2]. Thariq Hussan, M.I., and Kalaavathi, B., (2015). An Optimized User Behavior Prediction Model using Genetic Algorithm on Mobile Web Structure. *KSII Transactions on Internet and Information Systems,* Vol. *9*(5): 1963–1978.

[3]. Vassiliki Bamiatzi., Konstantinos Bozosa., and Neophytos Lambertides., (2016). Mapping the Trading Behavior of the Middle Class in Emerging Markets: Evidence from the Istanbul Stock Exchange. *International Business Review,* Vol. *25*(3): 679–690.

[4]. Khaleel, M.A., Dash, G.N., Choudhury, K.S., and Khan, M.A., (2015). Medical Data Mining for Discovering Periodically Frequent Diseases from Transactional Databases. *Computational Intelligence in Data Mining,* Vol. *1*: 87–96.

[5]. Ankita Katare, and Shubha Dubey, (2017). A Study of Various Techniques for Predicting Student Performance under Educational Data Mining. *International Journal of Electrical, Electronics and Computer Engineering,* Vol. *6*(1): 24–28.

[6]. Sara Day Thomson., (2016). Preserving Transactional Data. *International Journal of Digital Curation,* Vol. **11**(2): 126–137.

[7]. Elham Akhond Zadeh Noughabi., Behrouz H. Far, and Bijan Raahemi, (2016). Predicting Students Behavioral Patterns in University Networks for Efficient Bandwidth Allocation: A Hybrid Data Mining Method. *IEEE International Conference on Information Reuse and Integration, Pittsburgh, PA,* 102–109.

[8]. Kalli Srinivasa Prasad., and Seelam Ramakrishna., (2014). An efficient traffic forecasting system based on spatial data and decision trees. *International Arab Journal of Information Technology,* Vol. *11*(2):186-194.

[9]. Jiri Klema., Frantisek Malinka., and Filip Zelezny., (2017). Semantic Biclustering for Finding Local, Interpretable and Predictive Expression Patterns. *BMC Genomics,* Vol. **18**, Article number: 752.

[10]. Nazli Mohd Khairudin., Aida Mustapha., and Mohd Hanif Ahmad., (2014). Effect of temporal

relationships in associative rule mining for web log data. *Hindawi Publishing Corporation, The Scientific World Journal,* Article ID. 813983.
[11]. Nima Asadi., Jimmy Lin., and Arjen P. de Vries., (2014). Runtime optimizations for prediction with tree-based machine learning models. *IEEE Transactions on Knowledge and Data Engineering,* Vol. *26*(9): 2281-2292.
[12]. Yurong Zhong., (2016). The Analysis of Cases Based on Decision Tree. *IEEE International Conference on Software Engineering and Service Science, Beijing,* 142–147.
[13]. Lyudmyla Kirichenko., Tamara Radivilova., and Vitalii Bulakh., (2018). Machine Learning in Classification Time Series with Fractal Properties. *IEEE Second International Conference on Data Stream Mining and Processing, Ukraine,* 198–201.
[14]. Wajdi Alghamdi., Daniel Stamate., Katherine Vang., Daniel Stahl., Marco Colizzi., Giada Tripoli., Diego Quattrone., Olesya Ajnakina., Robin M. Murray., and Marta Di Forti., (2016). A Prediction Modelling and Pattern Detection Approach for the First-Episode Psychosis Associated to Cannabis Use. *IEEE International Conference on Machine Learning and Applications, Anaheim, CA,* 825–830.
[15]. Krishan Kumar Sethi., and Dharavath Ramesh., (2017). HFIM: A Spark-Based Hybrid Frequent Itemset Mining Algorithm for Big Data Processing. *The Journal of Supercomputing,* Vol. **73**(8): 3652–3668.
[16]. Wenjing Zhang., and Xin Feng., (2013). Event characterization and prediction based on temporal patterns in dynamic data system. *IEEE Transactions on Knowledge and Data Engineering,* Vol. *26*(1):144-156.
[17]. Hussan, MIT., Kalaavathi, B., (2013). Multi-Cluster Based Temporal Mobile Sequential Pattern Mining Using Heuristic Search. *WSEAS Transactions on Computers,* Vol. *2*(11): 432–439.
[18]. Erlandsson, F., Borg, A., Johnson, H., and Brodka, P., (2016). Predicting User Participation in Social Media. *Advances in Network Science, Lecture Notes in Computer Science,* Vol. *9564*: 126–135.

How to cite this article: Thariq Hussan, M.I. and Saini, H.S. (2019). Predictive Mining Model for Transactional Data Pattern using Probabilistic Based Decision Tree Model. *International Journal of Emerging Technologies*, **10**(3): 38–44.