ISSN 2347 - 3983

Volume 8. No. 1.1, 2020

International Journal of Emerging Trends in Engineering Research

Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter3181.12020.pdf

https://doi.org/10.30534/ijeter/2020/3181.12020

Machine Learning Supervised Analysis for Enhancing Incident Management Process

Aida Mustapha[1], Salama A. Mostafa[1], Marwan Hamid Hassan[2],Mohammed Ahmed Jubair[1], Shihab Hamad Khaleefah[1] and Mustafa Hamid Hassan[1]
[1]Faculty of Computer Science and Information Technology,
UniversitiTun Hussein Onn Malaysia, Parit Raja 86400, BatuPahat, Johor, Malaysia,
Email: {aidam, salama}@uthm.edu.my {mohamed.a.jubair, shi90hab, mustafa.hamid.alani}@gmail.com
[2]Department of Da'wa and Rhetoric, Imam Al-adham College, 31001, Anbar, Iraq,
marwanhh80@yahoo.com

ABSTRACT

This paper presents the descriptions of the attributes which are composed of an enriched incident management process event log. The degree of accuracy of the models for prediction depends on the usefulness of the log attributes utilized in building such models are and by using a machine learning algorithm approaches to the attributes gives a better comprehension of the background (underlying) process. This paper studies the classification method that applies for deciding which best feature among the features in an enriched event log dataset for the incident management process. The classification method compared to other methods used in related work. The result will state which one algorithm is the highest and the best one to be selected. The selection of an attribute is vital to building or creating a model with completion time prediction capability by determining concept description features to be learned and ways of feature combination. In this paper, a classification method was applied. Depending on the feature-target feature association (correlation), every attribute is separately analyzed. The outcomes indicate that the technique used outperformed human expert's decision making. We conducted predictions over different periods and achieved satisfactory performance in terms of accuracy, whereas, the best performance for classifying is achieved by the Bayes Net algorithm of 85.2760% accuracy.

Key words: Incident Management, Event Log, Machine Learning, Classification.

1. INTRODUCTION

An occurrence able to bring about halt or disruption to or loss of an organization's operational capabilities, services or functions is called an incident [1]. A term used to describe the organization's activities to determine, evaluate, and mitigate risks or hazards and avoid a future recurrence is Incident Management (IcM). An unmanaged incident can lead to an emergency, crisis or a disaster. A process, therefore, of restricting the potential disruption initiated by such an occurrence, followed by a return to normal business operation, is Incident Management. Business functions, employees, IT systems [2], information security [3], clients [4], or several essential business operations can be disrupted by an incident without existing effective Incident management [5]. In a process enriched event log dataset for incident management, data is collected from the Service Now platform's event or instance audit system. IT companies use Service Now TM platforms that focused on the transformation of information technologies by automating and standardizing business processes, as well as their integration across the enterprise [6]. Loaded from a relational database underlying a corresponding process-aware information system is the log of event data. In completion time prediction, an enriched log of event dataset of the Incident Management process contains 32 attributes that are applied in determining the dependent variable.

In terms of build a prediction model, considering both descriptive attribute sets and the event log is essential. This is needed to allow a detailed exploration of the event logs' huge data record [7]. For estimate completion time, conduct prediction focuses only on using the performance of naive and the actual process's superficial abstraction resulting in lower quality estimates [8]. By using process mining, large data can infer the data into a realistic process model [9]. Generally, using data mining methods and other similar techniques have different ways of improving the performance of Process mining and are defined as obtaining novel information and useful knowledge to essentially bolster decision-making procedures [10]. In data mining, there are specific tasks, and a different technique can always be applied to find the solution or method used to find the pattern. For the mining process, the specific context data will be gathered, transformed, and finally, organized before undergoing mining. For the human view, the mining will give a result that is already organized in the structured form [11, 12].

This paper studies the classification methods applied for deciding which best feature among the features in the enriched event log dataset for the incident management process. The classification method will be compared to other methods used in related work. The result will state which one algorithm is the highest and the best one to be selected.

The rest of this paper organization has section 2 surveying all related incident management process enriched log works. Section 3 contains the classification methodology in carrying out the data mining activity including the datasets and the evaluated parameters (metrics). Results are in Section 4 and lastly, a conclusion with some recommendation for future work is presented in Section 5.
submission.

2. RELATED WORK

various researches and studies have carried out on classification methods such as the work of Ros et al. [13] which implemented an attribute selection techniques using two wrapper methods and filter method; and a filter-incorporated-ranking and the wrapper with hill-climbing with best-first as heuristic search techniques. To subsequently be supplied to the prediction model, attributes for description in an authoritative subset is automatically determined by implementing these classical attribute selection techniques. Annotated Transition Systems (ATS) are a remarkable prediction model example as they in context extensively rely on the utilized attributes. For comparing the different techniques used, ATs are the prediction model choice. For the experimental report and analysis present, e is mean error on time prediction (in seconds), θ is ATS-implemented and e' is instances of a practical process of the incident management event log.

The methodology in the discussion here was created to resolve a real issue of time prediction that Information Technology (IT) organizations face. The event log collection and a series of incident description attribute in this organization are enabled by the Service Now TM platform-bolstered incident management process. It is an experimental application because there is a nonexistent initiative (scenario) for comparison and human experts performed attributes selection which was utilized as baselines to get around this challenge. In every case, the evaluation of semantic reasonability of the selected attributes in this practical process for management of incident is performed. Only the wrapper dependent solution can potentially surpass human experts as shown in the outcomes. The goal of the work is discovering an attribute subset during its resolution process that permits creating models or prototypes that are able to minimize the incident completion time's prediction error.

Weinzierlet al. [14] presented a prediction models building approach that applies a cross-validation technique with 5 folds to the enriched log of the event. It is implemented with the used event log, execution details and experiments set up. The mean MAPE of test folds in terms of median and mean of the incident's completion time is the ATS accuracy. Furthermore, the evaluation of the ATS completeness (or non-fitting) is determining the number of event records lacking a corresponding state in ATS. So, in an incident occurs, a caller identifies and reports it. Thereafter, knowing the incident completion time is a major (primary) desire (expectation).

The normal approximate adhere to Information Technology Infrastructure Library (ITIL) standards, which are dependent on some specific attributes of incident e.g. category, urgency etc. As it totals a vast amount of various scenarios and corresponding target completion times, this technique is fairly general and thus, inaccurate. Updates of some attributes are done, and new ones are included during process evolution from starting-phase to initial support and investigation. This can normally result in a value nearer to 100 attributes based on the utilized system. This process begins with progressively planned activities to produce the improved log of event needed in inducing the predicting model. Thereafter, applying the three attribute selection techniques examined in this study becomes possible; firstly, an expert-driven selection, secondly, a ranking based filter and finally, wrappers incorporating hill-climbing and best first search approaches. The MAPE for ATS being evaluated is determined when the evaluation function is implemented. The model index that brings out the least MAPE when applied to the testing log is returned when the application of minimization function to the ATS assessment is carried out. Conclusively, it returns the ATS with the least MAPE in the ATSs' set being evaluated. ATS application is the prediction model with the responsibility of producing the incident completion times' estimates, as well as acting as the wrapper search spaces' state evaluator in all the selection approaches. For a real-world application, the fundamental concept is that ATS can be produced from an attribute subset which appropriately depicts the recently finished incidents. For this reason, ATS is suitable in predicting the completion time of novel incidents as they run.

Sarnovsky et al (2018) [15] utilized the associated changes and the dataset of incident records to specify two major domains: (1) the correlation in the reported element of the infrastructure and affected one, (2) the correlation between associated changes and the incident. The use of various techniques in measuring the testing set's model accuracy. A Receiver Operator Characteristic Area Under the Curve (ROC AOC), normally utilized to express results for ML binary decision problems is used as the main metric. The best model (Random Forest) achieved the best results.

The confusion matrix shows the arrangement into specific classes and classification. On the other side in this task and more importantly is to verify the fact, if the reported CI is responsible for the incident. More precise from the task viewpoint is the classification of this class and, the error rate on this class can be more telling compared to the other one. In both cases, the development of predictive models resolves both presented problems using Random Forest and GBM models. Finding the best models possible was a major aim, and using the ROC curve, all models were assessed on the testing set. The CRISP-DM methodology guides the process.

3. METHODOLOGY

This section explains the method used, proposed solutions and the basis for the experiments. Data Mining denotes mining or digging deep into data, which is in several forms in creating patterns, and gaining knowledge on that pattern. In the data mining process, sorting huge data sets, patterns identification, and establishing associations are carried out for data analysis and problem-solving.

Type of method used for this experiment is classification. It is about a data analysis task. The problem of identifying to what set of class (subpopulations) a discovery belongs, depending on a training set of data comprising of findings and whose category membership is known is classification. The experiments were performed utilizing a 10-fold validation technique Weka toolset for training and testing. Weka is generally, an assembly of ML algorithms for data mining activities. The algorithms are called from either via a Java code or used directly on a dataset. Tools for data pre-processing, regression, clustering, classification, association rules and visualization are also found in Weka.

3.1 Dataset

The Dataset that is being used for this project is obtained from data collected from an IT company-utilized audit system of an instance from the Service Now TM framework. This extracted dataset is referred to as the Incident Management Process Event Log Dataset. There are 36 attributes and characterized as an integer. For this dataset, the number of instances is 141712. This dataset provides from UCI Machine Learning Repository. The dataset description is in Table 1 and samples of Management Process Event Log Dataset is in Table 2.

Table 1: Description Event Log Dataset

data Set Characteristics:	Multivariate, Sequential
Attribute Characteristics:	Integer
Associated Tasks:	Regression, Clustering
Number of Instances:	141712
Number of Attributes:	36
Missing Values?	Yes
Area:	Business
Date Donated	2019-07-14
Web Hits:	18966

Table 2: Samples of Management Process Event Log Dataset

numberINC	**incident**	**active**	**caller_id**	**opened by**	**opened at**
0000045	New	TRUE	2403	8	29/2/2016 01:16
0000045	Resolved	TRUE	2403	8	29/2/2016 01:16
0000045	Resolved	TRUE	2403	8	29/2/2016 01:16
0000045	Closed	FALSE	2403	8	29/2/2016 01:16
0000047	New	TRUE	2403	397	29/2/2016 04:40
0000047	Active	TRUE	2403	397	29/2/2016 04:40
0000047	Active	TRUE	2403	397	29/2/2016 04:40
0000047	Active	TRUE	2403	397	29/2/2016 04:40

3.2 Materials and Methods

There are five algorithms used in this project, which are consist of Trees Decision Stump, Bayes Bayes Net, Bayes Naïve Bayes, Rules One R, and rules. Zero R.

Decision stump Algorithm: It is a model in ML containing a one-level decision tree which is a decision tree that has a single internal node (roots) which is immediately linked to the terminal nodes (leaves) see figure1. A decision stump predicts relying on the value of only a single input feature see figure. The formula is shown in Equation 1.

$$\boldsymbol{f(x|j,t)} \coloneqq \{^{\boldsymbol{+1} \quad \boldsymbol{x}^{(j)} > t}_{\boldsymbol{1 - otherwise}} \tag{1}$$

where j ∈ {1, . . ., d} indexes an axis in Rd

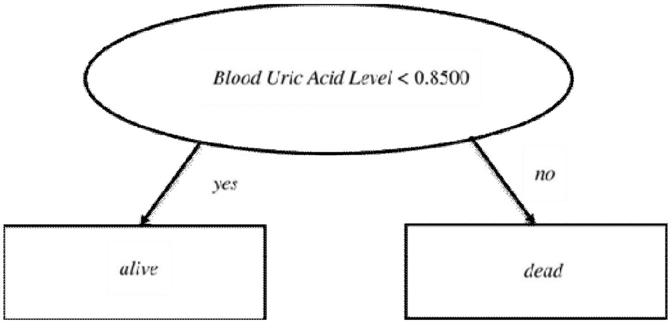

Figure1: an example of a decision stump

Bayes Net Algorithm: It is a probabilistic graphical model type that applies Bayesian deduction for computing probabilities. Bayesian networks tend to map conditional reliance, and hence, causes, by representing conditional reliance by edges in a coordinated graph. It is via these connections that one can proficiently and effectively make deductions on the arbitrary variables in the graph via the use of factors. The formula is shown in Equation 2.

$$\mathbf{Pr}(\boldsymbol{G}, \boldsymbol{S}, \boldsymbol{R}) = \mathbf{Pr}(\boldsymbol{G}|\boldsymbol{S}, \boldsymbol{R})\, \mathbf{Pr}(\boldsymbol{S}|\boldsymbol{R})\, \mathbf{Pr}\,(\boldsymbol{R}) \qquad (2)$$

where G = "Grass wet (true/false)", S = "Sprinkler turned on (true/false)", and R = "Raining (true/false)”.

Naive Bayes Algorithm: This class of simple "probabilistic classifiers" utilizes Bayes' theorem with strong (naïve) independence assumptions between the features. They are among the simplest Bayesian network models [1]. Extensively studied since the 1960s, Naïve Bayes was introduced (though not under that name) into the text retrieval community. It remains a fundamental (baseline) technique using word frequencies as the features for text categorization, providing solutions on documents judging problems; determining the category a document belongs (document categorization) (e.g. politics or sports, legitimate or spam, etc.). Its competitiveness in this area with suitable pre-processing compared with more advanced techniques such as support vector machines is remarkable. [3] In medical diagnosis automation, it is also highly applicable. [4] The formula is shown in Equation 3. Where A = Class, B = Data

$$\boldsymbol{P}(\boldsymbol{A}|\boldsymbol{B}) = \frac{\boldsymbol{P}(\boldsymbol{B}|\boldsymbol{A})\boldsymbol{P}(\boldsymbol{A})}{\boldsymbol{P}(\boldsymbol{B})} \qquad (3)$$

Rules.OneR Algorithm: It is an accurate classification algorithm that simply chooses the rule with the least total error as its "one rule" by creating a rule for every predictor in the data. In generating a rule for a predictor, a frequency table is produced for every predictor against the target. OneR creates rules sparingly less accurate than sophisticated classification algorithms while generating rules that are easily interpreted by as has been shown. The example formula is shown below.

```
For every predictor,
For every value of that predictor, make the following rule;
    - Count frequently every target (class) value is showing up
    - Look for the most recurring class
    - Ensure the rule assignclass to the value of the predictor
    -Evaluate the total error of the rules of every predictor
-Select the predictor having the least total error.
```

Rules.ZeroR Algorithm: It is referred to as 0-R orZeroR. It is a majority class classifier and it is the simplest of the rule-based classifiers. The 0-R (zero rule) classifier looks out for the target attribute and its possible values. In the given dataset, the most recurring value discovered for target attribute is always output from the 0-R.

3.3 Evaluation Metrics

There are three experimental evaluation metrics utilized for this work. They are accuracy, precision, recall and F measure [16], [17], [18].

Accuracy: In a general term, a measure of the degree of closeness of value from measurement or calculation to the real values is accuracy. Equation 4 is the formula used to evaluate accuracy.

$$CR = \frac{C}{A} \qquad (4)$$

where CR = The correct rate, C = The number of correctly recognized samples, A = The number of all sample.

Precision: Precision is the ratio of True Positives and the sum of True Positives and False Positives. In other words, it is the proportion of correctly predicted values with the total number of positive class values predicted. Precision calculation expression in Equation 5.

$$Precision = \frac{True\ Positive}{True\ positive + False\ Positive} \qquad (5)$$

Recall: The recall is the ratio of True Positives and the sum of True Positives and the number of False Negatives. Put another way it is the number of positive predictions divided by the number of positive class values in the test data. The formula for calculation of Recall is shown in Equation 6.

$$Recall = \frac{True\ Positive}{True\ positive + False\ Negative} \qquad (6)$$

F Measure:Generally, the balance between the recall and the precision is F Measure. It is additionally referred to as F1 Score. Equation 7 shows the F Measure calculation.

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \qquad (7)$$

4. RESULTS AND DISCUSSION

In this section, we will discuss the result that outcome from several apothems that applied in Incident Management Process data classification such as Trees.DecisionStump, Bayes.BayesNet, Bayes.NaiveBayes, Rules.OneR, and rules.ZeroR,which are used in the process for Incident management's improved event log dataset for completion time predictor experiments with the purpose of carrying out a performance comparison. Table 3 presents the outcomes of several algorithms of Machen learning and compare the result in terms of Accuracy Precision, Recall and F-Measure.

Table 3: Experimental results

Algorithm	Attributes	Accuracy	Precision	Recall	F
Decision Tree Stump	Incident state	99.0714 %	0.332	1.000	0.498
	Category	21.3459 %	0.753	0.735	0.744
	Priority	96.0568 %	0.986	1.000	0.993
Bayes BayesNet	Incident state	63.0906 %	0.731	0.617	0.669
	Category	95.8068 %	0.898	0.999	0.946
	Priority	96.9304 %	1.000	0.968	0.984
Bayes NaiveBayes	Incident state	63.5592 %	0.676	0.670	0.773
	Category	94.3015 %	0.872	0.984	0.924
	Priority	96.4865 %	1.000	0.964	0.981
Rules.OneR	Incident state	45.1994 %	0.365	0.624	0.461
	Category	96.9322 %	0.569	0.981	0.720
	Priority	99.0714 %	0.997	0.996	0.996
Rules.Zero R	Incident state	27.3202 %	0.273	1.000	0.429
	Category	13.0287 %	0.130	1.000	0.231
	Priority	93.4656 %	0.935	1.000	0.966

The results showed that there are two approaches employing a higher number of algorithms return the best results. These results use Bayes Algorithm. The best performance for classifying is achieved by the Bayes.BayesNet Algorithm and it produces an 85.2760% accuracy. This algorithm is taken as the best method because it shows higher value accuracy than other algorithms, such as Trees.DecisionStump, Bayes.NaiveBayes, Rules.OneR, and rules.ZeroR. It means that the correctly classified feature is near to the total number of instances which contains 141712 instances of the dataset.

5. CONCLUSION

Under the control of problem management and existing (registered) in the known-error database, incidents may match existing problems (having an unknown underlying cause) or 'known errors' (having a known underlying cause). This paper presents the descriptions of the attributes which are composed of an enriched incident management process event log. The focus of this paper is the study of the classification method appropriate for deciding which best feature among the features in the incident management process enriched event log dataset. The classification method is compared to other methods used in previous works. As a conclusion and based on the algorithm, the best feature classifier is Bayes Net Algorithm. The best performance for classifying is exhibited by the Bayes Net Algorithm and achieving 85.2760% accuracy. The future work is to study more about the other data mining technique in the machine learning algorithm. We are also interested in studying more about various machine learning software that has been used on large data.

ACKNOWLEDGEMENT

This paper is supported by Research Fund E15501, Research Management Centre, UniversitiTun Hussein Onn Malaysia.

REFERENCES

1. R.Mühlberger, S Bachhofner, C. Di ,L.García-Bañuelos, &O. López-Pintado. **Extracting Event Logs for Process Mining from Data Stored on the Blockchain.***In International Conference on Business Process Management* 2019, pp. 690-703.
2. Di Ciccio Claudio. **Towards a Process-oriented Analysis of Blockchain Data**.*In Modellierung (Companion)*, 2020, pp. 42-44.
3. A.Mandal, S.Agarwal, N.Malhotra, G.Sridhara, A.RayandD. Swarup.**Improving IT Support by Enhancing Incident Management Process with Multi-modal Analysis**. *In International Conference on Service-Oriented Computing,*2019, pp. 431-446.
4. A.Ray andD. Swarup.**Improving IT Support by Enhancing Incident Management Process with Multi-modal Analysis**. *In Service-Oriented Computing: 17th International Conference, ICSOC* 2019, Vol. 11895, p. 431.
5. S. A. Mostafa,M. S.Ahmad, A., AhmadandM. Annamalai.**Formulating situation awareness for multi-agent systems.***In International Conference on Advanced Computer Science* Applications and Technologies, 2013, pp. 48-53.
6. M.Atzmueller, S.BloemheuvelandB. Kloepper.**A Framework for Human-Centered Exploration of Complex Event Log Graphs**. *In International Conference on Discovery Science* ,2019, pp. 335-350.
7. N.Razali, S. A.Mostafa, A.Mustapha, M.H.AbdWahab andN. A. Ibrahim. **Risk Factors of Cervical Cancer using Classification in Data Mining**. In Journal of Physics: Conference Series,2020, Vol. 1529, No. 2, p. 022102. IOP Publishing.
8. M. Marrone, F. Gacenga, A. Cater-Steel and L. Kolbe. **Commune of the Association for Inform**. *Sys*., 2014, vol. 34, pp. 49.1–49.30.
9. N. A. S.Mohammed, Zaidi, A.Mustapha, S. A.MostafaandM. N. Razali.). **A Classification Approach for Crime Prediction**. In Communications in Computer and Information Science, 2019, pp. 68-78, Springer, Cham.
10. A. R. C.Maita, L. C Martins, C. R. L.Paz, L.Rafferty, P. Hung, S. M.Peres andM.Fantinato.**A systematic mapping study of process mining**. *Enterprise Information Systems,* 2018.v. 12, n. 5, pp. 505-549.
11. M. H. Hassan, S. A.Mostafa, A.Mustapha, M. H. A.Wahab and D. M. Nor.**A survey of multi-agent system approach in risk assessment**. *In International Symposium on Agent, Multi-Agent Systems and Robotics*, 2018, pp. 1-6.
12. C. A. L.Amaral, M.Fantinato, H. A Reijers andS. M.Peres.**Enhancing Completion Time Prediction**

Through Attribute Selection. *Lecture Notes in Business Information Processing*, 2019, v. 346, pp. 3-23.

13. F.RosandS. Guillaume.**From Supervised Instance and Feature Selection Algorithms to Dual Selection: A Review.***In Sampling Techniques for Supervised or Unsupervised Tasks*, 2020, pp. 83-128.
14. S.Weinzierl, S.Zilker, J.Brunk, K.Revoredo, A.Nguyen, M.Matzner andB. Eskofier. **An empirical comparison of deep-neural-network architectures for next activity prediction using context-enriched process event logs.** *arXiv preprint arXiv*,2020.01194.
15. P. T.Figueira, C. L.Bravo andJ. L. R. López.**Improving information security risk analysis by including threat-occurrence predictive models**. *Computers & Security*, 88, 101609.2020.
16. A. B. Annasaheb and V. K. Verma.**Data mining classification techniques: A recent survey**. *International Journal of Emerging Technologies in Engineering Research*, 4(8), pp 51-54,2016.
17. B. J. Chelliah, S. Kalaiarasi, A. Anand, G. Janakiram, B.Rathiand N. K. Warrier.**Classification of Mushrooms using Supervised Learning Models**. *International Journal of Emerging Technologies in Engineering Research (IJETER)*, 2018, 6(4),.
18. N. Razali, A.Mustapha, M. H. AbdWahab, S. A.Mostafa andS. K. Rostam,. **A Data Mining Approach to Prediction of Liver Diseases**. *In Journal of Physics: Conference Series, 2020,Vol. 1529, IOP Publishing.*