The Application of Machine Learning Techniques in Software Project Management- An Examination

Mahesh Babu Pasupuleti

Data Analyst, Department of IT, iMinds Technology Systems, Inc., 1145 Bower Hill Rd, Pittsburgh, PA 15243, **USA**

ABSTRACT

Planning and evaluating project management are key parts of project performance that should not be overlooked. It is difficult to succeed at project management unless you have a realistic and logical plan in place. This paper provides a comprehensive overview of papers on the application of machine learning in software project management, covering a wide range of topics. Apart from that, this study examines machine learning, software project management, and methodologies. Papers in the first category are the results of software project management studies or surveys. Papers in the third category are based on machine-learning methods and strategies applied to projects; studies on the phases and tests that are the parameters used in machine-learning management; and final classes of study results, contribution of studies to production, and promotion of machine-learning project prediction. Our work also provides a larger perspective and context, which could be useful for future project risk management research, among other things. To summarize, we have demonstrated that project risk assessment using machine learning is more effective in minimizing project losses, increasing the likelihood of project success, providing an alternative method for efficiently reducing project failure probabilities, increasing the output ratio for growth, and facilitating accuracy-based analysis of software fault prediction.

Keywords: Machine Learning, Project Estimation, Software, Project Management

11/20/2018 **Source of Support:** Nil, **No Conflict of Interest:** Declared

INTRODUCTION

Project managers must overcome challenges such as increasing the efficiency of a software project and ensuring the project's long-term viability and sustainability. A project's likelihood of failure is often caused by a lack of knowledge, skills, resources, and technological capabilities during the project's implementation.

The knowledge gained from historical project data sets can be applied to the development of predictive models, either through the application of mathematical methodologies such as linear regression and study of association, or through the application of machine learning (ML) approaches such as Artificial Neural Networks (ANN) and Support Vector Machines (SVM) techniques (SVM). Methods for forecasting the future of a project that are based on current and past project evidence are referred to as predictive methodologies. Because there are so many different machine learning algorithms, it has not been possible to study them all. According to the findings of the literature, the rationale for adopting automated projects, the difficulties of project management evaluation, and the development of ML methodology are all addressed. The empirical findings would be analyzed and assessed.

Despite the fact that project literature covers the success and failure of a project, there is a long-running argument on how project improvements can be assessed and reported. Different people have different perspectives of project performance and different assessments of project success. Hughes and other members of the Project Management Institute (PMI) distinguish between variables that influence project success and variables that influence project performance.

The project progress thresholds are examined in order to determine if a project is a success or a failure. The feedback received is also taken into consideration for the project's advancement. A successful project that is completed within a specified project period has historically been differentiated by the achievement of the desired results and the efficient usage of the available resources. PMI recognizes projects that are successful in achieving the project goals, criteria, and ambitions of the stakeholders involved. Researchers such as Como Aladwani, Cates and Mollaghasemi, Parsons, and Rosenfeld describe the effects of classical objective outcome metrics, such as project expense (above and below budget), project time (early, on, or late), and the project outcomes output, on the project's overall performance and outcomes (with less or better than required properties and functions).

Research Questions

When doing a systematic literature review, the research questions are crucial in determining the search approach and analysis. For this study, we developed the following research questions (RQs):

- What does the existing research literature reveal regarding the use of machine learning techniques in software project management?
- Can we construct better machine learning-based models in terms of accuracy prediction by using feature transformation and feature selection to efficiently reduce project failure probabilities?
- What are the current research gaps in the subject of Software Project Management?
- What are the prediction metrics, and what is the present level of accuracy demonstrated by various estimating techniques?

SOFTWARE PROJECT MANAGEMENT ESTIMATION BASED ON ML

Methodology for Developing Software Project Management Estimation: Steps in the Process.

1. Obtaining information is the first step in the project. Token extraction and word tokens are crucial qualities for assessing project work. Tokens are responsible for defining a significant portion of the project's material. A key component of the estimating model was the use of Unigram language modeling in terms of tokens, which served as a foundation for the model.

2. Following the extraction of tokens, the following project characteristics were selected for future investigation: For each token, the term frequency–inverse document frequency (TF–IDF) of the phrase is calculated. The term frequency–inverse document frequency (TF–IDF) of a phrase is a tool for calculating the sense of a phrase by taking into account word frequency and the reverse counts of records that include this word. This is a way for determining the meaning of a word.

3. For each of the different classification procedures, the resulting features are used as inputs to the Weka classification algorithm. The following is a summary of our decision-making process when selecting learning algorithms: It is a probabilistic classification method focusing on the Bayes theorem that takes independent features from a classroom and uses them to make a classification. Simple classification techniques such as Nave Bayes tackle large dimensionality of data with subjective assertion of freedom, which can outperform more complex classification systems. J48 Decision tree: Java's open source C4.5, which is a decision tree-generating method for situations where the tanning collection is not linearly separable. Decision trees are highly suited for these situations since they are fast and accurate. There are numerous decision trees in the Random Forest, and the class outputs are a statistical model of the individual tree output groups. The Random Forest is a category classifier that is composed of several decision trees and the class outputs.

REVIEW AND SURVEY ARTICLES

The study and research documents provided an overview of the most recent perceptions of machine learning technologies in SPM creation and evaluation, as well as the implementation of machine learning algorithms. Studies on Machine Learning and Their Application in SPM have been carried out:

The ML procedures are discussed and shown in this segment. Each of these articles was divided into a number of different subjects and implementations. Selected studies have been classified into large groupings, with the majority of the studies focusing on the ML approaches of manufacturing techniques. There were three subcategories for each of the six publications in this group, with three publications in each category. The K-Nearest Neighbor Algorithm was used to construct this sub-cluster (KNN). It was determined and analyzed in (Adusumalli, 2016a) that the observations, metrics, data sets, calculation measures, machine learning problems, multiple forecasting models, and ensemble models employed in the region of the maintenance prediction were all valid and reliable. ML technologies are becoming increasingly important, as evidenced by the use of KNN for the handling of missing values in information engineering data structures in one study (Pasupuleti, 2015a).

Several other categorization experiments were been out using Regression as well. The methodologies and predictors of volatility forecasting, as well as the classification criteria, were identified in the work (Adusumalli, 2016b). Characteristics that were employed as indicators of literary volatility parameters and forecasting methodologies that were used

to improve the precision of literature volatility parameters and prevision needs have been identified. Specifications with volatility are crucial for software projects since they directly affect costs and the length of the program's development cycle. In Pasupuleti (2015b), it was argued that the SLR could aid in the development of a formal method of reproducible discoveries. The organization of a data set, like other data sets, cannot be used to determine the specific application of the study's findings. Another (Adusumalli, 2017a) dealt with the application of machine learning algorithms for calculating the program effort. Based on the systemic analysis, it was discovered that machine learning methodologies, size scales, comparison data sets, assessment procedures, and other factors were important. One article on the Fuzzy Logic Studies (Pasupuleti, 2015c) looked into the usage of machine learning approaches to test the effort put into a software. Also included were examples of software work and cost analyses of systems-functioning approaches, with the major conclusion being that no alternative methodology should be selected by the process and model, as opposed to the traditional approach.

CASE STUDY

With the ML technique, this stage examines a project, campaign, or firm to determine the current situation, recommended remedies, and implementation measures, as well as the reasons that contributed to the failure or success of SPM development techniques. Articles in this part are mostly concerned with machine learning methods, and selected works have been organized into broad groups based on the use of ML approaches in SPM development procedures. The seven entries in this category were separated into three subcategories, each of which contained a single article.

Studies on Bayesian Networks Algorithm are covered in the articles in this category. Using a combination of qualitative and machine learning solutions, a solution for value estimation is provided in Pasupuleti (2016a). A probabilistic model encompassing the knowledge from various stakeholders will be used to predict the overall value of a given decision relating to product management and development. According to the authors of (Adusumalli, 2017b), an intelligent decision support system was used to construct a model that automatically determines the link between risk factors and mitigation measures (DSS). A number of commonly noted current risk management limitations are addressed in the technique, including a lack of a standardized DSS and the relationship between software risks and their mitigation.

The research for the articles in this category was done using the Fuzzy algorithm. The articles (Pasupuleti, 2016a; Fadziso et al., 2018) presented a fuzzy mathematics method into parametric modeling of risk influence diagrams in order to overcome the severe problem of obtaining probabilities of key occurrences, which is a difficult challenge to solve. Through the establishment of a topology structure of risk factors, the work describes the interaction between different influence variables in the risk management process of information technology projects. The data are compared to a number of different assessment metrics.

STUDIES CONDUCTED ON MACHINE LEARNING METHODS

The selected studies were divided into major groups based on the ML algorithms used in the SPM software. In this category, there are forty-three articles, each of which is broken into nine subcategories.

This section contains fifteen ML process studies that made use of a variety of SPM algorithm variations. Selected works are categorized into big groups using the methodology of machine learning in software development (ML in software production).

The first category of studies (Pasupuleti & Amin, 2018) analyzed applications from the field of SPM, categorizing their behaviors into two categories: working behavior and pertinence. It revealed that the model for calculating the system effort had a multi-target learning problem when it was designed. This aids in the understanding of the trade-offs between multiple performance measures by developing SEE models that were simultaneously automated by many objective evolutionary algorithms, such as genetic algorithms and evolutionary algorithms. The techniques that were used in this analysis were Naive Bayes, Logistic Regression, and Random Forests. Naive Bayes is a statistical technique that is used to detect patterns in data.

Using task text as input, two studies (Pasupuleti, 2017) developed an automated machine learning-based technique of assessing software work. An artificial neural network (ANN) is used to simplify effort estimating routines. It is possible to obtain results that exceed those obtained from the literature, and a system that promises to be much easier to integrate into any software SPM tool that stores textual task descriptions is based primarily on the use of textual descriptions of tasks that, unlike various other methods, are almost always readily available.

In (Fadziso et al., 2018), the authors presented the results of a reflection on applied data mining work, which included social metrics, effort estimation, test case production, and other aspects of the work. This informal examination yielded the conclusions that were later codified and systematized into the seven principles and a dozen additional recommendations. The purpose of this paper is to present techniques to achieving successful industrial data mining outputs; however, it should be noted that some of these principles may also be applicable to academic data mining.

The research on (Adusumalli & Pasupuleti, 2017) has produced a new hybrid model that is more accurate. As a result of its ability to be used in a single database, the model is excellent for a wide range of operations. Two machine learning algorithms, the ANN and the SVM, are used to evaluate the performance of our model. The experiments demonstrate that our SVM threat forecasting model is a more robust version of itself.

DISCUSSION

This paper examines key studies in modern project management using ML technology. This study's purpose is to highlight research trends in this field. This study is old and doesn't include implementation, but the literature. This study differs from others. The literature is provided as taxonomy. Creating a literature taxonomy might have several advantages in a research field. On one hand, literary taxonomy is widely publicized.

What are the research gaps in Software Project Management?

A fresh researcher researching a software project evaluation may be overwhelmed by the amount of information available, the lack of a framework, and the number of publications addressing developing trends in project management. Several studies have developed ML models and implementations.

A literature taxonomy helps arrange and categorize these different works and situations. However, taxonomy methodology provides useful information to scholars. It then

describes potential research areas. The present software evaluation taxonomy suggests, for example, that researchers appear to recommend strategies for program development and operation. The newest project evaluation and usage of ML technology are also presented.

The literature mapping shows weak and excellent study coverage into project review proposals. For example, the taxonomy emphasizes individual claim examination and assessment over consolidated procedures, structures, and growth activities (being expressed in the abundance of their categories). The taxonomy also revealed a lack of study on the development of project reviews. The research literature is vital. Studies in this field strive to improve and exchange ML. Industries participating in ML approach to deal with new innovations and enhance dormant fields are classified statistically by taxonomy divisions. This study provides a taxonomy for scholars to collaborate and examine new technologies, such as developments, comparative studies, and project appraisals using ML technology. The analysis demonstrates three aspects of the literature: the drivers driving the rise of automated project management using ML technologies, the obstacles to successful implementation, and solutions.

Challenges

Although computer-based training techniques, such as those used in SPM evaluation, have a number of advantages, they are not considered to be the best choice for projects that are in the process of evolving. According to the results of the surveys, academics are interested in evaluating projects and their application of machine learning technologies. The most significant challenges to the deployment of machine learning techniques are given below, along with some other topics. The challenges have been identified (Figure 7).

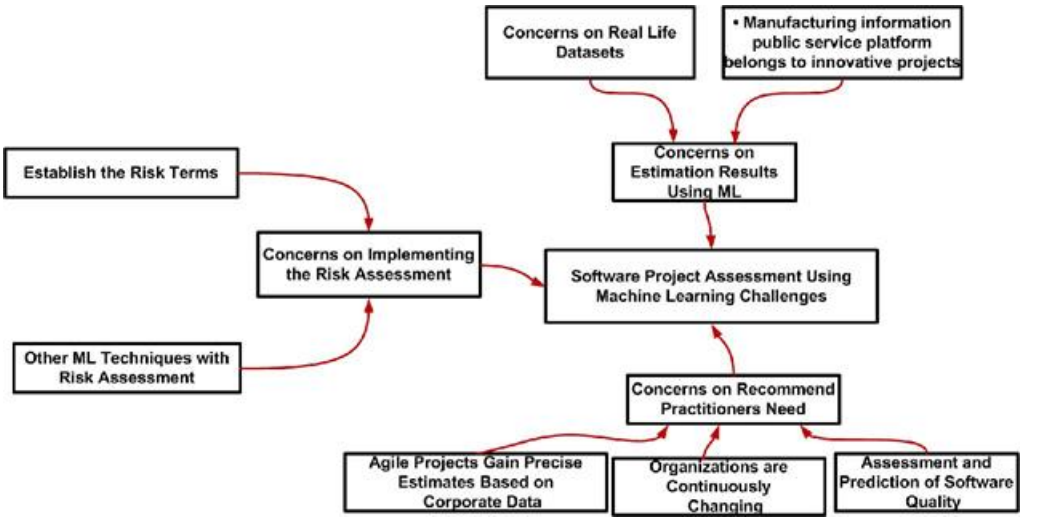

Figure 7: Machine Learning Project Evaluation.

ML Estimation Results: Concerns

Real-world data sets lack fundamental software development methodologies and require other reliable metrics to evaluate effort. More study is necessary to verify the conclusions of program effort projections. Iterative and cross-validation methods are mostly validated. Also, research trends show that calculating procedures need to be investigated and improved. Regression trees, size metrics, and other ML algorithms can be studied using

real-world data. The public service portal is part of knowledge manufacturing initiatives. This means our project team must update development practices more frequently, and engineers must monitor the development platform and infrastructure. Machine Learning for Software Project Evaluation.

Concerns with Risk Assessment

Risk assessment concerns: when the project begins. First, define the project hazards and risk elements, including demand shifts, infrastructure threats, staff coordination, and device protection strategies. The risk evaluation network focuses on current instances to construct the risk case learning process. Given the effectiveness of the two systems, it was required to assess the degree of risk of one condition, which led to the most suited framework for Bayesian classification. It is recommended in this discipline since it improves estimation outcomes for techniques like ML. Data from a single entity dataset are extracted into multiple homogenous groups, depending on the organization or sector. Stratification improved the accuracy of effort estimates.

Practitioners Needs Concerns

Be cautious about the calculating methods and data sets they employ for their initiatives. Agile initiatives get better estimates from corporate data. Similar to public data, private enterprises benefit from restricted internal project data. Further reporting on XP projects are encouraged when the organization implements more productive and agile steady growth strategies like XP. Despite the same productivity, recent research shows that XP projects have better effort management. Adaptive-project-control systems will be required as organizations change. Project managers are used to adding or removing KPIs and altering expert evaluation procedures. PMIS must manage such situations. Construct well-defined automated processes between top management and PMIS internal settings, promising novel relationship aspects in how these two entities connect.

The main basis for applying ML approaches in large computer firms is for evaluating and estimating product quality. The numerous product indicators can be utilized to create a software quality model utilizing ISO 15939 to measure and anticipate software output as well as meet these organizations' quality information criteria. The utility of ML approaches for such evaluation is documented. ISO 9126 defines quality as “the completeness of a software product's features and features that can satisfy stated or implicit needs.” According to ISO 25000, “the ability of software products to meet stated, indicated specifications under certain circumstances”. Assessing software quality early in the development process requires locating and allocating resources.

RECOMMENDATIONS FOR SOFTWARE EFFORT ESTIMATION

Software involvement measurement is critical for software success. In the software industry, precise computing effort estimates are required for appropriate project planning and budgeting. Price adjustments can cause overestimated company gains. Underestimation will overrun schedule and cost, costing the company a lot of money. Because the effort is large, the literature instead uses software effort and software cost terminology that relates to the estimated value.

What are the forecast metrics and how accurate are they currently?

The Hybrid Model has long been an important subject of study in program completion estimate. So we needed numerous high-performing independent models to improve

prediction precision and reliability. A hybrid proto-type has three distinct sets of attributes: a text-based attribute called summation; a text-based attribute called a definition; and metadata-based attributes. Unlike prior models in, this model is not dependent on a single data source that is not always available. It is not clear whether the same hybrid tactics apply to completion time estimation models.

Data Homogeneity the cost results homogeneity. This issue has never been addressed in machine estimation literature. An important aspect investigated in another study is the impact of the prediction success training data size. Because of disparities in data quality and prediction methods, studies have yet to demonstrate that this is an unresolved question. The latest research will help project managers assess how much data is needed to train the algorithm. Use ML to measure program expenditures and analyze performance. It may also inspire and guide other scholars to do similar studies. The suggested work method could target domain-specific variables to improve data quality and predictability. Data analysis in the embedded device realm may have done this.

Clustering Methods can take many forms. A more organized update mechanism for CC projects, as well as other clustering algorithms and simple students may be addressed. Dycom used clustering to create the CC subgroups. There are three ways of clustering: hierarchical clustering, K-means, and preference maximization. The original Dycom Clustering uses four CC sub-sets of four different sizes SEE tables. The study includes a toilet arrangement. K-mean clustering Dycom helps separate CC programs, with delivery equivalent to or greater predictive efficacy than Dycom. The number of CC subsets must also be predetermined, and an incorrect decision may harm the predicted results.

Such issues urge one to suggest an interactive CBR for managers using Case-Based Reasoning (CBR) Technology. The method aims to improve managers' understanding of the cost estimation mechanism. The CBR cycle is easy to grasp since it replicates human memory storage and retrieval. With the procedures performed, managers can better comprehend how the calculation was rendered. The CBR-C approach outperformed ANGEL. The current CBR-C technique has an indexing function that reduces prediction error. The CBR technique addresses current challenges by resolving historical difficulties. Detailed CBR commitment estimation. Here are the basic CBR regulations. The most comparable past projects are chosen to estimate the current project's cost. Are there any function weighting algorithms that improve on the present function selection methods?

RECOMMENDATIONS FOR RISK PREDICTION

Risk management is required to improve the efficiency of software projects regardless of their business industry. Consumer expectations are still not met in tested systems. Identifying project risks is crucial to determining project success or failure. Almost every company uses sophisticated tools to assess, mitigate, and eliminate damage.

Will a high-risk initiative be recognized in time? However, existing models assume that all error classification costs are equal and that probability estimation is in the software project. The cost of anticipating a failed project as a successful project differs from predicting a successful project as a failed project. To our knowledge, the cost-sensitive learning technique is still not used in the context of outsourcing software project risk management. There are two key research gaps in the software project risk prediction model. First, new risk prediction techniques for outsourced software projects are rarely studied. Second, despite extensive research on software project risk prediction, no one has used cost-sensitive

learning algorithms. Recently, the project's Agile-based software management was productive. Schedule-based hazards are time-related threats that effect deliverables. The cash, time projections, and the project manager's good efforts are not correctly distributed. Budget threats highlight financial risks from funds' swarming. Unintentional project reach expansion, low usage of existing outlets, and inadequate management. Regular project procedures include operational danger forms. These threats stem from incorrect procedures, poor planning, and teamwork. Discuss the value of agile approaches and contemporary frameworks in risk management. In the future, the team might use the risk factors to assess the threat's impact. Using such criteria also increases the likelihood of good production. Risk prediction models: A 50% estimate of the software venture delays variant reduces the amount of 'keyword' scan questions. Several categorization scales are used to compare the correctness of the received Bayesian models. The tree's improved network architecture shows successful experimental results for all data sets. The link between factors obtained by necessity engineers to determine the situation's hazard degree. Bayesian networks are useful in risk management automation. Risk management in software development seeks to identify, assess, prepare for, and respond to potential risks. The case findings validate the technique principle and predict the software project evaluation (Pasupuleti & Amin, 2018). To estimate the risks and implications of hazards, as well as to characterize the primary risk elements, is to model threats. However, the long development cycle, high product sophistication, and method instability make it impossible to predict and assess the project's risk. It is important to note that the present software project risk assessment philosophy often evaluates risks from general project management.

CONCLUSION

According to the findings of the literary analysis, an extensive study in software project management using ML methodologies had been conducted. There has been a consistent spread of employment opportunities across time. The ANN, Fuzzy Logic, Genetic, and Regression Algorithms are the most important machine learning approaches for autonomous effort estimation, according to the authors. One of the most important aspects in software development is the accurate estimation of effort requirements. The program was specifically influenced by the amount of time and difficulty available to it. Basic themes in software project management can be gleaned from a variety of machine learning (ML) works. Generally speaking, these endeavors can be divided into four categories: the first category includes reviews and surveys related to software project management; the second category includes papers that focus on case studies of software project management methods; the third category includes experimental publications that have been used in the management of ML, a type of structure or architectural model; and, the final category of the research contribution study is an analysis of a project, form, or structure. An in-depth analysis of these publications would assist software project management in reviewing machine learning methodologies in order to describe and explain the hazards, advantages, and suggestions that are associated with them. However, due to the large number of machine learning algorithms available, various machine study algorithms have remained unstudied. Following that, based on the findings of the literature, the reasons for adopting automated SPM, the issues of project preparation assessment, and machine learning engineering technologies are studied. However, despite the fact that project performance and loss are explained in the literature on SPM, there has been a long legacy of debates over whether project progress should be calculated. There are differences of opinion about what constitutes project development and how it should be estimated; these recommendations

will help to settle the issues that arise while developing software using machine learning methods and will offer up employment opportunities in this field. Research is still underway to investigate the estimation of effort based on machine learning algorithms that are primarily concerned with risk assessment. In addition, normal filtering methods are used to minimize the problem by forming districts with the same stakeholders and forecasting whether or not a stakeholder is aware of the situation. Preliminary answers to critical concerns about Software Project Management Estimation based on machine learning (ML) were presented by this literature review.

REFERENCES

Adusumalli, H. P. (2016a). Digitization in Production: A Timely Opportunity. *Engineering International*, *4*(2), 73-78. https://doi.org/10.18034/ei.v4i2.595

Adusumalli, H. P. (2016b). How Big Data is Driving Digital Transformation?. *ABC Journal of Advanced Research*, *5*(2), 131-138. https://doi.org/10.18034/abcjar.v5i2.616

Adusumalli, H. P. (2017a). Mobile Application Development through Design-based Investigation. *International Journal of Reciprocal Symmetry and Physical Sciences*, *4*, 14–19. Retrieved from https://upright.pub/index.php/ijrsps/article/view/58

Adusumalli, H. P. (2017b). Software Application Development to Backing the Legitimacy of Digital Annals: Use of the Diplomatic Archives. *ABC Journal of Advanced Research*, *6*(2), 121-126. https://doi.org/10.18034/abcjar.v6i2.618

Adusumalli, H. P., & Pasupuleti, M. B. (2017). Applications and Practices of Big Data for Development. *Asian Business Review*, *7*(3), 111-116. https://doi.org/10.18034/abr.v7i3.597

Fadziso, T., Adusumalli, H. P., & Pasupuleti, M. B. (2018). Cloud of Things and Interworking IoT Platform: Strategy and Execution Overviews. *Asian Journal of Applied Science and Engineering*, *7*, 85–92. Retrieved from https://upright.pub/index.php/ajase/article/view/63

Pasupuleti, M. B. (2015a). Data Science: The Sexiest Job in this Century. *International Journal of Reciprocal Symmetry and Physical Sciences*, *2*, 8–11. Retrieved from https://upright.pub/index.php/ijrsps/article/view/56

Pasupuleti, M. B. (2015b). Problems from the Past, Problems from the Future, and Data Science Solutions. *ABC Journal of Advanced Research*, *4*(2), 153-160. https://doi.org/10.18034/abcjar.v4i2.614

Pasupuleti, M. B. (2015c). Stimulating Statistics in the Epoch of Data-Driven Innovations and Data Science. *Asian Journal of Applied Science and Engineering*, *4*, 251–254. Retrieved from https://upright.pub/index.php/ajase/article/view/55

Pasupuleti, M. B. (2016a). Data Scientist Careers: Applied Orientation for the Beginners. *Global Disclosure of Economics and Business*, *5*(2), 125-132. https://doi.org/10.18034/gdeb.v5i2.617

Pasupuleti, M. B. (2016b). The Use of Big Data Analytics in Medical Applications. *Malaysian Journal of Medical and Biological Research*, *3*(2), 111-116. https://doi.org/10.18034/mjmbr.v3i2.615

Pasupuleti, M. B. (2017). AMI Data for Decision Makers and the Use of Data Analytics Approach. *Asia Pacific Journal of Energy and Environment*, *4*(2), 65-70. https://doi.org/10.18034/apjee.v4i2.623

Pasupuleti, M. B., & Amin, R. (2018). Word Embedding with ConvNet-Bi Directional LSTM Techniques: A Review of Related Literature. *International Journal of Reciprocal Symmetry and Physical Sciences*, *5*, 9–13. Retrieved from https://upright.pub/index.php/ijrsps/article/view/64

--0--