Machine Learning Integration of Herzberg's Theory using C4.5 Algorithm

Alexen A. Elacio[1], Luisito L. Lacatan[2], Albert A. Vinluan[3], Francis G. Balazon[4]
[1]AMA University, Philippines, alexen_elacio@hotmail.com
[2] AMA University, Philippines, lllacatan@amaes.edu.ph
[3] AMA University, Philippines, Albert.vinluan@gmail.com
[4] AMA University, Philippines, fbalazon@yahoo.com

ABSTRACT

The emergence of the global economy, particularly to the ASEAN Community in 2015, marked a significant milestone on its path towards becoming a highly competitive region in terms of the labor force integrated into the global economy. Human Resource Management Society revealed that businesses would spend the equivalent of six to nine times the monthly wage of an employee to locate and train a replacement. The Philippines is recognized to have a high aptitude skilled workforce and top talent. The I.T. industry became diversified, where sophisticated skillset is part of job hiring requirements. Retention of employees became a critical success factor of an organization. In the 1950s, Fredrick Herzberg, a psychiatrist, carried out a study of employee satisfaction. Herzberg proved in his experimentation test that job satisfaction and dissatisfaction are related to workplace environment factors. Some situations make an employee engaged and discouraged, which affects the attitude towards the job. The Machine Learning, as part of Data Mining classification tasks towards a knowledge acquisition process. It is categorized and carried out through the analysis. This method is intending to use talent databases to conduct the process of acquiring talent expertise and retention pattern in Human Resource (H.R.). The decision tree family C4.5 classifier algorithm suggests an appropriate dataset classifier. The classification method performed is using the management of talent for the identification or estimation of retention of employees. The decision tree induction adopts a top-down approach, which begins with a tuple training set and related class labels. As the tree constructs, the training set is repetitively dividing into smaller subsets. The result of this study affirms Herzberg's theory of satisfaction through a mathematical technique C4.5 Algorithm. The employment retention model using the machine learning technique identifies the dataset properties that can avoid business disruptions in multiple ways. It is measuring the robustness of the datasets using WEKA, giving a 99.06% data performance.

Key words: C4.5 Algorithm, Decision Tree, Herzberg Motivation Theory, Human Resource Management, Machine Learning

1. INTRODUCTION

The 21st-century debate on Human Resources Management assumes new complexities due to the current realities concerning the needs of employees, the aspirations of corporate stakeholders, the nature of the workforce, and other changing issues. Almost all recognized principles and assumptions of management practice have involvement in probing in recent times. [1][2]. In the last 15 years, Southeast Asia has experienced impressive growth and is one of the most competitive regional economies in the world. Studies by the Human Resources Management Society showed that companies should spend six to nine times the monthly salary of workers to find and train their replacements [3]. Highly skilled labor is becoming increasingly important for economies in the region to remain competitive and meet their growth targets. The I.T. industry is becoming more diverse, and more complex and sophisticated skills are challenging job requirements [4].

Maintaining a productive workforce is one of the top priorities for organizational success in today's business environment. The lack of satisfaction of the employee may have severe implications for the organization. The turnover of employees can have a domino effect on the quality of service, productivity, and overall success of the organization. [5]. The exploratory multi-case analysis has sought to define techniques that business leaders consider to be necessary to maintain I.T. expertise as presented in the "Herzberg Motivation-Hygiene Theory" and the "March and Simon Process Change Model."[3][6]. The Hygiene considerations are also referring to as maintenance factors. The biological, security, and love needs, which are variables that are not directly related to the job but to the circumstances that surround doing the work. Motivator considerations pertain to the quality of the situation, are intrinsic to the job itself, and do not benefit from "carrot and stick opportunities." These include the physiological need for development and recognition. The lack of these factors does not prove to be extremely dissatisfactory, but, at present, they build up

healthy motivation levels that result in good job results. Consequently, they are called satisfiers or motivators. Such considerations include accomplishment, appreciation, progress, the job itself, capacity for personal growth, and transparency.

Numerous motivational theories have influenced how organizations manage employees to create a motivated workforce. It gives perspective and explanation on how people behave in the way they do. Providing advice on the reasons and strategies that can best benefit workers as they work in terms of their dedication to work. Globally, organizations are spending millions of dollars on replacing I.T. experts. I.T. experts who possess technical abilities and competencies that combine business processes are expensive to replace. There are associated direct and indirect costs associated with quitting I.T. staff, such as marketing charges, head-hunting fees, and project delays. Many company rulers do not have approaches to reduce turnover rates, lacking a strong knowledge of the reasons why I.T. experts leave their jobs.

The most popular and widely used technique in machine learning is the decision-making algorithm. After Quinlan implemented ID3, these algorithms and variants have been the subject of numerous research papers. Quinlan addresses a wide range of issues relevant to decision trees, from a critical algorithm for building an original tree to pruning strategies, converting trees to regulation, and resolving multiple issues such as missing attribute values [7]. A huge amount of data is available within the organization to help establish an effective recruitment management plan. Developing a strategic strategy to determine the extent to which turnover affects the business and understanding the factors that affect turnover and establish retention strategies. Data mining can be useful for human resources in identifying the characteristics of their most productive employees, in particular in helping to recognize high turnover potential workers. [8], [9]. Human Resources (H.R.) management systems maintain records of all changes in employment status, including voluntary dismissal, employee history of work-action, length of time in a position, and payment history. The extensive collection of employee information within the company maintained by Human Resources (H.R.) in evaluating successful employee turnover forecasts.

This paper aims to present the Machine Learning Approach of Herzberg's Theory using the C4.5 Algorithm. This study aims to improve human resource management practices and workplace approaches. Promote employee retention, which can assist businesses to decrease their expenses of turnover that disrupts business in multiple ways and effects are adverse. This following are the specific objectives of this research:

1. To create an employee retention model of IT Job-related functions using the C4.5 Algorithm.

2. To measure the robustness and performance of the data sets using the C4.5 Algorithm.

3. To corroborate the Machine Learning approach using the C4.5 result to Herzberg's theory.

2. METHODOLOGY

The combination of Applied and Experimental research utilizing the evaluative approach applies in this study. It allows the researcher to weigh the outcome in distinguishing the Algorithm's performance in achieving the desired output. According to Baimyrzaeva, "Applied research aims to find a solution for an immediate problem facing business entities, whereas fundamental research is mainly concerned with generalizations and with the formulation of theory [10]. Applied research is considered a non-systematic inquiry, and it usually launched by a company, agency, or an individual to address a specific problem".

A. Herzberg Motivation Theory

Behavioral psychologist Frederick Herzberg proposed a two-factor or motivational-hygiene concept in 1959. According to him, job factors could either improve performance or discourage performance. "Satisfaction" is the opposite of "No Satisfaction," while "No Dissatisfaction" is the opposite of "No Dissatisfaction." [11]. It describes the factors that employee engagement find their employment satisfactory and dissatisfying and identified as hygiene factors and motivators. Hygiene factors can lead to discontent in the workplace, but when treated and nurtured can be adequate to satisfy employees that lead to increased motivation[12]. The theory of Herzberg has widely accepted and considered relevant across industries motivating employees to perform well. Countless strategic ideas have affected how companies treat employees to create a motivated workforce. The theory intends to explain how people behave in the way they do and to provide advice on the reasons and strategies that can best benefit workers as they work in terms of their willingness to work.
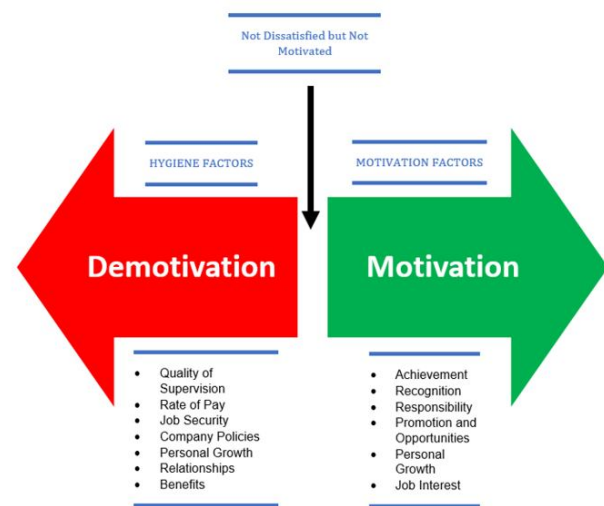

Figure 1: Herzberg Theory of Motivator-Hygiene

Hygiene considerations are alluding to as maintenance considerations that include natural health and relationship needs. These are conditions that are not directly linking to the job but the situation surrounding the work. Motivational factor considerations rigidly adhere to the quality of the job,

are innate to the job as a whole, and do not gain from' carrot and stick opportunities.' It encompasses the physiological need for development and appreciation. The lack of these factors does not prove to be particularly dissatisfactory, but they are currently building up a healthy level of motivation, which leads to excellent job performance. As a result, they are called satisfiers or motivators. Those attributes include achievement recognition, growth, jobs itself, the potential for personal growth, and openness.

B. Employee Retention Model

The best performance to accomplish corporate goals is the I.T. Investments comply with organizational skills and regulatory processes within the context of the business strategy. The weight of the services sector is growing in all economies and is a central component of the gross domestic product in developed countries. The purpose of this contribution is to investigate the potential direct and indirect effects of I.T. on the service sector through the employee retention performance of companies. The I.T.'s calculating difficulties impacts on organizational efficiency and constraints on the variety of products and the rapid growth of their demand[4]. The theoretical and practical studies have already shown the ability of information systems to be introduced and applied to improve the quality of the organization in terms of efficiency, profitability, the competitiveness of the enterprise, and growth. The Information Technology is one of the most critical factors that contribute to technical progress in the development, production, and delivery of all types of services.
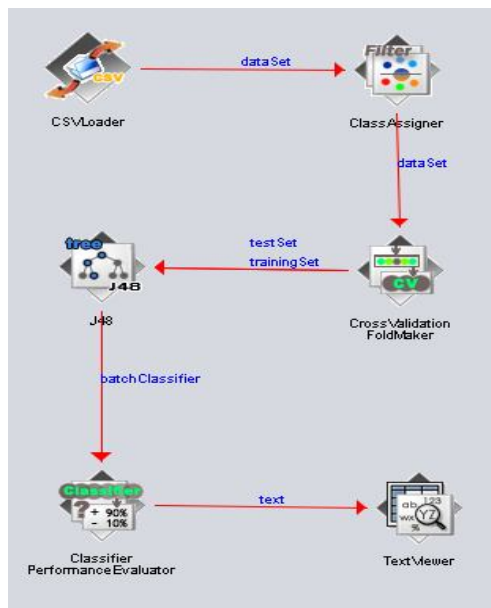

Figure 2: Retention Model using Knowledge Flow

The Government of the Philippines uses the Strategic The Performance Management System (SPMS) aligns employee performance with corporate performance and strengthens the

focus of the compensation system. The case of private companies also similar in terms of establishing the link between organizational performance and Strategic Planning and organizing performance indicators. The assurance and guarantee of an employee achieve the goals set by the organization. Likewise, the organization is meeting the goals set as its strategic plan is feasible through the enablement of efficiency by cascading organizational transparency at the various levels of the enterprise and connecting performance management to other H.R. Unit Systems. The absolute standards are implementing the basis of methods of assessment. The system compares workers to a norm, and the measurement is independent of any other worker in the organization. The comprehensive standard approach is convoying by a variety of techniques, such as the creation of feedback through essay evaluation. The verification of critical incidents that involve the employee, checklist of the vital competencies, a visual rating scale of the different performance indicators, forced ranking policy such as the bell curves, and behavioral rating scales. [13]

C. CRISP-DM

The CRISP-DM approach or Cross-Industry Standard Process is an industry-proven way to direct data mining activities and delivers an overview of the typical phases of the project. The CRISP-DM as a system model is providing an outline of the data mining life cycle. It consists of six steps, with arrows showing the most significant and frequent inter-step dependences. It is not a simple sequence of phases. Many projects move back and forth as needed between steps.
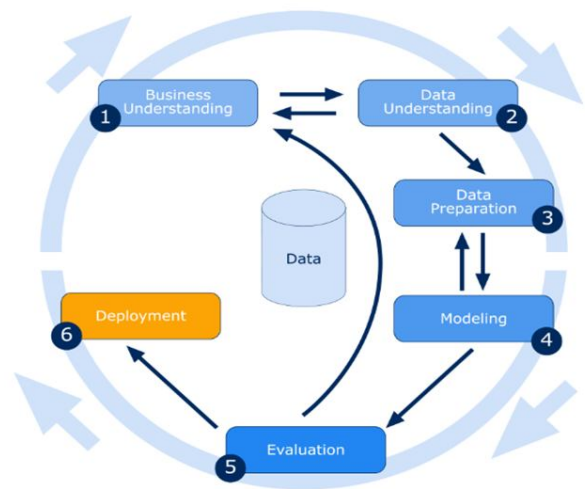

Figure 3: CRISP-DM Model

The Business Understanding addresses the business use case and engagement parameters. A useful calibration framework benefits the organization prominently. The employment perspective in managing the expectations of an employee provides information and tactics in building

employment-related decisions. These include the pay raises, promotions, or releases. The presence of a domain expert can facilitate in identifying the attributes of the use case to be included in the model. The dataset consideration for processing in this study are from the three largest IT Multi-National Companies in the Philippines and performed between the months of December 2018 to August 2019. In compliance with ISO 27001 – "Information Security Management System," BS7799 – "Information Security Management" and Republic Act 10173 – "Data Privacy Act of the Philippines"[14][15]. The researcher can not publicly disclose the company name and data details unless a non-disclosure agreement is serving by the interested party.

The Data Understanding phase is the exploration of the data in hand that involves analyzing dependent and independent variables through the identification of the variations of every independent variable and the combination of two or more variables using statistics. Selecting from the employee pool for this research involves two fundamental processes of identifying an acceptable employee and deciding whom the employee is convincingly and truthfully benefiting from the assumption that he or she will become the most effective and successful[16]. Decision making on the best retention choice is continuous to be carried out as a sequence of selection procedures and consists of attributing selection criteria with predictors that validate candidates ' abilities, talents, experience, expertise, and overall innovative and competent knowledge. Most organizations or firms have a formal calibration guide that grades performance.

The Data Preparation stage sets the activities from the initial raw data in building the final dataset for processing. The data category converts into analytical datasets via a reconsideration of criteria for data selection, determining of which dataset to use, collection of appropriate additional internal or external data, consideration of sample techniques, and explanation of why specific data should include or exclude.

The Modeling phase of this study is a pattern from the Software Engineering best practices "DAR" process. The "DAR" or Decision Analysis and Resolution is a process area (P.A.s) in the Capability Maturity Model Integration framework [17]. The goal of this P.A. is to ensure that vital decisions making is systematical. The DAR process is a formal method for evaluating critical program decisions and proposed program solutions. It provides a decision-making process that is controlled rather than reactionary. Data mining techniques that we use to derive essential and exciting data information have several tasks for this work, such as mining rules for associations, clustering, classification, and data forecasting[16], [18], [19].

The Evaluation stage is the preparation and cleaning phase of the provided data before pre-processing. This step commences immediately after the model creation, testing, and evaluation using WEKA- a product of Waikato University in New Zealand [8], [8], [19]–[21]. The "WEKA" or Knowledge Analysis Waikato Environment is a workbench collection of trailblazing machine learning

algorithms and pre-processing tools[22]. The design is to allow us to test existing methods in flexible ways on new datasets quickly. It provides extensive provision for the complete experimental data mining process. Preparation of input data, statistical analysis of learning schemes, and visualization of input data and learning outcomes are the primary activities pursued at this point. A wide range of pre-processing tools and a variety of learning algorithms. The "Raw" classification to employee's calibration category without the dimensionality reduction technique.

The Deployment stage is gathering recommendations. The knowledge gain will organize and present a way of accumulated recommendations user can process it. The generation of a report will remain for evaluation of calibrating employee retention[13], [23].

D.C4.5 Algorithm Implementation

The augmented data for training given a value of the vector
The algorithm of C4.5 chooses a data

$$C = c_1, c_2, ..., \text{ where } c_1, c_2 \text{ represents the} \tag{1}$$

class of which data the sample belongs.

attribute at each tree node that splits its sample set most effectively into subsets enriched in either class[24]. Its criterion is the standardized gain of information (difference in entropy) that results from the selection of the data split attribute. The attribute with the most meaningful information is selected to create the decision.

The C4.5 Algorithm reappears on the narrower sub-lists and few primary instances in this algorithm. All samples belong to the same class on the list. When this occurs, the choice tree merely generates a leaf node to choose that class. There is no information gain from any of the characteristics. In this case, C4.5 uses the expected value of the class to create a decision node higher up the tree[12], [25].

3. RESULTS AND DISCUSSION

The data mining approach is a methodology used to extract and classify the considered information according to the requirements of the measures and thresholds. The Creation of a target data set that focuses on a subset of variables and data illustrations are is essential to the carried work.

A.Employee Retention Model

Effective and manageable retention policies for employees in any organization are essential and unavoidable. The implementation of strategies includes a strong commitment on the part of management and, in particular, corporate stakeholders that would undoubtedly generate rich productivity dividends [26]. The importance of employee Retention programs cannot be appreciated unless a

calculation of operations cost savings and turnover cost to galvanize as a business issue. The findings of the research showed that jobs that are very complex and with specialized skills need more considerable attention for retention to achieve. Typically, this is the highly paid staff who are at the senior-level, such as Advisor Developer and Senior Actuarial. These positions tend to have disproportionately high turnover costs as a share of salary rises.

Table 1: Attribute Evaluator 24 Retain to Work):
Information Gain Ranking Filter

Ranked		Attributes
0.0458	17	TotalWorkingYears
0.043792	1	Age
0.043629	8	CareerLevel
0.039945	12	OverTime
0.037122	20	YearsAtCompany
0.035941	23	YearsWithCurrManager
0.026495	21	YearsInCurrentRole
0.023023	4	MaritalStatus
0.01836	7	CostPerManday
0.012598	5	Job Satisfaction
0.011659	6	OnshoreOpportunity
0.010702	11	NumCompaniesWorked
0.010097	2	ManagementSatisfaction
0.010064	9	OfficeDistanceFromHome in in KM
0.006906	19	WorkLifeBalance
0.006466	18	TrainingTimesLastYear
0.002717	22	YearsSinceLastPromotion
0.002603	10	MonthlyRate
0.002403	15	RelationshipSatisfaction
0.000631	3	Gender
0	13	AvePercentSalaryIncrease
0	16	StandardHours
0	14	PerformanceRating

The model presented after data processing and analysis that high engagement and retention came from employees who have been given economic growth in terms of monetary and career advancement and as shown in Table 1.

B.Data Performance

Weka is a brand produced at Waikato University in New Zealand. The workbench architecture flexibly evaluates existing methods on new datasets that can be easily integrated into the process using a collection of machine learning algorithms and pre-processing procedures[19], as shown in Table 1. Data output is broad and significant support for the entire experimental data mining process. It involves a wide range of pre-processing methods and several learning algorithms as shown in Table 2.

Table 2: Dataset Performance of Retention

Correctly Classified Instances	5825	99.06%
Incorrectly Classified Instances	55	0.94%
Kappa statistic	0.9649	
Mean absolute error	0.0112	
Root mean squared error	0.0928	
Relative absolute error	4.12%	
Root relative squared error	25.24%	
Total Number of Instances	5880	

A decision tree is less time-consuming and has fast classification speed, which can avoid decision error and deviation of all kinds. It is easy to describe, and it is easy to understand the tree structure of the rules. Most of the human resources personnel have limited knowledge of data mining, but in the way IF-THEN, they can follow the regulations made
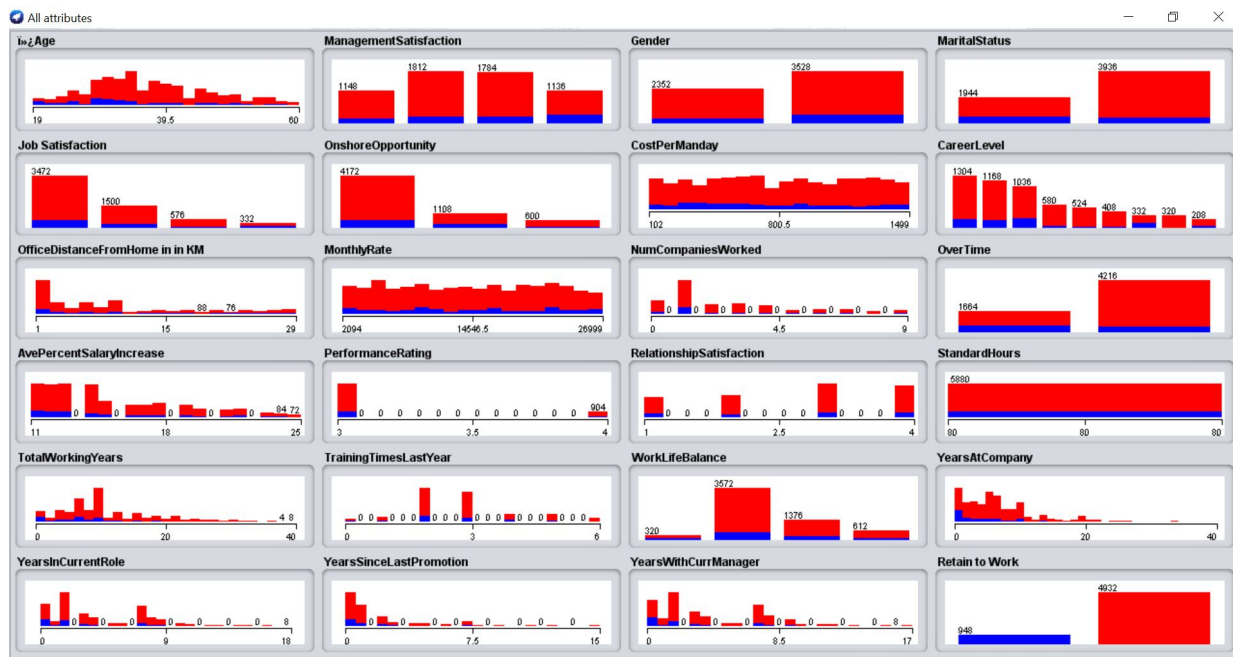

Figure 4: Dataset Visualization of Preprocessed Attributes

by the decision tree [13]. A decision tree algorithm has a good forecast of independent features that can perform well in different data on human resources. A decision tree is to use entropy as the metrics to judge which element will be the tree root and then divide by gain, Gini gain, or Gini ratio. Finding key features that influence the outcome of classification is easy. Furthermore, the importance of characteristics is reflecting in the decision tree layer. If the tree layer is higher, more important is this feature is undeniable [23].

A decision tree is a flow diagram-like tree design. The node in the tree signifies an attribute for testing. Each branch is the test output, and each leaf is a class or class distribution [27]. The top node serves as the root and can be a grouping from the root node to a single leaf. The application of classification rules in the decision tree is straightforward. It has several algorithms, but the main idea is to use an induction approach from top to bottom. Besides, an essential part is to choose which node attribute and determine whether the tree is correct. The decision-making tree consists mainly of two stages, the creation of a decision tree and pruning[27]. The design for a decision-making tree is called CLS. The variable "S" is known as the data set while the attribute array is variable "A," and the decision attribute as D.

The whole process proceeds by having "S" as the root node, transforming the node to the tree node if all the information in "S" belongs to the same group. Designate an attribute subset of "A" and divide the nodes by different attribute values if "S" has a different class. "S" has the number of lower layer nodes, and the branches were a different value case. Induct the previous steps from the nodes of the branch.

The decision trees with linear regression function at the nodes of the leaf have a classification of binary model trees[28]: Consequently, they can serve any rough direct calculation to an unknown purpose. There are two stages in creating a tree template. The first stage is the construction of the ordinary decision tree, which uses the maximization of the target value of the intra-subset variance as a splitting criterion. By replacing sub-trees with linear regression functions, wherever it seems possible, the second prune allows for high-pitched gaps that will inevitably follow between adjacent linear models on the leaves of the prune tree. The original design of the model trees, while linear models are used during the smoothing process at internal nodes, whenever the model is used to simulate a smoothing system that invokes compensation [28].

The decision tree algorithm is a set of no order; no sample data rules set to infer the decision tree representation classification, rules. A decision tree is a tree structure that provides automatic data classification. The use of information theory to analyze and summarize the generated properties to a large number of samples [24]. The purpose of the data mining approach by algorithms is a methodology used to extract and classify the considered information according to the requirements of the measures and thresholds. The algorithm-based data mining method aims to extract and classify the found understanding according to the specifications of standards and limits. Focus on the specific subset of data samples to be detected.

3. CONCLUSION

The C4.5 Algorithm is a useful method for creating an employee retention model for IT work-related functions. The retention model promotes engagement through sustainable and personalized campaigns. The WEKA application software is a tool for knowledge analysis that can facilitate data analysis and the creation of a predictive model for retention. The result identified using the machine learning approach that salary, office distance, training, career growth, and tenure are significant considerations in building a retention program. It corroborates to Herzberg's Theory as a substantial motivator for retention. The findings of the research showed that jobs that are very complex and specialized skills need more considerable attention for conservation to achieve. Typically, this is the highly paid staff who are at the senior-level, such as Advisor Developer and Senior Actuarial. These positions tend to have disproportionately high turnover costs as a share of salary rises. Employee turnover costs are comparatively high for companies, regardless of the value of salaries paid to depart or incoming staff. The model presented after data processing and analysis that high engagement and retention came from employees who have been given economic growth in terms of monetary and career advancement.

4. RECOMMENDATION

Developing a Human Capital Intelligence System using the C4.5 Algorithm and calibrated datasets transforms the manual approach of employee retention to the digital-agile environment. The design implementation of the system covers a workplace engagement module that is allowing collaboration between HR and Project Managers with a goal of continuous learning and an upscaling organizational culture that includes monitoring of Total Working Years, Age, Career Level, Over Time, Years at Company, Years with current Manager, Years in current role, Marital Status, Cost per manday, Job Satisfaction, Onshore Opportunity, Number of companies worked, Management Satisfaction, Office Distance from home, Work-Life Balance, Training Times in a year, Promotion, Monthly Rate, Relationship Satisfaction with co-workers, Gender, Ave Percentage of yearly salary increase, standard working hours and Performance Rating. The system that is accessible to all and will strengthen confidence and develop programs that promote integrity for employees.

REFERENCES

[1]   J. E. Duke II and E. N. Udono, "A New Paradigm in Traditional Human Resource Management Practices," *J. Manag. Sustain.*, vol. 2, no. 2, 2012. https://doi.org/10.5539/jms.v2n2p158

[2]   M. H. Ismail, M. Khater, and M. Zaki, "Digital Business Transformation and Strategy: What Do We Know So Far?," *Manuf. Artic.*, no. November 2017,

p. 36, 2017.

[3] R. Punnoose *et al.*, "Prediction of employee turnover in organizations using Machine Learning algorithms: a case for Extreme Gradient Boosting," *Proc. - 12th IEEE Int. Conf. Data Min. Work. ICDMW, 2012*, vol. 1, no. 3, pp. 344–353, 2016.

[4] B. Liviu, "Information Technology and the Company Performance in the Sector of Services," *Constantin Brancusi*, pp. 127–133, 2015.

[5] M. H. Joarder and M. A. Ashraf, "Work Satisfaction and Employee Intentions: An Empirical Study," *East-West J. Bus. Soc. Stud.*, vol. 3, pp. 19–35, 2012.

[6] G. Balamurugan, "a Study on Employee Retention Strategies in Leading," *Int. J. Hum. Resour. Rev.*, vol. 4, no. 3, pp. 42–48, 2016.

[7] Z. Mok, "Book Review Language Disorders in Speakers of Chinese (Series: Communication Disorders Across Languages). Edited by Sam-Po Law, Brendan Stuart Weekes, and Anita Mei-Yin Wong, 2008 (Bristol : Multilingual Matters) [p. 320]. Price £69.95/$139.95. ISB," *Clin. Linguist. Phon.*, vol. 25, no. 1, pp. 80–83, 2011.

[8] D. Alao and A. B. Adeyemo, "Analyzing employee attrition using decision tree algorithms," *Comput. Inf. Syst. Dev. Informatics Allied Res. J.* vol. 4, no. 1, pp. 17–28, 2013.

[9] R. Punnoose and C. Xlri -Xavier, "Prediction of employee turnover in organizations using Machine Learning algorithms: a case for Extreme Gradient Boosting," *Int. J. Adv. Res. Artif. Intell.*, vol. 5, no. 9, pp. 22–26, 2016. https://doi.org/10.14569/IJARAI.2016.050904

[10] M. Baimyrzaeva, "Institute of Public Policy and Administration Beginners' Guide for Applied Research Process: What Is It, and Why and How to Do It?" 2018.

[11] Lawrence Roy Arnold, "Strategies for Reducing High Turnover Among Information Technology Professionals," *ProQuest Diss. Theses*, no. December, pp. 212-n/a, 2016.

[12] J. Rabcan, M. Vaclavkova, and R. Blasko, "Selection of appropriate candidates for a type position using C4.5 decision tree," *Proc. Int. Conf. Inf. Digit. Technol. IDT, 2017*, pp. 332–338, 2017. https://doi.org/10.1109/DT.2017.8024318

[13] M. Jafari, A. Bourouni, and R. H. Amiri, "A New Framework for Selection of the Best Performance Appraisal Method," *Eur. J. Soc. Sci.*, vol. 7, no. 3, 2009.

[14] N. P. COMMISSION, "IMPLEMENTING RULES AND REGULATIONS OF REPUBLIC ACT NO. 10173, KNOWN AS THE 'DATA PRIVACY ACT OF 2012,'" 2012. [Online]. Available: https://www.privacy.gov.ph/wp-content/uploads/IRR-of-the-DPA.pdf.

[15] I. O. for Standardization, "ISO/IEC 27001:2013 Information technology — Security techniques — Information security management systems Requirements," 2008. [Online]. Available: https://www.iso.org/standard/54534.html.

[16] J. M. Kirimi and C. A. Moturi, "Application of Data Mining Classification in Employee Performance Prediction," *Int. J. Comput. Appl.*, vol. 146, no. 7, pp. 28–35, 2016. https://doi.org/10.5120/ijca2016910883

[17] The Process Group, "DECISION ANALYSIS AND RESOLUTION PROCESS," *October*, vol. 12, no. 2, 2005.

[18] H. Jantan, A. R. Hamdan, and Z. A. Othman, "Talent knowledge acquisition using data mining classification techniques," *Conf. Data Min. Optim.*, no. June, pp. 32–37, 2011. https://doi.org/10.1109/DMO.2011.5976501

[19] S. Kalmegh, "Analysis of WEKA Data Mining Algorithm REPTree, Simple Cart and RandomTree for Classification of Indian News," *IJISET -International J. Innov. Sci. Eng. Technol.*, vol. 2, no. 2, pp. 438–446, 2015.

[20] N. Bhargava, S. Dayma, A. Kumar, and P. Singh, "An approach for classification using simple CART algorithm in WEKA," *Proc. 2017 11th Int. Conf. Intell. Syst. Control. ISCO, 2017*, pp. 212–216, 2017. https://doi.org/10.1109/ISCO.2017.7855983

[21] M. Nasr, E. Shaaban, and A. Samir, "A proposed Model for Predicting Employees' Performance Using Data Mining Techniques: Egyptian Case Study," no. February 2019.

[22] G. S. Thakur, A. Gupta, and S. Gupta, "Data Mining for Prediction of Human Performance Capability in the Software Industry," *Int. J. Data Min. Knowl. Manag. Process*, vol. 5, no. 2, pp. 53–64, 2015. https://doi.org/10.5121/ijdkp.2015.5205

[23] N. Magesh, P. Thangaraj, S. Sivagobika, S. Praba, and R. M. Priya, "Evaluating The Performance Of An Employee Using Decision Tree Algorithm," *Int. J. Eng. Res. Technol.*, vol. 2, no. 4, pp. 2814–2830, 2013.

[24] X. Long and Y. Wu, "Application of decision tree in student achievement evaluation," *Proc. - 2012 Int. Conf. Comput. Sci. Electron. Eng. ICCSEE, 2012*, vol. 2, pp. 243–247, 2012. https://doi.org/10.1109/ICCSEE.2012.169

[25] W. Jia and L. J. Huang, "Improved C4.5 decision tree," *Int. Conf. Internet Technol. Appl. ITAP 2010 - Proc.*, vol. 3, no. 3, pp. 341–345, 2010.

[26] A. H. Lewis, Andrea & Sequeira, "Effectiveness of Employee Retention Strategies in Industry," no. July 2018.

[27] G. Ying and V. Valkovsky, "Using Decision Tree to Analyze the Turnover of Employees," no. May 2017.

[28] E. Frank, Y. Wang, S. Inglis, G. Holmes, and I. H. Witten, "Using model trees for classification," *Mach. Learn.*, vol. 32, no. 1, pp. 63–76, 1998.