

A Model Based Prediction of Desirable Applicants through Employee's Perception of Retention and Performance

Eduardo B. Santiago Jr.
Information Technology Education Program
University of the Immaculate Conception
Davao City, Philippines
esantiago_0202058@uic.edu.ph

Glenn Paul P. Gara
Information Technology Education Program
University of the Immaculate Conception
Davao City, Philippines
ggara@uic.edu.ph

Abstract—Selecting desirable applicants for an organization who has less tendency to quit their jobs are one of the major challenges in human resource management. Proper selection during pre-employment has a significant impact in an organization's productivity and performance. However, due to a prevalence of labor market competition, some companies are suffering to high turnover rate. It occurs when an applicant decides to quit and transfers to an organization which offers attractive benefits and compensation. This study helps human resource personnel to understand psychological climate and supports decisions in resolving employment turnover by selecting desirable applicants who has high probability of staying longer in an organization. Using the 12 dimensions of retention applied in generating association rules and naive bayes classifier, the researchers develop a custom application that can support organization's decision in relation to hiring process. Based on the results of 12 dimensions of retention, the system evaluated seven dimensions with poor psychological climate response resulting to higher possibility of employees to voluntarily quit his or her job. Moreover, upon generating the model for job position counter cashier, results show that age older than 20 years old and living away from workplace has higher probability of staying long in comparison to other applicants. This study employs PHP-ML Library in generating association rules and classification to analyze the profile of past and current employees in the company. The results were validated using the RapidMiner software to ensure the accuracy of the implementation of association rule mining and naive bayes classifier.

Index Terms—applicant prediction, applicant selection, psychological climate, association rule mining, naive bayes classification

I. INTRODUCTION

Human Resource Management plays a crucial role on organizational performance. They handle several matters pertaining to the employees in a company. Despite of having competent resource personnel, selecting best fit candidates with top talent remains a primary concern for many organizations today. Critical analysis of workforce trends point to an impending shortage of highly-skilled employees possessing the requisite knowledge and ability to perform at high levels. It only means that organizations failing to select and retain high performers

will be left with an understaffed, less qualified workforce that ultimately hinders their ability to remain competitive [1]. In a national setting an important aspect of strategically managing the HR function is to assess its effectiveness. Supangco in 2012 stated that there was a significant increase of annual staff turnover rate from 8.3% in 2003 to 16.0% in 2008. The doubling of turnover rate can be a cause for concern from HR effectiveness perspective, especially when the said rate is of the voluntary type, which means the employee itself decided to quit [2].

Studies conducted for challenges and trends in personnel selection [3] shows that labor market shortages, technological developments, applicant perceptions of selecting procedures, construct driven approaches are four major concerns. First, a survey conducted resulted with shortage of talent in workplace causing struggle for organization to find talented workers, this creates a dilemma on how to retain knowledgeable workers to replace nearly retiring employees to fill in positions and also to avoid high turnover, failure to do so will result to negative impact in organization which causes direct cost (e.g. recruitment, induction and training of employees) and indirect cost (e.g. learning, reduced morale, pressures and loss of social capital) [4], [5]. Second, technological developments which aim primarily to incorporate information technology to collect, transmit and process data in an organization for information. Several existing studies using computerization to conduct cognitive ability test [6], internet-based job recruitment and selection [7] shares the same intention to help resolve the issues of selecting best candidates and ease the hiring process in relation to job. Third, applicant perception refers to a link between applicant and organization in which both parties exchanges information with each other prior to deciding whether or not to pursue a relationship [8]. Studies suggested that psychological, career attitudes, self-esteem, self-efficacy has an impact also in harmonization's employee turnover in which researchers have begun studying selection through socio-psychological perception. An open-ended question conducted by Judge and Ferris stating "What is it you are looking for in a high quality candidate?". Interestingly, after several requests

for clarification from these decision makers, they found an amazing degree of convergence on the responses provided. Inevitably, it comes down to a statement of 'fit' - that is, they suggest that they are looking for a candidate who fits. Sometimes they elaborate to specify someone who fits the culture, the value of system, and so forth, but frequently it is merely stated as someone who fits. When pressed as to more precisely what it means to fit, many are initially hard pressed to provide a response [9]. Lastly, construct-driven approach refers to predicting how efficient an applicant would be in the future work performance.

The trends show a common link pointing out directly towards applicant selection correlated to matching competent employee which was seen as a challenge in human resource management. Clearly it shows that selecting best applicants and to predict if they will retain are at its infancy stage and needs further improvement.

The objectives of the study is to generate a knowledge in understanding employee perception, a knowledge to support organization's administration in planning strategically for employee management and a model to support organization's human resource to determine applicant's fitness.

II. RELATED WORK

Dimensions of retention by Hausknecht, Rodda and Howard consists of Advancement of Opportunities, Constituent Attachments, Extrinsic Rewards, Flexible work arrangements, Investments, Job Satisfaction, Lack of alternatives, Location, Non-work influences, Organizational commitment, Organizational justice and Organizational prestige. These dimensions demonstrates a significance in understanding psychological climate in relation to retention.

Association rule mining is long been used for discovering frequent patterns particularly the apriori algorithm. It has been used in several studies involving employees to improve personnel selection. In relation to this study, apriori was used to mine employees profile to discover a set of rules used to filter out similar employee profiles [10].

Similar research done by [11] uses association rules in extracting frequent patterns of personality types for suitable course of a student. Nave Bayes serves as a means of predicting best fit applicant profile with high probability of staying longer in an organization. The relative simplicity of the algorithm and the independent features assumption of Naive Bayes make it a strong performer for classifying texts and usually takes orders of magnitude less time to train when compared to models like support vector machines [12].

A research closely related to this study uses Person Organization Fit (P-O Fit) and Person Job Fit (P-J) approach to study work outcome and fitness based on gender for newly hired employees [13] with objectives to develop a model of person-organization and person-job fit that accounts for gender differences, validate the model through empirical studies among entry-level IT workers and examine the model created.

The study emphasize trends and challenges in personnel selection with an effort to apply a model that will project applicant fitness through performance and retention. Using 12 dimensions of retention [14] the researcher put emphasis in job satisfaction in relation with turnover intention and person organization fit [15]. Job Satisfaction result serves as a threshold value for association rule. It identifies strong rules based on some measures of interestingness or support. Profiles of employees that matches the rules were used for classifying new applicant profile to predict probability of applicant's number of years will stay in an organization.

III. METHODS

A survey was conducted in a small group of employees within a small company with at least two years and above working experience as per recommendation of HR Management to ensure better psychological climate response. Template containing fields of age, sex, religion, marital status, educational attainment, location by region, province, city and barangay with limits to employee location residing in Region 10, 11 and 12 were selected to generate dimension results.

Human resource department designates a personnel to conduct an opinion survey which was done through an online survey system. Survey coordinators instructed each participant to fill in basic profile followed by proper administering of survey questions. It was administered in a small room with one computer in order for the employee to think well without distractions. Collected data from employees perception were classified according to its corresponding dimension and percentage values. The values were used to determine support threshold for association rule mining while the confidence threshold is user input dependent.

Mining rule extracts frequent employee profile fields which determines employees with matching profile. Matching employee profile were used as training data for Naive Bayes classification and will be utilized to predict applicant's probability of number of years based on the job position application. Classification parameters were set using the default parameters provided by PHP-ML Library.

Human resource personnel can evaluate an applicant by comparing the results in a form of chart helping them to decide whether an applicant fits better in the organization or not.

IV. ARCHITECTURAL DESIGN

The system is intended for Employee, HR Management and Top Management to provide support in decision-making on improving psychological climate. The applicant model allows the recruitment team to help decide in selecting desirable applicant with a high probability of staying longer as illustrated in figure 1.

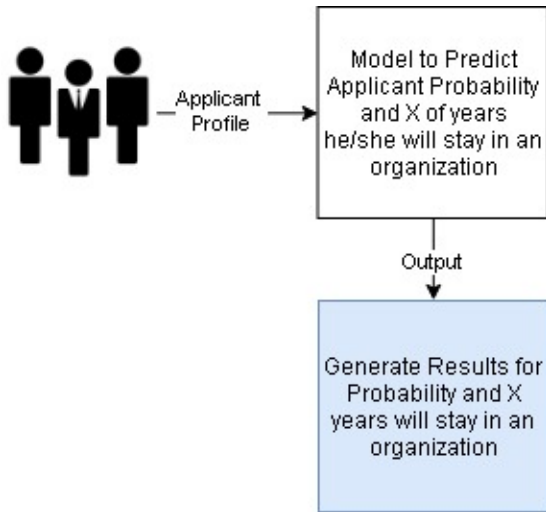


Fig. 1. Conceptual Design

Figure 2 shows overall architectural design of the study while figure 3 illustrates the detailed processes. Inevitably, employees will take part of psychological climate survey. A dashboard for HR Managers was designed to easily analyze the psychological climate results. Generated dimension results and frequent patterns allows both HR and Top Management to seek knowledge and support the decision-making process particularly in identifying areas of dimensions that needs immediate attention (negative employee perception). The HR can initialize values for support, confidence and discretization to mine association rules. A prediction process using Naive Bayes method computes probability of applicant's retention.

In figure 3, a detailed illustration shows how the application processed the generation of dimension, association rules and applicant prediction.

A. Dimension Results

In generating dimension results, employees are required to take the online survey and all responses are consolidated. The result helps human resource and top management to assess and improve dimensions with poor survey ratings.

B. Association Rules

In preparation for data mining and pattern analysis, HR Management must initialize support and confidence value in order to generate association rules. The system extracts and filter employee profile based on the rules generated. Filtered employee profiles were used to train classification.

C. Naive Bayes Classification

This probabilistic classifier uses training data from filtered employee profile and paired with input data from applicant profile in order to compute and predict applicant's probability of N years of retention in an organization.

V. SYSTEM OVERVIEW

The system is a web application and has two separate main page namely the employee and administration pages. Employee page is dedicated for survey while the administration page is for HR Personnel giving them full control to manage overall functions of the system.

A. Employee Page

The page contains all the necessary functions allowing an employee to participate in psychological climate survey. Employees must provide personal information details specifically the employee ID number, age, sex, religion, marital status, educational attainment, religion, province, city and barangay. Once done, the application will provide a set of survey question for the employee to answer.

1) *Employee Survey*: The survey is divided into 12 dimensions of retention where each is mapped to a set of questions that requires the employee to rate. Questions can be rated using a five point scale with a corresponding descriptive value Strongly Disagree, Neutral, Somewhat Agree, Strongly Agree and N/A (Not Applicable). They may also opt to include reasons to express their personal thoughts and opinions for each survey question.

B. Administration Page

The page contains functions that allow HR Management users to setup, process and control the system. It contains three modules dedicated for employee, applicant and the survey.

1) *Employee Module*: The module is used to generate dimension results from psychological climate survey. As illustrated in figure 3, it requires steps in consolidating and percentage rating.

User must create a template and should provide information in the required fields. The template will filter survey responses that matches the criterion and consolidates the survey when executed. The generated dimension results are used to set support values for association rule mining of employees profile. Job satisfaction value is important due to its relevance to PO (Person-Organization) fit and employee retention [16] which can serve as a basis in setting support values for association rules.

2) *Applicant Module*: The module was designed to mine and generate frequent patterns using association rules, probabilistic classification and chart generation for applicant comparison.

As seen in figure 3, it requires steps in generating itemsets and frequent itemsets by setting the support and confidence thresholds, and a discretization value. The support threshold indicates how frequently the itemset appears in a dataset and uses the equation 1.

$$supp(X) = |\{t \in T; X \subseteq t\}|/|T| \quad (1)$$

The equation states that the support of X with respect to T is defined as the proportion of transactions t in the dataset which contains the itemset X .

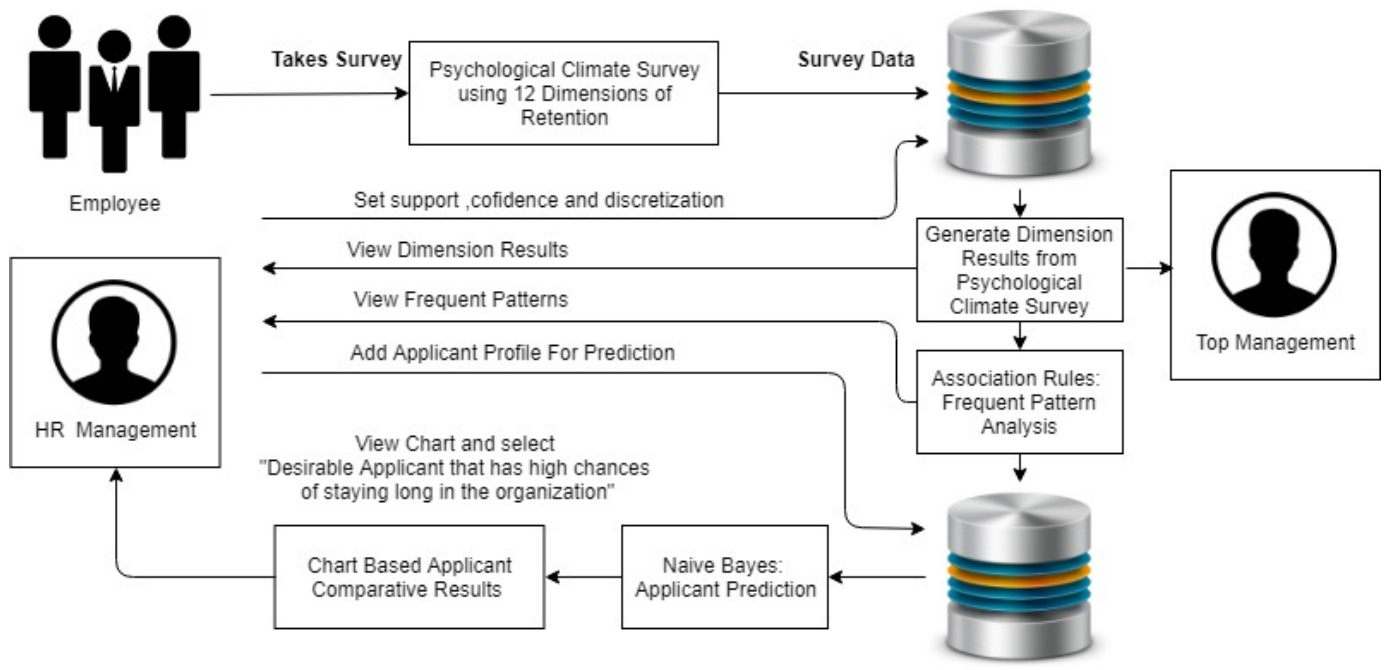


Fig. 2. System Architecture

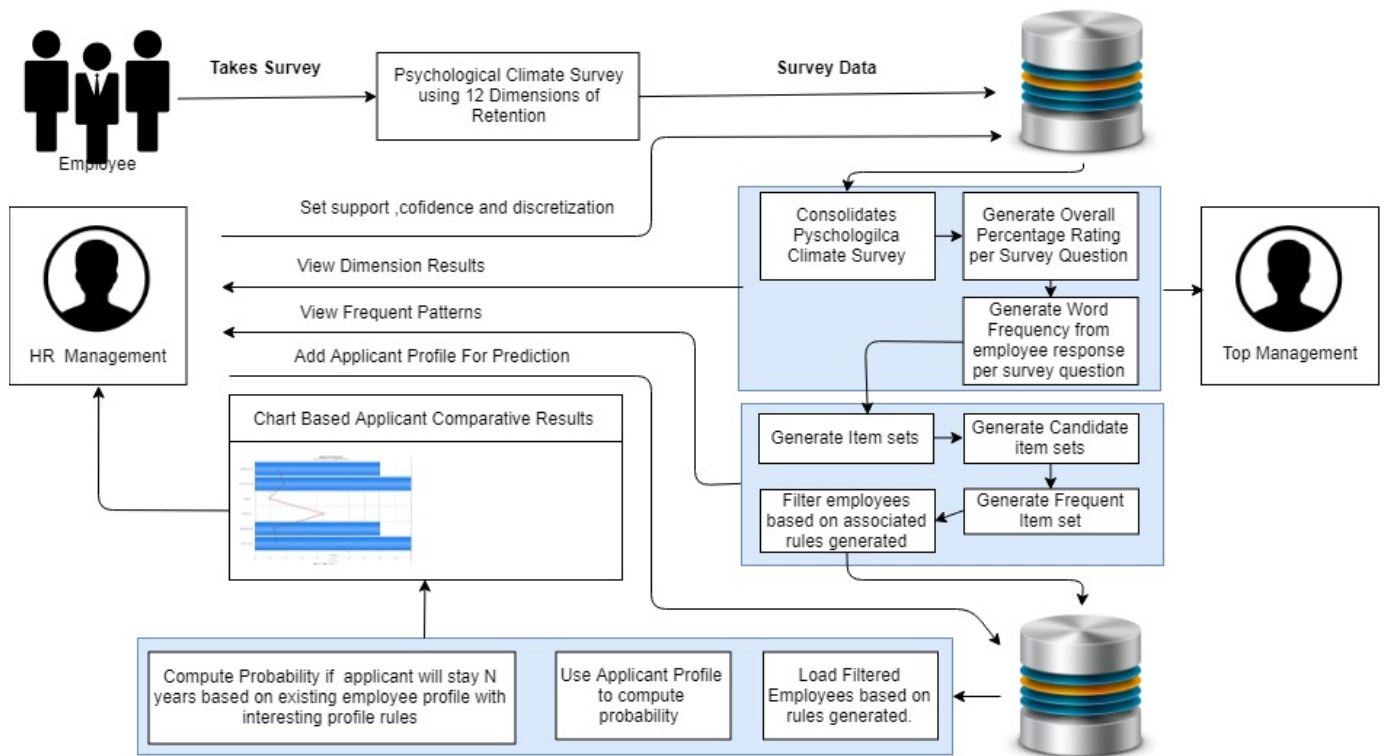


Fig. 3. Association Rules, Classifier and Dimension Result

The confidence is an indication as to how often the rule has been found to be true and uses the equation 2.

$$conf(XY) = \frac{supp(X \cup Y)}{supp(X)} \quad (2)$$

The confidence value of a rule, $X \cup Y$, with respect to a set of transactions T is the proportion of the transactions that contains X which also contains Y .

Apriori algorithm mines frequent pattern of employees' profile stored in a database using the support and confidence thresholds to identify the most important relationship. The application filters employee with similar rules that matches their profile per job position.

Probabilistic classification predicts the applicant's retention and its associated probability value. This study implements Nave Bayes Algorithm [16] due to its relative simplicity and independent features assumption and usually takes orders of magnitude less time to train when compared to models like support vector machines as stated in the related work section. The classifier needs a training data from filtered list of employees' profile in order to predict the applicant's retention.

3) *Survey Module*: The module is used to manage survey questionnaire for employee psychological climate survey. Only authorized users can create and update questions per dimension.

VI. RESULTS

A Summary of 12 dimension of retention shown in Figure 4, has an overall retention of 46.82%, leading to 50% more chances of employee turnover. Results shows that 5 out of 12 dimensions; Constituent Attachments, Organizational Prestige, Lack of Alternatives, Investments and Location receives 50% more positive employee response while 7 out of 12 dimensions; Advancement of Opportunities, Extrinsic Rewards, Flexible Work Arrangement, Job Satisfaction, Non-work Influences, Organizational Commitment, Organizational Justice received negative ideal response that needs to be addressed to increase overall retention.

In generating association rules, support values were set to 42% referred from dimension result particularly job satisfaction with confidence value of 100% including sex, civil status, city, province, citizenship, position, department, work history, educational attainment and age as part of items to be used in data mining shown in figure 5. The system mined set of rules per employee position, frequent items including list of filtered employees by rules and verified our results using Rapidminer Studio Software and it turned out to have similar results.

Dimensions	[+]Response %	Dimension %	Response %...
Job Satisfaction	42.86	24.52	10.51
Extrinsic Rewards	40.00	19.71	7.88
Constituent Attachments	60.00	16.35	9.81
Organizational Commitment	28.57	8.17	2.34
Organizational Prestige	50.00	6.25	3.13
Lack of Alternatives	100.00	4.81	4.81
Investments	50.00	4.33	2.16
Location	100.00	3.85	3.85
Advancement Opportunities	37.50	3.85	1.44
Flexible Work Arrangements	0.00	3.37	0.00
Organizational Justice	12.50	3.37	0.42
Non-work Influences	33.33	1.44	0.48
Total 100.00			Retention:4...

DETAIL	DESCRIPTION
✓	:Dimension used for support reference
[+]Response %	:Percentage of Positive Response (e.g 3 out of 7 questions per dimension)
Dimension %	:Percentage ratio in 12 dimensions
Response % Per Dimension	:Percentage of Dimension by Positive Response

Fig. 4. Dimension Results

Support	Confidence	Status
50	100	Inactive
42	100	Active

Sex	Yes
Civil Status	Yes
City	Yes
Province	Yes
Citizenship	Yes
Position	Yes
Department	Yes
Work History	Yes
Educational Attainment	Yes
Age	Yes
Dimension 1	No
Dimension 2	No
Dimension 3	No
Dimension 4	No
Dimension 5	No
Dimension 6	No
Dimension 7	No
Dimension 8	No
Dimension 10	No
Dimension 11	No
Dimension 12	No
Discretize	1

Fig. 5. Association Rules Settings

The counter cashier was selected as job position to predict applicant profiles that has high probability of staying long in an organization. The system uses frequent employee profile to train naive bayes classifier and uses applicant profile to predict the probability of work years. The comparative chart shows that applicants residing farther from workplace has better chances in staying long with the organization compared to applicants who live nearby. Hence, it helps HR Management to decide and select applicant in a smart way. To validate our results we use RapidMiner to generate classification results using naive bayes.

VII. CONCLUSION

The Applicant model for desirable applicants provides inexpensive, mobile and effective way of understanding employees. Based on results, applicant model is proportional to overall retention factor. This also suggests that applicant requirement is dependent to organizational status. The higher the percentage of overall retention in dimension results, the higher chances employees will stay.

This study covered and investigated a retail organization for psychological climate survey only. An additional insight can be provided by investigating other organization with large group of employees as well as other economic sectors to further improve the study.

REFERENCES

- [1] U. Kingdom and E. Brako, "Insights Into Performance - Based Rewards .," vol. II, no. 11, pp. 1–7, 2014.
- [2] V. T. Supangco, "Strategic HR Practices in Some Organizations in the Philippines," Tech. Rep., 2012.
- [3] F. Lievens, K. V. Dam, and N. Anderson, "Recent trends and challenges in personnel selection," 2002.
- [4] V. Oladapo, "The Impact of Talent Management on Retention," *Journal of Business Studies Quarterly*, vol. 5, no. 3, pp. 20–36, 2014.
- [5] N. Ali, "Factors affecting overall job satisfaction and turnover intention," *Journal of Managerial Sciences*, vol. 2, no. 2, pp. 239–252, 2010.
- [6] A. K. Mayer and G. Krampen, "Equivalence of computerized versus paper and pencil testing of information literacy under controlled versus uncontrolled conditions : An experimental study," 2015.
- [7] P. Cappelli, "On-Line Recruiting."
- [8] "Recruiting Employees: Individual and Organizational Perspectives," 1998. [Online]. Available: <http://sk.sagepub.com/books/recruiting-employees>
- [9] T. A. Judge and G. R. Ferris, "The Elusive Criterion of Fit in Employment Interview Decisions," *Center for Advanced Human Resource Studies*, pp. 91–26, 1991.
- [10] B. Kumar and K. Rukmani, "Implementation of Web Usage Mining Using APRIORI and FP Growth Algorithms," *Int. J. of Advanced networking and Applications*, vol. 404, no. 06, pp. 400–404, 2010. [Online]. Available: <http://ijana.in/papers/6.11.pdf>
- [11] G. P. P. Gara and F. R. F. Padoa, "Mining Association Rules on Students Profiles and Personality Types," *International MultiConference of Engineers and Computer Scientists*, vol. I, pp. 1–6, 2015.
- [12] I. Rish, "An Empirical Study of the Naïve Bayes Classifier," no. October, 2016.
- [13] V. Venkatesh Distinguished Professor, B. Billingsley Chair, J. B. Windeler Assistant Professor, K. M. Bartol Robert H Smith Professor, R. H. Smith, and I. O. Williamson, "PERSON-ORGANIZATION AND PERSON-JOB FIT PERCEPTIONS OF NEW EMPLOYEES: WORK OUTCOMES AND GENDER DIFFERENCES of Leadership and Innovation Co-Director, Center for Leadership, Innovation and Change 4538 Van Munching Hall," *Academy of Management Journal Journal of Applied Psychology, the Academy of Management Review Personnel Psychology Journal of Personality and Social Psychology, and MIS Quarterly*.
- [14] J. P. Hausknecht, J. Rodda, and M. J. Howard, "Targeted Employee Retention: Performance Based and Job-Related Differences in Reported Reason for Staying," *Human Resource Management*, vol. 48, p. 269 288, 2009. [Online]. Available: <http://digitalcommons.ilr.cornell.edu/cahrswp>
- [15] M. H. Jin, B. McDonald, and J. Park, "PersonOrganization Fit and Turnover Intention: Exploring the Mediating Role of Employee Followership and Job Satisfaction Through Conservation of Resources Theory," *Review of Public Personnel Administration*, vol. 38, no. 2, pp. 167–192, 2018.
- [16] M. Hassan, A. Akram, and S. Naz, "The Relationship between Person Organization Fit, Person-Job-Fit and Turnover Intention in Banking Sector of Pakistan: The Mediating Role of Psychological Climate," *International Journal of Human Resource Studies*, vol. 2, no. 3, p. 172, 2012. [Online]. Available: <http://www.macrothink.org/journal/index.php/ijhrs/article/view/2286>