**A STUDY ON EMPLOYEE ATTRITION AND RETENTION USING MACHINE LEARNING**



**INSTRUCTOR: SUBMITTED BY:**

Dr. RUCHIKA MALHOTRA KISHAN KASHYAP

(2K16/CO/162)

**Table of Contents**

Introduction 3

[Research Problem Statement](#_Data_Overview:)  4

Review of Literature 5

Experiment Design 6

1. Obtaining the dataset 6
2. Scrubbing the dataset 7
3. Exploring the dataset 7
4. Modelling the dataset 11

[Data](#_Dimension/Variable_Reduction) Interpretation 13

1. Results 13
2. Conclusions 13

[References](#_Data_sampling_(partition) 14

**INTRODUCTION**

Attrition or employee turnover is a natural reduction in the workforce of an organization for various reasons like retirement, resignation, sickness, death or for reasons that cannot be disclosed. Many of the undisclosed factors which can be perceived are, anticipating higher pay, lack of job security, lack of career advancement, non-compatibility with fellow employees, desire for change in search of better opportunities and family reasons.

A company never like to lose their loyal employees. In a way, if we look at it, it’s the employees that go a long way in making the company a huge success and in turn they expect that the company should reward them based on their work and behavior. Not in every possible scenario, this could happen, because it’s not only the employees that make a firm, rather the cultural values, leadership team, and several other factors go into the picture which made a firm into a success.

But there are certain scenarios, where an employee is not satisfied with the firm and it results in frustration and eventually leaving the firm. We have pulled up a data set from Kaggle to better understand the underlying reasons so as to why employees leave the firm and have tried to give recommendations based on our analysis.

An employee leaving a firm leaves a great impact on the firm and it’s associated business value:

* Brand of the firm: When multiple employees leave the firm, it’s the brand value of the firm which gets degraded as in the market, the talks of an employee leaving the firm to create abuzz and hence hampering the business value
* Reduction in Quality Manpower: When an employee leaves a firm, a good employee goes, because he or she is getting paid better in the outside market for the work that he/she is doing in your company, hence it impacts the firm/department’s deliverables as well.
* Impact on Revenue: A firm invests a lot of money in hiring and training the new talent and when that employee leaves the firm, the loss is beared by the company only. Consider this situation for many employees and imagine the amount lost in it.

**RESEARCH PROBLEM STATEMENT**

**ABSTRACT**

Employee turnover has always been a matter of concern for organizations. A large degree of employee turnover is highly detrimental to both the organization as well as the employees. How to reduce employee’s turnover intention is a very pivotal challenge for today’s HR managers.

Companies that maintain a healthy organization and culture are always a good sign of future prosperity. Recognizing and understanding what factors that were associated with employee turnover will allow companies and individuals to limit this from happening and may even increase employee productivity and growth. These predictive insights give managers the opportunity to take corrective steps to build and preserve their successful business.

High attrition rates result in escalating recruitment and training costs and lot of time involved in new employee adjustment to the work environment and thereby enhance their morale. This study can be helpful in knowing, why the employees prefer to change their job and which factors make employee dissatisfy using different machine learning techniques. Since the study is critical issue, it is needed by the originations in order to assess the overall interest and the feelings of the employees towards their Retention. This study can serve as a basis for measuring the organization’s overall performance in terms of employee satisfaction.

KEYWORDS: Attrition, HR Strategies, Congenial Work Climate, Retention, Reasons, Machine Learning.

**OBJECTIVE OF THE STUDY**

1. To predict whether an employee will leave the organization or not and how accurately can different machine learning algorithms predict it.

2. To obtain insights on the underlying features and how these features will influence employees’ attrition.

3. To discuss potential solutions that can help in employee retention.

**REVIEW OF LITERATURE**

The decision of leaving the Organization is not easy for an individual employee as well as significant energy is spent on finding new jobs, adjusting to new situations, giving up known routines and interpersonal connection and is so stressful (R1.Boswell, Boudreau and Tichy, 2005). Therefore if timely and proper measures are taken by the Organizations, some of the voluntary turnover in the Organization can be prevented.

The factors that influence employee turnover in Organizations include the individual work variables like demographic variables, integrative variables like job satisfaction, pay, promotion and working condition (R2.Pettman, 1975; R3.Mobley 1982;) and the individual nonworking variables such as family related variables (R2.Pettman, 1975; R3.Mobley, 1982;).

Any of the above factors could be the reasons, but the decision process to leave or stay in the Organization is to be periodically examined to understand the specific reasons that prompted them to take such a step and the Organizations should be mainly concerned about voluntary turnover and not involuntary turnover as it is within their control. Also it is found that employees who perform better and are intelligent enough have more external employment opportunities available compared to average or poor performance employees and thus they are more likely to leave (R4.Trevor, 2001). High rates of voluntary turnover of such employees are often found to be harmful or disruptive to firm’s performance (R5.Glebbeck & Bax, 2004). Organizations failing to retain high performers will be left with an understaffed, less qualified workforce that ultimately hinders their ability to remain competitive and tends to lose their existence (R6.Rappaport, Bancroft, & Okum, 2003).

(References at the end.)

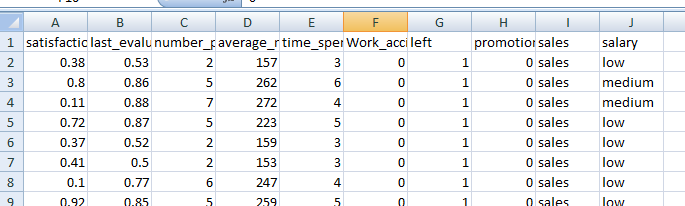
**EXPERIMENT DESIGN**

1. **Obtaining** the dataset.
2. **Scrubbing** or cleaning the dataset if required.
3. **Exploring** the data: It will follow right after and allow further insight of what our dataset contains. Looking for any outliers or weird data. Understanding the relationship each explanatory variable has with the response variable resides here and we can do this with a correlation matrix.
4. **Modeling** the data: It will give us our predictive power on whether an employee will leave.
5. **Interpreting** the data: It states the results, conclusions/recommendations made involving factors contributed most to employee turnover.

**1. OBTAINING THE DATASET**

Dataset used: HR\_comma\_sep.csv available at <https://www.kaggle.com/>.

Target-Variable: left (Turnover).

FIGURE 1: 

Statistical Overview of the dataset used:

1. About 15,000 employee observations and 10 features.
2. The company had a turnover rate of about 24%.

Dataset features names (renamed for better readability):

1. Satisfaction\_level to satisfactoin
2. last\_evaluation to evaluation
3. number\_project to projectCount
4. average\_montly\_hours to averageMonthlyHours
5. time\_spend\_company to yearsAtCompany
6. Work\_accident to workAccident
7. promotion\_last\_5years to promotion
8. sales to department
9. left to turnover

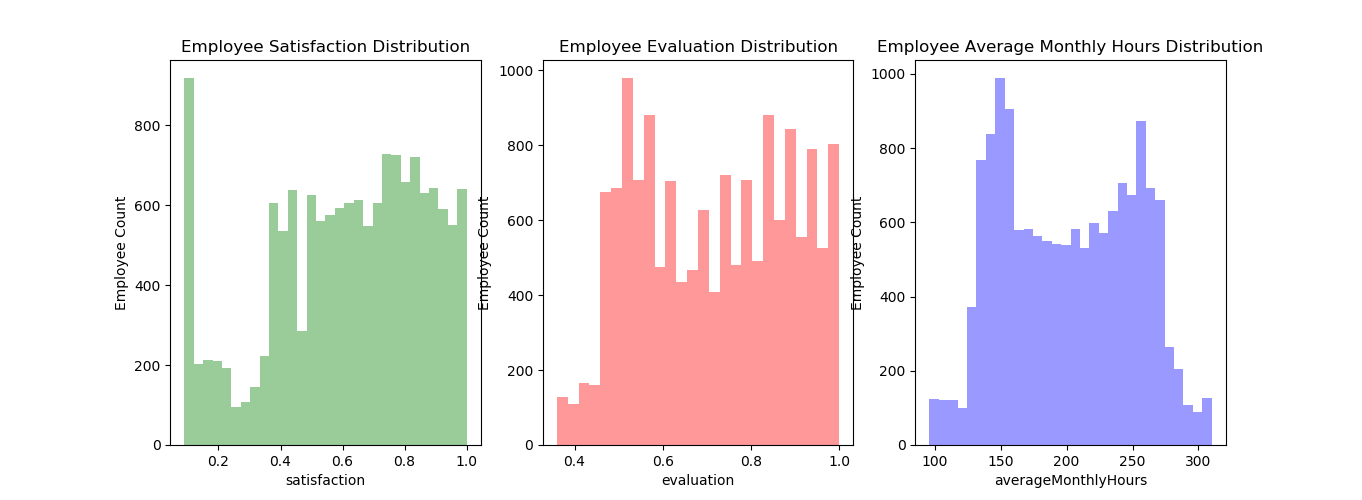
10. Salary

**2. SCRUBBING THE DATASET**

There were no missing values (NAs) in the dataset, hence no data-cleansing required.

**3. EXPLORING THE DATASET**

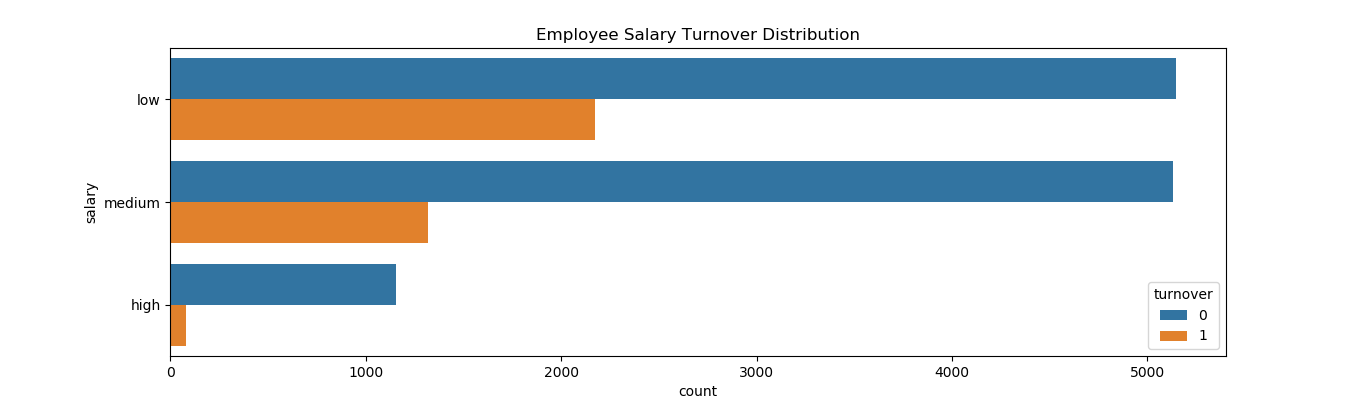
FIGURE 2: Distribution Plots (satisfaction, evaluation and average monthly hours VS employee count)



INSIGHTS OBTAINED:

1. **Satisfaction** - There is a huge spike for employees with low satisfaction and high satisfaction.
2. **Evaluation** - There is a bimodal distribution of employees for low evaluations (less than 0.6) and high evaluations (more than 0.8)
3. **AverageMonthlyHours** - There is another bimodal distribution of employees with lower and higher average monthly hours (less than 150 hours & more than 250 hours)
4. Employees with lower average monthly hours were evaluated less and vice versa.

FIGURE 3: turnover VS salary



INSIGHTS OBTAINED:

1. Majority of employees who left either had **low** or **medium** salary.
2. Barely any employee left with **high** salary
3. Employees with low to average salaries tend to leave the company.

FIGURE 4: department VS employee-count

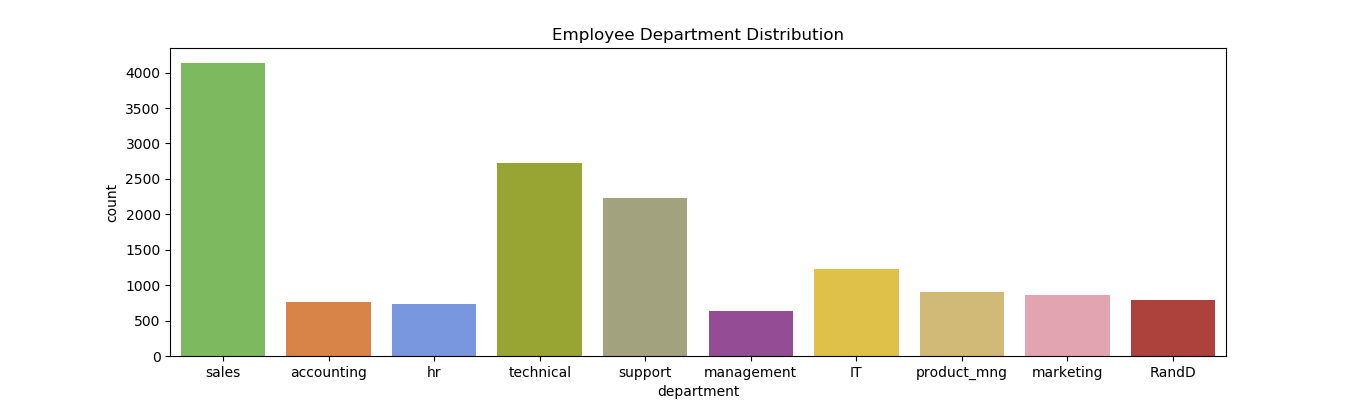
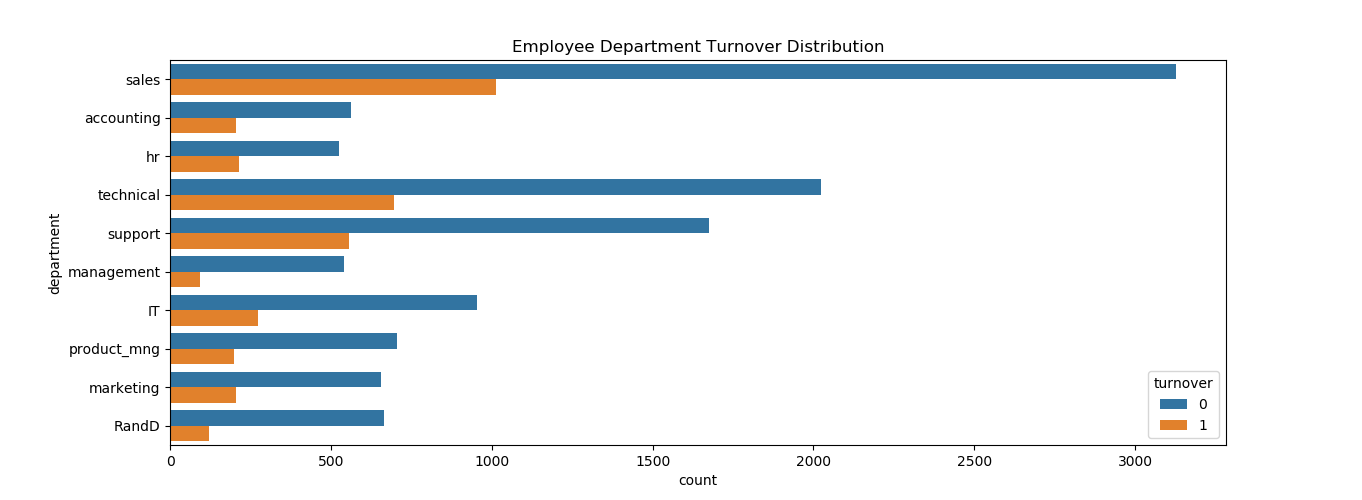


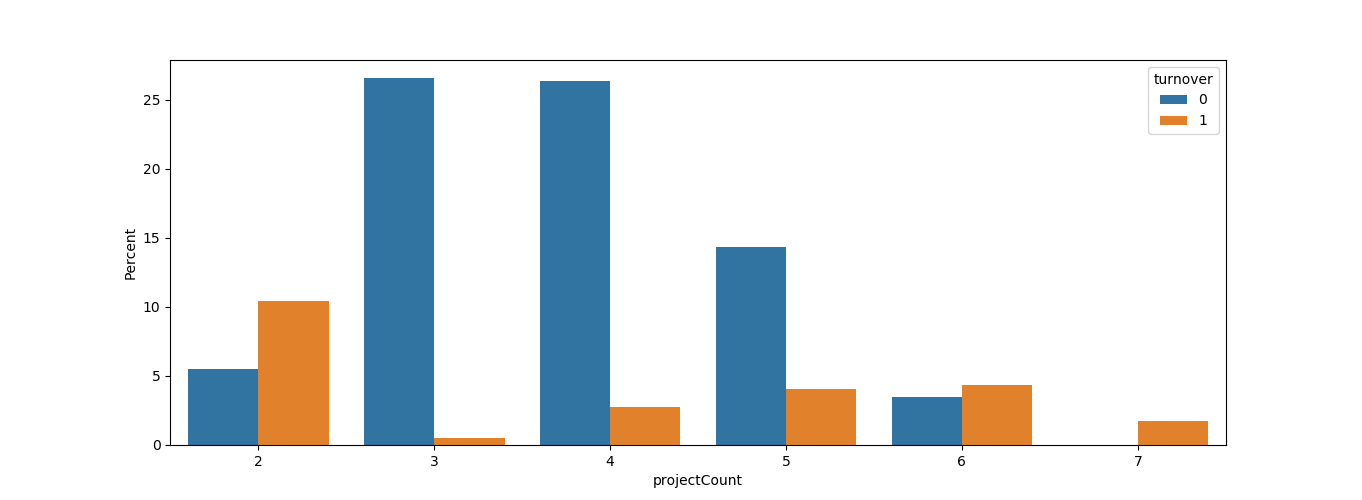
FIGURE 5: turnover VS department



INSIGHTS OBTAINED:

1. The **sales, technical, and support department** were the top 3 departments to have employee turnover.
2. The management department had the smallest amount of turnover.
3. Departments with highest employee count have highest turnover as well.

FIGURE 6: projectCount VS turnover



INSIGHTS OBTAINED:

1. More than half of the employees with **2,6, and 7** projects left the company
2. Majority of the employees who did not leave the company had **3,4, and 5** projects
3. All of the employees with **7** projects left the company
4. There is an increase in employee turnover rate as project count increases

FIGURE 7: yearsAtCompany VS turnover



INSIGHTS OBTAINED:

1. More than half of the employees with **4 and 5** years left the company.
2. Employees with **5** years should **highly** be looked into.

**4. DATA-MODELLING**

Since it is one of the classification-kind problems (moreover binary classification) so the machine learning techniques used were:

1. Decision Tree Classifier,
2. Random Forest Classifier, and
3. Gaussian Naïve Bayes Classifier

FIGURE 8: Comparison between these models:

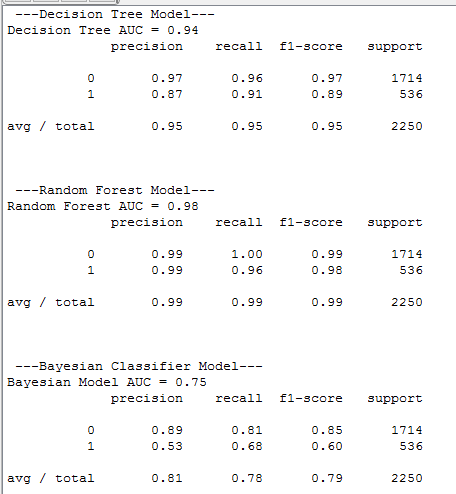
****

FIGURE 9: Receiver Operating Characteristic Curve (ROC curve)

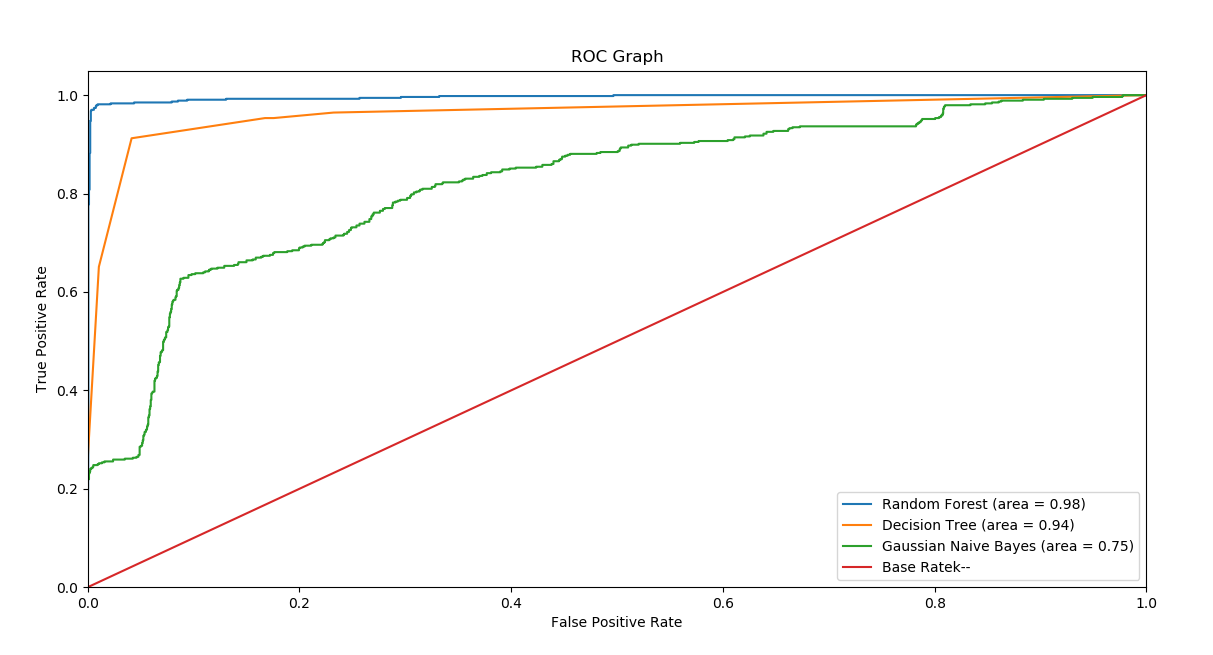
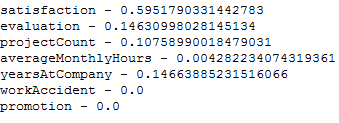


FIGURE 10: Importance of different features as per Decision tree classifier



INSIGHTS OBTAINED:

1. Among the three classifiers, the best one is Random Forest Classifier with an precision and recall of 0.99.
2. The poorest performance was of Gaussian Naïve Bayes Classifier with an accuracy of 0.7

**5. INTERPRETING THE DATA.**

**RESULTS:**

Combining all the insights we get:

1. Employees generally left when they are **underworked** (less than 150hr/month or 6hr/day)
2. Employees generally left when they are **overworked** (more than 250hr/month or 10hr/day)
3. Employees with either **really high or low evaluations** should be taken into consideration for high turnover rate
4. Employees with **low to medium salaries** are the bulk of employee turnover
5. Employees that had **2,6, or 7 project count** was at risk of leaving the company
6. Employee **satisfaction** is the highest indicator for employee turnover.
7. Employee that had **4 and 5 yearsAtCompany** should be taken into consideration for high turnover rate
8. Employee **satisfaction**, **evaluation** and **projectCount** were the three biggest factors in determining turnover.

**CONCLUSIONS:**

As per the understanding of the situation:

1. Employee engagement can go a long way in reducing the attrition rate, as it helps in increasing the satisfaction level of the employee which intern helps in increasing the loyalty towards the firm which intern can transform into productive work which will intern will impact the revenue.
2. Recognition can go a long way in increasing the confidence of the employee which in turn will impact the revenue of the firm which will further decrease attrition rate.
3. Growth is a misconceived term; growth can be horizontal as well, which can motivate the employee to stay in the firm and hence reducing the attrition rate.
4. Root cause analysis: every employee has its own concern which can be resolve through one on one conversation and can help the firm in keeping their employees.

**REFERENCES**

1. Boswell, W.R., Boudreau, J.W., &Tichy, J., (2005), “The relationship between employee job change and job satisfaction: The honey moon-hangover effect,” Journal of Applied Psychology, Vol.47, pp.275-301.
2. Pettman, B. O., (1975), “Labour Turnover and Retention,” John Wiley & Sons, New York.
3. Mobley, W. H. (1982). Employee Turnover: Causes, Consequences, and Control. Addison-Wesley Publishing, Philippines.
4. Trevor, C.O. ((2001). “Interactions among actual ease-of-movement determinants and job satisfaction in the prediction of voluntary turnover,” .Academy of Management journal, Vol 44, pp.621-638.
5. Glebbeek, A.C., & Bax, E.H.,(2004), “ Is high employee turnover really harmful? An empirical test using company records,” Academy of Management Journal, Vol.47, pp. 277-286.
6. Rappaport, A., Bancroft, E., &Okum, L. (2003), “The aging workforce raises new talent management issues for employers” Journal of Organizational Excellence, 23, 55-66.

**LINK TO CODE:**