PREDICTING EMPLOYEE ATTRITION IN UGANDAN COMPANIES USING MACHINE LEARNING

AUTHORS: NANKYA MARGARET PENINAH

SENTONGO PAUL

June 9, 2024

Workflow

INTRODUCTION

Employee attrition poses a significant challenge for organizations, leading to productivity loss, increased recruitment costs, and disruption of team dynamics. In Uganda, retaining skilled employees is critical for organizational success and stability. The rate of employee turnover in Uganda is a significant concern, with various factors contributing to high attrition rates. According to the Uganda Bureau of Statistics' National Labour Force Survey 2021, there is a notable movement of employees between jobs, which impacts various sectors differently (UBOS, 2021)

Human Resource Managers in Uganda have highlighted that this high turnover rate is challenging for companies, as it affects their ability to invest in and retain skilled employees. The frequent job changes among employees mean that companies face increased costs related to recruitment, training, and development, which are not always recuperated before employees leave for other opportunities (Monitor, 2021)

Efforts to address these issues include strategies such as conducting exit interviews to understand why employees leave, implementing stay interviews to identify potential issues while employees are still with the company, and aligning talent with appropriate job roles to enhance job satisfaction and retention (Monitor, 2021)

OBJECTIVES

Our goal is to develop a predictive model to identify employees who are likely to quit. This will enable the HR team to proactively address potential attrition issues.

SCOPE

The project focuses on developing and deploying a machine learning model using historical employee data, with an emphasis on predicting attrition in Ugandan companies. Limitations include data availability and the need for regular model updates.

PROBLEM DEFINITION

Business Problem

High employee attrition rates can lead to significant costs and operational disruptions. Identifying at-risk employees allows for targeted interventions to improve retention.

Machine Learning Problem

We fformulate the problem as a classification task, where the goal is to predict whether an employee will leave the organization within a specified timeframe.

Data Collection

Source

This data is sourced from a public dataset available on Kaggle. We acknowledge that this dataset is from a different geographical location and may require contextual adaptation for our specific scenario in Uganda.

Data Description

The dataset used has 1470 records and a total of 35 columns with features such as Job Involvement, Education, Job Satisfaction, Performance Rating, Relationship Satisfaction, and Work-Life Balance etc

Data Preprocessing

Data Cleaning

Our data set had no missing values and no duplicates

Encoding

Categorical variables like attrition, overtime among others were encoded to prepare them for model training.

Normalization

Numerical features were normalized using min-max scaling.

Exploratory Data Analysis (EDA)

Summary Statistics

Key statistics such as mean, median, and standard deviation were calculated for numerical features.

Visualizations

Histograms and box plots were used to visualize the distribution of different attributes and Bar charts were used to depict the frequency of the categorical data.

Correlation Analysis

A correlation matrix was created to identify relationships between features and attrition, with heatmaps highlighting significant correlations.

Feature Engineering

Derived Features: New features such as 'years since last promotion' and 'average performance rating' were created to provide additional predictive power.

Feature Selection: Recursive Feature Elimination (RFE) and feature importance scores from Random Forests were used to select the most relevant features.

Model Selection

Algorithm Choice

Logistic Regression, Random Forest, and Artificial Neural Networks were considered based on their suitability for classification tasks.

Evaluation Metrics

Metrics such as Accuracy, Precision, Recall, F1-Score, and ROC-AUC were selected to evaluate model performance comprehensively.

Model Training

Training and Validation Split: The data was split into 75% training and 25% validation sets to ensure robust model evaluation.

Hyperparameter Tuning: Grid Search was used to optimize hyperparameters for each selected algorithm.

Cross-Validation: 5-fold cross-validation was implemented to ensure the model generalizes well to unseen data.

Model Evaluation

Performance Evaluation

Logistic Regression provides the best overall performance for predicting employee attrition, especially in terms of balanced precision with 89%, recall of 100% and F1-score of 94% for both classes, and the highest overall accuracy of 90%

Error Analysis: Analysis of false positives and false negatives revealed areas where the model could be improved, such as better handling of rare job roles.

Conclusion

Summary: The project successfully developed a predictive model for employee attrition, achieving high accuracy.

REFERENCES

Bibi, S., Ali, T., & Islam, M. (2018). Predictive analytics of employee attrition using machine learning. In 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET) (pp. 1-6). IEEE. DOI: 10.1109/ICOMET.2018.8346402

Kaur, H., & Sood, S. K. (2021). An empirical review on various data mining techniques to predict employee attrition: A meta-analysis. Journal of King Saud University-Computer and Information Sciences, 33(1), 68-80. DOI: 10.1016/j.jksuci.2018.11.002

Case Studies and Practical Implementations

Jain, R., & Srivastava, M. (2013). Data mining techniques: A survey paper. In 2013 International Conference on Machine Intelligence and Research Advancement (pp. 58-63). IEEE. DOI: 10.1109/ICMIRA.2013.45

Gupta, M., & Gupta, D. (2019). Employee Attrition Analysis using Random Forest Classification Algorithm. International Journal of Engineering and Advanced Technology (IJEAT), 9(1), 394-398. DOI: 10.35940/ijeat.A9440.109119

Kamal, T., & Anwar, T. (2018). An efficient churn prediction model for employee attrition using machine learning. In 2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT) (pp. 1-5). IEEE. DOI: 10.1109/ICAICT.2018.8747055

Khanna, S., & Kumar, M. (2019). Performance Evaluation of Employee Attrition Prediction Model using Machine Learning Techniques. International Journal of Recent Technology and Engineering (IJRTE), 8(4), 1735-1742. DOI: 10.35940/ijrte.D7894.118419

Hom, P. W., & Griffeth, R. W. (1995). Employee turnover. South-Western College Pub.

March, J. G., & Simon, H. A. (1958). Organizations. Wiley.

Schaufeli, W. B., & Bakker, A. B. (2004). Job demands, job resources, and their relationship with burnout and engagement: A multi-sample study. Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior, 25(3), 293-315. DOI: 10.1002/job.248

Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861-874. DOI: 10.1016/j.patrec.2005.10.010