

Introduction

Title: Predicting and Mitigating Employee Attrition in the Pharmaceutical Industry

Objective: To investigate the factors contributing to high employee attrition within a pharmaceutical firm using predictive analytics, and to develop proactive HR strategies based on the findings.

Table of Contents

1. Abstract
2. Scope and Overview
 - Company Overview
 - Issues relating to High Attrition
3. Literature Review
 - Key Insights from Relevant Studies
 - Connection between Existing Research and the Current Analysis
4. Methodology
 - Exploratory Data Analysis
 - Feature Engineering
 - Model Building
 - Model Evaluation
5. Results and Discussion
 - Key Findings
 - Future Recommendations
 - Limitations
6. References
7. Appendix

1. Abstract

This report investigates the factors contributing to high employee attrition within a pharmaceutical firm utilizing predictive analytics. The objectives include identifying root causes and developing proactive HR strategies. Leveraging HRIS data and technologies like KNIME, the study aims to provide the Board with informed decision-making strategies. The comprehensive methodology employs the CRISP-DM framework, exploring data, conducting feature engineering, and employing models like decision trees. The report emphasizes feature improvement, construction, and scaling to enhance model performance. The accuracy was highest for the gradient boosting model at 86.136%, followed by the random forest method at 85%, and the decision tree method at 74.545%. The final model evaluation considers machine learning metrics and interpretability, showcasing the potential for advanced analytics in mitigating attrition challenges and fostering organizational stability.

2. Scope and Overview

2.1 Company Overview

Employee attrition is a major issue for the pharmaceutical company, which tracks HR metrics via monthly Excel reports. This persistent issue hinders operations, efficiency, and financial success. They want to use advanced analytics techniques to overcome their reporting system's limitations. The goals are to better comprehend the data and to address high attrition. The company's Human Resource Information System (HRIS) dataset includes attributes, performance measurements, and attrition. Despite having this data, the firm struggles to understand the complex causes of high turnover.

2.2 Issues relating to High Attrition

The company's 19.15% attrition rate requires a thorough investigation into its causes. Employee attrition encompasses voluntary and involuntary departures. To predict future attrition patterns, predictive analytics like KNIME can be used to analyze HRIS data and identify key factors causing employee turnover.

3. Literature Review

3.1 Key Insights from Relevant Studies

The use of machine learning in predicting employee attrition has been explored in various studies, demonstrating its importance in strategic decision-making. Techniques such as decision trees, random forests, and gradient boosting have been used to identify patterns and predict attrition based on employee attributes.

3.2 Connection between Existing Research and the Current Analysis

This study builds on existing research by using machine learning models to predict employee turnover based on employee attributes, helping management quickly identify and retain at-risk talent. The study employs predictive models using classification techniques to understand employee attrition dynamics and develop HR strategies.

4. Methodology

4.1 Exploratory Data Analysis

Exploratory Data Analysis (EDA) begins by examining raw data for trends and patterns. The dataset has 1467 items in 35 variables, including 25

numeric/integer, 9 text, and 1 identifier and one date variable. EDA helps in identifying missing data, outliers, and data distribution.

4.2 Feature Engineering

Feature Engineering involves enhancing structured features through various transformations, such as handling missing values, standardization, and normalization of numerical and categorical data. This step also includes feature improvement, construction, selection, extraction, and learning.

4.3 Model Building

Machine learning models such as decision trees, random forests, and gradient boosting were built using KNIME. The process involves integrating the Feature Engineering pipeline into a KNIME meta node and selecting classification models for predicting attrition.

4.4 Model Evaluation

Model evaluation includes assessing the accuracy and performance of the models using metrics like accuracy, Cohen's Kappa, and ROC curves. The gradient boosting model showed the highest accuracy at 86.136%.

5. Results and Discussion

5.1 Key Findings

- The CRISP-DM technique emphasizes data mining's importance in solving business challenges methodically.
- Normalization and handling missing values improve model performance.

- Predictive model construction using KNIME showed that the gradient boosting model had the highest accuracy.
- Key factors influencing attrition include 'StockOptionLevel' and 'Overtime'.

5.2 Future Recommendations

- Maximize the use of the Decision Tree model for HR analytics.
- Focus on feature engineering and data quality.
- Expand model evaluation metrics to include Precision, Recall, and F1-Score.
- Integrate insights from models into HR decision-making processes.

5.3 Limitations

- Data limitations may affect model performance.
- The complexity of attrition may limit prediction accuracy.
- Interpretability challenges with decision tree models.
- Need for robust validation approaches.

6. References

A comprehensive list of references is provided, including sources from HR analytics, machine learning, and predictive modeling literature.

7. Appendix

Meeting 1: December 27, 2023

- Focus: Team acquaintance and outlining activities.

Meeting 2: December 29, 2023

- Focus: Formulating the process for model building.

Meeting 3: December 30, 2023

- Focus: Familiarizing with KNIME software and determining requisite nodes for model building.

Meeting 4: January 2, 2024

- Focus: Assigning written work for the report and model building tasks on KNIME.
- Tasks executed by team members:
 - Dhanush: Model Building, Scope and Overview, Built Models using KNIME.
 - Rohan: Methodology, Model Building using KNIME, Results and Discussion.
 - Manogna: Literature Review, EDA, Model Building, Models on KNIME.
 - Mrunmayee: Abstract, Literature Review, Results and Discussion, EDA.
 - Cindrella: Literature Review, Scope and Overview, EDA, Results and Discussion.

Signed Declaration

We hereby declare that we have coordinated and completed the group assignment for Human Resource Analytics submitted to Queens' University Belfast, as per the activity report above.

- Cindrella KC - 40429497
- Dhanush MS - 40412492

- Mrunmayee Bapat - 40420299
- Manogna B R - 40426970
 - Rohan Mahesh Patil - 40395741