

# Predicting Employee Attrition with Machine Learning

Employee attrition (churn) is a significant cost to an organization. A high attrition rate can lead to increased tangible costs such as training, recruitment, and on-boarding, as well as intangible costs such as project management and customer relationships. Attrition analysis and support in defining an optimal talent retention strategy. However, this requires a deep understanding of employee behavior. If HR directors designed talent retention strategies based on data and machine learning, companies could significantly reduce attrition and see an increase in employee productivity and company profitability. Together with attrition, the quantification of employee value and attrition cost is key to making optimal HR decisions.

A team of data scientists builds an analytics solution for the HR department to help predict which employees are most likely to leave. A KNIME workflow collects employee data provided by the HR department (age, gender, and education, as well as salary, salary increases, business travel frequency, etc.), and estimates the probability of each employee leaving the company.

For the entire workflow, the dataset should be captured at a fixed window depending on the operational process of the organization. The dataset is then cleaned up for outliers, erroneous values, and/or representational structure. Employee attrition datasets are usually imbalanced for the attrition category; hence a rebalancing exercise is performed. Several models such as Random Forest, Logistic Regression, Naïve Bayes, and Gradient Boosted are trained, with the best performing model being chosen to score current employees.

The workflow is then deployed on KNIME Server and can be executed on demand or using the scheduling option, to update attrition probabilities. A dataset is generated with the model output, which the HR department uses for analysis on the profiles of the employees likely to leave i.e. how they distribute among departments, years spent at company, years since last promotion, last salary increase, etc. With the native Tableau integration within KNIME Analytics Platform, the results are exported directly to a Tableau dashboard. This provides business users with access to informative and useful visualizations quickly and easily.

A Guided Analytics application, made accessible via the KNIME WebPortal, enables business users to make parametric changes to understand their impact on the attrition process. Business users can also simulate scenarios. For example, whether the likelihood of an employee staying increases if their salary increases. This empowers the business users and promotes a data-driven decision-making culture.

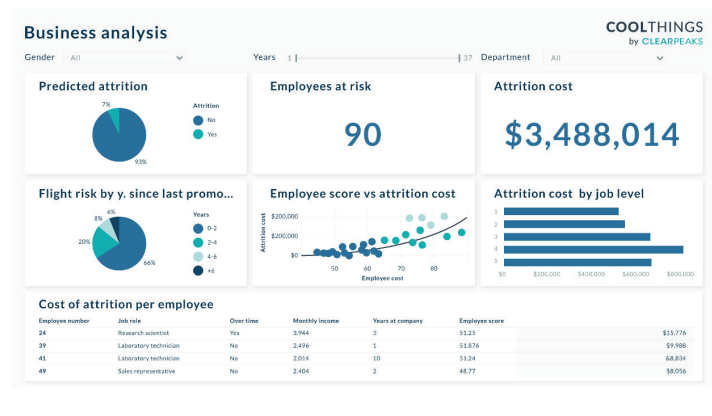


Fig 1. Visualization of results using the native Tableau Integration in KNIME showing estimated attrition cost along with key KPIs and analysis.

## Why KNIME Software

KNIME provides the ideal environment for building a classification machine learning model, and conducting an attrition analysis. It's possible to build workflows using the appropriate connector for different data sources, transform data, and train machine learning models. KNIME Server enables these workflows to be executed at predefined times. The KNIME WebPortal allows business users to interact with these workflows by uploading their own data or updating the attrition probabilities on demand. The seamless integration with Tableau, enables users to create dashboards in order to dive deeper into the results.

### Results:

With this solution, HR directors have access to the attrition probability for each of their employees, enabling them to:

- Create successful talent retention strategies
- Increase the retention of valuable employees and reduce overall attrition rate
- Reduce attrition-related costs
- Understand employee behavior

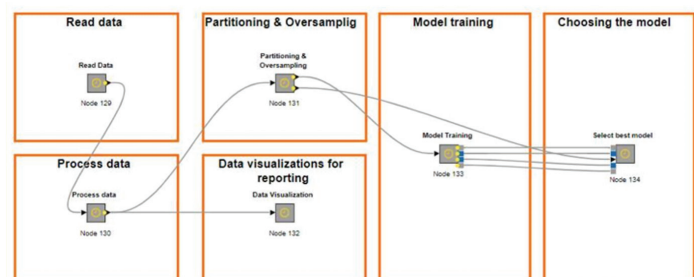


Fig 2. High level KNIME workflow.

