Predicting Employee Attrition

Author: Nolan Vegh
Faculty Mentor: Dr. Christoforos Christoforou

Introduction

The most important asset to a company is its' employees; therefore, minimizing employee attrition rates is top priority for large companies to avoid employee replacement costs and operational delays. Employee surveys could provide aggregate measures of the overall satisfaction of employees; however, anticipating when an employee might leave the company within a running year is a challenging task.

Study Objectives

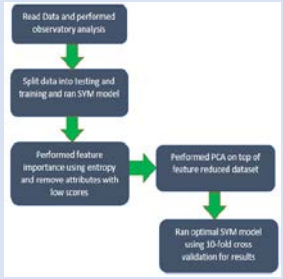
Explore the use of building machine learning models to predict the likelihood an employee leaving their company based on employee satisfaction data.

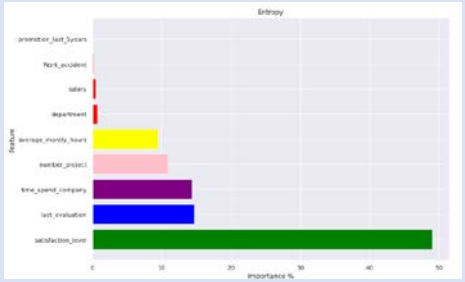
Key Findings

The Support Vector Machine (SVM) classifier applied on factors extracted using Principal Component Analysis (PCA) on employee satisfaction data predicts 97% of an employee's attrition likelihood.

Methodology

This analysis was performed on an online dataset which captures employee satisfaction survey data, performance data, and workload. The dataset is comprised of 14,999 records and 9 features.





The flowchart (left) shows the steps taken in this study. As part of the feature reduction phase, the entropy was calculated for each of the features in the dataset (right); and based on their entropy scores features *Department, Salary, Work_Accident, and Promotion_last_5_years* were removed.

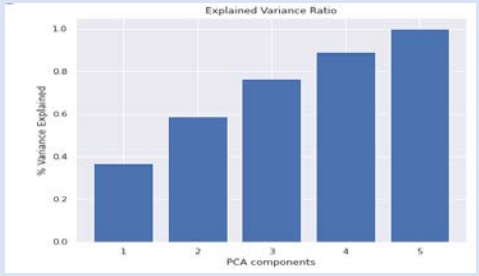


Figure 2. Principal Component Analysis was employed for dimensionality reductions. We considered using top two, and top four PCA components as features, capturing 59% and 89% of the variance, respectively.

Neuro-Intelligence Innovation Lab

Results

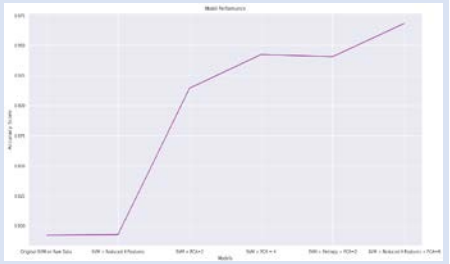


Figure 3. Shows 10-fold cross validations accuracy scores of the SVM model using the different feature sets. Model accuracy on the raw feature-set was 79% (baseline accuracy); PCA features improve model accuracy to over 94%, which corresponds to 19% increase in accuracy; combined PCA+extra features achieved 97% accuracy, which corresponds to 23% increase in accuracy.