

# PREDICTIVE HUMAN RESOURCE MANAGEMENT MODEL USING ATTRITION ANALYSIS

Amit Patti - PES1UG19CS054

Anisha Ghosh - PES1UG19CS067

Anupama Nhavalore - PES1UG19CS071

Sai Charan Poilishetty - PES1UG19CS329



## Team Attrition

# Overview

---

- **Introduction**
- **Problem Statement**
- **Objectives**
- **Solution Approach**
- **Models**
- **Implementation**
- **Results & Analysis**
- **Conclusion**
- **Contribution**

# Introduction



- The key to the survival, growth and development of an organisation stems from the way it structures its human resources and deals with attrition.
- The major pain points seen in Human Resource Management (HRM):
  - employee commitment
  - productivity
  - retention in the long run.
- The success of an organisation is driven by the balance maintained between the company's structural paradigm and its employee work environment.

# Problem Statement

Attrition is a widespread issue in today's fast-changing scope of business affecting both cost and retention of experienced talent.

Human Resource Management is a field dominated by the subjective perspective of every individual involved – employee and employer.

It proves to be time inefficient for an industry that relies on the current market situation and timing, which can prove to be very costly in the long run.

The dataset we're using to analyse and predict employees' attrition is provided by IBM analytics. This dataset includes 35 features and close to 1500 samples.



# Objectives

## Abstracting analysis

Our solution looks to aid the HRM industry to create a predictive analyzer to speculate attrition given the attributes as input.

## Optimise Process

- Improve in-house diagnostics & hiring process.
- suggest and bring about changes in company culture
- make specific recommendations to promote the retention of valuable employees.

## Impact

Bring in objectiveness into the HRM industry in terms of evaluating attrition.

Create some improvement by automating the process of attrition analysis.

# Solution Approach



In the sample of 1470 employees, 16% left their jobs, i.e. they underwent attrition, implying the dataset is imbalanced.

Pertaining to our problem and dataset, we had to use and augment classification algorithms to predict attrition.

We explored 6 different models spread across independent and ensemble classifiers for a comparative analysis to build a robust model for attrition analysis.

# Models

---



## Classification Models

We chose 4 classification models of varying complexities:

- Logistic Regression
- Naive Bayes Classifier
- K-Nearest Neighbours (KNN)
- Support Vector Machines(SVM)

## Ensemble Models

We chose one of each type of ensemble classifier -

- XGBoost (extreme gradient boosting)
- Random Forest Classifier (A bagging technique)

# Implementation

## Logistic Regression

PCA was done to represent data in 2 components and use it in Logistic Regression. The model had an accuracy of 82.319% after PCA.

## Naive Bayes

As it's based on Bayes theorem, it performs well with high dimensional data due to independence assumption of attributes. With this we achieved an accuracy of 79.591%.

## KNN

After reducing the dimensionality and using oversampling techniques, the optimal 'k' value found was  $k=9$ . The final model had an accuracy of 82.993%.



# Implementation

## SVM

SVM is effective for high dimensional data. The params for SVM; RBF kernel with C value = 1, gamma was scaled according to the dataset. The approach gave us 83.333% accuracy

## XGBoost

XGBoost uses gradient descent to minimise the loss function onto which the hypothesis is fit . It builds a tree using the greedy method, and prevents overfitting by penalizing trees that are too deep using regularization. The model gave an accuracy of 87.074%

## Random Forest Classifier

Random Forest classifiers work on aggregating independent, uncorrelated diverse decision trees. The features are selected based on entropy. In our model, we used 10 estimators and got an accuracy of 92.52% which is best among all models.

# Results & Analysis

Performance Metrics/Model	Logistic Regression	Naive Bayes	SVM	KNN	XGBoost	Random Forest Classifier
Accuracy (%)	82.3129	79.5918	83.3334	82.9932	87.0748	92.5200
Precision	0.8339	0.9186	0.8334	0.8328	0.8892	0.8643
Recall	0.9837	0.8286	1.0	0.9959	0.9755	0.9878
F1 score	0.9026	0.8712	0.9091	0.9071	0.9264	0.9219
Time taken for execution(s)	0.2322	0.01501	0.1681	0.0070	90.6854	0.1661

- Random Forest Classifier and XGBoost gave us the best accuracy values -These being ensemble models also show that they're not overfitting the data and account for the reproducibility of results for any test dataset
- Naive Bayes - yields the lowest accuracy due to independence and equal contribution assumption of attributes.
- Logistic Regression and SVM - yielded good accuracy but fell short due to the imbalance in the dataset
- KNN - yields a high accuracy, but due to the curse of dimensionality it has bias towards the majority class

- Overall, we see that the ensemble models performed much better than the classification models as ensemble models were able to deal with imbalance in the dataset
- Other models like SVM, and XGBoost can work better with a more balanced dataset.
- The attributes that contribute the most to attrition are monthly income, age, distance from home, years with current manager, percent salary hike, years at the current company, and Job role.

# Conclusion

- The Random Forest Classifier being the most accurate model, can be used for making relevant tools in the HRM industry for attrition analysis.
- The model doesn't need a lot of parameter tuning, it's simpler to visualise and improve with changing times.

# Contribution

## - Team Attrition

---

### **01 Amit Patti - PES1UG19CS054**

EDA, research for literature survey, SVM, Random Forest Classifier, overall code compilation

### **02 Anisha Ghosh - PES1UG19CS067**

EDA, Research Paper (Phase1&2), KNN, XGBoost, Logistic Regression with PCA, Presentation

### **03 Anupama Nhavalore - PES1UG19CS071**

EDA, Research Paper (Phase1&2), Naive Bayes Classifier, Random Forest Classifier, Presentation

### **04 Saicharan Polishetty - PES1UG19CS071**

EDA, research for literature survey, PCA, Simple Logistic Regression

THANK  
YOU

**TEAM ATTRITION**