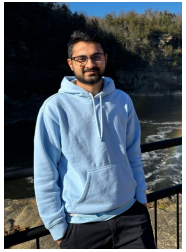
Employee Attrition

A case study to understand factors impacting employee attrition problem from data science lens

Our Team

Bhavika Bhutani

Neha Kulkarni



Rohan Deshpande

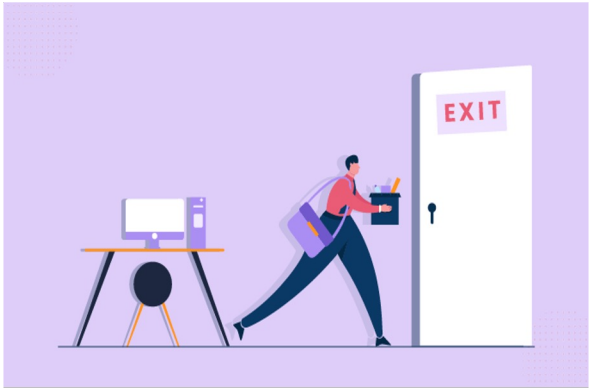
Shreyash Shinkar

Vedant Bhatambrekar

Agenda

- Introduction
- Literature Review
- Research Questions
- Methodology
- Results
- Limitations & Future work
- Conclusion

Introduction



What is Attrition?
Attrition is *the reduction of workforce by employees' leaving*

Why Attrition Data Matters:
Helps understand employee feelings

Impact on Companies:
Affects team morale & productivity

Future Effects:
Losing valuable expertise, disruptions in workflow, and spending resources on recruitment and training

Literature Review

[1] Fallucchi, F., Coladangelo, M., Giuliano, R., & De Luca, E. W. (2020). Predicting Employee Attrition Using Machine Learning Techniques. Computers, 9(4), 86. https://www.mdpi.com/2073-431X/9/4/86

Utilized correlation analysis and algorithms like Naïve Bayes to identify key factors contributing to attrition

[2] Reddy Dereddy, A. (2022). Predictive analytics for employee attrition & performance [Master's thesis, Iowa State University]. Iowa State University Digital Repository. https://lib.dr.iastate.edu/etd

Performed exploratory analysis and used machine learning models to predict attrition

[3] Yahia, N. B., Hlel, J., & Colomo-Palacios, R. (2021). From big data to deep data to support people analytics for employee attrition prediction. IEEE Access, 9, 60447–60458. https://doi.org/10.1109/access.2021.3074559

Discusses data imbalance and sampling techniques to overcome it

Research Question

What factors tell us most about why employees leave?

Application

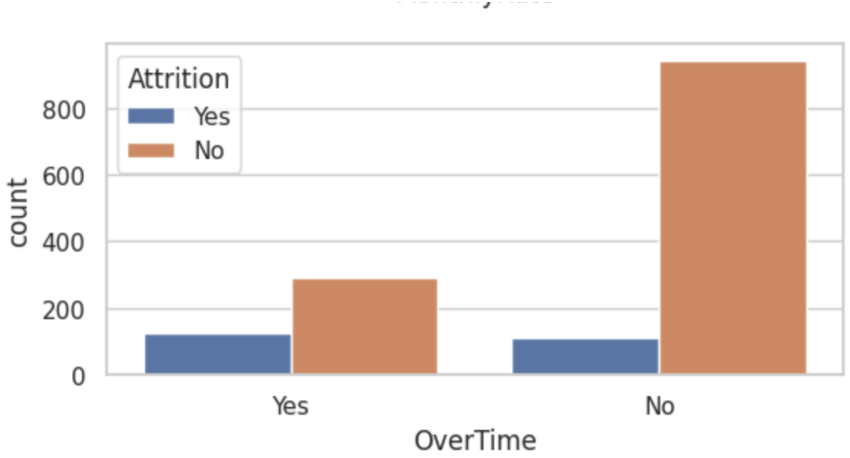
This research question holds substantial importance for both individual companies and the broader realm of workforce management.
This is a critical factor to understand that can help the company in following ways:

- Business Impact
- Productivity and Performance
- Talent Retention
- Strategic Decision-Making
- Employee Engagement and Satisfaction
- Competitive Advantage
- Forecasting and Planning

Analysis

Based on the decision trees and random forest implementation, following are the important factors:

- Over Time
- Switch Rate
- Monthly Income
- Years with current manager
- Work Life Balance

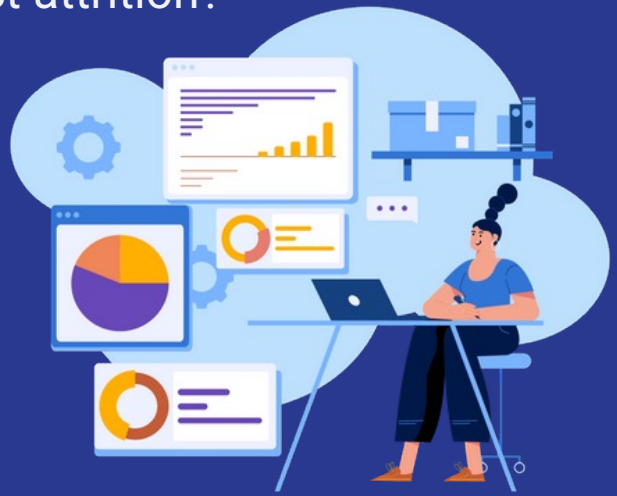


Results

- High levels of overtime might indicate potential lead in contributing to attrition.
- High switch rates might employ that employees are seeking better prospects elsewhere due to unmet expectations or limited career advancement.
- Poor work-life balance can lead to stress, dissatisfaction, and a desire to seek opportunities elsewhere that offer better balance.
- The companies can focus on Addressing Workload, Salary and Benefits, Managerial Training & Work-Life Balance Initiatives

Research Question

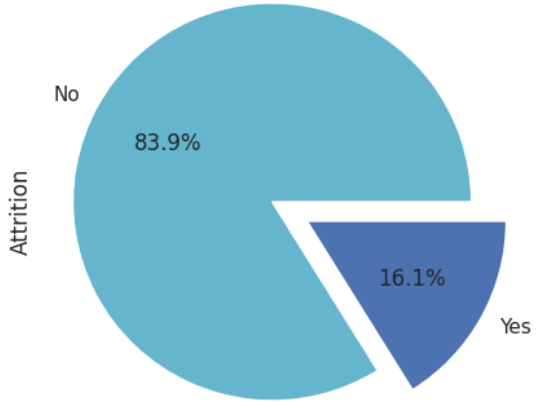
Which classification model can be used to best predict attrition?



Methodology

Data Preprocessing:

1. Elimination of redundant columns
2. Type conversion
3. Application of one-hot encoding for categorical data
4. Introduced new variable – SwitchRate



Modelling:

- Addressing class imbalance using SMOTE (Synthetic Minority Oversampling)
- Classification using
  - **Logistic Regression**
  - **Decision Trees**
  - **Random Forest**

Comparison of Classification Models

Model	Accuracy	Precision	Recall	F-1 Score
Decision Trees	0.8458	0.5365	0.3098	0.3928
Random Forest	0.8412	0.5384	0.0985	0.1667
Logistic Regression	**0.9247**	**0.9431**	**0.9016**	**0.9283**

- Logistic Regression outperforms other models
- Decision trees and random forest provide decent accuracy, but struggle with precision, recall
- The dataset being imbalanced, recall becomes an important metric
- Logistic Regression will be the best model to predict attrition

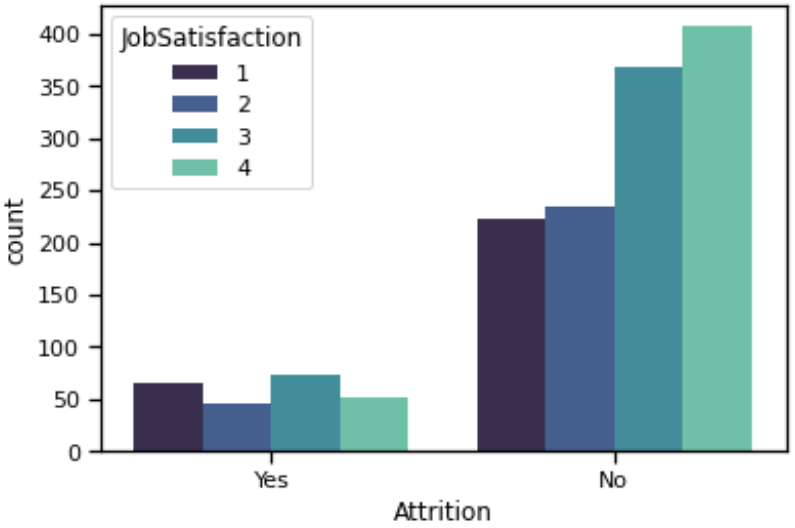
Research Question

Can we identify the factors that affect Job Satisfaction Rating?

Application

It is crucial to understand factors affecting job satisfaction for businesses -

- Employee Retention and Reduced Turnover
- Employee Productivity and Performance
- Positive Impact on Company Culture and Reputation



Methods and Analysis

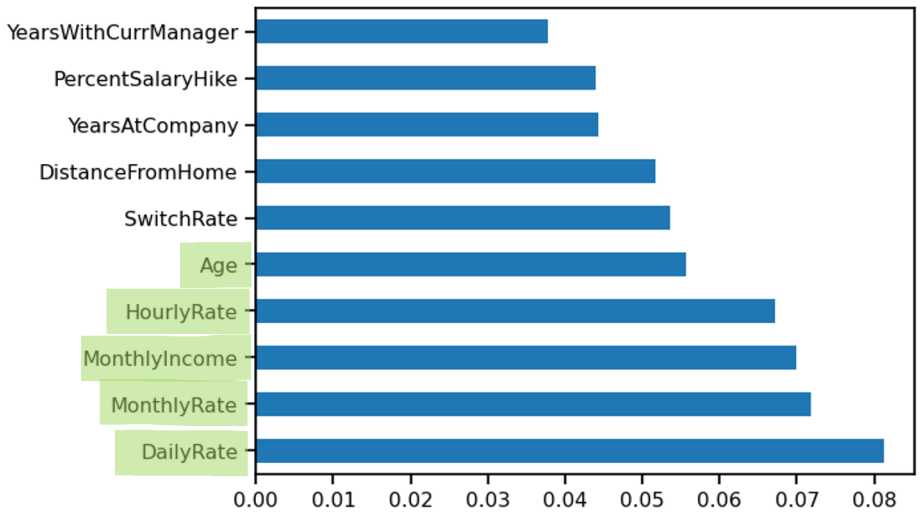
Significant attributes (chi-squared test)

Attribute	P Value
MonthlyRate	0.00E+00
DailyRate	7.81E-233
MonthlyIncome	9.29E-60
HourlyRate	2.45E-11
EducationField_Marketing	8.28E-02

Mutual Information Scores

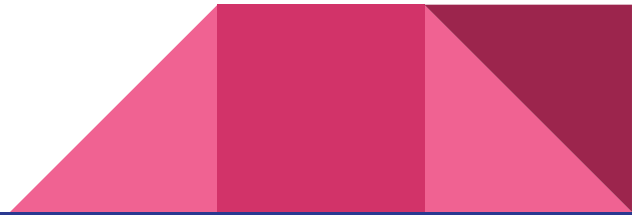
Attribute	Importance Score
MonthlyIncome	0.0317
Age	0.0266
MaritalStatus_Married	0.023
EducationField_Medical	0.0214
MonthlyRate	0.0199
JobInvolvement	0.0158
YearsInCurrentRole	0.0156

Random Forest Feature Importance Plot



Results

- Job Satisfaction rating seems to be highly dependent on Compensation (Monthly Rate, Hourly Rate, Daily Rate, Monthly Income)
- Age is also a factor affecting the Job Satisfaction rating
- No conclusive results can be drawn about the Job Satisfaction rating due to a lack of significant dependency on the other attributes
- A more holistic study of a larger sample set and personalized factors could provide more effective results



Research Question

Is there a significant relationship between salary increments and employee performance?

Can we derive new attributes that are significant in understanding employee attrition?

Research Questions & Applications

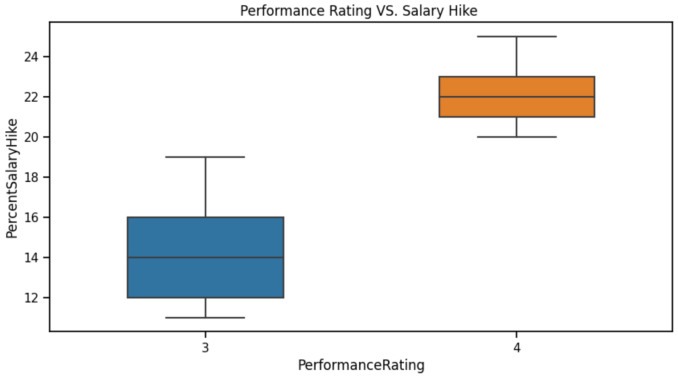
Is there a link between how much employee's salaries increase and their performance ratings?

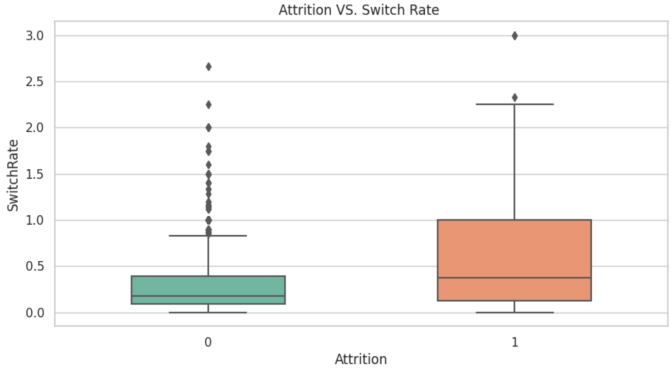
- Help in drafting retention strategies
- Aid in performance management
- Provides competitive advantage to companies

Is it important to see if how often an employee switch companies is related to Attrition

- Understand cost implication patterns
- Retention Strategies
- Market Intelligence

Implementation Methods

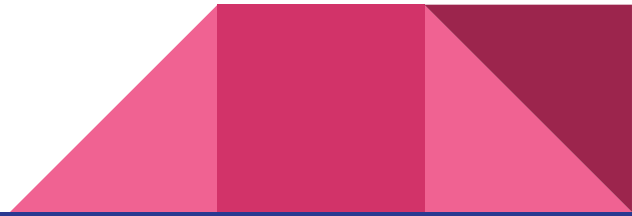




1. **Derived Variables**
   Switch Rate = #of companies switched/ years of experience
2. **Point Biserial Correlation**
3. **Spearman's Correlation**
4. **Boxplot Visualizations**

Results

- Through Spearman’s correlation and Point Biserial Correlation we were able to prove a significant relationship between Performance Rating and Percent Salary Hikes.
- Furthermore, through visualizations we derived that employees with higher percent hikes tend to have higher performance rating
- Calculated a novel metric Switch rate, which through Point Biserial Correlation proved out be significantly correlated to Attrition
- In today’s world too it’s considered people who tend to switch more jobs should be more prone to attrition

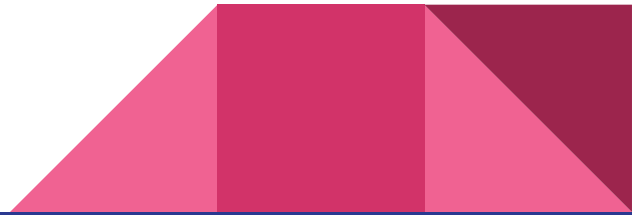


Limitations

- Lack of sufficient open data sources due to privacy concerns to conduct attrition analysis
- Bias in the data, lack of diversity in the employee groups will lead to biased results not applicable to general workforce
- Data imbalance significantly affects the performance of predictive models making it challenging to model results
- External factors are often overlooked in such studies due to the dynamic nature of the market conditions

Future Work

- Human-centric approach using qualitative analysis on data collected from exit interviews or surveys
- Feature engineering and consulting with HR experts to identify key-metrics and drivers of attrition may help improve model performances
- Utilizing Deep learning and AI techniques to model attrition analysis



Conclusion

1. Logistic Regression proved out to be the most efficient in predicting attrition among IBM employees
2. Among multiple predictive models we found features like switch rate, over time, work life balance and monthly income among the common significant features
3. The data is more fitted for attrition analysis rather than predicting jobs. Numerous patterns found among the data can be used by HR analytics sytems to keep a hold of Attrition
4. Data is imbalanced and not sufficient to come to any strong conclusion for predicting Attrition

References

Fallucchi, F., Coladangelo, M., Giuliano, R., & De Luca, E. W. (2020). Predicting Employee Attrition Using Machine Learning Techniques. Computers, 9(4), 86. https://www.mdpi.com/2073-431X/9/4/86

Reddy Dereddy, A. (2022). Predictive analytics for employee attrition & performance [Master's thesis, Iowa State University]. Iowa State University Digital Repository. https://lib.dr.iastate.edu/etd

Subhash, P. (2017). IBM HR Analytics Employee Attrition & Performance. Kaggle. https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset

Yahia, N. B., Hlel, J., & Colomo-Palacios, R. (2021). From big data to deep data to support people analytics for employee attrition prediction. IEEE Access, 9, 60447–60458. https://doi.org/10.1109/access.2021.3074559