Predicting Employee Attrition in the Workplace Using Predictive Models

1. Background of the Study

Employee attrition has emerged as a key concern for organizations across industries hence affecting their financial stability, productivity, and overall effectiveness (Hasan, 2017). Attracting and maintaining top personnel is more important than ever in today's business. Understanding the underlying reasons and anticipating staff attrition is critical for businesses to execute successful solutions and prevent the negative effects of high attrition rates.

Historically, attrition was seen as inevitable and managed reactively. With data analytics and employee behavior insights, attrition can be predicted before it becomes a major issue (Erkkilä, 2020). Sustainable and successful organizations need to transition from reactive to proactive management. The attrition challenge is complicated by many factors. The competitive job market, changing employee expectations, and changing workplace dynamics all contribute. Economic conditions and industry trends also affect talent flow in organizations.

This research is essential to identify employee attrition patterns, trends, and indications (Emerald Insight, 2019). By analyzing historical data, employee surveys, and external factors, researchers can find early warning indicators and build predictive models. This move towards predictive analytics helps retain valuable personnel and plan for workforce changes. Some reasons employees leave include the workplace environment, career advancement, work-life balance, and relationships with managers.

Recent studies have shown that employee attrition affects company culture and reputation (Holtom et al., 2008). High attrition rates can indicate internal issues and negatively impact current and prospective employees. The relationship between attrition and organizational health highlights the need for predictive models to guide strategic decision-making.

2. Business Issue/Problem

Employee attrition is a major issue for firms, damaging their bottom line and overall operational efficiency. The expenditures of recruiting, onboarding, and training new staff are significant (Chhavi, 2022). Aside from the financial implications, attrition damages team relationships, resulting in lower morale and productivity among surviving employees. When senior staff leave, businesses risk losing institutional knowledge and expertise. To address this issue, a proactive approach beyond reactive measures is required to predict attrition tendencies and apply targeted retention efforts (Grotto et al., 2017).

3. Objective/Goals of the Project

This project aims to create a prediction model for staff attrition to help companies avoid losing employees. It also seeks to discover significant attrition traits and patterns for a clear understanding and retention program creation.

4. Data Description

I acquired the dataset from the Kaggle website, here is the link to the dataset https://www.kaggle.com/datasets/rishikeshkonapure/hr-analytics-prediction/data. There are 35 variables and 1,470 employee records. The purpose of this data collection is to develop a model to forecast employees who are likely to quit based on the information about these 1,470 employees and the following characteristics:

- Age: Age in years of the employee
- Attrition: People who leave
- BusinessTravel: How often an employee embarks on a job-related travel

- DailyRate: Daily rate at which an employee is paid
- Department: Department where the employee works
- DistanceFromHome: Distance an employee travels from home to work
- Education: Level of education of the employee
- EducationField: What field the employee studied in school
- EmployeeCount: Count of employee
- EmployeeNumber: EMployee number
- EnvironmentSatisfaction: Employee environment satisfaction
- Gender: Gender of the employee
- HourlyRate: Hourly rate of pay of the employee
- JobInvolvement: Employee job involvement ratings
- JobLevel: Employee Job level
- JobRole: Employee Job role
- JobSatisfaction: Employee Job Satisfaction
- MaritalStatus: Employee Marital Status

5. Data Exploration/Data Visualization

In this section, we are going to discuss the findings of data visualization. This section is very crucial because it will enable us to identify hidden patterns and structure of the dataset (Wilke, 2019).

Table 5.1

Descriptive Statistics

Variable	count	mean	std	min	25%	50%	75%	max
Age	1470	36.92	9.14	18	30	36	43	60
Daily Rate	1470	802.49	403.51	102	465	802	1157	1499
Distance From Home	1470	9.19	8.11	1	2	7	14	29
Environment Satisfaction	1470	2.72	1.09	1	2	3	4	4
Hourly Rate	1470	65.89	20.33	30	48	66	83.75	100
Job Involvement	1470	2.73	0.71	1	2	3	3	4
Job Satisfaction	1470	2.73	1.10	1	2	3	4	4
Years At Company	1470	7.01	6.13	0	3	5	9	40
Number Companies Worked	1470	2.69	2.50	0	1	2	4	9
Monthly Income	1466	6504.96	4708.05	1009	2914.75	4933	8379	19999
Work Life Balance	1470	2.76	0.71	1	2	3	3	4
Performance Rating	1467	3.15	0.36	3	3	3	3	4
Years Since Last Promotion	1470	2.19	3.22	0	0	1	3	15

Table 5.1 shows that, on average, employees are 37 years old with ages ranging from 18-years old to 60 years old. Daily pay rates vary considerably, averaging $802 with a standard deviation of $404. Employees typically travel 9 miles from home to work, ranging from 1 to 29 miles. Work environment satisfaction is moderate since the average rating is 3 on a scale of 1 to 4. Job involvement and satisfaction are moderate also since the average rating is 3. The average

tenure with the company is 7 years with a wide range from 0 to 40 years. On average, employees have worked for three different companies. Monthly incomes show wide variability, averaging $6505 and ranging from $1009 to $19999. Work-life balance is moderate since the average rating is 3. Promotions are not frequent because it occurs 2 times on average with a wide range from 0 to 15 years since the last promotion.

Figure 5.1

Comparison of Number of Employees and Percentage of Attrition Across Different Categories of Department

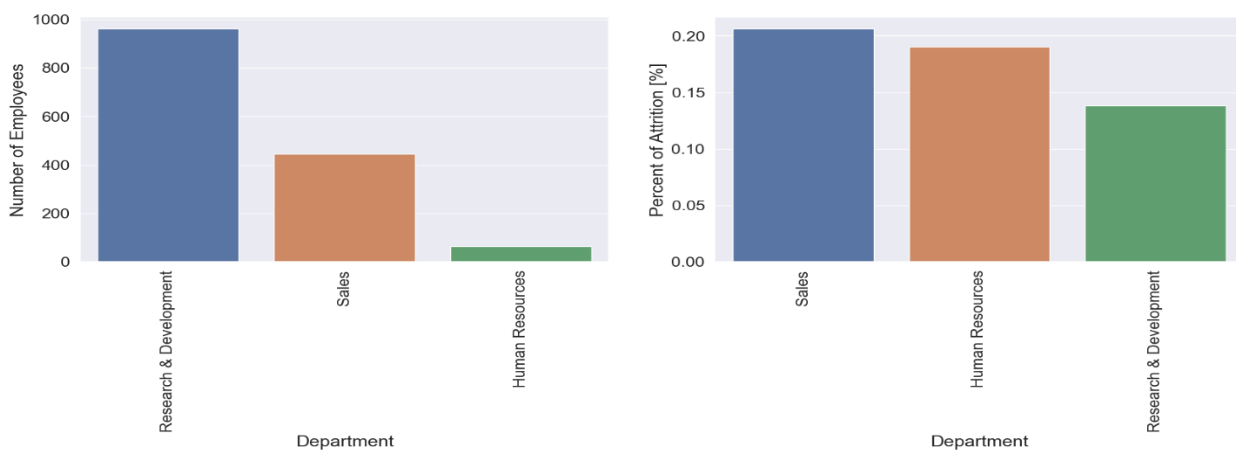

Figure 5.1 shows that sales and human resources employees have a higher likelihood of leaving the company than research and development employees. This could be due to a variety of factors such as high-pressure environments, demanding targets, which could lead to increased stress and potential dissatisfaction. Human resources may face challenges related to employee relation such as conflict resolution and other organizational changes hence contributing to a higher attrition rate.

Figure 5.2

Comparison of Number of Employees and Percentage of Attrition Across Different Categories of Business Travel

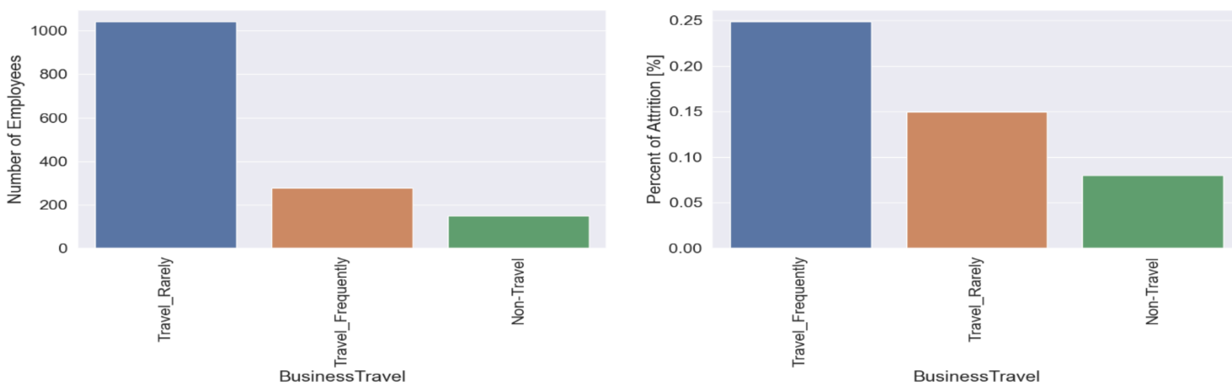

Figure 5.2 shows that frequent travelers' employees are more likely to leave the company. Frequent business travel can cause work-life imbalance and dissatisfaction due to extended time away from home.

Figure 5.3

Comparison of Number of Employees and Percentage of Attrition Across Different Categories of Education Field

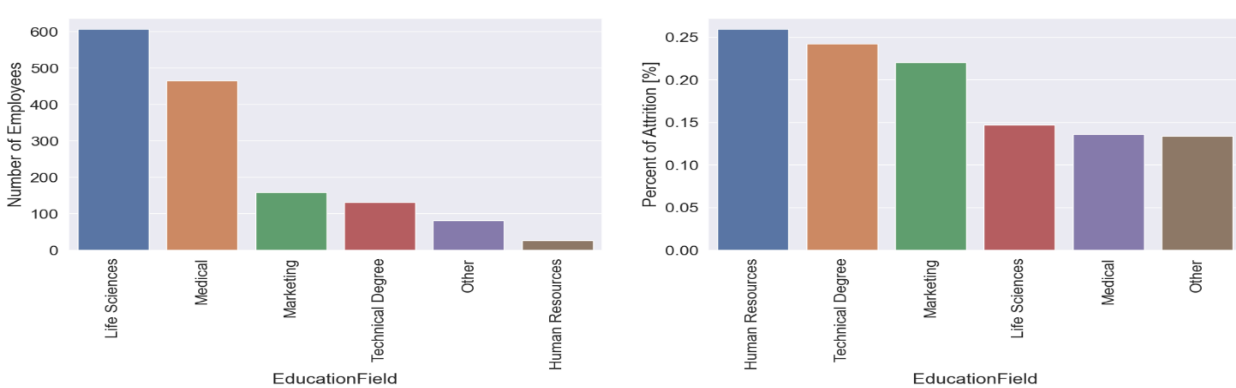

Figure 5.3 shows that employees with a background in human resources show a higher likelihood of leaving the company as compared to those who studied medical and other professions. This could be due to a variety of challenges such as dealing with employee's relations, workplace conflicts and other demanding tasks related to HR.

Figure 5.4

Comparison of Number of Employees and Percentage of Attrition Across Different Categories of Gender

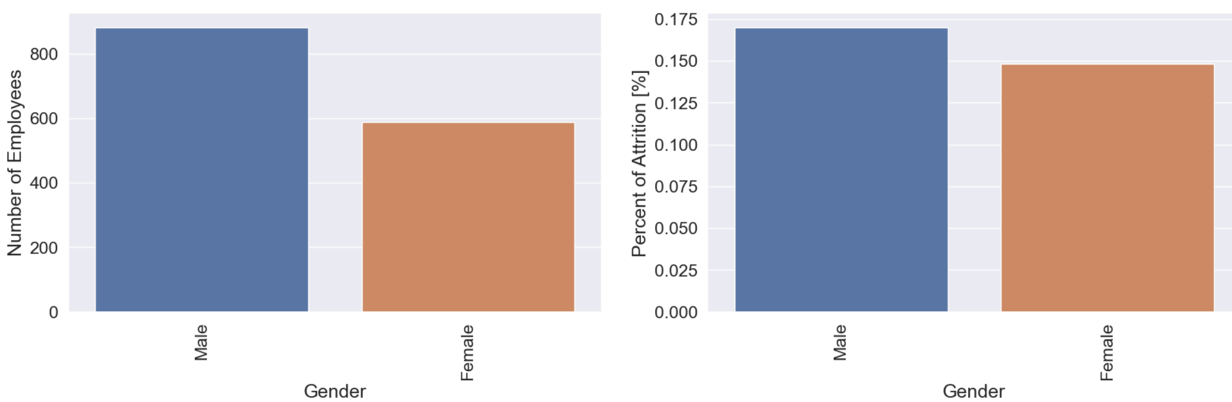

Figure 5.4 shows that male employees show a higher likelihood of leaving the company than female employees. Some of the factors are imbalance work-life, job dissatisfaction, and career advancement opportunities.

Figure 5.5

Comparison of Number of Employees and Percentage of Attrition Across Different Categories of Job Role

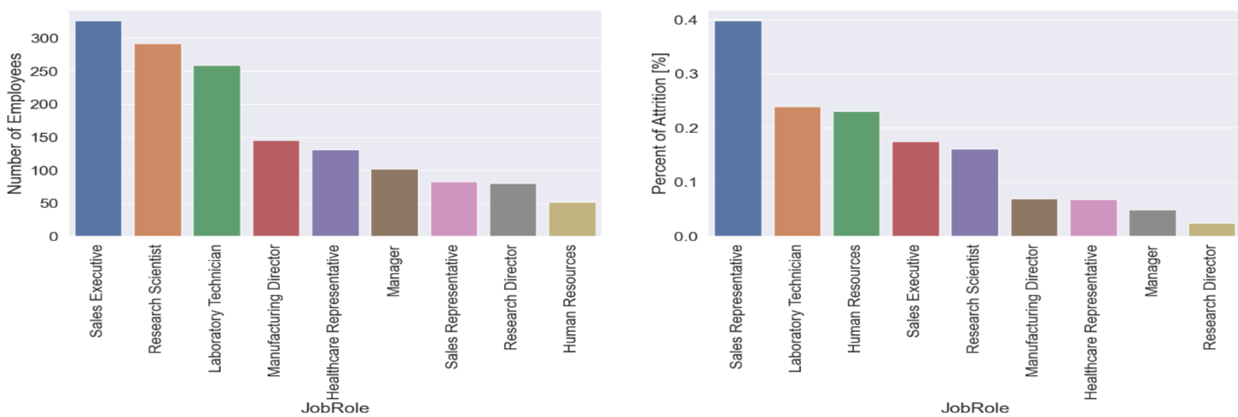

Figure 5.5 shows that sales representatives are more likely to leave the company than managers and research directors. High-pressure targets, intense competition, and frequent client interactions may cause this. Sales positions are stressful and dissatisfying which may increase attrition compared to managers and research directors.

Figure 5.6

Comparison of Number of Employees and Percentage of Attrition Across Different Categories of Marital Status

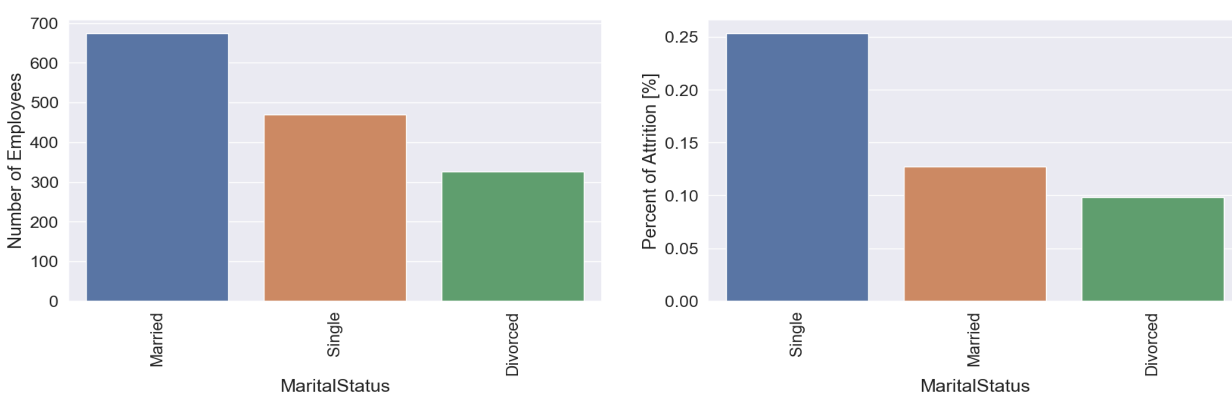

Figure 5.6 shows that single employees are more likely to leave the company than married or divorced employees. This could be due to a variety of factors such as different priorities and responsibilities. Single employees have more freedom in their personal lives, which leads to job changes.

Figure 5.7

Comparison of Number of Employees and Percentage of Attrition Across Different Categories of Over Time

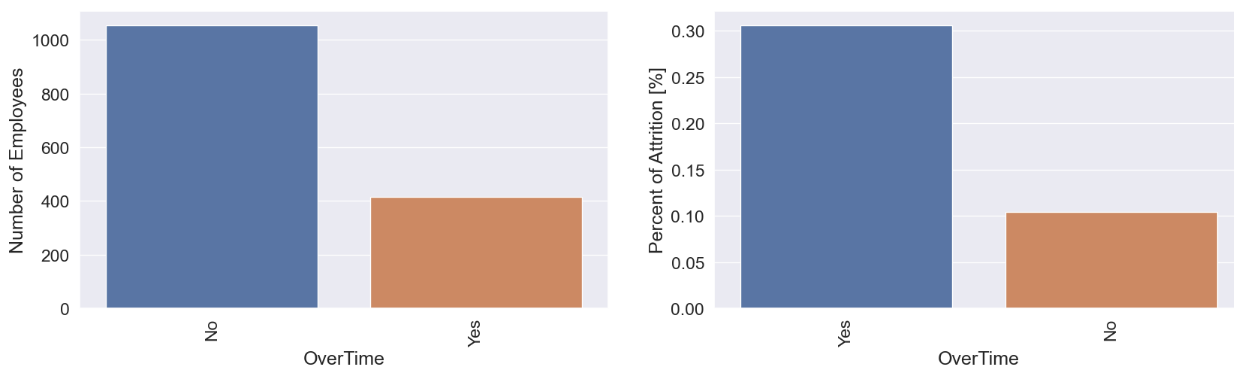

Figure 5.6 shows that single employees are more likely to leave the company than married and divorced employees. This could be due to several factors such as differing priorities and responsibilities. Single employees have more personal freedom, which leads to job changes.

Figure 5.8

Age distribution by Attrition

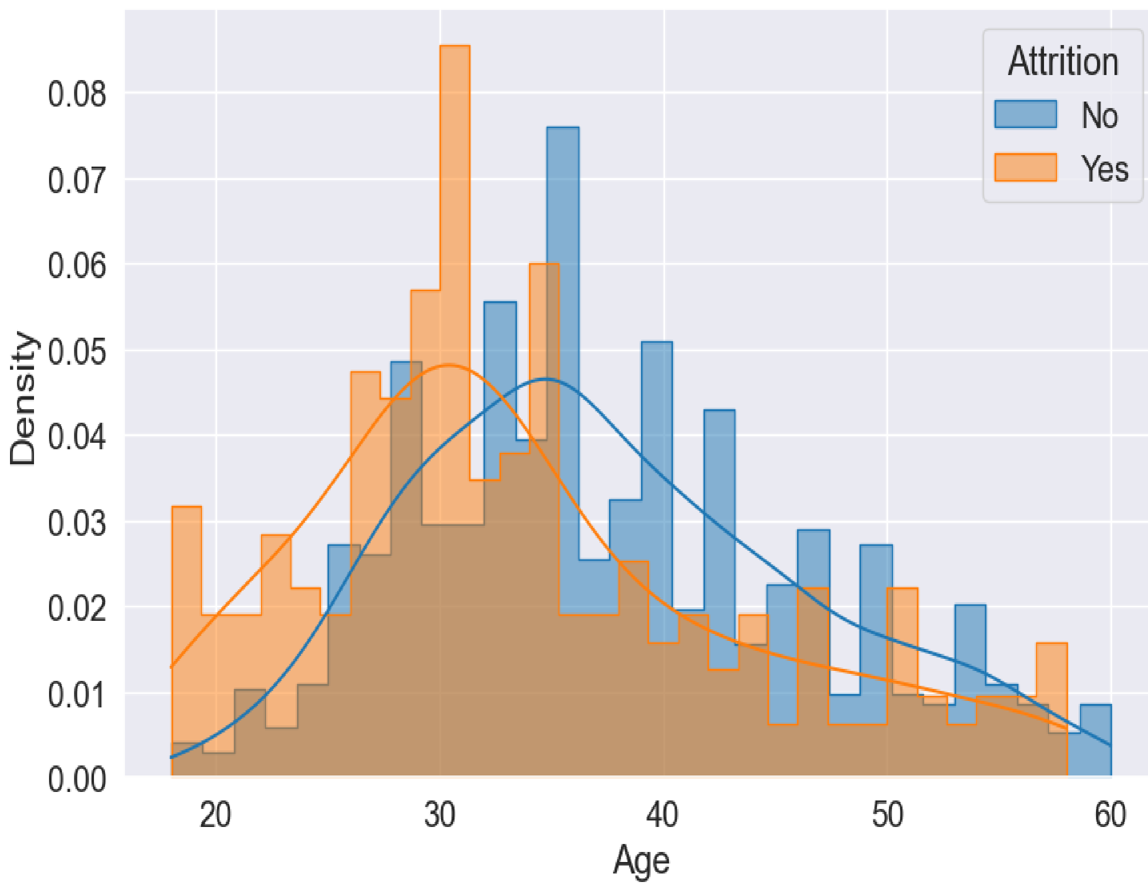

Figure 5.8 shows that the age range of around 25 to 35 years old stands out with the most remarkable attrition rate as compared to 40 years and older. Some of the reasons can be career development and advancement opportunities.

Figure 5.9

Daily Rate by Attrition

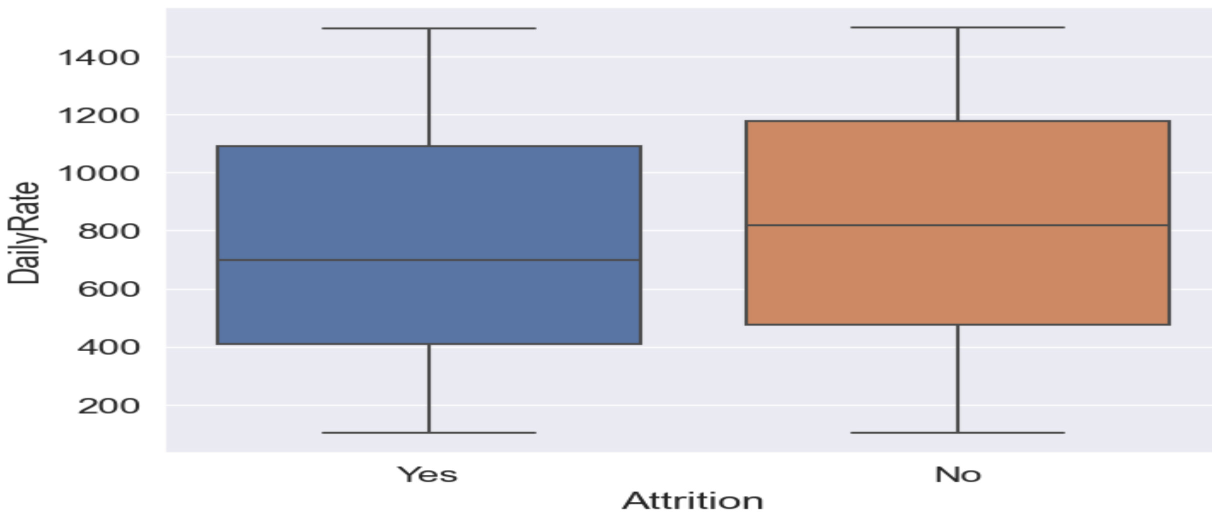

Figure 5.9 indicates that employees who are paid less per day are more likely to leave the firm than those who are paid more per day. Employees who believe they are underpaid for their skills, responsibilities, or market standards may be more likely to seek employment elsewhere.

Figure 5.10

Distance From Home to Workplace by Attrition

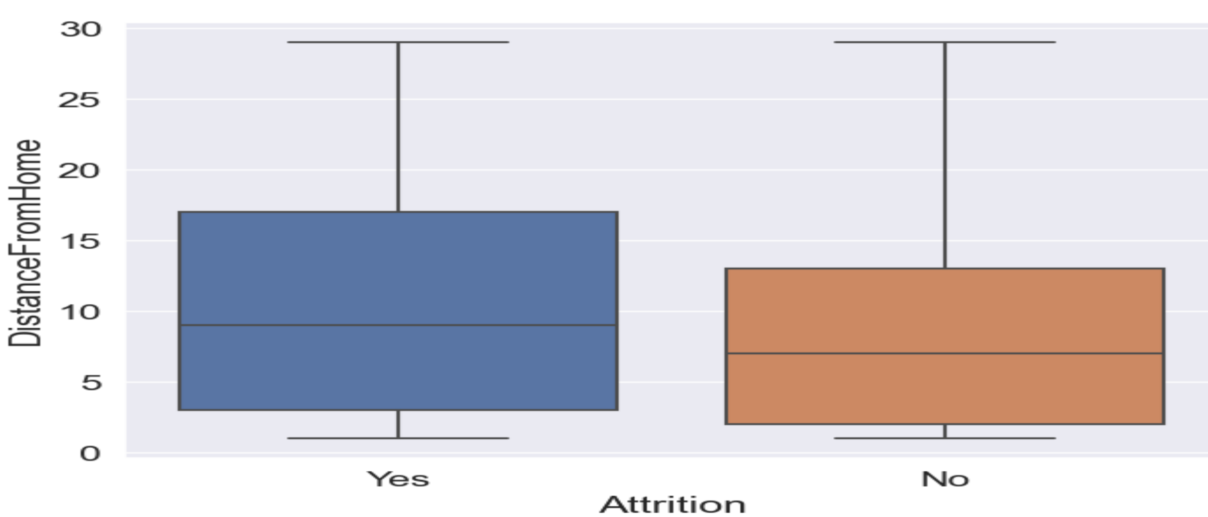

Figure 5.10 shows that employees who live far from work are more likely to leave. Commuting may affect job satisfaction. Long commutes can cause stress, fatigue, and work-life imbalance, which may make employees look for jobs closer to home.

Figure 5.11

Correlation Heatmap

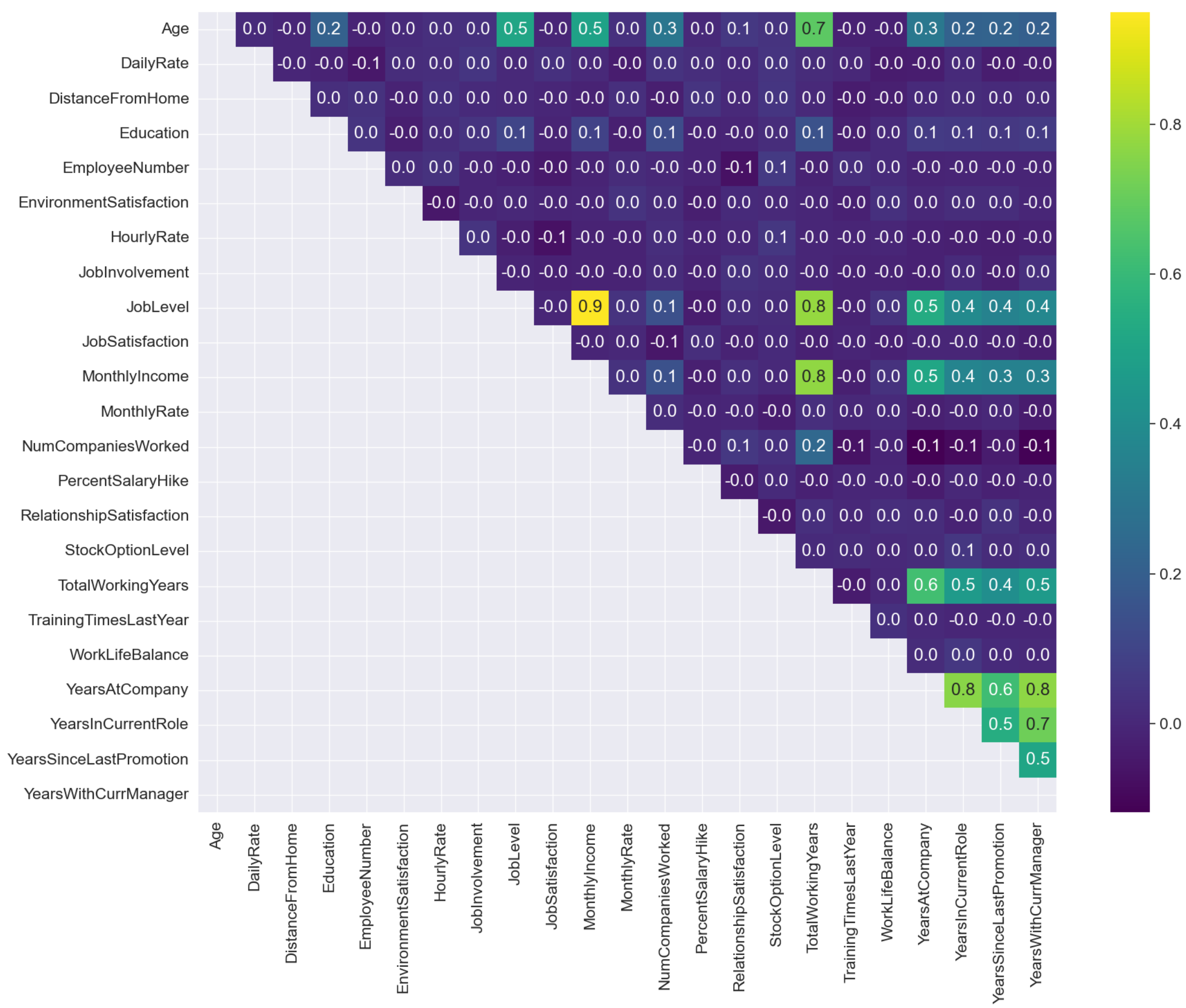

Figure 5.11 shows that education, age, job level, monthly income, number of companies worked, years at company, years in current role, years since last promotion, and years with the current manager are strongly correlated. These columns show multicollinearity among independent

variables, but I will use decision tree random forest and gradient boosting models that are unaffected by it.

6. Building Models

In this section, I converted categorical features to numerical features using label encoding to reduce features (Sree, 2021). Four models were fitted to training and testing sets. The table below presents the performance of each model after evaluation using the testing dataset.

Table 6.1

The Performance of Each Model

Model Name	Accuracy	ROC
Logistic Regression	89.12%	65.49%
SVC	88.78%	57.69%
Gradient Boosting	88.10%	61.64%
Random Forest	87.42%	54.74%
Naive Bayes	84.01%	73.41%

Table 6.1 shows that Logistic Regression predicts employee attrition very well as compared to other models. In contrast, Naive Bayes' high ROC score shows great predicting ability for identifying leaving employees. The Naïve Bayes model is appropriate for identifying employees who are likely to leave and those who are likely to stay since it has a ROC of 73.41%. Naïve Bayes effectively reduces false negatives hence making it ideal for predicting employee

leaves. Random Forest models are good in capturing complex relationships and managing multicollinearity hence can be best option since our data had multicollinearity.

Figure 6.1

Feature Importance

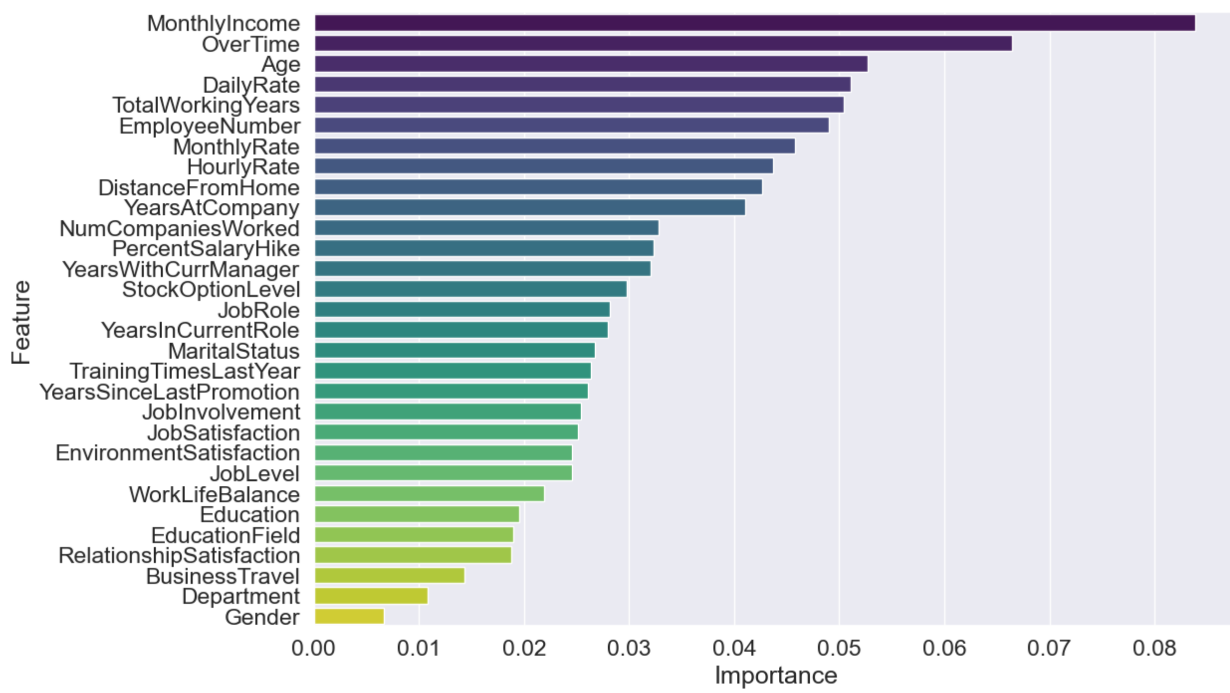

Figure 6.1 shows that monthly income, overtime, age, daily rate, total working years, number of employees, monthly rate, hourly rate, distance from home are top 11 features which influence the likelihood of leaving the company. The least importance features are gender, department, and business travel.

7. Recommendations

- Focus on Compensation and Benefits: MonthlyIncome, Overtime, and HourlyRate are crucial. Consider competitive salary packages and manage overtime work effectively to reduce attrition.
- To pay attention to the age factor: Younger employees may be more prone to leave, so strategies targeting this age group could be beneficial.
- Work-Life Balance and Satisfaction: TotalWorkingYears, JobSatisfaction, and WorkLifeBalance play roles. Ensure a healthy work-life balance, job satisfaction, and overall employee well-being.
- Individual Employee Engagement: EmployeeNumber and MonthlyRate suggest individual variations. Understand the needs and concerns of specific employees.
- Consider Retention Strategies: While BusinessTravel, Department, and Gender are less impactful, they should not be ignored. Evaluate the impact on a case-by-case basis and implement targeted retention strategies where needed.
- Best Machine Learning Model: The Random Forest model demonstrated good performance in predicting employee attrition. The Random Forest models are less prone to overfitting, ability to handle complex relationship, and manage multicollinearity since our data set had multicollinearity. Deploy this model as a key tool for retention strategy.

8. Conclusion

In conclusion, our project successfully addressed the challenge of employee attrition by developing predictive models and identifying key factors influencing attrition. The Logistic Regression model demonstrated high accuracy hence making it reliable for overall predictions. However, Naive Bayes outperformed in terms of ROC score hence it can minimize false negatives thus making it a suitable choice for identifying employees likely to leave. The feature importance analysis shown importance factors such as monthly income, overtime, age, and work-related aspects. My recommendations focus on compensation, considering age-related strategies, promoting work-life balance, and tailoring retention efforts based on individual engagement. Additionally, the Random Forest model emerged as the best machine learning tool because of its performance in handling complex relationships and multicollinearity. By implementing these findings, organizations can proactively address attrition challenges and encourage a more stable and productive workplace hence reducing financial losses.

References

Chhavi, K. (2022). *EBSCO Publishing Service Selection Page*. Web.p.ebscohost.com. https://web.p.ebscohost.com/

Emerald Insight. (2019). *Discover Journals, Books & Case Studies | Emerald Insight*. Emerald.com. https://www.emerald.com/insight/

Erkkilä, S. (2020). Managing voluntary employee turnover with HR analytics. *Lutpub.lut.fi*, *31*(211). https://lutpub.lut.fi/handle/10024/161686

Grotto, A. R., Hyland, P. K., Caputo, A. W., & Semedo, C. (2017). Employee Turnover and Strategies for Retention. *The Wiley Blackwell Handbook of the Psychology of Recruitment, Selection and Employee Retention*, *3212*(443), 443–472. https://doi.org/10.1002/9781118972472.ch21

Hasan, Md. N. (2017). Factors affecting employee turnover and sound retention strategies in business organization: a conceptual view. *Problems and Perspectives in Management*, *15*(1), 63–71. https://www.ceeol.com/search/article-detail?id=620104

Holtom, B. C., Mitchell, T. R., Lee, T. W., & Eberly, M. B. (2008). Turnover and Retention Research: A Glance at the Past, a Closer Review of the Present, and a Venture into the Future. *The Academy of Management Annals*, *2*(1), 231–274.

Sree, S. (2021, November 13). *Optimized Conversion of Categorical and Numerical Features in Machine Learning Models | IEEE Conference Publication | IEEE Xplore*. Ieeexplore.ieee.org. https://ieeexplore.ieee.org/abstract/document/9640967

Wilke, C. O. (2019). Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures. In *Google Books*. "O'Reilly Media, Inc." https://books.google.co.ke/books?hl=en&lr=&id=XmmNDwAAQBAJ&oi=fnd&pg=PP1&dq=Data+Visualization+In+this+section