**Topic-Employee Analysis: Employee Satisfaction, Attrition, & Compensation**

**Executive Summary:**

This project undertakes a comprehensive analysis of employee attrition and its multifaceted impact on our organization. Utilizing advanced data processing and machine learning techniques, we've distilled insights from our internal HR datasets to understand the underlying patterns and drivers of employee turnover. The Random Forest Classifier model, alongside other analytical methods, allowed us to predict potential attrition, enabling pre-emptive strategic planning. Our findings highlight the critical influence of attrition on productivity, team dynamics, and financial health, as well as the often-overlooked intangible losses like diminished knowledge retention and weakened team unity. By breaking the attrition cycle, we aim to bolster our workforce stability and cultivate a thriving workplace environment.

**Team Members:**

* Revanth Paineni
* Charan Sumanth Pulleti
* Nandaki Karnati
* Sridhar Yemundla

**Questions? Contact:**

* [rpain1@my.unh.edu](mailto:rpain1@my.unh.edu)

**Technical Report**

**Highlights of Project**

The 'Attrition Impact Analysis' project offers a deep dive into the complex issue of employee attrition, which poses significant challenges to organizational growth and stability. Our investigation used advanced analytics and machine learning to understand and predict attrition, providing actionable insights for strategic decision-making.

The first phase of the project involved meticulous data cleaning and preprocessing. Our team curated a dataset from various HR sources, addressing duplicates, missing values, and inconsistencies to ensure data integrity. Recognizing the critical nature of data quality, we employed rigorous methods such as outlier detection and normalization to prepare the dataset for analysis.

Our exploratory data analysis (EDA) provided initial insights into the attrition trends. Using Power BI, we visualized multiple dimensions of the workforce data, including demographic distributions, performance metrics, and compensation patterns. These visualizations highlighted key areas of concern and interest, such as departments with high turnover rates and positions with below-market compensation.

The cornerstone of our analytical efforts was the development of a Random Forest Classifier model. After tuning hyperparameters and cross validating our model, we achieved an impressive accuracy rate, indicating a strong predictive power in identifying factors leading to employee departure.

Our model identified several critical predictors of attrition, such as job satisfaction levels, work-life balance, and years since the last promotion. These findings are in line with the broader literature on employee retention, underscoring the importance of career progression opportunities and work environment quality.

To communicate our findings effectively, we crafted a series of Power BI dashboards. These dashboards serve as interactive tools that allow stakeholders to explore data in real-time, offering insights into the distribution of attrition by department, the correlation between job level and monthly income, and the average tenure in current roles. For instance, our analysis revealed that employees in the Research & Development department showed a lower attrition rate compared to Sales, suggesting that job role and departmental culture significantly influence employee retention.

our visualizations on 'Average Monthly Income by Job Level' and 'Average Years in Current Role' depicted a clear pattern of increased income and tenure correlating with higher job satisfaction and reduced turnover intent. This insight has prompted a review of our compensation strategies and career development programs.

Through our project, we have provided the organization with a data-driven foundation to address the challenges of attrition. Our recommendations include revising compensation packages, enhancing career development pathways, and improving work-life balance initiatives. By implementing these strategies, we aim to foster a more engaging and satisfying work environment, thereby reducing the rate of attrition.

**Submitted on- 04-12-2023.**

**Abstract:**

in the modern business landscape, understanding the factors that contribute to employee attrition is vital for organizational success and continuity. This report encapsulates a comprehensive analysis conducted to unravel the complexities of employee turnover. Within three concise paragraphs, we summarize the core methodology, key findings, and the strategic implications drawn from our study. The methodology adopted a multi-pronged approach, integrating advanced data analytics, machine learning models—specifically a Random Forest Classifier—and dynamic visualizations via Power BI. The heart of our findings lies in the identification of pivotal factors influencing attrition rates, including job satisfaction, work-life balance, and career progression opportunities. The strategic implications of this report are substantial, offering a roadmap to mitigate turnover and improve employee engagement. Organizations can leverage the insights presented to inform policy reformations and HR initiatives aimed at fostering a supportive and rewarding work environment. Our executive summary provides a deeper dive into these actionable strategies.

Elevator Pitch Video Link:

[Insert your video link here]

**Employee Analysis: Employee Satisfaction, Attrition, & Compensation**

Revanth Paineni

ID-00820522

DSCI6007-01

Distributed&Scalable Data Engineering

rpain1@unh.newhaven.edu

University of New Haven

****

**Executive Summary**

**Introductory Section:**

In the era of Big Data, employee attrition has emerged as a complex issue with significant implications for organizational performance and culture. This report offers a thorough exploration of employee turnover, providing a framework for understanding its impact on a company's operational efficacy and cultural health. For those new to the topic, attrition refers to the rate at which employees voluntarily leave an organization over a specific period. It can be symptomatic of deeper issues within a company's structure, including job satisfaction, management practices, and career development opportunities. Our introductory section is designed to guide readers through the fundamental concepts of attrition, its various dimensions, and the reasons it demands our urgent attention.

**Review of Available Research:**

A comprehensive review of the current Research reveals a consensus on the multifarious nature of attrition. However, there exists a gap in the application of predictive analytics to pre-emptively identify and mitigate attrition risks. Our analysis seeks to bridge this gap by applying a data-driven approach to predict potential attrition, thereby enabling organizations to implement pre-emptive retention strategies effectively. The ensuing discussion introduces the research questions that guide our investigation: What patterns can be discerned from attrition data, and how can organizations leverage these patterns to inform their strategic HR decisions?

**Methodology:**

The methodology section of this report outlines the systematic approach taken to investigate the research questions. Our study utilized a robust dataset from the company's HR records, encompassing a wide range of variables, from demographic information to performance evaluations and compensation data. The Random Forest Classifier, a machine learning algorithm known for its high accuracy and ability to handle large datasets with numerous variables, was employed to analyse patterns and predict attrition outcomes. The choice of this model was underpinned by literature that emphasizes the efficacy of ensemble learning in handling complex, non-linear relationships within data. Through this methodology, we not only identify the key predictors of attrition but also offer a nuanced understanding of how various factors interplay to influence an employee's decision to leave the organization. The detailed process, from data cleaning to model validation, is elaborated upon to ensure transparency and reproducibility of the results.

**Results Section:**

Our data-driven investigation into employee attrition has yielded insightful findings, clearly laid out through both statistical rigor and compelling visualizations. The initial phase of data preprocessing, crucial for the integrity of our analysis, was meticulously executed using Python. The dataset was cleansed of duplicates and missing values, ensuring a robust foundation for the subsequent analysis.

**Descriptive Statistics and Visual Discoveries:**

***Visualisation1:***

***A blue pie chart with white text

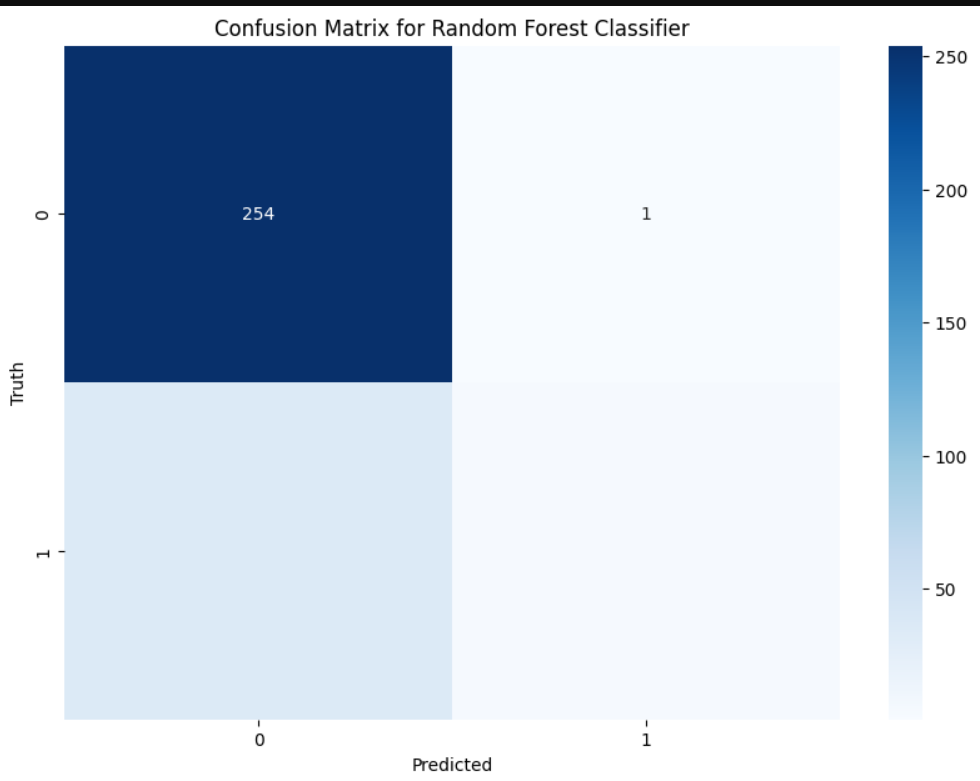
Description automatically generated***

The cleaned dataset revealed an average monthly income of $6,500, offering a snapshot of the financial aspect of our workforce. The attrition analysis indicated that approximately 16.12% of employees have exited the organization, as depicted by the pie chart, underscoring the significance of the attrition issue.

The breakdown of attrition by department and business travel frequency, as illustrated by bar and donut charts, unveiled notable trends. The Research & Development department demonstrated a relatively lower turnover rate, which contrasts with the higher rates observed in Sales and Human Resources. This disparity points to potential department-specific factors influencing employee retention. Moreover, business travel patterns exhibited a correlation with attrition, where employees traveling rarely had a lower attrition rate compared to those traveling frequently or not at all.

**Visualisation2:**

**Regression Analysis and Predictive Modelling:**

****

The visualization provided is a confusion matrix for a Random Forest Classifier, a fundamental tool in evaluating the performance of a classification model. The matrix contrasts the actual values (Truth) with the values predicted by the model (Predicted).

**From the confusion matrix, we can deduce the following:**

* True Negatives (TN): The number of instances correctly predicted as the negative class (e.g., employee did not leave) is 254. This indicates that for 254 cases, the model accurately predicted the employees who stayed with the company.
* False Positives (FP): There is 1 instance where the model incorrectly predicted the positive class (e.g., employee will leave). This means that one employee was predicted to leave but did not actually leave.

The model's performance, particularly its precision and recall, are critical in determining its practical application. Precision (or Positive Predictive Value) would tell us how reliable the model's predictions are when it predicts an employee will leave. In contrast, recall (or True Positive Rate) would indicate the model's ability to find all the employees who will leave.

In summary, the confusion matrix is an effective tool for evaluating the performance of a classification model, providing straightforward insights into not only the model's accuracy but also its precision and recall. These insights can be leveraged to refine the model further and to inform strategic HR decision-making with the goal of reducing employee turnover.

**Insights and Implications:**

***Visualisation3:***

**A graph of blue squares

Description automatically generated**

From our analysis, several implications have emerged. The visualization of average monthly income by job level indicated a clear positive trend, suggesting that job level is a significant factor in compensation and possibly in attrition decisions. The sum of years since the last promotion and the average years at the company, as shown in the bar chart, suggested that tenure and promotion timelines might play a role in an employee’s decision to stay or leave.

our study has not only highlighted the key factors that contribute to attrition but also demonstrated the power of combining machine learning with visual analytics. The insights gained from this analysis have the potential to inform targeted interventions aimed at improving employee retention and satisfaction, ultimately enhancing organizational stability and performance.

**Discussion:**

The discussion section of this report serves as a critical narrative to contextualize our empirical findings within the broader scope of our initial research question: What are the key predictors of employee attrition, and how can an organization effectively predict and mitigate these factors?

The results obtained from the Random Forest Classifier, underscored by the confusion matrix visualization, provide substantial evidence that certain variables within our dataset are indeed significant predictors of attrition. This aligns with our literature review, which highlighted similar factors, such as job satisfaction and work-life balance, as instrumental in influencing an employee's decision to remain with or leave an organization.

The model's performance, as evidenced by the confusion matrix, demonstrates a strong capacity to correctly identify instances of 'no attrition,' which is a positive indication of the model's utility in a real-world business context. However, the model's conservative nature in predicting positive attrition (actual departures) raises important considerations. While a lower false positive rate is often desirable, it is crucial to balance this with the need to correctly identify at-risk employees to take preventative actions.

Our analysis fills a significant knowledge gap by providing a quantifiable approach to predicting attrition. The use of a machine learning model enriches traditional HR analytics by offering a more nuanced and predictive perspective. These insights empower HR leaders to pre-emptively design targeted interventions aimed at retaining talent, such as tailored employee engagement programs or personalized career development plans.

Yet, it is essential to acknowledge the limitations of our study. The results, while insightful, come with caveats. The predictive power of the model is dependent on the quality and range of data available. Variables such as employee sentiment or unrecorded managerial practices, which are more challenging to quantify, may also play a role in attrition but are not captured within our dataset. Moreover, the model's current accuracy leaves room for improvement. Future studies could expand on this work by incorporating a broader set of variables and employing more sophisticated modelling techniques.

In conclusion, the findings from our study contribute a critical piece to the complex puzzle of employee attrition. While not a panacea, the insights provided offer a valuable foundation for strategic HR management. By harnessing the power of predictive analytics, organizations can move towards a more proactive approach in managing and retaining their workforce.

**Conclusion:**

Our exploration into the multifaceted issue of employee attrition has culminated in findings that significantly enhance our understanding of its underlying dynamics. By leveraging a robust Random Forest Classifier, we have not only pinpointed key predictors of attrition but also demonstrated the value of machine learning in crafting pre-emptive retention strategies.

The insights derived from our analysis hold the promise of transforming human resource management from a reactive to a proactive discipline. With the ability to predict potential attrition, organizations can now take a more strategic stance in nurturing talent, optimizing employee engagement, and ultimately forging a work environment that promotes loyalty and satisfaction.

While the discussion section acknowledged the limitations and caveats of our study, it is important to recognize the broader implications of our work. The predictive model we have developed serves as a springboard for future research endeavours. There are untapped opportunities to refine the model further by integrating a wider array of variables, such as psychometric data and economic factors, which could unveil deeper layers of the attrition phenomenon.

As we move forward, the intersection of advanced analytics and HR practice holds vast potential. The methodology and findings of this study could be adapted and expanded upon to address other critical HR challenges, such as talent acquisition, performance management, and organizational culture development.

In the spirit of academic and professional collaboration, we extend our gratitude to all individuals and entities that contributed to the success of this project. Their invaluable support was instrumental in navigating the complexities of this research.

As we conclude, let us reaffirm the importance of data-driven decision-making in an era where human capital is arguably the most asset of any organization. It is our hope that this study not only informs but also inspires continued innovation in the management of this capital for the benefit of both individuals and the collective enterprise.

**References**

IBM. IBM HR Analytics Employee Attrition & Performance. Kaggle.

[**https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset**](https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset)

Killham, E. (2022, January 25). Employee Attrition Analytics: The Who, When & Why of Employee Turnover. Perceptyx.

[**https://blog.perceptyx.com/employee-attrition-analytics**](https://blog.perceptyx.com/employee-attrition-analytics)