**SALIFORT MOTORS EMPLOYEE TURNOVER**

**Plan Stage**

- Audience: Senior Leadership Team at Salifort Motors

- Goal: Predict employee turnover based on various factors

- Impact: Improve employee retention, reduce costs associated with hiring and training new employees

- Questions: What factors influence employee turnover? Which employees are likely to leave?

- Resources: Dataset, Python, Machine Learning Libraries (scikit-learn, pandas, numpy)

- Deliverables: Prediction Model, Report detailing findings and recommendations

**Analyze Stage**

- Sufficient Information: Yes, the dataset contains relevant variables for predicting employee turnover

- EDA Steps: Descriptive Statistics, Visualizations (Histograms, Correlation Matrix), Handling Missing Data

- Hypothesis: Null Hypothesis - There is no relationship between the variables and employee turnover. Alternative Hypothesis - There is a relationship between the variables and employee turnover.

- Purpose of EDA: Understand the data, identify patterns, detect anomalies, test hypotheses

**Construct Stage**

- Unusual Variables: Check for outliers in the variables

- Vendor Groupings: Not applicable

- Visualizations: Bar Chart, Scatter Plot, Confusion Matrix

- Processes: Data Cleaning, Feature Engineering, Model Training

- Missing Data: Handle appropriately (Imputation, Deletion)

- Hypothesis Test: Formulate and conduct

- Regression Assumptions: Check if the data breaks the assumptions of the model

- Independent Variables: Select based on EDA and domain knowledge

- Model Performance: Validation Score, Cross-Validation

**Execute Stage**

- Initial Recommendations: Identify variables with high correlation with 'left' variable

- Anomalies: Check for outliers in the variables

- Strengthening Data: Additional data could include employee feedback, company policies, etc.

- Key Insights: Interpret EDA and model results

- Business Recommendations: Based on insights, suggest strategies to improve employee retention

- Other Questions: Identify areas for further investigation

- Sharing Visualizations: Use presentation tools, dashboards, reports

**KEY INSIGHTS**

Following rigorous exploratory data analysis, we find that a staggering 84.27% of our employees have tenure less than 5 years. This tells us that a large portion of our workforce is relatively new, which could influence their level of commitment and loyalty to the company. Interestingly, we find little effect of the department on whether or not employees leave or are retained. Most employees belong to the sales department, and the least number of employees are part of the management.

Next, we delve into the specifics of three distinct groups of employees. Group 1 consists of employees who work more than 170 hours per month, have a satisfaction level less than 0.5, and their last evaluation score is greater than 0.5. Despite working more hours, these employees have a lower satisfaction level and a higher last evaluation score compared to the overall population. This suggests that while long working hours might be a concern, a high last evaluation score can offset this. However, despite having a higher last evaluation score, these employees have a higher chance of leaving the company. This could be due to other factors not captured in this group, such as personal circumstances, lack of career progression, or issues with management.

Our investigation leads us to another intriguing finding. In group 1, only 1.01% of employees were promoted in the last 5 years. This is significantly lower than the overall average of 1.69%. This could suggest that employees in group 1 face challenges in advancing their careers within the company. This could be due to a variety of factors, such as lack of opportunities for growth, perceived unfairness in promotions, or issues with management.

We then turn our attention to group 2. This group includes employees who work less than 170 hours per month, have a satisfaction level less than 0.5, and their last evaluation score is less than 0.5. Despite working fewer hours, these employees have a lesser chance of leaving. This seems counterintuitive! There might be other factors at play that are not included in the current analysis. For example, employees in group 2 might have unique circumstances that make them more likely to stay, despite their lower satisfaction levels and last evaluation scores. This could also mean that workload is a single most important factor!

Finally, we look at group 3. This group includes employees who work less than 170 hours per month, have a satisfaction level greater than 0.5, and their last evaluation score is greater than 0.5. These employees get more projects and have a slightly higher chance of leaving. This suggests that while these employees are satisfied with their jobs, they might prefer to have fewer responsibilities, which could contribute to their decision to leave.

Modal salary was consistent across all groups and was low.

In conclusion, while high satisfaction levels can contribute to lower turnover, other factors such as career progression opportunities and workload can also play a significant role. Companies should consider these factors when developing strategies to reduce turnover. As we continue our journey through the world of corporate data, we hope these insights will serve as a guiding light, illuminating the path to a healthier, more stable workforce.

Logistic Regression model results: This implies that for each project added to the number of projects an employee has, there is a 6.8% increased odds of leaving the company and for every promotion within a five year period, there is a decreased chance (68.5%) of the employee leaving the company.

Random forest has shown a very high capacity of predicting the outcome variable. Area under curve is 96% for random forest, a much better improvement than the logistic regression model.

Evaluation metrics.

Accuracy: 98.61%

Precision: 98.75%

Recall: 92.80%

F1 Score: 95.68%

Based on the information provided about your Random Forest model and its results, here are some business recommendations and potential suggestions for your manager/company:

Business Recommendations:

1. **Improve Employee Satisfaction**: Since satisfaction level is a significant factor affecting employee turnover, invest in employee satisfaction programs. This could include improving working conditions, providing opportunities for professional growth, or enhancing communication channels.

2. **Encourage Project Participation**: Encouraging employees to participate in projects can increase their engagement and reduce the likelihood of them leaving the company. Consider creating more collaborative environments or offering incentives for active participation in projects.

3. **Optimize Work Hours**: If employees are spending excessive time at work, it could be indicative of burnout or dissatisfaction. Review workload management practices and consider implementing flexible work arrangements or work-life balance initiatives.

4. **Regular Performance Reviews**: Last score at the last performance review is a significant factor. Regular performance reviews can help identify areas for improvement early and provide timely feedback to employees, which can boost morale and job satisfaction.

5. **Promotion Opportunities**: While promotions contribute the least to employee retention, they are still important. Offering clear career paths and opportunities for advancement can motivate employees to stay with the company longer.

**Potential Recommendations for Manager/Company:**

1. **Data-Driven Decision Making**: Use the insights from the model to inform decision-making processes. For example, prioritize actions that have been shown to reduce turnover, such as improving job satisfaction or reducing work hours.

2. **Employee Development Programs**: Based on the model's findings, focus on developing programs that address the most critical factors for employee retention. This could involve training, mentorship, or resources to support personal growth.

3. **Performance Metrics**: Align company performance metrics with the model's findings to encourage behaviors that promote employee retention. For instance, rewarding employees for staying with the company or recognizing long-term contributions.

4. **Communication Strategies**: Develop strategies to communicate effectively with employees about the company's commitment to their well-being and career growth. This can help build trust and loyalty.

**Model Improvement:**

The model already shows a strong performance with a high Area Under the Curve (AUC) of 96%. However, there are always ways to potentially improve a model:

1. **Feature Engineering**: Investigate additional features that could be relevant to employee turnover. This could include demographic information, company culture, or employee benefits.

2. **Hyperparameter Tuning**: Optimize the hyperparameters of the Random Forest model to potentially improve its performance. This could involve adjusting the number of trees, the maximum depth of trees, or other parameters.

3. **Model Selection**: Experiment with other machine learning algorithms that might be better suited to the problem or offer different trade-offs in terms of bias and variance.

4. **Ensemble Methods**: Combine your Random Forest model with other models using ensemble methods to create a more robust prediction system.

5. **Addressing Class Imbalance**: If the dataset is imbalanced, techniques such as oversampling the minority class or undersampling the majority class can be applied to improve the model's performance on the minority class.

**Summary of Model Results:**

The model's results indicate that several factors, such as job satisfaction, number of projects, work hours, performance review scores, and years in the company, are significant predictors of employee turnover. Promotions have the least impact on retention rates. The model's high AUC and cross-validation score suggest that it is a reliable tool for predicting employee turnover and can be used to guide retention strategies.