Improving Employee Retention by Predicting Employee Attrition Using Machine Learning

Supported by:
Rakamin Academy
Career Acceleration School
www.rakamin.com

Rakamin Academy

Created by:
Bagus Ariobimo
bagusariobimo7@gmail.com
https://www.linkedin.com/in/bagus-ariobimo
https://github.com/riyouuyt

"A data enthusiast who has completed a course in this field and is ready to start his career. Have a excellent understanding in statistics, programming, and data processing. Proficient in using tools such as Python, R, and SQL. Able to collect, clean and analyze data with the necessary techniques. Have skills in data visualization and simple statistical modeling. Creative problem solver and has a passion for learning. Ready to contribute to data-driven projects and collaborate in teams. Committed to further developing data science skills and achieving significant results in data analysis."

Overview

Project Overview

Human resources (HR) is the key asset that needs effective management to help companies achieve their business goals. In this project, we are faced with an issue related to human resources within a company. Our focus is to understand how to retain employees to prevent the swelling of recruitment and training costs for new hires. By identifying the primary factors causing employees to leave, we can promptly address these concerns by creating relevant employee programs.

Project Objective

The objective of this project is to identify the key factors that influence employee retention within a fictional company. Through the analysis of historical employee data, including various attributes, resignation reasons, and tenure, we aim to:

- Understand the primary drivers of employee turnover within the organization.
- Provide actionable insights for the company to enhance employee retention.
- Develop predictive models to forecast potential resignations and proactively address employee concerns.

Goals

- Data Collection and Preparation: Gather historical employee data and preprocess it for analysis.
- Retention Insights: Identify key factors influencing employee retention and develop predictive models.
- Actionable Recommendations: Translate model results into actionable strategies for improving employee retention.

Data Understanding

Rakamin Academy

Project Overview

Human resources (HR) is the key asset that needs effective management to help companies achieve their business goals. In this project, we are faced with an issue related to human resources within a company. Our focus is to understand how to retain employees to prevent the swelling of recruitment and training costs for new hires. By identifying the primary factors causing employees to leave, we can promptly address these concerns by creating relevant employee programs.

Project Objective

The objective of this project is to identify the key factors that influence employee retention within a fictional company. Through the analysis of historical employee data, including various attributes, resignation reasons, and tenure, we aim to:

- Understand the primary drivers of employee turnover within the organization.
- Provide actionable insights for the company to enhance employee retention.
- Develop predictive models to forecast potential resignations and proactively address employee concerns.

Goals

- Data Collection and Preparation: Gather historical employee data and preprocess it for analysis.
- Retention Insights: Identify key factors influencing employee retention and develop predictive models.
- Actionable Recommendations: Translate model results into actionable strategies for improving employee retention.

Overview

Project Overview

Human resources (HR) is the key asset that needs effective management to help companies achieve their business goals. In this project, we are faced with an issue related to human resources within a company. Our focus is to understand how to retain employees to prevent the swelling of recruitment and training costs for new hires. By identifying the primary factors causing employees to leave, we can promptly address these concerns by creating relevant employee programs.

Project Objective

The objective of this project is to identify the key factors that influence employee retention within a fictional company. Through the analysis of historical employee data, including various attributes, resignation reasons, and tenure, we aim to:

- Understand the primary drivers of employee turnover within the organization.
- Provide actionable insights for the company to enhance employee retention.
- Develop predictive models to forecast potential resignations and proactively address employee concerns.

Goals

- Data Collection and Preparation: Gather historical employee data and preprocess it for analysis.
- Retention Insights: Identify key factors influencing employee retention and develop predictive models.
- Actionable Recommendations: Translate model results into actionable strategies for improving employee retention.

Data Understanding

- **`Username`:** The username or identifier of the individual employee.
- **`EnterpriseID`**: A unique numerical identifier for each employee within the enterprise.
- **`StatusPernikahan`:** The marital status of the employee, e.g., "Single," "Married," "Divorced," etc.
- **`JenisKelamin`:** The gender of the employee, e.g., "Male" or "Female."
- **`StatusKepegawaian`:** The employment status of the employee, e.g., "Full-time," "Part-time," "Contract," etc.
- **`Pekerjaan`:** The job or occupation of the employee, e.g., "Software Engineer," "Data Analyst," etc.
- **`JenjangKarir`:** The career level or stage of the employee within the organization.
- **`PerformancePegawai`:** Employee performance rating or evaluation, e.g., "Excellent," "Satisfactory," etc.
- **`AsalDaerah`:** The place of origin or hometown of the employee.
- **`HiringPlatform`:** The platform or method through which the employee was hired.
- **`SkorSurveyEngagement`:** A numerical score representing employee engagement based on survey responses.
- **`SkorKepuasanPegawai`:** A numerical score representing employee satisfaction, if available.
- **`JumlahKeikutsertaanProjek`:** The number of projects in which the employee has participated.

Data Understanding

- **`JumlahKeterlambatanSebulanTerakhir`:** The number of instances of tardiness by the employee in the last month.
- **`JumlahKetidakhadiran`:** The total number of days of absence from work by the employee.
- **`NomorHP`:** The employee's phone number.
- **`Email`:** The employee's email address.
- **`TingkatPendidikan`**: The highest level of education achieved by the employee, e.g., "Bachelor's Degree," "Master's Degree," etc.
- **`PernahBekerja`:** Indicates whether the employee has worked previously, e.g., "Yes" or "No."
- **`IkutProgramLOP`:** Indicates if the employee has participated in the "LOP" program, possibly with a numerical value for the number of times participated.
- **`AlasanResign`:** The reason for the employee's resignation, if applicable.
- **`TanggalLahir`:** The employee's date of birth.
- **`TanggalHiring`:** The date when the employee was hired.
- **`TanggalPenilaianKaryawan`:** The date when the employee's performance or evaluation was assessed.
- **`TanggalResign`:** The date of the employee's resignation.

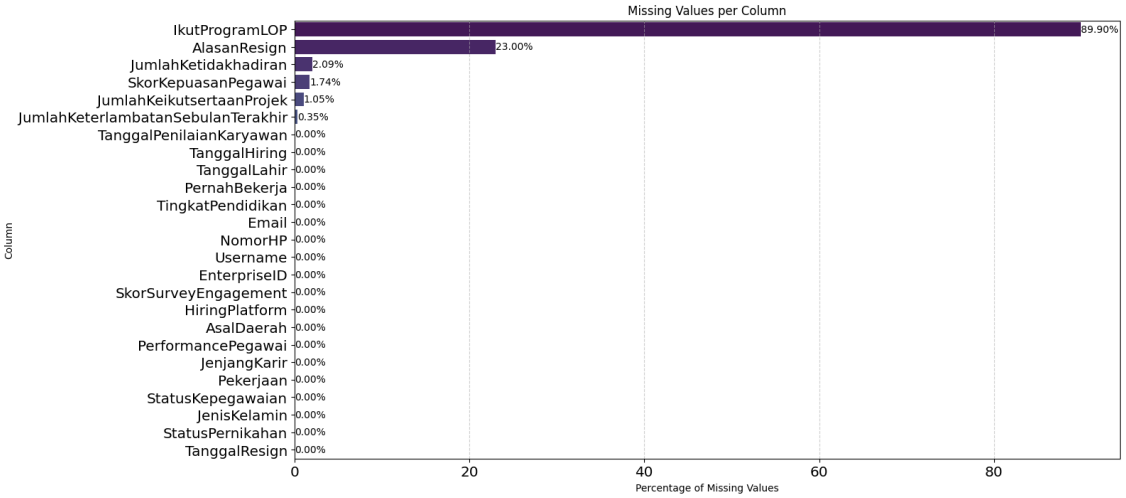
Data Preprocessing

Handling Missing Values:

- **SkorKepuasanPegawai:** we planning to fill missing values with the mean. This means we'll calculate the mean of the existing values in this column and replace missing values with that mean. This ensures that missing values are imputed with a central value that doesn't significantly affect the data's distribution.
- **JumlahKeikutsertaanProjek:** we'll fill missing values with the median. Similar to step 1a, this will provide a central value for the missing entries but using the median, which is robust to outliers.
- **JumlahKeterlambatanSebulanTerakhir:** we'll fill missing values with the mean. Again, this provides a central value for missing entries.
- **JumlahKetidakhadiran:** Rows with missing values in this column will be removed entirely. This is done when we decided believe there is no meaningful way to impute these missing values.

Handling Columns:

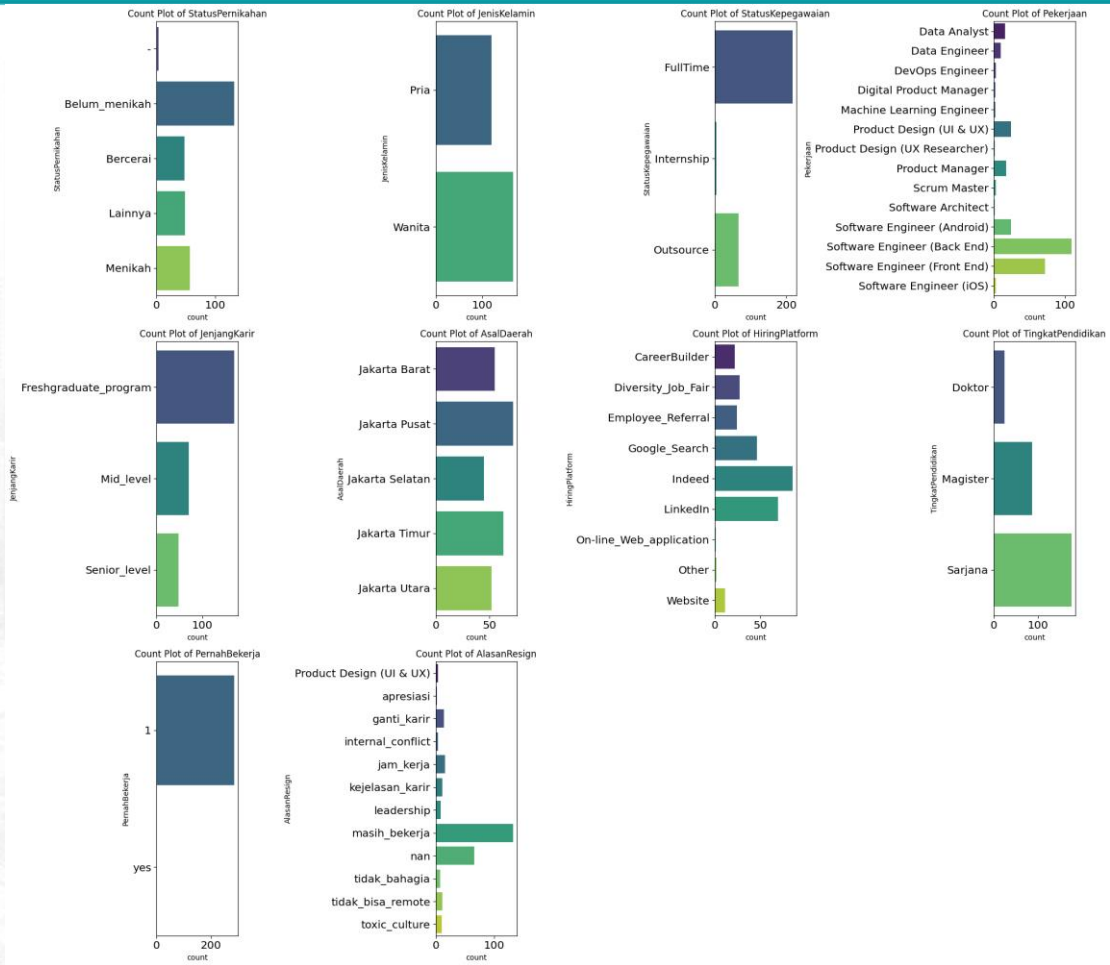
- **IkutProgramLOP:** we planning to remove this column entirely. This is done when a column doesn't provide valuable information for analysis or modeling.
- **AlasanResign:** Missing values in this column will be filled with "Dll" (lainnya). This helps in providing a placeholder for the missing data and avoids data loss.



Data Preprocessing

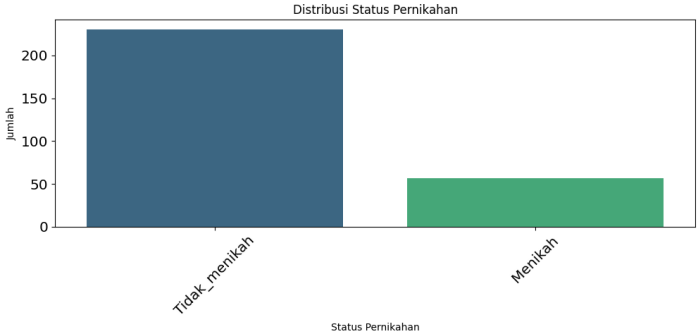
Handling Inconsistent Data:

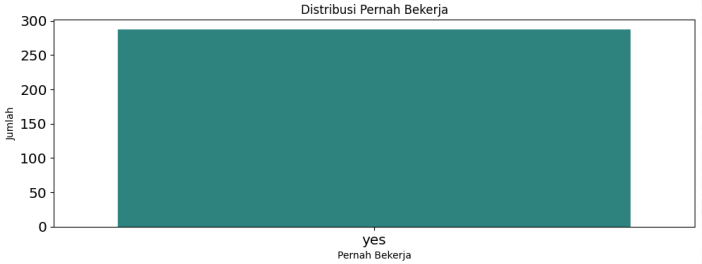
- **StatusPernikahan:** You intend to merge the value '-' into 'Lainnya'. This step helps in standardizing the data, making it more consistent for analysis.
- **PernahBekerja:** You plan to merge the value '1' into 'Yes', which standardizes the representation of this binary variable.
- **AlasanResign:** You intend to merge the value 'Product Design (UI & UX)' into 'Lainnya'. This again standardizes the data and groups similar categories together.

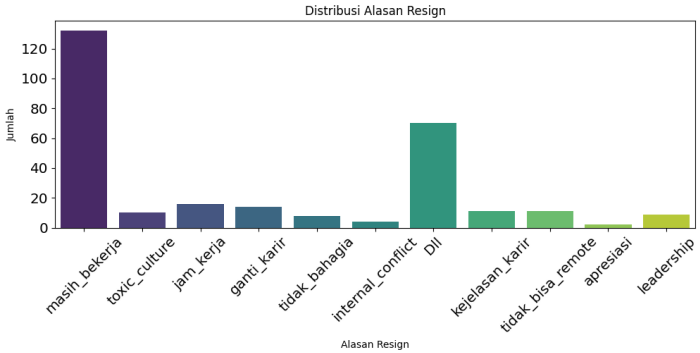


Data Preprocessing

After Handing Incosistent Data:







Data Preprocessing

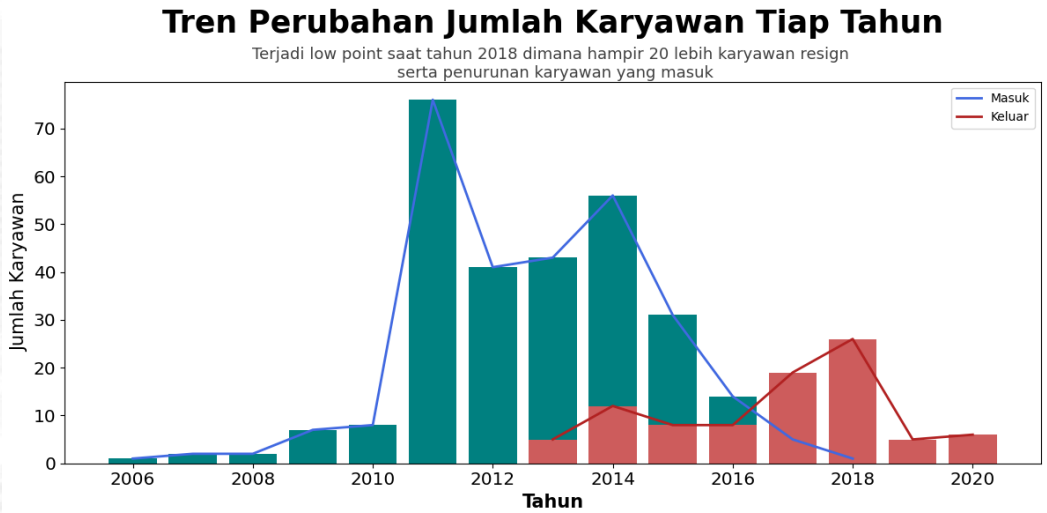
Feature Engineering 🛠:

a. MasihBekerja: You're creating a new binary column to indicate whether an employee is still working. This can be determined by comparing the resignation date with the current date or a specific cutoff date. If they haven't resigned, the value in this column is 'Yes'; otherwise, it's 'No'.

b. TahunHiring: You plan to create a new column to extract the hiring year from the 'TanggalHiring' column. This allows for easy analysis based on hiring years.

c. TahunResign: Similar to 'TahunHiring,' you'll create a new column to extract the resignation year from the 'TanggalResign' column.

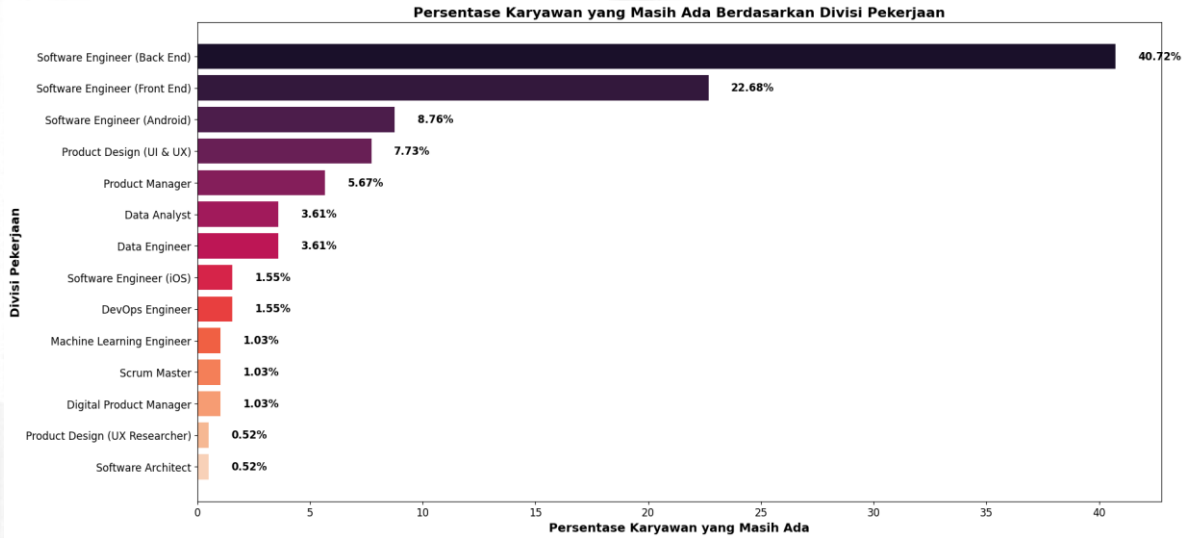
Annual Report on Employee Number Changes



Yearly Annual report on employee:

- **2006-2013:** During these years, the number of incoming employees is relatively low, and there is a gradual increase in outgoing employees. The total remaining employees are high, indicating that the company maintains its workforce.
- **2013-2015:** In 2013, a significant number of employees resigned, resulting in a sharp drop in the total remaining employees. This trend continues in 2014 and 2015. The increase in outgoing employees might be due to various factors such as job dissatisfaction, better opportunities elsewhere, or organizational changes.
- **2015-2018:** The number of incoming employees remains moderate, but the number of outgoing employees remains high. The total remaining employees continue to decrease, indicating a steady decline in the workforce.
- **2018-2019:** In 2018, the number of incoming employees is very low, while the number of outgoing employees is significantly higher. This leads to a drastic decrease in the total remaining employees. The trend of increasing resignations might be attributed to internal issues or external factors affecting job stability.

Annual Report on Employee Number Changes



Annual Report on employee based on work division:

- Software Engineering are the most dominant division on the company with total of **73,71%** of the percentage of the whole division shows how really are important Software Engineering roles in our company, with probably higher salaries and as a fundamental job on every IT division is really a good reason on why lots of Software Engineering are stil stay on our company.
- Followed by UI & UX with **7,73%** , While the rest division are below 6% show some assumptions that this maybe there are some structural hierarchy of importance role for the company like the software engineering.
- Beside the role of importance, higher education and experience can be a solid reason on why that happen like the DevOps, Scrum master, etc.

Resign Reason Analysis for Employee Attrition Management Strategy

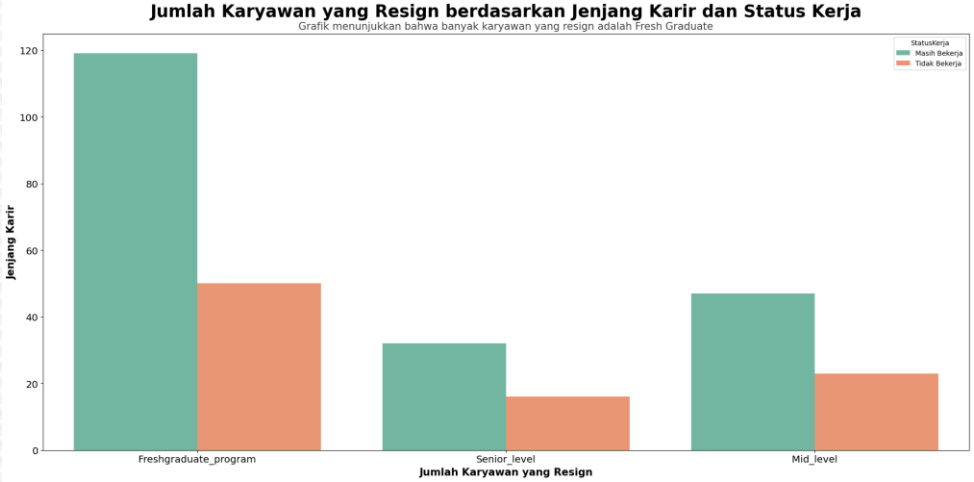
Total Employee Retention Based on Work status

Observation:

- **Fresh graduate programs** have consistently shown the highest retention rates in our company over the past few decades. Several plausible explanations for this trend include the possibility of new employees having false expectations, seeking job experience, or encountering job mismatches.

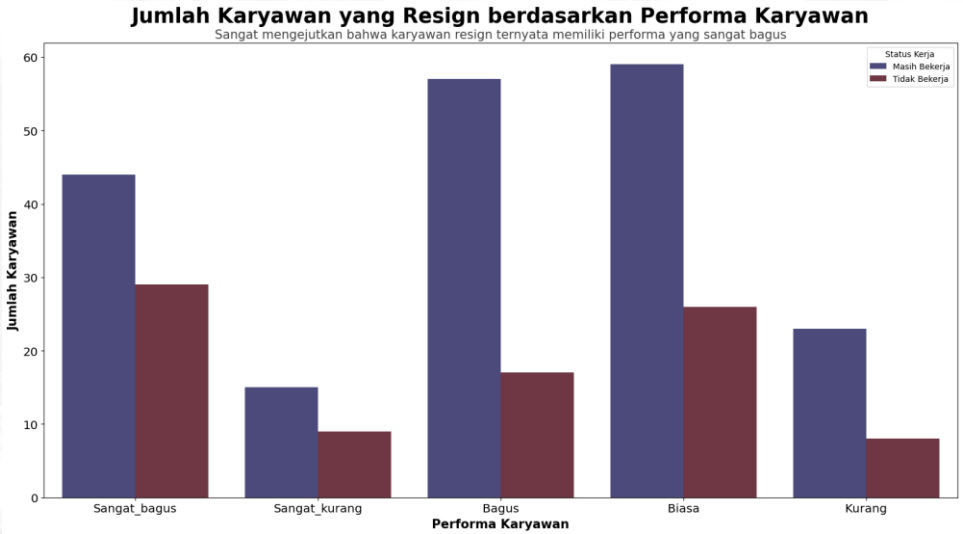
- **Mid-level employees**, with several years of experience, may consider leaving for better career growth opportunities in more senior roles. Factors like limited promotion prospects or lack of skill development within the company could influence their decision to resign.

- **Senior employees**, who have accumulated significant experience, may seek retirement, a change in career path, or entrepreneurial ventures. An increased desire for work-life balance or dissatisfaction with company policies may lead to higher resignation rates among senior staff.



For More Information About The Project, Click Here

Resign Reason Analysis for Employee Attrition Management Strategy

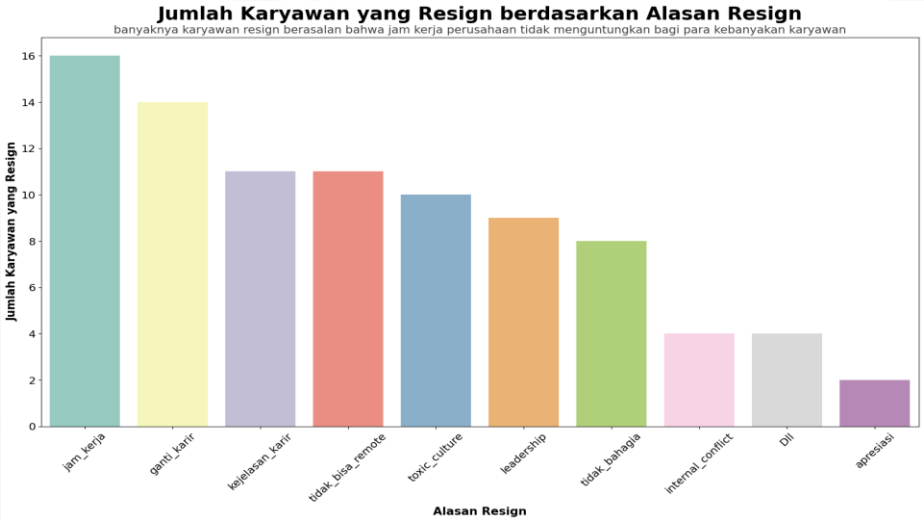


Total Employee Retention Based on Work Performance

Observation:

- **Very Good Performance**: Surprisingly, employees with consistently "very good" performance ratings show significantly higher resignation rates. While it's expected that top performers are retained, the occurrence of resignations in this group raises questions. This could be due to employees being headhunted for better opportunities or personal career development goals.
- **Mediocre Performance:** Employees with mediocre performance ratings have relatively high resignation rates, which may be attributed to a lack of motivation or challenges in their roles. The organization might consider addressing factors affecting the performance of this group.
- **Good Performance:** Employees with consistently good performance ratings show moderate resignation rates. It's crucial to explore the reasons behind these resignations, which could involve factors like limited career growth or a desire for more challenging roles.
- **Very Low Performance:** Employees with persistently low performance ratings also experience high resignation rates. Poor performance is often linked to unsuitability for roles or a lack of job satisfaction. Addressing these underlying issues is essential to improve retention.
- **Low Performance:** Those with consistently low performance have the highest resignation rates. It is vital for the organization to understand the reasons behind the poor performance and resignations, which might include job-role mismatches, lack of motivation, or other workplace challenges.

Resign Reason Analysis for Employee Attrition Management Strategy



Total Employee Retention Based on Resign Reason

Observation:

- **Working Hours (Jam Kerja):** A significant number of resignations being attributed to working hours may indicate that employees are struggling with work-life balance. Long working hours, excessive overtime, or inconsistent schedules can lead to burnout and negatively impact an individual's quality of life. When work demands become overwhelming and unsustainable, employees may decide to seek positions with more reasonable working hours.

- **Career Change (Ganti Karir):** This reason could suggest that employees are either seeking career growth or transitioning to different fields that align better with their long-term aspirations. It could also reflect that employees are not finding opportunities for advancement or skill development within the company. When such opportunities are limited, employees opt for external career changes.

- **Career Clarity (Kejelasan Karir):** A lack of career clarity within the organization might be leading to employee dissatisfaction. When employees are uncertain about their career paths, promotion criteria, or professional development opportunities, it can create frustration and hinder their commitment to the company. As a result, they may decide to pursue opportunities that provide clearer career trajectories and personal growth prospects.

For More Information About The Project, Click Here

Build an Automated Resignation Behavior Prediction using Machine Learning

Rakamin Academy

📊 Data Transformation and Enrichment Journey 🚀

In our quest for data-driven insights, we've meticulously transformed and enriched our dataset, breathing life into it! Let's explore the steps we've taken:

- **StatusKerja - The North Star** ✨: We introduced the compass of "StatusKerja" to help us navigate the employee universe. This column reveals whether an employee is "Masih Bekerja" or "Tidak Bekerja."
- 🧠 **Age Categories - Shaping the Generations** 🔍**:** We ingeniously categorized our workforce into age groups, providing insights into generational dynamics. Understanding age groups is essential for tailored HR strategies.
- 📅 **Date Extraction - Unveiling Temporal Insights** ⏳: We've harnessed the power of date extraction from columns like **"TanggaLahir,"** **"TanggaHiring,"** and more. This allows us to uncover trends and patterns over time.
- 🔠 **Encoding Magic - Transforming Words into Numbers** 🔢: We've employed one-hot encoding for categorical variables, such as **"StatusPernikahan,"** **"JenisKelamin,"** and more. This transforms text-based categories into numerical values for our models.
- 🏷️ **Label Encoding - Translating Insights into Action** 🚩**:** Label encoding breathed life into categorical variables like **"Pekerjaan,"** **"AlasanResign,"** and **"PerformancePegawai."** We assigned unique labels to each category for deeper analysis.

Our data transformation and enrichment journey equips us with the tools to explore and analyze our workforce data comprehensively. Now, let's set sail towards data-driven insights and HR strategies! 🚢✨

Build an Automated Resignation Behavior Prediction using Machine Learning

Rakamin Academy

Data Split and Feature Selection Analysis

We embarked on our data journey with a total of 287 employee records. For robust model training and evaluation, we decided to divide our dataset into training and testing subsets.

- Total Samples: 287
- Training Samples: 229 (79.79%)
- Testing Samples: 58 (20.21%)

To unlock valuable insights, we employed various machine learning models, each revealing its unique strengths.

Model Comparison:

No	ML Method	Accuracy	Precision	Recall	F1-score	ROC
1	**Support Vector Machine**	0.67	0.0	0.0	0.0	0.5
2	**Gradient Boosting**	0.95	0.9	0.95	0.92	0.95
3	**Decision Tree**	0.93	0.86	0.95	0.9	0.93
4	**Random Forest**	0.91	1.0	0.74	0.85	0.87
5	**Linear Regression**	0.67	0.5	0.11	0.17	0.53

In light of its remarkable performance, we've chosen **Gradient Boosting** as our preferred model. With an accuracy of **94.83%** and an AUC score of **94.80%,** it demonstrates an exceptional ability to predict employee retention.

Build an Automated Resignation Behavior Prediction using Machine Learning

Rakamin Academy

Cross-Validation and Hyperparameter Tuning Analysis

Through Cross-Validation, we fine-tuned our model and gauged its performance on our employee retention prediction task.

- **Mean Precision**: 0.97 (±0.05)
- **Mean Recall:** 0.96 (±0.07)
- **Mean ROC-AUC**: 0.97 (±0.03)

These metrics are a testament to the reliability and accuracy of our model. It effectively identifies employees at risk and exhibits robust predictive capabilities.

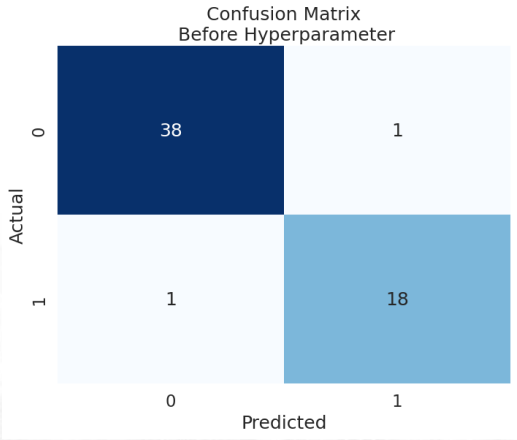
Hyperparameter Optimization:

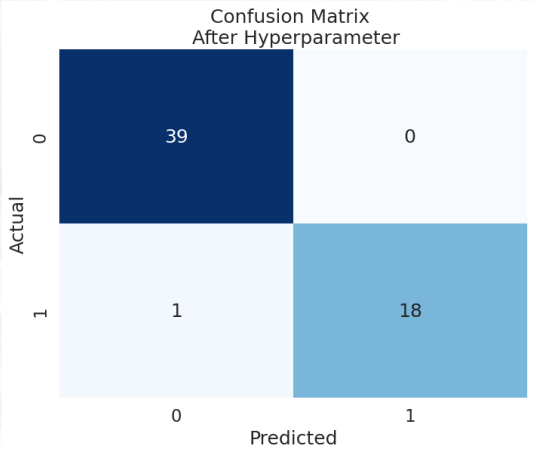
To take our model to the next level, we optimized its hyperparameters. The result? An exceptional AUC score of 1.00!

Best Hyperparameters:

- **Learning Rate:** 0.01
- **Max Depth:** 4
- **Min Samples Leaf:** 5
- **Min Samples Split:** 3
- **Number of Estimators:** 50
- **Subsample:** 0.8

These parameters allow our model to perform at its peak and confidently predict employee retention.





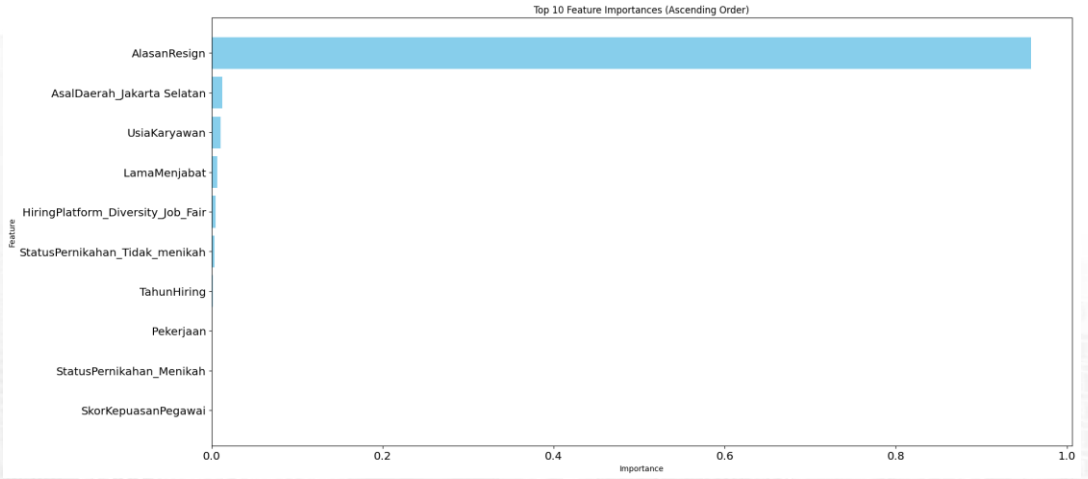
Build an Automated Resignation Behavior Prediction using Machine Learning

Rakamin Academy

Feature Importance Analysis:

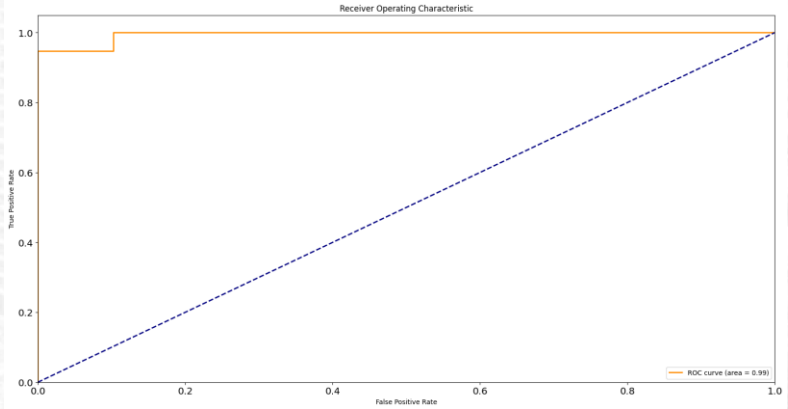
- **AlasanResign** 🔑**:** The reason for resigning is a crucial factor that affects employee retention. Understanding the specific reasons why employees leave can help organizations make targeted improvements to their work environment, which can ultimately reduce employee turnover.

- **AsalKota Jakarta Selatan** 🏢**:** The location from which employees come can be an important factor. In this case, employees from South Jakarta might have different commuting experiences or preferences that could impact their job satisfaction and, in turn, their likelihood of resigning.



ROC Curve Analysis:

The ROC curve for the model achieved an outstanding area under the curve (AUC) of 0.99, which means the model has a high ability to distinguish between employees who are likely to resign and those who are likely to stay. This high AUC score is a promising sign that the model is effective in identifying potential resignations, which can be invaluable for employee retention strategies. 🚀☑



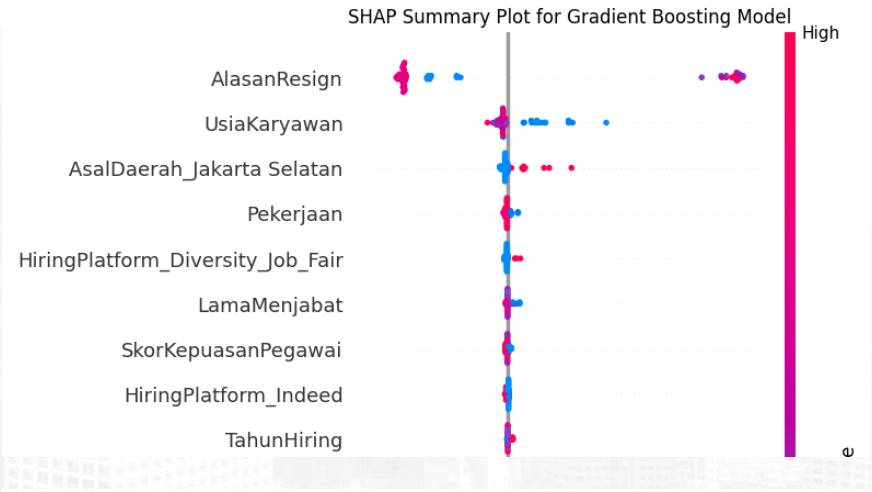
Presenting Machine Learning Products to the Business Users

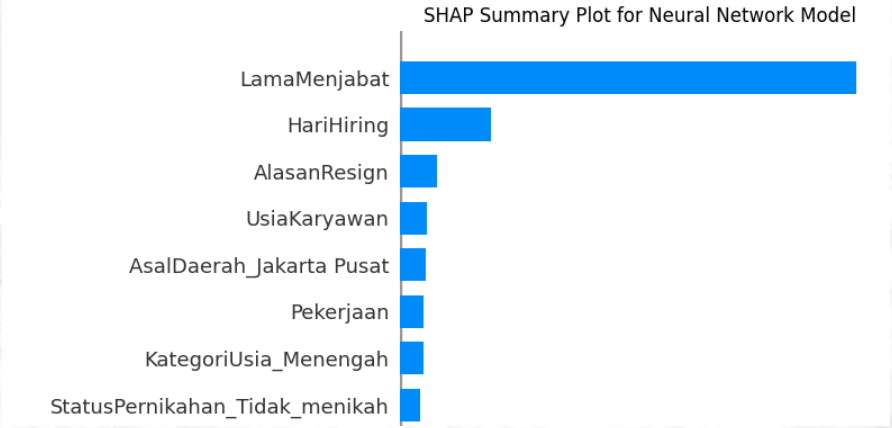
Rakamin Academy

SHAP Values Analysis

Both the Gradient Boosting and Neural Network models agree on the importance of factors such as **"AlasanResign"** and **"UsiaKaryawan"** as determinants of employee retention. Resignation reasons should be closely monitored, and employee tenure needs to be nurtured and rewarded. However, the neural network emphasizes the significance of **"LamaMenjabat"** and **"HariHiring,"** indicating that nurturing employee growth and optimizing onboarding processes are essential for retention.

Incorporating insights from both models can lead to a more comprehensive approach to employee retention. By addressing these key features, companies can proactively reduce turnover, boost employee satisfaction, and ultimately enhance their business performance.





Business Simulations

Business Simulation

Company Name Profile: **InnovateTech Solutions**

InnovateTech Solutions, a cutting-edge tech company, has successfully developed an Employee Retention Enhancement System (ERES) to address the ever-persistent issue of employee turnover. Using advanced machine learning techniques, ERES offers a comprehensive solution to predict and improve employee retention.

Storytelling

InnovateTech Solutions' journey towards developing ERES began with a mission to empower companies to reduce employee turnover and foster a more stable, productive workforce. Through years of research and development, our team of data scientists and engineers meticulously crafted ERES to offer insights and predictive power like never before.

Business Simulation

The insights we have gained from ERES are incredibly valuable for businesses in several ways:

- **AlasanResign (Resignation Reasons)**: By analyzing employee resignations, ERES identified that certain reasons significantly contribute to attrition. Addressing these concerns, such as job dissatisfaction or salary concerns, can directly impact retention rates. The key is to identify and address these issues proactively.

- **UsiaKaryawan (Employee Tenure)**: ERES revealed that employee tenure plays a crucial role in retention. It suggests that implementing programs or policies aimed at retaining long-term employees is essential. This could include offering career development opportunities, mentorship programs, or performance recognition for loyal employees.

- **AsalDaerahxJakartaSelatan (Originating from South Jakarta)**: The data showed that employees originating from South Jakarta tend to have higher retention rates. This could indicate that local hires may have stronger ties to the community or are more comfortable with the work environment. Companies could explore building stronger local connections to improve overall retention.

- **LamaMenjabat (Tenure in the Current Role)**: The neural network model placed high importance on the tenure in the current role. This suggests that job roles and responsibilities should be periodically reviewed and adjusted to keep employees engaged and motivated.

- **HariHiring (Day of Hiring)**: The day of the week when an employee is hired has an impact on retention. Further exploration into this variable may lead to insights about the onboarding process and how to make it more effective.

Business Simulation

Rakamin Academy

Business Recommendations ✨

Based on these findings, InnovateTech Solutions recommends the following strategies for businesses looking to enhance employee retention:

- **Resignation Reason Analysis**: Conduct regular exit interviews to understand why employees are leaving the company. Address the root causes, such as job satisfaction, compensation, or working conditions.
- **Tenure-Based Programs**: Develop initiatives that cater to employees at different stages of their careers within the organization. This includes mentorship programs, skill development opportunities, and a clear growth path.
- **Community Building**: Encourage stronger local connections within the workplace. Consider local events, community involvement, or social initiatives to strengthen the sense of belonging.
- **Regular Role Evaluation**: Periodically evaluate and update job roles and responsibilities to keep employees engaged and motivated. Ensure that the tasks align with their skills and interests.
- **Optimize Onboarding**: Investigate the significance of the day of hiring and fine-tune the onboarding process accordingly to create a more positive and lasting impression.

InnovateTech Solutions' Employee Retention Enhancement System (ERES) is the answer to reducing turnover, improving employee satisfaction, and ultimately boosting business performance. Let us help you unlock the full potential of your workforce for a brighter future.

Thank You For Participating✨