

Improving Employee Retention by Predicting Employee Attrition Using Machine Learning

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



Created by:

Philipus Dima Wira Pratomo

pwirapratomo@gmail.com

<https://www.linkedin.com/in/philipus-dima-wira-pratomo-a221391ab/>

Graduated from Pertamina University majoring in petroleum engineering with a GPA of 3.78. Currently I want to shift career to data scientist by attending a data science bootcamp. where I am developing expertise using data visualization tools such as Tableau and Looker Studio, SQL and several Python libraries to create machine learning. My objective is to become a professional data scientist and leverage my skills to contribute to data-driven decision-making in diverse industries.

"Human resources (HR) are the primary assets that need to be effectively managed by companies to achieve their business objectives efficiently. In this opportunity, we will be addressing an issue regarding the human resources within the company. Our focus is to understand how to retain employees in the current company, which can result in the swelling of costs for recruiting and training new hires. By identifying the key factors that cause employees to not feel engaged, the company can promptly address these issues by creating relevant programs tailored to the employees' concerns."

Handling Missing Value

- Handling missing values will be carried out on Mode in **IkutProgramLOP, AlasanResign, JumlahKetidakhadiran, SkorKepuasanPegawai, JumlahKeikutsertaanProjek, JumlahKeterlambatanSebulanTerakhir** Features

Handling Type Data

- **SkorKepuasanPegawai, JumlahKeikutsertaanProjek, JumlahKeterlambatanSebulanTerakhir, JumlahKetidakhadiran, IkutProgramLOP** can be change to INT64
- **TanggalLahir, TanggalHiring, TanggalPenilaianKaryawan, TanggalResign** can be change to DataFrame

Handling Incompatible Data

- The phrase "yes" in "PernahBekerja" can be changed to 1
- The phrase "-" in "StatusPernikahan" can be changed to "Lainnya"
- The phrase 'Product Design (UI & UX)' in "AlasanResign" can be changed into modus in "AlasanResign"

Feature Engineering

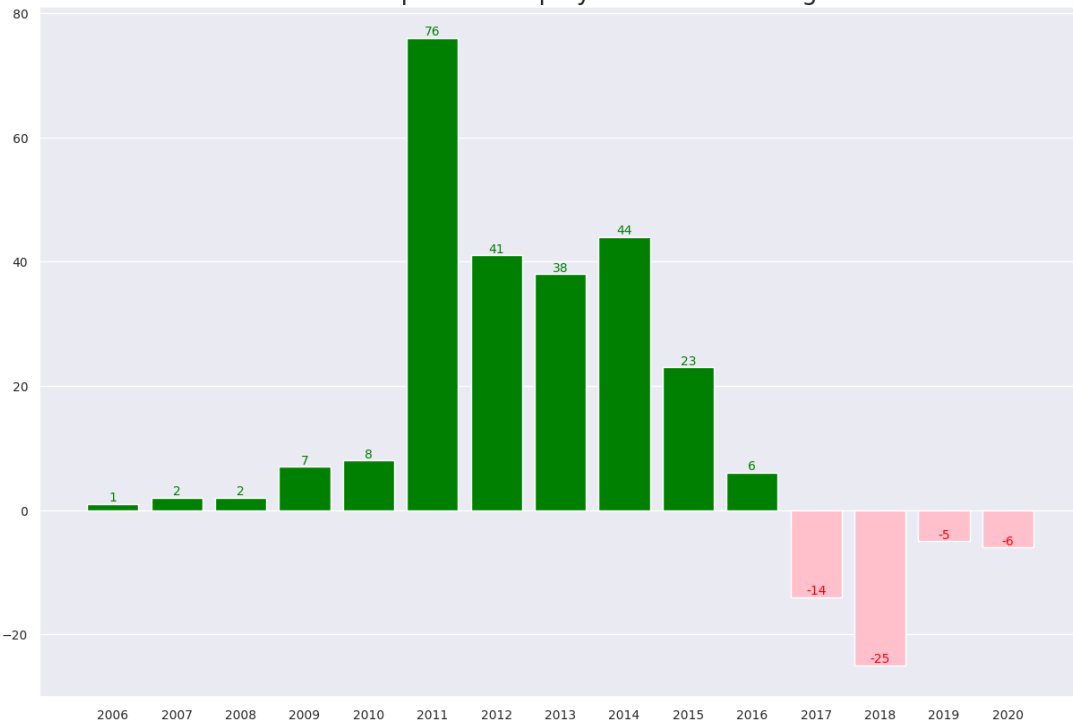
- Adding new features: TanggalLahir, TanggalHiring, TanggalPenilaianKaryawan and TanggalResign to extract the year.
- Creating a new feature where if the ResignationDate is null, it represents 'no,' while if there is a date, it signifies 'yes.'

Delete Unnecessary Feature

- The feature that will be deleted is because it only contains 1 data :
PernahBekerja

Annual Report on Employee Number Changes

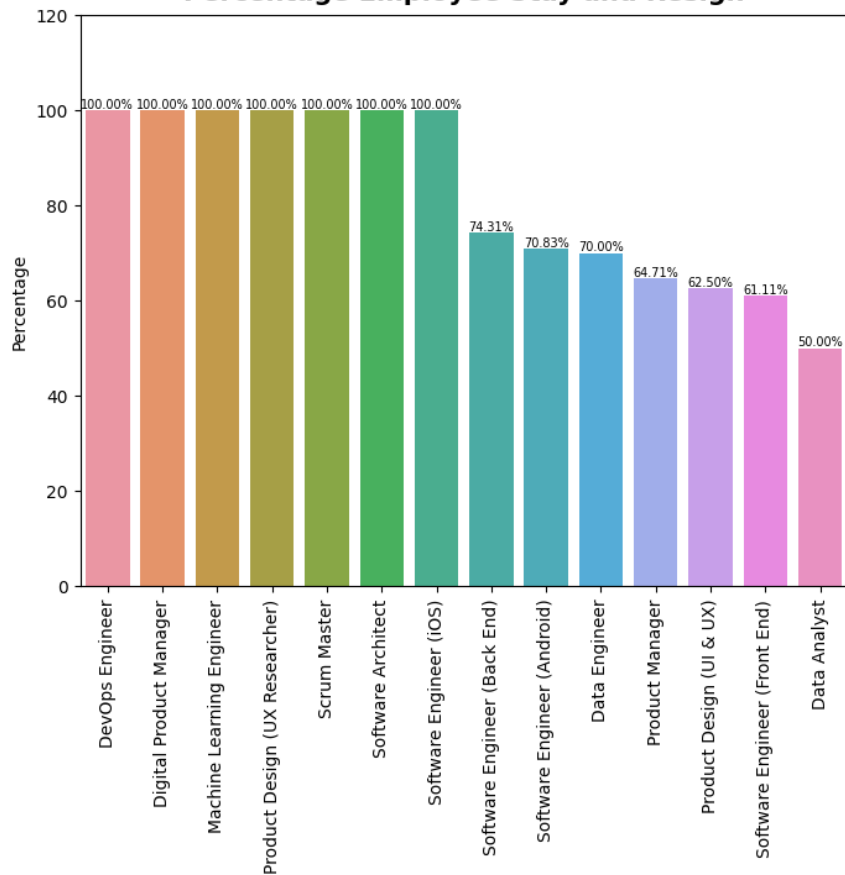
Annual Report on Employee Number Changes



INSIGHTS

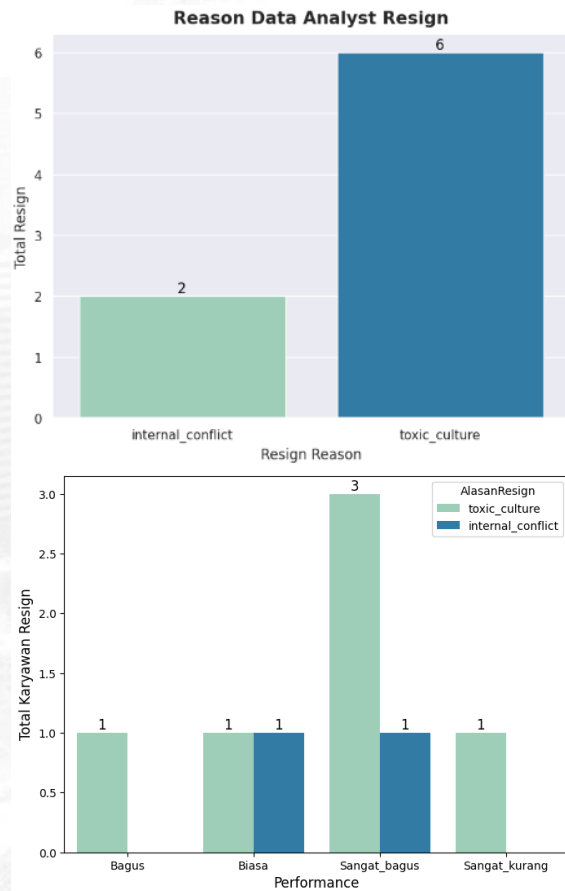
- In 2011, the number of employees who entered was the highest compared to those who left, namely 76
- In 2015, many employees resigned until the following year. Further analysis needs to be done as to why many employees resigned in that year.
- Employees continued to decline from 2015 onwards, peaking in 2018, far more employees left than entered
- In that range year the company was in a bad condition so it needed to improve itself so that other employees could survive so that the long-term strategy design could still be implemented without problems.

Percentage Employee Stay and Resign



INSIGHTS

- The following data represents the overall percentage of employees within each existing department, in comparison to those who have resigned, up until the year 2018.
- Notably, none of the employees in the roles of DevOps Engineer, Digital Product Manager, Machine Learning Engineer, Product Design, Scrum Master, and Software Architect have submitted their resignations.
- Data Analysts stand out as the department with the highest resignation rate, with up to 50% of its employees having resigned.
- It is essential to delve deeper into the underlying reasons behind the substantial turnover among Data Analysts.



INSIGHTS

- The graph illustrates the correlation between the reasons for resigning and the performance of data analysts.
- It's worth noting that all the data analyst employees who resigned were recent graduates.
- Interestingly, despite having outstanding performance, the presence of a toxic culture compels employees to resign.
- An in-depth examination of why data analyst employees perceive the work environment as toxic is warranted.

Split Data Test and Train

- 80% data Train and 20% Data Test

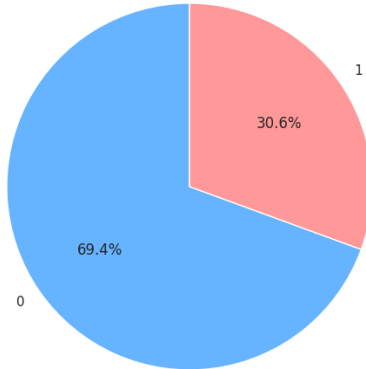
Label Encoding and One Hot Encoding

- There will be use label encoding in 'StatusKepengawain','Pekerjaan','JenjangKarir','AlasanResign'
- There will be use one hot encoding in 'PerformancePegawai'

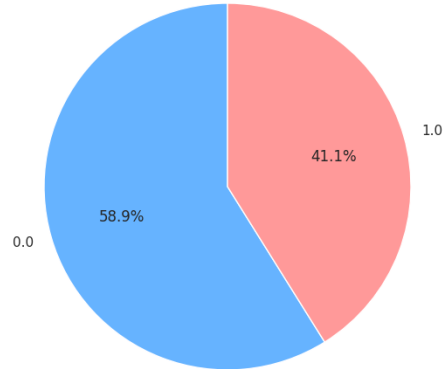
Class Imbalanced

- Using SMOTE to balanced

Before SMOTE



After SMOTE

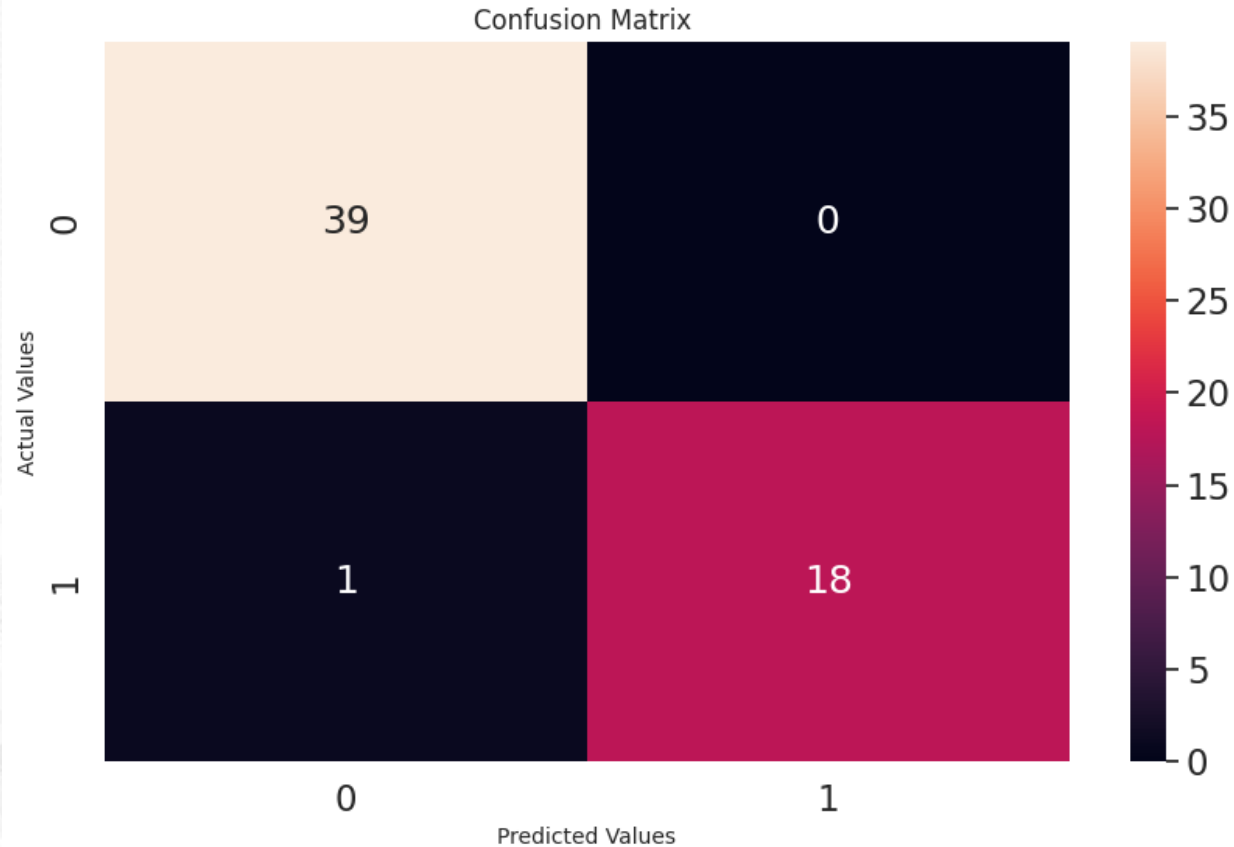


Selected Model

Model	Gradient Boosting	Random Forest	Decision Trees	SVM	KNN	XGBoost
Accuracy Train (%)	100%	100%	100%	99%	93%	99%
Accuracy Test (%)	98%	98%	97%	98%	79%	98%
Precision Train (%)	100%	100%	100%	100%	96%	100%
Precision Test (%)	100%	100%	95%	100%	77%	100%
Recall Train (%)	100%	100%	100%	96%	86%	97%
Recall Test (%)	95%	95%	95%	95%	53%	95%
F1 Score Train (%)	100%	100%	100%	98%	91%	99%
F1 Score Test (%)	97%	97%	95%	97%	63%	97%
ROC AUC Train (%)	100%	100%	100%	98%	92%	99%
ROC AUC Test (%)	97%	97%	96%	97%	72%	97%
CV Accuracy (%)	97%	99%	98%	99%	85%	99%
CV Precision (%)	97%	99%	97%	100%	91%	100%
CV Recall Test(%)	96%	97%	97%	96%	70%	97%
CV Recall Train(%)	100%	100%	100%	96%	84%	97%
CV F1 Score (%)	97%	98%	97%	98%	79%	99%
CV ROC AUC (%)	99%	99%	98%	99%	93%	98%

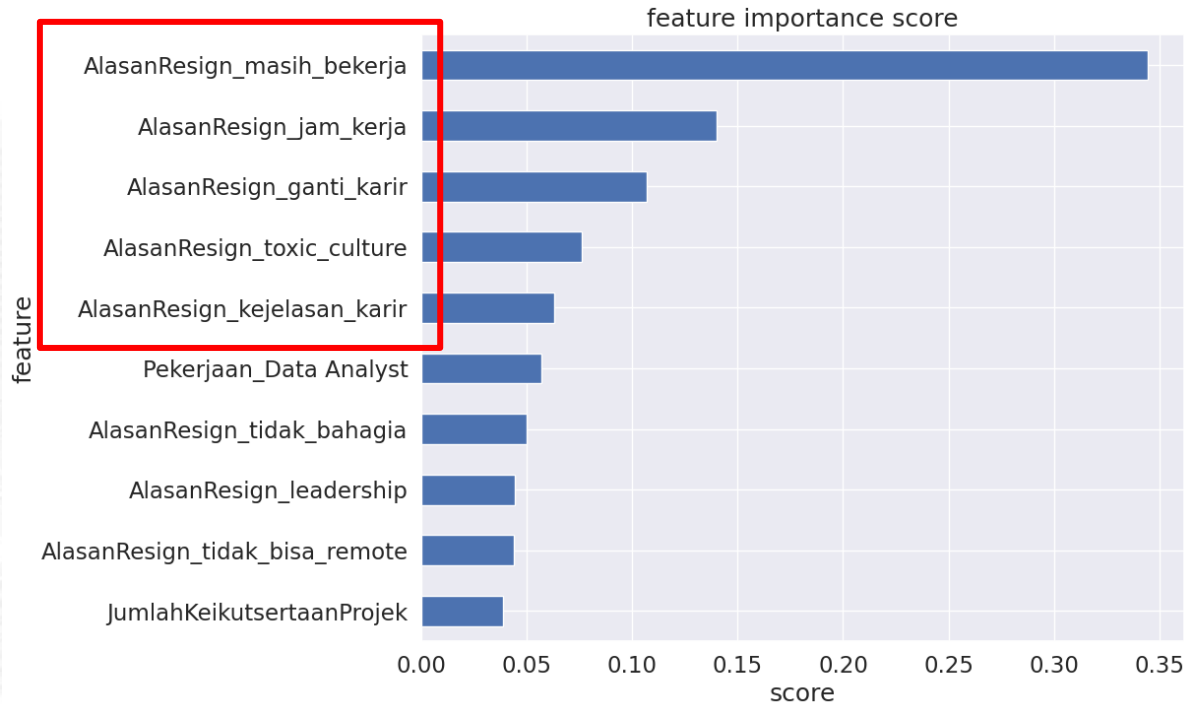
Because we are focusing in Recall, XGBoost is the best model because Evaluation Matrix in Train and Test is not to far

Confusion Matrix



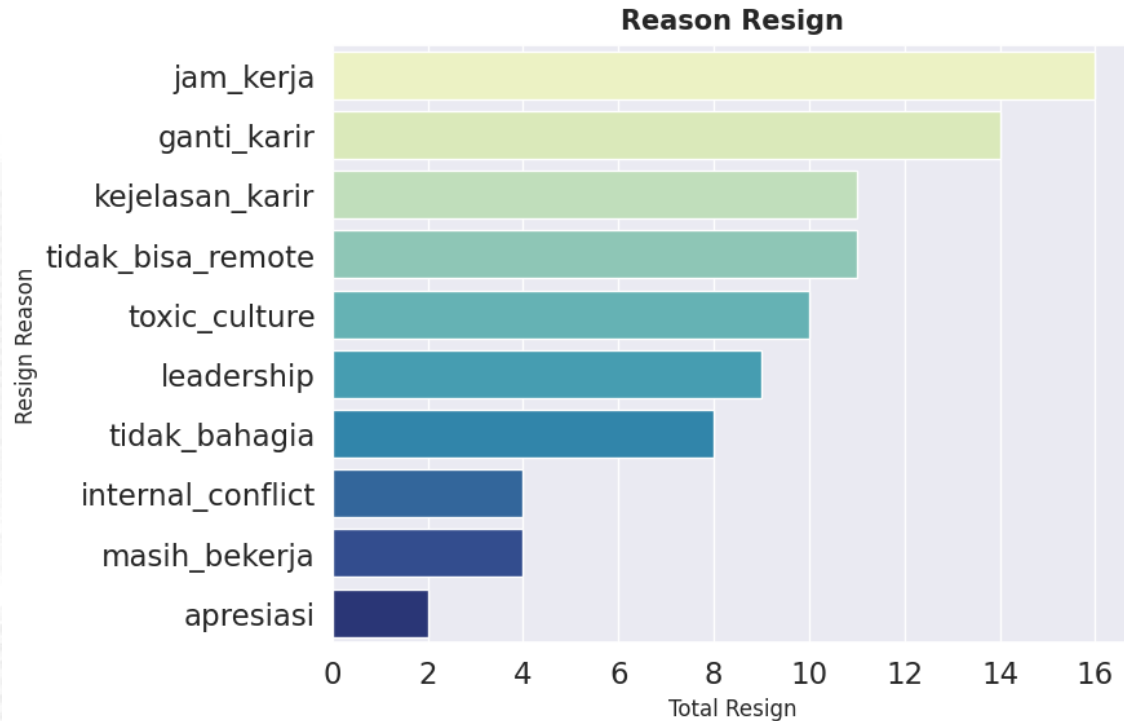
For details, you can see the jupyter notebook [here](#)

Feature Importance



AlasanResign is
the top of feature
importance

Reason Employee Resign



From the chart adjacent, it can be concluded that the highest resignation reasons are related to working hours, career change, career advancement uncertainty, and the inability to work remotely.

1. **Flexible Working Hours:** To address the issue of resignations related to working hours, consider implementing flexible working hours or remote work options where feasible. This can improve work-life balance and employee satisfaction.
2. **Career Development and Clarity:** Provide employees with clear paths for career advancement within the organization. Offer training, mentorship, and opportunities for skill development to help them progress in their careers.
3. **Remote Work Options:** Explore the possibility of allowing employees to work remotely, at least in part. This can be particularly important for those who value the flexibility of working from different locations.
4. **Skills Development:** Invest in training and development programs to help employees acquire the skills they need to advance within the company. This can boost loyalty and engagement.