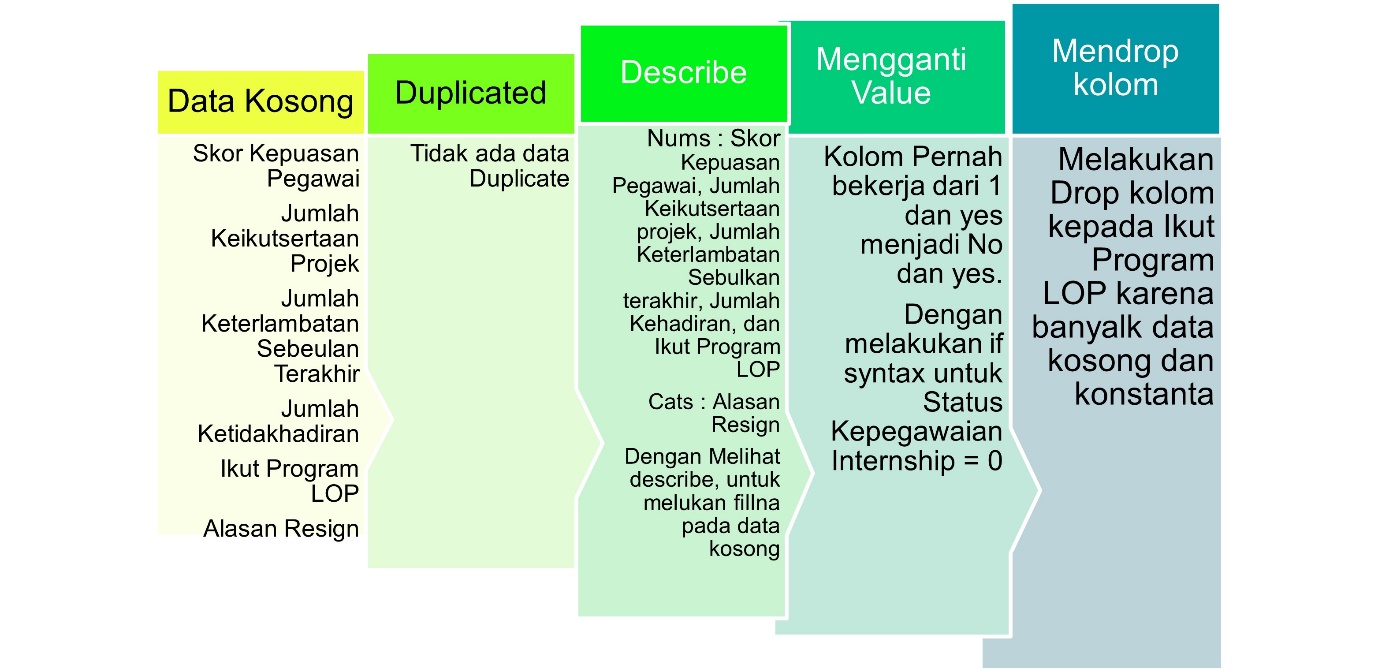
**Improving Employee Retention by Predicting Employee Attrition Using Machine Learning: Overview**

Human resources (HR) are the main assets that need to be managed properly by a company so that business objectives can be achieved effectively and efficiently. On this occasion, we will face a problem regarding human resources in the company. Our focus is to find out how to keep employees afloat in the existing company which can result in increased costs for employee recruitment and training for those who have just joined- programs that are relevant to employee problems.

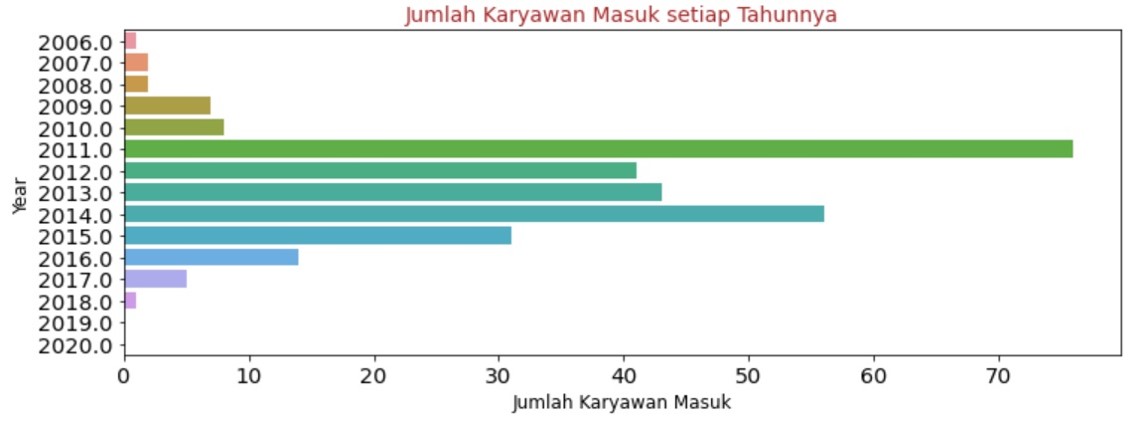
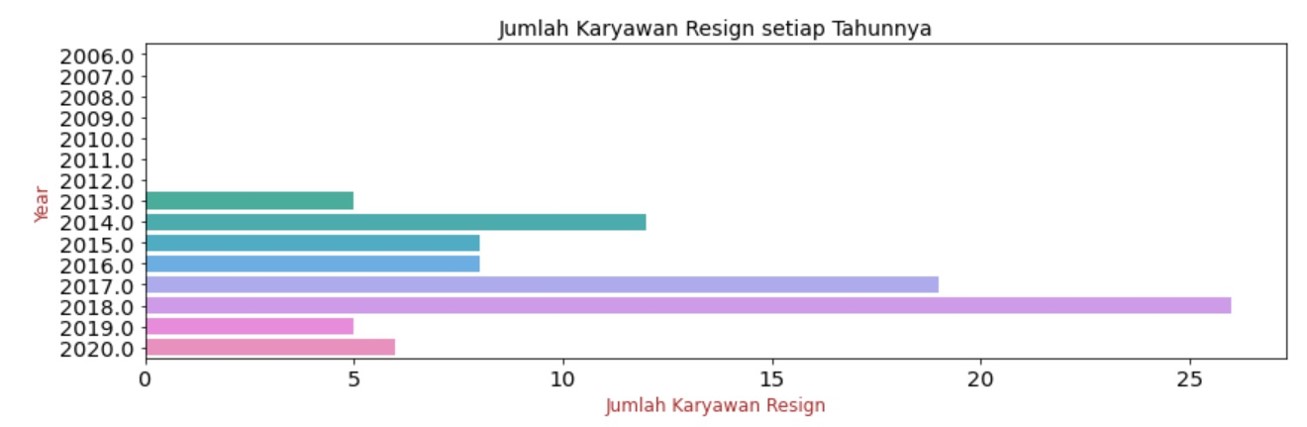
**Data Preprocessing**

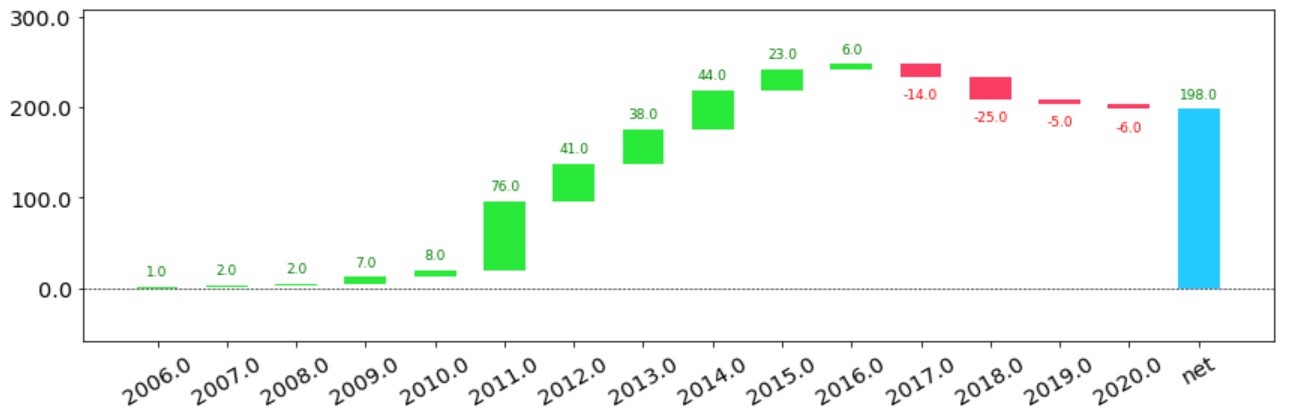


* In this dataset, there is no duplicated data. In other hands, there are null data.
* Change the value ‘Pernah bekerja’ column to 1 or 2 to yes or no. Doing a if syntax for ‘Status Kepagawaian’ column value Internship to 0.
* Drop column ‘Ikut Program LOP’ because there are so many null data.

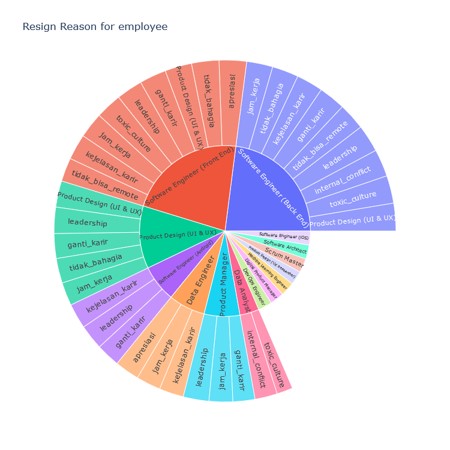
**Analyst**

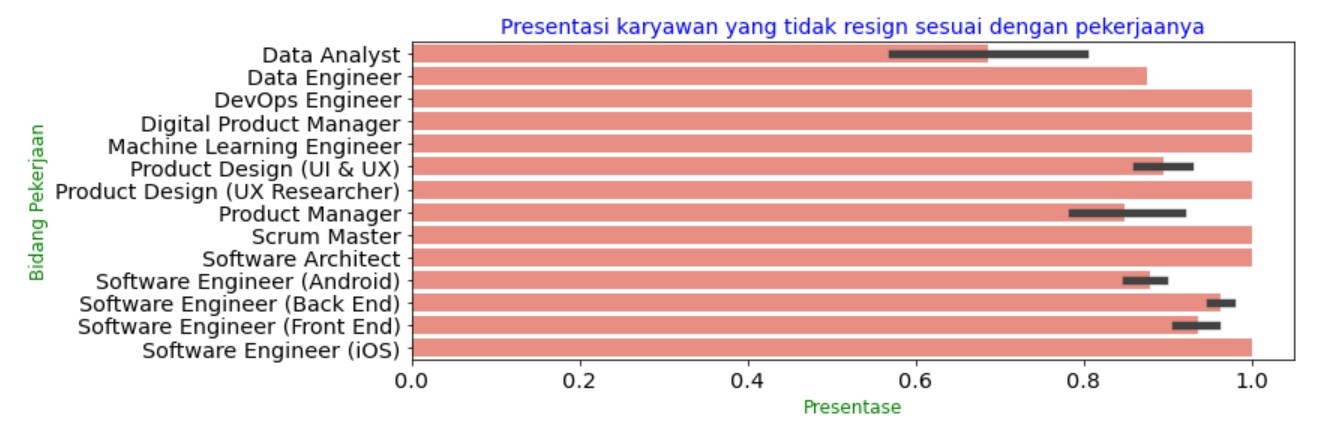
* Annual report on Employee Number Changes

From the both of the graphic, In 2011 the highest accepted new comer. In other hand, In 2018 the highest employee got their resign.

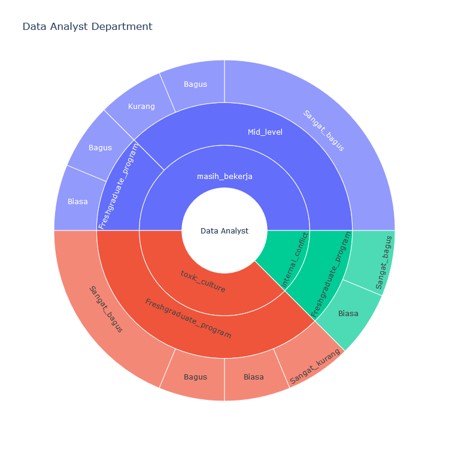
From the graphic above, between 2011 and 2016, a sum of people get accepted highest rather than employee resign. Following year, it’s opposite from before.

* Resign Reason Analysis for Employee Attrition Management Strategy



Chart above it’s about the reason employee get their resign

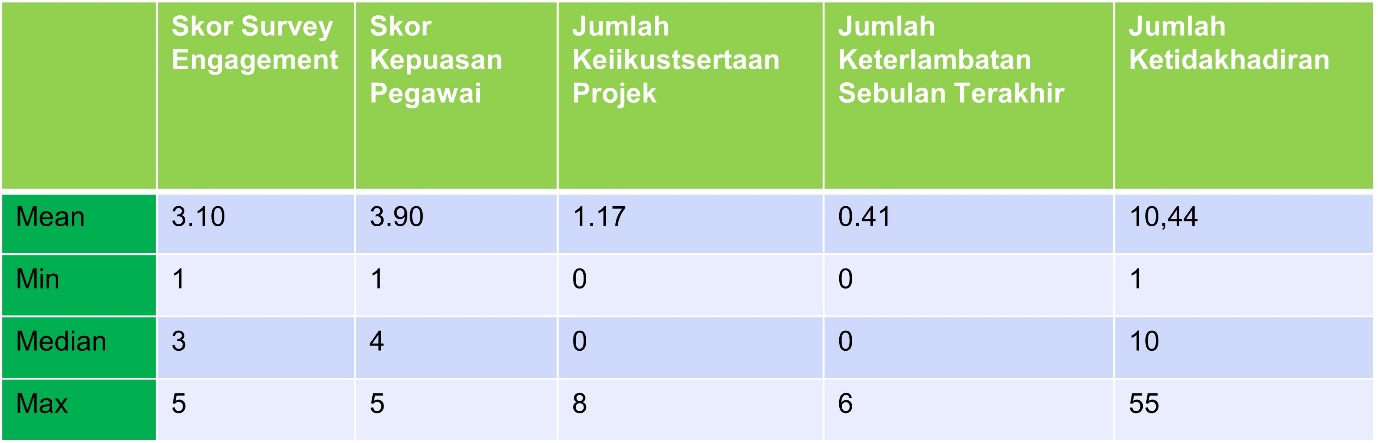
Data Analyst is the most percentage employee resign from their department.



The graph above, tell us about the reason resign, employee performance, and career path. From this graph, employee resign because of toxic culture and internal conflict, career path for the fresh graduate and good performance.

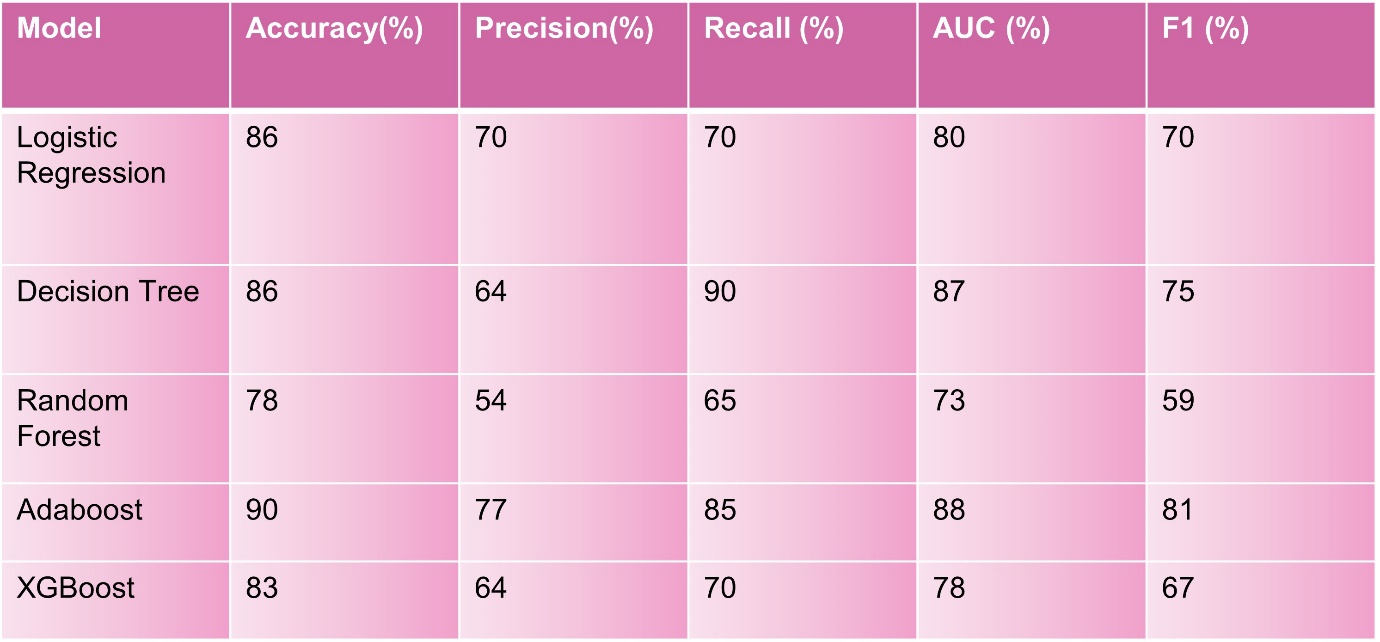
**Data Preprocessing**

* There is no duplicated data
* For null data, replace it with 0.
* There are 5 columns for nums and 11 columns for cats.
* Described for nums categories.

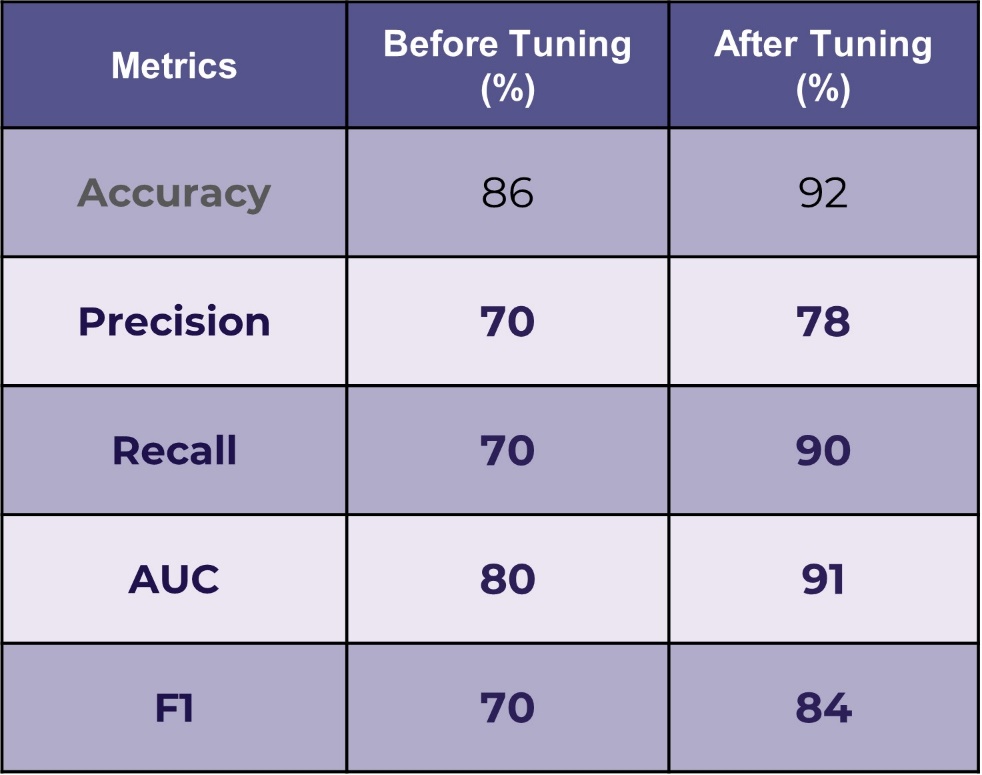


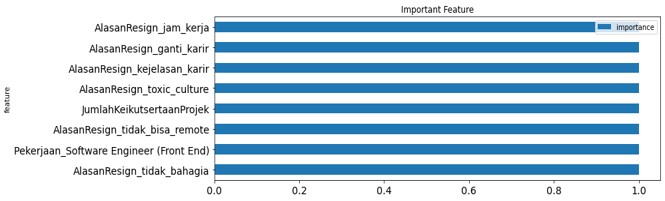
* Describe for cats categories.
* For the target, use ‘yes’ or ‘no’ from masih bekerja column.
* Using Log, zscore, label encoding, and one hot encoding.
* Split data 70:30 and having a 5 model with Resign coulumn as a target.
* Handling imbalance class SMOTE data train (1:1).

**Modelling**

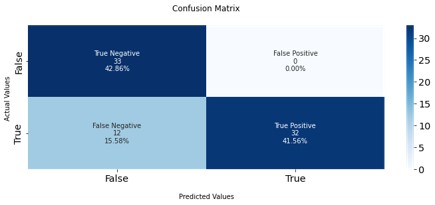
****

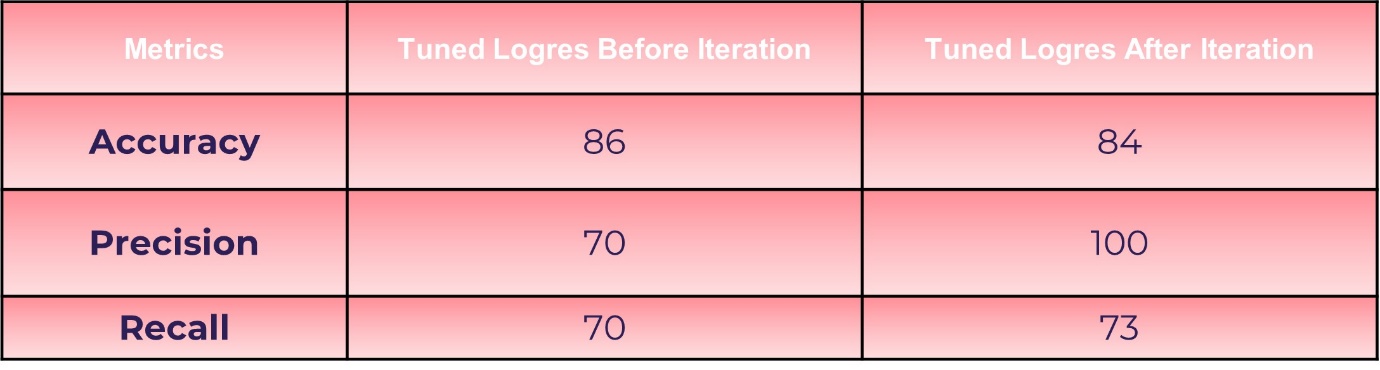
The best modelling is Logistic Regression, so this modelling recalculate with hyperturning. For the result:

****

After hyperturning, Logistic Regression get the increase result.

This is the best feature for improving employee retention.





**Result**

* From the matrix, found that employees who resign and employee who do not resign have almost the same input. The main reason why employee resign is a problem hours of work. If 1 employee working 8 hours/day, due to a small amount of people, so the employee has to work >8 hours/day.
* With the addition of adequate employees, create employee who in there have an exceeding working hours.
* Companies can provide challenges to their employees so that employees are not saturated with their existing jobs.
* There needs to build a togetherness activities so that between employee can build a sense of trust in every fields.
* Because there are still many freshgraduate employees, companies need to provide guarantees to them about their career path ahead.