Salifort Motors

IMPROVING EMPLOYEE RETENTION

Ritoja Kundu

PROJECT OVERVIEW

Company Context: Salifort Motors, a French-based alternative energy vehicle manufacturer, aims to address employee retention challenges.

Business Motivation: High employee turnover is costly due to the time and resources required for hiring and onboarding replacements.

Survey Insights: A recent employee survey highlighted potential issues influencing employee departures.

PROJECT OVERVIEW

Project Goal: Develop a predictive model to identify employees likely to leave the company.
Key Predictors Used:

- Department
- Number of projects
- Average monthly working hours
- Satisfaction level
- Tenure at the company
- Promotion status in the past 5 years

Strategic Benefit: Early identification of at-risk employees enables targeted interventions to improve retention and reduce turnover costs.

The Process

1. Exploratory Data analysis performed to check for null values, duplicate rows, outliers and to correct column names to make the dataset uniform.

2. Visualizations plotted using Matplotlib library to identify how certain columns are related and the effect of each column on the Target.

3. Two approaches tried out - Logistic Regression and Tree-based models, namely Decision tree and Random Forest.

4. The best model identified based on precision, recall, accuracy and ROC-AUC scores.

A GLIMPSE OF THE DATASET

Hr Analytics Job Prediction

Data Set of Hr Department to Predict Employees Behaviour

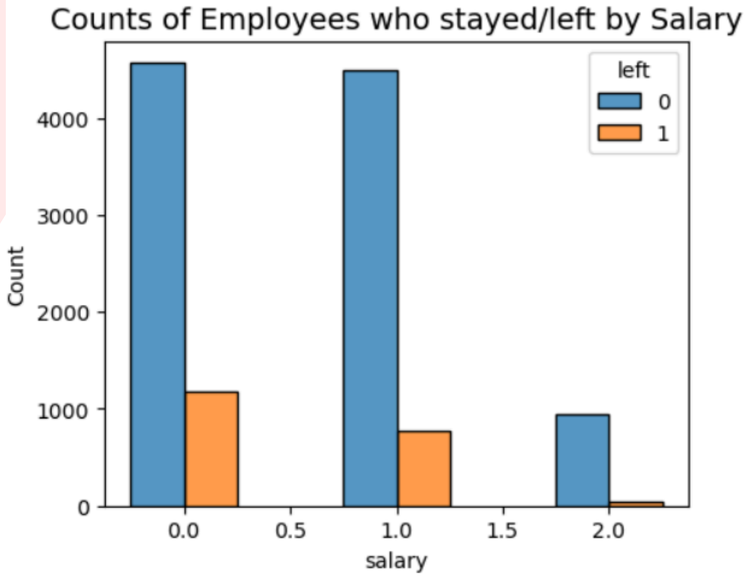
kaggle.com

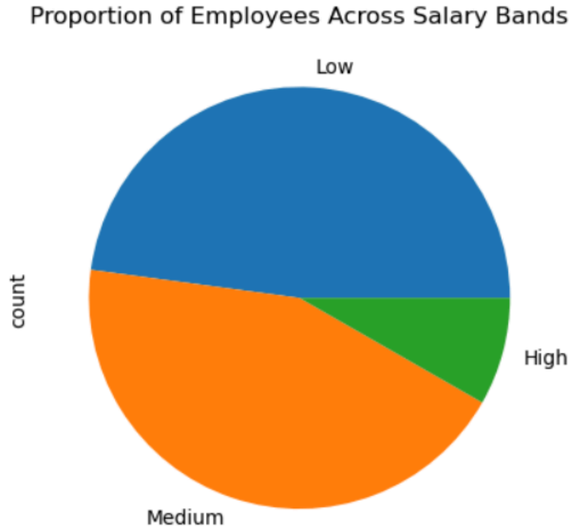
	satisfaction_level	last_evaluation	number_project	average_montly_hours	time_spend_company	Work_accident	left	promotion_last_5years	Department	salary
0	0.38	0.53	2	157	3	0	1	0	sales	low
1	0.80	0.86	5	262	6	0	1	0	sales	medium
2	0.11	0.88	7	272	4	0	1	0	sales	medium
3	0.72	0.87	5	223	5	0	1	0	sales	low
4	0.37	0.52	2	159	3	0	1	0	sales	low

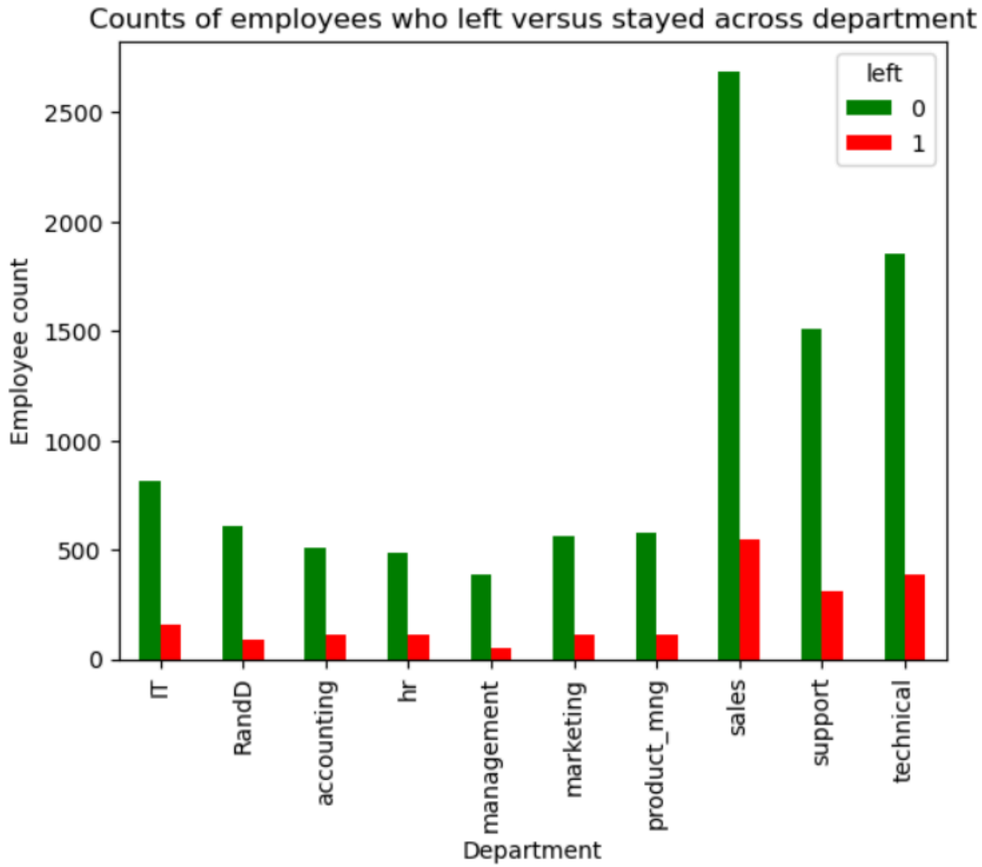
Target Variable : 'Left' Left(1) - employee left, Left(0) - employee did not leave.
'Department' and 'Salary' converted to numerical columns.

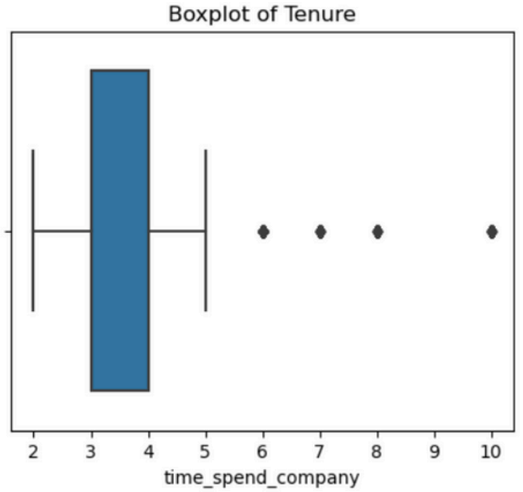
Logistic Regression - Outliers removed (Mean +- 1.5*IQR)
Decision Tree and Random Forest - Round 1 run with all the independent variables; Round 2 run after excluding 'Satisfaction level' and 'monthly average hours' while creating a new column 'Overworked'.
Average monthly hours : Min = 90, Max = 310, average = 166.67
If Average monthly hours > 170 then the employee is overworked (1).

VISUALIZATIONS

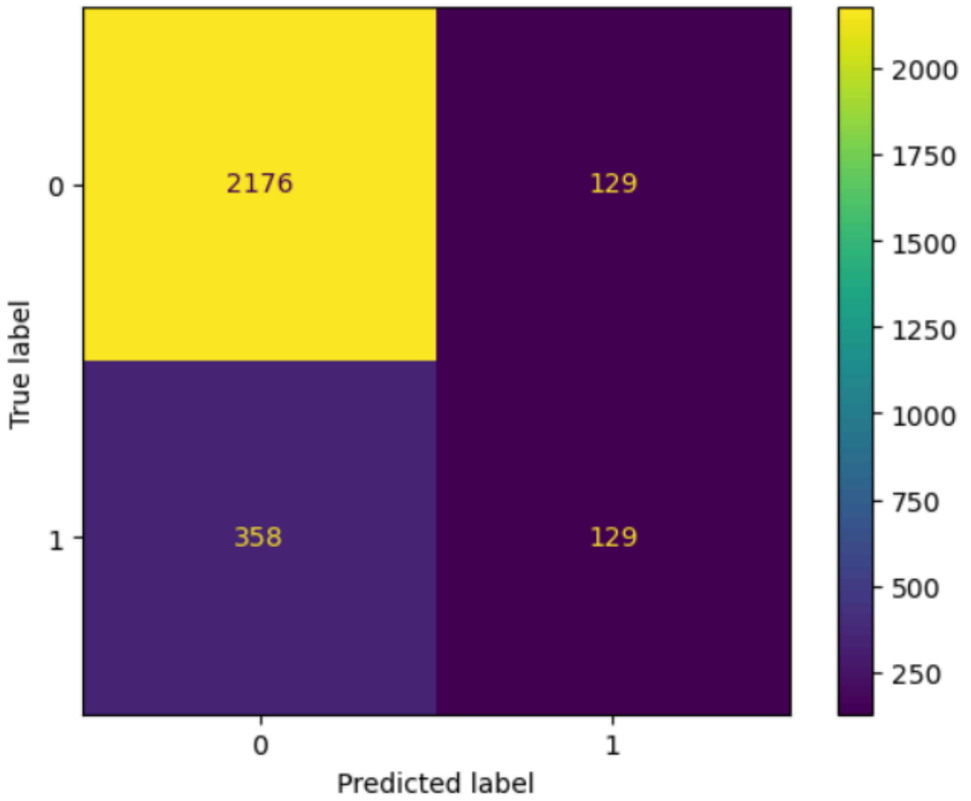








CONFUSION MATRIX FROM LOGISTIC REGRESSION MODEL



Upper-left quadrant : True negatives.
Upper-right quadrant : False positives.
Bottom-left quadrant : False negatives.
Bottom-right quadrant : True positives.

True negatives: The number of people who did not leave that the model accurately predicted did not leave.
False positives: The number of people who did not leave the model inaccurately predicted as leaving.
False negatives: The number of people who left that the model inaccurately predicted did not leave
True positives: The number of people who left the model accurately predicted as leaving

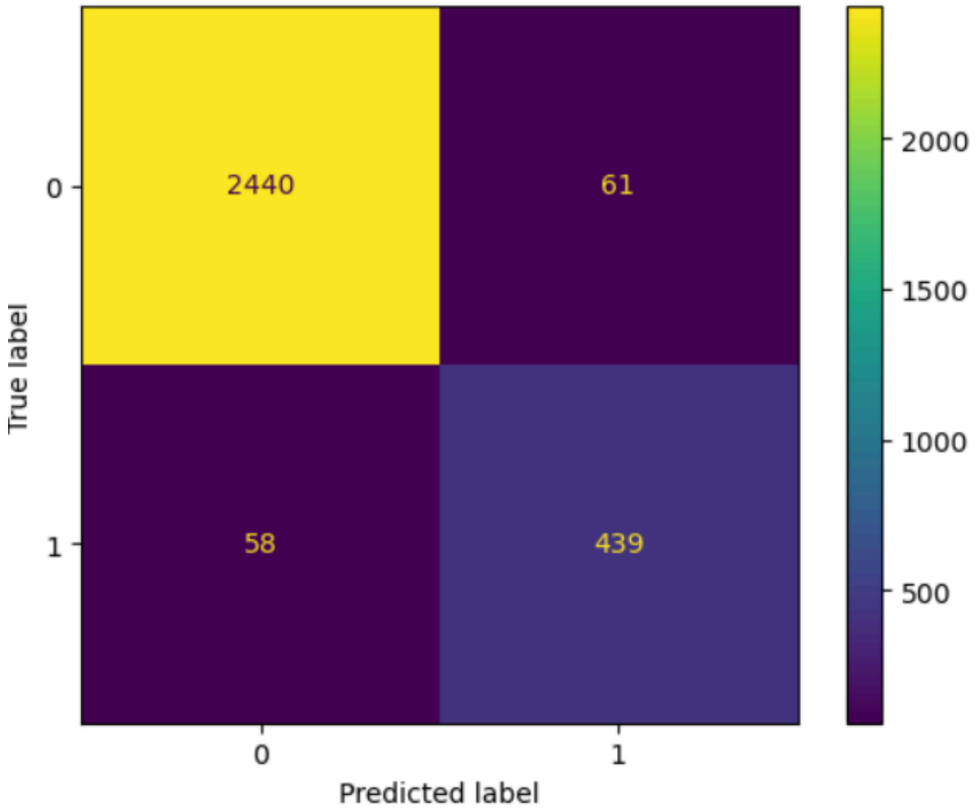
Model predicted more true negatives than true positives.

LOGISTIC REGRESSION MODEL - RESULTS

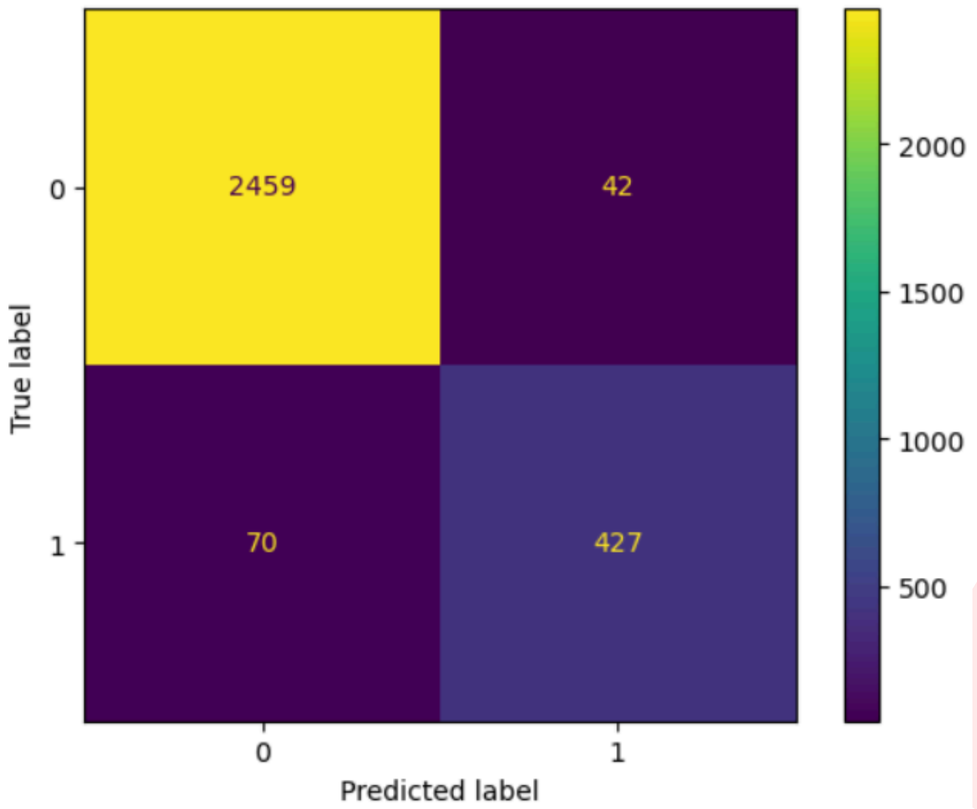
	precision	recall	f1-score	support
Predicted to not leave	0.86	0.94	0.90	2305
Predicted to leave	0.50	0.26	0.35	487
accuracy			0.83	2792
macro avg	0.68	0.60	0.62	2792
weighted avg	0.80	0.83	0.80	2792

The model achieved a precision score of 80% and a recall score of 83%. Although if it is more important to predict whether an employee would leave the company, then the scores were a bit low.

CONFUSION MATRIX FROM TREE-BASED MODELS



Predictions of Decision Tree Classifier

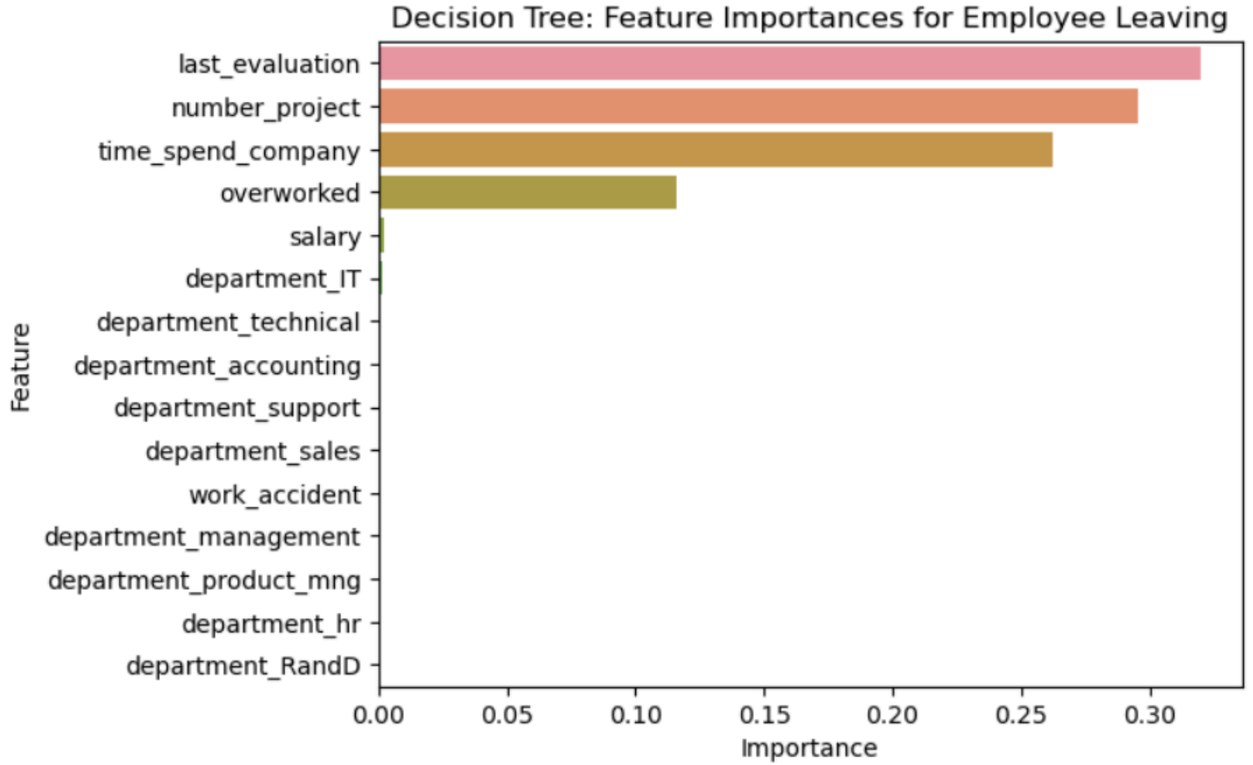


Predictions of Random Forest Classifier

DECISION TREE MODEL - RESULTS

	model	precision	recall	F1	accuracy	auc
decision tree cv		0.963198	0.922352	0.942266	0.981208	0.968911

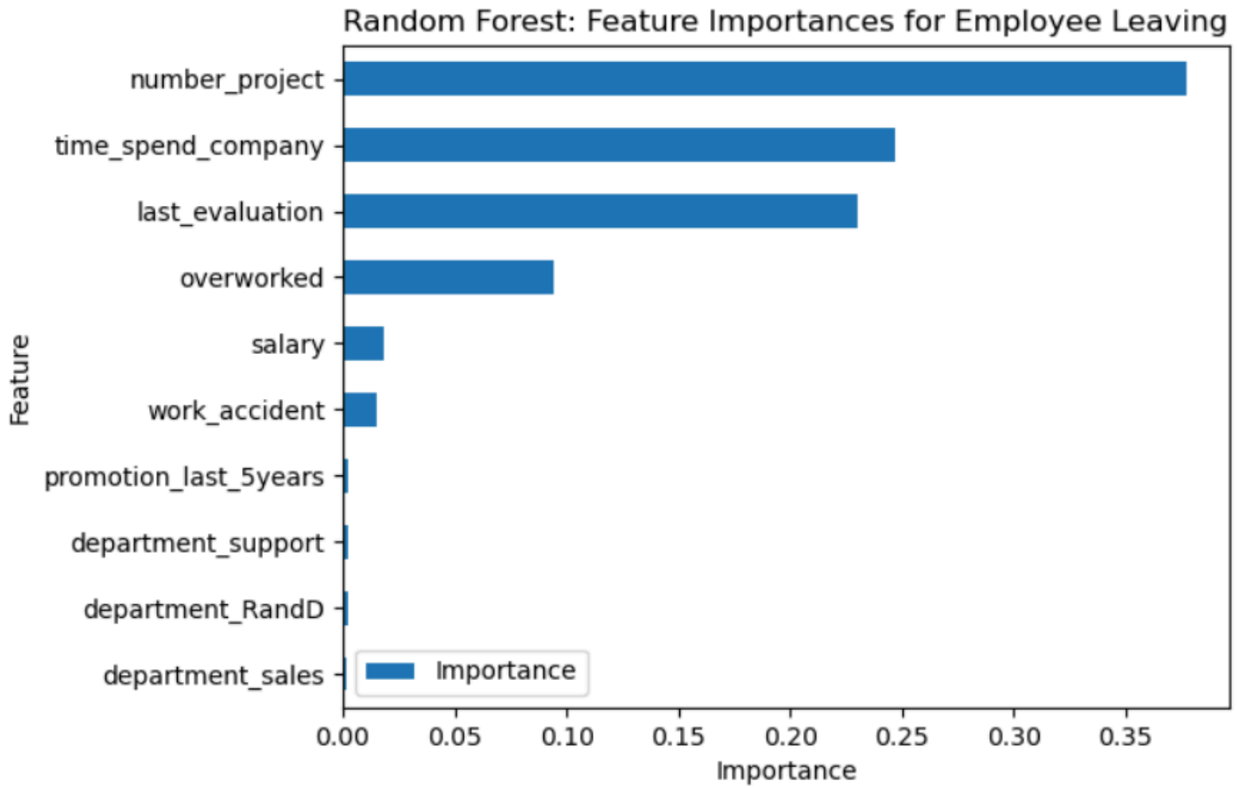
	model	precision	recall	F1	accuracy	auc
decision tree 2 cv		0.886223	0.902263	0.894022	0.964416	0.957847



RANDOM FOREST MODEL - RESULTS

model	precision	recall	F1	accuracy	auc
random forest cv	0.984917	0.91365	0.947918	0.98332	0.981634

model	precision	recall	F1	accuracy	auc
random forest 2 cv	0.917926	0.866807	0.891315	0.964861	0.968666



OUTCOMES

Random Forest turned out to be the model of choice for the prediction.
The feature importance extracted from the models confirm that employees at the company are overworked.

To retain employees, the following recommendations may be presented to the stakeholders :

- The number of projects that employees can work on can be capped.
- It may be considered promoting employees who have been with the company for at least four years, or conducting further investigation about why four-year tenured employees are so dissatisfied.
- Either employees can be rewarded for working longer hours, or not required to do so.
- If employees aren't familiar with the company's overtime pay policies, they should be informed about the same. If the expectations around workload and time off aren't explicit, they can be made clear.
- Company-wide and within-team discussions can be held to understand and address the company work culture, across the board and in specific contexts.