

Employee Attrition Analysis.



ARISE
Rakamin Academy Data Science Batch 33



Company Background.

ARISE is a long-established manufacturing company known for producing high-quality products. Currently, the company is facing a significant challenge related to a high employee attrition rate. As a response, the management has tasked the data and BI team with understanding the key factors to focus on in efforts to reduce the employee attrition rate.



ARISE members.

Zulfikri Syafnur

Project Lead



Anik Maulia Tri H.

Data Scientist



Jennifer Greyselda

Business Intelligence



Wahyu Fajrin M.

Data Analyst





OUTLINE.

01

**Business
Understanding**

02

**Exploratory Data Analysis
(EDA)**

03

Data Preprocessing

04

Modeling & Evaluation

05

**Business
Recommendation**

06

Business Simulation



A group of four business professionals (three women and one man) are walking and talking in a modern office hallway. The man on the left is holding a coffee cup. The woman next to him is clapping her hands. The woman next to her is gesturing with her hands. The woman on the right is holding a clipboard and a bag. A large orange circle with the number '01' is overlaid on the image.

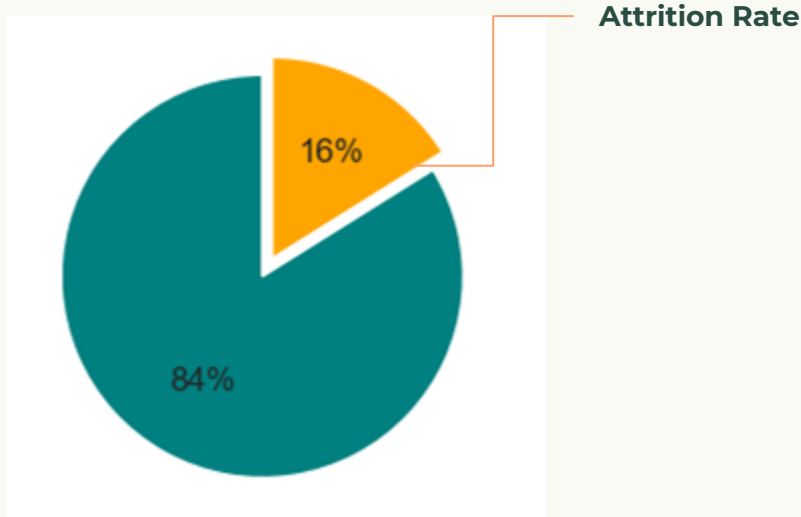
01

Business Understanding.

Problem Statement, Goals, Objective & Business Metrics.

Problem Statement.

1. The number of employees experiencing attrition



Attrition
237 employees

No Attrition
1,233 employees

Total Employee
1470 employees

2. Loss

Hiring Cost:

The replacement cost reaches 4700 USD per person (*SHRM).

Total Hiring Cost = 1,113,900 USD

Opportunity Cost:

An employee who resigns has the potential to contribute more optimally to the company.

Lower Productivity:

Disrupting the positive relationships that have been built within the team.

*SHRM = The Society for Human Resource Management (SHRM) is the largest HR association in the world, with the goal of creating a better work environment.



Goals, Objective & Business Metrics.



Goals

Reducing the yearly attrition rate from 16% to less than 10% and lowering costs incurred for employee recruitment, so doing will contribute to improved workforce stability and financial efficiency.



Objective

Create a classification model to predict employees prone to attrition.



Business Metrics

1. Employee Attrition Rate
2. Cost for Hiring Process



02

Exploratory Data Analysis (EDA)

Informasi Dataset dan Preliminary Insight



Data Understanding.

Features of Employee Profile:

1. Age
2. DistanceFromHome
3. Education
4. EducationField
5. Gender
6. MaritalStatus
7. Over18

Features of Current Company:

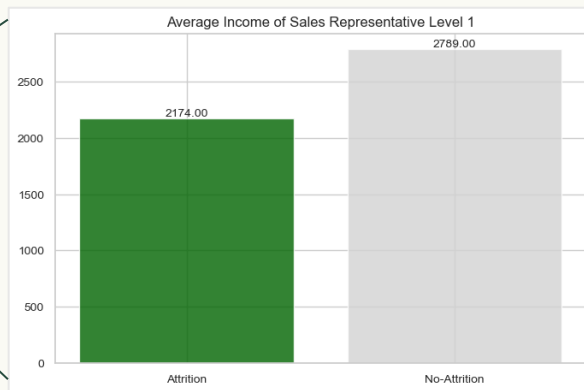
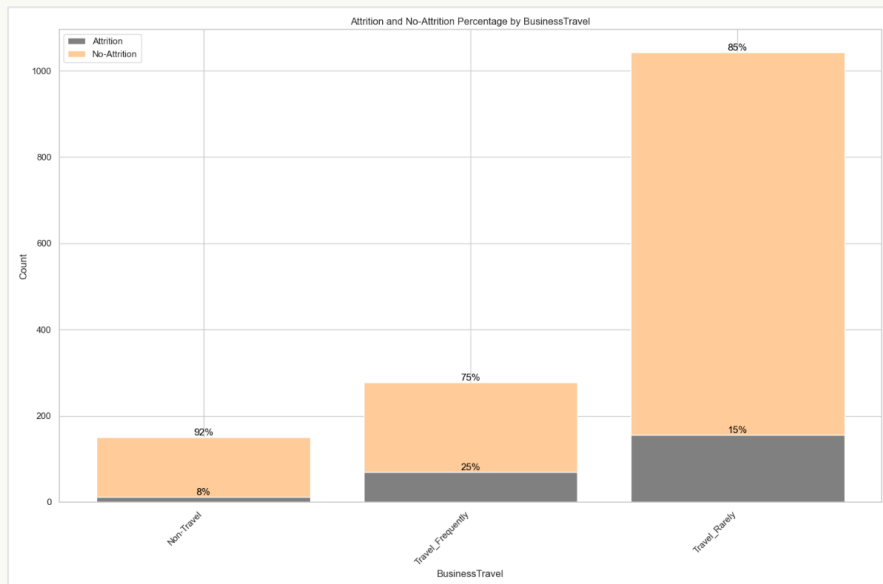
- | | |
|----------------------------|------------------------------|
| 1. BusinessTravel | 14. OverTime |
| 2. DailyRate | 15. PercentSalaryHike |
| 3. Department | 16. PerformanceRating |
| 4. EmployeeCount | 17. RelationshipSatisfaction |
| 5. EmployeeNumber | 18. StandardHours |
| 6. EnvironmentSatisfaction | 19. StockOptionLevel |
| 7. HourlyRate | 20. TrainingTimesLastYear |
| 8. JobInvolvement | 21. WorkLifeBalance |
| 9. JobLevel | 22. YearsAtCompany |
| 10. JobRole | 23. YearsInCurrentRole |
| 11. JobSatisfaction | 24. YearsSinceLastPromotion |
| 12. MonthlyIncome | 25. YearsWithCurrManager |
| 13. MonthlyRate | |

Features of Previous Work Experience:

1. NumCompaniesWorked
2. TotalWorkingYears

Target: Attrition

- The dataset consists of 1.470 rows and 35 columns
- The dataset contains no missing values and no duplicates



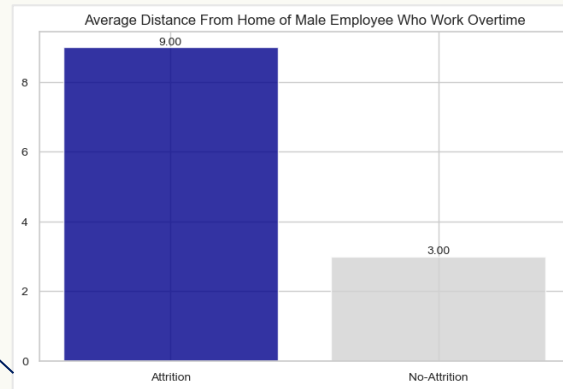
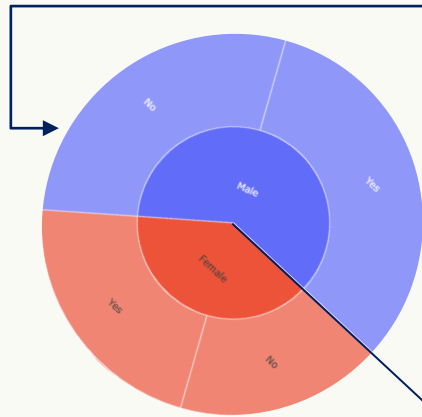
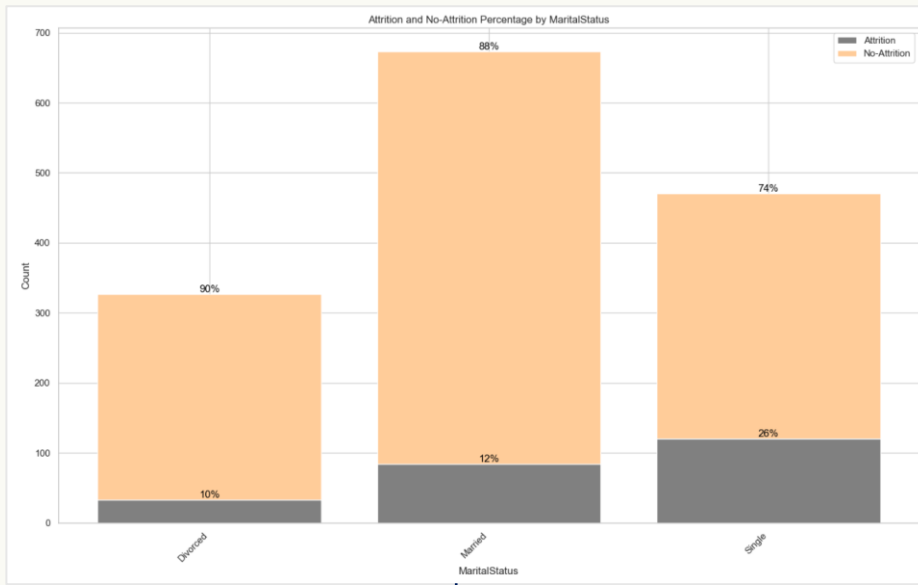
Business Insight 1

Attrition Based on BusinessTravel, JobRole, JobLevel, and MonthlyIncome

Sales Representative Level 1 employees with frequent business travel face high attrition due to a \$615 monthly income gap compared to their peers.

Recommend reviewing and adjusting monthly income for fairness and retention.





Business Insight 2

Attrition Based on MaritalStatus, Gender, OverTime, and DistanceFromHome

High attrition among single male employees is linked to frequent overtime and a 6 KM difference in commuting distance compared to those who stay.

Recommend a review of employee working hours. This could involve setting a maximum overtime limit, such as not exceeding 80 hours per week, or offering additional compensation if employees work beyond the standard 80 hours, with a focus on enhancing retention.



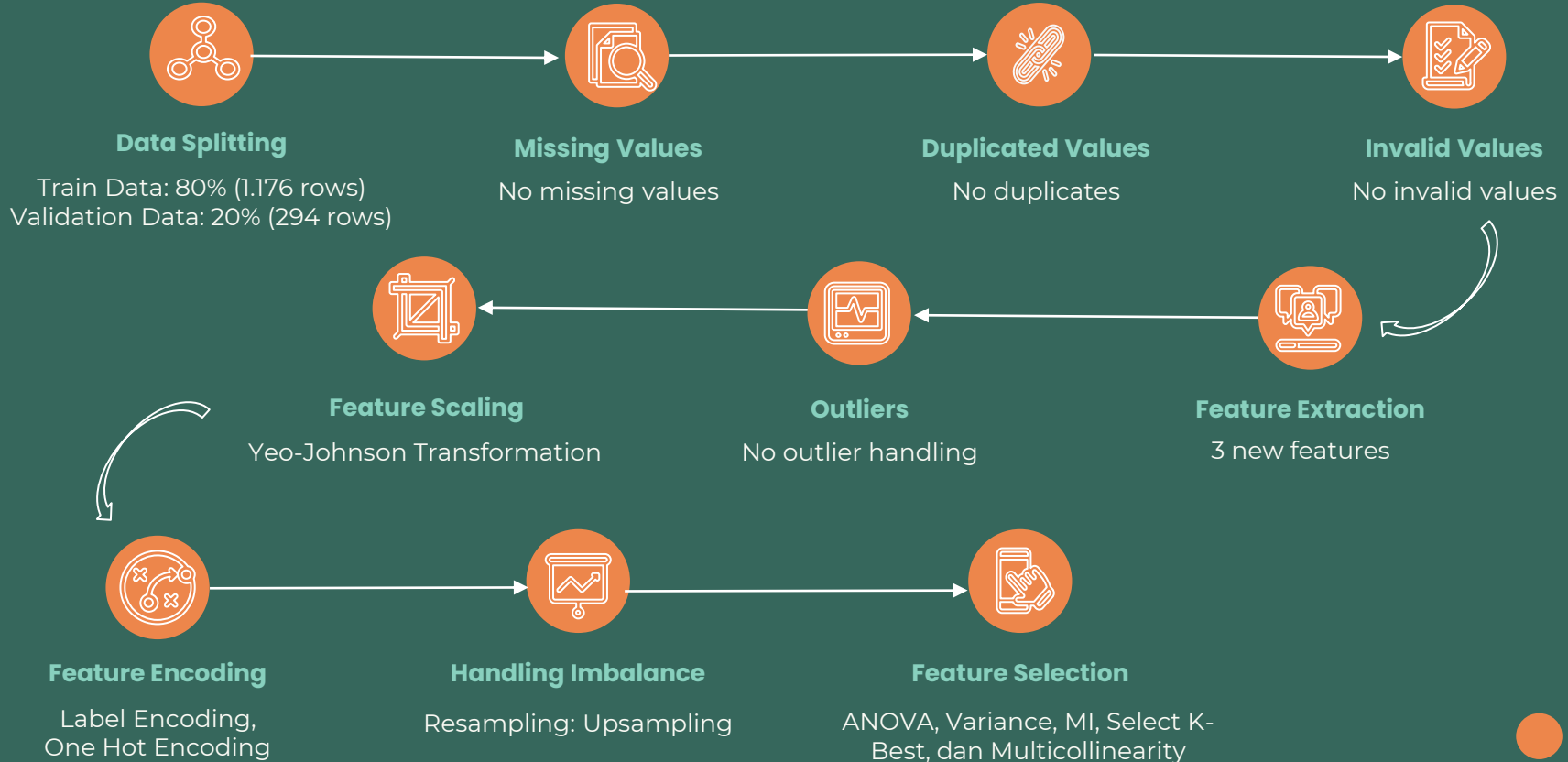
A man and a woman are shaking hands in front of a large window. The man is wearing a red shirt and the woman is wearing a white turtleneck and jeans. The background shows a modern building with large glass windows.

03

Data Preprocessing.

Handling Missing, Duplicated, dan Invalid Values; Handling Outliers; Feature Extraction, Transformation, Encoding, dan Selection; Handling Imbalance Data

Data Preprocessing.





Data Preprocessing.

15 features will be used for modeling:

- | | |
|--------------------------------|-----------------------------|
| 1. AverageTenurePerCompany (*) | 9. MaritalStatus_Single |
| 2. OverTime | 10. MonthlyRate |
| 3. Age | 11. DailyRate |
| 4. YearsWithCurrManager | 12. HourlyRate |
| 5. JobLevel | 13. DistanceFromHome |
| 6. YearsInCurrentRole | 14. YearsSinceLastPromotion |
| 7. ExperienceSalaryRatio (*) | 15. AverageSatisfaction (*) |
| 8. StockOptionLevel | |

- Features marked with (*) are the result of feature extraction.

A background image showing three people in a professional setting. A woman on the left is smiling and looking at a man in the center, who is also smiling and looking at a man on the right. The man on the right is holding a cup. The background is slightly blurred, showing some colorful shapes.

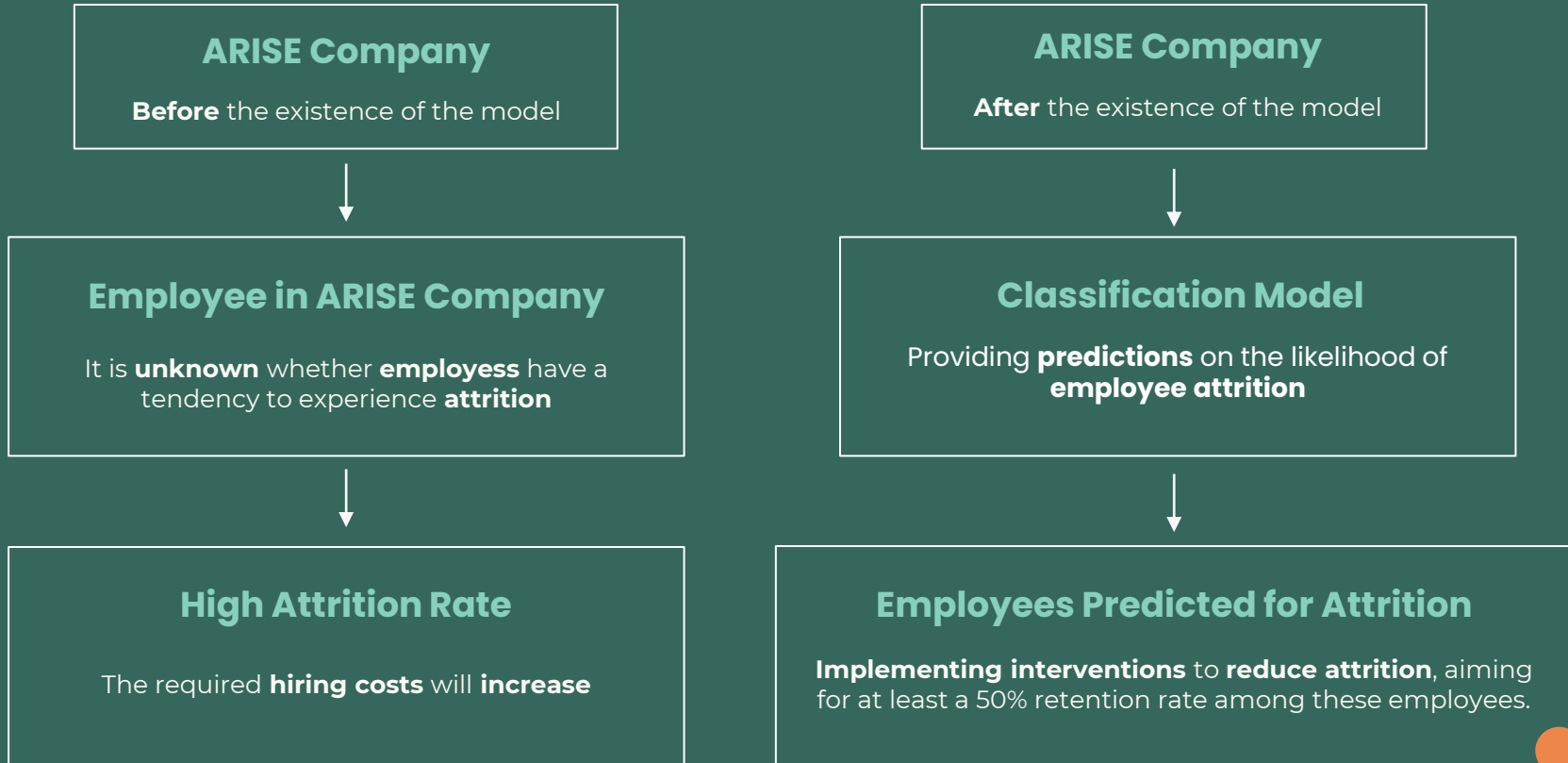
04

Modeling & Evaluation.

Model Evaluation Parameters, Machine Learning
Techniques, Model Comparison, and Feature Importance



Business Flow Simulation.





Model Evaluation Metrics.

Recall

Primary Metrics Evaluation

Reduce False Negative

Employees predicted as No Attrition, but actually Attrition.

- Minimizing employees experiencing attrition.
- Reducing the costs associated with hiring new employees.

Precision

Secondary Metrics Evaluation

Reduce False Positive

Employees predicted as Attrition, but actually No-Attrition.

- Minimize costs associated with employee intervention or retention efforts (if necessary).

Optimizing ROC AUC

Balancing Trade-offs

Assessing the model's ability to distinguish between positive and negative instances while optimizing the balance between recall and precision for effective performance.

Machine Learning Models

1. Decision Tree
2. Random Forest
3. Adaboost Classifier
4. XGBoost Classifier
5. Gradient Boosting Classifier



Model Comparison.

	Models	Recall (Train)	Recall (Val)	Precision (Train)	Precision (Val)	Accuracy (Train)	Accuracy (Val)	ROC AUC (Train)	ROC AUC (Val)
0	Adaboost Classifier	0.820000	0.760000	0.830000	0.830000	0.830000	0.760000	0.830000	0.720000
1	Decision Tree	1.000000	0.770000	1.000000	0.760000	1.000000	0.770000	1.000000	0.560000
2	Gradient Boosting Classifier	0.930000	0.780000	0.920000	0.820000	0.920000	0.780000	0.920000	0.680000
3	Random Forest	1.000000	0.810000	1.000000	0.780000	1.000000	0.810000	1.000000	0.580000
4	XGBoost Classifier	1.000000	0.820000	1.000000	0.800000	1.000000	0.820000	1.000000	0.620000

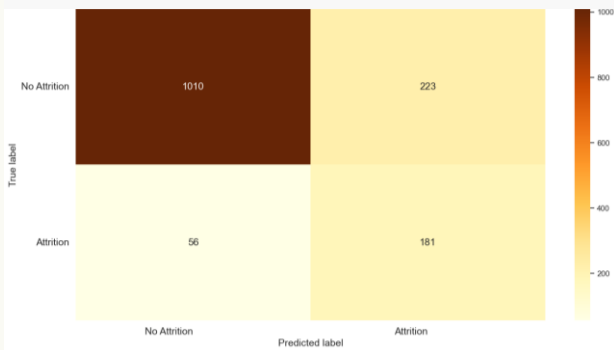
The model with the best and consistently good performance is the **Adaboost Classifier**, while the other four models show indications of overfitting as they only perform well on their training data.

Model Comparison.

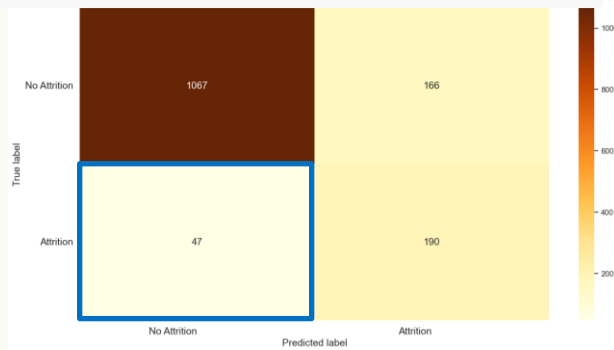
Adaboost Classifier Before and After Tuning

	Models	Recall (Train)	Recall (Val)	Precision (Train)	Precision (Val)	Accuracy (Train)	Accuracy (Val)
0	Before Tuning	0.820000	0.760000	0.830000	0.830000	0.830000	0.760000
1	Tuning using GridSearch	0.900000	0.770000	0.880000	0.830000	0.890000	0.770000
2	Tuning using RandomSearch	0.870000	0.770000	0.860000	0.830000	0.870000	0.770000

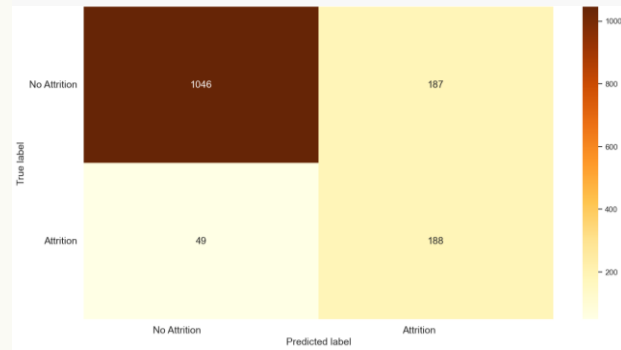
Before Tuning



After Tuning using GridSearch

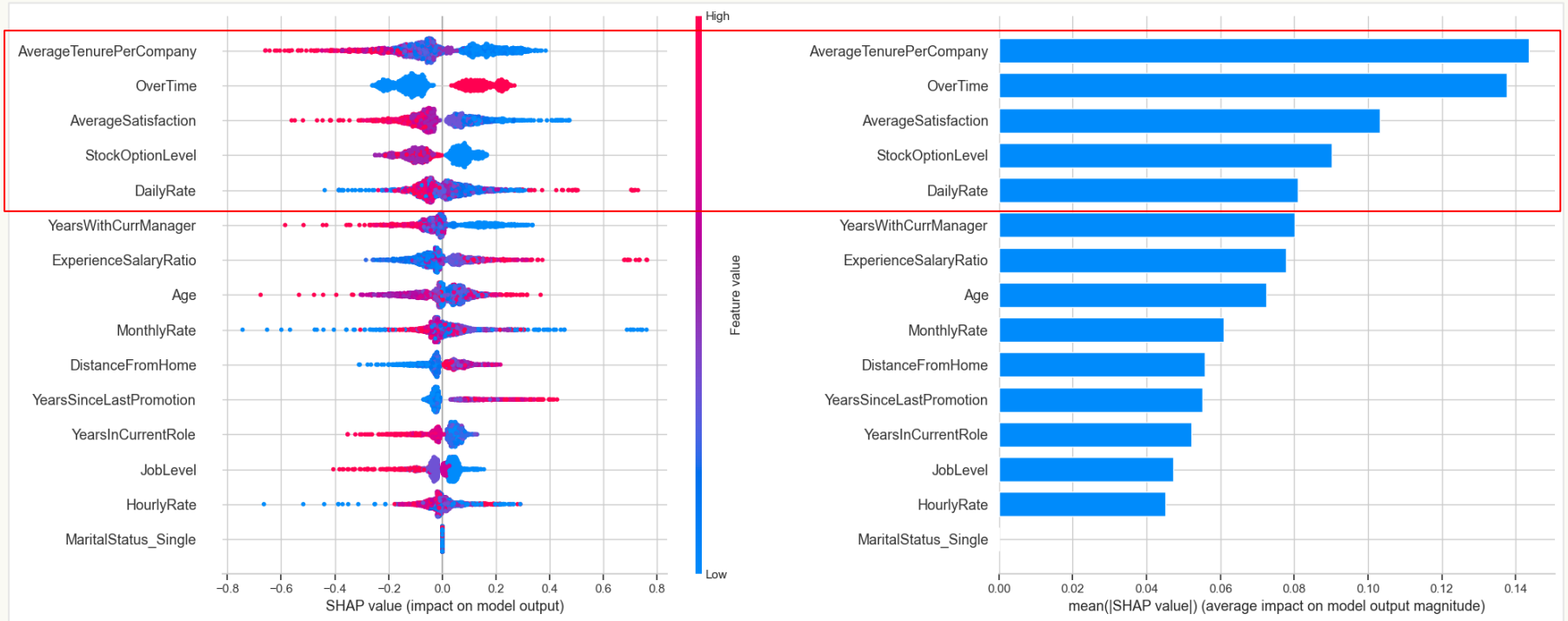


After Tuning using RandomSearch



Adaboost Classifier After Tuning using GridSearch exhibits the lowest cases of False Negatives (predicting No Attrition but actual Attrition), making it the selected model.

Feature Importance.



- Negative = No Attrition
- Positive = Attrition

A group of four business professionals (three women and one man) are gathered around a wooden table in a modern office setting. They are smiling and engaged in a discussion. A laptop is open on the table, and there are some documents and a small potted plant. In the background, a whiteboard with a target diagram and handwritten notes is visible. The scene is framed by a thin orange border with small orange circles at the top and bottom center.

05

Business Recommendations.

Business Recommendations.

1. Smart Recruitment

- Consider AverageTenurePerCompany and inquire about job changes or short durations.
- Ensures understanding of candidate expectations, fostering a better match and potential turnover reduction.

2. Optimizing Work Hours

- Explore a maximum overtime limit, e.g., 80 hours per week.
- Provide extra compensation beyond 80 hours.

3. Proactive Feedback

- Actively listen to employee input and concerns.
- Understand issues leading to resignations.
- Strategy to reduce attrition and enhance overall satisfaction.

4. Reviewing Compensation Systems

- Evaluate stock distribution and daily income alignment with employee contributions.
- Identify areas for improvement.



A background image showing a group of people in an office setting, some with their arms raised in celebration. A large, semi-transparent teal rectangle is overlaid on the image, containing the main title and a definition. Several orange circles of varying sizes are scattered around the teal rectangle. The title 'Business Simulation' is written in a large, bold, white sans-serif font.

Business Simulation

Business Metrics are measurable indicators used to track business processes for assessing the level of business performance

Attrition Rate

Hiring Cost

Attrition Rate

TP = 190
FN = 47
Total Employees = 1,470

BEFORE MODEL

16.1%

Assumption: 50% of TP
receiving treatment will
not experience attrition.

$$AR = 0.5(TP) + FN / \text{Total } e^*$$

AFTER MODEL

9.6%

DECREASE

6.5%

Description:
TP = Employees predicted for Attrition and actual Attrition
FN = Employees predicted No Attrition but actual Attrition
e = Employee
AR = Attrition Rate

Hiring Cost

\$4700 per person



BEFORE MODEL

Total Employees	1,470 employees
Total Employee Attrition	237 employees
Cost for Hire	\$1,113,900

\$446,500

AFTER MODEL

Total Employees	1,470 employees
Employee Attrition (50% TP + FN)	142 employees
Cost for Hire	\$667,400



Thanks.

Do you have any questions?



CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**

Please keep this slide as attribution