Homework Solution - Stage 3 (Modelling)

Team:
1. Virgo Gilang Pratama (Leader)
2. Faldo Aditya
3. Putri

Submission:
1. Report: (https://github.com/zerobase-one/Pengabdi-FinPro.git)
2. Notebook: (https://colab.research.google.com/drive/1PFYTh519RwDpR6SQikEsxsPCH0LNmR8g?usp=sharing)

Pre-Modelling

- **Split Data Train & Data Test**
- **Scale Data Train & Data Test**
  Menggunakan:
  - `X_train_scaled = scaler.fit_transform(X_train)`
  - `X_test_scaled = scaler.transform(X_test)`
- **Handle Class Imbalance dengan RandomOverSampler**
- **Feature Selection (corr threshold: >0.07)**

Age	**-0.188209**
DailyRate	**-0.102016**
DistanceFromHome	**0.094112**
JobLevel	**-0.244937**
MonthlyIncome	**-0.277302**
OverTime	**0.295762**
StockOptionLevel	**-0.142361**
TotalWorkingYears	**-0.292519**
TrainingTimesLastYear	**-0.081083**
YearsAtCompany	**-0.242391**
YearsInCurrentRole	**-0.216552**
YearsWithCurrManager	**-0.231257**
Accumulated_Satisfaction	**-0.269434**
YearsWorkingPerCompany	**-0.268648**
BusinessTravel_Non-Travel	**-0.151371**
BusinessTravel_Travel_Frequently	**0.214879**
BusinessTravel_Travel_Rarely	**-0.114497**

```
Department_Research & Development   -0.130844
Department_Sales                     0.139799
MaritalStatus_Divorced              -0.128017
MaritalStatus_Married               -0.091326
MaritalStatus_Single                 0.198954
JobRole_Healthcare Representative   -0.167948
JobRole_Laboratory Technician        0.154243
JobRole_Manager                     -0.130960
JobRole_Manufacturing Director      -0.122460
JobRole_Research Director           -0.187044
JobRole_Sales Representative         0.184594
```

Modelling

- **Train model dengan algoritma:**

  - Logistic Regression
  - Linear SVC
  - SVM
  - KNN
  - Gaussian Naive Bayes
  - Perceptron
  - Stochastic Gradient Descent
  - Decision Tree
  - Gradient Boosting Trees
  - Random Forest

- **Hasil Training Model**

```
Model  Precision Score
7               Decision Tree           100.00
9               Random Forest           100.00
8     Gradient Boosting Trees            94.67
3                         KNN            90.64
1                         SVM            87.12
5                  Perceptron            84.42
0         Logistic Regression            77.32
2                  Linear SVC            76.50
6  Stochastic Gradient Decent            68.76
4                Naive Bayes             58.06

     Model  Precision CV 10-Fold
9               Random Forest                 96.20
8     Gradient Boosting Trees                 89.16
7               Decision Tree                 88.60
```

```
3                          KNN                 82.89
1                          SVM                 82.53
2                   Linear SVC                 76.01
0         Logistic Regression                 75.96
6  Stochastic Gradient Decent                 73.68
5                   Perceptron                 73.46
4                 Naive Bayes                  58.13
```

- **Model Evaluasi 1**

```
Model: Logistic Regression
Precision Scores: [0.78378378 0.69444444 0.7625     0.78082192
0.734375   0.703125
 0.78571429 0.85483871 0.71014493 0.78666667]
Mean Precision Score: 0.760
----------------------------------------
Model: SVM
Precision Scores: [0.82608696 0.81428571 0.80246914 0.85333333
0.84285714 0.8028169
 0.80519481 0.89393939 0.79710145 0.82278481]
Mean Precision Score: 0.826
----------------------------------------
Model: Linear SVC
Precision Scores: [0.80555556 0.70422535 0.7625     0.78082192
0.72131148 0.67692308
 0.77464789 0.88333333 0.72727273 0.76623377]
Mean Precision Score: 0.760
----------------------------------------
Model: KNN
Precision Scores: [0.79487179 0.7654321  0.77011494 0.77272727 0.75
0.80246914
 0.81176471 0.825      0.7875     0.78823529]
Mean Precision Score: 0.787
----------------------------------------
Model: Decision Tree
Precision Scores: [0.84337349 0.8625     0.8452381  0.87654321
0.91025641 0.8875
 0.91025641 0.89873418 0.91025641 0.8452381 ]
Mean Precision Score: 0.879
----------------------------------------
Model: Gradient Boosting Trees
Precision Scores: [0.89189189 0.88       0.85365854 0.92
0.88311688 0.86419753
 0.93243243 0.93150685 0.90789474 0.86075949]
Mean Precision Score: 0.893
----------------------------------------
Model: Random Forest
```

```
Precision Scores: [0.97222222 0.93333333 0.94666667 0.97260274
0.93421053 0.97260274
 1.         0.97260274 0.98611111 0.95945946]
Mean Precision Score: 0.965
----------------------------------------
```

- **Model Evaluasi 2**

```
Model: Logistic Regression
Precision Scores: [0.78378378 0.69444444 0.7625     0.78082192
0.734375   0.703125
 0.78571429 0.85483871 0.71014493 0.78666667]
Mean Precision Score: 0.760
Standard Deviation of Precision Scores: 0.047
----------------------------------------
Model: SVM
Precision Scores: [0.82608696 0.81428571 0.80246914 0.85333333
0.84285714 0.8028169
 0.80519481 0.89393939 0.79710145 0.82278481]
Mean Precision Score: 0.826
Standard Deviation of Precision Scores: 0.029
----------------------------------------
Model: Linear SVC
Precision Scores: [0.80555556 0.70422535 0.7625     0.78082192
0.72131148 0.67692308
 0.77464789 0.88333333 0.72727273 0.76623377]
Mean Precision Score: 0.760
Standard Deviation of Precision Scores: 0.055
----------------------------------------
Model: KNN
Precision Scores: [0.79487179 0.7654321  0.77011494 0.77272727 0.75
0.80246914
 0.81176471 0.825      0.7875     0.78823529]
Mean Precision Score: 0.787
Standard Deviation of Precision Scores: 0.022
----------------------------------------
Model: Decision Tree
Precision Scores: [0.86419753 0.85185185 0.86585366 0.87654321
0.91025641 0.91025641
 0.89873418 0.91025641 0.93421053 0.8875    ]
Mean Precision Score: 0.891
Standard Deviation of Precision Scores: 0.025
----------------------------------------
Model: Gradient Boosting Trees
Precision Scores: [0.89189189 0.88311688 0.85365854 0.92
0.88311688 0.86419753
 0.93243243 0.93150685 0.90789474 0.86075949]
Mean Precision Score: 0.893
Standard Deviation of Precision Scores: 0.028
```

```
----------------------------------------
Model: Random Forest
Precision Scores: [0.98591549 0.93333333 0.95945946 0.98611111
0.93421053 0.97260274
 0.98611111 0.95945946 0.98611111 0.97260274]
Mean Precision Score: 0.968
Standard Deviation of Precision Scores: 0.020
----------------------------------------
```

Based on the precision score, model evaluation 1 & model evaluation 2, we decided to choose the model with lowest standard deviation of precision score and highest Precision Score. So, we decided to choose Random Forest Classifier.

- **Hyperparameter Tuning RandomForestClassifier dengan RandomizedSearchCV**
  Best parameters: {'n_estimators': 100, 'min_samples_split': 5, 'min_samples_leaf': 1, 'max_depth': 50, 'criterion': 'gini'}
  Best precision score: 0.9610206865120349
- **Check Model with Best Params**
  Random Forest Precision Default Before (Training): 1.0

  Random Forest Precision CV 10-Fold Default Before (Training): 0.959349593495935

  Random Forest Precision Best After (Training): 1.0

  Random Forest Precision CV 10-Fold Best After (Training): 0.9528301886792453

  Random Forest Precision Best After (Test): 0.6666666666666666

  Random Forest Precision CV 10-Fold Best After (Test): 0.6428571428571429
- **Learning Curve**

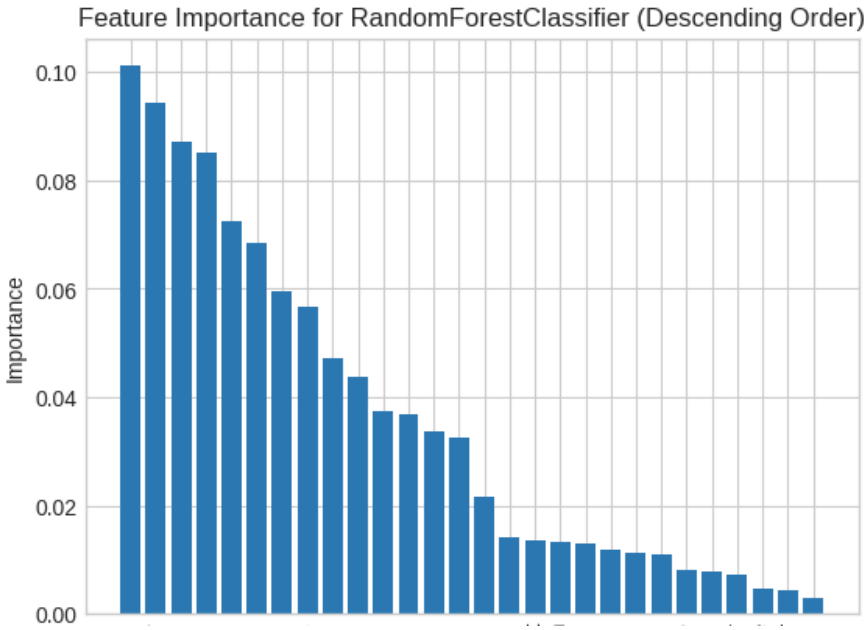
- **Predict Data Test using model with best params**

```
[88] print("Precision: %s" % score)

     Precision: 80.0
```

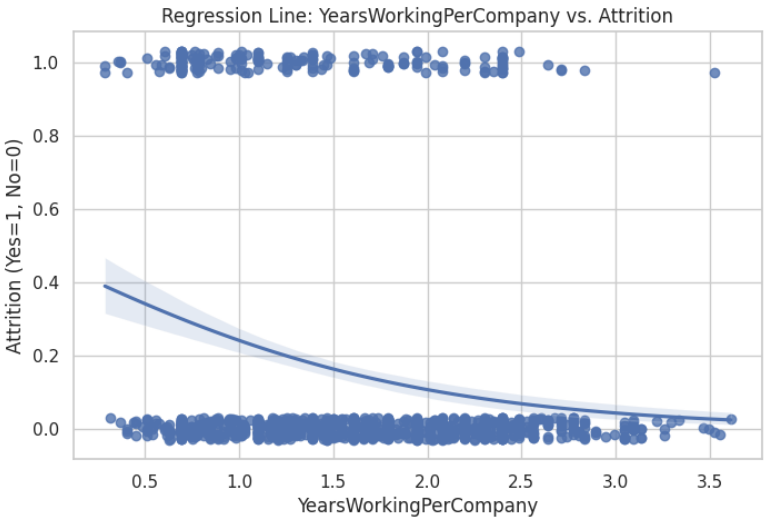
```
[89] print(classification_report(y_test, predictions))

                   precision    recall  f1-score   support

                0       0.86      0.99      0.92       308
                1       0.80      0.14      0.24        58

         accuracy                           0.86       366
        macro avg       0.83      0.57      0.58       366
     weighted avg       0.85      0.86      0.81       366
```

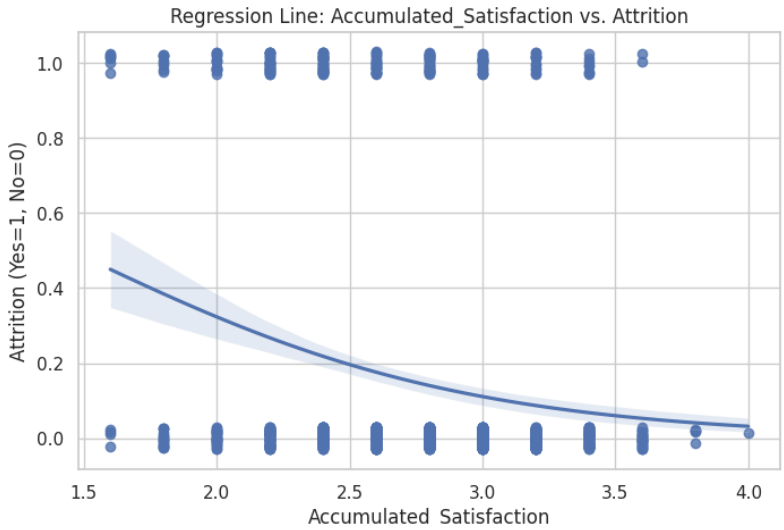
Feature Importance

1. Feature: MonthlyIncome, Score: 0.10109
2. Feature: YearsWorkingPerCompany, Score: 0.09413
3. Feature: Accumulated_Satisfaction, Score: 0.08709
4. Feature: Age, Score: 0.08498
5. Feature: DailyRate, Score: 0.07232
6. Feature: OverTime, Score: 0.06828
7. Feature: DistanceFromHome, Score: 0.05947
8. Feature: TotalWorkingYears, Score: 0.05674
9. Feature: YearsAtCompany, Score: 0.04714
10. Feature: YearsWithCurrManager, Score: 0.04364
11. Feature: YearsInCurrentRole, Score: 0.03737
12. Feature: JobLevel, Score: 0.03689
13. Feature: TrainingTimesLastYear, Score: 0.03364
14. Feature: StockOptionLevel, Score: 0.03251
15. Feature: BusinessTravel_Travel_Frequently, Score: 0.02153
16. Feature: MaritalStatus_Single, Score: 0.01420
17. Feature: Department_Research & Development, Score: 0.01361
18. Feature: JobRole_Laboratory Technician, Score: 0.01320
19. Feature: Department_Sales, Score: 0.01300
20. Feature: MaritalStatus_Divorced, Score: 0.01189
21. Feature: MaritalStatus_Married, Score: 0.01120
22. Feature: BusinessTravel_Travel_Rarely, Score: 0.01107
23. Feature: JobRole_Healthcare Representative, Score: 0.00798
24. Feature: BusinessTravel_Non-Travel, Score: 0.00791
25. Feature: JobRole_Manufacturing Director, Score: 0.00713
26. Feature: JobRole_Sales Representative, Score: 0.00467
27. Feature: JobRole_Research Director, Score: 0.00438
28. Feature: JobRole_Manager, Score: 0.00293



Menarik untuk dianalisis lebih lanjut, bahwa kedua fitur baru yang dihasilkan dari ekstraksi fitur-fitur yang sudah ada menempati posisi Top 3 feature importance. Selanjutnya kedua fitur ini menjadi objek menarik untuk ditelaah lebih lanjut untuk diberikan insight dan rekomendasi.





Grafik di atas menunjukkan bahwa hubungan antara YearsWorkingPerCompany dengan Attrition menunjukkan bahwa semakin besar nilai YearsWorkingPerCompany maka kecenderungan untuk Attrition akan menurun. Hal ini menjelaskan bahwa semakin besar nilai, semakin loyal karyawan tersebut. Dalam upaya meningkatkan loyalitas karyawan, diperlukan aksi nyata oleh perusahaan. Salah satunya dengan melihat fitur Accumulated_Satisfaction yang juga menunjukkan semakin tinggi angka kepuasan karyawan, semakin rendah kecenderungan Attrition dari karyawan itu sendiri. Oleh karena itu, aksi nyata peningkatan kepuasan karyawan secara bersamaan juga akan meningkatkan nilai loyalitas karyawan. Maka, rekomendasi aksi yang dapat dilakukan oleh perusahaan adalah yang berorientasi pada peningkatan faktor kepuasan akumulasi yaitu, Environment Satisfaction, Job Satisfaction, Job Involvement, Relationship Satisfaction dan Worklife Balance karyawan.

RECOMMENDATIONS:

Rekomendasi aksi/program:

1. **Environment Satisfaction:**
    - **To ensure that the compensation and benefits offered to employees are attractive in order to keep employee morale high. The role of compensation and benefits is critical in hiring and retaining qualified and talented employees. (Sinta, Azmieti K., Setiadi, B., Jumawan, Damayanti, Endah S. and Soehaditama, Josua Panatap. (2023). "Employee Retention Strategy: Analysis Path Career, Compensation Benefit, Organization Commitment & Reward System". Retrieved from: https://journal.formosapublisher.org/index.php/eajmr/article/view/3672/3404.)**
    - **Uses various rewards or incentives to attract, keep people and motivate them to achieve personal goals and goals organization. (Sinta, Azmieti**

K., Setiadi, B., Jumawan, Damayanti, Endah S. and Soehaditama, Josua Panatap. (2023). “Employee Retention Strategy: Analysis Path Career, Compensation Benefit, Organization Commitment & Reward System”. Retrieved from: https://journal.formosapublisher.org/index.php/eajmr/article/view/3672/3404.)

2. Job Satisfaction:
    - Focus on training and development. Identifying the training need and giving appropriate training to the employees make them to enhance their skills and motivate them to reach higher positions. (Vetrivel, T., Ramesh, S. and M.S, Kamalaveni. (2019). “A REVIEW OF LITERATURE ON EMPLOYEE RETENTION”. Retrieved from: https://www.researchgate.net/publication/335677274.)
    - It is essential to conduct a survey to measure the satisfaction level of employees in organizations. The reason behind this is that this would help in understanding the satisfaction level of the employees and steps can be taken to improve the satisfaction level. (Vetrivel, T., Ramesh, S. and M.S, Kamalaveni. (2019). “A REVIEW OF LITERATURE ON EMPLOYEE RETENTION”. Retrieved from: https://www.researchgate.net/publication/335677274.)
    - A career path or career path is defined as a set of positions that each employee must go through to reach a certain level of position within a company. (Sinta, Azmieti K., Setiadi, B., Jumawan, Damayanti, Endah S. and Soehaditama, Josua Panatap. (2023). “Employee Retention Strategy: Analysis Path Career, Compensation Benefit, Organization Commitment & Reward System”. Retrieved from: https://journal.formosapublisher.org/index.php/eajmr/article/view/3672/3404.)
    - Security at work place especially in the case of women employees may help in retaining pool of women workforce. (Karumuri, Venkateswarlu and Singareddi, Sriprasanthi. (2014). “EMPLOYEE ATTRITION AND RETENTION: A THEORETICAL PERSPECTIVE”. Retrieved from: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=8a8ba9c73f3e624032e71c31a3cbf1a94f935905.)
3. Job Involvement:
    - The management should consider the ideas and opinions of employees while any decisions are taken with related to work and other matter. Active participation of employees in the decision making process increases the morale of the employee. (Karumuri, Venkateswarlu and

Singareddi, Sriprasanthi. (2014). “EMPLOYEE ATTRITION AND RETENTION: A THEORETICAL PERSPECTIVE”. Retrieved from: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=8a8ba9c73f3e624032e71c31a3cbf1a94f935905.)

- The employees can be empowered in their works by giving autonomy in performing their jobs without more involvement of superiors. (Karumuri, Venkateswarlu and Singareddi, Sriprasanthi. (2014). “EMPLOYEE ATTRITION AND RETENTION: A THEORETICAL PERSPECTIVE”. Retrieved from: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=8a8ba9c73f3e624032e71c31a3cbf1a94f935905.)
- To avoid monotony of work in some jobs, fun related activities, games, get together, stress management programs can be organized. (Vetrivel, T., Ramesh, S. and M.S, Kamalaveni. (2019). “A REVIEW OF LITERATURE ON EMPLOYEE RETENTION”. Retrieved from: https://www.researchgate.net/publication/335677274.)

4. Relationship Satisfaction:
    - Equality and fairness treatment among the employees would create a sense that everyone is treated as one and the same i.e. there is no bias among the employees. (Vetrivel, T., Ramesh, S. and M.S, Kamalaveni. (2019). “A REVIEW OF LITERATURE ON EMPLOYEE RETENTION”. Retrieved from: https://www.researchgate.net/publication/335677274.)
    - It is highly important to conduct exit interview to identify the reasons behind leaving of employees in organizations. This would help in strengthening the strategies for employee retention. (Vetrivel, T., Ramesh, S. and M.S, Kamalaveni. (2019). “A REVIEW OF LITERATURE ON EMPLOYEE RETENTION”. Retrieved from: https://www.researchgate.net/publication/335677274.)
    - Mutual trust between the employer and employee must be created so as to bring a feel that employees are part of the organization and they are working not only for their benefit but also for the benefit of the organization and as well as for the benefit of the society. (Vetrivel, T., Ramesh, S. and M.S, Kamalaveni. (2019). “A REVIEW OF LITERATURE ON EMPLOYEE RETENTION”. Retrieved from: https://www.researchgate.net/publication/335677274.)
5. Worklife Balance:
    - Work from home can be encouraged in organizations where there is no crèche facility which would boost the morale of the female employees. (Vetrivel, T., Ramesh, S. and M.S, Kamalaveni. (2019). “A REVIEW OF

LITERATURE ON EMPLOYEE RETENTION". Retrieved from: https://www.researchgate.net/publication/335677274.)

- **Fair and competitive salaries, performance related incentives may help in retaining the talented employees with the organization for a long period. (Karumuri, Venkateswarlu and Singareddi, Sriprasanthi. (2014). "EMPLOYEE ATTRITION AND RETENTION: A THEORETICAL PERSPECTIVE". Retrieved from: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=8a8ba9c73f3e624032e71c31a3cbf1a94f935905.)**

Reference:

1. **Vetrivel, T., Ramesh, S. and M.S, Kamalaveni. (2019). "A REVIEW OF LITERATURE ON EMPLOYEE RETENTION". Retrieved from: https://www.researchgate.net/publication/335677274.**
2. **Karumuri, Venkateswarlu and Singareddi, Sriprasanthi. (2014). "EMPLOYEE ATTRITION AND RETENTION: A THEORETICAL PERSPECTIVE". Retrieved from: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=8a8ba9c73f3e624032e71c31a3cbf1a94f935905.**
3. **Sinta, Azmieti K., Setiadi, B., Jumawan, Damayanti, Endah S. and Soehaditama, Josua Panatap. (2023). "Employee Retention Strategy: Analysis Path Career, Compensation Benefit, Organization Commitment & Reward System". Retrieved from: https://journal.formosapublisher.org/index.php/eajmr/article/view/3672/3404.**