# Section 1: Assignment P3 - Groupings

To objectively define groups, we can use available employee attributes from SFB's data. Additionally, since the goal is to offer RCCs to employees likely to leave, grouping them by predicted acceptance probability is a reasonable approach. This should reduce the number of RCC offers needed, consequently speeds up negotiations and allows for a faster transition to BAP offices. This directly relates to lowering the operational of the merging companies

Before using predicted probabilities, employees should first be group by departments (HR, Sales, R&D). This ensures that no department experiences overwhelming attrition, by providing number of RCCs across departments relative to their size. In 'Part 2' of the assignment, 8-HR, 60-R&D, and 38-Sales employees were predicted to accept RCC (probability > 50%). Although R&D has the most employees, losing all 40 (study requirement), could disrupt operations and even lead to re-hiring, both drawbacks are costly.

After grouping by departments, next step is to group employees based on their on their predicted RCC acceptance, using predicted probabilities and key features in *Figure 1* as guidance.

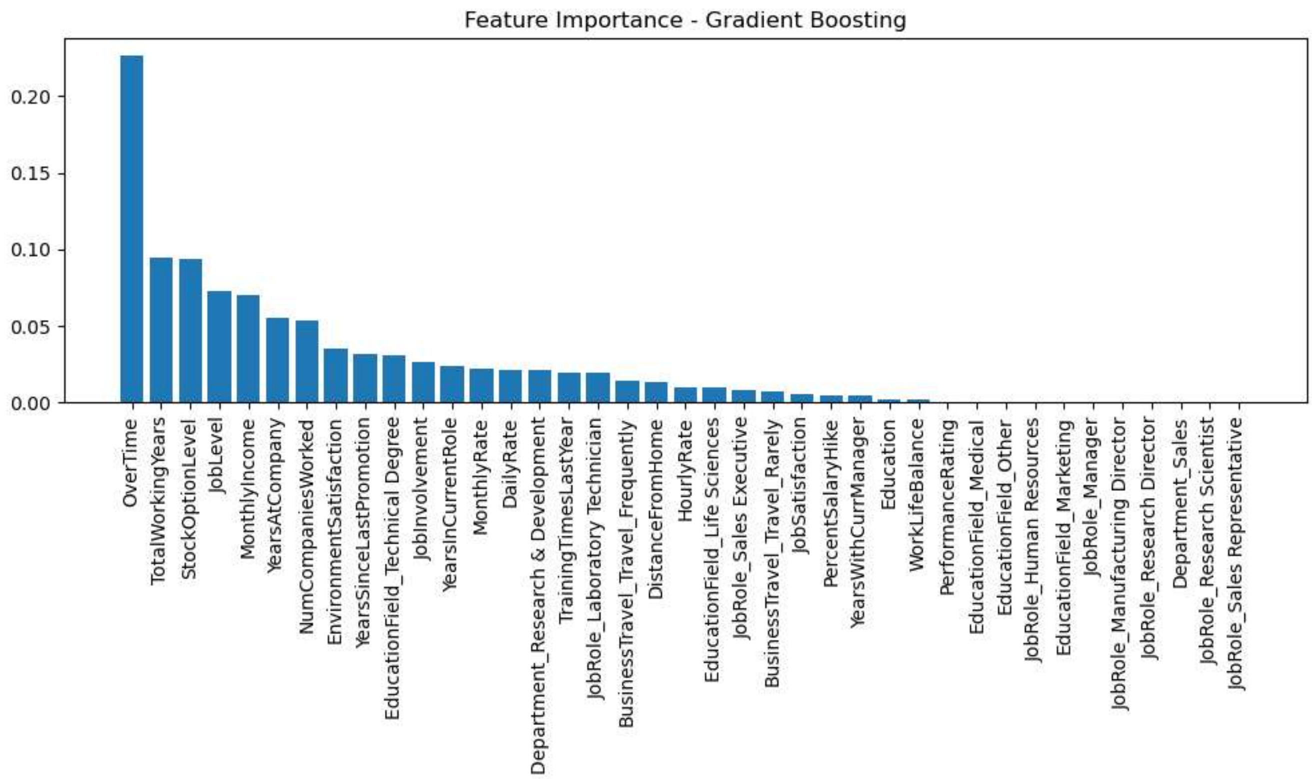


*Figure 1: Feature Importance for Gradient Boosting Model (best found)*

Feature and Partial Dependence plots can help grouping employees by key model features, that are representative of predicted leaving probabilities, allowing for a fair and objective grouping. For example, 'OverTime' feature can define initial group split, as we found employees working overtime, tend to accept RCCs. This results in 5 HR, 74 R&D, and 39 Sales employees. Given only 5-HR employees are likely to leave, grouping them by overtime can be sufficient. R&D and Sales can be further divided using more key features, until reaching controllable size (e.g., 5-10).

Small groups allow for allow for a controlled and precise targeting of employees but may lead to accidental discrimination by demographic (e.g., 5 in HR: 1 female, 4 males). The opposite, larger groups (e.g., 39 in Sales: 21female, 18 male), appear fairer at the and less prone to accidental discrimination, but harder to control. Thess risk must be considered when grouping. Group sizes of 6-10 should balance fairness and control, suitable for the goal of 40 leaves, supporting gradual yet safe approach. Additionally, employees offered RCC in an earlier grouping, should not be re-grouped.

To measure if a group is non-discriminatory, it is required to examine the distribution of protected characteristics within each group and focus on proportional representation. In overtime groups, gender distribution range is 46-80% for male, while overall dataset distribution is 60% male (i.e., splitting by overtime in HR may be considered unfair toward men). Applying this check to other protected characteristics (e.g., Age, MaritalStatus) helps ensure groupings are non-discriminatory (See 'prediction.ipynb', Section 6, for metric used in this section).

## Section 2: Assignment P4 - Excel Solver Optimization

First, my decision variables were binary decider whether to offer a Group an RCC. If its 1, the offer is given and the rest of the metric are calculated, otherwise they are not. The solver would iterate between allowing and not allowing to give offers to specific groups. Second, my objective function was to minimize the total RCC cost, which I calculated as Average RCC Cost * ExpectedLeavers. The RCC cost was calculated using the formula provided to minimum compensation RCC package in the SFB case study, Appendix A2. Finally, the constraints were:

1. Annual Salary Reduction (12 * AverageMonthlyIncome * Expected Leavers) should be at least €3 million,
2. Number of Leavers should at least 40 which references the total EmployeesLeaving,
3. Number of Employees Left in each department should be at least 80% of the original size, by summing the number of leavers from each department and subtracting from original, then comparing to original number of employees * 0.8 for each department,
4. The decision variables should be binary,
5. And number of Employees Leaving should be an integer, which I solved by rounding the number of expected leavers to zero decimal places.

I have defined groups the way I have described in Section 1 of the report. I have implemented a code, that divided the data first by the departments, then in order of importance (from *Figure 1*) divided the data by the features. While doing so, ensuring that each group has 5 to 20 employees in it. The algorithm iterates by the top 6 features as follows, it attempts to divide them by best feature 'OverTime=Yes' if it yields a group with 5-20 employees, it keeps it, if it gets a group with less than 5 employees, it will select the feature as 'OverTime=No' and try the next best feature (i.e. TotalWorkingYears). At the end, the out put should be several groups with 5-20 employees, and several with above 20 employees which would need to use the next best features (i.e., after YearsAtCompany) to divide them further. After obtaining all of the groups, I apply a very simple check for discriminatory grouping, by checking if the group has at least 1 male and 1 female.

For this assignment I have only kept the groups with 20 or less ensuring that at least 40 employees are going to be predicted to leave to use for the optimization. However, using these groups, I was not able to satisfy all constrains. Using solver, the best solution is not able to reach at least €3 million annual salary reduction and violated the HR department constraint. This is likely solver model issue, but the groupings algorithm, if I have used more variables to further divide the large employee groups, it would have likely yielded a better result aligning with the constrains (due to time mismanagement, I am not able to do this).

# Section 3: Assignment P5 – Prediction & Optimization

Combination of ML-based probability prediction and cost optimization provide a data-driven framework for workforce restructuring, aligning employee RCC acceptance likelihood with RCC cost-minimization.

This approach offers several advantages. First, predictive likelihoods help target RCC offers to likely accepters, reducing response collection time. Second, allows for cost minimization, while satisfying set constraints (e.g., saving ≥€3 million). Third, excluding protected characteristics and ensuring fair groupings, reduces subjective bias and improved legal compliance. Finally, the framework can adapt faster to constraint changes (e.g., reduced savings targets) compared to manual processing.

However, there are also limitations. First the model is not perfectly accurate and over/underestimate RCC acceptance, potentially resulting in additional costs. Second, it cannot capture unpredictable human behaviour or external factors influencing decisions. Finally, model-based groupings may appear in-humane, creating tension during RCC negotiations.

The model relies on assumptions that may not always hold in practice. First, it assumes acceptance probabilities to be independent, though a peer effect may exist. Employees may influence each other through unions and relationships (second limitation), resulting in over/underestimate RCC acceptance. Additionally, probabilities are treated as expected values (e.g., $0.7 \times 10 = 7$), though for reasons discussed above, it may differ in practice. Finally, assumption of removing protected characteristics from modelling eliminates bias is not guaranteed, as it can stay with collinear features. This can be improved by reviewing groupings for fairness and consulting Fresh modelling law to ensure compliance.