Individual Coursework Submission Form

Specialist Masters Programme

Surname: Lai	**First Name: Duong**
MSc in: MSc Business Analytics	**Student ID number: 240034730**
Module Code: SMM750	
Module Title: Digital Technologies and Value Creation	
Lecturer: Dr. Philippe Blaettchen	**Submission Date: 08/04/2025**

Declaration:
By submitting this work, I declare that this work is entirely my own except those parts duly identified and referenced in my submission. It complies with any specified word limits and the requirements and regulations detailed in the coursework instructions and any other relevant programme and module documentation. In submitting this work, I acknowledge that I have read and understood the regulations and code regarding academic misconduct, including that relating to plagiarism, as specified in the Programme Handbook. I also acknowledge that this work will be subject to a variety of checks for academic misconduct.

We acknowledge that work submitted late without a granted extension will be subject to penalties, as outlined in the Programme Handbook. Penalties will be applied for a maximum of five days lateness, after which a mark of zero will be awarded.

Marker's Comments (if not being marked on-line):

Deduction for Late Submission:

Final Mark: %

1. Grouping employees

To define the groups, three clustering approaches were tested: K-means and hierarchical clustering within PCA-reduced space, and K-means clustering directly within the feature space. Prior to clustering, attributes such as Age, Marital Status, Gender, and Attrition Probabilities were excluded from both PCA and clustering procedures. This ensures that group formation remains objective and unaffected by these sensitive attributes. Among these methods, hierarchical clustering was selected due to its lowest projected severance package cost within the targeted scenario.

Regarding group size, increasing the number of clusters naturally decreases individual cluster sizes, and vice versa. Smaller groups, while potentially more homogeneous, carry an increased risk of accidental discrimination. Conversely, larger groups are simpler to define but tend to contain greater variance and risk, making outcomes harder to control and predict.

To assess whether the formed clusters are non-discriminatory, each clustering method was evaluated by examining the distribution of Gender, Age, and Marital Status across groups. Distribution plots revealed that all three clustering methods produced balanced demographic distributions, indicating the clusters could safely be used to offer redundancy compensation packages (RCCs) without significant risk of discrimination.

2. Optimization problem

To optimize severance costs, Excel Solver was employed to identify suitable employee groups for offering redundancy compensation packages (RCCs). The optimization problem comprised three main components: the Objective, Decision Variables, and Constraints.

Objective: The goal was to minimize the total severance cost, calculated by multiplying each cluster's base severance cost by its corresponding attrition probability.

Decision Variables: The decision of whether to offer an RCC to a specific group was represented by a binary variable, taking values of either 1 (offered) or 0 (not offered).

Constraints:

- Achieve a total annual salary reduction of at least EUR 3 million.
- Ensure the number of employees who attrite is above 40.
- Maintain departmental stability by limiting each department's attrition rate to a maximum of 20%, thereby preserving at least 80% of each department's original workforce.

The Excel Solver optimization indicated that offering RCCs exclusively to Group 1 (with 107 employees) satisfies all the specified requirements. With this approach, BAP can expect to have around 60 employees accepting the RCCs, saving over EUR 6.5 million a year in salary while paying only a little less than EUR 450,000 for the severance package. The ratio of employees remaining in the HR, R&D and Sales Department is expected to be 87%, 88% and 83% respectively.

3. Pros and cons of prediction-and-optimization approach

The prediction-and-optimization approach employed has several advantages. Firstly, it separates prediction and optimization into distinct processes, simplifying individual optimization. Secondly, this separation offers flexibility, as predictive models can be interchanged or retrained independently when data or conditions change, without affecting the optimization logic. Additionally, this structure enhances scalability by allowing each component to be independently scaled according to its resource requirements, such as increasing computational resources for model training or employing more powerful solvers for optimization tasks.

However, the approach also has notable disadvantages. Predictive models typically minimize error metrics like mean squared error or mean absolute error, whereas optimization processes aim to minimize or maximize different objectives, such as severance costs. Consequently, models optimized purely for predictive accuracy may not necessarily yield optimal outcomes when their predictions are used for decision-making, as minor prediction errors can significantly impact final decisions. Moreover, the predictive model is trained independently from the optimization step, potentially introducing systematic biases that propagate into suboptimal results.

A key assumption underlying this approach is that severance package costs and employee attrition numbers can be accurately estimated by multiplying the total severance and employee counts in each group by their respective attrition probabilities. The same assumption extends to calculating attrition rates across departments within clusters. Although the assumption does not fully incorporate performance metrics (such as accuracy or recall) or the fact that employees decision may be based on other factors not included in the data set features, reducing its likelihood of complete fulfillment, the high overall predictive performance provides a reasonable level of confidence that the optimization will remain effective and beneficial.

References:

- Cooper, A. (2021). *Explaining Machine Learning Models: A Non-Technical Guide to Interpreting SHAP Analyses*. [online] Aidan Cooper. Available at: https://www.aidancooper.co.uk/a-non-technical-guide-to-interpreting-shap-analyses/ [Accessed 7 Apr. 2025].
- shap.readthedocs.io. (n.d.). *Welcome to the SHAP Documentation — SHAP latest documentation*. [online] Available at: https://shap.readthedocs.io/en/latest/ [Accessed 7 Apr. 2025].