Employee Attrition

A SAP project

Roman Poletukhin
13.02.2022

Presentation Plan

— Task Statement
— Working Process
— Results & Analysis
— Overall Conclusion

Task Statement

— The task is regarded as a problem of binary classification since all employees of the company N fall into one of 2 groups: whether the particular employee has attrition or not.

— 4 algorithms of supervised machine learning are used to solve this problem of binary classification with the data given in the task.

— The algorithms were evaluated by several ML-metrics.

— The weights for each feature were analysed.

— As the bigger weight indicates the bigger impact in making an employee attrited, I have a sequence of attrition factors in descending (by significance) order as a result.

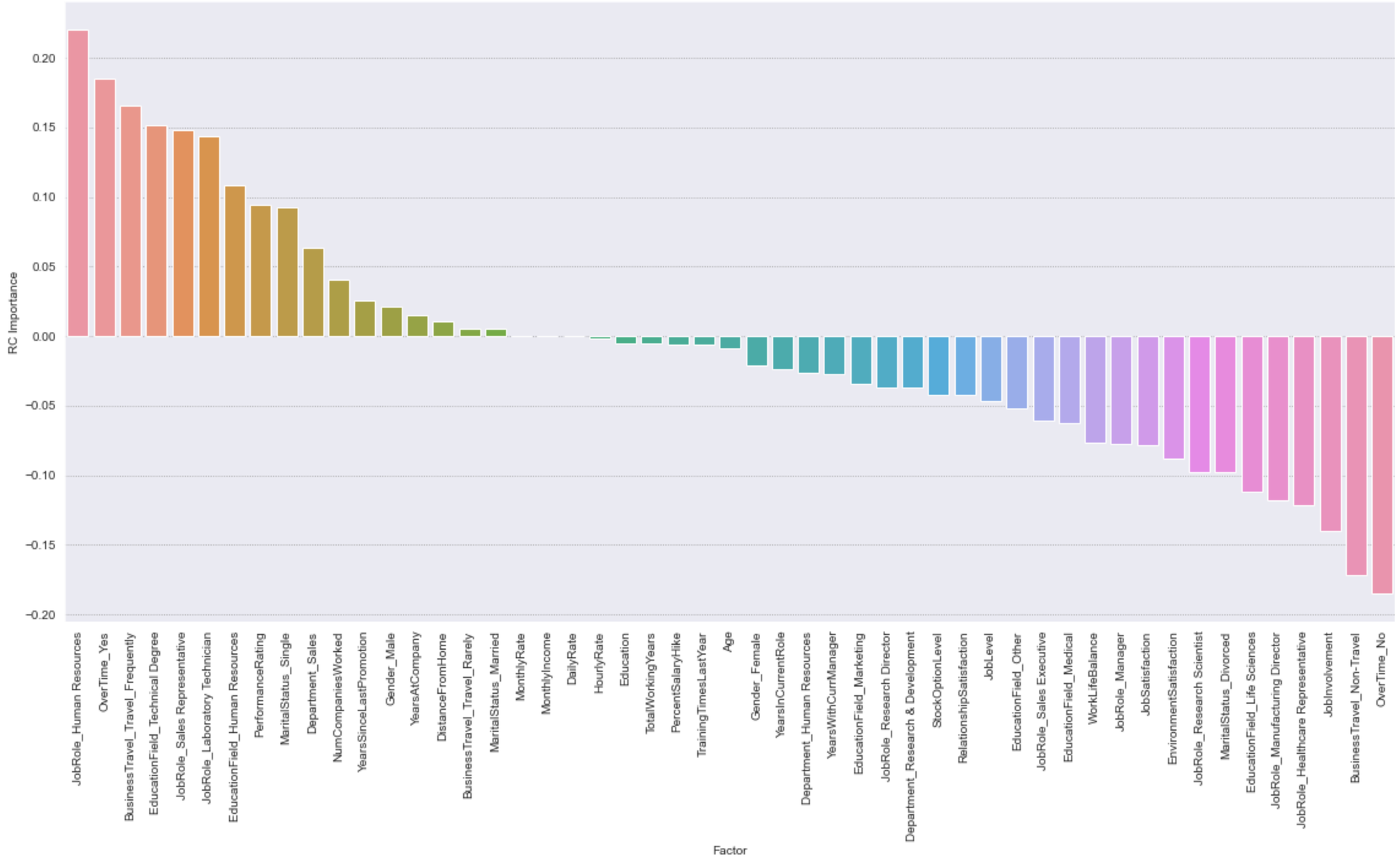
Working Process

The work is described and completed in Jupyter Notebook.

PDF and .ipynb versions are attached to the email.

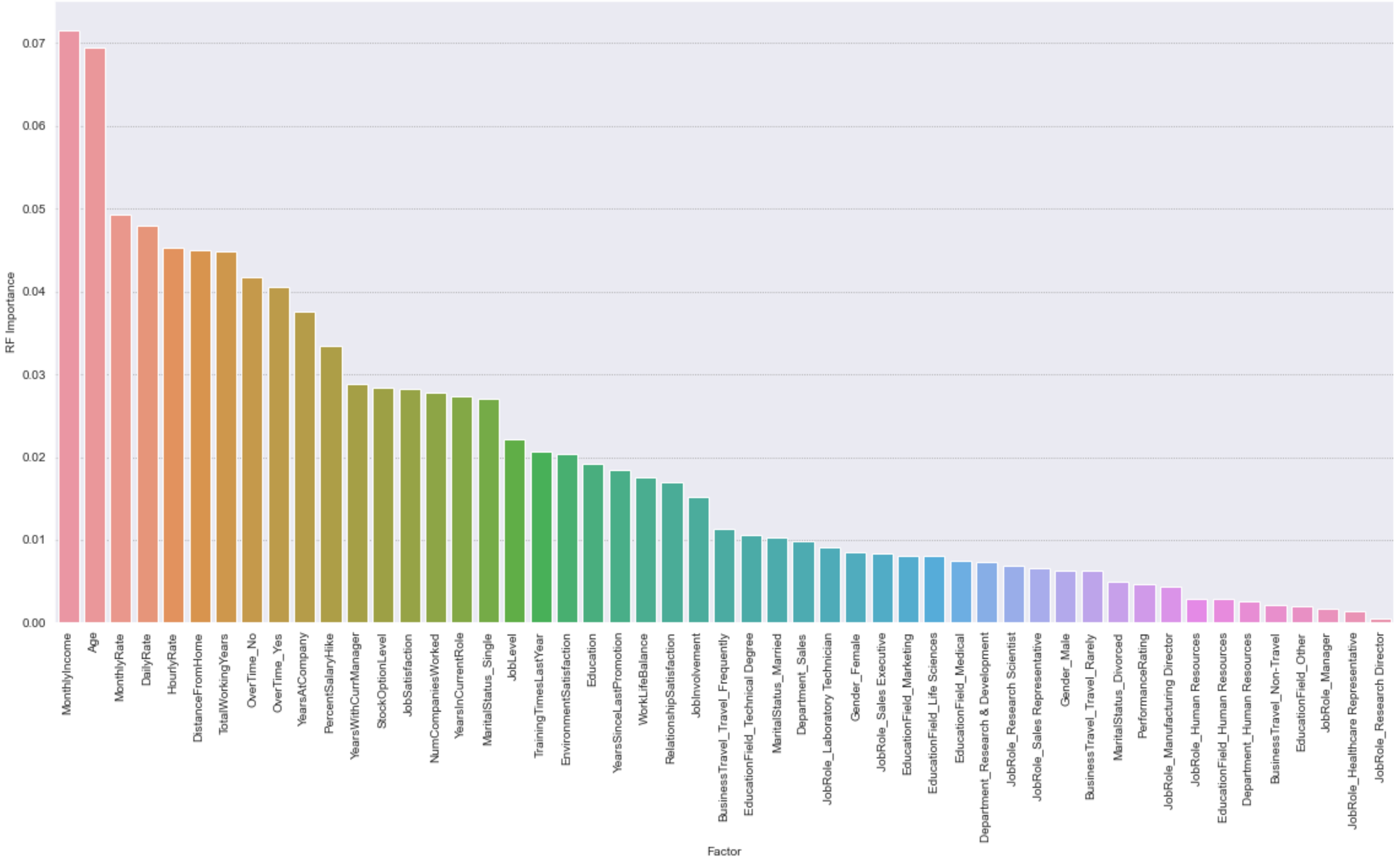
Results

Linear model: Ridge Classifier



Even though the Ridge Classifier Model has high scores (accuracy = 0.87, f1-score = 0,84), the model is not suitable for the purposes of the task: the model focuses on specific features of a person (e.g. Job Role or Education Field) rather than common features (e.g. Age, Distance From Home).
Therefore, if we have a different group of employees where there's no `Sales Representatives` or `Human Resources` present, the model’s most valuable weights will be meaningless.

Ensemble model: Random Forest Classifier



This model has a decent score (accuracy = 0.86, f1-score = 0,82) and doesn’t have the disadvantage of the linear model: all specific features have small weights, the model focuses on common features.

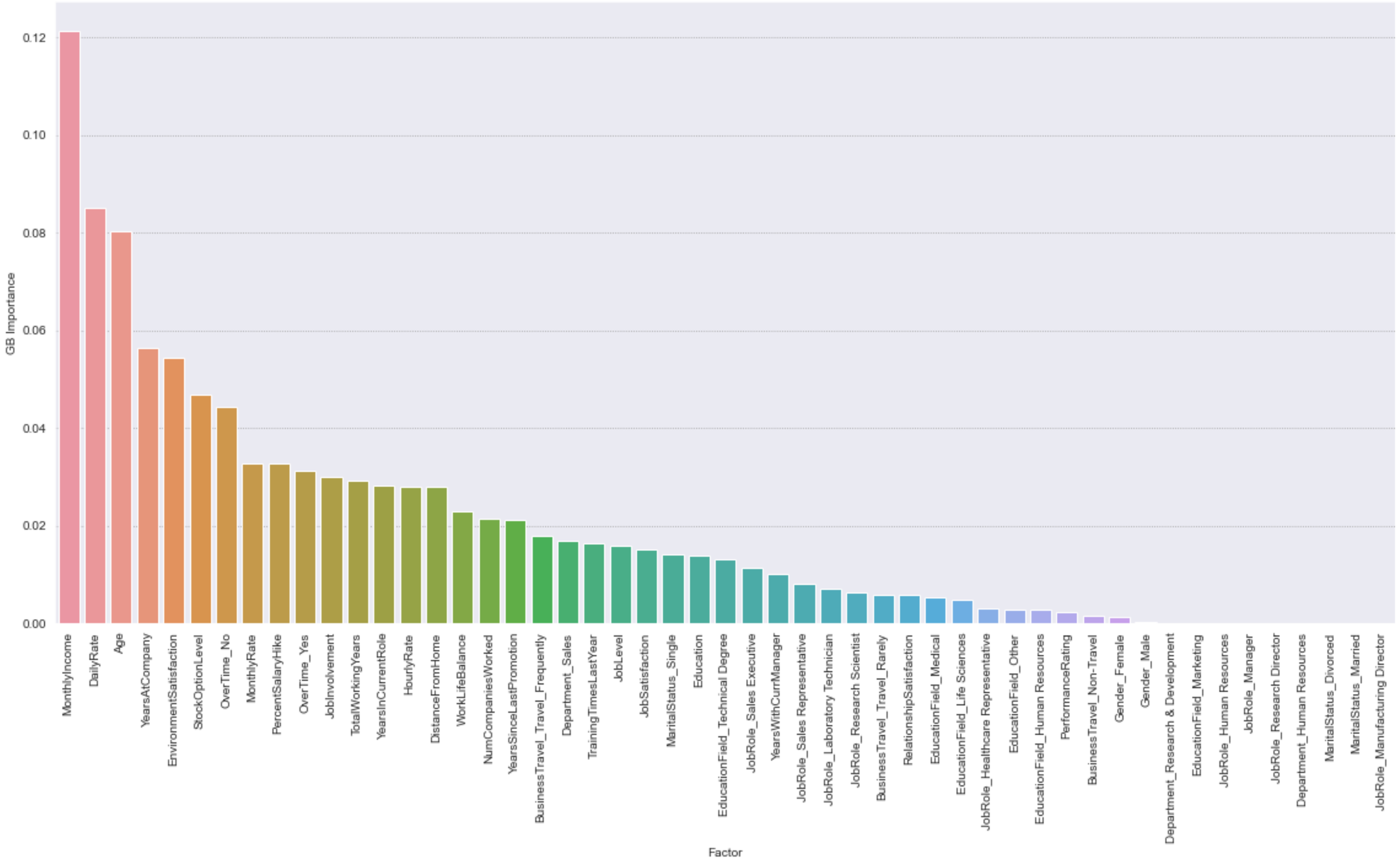
The most important factors in provoking attrition are:

— Money factors (MonthlyIncome, Monthly, Daily and Hourly Rates)

— Time factors (Age and TotalWorkingYears)

— Commuting/Environment factor (DistanceFromHome)

Ensemble model: Gradient Boosting Classifier



Similar situation with this model — unsurprisingly, since both of them are classic ensemble models.

The most important factors in provoking attrition are mostly the same with the following differences:

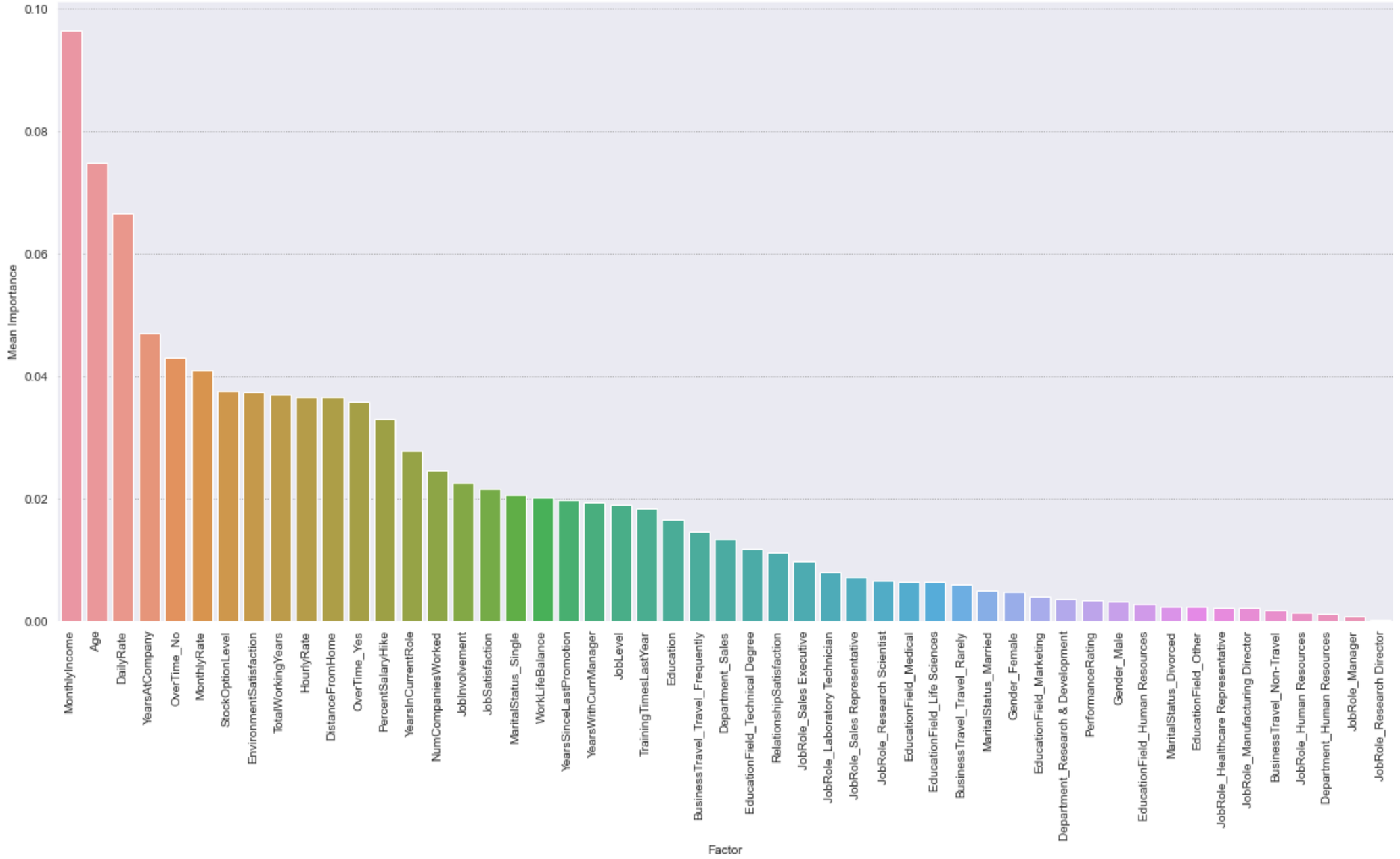
— YearsAtCompany is preferred to TotalWorkingYears

— EnvironmentSatisfaction is preferred to DistanceFromHome

But the overall description of factors remains the same: these are Financial, Time and Environment features.

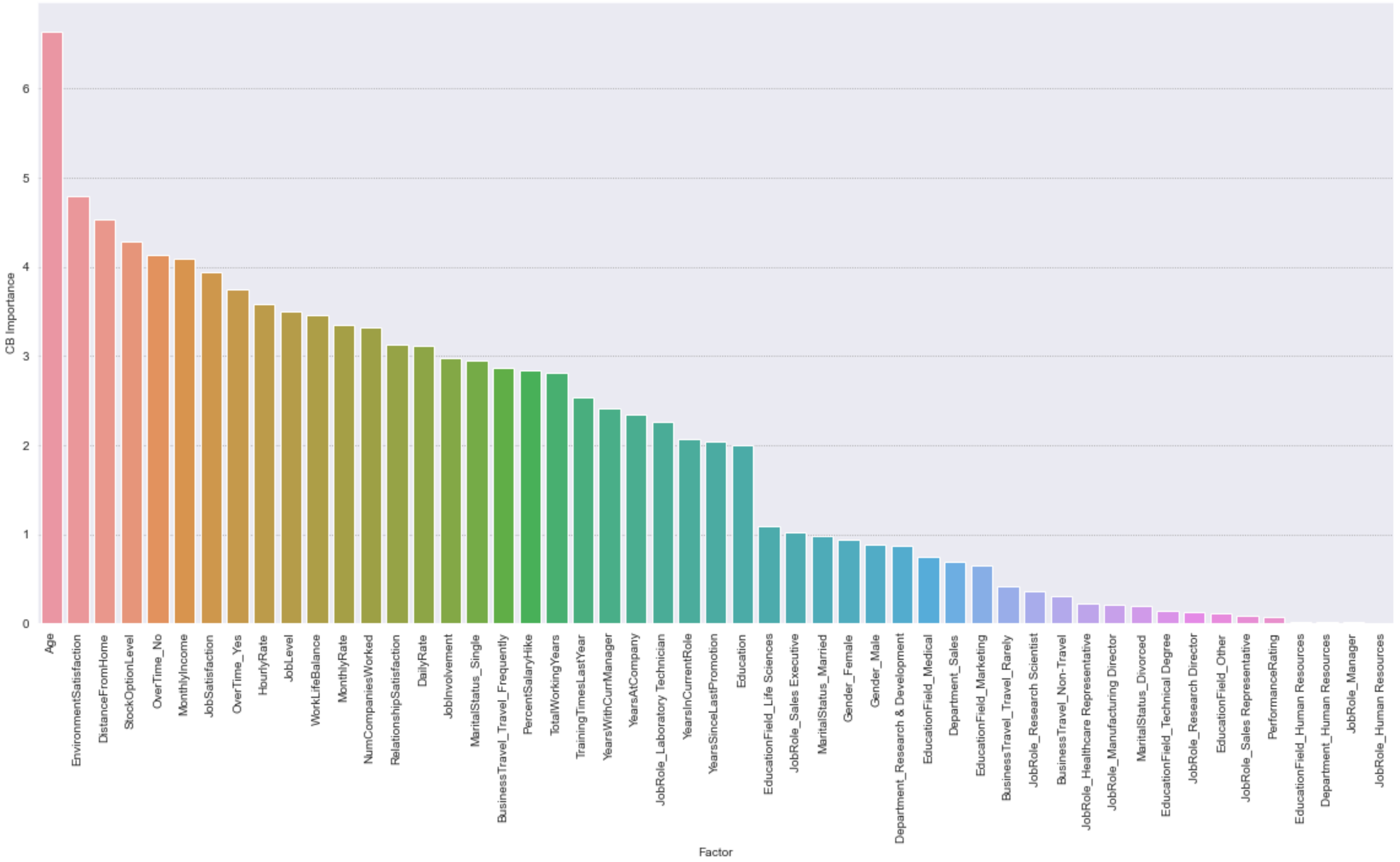
As a naive heuristic approach, let’s look at a mean hierarchy of these two models (since they showed quite similar results in terms of both accuracy metrics and feature importances).

Heuristic ensemble model: mean of GB and RF



Here we obtained a more stair-like graph where the margins between different groups of factors are more visible. The groups themselves, obviously, did not change.

Advanced ensemble model: CatBoost



CatBoost, as the most powerful (and heavy) algorithm of all listed in this presentation, also supported our naive hypothesis about specific features of an employee (EducationField, JobRole, Department): they don’t have a crucial impact in provoking an attrition.

An addition to key factors is StockOption, which is a way to encourage employees to work harder, therefore, increasing the price of company and wealth of employees.

The input of OverTime_Yes and OverTime_No in this model is similar, so their cumulative effect is not regarded.

Conclusion

The major provoking factors in attrition:

- Age of person (in a less degree — years in company or total working years): each valid model classified this factor as highly impactful. Therefore, we might suggest that primarily people of age are prone to experience an attrition, that it is the most vulnerable group.

- Financial factors: surprisingly, the better income, rates and stock options imply the higher vulnerability to attrition. Common sense suggests that the reason might be the higher level of responsibilities and stress, which eventually lead to attrition.

- Environmental issues: the necessity to spend a greater amount of time commuting (DistanceFromHome) tends to lead to attrition. Surprisingly, the higher EnvironmentSatisfaction also has a significant role in attrition: common sense might interpret it as a situation, where the person is already satisfied with his/her life and he/she needs new challenges or motivation to carry on.

Thanks for your attention