**Individual Write-Up  (Helmi K Abraham)**

As part of the Data Analytics assignment, I worked on a group assignment focused on building a data analytics model using RapidMiner to analyze customer churn in the telecommunications industry. This project provided a hands-on opportunity to apply data mining concepts, enhance technical skills, and deliver actionable insights. The objective of our project was to predict customer churn using a public dataset from Kaggle (Telecom Customer Churn). We explored the dataset, performed data preprocessing, and iteratively developed machine learning models.

**Our Approach and Contributions**

As a group of two, we collaborated on every aspect of the project, ensuring equal participation and shared decision-making. Below are the key tasks completed.

* We jointly defined the problem statement: identifying factors leading to customer churn, exploring the dataset for it and then building model which perform the best.
* During the exploration, my first thought was removing the columns with more missing values. Then on in-depth analysis, we understood, there is some influence for the other columns (like phone service and internet service) in that missing values. That is why we filled the missing values for many columns.
* Then created new columns from the existing ones, which are some indicators of internet services, payment methods, contract type with the company etc., which made the modelling easier.
* We collaborated on the preprocessing stage, handling missing values, encoding categorical variables, and normalizing numerical data. And ensured the data was prepared in a format suitable for machine learning algorithms.
* The best lesson learned after pre-processing was importance of backup data, as we lost some during the work.
* After joining tables using a left join to combine customer and zip code tables, we did all pre-processing.
* Then found the relationship between the attributes using visualizations such as pie chart, box plot, scatter plot, bar char, stacked bar chart and word cloud.
* Then after cross validation, compared multiple models like decision tree, random forest, gradient boosted trees by analysing the ROC curves and AUC scores. Created Lift charts for the Final model. Used Performance Operator to get the performance matrix of the model.
* The Decision Tree model was chosen after iteratively tuning parameters such as tree depth, split size, and leaf size and we achieved an accuracy of 86.31%.
* Outer cross-validation and optimization techniques were applied to fine-tune the model further.
* After evaluation and comparison of the final model, we collaborated closely on drafting the joint report, ensuring that all sections—problem definition, data exploration, model building, and evaluation—were clear, concise, and aligned with academic standards.

This project was a valuable learning experience that allowed me to apply theoretical knowledge in a practical setting. By working together at every stage, I can deepen my understanding of the data mining process, particularly the importance of iterative modeling and evaluation. This collaboration also strengthened the teamwork and communication skills, which are crucial for success in data analytics projects.