While pursuing my Master's in Data Analytics at TUS, I had the opportunity to work on a comprehensive project analysing customer churn in the telecom industry. Using the Telecom Customer Churn dataset from Kaggle, which included data on over 7,000 customers , our primary goal was to identify patterns contributing to churn and build a predictive model to assist in proactive customer retention strategies.

**Key Lessons Learned :-**

This project was an incredible learning experience, both technically and personally. Two important life lessons stood out:

* **Always have a plan for worst-case scenarios**: Unexpected challenges in data quality and modelling reinforced the importance of having contingency plans ready at every stage.
* **Always back up your work**: A hard-learned lesson came when critical work was lost, highlighting the importance of consistent documentation and backups.

**From an academic perspective**, the project deepened my understanding of data mining concepts and RapidMiner processes :

* Loading and Retrieving the datasets and joining tables using a left join to combine customer and Geographical information.
* Handling missing values logically (e.g., replacing nulls with "No" or "0" based on service subscription logic).
* Using operators such as Replace Missing Values, Discretize, shuffle, set role, and Generate Attributes, with conditional statements to refine the data for analysis.
* Have Used Visualizations to showcase the distribution & relationship between the attributes such as pie chart, box plot, scatter plot, bar char, stacked bar chart, word cloud and many more
* Explored the Visualization pane and Statistics pane in Rapidminer to get a thorough understanding of the dataset.
* Created dummy variables for categorical attributes to enhance model compatibility.
* Grouped attributes using the Discretize operator for more effective analysis.
* Selected key features for modelling to ensure the inclusion of impactful variables.
* Used Cross Validation technique for strength of the model.
* Compared multiple models (Decision Tree, Random Forest, Gradient Boosted Trees) by analysing their ROC curves and AUC scores.
* Created Lift charts for the Final model.
* Used Performance Operator to get the performance matrix of the model.
* Chose the Decision Tree model for its simplicity and interpretability, iteratively tuning parameters such as tree depth, split size, and leaf size to achieve an accuracy of 86.31%, outperforming an existing published model chosen from the web.
* Validated results using outer cross-validation and applied optimization techniques to fine-tune the model further.

The project was a mixed experience --- a blend of challenges and triumphs. While I encountered setbacks, they taught me resilience and adaptability. At the same time, I gained critical technical skills in data pre-processing, feature engineering, model evaluation, and optimization using RapidMiner. More importantly, comparing my model to published results and outperforming them was deeply rewarding, solidifying my understanding of predictive analytics and its real-world implications.

I believe this project exemplifies my ability to apply data analytics tools effectively while translating insights into actionable strategies, a skill I aim to bring to future endeavours in your company.