

Assignment 1

TelCom Solutions Churn Crisis

Context

TelCom Solutions is one of the leading telecom providers in the country, offering a wide range of services including internet, cable, and phone services to millions of customers. Despite its reputation for innovation and quality, the company has been facing increasing challenges in retaining its customers. Over the past year, customer churn rates have been rising steadily, and this trend has raised alarm bells at the executive level.

The situation reached a breaking point recently with a **major network outage** that affected thousands of customers across multiple regions. This outage led to widespread dissatisfaction, with customers voicing their frustrations on social media. Many have already started switching to competitors who are aggressively marketing their services as more reliable alternatives. **TelCom Solutions' COO highlighted that this churn is not only damaging revenue but also threatening the company's reputation as a market leader.**

Adding to this urgency, TelCom's **competitors have introduced attractive bundles and promotions targeted directly at dissatisfied customers.** Internal reports suggest that unless swift action is taken, the company's market share could face a significant decline.

At a recent board meeting, the CEO shared his "gut feeling" that **15% of customers churn are coming from** international plan. The VP of Marketing added that churn appears higher among customers with monthly contracts, but admitted that the company **lacks solid data to confirm** these assumptions. The board is unified in their demand for **actionable insights, emphasizing the need for a data-driven approach to tackle this crisis.**

The Customer Retention Taskforce

You are part of the newly formed **Customer Retention Taskforce** at TelCom Solutions. This taskforce is composed of cross-functional members, including marketing analysts, customer service specialists, and data scientists like yourself. **The team's primary mandate is to mitigate the impact of churn and stabilize customer retention by leveraging data-driven strategies.**

The taskforce operates under tight deadlines and is tasked with delivering high-priority solutions to the executive team. Your specific responsibilities include:

1. **Data Analysis:** Conduct a thorough examination of customer data to identify patterns and trends contributing to churn. Use this analysis to validate or disprove the assumptions made by the board.
2. **Predictive Modeling:** Build machine learning models to predict which customers are most at risk of churning, allowing the company to take preemptive action.
3. **Strategic Insights:** Provide the executive team with actionable recommendations based on your findings. These recommendations will form the foundation for customer retention campaigns and operational changes.

The work of this taskforce is critical not only to addressing the immediate crisis but also to positioning TelCom Solutions as a more customer-centric and resilient organization in the

future. You have access to detailed customer records, including demographics, service usage patterns, and billing information, which will serve as the foundation for your analysis.

Instructions

1. **Download the Dataset:** Use the **Telecom Churn Dataset** from **Moodle**.
2. **Complete All Tasks:** Follow the tasks outlined below step-by-step.
3. **Prepare Files for Submission:** Submit your work as a ZIP file named: `StudentNumber_Churn_Assignment.zip`. This ZIP file should contain:
 - o A Jupyter Notebook file (`StudentNumber_Churn_Notebook.ipynb`).
 - o A trained model saved as a Pickle file (`StudentNumber_Model.pkl`).
4. **Upload to Moodle:** Submit the ZIP file on Moodle under the appropriate assignment link before the deadline.

Deleted: here:

Question 1 4 points

The CEO claims that 15% of customers churn are coming from international plan. Using the provided dataset validate the CEO's statement and extract additional insights. (In the submitted notebook you should clearly state all the steps done to answer the questions)

Q1.1 – Is the CEO statement correct? [Yes, No] If not, what is the churn percentage? [Numeric] 1 points

Q1.2 – Provide three insights you extracted from the EDA analysis [Free text + code] 3 points

Question 2 4 points

The executive team needs a robust predictive solution to identify customers most at risk of churning. Machine learning models can provide the insights required for targeted retention strategies.

Q2.1. Which kind of machine learning problem is **Telecom** facing? [Multiple choice, regression, classification, recommendation systems, foundational model] 1 point

Q2.2. Train two models that are adecuated to the problem. Justify why you chose those. [Free text + code] In the code document the hyperparameter tuning, train test split and all the processing steps. 3 points why not 3?

Question 3 4 points

To ensure the predictive models are reliable, the taskforce must evaluate their performance using appropriate metrics. This evaluation will determine which model is best suited for deployment.

Q3.1. Which performance metric did you use to evaluate the performance? Why? [Free text + code] 2 points

Q3.2. Which model provided the best results? [Free text] 2 points

Question 4. 4 points

why only two points????

The executive team has provided an external validation dataset with the same structure as the original data. To test the best model, the team must ensure it works seamlessly on new data. The model should predict churn using binary outputs: 0 (No Churn) and 1 (Churn).

- Save the best-performing model as a Pickle file (`StudentNumber_Pipeline.pkl`). Please see the code `Assignment1_SaveAsPickle.ipynb` with the examples
- To ensure the code runs in every machine please save the requirements files with the name `StudentNumber_requirements.txt` - please see `Assignment1_SaveAsPickle.ipynb` on how to do it.
 - For this to happen you will also require submitting a `.py` file with your feature engineer function, please keep the following name `feature_engineering.py`
- If you are using a model outside the scikit-learn you will need to save the following files:
 - `StudentNumber_Preprocessor.pkl` - where the preprocessing steps are performed
 - `StudentNumber_Model.pkl` - where the model does the prediction task.
- Write code to reload the saved model and test it on the external validation dataset (please see `Assignment1_SaveAsPickle.ipynb` on how to do it. The validation set will be in the format from the file `2767ML_assignment1_externalvalidation_data_toStudents.csv`). If your model runs in this dataset it will run in the final validation set.
- Ensure the model outputs predictions in the required format 0 (No Churn) and 1 (Churn).

How This Question Will Be Evaluated:

1. **Top 25% Models:** The models that perform the best on the external validation dataset will receive **full marks for this task**. Performance will be ranked based on accuracy.
2. **Next 25% Models:** Students whose models perform in the following 25% will have 3 and so on.
3. **Non-Functional Models:** If your model fails to load, run, or provide predictions in the required format (0 for No Churn, 1 for Churn), you will receive **zero marks for this task**. Make sure your `.pkl` file and code are functional, tested, and well-documented.
 1. You can find a test set in the assignment **on moodle with the same format as the final** (`2767ML_assignment1_externalvalidation_data_toStudents.csv`).

Question 5. 4 points

The executive team requires actionable insights to guide the strategy to address customer churn effectively. Your analysis will directly inform their decisions

Q5.1 What customer characteristics most strongly influence churn? [Free text] 2 points

Q5.2 What actionable steps should the company take to reduce churn? Suggest two strategies. [Free text] 2 points why only 1.5

Deleted: Firstname_Lastname

Deleted: Model

Formatted: HTML Code, Font:

Formatted

Deleted: <#>

Formatted: HTML Code, Font:

Formatted: Font: Not Bold

Formatted: Font: Not Bold

