

# Customer Churn Prediction

*XYZ Telecom.* is one of the leading telecom companies in the world. They recently noticed that some of their customers are cancelling their subscription for other telecom providers.

To tackle this, they want to incorporate artificial intelligence technology to predict when customers are likely to leave their service. They have requested fusemachines, a *proof-of-concept* for **predicting if a customer will churn or not**.

Their team has provided us with the following historical dataset consisting of details of **customers** with **2499 samples** and whether they churned or not in the “**Churn**” column.

**Dataset:** <http://bit.ly/texam-2021>

## Task

Prepare a solution on the entire Machine Learning process, from data analysis to modeling as well as providing recommendations for the next steps to develop the system.

- Choice of metrics and evaluation setup should be detailed
- Report the performance of the model

Their research team has also requested easy setup to test the model on their holdout set. Which has not been shared to Fusemachines.

**Submission format:** Submission should be a zip file with code (notebook/ scripts), steps to reproduce results, report and model. The report and code can all be in the same notebook as well. It will be evaluated on a held-out set.

### Process and deadline:

Submit by email to [shreesha@fusemachines.com](mailto:shreesha@fusemachines.com), [amit@fusemachines.com](mailto:amit@fusemachines.com), [bijay@fusemachines.com](mailto:bijay@fusemachines.com), [rojesh@fusemachines.com](mailto:rojesh@fusemachines.com), [rajesh.titung@fusemachines.com](mailto:rajesh.titung@fusemachines.com) before 5:30 PM, February 11.

## Evaluation Criteria

Note: The weightage here is rough and not exact. They are only to show where the value is placed.

1. Justification of each step performed in the ML pipeline is desired(20%)
  - a. Project feasibility analysis
  - b. Data Exploration and Visualizations
  - c. Pre-processing
2. Baseline model (40%)
  - a. You are free to choose the metric and optimization criteria. However, a **valid justification** needs to be made for these choices.
  - b. Model Description to detailing how the model works.
  - c. Model Selection and Evaluation setup
  - d. Appropriate Hyperparameter Tuning
3. Documentation and Code-quality (10%)
  - a. Proper documentation (numpy/google)
  - b. Code quality and python guidelines
4. Reproducibility: Should be able to reproduce in other machines (20%)
  - a. Prediction Pipeline (With Model Persistence)
  - b. Environment files and other dependencies
  - c. Steps to reproduce results
5. Conclusion and Recommendations (10%)
  - a. Summaries the results in appropriate way to present to the stakeholders
  - b. Recommendation on the Next steps

# Rules

1. This is an open book exam, you may use the internet but you are forbidden to get help from another person.
2. You are not allowed to share or discuss about the dataset or your code with anyone else.