Customer Churn Prediction and Analysis Project

NAMES:

1- Basmala Said
2- Retag Tarek
3- Menna Samy
4- Nada Mwafi
5- Sandra Samer

Contents

Abstract

Customer churn is a major challenge in the telecommunication industry, leading to significant revenue loss and increased acquisition costs. This project focuses on predicting and analyzing customer churn using data science techniques to help telecom providers proactively retain customers. By leveraging machine learning and advanced analytics, the project identifies key factors influencing churn and enables companies to implement targeted retention strategies. A well-optimized churn prediction model allows businesses to reduce customer attrition, enhance service quality, and improve customer satisfaction. This data-driven approach empowers telecom providers to optimize marketing efforts, personalize customer engagement, and ultimately increase long-term profitability.

Chapter 1: Project planning and management

1.1 Project proposal:

Customer retention is crucial for sustainable business growth, and predicting customer churn is a key strategy for minimizing revenue losses. This project proposes the development of a machine learning-based churn prediction model for the telecommunications sector. The model will leverage historical customer data, analyze behavior patterns, and classify customers based on their likelihood of churning.

Key Objectives:

- Develop a predictive model to identify at-risk customers before they churn.
- Leverage data science techniques for customer segmentation and trend analysis.
- Deploy and monitor the model to ensure long-term accuracy and business impact.
- Provide actionable insights to enhance customer retention strategies.

Project Scope:

The project will encompass data collection, preprocessing, exploratory data analysis (EDA), feature engineering, model training, optimization, deployment, and monitoring. It will also include stakeholder engagement and a final report with recommendations. This project follows a structured data science methodology, combining business understanding, data processing, machine learning modeling, evaluation, and deployment. The methodology consists of the following **key phases**:

1. Methodology:

1.1. Business Understanding

- Define the problem of customer churn and its impact on telecom businesses.
- Identify key business objectives and expected outcomes.
- Align the project with real-world telecom industry needs.

1.2 Data Collection and Preprocessing

- Gather historical customer data from multiple sources (billing, call records, customer service interactions, etc.).
- Handle missing values, outliers, and inconsistencies through data cleaning techniques.
- Normalize and encode categorical features for better model performance.
- Split the dataset into training, validation, and testing subsets.

1.3 Exploratory Data Analysis (EDA)

- Perform statistical analysis and visualization to understand key customer behavior patterns.
- Identify correlations between features and churn rates.
- Detect trends, seasonality, and potential biases in the dataset.

1.4 Feature Engineering

- Select the most relevant features using statistical techniques (e.g., correlation analysis, feature importance ranking).
- Create new features that enhance predictive accuracy.
- Normalize and scale numerical variables to standardize input for machine learning models.

1.5 Model Selection and Training

- Train multiple machine learning models, including logistic regression, decision trees, random forests, and deep learning approaches.

1.6 Model Evaluation and Validation

- Evaluate models using performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC.
- Compare model performances and select the best one based on business requirements.
- Perform bias and fairness checks to ensure ethical AI deployment.

1.7 Deployment and Integration

- Develop an API-based solution or integrate the model into an existing CRM system.
- Implement real-time or batch processing for churn prediction.
- Design dashboards and reporting tools using Power BI/Tableau to provide actionable insights.

1.8 Monitoring and Continuous Improvement

- Establish feedback loops to refine the model based on new customer data.
- Schedule periodic model retraining to adapt to changing market conditions.
- Monitor concept drift and update the system accordingly.

1.2. Project plan:

1.2.1. Project Timeline Overview

This document outlines the essential tasks associated with the project development cycle along with corresponding deadlines for students and graduate participants. Adherence to these timelines is crucial for ensuring project success and timely graduation.

1.2.1. Key Milestones and Deadlines

Item	**Student Deadline**	**Graduate Deadline**
Project Planning & Management	March 21, 2025	February 24, 2025
Literature Review	March 21, 2025	February 24, 2025
Requirements Gathering	March 21, 2025	February 24, 2025
System Analysis & Design	March 21, 2025	February 24, 2025
Implementation (Source Code & Execution)	May 9, 2025	April 11, 2025
Final Presentation & Testing & Reports	May 9, 2025	April 11, 2025

1.2.3. Description of Each Task

1. **Project Planning & Management**
    - Involves the strategic planning phase including defining objectives, resources, and timelines.
2. **Literature Review**
    - A comprehensive review of existing research and materials relevant to the project topic to establish a foundation and rationale.
3. **Requirements Gathering**
    - Systematic collection of project requirements from stakeholders to ensure that all needs and expectations are documented.
4. **System Analysis & Design**
    - Analyzing gathered requirements and designing the system architecture and user interfaces to meet those requirements.
5. **Implementation (Source Code & Execution)**
    - The actual coding and deployment of the project, accompanied by structured testing to ensure functionality.
6. **Final Presentation & Testing & Reports**
    - Presenting the completed project to an evaluation panel, followed by the submission of reports detailing the methodology and findings.

1.2.4. Project Timeline and Milestones

1. **Milestone 1: Complete Data Cleaning**

    Deadline: End of Week 2

    Activities:
    - Data inspection and identifying missing values.
    - Handling outliers and noise.
    - Normalization and transformation.
2. **Milestone 2: Finalize Features for Modeling**

    Deadline: End of Week 3

    Activities:
    - Feature selection using statistical methods (e.g., correlation analysis).
    - Creating new features if necessary.
    - Finalizing the list of attributes for modeling.

3. **Milestone 3: Initial Model Built**
   Deadline: End of Week 5
   Activities:
   - Selecting appropriate modeling techniques (e.g., logistic regression, decision trees).
   - Initial model training and evaluation.
   - Hyperparameter tuning and validation.
4. **Milestone 4: Deliver Final Report**
   Deadline: End of Week 7
   Activities:
   - Compiling results from the analysis and modeling phases.
   - Writing the final report with findings and recommendations.
   - Preparing presentations for stakeholders.

1.2.5. Resource Allocation Based on Deadlines

1. **Project Planning & Management**
   Deadlines:
   Student: March 21, 2025
   Graduate: February 24, 2025
   Resource Needs:
   - Project manager (1)
   - Weekly meetings for updates.
2. **Literature Review**
   Deadlines:
   Student: March 21, 2025
   Graduate: February 24, 2025
   Resource Needs:
   - Research assistant (1-2).
   - Access to research databases.
3. **Requirements Gathering**
   Deadlines:
   Student: March 21, 2025
   Graduate: February 24, 2025
   Resource Needs:
   - Stakeholder interviews (2-3).
   - Documentation tools.

4. **System Analysis & Design**
   Deadlines:
   Student: March 21, 2025
   Graduate: February 24, 2025
   Resource Needs:
   - Data analyst (1).
   - Design software/tools.
5. **Implementation (Source Code & Execution)**
   Deadlines:
   Student: May 9, 2025
   Graduate: April 11, 2025
   Resource Needs:
   - Software developers (2-3).
   - Testing team (1-2).
6. **Final Presentation & Testing & Reports**
   Deadlines:
   Student: May 9, 2025
   Graduate: April 11, 2025
   Resource Needs:
   - Presentation designer (1).
   - Documentation assistant (1).

1.2.6. Project Deliverables

Each phase of the project will result in specific deliverables that track progress and ensure successful completion.

1. **Planning & Data Understanding**
   Deliverables:
   - Project scope document
   - Data dictionary
   - Business insights report

2. **Data Preparation & Exploration**
   Deliverables:
   - Cleaned dataset
   - EDA report
   - Visualizations

3. **Feature Engineering & Model Development**

   Deliverables:

   - Feature selection report
   - Trained models
   - Performance comparison

4. **Model Optimization & Validation**

   Deliverables:

   - Optimized model
   - Model validation report
   - Evaluation metrics

5. **Deployment & Reporting**

   Deliverables:

   - Model deployment strategy
   - Dashboards (Power BI/Tableau)
   - Final project presentation

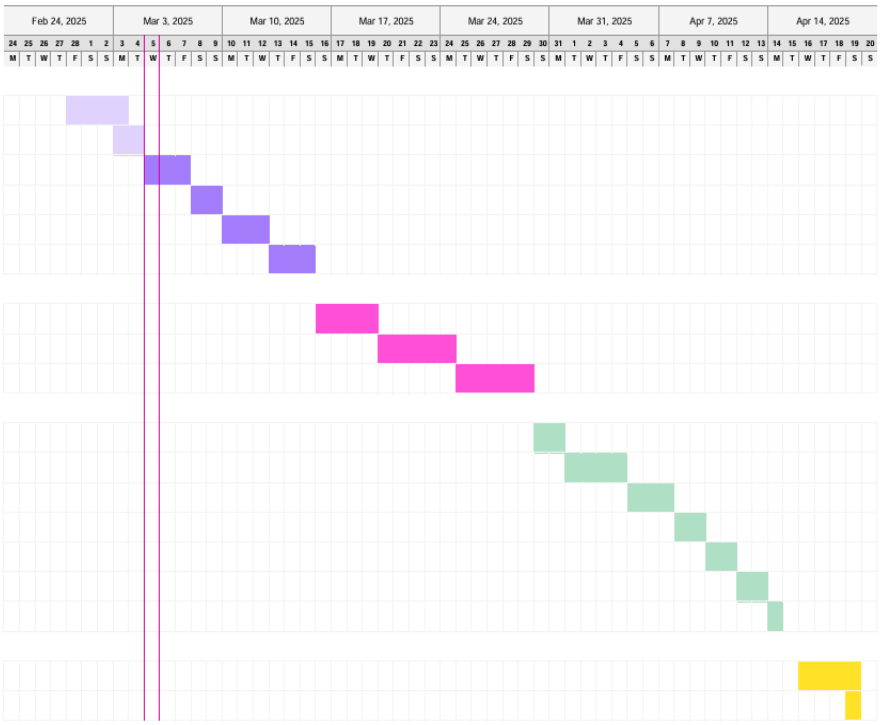
1.3 Task Assignment &roles (Gantt chart):

Customer Churn Prediction and Analysis

Project start: **Fri, 2/28/2025**

Display week: **1**

TASK	ASSIGNED TO	STATUS	START	END
Data Collection and Preprocessing				
Data Collection	Basmala & Menna	100%	2/28/25	3/3/25
Data Exploration	Sandra & Retaj	100%	3/3/25	3/4/25
Exploratory Data Analysis	Nada & Basmala	25%	3/5/25	3/7/25
EDA Report	Menna & Retaj	0%	3/8/25	3/9/25
Interactive Visualizations	Sandra & Nada	0%	3/10/25	3/12/25
Cleaned Dataset	Basmala & Menna	0%	3/13/25	3/15/25
Feature Engineering				
Data Analysis Report	Sandra & Retaj	0%	3/16/25	3/19/25
Enhanced Visualizations	Menna & Nada	0%	3/20/25	3/24/25
Feature Engineering Summary	Basmala & Sandra	0%	3/25/25	3/29/25
Machine Learning Model Development				
Model Selection	Retaj & Nada	0%	3/30/25	3/31/25
Model Training	Basmala & Menna	0%	4/1/25	4/4/25
Model Evaluation	Sandra & Retaj	0%	4/5/25	4/7/25
Model Comparison	Menna & Nada	0%	4/8/25	4/9/25
Model Evaluation Report	Basmala & Sandra	0%	4/10/25	4/11/25
Model Code	Retaj & Menna	0%	4/12/25	4/13/25
Final Model	Sandra & Nada	0%	4/14/25	4/14/25
Final Documentation and Presentation				
Final Presentation	group task	0%	4/16/25	4/19/25
Final Report	group task	0%	4/19/25	4/23/25



1.4 Risk assessment and mitigation plan:

1.4.1 Data Quality and Integration Issues:

Risk: Incomplete, inconsistent, or biased data from various sources (billing, usage, customer service) may lead to inaccurate model predictions.

Solution:

- Implement robust data preprocessing (e.g., data cleaning, normalization, imputation).
- Regularly audit and validate data sources.
- Use a centralized data platform to harmonize data integration.

1.4.2 Model Accuracy, Overfitting, and Concept Drift:

Risk: Models might overfit historical data or become less effective over time due to changing customer behavior (concept drift), leading to false positives or negatives.

Solution:

- Use cross-validation, regularization, and ensemble methods to improve robustness.
- Continuously monitor performance metrics (accuracy, recall, F1-score) and update the model periodically.
- Incorporate feedback loops to adapt to evolving market conditions.

1.4.3 Customer Privacy and Compliance:

Risk: Handling sensitive customer data can lead to privacy breaches or non-compliance with regulations (e.g., GDPR, CCPA).

Solution:

- Anonymize and encrypt customer data.
- Implement strict access controls and regular compliance audits.
- Maintain clear documentation of data handling practices.

1.4.4 Operational and Integration Challenges

Risk: Difficulties in integrating the churn prediction model with our existing technical infrastructure and business processes can hinder project progress.

Solution:

- Start with pilot testing and phased deployment.
- Develop modular, API-based integration solutions.

1.4.5 Dynamic Market Conditions (Concept Drift)

Risk: Shifts in customer behavior or competitive actions can render the predictive model outdated.

Solution:

- Continuously monitor market trends and customer data.
- Schedule regular retraining sessions using the latest data.
- Consider adaptive learning techniques to quickly adjust to new patterns.

1.5 KPIs (Key Performance Indicators)

To evaluate the success of the Churn Customer Prediction project, the following key performance indicators (KPIs) will be tracked:

1. **Model Performance Metrics**
    - Accuracy
    - Precision, Recall, and F1-score
    - ROC-AUC score
    - Log Loss
2. **Business Impact KPIs**
    - Customer churn rate (before and after implementation)
    - Retention rate improvement
    - Revenue saved by reducing churn
    - Increase in customer lifetime value (CLV)
3. **Operational KPIs**
    - Training time and inference time of the model
    - Data processing time
    - Percentage of correctly identified high-risk customers
4. **Adoption and Usability Metrics**
    - User adoption rate of the predictive system
    - Percentage of business decisions influenced by predictions
    - Feedback from stakeholders (e.g., marketing or customer service teams)