Analyzing Customer Churn

Customer churn is painful for all the stakeholders in a company. A business analyst must thus look for ways in which the customer churn rate can be minimised. Additionally, they have to identify the cause behind customer churn to improving business growth. Having a fair idea of which customer is likely to churn out will help a business analyst develop better strategies. Project Idea: In this project, you will be introduced to one of the popular classification machine learning algorithms, logistic regression. The goal is to use logistic regression for estimating the chances of churn for each customer.

**Business Requirements Document (BRD)**

**Project Title**: Analyzing Customer Churn  
**Project Summary**:  
This project aims to identify factors contributing to customer churn and minimize its rate through machine learning techniques like logistic regression. By predicting customer churn, the business will develop strategies for improving customer retention and business growth.

**1. Background**

Customer churn is a critical issue for companies. High churn rates result in revenue losses, and retaining existing customers is more cost-effective than acquiring new ones. Predicting which customers are likely to churn can help the company target those customers with retention strategies.

**2. Business Objectives**

The main objectives of this project are:

* To identify the primary factors causing customer churn.
* To predict which customers are at risk of churning using logistic regression.
* To develop strategies for customer retention based on the analysis.
* To provide detailed and actionable reports to stakeholders that will guide decision-making.

**3. Scope**

**In-Scope:**

* Collecting and analyzing customer demographic, behavioral, and transactional data.
* Using logistic regression to estimate churn probabilities.
* Creating reports and dashboards for stakeholders.

**Out-of-Scope:**

* Implementation of customer retention strategies post-analysis.
* Data collection from third-party sources not directly related to customer churn.

**4. Functional Requirements**

**4.1. Data Requirements**

* Customer demographic data (age, gender, location, etc.).
* Transactional data (purchase history, frequency, value of purchases).
* Customer interaction history (support tickets, complaints, etc.).

**4.2. Data Analysis**

* **Univariate Analysis**: Understanding individual variables affecting churn, such as average transaction value or customer service interaction frequency.
* **Bivariate Analysis**: Understanding relationships between two variables, such as how customer complaints correlate with churn.

**4.3. Predictive Modeling**

* Use **logistic regression** to predict the likelihood of customer churn.
* Evaluate the model using performance metrics:
  + **Confusion Matrix**: To understand the true positives, true negatives, false positives, and false negatives.
  + **Recall**: To measure how well the model identifies actual churners.
  + **Accuracy & Precision**: To assess the model’s overall correctness and relevance.
  + **F1-Score**: To balance precision and recall.
  + **AUC-ROC Curve**: To understand the model’s capability to distinguish between churners and non-churners.

**5. Non-Functional Requirements**

* **Performance**: Te model should be able to process large datasets efficiently.
* **Scalability**: The analysis system must be scalable to accommodate new customer data as it becomes available.
* **Security**: Ensure that customer data privacy is protected in compliance with GDPR and other relevant regulations.

**6. Assumptions**

* Data provided is clean and without significant missing values.
* Stakeholders will provide prompt feedback on reports and model outputs.
* Logistic regression is an appropriate method for this problem, based on the data available.

**7. Key Stakeholders**

* **Senior Management**: Requires overall insights into customer churn and its impact on revenue.
* **Sales and Marketing Teams**: Needs information on which customers to target for retention.
* **Customer Service Team**: Needs insights on how customer interactions contribute to churn.

**8. Reporting Requirements**

The following reports will be generated:

1. **Churn Prediction Summary Report**:
   * A high-level report showing the overall customer churn rate and the number of customers likely to churn.
   * Key metrics: Overall churn rate, customers at risk, precision, recall, accuracy.
2. **Customer Segmentation Report**:
   * Grouping of customers by demographic, transaction value, and likelihood of churn.
   * Visuals like bar graphs and pie charts to display customer segments most likely to churn.
3. **Churn Drivers Report**:
   * A detailed analysis of the factors most contributing to customer churn (e.g., complaints, transaction frequency).
   * SHAP values or similar interpretability metrics to explain why certain customers are predicted to churn.
4. **Churn Model Performance Report**:
   * A technical report showing the performance of the logistic regression model using metrics such as confusion matrix, precision

**9. Project Timeline**

**Phase 1: Data Collection and Cleaning – 2 weeks**

* Collect data from CRM, transactional databases, and customer service logs.

**Phase 2: Data Analysis & Model Building – 4 weeks**

* Perform univariate and bivariate analysis.
* Build and optimize the logistic regression model.

**Phase 3: Reporting & Recommendations – 2 weeks**

* Generate reports and present findings to stakeholders.

**10. Risk Analysis**

* **Data Quality Issues**: Unclean or incomplete data could result in inaccurate churn predictions.
* **Model Bias**: If the dataset is imbalanced (e.g., too few churned customers), the model could be biased toward predicting non-churn.
* **Stakeholder Misalignment**: If stakeholder expectations are not met, it could affect the acceptance of the model and reports.

**Reports for Stakeholders**

**1. Churn Prediction Dashboard**

A visual dashboard that provides:

* The percentage of customers predicted to churn.
* Key churn indicators such as customer demographics, transaction activity, and support history.
* Drill-down capabilities to explore individual customer profiles.

**2. Retention Strategy Insights Report**

A report that includes:

* Insights from churn drivers analysis.
* Suggested retention strategies based on identified risk factors.
* Prioritized customer segments for targeted retention efforts.

A screenshot of a computer

Description automatically generated

A blue pie chart with a few dots

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated