Low-Level Design (LLD)

Churn Analytics

Revision Number – 1.0

Last Date of Revision: 12 – 05 -2023

Ankit Kumar

Document Version Control

Date	Version	Description	Author
25-02-2023	1.0	First Version of Complete LLD	Ankit Kumar

Contents

Abstract

Churn analytics and customer churn are critical concepts in business analytics and customer relationship management. Churn analytics refers to the process of analyzing customer attrition, also known as churn, which occurs when customers cease their engagement or terminate their relationship with a business. Understanding and predicting customer churn is essential for businesses to develop effective retention strategies, enhance customer satisfaction, and maintain long-term profitability. By employing advanced data analysis techniques and machine learning algorithms, churn analytics enables businesses to identify patterns, factors, and early warning signs associated with customer churn. This information helps businesses take proactive measures to retain valuable customers, improve products or services, and optimize marketing and customer engagement strategies. Ultimately, churn analytics plays a pivotal role in fostering customer loyalty, driving business growth, and maintaining a competitive edge in today's dynamic marketplace.

1. Introduction

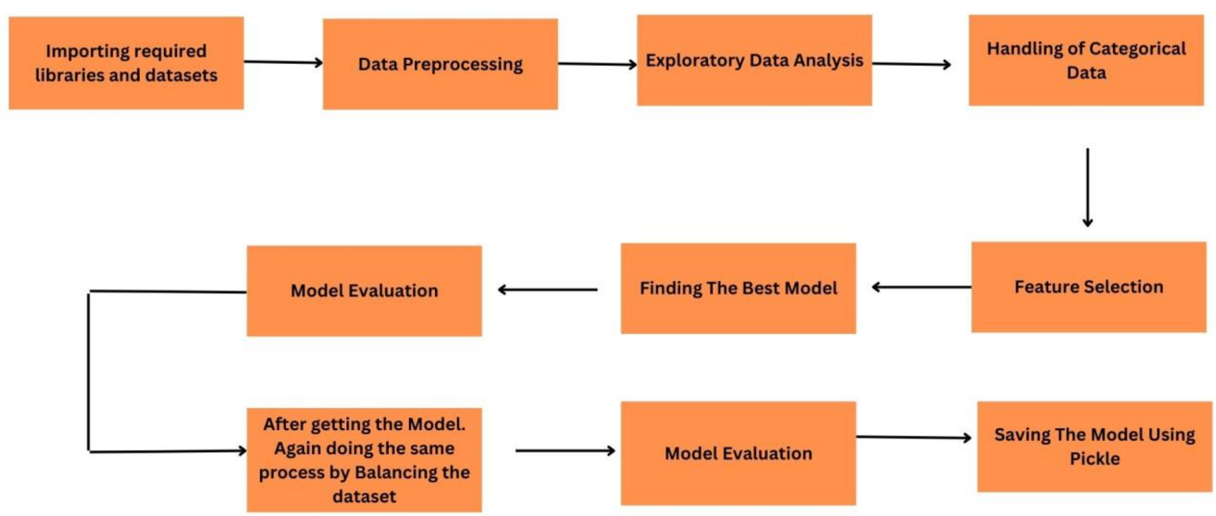
1.1 Why is the LLD Document?

The main goal of the LLD document is to give the internal logic design of actual code implementation and supply the outline of the machine learning model and its implementation. Additionally, it provides the description of how our project will be designed end-to-end.

1.2 Scope

Low-level design (LLD) is a component-level design process that follows a step-by-step refinement process. This process can be used for designing data structures, required software architecture, source code, and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work.

2. Architecture



3. Architecture Design

This project is designed to create a model for the business to prediction of customer churn for different categories.

3.1 Data Collection

The data for these project is provided by the company in which I am doing internship.

3.2 Data Description

Churn Analytics data has 7043 rows and 21 columns.The columns contains information such as customer ID, Gender, Senior Citizen, Dependents, Tenure etc.

A1 customerID

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	customer	gender	SeniorCiti	Partner	Depender	tenure	PhoneSer	MultipleLi	InternetS	OnlineSec	OnlineBac	DevicePrc	TechSupp	Streaming	Streaming	Contract	Paperless	PaymentM	MonthlyC	TotalChar	Churn
2	7590-VHV	Female	0	Yes	No	1	No	No phone	DSL	No	Yes	No	No	No	No	Month-to-	Yes	Electronic	29.85	29.85	No
3	5575-GNV	Male	0	No	No	34	Yes	No	DSL	Yes	No	Yes	No	No	No	One year	No	Mailed ch	56.95	1889.5	No
4	3668-QPYI	Male	0	No	No	2	Yes	No	DSL	Yes	Yes	No	No	No	No	Month-to-	Yes	Mailed ch	53.85	108.15	Yes
5	7795-CFO	Male	0	No	No	45	No	No phone	DSL	Yes	No	Yes	Yes	No	No	One year	No	Bank trans	42.3	1840.75	No
6	9237-HQIT	Female	0	No	No	2	Yes	No	Fiber opti	No	No	No	No	No	No	Month-to-	Yes	Electronic	70.7	151.65	Yes
7	9305-CDSI	Female	0	No	No	8	Yes	Yes	Fiber opti	No	No	Yes	No	Yes	Yes	Month-to-	Yes	Electronic	99.65	820.5	Yes
8	1452-KIOV	Male	0	No	Yes	22	Yes	Yes	Fiber opti	No	Yes	No	No	Yes	No	Month-to-	Yes	Credit car	89.1	1949.4	No
9	6713-OKO	Female	0	No	No	10	No	No phone	DSL	Yes	No	No	No	No	No	Month-to-	No	Mailed ch	29.75	301.9	No
10	7892-POO	Female	0	Yes	No	28	Yes	Yes	Fiber opti	No	No	Yes	Yes	Yes	Yes	Month-to-	Yes	Electronic	104.8	3046.05	Yes
11	6388-TAB	Male	0	No	Yes	62	Yes	No	DSL	Yes	Yes	No	No	No	No	One year	No	Bank trans	56.15	3487.95	No
12	9763-GRSI	Male	0	Yes	Yes	13	Yes	No	DSL	Yes	No	No	No	No	No	Month-to-	Yes	Mailed ch	49.95	587.45	No
13	7469-LKBC	Male	0	No	No	16	Yes	No	No	No intern	No intern	No intern	No intern	No intern	No intern	Two year	No	Credit car	18.95	326.8	No
14	8091-TTVA	Male	0	Yes	No	58	Yes	Yes	Fiber opti	No	No	Yes	No	Yes	Yes	One year	No	Credit car	100.35	5681.1	No
15	0280-XJGE	Male	0	No	No	49	Yes	Yes	Fiber opti	No	Yes	Yes	No	Yes	Yes	Month-to-	Yes	Bank trans	103.7	5036.3	Yes
16	5129-JLPIS	Male	0	No	No	25	Yes	No	Fiber opti	Yes	No	Yes	Yes	Yes	Yes	Month-to-	Yes	Electronic	105.5	2686.05	No
17	3655-SNQ	Female	0	Yes	Yes	69	Yes	Yes	Fiber opti	Yes	Yes	Yes	Yes	Yes	Yes	Two year	No	Credit car	113.25	7895.15	No
18	8191-XWS	Female	0	No	No	52	Yes	No	No	No intern	No intern	No intern	No intern	No intern	No intern	One year	No	Mailed ch	20.65	1022.95	No
19	9959-WOF	Male	0	No	Yes	71	Yes	Yes	Fiber opti	Yes	No	Yes	No	Yes	Yes	Two year	No	Bank trans	106.7	7382.25	No
20	4190-MFLU	Female	0	Yes	Yes	10	Yes	No	DSL	No	No	Yes	Yes	No	No	Month-to-	No	Credit car	55.2	528.35	Yes
21	4183-MYFI	Female	0	No	No	21	Yes	No	Fiber opti	No	Yes	Yes	No	No	Yes	Month-to-	Yes	Electronic	90.05	1862.9	No
22	8779-QRD	Male	1	No	No	1	No	No phone	DSL	No	No	Yes	No	No	Yes	Month-to-	Yes	Electronic	39.65	39.65	Yes
23	1680-VDC	Male	0	Yes	No	12	Yes	No	No	No intern	No intern	No intern	No intern	No intern	No intern	One year	No	Bank trans	19.8	202.25	No
24	1066-JKSG	Male	0	No	No	1	Yes	No	No	No intern	No intern	No intern	No intern	No intern	No intern	Month-to-	No	Mailed ch	20.15	20.15	Yes
25	3638-WEA	Female	0	Yes	No	58	Yes	Yes	DSL	No	Yes	No	Yes	No	No	Two year	Yes	Credit car	59.9	3505.1	No

3.3 Data Preprocessing

- Checked for info on the Dataset, to verify the correct datatype of the Columns.
- Checked for Null values, because the null values can affect the accuracy of the model.
- Performed Data Visualization using Matplotlib and seaborn to get the impact of features on Customer Churn.
- Performed Feature Engineering by removing the unnecessary columns and also to convert the categorical columns to numerical columns.
- Checking the distribution of the columns to interpret their importance and found that data is imbalanced so decided to make model before balancing the dataset and then after balancing it.

- Now, the info is prepared to train a Machine Learning Model.

3.4 Model Creation

The Preprocessed info is now envisioned and drawn insights help us to select the feature that improves the accuracy of the model. The info is randomly used for modeling with different machine learning algorithms to create a model to predict Customer Churn. After performing on different algorithms, we use Naïve Bayes.