PROJECT : PREDICTING CUSTOMER CHURN FOR A TELECOM COMPANY

PROBLEM TYPE: CLASSIFICATION

1. **Domain: Telecommunication**

2. Problem Statement

The major problem of Telecom companies is customer churn (loss of a customer to a competitor). The acquisition of a new customer is very expensive compared to retention of existing customers. Small percentage of reduction in churn improves huge margins to Telecom companies. The companies perform various target marketing activities or reward customer through offers and incentives, to retain a customer if he is identified before he churns. The current challenge, is to develop a deep learning based engine that predicts customers who are likely to churn.

Metric: Accuracy

3. Data Description

Data presented contains attributes related to users' information, the usage plans, charges billed, payment method etc, and the target column of interest is if the user has churned out or not. The task is to build a predictive model that can predict user ratings with reasonably good accuracy and sensitivity.

Data Definition: This dataset contains about 7k instances and 20 features. Each row corresponds to a user's historical data with usage, payment information, etc., and includes the following variables:

- customerID : Unique ID of the customer
- gender : Whether the customer is a male or a female
- SeniorCitizen : Whether the customer is a senior citizen or not (1, 0)
- Partner : Whether the customer has a partner or not (Yes, No)
- Dependents: Whether the customer has dependents or not (Yes, No)
- tenure: Number of months the customer has stayed with the company
- PhoneService: Whether the customer has a phone service or not (Yes, No)
- MultipleLines: Whether the customer has multiple lines or not (Yes, No)
- InternetService: Customer's internet service provider (DSL, Fiber optic, No)
- OnlineSecurity: Whether the customer has online security or not (Yes, No)
- OnlineBackup: Whether the customer has online backup or not (Yes, No)
- DeviceProtection: Whether the customer has enabled device protection or not (Yes, No)
- TechSupport: Whether the customer has assisted technical support or not (Yes, No)
- StreamingTV:Whether the customer has streaming TV or not (Yes, No)
- StreamingMovies:Whether the customer enabled the streaming services or not (Yes, No)
- Contract: Whether customer opted for short-term or long-term contracts(Month-to-Month, One year, Two year, Credit card (automatic), Bank transfer (automatic))

- PaperlessBilling: Whether the customer has opted for paperless billing or not (Yes, No)
- PaymentMethod: Modes of payment(Electronic check, Mailed check,
- MonthlyCharges: Monthly charges
- TotalCharges: Total Charges

6. Approach to be followed:

i. Load the data and understand it; You will observe that its predictors belong to three different types, numeric/integer, categorical.

ii. Exploratory analysis to analyse the data.

iii. Do necessary type conversions

iv. Columns like CustomerID can be removed from the analysis

v. Split the data into train and validation sets and performing prepro-cessing appropriately on each of them.

- Deal with missing values if any
- On numeric data : apply a standardisation technique, preferably using standard scaler
- On categorical data: Apply one-hot encoding/label encoding as appropriate

vi. Build deep neural net model, compile and fit the model. Tune it to improve validation accuracy/recall. Observe the performance

vii. Using auto encoders, get deep features for the same input, and using the deep features, build and tune to a good model and observe the performance

viii. Also, as there is class imbalance in the data, and recall, being an im-portant metric for this problem is highly effected by the imbalance,

try to work on mitigating the effect of class imbalance. Explore parameters like class weight while fitting the model, and analyse the performance