AIFL | MODULE PROJECT

Supervised Learning

TOTAL SCORE 30

Part A - 15 Marks

- **DOMAIN:** Telecommunications
- **CONTEXT:** Singapore Telecommunications Limited (commonly abbreviated as Singtel, and previously stylised as SingTel) is a Singaporean multinational telecommunications conglomerate and one of the four major telcos operating in many countries of North America and Europe.
  As a data scientist we have been asked to analyze the data and find some intuitive business information out of which so that they can reduce the customer churn rate. Our task is to look into the data and find out some intuitive information and run a ML model to predict the customer who is likely to get churn.
- **DATA DESCRIPTION:** Recently They have done a customer survey on many parts of USA to find out the customers who are likely to churn. They have collected various parameters like regarding calls like total evening minutes, total day charge, states, area-code, international plans etc.

• **State** - state of the customer	• **Total day minutes** -Total day minutes used
• **Account Length** - how long the account has been active	• **Total day calls** - Total day calls made
• **Area Code** - Area Code	• **Total day charge** - Total day charges
• **International Plan** - Does the customer have any international plan or not	• **Total eve minutes** - Total Evening Minutes
• **Voice mail Plan** - Does the customer have any voice mail plan or not	• **Total Eve Calls**:Total Evening Calls
• **Number vmail messages** - Number of voice mail messages	• **Total Eve charge**:Total Evening Charge
• **Number Customer Service calls**: Number of Customer Service calls	• **Total Night Charge**: Total Night Charge
	• **Total Night Minutes**: Total Night Minutes
	• **Total Night Calls**: Total Night Calls
	• **Total International Minutes**: Total International Minutes
	• **Total Intl Calls**: Total International Calls
	• **Total Intl Charge**: Total International Charge
	• **Churn(Target)**: Whether the customer is churned or not

- **OBJECTIVE:**
  To use supervised learning techniques such as logistic regression in order to identify customers along with some intuitive findings from the data.
- **STEPS AND TASKS:**
  1. **Univariate Analysis: [4 Marks]**
     - A. Import and Read 'Telecom Churn Data.csv'. [1 Mark]
     - B. Check the target variable. State whether the data is imbalance or balance. [1 Mark]
     - C. Check the distribution of Total international Calls and share insights for the same.[2 Mark]
  2. **Bivariate Analysis: [6 Marks]**
     - A. Study the relationship between International Plan Vs Churn & Churn vs Number of Customer Calls using Bivariate Analysis and state the inference clearly. [2 Marks]
     - B. Study the relationship between Churn vs Total Day Charge & Churn vs Total Night Charge using boxplot. State the inference Clearly. [3 Marks]
     - C. Split the dataset into 80:20 (i.e. 80% train and 20% test) [1 Marks]
  3. **Model Training: [5 Marks]**
     - A. For further analysis train a logistic regression model and Naive Bayes model.
       Evaluate the model using Accuracy, ROC-AUC curve(optional),[5 Marks]

Part B - 15 Marks

- Refer attached problem statement 'AI POC - Template for.pptx' for Part B
- A Case study has been given.
- Learners are supposed to create a sample POC for the given case study.
- Expected submission format: '.pptx' and '.pdf' of the same 'pptx'