AAM IPL Week-4
Project - Cross-Sell Vehicle Insurance Prediction

B.Tech – CSE(AIML)
V Semester - ML and AIUP, Aug-Oct 2024
Department of Computer Science Engineering – AI and ML (CSM)
G.Pulla Reddy Engineering College (Autonomous), Kurnool, AP

Algorithm of Application
Logistic Regression

Project Title
Cross-Sell Vehicle Insurance Prediction

Object and Project Statement
A healthcare insurance company that has a sizeable existing customer for its health insurance product needs your help in building a ML to predict whether the existing health insurance customers/policyholders will also be interested in buying vehicle/automobile insurance provided by the company (2,3, and 4-wheelers) – such selling of another product to an existing customer is called cross-selling.

Vehicle insurance is a mandatory insurance that needs to be bought by each vehicle owner and usually purchased/renewed annually.

Building an ML model to predict whether a customer would be interested in Vehicle Insurance is extremely helpful for the company because it can then accordingly plan its communication strategy to reach out to those customers and optimise its business model and increase revenue.

Dataset
- Description
  - The provided dataset - train and test CSVs – contain information about demographics (gender, age, region code type), vehicles (vehicle age, damages), healthcare policy (premium, sourcing channel) etc. The train CSV contains values for the response column whereas the values for this column are blank and need to be predicted and filled in the test CSV file.
  - The description of fields in train and test data sets are provided below.

Variable	Definition
id	Unique ID for the customer
Gender	Gender of the customer
Age	Age of the customer

Driving_License	0 : Customer does not have DL, 1 : Customer already has DL
Region_Code	Unique code for the region of the customer
Previously_Insured	1 : Customer already has Vehicle Insurance, 0 : Customer doesn't have Vehicle Insurance
Vehicle_Age	Age of the Vehicle
Vehicle_Damage	1 : Customer got his/her vehicle damaged in the past. 0 : Customer didn't get his/her vehicle damaged in the past.
Annual_Premium	The amount customer needs to pay as premium in the year
Policy_Sales_Channel	Anonymized Code for the channel of outreaching to the customer ie. Different Agents, Over Mail, Over Phone, In Person, etc.
Vintage	Number of Days, Customer has been associated with the company
Response	1 : Customer is interested, 0 : Customer is not interested

- Dataset Details:
  - Number of rows: 3,81,109
  - Number of Features: 10
  - Response Variable (Target): Customer is interest or NOT interested (1 or 0)
- Features:
  - Gender
  - Age
  - Driving_License
  - Region_Code
  - Previously_Insured
  - Vehicle_Age
  - Vehicle_Damage
  - Annual_Premium
  - Policy_Sales_Channel
  - Vintage
- Usage in Machine Learning:
  - The dataset is used to predict the binary response/interest of the customer in buying another product from the same company.
- Data Source and Published By:
  - The data is published by Anmol Kumar on Kaggle - Health Insurance Cross Sell Prediction (kaggle.com)
- Data Download Link
  - Kaggle - Health Insurance Cross Sell Prediction (kaggle.com)

Implementation Steps

1. Load the train.csv and test.csv datasets using pandas
2. Print the number of rows in the training and test datasets
3. Check for any missing values in the datasets
4. Encode categorical features
5. Standardise the data and train the model
6. SMOTE technique may be used optionally for handling data imbalance
7. Optional hyper parameter tuning can be employed
8. Evaluate the model using Metrics – Classification report, confusion matrix etc.
9. Plot RoC and Precision Curves
10. Plot Feature Importance
11. Prediction on Test Data and Create Submission File

Project Files Provided

- Project shell code file - ***AAM-IPL-Wk-4-LogiReg-Cross-Sell-Vehicle-Insurance-Prediction-Shell-Code-V2.ipynb***
- Vehicle Insurance Cross-Sell Training Data - ***train.csv***
- Vehicle Insurance Cross-Sell Test Data for Prediction - ***test.csv***
- Watermark image for plots - ***AAM-IPL-Header-6.png***

Project Overview, Implementation and Submission Timeline

05-09-2024 – Saturday 10:00 AM – Next Week Project Details Announcement – Topic, Data Set, Shell Code etc. Announcement Channels – Google Class, Industry Projects WhatsApp Group.			
6	05-10-2024 - Saturday Duration: 1.5 Hrs	Logistic Regression – Model Overview/Recap, Project Description, and Interactive Q&A	Online – Google Class
7	06-10-2024 - Sunday Duration: 1.5 Hrs	Logistic Regression – Model Building, Output Demonstration, Q&A	Online – Google Class
10-10-2024 – Sunday 11:59 PM - Deadline to upload the project code submission by all students in Google Class.			

Guest Lecture Timings:
Saturdays: 10:30 AM IST – 12:00 Noon IST
Sundays: 10:30 AM IST – 12:00 Noon IST
Mondays: 6:30 PM IST – 8:00 PM IST

Development Environment

- Computing Language – Python
- IDE – Visual Studio Code with Jupyter Notebook

Instructor

Instructor	
Venkateswar Reddy Melachervu (alumnus of GPREC, ECE Class of '92) CTO, Brillium Technologies, Bengaluru Email: **venkat.reddy.gf@gprec.ac.in** Profile: LinkedIn	Visiting Faculty

Coordination

All the activities of this programme – lecture venues, weekly projects details announcements, general announcements, changes in lecture timings, etc. will be coordinated by CSM faculty member Sri V.Suresh.
Channels of Communication and Announcements

- Google Classroom
- Whatsapp group **- Applied AI & ML Industry Projects Lab**
- Emails (Strictly GPREC email addresses only)

Programme Coordinator	
Prof. V.Suresh Email: **vsuresh.ecs@gprec.ac.in**	Faculty Member, CSM

© "...dare to dream; care to win..."

Reference Books

- Pattern Recognition and Machine Learning by Chris Bishop, 2006 – PDF Link
- Machine Learning using Python by Manaranjan Pradhan and U Dinesh Kumar, Wiley 2019 – PDF Link

Policies

- **Attendance:** All sessions are expected to be attended by all the enrolled students. In case of inability to attend, prior information is expected to be provided by the student to the coordinator with a copy to the visiting faculty
- **Project Submissions:** Duly completed projects (Jupyter NB file and a PDF of the Jupyter NB file) are expected to be submitted through google class prior to the deadline. In case of inability to complete due various unforeseen circumstance, students are expected to seek extension for the submission deadline.
- **Academic Integrity:** Students are expected to uphold the highest standards of academic integrity in all assignments for the Applied AI & ML Industry Projects Lab. Each assignment must be the student's own work, and all sources and collaborators must be properly acknowledged. By submitting their completed project source code, students confirm that they have adhered to this integrity policy and completed their work in an honest and ethical manner.

Additional Information

For students interested in engaging with special projects in the field of Gen AI, please reach out to the visiting faculty at **venkat@brillium.in** for further details and opportunities.

Contact Information

For any questions or concerns or further details on this programme, please contact **Program Coordinator** during office hours or via email.

---------- End of the Document----------

C-501, Salarpuria Serenity, 5th Main, Sector 7, HSR Layout, Bengaluru 560102 KA India
Mobile: +91 97012 22130, Email: vmelachervu@gmail.com, vmela23@iitk.ac.in, Website: www.linkedin.com/in/vmelachervu