**Capstone Project-3 Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| 1. Name: - Kartik Subhashrao Dhande. 2. Email ID:- [dhandekartik07@gmail.com](mailto:dhandekartik07@gmail.com)  * Contributed In notebook helped with Google diver data connectivity and data cleaning, data manipulation, and in EDA Visualization * Contributed for the contents of PPT. * Contributed in Technical Documentation in content of problem statement goal of project and steps involved.  1. Name: - Pranav Vilasrao Balpande. 2. Email ID: - [pranav.balpande@gmail.com](mailto:pranav.balpande@gmail.com)  * Contributed in notebook for data cleaning, data manipulation, and in EDA Visualization and finalizing the conclusion. * Contributed in PPT by making sure all the points to be covered. * Contributed in Technical Documentation in content of problem statement goal of project and steps involved      1. Name: - Sanket Rajendra Bhosale. 2. Email ID: - [sanketbhosale0023@gmail.com](mailto:sanketbhosale0023@gmail.com)  * Contributed in notebook for data cleaning, data manipulation, and in EDA Visualization and finalizing the conclusion. * Contributed in Technical Documentation in content of problem statement goal of project and steps involved  1. Name: - Kartik Anilrao Pisudde. 2. Email ID: - [pisuddekartik@gmail.com](mailto:pisuddekartik@gmail.com)  * Contributed In notebook helped with Google diver data connectivity and data cleaning, data manipulation, and in EDA Visualization * Contributed for the contents of PPT. |
| **Please paste the GitHub Repo link.** |
| Kartik Anilrao Pisudde GitHub Link: -  <https://github.com/kartik-pisudde/capstone-project-playstore-review-analyze> |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| * Our client is an Insurance company that has provided Health Insurance to its customers now they need your help in building a model to predict whether the policyholders (customers) from past year will also be interested in Vehicle Insurance provided by the company. * Building a 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 revenue. * Now, in order to predict, whether the customer would be interested in Vehicle insurance, you have information about demographics (gender, age, region code type), Vehicles (Vehicle Age, Damage), Policy (Premium, sourcing channel) etc.   From the above data analysis we clearly saw that there is a huge difference between the  data set.  Standard ML techniques such as Decision Tree and Logistic Regression have a bias towards the majority class, and they tend to ignore the minority class. So solving this issue we use Random Over Sampling technique.  After Random Over Sampling Of Minor Class Total Samples are : 668798 Original dataset shape Counter({0: 334399, 1: 46710})  Resampled dataset shape Counter({1: 334399, 0: 334399})  For modeling, we tried the various classification algorithms like:   * Logistic Regression * RandomForest Classifier * XGBoost |