**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

| **Team Member’s Name, Email, and Contribution:** |
| --- |
| **Team Member’s Role: -**   * **Debashish Das**   **E-Mail-** devashishdas40@gmail.com   * + Data Understanding   + Feature Analysis   + Feature Engineering   + Linear Regression modeling   + Random forest Classifier   + XGB Classifier   + Gaussian Naïve Bayes.   + Gradient Boosting   + Hyper parameter tuning * **Lucky Jain**   **E-Mail-jainluckycool@gmail.com**   * + Data Understanding   + Feature Analysis   + Data Visualization   + Multivariate Analysis   + Handling imbalanced data   + Logistic Regression   + XGB Classifier   + Hyper parameter tuning   + Technical Documentation.      * **Vivek katolkar**   [**E-Mail-**vivekkatolkar7@gmail.com](mailto:E-Mail-vivekkatolkar7@gmail.com)   * + Data Understanding   + Data Visualization   + Exploring the Numerical features   + Multivariate Analysis   + Label Encoding   + Research Analytics   + Conclusion |
| **Please paste the GitHub Repo link.**  GitHub Link:- |
| **Please write a summary of your Capstone project and its components. Describe the problem statement, your approaches, and your conclusions. (200-400 words)** |
| An insurance policy is an arrangement by which a company undertakes to provide a guarantee of compensation for specified loss, damage, illness, or death in return for the payment of a specified premium. A premium is a sum of money that the customer. Just like medical insurance, there is vehicle insurance where every year customer needs to pay a premium of certain amount to insurance provider company so that in case of unfortunate accident by the vehicle, the insurance provider company will provide a compensation (called ‘sum assured’) to the customer. So we tried our best to build a model to predict whether a customer would be interested in Vehicle Insurance or not.  From our attempt to glean insights from the data at hand through exploratory data analysis.  They are.  1.Customers between the ages of 30 and 57 are the most likely to get insurance.   * Customers with driving privileges are more likely to get insurance. * Insurance is likely to be purchased by clients who have vehicle damage. * We learn which characteristics have an impact on the target variable and which do not. * The response column's target variable is quite unbalanced.   2.On the columns for vehicle damage, vehicle age, and gender, use level encoding.  3.We use Extra Trees Classifier Driving License to choose features, and since gender makes a very small contribution, I'm eliminating those columns.  4. Some of the features in our dataset are highly imbalanced, hence to avoid this error, the dataset is balanced using a technique called SMOTE (Synthetic Minority Oversampling Technique).  5.I utilized Logistic Regression, Random Forest Classifier, XGBClassifier, and Random Forest with GridSearchCV (Hyper parameter Tuning) to develop my models.    6. The ideal model is the **XGBClassifier.** |
| Drive link: |