

# Capstone Project Submission

## Instructions:

- i) Please fill in all the required information.
- ii) Avoid grammatical errors.

### **Team Member's Name, Email and Contribution:**

#### **1) Balaji B. Jadhav**

**E-mail:** [bjadhav015@gmail.com](mailto:bjadhav015@gmail.com)

- Data sorting.
- Approach towards plan.
- Graphical representation.
- Bar plot and Heat map.
- Implementation of various Models.
- Model selection and implementation.

#### **2) Anant M. Patil**

**E-mail:** [mr.anantmpatil@gmail.com](mailto:mr.anantmpatil@gmail.com)

- Data visualization.
- Sorting of values.
- Pi-plot and Heat map.
- Various model implement.
- Random search parameters.
- Project summery template.

#### **3) Nishigandha Ingale**

**E-mail:** [Nishiingale28@gmail.com](mailto:Nishiingale28@gmail.com)

- Data analysis.
- Approach towards multiline graph.
- Frame work of project.
- Histogram plot.
- Model presentation.
- Analyzing results of model.

**Problem definition:**

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. The goal of this study is to maximize the profits of insurance companies by devising communication strategies that can optimize business models and profits for customers.

**EDA on given Data set:**

Digging into data we understand that

- There is no null value in the data set.
- Independent Variables are: Id, Gender, Age, Driving License, Previously Insured, Vehicle Age, Vehicle Damage, Annual Premium, Policy Sales Channel, Vintage
- Dependent variable is Response either 0 or 1 for 0 – Not interested, 1- interested.
- Graphical representation according to various columns and with manipulation of columns.

**Model selection and implementation:**

After implementing various models on the given data such as Logistic Regression, Decision Tree Classifier, XG boost, Naïve Bays Classifier, SGD Classifier, Cat Boost Classifier, KNN Classifier, Gradient Boosting Classifier, AdaBoost Classifier, Random Forest Classifier. We get maximum accuracy with Decision Tree Classifier, Random Forest Classifier, XG boost but in case of Decision Tree and Random forest accuracy is decreased for testing data it is the case of Over fitting. In case of XG Boost accuracy decreases but with less percentage here there is no problem of overfitting.

Model	Train	Test	Model	Train	Test
Logistic Regression	0.76	0.59	Cat Boost Classifier	0.5	0.87
Decision Tree Classifier	0.99	0.80	KNN Classifier	0.86	0.73
XG boost	0.88	0.81	Gradient Boosting Classifier	0.82	0.72
Naïve Bays Classifier	0.76	0.58	AdaBoost Classifier	0.81	0.70
SGD Classifier	0.76	0.59	Random Forest Classifier	0.99	0.81

**Conclusion**

There will be more profit if company sells both health and vehicle insurance. Previously insured is important feature for cross sell. After implementation of various models, we got best results from Decision tree classifier, Random forest classifier and XG boost classifier. Decision tree and Random forest are over fitting so finally we decide to go with XG boost. XG boost with train accuracy as 0.88 and test accuracy of 0.81. If we consider a case of discount for package if both the insurance health as well as vehicle then there are more chances of selling of insurance to the customers, if the vehicle is previously insured has more impact on selling of insurance.

**Please paste the GitHub Repo link.**

Github Link:- <https://github.com/BalajiJadhav015/Machine-learning-Classification.git>

**Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)**