**Approach Document**

**Health Insurance Lead Prediction**

**Problem Statement**

Client FinMan is a financial services company that provides various financial services like loan, investment funds, insurance etc. to its customers. FinMan wishes to cross-sell health insurance to the existing customers who may or may not hold insurance policies with the company. The company recommend health insurance to its customers based on their profile once these customers land on the website. Customers might browse the recommended health insurance policy and consequently fill up a form to apply. When these customers fill-up the form, their Response towards the policy is considered positive and they are classified as a lead.

Once these leads are acquired, the sales advisors approach them to convert and thus the company can sell proposed health insurance to these leads in a more efficient manner.

Now the company needs your help in building a model to predict whether the person will be interested in their proposed Health plan/policy.

**Business Goal**

Building a model to predict whether a customer would be interested in Health plan/policy which 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 Health insurance, you have information about demographics (city, age, region etc.), Information regarding holding policies of the customer, Recommended Policy Information.

### **Approach followed:**

1. Explore distributions of numerical and categorical features and their relationships with the target feature, Response (whether the customer responded to an offer about buying Health insurance)
2. Pre-process data in order to model it Accurately
3. Handle the missing values in our data
4. Handle the Outliers in our data
5. Normalize and Standardize the data frame
6. Binning method is also used to create bins for continuous variables where continuous variables are converted to categorical variables.
7. Test out different Classification models by tuning their hyperparameters and comparing their performance
8. Explore which features were the most impactful
9. Explore potential interactions between features

**Final Model:**

**StackingClassifier** model is used in this approach.

So, basically what stacking is Stacking is an ensemble machine learning algorithm that learns how to combine the predictions from multiple well-performing machine learning models and use it to give better predictions.

The get\_stacking() function defines the StackingClassifier model by first defining a list of tuples for the eight base models as estimators, which are LogisticRegression, DecisionTreeClassifier, GaussianNB, RidgeClassifier, SGDClassifier, RandomForestClassifier, AdaBoostClassifier, LGBMClassifier and then defining the logistic regression model as final estimator to combine the predictions from the base models using 5-fold cross-validation.

Then we will fit our training data in our stacking model and finally predict our test data and then generate a csv file having the predictions according to the ID.