**GROUP 7 REPORT ON THE DATASET: PERSONAL BANK LOAN MODELLING**

Thera Bank is on the set for subscription of their loans to see the number of subscribers if the loan policies are favorable. We took our dataset from Kaggle.com which has its headquarters in San Francisco in United States of America. Thera Bank aims to expand its customer base, primarily liability customers, to increase loan business and earn more interest on loans. The bank's retail marketing department aims to target customers with a higher probability of purchasing loans, reducing campaign costs and increasing the success ratio. A successful campaign for liability customers demonstrated a healthy conversion rate of over 9%.

These are some questions to be answered in our report.

Personal Loan: did this customer accept the personal loan offered in the last campaign? Securities Account, Does the customer have a securities account with the bank? CD Account, Does the customer have a certificate of deposit (CD) account with the bank? Online, Does the customer use internet banking facilities? Credit Card, Does the customer use a credit card issued by Universal Bank? We used python: jupyter notebook as our tool and matplot, seaborn, numpy, pandas and cufflinks as some libraries for our work. For the details of our data, we checked for the description, the first five rows and some information derived from the data. From our data, we found that out of 100%, 41.92% of the bank’s loan subscribers are undergraduate students, 30.02% are graduate students and 28.06% are advanced graduates. The average income of the customers was about $74, that resulted in the loan subscription.

**Methodology:**

Data cleaning, exploratory data analysis, model selection, and model evaluation are all part of the project's methodology. A total of 5000 observations with 14 variables divided into four measurement categories made up the dataset. Data cleaning includes eliminating any uninteresting variables, outliers, and missing values. To understand the distribution of variables in the dataset, exploratory data analysis was done. Using performance criteria like recall, accuracy, and F1 score, we created and assessed a variety of classification models, such as logistic regression, decision trees, random forests, and linear regression. Based on how well it performed on the test set, the top-performing model was chosen.

**Conclusion:**

* Multiple linear regression was used to determine our parameters using OLS estimator linear. The regressors of the regressed in the OLS estimation explained about 75% of the model and most of the regressors are statistically significant. Durbin-Watson of about 2 also is a goof indicator of no autocorrelation.
* An indicator of how well a statistical model fits the data is the log-likelihood. It is, specifically, the natural logarithm of the likelihood function, which, given the model's parameters, is the probability of witnessing the supplied data.

(The likelihood function assessed at its maximum is likely to be a high number (since the logarithm of a number between 0 and 1 is negative), a log-likelihood of -15994 in your example indicates that the model fits the data reasonably well. We made use of the residual plot to check for our error terms in the dataset.)

* the selected model (random forest) has the highest accuracy of about 98%

Though all models used perform marvelously, random forest score the highest.

* The campaign initiated by Thera Bank did not yield a highly desirable result since little or no attention had been given to the key drive of the personal loan. Thus, CD Account, CC Avg, and the income of the customers.

**Recommendation**

* The undergraduates constituted a great percentage of the customers hence the next campaign should factor their interest. (Student loan can be an initiative.
* Most of the customers earn below the average of $74 therefore decreasing interest rate but to a profitable level would induce customers purchase of the loan.
* Securing Credit Card Account is seen to play a key role in purchase of Personal loan so, if this is provided at a discounted rate, purchase of the loan would increase.