JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Application of AI in Banking Industry

Dr. Pragati Priyadarshinee

Chaitanya Bharathi Institute of Technology(A), Telangana,India
Email: pragatipriyadarshinee_it@cbit.ac.in

Abstract : Banks or any other organizations find it more profitable to have long term customers rather than short term customers. In banks, long term customers profit them by depositing their money for a longer period of time, and hence increasing the bank's revenue through interests. Hunting for new customers and putting efforts towards them to make them stay, is more expensive than convincing an existing customer to opt for term deposits. With the introduction of Artificial Intelligence algorithms for Data handling, the efforts put towards campaigning can be pin pointed towards customers who show more probability of success in the sales. Such ease towards successful campaigns will also leave the bank with more resources and time for the development of other aspects of their business

***IndexTerms* - Component,formatting,style,styling,insert.**

I. INTRODUCTION

Requirement specification is a highly technical which explains about the software that have been used in the development process. This needs to be specified very accurately. A beginning step of requirement analysis is the requirement specification part. It provides a listing of software systems that are required for the proper functioning of the project. It includes functional aspects, performance and security aspects of the project. The requirements also generate perspectives from the users or a test group or an admin point of view. The goal of software requirement specification is to enlighten groups of individuals about the working of the software system, its arguments and its purposes. It provides an idea to the user about the software and hardware that is required to run the project successfully. It also provides an idea about the extent of the project's functionalities.

A Portuguese Banking institution came to notice a decline in their revenue, and they urged their analytics team to investigate into the problem and to come up with actions that can be taken to tackle this problem. After thorough research, the team learned that the main issue arose because most of their clients were not choosing long term plans to deposit their money. The investment sect of the bank had taken a hit. People were investing their money less frequently and for lesser periods. Banks utilize the deposits from their customers to attain trades with more profitable businesses. Knowing all this, the Portuguese banking institution would like to identify existing customers who have a higher probability of subscribing for a term deposit. Being able to achieve this makes the bank to put more focus and effort towards marketing on such clients.

II. LITERATURE SURVEY

In this study, the authors researched about a Portuguese bank's problem which was to make a prediction if a customer will opt for term deposits. They introduced a comparison between the classical Artificial Intelligence algorithms to find the most suitable algorithm that achieved the highest accuracy and took the least time in the process of training and prediction. The authors of this paper suggested using Data Mining (DM) algorithms to analyses and guess the success of telemarketing for making the customers choose long term deposits. A Portuguese retail bank was chosen, with data collected through 5 years. That period also included the impacts of a financial crisis in Portugal. The authors made an analysis of a data set with more than a hundred features. This data set provided the information related to the bank's customers. A feature selection, which was not entirely automatic was chosen in the modelling phase, executed with the data prior to July 2012 which allowed them to make a selection of around 20 features from the total number of features. The authors also made a comparison of the 4 Data Mining models, viz., Logistic Regression (LR), Decision Trees (DT), Neural Network (NN) and Support Vector Machines (SVM). The two metrics used were: Area Under the Curve (AUC) and Area of the LIFT cumulative curve (ALIFT). Using these two metrics, the four chosen models were put to test on a deducing set, using the most recent data and a rolling window schematic. The Neural Network resulted in the best accuracy (AUC = 0.8 and ALIFT = 0.7), allowing to reach 79 percent. Two knowledge extraction methods, a sensitivity analysis and a Decision Tree, were passed to the Neural Network model which

gave out many important attributes (e.g., Euribor rate, Outcome of the previous call and experience of the agent). Such knowledge extraction confirmed the obtained model as credible and valuable for telemarketing campaign managers [1][2].

A set of data gathered by collecting information from various campaigns run by a bank was explored. It was observed that the aim of the bank was to improve the subscription rates towards a lengthier term deposit. A bunch of ML techniques/algorithms were applied to get a realistic solution to the problem. The problem being the question, 'In what way can institutions advertise their products such that those products are sold in a profitable manner?'. The goal was to acquire an efficient, accurate and best solution to the problem. Social media has increased the growth in ease of communication and sharing of knowledge. It is really simple and easy to gather information. And hence, people are easily approachable. With the information gathered from a past campaign, the attributes of the client, the campaign, and economic conditions were examined. Depending on the data, Artificial Intelligence predictors will come up with the list of clients who are more probable of taking a chance on various subscriptions, thereby answering the question, 'what can banks do to improve the subscription rate?'. A lot of data structures like, arrays, vectors, dictionaries, are utilized according to the need. The beginning steps include, loading the data frame with a dataset for ease of data pre-processing using the distinct and various python libraries and packages. One of the attributes was excluded because there was a risk of leakage. The attribute which was dropped was the duration of the previous campaigning call made to the customer. This attribute was noticed to be too random which makes our data set very sparse. A dense data set is seemed to provide the best results. Such sparse data set would require more than one model to train the data. The further steps were to apply data cleaning methodologies to convert all the values into numeric type, so that it is easy to get a better accuracy from the predictive model. Graphs from each predictive model were generated and the most accurate model was chosen [3][4].

The interesting thing about this paper is the research done about the role of internet and social media in Relationship Marketing (RM). Their research was done particularly in the banking agenda. The purpose of the authors was to get an idea as to why a few banks don't feel interests in the social media trend. Another purpose was to learn about the methodologies used by such banks for improving relations with the customers. The authors focused upon the innovative ideas of banking institutions in quickly developing countries and regions, particularly in Europe. A good approach towards case-studies were used for this research. On the whole 3 case-studies were built, which explained the methods and ideologies of Relationship Marketing of retail banks in Europe. The data set utilized for building the case studies are collected via in-detail examinations of the top management, research and websites of the banks. Major reasons for abstaining from social media were: The form of approach was really famous among the customers; security issues about the internet and social media for banking; and lack of alignment with latest RM strategies. Social media seemed to be mor relatable to younger generation banks, whereas these days social media is very dependable for gathering information about innovative ideas for campaigning activities. The paper identifies significant requirements for the adaption of social media in Banking Industry and provides information on possibilities for alternative Relationship Marketing techniques which merges electronic sources with a more intimidate approach to banking. Case studies provide insights on marketing strategies of banks in the European region. The paper presents challenges banks come across in their Relationship Marketing efforts and future vision of Relationship Marketing in a contemporary online setting [5][6].

III. EXPERIMENTAL RESULTS

The further steps include making a data frame with the outcome of every base-line model and make a graph of the results on a graph using the 'seaborn' library. We are going to make use of the AUC (Area Under the Curve) to calculate and select a finest model. AUC is one of the best metrics of data science performance-wise for picking the best model as it catches the trade-off among the actual positives and fake positives. We are not even required to set a threshold for this to function.

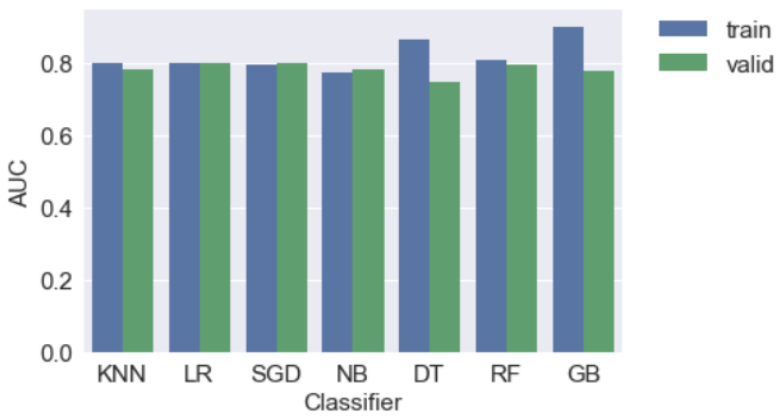

Figure 1: Classifiers.

In this same way, the success rates and accuracies of the rest of the rest of the classifiers are compared and analysed. These comparisons were plotted on a graph to get a clear picture.

Looking at the graph, we can make an observation that the Gradient Boosting Classifier with optimized hyperparameters has a higher AUC value when compared to the Baseline Model. It is obvious that the best model is 'Gradient Boosting Classifier'.

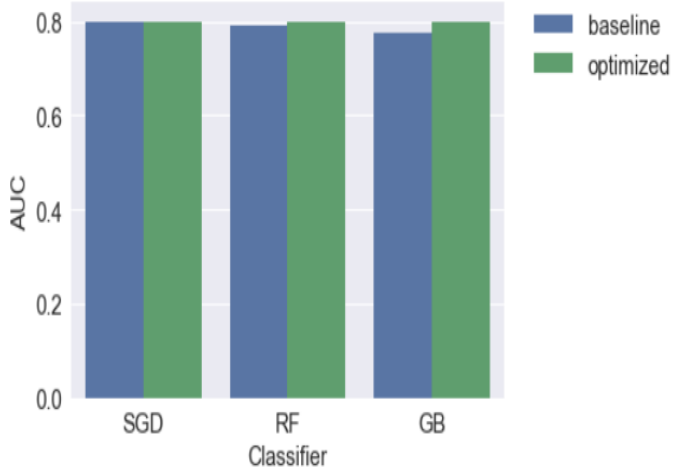

Figure 4: Comparing various classifiers.

We have chosen the 'Gradient Boosting Classifier' as the best model for our project, because it had the best AUC values when implemented on the validation set. In order to make our prediction model reusable, we will save the classifier using the 'pickle' package.

At last, we will test the performance of our best classifier upon the test sample.

In [115]:
```
# load the model, columns, mean values, and scaler
best_model = pickle.load(open('best_classifier.pkl','rb'))
cols_input = pickle.load(open('cols_input.sav','rb'))
df_mean_in = pd.read_csv('df_mean.csv', names =['col','mean_val'])
scaler = pickle.load(open('scaler.sav', 'rb'))

```

In [116]:
```
# load the data
df_train = pd.read_csv('df_train.csv')
df_valid= pd.read_csv('df_valid.csv')
df_test= pd.read_csv('df_test.csv')
```

In [117]:
```
# fill missing
df_train = fill_my_missing(df_train, df_mean_in, cols_input)
df_valid = fill_my_missing(df_valid, df_mean_in, cols_input)
df_test = fill_my_missing(df_test, df_mean_in, cols_input)

# create X and y matrices
X_train = df_train[cols_input].values
X_valid = df_valid[cols_input].values
X_test = df_test[cols_input].values

y_train = df_train['OUTPUT_LABEL'].values
y_valid = df_valid['OUTPUT_LABEL'].values
y_test = df_test['OUTPUT_LABEL'].values

# transform our data matrices
X_train_tf = scaler.transform(X_train)
X_valid_tf = scaler.transform(X_valid)
X_test_tf = scaler.transform(X_test)
```

Figure 5: Evaluating the performance.

As a closing statement, let us preview the concluding evaluation of the performance on test sample.

IV. CONCLUSION

In this article a Artificial Intelligence based technique is used to predict if the customer is willing to opt for term deposits or not. We successfully created a predictive model based on Artificial Intelligence. After analysing all the classifiers, through graphs and plots, we arrived at the conclusion that Gradient Boosting Classifier is the most accurate one. This is bases on the testing done with optimized hyperparameters [7]. The test performance (AUC) is nearly 80%. The ratio of precision and prevalence is 1.6. This ratio indicates that the selected model is 1.6 times better at predicting the outcome compared to guesses made by the marketing agents. The selected model i.e., Gradient Boosting Classifier has correctly predicted that more than 60% of customers are going to subscribe for term deposits. During the development of this project, most of the focus was put on customers who had a significant con_price_idx (consumer price index) and euribor3m (3-month period for clearing the loans) because these features were deemed as important by the feature-importance algorithm. A lot of time, efforts, and resources can be saved by analysing the data set in this manner.

V. References

[1] Guyon, I. and Elisseeff, A., 2003. An introduction to variable and feature selection. *Journal of Artificial Intelligence research*, *3*(Mar), pp.1157-1182.

[2] Martens, D. and Provost, F., 2014. Explaining data-driven document classifications. *MIS quarterly*, *38*(1), pp.73-100

[3] Cortez, P. and Embrechts, M.J., 2013. Using sensitivity analysis and visualization techniques to open black box data mining models. *Information Sciences*, *225*, pp.1-17.

[4] Hossein Javaheri, S., 2008. Response modeling in direct marketing: a data mining based approach for target selection.

[5] Phillips, R., 2013. Optimizing prices for consumer credit. *Journal of Revenue and Pricing Management*, *12*(4), pp.360-377.

[6] Moro, S., Cortez, P. and Rita, P., 2014. A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*, *62*, pp.22-31.

[7] Mitic, M. and Kapoulas, A., 2012. Understanding the role of social media in Banking Industry. *Marketing Intelligence & Planning.*