Machine Learning in Marketing: Banking Domain

Name: Pranali Mohan Kanade

U ID: U70566891

Date: 31 October 2019

Abstract:

Marketing organisation are implementing and expanding AI and Machine learning. 75% of enterprises using them have enhanced satisfaction by more than 10%. Measuring marketing’s contribution to revenue growth is becoming more accurate and real time because of analytics and Machine Learning. Marketers are using machine learning to understand, anticipate and act on the problem.

In banking domain Machine learning came into picture for managing portfolios, conducting high frequency trading, detecting frauds and threads to financial system, personalised customer service in digital world.

Credit checking, banking, loans and mortgages have facilitated Machine learning and Artificial Intelligence provides solution to Banking problems and requirements. Applicants having lacking credit history eventually lack in their ability to pay. Providing customers with chatbots and other interfaces responding to customers need and queries can be answered without intervention needed from human. We will be discussing implementing Machine learning for customer related services.

In marketing industry there a lot of scope for machine learning like identifying and defining sales projection, optimising the marketing, demand forecasting. I will like to explore various areas in Marketing and work on Algorithm part of Machine Learning and find suitable solution to achieve better sales projection and demand forecasting.

Table of Content

Introduction:

The banking industry is continuously innovating and evolving. Industry evolution is happening at rapid pace with help of innovating new technology, customer dynamics and geopolitical movements and evolving demographics.

Software challenges in Data entry, product recommendations, customer segmentation and predictions are requiring a lot of enhancements. ML in current market is not that optimized. Challenges lies in Data, human contributions in analysis etc. Focus of the report will be on how good algorithm which is able to perform number crunching effectively on available data and provide required outcomes.

In order to build and implement Machine learning and Artificial Intelligence to system data is initial and important step which is base for all the work following. Sometimes data provided might not be sufficient and accurate enough for applying algorithms. We have to think and might have to request new data to predict and build better model. After getting high quality data it needs pre-processing and transformation so that it is good for applying algorithm and build the model as per needs. Data can be in many forms like excel, soiled data set, stale data set.

Social media uses image classification for recommendation and customer classification and identification. It will help them in marketing better as more information is available in form of images in applications like Facebook, Instagram and Snapchat. Tagged images and images with text have both images and text so applying separate algorithm for images and text is better solution. Later combining the predictions for getting the final outcome help for better predictions.

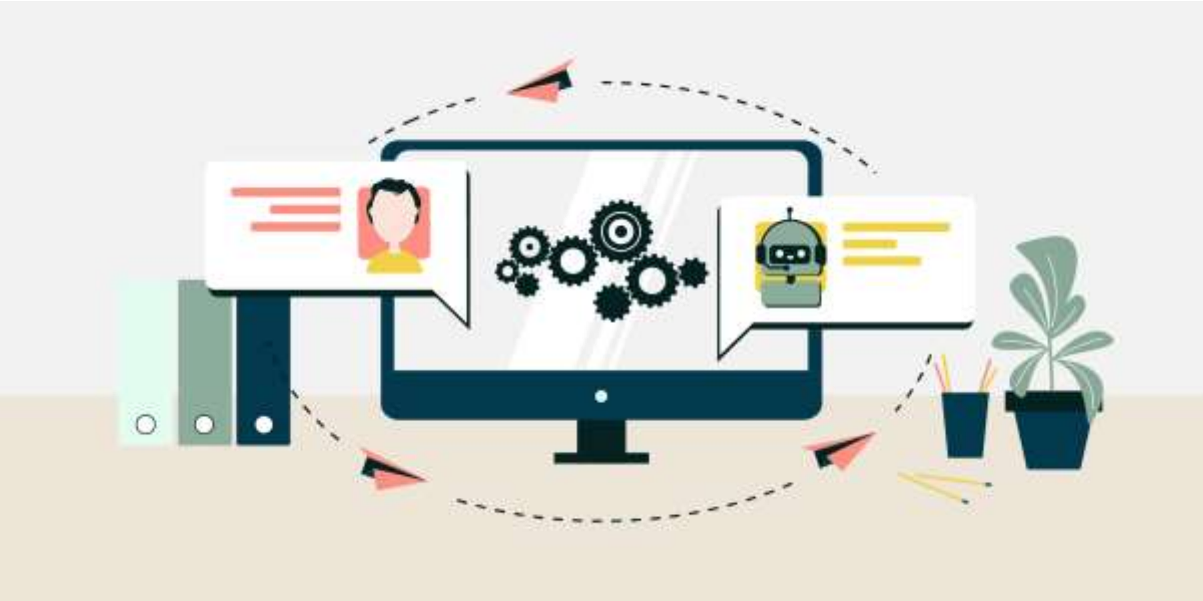
Machine learning in Marketing

Clustering of customer behaviour for behaviour prediction is made possible because of machine learning. Amazon using Collaborative filtering for recommendation for its customer is good example of using machine learning for marketing. Amazon can recommend product based on your purchase and selections by running algorithm on data of the customer. It can also recommend to customers who are not logged in the system with the help of item to item filtering. Here recommendations are made based on item selected and finding customer who bought the same item and then recommending the item that customer bought. Here Machine learning has helped providing new feature to Amazon that has considerably improved its sale.

AI is making important contributions in account based marketing. DemandBase a Targeting and Personalisation Company found that AI could be used to filter out companies from the long list of prospects that would lose the company money. Company's VP stated

“We’re a stats-based company and if they churn from us in less than a year we lose money on them. So, we took the elements that made those customers churn and removed them from our model,”

AI helped DemandBase to identify ‘timely intent’ highlighting the accounts having window of opportunity to approach before commitment to a competitor is made.



Bank Customer Term Deposit Subscription Prediction

Term deposit provides bank certain level of assurance where customer makes long agreement with bank and if they are requesting amount before contract ends then bank deducts charges based on the prior notice rate of interest is not same as the agreement.

Bank makes money from term deposit by using money deposited in investment in other financial products that gives higher rate of interest. Part of this income is given as interest to the customer. For example bank offers term deposit with 6% interest to lender. This is then given as loan with 10% interest rate to its other customer. Here bank is making money with extra interest rate that it is applying on loan. Here bank having lot of term deposit help in having good cash inflow which they can use for investment which in turn will be profit to the bank. Marketing to promote term deposit so that bank will be profited are not limited to advertisement and promotions alone. We can apply machine learning for the marketing in banking.

In this report focus will be on how machine learning helps in increasing sales in banking domain by predicting weather the customer contacted will subscribe for the term deposit. Goal here is to demonstrate how applying machine learning algorithm will help increasing the sales by knowing the target customer who are more likely to subscribe to term deposit. In this report we will be discussing various algorithm and its impact on prediction with a sample experiment conducted.

Demonstrating the Machine learning application on banking dataset imported from Kaggel is used. Here we are trying implementation of the Machine Learning algorithm on the cleaned dataset to predict weather customer contacted will be subscribing for term deposit. The process will be divided into problem understanding, hypothesis generation, Data gathering, Data exploration, Model training, model evaluation, model testing. In general software lifecycle for implementing includes above mentioned steps

Problem Understanding

Dataset used here has details of client's phone calls and whether it will subscribe for the tem deposit or not. People deposit their money for fixed duration. It has maturity ranging from a month to few years. When account holder deposit money to account that money is used by bank to lend it to other customers. Bank charges interest rate on the amount that they have lent and provides customer who has deposited the amount with bank. Depositor can withdraw their money at any time making it difficult for bank to plan on lending money.

Predicting and classifying the target customer with the help of Machine learning algorithm can help bank to focus better on their marketing strategy and treating customer who are more likely to opt for term insurance after promotions. Here for example we consider telephone or cellular call made to its customer whose data is already available with bank consisting of their details like job, marital status, housing, loan, duration, etc.

Based on this model will try to predict weather customer will opt for term deposit or not. Knowing weather customer is more likely to opt for term deposit before the call bank can accordingly approach them differently. For example if bank knows customer is likely to opt for term deposit then focus will be more on providing details of plan and how it is better in market than other bank and how they can invest accordingly. Secondly if bank knows that customer is less likely to opt for term deposit then focus will be on convincing customer the benefits and trying to convince them so that they can consider investing into term deposits.

If the model is able to predict what is the likely reason behind customer from choosing or avoiding investment in term deposit then it is simpler for bank to classify them and apply its marketing strategy. Investing more time and money on more likely customer will add obvious revenue to bank. However by focusing on less likely customers bank will be actually be able to convert them to opt for term deposit. This can be facilitated by providing some offers and discount to attract and increase the profit. It is also important to know how much money made with such investment. If bank is paying more money in marketing and model building than simple telephonic call then and profit from enhanced model is same or with small improvement then machine learning algorithm is actually expensive. As it has involvement of time and effort.

Building Hypothesis

Machine learning expertise should have good knowledge about factors and characteristics that are more likely to have influence on customer's decision about opting for term deposit or not. For example credit score of customer is prime factor when deciding whether customer will be eligible for loan sanction or not. For predicting likely customer opting for term deposit factors like their job, loan status can be primary factors. If the customer is already having loan with bank or has job with relatively less income then probability that they invest in term deposits is very less. However if the customer does not have any loan with bank has stable high income job and has previously deposited in bank chances are high that they will like to deposit with bank for term deposit.

In first scenario if the customer is already having loan with bank then there are absolutely no chances that they will fix deposit in bank. They will first want to clear off the loan then plan to do investment. Here it is easy for bank to exclude them or invest minimum possible for marketing knowing that it has less chance for investment and bank can save money by not investing for such customers money saved because of this is actually bank is making profit by removing such customers from marketing and saving money on marketing. Here having knowledge of both customer behaviour and technology is the key.

Data Gathering, Data Exploration and Data Transformation

Data available with bank is as is provided for applying algorithm. Data gathering, Data exploration, and Data transformation are subtopics associated with Data. The raw data cannot be directly used for applying Machine learning algorithm as it is possible that data is not clean and sometimes not good enough for building efficient model. After analysing the data provided next step is make changes for cleaning of data and apply transformation to it if needed. So that data is now consistent for better prediction model.

Data exploration helps in getting insight from data. Data provided is huge and has a lot of inconsistency that needs to be detected and handled for better result. There are a lot of exploration strategies like Univariate analysis, Bivariate analysis, Multivariate analysis, Cross table which help in solving complicated problem in less time. Analysing every aspect of customers column using data exploration technique help understand the statistical aspects of data available and result from analysis help determining better deciding the algorithm for better model prediction.

Feature engineering

Feature engineering is part where experimentation and domain knowledge plays critical role. In Machine learning single solution cannot be best solution for all or most of the problem. Here innovation and creative ideas help develop the ideas for new features during the data exploration and hypothesis generation stage. The goal of feature engineering is to achieve better prediction models by creating new features. For example mostly categorical variables have near zero variance distribution. Near zero variance distribution is the category in variable having >90% of the values. Binary variables are created for depicting presence or absence of category. This new feature will now contain 0 and 1 values which will help better, efficient and less expensive algorithm.

Model training and evaluation

By this stage we will have data ready for training model. Here we can experiment with different algorithms like Adaboost, Neural network, decision tree, regression, ensemble etc. Then the data is trained with available data with different algorithms and various parameters like accuracy, confusion matric, score and mean absolute error etc. are checked to determine best suitable models. While deciding model simplicity is the key as many times simpler models give better accuracy than complex model. If your data is relatively simple then simpler models like Decision trees can help you get better result. Complex algorithms like Neural Network has high accuracy model. But they are expensive models and might perform poor for easy data like bankers accounts details from database. However this model will perform better for complex data like images.

In banking domain data is in the form of categories and with simpler algorithm it is possible to achieve better accuracy and lesser absolute error. For example we will be getting data with different categories like age, job, education, loan etc. So model like decision tree, regression can work effectively using algorithms like neural network can add complexity to model and might result in issues like over fitting where model perform poor on new data. As model learns the quite too well to provide high accuracy and later when it is tested on new data it performs poor as model has also learned the noise from the data set and makes poor prediction on test dataset. It is good idea to have different training and test data for most of the models for verification of performance on new data and ensuring there is not over fitted model.

Algorithms

In process of building hypothesis we are assuming some criteria for customer classification. Study of customer behaviour on the bank details of customer and determining critical factor will help in building effective model. In Machine learning experimentation and innovation are crucial. This domain is no different as well. It is possible to have some criteria discovered with help of algorithm that prove to be more relevant when making prediction. For example we are considering job as primary criteria for classification but it was found with experimentation that education and number of children are more effective criteria for classification of customer.

Machine Learning Algorithm is crucial weapon to understand the problem and providing solution which will help bank to take actions before possibly problem could occur. This way Bank and Customers relation are improved as issues and concerns are possible to detected and corrected before they could actually occur.

Machine learning algorithms are categories in three types viz. Supervised, Unsupervised and Reinforcement. In Supervised learning, algorithm model is trained with the dataset and continuously corrected to achieve the acceptable performance level. It predicts the result based on training data using set of variable. Function is built to map input to yield desired output. Decision tree is example of supervised learning algorithm.

The unsupervised learning algorithm is not trained however algorithm is left to discover the interesting structure in the data. With the help of the unsupervised algorithm, patterns can be detected based on characteristics of data. Clustering is an example of the unsupervised learning model. The machine learns the data provided and then it groups similar data type to form different clusters within the data.

1. Decision Tree

Decision tree is easy yet powerful classic algorithm. Decision tree is basically a set of binary rules repeatedly splitting data sets into categories and giving final predictions at each leaf node. The intuitive idea of decision tree learning is building disjunction of conjunction of constraint so that each conjunction will give one prediction. Decision are simple for humans to understand them can be easily explained to them. Decision tree based on a training data set does recursive pick of feature and split the data set by the attribute of this specific feature. As soon as a node is created, the data set is split into two parts and we can further create nodes with the same method. Information gain measures how much entropy can be decreased for given answer of specific feature.

Decision tree algorithm has scope for overfitting. Overfitting is phenomenon when a model performs better and better in training data set it tend to learn the noise of training dataset which result in poor performance on testing data set.

In order to avoid overfitting there are a lot of methodologies which help in dealing with overfitting problem. One such method is error reduced pruning. Here idea is to split a validation set outside of the training set. Deepest decision tree is build first then in pruning tree is folded node by node and validation error reduction is checked for pruned. Smaller decision tree can give comparable results as compared to deep and complex decision tree.

Simple decision tree models are less expensive and if the difference in accuracy is not huge then opting for simple decision is good idea as simpler model has less risk of overfitting. Another way to avoid overfitting is by having large training data set. Small training dataset has high risk of overfitting as ML algorithm will be learning tries to learn all the details and end up learning noise and discontinuities in data which can be solved with the help of large training data set.

2. Regression

Under regression we will be discussing linear regression and logistic regression. Regression models relation between features and labels as weighted summation. The model believes every feature makes some contribution in final classification. Basic linear regression model is represented as below:

$Y = \beta 0 + \beta 1X + e$,

Where X is representing feature, Y is representing corresponding label, β0 and β1 are two constants the intercept and slope respectively, e is the error term representing the difference between real label and estimated label. Using linear regression we have predict value of β0 and β1.

Logistic regression formula is given as

$$p(x) = \frac{e^{\beta_0 + x\beta}}{1 + e^{\beta_0 + x\beta}} = \frac{1}{1 + e^{-(\beta_0 + x\beta)}}.$$

Where p(x) represents probability of the features give a positive label. Similar trick can be applied to gain the optimal solution of parameters β0 and β1. The reason has been explained in previous text. In summary, the general idea of logistic regression is to use linear function to approximate the logit odds of the probability that such feature gives positive label. The training algorithm will assign appropriate weights to each feature. The larger the absolute value of a feature, the more it contributes to the labelling decision. Positive weight means this feature indicates positive label and vice versa.

3. Boosting

In Boosting a weak learner which is slightly better than random guess is boosted into an arbitrary accurate strong learner. AdaBoost which stands for Adaptive Boosting trains bunch of weak learner and the final strong learner is a weighted combination of them with appropriate weights. Higher weights are assigned to learners with lower rates. Samples in training dataset have weights. Higher weights are assigned to previously poor prediction sample so that later learners are able to concentrate on difficult samples.

Many theoretical and experimental studies have proven the power of AdaBoost as a "boosting" algorithm turning weak learners into a strong learner. For boosting the learners chosen should be diverse and perform moderately well. Combination of diverse algorithm provide better prediction and multiple model which are performing good on same functionality are less likely to perform better in Adaboost.

Risks

Applying machine learning model for marketing in banking domain has increasing influence. Application of machine learning in banking has risks associated with credit risk, market risk, operational risk and liquidity risks. The prominent global financial crisis, risk management has focus on how risk are detected measured reported and managed.

Credit risk

Credit is risk of potential loss to bank when borrower fails to meet obligation such as interest and principal amounts. Credit risk is prominent risk associated that bank can face. Key parameter is to estimate default, loss given default, and exposure at default.

Liquidity risk

Liquidity risk has two forms viz. asset liquidity risk and funding liquidity risk. A bank faces asset liquidity risk when transaction cannot be executed at prevailing market prices which has consequence of the size of position relative to normal trading lot size. Funding liquidity risk refers to inability to meet cash flow obligation and known as cash flow obligation. Banks need to establish a robust liquidity risk withstanding range of stress event. A solid process for identification, measurement, monitoring and control of liquidity risk implementation is critical.

Operational risk

Operational risk is defined by BCBS as the risk of loss resulting from inadequate or failed management at the bank. It includes legal risk and excludes strategic and reputational risk. Operation risk includes fraud risk, cyber security, client's product and business practices, information and resilience risk, money laundering and financial crime risks, vendor and outsourcing risks, technology risk, business disruption risks. Bank report compliance and legal risk under operational risk.

Application

Banking is more traditional domain where banks have standard processes and stringent process. Leaders are often reluctant when applying upgradation in the system or changing the technology process as it has high risks associated which are separately discussed in the report. With the evolving technology and technological advancement latest technologies like Machine learning and Artificial intelligence has allowed bank to rely less on human experts and employees can focus better other advancements and customer satisfaction. Some of the prominent applications that bank has with Machine learning as below.

Credit decisions

Credit scoring can be more sophisticated than traditional credit scoring processes with help of Artificial Intelligence. It allows fast, accurate assessment of potential borrower with less cost than traditional methods. Additionally technology eliminates bias as machines are more objective than humans. Bank can determine applicants default risk and its credit worthiness even without their credit history.

Bank has a lot of data of its customer including vast amount of historical data. Machine learning model can be trained to perform credit scoring over and over again to learn from mistakes and improving on its own in continuation. The result is faster, more accurate credit scoring system that bank can trust. Customers are able to receive faster responses from institutions and have better understanding of their finances.

Risk Assessment and Management

Automating credit risk testing allows bank mitigate risk as they receive accurate reporting which is not prone to human errors. AI does more than just reducing risk to bank and customer. It helps bank forecasting issues and deciding steps to avoid problems.

Machine Learning Algorithms are much faster evaluation risk and it has ability to process huge amount of data that human has limitation over. It helps individual portfolio holders assessing risk that help in making better financial decisions.

Fraud Prediction

Fraud has prominent impact and serious implication on financial domain and one where AI and ML have substantial impact. Analysis of spending pattern, location and client's behaviour allows machine learning detecting anomalies in spending and alert cardholder help in reduction of credit card fraud. It is almost impossible for human to analyse thousands of transaction simultaneously in real time. System can flag suspicious behaviour and can request for additional information from users or block transaction completely in fraction of seconds. It allows bank to catch fraud in real time than waiting for it happen followed by rectification and resolution.

High Frequency trading

Trading has potential to accelerate at faster pace in association with AI and Machine learning algorithms. AI and ML algorithms can be used for trading decisions and adaption of real time changes prove to be instrumental for conducting efficient and successful business. Financial institutions have capability to create AI and Machine learning technology account for most current nature of market before conducting trades.

Personalized customer service

With technological advancement in AI and Machine learning overseeing financial matter relates to customers expectation in interactive Chatbot and algorithmic interfaces responding to customers. Algorithmic interfaces respond to customer's needs, enquiries and complain with help of Chatbot, automated customer care. There is technological progress in language recognition for AI and Machine learning processes which help in answering unique questions by customers more effectively. This saves a lot of time and resources required for in person consultation to customer.

Personalised Customer service

Machine learning algorithm has ability to analyse individual's data and monitor anomalies that might occur. For example notifying customer if the customer was charged twice for expense or if they tipped excess amount at restaurant. Such daily hassles of customer can be tracked and monitored to provide assistance to customer in faulty transactions and their expenses. With help of Machine Learning it is possible to provide recommendation so that customer can take better financial decision. Big banks has personalised functionalities like reminders to pay bills, financial planning, and other perks which helps its customer to finance better easy understanding of expenses and tracking. Such personalised services help providing experience and recommendation to keep customers happy and loyal.

Process Automation

Automating daily mundane task is easily possible with automation technology like robotic process automation. It has ability to building model which is faster and error free as compared to human intervention. Banks like JP Morgan Chase has automated processing of legal documents, extraction of data and review of certain type of legal contracts Machine learning algorithm uses image recognition for identifying patterns in agreements. It reduces a lot of which was needed with manual process.

Software Challenges

AI powered marketing platform is embracing new opportunities in banking. AI marketing has become simpler for use but Knowing challenges

Insufficient IT Infrastructure

AI driven marketing strategy needs a robust IT infrastructure. AI technology processes need to deal with vast quantities of data for that it needs high performing hardware. Computer system needed for same are very expensive additional it requires frequent updates and maintenance ensuring smooth working of application.

Lack of Data or Poor data Quality

AI has to process high quality data for efficient functioning of application. Insufficient data or poor quality data lead to poor results. Companies are having Big Data which is allowing them for collecting increasing amount of data however data is not right data needed for successful AI marketing strategy.

Lack of Trust in Software

Machine Learning and AI are relatively new technology in market and also complex. There is lack of technically strong employees who are trained in AI. The general public are suspicious about because same.

Insufficient Budget/Investment for Implementation

Companies are investing handsome amount for AI solution however it is insufficient spending for efficient system. AI solution provides good return of investment but it needs investment strategically and significantly done. Technology needs complex software and high performance hardware which is expensive in cost and maintenance. It is difficult in small scale companies where budget is very tight.

Lack of In-House Talent

Currently there is AI skill gap impacting businesses willingness to develop in-house AI marketing solution. Employees do not have exposure to up to the mark training. Even after getting readymade AI marketing there is lack of sufficiently skilled employees who can deploy and manage same.

Summary

In reinforcement, a learning machine is trained to make the decision based on hit and trial. The machine is continuously trained with experience and tries to capture the best possible knowledge which can be used for accurate business decision.

To summarise it can be very well said that AI is expanding its domain in every dimension. Though it is in very beginning phase it is already resolving many complex issues. But along with that there are some serious concerns and risk that comes with it. More important than resolving concerns is to address them so that there resolution can be worked on in as effective and efficient way

Referances

1. https://www.forbes.com/sites/forbesfinancecouncil/2018/10/05/what-role-can-machine-learning-and-ai-play-in-banking-and-lending/#7f484e044122
2. https://www.techradar.com/news/how-ai-and-machine-learning-our-improving-the-banking-experience
3. https://www.techradar.com/news/how-ai-and-machine-learning-our-improving-the-banking-experience
4. https://www.concured.com/blog/biggest-challenges-to-ai-marketing-success
5. https://aliz.ai/machine-learning-in-digital-marketing-apply-ml-for-marketing-like-a-pro/
6. https://pdfs.semanticscholar.org/a3f6/878eb00901500600a8760046c527e077b33d.pdf
7. https://www.publicissapient.com/insights/anticipatory-banking?gclid=EAIaIQobChMIiPOnt6bH5QIVEdvACh2XCQOKEAAYASAAEgIEFvD_BwE
8. https://www.canstar.com.au/opinions/how-banks-work/
9. https://www.kaggle.com/henriqueyamahata/bank-marketing
10. https://pdfs.semanticscholar.org/8080/b472daccb9d4fd54165e6c05d762ccf37380.pdf