BANK DIRECT MARKETING

USING

CRISP DM METHODOLGY

FINAL PROJECT REPORT

Submitted by:- To:-

Gurpreet Singh Dr. Changhe Yuan

Jianglin Wu

CS 3813(Machine Learning)

Contents

* Keywords
* Abstract
* Background
  + CRISP-DM Methodology
  + Business Understanding
  + Bank Direct Marketing
* Methods: Data Mining on Marketing Campaigns
  + Decision Tree (Gurpreet Singh)
  + Naïve Bayes Classifier(Jianglin Wu)
* Data Understanding
  + Bank Direct Marketing Data
* Modeling
  + First Model: Decision Tree
    - Experiments and Results
  + Second Model: Naïve Bayes
    - Experiments and Results
* Conclusions
* References

KEYWORDS

Classification, Decision Tree, Naïve Bayes, Directed Marketing, Data Mining, CRISP-DM,

Contact Management.

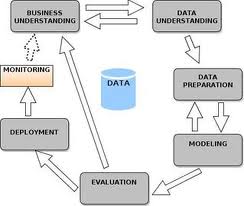
ABSTRACT

The impact of Mass Marketing Campaigns on general public is decreasing due to educated and more informed audience. Furthermore, financial crisis and economic pressures has “led marketing managers to invest on directed campaigns with a strict and rigorous selection of contacts” (Moro, Sergio). The efficiency and performance of such direct marketing campaigns can be enhanced through the use of Business Intelligence (BI) and Data Mining (DM) Techniques. This paper describes the implementation of two data mining techniques: Naïve Bayes and Decision Tree and how they are implemented keeping CRISP-DM methodology in mind. The training and testing data is actually Real World Data which is collected from a Portuguese marketing campaign related with bank deposit subscription. The goal of implementing bank direct marketing project is “to find a model that can explain the success of a contact, i.e. if the client subscribes the deposit” (Moro, Sergio). Such a model will lead to better allocation and use of available resources, saves money and increase campaign efficiency.

BACKGROUND

METHODOLGY USED

The methodology used to implement this bank direct marketing project is CRISP-DM methodology which stands for Cross-Industry Standard Process for Data Mining. This is a popular methodology to increase the success of Data Mining projects. This is a non-rigid sequence of Six Phases:



**Six Phases of CRISP-DM Methodology**

1. Business Understanding
2. Data Understanding
3. Data Preparation
4. Modeling
5. Evaluation
6. Deployment (**NOT USED**)

FIRST PHASE: BUSINESS UNDERSTANDING

The first step in CRISP-DM methodology is Business Understanding. As mentioned in Keywords in beginning, the business understanding starts with the word “Directed Marketing.” As we know, marketing campaigns are used by companies to increase customer base, enhance brand recognition and to boost business profits. There are two main approaches: Mass Marketing and Direct Marketing. The Mass Marketing targets general larger audience and companies spends millions of dollars to promote their products via mass marketing campaigns. The problem is despite of spending millions on Mass Marketing, the positive response rate is very low, typically less than 1%. On the other hand, direct marketing campaigns are very efficient and effective in targeting specific set of contacts, those who are assumed to be more interested in certain products and services. In bank marketing, significance of direct marketing campaigns increases many times as it is about winning client’s confidence in specific services. So, the direct marketing is the key area of focus in this project.

BANK DIRECT MARKETING – THE ACTUAL PROJECT

Portuguese Banking Institution

Offers long-term Deposit Applications with good Interest Rates to their Clients

Use Direct Marketing Campaign (Phone Calls) to target Clients

**Business goal:** To Increase Efficiency of Direct Marketing Campaign by calling only those Clients who are more likely to subscribe a Term Deposit

DATA MINING TECHNIQUES & CONTRIBUTIONS

* Decision Tree: Fully Implemented by Gurpreet Singh
* Naïve Bayes Classifier: Fully Implemented by Jianglin Wu

DATA UNDERSTANDING

To build and train a Decision tree and Naïve Bayes model, real world data was collected from Portuguese Marketing Campaign related with bank deposit subscription.

General Information about data:

There are 45211 data points and each data point has 17 attributes

* + - Each data point represents a Client
    - 16 + Output attribute
    - 10 Categorical + 7 Numeric

The actual success rate is just 11.5 % i.e. only 5269 Clients has subscribed term deposit out of 45211 Clients. So, it is important to find a Data Mining Model that could predict whether a Client will subscribe a term deposit or not. Bank representative will call only those clients for whom the model will predict “yes”. This will increase the efficiency of Direct Marketing campaign and allows us to target more clients with the same resources.

Each Data Point has 17 attributes. These 17 attributes can be further divided into 4 subsets.

* + Bank Client Data
  + Last Contact of the Current Campaign
  + Output Variable
  + Other

BANK CLIENT DATA (FIRST 8 ATTRIBUTES)

1 - age (numeric)   
2 - job: type of job (categorical: 'admin.', 'unknown', 'unemployed‘,

…….. , ‘ technician', 'services').

3 - marital: marital status (categorical: 'married', 'divorced', 'single‘)   
4 - education (categorical: 'unknown', 'secondary', 'primary', 'tertiary')   
5 - default: has credit in default? (categorical: 'yes', 'no')   
6 - balance: average yearly balance, in euros (numeric)   
7 - housing: has housing loan? (categorical: 'yes', 'no')   
8 - loan: has personal loan? (categorical: 'yes', 'no')

LAST CONTACT OF THE CURRENT CAMPAIGN (ATTRIBUTES: 9 to 12)

9 - contact: contact communication type (categorical: 'unknown‘, 'telephone' , 'cellular')   
10 - day: last contact day of the month (numeric)   
11 - month: last contact month of year (categorical: ‘jan', 'feb', ..., 'nov', 'dec')   
12 - duration: last contact duration, in seconds (numeric)

OTHER (ATTRIBUTES: 13 to 16)

13 - campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)   
14 - pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric, -1 means client was not previously contacted)   
15 - previous: number of contacts performed before this campaign and for this client (numeric)   
16 - poutcome: outcome of the previous marketing campaign (categorical: ‘ unknown', 'other', 'failure', 'success')

OUTPUT VARIABLE

**17 - y - has the client subscribed a term deposit? (categorical: 'yes', 'no')**

**Sample Data Point**

33, services, married, secondary, no, 4789, yes, yes, -cellular, 11, may, 220, 1, 339, 4, failure, **no (Output)**

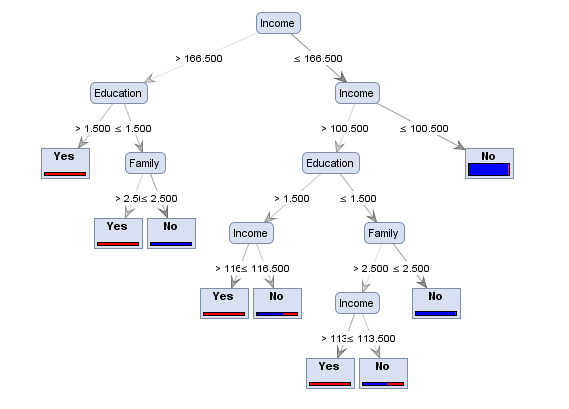
|  |  |  |
| --- | --- | --- |
| S.no. | Attributes | Sample Data Point Values |
| 1 | Age | 33 |
| 2 | Job | Services |
| 3 | Marital | Married |
| 4 | Education | Secondary |
| 5 | Default | No |
| 6 | Balance | 4789 |
| 7 | Housing | Yes |
| 8 | Loan | Yes |
| 9 | Contact | Cellular |
| 10 | Day | 11 |
| 11 | Month | May |
| 12 | Duration | 220 |
| 13 | Campaign | 1 |
| 14 | pdays | 339 |
| 15 | previous | 4 |
| 16 | poutcome | Failure |
| 17 | Y | No (Output) |

MODELING

The 4th phase of CRISP-DM methodology is Modeling in which we build and train a model.

**FIRST MODEL**

**DECISION TREE BY GURPREET SINGH**



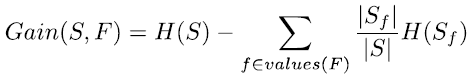
**Fig. above: Sample Decision Tree**

A decision tree is a decision support tool that uses a tree-like [graph](http://en.wikipedia.org/wiki/Diagram) or [model](http://en.wikipedia.org/wiki/Causal_model) of decisions and their possible consequences, including [chance](http://en.wikipedia.org/wiki/Probability) event outcomes (Wikipedia). In Machine Learning, Decision Tree is used as a classification model which predicts output y out of many choices. Decision tree can be build by learning from data. Decision Tree Model can be evaluated by percentage accuracy, Sensitivity, ROC curve etc. The basic building blocks of Decision tree are nodes which can be further classified as Root Node, Internal Node and Leaves. Each Internal Node Corresponds to a feature whereas leaves associated with target values.

* Nodes with continuous features
  + have Two Children for value ˂ and ≥ a break point
* Nodes with Nominal features have N children, where N is the number of nominal values

**Building a Decision Tree** requires data and usually a recursive algorithm.

* What feature to use?
  + Use Information Gain Formula



* + Use the feature with best Information Gain

RESULTS & EVALUATION

The fifth phase of CRISP-DM is Results and Evaluation.

1. **Testing Set Accuracy Rates** at various Training percentages

Below is a table showing the Training Set Accuracy rates:

These accuracy rates are recorded at various training percentages

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Testing Set Accuracy Results | | | | |
| S.no. | Training Set Percentages | | | |
| 20% | 40% | 60% | 80% |
| 1 | 80.55 | 81.05 | 81.46 | 82.16 |
| 2 | 80.13 | 80.30 | 82.26 | 82.75 |
| 3 | 79.80 | 80.89 | 81.34 | 81.65 |
| 4 | 79.54 | 81.40 | 81.59 | 82.01 |
| 5 | 79.61 | 80.34 | 82.29 | 81.99 |
| 6 | 80.19 | 80.56 | 81.13 | 81.77 |
| 7 | 80.41 | 81.10 | 81.16 | 82.23 |
| 8 | 79.72 | 81.05 | 82.10 | 82.77 |
| 9 | 80.15 | 80.11 | 81.39 | 81.95 |
| 10 | 79.95 | 81.19 | 82.02 | 82.69 |
| Average  (approx.) | 80.10% | 81.05% | 81.95% | 82.25 % |

**Table 1: Testing Set Accuracy Rates at Different Training Percentages**

**Graphical Analysis of Testing set accuracy rates from Table 1**

**Graph🡪Table1: Training Set Percentage vs. Testing Set Accuracy Rate**

**Analysis:** The Testing Set accuracy rate increases with percentage of Training data set. Since the total number of data points are very high (45,000 data points). Even the smaller training percentage like 20% (have 9,000 data points—a good number) led to good testing set accuracy rate 80%. This is the reason why even on increasing training set percentage from 20% to 80%, the increase in testing set accuracy is low from 80% to 82.5%. In this case, as accuracy rate increase is low, it would be good to build model on smaller training set (20%), as it will take less time to run and use less training data (9,000 data points) compare to 36,000 data points(80%).

1. **Training Set Accuracy Rates** at various Training set percentages

Below is a table showing the Training Set Accuracy rates:

These accuracy rates are recorded at various training percentages

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Training Set Accuracy Results | | | | |
| S.no. | Training Set Percentages | | | |
| 20% | 40% | 60% | 80% |
| 1 | 94.20 | 95.89 | 96.68 | 98.79 |
| 2 | 94.19 | 95.81 | 96.99 | 98.62 |
| 3 | 94.13 | 95.62 | 96.98 | 98.70 |
| 4 | 94.07 | 94.95 | 97.01 | 98.99 |
| 5 | 93.90 | 95.50 | 96.25 | 98.25 |
| 6 | 94.50 | 95.25 | 96.50 | 98.37 |
| 7 | 94.25 | 95.69 | 96.61 | 98.14 |
| 8 | 94.05 | 94.99 | 96.33 | 98.85 |
| 9 | 93.95 | 96.10 | 96.11 | 98.93 |
| 10 | 94.01 | 95.27 | 97.12 | 97.99 |
| Average  (approx.) | 94.25% | 95.50% | 96.93% | 98.65 % |

**Table 2: Training Set Accuracy Rates at Different Training Percentages**

**Graphical Analysis of Training set accuracy rates in Table 2**

**Graph🡪Table2: Training Set Percentage vs. Training Set Accuracy Rate**

**Analysis:** The Training Set accuracy rate increases with the increase in percentage of Training set. The model gives very high Training set accuracy rates even at low Training Set Percentages. The model gives an average Training set accuracy of 93.5% on 20% training set percentage. There is an over 5% increase in training set accuracy rate on increasing training set percentage to 80%. As one can expect that building tree on 80% training set is much slower than building tree on 20% training set, so for large data sets it is good to train model on small training set especially when it gives good accuracy rate of more than 90% even on small training percentages.

1. **Full Data Set Percentage Accuracies** at various Training set percentages

“Full Data Set” refers to complete data set. One can say full data set is the actual data set or union of testing and training data set. We can also say “Full Data set” is the real data set before partitioning into testing and training sets. Full data set accuracy rate can be obtained from Testing and Training set accuracy rates, by adding their accuracy rates according to their percentages. So, here we trained model at some % of Training set, but we test every single data point using the trained model.

Below is a table showing the Full Data Set Accuracy rates:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Full Data Set Accuracy Rates | | | | |
| S.no. | Training set Percentages | | | |
| 20% | 40% | 60% | 80% |
| 1 | 85.48 | 87.55 | 89.95 | 91.34 |
| 2 | 85.92 | 88.15 | 89.61 | 91.38 |
| 3 | 85.06 | 87.59 | 89.73 | 91.26 |
| 4 | 84.98 | 87.19 | 90.05 | 91.16 |
| 5 | 85.11 | 87.27 | 90.12 | 90.88 |
| 6 | 85.77 | 88.20 | 89.77 | 91.56 |
| 7 | 84.76 | 87.32 | 89.79 | 91.29 |
| 8 | 85.38 | 87.19 | 90.17 | 90.93 |
| 9 | 85.31 | 88.33 | 89.83 | 90.84 |
| 10 | 84.89 | 87.63 | 90.05 | 91.11 |
| Average  (approx.) | 85.25% | 88.37 % | 90.01 % | 91.17 % |

**Table 3: Full Data Set Accuracy Rates at Different Training Percentages**

**Graphical Analysis of Full Data set accuracy rates**

**Graph🡪Table3: Training Set Percentage vs. Full Data Set Accuracy Rates**

**Analysis:** As seen before, Testing and Training set accuracy rates are increasing at various training percentages. One can intuitively say that Full Data set accuracy rate may also be increasing and that he/she would be correct in saying that. The combined accuracy rate increases slightly over 5% on increase in Training set percentage from 20% to 80%. On moving right along the x-axis from 20% to 80%, the rate of growth of accuracy rate is initially

high but decrease with the increase in training set percentage. This is due to add of complexity with larger decision tree models. The net result is that the tree generalizes very well.

**Short summary** of Percentage Accuracy Results and Analysis:

When Training set is **80%**

* Avg. Training Set Accuracy Rate: 98.65 %
* Avg. Testing Set Accuracy Rate: 82.25 %

When Training set is **60%**

* Avg. Training Set Accuracy Rate: 96.93 %
* Avg. Testing Set Accuracy Rate: 81.95 %

When Training set is **40%**

* Avg. Training Set Accuracy Rate: 95.50 %
* Avg. Testing Set Accuracy Rate: 81.05 %

When Training set is **20%**

* Avg. Training Set Accuracy Rate: 94.25 %
* Avg. Testing Set Accuracy Rate: 80.10 %

EVALUATION METHOD 2: SENSITIVITY AND PRECISION

* Accuracy: Correct prediction / Size of testing set
  + We discussed percentage accuracy above.
* Sensitivity: Percentage of actual positive outputs that have been detected. TP/(TP+FN)
* Precision: Percentage of calculated positive outputs that are actual positive. TP/(TP+FP)

|  |  |  |  |
| --- | --- | --- | --- |
| Percentage of Training Data | Average on Full Data set  (Main Data Set: 45211 data points) | | |
|  | Accuracy | Sensitivity | Precision |
| 20% | 0.85 | 0.48 | 0.401 |
| 40% | 0.88 | 0.63 | 0.477 |
| 60% | 0.90 | 0.76 | 0.544 |
| 80% | 0.91 | 0.87 | 0.587 |

DECISION TREE: OBSERVATIONS

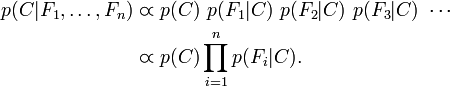
* Despite of high accuracy at 20% training set, Sensitivity is low. Actual sensitivity increases at a very high rate with increase in training set percentage. This clearly explains the impact of higher of training set on actual prediction rate.
* Precision also increases on increase of training set
* Accuracy increases but not really fast with increase in training set percentage

DECISION TREE MODEL ENDS HERE

SECOND MODEL ON NEXT PAGE

SECOND MODEL

NAÏVE BAYES BY JIANGLIN WU

Naïve Bayes classifier is a probabilistic classifier based on Bayes’ theory. The naïve Bayes probabilistic model is a conditional model. Based on Bayes’ theorem, C:\Users\Jianglin\Desktop\bayes.png where F1,…,Fn are the inputs features and C is the categorical value. The goal of naïve Bayes classifier is to find the categorical value, at which P is the largest. Since denominator of the equation does not depend on C, so we are more interested in the numerator which can be written as joint probability C:\Users\Jianglin\Desktop\044dfd3e1b822193af59618fa30640b2.png. We assume that each input feature Fi is conditionally independent. Then, the joint model can be expressed as 

Therefore, We only need to calculate prior P(C) and independent probability distributions P(Fi|C). P(C=c)=(total number of data belong to Class c/total number of data) and P(Fi=f|C=c)=total number of data with feature Fi=f in Class c/total number of data in Class c, if the feature is discrete. Otherwise, we have to assume a distribution models for the features from data. Typically, most data are distributed according to a Gaussian distribution or uniform distribution.

In this project, feature Age, Balance, Day, Month, Duration, Campaign, Pdays and Previous are numeric values. We will use two solutions to deal with these features. 1. Discretize numeric variables. For instance, in Age feature, set age between 19 and 32 as category 0, etc. 2. Use Probability Density functions. We assume all of these features follow Gaussian or uniform distributions. We are trying to find a better Naïve Bayes classifier for the data we have.

Discretize Numeric Variables: There are 16 input features: Age, Job, Marital, Education, Default, Balance, Housing, Loan, Contact, Day, Month, Duration, Campaign, Pdays, Previous, Poutcome and 1 output: if client subscribed a term deposit.

TRIAL 1

Based on the data, we initially set 5,12,3,4,2,10,2,2,3,31,12,10,5,5,5 and 4 categories to each feature respectively. For some numeric features, we set break point according to median value. Other numeric and Boolean type features have same category numbers to the number of their values.

|  |  |  |
| --- | --- | --- |
| Feature | Number of Categories | Break points |
| Age | 5 | 32,36,43,51 |
| Balance | 10 | -1,22,128,263,425,725,1147,1956,3919 |
| Duration | 10 | 58,89,119,149,180,226,285,383,579 |
| Campaign | 5 | 1,2,3,4 |
| Pdays | 5 | -1,135,188,325 |
| Previous | 5 | 0,1,2,3 |

Result:

Size of Dataset=45211.

Test 10 times.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Percentage Of Training Data | Average On Training Data | | | Average On Testing Data | | |
|  | Accuracy | Sensitivity | Precision | Accuracy | Sensitivity | Precision |
| 20% | 0.88982 | 0.44919 | 0.54317 | 0.88845 | 0.43348 | 0.52738 |
| 40% | 0.88977 | 0.44843 | 0.53990 | 0.88788 | 0.43398 | 0.52799 |
| 60% | 0.88968 | 0.43999 | 0.53348 | 0.88941 | 0.44225 | 0.53197 |
| 80% | 0.88977 | 0.44187 | 0.53693 | 0.89007 | 0.44092 | 0.53701 |
| 100% | 0.88991 | 0.43959 | 0.53596 | 0.88517 | 0.39423 | 0.50122 |

For the 100% of training data set, randomly select 10% of total data as testing data.

Accuracy: Correct prediction / Size of testing set,

Sensitivity: Percentage of actual positive outputs that have been detected,

Precision: Percentage of calculated positive outputs that are actual positive.

At any percentage of training data set, the accuracy, sensitivity and precision don’t change much. Therefore, for the reaming trails in this project, we will use all data as training set and same 10% of total data as testing set.

TRIAL 2

Then, filter training and testing data by features. Train the Naïve Bayes Classifier by the same training set and test on each filtered data set. Ignore those category values which have less than 5% of total number of training or testing set.

In most cases, filtered data will increase the accuracy sensitivity and precision. (See Trial2Output.txt)

TRIAL 3

Since many features don’t have enough training or testing data, we merge some of the feature categories.

|  |  |  |
| --- | --- | --- |
| Feature | Merged categories | Number of categories after mergence |
| Age | None | 5 |
| Job | 035, 4781011, 1269 | 3 |
| Marital | None | 3 |
| Education | 01, 23 | 2 |
| Default | None | 2 |
| Balance | 012, 345, 6789 | 3 |
| Housing | None | 2 |
| Loan | None | 2 |
| Contact | None | 3 |
| Day | 0-9, 10-19, 20-30 | 3 |
| Month | 012, 345, 678, 91011 | 4 |
| Duration | 012, 345, 6789 | 3 |
| Campaign | None | 5 |
| Pdays | 0, 1234 | 2 |
| Previous | 0, 1234 | 2 |
| Poutcome | 0, 1234 | 2 |

The overall and almost all the splitting result after mergence becomes worse.

Overall:

Training size: 45211,

Testing size: 4520

Correct Prediction: 3921

Accuracy: 0.8674778761061946

Sensitivity: 0.29423076923076924

Precision: 0.3974025974025974.

TRIAL 4

Filter by Feature value after mergence: (See Trial4Output.txt)

The result shows that merge categories will not help increase accuracy, sensitivity or precision. Compare to the unmerged data set (Trial 3 vs. Trial 1, Trial 4 vs. Trial 2), merged data decreases results.

Use Probability Density functions.

TRIAL 5

Assume the numeric values Age, Balance and Duration follow Gaussian distribution and Day and Month follow uniform distribution. From previous results, we also know that most common category values (more than 80%) of Pdays and Previous are -1. Therefore, we change numeric Pdays and Previous to categorical (“-1” or “>0”). Similar solution for Campaign, it’s difficult to find the distribution of Campaign, so we keep Campaign as categorical feature.

Overall:

Training size:45211,

Testing size:4520

Correct Prediction: 3976

Accuracy: 0.879646017699115

Sensitivity: 0.425

Precision: 0.4742489270386266

Compare the results to Trial 1, the accuracy, sensitivity and precision by using probability density functions decrease. In this set of data, discretize numeric variables will have better results than using probability density functions.

There are also some interesting results during the test.

In trial 2,

|  |  |
| --- | --- |
| Filter By Feature, Category Value | Precision |
| Job, 11 (Service) | 0.73684 |
| Balance, 0 (<0) | 0.84210 |
| Loan, 1 (Yes) | 0.64 |
| Contact, 1 (Telephone) | 0.61764 |
| Day, 17 (18th) | 0.78947 |
| Month, 10 (November) | 0.63333 |
| Duration, 0 (<=58 seconds) | 1.0 |
| Duration, 7 (>285 and <=383) | 0.66666 |
| Campaign, 3 (4) | 0.60869 |
| Campaign, 4 (>=5) | 0.68965 |

Filter results have more than 60% precision in these situations above. That means the banks will really get a deposit if their conclusion is “Yes” based on their calculations. If the Duration category value=0 which is “<=58 seconds”, the precision are all 100%. If the Balance category value=0, which is “No balance”, then the precision is about 82%. The results tell the bank that they should focus on the clients who have these characteristics, and try to make phone call duration less than 58 seconds. They can also have more bank phone representatives on 18th each month and in November.

CONCLUSIONS

In this paper, we used two Data Mining models to increase the efficiency of Direct Marketing Campaigns. We used real world data collected from a Portuguese bank. We implemented two models Decision Tree and Naïve Bayes Classifier and analyze their results in different settings and situations. Before implementation, we preprocessed the data to make it suitable for Input and also removed the data containing missing values.

In conclusion, we find out that **based on Average**:

1. **Sensitivity: Decision Tree Model performed better than Naïve Bayes Classifier**.

Decision Tree Model gives maximum Sensitivity of 0.87 at 80% Training set. On the other hand, Naïve Bayes gives maximum Sensitivity of 0.44 at 80% Training set.

1. **Accuracy**: **Naïve Bayes Classifier performed better than Decision Tree.**

Naïve Bayes Classifier gives avg. accuracy of 0.88 on Testing data whereas Decision Tree gives 0.82 accuracy on Testing data.

1. **Precision**: **Decision Tree Model performed better than Naïve Bayes Classifier.**

In terms of Precision, the maximum precision value generated by Decision Tree is 0.587 whereas the maximum Precision value generated by Naïve Bayes is 0.54.

Some other problems

Since we implement different models by using same set data, the results should be close to each other. We get similar Accuracy and Precision in two models, which prove that both models work properly. Accuracy and Precision in two models do not increase or decrease too much, they remain at about 88% and 50%. However, there is huge difference between Sensitivities in the two models. The Sensitivity in Decision Tree increases from 48% to 87%, whereas it stays at about 44% in Naïve Bayes.

Feature Study

1. We still have the problem that why the Sensitivities are unstable.
2. In this paper, we found that filtered data will increase the accuracy, sensitivity and precision. There are many different ways to filter data. However, due to the complexity, we only tried one of the possible filters. We are interested in finding the best filters that will produce highest results and the features that contribute the most in Naïve Bayes model.

REFERENCES

"UCI Machine Learning Repository: Bank Marketing Data Set." *UCI Machine Learning*

*Repository: Bank Marketing Data Set*. N.p., n.d. Web. 21 Dec. 2012.

Wikipedia contributors. "Decision tree." *Wikipedia, The Free Encyclopedia*. Wikipedia, The

Free Encyclopedia, 19 Dec. 2012. Web. 21 Dec. 2012.

Moro, Sergio, Raul M. Laureano, and Paulo Cortez. *http://repositorium.sdum.uminho.pt*. Rep.

N.p., n.d. Web. Dec.-Jan. 2012.