

SEGMENTATION OF CREDIT USERS FOR BETTER BANKING CUSTOMER RELATIONSHIP MANAGEMENT

¹V. Priyanka, ²Prof. K. Mallikarjun Reddy

Research Scholar, Dept. of Business Management, Osmania University.
Retd. Professor, Dept. of Business Management, Osmania University.

Abstract:- Building a strong, enduring, and profitable connection with consumers is contingent upon having a thorough understanding of them, striving to fulfill their needs and preferences to the fullest extent possible, and doing all of this in today's cutthroat business environments. The foundation of customer relationship management is this. A strong knowledge of the customer is the cornerstone for increasing the customer lifetime value, which includes customer segmentation. In order to build lucrative and expanding customer groups that will allow businesses to target each category with specialized products, customer segmentation aims to group consumers based on shared criteria. This cannot be accomplished without the use of clever data analysis strategies and methodologies. With an emphasis on credit users' segmentation tasks in the banking sector, this study focuses on business strategy-driven customer segmentation in an effort to optimize customer potential, which is the most valuable resource in company. The case study that is being presented shows how a multilayer feed

forward neural network may be used to divide bank clients into two categories: those who experience payment issues and those who do not.

1. INTRODUCTION

Companies must now transition from the outdated "designbuild-sell" marketing approach to the customer-oriented perspective of relationship marketing in order to compete in today's competitive marketplaces. This involves focusing on consumers' requirements, desires, attitudes, behaviors, preferences, and perceptions. Due to many interrelated factors, including: (1) shortened marketing cycle times; (2) higher marketing expenses; (3) streams of new product offers; and (4) specialty rivals, customer relationship management is an extremely difficult subject [9]. In order to respond to each and every one of these needs, successful businesses must move quickly. This requires obtaining comprehensive consumer information in order to identify both the most lucrative and unworthy clients for the company and to create the ideal offer. CRM refers to these

endeavors as customer relationship management.

The collection of client data is the first step in CRM. A marketing data mart and certain feeds from other business information systems may include all the data needed for CRM.

CRM must use statistical, OLAP, and/or data mining technologies to examine the data. Market segmentation is the ultimate outcome. By using customer segmentation, a business may find client groups with distinctive qualities that can help it manage and target consumers more successfully. A significant segment is a subset of consumers who, according to [2], satisfy three requirements: they are lucrative, observable via shared traits, and experiencing growth.

One method for deriving meaning from data is data mining. When data mining techniques are used for consumer segmentation, they utilize information from corporate databases, data warehouses, or user-generated content from websites to create models that characterize patterns and correlations in the data. It locates and gathers information that may help enhance the comprehension and use of the data to customize specific marketing plans for market niches that are appealing to companies and enhance customer service.

We described our effort to segregate bank clients' data in this paper. The objective was to use neural networks to split clients into groups based on borrower solidity as determined by credit score. The binary answer variable "creditability" is accessible for each customer, representing both creditworthy and non-creditworthy consumers.

2. CRM AND CUSTOMER SEGMENTATION

"Managerial efforts to manage business interactions with customers by combining business processes and technologies that seek to understand a company's customers, i.e. structuring and managing the relationships with customers," is how customer relationship management, or CRM, is described [6]. Dividing the client base into groups or segments to find the customers with the highest profit potential is the quickest approach to create a successful customer-focused company. According to [7], customer comprehension is the foundation of CRM. Increased client lifetime value is the result of appropriate customer comprehension and actionability. Inaccurate comprehension by the client may result into hazardous actions. The law of diminishing return states that misdirected behaviors, such endless attempts to acquire or keep every client, may also result in a

decline in customer lifetime value. Therefore, proper consumer comprehension and coordinated actions resulting from it should be prioritized.

"The process of dividing customers into distinct, meaningful, and homogeneous subgroups based on various attributes and characteristics" is the definition of customer segmentation given in [8, pp. 189]. It serves as a technique for marketing differentiation. It helps businesses to comprehend their clientele and develop unique tactics based on their traits. This idea is further clarified in [10] as "a procedure that splits clients into more manageable groupings known as segments. Within a segment, homogeneity is required, although heterogeneity across segments is preferred. Put another way, clients in the same segments have a comparable or same set of characteristics. However, consumers in various markets have unique sets of characteristics.

One of the best things about CRM is that it helps a business identify its best customers and give them extra attention. However, as [4] points out, this could work against the company, so it's important to exercise caution when developing relationships with best customers and letting go of customers who are labeled as least valuable.

People with similar traits tend to exhibit similar behavioral patterns, which is why customer segmentation is often used. These behavioral patterns are especially significant in customer relationship management, marketing, and the credit and insurance industries.

There are several methods for segmenting a consumer base based on their goals and explicitly stated motivations. Techniques for segmenting data that may be used are either focused on trend monitoring and forecasting or are motivated by business strategy. The former assigns customers to business strategies based on strong predictive business knowledge, while the latter forecasts over well-designed customer segmentation and employs trend analysis to track and identify freshly developing trends.

Our study focuses on customer segmentation driven by company strategy in an effort to optimize the use of customer potential, which is the most valuable resource in business. Predictive models are used to calculate customer potentials based on previous data. Customer metrics are used to measure them. In order to manage customers proactively, corporate plans and actions are then determined by predictive customer data.

Numerous customer care tasks are included in predictive CRM, including market research and customer acquisition, customer conversions, debt collection management, fraud detection, transaction audit, customer loyalty programs, cross-selling, up-selling, reselling, customer attrition and churn management, trend monitoring, and call center analytics [10].

The segmentation task of credit consumers in the banking business is the main topic of our article. Banks must to categorize their clientele and use distinct marketing techniques for every group that they identify. Customers are often evaluated, categorized based on values, and their unique traits are identified as part of the categorization process

3. DATA USED FOR CUSTOMER SEGMENTATION

The preceding section presented several forms and uses of client segmentation, contingent upon a firm.

Objective: Behavioral, attitudinal, or loyalty traits, customer value, and sociodemographic and lifestyle data are some of the segmentation factors they use. As per [10], the data that is most often used for client segmentation may be categorized into the following groups: Customer characteristics such as age, gender, race,

education, occupation, and income are examples of demographic segmentation variables. Geographic variables include zip code, state, climate, population, and other information that can be gathered from national census data. Psychographic segmentation variables include personality, values, attitudes, and other information on a person's lifestyle. Behavioral segmentation variables include product usage rate, brand loyalty, ready-to-buy stage, and so on. Past business history refers to customers' previous business track records.

Numerous studies have been conducted to determine customer value using customer lifetime value (LTV) as a segmentation criteria. It is suggested in [6] that there are several shortcomings with the conventional LTV calculation. As a result, a unique method for calculating lifetime value (LTV) is suggested, one that takes into account the existing worth of the customer (derived from past profit contribution), prospective value (derived from the customer's future financial contribution, potential for profit production, and anticipated service durations), and customer loyalty. The authors provided examples of how to apply the suggested LTV to a wireless communication company's client segmentation.

Sociodemographic data and information on consumer behavior—such as a customer's gender, whether or not they use SMS, how often their payment method was changed, how they pay for deposits and registration fees, what their occupation is, how many optional services they use, etc. formed the basis of the research.

You may examine the customer segmentation input data based on the data source. Internal data sources include summary tables that provide information on customers from marketing, HR, or finance perspectives; customer surveys that ask specific questions of a subset of the customer base; and behavioral data from transaction systems like credit card records and browser logs. Purchased databases, phone book or address lookups, household hierarchies, Fair Isaac credit ratings, and webpage browsing profiles are examples of external data sources [3, pp. 492].

The information that is accessible about individuals with whom one does not already have a connection is much more restricted than that of one's current clientele, which consists only of a few broad characteristics like city, state, and ZIP code. As a result, there is little predictive potential in such data. This is due to the fact that client data is often gathered via transactional or self-provided means. General descriptive information is provided

by the customers themselves, while transactional information includes information on the product name, amount, location, and time of purchase. Many studies concentrate on keeping current customers rather than acquiring new ones since it is very difficult to do adequate research on possible client preferences due to a lack of knowledge about new customers [1].

In situations when client data is scarce or nonexistent, a marketing survey might be used to gather pertinent data.

The secret to this process is to establish a connection between one's knowledge and actions want to act like a model.

The amount of information that Serbian banks gather about their clients is extensive and includes personal information, information about where they live, qualifications, employment, position, and industry; information about the spouse's work and income; information about the marital status; information about the family; information about other credits; information about credit cards; information from the employer; information about the employment organization; credit and membership payments; and authorized credit card users/warranties. These statistics, however, are not accessible to the

general public since they are regarded as business information.

As a result, we had no choice but to use the German bank's sample dataset, which you can get at http://www.stat.unimuenchen.de/service/datenarchiv/search_e.html. The research we conducted, which is detailed in the sequel, involved the use of the following data to segment bank customers through a data mining process: demographic data (age, sex, marital status, type of apartment, occupation, and so on); current account data (balance, percentage of available income, value of savings or stocks, etc.); and previous credit history (payment history, number of prior credits at this bank, etc.).

4. CRM DATA MINING TECHNIQUES

Because so much customer data is gathered, it may be difficult to extract the valuable insights that marketers need to build proactive, intelligent paths back to customers. This makes the process of customer segmentation very complex and difficult. It is no longer feasible to wait until the indications of consumer unhappiness are clear before taking action, as mentioned in [9]. Businesses need to be proactive and anticipate consumer needs if they want to flourish.

Therefore, to best expose this information, apply sophisticated analytical methods. Businesses may sift through layers of apparently unconnected data to find important links where they can predict, rather than just respond to, client requirements with the use of data mining (DM) technology and methodologies.

In [1], the following justification is provided for the use of data mining in CRM: Even if a seasoned marketer can often choose relevant demographic selection criteria, the procedure becomes more challenging as data volumes rise. The more consumers that are taken into consideration and the more information that is available for each client, the more complicated the patterns become. Given the explosive increase of consumer datasets in recent years, the task of categorizing potential clients is becoming more and more difficult. Organizations may identify important customers, forecast future behaviors, and make knowledge-driven choices by using data mining to uncover predictive information buried in massive datasets. This allows CRM to go beyond assessments of historical events and enable future-oriented CRM.

Finding pertinent patterns in a database may be automated via data mining. It creates models that forecast consumer behavior by using tried-and-true statistical

and machine learning methodologies. [6] describes the use of logistic regression, decision trees, and artificial neural networks to assess customer value in the wireless communication sector. Since numerous lucrative optional services were offered to clients, the authors defined the potential value of customers as predicted earnings that may be earned from a specific customer when the customer employs the extra services of a wireless communication business. Using data on sociodemographic characteristics and service-related buying patterns, the authors first estimated the potential customer value by analyzing the relationship between input factors and the reality of whether or not the extra service was actually utilized. The most potent data mining technique was used for every extra service that was provided. Neural networks and decision trees both performed better in this investigation than the logarithmic regression data mining technique. Using data mining methods, the authors of the same research calculated the churn rate by measuring the chance of each client leaving. Decision tree, neural network, and logistic regression models were compared to determine the likelihood of adopting a provided service. Based on the outcomes of comparison tests using the lift chart approach and misclassification rate, the best model was then chosen. It was found that the logistic regression model was the

most accurate in predicting the customer attrition rate.

A retailer is described in [3] as being able to automatically analyze its point-of-sale data and correlate shop groupings with sales patterns via the use of clustering and neural network technologies. According to reports, the acquired findings allowed retail management to estimate demand for stock holding units, forecast seasonal changes at the store-item level, and run a just-in-time inventory program much more successfully. Additionally, a case study is presented in which a home equity marketer used logistic regression response probability modeling and chi-squared automatic interaction detector (CHAID) segmentation to increase the efficiency of targeting current mortgage customers who might be interested in the client's service by at least 10%. The unique Internet technology was used for the deployment of the predictive models.

In order to create forecasting models, the authors in [12] used neural network technology, the C5.0 decision tree model, the classification and regression (C&R) tree, and CHAID. Following an analysis of the forecasting performance of the four models, the C5.0 algorithm was selected and used to develop all of the "if-then" rules that delineate the attributes of various bank client types.

In order to more precisely profile client groupings, we presented in section 6 the use of supervised neural network technology in conjunction with the unsupervised Fuzzy C-means clustering approach for bank customer categorization.

Analysts are still required to evaluate model findings and confirm the validity of the model predictions, even when DM tools offer the results in a form that is meaningful for business users. The pearls that data mining seeks to uncover are non-intuitive relationships between variables and client behaviors. Additionally, the output of the data mining process should be placed in a context that the user is familiar with in order to immediately root the findings in reality, since the user is unaware of what the process has uncovered previously.

5. DATA MINING PROCEDURE FOR CRM

The process of managing a customer's relationship involves several following phases. An business has to have the necessary client data before it can do CRM. As previously said, raw data is collected and processed from a variety of internal and external databases, marts, or warehouses. The data transformation step is the gathering of all of this data into a single location where it can be seen and examined.

Additionally, the raw data must be cleaned up and normalized in order to be ready for data mining. From a huge dataset consisting of 16,384 customer records, the authors of [6] randomly selected roughly 2000 records for the research. Of the 200 data fields, they selected 101 data fields based on the outcome of eliminating unnecessary data fields. Missing values were replaced with the mean value for continuous values and the mode value for class variables. To obtain equal scales, the original data must be normalized, which is the process of transforming the data into real values between 0 and 1. The outcome of customer segmentation is shown in a three-dimensional space of present value, projected value, and customer loyalty in [6] by the application of normalization. The three axes' original scales had to be transformed to the interval [0,1] since they were not the same.

In order to identify market segments and make specific judgments about them, the data must next be evaluated using a CRM technique. When data mining methods are used, consumer knowledge is automatically found. The campaign management process utilizes the customer segmentation data, often in the form of a predictive model, to choose targets with the biggest potential for profit.

The additional phases in client segmentation that are suggested in [5] are as follows: 1. the creation of customer segments and the completion of the business rules governing the division of consumers into segments; 2. Targeting client segments: The objective is to choose important segments for promotional endeavors.

More market research is appreciated as it may provide more qualitative data to certain groups; 3. Segment positioning - To do this, relevant KPIs (key performance indicators) must be set up to track segment performance; 4. Establishing dedicated customer segment management teams that will collaborate and operate independently to develop an integrated strategic marketing plan based on the unique needs of the segments, with an emphasis on tasks like budget and planning, channel management, brand management, product management, and marketing services management

6. CASE RESEARCH

The segmentation of bank clients is the issue that this case study focuses on. Our case study focuses on merging the analytical findings from various tools and DM approaches in order to get enhanced outcomes, just as each data mining endeavor is distinct to either the applied methods and techniques or the employed

tools. Therefore the analysis is partially undertaken in Weka2 tool for data mining, where the classification and prediction are conducted using the technique of a multilayer feed forward neural network – Multilayer Perceptron (MLP). The Fuzzy C-means clustering technique combines the outcomes of further data analysis carried out in DataEngine 3 with the findings gained in Weka. These technologies operate together to greatly simplify the job of analysts in knowledge discovery by supporting the interpretation of data and making it easier to draw conclusions that are unambiguous.

The two distinct categories of bank clients are those who are credit worthy and those who are not. This categorization is reflected in the binary answer variable "creditability" for each client.

The dataset consists of 1000 customer credits from a German bank. In the first study, we sought to determine if past credit attribute payment history, with values indicating no payment issues (value 0) and payment issues (value 1), could be used to forecast future credit users, both good and bad. The following additional attributes were used in this analysis: balance of the current account, with values of 1 denoting no running account, 2 denoting no balance or debit, 3 denoting balance between 0 and 200 currency units, 4 denoting balance

higher than 200 currency units, or checking account for at least a year; percentage of available income, with values of 1 denoting more or equal to 35, 2 between 25 and 35, 3 denoting between 20 and 25, and 4 denoting less than 20; further running credits, with values of 1 at other banks, 2 at department stores or mail order houses, and 3 denoting no further running credits; amount of credit (metric); number of previous credits at the specified bank (including the running one), with values of 1 denoting one, 2 - two or three, 3 - four or five, and 4 - six or more.

The distribution of values for all characteristics, including the class attribute, when analyzed, reveals a far higher percentage of customers who make their debt payments on time. Weka, a data mining tool, is used for the study. The intended usage of the constructed model is to categorize current customers into those who have and do not have credit payment issues. This allows for the prediction of consumers or credit users who may pose a risk to the bank based on certain variables. The Multilayer Perceptron (MLP), a multilayer feed forward neural network, is the method used for classification and prediction.

Since the data was sorted by the column payment of prior credits and the goal was to include a certain number of customers with payment issues in the test set, during the

preparation phase, we checked to make sure there were no missing values in the data and applied a Randomize filter to the whole set. The split percentage between the test and training sets of data was 66.

The following are the parameters of the optimal MLP model:

Training time was set to 500 epochs; normalization of attributes was set to True because the amount of credit is metric and significantly out of range of other attributes; number of hidden layers was set to "a," meaning that the number of nodes in a hidden layer is calculated as one half of the sum of the number of attributes and classes. Learning rate was set at 0.3 and momentum at 0.2. As a result, in our instance, the buried layer has 3 nodes. This neural network contains two nodes, one for each class value, as would be anticipated.

Fig. 1 shows the architecture of this neural network. 0.0683878 was the computed error per epoch.

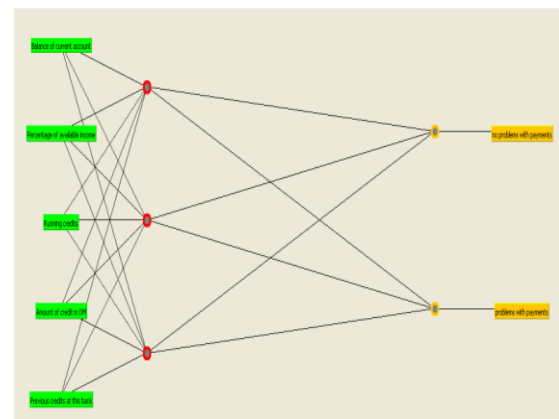


Fig 1. The neural network architecture

Of the test set, 92.0588% of the occurrences were properly categorized, while 7.9412% of the instances were wrongly classified. The root mean squared error was 0.2726 and the mean absolute error was 0.1311. All of this suggests that the model's categorization and prediction were accurate.

```

=== Confusion Matrix ===
      a    b  <-- classified as
306    8 |   a = no problems with payments
  19    7 |   b = problems with payments
  
```

Fig 2. The confusion matrix

The confusion matrix used to classify the German bank dataset is shown in Fig. 2. As can be shown, the neural network's configuration incorrectly identified customers experiencing payment issues as those who do not. The next objective was to attempt to enhance the neural network's performance since it is preferable to anticipate customers without issues as ones who have issues rather than the other way around.

Setting a decay feature to True was the first step in improving the performance of the neural network. When decay is enabled, each time a new epoch is performed, the learning rate is decreased because the value of the learning rate is divided by the number of epochs. In this manner, early in the training process, neural network training will begin with somewhat high weight

changes and gradually lower the learning rate as it approaches the ideal weight values.

The second modification concerned the discontinuation of network training. The stop condition was changed from 500 epochs finished, as it was in the prior setup, to when the error falls below the stated value.

The validation Threshold the number of consecutive epochs during which the error on the validation set is permitted to worsen before training is stopped was also set to its default value of 20, as was the validation Set Size, which represents the proportion of the dataset utilized for validation. In this training approach, overfitting is indicated by a rise in error values for the validation set, rather than by a predetermined error threshold being attained.

Weka may continue to refine the network using this technique until overfitting occurs. A few nodes were eliminated from the neural network's topology in order to enhance the results, but not much. This was done since altering the decay parameter and stop condition had no effect on the outcomes. Given that the model did an excellent job of categorizing bank customers who are problem-free and that other model quality metrics were also good, we deduced that the findings were impacted

by the fact that a very tiny percentage of credit users experienced payment issues.

The study presented here combines the findings from Weka with the outcomes of further data analysis performed in DataEngine using the fuzzy C-means clustering technique. This additional research provided a more thorough explanation of the composition of the two client segments: those who experienced credit payment issues and those who did not. Cluster center features are shown in Fig. 3. Because of the large differences in the data ranges, attribute values are standardized. In Table 1, the denormalized data are shown.

TABLE I
Denormalized Attribute Values

Cluster label Attribute	NoProblem	Problem
Balance of current account	2.161571946	2.140222608
Percentage of available income	2.1762141836	3.6225730707
Amount of credit	3954.05600934	2796.78071686
Previous credits at this bank	1.3831241008	1.3523150874
Running credits	2.6308668224	2.7403758108

The amount of free income that separates the two categories of bank clients is the primary distinction. The bulk of customers who do not have credit payment issues often have free incomes between 25 and 35%, while the majority of customers who have credit payment issues typically have smaller free incomes between 20 and 25%.

Additionally, compared to clients in the other sector, those who struggle with credit

payments often seek for smaller credit amounts. Further data analysis revealed that the demographic information of clients with and without credit payment issues does not vary much, but at the same time the majority of consumers having credit payment troubles have short or long term deposits between 100 and 500 currency units, and own shares/stocks equivalent to 500 to 1000 currency units.

The quickest approach to classify clients into different business strategies and determine which customers have the most profit potential is to divide a customer base into groups or segments. "Marketing managers can combine many basic strategies... to develop a customized strategy according to each customer segment," according to [6]. Also in [2] the authors illustrated via an example of customer segmentation for wholesaler-distributors how segmentation can identify strategies to better deploy sales reps. In the credit and insurance industries, where effective customer segmentation may reduce exposure to risk associated with credit and insurance, Reference [10] provides an example of business strategies based on the outcomes of customer segmentation. The authors of [12] identified the shared traits of many customer types and demonstrated how marketing professionals may locate

possible clients and use various marketing approaches for various clientele.

According to our case study on banking customer segmentation, bank policies may be tailored to possibly "problematic" consumers based on the aforementioned distinguishing features. For example, one measure of precaution in credit approval should be to limit the amounts of total credits approved for a customer to a predetermined percentage of his continuous income, and/or to introduce additional evaluation measures for credit claims of lower amounts for customers that comply with the profile - have a certain range of short or long term deposits and possess shares/stocks whose value belongs to the "critical" range.

7. CONCLUSION

Modern firms are aware that companies with a higher likelihood of success and increased profit are those who use client data efficiently. As a consequence, it is not surprising that attempts to segment customers are growing and have a wide range of uses. In the article, we highlighted a particular usage of credit user segmentation for enhanced banking customer relationship management. Using the multilayer feed forward neural network approach, a classification model was created in the Weka tool to categorize

current customers into those who have and do not have credit payment issues. The fuzzy C-means clustering technique was used in DataEngine to do further data analysis. Given certain covariates that should be regarded as a possible risk to the bank, the classification model could be used to predict which customers or credit users would be at risk. On the other hand, the clustering model, which rejected customer characteristics that set them apart, could be helpful for customizing bank policies for each customer segment.

8. REFERENCES

- [1] Alex Beerson, Stephen Smith, Kurt Thearling, Customer Acquisition and Data Mining, [Online] Available http://www.thearling.com/text/chapter10/c_hapter10.htm, December 21, 2010.
- [2] Brent R. Grover, How To Segment Customers, Modern Distribution Management - the Newsletter for the Wholesale Distribution Channel, 2004, [Online] available at http://www.mdm.com/issues/cgibin/udt/im_display.printable?client_id=mdm&story_id=2155
- [3] Chris Rygielski, Jyun-Cheng Wang, David C. Yen, Data Mining Techniques for Customer Relationship Management, Technology in Society, No.32, Vol. 24., 2002, pp 483-502.

[4] Christopher J. Bucholtz, How Customer Segmentation Can Unravel CRM, CRM Buyer, 2010.

[5] Customer DNA, Data Mining & Market Research in Segmentation Management, September 2010, [Online] Available [http://www.customersdna.com/index.php?page=shop.product_details&flypage=tpflypage. tpl&product_id=51&category_id=14&option=com_virtuemart& Itemid=53](http://www.customersdna.com/index.php?page=shop.product_details&flypage=tpflypage.tpl&product_id=51&category_id=14&option=com_virtuemart&Itemid=53)

[6] Hyunseok Hwang, Taesoo Jung, Euiho Suh, An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry, Expert Systems with Applications 26, 2004, pp. 181–188.

[7] Jaideep Srivastava, Data Mining for Customer Relationship Management, [Online] Available <http://www.dtc.umn.edu/ddmc/resources/crm.pdf>, January 15, 2011.

[8] Konstantinos Tsipis, Antonios Chorianopoulos, Data Mining Techniques in CRM Inside Customer Segmentation, John Wiley & Sons, Ltd, 2009.

[9] Kurt Thearling, Data Mining and Customer Relationships, [Online] Available

<http://www.thearling.com/text/whexcerpt/whexcerpt.htm>, January 15, 2011.

[10] Rosella Predictive Knowledge & Data Mining, Customer segmentation, [Online] Available <http://www.roselladb.com/customersegmentation.htm>, November 28, 2010.

AUTHORS



1. Mrs. V. Priyanka, Research Scholar at Osmania University, pursuing PhD in Management and MBA from Aurora's PG college. She has 10 years of teaching Experience.



2. Prof. K. Mallikarjun Reddy, Retired Professor and Ex. Dean, Head of Department of Business Management, Osmania University.