© 2018 IJSRSET | Volume 4 | Issue 8 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section : Engineering and Technology

Analyze of Mobile Shopping Using Naivebayes Classifier

T. Lucia Agnes Beena[1], A. Irudhaya Amala Benitta[2]
[1]Head of the Department, Department of IT, St. Joseph's College, Trichy, Tamil Nadu, India
[2]M.Sc., Computer Science, Department of IT, St. Joseph's College, Trichy, Tamil Nadu, India

ABSTRACT

Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential used in various commercial applications including retail sales, e-commerce, bioinformatics etc. There are varieties of popular data mining task within the data mining e.g. classification, clustering, outlier detection, association rule, prediction etc. This paper applies Naive Bayes classifier to analyze the behavior of youngsters in mobile purchase.

Keywords: Data Mining, Weka Tool, Classifier, Naivebayes

I. INTRODUCTION

In recent years, the advent of information technology has transformed the way marketing is done and how companies manage information about their customers [1].

Data mining can be used in various applications:

Banking: loan/credit card approval, predict good customers based on old customers, view the debt and revenue changes by month, by region, by sector, and by other factors , access statistical information such as maximum, minimum, total, average, trend, etc [2].

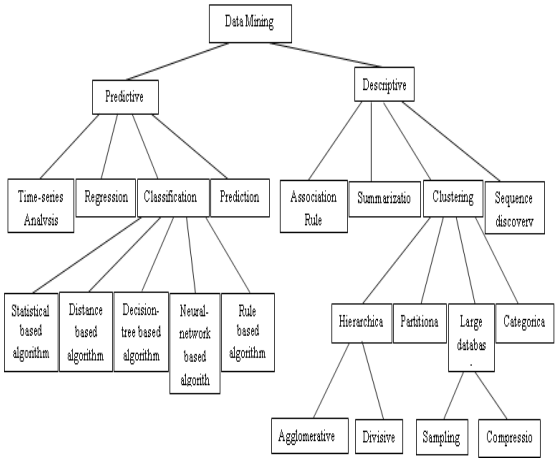

Figure 1. data mining model & Task

Telecommunication industry: To identify potentially fraudulent users and their a typical usage patterns, detect attempts to gain fraudulent entry to customer accounts, discover unusual patterns which may need special attention, find usage patterns for a set of communication services by customer group, by month, etc., promote the sales of specific services, improve the availability of particular services in a region.

Retail Industry: To Identify customer buying behaviors, discover customer shopping patterns and trends, to improve the quality of customer service, achieve better customer retention and satisfaction, enhance goods consumption ratios, design more effective goods transportation and distribution policies.

Marketing: Data mining facilitates marketing sector by classifying customer demographic that can be used to predict which customer will respond to a mailing or buy a particular product and it is very much helpful in growth of business [3].

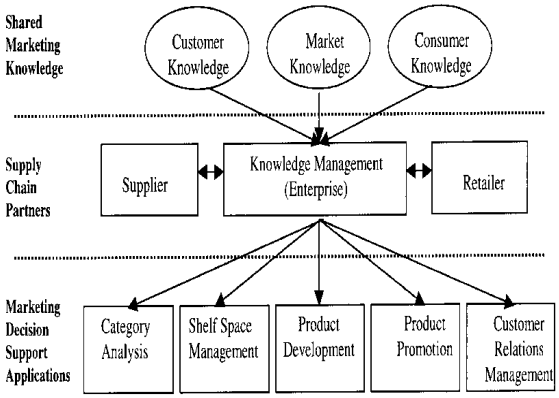

Figure 2. Integrated knowledge management system for marketing

II. MOBILE PURCHASE

Mobile commerce industry is getting more and more competitive because of the explosive increase in the number of mobile users, especially mobile shoppers. Cell phone manufacturers today offer customers more choice than ever when purchasing new mobile phones. But developing a new phone is an expensive undertaking, meaning market researchers must help manufacturers determine exactly what customers want in a phone. According to "Factors Affecting Customer Choice of Mobile Phones: prize, brand, size, Interface, Technical Features. The development of mobile phones and technologies has been an extended history of innovation and advancements cropped up due to dynamic changes in consumers' needs and preferences. This paper analyze mobile purchase pattern of the youngsters which mobile phone best to buy in particular data set.

The future success of cell phone companies rest on marketers' understandings of how we project our personalities on their products. By harnessing this information, analyzing personality factors, and developing advertisements that reflect this understanding, these companies will edge out the competition and dominate the marketplace. The development of mobile phones and technologies has been an extended history of innovation and advancements cropped up due to dynamic changes in consumers' needs and preferences. Among these developments, mobile phone devices have had one of the fastest household adoption rates of any technology in the world's modern history now a days, mobile handsets have become an integral part of human daily life and personal communication across the globe [4].

Main Trends in cell phone features

The youngsters use the smart phone for the Communication services of the voice, Text, Email and Online Video.

Weka Tool

Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes. Machine learning is nothing but a type of artificial intelligence which enables computers to learn the data without help of any explicit programs. Machine learning systems crawl through the data to find the patterns and, when these are found, adjust the program's actions accordingly. Data mining analyses the data from different perspectives and summarizes it into parcels of useful information. The machine learning method is similar to data mining. The difference is that data mining systems extract the data for human comprehension. Data mining uses machine language to find valuable information from large volumes of data.

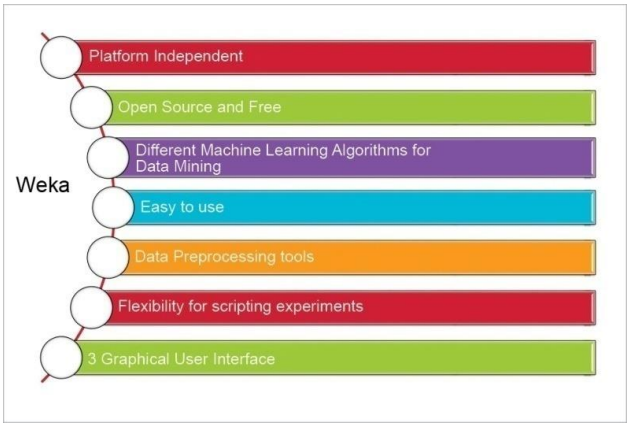

Figure 3. Use of weka tool

Preprocessing

Data preprocessing is a must. There are three ways to inject the data for preprocessing:

Open File – enables the user to select the file from the local machine

Open URL – enables the user to select the data file from different locations

Pen Database – enables users to retrieve a data file from a database source.

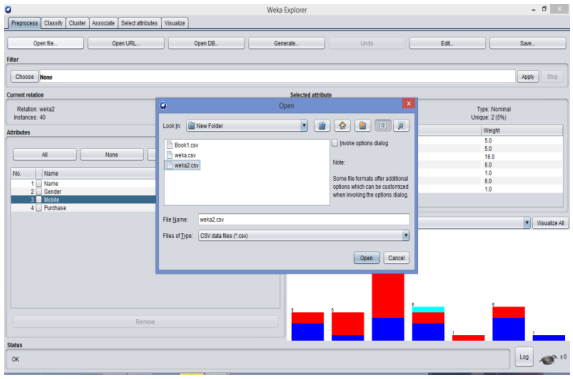

Figure 4. Preprocess weka tool

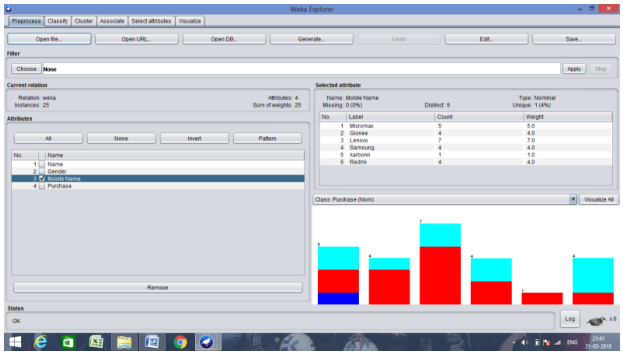

Figure 5. Output preprocess

Classification

To predict nominal or numeric quantities, we have classifiers in Weka. Available learning schemes are decision-trees and lists, support vector machines, instance-based classifiers, logistic regression and Bays' nets. Once the data has been loaded, all the tabs are en1abled. Based on the requirements and by trial and error, we can find out the most suitable algorithm to produce an easily understandable representation of data.

Before running any classification algorithm, we need to set test options. Available test options are listed below.

Use training set: Evaluation is based on how well it can predict the class of the instances it was trained on.

Supplied training set:

Evaluation is based on how well it can predict the class of a set of instances loaded from a file.

Cross-validation:

Evaluation is based on cross-validation by using the number of folds entered in the 'Folds' text field.

Split percentage:

Evaluation is based on how well it can predict a certain percentage of the data, held out for testing by using the values entered in the '%' field.

To classify the data set based on the characteristics of attributes, Weka uses classifiers.

```
Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances           5               35.7143 %
Incorrectly Classified Instances         9               64.2857 %
Kappa statistic                          0.0156
Mean absolute error                      0.2268
Root mean squared error                  0.3358
Relative absolute error                101.3236 %
Root relative squared error            101.7248 %
Total Number of Instances               14

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               0.000    0.333    0.000      0.000   0.000      -0.258   0.208     0.121     Micromax
               0.000    0.000    ?          0.000   ?          ?        0.771     0.268     Gionee
               0.833    0.625    0.500      0.833   0.625      0.228    0.552     0.476     Lenovo
               0.000    0.000    ?          0.000   ?          ?        0.654     0.143     Samsung
               ?        0.000    ?          ?       ?          ?        ?         ?         Karbonn
               0.000    0.000    ?          0.000   ?          ?        0.439     0.262     Redmi
               ?        0.000    ?          ?       ?          ?        ?         ?         Asus
Weighted Avg.  0.357    0.315    ?          0.357   ?          ?        0.517     0.326
```

Figure 6. Screen shot of naive bayes classifier

III. RESULT AND DISCUSSION

The date for this analysis is called from the age group of students 20 to 22. The purchase details, mobile detail and features that made them to purchase that mobile were collected. The weka tool is used to analyses these data. The naïve bayes classifier is applied to predict the result.

Figure 6 the screenshot of Naive bayes classification of mobiles. The PRC area & ROC area values for Lenovo mobile is high compared to the other mobiles.

From this result, it was found that the Lenovo mobile is preferred by most youngsters.

Clustering:

The cluster tab enables the user to identify similarities or groups of occurrences within the data set. Clustering can provide data for the user to analyze [9]. The training set, percentage split, supplied test set and classes are used for clustering, for which the user can ignore some attributes from the data set, based on the requirements.

Association

The only available scheme for association in Weka is the Apriori algorithm. It identifies statistical dependencies between clusters of attributes, and only works with discrete data. The Apriori algorithm computes all the rules having minimum support and exceeding a given confidence level.

Attribute selection

Attribute selection crawls through all possible combinations of attributes in the data to decide which of these will best fit the desired calculation—which subset of attributes works best for prediction. The attribute selection method contains two parts.

Search method: Best-first, forward selection, random, exhaustive, genetic algorithm, ranking algorithm.

Evaluation method: Correlation-based, wrapper, information gain, chi-squared.

Visualisation

The user can see the final piece of the puzzle, derived throughout the process. It allows users to visualise a 2D representation of data, and is used to determine the difficulty of the learning problem. We can visualise single attributes (1D) and pairs of attributes (2D), and rotate 3D visualisations in Weka. It has the Jitter option to deal with nominal attributes and to detect 'hidden' data points.

IV. CONCLUSION

In this paper, data mining is applied to mobile shopping. The data is collected from the youngsters recording mobile shopping. The data mining technique is applied to discover knowledge. Lenovo is one the best choices for mid range and budget series phone. From the analyses it was found most of the purchase Lenovo mobile. It was found that Lenovo has been selected for its sound quality and video clarity and it's also available in attractive.

V. REFERENCES

[1]. Han,J. and Kamber, M., "Data Mining: Concepts andTechniques", 2nd edition.

[2]. Data Mining Introductory and Advanced Topics" by Margaret H. Dunham

[3]. M.J.A. Berry, G. Linoff, Data Mining Techniques for Mar- keting, Sales, and Customer Support, Wiley, New York, 1997.

[4]. Blackwell, RD, Miniard, PW & Engel, JF (2006). Consumer behavior, 10th edn, Thomson South-Western, Boston. Cochran, W. G. (1977). Sampling techniques (3rd ed.). New York: John Wiley & Sons. Comer and T. A. Wikle (2008), Worldwide diffusion of the cellular telephone, (1995-2005). The Professional Geographer, 60(2), 252-269.

[5]. Palak shah, opensourceforu.com (2017/01) an-introduction-to-weka.

[6]. H. Holtz, Database Marketing-Every Manager's Guide to the Super Marketing Tool of the 21st Century, Wiley, New York, 1992.

[7]. D.A. Keim, Visualization techniques for mining large databases: a comparison, IEEE Transactions on Knowledge and Data Engineering 8 Ž6. Ž1996. 923–937

[8]. D. Peppers, M. Rogers, Is your company ready for one-to-one marketing? Harvard Business Review Ž1999. 151–160.

[9]. Kifaya (2009) Mining student evaluation using associative classification and clustering Communications of the IBIMA vol. 11 IISN 1943-7765