

Use of Association Mining to Analyse Customer Behaviour for Small Scale Food Joints -A Case Study Approach

Priti Kulkarni

Asst. Prof
Symbiosis Institute of
Computer Studies and Research
Symbiosis International University,
Pune, Maharashtra—411016, India
pritiap@gmail.com

Adya Ranjan

student
Symbiosis Institute of Computer
Studies and Research
Symbiosis International University,
Pune, Maharashtra—411016, India
15030122009@sicsr.ac.in

Nishigandha Bahirgonde

Student
Symbiosis Institute of Computer
Studies and Research
Symbiosis International University
Pune, Maharashtra—411016, India
15030122055@sicsr.ac.in

Abstract—*This paper on association rules in data mining aims to solve key product marketing problems of small businesses with the help of Market basket analysis. Market basket analysis help to identify which items are purchased together. This has been done by collecting sales data from a food joint. The goal of data analysis is to determine association rules between sold items on the menu. The findings in the study help business to understand customer choices and provide solution to improve sales of food items.*

Keywords—*Association, Apriori, data mining, market basket analysis*

I. INTRODUCTION

In the fast paced and the dynamic world that we live in, predicting the future becomes an impossible task. Change may not affect an individual in the short run. It can make or break an enterprise. Therefore it becomes imperative for a business to prepare for contingencies for the survival of one's business.

One of the most important focuses is customer satisfaction. Businesses are becoming people centric rather than product oriented. Business through its marketing efforts tries to analyse the needs, preferences and dislikes of a customer and aims to satisfy these.

In the food and beverage industry, the customer churn rate is higher than any other which is easily susceptible to incurring losses. A highly competitive industry such as the food industry has to continuously make efforts in the right direction and understand the ever changing trends of its consumers. Keeping up with changes might be difficult but early detection with the help of technology can bring out quick responses from the businesses.

Large scale businesses are already using data mining to deal with big data which gives them a competitive edge. In the era of emerging small business and start-ups, information about the market becomes critical to them. Customer

behaviours captured in corporate data are not random, but reflect the differing needs, preferences, propensities, and treatments of customers' (Hasan Ziafat, 2014). Knowledge about the customer choice can help a business to make and review its current prices and offers.

In our study, we have considered a small scale business. Small scale retail store or cafes can make use of data mining tools to perform some of the analysis relevant to their own business. These businesses can use data mining tools to gain knowledge that can potentially improve their operations and marketing efforts. This knowledge can be further used to expand their business.

Therefore, the business needs to analyse -What are the items that have the highest and lowest sales turnover? What is the feasible marketing mix that can be applied to the items? How can sales be increased for non-selling and introductory items.

II. LITERATURE REVIEW

A. DATA MINING

A Data mining is the process of discovering insightful, interesting, and novel patterns, as well as descriptive, understandable, and predictive models from large-scale data (Mohammed J. Zaki, 2014). Data mining can be regarded as an algorithmic approach that takes input data and yields patterns such as classification rules, association rules, or summaries as output (L Geng, 2006). The classification technique in data mining can divide customers into different segments. In customer segmentation, a common strategy is to use individual differences as the predictor of future behaviour (Spender). Targeted campaigns save money by directing offers only to those likely to want the product (Ian H. Witten, 2011). By studying trends in purchasing, loyal customers can be identified. In some cases identifying profitable customers beforehand and making targeted promotions towards them can

also build loyalty. Linoff and Berry describe loyalty as not an accident but ones accumulated knowledge of their customers taste and their price ranges which keeps them coming back (Gordon S. Linoff, 2011).

Predictive analysis uses statistical analysis as well as data mining techniques such as clustering, classification and segmentation as well as pattern detection (Loshin, 2012). In predictive analysis, data can be used to predict future patterns through forecasting and modelling. Algorithms that are made for prediction make predictive analysis for marketing possible.

According Omer and Levin, the rise in predictive marketing has been triggered by as customers now are demanding more personal and integrated approach through marketing channels (Omer Artun, 2015). Early adoption of this technology delivers an enormous value to its early adopters (Omer Artun, 2015).

Another use of data mining is to retain valuable customers or reduce customer churns. The article Predictive Accuracy of Customer Churn tests predictive models for customer churns and finds out the method which works the best(SCOTT A. NESLIN, 2006).

B. INTRODUCTION TO ASSOCIATION MINING

Association is one of the many techniques of data mining that establish co-relations between two or more different items in a data set. These relations are used to form patterns for prediction analysis. In the case of a retail store, if often denotes that if one item is purchased then there is a good chance that the other item is also purchased.

Association rule is an implication of the form $X \rightarrow Y$, where X and Y are nonintersecting sets of items (L Geng, 2006). These nine aspects like conciseness, coverage, reliability, peculiarity, diversity, novelty, surprisingness, utility, and actionability, determine whether a pattern is interesting (L Geng, 2006).

Market basket analysis is the general name for understanding product purchase patterns at the customer level (Linoff, 2007). Market basket analysis to identify which items are purchased together and, more importantly, whether the purchase of one item affects the probability of another item being purchased (Qualls, 2013). A retail store chain in the US placed the beer section beside the diapers when they observed that its young male customers who buy diapers also buy beer to increase sales.

C. APRIORI Algorithm

Apriori is an algorithm propounded by R. Agrawal and R. Srikant in 1994 for mining frequent item sets for Boolean association rules (Shailja N.UKE, 2014). An Apriori algorithm is an effective way to eliminate some of the candidate item sets without counting their support values (Tank, 2014). Apriori is the first association rule mining algorithm that pioneered the use of support based pruning to systematically grow the exponential growth of candidate item sets (Pang-Nin Tan, 2006).

Candidate Generation: Candidate item sets are generated based on frequent item sets after every iteration.

Candidate Pruning: It helps eliminate candidate item sets which have infrequent subsets or support.

Apriori algorithms works in the following steps-

- The algorithms goes through the data set and checks the support of each item.
- The algorithm iteratively generates new candidate k-item sets. Candidate generation is implemented here (Pang-Nin Tan, 2006).
- The algorithm eliminates all the items whose support is less than minimum support.

The total number of item sets necessary for the algorithm is k_{mx+1} , where k is the length of item sets and k_{mx} is the maximum size of frequent item sets.

Key terms in Association

Support is the number of times an item occurs in the total item set.

Confidence is the frequency of co-occurrence of associated item (M. S. Murali Dhar, 2013).

Confidence (LHS \rightarrow RHS) = Support of LHS and RHS/ Support of LHS

Lift compares confidence of a rule against expected confidence (DnI Consulting, 2014). So, a rule with higher value of Lift is the better (DnI Consulting, 2014).

Lift (LHS \rightarrow RHS) = confidence (LHS \rightarrow RHS)/ Support (RHS)

III. METHODOLOGY

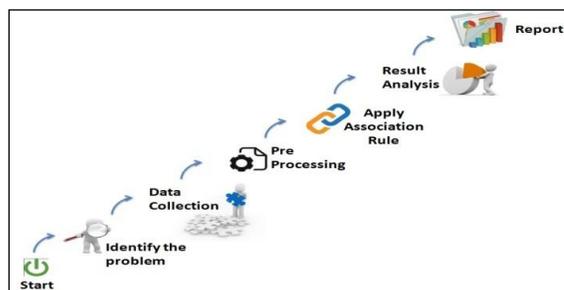


Fig1.steps for data analysis

A. Data Collection

The data is collected from a café called Shubham Coffee House in Model Colony, Pune. It is a business that aims to serve the students and working class in the neighbouring areas with everyday meals.

It is running as small business with an average of 100 transactions throughout the day. The sales data was collected as bills generated by the system. The recorded transactions with information about product purchased were entered in an Excel sheet. A total of 400 records were collected containing 12 items. Transactions which were made in an invoice are denoted with “T” (i.e. true) and the rest of non-incurring items in the invoice were denoted with “F” (i.e. false).

B. Data pre-processing

Data pre-processing deals with detecting and removing errors and inconsistencies, missing values from data in order to

improve the quality of data. After calculating the support of each item, the threshold value is set to 20. Items with low support were removed and 120 receipts with 9 items- pav bhaji, Batata wada, Dosa, Udid Wada, Pohe, Pav, Samosa, Upma and Tea were considered for data analysis.

C. Experiment

The collected data was analysed using open source data mining software, WEKA. Weka is popularly used for data mining. Its tools can perform various data mining techniques like pre-process, association, clustering, classification, regression and visualisation. To derive the association rules in the purchasing pattern of the customer Apriori algorithm is used with a support threshold for the best 'n' rules.

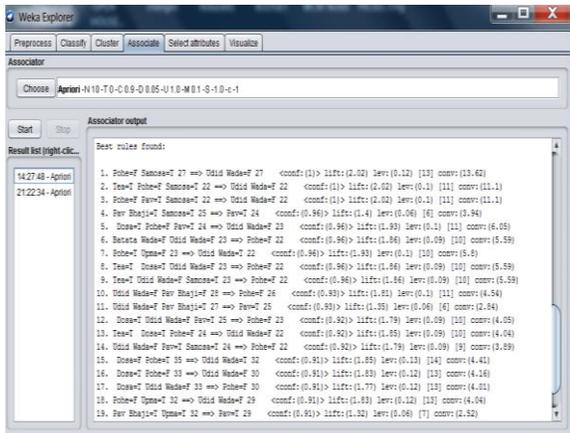


Fig.2 Result of association mining

We have found interesting patterns in customer choices in various combinations of food. The most popular and feasible combination for a single persons meal is Samosa and Tea. Another popular combination with Pav is Pav Bhaji and Batata wada. The two items that were found to complement the item tea are Samosa and Dosa. Confidence of 92% indicates that the most preferred combinations of South Indian dishes are Dosa and Udid wada.

Preferred meals combinations of two items consist of the combinations are,

- Pohe, Upma or Udid wada with a confidence of 96%.
- Dosa, Pav, Udid wada or Pohe with a confidence of 92%.
- Preferred meal combinations of three items consist of the combinations-
- Dosa, Pav and Pohe with a confidence of 92%.
- Pav Bhaji, Pav and Upma with a confidence of 91%.

Based on the results, if the cafe wishes to launch a new product it can be bundled with Pav or tea since they are the most preferred combinations. The products can be sold at discounted price till it becomes known among the customers. So while introducing dish in the menu, this factor can be taken into consideration and the new product can be clubbed with Pav or designed in such a way that it complements the dish and thus, increase its sales potential.

Our results points towards meal combinations that can be classified. The cafe can list out some popular meal

combinations and give promotional offers. Referring to these results, a price revision in the menu can also be made by the owner. The possible discount options can be-

- 1) Discount can be provided on Poha, if the customer purchases Udid Wada and thus, increasing its confidence level even more
- 2) Discount on PavBhaji and Pav on purchase of Upma

IV. CONCLUSION

After analysing the results it was observed that the lowest sold product is Samosa and the Highest sold product is Pav and Tea. So we propose combo offer in order to increase the sales of Samosa to bundle it together with Pav and Tea and sell it at discounted price.

The combinations in the output like Pohe with Udidwada or Upma with PavBhaji, is expected to be rare and is hardly noticed. In an overall view association rules in data mining have proved to be a resource that would solve one aspect of the owner’s dilemma. This study would help a general owner of a small scale enterprise to help them to refine their strategy and boost their business.

REFERENCES

1. Dnl Consulting. 2014. Market Basket Analysis : Step by Step using R.
2. Gordon S. Linoff, M. J. (2011). Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management. John Wiley and Sons.
3. Hasan Ziafat, M. S. (2014, September). Using Data Mining Techniques in Customer Segmentation. Int. Journal of Engineering Research and Applications.
4. Ian H. Witten, E. F. (2011). Data Mining : Practical Machine Learning Tools and Techniques. Elsevier.
5. L Geng, H. H. (2006, September). Interestingness Measures for Data Mining: A Survey. ACM Computing Surveys, 38.
6. Linoff, G. S. (2007). Data Analysis Using SQL and Excel. Wiley.
7. Loshin, D. (2012). Business Intelligence : The Savvy Managers Guide. Ed.: 2nd ed. Waltham, MA : Morgan Kaufmann.
8. Mlambo, N. (2016). Classification-Based Data Mining in Target Marketing. International Journal of Advanced Research in Computer Science and Management Studies.
9. MOHAMMED J. Zaki, W. M. (2014). Data Mining and Analysis. USA: Cambridge University Press.
10. M.S. Murali Dhar. (2013, December). A Survey on Algorithms for Market Basket Analysis. International Journal of Advanced Research in Computer Science and Management Studies.
11. Omer Artun, D. L. (2015). Predictive Marketing: Easy Ways Every Marketer Can Use Customer Analytics and Big Data. Wiley.
12. Pang-Nin Tan, M. S. (2006). Introduction to Data Mining.
13. PREDICTIVE ANALYTICS AND DATA MINING. (n.d.). Retrieved from Web Data Mining: <http://www.web-datamining.net/analytcs/>
14. Qualls, B. (2013). Introduction to Market Basket Analysis .
15. Rajagopal, D. S. (2011). CUSTOMER DATA CLUSTERING USING DATA. International Journal of Database Management Systems .
16. SCOTT A. NESLIN, S. G. (2006, April). Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models. Journal od Marketing Research.
17. Shailja N,UKE, P. S. (2014). Advanced Database Management. Pune: Nirali Prakashan.
18. Spendler, L. I. (n.d.). Data Mining and Management. New York, USA: Nova Science Publishers.Inc.
19. Tank, D. M. (2014). Improved Apriori Algorithm for Mining Association Rules. I.J. Information Technology and Computer Science.