PRINCIPAL COMPONENT ANALYSIS FOR INITIAL CUSTOMER IDENTIFICATION AT SMALL AND MEDIUM ENTERPRISES

Weerakoon W.M.H.G.T.C.K.[1*] and Rathnayake R.M.K.T.[1]

[1]Department of Physical Sciences and Technology, Faculty of Applied Sciences, Sabaragamuwa
Univeristy of Sri Lanka, Sri Lanka
*thariweera@gmail.com

Almost every industry, organization use Information Technology for their core business activities. Similarly, Small Medium Enterprises (SME(s)) try to use technologies for their businesses and this usage has increased data generation. Like other business organizations, SMEs keen on customer-oriented marketing and in order to foresee such a marketing strategy customer information needed to be gathered. Hence, adopting data mining and machine learning will increase the information gathering in terms of customers at SMEs. Information extracted from data mining in SMEs can be used to identify SMEs' customers and thereby, engage in target marketing campaigns for customer bases of SMEs. SME at the initial stage of data mining lacks high-pitched objectives due to inexperience with the concept. Since, at data repositories, there can be data under numerous attributes, which will lead to complex data mining analytical approaches at SMEs. In order to avoid that extreme course, Principle Component Analysis (PCA) can be practiced, so SMEs can identify which attributes significantly facilitate accurate and reliable initial customer identification. The cloth retailing dataset which comprises of demographic features; Gender, Age, Residency information, Marital Status and Occupation first subjugated to preprocessing and feature engineering. Then, PCA was carried out to investigate the most significant and prominent data features that needed to be focused on when performing customer identification. Once the PCA has done the total explained variance percentage for the first two dimensions are 10%. With the complexity of the data dimensionality, the extension of PCA, the Multi Component Analysis (MCA) was carried out, which improved the degree of customer identification in terms of demographic attributes. According to PCA, it has identified the Age and Occupation statuses were principle components and with MCA the Age group of 0-17 and Occupation category of 10 have the greater prominence with respect to customer deification. Considering the above facts, a classification of customers carried out with 4 definitive classes. With this initial, sound customer identification, SMEs can focus on effective target marketing based on customer segmentation and better customer relationship management.

Keywords: *Principle Component Analysis, Small Medium Enterprises, Data Mining, Customer Identification, Demographic Analysis*