## **Comparative Analysis of : PCA to other data mining techniques**

1. **Introduction:**

Market Basket Analysis (MBA), also referred to as association rule learning or affinity analysis, is a data mining approach commonly applied in fields such as education, nuclear science, bioinformatics, and marketing. This technique involves examining customer purchasing patterns to provide insights that help retailers make more informed business decisions (Kaur and Kang, 2016)**.** It is challenging for organizations to maintain a strong position in competitive markets, as this depends heavily on their decision-making abilities and understanding of customer behaviour (Raorane, Kulkarni and Jitkar, 2012). For this reason, studying customers' buying habits is vital for any organization's success (Gupta and Mamtora, 2014). Data mining aids in business to make correct business decisions and to understand the consumer behaviour (Berry and Linoff, 2004). The methodology used to discover the associations between different products the buyer has purchased is Association Rule Mining (Hand, Mannila & Smyth, 2001). Association is one of the best methods used. In this method a pattern is revealed based on the connection of an item and other items in a transaction. For instance, in MBA, it is used to detect the products that consumers often purchase together (Bharati & Ramageri, 2010). In order to understand consumer behavior, the data collection of past transactional knowledge is analyzed. Only the information about the transactions over a certain period, such as a day or a week etc., was available on personal computers back in time. Change in product scanning and bar code technology, however, enables data on transactions to be stored on a regular basis. As a result, vast volumes of data are obtained and processed. Due to restricted database functionality, such data sets are typically stored in higher-level storage. For example, to improve the functionality of the database and to process the data, if anyone buys a bottle of milk, they also prefer to buy a packet of oats at the same time. So, Milk => Oats. MBA is used in determining the location of goods within a store. If a client purchases a bottle of milk, he is much more likely to purchase a packet of oats. It would make consumers tempted to purchase one item with another to hold the milk and oats next to each other in a supermarket (Dubey et al., 2021)

.

1. **Background:**
2. **Literature Review:**
   1. **Machine Learning Models for Body Fat Prediction ........................................................................**
   2. **5 3.1.1 Non-Linear Machine Learning Models ................................................................................**
   3. **5 3.1.2 Hybrid and Ensemble Approaches .......................................................................................**
   4. **6 3.2 Comparative Analysis of Existing Studies ...................................................**

**.. 6 3.2.1 Strengths of Existing Models ...............................................................................**

**3.2.2 Limitations of Prior Studies .............................................................................. 7**

**3.3 Identification of Gaps ................................................................................................................... 7 3.4 This research’s novel Contributions................................................**

**3.1 Machine Learning Models for Body Fat Prediction ........................................................................ 5 3.1.1 Non-Linear Machine Learning Models ................................................................................ 5 3.1.2 Hybrid and Ensemble Approaches ....................................................................................... 6 3.2 Comparative Analysis of Existing Studies ..................................................................................... 6 3.2.1 Strengths of Existing Models ............................................................................................... 7 3.2.2 Limitations of Prior Studies ................................................................................................. 7 3.3 Identification of Gaps ................................................................................................................... 7 3.4 This research’s novel Contributions........................................................................................................ 8**

**3.1 Machine Learning Algorithms:**

**3.2 Apriori Algorithm:**

According to (Gupta and Mamtora, 2014). The intuition behind the apriori algorithm is described by,

* Only a large item set can be an item set if all its subsets are large item sets.
* It is possible to accept collections of products that have minimal support.
* From frequent item sets, association rules can be created.
  1. **Collaborative Filtering Algorithm for Product Recommendation:**

Collaborative filtering is a way of predicting (filtering) the tastes of a consumer automatically by collecting preferences or taste data from many users (collaborating). It suggests items by identifying other users with a similar preference; it uses their opinion to suggest items to the active user (Gong, 2010).

**3.5 Principal Component Analysis:**

Principal Component Analysis (PCA) is a well-known method in statistical multivariate data analysis, and it has been applied widely. It is mainly utilized to reduce the dimensionality of the dataset (Levin, Ken and Moody, 1994). The PCA was employed to project the data onto a transformed space that is invariant to certain linear transformations, to better measure the similarity or dissimilarity of two points. It has been implemented for information retrieval and is called latent semantic indexing (Aggarwal, 2001).

* 1. Research Question:

RQ1) How effective is Principal Component Analysis in performing Market Basket Analysis?

RQ2) What major challenges do retailers face when implementing Association rule mining techniques?

* 1. Search String
  2. Inclusion/Exclusion criteria:

1. Bibliography

**Kaur, M. and Kang, S. (2016).** Market basket analysis: Identify the changing trends of market data using association rule mining. Procedia Computer Science, 85, pp.78–85.

**Raorane, A.A., Kulkarni, R.V. and Jitkar, B.D. (2012).** Association rule – extracting knowledge using market basket analysis. Research Journal of Recent Sciences, [online] ISSN 2277-2502.

**Levin, A.U., Ken, T.K. and Moody, I.E. (1994).** Fast pruning using principal components. Advances in Neural Information Processing Systems, 6, pp.35–42.

**Aggarwal, C.C. (2001).** On the effects of dimensionality reduction on high dimensional similarity search. In: Proceedings of the ACM Symposium on Principles of Database Systems (PODS). ACM.

**Berry, M.J.A. & Linoff, G.S. (2004).** Data Mining Techniques. 2nd ed. Hoboken, NJ: Wiley

**Hand, D., Mannila, H. & Smyth, P. (2001).** Principles of Data Mining*. Cambridge: MIT Press.*

**Bharati, M. & Ramageri, M. (2010).** Data mining techniques and applications [online]. Available at: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.898.6657&rep=rep1&type=pdf [Accessed 16 Jun. 2025].

**Gong, S. (2010).** A collaborative filtering recommendation algorithm based on user clustering and item clustering. Journal of Software, 5.

**Dubey, S.K., Mittal, S., Chattani, S. & Shukla, V.K. ( 2021).** Comparative analysis of market basket analysis through data mining techniques. In: 2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), Amity University Dubai, UAE, 17–18 March 2021. IEEE. doi:10.1109/ICCIKE51210.2021.9410737.