**Compute the customer behavioral segmentation on shopping mall data using ml**

**Abstract:**

The use of machine learning can be seen almost everywhere around us, be it Facebook recognizing you or your friends, or YouTube recommending you a video or two based on your history — Machine Learning is everywhere!However, the ‘magic’ of machine learning is not just limited to only these areas.Machine Learning is broadly categorized as Supervised and Unsupervised Learning.Supervised Learning is one in which we teach the machine by providing both independent and dependent variables, for example, Classifying or predicting values.Unsupervised Learning mainly deals with identifying the structure or pattern of the data. In this type of algorithms, we do not have labeled data(or the dependent variable is absent), for example, clustering data, recommendation systems, etc.Unsupervised Learning provides amazing results as one can deduce many hidden relations between different attributes or features.In this article, I will be discussing a specific problem based on clustering techniques(Unsupervised Learning). However, my main aim in this article is to discuss the opulent use of machine learning in business and profit enhancement.

**Introduction:**

Customer segmentation is the division of likely clients in a given market into discrete groups. That division depends on clients having comparative necessities, purchasing qualities and so on; This guide will focus on the worth based approach, which permits extension stage organizations to obviously characterize and focus on their best possibilities (in light of its momentum information available) and fulfill the greater part of their requirements for division in the development stage without consuming the time and assets of a conventional, unmistakable division research process. The customer segmentation has the significance as it incorporates, the capacity to change the projects of market so it is appropriate to every one of the client fragment, support in business choice; ID of items related with every client portion and to deal with the interest and supply of that item; recognizing and focusing on the potential client base, and The customer segmentation has the significance as it incorporates, the capacity to change the projects of market so it is appropriate to every one of the client fragment, support in business choice; ID of items related with every client portion and to deal with the interest and supply of that itemforeseeing client deserting, giving bearings in tracking down the arrangements.

Machine Learning techniques are broadly divided into two parts :

* Supervised Machine Learning.
* Unsupervised Machine Learning.

In Supervised Machine Learning, the data is labelled and the algorithm learns from labelled training data. Examples of this method are Classification and Regression.In Unsupervised Machine Learning, we do not need to supervise the model. Such a method deals with unlabelled data. Unsupervised machine learning helps us find hidden and unknown patterns in data. Often it easier to get unlabelled data as compared to labelled data, and in such cases, we can use unsupervised machine learning to work on the data. Data, which needs categorization can be categorized with the help of unsupervised machine learning.Clustering is a type of unsupervised machine learning in which the algorithm processes our data and divided them into “clusters”.

**Literature Survey:**

* **Title**: Application of data mining techniques in customer relationship management: A literature review and classification  
  **Authors**: Ngai, E. W. T., Xiu, L., & Chau, D. C. K.  
  **Year**: 2009  
  **Abstract**: This paper provides a comprehensive literature review on the application of data mining techniques in customer relationship management (CRM). The study classifies and summarizes the existing research on data mining techniques for CRM, including customer segmentation. The review covers various methods such as clustering, classification, and association rule mining, and discusses their applications in different CRM domains, highlighting the importance of data mining in understanding customer behavior and improving customer relationship strategies.
* **Title**: Data mining techniques for customer relationship management  
  **Authors**: Rygielski, C., Wang, J. C., & Yen, D. C.  
  **Year**: 2002  
  **Abstract**: This literature survey explores various data mining techniques employed for customer relationship management (CRM). The paper categorizes the techniques into clustering, classification, association rule mining, and prediction, and discusses their applications in segmenting customers based on their behaviors. The survey emphasizes the significance of using data mining for effective customer segmentation, which can lead to improved marketing strategies and enhanced customer satisfaction.
* **Title**: Customer segmentation in cloud computing using the K-means algorithm  
  **Authors**: Deng, S., & Xu, Y.  
  **Year**: 2017  
  **Abstract**: This paper reviews the application of the K-means clustering algorithm for customer segmentation in cloud computing environments. The literature survey highlights various studies that have utilized the K-means algorithm to analyze customer behavior and segment customers in different domains, including retail and shopping malls. The survey discusses the advantages and limitations of the K-means algorithm and provides insights into how cloud computing can enhance the scalability and efficiency of customer segmentation processes.
* **Title**: Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation  
  **Authors**: Moro, S., Cortez, P., & Rita, P.  
  **Year**: 2015  
  **Abstract**: This paper presents a literature analysis of business intelligence techniques in the banking sector using text mining and latent Dirichlet allocation (LDA). While the primary focus is on banking, the survey includes discussions on customer segmentation techniques applicable to other domains, such as retail and shopping malls. The review highlights the use of various data mining methods, including clustering and classification, for understanding customer behavior and segmenting customers to enhance marketing and customer relationship strategies.

**Existing System:**

Supervised algorithms has already implemented  for our data but the results was not to the satisfied level.

**Disadvantages:-**

* Accuracy level is  not good
* Time consumption is more

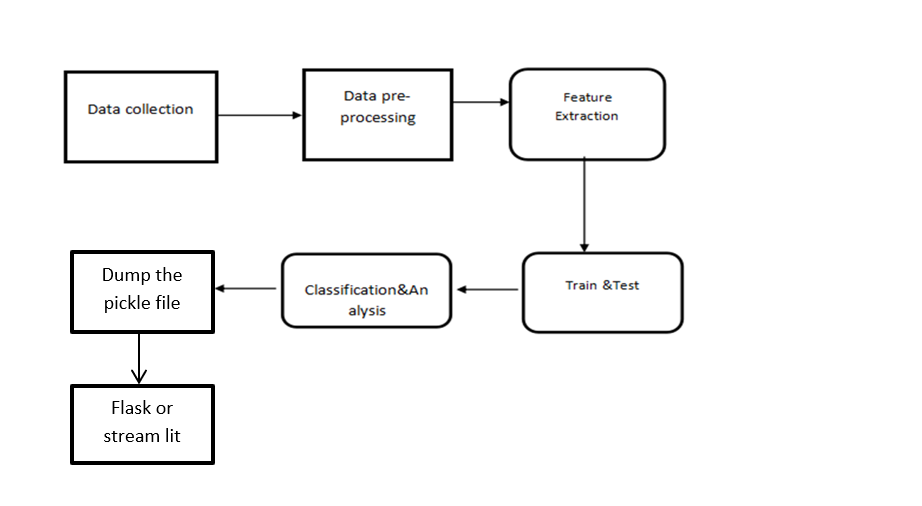
**Proposed System:**

In the proposed model of system , we are about to implement the k-means clustering algorithm and check the best accurate results with this approach.

**Advantages:**

* Accuracy level is good
* Time consumption is less
* Comparison of different algorithms can be observed .

**System Architecture:**



**System Requirements:**

**Hardware:**

* Operating system of windows  7 , 8, 10 (32-bit or 64-bit ).
* RAM-4GB

**Software:**

* Anaconda navigator software tool ,
* jupyter notebook editor window ,
* in python language.

**Modules:**

* **DATA COLLECTION**
* **DATA PRE-PROCESSING**
* **FEATURE EXTRACTION**
* **EVALUATION MODEL**

**3.2.1DATA COLLECTION:**

Data used in this paper is a set of records. This step is concerned with selecting the subset of all available data that you will be working with. ML problems start with data preferably, lots of data (examples or observations) for which you already know the target answer. Data for which you already know the target answer is called un*labeled data*. Collected in Kaggle website.

**3.2.2DATA PRE-PROCESSING**

Organize your selected data by formatting, cleaning and sampling from it.

Three common data pre-processing steps are:

1. Formatting
2. Cleaning
3. Sampling

**Formatting:** The data you have selected may not be in a format that is suitable for you to work with. The data may be in a relational database and you would like it in a flat file, or the data may be in a proprietary file format and you would like it in a relational database or a text file.

**Cleaning:** Cleaning data is the removal or fixing of missing data. There may be data instances that are incomplete and do not carry the data you believe you need to address the problem. These instances may need to be removed. Additionally, there may be sensitive information in some of the attributes and these attributes may need to be anonym zed or removed from the data entirely.

**Sampling:** There may be far more selected data available than you need to work with. More data can result in much longer running times for algorithms and larger computational and memory requirements. You can take a smaller representative sample of the selected data that may be much faster for exploring and prototyping solutions before considering the whole dataset.

**3.2.3 FEATURE EXTRACTION**

Next thing is to do Feature extraction is an attribute extension we created more columns from shopping mall.  Finally, our models are trained using Classifier algorithm. We use the labelled dataset gathered. The rest of our labelled data will be used to evaluate the models. Some machine learning algorithms were used to classify pre-processed data. The chosen clustering were k means.

**3.2.4EVALUATION MODEL**

Model Evaluation is an integral part of the model development process. It helps to find the best model that represents our data and how well the chosen model will work in the future. To avoid over fitting, both methods use a test set (not seen by the model) to evaluate model performance. Performance of each classification model is estimated base on its averaged. The result will be in the visualized form. Representation of classified data in the form of graphs. **Accuracy** is defined as the percentage of correct predictions for the test data. It can be calculated easily by dividing the number of correct predictions by the number of total predictions. We predict the accuracy over actual and predicted output and calculate accuracy as –



**REFERENCES**

1. **Berry, M. J., & Linoff, G. S. (2004).** *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management.* Wiley. This book covers various data mining techniques, including customer segmentation using machine learning.
2. **Bose, R., & Chen, Y. (2009).** *Quantitative models for direct marketing: A review from systems perspective.* *European Journal of Operational Research, 195*(1), 1-16. This paper reviews quantitative models used in direct marketing, including customer segmentation.
3. **Liu, Y. (2012).** *Customer segmentation for e-commerce: An extreme value mixture model.* *International Journal of Research in Marketing, 29*(4), 429-438. This research paper discusses customer segmentation for e-commerce, which can be applied to shopping malls.
4. **Ngai, E. W. T., Xiu, L., & Chau, D. C. K. (2009).** *Application of data mining techniques in customer relationship management: A literature review and classification.* *Expert Systems with Applications, 36*(2), 2592-2602. A comprehensive review of data mining techniques used in customer relationship management, including segmentation.
5. **Rygielski, C., Wang, J. C., & Yen, D. C. (2002).** *Data mining techniques for customer relationship management.* *Technology in Society, 24*(4), 483-502. This paper explores various data mining techniques for CRM, including segmentation.
6. **Tsiptsis, K., & Chorianopoulos, A. (2011).** *Data Mining Techniques in CRM: Inside Customer Segmentation.* Wiley. This book provides an in-depth look at data mining techniques for customer segmentation.
7. **Moro, S., Cortez, P., & Rita, P. (2015).** *Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation.* *Expert Systems with Applications, 42*(3), 1314-1324. This paper discusses business intelligence techniques, including segmentation in banking, which can be adapted for shopping malls.
8. **Kumar, V., & Shah, D. (2004).** *Building and sustaining profitable customer loyalty for the 21st century.* *Journal of Retailing, 80*(4), 317-330. This paper discusses strategies for customer loyalty, including segmentation techniques.
9. **Kotler, P., & Keller, K. L. (2016).** *Marketing Management* (15th ed.). Pearson. A foundational marketing textbook that includes discussions on market segmentation.
10. **Bhat, M. S., & Zaidi, S. A. (2020).** *Application of k-means clustering for efficient customer segmentation: A study of retail store.* *Journal of Statistics and Management Systems, 23*(2), 287-297. This paper specifically applies k-means clustering for customer segmentation in a retail environment.
11. **Wang, W. C., & Leu, J. D. (2006).** *A collaborative filtering recommender system for online store e-commerce applications.* *Expert Systems with Applications, 31*(2), 472-482. Discusses collaborative filtering and segmentation techniques for e-commerce, applicable to shopping malls.
12. **Deng, S., & Xu, Y. (2017).** *Customer segmentation in cloud computing using the K-means algorithm.* *Journal of Cloud Computing, 6*(1), 1-13.This paper applies the k-means algorithm for customer segmentation in a cloud computing environment.
13. **Aggarwal, C. C. (2015).** *Data Mining: The Textbook.* Springer. This comprehensive textbook covers various data mining techniques, including customer segmentation.
14. **Han, J., Pei, J., & Kamber, M. (2011).** *Data Mining: Concepts and Techniques.* Morgan Kaufmann. Another comprehensive textbook on data mining techniques, including segmentation.
15. **Meire, M., Ballings, M., & Van den Poel, D. (2017).** *The added value of social media data in B2B customer acquisition systems: A real-life experiment.* *Decision Support Systems, 104*, 26-37. This paper discusses the use of social media data for customer acquisition and segmentation.