**Customer segmentation on online retail dataset**

**Introduction:**

Customer segmentation is crucial for all businesses, as it helps understand and connect with customers effectively. Most traditional approaches fail to keep up with the evolving needs of the companies which limits personalization and optimization. These traditional approaches depended on demographic and geographic characteristics of customers failing to take into consideration behavioral characteristics which is not the case in current modern approaches. innovative solutions analyzing data are essential to derive actionable insights for better targeted marketing strategies.

**Objectives:**

* Develop a customer segmentation model using k-means clustering and other clustering algorithms that consider behavioral characteristics such as spending and income.
* Enhance customer understanding and identification by creating distinct customer segments based on their shared characteristics.
* Provide actionable insights and recommendations for marketing and business strategies based on the identified customer segments increasing profit and enhancing customer experience.

**Background & Literature Review:**

Previous research in customer segmentation has explored various techniques, including demographic, geographic, and psychographic approaches. However, these traditional methods often overlook the importance of behavioral characteristics in understanding customer preferences and purchasing behavior. The k-means clustering algorithm has emerged as a powerful tool for customer segmentation, as it enables the identification of distinct groups based on behavioral attributes.

**Methodology:**

**Data Collection and Understanding:**

* Identify the source of the data, in this case, An Online Retail Dataset containing one-year worth of clients' transactions.

**Data Cleaning and Preprocessing:**

* Handle missing values, outliers, and any other data quality issues.
* Transform and preprocess the data to make it suitable for analysis.

**Exploratory Data Analysis (EDA):**

* Perform descriptive statistics to gain initial insights into the dataset
* Visualize the data using charts, graphs, and plots to identify patterns, trends, and relationships between variables.

**Feature Engineering and Selection:**

* Derive new features from the existing dataset that could provide valuable insights or improve model performance.
* Select relevant features for the analysis based on domain knowledge, statistical significance, and feature importance techniques.

**Model Development:**

* Determine the appropriate clustering algorithms based on the problem and the nature of the data. This could include algorithms such as K-means clustering, spectral clustering, Gaussian mixture model, DBSCAN and others.
* Split the dataset into training and testing sets to assess model performance.
* Train and tune the selected models using appropriate algorithms and hyperparameter optimization techniques.

**Model Evaluation and Validation:**

* Validate the trained models using the testing dataset to assess their performance and generalization capabilities.
* Compare and select the best-performing model based on evaluation metrics and business requirements.

**Results Communication:**

* Visualize and present the results using appropriate charts, graphs, and reports.
* Provide actionable insights and recommendations to stakeholders based on the analysis.

**References:**

[1] Alam, M. F. “Customer Segmentation Using K-Means Clustering in Unsupervised Machine Learning.” 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N) /, 2021, pp. 94–98, https://doi-org.proxy.bib.uottawa.ca/10.1109/ICAC3N53548.2021.9725644.

[2] Regmi, S. R. “Customer Market Segmentation Using Machine Learning Algorithm.” 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), 2022, pp. 1348–54, https://doi-org.proxy.bib.uottawa.ca/10.1109/ICOEI53556.2022.9777146.

[3] Patankar, N. “Customer Segmentation Using Machine Learning.” Advances in Parallel Computing., vol. 39, 2021, pp. 239–44, https://doi-org.proxy.bib.uottawa.ca/10.3233/APC210200.

[4] Maheswari, T. L. “Customer Segmentation Based on Sentimental Analysis.” 2022 International Conference on Advanced Computing Technologies and Applications (ICACTA), IEEE, 2022, https://doi-org.proxy.bib.uottawa.ca/10.1109/ICACTA54488.2022.9753207.

[5] Aktas, A. A. “Comparative Unsupervised Clustering Approaches for Customer Segmentation.” 2021 2nd International Conference on Computing and Data Science (CDS), 2021, pp. 530–35, https://doi-org.proxy.bib.uottawa.ca/10.1109/CDS52072.2021.00097.