

PROJECT TITLE:

CUSTOMER SEGMENTATION USING DATA SCIENCE

PROBLEM DEFINITION:

Customer segmentation is the process of dividing a company's customer base into distinct groups or segments based on shared characteristics or behaviors. These characteristics can include demographic data (age, gender, location), psychographic data (interests, values, lifestyles), and behavioral data (purchase history, online activity).

Data science plays a crucial role in customer segmentation by leveraging advanced analytical techniques and algorithms to identify meaningful patterns and segments within a large dataset.

In summary, customer segmentation using data science is a powerful approach to tailor marketing efforts and enhance customer experiences. However, it involves various challenges related to data quality, algorithm selection, interpretation, and ethical considerations that need to be carefully managed for successful implementation.

DESIGN THINKING:

Data Collection: Collect customer data, including attributes like purchase history, demographic information, and interaction behavior.

Data Preprocessing: Clean and preprocess the data, handle missing values, and convert categorical features into numerical representations.

Feature Engineering: Create additional features that capture customer behavior and preferences, such as total spending, frequency of purchases, etc.

Clustering Algorithms: Apply clustering algorithms like K-Means, DBSCAN, or hierarchical clustering to segment customers.

Visualization: Visualize the customer segments using techniques like scatter plots, bar charts, and heatmaps.

Interpretation: Analyze and interpret the characteristics of each customer segment to derive actionable insights for marketing strategies

Proposed system:

- . The goal is to assign each customer to one group, then you should reach for a K Means algorithm.
- . It is a popular method of unsupervised machine learning method that finds “different” clusters and group them together so you end up with the most possible customer segments to interpret.

PRE-PROCESSING:

STEPS:

- DATA CLEANING
- HANDLE MISSING VALUES
- CATEGORICAL TO NUMERICAL REPRESENTATIONS.

DATA CLEANING:

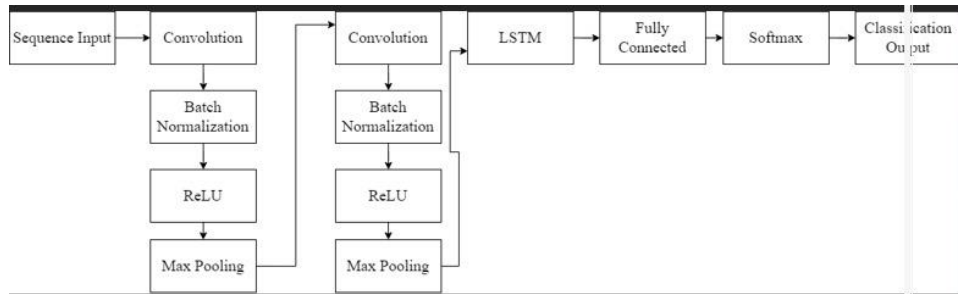
Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. When combining multiple data sources, there are many opportunities for data to be duplicated or mislabeled.

HANDLE MISSING VALUES:

- Deleting Rows with missing values
- Impute missing values for continuous variable
- Impute missing values for categorical variable
- Other Imputation Methods
- Using Algorithms that support missing values
- Prediction of missing values
- Imputation using Deep Learning Library — Datawig.

CATEGORICAL TO NUMERICAL REPRESENTATIONS:

- cat.codes Attribute
- replace
- Label Encoder



ALGORITHM:

- DEEP LEARNING TECHNIQUE : PCA
- ATTENTION MECHANISMS

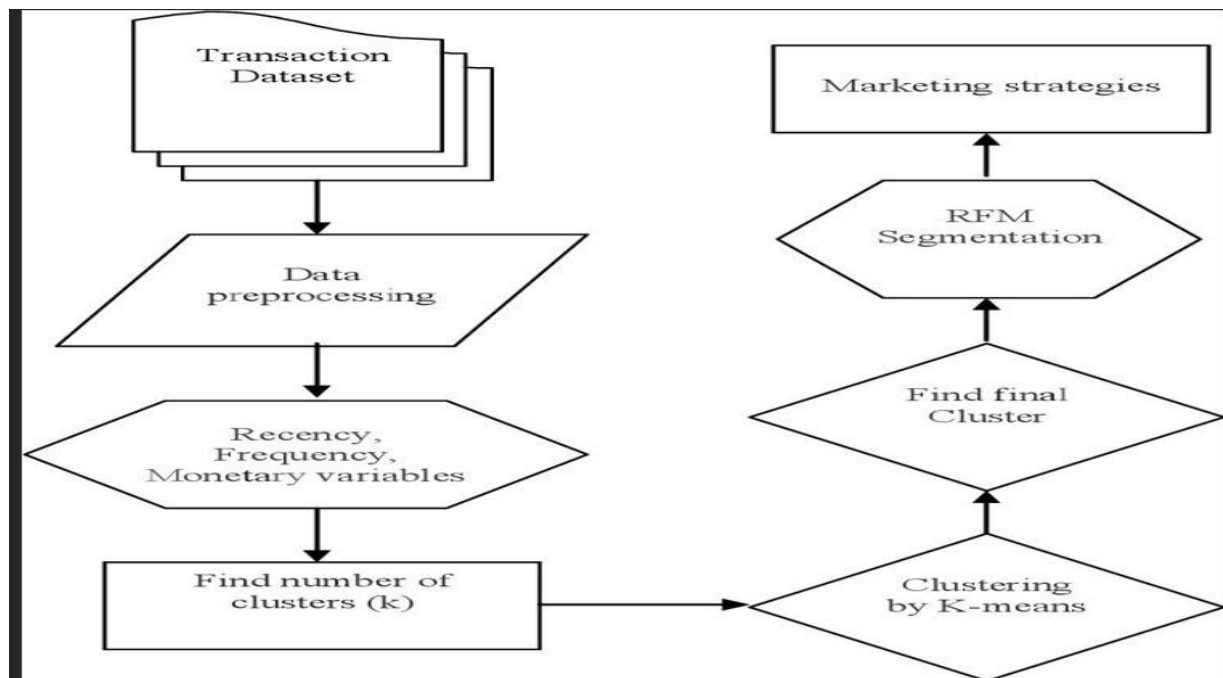
PCA :

Using unsupervised learning techniques of Principal Component Analysis (PCA) and KMeans to identify customer segments of the German population that were popular or less popular with a mail-order sales in Germany.

ATTENTION MECHANISM:

An attention mechanism is an Encoder-Decoder kind of neural network architecture that allows the model to focus on specific sections of the input while executing a task. It dynamically assigns weights to different elements in the input, indicating their relative importance or relevance.

PROJECT WORKFLOW



Conclusion

In conclusion, this project has successfully applied the K-means algorithm to perform customer segmentation on a dataset from a mall. Through this analysis, we have gained valuable insights into the distinct groups of customers within the mall's customer base. This segmentation enables us to understand the unique characteristics, behaviors, and preferences of each group, facilitating more targeted marketing strategies and personalized services.

By dividing customers into clusters, we have the potential to enhance the mall's operational efficiency and customer satisfaction. This project has highlighted the power of data science and machine learning techniques in transforming raw data into actionable insights that can drive business decisions and improve overall customer experience.

As businesses strive to adapt to changing market dynamics and increasing competition, customer segmentation remains a critical tool for staying relevant and competitive. The results and findings from this project not only contribute to a deeper understanding of the mall's customers but also pave the way for more effective marketing campaigns, improved customer retention, and ultimately, increased success for the mall in the ever-evolving retail landscape.

REFERENCES:

- Customer Segmentation and Profiling For Life Insurance Using K-Modes Clustering And Decision Tree Classifier. Shuzlina Abudl Rahman, Nurin Faiqah Kamal Arifin, Mastura Hanafiah Sofianita Mutalib

<https://www.researchgate.net/publication/355086360> Customer Segmentation and Profiling for Life Insurance using K-Modes Clustering_and_Decision Tree Classifier

- Customer Segmentation Using RFM Model and K-Means Clustering. Rahul Shirole. Laxmiputra Salokhe Saraswati Jadhav <https://www.researchgate.net/publication/352393770> Customer Segmentation using RFM_ Model and K-Means Clustering

- Customer Segmentation Analysis and Customer Lifetime Value Predictions Using Pareto /NBD and Gamma - Gamma Model . Kim-Giao Tran . Van-Ho Nguyen, Thanh Ho <https://www.researchgate.net/publication/359864311> Customer segmentation_analysis and customer_lifetime_value_prediction using ParetoNBD_and_Gamma-Gamma_model

- Customer Segmentation Analysis for Improving Sales Using Clustering. Pilli Sri Durga, J.A.Paulson,MarriSrinivasareddy <https://www.researchgate.net/publication/373513641> Customer segmentation_analysis_for_improving sales using clustering