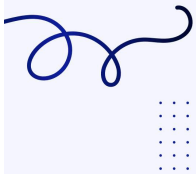
Cognitive Systems

Review-1
CSE 4059
Dr. Nivedita M



Title

- Customer behavior prediction

Team Members

Aaryan Mehta
20BAI1108

Sarvesh Chandak
20BAI1221

Aryan Singh Kushwaha
20BAI288

Absract

Customer retention refers to a company's ability to turn customers into repeat buyers and prevent them from switching to a competitor. It indicates whether your product and the quality of your service please your existing customers. So our project is about customer behavior retention and prediction where we aim to develop a algorithm which any business can use to increase the retention rate of their customers and increase the sales of the company. Our algorithm will predict the behavior of the customer which the business can further use to change their strategy and increase the retention rate. We intend to do basic exploratory analysis, followed by clustering customers into groups using Kmeans clustering, and analyzing some of those groups using various metrics. Finally we will apply Recurrent Neural Network to predict sales using Keras.

Introduction

The algorithm which we will be using is Simple RNN.
Recurrent Neural Network(RNN) is a type of Neural Network where the output from the previous step are fed as input to the current step.
RNN were created because there were a few issues in the feed-forward neural network like:

- **Cannot handle sequential data**
- **Considers only the current input**
- **Cannot memorize previous inputs**

The solution to these issues is the RNN. An RNN can handle sequential data, accepting the current input data, and previously received inputs. RNNs can memorize previous inputs due to their internal memory. Our algorithm will be able to predict R,F and M value for any business:
1. Recency (R) - How many days since customer's last purchase until the present
2. Frequency (F) - How many purchases the customer has done since their start of the time period
3. Monetary Value (M) - Measures how much the customer has spent since the start of the time period

Objective

1.To understand the purchasing behaviour pattern of customers in a better way
2. To help marketers segment the customers and target them based on their behaviour
3. To help businesses retain customers
4. To calculate RFM values.

Problem Statement

- Businesses need to classify the customers so as to segment them which would generate high revenues for the businesses in the future.
- Such customers should be retained based on the recency of their purchase, monetary value generated them and their frequency of purchase.
- This problem is usually faced by small businesses as they have less capital compared to well established MNCs.

Literature Survey

The Prediction of Consumer Behavior from Social Media Activities

The knowledge in the study was gained through the analysis of the relationship between social media activities and consumer behavior. The study used a dataset containing information on consumer behavior and social media activities, and applied machine learning algorithms, including decision trees and artificial neural networks, to make predictions about consumer behavior. The study found that social media activities are a significant predictor of consumer behavior, and that the combination of decision trees and artificial neural networks performed the best in predicting consumer behavior. The study also found that consumer behavior is influenced by several factors, including consumer demographics, consumer attitudes, and consumer preferences. These findings provide insights into the relationship between social media activities and consumer behavior, and how this relationship can be used to make predictions about consumer behavior.

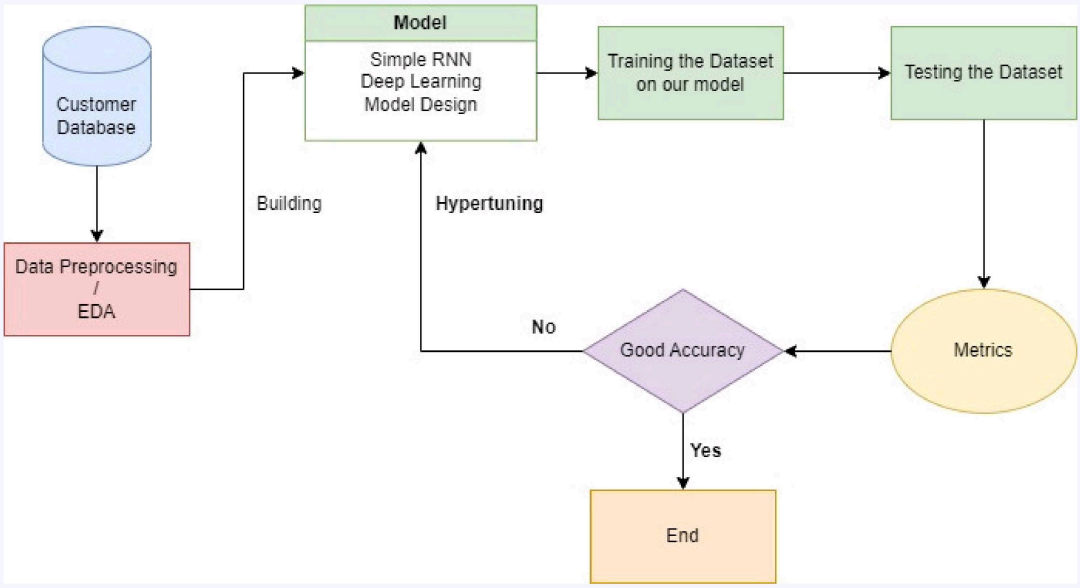
"Prediction of Customer Behavior using Machine Learning: A Case Study"

The knowledge in the study was gained through a combination of data analysis, experimentation, and evaluation. The authors used a dataset of customer transactions and demographic information from a Vietnamese bank to train and test machine learning algorithms for predicting customer behavior. They compared the performance of different algorithms, such as decision trees, random forests, and support vector machines, and evaluated their performance using metrics such as accuracy, precision, recall, and F1 score. The authors also performed sensitivity analyses to identify the most important variables in predicting customer behavior, and discussed the implications of their findings for improving customer service and decision-making in banks.

Prediction of customer engagement behaviour response to marketing posts based on machine learning

The paper presents a study that aims to predict customer engagement behavior towards marketing posts using machine learning methods. The authors collected data from a fashion brand's WeChat official account to form the dataset for their study. They applied various machine learning algorithms, including Logistic Regression, Random Forest, and Gradient Boosting, to predict customer engagement, which was defined as the number of likes, comments, and reposts on the marketing posts. The results showed that the Gradient Boosting algorithm performed the best, with an accuracy rate of 72.71%. The study highlights the potential of machine learning in predicting customer engagement and provides insights for fashion brands to improve their marketing strategies on social media platforms.

Architecture Diagram



Modules Explanation

1.Data Collection and Preprocessing:
This module involves collection of dataset(Ta Feng Dataset) and preprocessing the dataset which we will use for training and testing. This may include tasks such as remocing null values, cleaning data and removimg outliers and unwanted values.

2. Deep learning Model Design:
In this module, we are using RNN as the primary Deep Learning Model. This may include selecting the appropriate layers, number of filters and kernel size, optimizer, and loss function for the model.

1.

Modules Explanation

3. Training and Validation:
Our Deep Learning model will be trained on the preprocessed dataset and fine-tuned to improve its performance. This will be done by splitting the data into training and testing, and training the model on the training set.

4. Testing and Evaluation:
The trained models will be evaluated using the validation dataset and the test dataset. This will be done to check the accuracy, precision, recall, and other evaluation metric of the model on unseen data.

5. Prediction:
We will be predicting RFM values using the model we have created.

Dataset

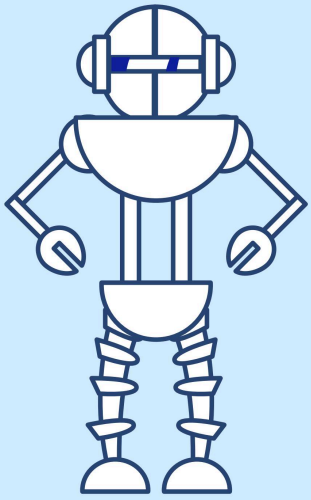
The Ta Feng Dataset is a Supermarket Dataset containing 817741 transactions from November 2000 until the end of February 2001. The dataset contains information about 119578 shopping baskets, belonging to 32266 users, where 1129939 items were purchased from a range of 23812 products.

Link of the dataset:

https://www.kaggle.com/datasets/chiranjivdas09/ta-feng-grocery-dataset

Tools used

1. Simple RNN and other ML/DL algorithms
2. Tensorflow
3. Sklearn
4. Pandas



5. L1 Regularization
6. Relu Activation
7. Hidden Layers
8. MSE Loss

Evaluation Metrics

For evaluation of the result, we will measure the count of active customers who purchased some product each week. We will also measure average traction price as well as unit price. Our main focus will be on measuring RFM (Recency, Frequency, Monetary value) metrics.

- **Recency (R) - Average number of days since the customer's last purchase until present**
- **Frequency (F) - Average number of purchases of customer in a given time period**
- **Monetary value (M) - Average spending of the customer in a given time period**

Based on this, we will segment the customers in different categories, for example, gold, silver and bronze level customers, with the help of any suitable un-supervised ML algorithm.

References

https://www.iise.org/uploadedfiles/iie/community/technical_societies_and_divisions/sems/abstract_909.pdf

https://www.mdpi.com/2076-328X/12/8/284/pdf

https://ceur-ws.org/Vol-3026/paper18.pdf

https://iopscience.iop.org/article/10.1088/1742-6596/1299/1/012041


Thank You