Customer Segmentation Using Machine Learning

Hibba Mehmood (w17bscs002) Mishaal Arbab (w17bscs007)

Data Science With Python Course Project 2020

Abstract

This project aims to identify the potential customer base for selling the product. So, by predicting the correct class of the customer using various machine learning algorithms such as neural network, SVM, Gaussian Naïve Bayes, Decision Trees, Random Forests, K- Nearest Neighbor, and logistic regression, we would achieve that. Our main goal is to identify the customer segment/class in order to sell the product.

Introduction

Whenever you need to find your best customer, customer segmentation is the ideal methodology. “Customer Segmentation is the process of division of customer into several groups or categories that share a similarity in different ways that are relevant to marketing such as gender, age, interests, and miscellaneous spending habits”.

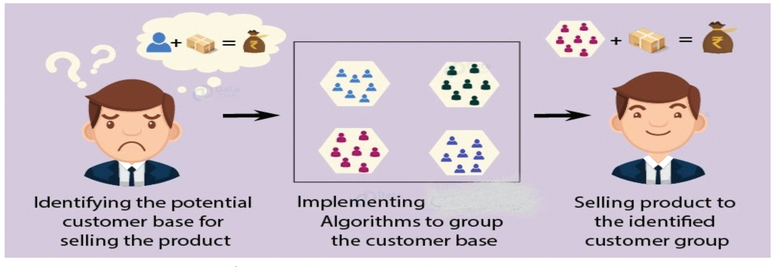
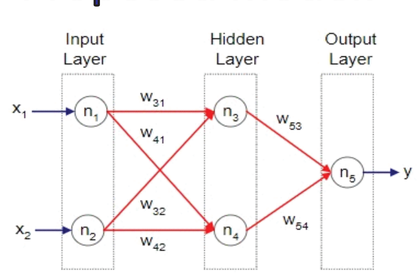


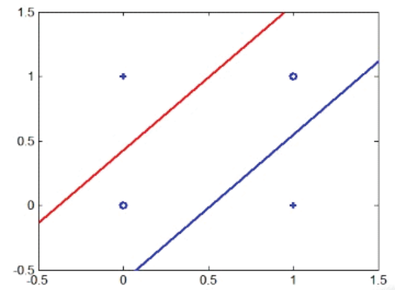
Fig. 1. Customer Segmentation

Problem Statement

Companies aim to gain a deeper approach of the customer they are targeting. Every customer has its own requirements and preferences therefore marketing strategies accompanied with machine learning are designed in such a way that would cater the customer of every segment and thus products are designed and develop according to the targeted audience. In this way companies would reap maximum profit in every domain. Many machine learning algorithms have been applied to effectively forecast the class of the customer. We propose to apply machine learning techniques i.e., Neural Network, SVM, Gaussian Naïve Bayes, Decision Trees, Random Forests, Multi-Layer Perceptron Classification, K- Nearest Neighbor, and logistic regression. It is crucial to correctly predict the class of the customer in order to maximize profit.

Proposed model:

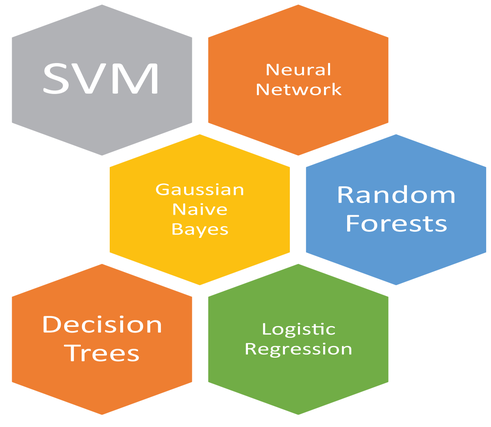




Methedology and Results

In this project, we went through the customer segmentation model. We developed this using a type of machine learning known as supervised learning. Specifically, we made use of neural network algorithm. We also trained our model through other machine learning algorithms. We analyzed and visualized the data and then proceed to implement the algorithm.

Techniques:



We are implementing algorithms to group the customer base. First we use NN and to increase the accuracy we play around with the parameters to get the maximum accuracy. Furthermore, we use decision trees, random forest, naïve bayes etc. that are giving maximum accuracy and performed well.

Activation Function	Hidden layers	solver	Accuracy
relu	100	**sgd**	92%
relu	100	**lbfgs**	80%
tanh	100	adam	96%
logistic	100	adam	93%
identity	100	adam	64%
relu	**800**	adam	94%
relu	**1500**	adam	72%
relu	**70**	adam	71%
relu	**5**	adam	34%

Fig. 2. NN(Neural Network) with Variations:

Main Results:

	Solver	Activation Function	Hidden Layers	Accuracy
Best Fit	lbfgs	tanh	500	100%

Algorithms	F1-Score	Recall	Precision	Accuracy
SVM	0.97	0.96	0.97	0.96
K-Nearest Neighbors	0.96	0.96	0.97	0.96
Gaussian Naïve Bayes	0.99	0.99	1.0	0.99
Decision Trees	**1.0**	**1.0**	**1.0**	**1.0**
Random Forests	**1.0**	**1.0**	**1.0**	**1.0**
Logistic Regression	0.95	0.94	0.95	0.94
Neural Network	**1.0**	**1.0**	**1.0**	**1.0**

Fig. 3 Comparison with other ML Algorithms:

Conclusion:

After training our model we calculated results and recorded accuracy and thus conclude that our model is best fitted for the given problem with the accuracy of 100% and is trained well. Our goal that is selling product to the identified customer segment/category is very much achieved.

Refrences:

[1]https://www.kaggle.com/vjchoudhary7/customer-segmentation-tutorial-in-python
[2]https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html
[3] https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html
[4] https://data-flair.training/blogs/r-data-science-project-customer-segmentation/
[5] https://www.datacamp.com/community/tutorials/introduction-customer-segmentation-python