2022

IOMP Project Report on

Customer Segmentation Using IBM Watson Machine Learning

SUBMITTED BY

SWATHI K	**19UK1A05H8**
RAMYASRI G	**19UK1A05J2**
MOHAMMED ALIUDDEN HYDER	**19UK1A05J9**
DONESH REDDY P	**19UK1A05G5**
ROHITH P	**19UK1A05H9**

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VAAGDEVI ENGINEERING COLLEGE

(Affiliated to JNTUH, Hyderabad)

Bollikunta, Warangal _ 506005

2019_ 2023

CONTENTS

CHAPTERS PAGE NO'S

CHAPTER 1 INTRODUCTION

OVERVIEW

In today's highly competitive world, the primal aim of any business is to grab potential customers who can generate profits for the organization. With increasing the number of organizations in the market, companies want to gain a competitive advantage over others. The primal task of Management is to identify potential customers from the rest. This will be simplified with the help of Machine Learning models to classify the customers into segments based on various attributes. The intervention of Data Science and AI helps the business to build such models to analyze the customers and their products in better decision making, to

improvise the business process, to formulate better strategies, and to improve the revenue. This project deals with understanding and segmenting the customers based on the data

1.2 PURPOSE

The goal of customer segmentation is to help you tailor your marketing techniques to meet the specific needs of each consumer group. And through this form of marketing, you get to interact with your clients more effectively. So, here's why customer segmentation offers effective interaction with your clients.

CHAPTER 2 LITERATURE SURVEY

2.1 EXISTING PROBLEM

A solution is proposed as distinguish the customers group into two groups named as premium and standard with the help of machine learning methods named as NEM, LiRM and LoRM . Tushar Kansal, Suraj Bahuguna, Vishal Singh, Tanupriya Choudhury. "Customer Segmentation using Kmeans Clustering", International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS).2018, In this paper customer segmentation on Telecom customers is achieved by using information such as age, interest, etc.
with the help of cluster analysis method

2.2 PURPOSED SOLUTION

The Model we built will be able to classify the customer's potentiality in purchasing power. We will be using classification algorithms such as H-clustering, k-means clustering Decision tree, Random forest, KNN, and xgboost. We will train and test the data with these algorithms. From this best model is selected and saved in pkl format. Once the model is saved, we integrate it with the flask application and also deploy the model in IBM.

The method or solution is Jupiter notebook and spyder we used to complete this project. and you will use this jupiter notebook for you recommended.

To build Machine learning models you must require the following packages

- **Numpy**- It is an open-source numerical Python library. It contains a multidimensional array and matrix data structures. It can be used to perform

mathematical operations on arrays such as trigonometric, statistical, and algebraic routines.

- **Pandas**- It is a fast, powerful, flexible, and easy-to-use open-source data analysis and manipulation tool, built on top of the Python programming language.
- **Seaborn**- Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.
- **Matplotlib**- Visualisation with python. It is a comprehensive library for creating static, animated, and interactive visualizations in Python
- **Sklearn** – which contains all the modules required for model building
- **Scipy** – which contains all the modules required for scientific and computing functions

CHAPTER 3 THEORETICAL ANALYSIS

Customer segmentation is the practice of dividing a company's customers into groups that reflect similarity among customers in each group. The goal of segmenting customers is to decide how to relate to customers in each segment in order to maximize the value of each customer to the business., we gone through lot of algorithms like K-means clustering, Hierarchical Clustering, Density Based Clustering, Affinity Propagation Algorithm , Customer segmentation has the potential to allow marketers to address each customer in the most effective way. Using the large amount of data available on customers (and potential customers), a customer segmentation analysis allows marketers to identify discrete groups of customers with a high degree of accuracy based on demographic, behavioral and other indicators.

.

3.1 BLOCK DIAGRAM

K-means clustering algorithm is one of the clustering algorithms based on division. It adopts a heuristic iterative process to re-divide data objects and re-update cluster centres. The basic idea of the algorithm is: suppose a set with element objects and the number of clusters to be generated. In the first round, a sample element is randomly selected as the initial cluster centre, and the distance between other sample elements and the centre point is analysed the clusters are respectively divided according to the distance. In each of the following rounds, the iterative operation of the above steps is continuously performed, and the average value of the element objects obtained this time is taken as the centre point of the next round of clustering until the condition that the clustering centre point no longer changes in the iteration process is met.

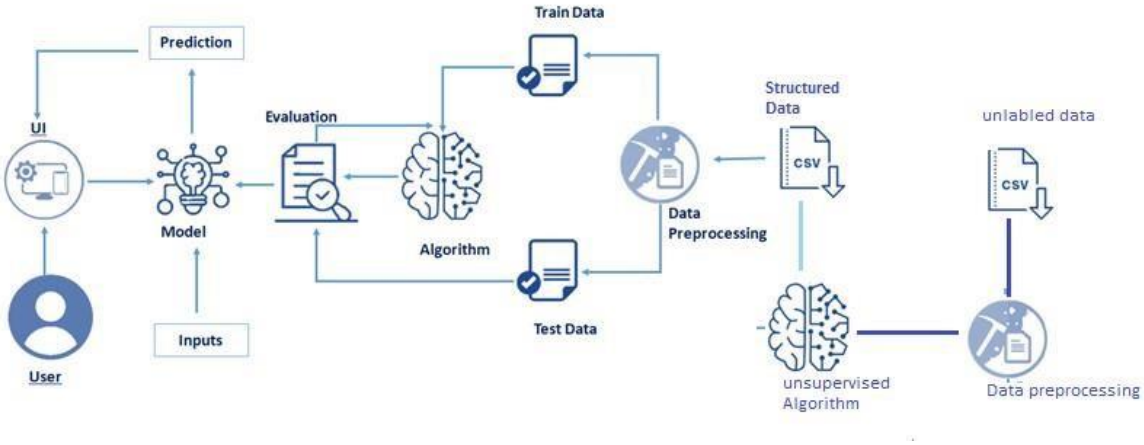

Fig 3.1. Block diagram

3.2 HARDWARE / SOFTWARE DESIGNING

The hardware required for the development of this project is:

Processor : Intel CoreTM i5-9300H
Processor speed : 2.4GHz
RAM Size : 8 GB DDR
System Type : X64-based processor

SOFTWARE DESIGNING:

The software required for the development of this project is:

Desktop GUI : Anaconda Navigator
Operating system : Windows 11
Front end : HTML, CSS
Programming : PYTHON

CHAPTER 4 EXPERIMENTAL INVESTIGATION

IMPORTING AND READING THE DATASET

Importing the Libraries

First step is usually importing the libraries that will be needed in the program.

Pandas: It is a python library mainly used for data manipulation.

NumPy: This python library is used for numerical analysis.

Matplotlib and Seaborn: Both are the data visualization library used for plotting graph which will help us for understanding the data. **csr_matrix() :**A dense matrix stored in a NumPy array

can be converted into a sparse matrix using the CSR representation by calling the csr_matrix() function.
Train_test_split: used for splitting data arrays into training data and for testing data.
Pickle: to serialize your machine learning algorithms and save the serialized format to a file.

Reading the Dataset

For this project, we make use of data set 'H-1B Visa Petitions 2011-2016 dataset'We will be selecting the important features from the dataset that will help us in predicting the h1b visa approvalThe next step is to read the dataset into a data structure that's compatible with pandas. Let's load a .csv data file into pandas. There is a function for it, called **read_csv().**

We will need to locate the directory of the CSV file at first (it's more efficient to keep the dataset in the same directory as your program).If the dataset in same directory of your program, you can directly read it, without any path. After the next Steps we made following bellow:

1.Data visualization
2.Collabrative and filtering
3.Creating the Model
4.Test and save the model
5.Buil Python Code
6.Build HTML Code
7.Run the Application
We are the following above sections we did and investigate it.

CHAPTER 5 VENN DIAGRAM

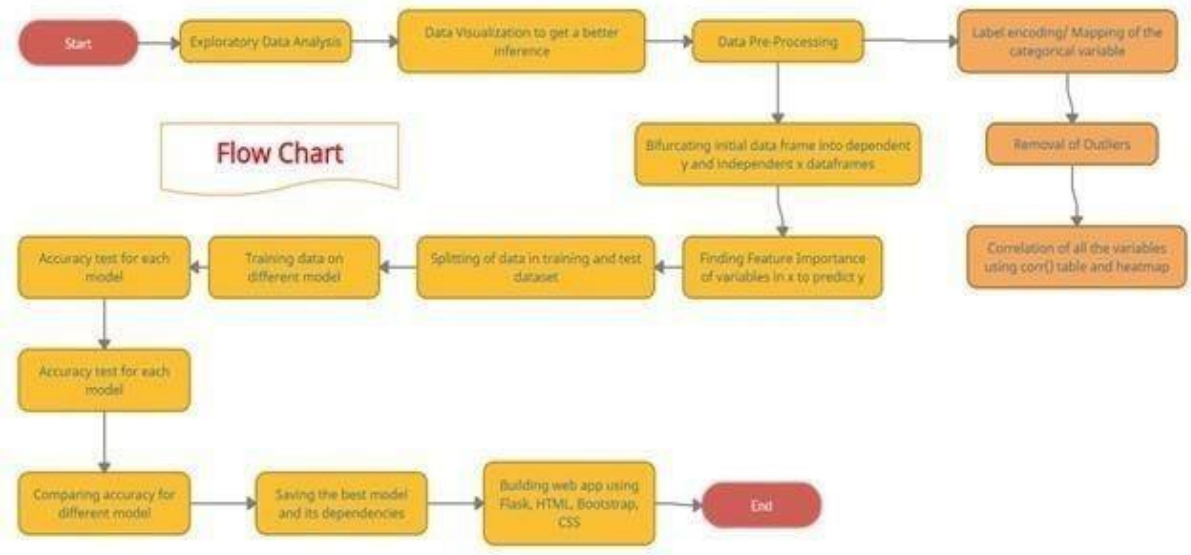

Fig 5.1 Venn diagram of the project

Project Flow:

- User interacts with the UI (User Interface) to upload the input features.
- Uploaded features/input is analyzed by the model which is integrated
- Once model analyses the uploaded inputs, the prediction is showcased on the UI.

1. Data Collection.

ML depends heavily on data, without data, a machine can't learn. It is the most crucial aspect that makes algorithm training possible. In Machine Learning projects, we need a training data set. It is the actual data set used to train the model for performing various actions. You can collect datasets from different open sources like kaggle.com, data.gov; UCI machine learning repository etc. The dataset used for this project was obtained from Kaggle.

2. Data Pre- processing.

Data Pre-processing includes the following main tasks Importing the libraries.

- Importing the data
- Analyse the data.
- Taking care of Missing Data.
- Data Visualisation.
- Splitting Data into Train and Test

3. Model Building

1.Unsupervised model Building

- Import the model building libra.
- Initialising the model.
- Fit and predict the clusters.
- Add the classes to the main data set and solve the dataset.
- Splitting x and y.
- Splitting train and test data.

2.Supervised model Building

Model building includes the folloeing main tasks

- Import the model building Libraries.
- Intializing and testing the model.
- Training and testing the model.
- Evaluation of Model.
- Save the Model

4.Application Building

In this section ,we will be building a web application that is integrated to the model we built. A UI is provided for the users where he has to enter the values for predictions. The enter values are given to the save model and prediction is showcased on the UI.

This section has the following tasks

- Build HTML Pages
- Building server-side script

CHAPTER 6 RESULT

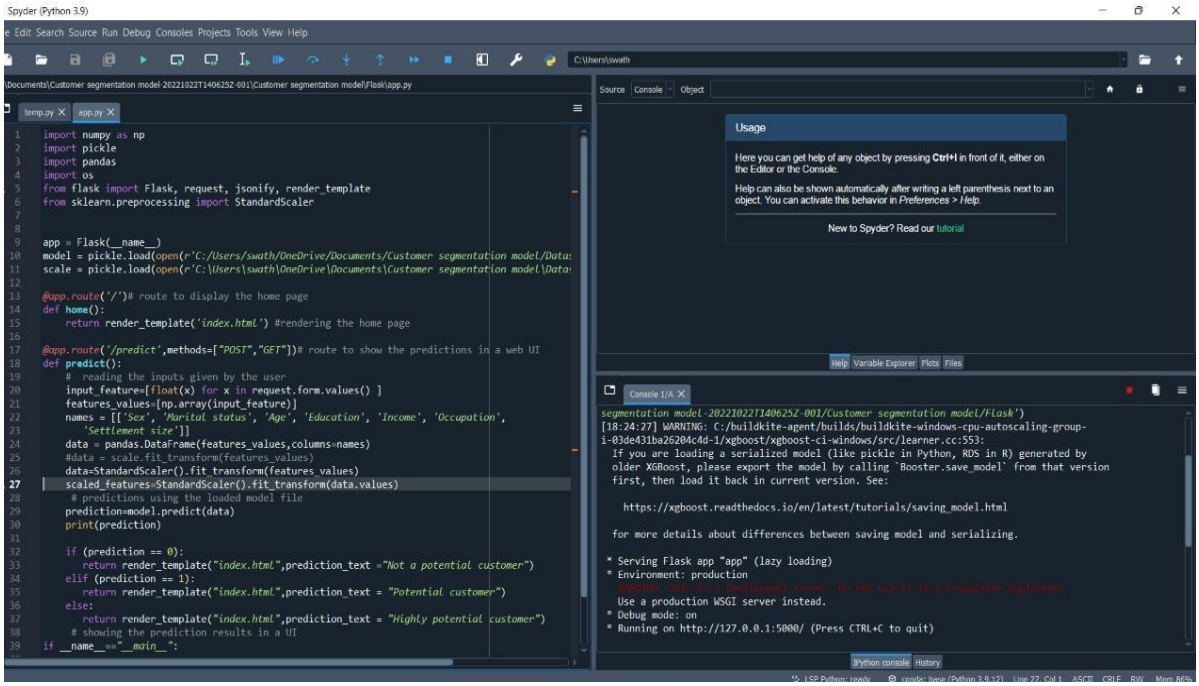

Fig 6.1 Flask App Code with Output Page

Fig 6.2.Home Page of customer segmentation

Fig.6.4 If the prediction belongs to class-2, it means that the customer is highly potential

Fig.6.5.If the prediction belongs to class-1,it means that the customer is potential

Fig.6.6.If the prediction belongs to class-0,it means the customer is a not potential

CHAPTER 7 ADVANTAGES AND DISADVANTAGES

ADVANTAGES

Implementing customer segmentation leads to plenty of new business opportunities. You can do a lot of optimization in:

- budgeting,
- product design,
- promotion,
- marketing,
- customer satisfaction

DISADVANTAGES

- Marketing will become expensive.

- Because of having less no. of customers in a segment problem of limited production occurs.

CHAPTER 8

APPLICATIONS

The areas where this solution can be applied:

- Can be applied in each and every individual's Daily Life.
- Finding optimal number of unique customer groups will help you understand how your customers are different, and help you give them exactly what they expect from your company. Employing

CHAPTER 9

CONCLUSION AND FUTURESCOPE

CONCLUSION

Nowadays the competition has been highly increased in every industry, retail is no exception. So every business either it may be a small supermarket or an ecommerce giant like amazon ,flipkart. Every business try to use some tools, approaches, marketing strategies to attract customers towards their business. One such approach used by the above mentioned is customer segmentation. It is obvious that each and every customer can't be served with same product model, SMS campaigns, emails, advertisements. Customers have different needs. Treating all customers equally might not benefit the company in long run. Customer segmentation is one such cure for this problem. Finding optimal number of unique customer groups will help you understand how your customers are different, and help you give them exactly what they expect from your company. Employing

Customer segmentation has high probability of increasing your company's revenue. This is the reason why segmentation can turn out to be a great technique by means you can surpass your competitors in terms of profits andCan get you more customers. Doing it with machine learning is definitely the right choice**.**

FUTURESCOPE

On our Dataset, we have applied Random Forest Regression and K-means algorithm.

Random forest has got the highest accuracy of 99%.

Enhancements that can be made in the future:

While this method proposes a step-by-step manner for identifying and focusing on your best customer segments, truly following it does not completely gurantee company's profits.. To be efficient, you have to put together and plan for the diverse demanding

situations and hurdles that may occur at every step, and continually make certain changes to the method to process any new incoming

data that would alternate its output. And only employing this method cannot produce expected results, this method can also be collaborated with some other techniques, tools to produce best possible results.

CHAPTER 10 BIBILOGRAPHY

1.https://www.kaggle.com/dev0914sharma/customer-clustering

2.Sukru Ozan, "A Case Study on Customer Segmentation by using Machine Learning Methods", IEEE, Year: 2018.

3.Jayant Tikmani, Sudhanshu Tiwari, Sujata Khedkar "Telecom customer segmentation based on cluster analysis An Approach to Customer Classification using k-means", IJIRCCE,Year: 2015.

4.Chinedu Pascal Ezenkwu, Simeon Ozuomba, Constance kalu Electrical/Electronics and Computer Engineering Department, University of Uyo, Akwa Ibom State, Nigeria "Application of K-Means Algorithm for Efficient Customer Segmentation: A Strategy for Targeted Customer Services", IJARAI,Year: 2015.

5.Potharaju, S. P., Sreedevi, M., Ande, V. K., & Tirandasu, R. K. (2019). Data mining approach for accelerating the classification accuracy of cardiotocography. Clinical Epidemiology and Global Health, 7(2), 160-164.

6.Yogita Rani and Dr. Harish Rohil "A Study of Hierarchical Clustering Algorithm", IJICT, Year: 2013.

APPENDIX

A Source Code of Flask:

```
import numpy as np
import pickle import
pandas import os
from flask import Flask, request, jsonify, render_template from
sklearn.preprocessing import StandardScaler



app = Flask(_name_)
model = pickle.load(open(r'C:/Users/swath/OneDrive/Documents/Customer segmentation model/Dataset/xgbmodel.pkl','rb'))
scale = pickle.load(open(r'C:\Users\swath\OneDrive\Documents\Customer segmentation model\Dataset\scale.pkl','rb'))


@app.route('/')# route to display the home page def
home():
    return render_template('index.html') #rendering the home page


@app.route('/predict',methods=["POST","GET"])# route to show the predictions in a web UI
def predict():
    #  reading the inputs given by the user
    input_feature=[float(x) for x in request.form.values() ]
features_values=[np.array(input_feature)]
    names = [['Sex', 'Marital status', 'Age', 'Education', 'Income', 'Occupation',
       'Settlement size']]
    data = pandas.DataFrame(features_values,columns=names)
   #data = scale.fit_transform(features_values)
data=StandardScaler().fit_transform(features_values)
scaled_features=StandardScaler().fit_transform(data.values)
```

```
    # predictions using the loaded model file
prediction=model.predict(data)     print(prediction)


    if (prediction == 0):
        return  render_template("index.html",prediction_text  ="Not  a  potential  customer")
elif (prediction == 1):
        return   render_template("index.html",prediction_text   =   "Potential   customer")
else:
        return render_template("index.html",prediction_text = "Highly potential customer")
     # showing the prediction results in a UI if
_name=="__main_":


   # app.run(host='0.0.0.0', port=8000,debug=True)    # running the app
port=int(os.environ.get('PORT',5000))
app.run(port=port,debug=True,use_reloader=False)
```