|  |  |
| --- | --- |
| **Epic** | **PREDICTIVE MARKETING ANALYTICS** |
| **Project status** | **Development** |
| **Client** | Internal – Sales and Business Development Team |
| **Project owner** | Jovita |
| **Developers** | Jovita |
| **Stakeholders** | Customers  Partners  Internal Sales and Business Development Team |
| **Design complete** | 02-MAR-2020 |
| **Code complete** | 08-MAR-2020 |
| **Testing complete** | 11-MAR-2020 |
| **Release date** | 12-MAR-2020 |

Background

Companies/Retail Stores invest in customers (acquisition costs, offline ads, promotions, discounts & etc.) to generate revenue and be profitable. Naturally, these actions make some customers super valuable in terms of lifetime value but there are always some customers who pull down the profitability .We need to identify these behavior patterns, segment customers and act accordingly.

In this project we have first used unsupervised technique and categorized customers into various clusters and then used predictive modelling technique to predict next purchase date of individual customers, this can then be used for analyses and have targeted marketing techniques.

Problem statement

How does an organization identify its Valuable or more loyal customers and how does it target its marketing strategy to its valuable customers to maximize their sales which in turn generates revenue to the organization.

Industry/ domain

The analyses from this project helps retail industry to maximize their sale. Data considered here as an example is from online sales from UCI. Predictive analytics are used to determine customer responses or purchases, as well as promote cross-sell opportunities. Predictive models help businesses attract, retain and grow their most profitable customers.

Stakeholders

Customers: whose information is used for analysis and the results are then applied on them. Other stakeholders include topmanagement, marketing managers, retail sales agents, field sales agents, dealers and product managers.

They care about the issue to maximize sales and generate revenue. And most stakeholders would want this model to predict accurate results to plan and increase their profits.

# Business question

What promotional offer to give to a particular customer in a given time?

When to give a promotional offer to a particular customer and for what value?

Business Value of answering this question would be as below:

1. For a small dataset considered in this example, if we are able to convert around 202 average customers into profitable customers this would increase sales by a minimum of 27,000 GBP, annually.
2. If we could drive a at least 5% of profitable customers who were supposed to make a purchase next month, to make purchases this month then it would increase our monthly sales by 8,500 GBP.

We would require our model to do at least 80% correct predictions as it then effects the pipelined activities like promotional offers and marketing. Inaccurate model might not reflect the actual business scenario and might lead to incorrect decisions that would later effect the business growth.

# Data question

Data should provide customer sales history like date of purchase, quantity and payment value. It would require to have all the customer transactions for a given period in order to make correct predictions. It is better to have date for a longer period to make better analysis and prediction.

Basic questions that would include while considering the data would be as follows:

1. Does the data include Purchase date?
2. Does the data include quantity purchased with unit price or payment value of each purchase?
3. Does the data include all the transactions for a customer in a given period?
4. What’s the date range of the sales history?

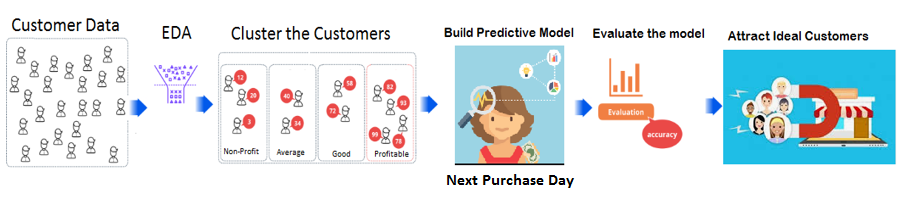
# Data

Data was sourced from UCI Machine Learning Repository. Data is about the online sales of a retail store. It had 541909 rows with attributes like Invoice No, Stock Code, Description, Quantity, Invoice Date, Unit Price, Customer ID and Country. Raw data was very clear and useful except for 133600

Rows which did not have Customer ID, since Customer ID very much essential for analysis we decided not to consider the rows which did not have Customer ID.This is a transnational data set which contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail.

# Data science process

Below steps shown in the diagram were used for Data Science process



# Assumptions:

1. Factors that affect Customer value are Recency, Frequency and Payment Value only.
2. Payment value only depends on Unit Price and Quantity.
3. Negative value for quantity means, goods are returned.
4. All Sales information for a customer for a given period is included in the dataset.

# Data analysis

Raw data was used to convert data into required format and below are the important steps carried out.

1. Recency Calculation: Calculating the Inactive period of the customer is been done from subtracting the last purchase date of the customer from the last date in the dataset. Do not consider the value with negative quantity as it is assumed to be returned order.
2. Frequency Calculation: Calculating number of visits made by the customer within the time period in the dataset. Do not consider the value with negative quantity as it is assumed to be returned order.
3. Payment Value: Multiply the unit price and quantity of each purchase. Consider the one’s with negative value also else it would show a positive amount for a returned order.
4. Convert the above 3 parameters Recency Calculation, Frequency Calculation and Payment Value into ranks with the range of 0 to 1 to normalize the data.
5. Split the data into learning period and testing period to predict the next purchase date in testing period.
6. Find the purchase trend in the training period by considering last 3 purchases made by the customer in training period.
7. Only consider those customer who have made at least 3 purchases in training period.

This process is reusable and the same could be done for future data sets to achieve similar results. The intermediary data structure use in pandas data frame.

# Modelling

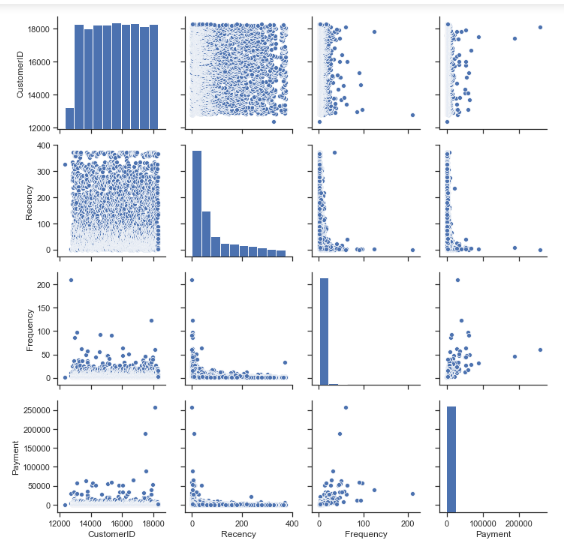
Main Features considered for Unsupervised learning as follows

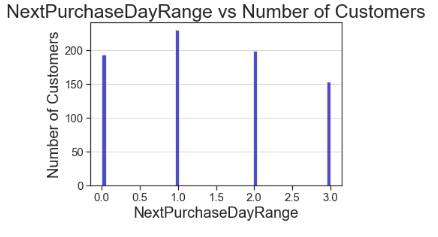
1. **Customer ID**
2. **Recency**
3. **Frequency**
4. **Payment**

Main Features considered for supervised learning as follows

1. **Customer ID**
2. **Recency**
3. **Frequency**
4. **Payment**
5. **Cluster**
6. **Invoice Day(n)**
7. **DayDiff1(n-1)**
8. **DayDiff2(n-2)**
9. **DayDiff3(n-3)**

The interesting interactions between features could be shown as below:



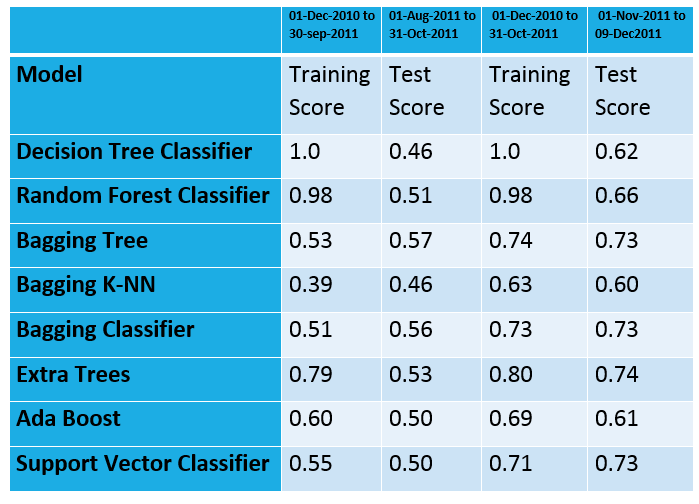


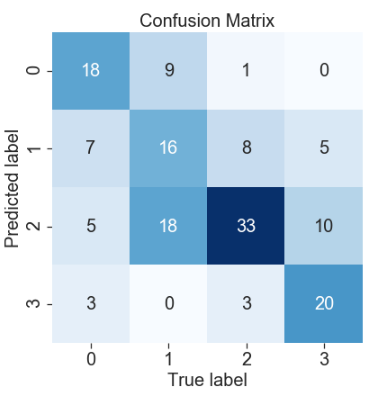
Analyzing the trend of the data helped us to choose the correct features for modelling and in addition to this we had to do feature engineering and calculate the columns mentioned above to obtain the best results.

Feature Engineering techniques used include

1. Ranking the data to normalize it.
2. Converting the rows into columns to get the intervals of customer visit.
3. Averaging the values to obtain better results.
4. Using unsupervised learning method of K means clusters to cluster the customers into value groups.

Below is the image showing models used and the accuracy score for the same .Bagging Classifier is the final model chosen.





# Outcomes

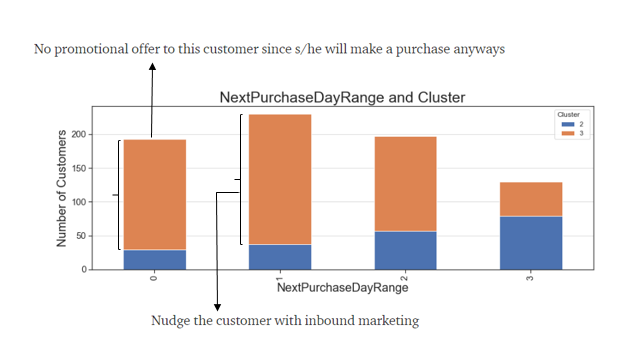
It could be found that customers could be clustered into different value groups and next purchase date for the customer could be predicted.

# Implementation

This code could be deployed in similar industry application and providing the recent sales data as input one could predict the next purchase day for a customer and what promotions to offer to the customer to increase sales.

# Business answer

Business answer could be as shown in the below figure:



# Response to stakeholders

Stakeholders could take business decisions based on the project and increase the sales by understanding customer trends.

# References

http://archive.ics.uci.edu/ml/datasets/Online+Retail

https://towardsdatascience.com/data-driven-growth-with-python-part-3-customer-lifetime-value-prediction-6017802f2e0f

https://support.treasuredata.com/hc/en-us/articles/360001458407-Predicting-Customer-Behavior

https://www.wsj.com/articles/predictive-analytics-give-a-boost-to-diageos-cost-savings-efforts-11556208946