Coupon Purchase Prediction
(Using Machine learning)

Shivam Agarwal

16/05/2023

1.ABSTRACT

The project is based on user data analysis which is used to predict whether a user will purchase coupons or not based on his previous purchases from the site. Issuing discount shopping coupons and barter purchase is a popular way to promote sales in E-commerce environments. In order to improve the coupon usage ratio, it is important to predict the probability in which a user will use a coupon. To treat the coupon redemption probability prediction problem as a binary classification problem, we use machine learning methods to analyze users' coupon usage behavior and coupon purchase prediction.

Keywords: Machine Learning, Coupons Code, Prediction, E-Commerce

2.INTRODUCTION

E-Coupons are used to reduce the price of a product or products that customers purchase from any online website, making it more useful for consumers as well. E-coupons are more attractive to the buyer, more efficient and more efficient at tempting new buyers. Coupons promise to make better flexibility, really fraud-protective, socialization incentives, advertisers, more agreeable, and more. In the paper that follows we present a summary of the scenery as well as strategies to optimize marketing sales by using digital coupons. Customers who use E-coupons are recommended group. Most online store websites use consumer buying history and provide consumers with different e-coupons which are used to lower the price of that product.

Why Coupons Purchase Prediction is Required?

- Coupons Purchase prediction allows advertisers to increase their market share, and raise sales volume often reactivates those customers you can have to lose.

- Because of this forecast, many new consumers are registers on online store websites. Since the coupons give discounts to consumers on those websites.

- This prediction answers plenty of consumer behavior problems against coupons or websites. A few questions do she or he visit the online stores on what days and times? How much time does he waste, or she? Will she or he use coupons to buy something from supermarket.

- It also helps businesses develop their coupon offerings according to the consumption actions of the customer coupon. Because of this prediction marketers know the consumer's percentage rate to purchase and show coupons.

3.PROBLEM STATEMENT

Designed a recommendation engine for Ponpare, Japan’s leading joint coupon site which predicts the 5 most likely coupons the customer will buy out of 310 available coupons.Performed the computation, log transformations, feature engineering and Data exploration to identify the insights and conclusions from complex data.Built an algorithm implementing the concept of Cosine Similarity which recommends the 5 coupons the customer will most likely to buy.

4.Market/Customer/Business need Assessment

The Project is based on user data analysis which is use to predict users coupons usage behaviour. Issuing discount shopping coupons is a popular way to promote sales in Ecommerce environments. In order to improve the coupon usage ratio, it is important to predict the probability in which a user will use a coupon. To treat the coupon usage probability prediction problem as a binary classification problem, and we use machine learning methods to analyze users coupon usage behaviours and coupon purchase prediction.

5.Target Specification

Coupons Purchase prediction help marketers grow their market share, increase sales volume also reactivate those consumers you might have to lost contender. Due to this prediction lots of new customers are registers on online store websites .because those websites provides the coupons offer discounts to customers. This prediction solve lots of questions about customers behaviour against coupons or websites.

6. Business Model

It helps businesses develop their coupon offerings according to the consumption actions of the customer coupon. Because of this prediction marketers know the consumer's percentage rate to purchase and show coupons.In the paper that follows we present a summary of the scenery as well as strategies to optimize marketing sales by using digital coupons.

Customers who use E-coupons are recommended group. Most online store websites use consumer buying history and provide consumers with different e-coupons which are used to lower the price of that product.

7. External Search

1- https://ieexplore.ieee.org/document/970058.
2- https://rstudio-pubs-static.s3.amazonaws.com/136006_19010ce2f5144f4c89556a3eee159c57.html

8. Technologies Used

We used Weka tool for feature engineering techniques like Data Feature Formatting, filling NANs appropriately by grouping with other attributes and label Encoding.
We have used Random Forest Model for classification of data points and parameters are tuned and optimized and tuned properly which predicts the Redemption Status.

A. Weka Tool
Weka is data mining software that uses a collection of machine learning algorithms. These algorithms can be applied directly to the data or called from the Java code.Weka is a collection of tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.

Only relevant features are identified and used and relationships between multiple features are observed. Cross-validation is done using Stratified K-Folds which is used along with Grid Search techniques.

10. Applicable Regulations (Government and Environmental)

- Data protection and privacy regulations(Customers).
- Govt Regulations for small businesses.
- Employment Laws .
- Antitrust Regulations.
- Regulations against false advertising.

11. Applicable Constraints

- Data Collection from Companies and Websites.
- Continuous data collection and maintenance.
- Lack of technical knowledge for the user.
- Taking care of rarely bought products
- Convincing the Companies to implement the system in their Workspace.

12. Concept Generation

We come up with the idea on basis of these top reasons:

- Convenience
- Saving Time

- Saving Money / Cost Control
- Comparison Shopping
- Greater Variety / More Options
- Personalization

13. Project Details

Pre-processing and supervised learning methods use in prediction:
For each pair (consumer, coupon code), we calculate the probability that the consumer will be able to purchase such coupons during the test period using a gradient boosting classifier that helps reduce a collection of parameters, coefficients in a regression equation. Through chance, we sorted the coupons for every single and very customer. We do training data for 45 "train cycles" to train gradient boosting classifier that simulated the test timing. Train timing 1 is the week from 2019-02-08 through 2019-02-14, which contains all coupon code with a date of DISPFROM-the dates on which they will be seen for the first time-in that week.Train duration 2 is the week from 2019-02-15 to 2019-02-21, which contains all coupons code in that week with a DISPFROM date. Train duration 5 is the week from 2019-03-17 to 2019-06-23, and that week contains all coupons code with a DISPFROM date. The only supervised method of learning which we are using was gradient boosting. At the beginning of our research, we cycled through other algorithms to get a feeling for their efficiency-technical regressions, SVMs, random forests as well as deep neural networks-but found that gradient boosting was the best fit for our approach.

1- At the beginning of our research, we cycled through other algorithms to get a feeling for their efficiency-technical regressions, SVMs, random forests as well as deep neural networks-but found that gradient boosting was the best fit for our approach. The goal for each observation is set at 1 if that coupon was purchased by the user during the week of training, and otherwise 0.

Simply counting the amount of times a consumer has seen a test set coupon is significantly helpful in predicting purchases from test set. As seen in the figure below in the left column, users often purchase a coupon code if they display it exactly once ordered to their DISPFROM, but that likelihood increases to 32 per cent if they display the coupon three or more times.

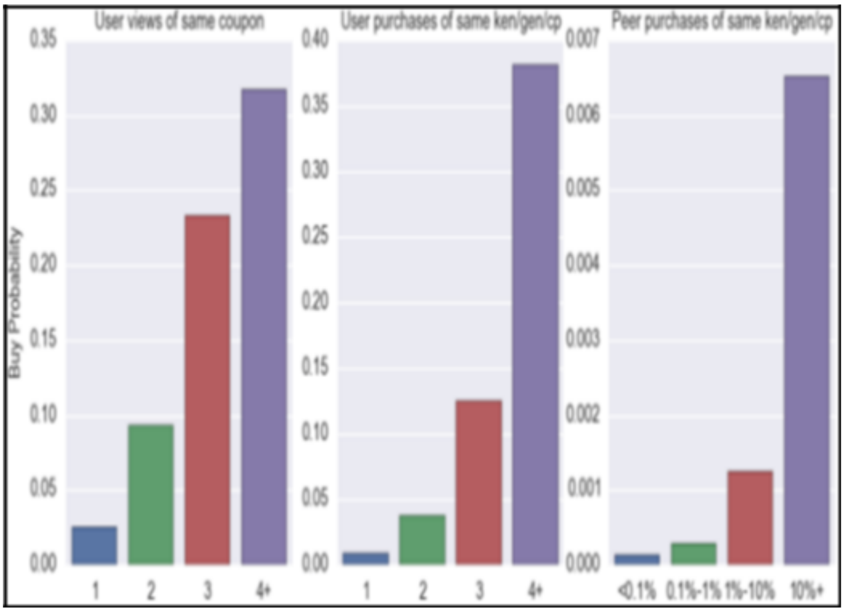


Fig. 1: Prediction

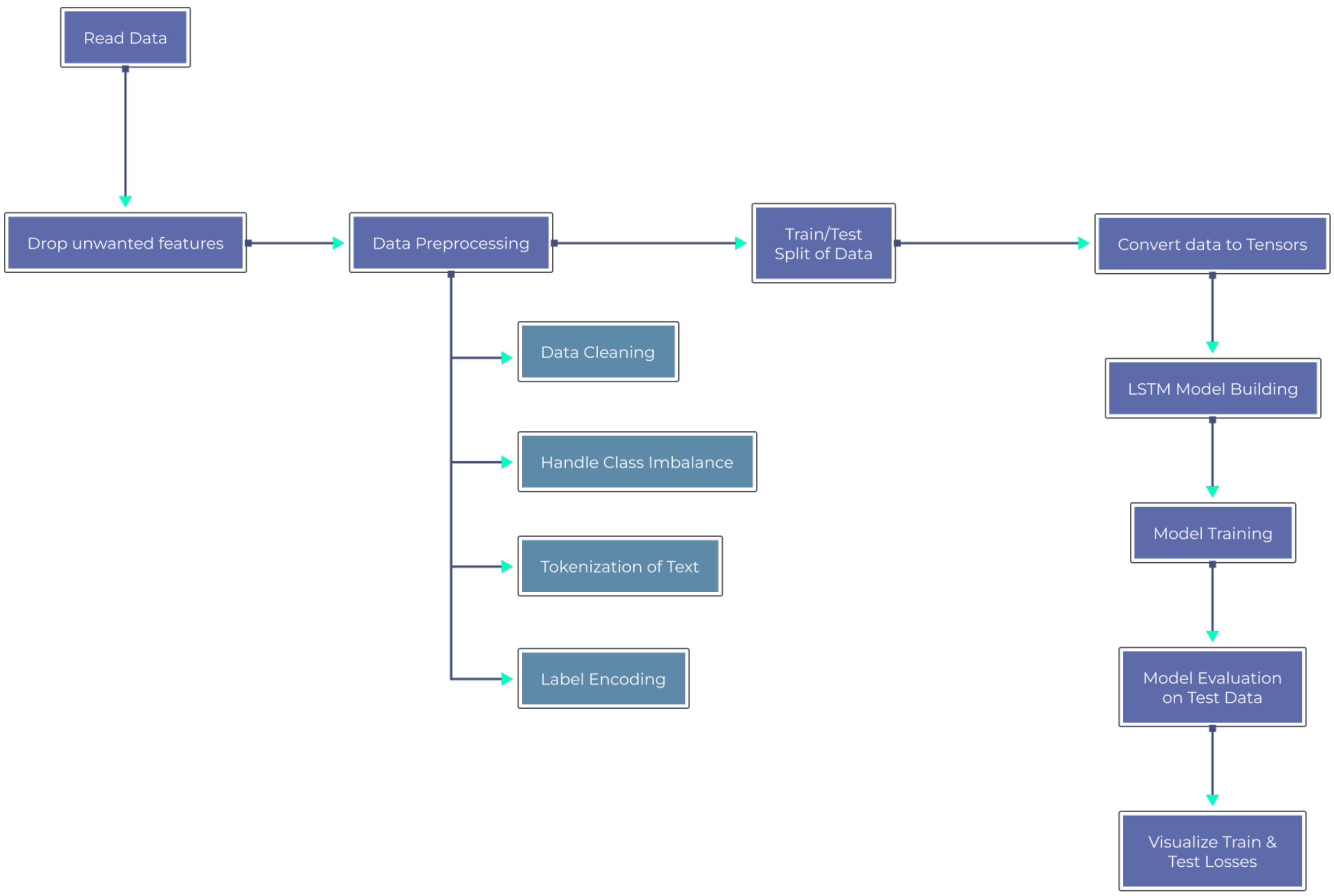
2 - Second, consumers tend to buy the same coupons time and time again. As shown in the middle panel of the above figure, if offered for sale, a customer who has purchased a coupon with a given prefecture, genre, and catalog cost four or more times has a 38 percent chance of purchasing a matched coupon again in the following week.

3-Third, common interests group averages can help forecast the activity of customer with little or no history data. The right panel of the above figure display that a user's probability of purchasing a coupons code increases from less than 0.1% to above 0.6% if more than ten percent of sex, age and geography-matched peers have purchased a coupon code with the same properties.

4- Fourth, it's very important to consider the geographic coverage of each and very coupon code. To be specific, a coupon code is real for the more prefectures listed incoupon_area_train.csv files, not just the single prefecture listed for that coupon in coupon_list_train.csv files

14. Future Scope

Everyone knows the Internet's growth is exponentially growing, people often go shopping online to buy everything they need for everyday use. Do you know that anyone can save money by using online coupon code that lowers the price of the product that purchases it? They search on your favorite shopping e-commerce website for available coupons, discounts, cashback on every good transaction or online purchase. There are more types of goods you can buy online today, but what many people on the planet don't realize is that you don't have to pay the real cost of those listed items. Online shopping is quick and easy to buy any items without going anywhere, as well as having many benefits such as coupon codes, promotional codes and coupons that give us so many product discounts. Because of Coupons Code, in future people often refer to online shopping websites.



15. ADVANTAGES AND DISADVANTAGES ONLINE COUPONS TO THE CUSTOMERS

A. Advantages

1) Easy to Search and Use Coupons Code -Coupons can be quickly identified online. Website grabaon.com, and several other forms of website code coupons. You're only going to the sites as a customer and picking up just one coupon. Yet keep in mind that various coupons are useful for different programs and different goods.

2) Save money- money is the most significant benefit. Once consumers first order every commodity they search for any relevant offers or not. If coupons apply then the price of the good is decreased.

B. Disadvantages

1) Over-Spent-Consumers waste more time on goods by utilizing discounts than they had expected.

2) May Feel Alienated Existing Consumers-Advertisers always know who their current customers and new customers are, as well as one-time buyers. They just want to include discounts for potential consumers so others who are current consumers would be left out. In this situation they would have to deliver promotions code in a way that suits all their customers ‘needs.

16.Code Implementation

```
from .timer import Timer

import pandas as pd
import numpy as np



class Model(object):

    def __init__(self, train, test, users, purchases, visits):
        """
        :param train: pandas.DataFrame of training coupon data
        :param test: pandas.DataFrame of test coupon data
        :param users: pandas.DataFrame of user data
        :param purchases: pandas.DataFrame of All user purchases
        :param visits: pandas.DataFrame of All user visits
        """
        self.users = users
        self.purchases = purchases

        self.visits = visits[visits.VIEW_COUPON_ID_hash.isin(train.COUPON_ID_hash)].copy(deep=True)
        self.visits = self.visits[self.visits["PURCHASE_FLG"] == 1]

        self.fields = ["COUPON_ID_hash",
                       "CAPSULE_TEXT", "GENRE_NAME",
                       "PRICE_RATE", "CATALOG_PRICE", "DISCOUNT_PRICE",
                       "VALIDPERIOD",
                       "USABLE_DATE_MON", "USABLE_DATE_TUE", "USABLE_DATE_WED",
                       "USABLE_DATE_THU", "USABLE_DATE_FRI", "USABLE_DATE_SAT",
                       "USABLE_DATE_SUN", "USABLE_DATE_HOLIDAY", "USABLE_DATE_BEFORE_HOLIDAY",
                       "large_area_name", "ken_name", "small_area_name"]

        # keep relevant coupon fields only
        self.train = train[self.fields].copy(deep=True)
        self.test = test[self.fields].copy(deep=True)
```

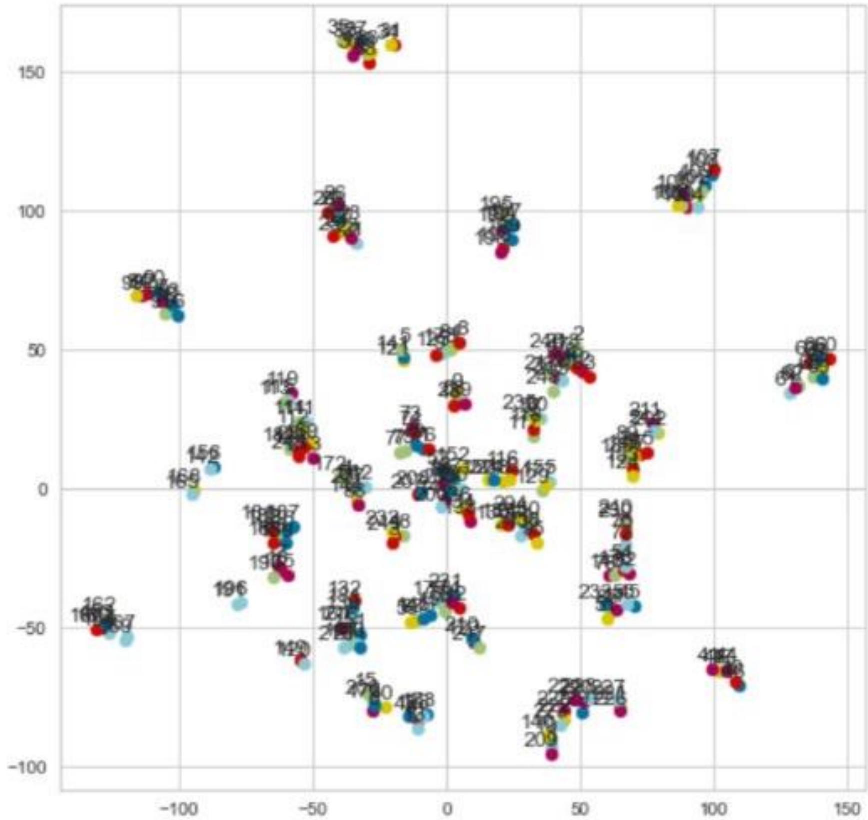
```python
        # expand categorical variables
        self.categorical = ["CAPSULE_TEXT", "GENRE_NAME",
                            "large_area_name", "ken_name", "small_area_name"]
        self.categorical_weights = [3.0, 3.0,
                                    3.0, 3.0, 3.0]
        self._expand(self.categorical_weights, transform=1.0)

        # scale numerical variables
        self.numerical = ["PRICE_RATE", "CATALOG_PRICE", "DISCOUNT_PRICE",
                          "VALIDPERIOD",
                          "USABLE_DATE_MON", "USABLE_DATE_TUE", "USABLE_DATE_WED",
                          "USABLE_DATE_THU", "USABLE_DATE_FRI", "USABLE_DATE_SAT",
                          "USABLE_DATE_SUN", "USABLE_DATE_HOLIDAY", "USABLE_DATE_BEFORE_HOLIDAY"]
        self.numerical_weights = [1.0, 1.0, 1.0,
                                  1.0,
                                  1.0, 1.0, 1.0,
                                  1.0, 1.0, 1.0,
                                  1.0, 1.0, 1.0]
        self._scale(self.numerical_weights, transform=1.0)

        # replace missing values
        self._replace_nan(0.)

        self.timer = Timer()

        # construct ItemProfile using finalized training and test sets
        self.item_profile = ItemProfile(self.train, self.test)

        # parameters
        self.num_purchases_w = 0.15
        self.purchase_date_w = 0.85
        self.purchased_w = 0.5
        self.visited_w = 0.5
```

```python
    @staticmethod
    def run():
        """
        Run model training process.
        """
        print "No training required..."


    def predict(self):
        """
        Predictions are made on the test set; returns DataFrame in Kaggle submission format.
        """
        submission = []

        print "Userlist size: ", self.users.shape
        print "Training coupons: ", self.train.shape
        print "Test coupons: ", self.test.shape

        e = 0
        for index in self.users.index:
            # get user
            user = self.users.ix[index]
            # get relevant coupons and corresponding weights
            purchased_coupons, purchased_weights, visited_coupons, visited_weights = self._coupon_filter(user)
            # get recommendations
            final_coupons = self._recommend(purchased_coupons, purchased_weights, visited_coupons, visited_weights)
            # add to submissions
            submission.append([user.USER_ID_hash, final_coupons])
            e += 1
            if e % 1000 == 0:
                print "At User: ", e

        return pd.DataFrame(submission, columns=["USER_ID_hash", "PURCHASED_COUPONS"])
```

```python

    def _coupon_filter(self, user):
        """
        :param user: row corresponding to user in user_list
        Takes the user and returns the purchased coupons and visited coupons along
        with their corresponding weights.
        """
        # get user purchases (note: not all users have made purchases)
        user_buys = self.purchases[self.purchases.USER_ID_hash == user.USER_ID_hash]
        # get corresponding purchased coupons
        purchased_coupons = self.train[self.train.COUPON_ID_hash.isin(user_buys.COUPON_ID_hash)]

        # get user visits (not including purchases)
        user_visits = self.visits[self.visits.USER_ID_hash == user.USER_ID_hash]
        # get corresponding visited coupons
        visited_coupons = self.train[self.train.COUPON_ID_hash.isin(user_visits.VIEW_COUPON_ID_hash)]

        final_purchased_weights = None
        if not purchased_coupons.empty:
            # get the frequency of purchase for each coupon
            bought_coupon_groups = user_buys.groupby(by='COUPON_ID_hash').groups
            purchased_weights = {}
            for key in bought_coupon_groups:
                new_key = purchased_coupons[purchased_coupons.COUPON_ID_hash == key].index[0]
                purchased_weights[new_key] = len(bought_coupon_groups[key])
            purchased_weights = pd.DataFrame.from_dict(purchased_weights, orient='index').sort_index()

            purchased_weights.columns = ["freq"]
            purchased_weights = Model._normalize(purchased_weights) # scale to [0,1]

            # get the most recent purchase date for each coupon
            pdates = user_buys[["COUPON_ID_hash", "NUM_DAYS"]].groupby(by='COUPON_ID_hash').max()
            pdates.columns = ["recent"]
            pdates = Model._normalize(pdates) # scale to [0,1]
            actual_index = []
```

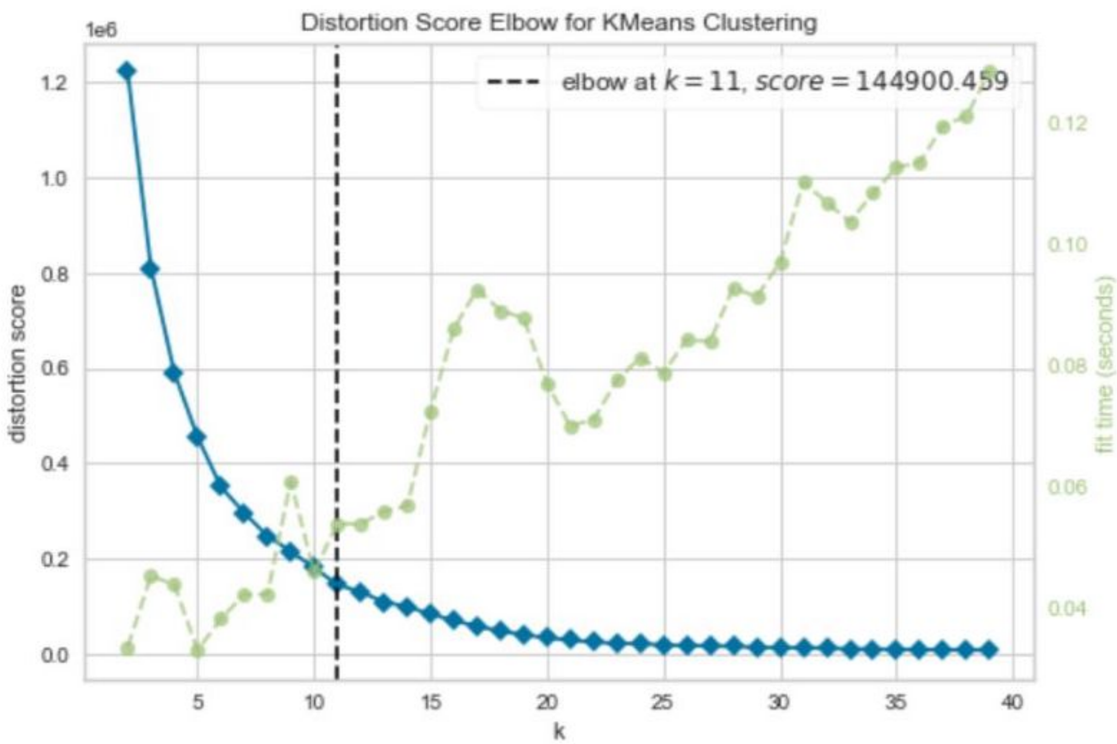
In [85]:

```
p2cluster_model = p2cluster(product_vector_model.p2v_model)
p2cluster_model.tsne_train()
p2cluster_model.tsne_plot()
```



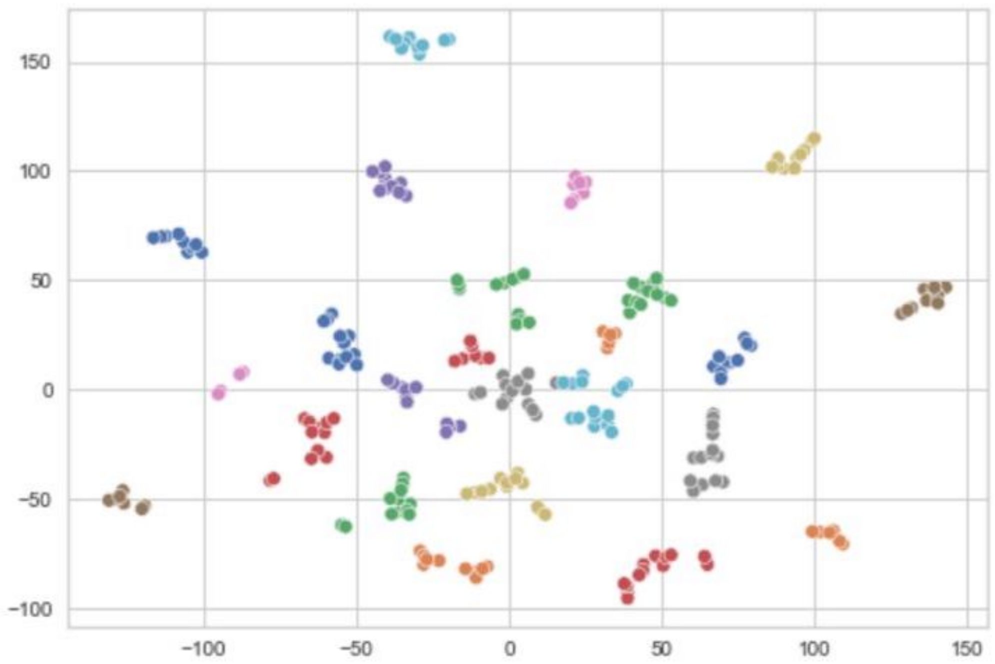
In [86]:

```
p2cluster_model.elbow_plot()
```

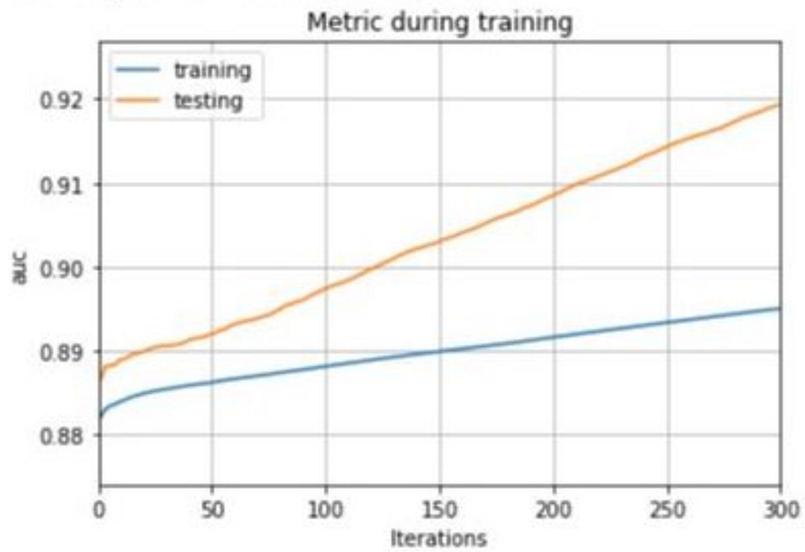


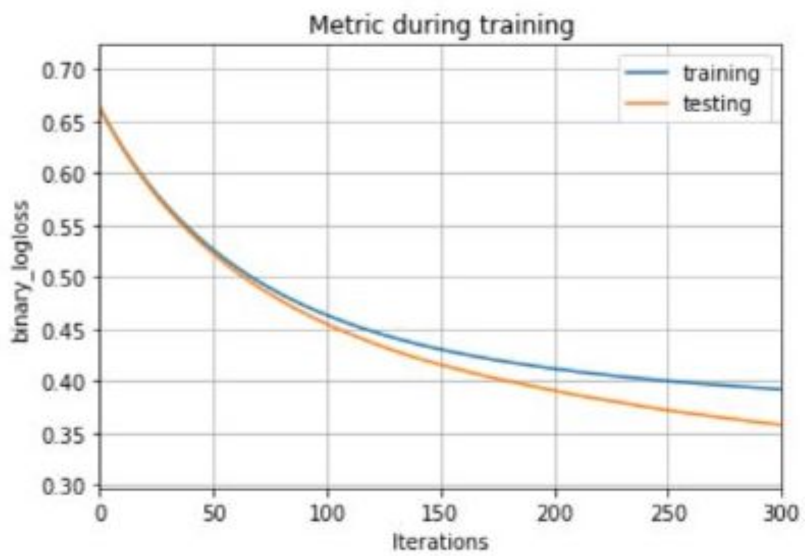
In [87]:

```
# set to 25 clusters
p2cluster_model.train_cluster(25)
p2cluster_model.clust_plot()
```

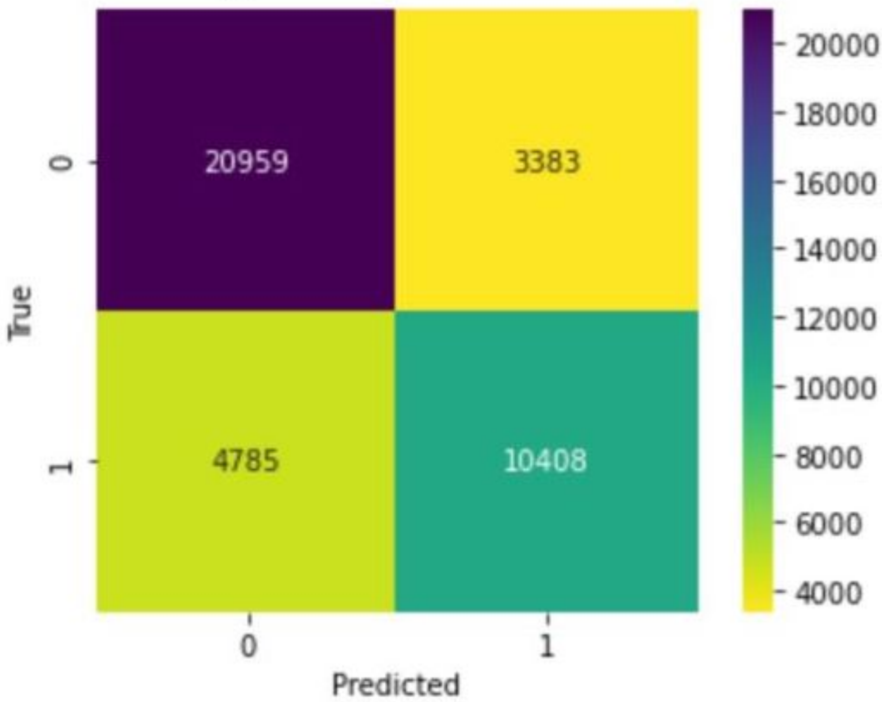



```
The computation took 9.42 minutes.
```





```
Classification Accuracy:  0.7933982547110155
Classification Error:    0.20660174528898445
Recall:                  0.6850523267294149
Precision:               0.7546950910013777
F1 score:                0.7181893458459839
AUC:                     0.8846233294727821
Cross Entropy:           0.41213648762970256
```



17. Conclusion

According to the expected performance of consumer data, e-commercemay use performance results to predict discount coupons code would be purchased by which cluster with higher probability and e-commerce company would decide to make specific rules of one particular form coupons code for specific clusters and peer groups of customers, concentrating in particular on the validity of coupons code to accurate marketing store and discount rates. And when every e-commerce company tries to lay down its coupon code sales planning, our concept is that e-company should first identify a peer-customer cluster and then focus on consumer purchasing behavior to tailor the suggested behavior