

**CUSTOMER SEGMENTATION BY USING MACHINE LEARNING AND
E-COMMERCE SOLUTION**

BY

**ANUP CHANDRA BEPARY
ID: 162-15-8057**

AND

**ZANNATUL FERDOUS
ID: 162-15-7703**

This Report Presented in Partial Fulfillment of the Requirements for the
Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Ms. Afsara Tasneem Misha
Lecturer
Department of CSE
Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY

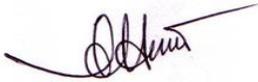
DHAKA, BANGLADESH

JULY 2020

APPROVAL

This Project titled “CUSTOMER SEGMENTATION BY USING MACHINE LEARNING AND E-COMMERCE SOLUTION” , submitted by **Anup Chandra Bepary, ID: 162-15-8057** and **Zannatul Ferdous, ID: 162-15-7703** to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 09 July 2020.

BOARD OF EXAMINERS



Dr. Syed Akhter Hossain
Professor and Head

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

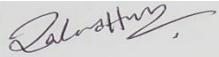
Chairman



Dr. Sheak Rashed Haider Noori
Associate professor & Associate Head

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

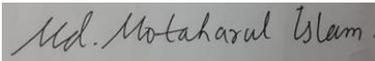
Internal Examiner



Md. Zahid Hasan
Assistant Professor

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Dr. Md. Motaharul Islam
Professor

Department of Computer Science and Engineering
United International University

External Examiner

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Ms. Afsara Tasneem Misha, Lecturer, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

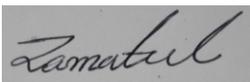


Ms. Afsara Tasneem Misha
Lecturer
Department of CSE
Daffodil International University

Submitted by:



(Anup Chandra Bepary)
ID: 162-15-8057
Department of CSE
Daffodil International University



(Zannatul Ferdous)
ID: 162-15-7703
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to **Supervisor Ms. Afsara Tasneem Misha, Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of “*Field name*” to carry out this project. His endless patience, scholarly guidance, continual encouragement ,constant and energetic supervision, constructive criticism, valuable advice ,reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to -----, -----, and Head, Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Customer segmentation is very important for online business like ecommerce. To grow a business a person should understand their customers very thoroughly. That's why customer segmentation is very important. It is also part of Business intelligence. This research is about clustering customers into groups. We made an ecommerce for farmers where customers will sign up and from there we collected the data of the customers for the segmentation. We used unsupervised learning for the segmentation. K Means algorithm is used for the clustering. Using the clustering we can get idea about our customers. Farmers can understand how much they should produce every year. Business man can also make off and coupons for their customers easily and which customer will need which ads. This research will help to run ecommerce business very efficiently with 94.44 % accuracy.

Keywords – Customer segmentation, Machine learning, unsupervised learning, k-means clustering

TABLE OF CONTENTS

CONTENTS	PAGE
Board of examiners	Ii
Declaration	Iii
Acknowledgements	Iv
Abstract	v

CHAPTER

CHAPTER 1: Introduction	1-3
1.1 Introduction	1
1.2 Motivation	1
1.3 Rationale of the Study	2
1.4 Research Questions	2
1.5 Expected Output	3
1.6 Report Layout	3
CHAPTER 2: Background	4-6
2.1 Introduction	4
2.2 Related Works	4
2.3 Research Summary	5

2.4 Scope of the Problem	5
2.5 Challenges	6
CHAPTER 3: Research Methodology	6-14
3.1 Introduction	6
3.2 Research Subject and Instrumentation	6
3.3 Data Collection Procedure	8
3.4 Statistical Analysis	9
3.5 Implementation Requirements	11
CHAPTER 4: Experimental Results and Discussion	15-19
4.1 Introduction	15
4.2 Experimental Results	15
4.3 Descriptive Analysis	15
4.4 Summary	19
CHAPTER 5: Summary, Conclusion, Recommendation and Implication for Future Research	20-22
5.1 Summary of the Study	21
5.2 Conclusions	21
5.3 Recommendations	22
5.4 Implication for Further Study	22
APPENDIX	22
REFERENCES	23-24

LIST OF FIGURES

FIGURES	PAGE NO
Figure 1: Research Process Flowchart	7
Figure 2: K-Means algorithm working Process Flowchart	9
Figure 3: Number of Cluster	10
Figure 4: Distplot data according to Age, Annual Income and Spending score.	16
Figure 5: Gender difference	16
Figure 6: Age based income and spending cluster	17
Figure 7: Age vs income w.r.t gender	17
Figure 8: Annual income and spending score	18
Figure 9: Clustering with the spending score	19

LIST OF TABLES

TABLES	PAGE NO
TABLE I: Dataset	8
TABLE II: Data Description	10
TABLE III: <i>K</i> -Means Algorithm	12
TABLE IV: Bisecting of <i>K</i> -Means Algorithm	13

CHAPTER 1

Introduction

1.1 Introduction

In this modern era of science and information technology E-commerce is an important issue. Which through not only simplifies of our standard living but also connects us with the world in a touch of modernity. That's one the example of modern technology. And the customer is one of the important parts of E-Commerce. But knowing the demands of a huge customer and what a customer purchase what product it's a random process. But it's an important thing that- Provide the product as per the demand of the customer and provide the service accordingly. So, we think that-by using machine learning to segmentation the customers and according to create an e-commerce site by clustering to given the E-commerce solution. In this case customers purchase information is very important for our research. That's indicate the most things for E-commerce Solution. For example -How would Like he wants to buy the product with the price? Which product he/she wants to buy more? Does he want to take the offer? What types of the products wants to take more? What types of customer? In this case we think that- the admin is the service provider which can add product update product ,price ,add user, view location and track ,transaction money, want to know the demand of customer on the other hand customer is the service recipient which can sign up, log in, check in, purchase product, order, pay bit .Overall for the segmentation admin can provide help and customer can take necessity from E-Commerce solution. This research not only help our E-commerce solution but also exhibit proper classification of customer segmentation.

1.2 Motivation

Many online market places and e-commerce sites suffers about their profit & less account and their future condition of business. Particularly our main priority is comparing and classifying our expected outcome with our agro based business online position and the demand and the percentage of risk factor of business profit.

According to this solution, we might be taken so many effective action for advance business and digital marketing. Sometimes we are able to segment customers.

Age, gender, yearly Income and find out which customer focus on the moment which product he/she wants. and what's product things interested they want to by basic getting offer. If the admin could see that the customer wants to her/his favourite product.

So, the admin provides their needs. This policy reduces the risk factor of agro based digital business. By using this policy we can figure out a good customer segmentation and get a proper e-commerce solution.

1.3 Rationale of the Study

If the customer has good income and they will get all the good opportunity according their needs like-good product, get offer then the buying and selling will be better for customers and E-commerce site owner. But if when the customer carry good income monthly and he/she get a poor service in the e-commerce site or what would happen if the demand of product is so high and if the supply is low then what will happen? And if that this site can afford the advantage, that's a good thing. As a result even if the products price will high the customer purchase the product. In that context will do the better customer segmentation, which is good for e-commerce solution. We do the customer segmentation by based on customer annual income, gender, age and their demand of purchasing product, their taken offer. Which used by unsupervised machine learning technique and k-means algorithm to clustering customers and probable matching in future result from some exiting attribute of e-commerce site.

1.4 Research Questions

By observation and monitoring the performance of E-commerce site and demand of customer we have to formulate some logical question. So, some question were formulate by us. These are

- which is the main and fundamental Factor for performing a good customer segmentation and proper e-commerce solution.

- what is the relation between the e-commerce site

- which data mining technique show we follow to predict and figure Out the particular event.

- Which algorithm Should we use for the highest accuracy

- How customer and our agro business Would be benefitted by digital marketing.

1.5 Expected Output

The main goal of this research is allocate performing agro based e-commerce Site. perfect e-commerce solution depends on many things. usually every site have two character. (1) Admin (2) use. By using this e-commerce site, we entry the data set and segmentation of the customer. -- in which class they belong (1) elite class (2) middle class (3) Only for browsing for the window shopping. How many customer purchasing. How many they accept the offer and which product they purchase much and More. Every product have some quantity, price and offer and a clean idea. The perform Mostly depends on the marketing attributes, like product price, customers age, customers income and customers demand.

1.6 Report Layout

The research constructed by focusing on some particular un-supervised machine learning algorithm. Like k-means algorithm, clustering, decision tree. We have also considered similar work for making e-commerce site. we used Python Django, framework. And background work in some in the same topic Literature review study. we collected data set from 'Kaggle'. Then the research methodology has shown by supplementation. Applying the respective algorithm We discussed the output.

CHAPTER 2

Background

2.1 Introduction

Our research is based on providing data set to clustering the customer diving the categories which they are belong to and by using E-commerce site we can collect data in future and segmentation the customer. When we looked for previous work of this research we found many similarities and dissimilarities with other research. Most of the research goal was same.

2.2 Related work

Jong Tak ,et al.[1] discussed as the population increase, health services also increase that can be accessed by personal computers and smart phones to provide health related problems. P2P is based on PBR (Personal Bio Er.Rupampreetkaur et al, International Journal of Advanced Research in Computer Science, 8 (5), May-June 2017,857-861 © 2015-19, IJARCS All Rights Reserved 859 Record) platform for increasing silver population. The health cluster model used to provide services in multi - platform environment and enhance health status of old aged patients having chronic diseases.

Noble, et al. [2] analyzed use three clusters to identify the latent class analysis. The methodology of clustering was performed on 377 participants that attend aboriginal community controlled health service (ACCHS) in Australia. In this cluster one survey the low fruit/ vegetable intake , cluster two includes younger unemployed males who have smoking, alcohol addiction problem whereas cluster three include depressed personalities.

Cluster three only include younger to mid aged women. Alzahrani , et al.[3] reviewed two clusters in which one cluster has low fruit consumption and second one include high sweet consumption. The author use hierarchical agglomerative cluster analysis (HASA). Main study was on school students whose age is between 13-14 and 17-19 years. The students were divided in different grades i.e. intermediate and secondary schools. Health related behaviors, demographics characteristics parents occupation was included in questionnaire.

[4]The standard k-means clustering was used by the author. The k-means clustering was divided in three groups i.e. poor, intermediate and good outcomes. The surrogate value is included in this survey. The cluster include the data points from the surrogate variable value. The main limitation in this paper is to find the surrogate value. The surrogate value unstably correlated with the health outcomes. So the result may be inappropriate. The main survey was based on socio demographic background and is considered for health service utilization.Brito.et.al.

[5]The investigation of the data mining approaches for customer segmentation was the main aim of this paper. k-medoids and CN2-SD are the data managerial approaches. These are used to segment problem and complement each other. The segmentation is further divided externally and internally. The external segment helps the company to redefine communication strategy for sales promotion and internally matching the product of customer preference that will help to redefine product design.

Azizpour, et al.[6]Implemented the sources of corporate default clustering in the United States. We reject the hypothesis that firms' default times are correlated only because their conditional default rates depend on observable and latent systematic factors. By contrast, we find strong evidence that contagion, through which the default by one firm has a direct impact on the health of other firms, is a significant clustering source. The amount of clustering that cannot be explained by contagion and firms' exposure to observable and latent systematic factors is insignificant.

Guha ,et al.[7] Clustering can be a useful and everywhere tool with data analysis. Commonly conversing, clustering is definitely the difficulty of group any data collection directly into several groups so that, underneath many specification of "likeness," very similar products are with the exact same group and dissimilar products are in different groups.

Tsuboi.et al.[8]Within this phase we center on clustering in a very internet streaming predicament in which limited data products are introduced during a period and now we won't be able to retail store most the data points. Therefore, our own algorithms will be limited by a single pass. Space stops is generally sub linear, $i(\ln)$, in which the quantity of feedback factors can be in.

2.3 Research Summary

After studying the previous work most of the work was based on un-supervised learning. We also used un-supervised learning. We used the k-means clustering. All the works based on k-means clustering algorithm and un-supervised learning method.

2.4 Scope of the Problem

It was a big problem for us to finding the accurate data. Also problem for us which algorithm will work perfectly to solve the problem. At first we had poor amount of data in that time our accuracy rate was really low. So that we try to find a huge amount of data set from the 'Kaggle' which is an machine learning online platform and also made a dummy website for customer segmentation and E-commerce solution.

2.5 Challenges

It was a big challenge for us to preparing a dataset and marketing an E-commerce site for segmentation the customers. First we didn't know how customers purchase the product by basis on their demand. Also looking the website for real data from survey was really tuff. Because this types of data really sensitive and confusing.

CHAPTER 3

Research Methodology

3.1 Introduction

Our raised methodology started with gathering data from our vegetable ecommerce site. We created the ecommerce site to sell vegetable and other things online. It was made to help the farmers. Our research part customer segmentation is based on the customer data from the ecommerce site. This is a classification problem. Gender, Age, Annual income, Spending score data will helped us to build our model

3.2 Research Subject and Instruction

Our research subject is about clustering the customers of the ecommerce site. If we can understand and cluster the customers into groups. Then we can make appropriate offers for all groups accurately.

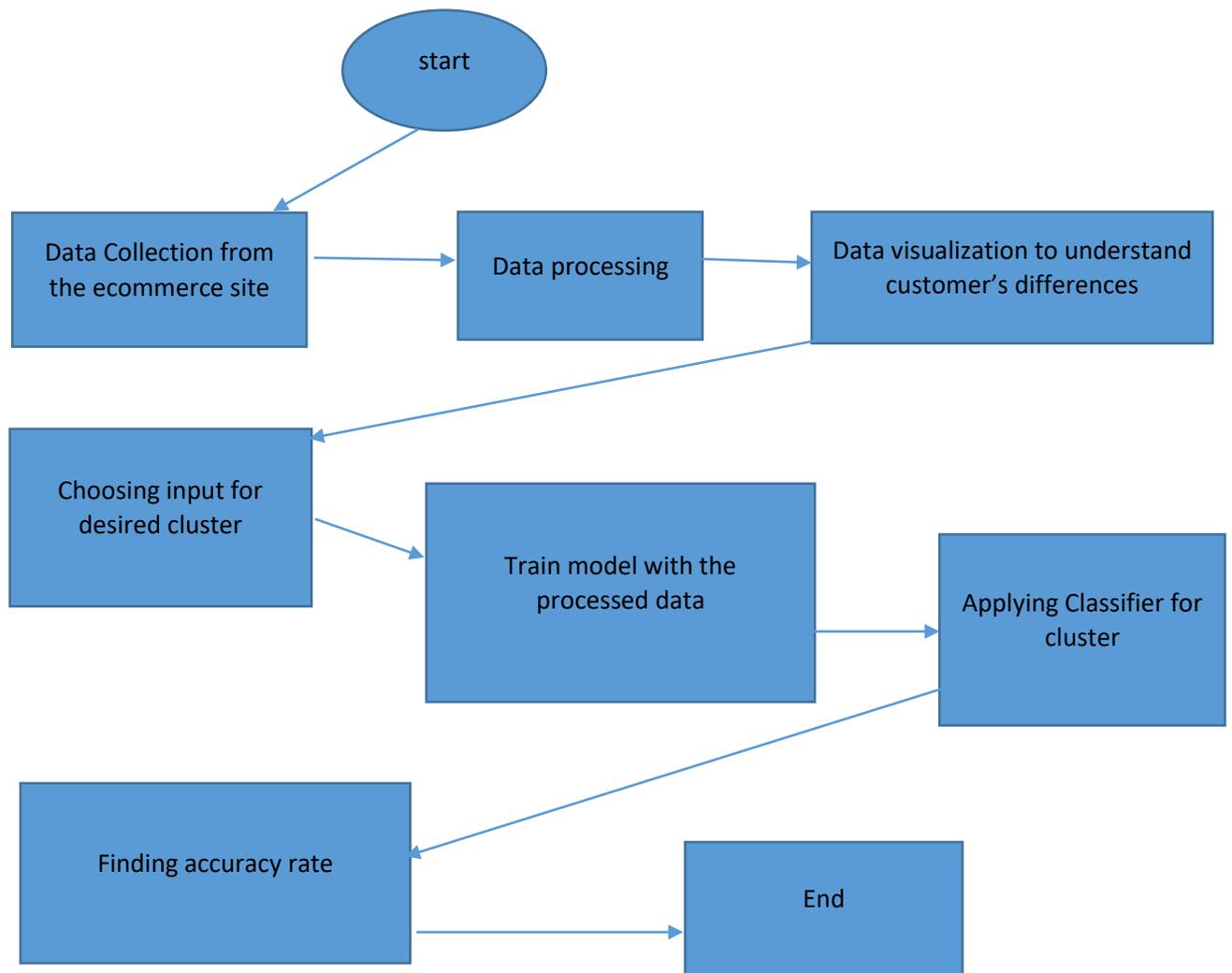


Figure 1: Research Process Flowchart

In the ecommerce site customers will pay in online to we monitor the IP address of every customer. We get some data collecting problems in the process. We also find out the customer interest from the ecommerce site. With the help of ecommerce site try to collect all the data we want

3.3 Data Collection Procedure

We used online shopping mall customer dataset to build our proposed model. For training we used 2000 customer data and 500 data to test our model. We collect our data set online site and e-commerce site by related real market place. But the common problem dataset is not fixed. It is uploading Like- product price, production, Customers demand, customers purchase rate etc.so real market place is to confidential for e-commerce solution.

TABLE I: Dataset

Customer ID	Gender	Age	Annual Income (K TAKA)	Spending score (1-100)
1	Male	19	20	39
2	Male	21	26	81
3	Female	20	52	60
4	Female	23	38	77
5	Female	31	65	100

working process of k-means algorithm

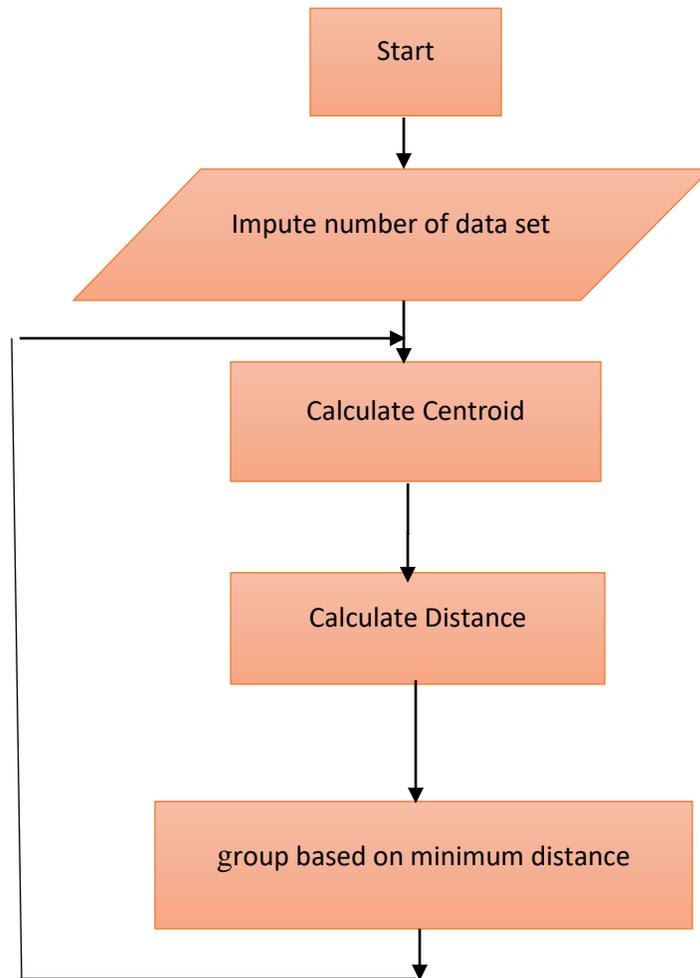


Figure 2: k-means algorithm working Process Flowchart

3.4 Statistical Analysis

Using dummy data is not a wise decision for this research. So we collected shopping mall customer data to train our model. Because when the ecommerce site will run then we will a lot more data to train our model. In this following table we describe all the data label.

TABLE II: Data Description

Column Label	Description
Customer ID	Unique id for every customer
Gender	Gender of every customer
Age	Age of every customer
Annual income	Annual income of every customer
Spending Score	The score is out of 100. How much a customer spend on the ecommerce.



Figure 3: Number of Cluster

From this data we will determine which customer we should put in which group. In the figure we describe how many cluster is happening in every inertia.

We transform our data into CSV file. We use pandas library in python to process our data. In this research data visualization is very important because that's how we can understand the dependencies. So we use matplotlib library in python for data visualization.

3.5 Implementation Requirement

Why it's called the K-Means? Because the letter represent the number of cluster chosen. To a particular observation assign which mean nearest cluster some special function participate the K-Means algorithm for clustering.

At first the initial set of means identify and then the consequent classification based on their distance to the center. Then the cluster means computed again and next the base on their center distances is done reclassification. This is the interaction unit cluster means does not mutation much between continuously repeat. At last, once again calculate the cluster and then all the cases are input the standing cluster [9].

This is iterations until cluster means don't change much between continuously repeat. Finally, the means of the clusters once again calculated and then all the cases are imputed to the permanent clusters. Given a set of observations (x_1, x_2, \dots, x_n) , where each

observation x_i is a d -dimensional real vector. The k -means clustering algorithm target division to partition the n observations into k groups of observations called clusters where $k \leq n$, so as to reduce the sum of squares of distances between a particular cluster within observations.

As shown in Table I, the sum of squares of the distance may be given by the equation $\arg \min S = \sum_{i=1}^n \sum_{j=1}^k |x_j - \mu_j|^2$, where μ_i is the mean of points in S_i . Given an initial set, k -means computes initial means $m_1(1), \dots, m_k(1)$ and it recognize k clusters is given by raw data set[10].

TABLE III: K-Means Algorithm

Simplified simulation flow of <i>k</i> -means algorithm
Begin
Inputs: $X = (x_1, x_2, \dots, x_n)$
Determine: <i>Clusters – k</i>
<i>Initial Centroids - C1, C2, ..., Ck</i>
Assign each input to the cluster with the closest centroid
Determine: <i>Update Centroids - C1, C2, ..., Ck</i>
Repeat: <i>Until Centroids don't change significantly (specified threshold value)</i>
Output: <i>Final Stable Centroids - C1, C2, ..., Ck</i>
<i>End</i>

Maximum K-Means algorithm working process is very slow to come together from different direction so as eventually meet. It takes some time to coverage exponential for mostly perfect condition. Together from different direction it an equable doorstep of specified to quick delivery output without make up much accuracy[11].

As shown in Table II, the Sum of Square of Errors (SSE) may be notably reduced to identify more number of clusters.

It's always desirable to improve SSE without rising number of clusters which is possible due to the fact that k-means converges to a local minimum. To decrease SSE, a cluster may be divide or may be introduced a new cluster of centroid.

TABLE IV: Bisecting of K -Means Algorithm

Bisecting sequence of k -means algorithm
Begin
Initialize clusters
Do:
<i>Remove a cluster from list</i>
<i>Select a cluster and bisect it using k-means algorithm</i>
<i>Compute SSE</i>
<i>Choose from bisected clusters one with least SSE</i>
<i>Add bisected clusters to the list of clusters</i>
Repeat:
<i>Until the number of cluster have been reached to k</i>
End

Increasing SSE, a cluster may be radiate or two clusters may be united. All observation points to obtain k -clusters from a set divide into the observation points are two clusters and again one of these clusters is divide further into two clusters. Initially a cluster of maximum size or a cluster with maximum SSE may be picked for intersection process. This process is periodically until the k numbers of clusters have been emanated[12].

Thus it is freely remarkable that the SSE can be changed by dividing or fully connecting the clusters. This appointed landmark of the k -means clustering is very much important & desirable for marketing segmentation research[13].

The new SSE is again calculate after updating cluster centroid. Without changing this further is repeated until SSE is reached to a minimum value or a condition similar to congruence becomes constant .

The SSE is represented mathematically by $SSE = \sum_{i=1}^k (\mu_i - x)^2$ where x is any point in the same cluster and μ_i is the centroid of i th cluster represented by μ_i and. A condition for achieving minimum SSE can be easily computed by differentiating SSE setting it equal to 0 and then solving the equation:

$$\begin{aligned} \frac{\partial}{\partial \mu_k} SSE &= \frac{\partial}{\partial \mu_k} \sum_{i=1}^k \sum_{x \in c_i} (\mu_i - x)^2 \\ &= \sum_{i=1}^k \sum_{x \in c_i} \frac{\partial}{\partial \mu_k} (\mu_i - x)^2 \\ &= \sum_{x \in c_k} 2 * (\mu_k - x_k) = 0 \\ m_k \mu_k &= \sum_{x \in c_k} x_k \end{aligned}$$

Here m_k is total number of elements and μ_k is centroid in k th cluster c_k . Further it can be simplified as –

$$\mu_k = \frac{1}{m_k} \sum_{x \in c_k} x_k$$

This concludes that the minimum SSE can be achieved under the condition of the centroid of the cluster being equal to the mean of the points in the k^{th} cluster c_k [15].

CHAPTER 4

Experimental Results and Discussion

4.1 Introduction

The study and research outcome is really good. The accuracy rate is the primary result of the experiment. Our model is unsupervised learning and we use K-means algorithm. And our accuracy rate is 94%. Our research is based on classifying the customer's category for the business man to understand their customer well.

We are using our model to segment our ecommerce customer. It is very important to understand our customer for ecommerce business. It's a first part for data science for business intelligence.

4.2 Experimental Results

Our model got 95% accuracy using K-means algorithm. It depends on 5 data of every customers. It will help every business man to understand their basic customers. By the problem analyzing result and compare with the prediction output it's about similar.

4.3 Descriptive Analysis

Thus figure we can see that the accuracy of using machine learning algorithms after implementing the model. The accuracy rate is so good. We see that the figure age, annual income and spending score graph plot, age on based income and spending cluster, annual income and spending cluster, spending score cluster etc.

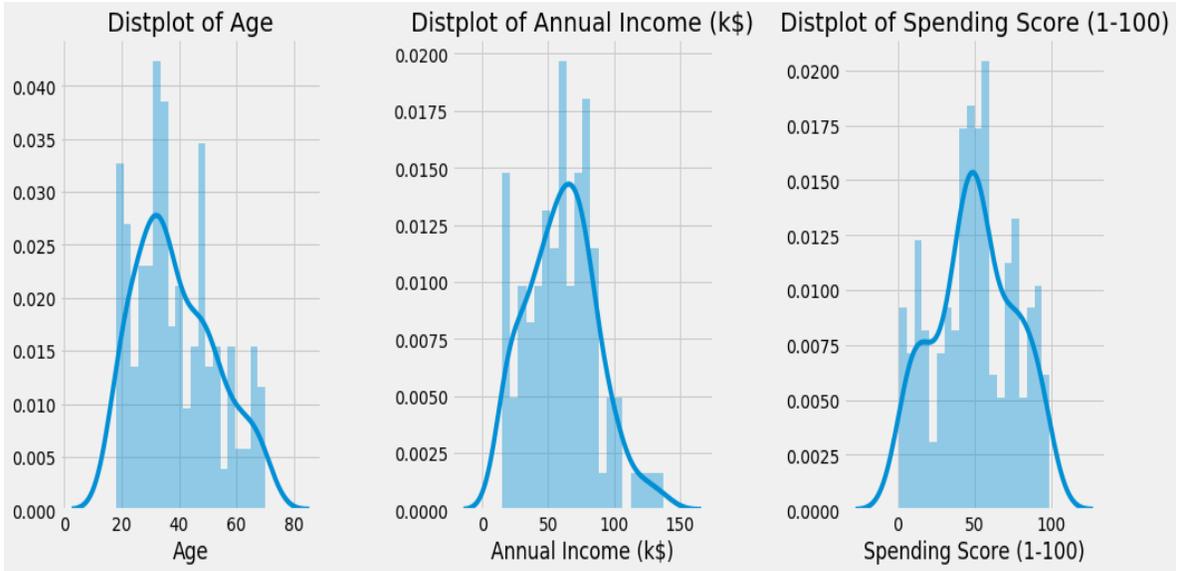


Figure 4: Distplot data according to Age, Annual Income and Spending score.

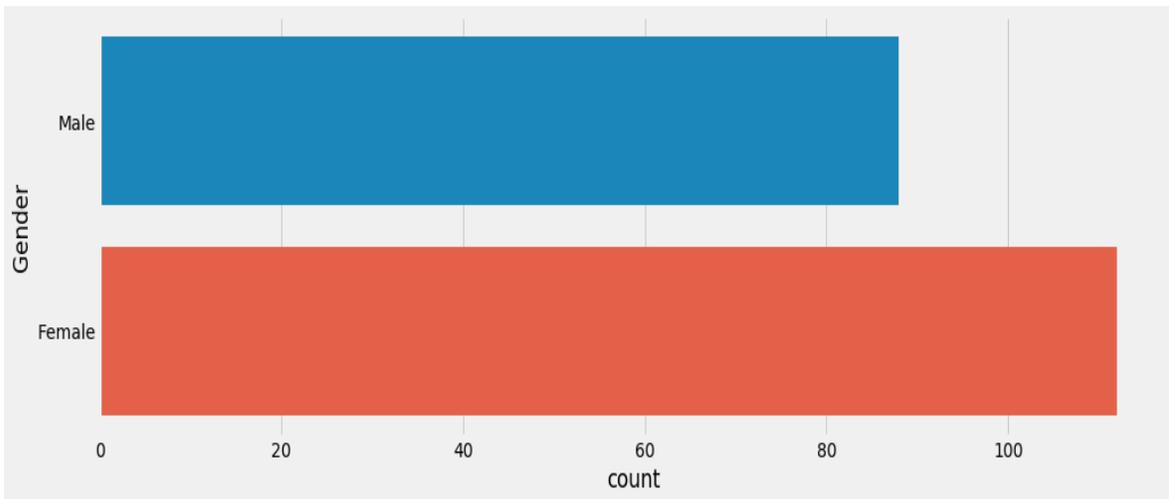


Figure 5: Gender difference

In this figure:5, we see that the purchase rate of two different customer male and female.

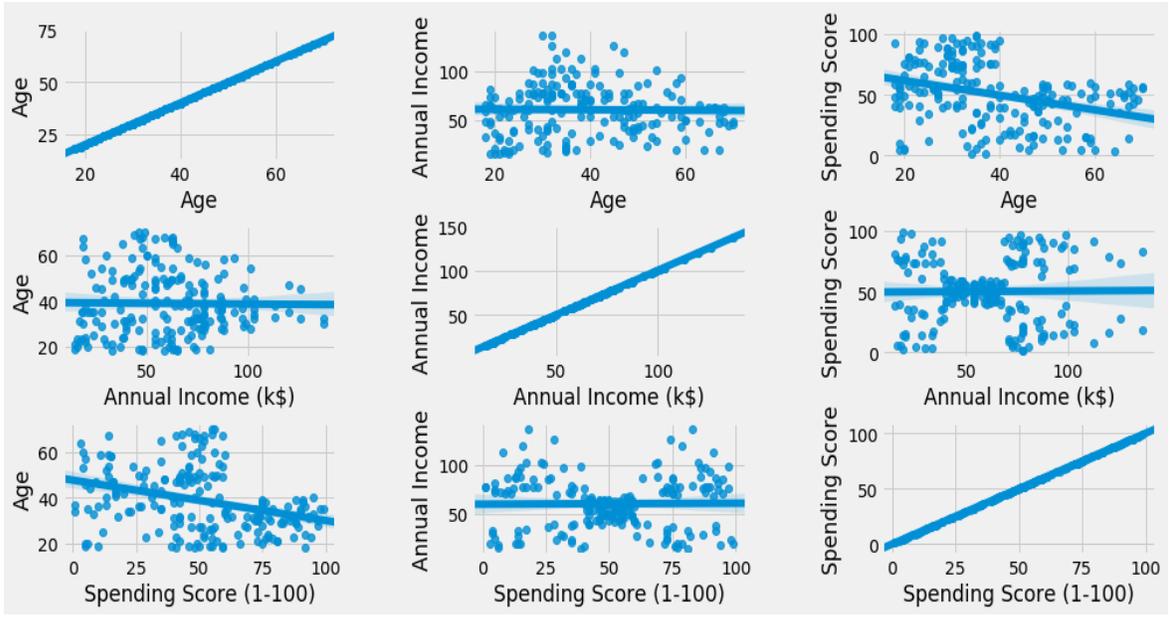


Figure 6: Age based income and spending cluster

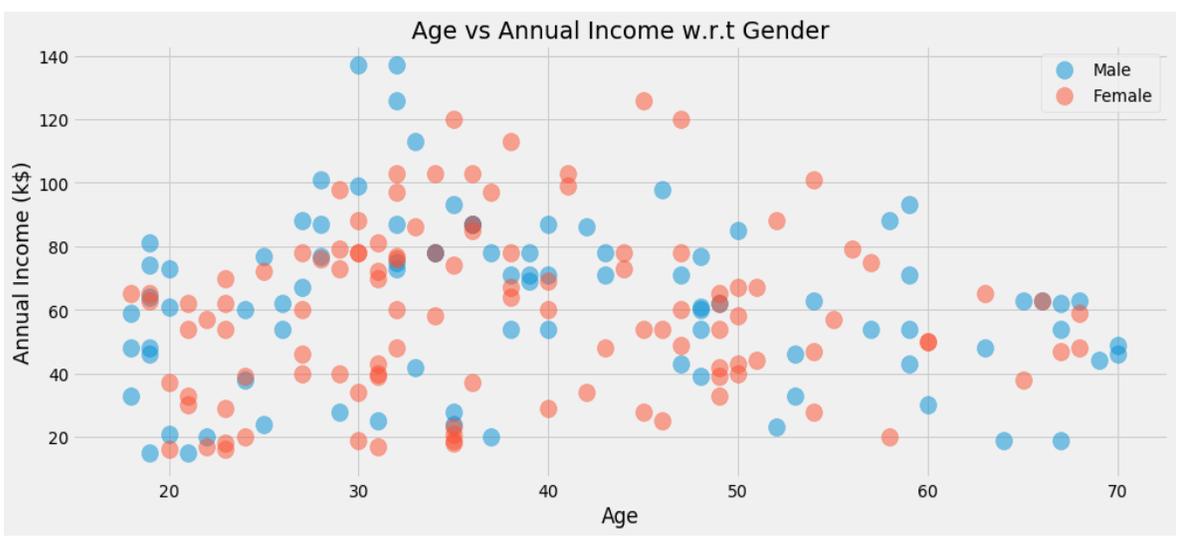


Figure 7: Age vs income w.r.t gender

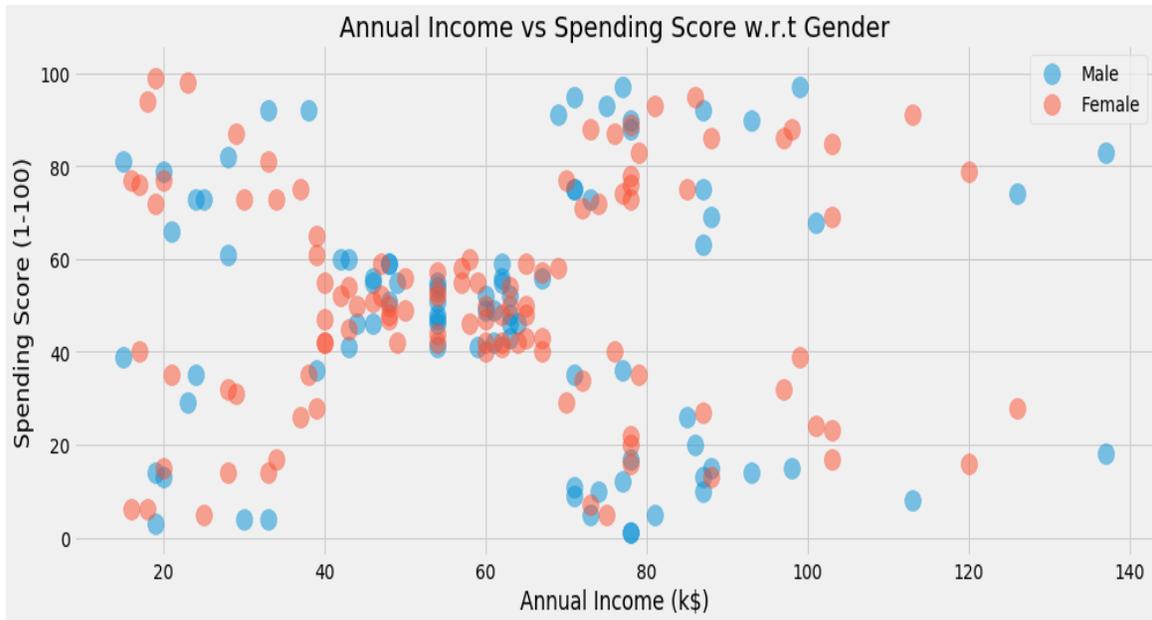


Figure 8: Annual income and spending score

The Figure:8 has given the clustering of the customer by buying on annual income and spending score.

Annual income and spending score is most important thing on this research. Because on based of customers income we can figure out the future situation of business sit. It's good for a customer if she/he have earn a healthy income she/he can parches more. That's why the customers spending score of is highly increase. Which is also good for e-commerce site.

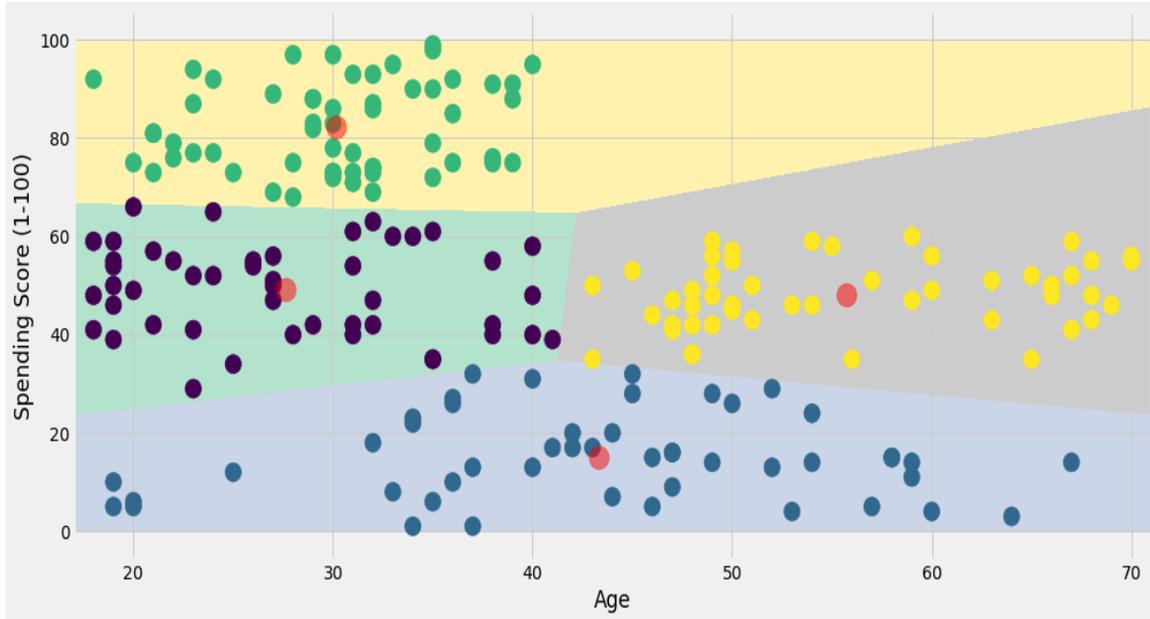


Figure 9: Clustering with the spending score

Finally the figure 9: We see that the clustering of the customer with spending score.

4.4 Summary

From above the result analyses most of the machine learning are perform good but k-means algorithm got the highest probable cluster accuracy rate 94.5%. Other algorithm is not fully perfect for clustering the customer. When we started our work with an online e-commerce site our actual result depends on the amount of accurate data.

Maximum data help the machine to learn more efficiently. Because the machine will know that what will do in the best case scenario.

CHAPTER 5

Summary, Conclusion, Recommendation and Implication for Future

5.1 Summary of the Study

The proposed model gives better performance Over all an annual profit and less an e-commerce site. The validation set timing gives the accuracy and optimizes result in this event. Finally we are able to get the most perfect accuracy by using k-means algorithm. The model allows to get the best result compare to the previous result. Whatever this work will help also our agro based marketing and e-commerce. For an e-commerce Site if it use a system. That could modify that how many profit they earn, how may they purchase and what types of customer demand what types of product. By using this site we can easily segmentation the customer and categorize them.

5.2 Conclusions

If we would get help or core information the use of popular at e-commerce site our default site would become more resource full with expect opinion we will Enrich our default set by our agro based e-commerce site. Then we will implement Unsupervised machine learning and data mining algorithm Like k-means algorithm, clustering etc.

Classifying and clustering algorithm to gain more effective attribute and accuracy. Then we will be able to classify rest of the events in according with the principals of customers. We have a plan to implement the model with an android application and extra features will be able which will be more helpful for our agro based e-commerce and customers.

Overall This segmentation and solution Can be help in various fields like computer at any kind of public, private field. The model will help the entrepreneur who Is involved in digital marketing as also agro based e-commerce.

5.3 Recommendations

Monitoring of this solution is it will only work for tis marketing system. For different kind of system to elegancy of process will cane, also can be to output and also even input due to won't be different. We created a website for cell the products Viewing customers performance and the condition of the site. In future also collecting data from customer. Overall this system will reduce the risk of business because in this site we give that. a predictable approximate output. Our machine will learn the process and give us a probable output which is similar as around the given output. If the data set change the output will be changed.to get highest accuracy We have used unsupervised learning in this research and k-means algorithm and clustering. Which play the main role for segmentation In unsupervised learning there has a special technique for machine learning by using information Where there is no leveling In data set. So we must have to do the segmentation Of the customer fast get e-commerce solution Through the algorithm By machine learning.

5.4 Implication for Further Study

we want to increase our data set more in future we will also focus on the accurate result. We have plan to update our individual in future and want to add more extra features in business purpose. We can implement this model in same e-commerce site predict on analyze. Customers purchase performance and profit rate of e-commerce site annually and monthly by this business site about the draw back to improve their productivity based on profit. This research also helps in many field.

This could be alert for the growers, buyers and also owner which the admin of this site who are very concern their daily shopping to lead their life by based on their income Profit loss, deceit are the vast problem for any business.

We can also predict the possibilities of profit rate and customer demand and e-commerce site provide the customers demand. For thereby no hassle, no risk, no profit loss. We also modify our website and add more fetchers to easy website.

We add our all popular Bangladeshi mobile and online payment system and also add the google tracking map so that we can tress our customer, service provider. If we can more data from others site then we will make our more efficient. To make it more accurate we need lot amount to train our wonder. We can also determine by this kind of approach in future. So we have to implement some new types of criteria same particular field in on data set.

APPENDIX

K-Means Algorithm– K-Means Clustering Algorithm.

SVC – Switch Virtual Circuit.

SVM– Support Vector Machine.

SSE– Sum of Square of Errors.

CSV–Comma-Separated Values.

REFERENCES

- [1] Kim, Jong Tak, Hee-Jun Pan, and Jonghun Kim. "P2Pbased u-health cluster service model for silver generation in PBR platform." *Peer-to-Peer Networking and Applications* . 14 July 2015 ,SpringerScience,Business Media New York 2015.
- [2] Noble, Natasha E., Christine L. Paul, Nicole Turner, Stephen V. Blunden, Christopher Oldmeadow, and Heidi E. Turon. "A cross-sectional survey and latent class analysis of the prevalence and clustering of health risk factors among people attending an Aboriginal Community Controlled Health Service."Noble et al. *BMC Public Health* (2015).
- [3] Alzahrani, Saeed G., Richard G. Watt, Aubrey Sheiham, Maria Aresu, and GeorgiosTsakos. "Patterns of clustering of six health-compromising behaviours in Saudi adolescents" volume14(5), Alzahrani et al. *BMC Public Health* 2014
- [4] Grosskreutz, Henrik, Mario Boley, and Maike KrauseTraudes. "Subgroup discovery for election analysis: a case study in descriptive data mining."volume 6332,In *International Conference on Discovery Science,Discovery Science* pp 57-71.
- [5] Brito, Pedro Quelhas, Carlos Soares, Sérgio Almeida, Ana Monte, and Michel Byvoet. "Customer segmentation in a large database of an online customized fashion business." *Robotics and Computer-Integrated Manufacturing* , Elsevier ,2015.
- [6]] Azizpour, Shahriar, Kay Giesecke, and Gustavo Schwenkler. "Exploring the sources of default clustering."June 15, 2011; this draft February 24, 2014
- [7] Guha, Sudipto, and Nina Mishra. "Clustering data streams:- In *Data Stream Management*". Springer Berlin Heidelberg, 2016.
- [8] Al-Qaed F, Sutcliffe A. Adaptive Decision Support System (ADSS) for B2C E-Commerce. 2006 ICEC Eighth Int Conf Electron Commer Proc NEW E-COMMERCE Innov Conqu Curr BARRIERS, Obs LIMITATIONS TO Conduct Success Bus INTERNET. 2006:492-503.
- [9] Mobasher B, Cooley R, Srivastava J. Automatic Personalization Based on Web Usage Mining. *Commun ACM*. 2000;43(8).
- [10] Cherna Y, Tzenga G. Measuring Consumer Loyalty of B2C e-Retailing Service by Fuzzy Integral: a FANP-Based Synthetic Model. In: *International Conference on Fuzzy Theory and Its Applications iFUZZY.*; 2012:48-56.
- [11] Baer D. CSI: Customer Segmentation Intelligence for Increasing Profits. SAS Glob Forum. 2012:1-13. <http://support.sas.com/resources/papers/proceedings12/1032012.pdf>.

- [12] Chan C, Swatman PMC. Management and business issues for B2B ecommerce implementation. Proc 35th Annu Hawaii Int Conf Syst Sci. 2002;00(c):1-11. doi:10.1109/HICSS.2002.994303.
- [13] Ma H. A Study on Customer Segmentation for E-Commerce Using the Generalized Association Rules and Decision Tree. 2015;(December):813-818.
- [14] Hua S, Xiu S, Leung SCH. Expert Systems with Applications Segmentation of telecom customers based on customer value by decision tree model. Expert Syst Appl. 2012;39(4):3964-3973. doi:10.1016/j.eswa.2011.09.034.
- [15] Birant D. Data Mining Using RFM Analysis. KnowledgeOriented Appl Data Min. 2011;(iii):91-108. doi:10.5772/13683.

CUSTOMER SEGMENTATION BY USING MACHINE LEARNING AND E-COMMERCE SOLUTION BY 162-15-8057 and 162-15-7703

ORIGINALITY REPORT

29%	26%	3%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.ijarcs.info Internet Source	13%
2	www.ijcte.org Internet Source	10%
3	baadalsg.inflibnet.ac.in Internet Source	2%
4	Submitted to University of Kent at Canterbury Student Paper	1%
5	Submitted to Sri Lanka Institute of Information Technology Student Paper	<1%
6	quppler.com Internet Source	<1%
7	Submitted to Queen Mary and Westfield College Student Paper	<1%
8	Submitted to New Zealand School of Education Student Paper	<1%

9	Submitted to Rungta International School Student Paper	<1%
10	Submitted to Oxford Brookes University Student Paper	<1%
11	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	<1%
12	open.library.ubc.ca Internet Source	<1%
13	trap.ncirl.ie Internet Source	<1%
14	Submitted to Hong Kong Baptist University Student Paper	<1%
15	Submitted to CSU, San Jose State University Student Paper	<1%
16	Submitted to City University of Hong Kong Student Paper	<1%
17	Submitted to City University Student Paper	<1%
