A Project Report on

Customer Classification Prediction Model For Online Store

Submitted in partial fulfillment of the requirements for the award of the degree of

Bachelor of Engineering

in

Department of Computer Engineering

by

Arhanti Gawde(16102026)
Arpita Hirlekar(16102018)
Snehal Surve(17202013)

Under the Guidance of

Prof. Pravin Adivarekar

Department of Computer Engineering
A.P. Shah Institute of Technology
G.B.Road,Kasarvadavli, Thane(W), Mumbai-400615
UNIVERSITY OF MUMBAI

Academic Year 2019-2020

Approval Sheet

This Project Report entitled ***"Customer Classification Prediction Model For Online Store"*** Submitted by ***"Arhanti Gawde"(16102026), "Arpita Hirlekar"(16102018), "Snehal Surve"(17202013)*** is approved for the partial fulfillment of the requirenment for the award of the degree of ***Bachelor of Engineering*** in ***Computer Engineering*** from ***University of Mumbai***.

(Prof. Pravin Adivarekar)
Guide

Prof. Sachin Malve
Head Department of Computer Engneering

Place:A.P.Shah Institute of Technology, Thane
Date:

CERTIFICATE

This is to certify that the project entitled ***"Customer Classififation Prediction Model For Online Store"*** submitted by ***"Arhanti Gawde"(16102026), "Arpita Hirlekar"(16102018), "Snehal Surve"(17202013)*** for the partial fulfillment of the requirement for award of a degree ***Bachelor of Engineering*** in ***Computer Engineering***,to the University of Mumbai, is a bonafide work carried out during academic year 2019-2020.

(Prof. Pravin Adivarekar)
Guide

Prof. Sachin Malve
Head Department of Computer Engineering

Dr. Uttam D.Kolekar
Principal

External Examiner(s)

1.

2.

Place:A.P.Shah Institute of Technology, Thane
Date:

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, We have adequately cited and referenced the original sources. We also declare that We have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

_____

(Signature)

_____

Arhanti Gawde (16102026)
Arpita Hirlekar (16102018)
Snehal Surve (17202013)

Date:

Abstract

This paper deals with classification of the customers as a part of the marketing strategy in order to expand the business profit by understanding the relationship between the customers and the products that they purchase. Classification of the customer proves to be helpful since it targets the specific customer for the specific product. This will help the company heighten the level of information to the appropriate customers. On the basis of what customers are buying, adding to cart or viewing, we will cluster the customers depending upon the similar parameters. Also due to this the business managers and analysts will be further able to provide recommendations to the customers belonging to a particular cluster only. While launching new products, services or releasing new versions of existing products / services the business can reach potential customers only. However, a real time big data will be used to perform the analysis on. This model will serve the purpose of identifying the customers who will most likely respond to the recommendations by the company based on their past purchasing history. This system would be useful for the companies in putting a marketing tactic for promotion of their new products. In this manner the business will stay focused and targeted.

Contents

List of Figures

List of Tables

List of Abbreviations

N.A.

Chapter 1

Introduction

Identifying the potential customers of certain products and promoting the products accordingly is an important concept of almost all businesses nowadays. There are two methods for advertisements and promotions, namely mass marketing and direct marketing. Mass marketing is a procedure wherein the promotion of the products is carried out by targeting the general public through media such as newspapers, radios and televisions. This results in high wastage and low response rate from the customers who will actually buy the product. Hence this is not an effective method to be applied in today's competitive market. Therefore rather than going forward with such an unreliable method, the marketers are now shifting their focus from traditional Mass Marketing to Direct Marketing. We are using ELK in our project for identifying different customers on the basis of their past purchases, their time of purchase, brand they purchase, etc. Our project develops predictive analysis using an analysis software known as ELK (Elasticsearch, Logstash and Kibana) in order to predict the bifurcate the customers according to their shopping time, history, choice, etc. This framework is computationally effective and can be connected by numerous degrees of opportunity. Intermediary analyzer demonstrates diverse methods for appearing of an intermediary logs, it will likewise indicate distinctive moves made by client. Principle work of intermediary analyzer is to screen and dissect each log for making report in comprehensible organization on account of it,It will accommodating for human to make further move.

1.1 How it works?

Front end: kibana works as a front end tool, which also shows result in visualized formats of data which analyze with the help of logstash and elasticsearch.

Elasticsearch provides a REST API over multiple indexes that can be searched and queried. Indexes are automatically created when you post a JSON document to an index scheme. The index scheme is composed of 3 parts:
• Index name
• Index type
• Document ID

This is an end-to-end stack that handles everything from data aggregation to data vi- sualization.

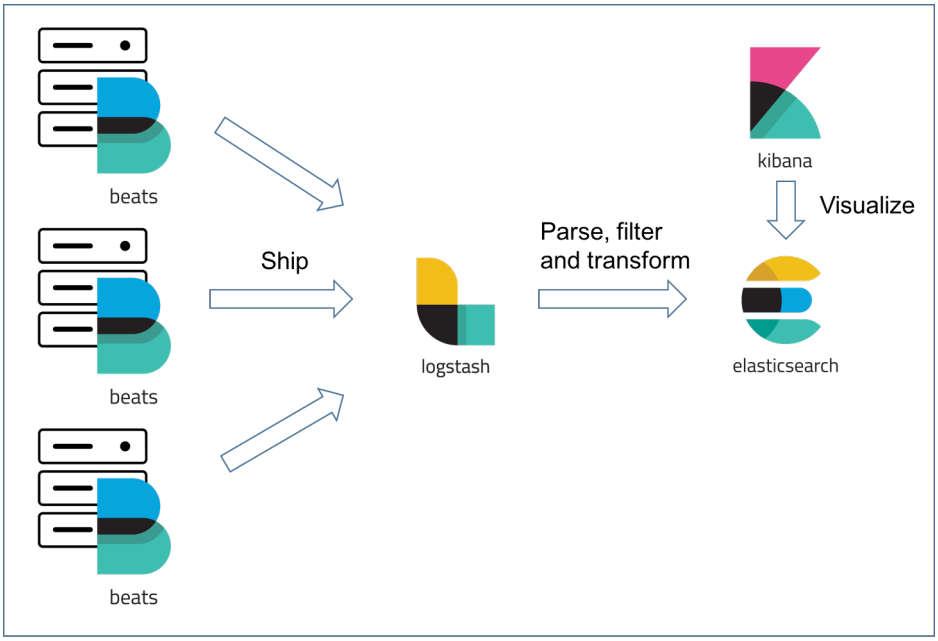


Figure 1.1: Working Overview

1.2 Objective

- The objective of this project is to build a model that will accept a big data and classify customers based on the basis of their purchase history, purchase time, brand, etc. and give the business a good and detailed report of the analysis.
- After customers are classified, the company or the business will not have to face the excess of expenditure in providing irrelevant recommendations for new products that will be launched on the website.
- The company will be benefited because it can send recommendation only to customers who are likely to revert to the recommended product.

1.3 Problem Statement

To generate a model that will accept a big data and classify customers on some constraints and make clusters of customers who portray some or the other common characteristics. Clusters will be having different customers on the basis of their past purchases, their time of purchase, the brand they purchase, etc. Also this model will provide a good and detailed report about the performed analysis. This process of classification is an integral part of the direct marketing scheme which is trending globally nowadays.

1.4 ElasticSearch

1.4.1 What is Elasticsearch?

Elasticsearch is a distributed, open source search and analytics engine for all types of data, including textual, numerical, geospatial, structured, and unstructured. Known for its simple REST APIs, distributed nature, speed, and scalability, Elasticsearch is the central component of the Elastic Stack, a set of open source tools for data ingestion, enrichment, storage, analysis, and visualization. Commonly referred to as the ELK Stack (after Elasticsearch, Logstash, and Kibana), the Elastic Stack now includes a rich collection of lightweight shipping agents known as Beats for sending data to Elasticsearch. Elasticsearch is completely free. The open source features of Elasticsearch are free to use under the Apache 2 license. Additional free features are available under the Elastic license, and paid subscriptions provide access to support as well as advanced features such as alerting and machine learning.

To get started, one should read the API conventions, learn about the different options that can be applied to the calls, how to construct the APIs and how to filter responses. A good thing to remember is that some APIs change and get deprecated from version to version, and its a good best practice to keep tabs on breaking changes.

Following are some API of Elasticsearch :

• Elasticsearch Document API:
This classification of APIs is utilized for taking care of reports in Elasticsearch. Utilizing these APIs, for instance, you can make reports in a file, update them, move them to another record, or evacuate them.
• Elasticsearch Search API:
As its name implies, these API calls can be used to query indexed data for specific information. Search APIs can be applied globally, across all available indices and types, or more specifically within an index. Responses will contain matches to the specific query.
• Elasticsearch Indices API:
This classification of APIs is utilized for taking care of reports in Elasticsearch. Utilizing these APIs, for instance, you can make reports in a file, update them, move them to another record, or evacuate them.
• Elasticsearch Cluster API:
Different, explicit ways. There are types, for instance, that include security usefulness, disclosure components, and examination abilities to Elasticsearch.

Elasticsearch supports a variety of languages and official clients are available for:

- Java
• JavaScript (Node.js)
• Go
• .NET (C)
• PHP
• Perl
• Python

• Ruby
Elasticsearch supports 34 text languages, from Arabic to Thai, and provides analyzers for each.

1.4.2 How does Elasticsearch work?

Raw data flows into Elasticsearch from a variety of sources, including logs, system metrics, and web applications. Data ingestion is the process by which this raw data is parsed, normalized, and enriched before it is indexed in Elasticsearch. Once indexed in Elasticsearch, users can run complex queries against their data and use aggregations to retrieve complex summaries of their data. From Kibana, users can create powerful visualizations of their data, share dashboards, and manage the Elastic Stack.

1.4.3 What is an Elasticsearch index

An Elasticsearch index is a collection of documents that are related to each other. Elasticsearch stores data as JSON documents. Each document correlates a set of keys (names of fields or properties) with their corresponding values (strings, numbers, Booleans, dates, arrays of values, geolocations, or other types of data).

Elasticsearch uses a data structure called an inverted index, which is designed to allow very fast full-text searches. An inverted index lists every unique word that appears in any document and identifies all of the documents each word occurs in.

During the indexing process, Elasticsearch stores documents and builds an inverted index to make the document data searchable in near real-time. Indexing is initiated with the index API, through which you can add or update a JSON document in a specific index.

1.4.4 Why use Elasticsearch?

- Elasticsearch is fast -
  Because Elasticsearch excels at full-text search. Elasticsearch is also a near real-time search platform, meaning the latency from the time a document is indexed until it becomes searchable is very short — typically one second. As a result, Elasticsearch is well suited for time-sensitive use cases such as security analytics and infrastructure monitoring.

- Elasticsearch is distributed by nature
  The documents stored in Elasticsearch are distributed across different containers known as shards, which are duplicated to provide redundant copies of the data in case of hardware failure. The distributed nature of Elasticsearch allows it to scale out to hundreds (or even thousands) of servers and handle petabytes of data.

- Elasticsearch comes with a wide set of features -
  In addition to its speed, scalability, and resiliency, Elasticsearch has a number of powerful built-in features that make storing and searching data even more efficient, such as data rollups and index lifecycle management.

- The Elastic Stack simplifies data ingest, visualization, and reporting -
  Integration with Beats and Logstash makes it easy to process data before indexing into Elasticsearch. And Kibana provides real-time visualization of Elasticsearch data as well as UIs for quickly accessing application performance monitoring (APM), logs, and infrastructure metrics data.

1.5 Logstash

1.5.1 What is Logstash

Logstash is a free and open server-side data processing pipeline that ingests data from a multitude of sources, transforms it, and then sends it to your favorite "stash." Logstash dynamically ingests, transforms, and ships your data regardless of format or complexity.
Logstash, one of the core products of the Elastic Stack, is used to aggregate and process data and send it to Elasticsearch. Logstash is an open source, server-side data processing pipeline that enables you to ingest data from multiple sources simultaneously and enrich and transform it before it is indexed into Elasticsearch.

Logstash is the "L" in the ELK Stack — the world's most popular log analysis platform and is responsible for aggregating data from different sources, processing it, and sending it down the pipeline, usually to be directly indexed in Elasticsearch.
Logstash gathers logs and occasions from different sources like HDFS, MySql, logs(framework logs, application logs, organize logs), twitter and so forth It changes the information and sends to the Elasticsearch database. In the meantime Logstash utilizes various data sources, channels and yield modules. It changes the crude information dependent on indicated channels in its design record. Logstash has begun the pipeline among Elasticsearch and Logstash and after that parsing the information to Elasticsearch has begun. On the off chance that we need to imagine the information, we will utilize Kibana, the representation apparatus.

Because of some intrinsic execution issues and configuration defects, Logstash has gotten an average measure of protests from clients throughout the years. Side undertakings were created to lighten a portion of these issues (for example Logger, Logstash-Forwarder, Beats), and elective log aggregators started contending with Logstash.
Yet despite these flaws, Logstash still remains a crucial component of the stack. Big steps have been made to try and alleviate these pains by introducing Beats and improvements to Logstash itself, ultimately make logging with ELK much more reliable than what it used to be.

1.5.2 What is Logstash used for?

Logstash, one of the core products of the Elastic Stack, is used to aggregate and process data and send it to Elasticsearch. Logstash is an open source, server-side data processing pipeline that enables you to ingest data from multiple sources simultaneously and enrich and transform it before it is indexed into Elasticsearch.

1.5.3 Logstash Inputs

Data is often scattered or siloed across many systems in many formats. Logstash supports a variety of inputs that pull in events from a multitude of common sources, all at the same time.
One of the things that makes Logstash so powerful is its ability to aggregate logs and events from various sources. Using more than 50 input plugins for different platforms, databases and applications, Logstash can be defined to collect and process data from these sources and send them to other systems for storage and analysis.
The most common inputs used are file, beats, syslog, http, tcp, ssl (recommended), udp, stdin but you can ingest data from plenty of other sources.
Inputs are the starting point of any configuration. If you do not define an input, Logstash will automatically create a stdin input. Since you can create multiple inputs, it's important to type and tag them so that you can properly manipulate them in filters and outputs.

1.5.4 Logstash Filters

As data travels from source to store, Logstash filters parse each event, identify named fields to build structure, and transform them to converge on a common format for more powerful analysis and business value.
Logstash supports a number of extremely powerful filter plugins that enable you to manipulate, measure, and create events. It's the power of these filters that makes Logstash a very versatile and valuable tool.

Logstash dynamically transforms and prepares your data regardless of format or complexity:
•Derive structure from unstructured data with grok
•Decipher geo coordinates from IP addresses
•Anonymize PII data, exclude sensitive fields completely
•Ease overall processing, independent of the data source, format, or schema

The possibilities are endless with the rich library of filters and versatile Elastic Common Schema.

1.5.5 Logstash Outputs

While Elasticsearch is our go-to output that opens up a world of search and analytics possibilities, it's not the only one available. Logstash has a variety of outputs that let you route data where you want, giving you the flexibility to unlock a slew of downstream use cases.

As with the inputs, Logstash supports a number of output plugins that enable you to push your data to various locations, services, and technologies. You can store events using outputs such as File, CSV, and S3, convert them into messages with RabbitMQ and SQS, or send them to various services like HipChat, PagerDuty, or IRC. The number of combinations of inputs and outputs in Logstash makes it a really versatile event transformer.
Logstash events can come from multiple sources, so it's important to check whether or not an event should be processed by a particular output. If you do not define an output, Logstash

will automatically create a stdout output.
Logstash has a pluggable framework featuring over 200 plugins. Mix, match, and orchestrate different inputs, filters, and outputs to work in pipeline harmony.

1.5.6 Logstash Configuration

Logstash configuration is one of the biggest obstacles users face when working with Logstash. While improvements have been made recently to managing and configuring pipelines, this can still be a challenge for beginners.
To configure Logstash, you create a config file that specifies which plugins you want to use and settings for each plugin. You can reference event fields in a configuration and use conditionals to process events when they meet certain criteria. When you run logstash, you use the -f to specify your config file.

1.6 Kibana

1.6.1 What is Kibana?

Kibana is a free and open user interface that lets you visualize your Elasticsearch data and navigate the Elastic Stack. Kibana is an opensource visualization tool which provides a beautiful web interface to visualize the Elasticsearch data. Kibana allows us to create real-time dashboards in browser based interfaces. Kibana has different visualization effects like bar charts, graphs, pie charts, maps, tables etc. It allows to save, edit, delete and share the dashboards.

Totally open source, Kibana is a program based UI that can be utilized to seek, break down and imagine the information put away in Elasticsearch records (Kibana can't be utilized related to different databases). Kibana is particularly prestigious and well known because of its rich graphical and perception abilities that enable clients to investigate vast volumes of information.
Kibana can be introduced on Linux, Windows and Mac utilizing .zip or tar.gz, archives or on Docker. Kibana keeps running on node.js, and the establishment bundles come worked in with the required pairs.

1.6.2 Kibana Searches

Searching Elasticsearch for specific log message or strings within these messages is the bread and butter of Kibana. In Kibanas query bar, you can enter Lucene query syntax or searches based on Elasticsearch Query DSL. Once entered, the main display area will filter the data displayed accordingly, showing matches in reverse chronological order.

Kibana querying is an art unto itself, and there are various methods for performing searches on your data. Here are the most common search types:
• Free text searches used for quickly searching for a specific string.
• Field-level searches used for searching for a string within a specific field.

• Logical statements used to combine searches into a logical statement.
• Proximity searches used for searching terms within a specific character proximity.

1.6.3 What is Kibana used for?

Kibana is a data visualization and management tool for Elasticsearch that provides real-time histograms, line graphs, pie charts, and maps. Kibana also includes advanced applications such as Canvas, which allows users to create custom dynamic infographics based on their data, and Elastic Maps for visualizing geospatial data.

1.7 Process Analysis

•Testing Analyzers : Especially when you are new to Elasticsearch, it is sometimes difficult to understand what is actually being tokenized and stored into your index. To better understand what is going on, you can use the analyze API to see how text is analyzed The token is the actual term that will be stored in the index. The position indicates the order in which the terms appeared in the original text. The start offset and end offset indicate the character positions that the original word occupied in the original string.
•Specifying Analyse : When Elasticsearch detects a new string field in your documents, it automatically configures it as a full-text string field and analyzes it with the standard analyzer. You dont always want this. Perhaps you want to apply a different analyzer that suits the language your data is in. And sometimes you want a string field to be just a string fieldto index the exact value that you pass in, without any analysis, such as a string user ID or an internal status field or tag.

Chapter 2

Literature Review

- Title: A Comparison of Different Classification Techniques for Bank Direct Marketing

  Author: K Wisaeng

  About The Paper: In this paper, the comparison of different classification techniques in open source data mining software which consists of a decision tree methods and machine learning for a set of bank direct marketing dataset are presented. All decision tree methods tested are J48-graft and LAD tree while machine learning tested are radial basis function network and support vector machine. The experiment results show are about classification sensitivity, specificity, accuracy, mean absolute error and root mean squared error.

  Lad Tree: Logical Analysis of data is a two class learning algorithm which integrates principles of combinatorics , Optimization and the theory of Boolean function.

- Title: Bank Direct Marketing Based on Neural Network

  Author: Hancy.A.Elsalamony, Alaa.M.Elsayad

  About The Paper: This paper introduces applications of recent and important models of data mining; Multilayer perceptron neural network (MLPNN) and Ross Quinlan new decision tree model (C5.0). The objective is to examine the performance of MLPNN and C5.0 models on a real-world data of bank deposit subscription. The experimental results demonstrate, with higher accuracies, the success of these models in predicting the best campaign contact with the clients for subscribing deposit. The performances are measured by three statistical measures; classification accuracy, sensitivity, and specificity.

  Multilayer Perceptron Neural Network: A multilayer perceptron (MLP) is a class of feedforward artificial neural network. An MLP consists of at least three layers of nodes: an input layer, a hidden layer and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP utilizes a supervised learning technique called backpropagation for training. Its multiple layers and non-linear activation distinguish MLP from a linear perceptron. It can distinguish data that is not linearly separable.

- Title: Data Mining Framework for Direct Marketing

Author: Lilian Sing'oei, Jiayang Wang

About The Paper: The objective of this paper is to provide a comprehensive framework to guide research efforts focusing on direct marketing strategy and aid practitioners in their quest to achieve direct marketing success using data mining methods like the Decision Tree Algorithm. The framework builds on the literature from direct marketing concepts and data mining methods that provides a systematic approach to users who have little knowledge in data mining in order to carry out effective marketing campaigns. A case study on bank marketing campaigns was used for evaluating the feasibility of the framework.

Decision Tree: Decision Tree algorithm belongs to the family of supervised learning algorithms, decision tree algorithm are used for solving regression and classification problems.

- Title: A Data Mining Based Response Model for Targeting Selection in Directing Marketing

  Author: Eniafe Festus Ayetrian, Adesesan Barnabas Adeyemo

  About The Paper: In this paper, using historical purchase data, a predictive response model with data mining techniques was developed to predict the probability that a customer in Ebedi Microfinance bank will respond to a promotion or an offer. To achieve this purpose, a predictive response model using customers' historical purchas data was built with data mining techniques. The data were stored in a data warehouse to serve as management decision support system. The response model was built from customers' historic purchases and demographic dataset. Bayesian algorithm precisely Naive Bayes algorithm was employed in constructing the classifier system. Both filter and wrapper feature selection techniques were employed in determining inputs to the model.

  Naïve Bayes: Naïve Bayes algorithm is a probabilistic classifier. It splits the attributes into independent ones. This model is easy to build and particularly useful for very large data sets.

- Title: A customer classification prediction model based on machine learning techniques

  Author: T. K. Das

  In this paper, the author has basically tried out experimenting the mining of data using various methods like Naïve Bayes, KNN and SVM to classify a bank customer data. Subsequently, he attempted to compare the effictiveness of all these techniques to find out which one produces the maximum accuracy for the existing data set.

  Naïve Bayes: Naïve Bayes algorithm is a probabilistic classifier. It splits the attributes into independent ones. This model is easy to build and particularly useful for very large data sets. Also, KNN refers to a non-parametric method that produces an algorithm for the process of classification and regression.

- Title: Decision Tree and Naïve Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing

Author: Masud Karim, Mohammad Rashedur

In this work, the author has compared the working of different types of data mining techniques known as the Naive Bayes and the C4.5 decision tree algorithms. The goal of this work is to predict whether a client will subscribe a term deposit. Publicly available UCI data is used to train and test the performance of the algorithms. Besides, they also extract actionable knowledge from decision tree that focuses to take interesting and important decision in business area.

- Title: Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology

  Author: Sérgio Moro, Paulo Cortez, Raul Laureano

  This paper makes use of use of Business Intelligence (BI) and Data Mining (DM) techniques to enhance the campaigns that have over the time reduced it's importance and also the economical pressures and competition has led marketing managers to invest on directed campaigns with a strict and rigorous selection of contacts.Real-world data were collected from a Portuguese marketing campaign related with bank deposit subscription. The business goal is to find a model that can explain success of a contact, i.e. if the client subscribes the deposit. Such model can increase campaign efficiency by identifying the main characteristics that affect success, helping in a better management of the available resources (e.g. human effort, phone calls, time) and selection of a high quality and affordable set of potential buying customers.

Chapter 3

Use Case Diagram and Class Diagram

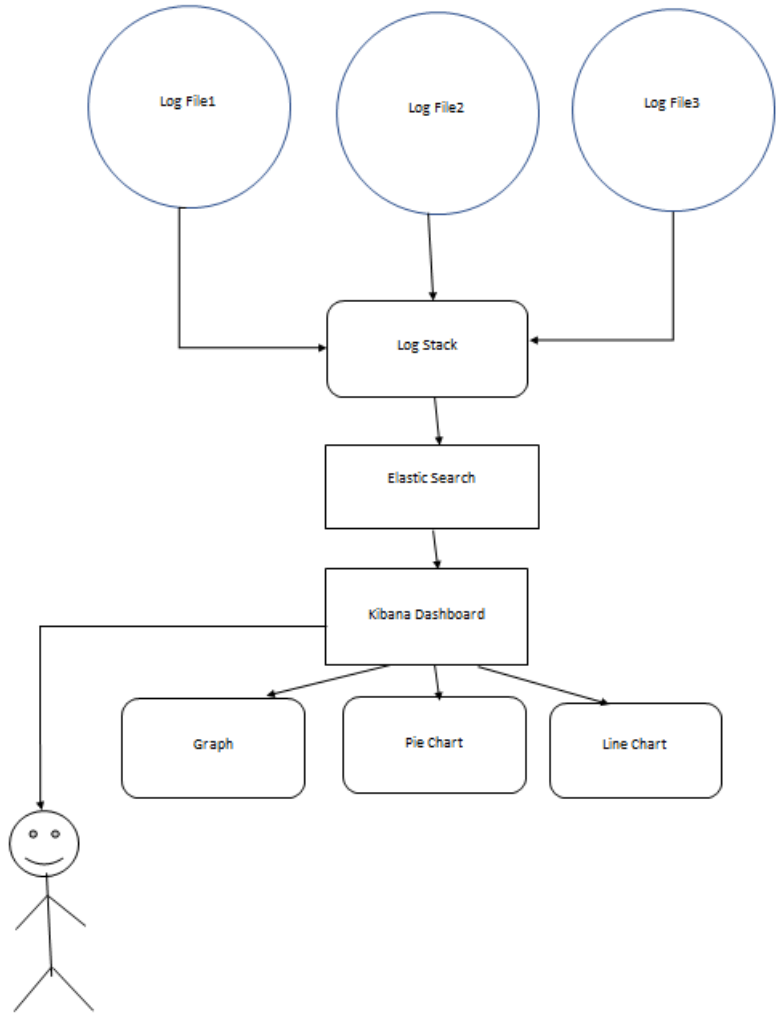


Figure 3.1: Use Case Diagram

- Log File - This files contain the data or information . There can be a number of log files.

- Logstash - It collects the log files data and transfer it to elastic search for the further processing.
- Elastic Search- Elasticsearch is the central component of the Elastic Stack, a set of open source tools for data ingestion, enrichment, storage, analysis, and visualization.
- Kibana Dashboard- All the visualized and managed data are represented in the form of graph, Pie-chart, Line chart, Histogram etc.
- User- The user is able to search and analyze the data . Here the user also can use this for many other purpose like monitoring, security analytics, business analytics etc.

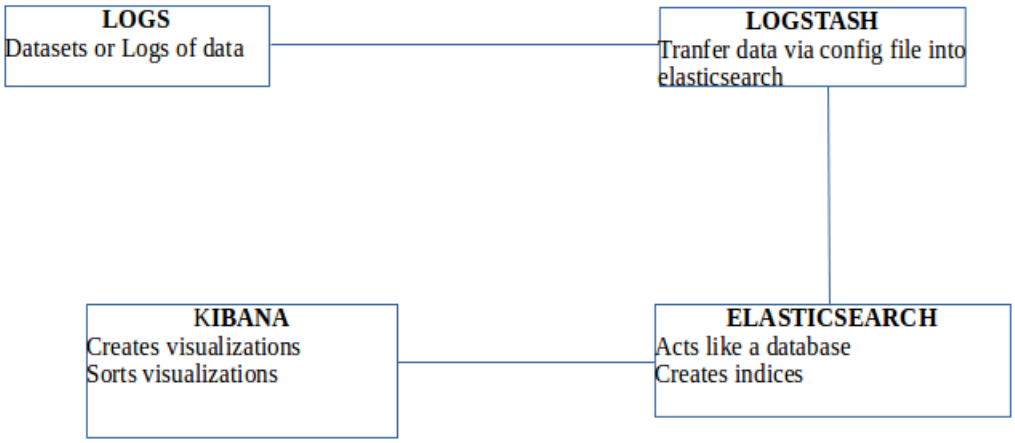


Figure 3.2: Class Diagram

Chapter 4

Implementation

4.1 Installation of ELK

The ELK Stack can be installed using a variety of methods and on a wide array of different operating systems and environments. ELK can be installed locally, on the cloud, using Docker and configuration management systems like Ansible, Puppet, and Chef. The stack can be installed using a tarball or .zip packages or from repositories.
Many of the installation steps are similar from environment to environment and since we cannot cover all the different scenarios, we will provide an example for installing all the components of the stack Elasticsearch, Logstash, Kibana, and Beats on Linux.

Following are steps to install ELK (Elasticsearch, Kibana , Logtsash) :

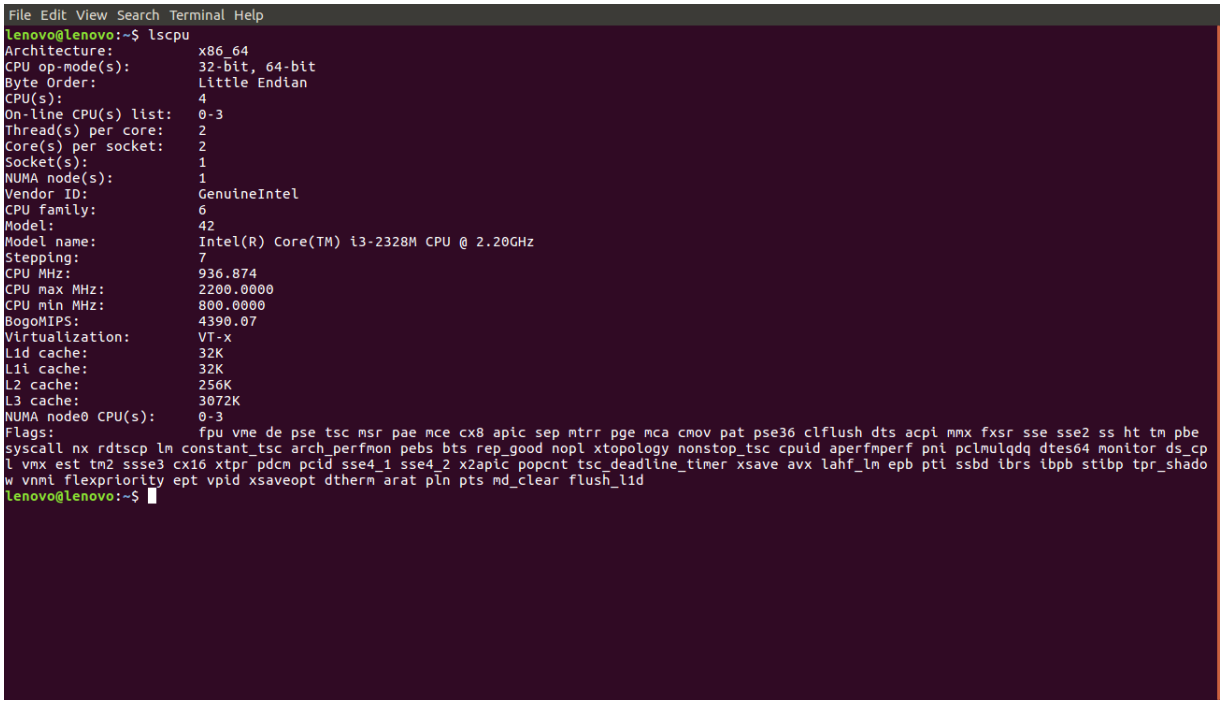


Figure 4.1: Check the Ubuntu System Architecture using the command

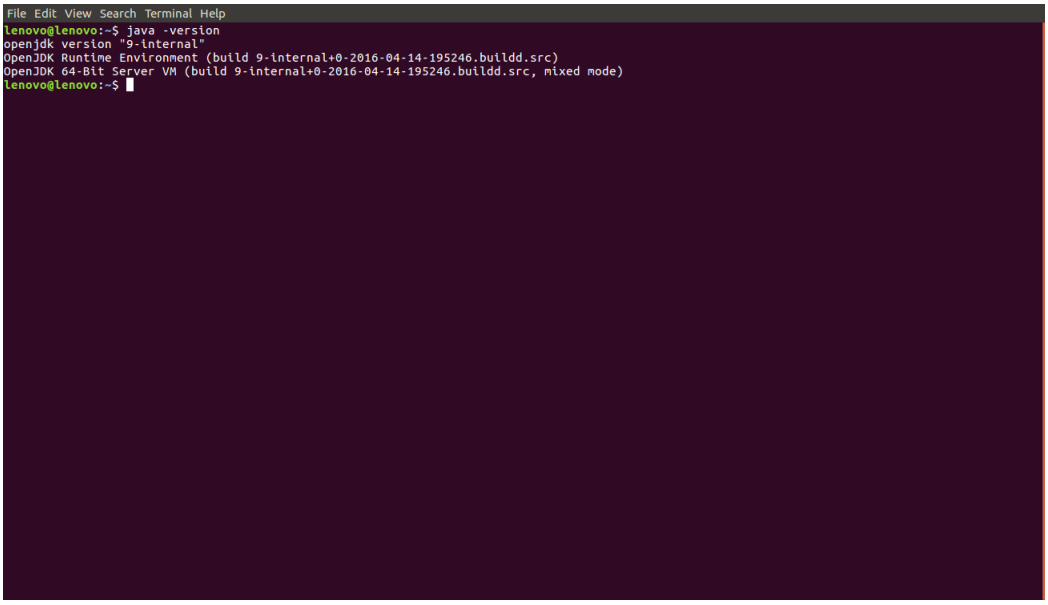


Figure 4.2: Check the Java Version as shown above

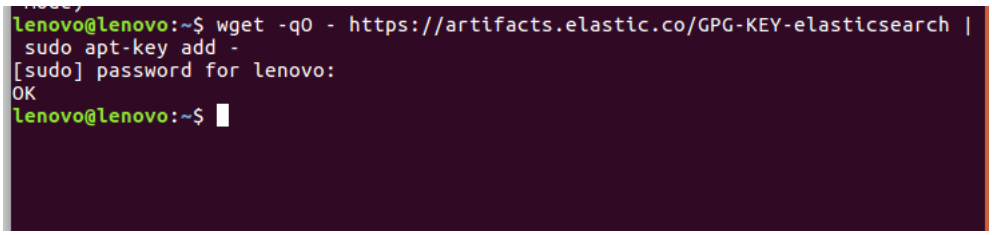


Figure 4.3: Importing Elasticsearch public GPG key into APT

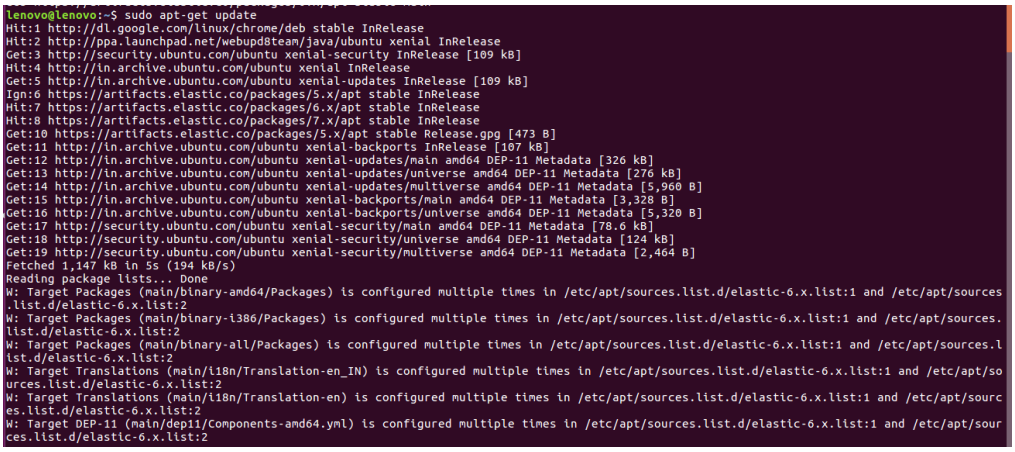


Figure 4.4: Updating the package lists so that APT can read the new Elastic source

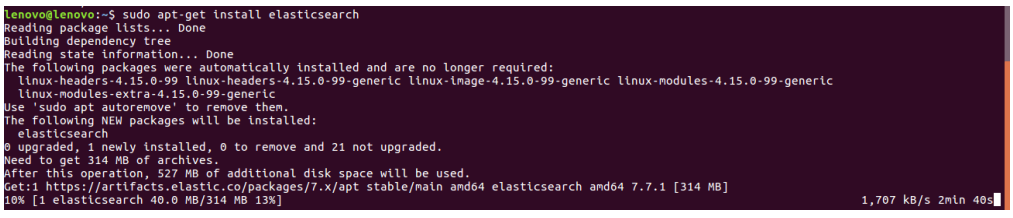


Figure 4.5: Installing Elasticsearch

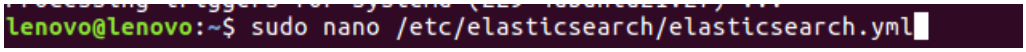


Figure 4.6: To edit Elasticsearch's main configuration file

Figure 4.7: Uncomment network host and http port

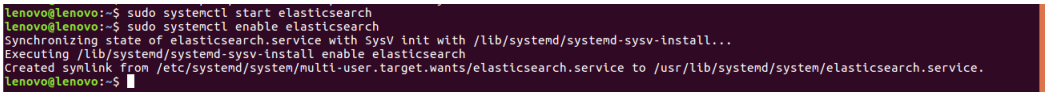


Figure 4.8: Start Elasticsearch

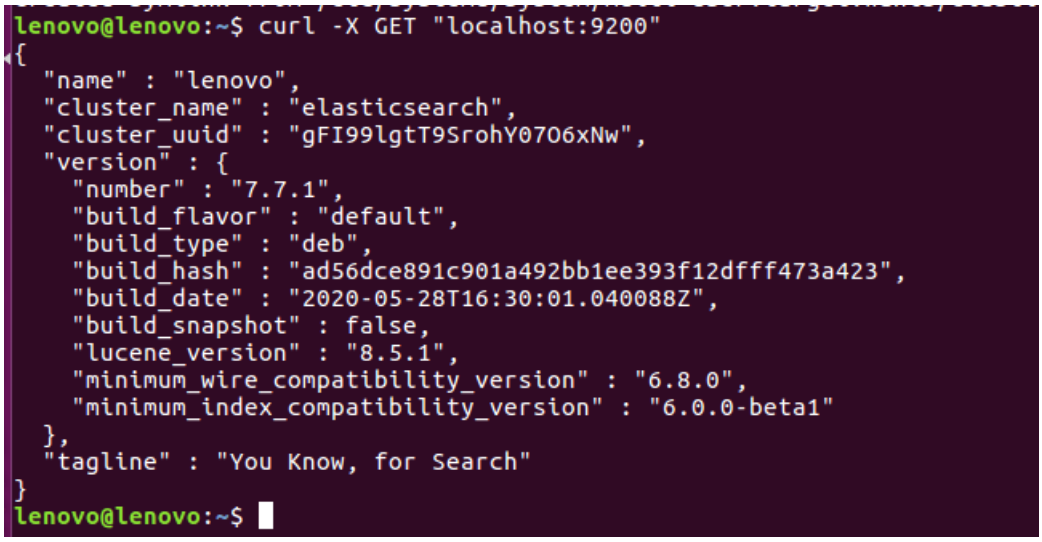


Figure 4.9: Testing the Elasticsearch

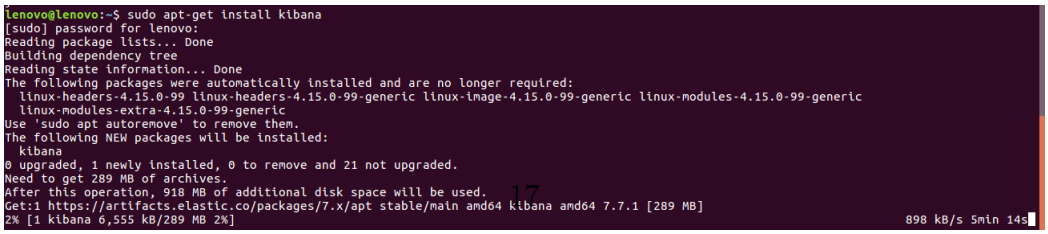


Figure 4.10: Installing Kibana

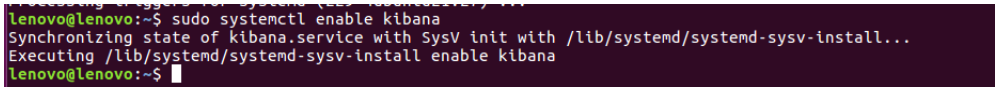


Figure 4.11: Enable Kibana Services

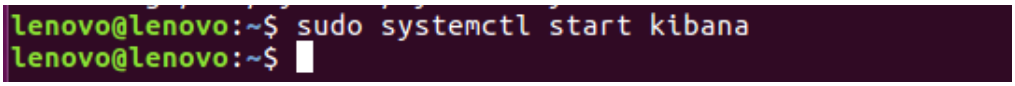


Figure 4.12: Start Kibana Services

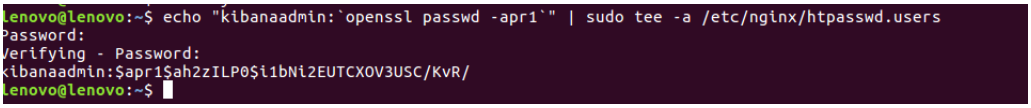
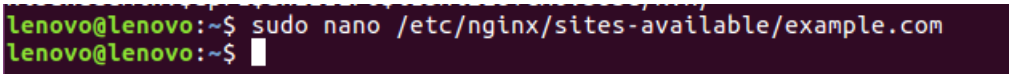


Figure 4.13: Create an administrative Kibana user



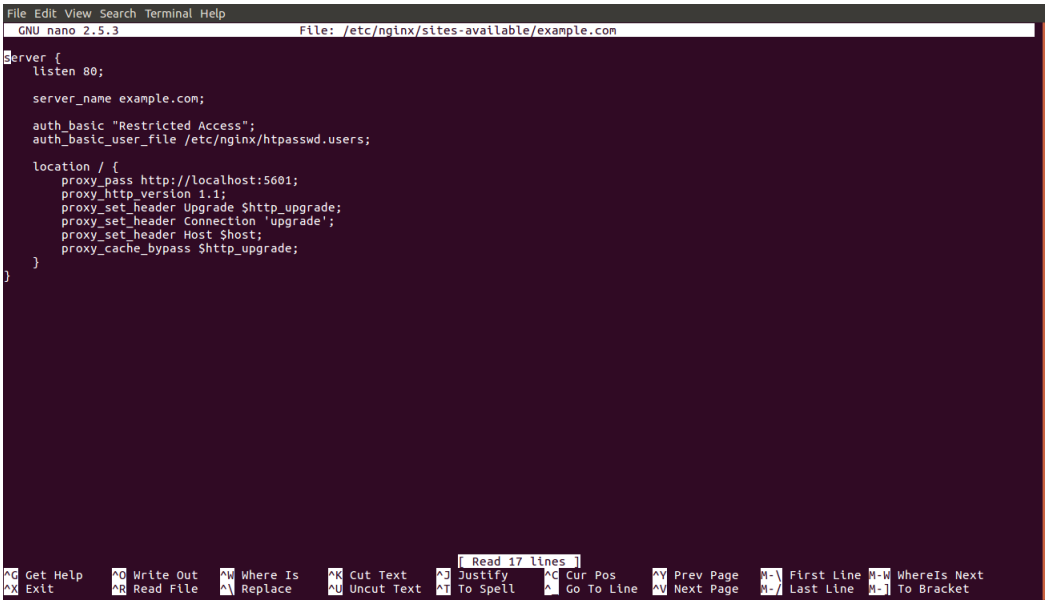


Figure 4.14: Check the Nginx

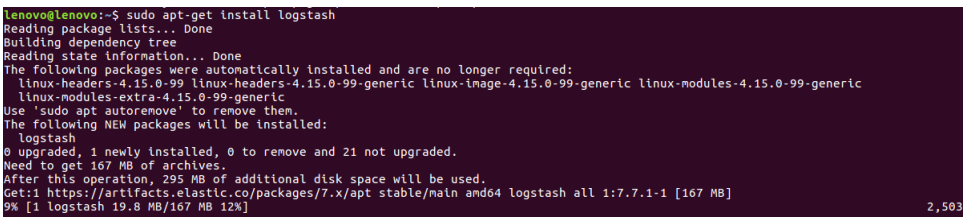


Figure 4.15: Installing Logstash

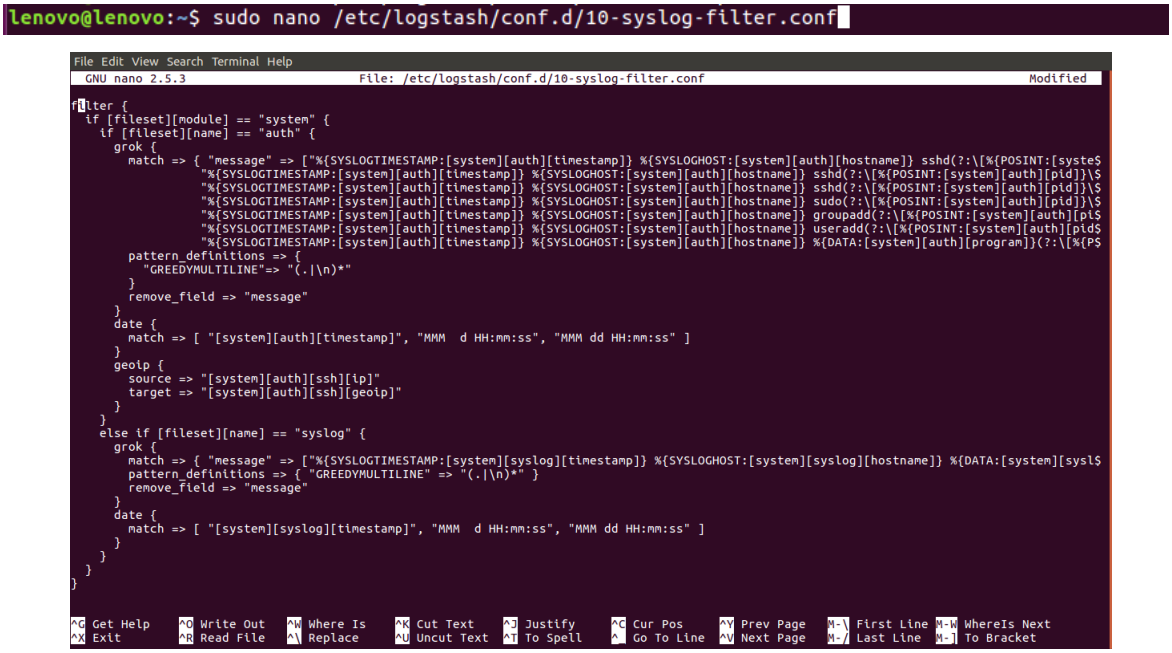


Figure 4.16: Add a Filter to the system logs

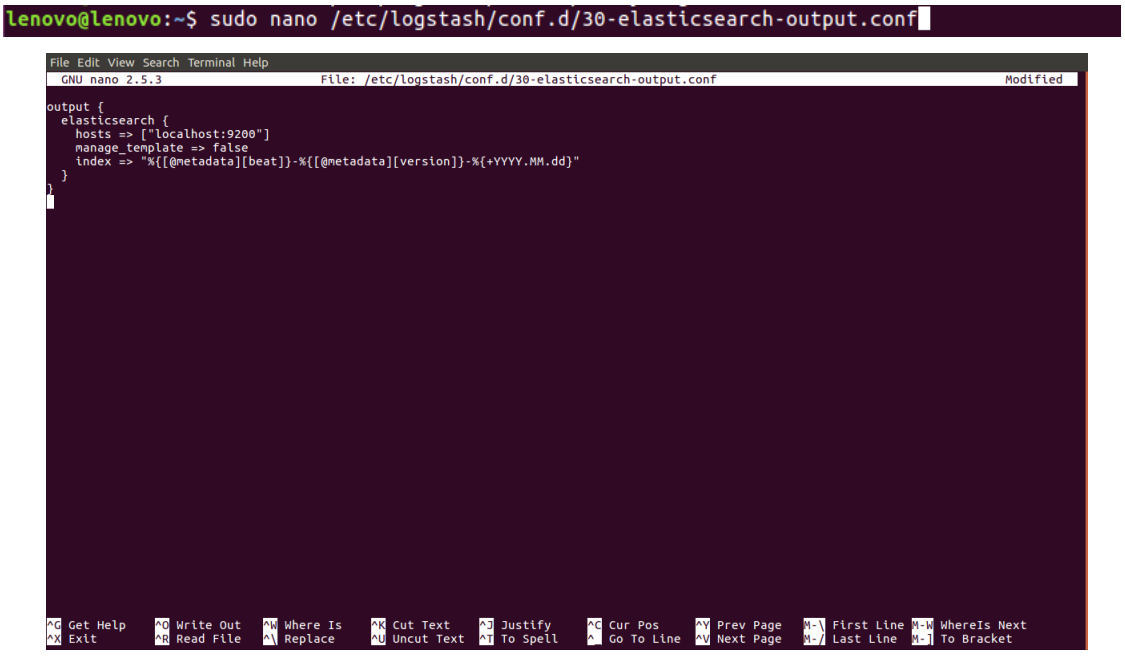


Figure 4.17: Create a config file

4.2 Implementation of Project

Utilizing various input, filter, and output modules, Logstash empowers the simple change of different occasions. In any event, Logstash needs an input and a output module determined in its configurational document to perform transactions. In our project too we have created one log file are as below. In this file are various types of fields :

```
Open                                                                                                    Save
input {
        file {
                path => "home/Desktop/flipkart_com-ecommerce_sample.csv"    //path of our csv file
                start_position => "beginning"
              }
      }

filter {
        csv {
                separator => ","
                columns => ["uniq_id","crawl_timestamp","product_url","product_name","product_category_tree","pid","retail_price",
                            "discounted_price","image","is_FK_Advantage_product","description","product_rating","overall_rating",
                            "brand","product_specifications"]
              }
        }
output {
        elasticsearch {
                        hosts => "http://localhost:9200"
                        index => "customer"
                      }
        }
stdout{}
}
Plain Text   Tab Width: 8   Ln 23, Col 2   INS
```

Figure 4.18: Config File

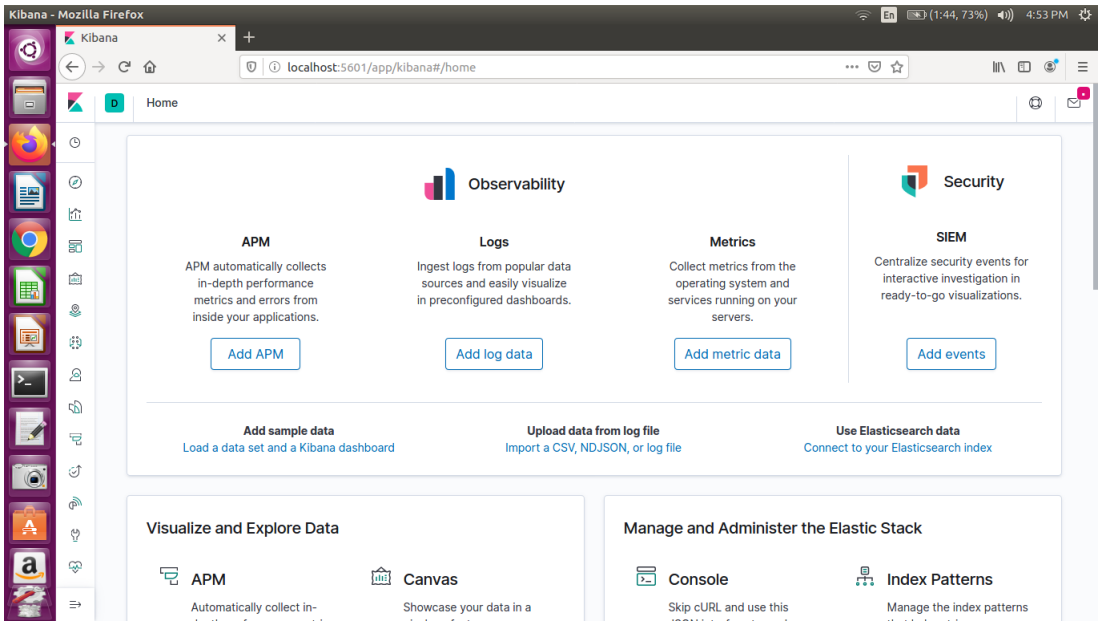


Figure 4.19: Kibana Home Page

The Kibana Visualize page is the place you can make, change, and view your own custom filterations and visualization. There are a few distinct kinds of perceptions, extending from Vertical bar and Pie outlines to Tile maps (for showing information on a guide) and Data tables. Representations can likewise be imparted to different clients who approach your Kibana occasion. In the event that this is your first time utilizing Kibana representations, you should reload your field list before continuing. Guidelines to do this are canvassed in the Reload Field Data subsection, under the Kibana Settings area. To create a visualization, first, click the Visualize menu item.Decide which type of visualization you want, and select it.

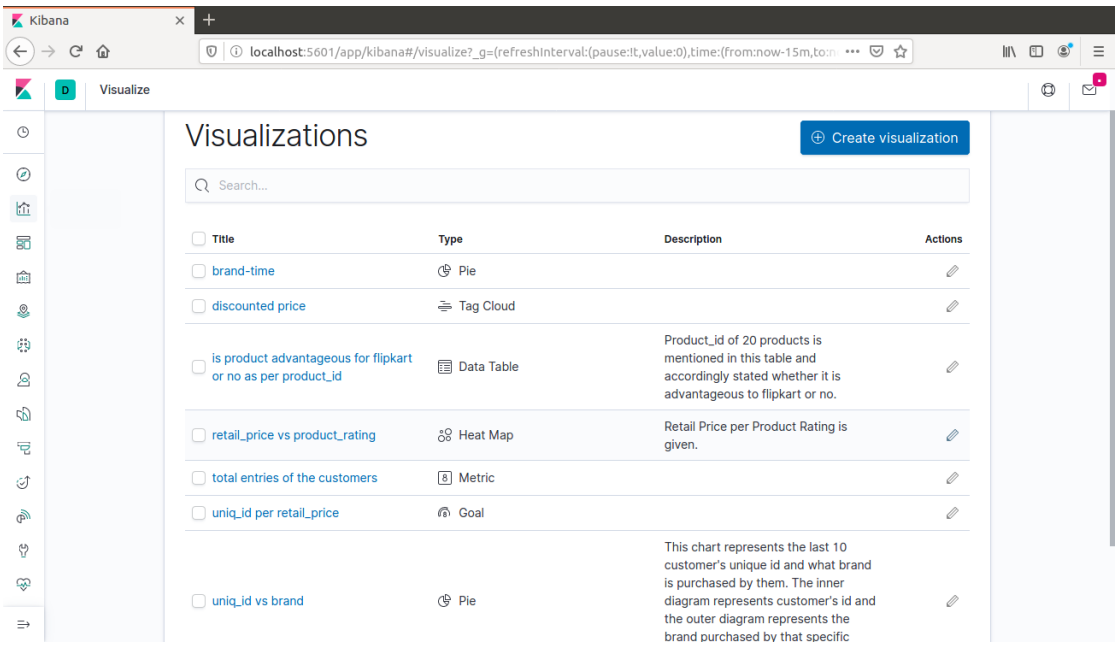


Figure 4.20: Visualizations page

Following are the options that you get in visualization :

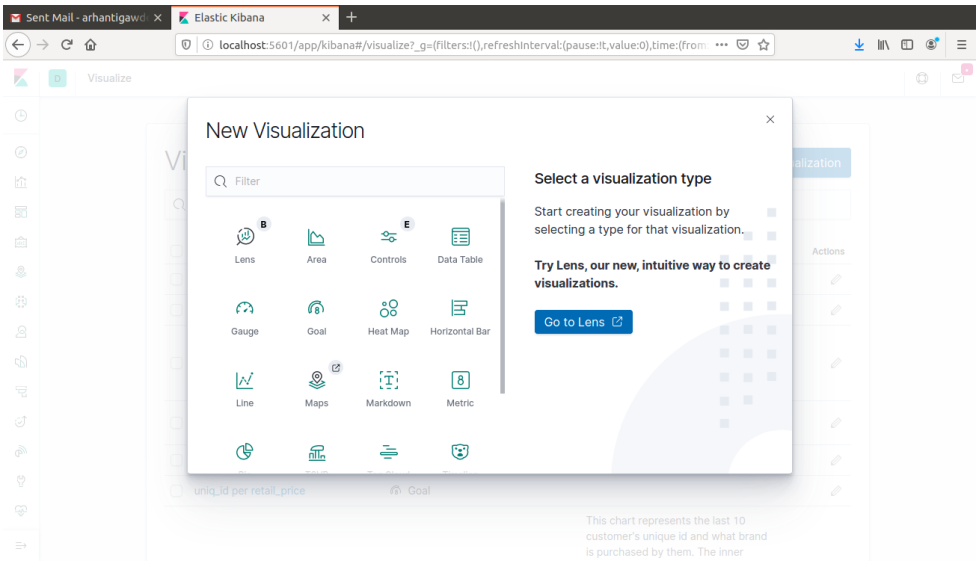


Figure 4.21: Types of visualizations

- Heat map:For showing statistical outliers and are often used for latency values.

- Horizontal line : Good for showing relationship between two fields

- Line :These are simple way to show time series and are good for splitting lines to show anomalies.

- Pie: Useful for displaying parts of whole.

- Vertical Bar: Great series data and for splitting lines across fields.

- Data Table: Best way to split multiple Fields in custom way.

- Gauge: A way to show status of specific metric using thresholds you defined.

- Goal: similar to previous one, useful for monitoring a specific metric defined as a goal.

- metric It is useful visualization for displaying calculation as a single number.

- region map : with the help of that we can mapping the data but for that we need Geopoints

- experimental : it allows you to add custom visualizations based on Vega and vegaLite.

- chart:it use for visualizing time series data and for splinting lines on fields.

In our project we first select pie chart, in that visualization we can apply diffrent filters and split slices for more filtration. For applying filter there is an option on top and for slicing there option is on left hand side we can also sub slice the content. After selection of fields for process we have to click on play button on that window. Then it can be seen like the following window. In these visualization we can see we filtered data by applying filter and slice the pie charts based on user and date. We can also identify some top user list. Also in that pie chart, if we want to get only one site information then can be done as well by just hovering cursor over it and after click on that we get filtered specific site data. When we are ready to save our visualization, click the Save Visualization icon, near the top, then name it and click the Save button.. We can select a source to import data in visualization.

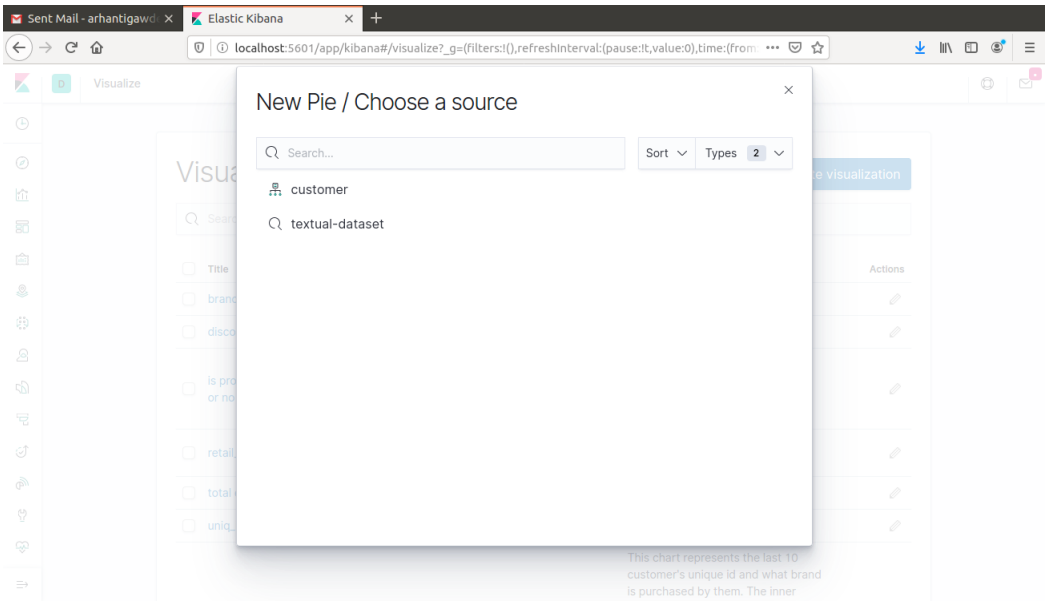


Figure 4.22: Pie chart visualization type

This is where you can start selecting your filters :

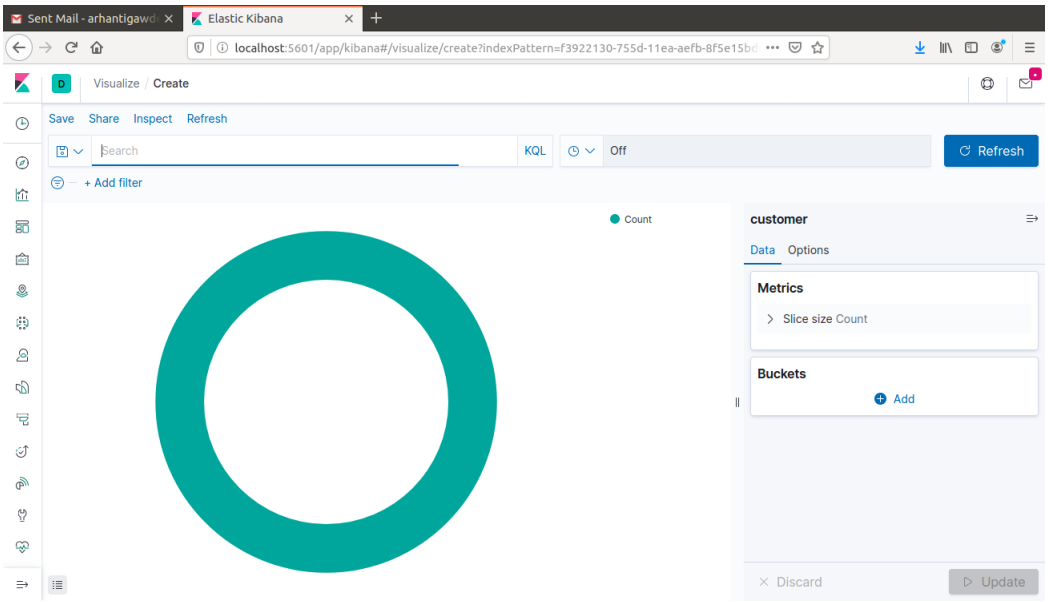


Figure 4.23: Start selecting filter

You can select the filter of your choice :

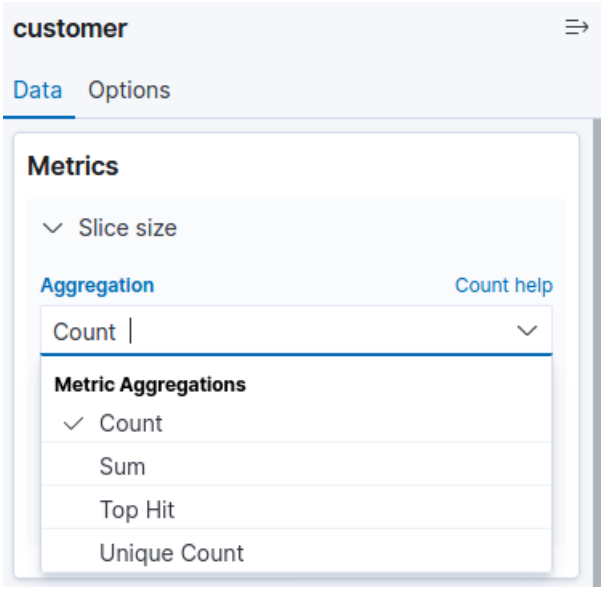


Figure 4.24: Select filter

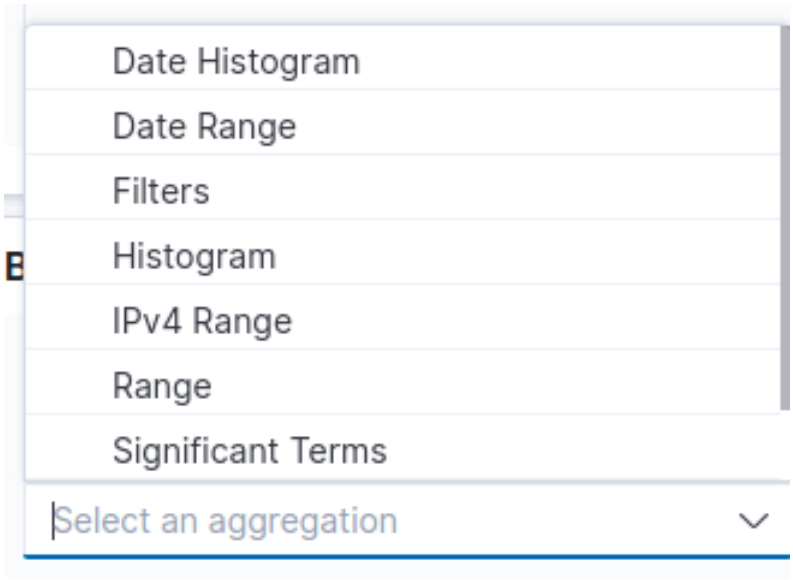


Figure 4.25: Select range to split the diagram

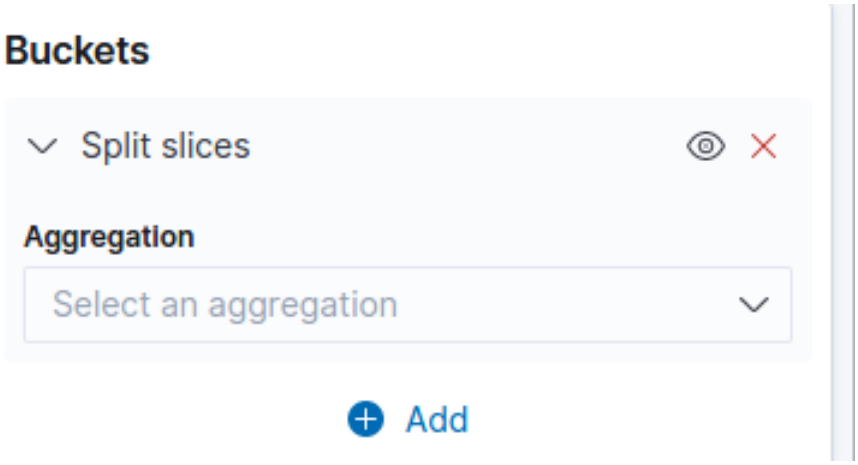


Figure 4.26: You can split diagram into various sections

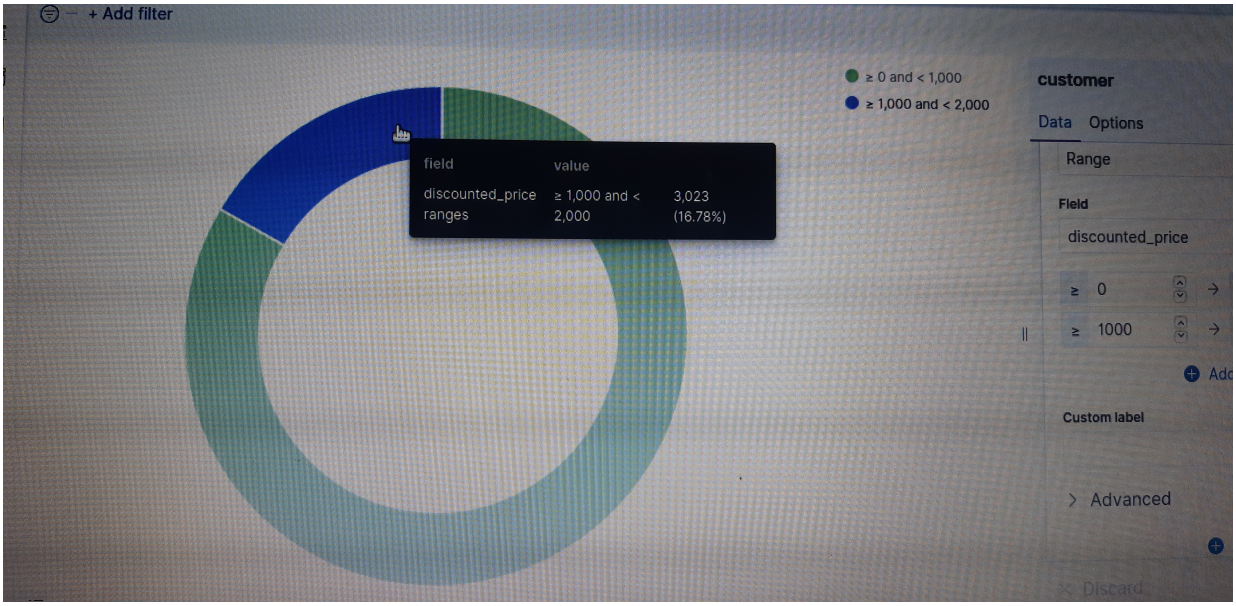


Figure 4.27: Diagram showing discounted price variation

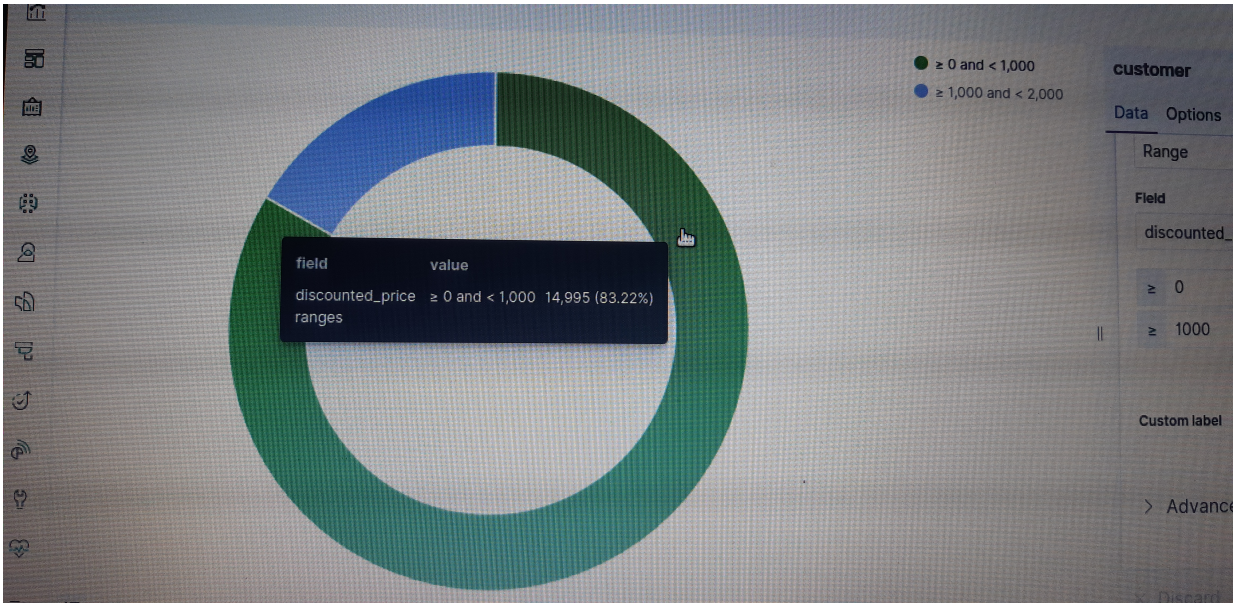


Figure 4.28: Diagram showing discounted price variation

For creating dashboard user has to click on dashboard which is present in kibana tool list.
For creating new dashboard press on New Dashboard icon (to the right of the search bar).

For creating Dashboard following steps are required:

• click on ”add visualize” icon

• then select name of visualizations,search and filtered items in search bar which is located on top of the screen

• After selection click on add button and Collapsed the Add Visualization menu

• Rearranged and re sized the visualizations on the dashboard for aesthetically pleasing to end user.

• click on save button for saving dashboard.

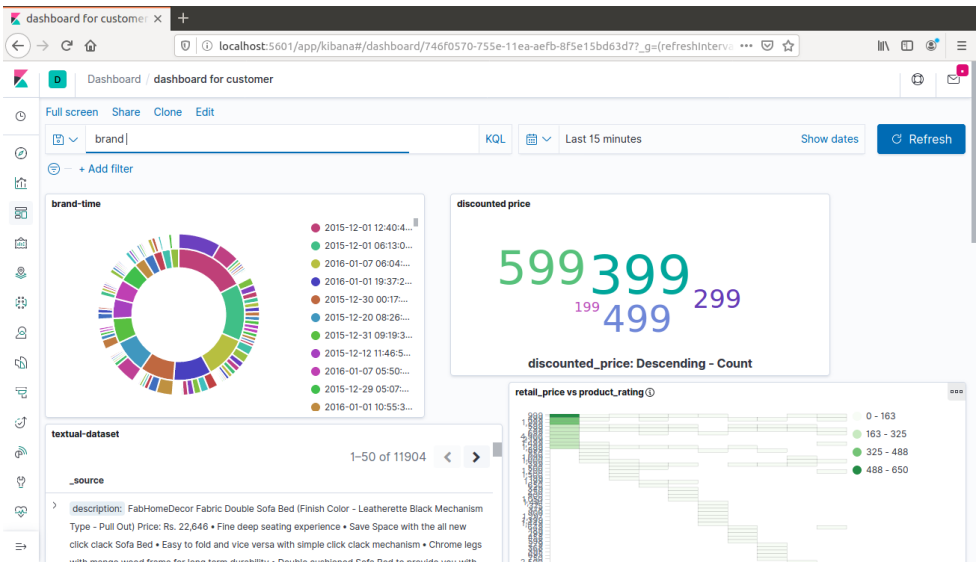


Figure 4.29: Dashboard

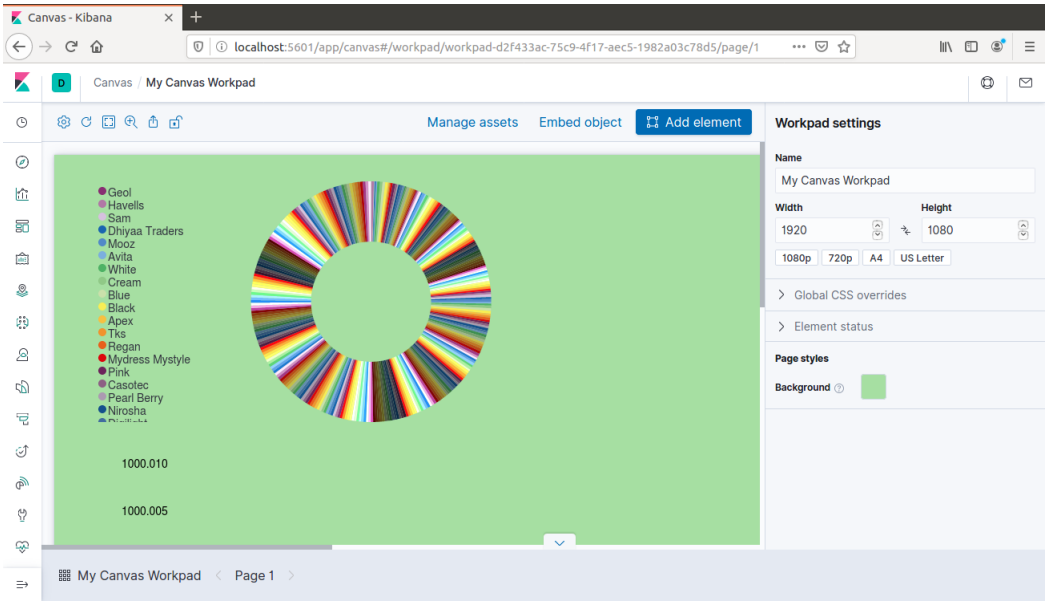


Figure 4.30: Canvas

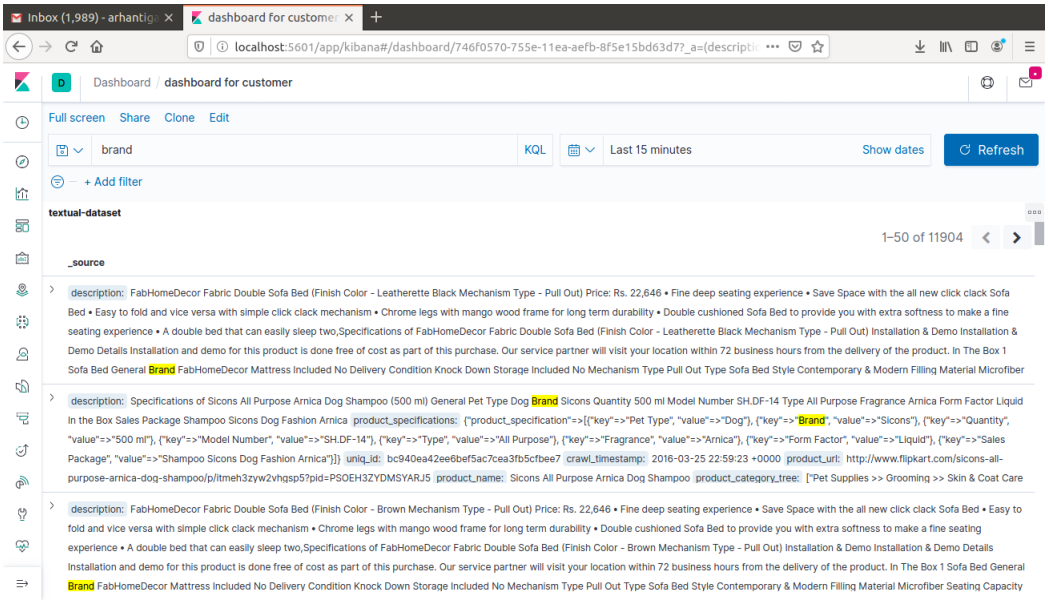


Figure 4.31: This is how the textual data looks like

Chapter 5

Result

The classification of customers will help to convey specific, helpful information about clothes to customers by giving them recommendations so that they get to know about new and trending products and our sale increases by attracting them towards such products. An online presence helps in terms of building and fostering a rapport among your customers.The results obtained will help the store/organization/company to plan effective marketing of their products and services by obtaining a guiding report on the status of their customers which will go a long way in assisting management in saving significant amount of money that could have been spent on wasteful promotional campaigns.

Chapter 6

Annexure A

6.1 Gantt Chart

GANTT CHART TEMPLATE

CUSTOMER CLASSIFICATION AND PREDICTION MODEL FOR ONLI

PROF.PRAVIN ADIVAREKAR

WBS NUMBER	TASK TITLE	TASK OWNER	START DATE	DUE DATE	DURATION(Weeks)	PCT OF TASK COMPLETE
1	**Project Conception and Initiation**					
1.1	Research paper search	ARHANTI,ARPITA,SNEHAL	7/13/19	7/26/19	2	100%
1.1.1	Research paper finalization	ARHANTI,ARPITA,SNEHAL	7/20/19	7/26/19	1	100%
1.2	Customer classification and prediction model for online store	ARHANTI,ARPITA,SNEHAL	7/29/19	8/2/19	1	100%
1.3	Abstract	ARPITA	7/29/19	7/29/19	1	100%
1.4	Objectives	SNEHAL	8/23/19	8/30/19	1	100%
1.5	Literature Review	ARHANTI	8/28/19	8/30/19	1	100%
1.6	Problem Definition	SNEHAL	3/23/18	8/30/19	1	100%
1.7	Scope	ARHANTI	8/12/19	8/16/19	1	100%
1.8	Technology stack	ARPITA	8/12/19	8/16/19	1	80%
1.9	Benefits for environment	SNEHAL	8/12/19	8/16/19	1	100%
1.1	Benefits for society	ARHANTI	8/12/19	8/16/19	1	80%
1.11	Applications	ARPITA	8/12/19	8/16/19	1	100%

Figure 6.1: Gantt Chart

1.8	Technology stack	ARPITA	8/12/19	8/16/19	1	80%
1.9	Benefits for environment	SNEHAL	8/12/19	8/16/19	1	100%
1.1	Benefits for society	ARHANTI	8/12/19	8/16/19	1	80%
1.11	Applications	ARPITA	8/12/19	8/16/19	1	100%
2	**Project Design**					
2.1	Proposed System	ARPITA	9/16/19	9/20/19	1	70%
2.2	Design(Flow Of Modules)	ARHANTI	9/26/19	9/30/19	1	70%
2.3	Activity Diagram	ARHANTI	9/25/19	9/27/19	1	50%
2.4	Use Case Diagram	SNEHAL	9/25/19	9/27/19	1	50%
2.5	Description Of Use Case	SNEHAL	9/28/19	10/4/19	1	30%
2.6	Modules	ARPITA	10/3/19	10/16/19	2	
2.6.1	Module-1	ARHANTI	10/4/19	10/17/19	2	
2.6.2	Module-2	ARPITA	10/4/19	10/17/19	2	
2.6.3	Module-3	ARHANTI	10/5/19	10/18/19	2	
2.6.4	Module-4	SNEHAL	10/5/19	10/18/19	2	
2.7	Preparation Of Report	ARHANTI,ARPITA,SNEHAL	10/24/19	10/30/19	1	
3	**Project Implementation**					
3.1	Module-1	ARHANTI,ARPITA,SNEHAL	01/13/20	01/17/20	1	100%
3.2	Module-2	SNEHAL	01/20/20	01/24/20	2	85%
3.3	Module-3	ARHANTI	01/27/20	01/31/20	1	50%
3.4	Module-4	ARPITA	2/3/20	2/21/20	3	70%
4	**Testing**					
4.1	Design of Test Cases	ARHANTI,ARPITA,SNEHAL	2/24/20	2/28/20	1	100%
4.2	Testing	ARHANTI,ARPITA,SNEHAL	3/2/20	3/13/20	2	90%

Figure 6.2: Gantt Chart

2.6.4	Module-4	SNEHAL	10/5/19	10/18/19	2	
2.7	Preparation Of Report	ARHANTI,ARPITA,SNEHAL	10/24/19	10/30/19	1	
3	**Project Implementation**					
3.1	Module-1	ARHANTI,ARPITA,SNEHAL	01/13/20	01/17/20	1	100%
3.2	Module-2	SNEHAL	01/20/20	01/24/20	2	85%
3.3	Module-3	ARHANTI	01/27/20	01/31/20	1	50%
3.4	Module-4	ARPITA	2/3/20	2/21/20	3	70%
4	**Testing**					
4.1	Design of Test Cases	ARHANTI,ARPITA,SNEHAL	2/24/20	2/28/20	1	100%
4.2	Testing	ARHANTI,ARPITA,SNEHAL	3/2/20	3/13/20	2	90%
5	**Results and Analysis**					
5.1	Analysis Of Results	ARPITA,SNEHAL	3/16/20	3/27/20	2	85%
5.2	Graphical Representation	ARHANTI	3/30/20	4/3/20	1	90%
5.3	Report Preparation	ARHANTI,ARPITA,SNEHAL	04/06/20	04/17/20	2	100.00%

Figure 6.3: Gantt Chart

Chapter 7

Conclusions and Future Scope

The model accepts a big data and classify customers based on the basis of their purchase history, purchase time, brand, etc. and give the business a good and detailed report of the analysis.Providing special offers to particular group of customers on the website lets them know we appreciate them for visiting and shopping from our website. Customers will get recommendations, different recommendations to different people, which will give customers products they might like resulting in increasing sales and thereby giving profit. This technique of direct marketing will prove to be helpful for not only online stores but also the ground shops. It can be applied to a variety of stores, industries, banks, etc..

Bibliography

[1] T. K. Das, “A Customer Classification Prediction Model Based on Machine Learning Techniques”

[2] K. Wisaeng, ”A Comparison of Different Classification Techniques for Bank Direct Marketing”

[3] Hany. A. Elsalamony, Alaa. M. Elsayad, ”Bank Direct Marketing Based on Neural Network”.

[4] Lilian Sing’oei1, Jiayang Wang, ”Data Mining Framework for Direct Marketing: A Case Study of Bank Marketing”

[5] Eniafe Festus Ayetiran, ”A Data Mining-Based Response Model for Target Selection in Direct Marketing”

[6] Karim Masud, Rashedur M. Rahman, ”Decision Tree and Naïve Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing”

[7] Paulo Cortez, ”Using Data Mining For Bank Direct Marketing: An Application Of The Crisp-DM Methodology”

[8] Customer Classification Based on The Historical Purchase Data - Leading India Paper

Appendices

N.A.

Acknowledgement

We have great pleasure in presenting the report on **Customer Classification Prediction Model For Online Store** We take this opportunity to express our sincere thanks towards our guide **Prof. Pravin Adivarekar**. Department of Computer Engineering, APSIT thane for providing the technical guidelines and suggestions regarding line of work. We would like to express our gratitude towards his constant encouragement, support and guidance through the development of project.

We thank **Prof. Sachin Malave** Head of Department,Computer Engineering, APSIT for his encouragement during progress meeting and providing guidelines to write this report.

We thank **Prof. Amol Kalugade** BE project co-ordinator, Department of Computer Enginering, APSIT for being encouraging throughout the course and for guidance.

We also thank the entire staff of APSIT for their invaluable help rendered during the course of this work. We wish to express our deep gratitude towards all our colleagues of APSIT for their encouragement.

Student Name: Arhanti Gawde
Student ID: 16102026

Student Name: Arpita Hirlekar
Student ID: 16102018

Student Name: Snehal Surve
Student ID: 17202013

Publication

Paper entitled **"A Customer Classification Prediction Model Based on Machine LearningTechniques"** is presented at **"2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)"** by **" T. K. Das".**