Computer Engineering Department

A.P. Shah Institute of Technology

G.B.Road, Kasarvadavli, Thane(W), Mumbai-400615

UNIVERSITY OF MUMBAI

Academic Year 2019-2020

A Project Report on

CUSTOMER CLASSIFICATION PREDICTION MODEL FOR ONLINE STORE

Submitted in partial fulfillment of the degree of

Bachelor of Engineering(Sem-8)

in

Computer Engineering

By

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

Under the Guidance of

Prof. Pravin Adivarekar

1.Project Conception and Initiation

1.1 Abstract

- This paper deals with classification of the customers as a part of the marketing strategy in order to expand the business profit by recommending customers the relevant products depending upon the category.
- 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.
- Customers belonging to a particular cluster only will get recommendations about a specific product they would likely be interested in. While launching new products, services or releasing new versions of existing products / services we will reach potential customers only. To address which customers to give the recommendation, details of existing customers will be maintained at all levels and will be analyzed henceforth.
- 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.

1.2 Objectives

- The objective of this project is to build a model that will classify customers based on the basis of some parameters and give them recommendations accordingly and provide a comprehensive framework for the efforts taking place in direct marketing strategy.
- The company will be benefited because it will send recommendation only to customers who are likely to revert to the recommended product.

1.3 Literature Review

- Base IEEE paper : T. K. Das, “A Customer Classification Prediction Model Based on Machine Learning Techniques”

- Papers referred :

- K. Wisaeng, “A Comparison of Different Classification Techniques for Bank Direct Marketing”.
- Hany. A. Elsalamony, Alaa. M. Elsayad, “Bank Direct Marketing Based on Neural Network”
- Lilian Sing’oei1 and Jiayang Wang,”Data Mining Framework for Direct Marketing: A Case Study of Bank Marketing”.
- Eniafe Festus Ayetiran,“A Data Mining-Based Response Model for Target Selection in Direct Marketing”.

1.4 Problem Definition

- To generate a model that will classify customers on some constraints and make clusters of customers to give appropriate recommendations to particular clusters. Every cluster will be having different customers on the basis of their past purchases and items they select in cart.
- All customers in a cluster will be given the same recommendations which will be relevant to their selection of clothes. This process of classification is an integral part of the direct marketing scheme which is trending globally nowadays.

1.5 Scope

- Direct marketing deals with an important issue of identifying customers who are more likely to respond to the new offers. Data mining is an extensively used method to carry out the process of target selection i.e. identifying the potential customers.
- This project develops predictive analysis using data mining techniques in order to predict the probability that a customer will respond to a promotion or an offer depending on his purchase/selection/ viewing history. All of the data mining steps will be carried out and out of several standard techniques the best classification technique will be selected.
- The aim of this project is to provide a comprehensive framework for the efforts taking place in direct marketing strategy. 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.

1.6 Technology stack

FRONT END-

- HTML 5- HTML is the standard markup language that allows users to create and structure sections, headings, links, paragraphs, and more, on a website using various tags and elements.

- CSS 3- CSS is the language that describes the style of an HTML document. we are using CSS3 which introduces new selectors and properties that allow for more flexibility with page layout and presentation.

- PHP- PHP is a widely-used, open source scripting language. PHP scripts are executed on the server.

BACK-END-

- PHPMYADMIN-phpMyAdmin is a free software tool written in PHP, intended to handle the administration of MySQL over the Web. phpMyAdmin supports a wide range of operations on MySQL and MariaDB.

- ELASTIC STACK-ELK stack has been re-branded as Elastic Stack.The ELK stack is an amazing and powerful collection of three open source projects- Elasticsearch , Logstach , Kibana. Despite each one of these three technologies being a separate project, they have been built to work exceptionally well together.

1.7 Benefits for environment & Society

Website creation, being a part of this project, is an important area which will help to lessen the environmental issues. Opting for a darker background for the website can surely be valued. Darker backgrounds are known to save as much as 4% of the total electrical energy as compared to lighter or white background. It also supports visual hierarchy thus reducing eye strain if browsed for hours. It offers depth in the reflection of the presented content. Using cool dark backgrounds for a website instantly transforms it into a night-mode one. Nowadays most of the devices feature an option which is called night-mode and turns the lighter colours on the device to black. Because of this particular technique the device produces less light due to which energy is saved which also prevents the end-user's eyes from getting sore quickly. Since this project is typically based on software and analysis, the probability of producing e-waste or the hardware waste has been brought down.

2. Project Design

2.1 Proposed System

- In our project we will be classifying customers who are using our website in divergent clusters based on their liking. Customers will be put in particular cluster based on some constraints like similar shopped items, viewed items or items in their cart.
- In the IEEE paper that we have selected, they focus on three mostly used classification algorithms namely Naïve Bayesian, K Nearest Neighbor and Support Vector Machines to a bank customer dataset.
- In the reference papers it can be seen that various authors have used different machine learning algorithms.
- On analyzing the reports and results from the reference papers, it was observed that Naïve Bayesian classification achieves the highest accuracy in the context of existing back customer data set. In our project we are doing it for an online shopping website related to clothing.
- After classification we come to the next important part which is recommendation of products. Here, recommendation about a new product will be given to all the customers in a cluster who will most likely respond to the new offering.

2.2 Design(Flow Of Modules)

- DOMAIN UNDERSTANDING

- In any business it is a challenge to figure out the exact knowledge of the business for it to be successful.
- The domain of this project involves online shopping and classification of its customers. Hence, it is important to understand the environment of the website and its marketing strategy.

DATA COLLECTION

- Data collection is the process of gathering and measuring data, information or any variables of interest in a standardized and established manner.
- Usually initial and integral component in data mining is data collection which depends on environment of the domain. The primary goal of any data collection is to capture quality data or evidence.
- Data collection in our project is the data of users which we will collect from our website.

- DATA PREPROCESSING

- Data cleaning- Adjustments are made according to the addition or deletion of items in the cart. Also, duplication of data is removed.
- ETL- First the extract function reads data from our website and loads the transformed data in the database.
- Attribute selection- Different categories of clothes are the various attributes that are to be selected.

MODEL CONSTRUCTION

- This step involves application of Data Mining techniques on the data to discover the interesting patterns.
- Since our project deals with classification of customers, we will be applying classification algorithms to the obtained dataset.
- Elastic Stack has been selected for this project.

ANALYSIS

- The analysis stage is the stage at which data is finally usable. It involves visualizing the data in a very clear and understandable way.
- The classification done in previous stages will result in various categories. In this step we will decide what and how recommendations have to be given to specific groups.

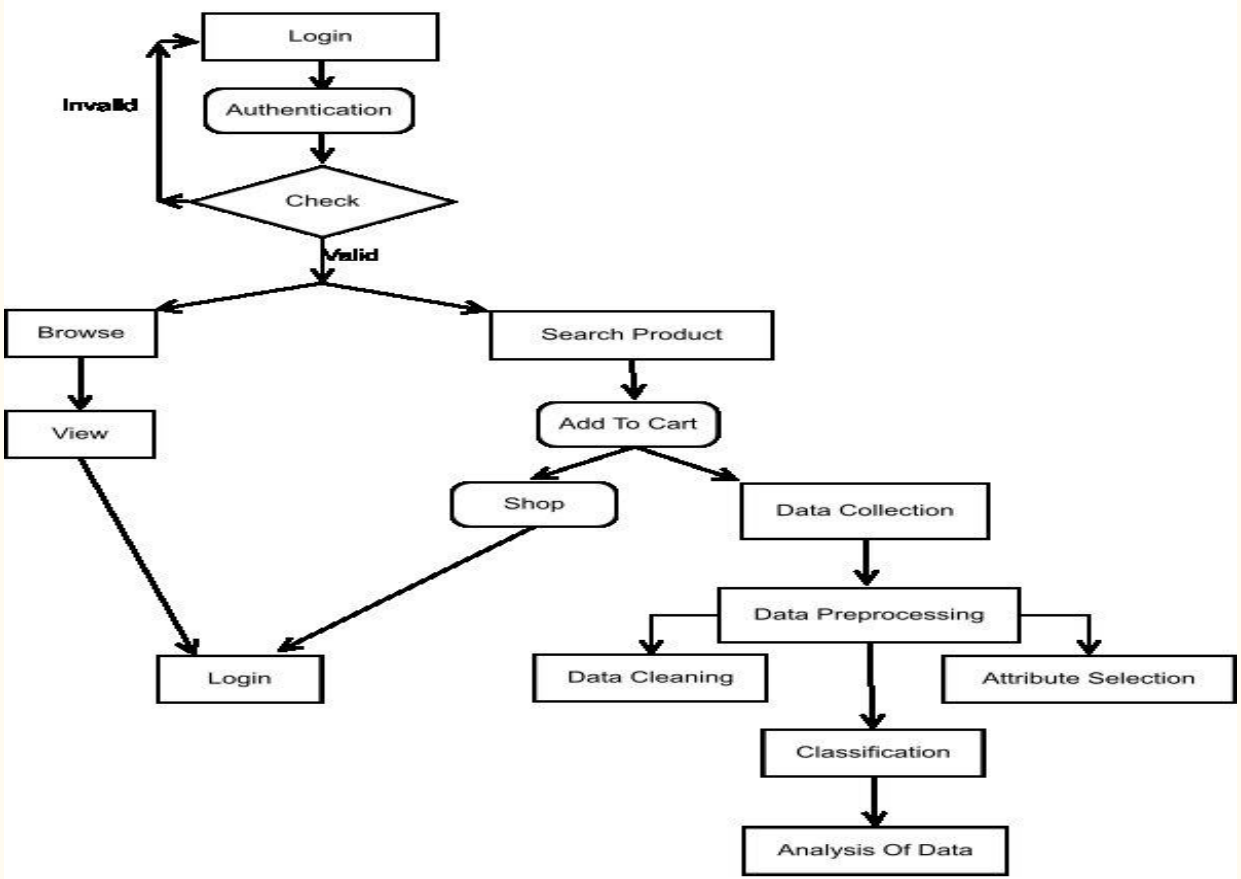
REPORTING

- Data reporting is the part of a system that reports key elements relating to an organizations' performance in order to improve different aspects.
- After observing data gathered for a certain period of time, we can make changes in what recommendations we give and the way we are giving it.

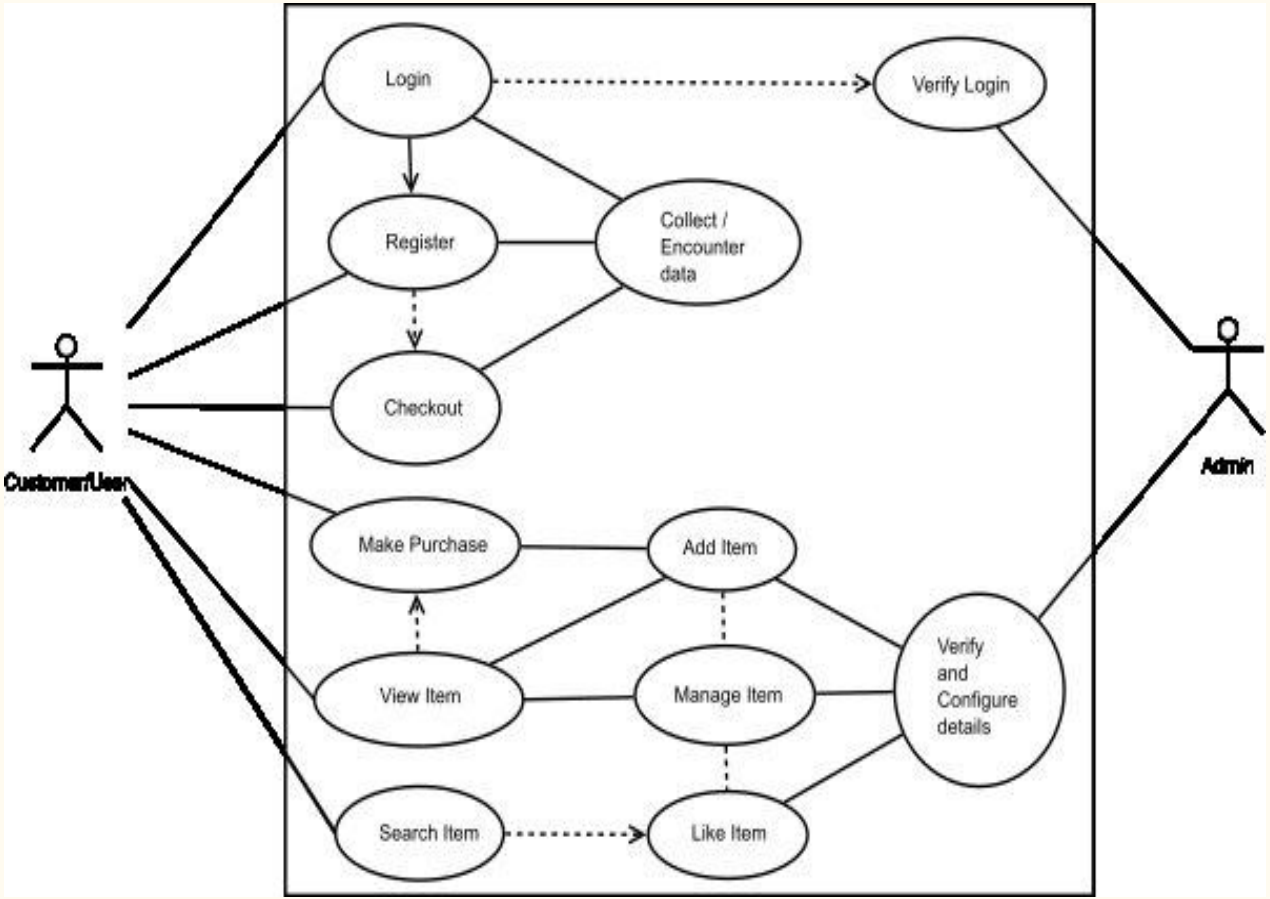
2.3 Description Of Use Case

Use cases used in this project are login, register, collection of the details, making purchase, viewing items, etc. Whenever a customer logs in or creates an account on the website created, then he/she will be directed to the items on the website. All these data is then collected in the databases. Now the appropriate users are eligible to view, shop, purchase the items they like. All these actions are hereby managed by the databases. The databases not only collect the user information but all the activities done by the user. Depending on the purchases, the customers are thus classified.

2.4 Activity diagram



2.5 Use Case Diagram



2.6 Module-1

- DOMAIN UNDERSTANDING

- In any business it is a challenge to figure out the exact knowledge of the business for it to be successful. To acquire the business understanding and project goals it is essential to have information regarding the customers, products, cost structure, distribution and service of the product and advertising and promotion of the product.
- The domain of this project involves online shopping and classification of its customers. Hence, it is important to understand the environment of the website and its marketing strategy.

Module-2

- DATA COLLECTION

- Data collection is the process of gathering and measuring data, information or any variables of interest in a standardized and established manner. This is an integral, usually initial, component in data mining.
- Data collection is concerned with the accurate acquisition of data although methods may differ depending on the field.
- The primary goal of any data collection endeavor is to capture quality data or evidence that easily translates to rich data analysis that may lead to credible and conclusive answers to questions that have been posed.
- Accurate data collection is essential to ensure the integrity of the research. Data collection in our project is the data of users we will collect from our website.

Module-3

- DATA PREPROCESSING

- Data cleaning- Adjustments are made according to the addition or deletion of items in the cart. Also, duplication of data is removed.
- ETL- First the extract function reads data from our website and loads the transformed data in the database.
- Attribute selection- Different categories of clothes are the various attributes that are to be selected.

Module-4

- MODEL CONSTRUCTION

- After all the above steps are applied, the transformed and processed data undergoes the step of model construction.
- This step involves application of Data Mining techniques on the data to discover the interesting patterns. Based on the business objectives, suitable modeling techniques are selected for the prepared data set.
- Since our project deals with classification of customers, we will be applying classification algorithms to the obtained data set.
- Elastic Stack is used in our project.

Module-5

- ANALYSIS

- The analysis stage is the stage at which data is finally usable. It involves visualizing the data in a very clear and understandable way.
- The classification done in previous stages will result in various categories. In this step we will decide what and how recommendations have to be given to specific groups.

Module-6

- REPORTING

- Data reporting is the part of a system that reports key elements relating to an organizations' performance in order to improve different aspects.
- After observing data gathered for a certain period of time, we can make changes in what recommendations we give and the way we are giving it.

2.7 References

- T. K. Das, “A Customer Classification Prediction Model Based on Machine Learning Techniques”
- K. Wisaeng, "A Comparison of Different Classification Techniques for Bank Direct Marketing"
- Hany. A. Elsalamony, Alaa. M. Elsayad, "Bank Direct Marketing Based on Neural Network"
- Lilian Sing'oei1, Jiayang Wang, "Data Mining Framework for Direct Marketing: A Case Study of Bank Marketing"
- Eniafe Festus Ayetiran, "A Data Mining-Based Response Model for Target Selection in Direct Marketing"
- Karim Masud, Rashedur M. Rahman, "Decision Tree and Naïve Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing"
- Paulo Cortez, "Using Data Mining For Bank Direct Marketing: An Application Of The Crisp-DM Methodology"
- Customer Classification Based on The Historical Purchase Data - Leading IndiaPaper

3.This semester work
—

Work till done now

- Front-end is fully completed.
- The back-end of the project is partially done.
- ELASTIC STACK has been installed.
- PHP and PHPMYADMIN has also installed.
- All software requirements are fulfilled.

ELASTIC STACK

Elastic Stack is a group of open source products from Elastic designed to help users take data from any type of source and in any format and search, analyze, and visualize that data in real time. The product group was formerly known as ELK Stack, in which the letters in the name stood for the products in the group: Elasticsearch, Logstash and Kibana. A fourth product, Beats, was subsequently added to the stack, rendering the potential acronym unpronounceable.

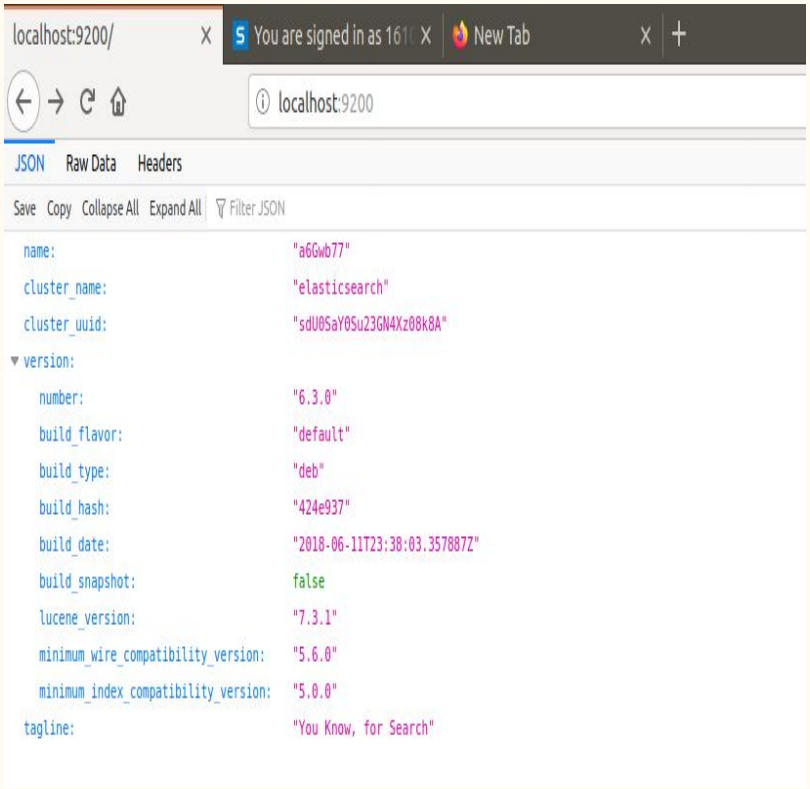
Elastic Stack components:

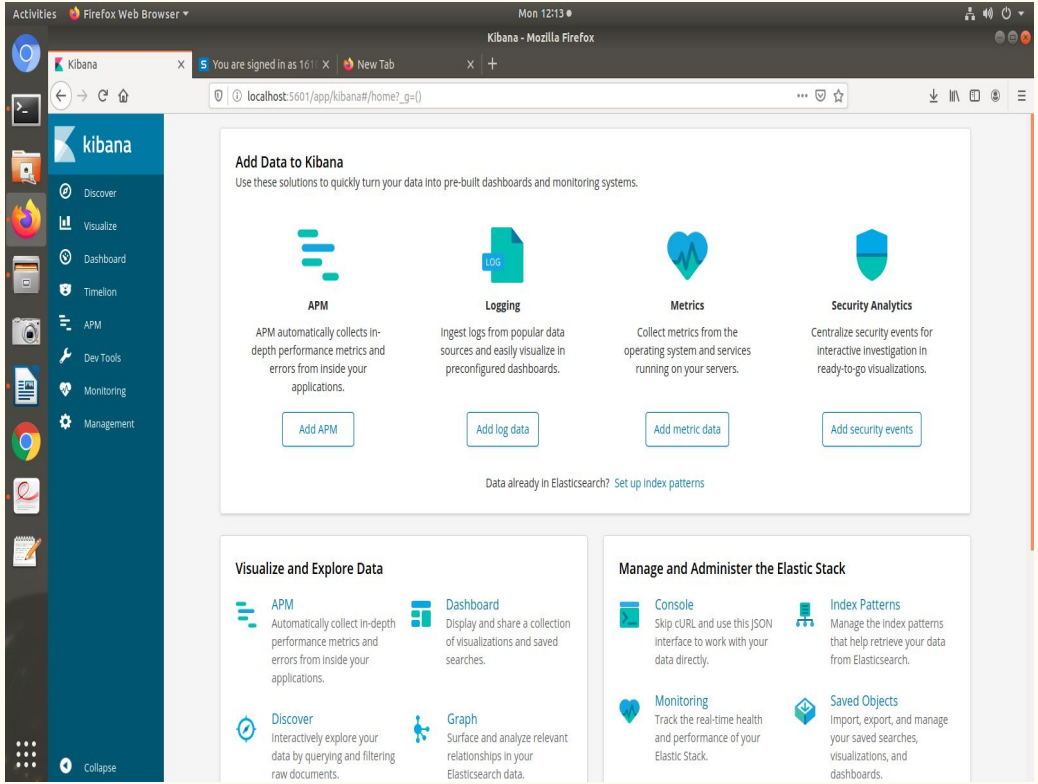
- Elasticsearch is a Restful distributed search engine built on top of Apache Lucene and released under an Apache license. It is Java-based and can search and index document files in diverse formats.

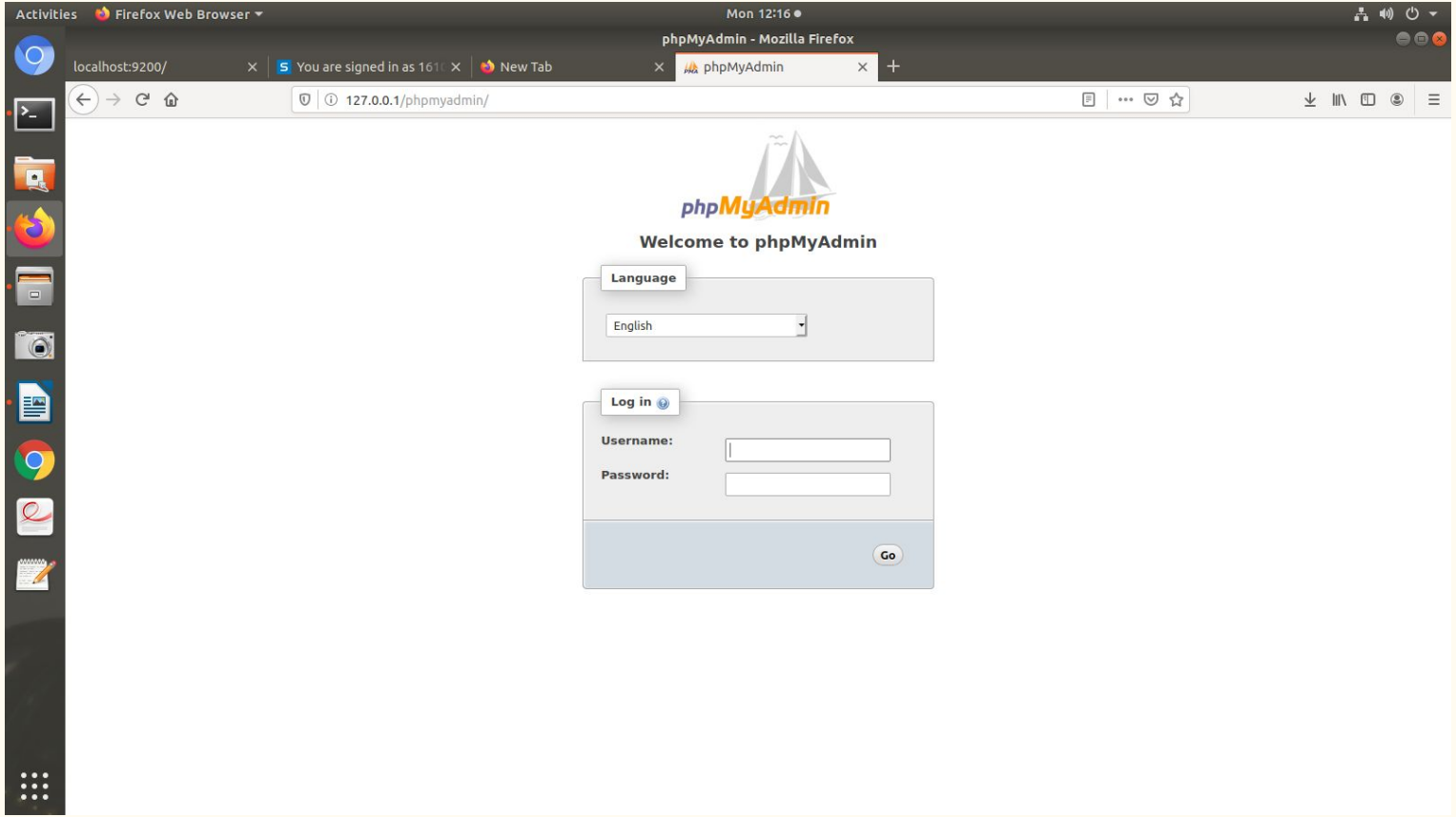
- Logstash is a data collection engine that unifies data from disparate sources, normalizes it and distributes it. The product was originally optimized for log data but has expanded the scope to take data from all sources.

- Kibana is an open source data visualization and exploration tool from that is specialized for large volumes of streaming and real-time data. The software makes huge and complex data streams more easily and quickly understandable through graphic representation.

SCREENSHOTS






Activities
Firefox Web Browser
Mon 12:16
phpMyAdmin - Mozilla Firefox
localhost:9200/
You are signed in as 161
New Tab
phpMyAdmin
127.0.0.1/phpmyadmin/
phpMyAdmin
Welcome to phpMyAdmin
Language
English
Log in
Username:
Password:
Go

Thank You