SAVITRIBAI PHULE PUNE UNIVERSITY

A PROJECT REPORT ON

Generic user activity analysis and prediction

SUBMITTED TOWARDS THE PARTIAL
FULFILLMENT OF THE REQUIREMENTS OF

BACHELOR OF ENGINEERING (Computer Engineering)

BY

Shreyas Kulkarni	B120054350
Kiran Mokashi	B120054362
Shailesh Bagade	B120054232
Shivam Bawane	B120054240

Under The Guidance of

Prof. M. S. Wakode

DEPARTMENT OF COMPUTER ENGINEERING
Pune Institute of Computer Technology
Dhankawadi, Pune-411043

Pune Institute of Computer Technology
DEPARTMENT OF COMPUTER ENGINEERING

CERTIFICATE

This is to certify that the Project Entitled

Generic user activity analysis and prediction

Submitted by

Shreyas Kulkarni	B120054350
Kiran Mokashi	B120054362
Shailesh Bagade	B120054232
Shivam Bawane	B120054240

is a bonafide work carried out by Students under the supervision of Prof. Guide Name and it is submitted towards the partial fulfillment of the requirement of Bachelor of Engineering (Computer Engineering).

Prof. M. S. Wakode
Internal Guide
Dept. of Computer Engg.

Dr. R. B. Ingle
H.O.D
Dept. of Computer Engg.

Dr. P. T. Kulkarni
Principal
Pune Institute of Computer Technology

Signature of Internal Examiner

Signature of External Examiner

PROJECT APPROVAL SHEET

Generic user activity analysis and prediction

Is successfully completed by

Shreyas Kulkarni	B120054350
Kiran Mokashi	B120054362
Shailesh Bagade	B120054232
Shivam Bawane	B120054240

at

DEPARTMENT OF COMPUTER ENGINEERING

PUNE INSTITUTE OF COMPUTER TECHNOLOGY

SAVITRIBAI PHULE PUNE UNIVERSITY,PUNE

ACADEMIC YEAR 2016-2017

Prof. M. S. Wakode
Internal Guide
Dept. of Computer Engg.

Dr. R. B. Ingle
H.O.D
Dept. of Computer Engg.

Abstract

This system presents a prototype of design and implementation of a system which carries out data analysis and prediction that allows clients (e-commerce sites, video streaming etc) to configure the system according to their application. The proposed scheme consists of event collector, database, search, analytics and recommendation engine. Data is collected in form of events from the user using RabbitMQ and then streamed through using SPARK and eventually stored in database used, MongoDB. The elastic search assists to build the search engine which helps us to parse through the data and KIBANA to perform various data analytics. Machine learning(J48 algorithm) using WEKA platform is used to build recommendation engine and analysis for vivid functionalities on available data set to provide suggestions, enhancing user experience. Analytics UI helps providing proper understanding of search patterns and user events to the clients to increase efficiency of the application.

Acknowledgments

It gives us great pleasure in presenting the preliminary project report on ***'GENERIC USER EVENT ANALYSIS AND PREDICTION'****.*

I would like to take this opportunity to thank my internal guide ***Prof.M. S. Wakode*** *for giving me all the help and guidance I needed. I am really grateful to them for their kind support. Their valuable suggestions were very helpful.*

I am also grateful to ***Prof.R. B. Ingle****, Head of Computer Engineering Department, Pune Institute of Computer Technology for his indispensable support, suggestions.*

Kiran Mokashi
Shailesh Bagade
Shivam Bawane
Shreyas Kulkarni
(B.E. Computer Engg.)

Contents

List of Figures

List of Tables

CHAPTER 1
SYNOPSIS

1.1 Project Title

Generic User activity analysis and prediction.

1.2 Project Option

Industry sponsored

1.3 Internal Guide

Prof. M .S. Wakode

1.4 Sponsorship and External Guide

G.S. Lab

1.5 Technical Keywords

1. Data Analysis
   (a) a. Sorting
   (b) b. Filtering
   (c) c. Searching

1.6 Problem Statement

To implement a system that performs data analysis and prediction for events (searches etc) performed by user on application interface using open source softwares. Predictions and recommendation will be made using suitable algorithms.

1.7 Abstract

- This system presents a prototype of design and implementation of a system which carries out data analysis and prediction that

allows clients (e-commerce sites, video streaming etc) to configure the system according to their application. The proposed scheme consists of event collector, database, search, analytics and recommendation engine. Data is collected in form of events from the user using RabbitMQ and then streamed through using SPARK and eventually stored in database used, MongoDB. The elastic search assists to build the search engine which helps us to parse through the data and KIBANA to perform various data analytics. Machine learning(J48 algorithm) using WEKA platform is used to build recommendation engine and analysis for vivid functionalities on available data set to provide suggestions, enhancing user experience. Analytics UI helps providing proper understanding of search patterns and user events to the clients to increase efficiency of the application.

1.8 Goals and Objectives

- To implement a system which will efficiently collect, analyze and predict output.
- To implement a system which is generic.

1.9 Relevant Mathematics Associated With The Project

System Description:

- Input:: User Events.(like searches made by user on application interface)
- Output:Prediction depending upon user events, analyzed output in form of charts.
- Success Conditions:proper recommendation is given
- Failure Conditions:The recommendation is not the most appropriate for the user.

1.10 Names of Conferences / Journals where papers can be published

- IEEE/ACM Conference/Journal 1
- Conferences/workshops in IITs
- Central Universities or SPPU Conferences
- IEEE/ACM Conference/Journal 2

1.11 Review of Conference/Journal Papers Supporting Project Idea

1 .Predictive Analytics Using Data Mining Technique Using Data Mining Technique: Prediction can be done by using data mining techniques on large data sets. Data mining is a broad concept that consists of series of steps. Firstly data is pre-processed and then mining techniques are applied. Results from mining techniques are evaluated and interpreted and the expected result is generated using prediction algorithm

2 .The Predictability of Data Values:

The predictability of data values is studied at a fundamental level. Two basic predictor models are defined : Computational predictors perform an operation on previous values to yield predicted next values. To understand the potential of value prediction we perform simulations with unbounded prediction tables that are immediately updated using correct data values.

3 .Feasibility Analysis of Big Log Data Real Time Search Based on Hbase and Elastic Search:

Elastic Search, which is based on Lucene, is the modern search engineer for cloud environment. This paper presents a real-time big data search method: First, Flume agent from the end user's machine collect log events, then Elastic Search according to the search conditions are needed row key list; finally Hbase using these row key directly from the database to get the data, the paper-based hardware to create a virtual machine environment, the experiment proved, with the search for more log events, the search response time does not increase linearly.

4 .Mining Modern Repositories with Elastic search:

Elastic search a distributed full-text search engine — explicitly addresses issues of scalability, big data search, and performance that relational databases were simply never designed to support . While Elastic search and traditional RDBMSs in many ways, at the higher-level many of the core concepts of Elastic search have analogues in the RDBMS world. All data in Elastic search is stored in indices. An index in Elastic search is like a database in a RDBMS: it can store types of documents, update them, and search for them. Each document in Elastic search is a JSON object, analogous to a row in a table in a RDBMS.

5 .Survey Paper on Elastic Search: Elastic search is a way to organize data and make it easily accessible. It is a server based search on Lucene. It is a highly scalable, distributed and full-text search engine. Elastic search is. Elasticsearch is a standalone database which is written in Java and using HTTP/JSON protocol,its takes data and optimized the data.

6 .A recommender system by using classification based on frequent pattern mining and J48 algorithm.:

Users behavior modeling on the web and extracting its patterns can be utilized for customizing search results without users specifications. Since offering a precise suggestion to users in search engines and e-commerce is desirable for users, precision is the most important factor is such systems.

7 .Kafka: a Distributed Messaging System for Log Processing:

Kafka, a distributed messaging system that we developed for collecting and delivering high volumes of log data with low latency. Kafka has superior performance when compared to two popular messaging systems. We have been using Kafka in production for some time and it is processing hundreds of gigabytes of new data each day.

8 .Performance of Elasticsearch in Cloud Environment: Elasticsearch is a distributed data storage system. It can store and fetch complex data structures serialized as JSON documents in real time . In other words, the instance in which a document has been indexed in Elasticsearch, it can be retrieved from any node in the cluster. When performing search, JSON objects are given to Elastic search and result obtained is also in JSON format. Elastic search, a full-text java based search engine, designed keeping

cloud environment in mind solves issues of scalability, search in real time, and efficiency that relational databases were not able to address.

9 .Design and Implementation of an Indexing Method Based on Fields for Elastic search:

Designing a search engine for the application and version of the nodes is of great importance for the Internet safe guard. In order to meet the users needs of searching the information of IP address and domain name, the paper proposed a method to convert IP address and analyze domain names. As more and more information needs to be indexed, it will take a longer time to query in the index, which will influence the users experience. A method to create the index based on fields, and the corresponding compressing algorithm is used to guarantee the compression efficiency.

10 .Classification and prediction based data mining algorithms to predict slow learners in education sector:

This paper is about identifying the slow learners among students and displaying it by a predictive data mining model using classification based algorithms. The database school is tested and applied various prediction algorithms. WEKA an Open source tool. As a result, statistics are generated based on all classification algorithms and comparison of all five classifiers is also done in order to predict the accuracy and to find the best performing classification algorithm among all. In this paper, a knowledge flow model is also shown among all five classifiers.

1.12 Plan of Project Execution

Using planner or alike project management tool.

CHAPTER 2

TECHNICAL KEYWORDS

2.1 Area of Project

Machine learning, Data analysis

2.2 Technical Keywords

(a) . Machine Learning
    i. Control Structure
        A. Reliability
        B. Efficiency

(a) Data Analysis
    i. Sorting
    ii. Filtering
    iii. Filtering
    iv. Indexing

CHAPTER 3

INTRODUCTION

3.1 Project Idea

- User Events Analysis.

3.2 Motivation Of The Project

- To design a system that gives relevant and more accurate recommendations to end user.
- To provide proper understanding of search patters and user events to increase efficiency of the application in which it is used.

3.3 Literature Survey

- Predictive Analytics Using Data Mining Technique Using Data Mining Technique: Prediction can be done by using data mining techniques on large data sets. Data mining is a broad concept that consists of series of steps. Firstly data is pre-processed and then mining techniques are applied. Results from mining techniques are evaluated and interpreted and the expected result is generated using prediction algorithm.
- The Predictability of Data Values:

  The predictability of data values is studied at a fundamental level. Two basic predictor models are defined : Computational predictors perform an operation on previous values to yield predicted next values. To understand the potential of value prediction we perform simulations with unbounded prediction tables that are immediately updated using correct data values.
- Feasibility Analysis of Big Log Data Real Time Search Based on Hbase and Elastic Search:

  Elastic Search, which is based on Lucene, is the modern search engineer for cloud environment. This paper presents a real-time big data search method: First, Flume agent from the end user's machine collect log events, then Elastic Search according to the search conditions are needed row key list; finally Hbase using these row key directly from the database to get the data, the paper-based hardware to create a virtual machine environment, the experiment

proved, with the search for more log events, the search response time does not increase linearly.

- Mining Modern Repositories with Elastic search:

  Elastic search a distributed full-text search engine — explicitly addresses issues of scalability, big data search, and performance that relational databases were simply never designed to support .

  While Elastic search and traditional RDBMSs in many ways, at the higher-level many of the core concepts of Elastic search have analogues in the RDBMS world. All data in Elastic search is stored in indices. An index in Elastic search is like a database in a RDBMS: it can store types of documents, update them, and search for them. Each document in Elastic search is a JSON object, analogous to a row in a table in a RDBMS.

- Survey Paper on Elastic Search: Elastic search is a way to organize data and make it easily accessible. It is a server based search on Lucene. It is a highly scalable, distributed and full-text search engine. Elastic search is. Elasticsearch is a standalone database which is written in Java and using HTTP/JSON protocol,its takes data and optimized the data.

- A recommender system by using classification based on frequent pattern mining and J48 algorithm.

  Users behavior modeling on the web and extracting its patterns can be utilized for customizing search results without users specifications. Since offering a precise suggestion to users in search engines and e-commerce is desirable for users, precision is the most important factor is such systems.

- Kafka: a Distributed Messaging System for Log Processing:

  Kafka, a distributed messaging system that we developed for collecting and delivering high volumes of log data with low latency. Kafka has superior performance when compared to two popular messaging systems. We have been using Kafka in production for some time and it is processing hundreds of gigabytes of new data each day.

- Performance of Elasticsearch in Cloud Environment: Elasticsearch is a distributed data storage system. It can store and fetch complex data structures serialized as JSON documents in real time . In other words, the instance in which a document has been indexed in Elasticsearch, it can be retrieved from any node in the cluster. When performing search, JSON objects are given to

Elastic search and result obtained is also in JSON format. Elastic search, a full-text java based search engine, designed keeping cloud environment in mind solves issues of scalability, search in real time, and efficiency that relational databases were not able to address.

- Performance of Elasticsearch in Cloud Environment: Elasticsearch is a distributed data storage system. It can store and fetch complex data structures serialized as JSON documents in real time . In other words, the instance in which a document has been indexed in Elasticsearch, it can be retrieved from any node in the cluster. When performing search, JSON objects are given to Elastic search and result obtained is also in JSON format. Elastic search, a full-text java based search engine, designed keeping cloud environment in mind solves issues of scalability, search in real time, and efficiency that relational databases were not able to address.

- Design and Implementation of an Indexing Method Based on Fields for Elastic search:

  Designing a search engine for the application and version of the nodes is of great importance for the Internet safe guard. In order to meet the users needs of searching the information of IP address and domain name, the paper proposed a method to convert IP address and analyze domain names. As more and more information needs to be indexed, it will take a longer time to query in the index, which will influence the users experience. A method to create the index based on fields, and the corresponding compressing algorithm is used to guarantee the compression efficiency.

- Classification and prediction based data mining algorithms to predict slow learners in education sector:

  This paper is about identifying the slow learners among students and displaying it by a predictive data mining model using classification based algorithms. The database school is tested and applied various prediction algorithms. WEKA an Open source tool. As a result, statistics are generated based on all classification algorithms and comparison of all five classifiers is also done in order to predict the accuracy and to find the best performing classification algorithm among all. In this paper, a knowledge flow model is also shown among all five classifiers.

CHAPTER 4

PROBLEM DEFINITION AND SCOPE

4.1 Problem Statement

To implement a system that performs data analysis and prediction for events performed by user.

4.1.1 Goals and Objectives

Goal and Objectives:

- To design a system which carries out data analysis and prediction that allows client to configure the system according to their application.
- To provide proper understanding of search patters and user events to increase efficiency of the application in which it is used.
- Applicable for different systems.

4.1.2 Statement of Scope

- System will collect different events performed by user and analyze it using machine learning algorithm like WEKA. It will collect all types of events performed by user, analyze it and then will predict result, generate charts.
- Software will be generic. It means that it will collect all type of data, analyze and predict output. It will also generate output in the form of charts.

4.2 Major Constraints

- All types of events will be collected, analyzing particular events based on vendors requirement.
- Events must be classified based on vendors requirement.

4.3 Methodologies of Problem Solving and Efficiency Issues

- Event collector will collect all events performed by user, out of which 10 percent will be used for analysis. Hence data must be sorted efficiently using tools like elastic search so that required data will be available for analysis and else will be removed.

4.4 Outcome

- Prediction based upon user events.
- Concise and precise summary and analysis of various events performed by user.

4.5 Applications

- Video Searching: Videos mostly viewed, analysis of videos based upon viewed in various locations, users etc .
- E-Commerce Sites: Finding products maximum searched.

4.6 Hardware Resources Required

Sr. No.	Parameter	Minimum Requirement	Justification
1	CPU Speed	2 GHz	to process the data
2	RAM	3 GB	to provide max speed

Table 4.1: Hardware Requirements

4.7 Software Resources Required

Platform :

(a) Operating System: Ubuntu 64 bit Opensource Operating System

(b) IDE: Eclipse Neon

(c) Programming Languages: : Java, Python

CHAPTER 5

PROJECT PLAN

5.1 Project Estimates

5.1.1 Reconciled Estimates

5.1.1.1 Cost Estimate

Open source software are used. Therefore, no need for licensing cost.

5.1.1.2 Time Estimates

Software will take 6 months to develop.

5.1.2 Project Resources

- 1. RabbitMQ
- 2. MongoDB
- 3. Elastic Search
- 4. SPARK
- 5. Kibana

5.2 Risk Management w.r.t. NP Complete analysis

This section discusses Project risks and the approach to managing them.

5.2.1 Risk Identification

For risks identification, review of scope document, requirements specifications and schedule is done. Answers to questionnaire revealed some risks. Each risk is categorized as per the categories mentioned in [**?**]. Please refer table 5.1 for all the risks. You can refereed following risk identification questionnaire.

(a) Have top software and customer managers formally committed to support the project?: No

(b) Are end-users enthusiastically committed to the project and the system/product to be built?: Yes

(c) Are requirements fully understood by the software engineering team and its customers?: Yes

(d) Have customers been involved fully in the definition of requirements?: Yes

(e) Do end-users have realistic expectations?: Yes

(f) Does the software engineering team have the right mix of skills?: Yes

(g) Are project requirements stable?: No

(h) Is the number of people on the project team adequate to do the job?: Yes

(i) Do all customer/user constituencies agree on the importance of the project and on the requirements for the system/product to be built?: Yes

5.2.2 Risk Analysis

The risks for the Project can be analyzed within the constraints of time and quality

ID	Risk Description	Probability	Impact		
			Schedule	Quality	Overall
1	Description 1	Low	Low	High	High
2	Description 2	Low	Low	High	High

Table 5.1: Risk Table

5.2.3 Overview of Risk Mitigation, Monitoring, Management

Following are the details for each risk.

Probability	Value	Description
High	Probability of occurrence is	$> 75\%$
Medium	Probability of occurrence is	$26 - 75\%$
Low	Probability of occurrence is	$< 25\%$

Table 5.2: Risk Probability definitions

Impact	Value	Description
Very high	$> 10\%$	Schedule impact or Unacceptable quality
High	$5 - 10\%$	Schedule impact or Some parts of the project have low quality
Medium	$< 5\%$	Schedule impact or Barely noticeable degradation in quality Low Impact on schedule or Quality can be incorporated

Table 5.3: Risk Impact definitions

Risk ID	1
Risk Description	Description 1
Category	Development Environment.
Source	Software requirement Specification document.
Probability	Low
Impact	High
Response	Mitigate
Strategy	Strategy
Risk Status	Occurred

Risk ID	2
Risk Description	Description 2
Category	Requirements
Source	Software Design Specification documentation review.
Probability	Low
Impact	High
Response	Mitigate
Strategy	Better testing will resolve this issue.
Risk Status	Identified

Risk ID	3
Risk Description	Description 3
Category	Technology
Source	This was identified during early development and testing.
Probability	Low
Impact	Very High
Response	Accept
Strategy	Example Running Service Registry behind proxy balancer
Risk Status	Identified

5.3 Project Schedule

5.3.1 Project Task Set

Major Tasks in the Project stages are:

- Task 1: Deciding input, output and scope
- Task 2: Acquiring the data-set
- Task 3: Pre-processing data-set and creating database
- Task 4: Deciding and implementation of algorithm
- Task 5: Creating GUI

5.3.2 Task Network

Project tasks and their dependencies are noted in this diagrammatic form.

5.3.3 Timeline Chart

A project timeline chart is presented. This may include a time line for the entire project.

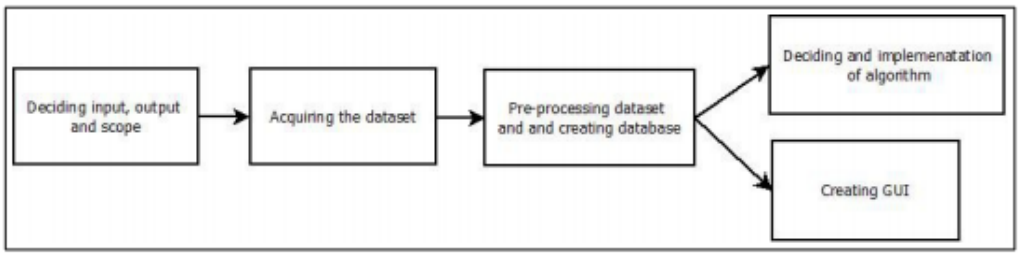


Figure 5.1: Task Dependency diagram

5.4 Team Organization

The manner in which staff is organized and the mechanisms for reporting are noted.

5.4.1 Team Structure

- We work in group of 4 members. We communicate with help of Apps like WhatsApp, Hangout. We have meeting every week with our guide to discuss the progress and hurdles which we face and try to find a feasible solution for the problem.

5.4.2 Management Reporting and Communication

Mechanisms for progress reporting and inter/intra team communication are identified as per assessment sheet and lab time table.

CHAPTER 6

SOFTWARE REQUIREMENT SPECIFICATION

6.1 Introduction

6.1.1 Purpose and Scope of Document

An SRS is basically an organizations understanding (in writing) of a customer or potential clients system requirements and dependencies at a particular point in time (usually) prior to any actual design or development work. Its a two-way insurance policy that assures that both the client and the organization understand the others requirements from that perspective at a given point in time.

- It provides feedback to the customer.
- It decomposes the problem into component parts.
- It serves as an input to the design specification.
- It serves as the parent document.

6.1.2 Overview of Responsibilities of Developer

- Initial research duties for a product development engineer include identifying the needs and goals for a new product, from function to aesthetics.
- They often coordinate with market researchers to evaluate market needs, existing competition and potential costs.
- The primary responsibility of development engineers is to create a product design that fulfills a company or clients strategic goals.
- They oversee research and design teams, lead testing procedures and draft specifications for manufacturing.
- They direct the creation of models or samples and fine-tune designs until they are ready for production.

6.2 Usage Scenario

This section provides various usage scenarios for the system to be developed.

6.2.1 User Profiles

Collect various events performed by user. Store it in database. Analyze data, predict result by using machine learning algorithm.

6.2.2 Use-Cases

All use-cases for the software are presented. Description of all main Use cases using use case template is to be provided.

Sr No.	Use Case	Description	Actors	Assumptions
1	Event collection	collect events performed by the user	-	All events performed by user are collected

Table 6.1: Use Cases

6.2.3 Use Case View

Use Case Diagram. Example is given below

6.3 Data Model and Description

6.3.1 Data Description

Data objects that will be managed/manipulated by the software are described in this section. The database entities or files or data structures required to be described. For data objects details can be given as below

6.3.2 Data Objects and Relationships

Data objects and their major attributes and relationships among data objects are described using an ERD- like form.

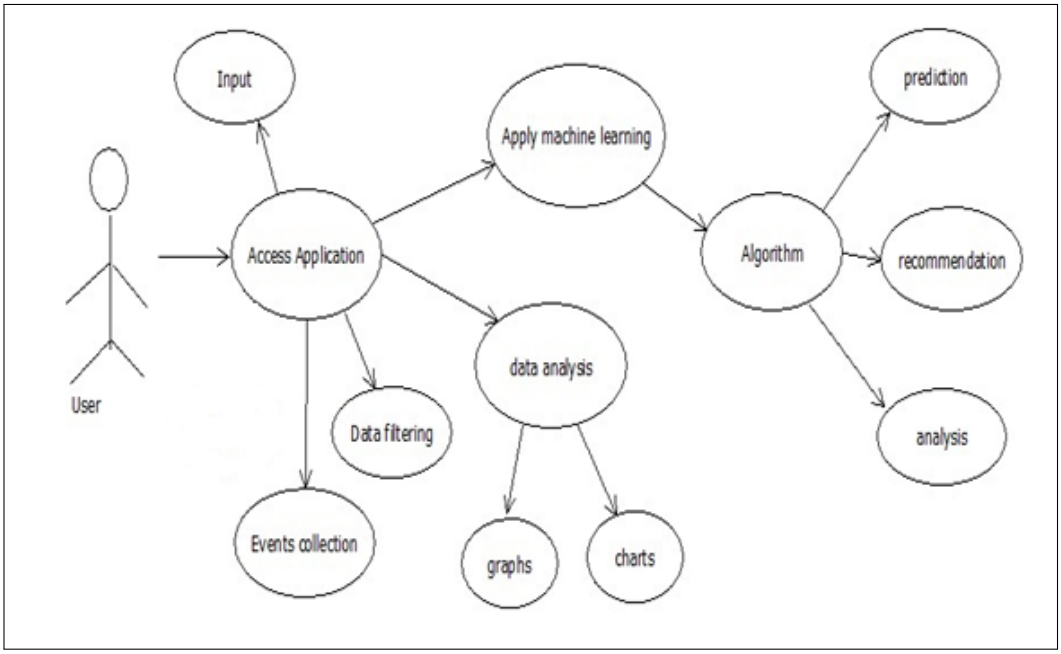


Figure 6.1: Use case diagram

6.4 Functional Model and Description

A description of each major software function, along with data flow (structured analysis) or class hierarchy (Analysis Class diagram with class description for object oriented system) is presented.

6.4.1 Data Flow Diagram

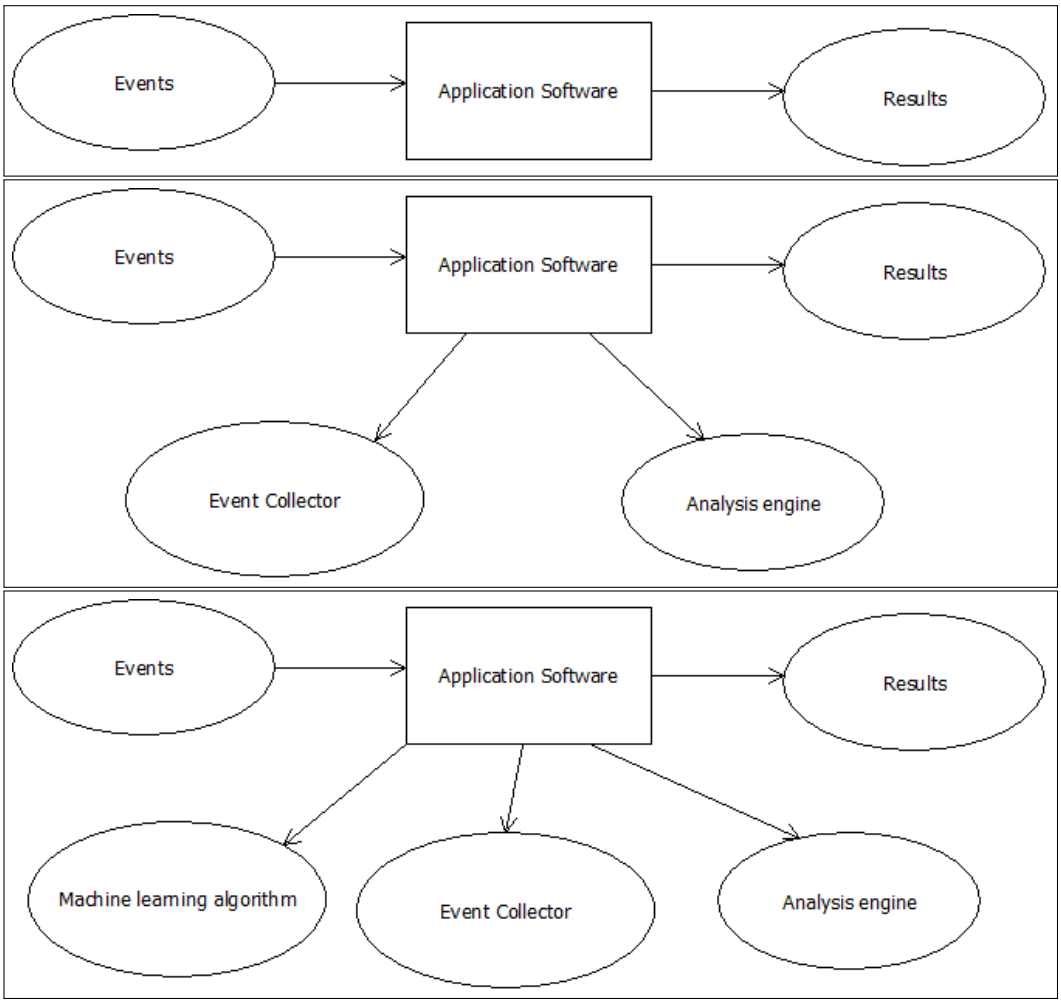


Figure 6.2: Data Flow diagram

6.4.2 Activity Diagram:

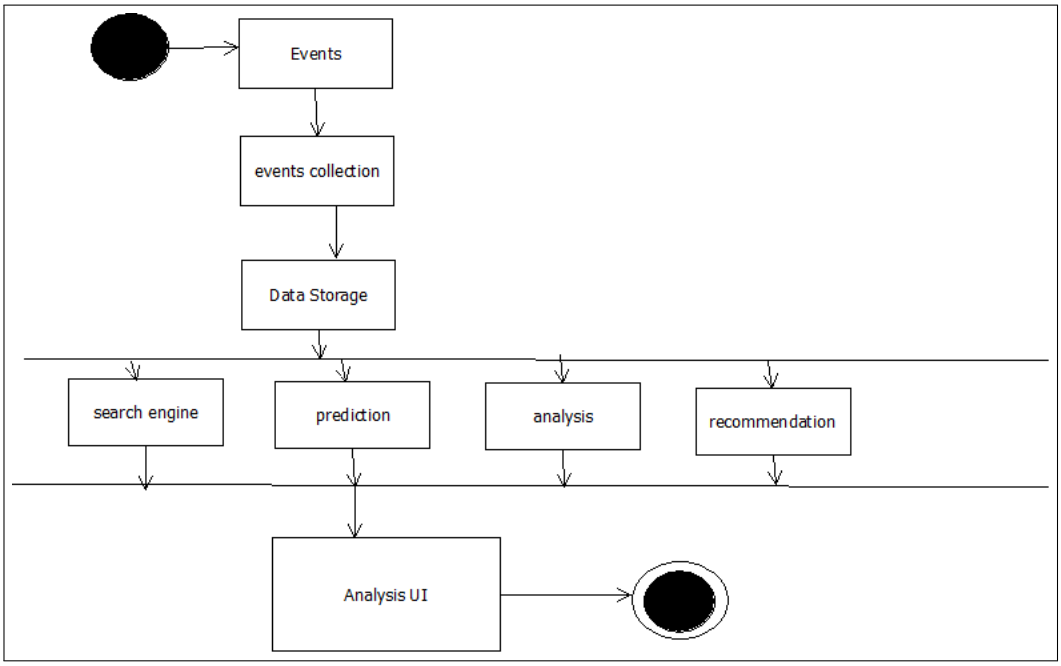


Figure 6.3: Activity Diagram

6.4.3 Non Functional Requirements:

- Interface Requirements
- Performance Requirements
- Software quality attributes such as availability [related to Reliability], modifiability [includes portability, reusability, scalability] , performance, security, testability and usability[includes self adaptability and user adaptability]

6.4.4 State Diagram:

State Transition Diagram
The states are represented in ovals and state of system gets changed when certain events occur. The transitions from one state to the other are represented by arrows. The Figure shows important states and events that occur while creating new project.

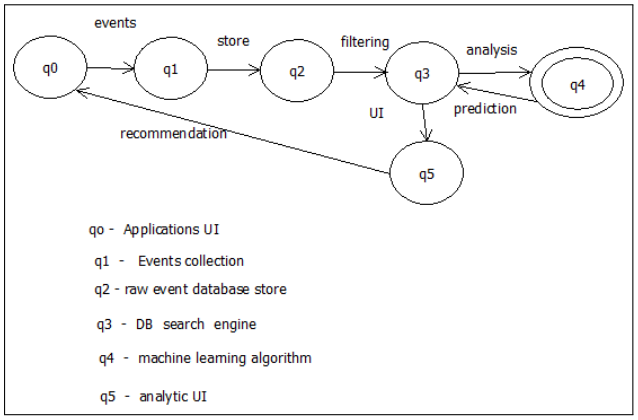


Figure 6.4: State transition diagram

6.4.5 Design Constraints

Any design constraints that will impact the subsystem are noted.

6.4.6 Software Interface Description

The software interface(s)to the outside world is(are) described.
The requirements for interfaces to other devices/systems/networks/human

are stated.

CHAPTER 7

DETAILED DESIGN DOCUMENT USING APPENDIX A AND B

7.1 Introduction

This document specifies the design that is used to solve the problem of Product.

7.2 Architectural Design

A description of the program architecture is presented. Subsystem design or Block diagram,Package Diagram,Deployment diagram with description is to be presented.

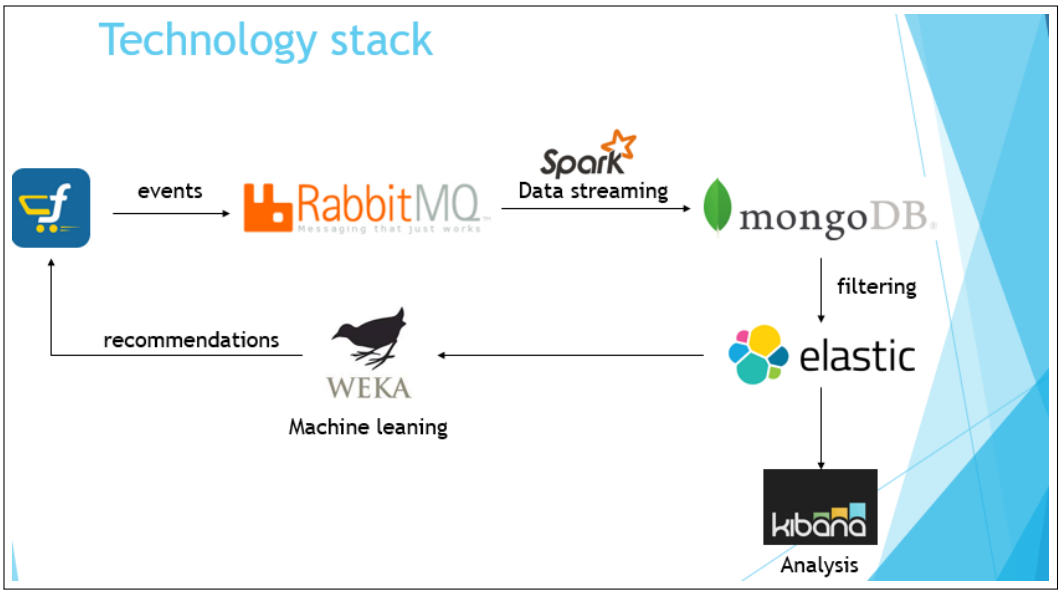


Figure 7.1: Architecture diagram

7.3 Data design (using Appendices A and B)

A description of all data structures including internal, global, and temporary data structures, database design (tables), file formats.

7.3.1 Data structure

Data structures can implement one or more particular abstract data types (ADT), which specify the operations that can be performed on a data structure and the computational complexity of those operations. In comparison, a data structure is a concrete implementation of the specification provided by an ADT Data structures used in project: arrays, lists, vectors

7.3.2 Database description

Each event performed by the user is collected and stored in MongoDB and then sent to Elastic search for analysis.

7.4 Compoent Design

Class diagrams, Interaction Diagrams, Algorithms. Description of each component description required.

7.4.1 Class Diagram

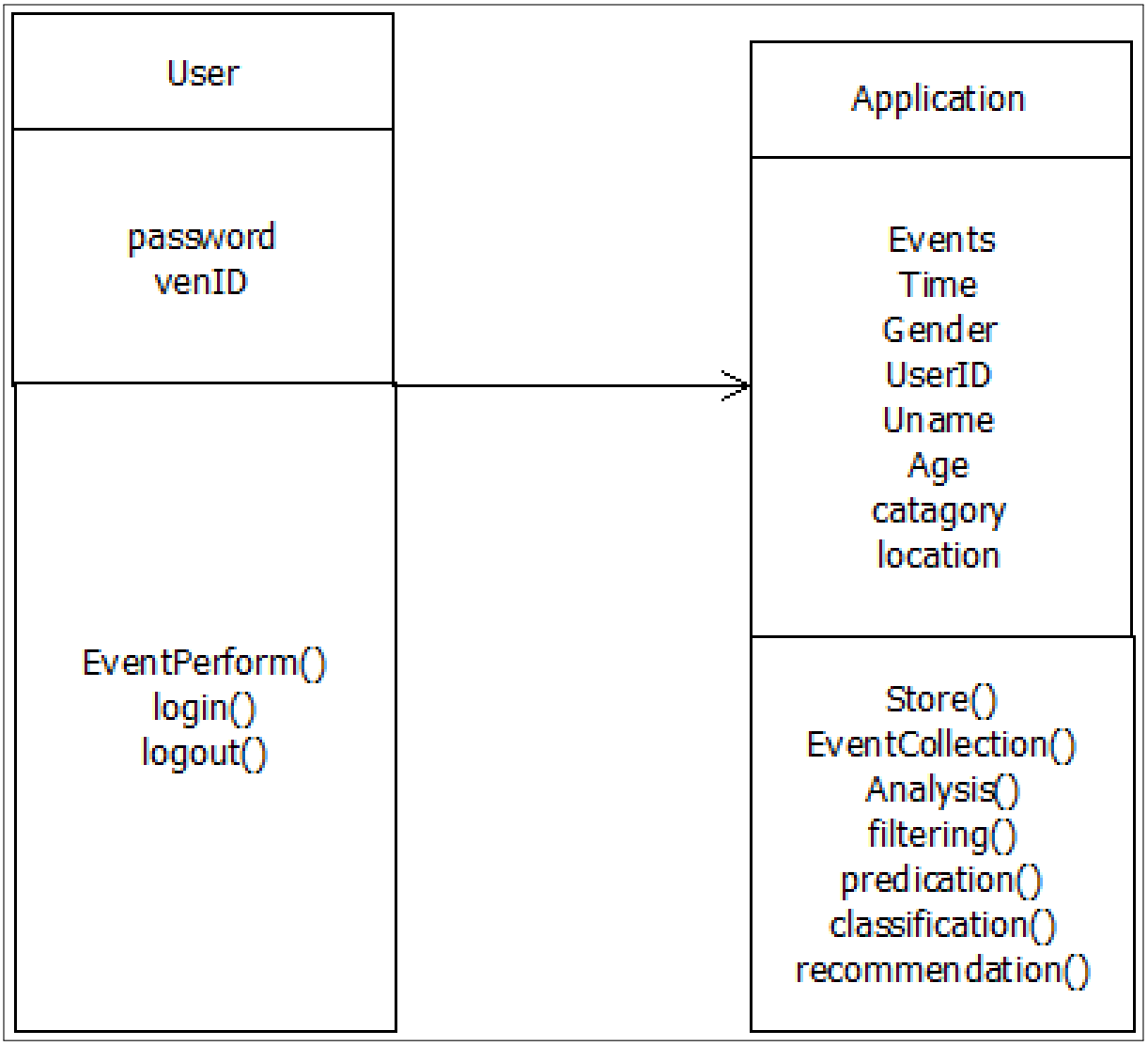


Figure 7.2: Class Diagram

CHAPTER 8

PROJECT IMPLEMENTATION

8.1 Introduction

Data is collected in form of events from the user using RabbitMQ and then streamed through using SPARK and eventually stored in database used, MongoDB. The elastic search assists to build the search engine which helps us to parse through the data and KIBANA to perform various data analytics. Machine learning(J48 algorithm) using WEKA platform is used to build recommendation engine and analysis for vivid functionalities on available data set to provide suggestions, enhancing user experience. Analytics UI helps providing proper understanding of search patterns and user events to the clients to increase efficiency of the application.

8.2 Tools and Technologies Used

- RabbitMQ
- Spark
- MongoDB
- Elastic search
- Kibana
- Weka tool

8.3 Methodologies/Algorithm Details

We have used J48 algorithm and the recommendation algorithm.

8.3.1 Algorithm 1/Pseudo Code

(a) Collect information of user clicks from database.
(b) Train data set from collected information.
(c) Generate decision tree from train data set.
(d) Collect test data set from user for recommendation.
(e) Generate recommendation by applying test data set on decision tree.

CHAPTER 9

SOFTWARE TESTING

9.1 Type of Testing Used

Unit Testing: Unit Testing is a level of software testing where individual units/ components of a software are tested. The purpose is to validate that each unit of the software performs as designed.

9.2 Test Cases and Test Results

Sr. No.	Test Case	Expected Output	Actual Output	Status
1	Database connection	System connected to the database	System connected to the database	Pass
2	event detection	event is caputred	event is caputred	Pass
3	Analysis	proper analysis is shown	proper analysis is shown	Pass
4	recommendation	recommended	recommended	Pass

Figure 9.1: Test Cases

CHAPTER 10

RESULTS

10.1 Screen shots

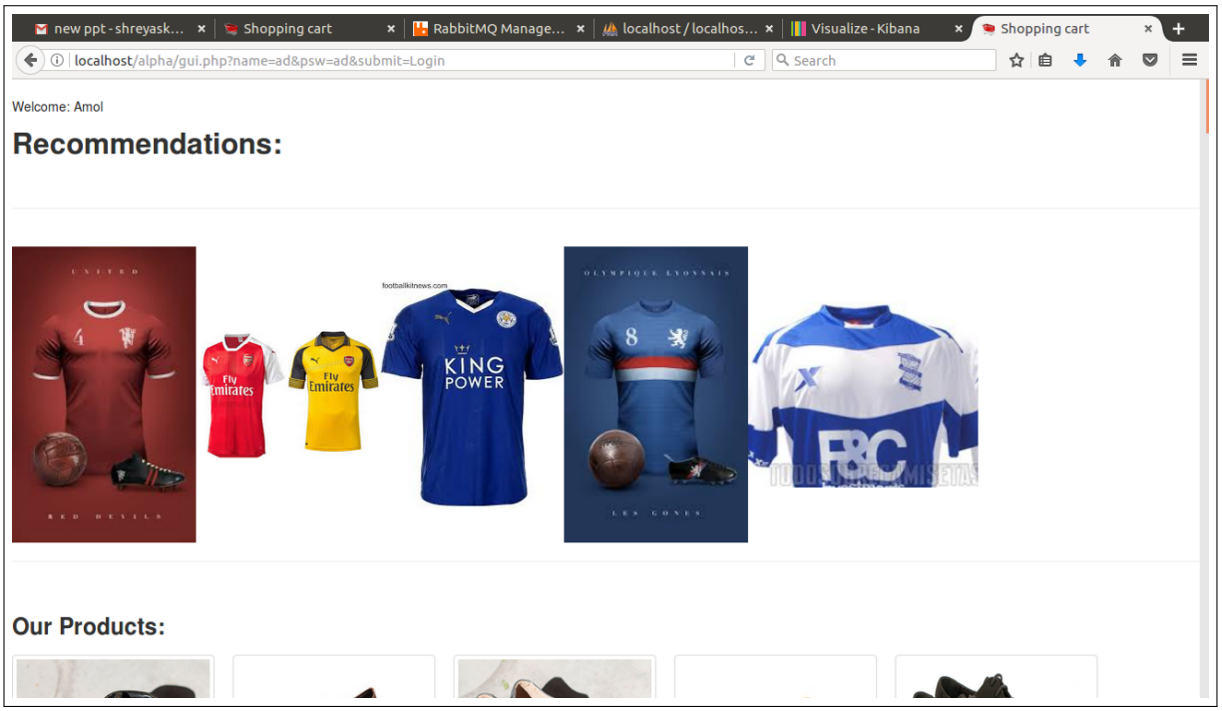


Figure 10.1: Recommendation

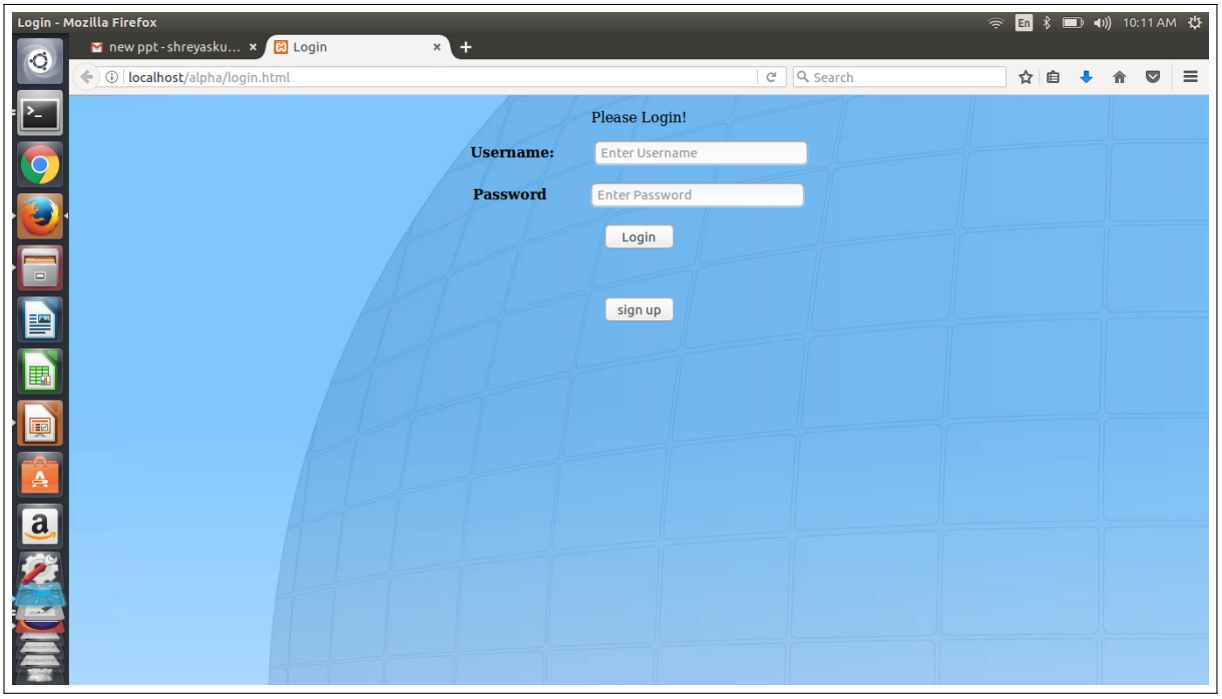


Figure 10.2: Login page

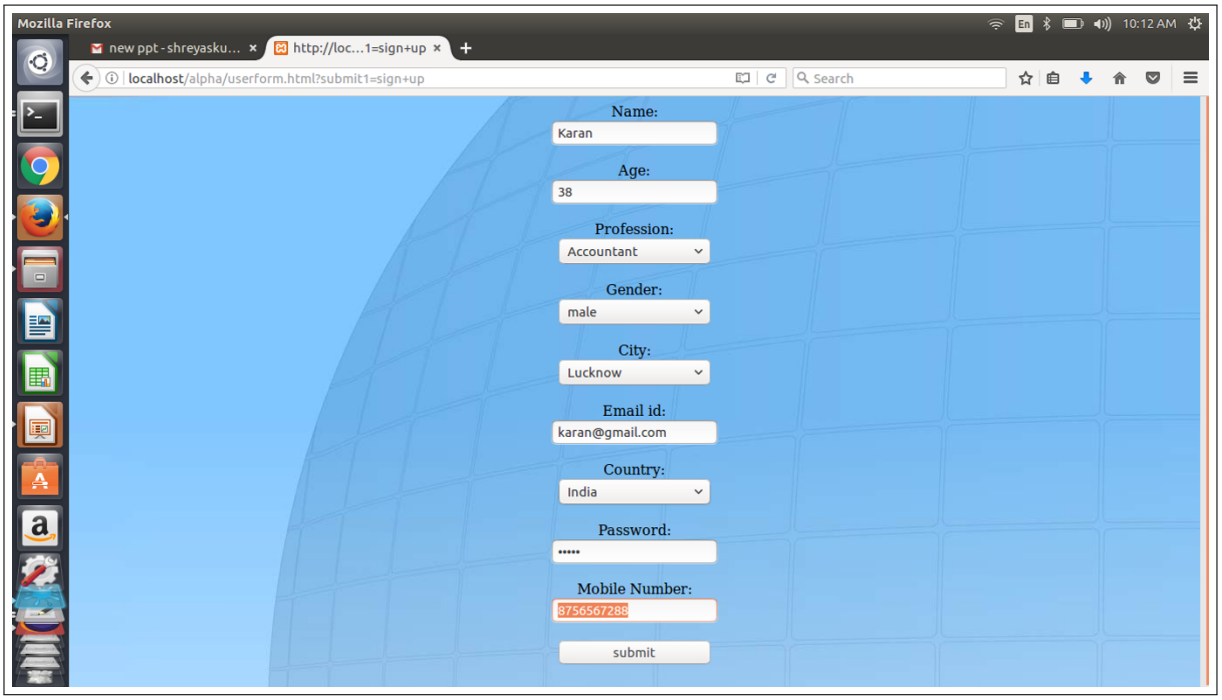


Figure 10.3: User form

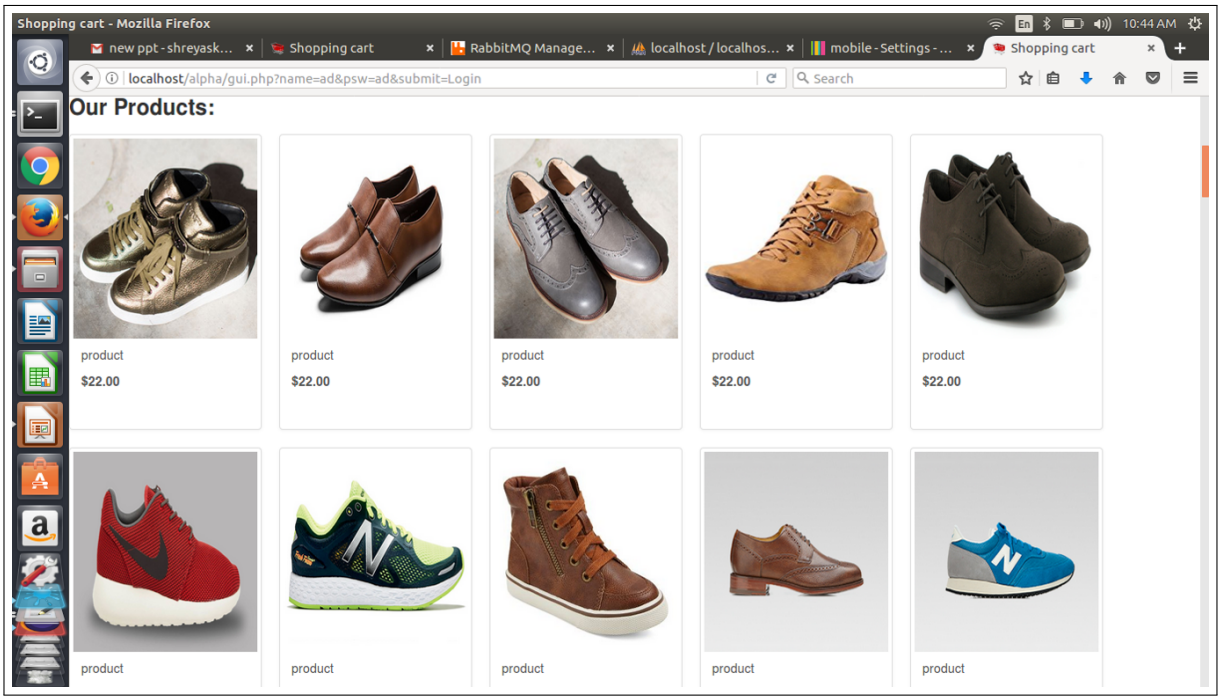


Figure 10.4: Demo website

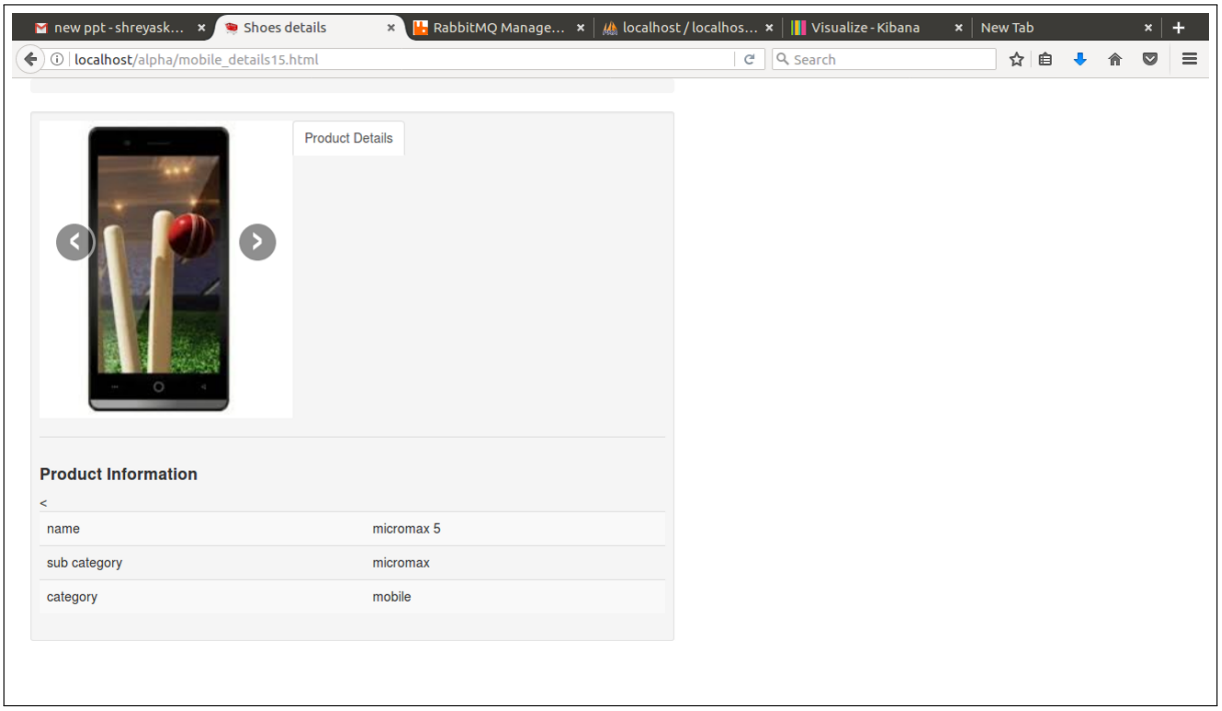
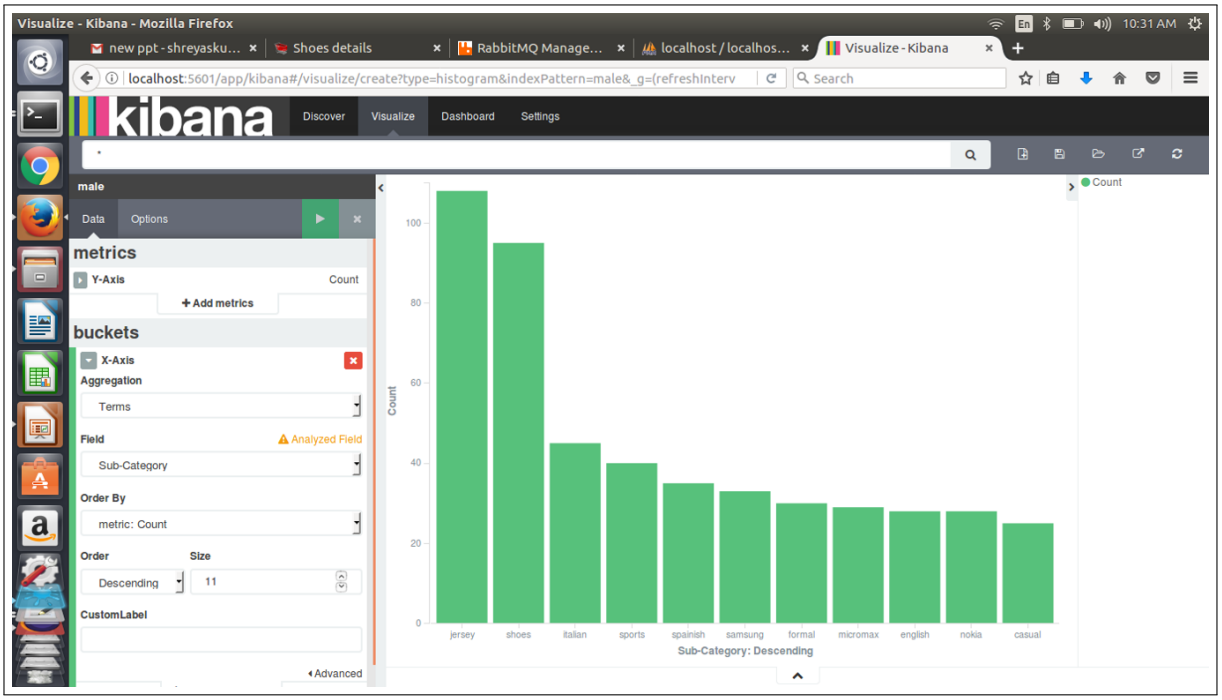


Figure 10.5: Image details


Visualize - Kibana - Mozilla Firefox
10:31 AM
new ppt - shreyasku...
Shoes details
RabbitMQ Manage...
localhost / localhos...
Visualize - Kibana
localhost:5601/app/kibana#/visualize/create?type=histogram&indexPattern=male&_g=(refreshInterv
Search
kibana
Discover
Visualize
Dashboard
Settings
male
Data
Options
metrics
Y-Axis
Count
+ Add metrics
buckets
X-Axis
Aggregation
Terms
Field
Analyzed Field
Sub-Category
Order By
metric: Count
Order
Size
Descending
11
CustomLabel
Advanced
Count
100
80
60
40
20
0
jersey
shoes
italian
sports
spainish
samsung
formal
micromax
english
nokia
casual
Sub-Category: Descending

CHAPTER 11

DEPLOYMENT AND MAINTENANCE

11.1 Installation and Un-Installation

(a) Since there are two types of users for our System Customer and Vendor.

(b) Customer just need to access websites of Vendors so no need of installation for Customer.He will only require a browser.

(c) Vendor will require to install Kibana for accessing analysis.

(d) When we uninstall kibana , all its dependencies automatically get removed.

11.2 User Help

For Users

(a) Fill all the details while signing up in the userform..

(b) Login with email-id and password.

(c) Click on the desired image so that in the backend user details and product details will be inserted into the database to perform analysis and display recommendation.

For Vendors

(a) In the backend, changes should be made so the javascript file will be called and details should fetched for every event performed by the user.

(b) Install kibana to perform appropriate queries so to perform the required analysis and customise the system.

(c) GUI should be dynamically changed so that proper recommendation will be displayed for every user.

CHAPTER 12

CONCLUSION AND FUTURE SCOPE

- In this project, we have designed optimization framework operates on various systems. The technology stack used can be used to improve vendors performance efficiently. Analysis can be correctly performed based on various factors and criterion which depends on the argument provided to this framework by vendors according to their requirements. So that vendor can optimize their system and thereby increase efficiency and profit. It is generic framework so performance depends on the application and their arguments. As dataset increases, accuracy of the results also increases. J48 algorithm using weka platform is used to recommend products to the users based on raw data collected. So, this framework can be used almost everywhere as recommendation engine for users and analysis for vendors displayed using various geomentric diagrams.
- Currently we are giving recommendation on user attributes , but we can also give recommendation based on both , user profiles and user events.Also we will be able to predict user attributes if the person logs in as guest based on the the events he performed.

ANNEXURE A

REFERENCES

[1]Hina Gulati, Predictive Analytics Using Data Mining Technique, 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom). Ticketing System 2015 International Conference on Green Computing and Internet of Things (ICGCIoT)
[2] Yiannakis Sazeides and James E. Smith Department of Electrical and Computer Engineering University of Wisconsin-Madison, The Predictability of Data Values, 14 15 Engr. Dr. Madison, WI 53706, ieee-paper 2014 .
[3] Jun Bai, Northern BeiJing Vocational Education Institute BeiJing, University of Wisconsin-Madison, Feasibility Analysis of Big Log Data Real Time Search Based on Hbase and ElasticSearch, 2013 Ninth International Conference on Natural Computation (ICNC)
[4]. Oleksii Kononenko, Olga Baysal, Reid Holmes, and Michael W. Godfrey David R. Cheriton School of Computer Science University of Waterloo, Waterloo, ON, Canada, Mining Modern Repositories with 2015Elaticsearch.
[5].Pragya Gupta1, Sreeja Nair, Survey Paper on Elastic Search, International Journal, 2014.
[6]. Dalip, Vijay Kumar Ph.D., GPS and GSM based Passenger Tracking System, International Journal of Computer Applications (0975 8887) Volume 100 No.2, August 2014.
[7].Jay Kreps, LinkedIn Corp, Neha Narkhede, LinkedIn Corp. Jun Rao, LinkedIn Corp. Kafka: a Distributed Messaging System for Log Processing, NetDB'11, Jun. 12, 2011, Athens, Greece. Copyright 2011 ACM 978-1-4503-0652-2/11/06.
[8]. Urvi Thacker , Manjusha Pandey, Siddharth S. Rautaray,School of Computer Engineering KIIT University Bhubaneswar, India Performance of Elasticsearch in Cloud Environment with nGram and non-nGram indexing, International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) - 2016.'
[9].Xue-meng Li, Yong-yi Wang Network Engineering Department,PLA Electronic Engineering Institute,Hefei, ChinaDesign and Implementation of an Indexing Method Based on Fields for Elasticsearch, 2015 Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control,ieee 2015,china.
[10].Parneet Kaura,Manpreet Singhb,Gurpreet Singh Josanc aScholar, Department of CSE, Punjab Technical University,Jalandhar 144603,India Assistant Professor, Department CSEIT, GNDEC, Ludhiana, Punjab, India Classification and prediction based data mining al-

gorithms to predict slow learners in education sector, 3rd International Conference on Recent Trends in Computing 2015(ICRTC-2015).

ANNEXURE B

LABORATORY ASSIGNMENTS ON PROJECT ANALYSIS OF ALGORITHMIC DESIGN

- To develop the problem under consideration and justify feasibilty using concepts of knowledge canvas and IDEA Matrix. Refer [**?**] for IDEA Matrix and Knowledge canvas model. Case studies are given in this book. IDEA Matrix is represented in the following form. Knowledge canvas represents about identification of opportunity for product. Feasibility is represented w.r.t. business perspective.

- Project problem statement feasibility assessment using NP-Hard, NP-Complete or satisfy ability issues using modern algebra and/or relevant mathematical models.
- input x,output y, y=f(x)

ANNEXURE C

LABORATORY ASSIGNMENTS ON PROJECT QUALITY AND RELIABILITY TESTING OF PROJECT DESIGN

It should include assignments such as

- Use of divide and conquer strategies to exploit distributed/parallel processing of the above to identify object, morphisms, overloading in functions (if any), and functional relations and any other dependencies (as per requirements). It can include Venn diagram, state diagram, function relations, i/o relations; use this to derive objects, morphism, overloading
- Use of above to draw functional dependency graphs and relevant Software modeling methods, techniques including UML diagrams or other necessities using appropriate tools.
- Testing of project problem statement using generated test data (using mathematical models, GUI, Function testing principles, if any) selection and appropriate use of testing tools, testing of UML diagram's reliability. Write also test cases [Black box testing] for each identified functions. You can use Mathematical or equivalent open source tool for generating test data.

ANNEXURE D

PROJECT PLANNER

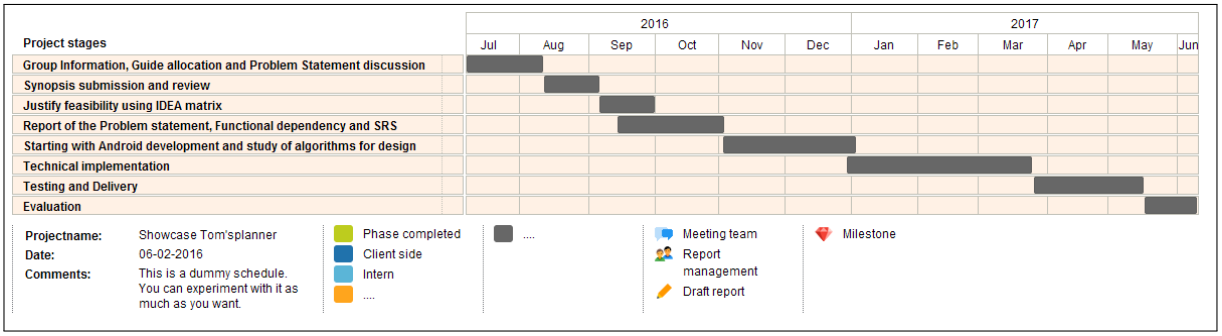


Figure D.1: Planner

ANNEXURE E

REVIEWERS COMMENTS OF PAPER SUBMITTED

(a) Paper Title:Generic user activity analysis and prediction.

(b) Name of the Conference/Journal where paper submitted : IRJET

(c) Paper accepted/rejected : accepted

(d) Review comments by reviewer : -

(e) Corrective actions if any : -

ANNEXURE F

PLAGIARISM REPORT

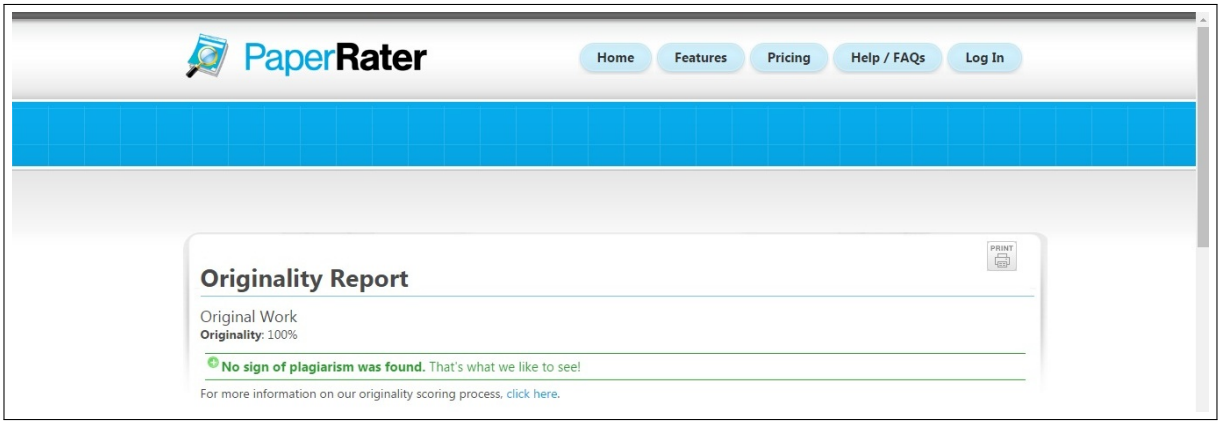


Figure F.1: Plagiarism

ANNEXURE G

TERM-II PROJECT LABORATORY ASSIGNMENTS

(a) Review of design and necessary corrective actions taking into consideration the feedback report of Term I assessment, and other competitions/conferences participated like IIT, Central Universities, University Conferences or equivalent centers of excellence etc.

(b) Project workstation selection, installations along with setup and installation report preparations.

(c) Programming of the project functions, interfaces and GUI (if any) as per 1 st Term term-work submission using corrective actions recommended in Term-I assessment of Term-work.

(d) Test tool selection and testing of various test cases for the project performed and generate various testing result charts, graphs etc. including reliability testing.

(e) Installations and Reliability Testing Reports at the client end.

ANNEXURE H

INFORMATION OF PROJECT GROUP MEMBERS

(a) Name : Kiran Mokashi
(b) Date of Birth : 27/03/1996
(c) Gender : Male
(d) Permanent Address : Govind Complex, Flat No-4, BalajiNagar, Pune-411043
(e) E-Mail : kiranmokashi27@gmail.com
(f) Mobile/Contact No. : 9767641743
(g) Placement Details : Placed at Veritas
(h) Paper Published : Yes

(a) Name : Shailesh Bagade
(b) Date of Birth : 21/03/1995
(c) Gender : Male
(d) Permanent Address : krushkunj, siddhivinayak nagar, ghatpuri road, khamgaon
(e) E-Mail : shaileshvbagade@gmail.com
(f) Mobile/Contact No. : 8275028550
(g) Placement Details : Placed at Cybage.
(h) Paper Published : Yes

(a) Name : Shivam Bawane
(b) Date of Birth : 27/02/1995
(c) Gender : Male
(d) Permanent Address : 59B, shankar nagar , nagpur-440010.
(e) E-Mail : shivambawane@gmail.com
(f) Mobile/Contact No. : 7774037664
(g) Placement Details : Not placed.
(h) Paper Published : Yes

(a) Name : Shreyas Kulkarni
(b) Date of Birth : 20/01/1996
(c) Gender : Male
(d) Permanent Address : Flat no 11,Ganadhish appt.,Narsinh nagar,Gangapur road,Nashik
(e) E-Mail : shreyaskulkarni20@gmail.com
(f) Mobile/Contact No. : 8308202837
(g) Placement Details : Placed at Accenture
(h) Paper Published : Yes