

SMART SHOPPING ENVIRONMENT USING CLOUD AND REGRESSION

Gaurav Patil*¹, Rushikesh Gorde*², Kunal Thakare*³, Akash Ghodke*⁴,

Mr. D.S.Rakshe*⁵

*^{1,2,3,4}Department Of Computer Engineering , PREC, Loni, Maharashtra, India.

*⁵Ass. Professor Department Of Computer Engineering, PREC, Loni, Maharashtra, India.

ABSTRACT

Today online shopping is a hot trend. With the increase in market share of online shopping entities many consumers are suggesting small time shopping entities to go online. But mere going online will not help a shop as it will be hard to handle the demand and sale using the existing shopping structure. Currently the shop knows what they have to order as the consumer is front of them. But in online structure it is not possible as impulse buying of consumers where a certain product will be in high demand causing shortage and out of stock of the products. Due to out of stock a consumer will go to other online sites and thus causing a shop to lose his customer. So, to handle the impulse buying shortage we are suggesting a framework with the combination of cloud computing, mobile and machine learning together to analyses and stop impulse buying shortage. Thus, in our project we will design an online shopping scenario where a shop admin will first enter product details for sale. The admin application will be standalone application which will also handle demand and sale using machine learning. Then a mobile application will be provided on the customer side as mobile is used by all and provides good medium for interaction in an online shopping scenario. Thus, a customer will order the product. For the communication between an admin and a customer we are going to use Google Cloud Platform as our cloud provider. On the admin side we will use machine learning algorithm Linear regression to predict which products will be sold more and cause a shortage. The training dataset used to train linear regression will be custom built according to our need. Thus, by using our system an online shop admin will have good insights on the demand and sale of products and he can handle the stock of such products efficiently.

Keywords: Online Shopping, Impulse Buying, Cloud Computing, Mobile Computing, Machine Learning, Linear Regression

I. INRODUCTION

Impulse buying is trend present in online buying of products, where a consumer has a specific taste or need of a product depending upon there sale and availability. As more and more consumers are opting for online shopping due to development and economical status, they tend to buy more and more products. These purchases give rise to impulse buying as within seconds a product can be sold out and a shop owner or entity would grapple to make the stock available. This new trend seen in online shopping has become a nuisance for a shop entity as that have to study which products will come under impulse buying. So, there is a need to study the impulse buying behavior so that a shop entity can understand which product he has to stock up and which not. As more and more shops go online to compete there arises a need to understand the online demand and sale of products. In online shopping scenario impulse buying of specific products causes time and money for a shop owner. If impulse buying not understood properly it may give rise to shortage of products or stock outage. Thus, stock outage creates a bad name for the shopping entity as customer will go to other sites for the product. To meet the demand after a stock outage will take some time. This will cost the business dearly. So there arises a need to understand impulse buying scenario and be prepared for the same by scaling of the product stock. If the impulse buying scenario is wrong than a shop owner loses money by ordering wrong products instead of right ones, so this kind of buying error has also be considered during designing the system. So to over come the existing system and create a new smart shopping environment system we are going to develop a project with a combination of cloud computing, mobile computing and desktop computing together. We are going to use machine learning algorithm Linear regression to help a shop owner which products will come under impulse buying.

II. METHODOLOGY

1. System Architecture

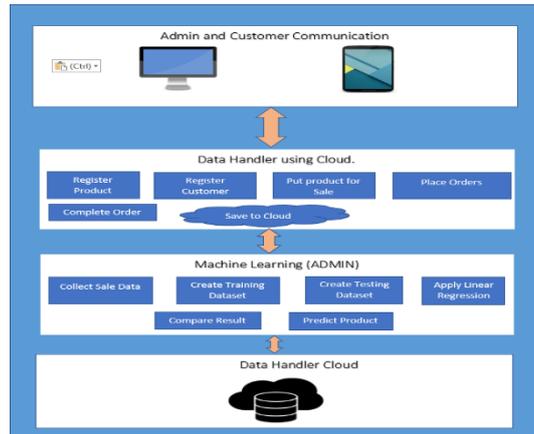


Figure 1: System Architecture

2. Mathematical Model

Our projects mathematical perspective can be put and described as given below. 1 Set Theory Applied to the Project

1. Data Handler: -

Set(D)={D0, D1, D2, D3, D4, D5} D0 ∈ D = Enter product details. D1 ∈ D = Register customer on mobile. D2 ∈ D = View products. D3 ∈ D = Place order. D4 ∈ D = Complete payment. D5 ∈ D = Get Acknowledgement. 2. Machine Learning: - Set(M)={M0, M1, M2, M3, D2} M0 ∈ M = Get product sale data. M1 ∈ M = Create training dataset. M2 ∈ M = Apply Linear Regression. M3 ∈ M = Predict products in impulse buying range. D2 ∈ M = View products.

2. Probability

NP-Hard and NP-Complete So, by studying the sets as defined above we come to notice that a element D2 is common in both modules and used in coordination in both sets which can be placed as $x \in D \cap M$ if $x \in D$ and $x \in M$ (4.1) Thus, the probability of intersection of element in both modules can be given as $P(D \cap M) = P(D) + P(M)$ (4.2) So, intersection of common element can be shown as $D \cap M = \{D2\}$ (4.3) The conditional probability of both modules using the same element can be shown as $P(D|M) = P(D \cap M) P(M)$ (4.4) Thus, we conclude that our project "Smart Shopping Environment Using Cloud and Regression" success and failure will depend upon the internet as our shop admin has to enter product info and save to cloud for a customer to buy, i.e., if the internet connection is not good or not present the product info cannot be saved and will not be fetched thus the project won't work, this is a case of failure, so our project supports NP-Hard and not NP Complete.

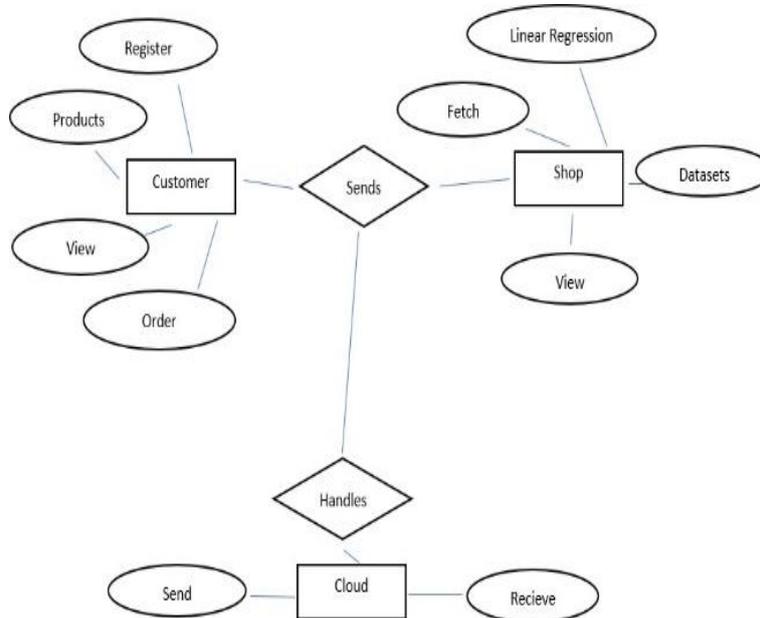


Figure 2: E-R diagram

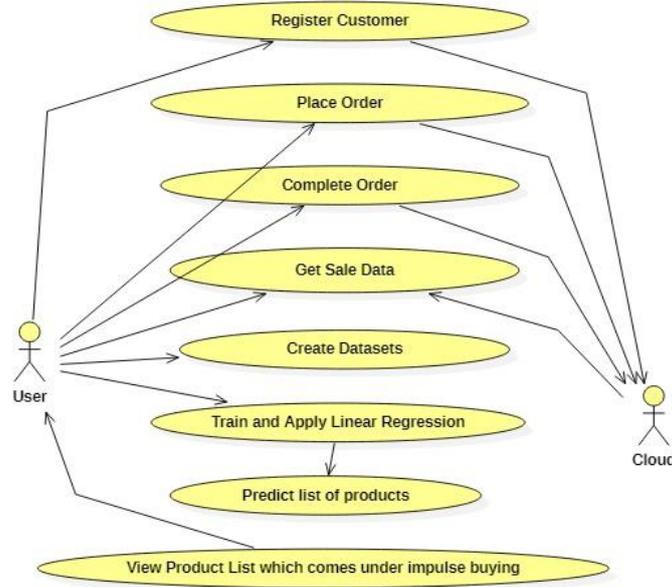


Figure 3: Use Case

III. MODELING AND ANALYSIS

Cost Estimates :-

The number of lines required for implementation of various modules can be estimated as follows:

Table 1: KLOC model for project

Sr.No.	Modules	KLOC
1.	Customer app GUI	0.70
2.	Admin desktop app GUI	0.50
3.	Mobile and Cloud interface code	0.60
4.	Desktop and cloud interface code	0.70
5.	Machine learning code	0.30

Thus total number of lines required is approximately estimated as 2.80 KLOC.

Time Estimates

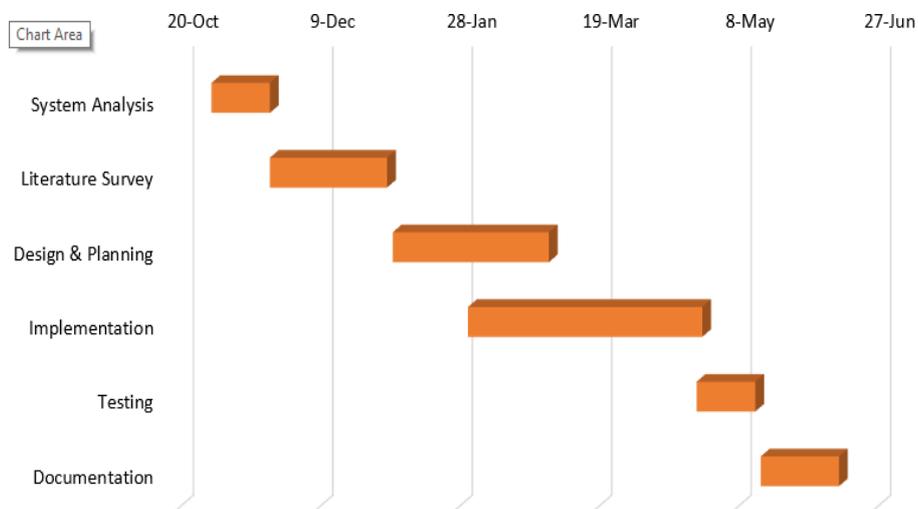


Figure 4: Phase wise Gantt Diagram

Project Task Set

Major tasks in the project can be explained as:

1. Task-1 :- Register customer.

2. Task-2 :- Enter product.
3. Task-3 :- Purchase product.
4. Task-4 :- dispatch order.
5. Task-5 :- Apply machine learning.
6. Task-6 :- View regression results

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 for all the risks. You can have refereed following risk identification questionnaire.

1. Have top software and customer managers committed to support the project? Ans: Yes
2. Are end-users enthusiastically committed to the project and the system/product to be built? Ans: Yes
3. Are requirements fully understood by the software engineering team and its customers? Ans: Yes
4. Have customers been involved fully in the definition of requirements? Ans: Yes
5. Do end-users have realistic expectations? Ans: No
6. Does the software engineering team have the right mix of skills? Ans: Yes
7. Are project requirements stable? Ans: Yes

Risk Analysis

Risk analysis and management are a series of steps that help a software developer to understand and manage uncertainty. The software developer establishes a plan for managing the risks. The primary objective is to avoid the risk. The developer works to develop a contingency plan that will enable it to respond in a controlled and effective manner.

Timeline Chart

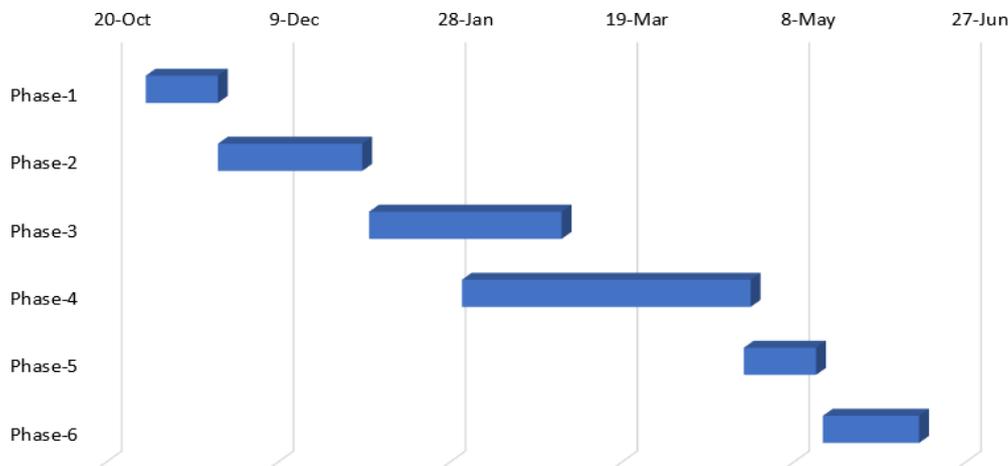


Figure 5: Timeline Chart Gantt Diagram

Project Implementation

Customer module :- This module is a mobile application. In this module a customer will register first. Then he will login using the credentials and view menu. From menu first he will view the products for sale.

Admin module :- This module is a desktop application. In this module an admin will view products, customers in a table view. The admin will then create a training dataset using the qty, rate and amount as parameters. The training dataset will be used to train the Linear regression algorithm. The admin will retrieve the product sale details and then create a training dataset for analysis.

Google cloud module :- This module is used as a communication medium between customer and admin. We are using Google sheets as backend which is a non-SQL database. It is free without charges and available 24X7.

Tools and Technologies Used

Java Development Tool Kit:- (JDK) The Java Development Kit (JDK) is an implementation of either one of the Java SE, Java EE or Java ME platforms released by Oracle Corporation

Python :- Python is an associate degree taken high-level general artificial language. Its language constructs in addition as its object-oriented approach aim to assist programmers write clear, logical code for little and large-scale comes. Python is dynamically-typed and garbage-collected.

NetBeans :- NetBeans is an associate degree integrated development atmosphere (IDE) for Java. NetBeans permits applications to be developed from a group of standard code elements known as modules..

Android :- Android may be a mobile software package (OS) initially developed by a geographical area company by the name of mechanical man opposition. A collaboration spearheaded by Google in 2007 through the Open phone Alliance (OHA) gave mechanical man a foothold in delivering an entire software package set, which has the most OS, middleware and specific mobile application, or app. Development for the mechanical man could also be done through Windows, UNIX system or waterproof..

Google Spreadsheet :- Google Sheets may be a programme program enclosed as a part of a free, web-based software package workplace suite offered by Google among its Google Drive service. The service conjointly includes Google Docs and Google Slides, a word processing system and presentation program severally. Google Sheets is accessible as an internet application, mobile app for mechanical man, iOS, Windows, BlackBerry, and as a desktop application on Google's ChromeOS.2. Algorithm Used :-

Linear Regression:- In statistics, regression toward the mean may be a linear approach to modelling the link between a scalar response and one or a lot of instructive variables (also called dependent and freelance variables). The case of 1 instructive variable is named easy linear regression; for over one, the method is named multiple regression toward the mean. This term is distinct from variable regression toward the mean, wherever multiple correlate dependent variables area unit foreseen, instead of one scalar variable. In regression toward the mean, the relationships area unit shapely mistreatment linear predictor functions whose unknown model parameters area unit calculable from the information. Such models area unit known as linear models. most ordinarily, the conditional mean of the response given the values of the instructive variables (or predictors) is assumed to be an associate degree affine operate of these values; less unremarkably, the conditional median or another quantile is employed. Like all styles of multivariate analysis, regression toward the mean focuses on the probability distribution of the response given the values of the predictors, instead of on the probability distribution of all of those variables, that is that the domain of statistical method.

Testing

Integration Testing :- Integration testing is the software testing process where individual units are combined and tested as a group. The purpose of this testing is to expose faults in the interaction between integrated units.

Acceptance Testing :- It is performed with realistic and various image data for training to demonstrate that the software is working satisfactorily in testing phase. User acceptance is a critical phase of any project and requires significant participation by the end user. It ensures that system meets the functional requirements. Acceptance of end users is done by proper images output.

Validation Testing :- Validation testing will be done to ensure validation of the client's requirements given to the software team for the system compared to the performance and results from the software system. Given that the system has passed all the previous unit and integration testing, validating the system will measure an aspect of an implementations behavior against an expectation and check whether the software complies with that expectation.

Performance Testing :- Performance testing is the testing, which is performed, to ascertain how the components of a system are performing, given a particular situation. In proposed system it is mainly focused on increasing the performance of system to increase the accuracy of the output and analysis of result is done by using accuracy and time parameters.

Test Case And Test Result

Table 2 : Test Case For Customer Module

Test Id	Description	Expected result	Remark
TC-01	Register customer	Registration success message	Pass
TC-02	Login customer	Menu screen after success	Pass

TC-04	View product info	Seen in table view	Pass
TC-03	Place order	success message	Pass
TC-04	View order Status	Seen in table view	Pass

Table 3: Test case For Desktop Module

Test Id	Description	Expected result	Remark
TD-01	Login admin	Menu screen after success	Pass
TD-01	Enter product info	Success message	Pass
TD-02	View product info	Seen in table	Pass
TD-03	View product sale info	Seen in table	Pass
TD-04	View pcustomer info	Seen in table	Pass
TD-05	Fetch and apply Linear re- gression on sale dataset	View regression results i table	Pass

IV. RESULT

The final outcome of the project is the shop sales data analysis using linear regression.

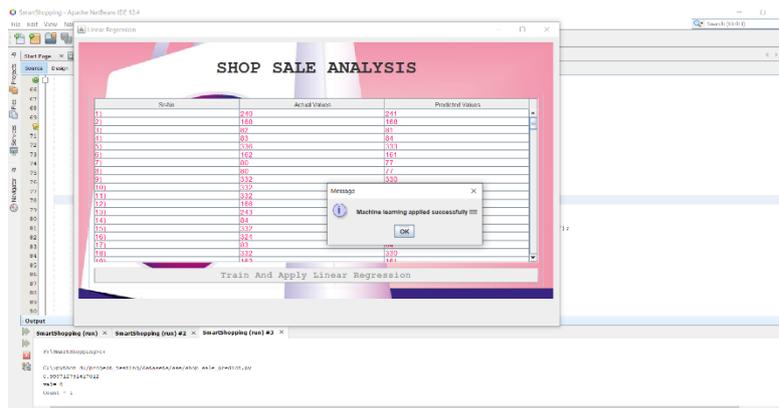


Figure 6 : The Final Outcome

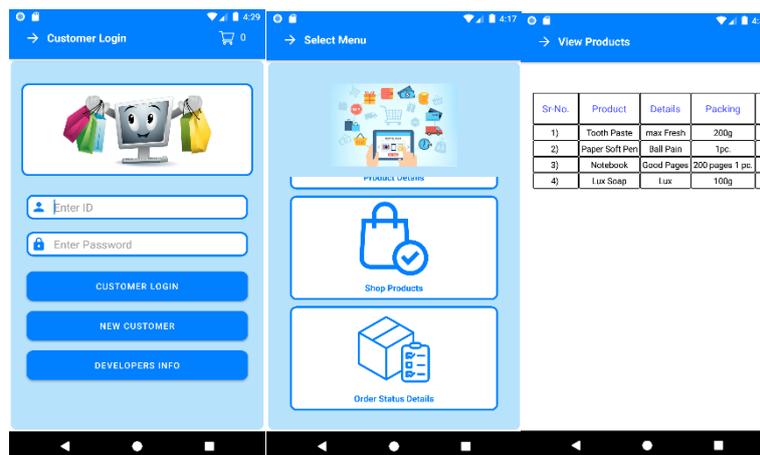


Figure 7: Login Screen **Figure 8 :** Home Menu **Figure 9 :** Product Info

V. CONCLUSION

In this project, we are developed novel collaboration of machine learning, cloud computing, mobile computing and desktop computing together to create a more reliable shopping which will successfully handle impulse buying of products. The basic idea of the project was to design a reliable prediction system for a shopping environment which will predict products which would come under impulse buying. While designing the project we incorporated ideas from [1][2][3] to fit different ideas in one framework. We also have used Linear

Regression algorithm to predict list of products which will come under impulse buying so that a shop can stock up the products. We have also made effective use of mobile computing to design a mobile app on customer side as mobiles are used by everybody today. We made use of Google Cloud Platform as our cloud provider. Thus, we conclude that we will save time and money of a shop and make stock management easy during impulse buying.

Future Work

In our project's future work we will try to design a iOS based mobile application for Apple mobiles and try to use more machine learning algorithms for Agri product analysis.

Applications

Thus applications of the system that can be used for development of other ideas can be given as.

1. Online food ordering system.
2. Online grocery sales.
3. Online Agricultural products sales.
4. Online warehousing system

VI. REFERENCES

- [1] Kai Xiong, "Research on Influencing Factors of Impulse Buying in Online Shopping Environment.", in IEEE-2020.
- [2] Md. Ariful Islam Arif, Saiful Islam Sany, Faiza Islam Nahin and AKM Shahariar Azad Rabby, "Comparison Study: Product Demand Forecasting with Machine Learning for Shop.", in IEEE-2019.
- [3] Nitin Harale, Sebastien Thomassey, Xianyi Zeng, "Supplier Prediction in Fashion Industry Using Data Mining Technology.", in IEEE-2019.
- [4] Zeynep Hilal Kilimci, A. Okay Akyuz, Mitat Uysal, An Improved Demand Forecasting Model Using Deep Learning Approach and Proposed Decision Integration Strategy for Supply Chain, Journal of Hindawi, vol. 2019, no. 1, pp. 1-15, 2019.
- [5] Majed Kharfan, Vicky Wing Kei Chan, Forecasting Seasonal Footwear Demand Using Machine Learning, Publisher Massachusetts Institute of Technology, 2018.
- [6] Marsland, S., "Machine learning: an algorithmic perspective," 2015, CRC press.
- [7] James, G., Witten, D., Hastie, T and Tibshirani, R., "An introduction to statistical learning," 2013, (Vol. 6), Springer.
- [8] Loh, W.-Y., "Classification and regression trees. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery," 2011, 1 (1), 14-23.
- [9] D. Grewal, A. L. Roggeveen, and J. Nordfalt, The Future of Retailing, Journal of Retailing, vol. 93, no. 1, pp. 16, 2017.
- [10] M. A. A. Hasin, Shuvo Ghosh, and Mahmud A. Shareef, An ANN Approach to Demand Forecasting in Retail Trade in Bangladesh, International Journal of Trade, Economics and Finance, Vol. 2, No. 2, pp. 154-160, April 2011.