



e-ISSN: 2278-8875
p-ISSN: 2320-3765

International Journal of Advanced Research

in Electrical, Electronics and Instrumentation Engineering

Volume 10, Issue 3, March 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.122

9940 572 462

6381 907 438

ijareeie@gmail.com

www.ijareeie.com



Price Comparison System Based on Web Scrapping and Spark

Prof. Shalini Nigam¹, Sayli Jadhav², Kshitija Gholap², Manish Mowall²

Head of Department, Dept. of I.T., G.S.M.C.O.E, Savtribai Phule Pune University, Pune, Maharashtra, India¹

B. E Students, Dept. of I.T., G.S.M.C.O.E, Savtribai Phule Pune University, Pune, Maharashtra, India²

ABSTRACT: E-Commerce nowadays plays a vital role in our daily lives. It is redefining commercial activities around the world. Over the years, E-Commerce has evolved in profound ways. So, Online Shopping with price comparison offers with best deals and trends in different shopping sites is in business competition and is beneficial for consumers. This paper presents a price comparison system for comparing prices, offers, deals and shows trends in products of different shopping sites. In this proposed system scraper, crawler with bot is implemented for data indexing and retrieval with spark for efficient dynamic data processing and fast recommendation.

KEYWORDS : System scraper; crawler with bot; efficient dynamic data processing; data indexing and retrieval. Price Comparison System(PCS).

I. INTRODUCTION

With changing communication technologies the manner in which individuals access and acquire information from various information sources has also changed. Due to information convergence, Web users enjoy a wide access to abundant information from various sources through Web sites and services. It's very difficult to gather low price and good quality products by an individual by searching on different shopping sites. So, many of the sites remain unsearched and very low users get to know them, users also have to compromise with limited shopping sites deals and offers. It allows consumer to see different list of prices for the product chosen by user and it helps consumer to make an informed decision about which to choose in order to save money. It also act as a tool to help consumers increase their price consciousness so that they will not feel cheated by the advertisement from the retailers that claimed they are offering the cheapest price but the reality happened to be otherwise.

The price comparison system (PCS) provides online shoppers a wide range of information on various products. It reduces the amount of time or effort required for buying a product online. Conventional PCSs are generally suitable for direct searches, which focus on the required information on specific product searches.

For this study, an optimal PCS is devised to support accommodating and recommending user-defined price ranges. Price ratings and product clusters are devised that employ Business logic and data mining, which have emerged as useful tools for processing information collected from Web sites and providing personalized Web services.

II. COMPUTATIONAL TECHNOLOGY

The whole system is divided into three main modules:

2.1 Data Extraction

In this module about the products will be extracted from the sites that we are going to use in our further operations. Data extraction happens continuously for the duration of the event in order to periodically capture the live data of the event. There are many different ways of gathering the required information in web scraping but to parse HTML pages and to retrieve and transform page content HTML parsing is used. scrapping of the data is done for storage and computation. Many Languages used for web scraping and crawling with mining but the best for use is Python in project as it has many libraries which provide many functions to perform. Different libraries to be used is: Request - For processing request for Data

BeautifulSoup - Getting the required information in tidy way

Scrapy - For indexing many pages and scrape data without more human intervention [5][9].

2.2 Data Processing

Processing may include aggregation, prediction, clustering, classification, etc. Computational algorithms will be used to



perform this processing on the available data. This will be done on real time data ,as nowadays 70% of data produced is real time data. Spark is used as a Framework for Data processing and gets final info. requested by the user. Spark Streaming with spark SQL to be used form many components of Spark. As spark is an in memory mechanism so the process data has to be stored, for this cloud service is used.[7].

2.3 DataDelivery

Processed data is received by the delivery module which is an UI. There are prevalent types of user interface but here Graphical User Interface (Desktop version) is used.Where the various interests of individuals will be delivered through suggestions on UI. The processed response are posted on the UI automatically. For smooth and user friendly interconnection Angular is used as it is a platform and framework for building single-page client applications using HTML and TypeScript. Angular is written in TypeScript[8]. It implements core and optional functionality as a set of TypeScript libraries that you import into your apps.

Graphs:

By graphs we can ace two things as business owners or data analytic of companies can decide about their strategies on product and services by the graphs efficiently without extra surveys.Secondly customer can view the trends of different product on which site product sold is high to buy and quality graphically.as human grab thing faster without bore by colors and graphs. And these graphs will be real time grapes which will be shown on UI with one click. For this chart.js [8]and Graph-X is being used.[7]

III. SYSTEM ARCHITECTURE

As the system is divided into three main modules. The working of each module is independent but is dependent on each other's data generated.

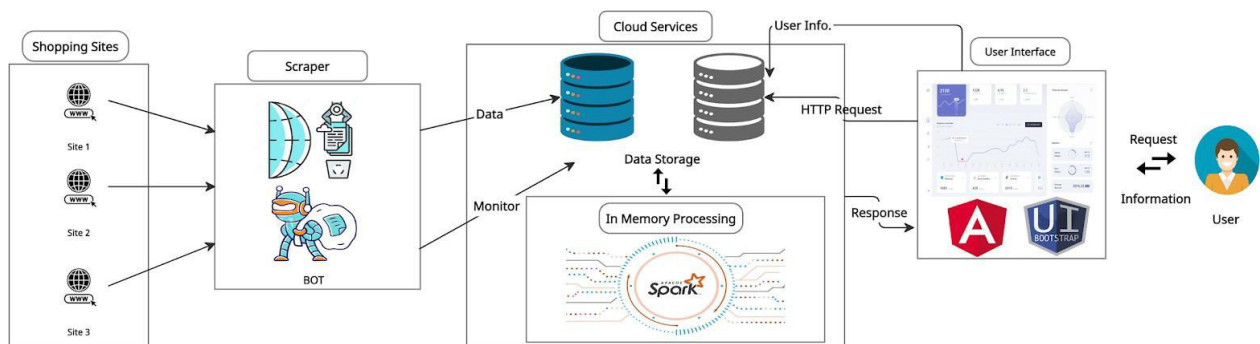
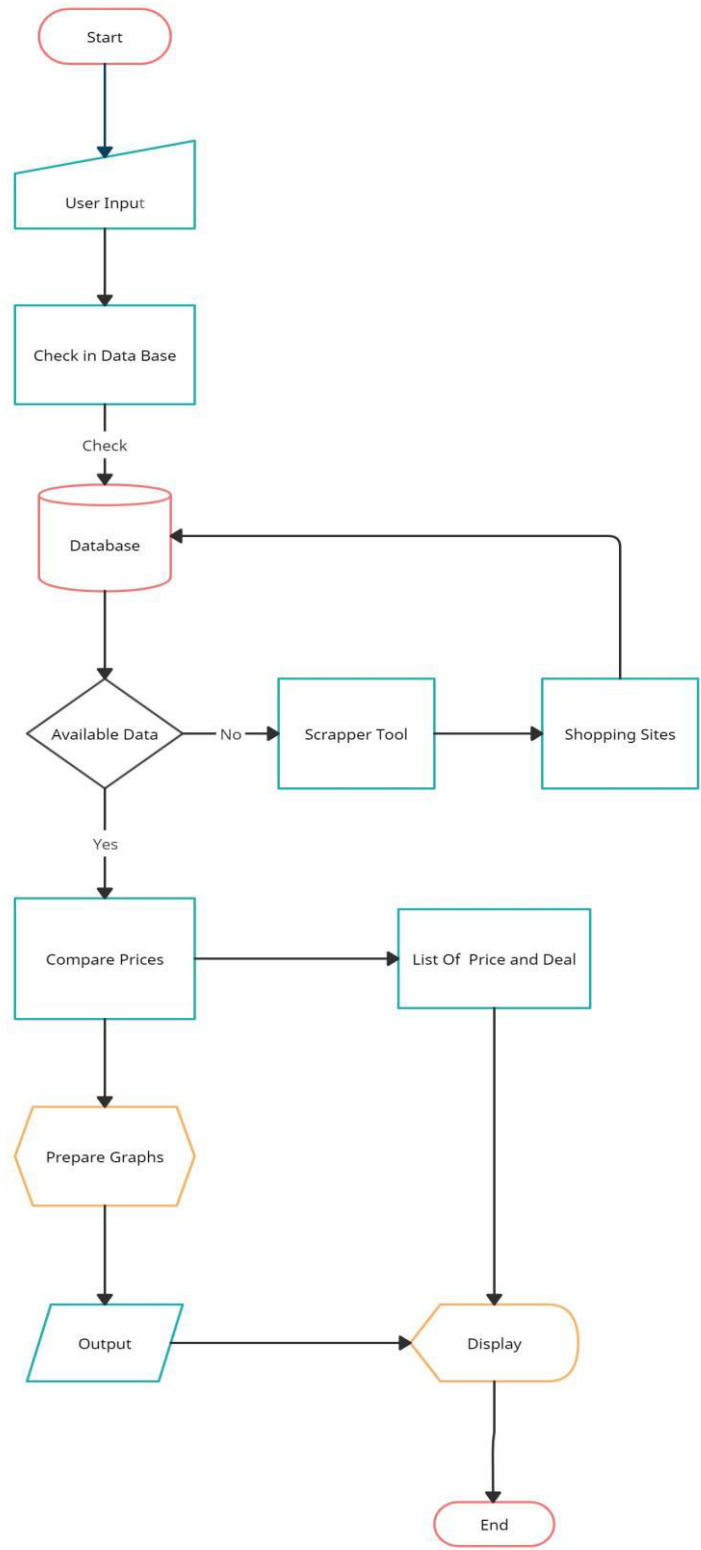


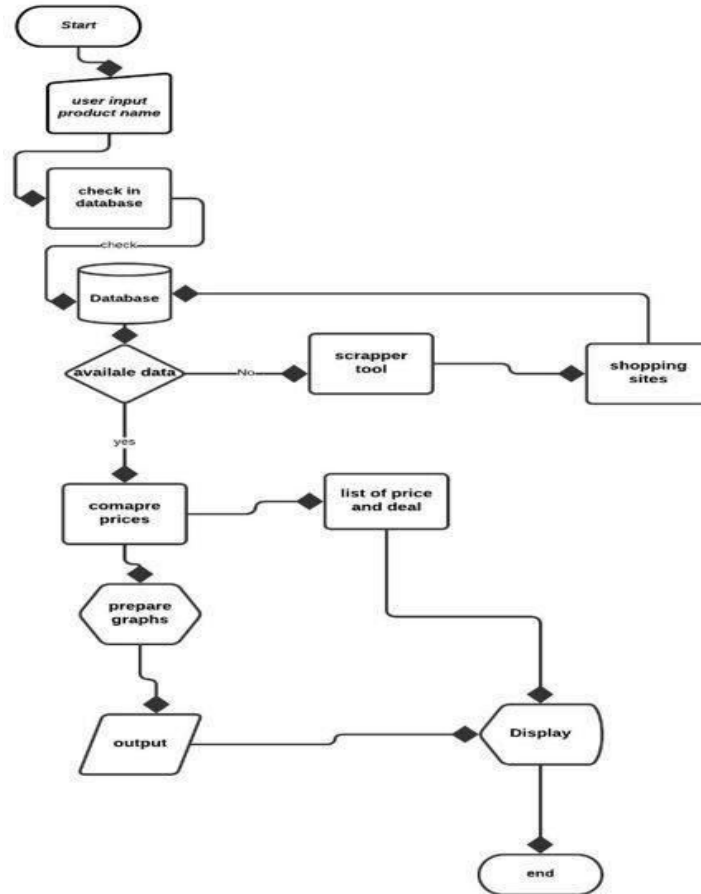
Fig. 1 - System Architecture

Figure 1 describes system architecture and its detailed working procedure. The front end system provides a graphical user interface (GUI) in the form of website where clients interact with the system whereas the backend consists of scrapping techniques in order to extract product information from different e-commerce websites. The extracted information of e-commerce products is stored in Store in a spark which is located in a cloud services . Client requests for desired product from main website and query is fired in local database. Product Information is displayed on main web page. Client can see prices of required product at one place present on different E-commerce firms. Another feature is Feature is provided that user can see the comparison through the life charts which is more easy to understand for the user. Users may also analyze the product for its details and specifications. a user can get the URLs for the different products which they are searching,



The flow of the data with working of different module is :





IV. WORKING

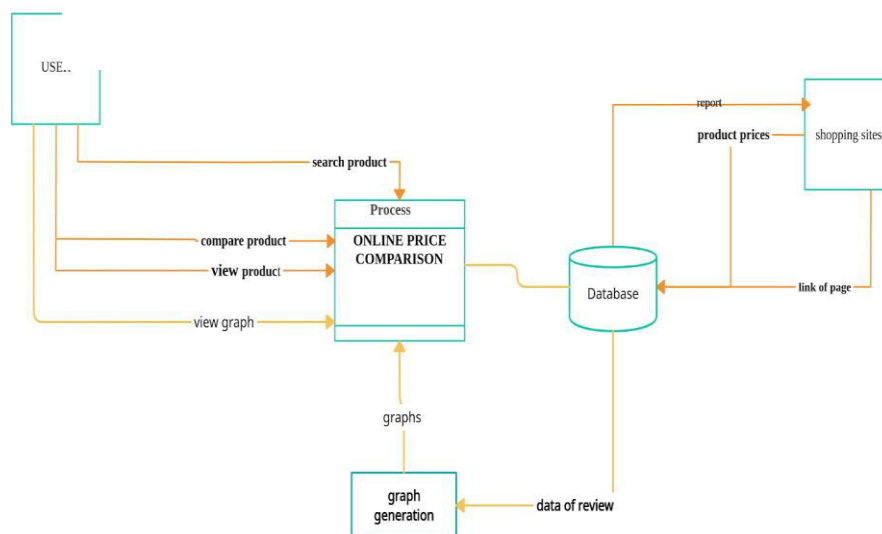


Fig. 3 -Data Flow Diagram

A system is as follows the backend system consists of web scrapping[4] as an important techniques. Is a technique that



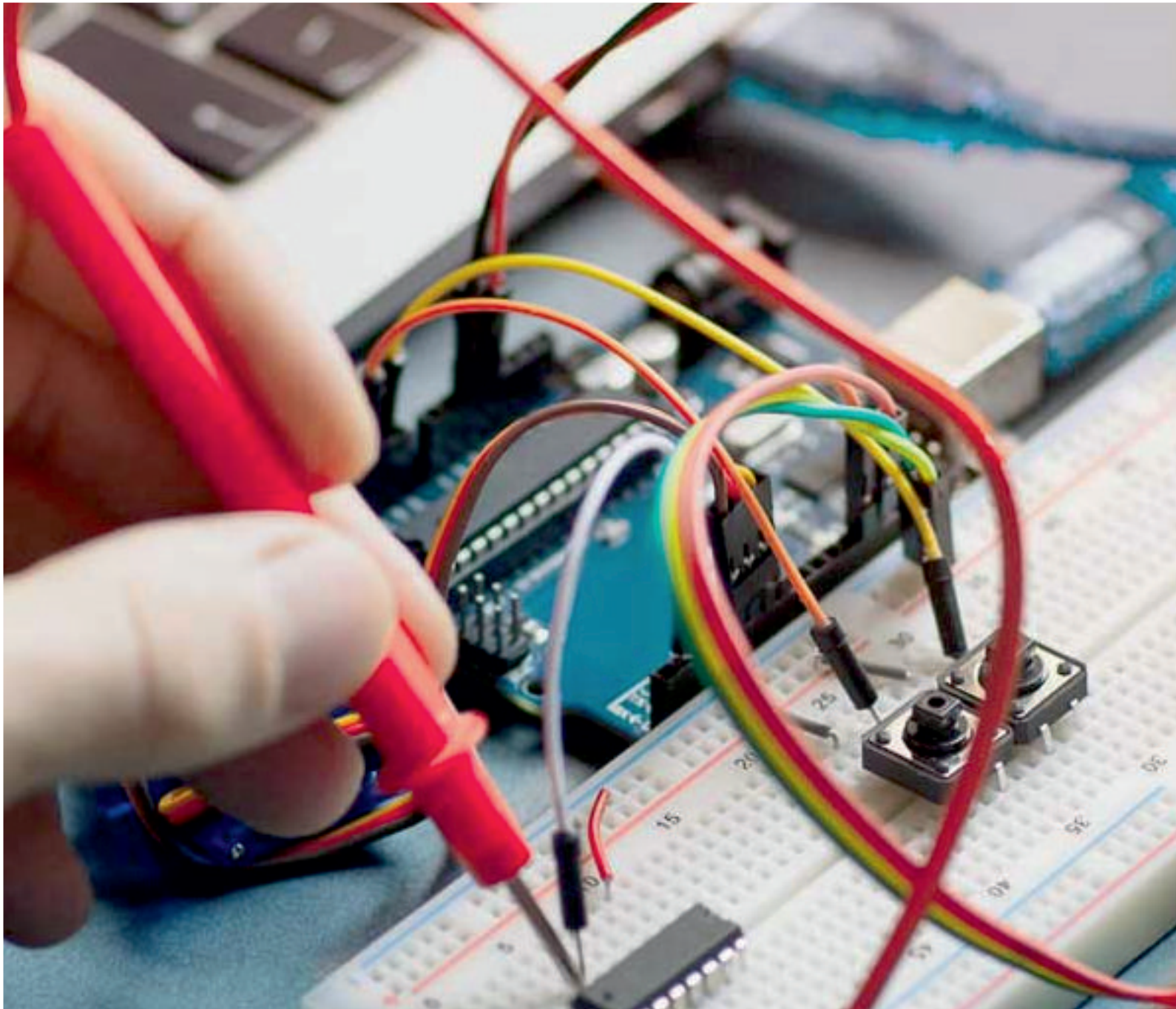
is used to extract the information in human readable format and display it destination terminal .Interesting tracking is done using the Python at is provide the reset of libraries to address this task after scrapping the product information from different E-commerce websites that are stored in the spark. The frontend consists of main website whenever the client searches for the required product in a search bar and the queries fired in database .If it is said that it is present in a databaseornotifitispresentindatabaseitisgoingtothetowardsthesparkandperformtheseveral tasksliketheprice comparison and the generation of the graph [7]and its shows to the user the desired output if it is not available in the database it again goes to the web crawler and scrap the required data from the website. It will present the output to the user and its on the user which site he prefers and then the user will be redirected to the original E-commerce website.

V. CONCLUSION

The proposed system will be useful tool that will provide suggestions to the user as per there recommendation. ThisSystemusesWebScrappingandSparktoFacilitatetheBuyerstocompareProductsfromdifferentE-Commerce Websites for the best deal. This way the buyer have more power in his/her hand and can make better decisions. Thus this System helps buyers to save Time and Money and also helps users to analyze through graphs.

REFERENCES

1. LadislavBeranekandRadimRemes(2019)'E-commercenetworkwithpricecomparatorsites',2019 9th International Conference on Advanced Computer Information Technologies(ACIT),10.1109/ACITT.2019.8779865), pp. [Online]. Available at: <https://ieeexplore.ieee.org/>(Accessed: 5-7 June 2019).
2. Vidhi Singrodia ; Anirban Mitra; Subrata Paul (2019) 'A Review on Web Scrapping and its Applications', 2019InternationalConferenceonComputerCommunicationandInformatics(ICCCI),(),pp. [Online]. Available at:<https://ieeexplore.ieee.org/>(Accessed: 23-25 Jan.2019).
3. K. Jayamalini and M. Ponnaivaikko (2017) 'Research on web data mining concepts, techniques and applications', 2017 International Conference on Algorithms, Methodology, Models and Applications in EmergingTechnologies(ICAMMAET),(),pp.[Online]. Availableat:<https://ieeexplore.ieee.org/>(Accessed: 16-18 Feb. 2017).
4. BoZhao(2017)'Web Scraping',*EncyclopediaofBigData*,(10.1007/978-3-319-32001-4_483-1),pp.1-3 [Online]. Available at:<https://www.researchgate.net/>(Accessed: 08 may2017).
6. JawahireNakash*,ShaikhAnas,SiddiqiMuzammilAhmad,AnsariMohd.Azam,ProfTabrezKhan(2015) 'Real Time Product Analysis using Data Mining', *International Journal of Advanced Research in Computer Engineering& Technology (IJARCET)*,4(3),pp.[Online]. Availableat:<http://ijarcet.org>(Accessed:).
7. Riya Shah * , Karishma Pathan , Anand Masurkar , Shweta Rewatkar , Prof. (Ms.) P.N.Vengurlekar (2016) 'Comparison of E-commerce Products using web mining', *International Journal of Scientific and Research Publications*, 6(5), pp. [Online]. Available at:www.ijsrp.org(Accessed: may2016).
8. Eman Shaikh; Iman Mohiuddin; Yasmeen Alufaisan; Irum Nahvi (2020) 'Apache Spark: A Big Data Processing Engine', 2019 2nd IEEE Middle East and North Africa COMMunications Conference (MENACOMM),(),pp.[Online]. Availableat:<https://ieeexplore.ieee.org/document/8988541>(Accessed:).
9. Mohamed Sultan (2018) 'Angular and the Trending Frameworks of Mobile and Web-based Platform Technologies:AComparativeAnalysis',*FutureTechnologiesConference(FTC)2017*,(),pp.[Online]. Available at:<https://saiconference.com/>
10. SANYA GOEL; MUDIT BANSAL; ATUL KUMAR SRIVASTAVA; NEHA ARORA (02 September 2019) 'WebCrawling-basedSearchEngineusingPython',20193rdInternationalconferenceonElectronics, Communication and Aerospace Technology (ICECA), (), pp. [Online]. Available at:<https://ieeexplore.ieee.org/document/8821866>



INNO  **SPACE**
SJIF Scientific Journal Impact Factor

Impact Factor:
7.122

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



International Journal of Advanced Research

in Electrical, Electronics and Instrumentation Engineering

 **9940 572 462**  **6381 907 438**  **ijareeie@gmail.com**



www.ijareeie.com

Scan to save the contact details