iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IV **Month of publication:** April 2023

DOI: https://doi.org/10.22214/ijraset.2023.50679

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

AI Based E-Mail Scraper and SendingTool

Ketan Nimase[1], Siddhesh Thakur[2], Sahil Gaonkar[3]

Abstract: ***Our research 'AI BASED E-MAIL SCRAPER AND SENDING TOOL' is basically a fast, affordable and easy-to-find marketing and communication solution. Using email can greatly help businesses as it provides an efficient and effective way to advertise variety of electronic information. Email extractor is a type of software used to extract email address from online bases which generate a huge list of email addresses. Even though these extractors can assist multiple genuine purposes such as marketing campaigns, unfortunately they are mostly used to direct spamming and phishing emails. Email filter is a tool to extract emails based on specified criteria. It split up all types of emails such as Gmail, Yahoo Mail in various text files according to their name automatically. Email validator is a tool to check the validation of email existence these means it checks that given mail is originally exist. It checks the username of email in all the mail facility for the existence. Bulk mail sender tool is work on distribution of a lot of mail at once. You drop an email list to director to send the letter to the user. You can send mail to a multiple user at once. There is no restriction for this to send the mails.***

Keywords: Scraper, Validator, Extractor, Email Filter, Marketing, inexpensive

I. INTRODUCTION

If you are working at a startup and want to reach out to more potential leads, you may need to collect as many email addresses as possible. Email extractor or scraper is a type of software used to extract email address from online bases which generate a huge list of email addresses [1]. Email filter is a tool to extract emails based on specified criteria. It split up all types of emails such as Gmail, Yahoo Mail in various text files according to their name automatically. Email validator is a tool to check the validation of email existence these means it checks that given mail is originally exist [2]. Bulk mail sender tool is work on distribution of a lot of mail at once You drop an email list to director to send the letter to the user. A bulk email sender is a facility that allows its customers to send bulk emails to multiple lists of recipients at the same time [1]. With this service, you can send messages to thousands of people on your mailing list or send personal e-mails to everyone on your list. our e-mail service can send e-mail to lists of any size. Most of these service providers price their products based on the volume and frequency of emails people want to send. With our system, users can send e-mails to thousands of peoples with different subscribing plans. [1]. As you may have read in one of our past blogs, data scrapers are your best pals when it comes to exerting valuable data from the Internet, building a meaningful relationship with your clients and taking your outreach game to leap and bound [1]. However, I realize that email data extraction can be difficult and massive but if you have good data scraping mechanism, you can get good data right away with a few mouse clicks. Finding the right data is the bridge between free Excel files and increased sales and/or productivity. In today's world internet is more used than medical antiseptic gels (this past year) – people produce a staggering 2.5 quintillion bytes of data on a daily level. Whether you're about to start your dream business or have owned it for decades, the information in your database can help you attract customers and keep them coming back from your competitors [2]. Scraping data or extracting useful information from the Internet and converting it into a useful format such as a spreadsheet is an essential part of building advanced B2B database trust. The website information tells you almost everything you need to know about these customers, from the average price they pay to the must-have features of the period. But not every SME has the time or budget to spend hours on manually extracting and validating data. This is where web or data scraping mechanism come into play, and the process can be quite complex. It is difficult to say what factors should be considered when choosing the appropriate data scraping tool [2]. Of course, different users have very different needs and there are different tools for each.

II. LITREATURE REVIEW

A. Background

Email scraping helps you to collect email addresses that are publicly available to scrap. What makes it great is that you can control where to get your email list and who can sign up. Also, you don't have to rely on second-hand system equipment. [6]. Usually, email scraping is done with the help of email parsers or scrapers (services designed to extract email addresses from web pages). these programs can usually extract email addresses from web pages and export the results to a suitable file such as Excel or csv files. Professional scanners often analyze information from social networks (Twitter, from the web [2]. If a company needs to find e-mail addresses for legal entities, the company collects the necessary information from the companies' corporate websites.

©IJRASET: All Rights are Reserved

B. Basic Terminologies

1) E-Mail Scraper and Filter

Emails scraping is also called email extraction, email harvesting, or email collection. Email scraping is the process of getting emails from prospects from websites to Excel or CSV formats for marketing purposes by using email extracting technique. Email extractor can help you to gather the emails of targeted customers and businesses for email marketing using an email scraper [10]. Email filter is a tool to extract emails based on specified criteria. It split up all types of emails such as Gmail, Yahoo Mail in various text files according to their name automatically

2) E-Mail validator

Email validator is a tool to check the validation of email existence these means it checks that given mail is originally exist. Bulk Email List Verification - an essential tool for bulk email companies to determine if a list of email addresses is valid and deliverable [8]. The process includes verifying each email in the upload and export list based on custom software by address. Below we explain in detail how the email verification process works.

3) Bulk Email Sending Tool

A bulk email sending service is an e-mail marketing business that allows its customer to send mass emails to multiple recipients at the same time. By using this service, you can send a message to thousands of people in your mailing list or send an e-mail to all the addresses in your list. Major e-mail services can send e-mail to lists of any size [7]. Most of these service providers charge for their products based on the volume and frequency of emails people want to send. With this system, customers can send e-mails [3] to thousands of users with different estimating plans.

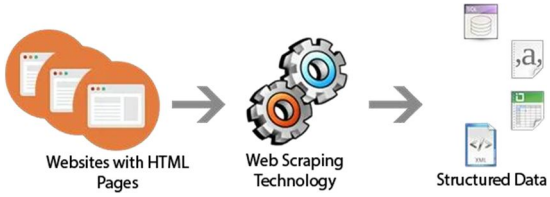


Fig 3.1 Email Scrapping [4]

4) Steps Involved in Email scrapping

a) Identify Your Sources: The first step in email scraping is to identify the sources from which you want to collect email addresses.

b) Choose Your Email Scraper Tool: Next, you will need to choose an email scraper tool that can help you extract email addresses from your identified sources.

c) Configure Your Scraper: Once you have your email scraper tool, you will need to configure it to collect email addresses from your identified sources.

d) Start Scraping: After configuring your email scraper tool, you can start scraping email addresses from your identified sources.

e) Clean and Verify Your Data: After scraping email addresses, it's important to clean and verify your data to remove duplicates, invalid email addresses, and other irrelevant data.

f) Store and Use Your Data: Finally, you can store your email list in a secure database or CRM system and use it for email marketing campaigns or other outreach efforts.

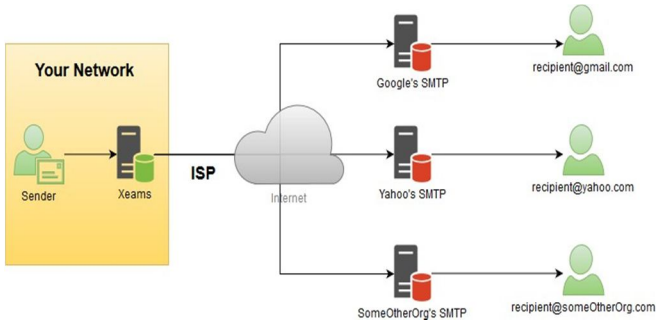


Fig 3.2 Bulk Emailing [5]

5) *Here are Some of the Steps Involved in Setting up a Bulk Emailsending System*

a) *Email List Management:* The first step in using a bulk email sending tool is to upload your email list. This can be done by importing a CSV file.

b) *Creating Email Content:* The next step is to create youremail content. You can typically use a drag-and-drop emailbuilder or HTML editor to design your email templates. You can add text, images, links, and other elements to create visually appealing emails that are relevant to your audience.

c) *Personalization:* Many bulk emails sending tools allow you to personalize your emails by inserting dynamic fields such as recipient name, company name, and other relevant data. This can help improve engagement and increase the likelihood of conversions.

d) *Email Scheduling:* Once you have your email content ready, you can schedule your email campaign to send at a specific time and date.

e) *Analytics:* After the emails are sent, you can track the performance of your campaign using analytics.

C. *Existing System*

Web scraping or extracting has been around for a long time and is a big part of the internet. Malicious bots extract content from websites for purposes beyond the website owner's control. Malicious bots make up 20% of all websites and are used for various crimes such as data mining, online fraud, theft, data theft, illegal search [3]. However, the law does not give web browsers the freedom to use the data theyfind for unlimited commercial purposes. A website thatrequires user authentication and the user must accept the website's terms of use to access the site. These terms of use are generally limited to automatic data collection. On public websites, users can use a web browser to collect informationfrom the website, as users are not required to accept the terms of use before accessing the information. [3]. We used E-mailExtractor, E-mail validator and Mass E-mail sender tool to build this system and to generate legitimate E-mails in Bulkquantity.

D. *Methodology*

This phase started at the beginning of our project. We formedgroups and modularized the project. Important points ofconsideration were

1) *Planning:* Describe the project goals and scope, identify websites to be scraped, and determine the data points to be collected.

2) *Design:* Develop a plan for how the scraper will collect and store data, and create a user-friendly interface for managing the scraper.

3) *Development:* Build the scraper using a programming language such as Python or Node.js, and test it to ensure that it is functioning correctly.

4) *Deployment:* Install the scraper on a server or cloud-basedplatform, and ensure that it is able to run continuously.

5) *Maintenance:* Monitor the scraper's performance, identify and fix any issues or errors, and make updates and improvements as necessary.

6) *Compliance:* Ensure that the scraper is compliant withrelevant laws and regulations, and obtain the necessary permissions and consent to collect and use the data.

7) *Reporting:* Generate reports on the data collected by the scraper, and use this information to inform business decisions and improve performance.

III. PROJECT ANALYSIS

A. *Dataset*

Before we begin, we need to know which pages we want to target for this project. We didn't create a database schema from scratch. This is something we think about later. This is a top-down approach where we first think about the type of information we're looking for and then build the informationaround the initial information.

B. *Algorithms*

1) *Data Cleaning:* This step focuses on cleaning and filtering of extracted email data to ignore all information that cannot be classified as email. If any of the required parameters are missing from the email, it will be ignored during cleaning. This step ensures that the next process does not have to deal with incorrect data. This improves data quality.

2) *Data Extraction:* This stage focuses on extracting data generated from emails. It includes key point extraction, sentiment analysis, regular expression, site extraction, and content extraction. Each of these methods focuses on some ofthe data extracted using text mining and NLP techniques.

a) *Keyword Extraction:* Context plays an important role in determining the meaning of irrelevant information. Keywordextraction technology is the automatic extraction of keywords. It is used to extract the most important details thatdescribe the email.

b) *Regular Expression:* Regular expressionis essential to extract information from the text. Date, phone number, email address, etc. It extracts certain data with certain patterns, such as from text using the search model. Each email is regularly passed through the library to extract email (to reply), email (from reply), date, URL, and phone number.

c) *Entity Extraction:* Entity extraction is one of the best ways to extract names from text.

IV. PROJECT DESIGN

The Project Design section in a report is a detailed description of the proposed project, including a managementplan and methods for quantifying the proposed project and shall include all appropriate, relevant and required documentation and materials necessary for the validation of the proposed project requirements.

A. Project Outline

These parts are considered as follows:

- Machine Learning Model
- Required Tools

B. Machine Learning Model

If you are using machine learning to develop an emailscraper for a legitimate purpose, there are several approaches you can take. One common method is to use natural language processing (NLP) algorithms to extract emails from websites. This involves analysing the text on a page and identifying patterns that match the structure of email addresses, such as strings of characters separated by "@" and "." symbols.

Another approach is to use machine learning algorithms to classify web pages based on their content and determine whether they are likely to contain email addresses. This can involve training a classifier on a dataset of web pages that areknown to contain email addresses, and then using this model to classify new pages. It is worth noting that the effectivenessof machine learning models for email scraping can differ widely liable on aspects such as the quality of the data, the complexity of the web pages being scraped, and the specific algorithms used. Additionally, it is important to ensure that your email scraping tool is compliant with anti-spam laws and respects the privacy of individuals whose email addresses are being collected.

C. Beautiful Soup

For extracting data, we need a python package which is beautiful soup. It integrates with your favourite parser to provide easy navigation, search and editing of parse trees. It's common for programmers to save money by working hours or days. To scrape the web with Beautiful Soup, we need to use the requests library to send requests to the weband receive the responses, and also extract the HTMLcontent from the response and pass it to Beautiful Soup forparsing.

D. Selenium

Selenium Python Connect provides a simple API for writingSelenium WebDriver function/acceptance tests. You caneasily access all aspects of Selenium WebDriver using Selenium Python API. The Selenium scheme is used to scrape websites like Facebook and Twitter that load contentdynamically, or when we need to click or scroll to enter the page or sign up to access the page to scrap. Once the websitehas loaded the dynamic build, we can use Selenium to accessthe HTML website and feed it to scrappy or Beautiful Soup to do the same job.

E. Pandas

Pandas is a database management and system analysislibrary. It is used to extract the data and you can save it in the format you want.

F. Activity Diagram

The Activity Diagram plays an essential part in our design. Itshows the inflow of conditioning and the connections between different tasks in a visual manner, which is essential for understanding our design methodology. In this section, we will offer a complete explanation of the exertion Diagram, including its end, symbols, and significance. By doing so, wehope to give a clear and terse understanding of our design methodology and the way involved. The illustration andexplanation will help the anthology grasp our approach and results more thoroughly.

As Fig 5.1 shows, the section will give a summary of the system architecture and also tell about the steps used while implementing the project idea. It also tells us about theworking of the project and what tools are used for developingthe project.

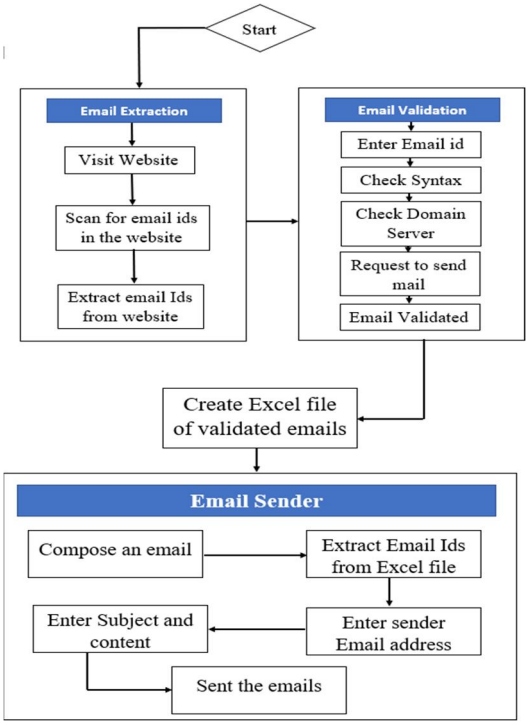


Fig 5.1 Activity Diagram

1) Email Extraction

Facial Email extraction using a web scraper involves the following steps: Identify the websites or web pages from which email addresses need to be extracted. Determine the web scraper tool to be used. There are several web scraper tools available in the market, including Beautiful Soup, Scrapy, and Selenium. Configure the web scraper tool to extract email addresses. This typically involves defining the webpages to be scraped, identifying the location of email addresses on the webpages, and defining the data extraction rules. Run the web scraper tool to extract email addresses. This can be done manually or automated through a script. Clean and validate the extracted email addresses. The extracted email addresses may contain invalid or duplicate email addresses, so it's important to validate and clean the data before using it.Store the extracted email addresses in a database or file. The extracted email addresses can be storedin a database or file for further processing, such as sending marketing emails or conducting research.

2) Collecting the Extracted Data

The steps involve the extracted emails from the web scrapertool. The extracted emails are typically stored in a file or database by the web scraper tool. Filter and clean theextracted emails. The extracted emails may contain invalidor duplicate email addresses, so it's important to filter and clean the data to remove any invalid or duplicate email addresses. Validate the extracted emails. The extractedemails should be validated to ensure that they are in the correct format and are deliverable. This can be done using email validation tools or services. Use the extracted emails for the intended purpose. The extracted emails can be used for a variety of purposes, such as email marketing, customer outreach, or research. It's important to ensure that the intended use of the extracted emails complies with allapplicable laws and ethical considerations.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

3) Cleaning and Validating Data

Cleaning the extracted email addresses is an important stepto ensure that the data is accurate, relevant, and usable. Hereare some steps to clean the extracted emails using a web scraper: 1) Remove duplicate email addresses: Scraper tools may extract duplicate email addresses, so it'simportant to remove them. This can be done by comparingthe extracted email addresses with a list of existing email addresses and removing any duplicates. 2)Remove invalid email addresses: Some extracted email addresses may not be valid, such as misspelled email addresses or email addresses with incorrect formatting. You can use an email validation tool or regex to identify and remove invalid email addresses.

4) Bulk E-Mailing

Bulk Emailing involves the following steps: 1) Creating theemail content: Once you have collected the email addresses, you need to create the email content, including the subject line, body text, and any attachments. 2) Setting up the emailcampaign: This involves using a bulk email sending tool toupload the email addresses and email content. You can alsosegment your email list based on demographics, location, interests, or other criteria to make the email more targeted. 3) Sending the emails: Once the email campaign is set up, you can send the emails to the scraped email addresses.

5) User Interface

The user interface component provides a graphical interface that allows users to interact with the web scraping system. The user interface can include a web-based interface or a desktop application.

V. RESULTS

An email scraper tool is designed to automatically extract email addresses from various sources such as websites, social media platforms, directories, and other online sources. Once the email scraper tool has finished running, it typically generates a list of email addresses that it has extracted. The next steps after running an email scraper tool will depend onthe purpose for which the tool was used. Once the validation process is complete, the list of email addresses is typically divided into two categories: valid and invalid. Valid email addresses are those that have been verified as active and deliverable, while invalid email addresses are those that have been identified as fake, spam, or no longer in use.

VI. CONCLUSION

In conclusion, adding patterns to the extracted email files willprovide important details with important information. Ignoring this information in the business world can result in the loss of important products. This information will help businesses make decisions by collecting customer preferences, opinions and behaviors regarding certain products or services. In summary, while email scrapers can be useful for collecting email addresses, it is important to usethem ethically and responsibly. Meanwhile, a bulk email sending system can be an effective tool for businesses to sendtargeted and personalized emails to a large audience while maintaining compliance with email regulations.

This study focuses on discussing the most important advantages and disadvantages of Email marketing and analyzing the success factors of this marketing by avoiding these disadvantages andbenefiting all these advantages.

VII. ACKNOWLEDGMENT

We express our gratitude to our project guide Dr. B.K Sarkar,who provided us with all the guidance and encouragement.

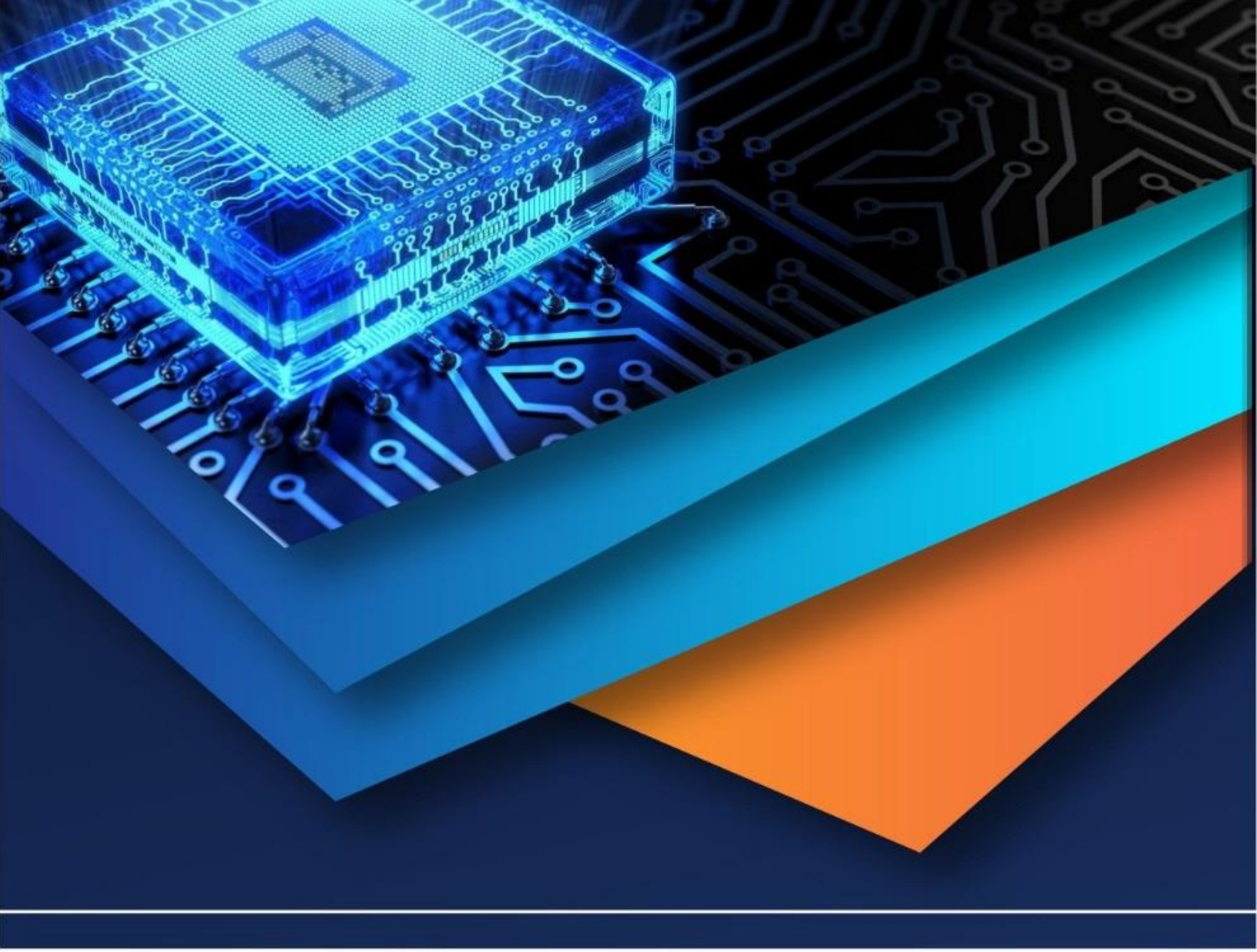
We are also thankful to her for providing us with the neededassistance, and detailed suggestions to do the project. We alsowould like to deeply express our sincere gratitude to the Project coordinators. We are eager and glad to express our gratitude to the Head of the Computer Department.

Prof. Rohini Bhosale for her approval of this project. We would like to deeply express our sincere gratitude to our respected principal Dr. J.W.Bakal and the management of Pillai HOC College of Engineering and Technology for providing such anideal atmosphere to build up this project.

REFERENCES

[1] https://sendpulse.com/support/glossary/bulk-email
[2] https://www.linkedin.com/pulse/3-case-studies-scraping-solutions- help-you-build-solid-dancho-dimkov/?trk=pulse-article_more- articles_related-content-card
[3] https://www.quora.com/Is-Web-Scraping-legal-if-the-link-to-the- website-scraped-is-provided

[4] https://www.price2spy.com/blog/case-study-web-scraping-data- extraction-for-ecommerce
[5] https://www.xeams.com/bulkmail.htm
[6] https://www.octoparse.com/blog/best-email-scraping-tools-for-sales- prospecting-in-2019
[7] https://www.accuwebhosting.com/blog/top-10-bulk-email-list- verification-validation-services-compared/
[8] https://www.quora.com/What-is-Email-Scrapping



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)