



# VENDOR SELECTION THROUGH REVIEW ANALYSIS

<sup>1</sup>Aliya Abdullah, <sup>2</sup>Anushka Bagde, <sup>3</sup>Aadarsh Kapoor, <sup>4</sup>Dr. Varsha Shah, <sup>5</sup>Shiburaj Pappu

<sup>1,2,3</sup>Student, <sup>4</sup>Principal, <sup>5</sup>Head of Department  
Department of Computer Engineering,  
Rizvi College of Engineering, Mumbai, India

**Abstract :** The paper presents the finished website for Vendor Selection using Review Analysis. It is a website aimed at helping small businesses make smart choices about selecting vendors. Because of the wide range of vendors and products that are now available, the decision-making process is complicated. Our website aims to help this problem by acting as a tool to find all relevant information about vendors in one place. The user, while being presented with a variety of products, can choose one to his liking. The website aggregates ratings, reviews, and sentiment analysis of tweets regarding the vendors to present an overview of crucial information regarding the vendor's reliability and quality of service.

**IndexTerms -** Web scraping, Sentiment analysis, Twitter, Amazon, Support Vector Machine.

## I. INTRODUCTION

In this report we present our finished project based on "Vendor Selection through Review Analysis". This system aims to help the user find the best product based on people's reviews. The user can search for any product of its choice on the website. The model then attempts to gather all information (reviews, ratings, etc) relating to its quality in one place. The model accomplishes this by mainly two methods, scraping reviews and using Sentiment Analysis on Twitter. It collects reviews and ratings of the entered product from e-commerce websites and also scans through various tweets from the microblogging website, Twitter. Then, using Natural Language Processing, it processes these tweets to extract sentiment information. This means it can categorize the tweets as positive or negative. The results of these two methods are displayed to the user, so that they may make a more well-informed decision about whether or not to buy the product.

## II. LITERATURE SURVEY

### 2.1. Laravel – A Trending PHP Framework

Laravel is an open-source PHP framework, which is robust and easy to understand.[7] It follows a model-view-controller design pattern.[1][2] Laravel reuses the existing components of different frameworks which helps in creating a web application.[1] The web application thus designed is more structured and pragmatic. Laravel offers a rich set of functionalities which incorporates the basic features of PHP frameworks like CodeIgniter,[4] Yii and other programming languages like Ruby on Rails.[1] Laravel has a very rich set of features which will boost the speed of web development.[5]

Features:

**Modularity:** [9] Laravel provides 20 built in libraries and modules which helps in enhancement of the application. Every module is integrated with Composer dependency manager which eases updates.

**Testability:** [9] Laravel includes features and helpers which helps in testing through various test cases. This feature helps in maintaining the code as per the requirements.

**Routing:** [9] Laravel provides a flexible approach to the user to define routes in the web application. Routing helps to scale the application in a better way and increases its performance.

**Configuration Management:** [9]A web application designed in Laravel will be running on different environments, which means that there will be a constant change in its configuration. Laravel provides a consistent approach to handle the configuration in an efficient way.

**E-mail:** Laravel includes a mail class which helps in sending mail with rich content and attachments from the web application.

**Authentication:** User authentication is a common feature in web applications. Laravel eases designing authentication as it includes features such as register, forgot password and send password reminders

## 2.2. PHP

PHP The PHP Hypertext Preprocessor (PHP) is a programming language that allows web developers to create dynamic content that interacts with databases.[3] PHP is basically used for developing web-based software applications.[3] This tutorial helps you to build your base with PHP. Five important characteristics make PHP's practical nature possible –

- Simplicity
- Efficiency
- Security
- Flexibility
- Familiarity

## 2.3 MVC Architecture

MVC is an acronym for 'Model View Controller'. It represents architecture developers adopt when building applications. With the MVC architecture, we look at the application structure with regards to how the data flow of our application works. [5] MVC is a software architecture...that separates domain/application/business...logic from the rest of the user interface. It does this by separating the application into three parts: the model, the view, and the controller. [5] The model manages fundamental behaviors and data of the application. It can respond to requests for information, respond to instructions to change the state of its information, and even notify observers in event-driven systems when information changes. This could be a database or any number of data structures or storage systems. In short, it is the data and data-management of the application. The view effectively provides the user interface element of the application. [5] It'll render data from the model into a form that is suitable for the user interface. The controller receives user input and makes calls to model objects and the view to perform appropriate actions.

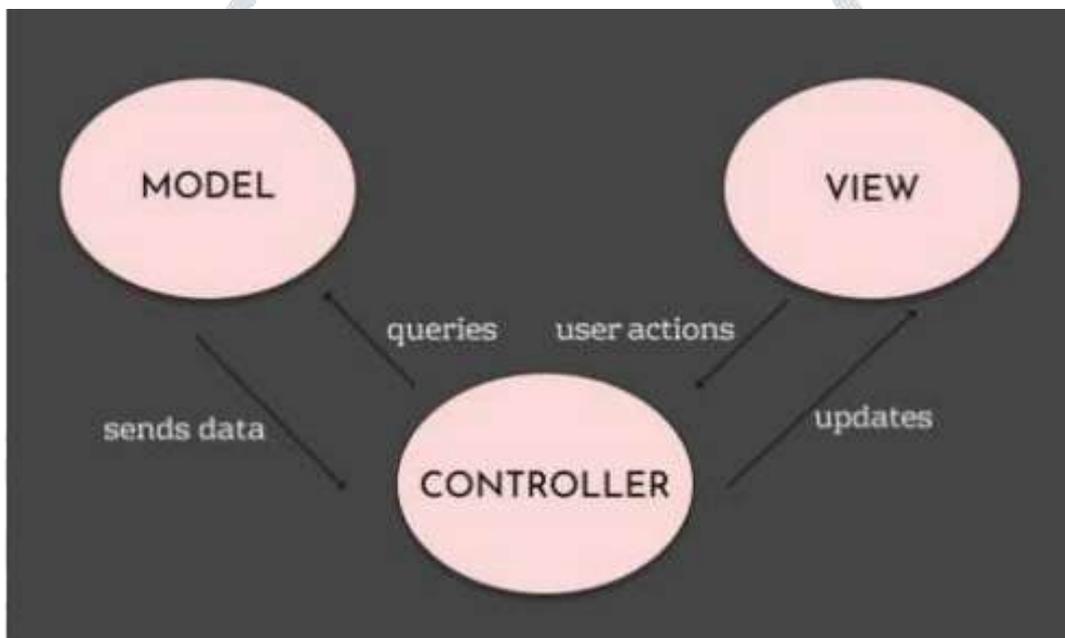


Fig 1 MVC Architecture

## 2.4. MySql

In today's world, for data to be useful, it needs to be structured, relevant, timely and accurate. Almost all databases provide functionality for report generation as it is a closely related feature. This review focuses on generating web-based report using MS SQL Server and MySQL. [8]

Databases are coherently displayed bunches of data, or information. A database the board framework (DBMS), then again, is a PC program that interfaces with a database. A DBMS permits you to control access to a database, compose information, run questions, and play out some other undertakings identified with database the board. Despite the fact that database the board frameworks are frequently alluded to as "databases," the two terms are not exchangeable. A database can be any assortment of information, not only one put away on a PC, while a DBMS is the product that permits you to collaborate with a database. [9]

MySQL has been the most well known open-source RDBMS since the site started following database fame in 2012. It is a component rich item that powers a large number of the world's biggest sites and applications, including Twitter, Facebook, Netflix, and Spotify. Beginning with MySQL is moderately clear, thanks in enormous part to its thorough documentation and huge network of designers, just as the bounty of MySQL-related assets on the web. MySQL was intended for speed and unwavering quality, to the detriment of full adherence to standard SQL. The MySQL engineers consistently progress in the direction of closer adherence to standard SQL, however it despite everything lingers behind other SQL executions. It does, notwithstanding, accompany different SQL modes and augmentations that carry it closer to consistence. Not at all like applications utilizing SQLite, applications utilizing a MySQL database get to it through a different daemon process. Since the server procedure remains between the database and different applications, it takes into account more noteworthy power over who approaches the database. MySQL hosts propelled an abundance of thirdgate

Advantages of MySQL:

- Popularity and Ease of use: As one of the world's most mainstream database frameworks, there's no lack of database executives who have experience working with MySQL. In like manner, there's a wealth of documentation in print and online on the most proficient method to introduce and deal with a MySQL database, just as various outsider apparatuses, for example, phpMyAdmin — that plan to streamline the way toward beginning with the database. •
- Security: MySQL comes introduced with a content that encourages you to improve the security of your database by setting the establishment's secret word security level, characterizing a secret word for the root client, evacuating mysterious records, and expelling test databases that are, of course, open to all clients. Likewise, not at all like SQLite, MySQL supports client the board and permits you to give get to benefits on a client byclient premise.
- Speed: By deciding not to actualize certain highlights of SQL, the MySQL designers had the option to organize speed. While later benchmark tests show that different RDBMSs like PostgreSQL can coordinate or possibly approach MySQL as far as speed, MySQL still holds a notoriety for being an exceedingly quick database arrangement.
- Replication: MySQL underpins various sorts of replication, which is the act of sharing data across at least two hosts to help improve dependability, accessibility, and adaptation to non-critical failure. This is useful for setting up a database reinforcement arrangement or on a level plane scaling one's database.

## 2.5 Selenium Web Driver

Selenium is composed of multiple software automation tools such as, Selenium IDE, Selenium RC (selenium 1.0), and Selenium WebDriver (selenium 2.0) [10]. Selenium IDE is an integrated development environment to build the test scripts. It is a Firefox plugin allow you to record edit and debug the selenium test cases [12]. It records all actions performed by the end user and generate the test scripts. Selenium remote control (RC) was main selenium project for long time. Selenium RC is slower than the selenium webdriver because it uses the java script program called selenium core [10]. Selenium RC requires to start the server before executing the test scripts. It doesn't support the Ajax applications. To avoid the limitations of selenium RC, selenium webdriver has been invented by merging selenium and webdriver.

Selenium webdriver is also known as selenium 2.0 [10]. Selenium webdriver directly communicate with the browser, so selenium webdriver is faster than selenium RC. Selenium webdriver supports multiple web browsers and also support for Ajax applications. The main goal of the selenium webdriver is to improve support for modern web application testing problems. Selenium webdriver supports multiple languages to write the test scripts. Selenium webdriver's API is simpler than the selenium RC's [11]. However, despite all advantages of selenium webdriver, it has some limitations when testing the web applications. Selenium webdriver does not have built in functionality to generate the screenshots for failure test cases. Selenium webdriver does not have built in capability to generate the test results. It depends on third party tools to generate the test reports. This limitation can be avoided by using TestNG framework

## III. RESEARCH METHODOLOGY

### 3.1. User-Facing Website

The application is built on Laravel framework. It is a convenient open-source PHP web framework. It uses the model–view–controller architectural pattern (MVC model) which makes developing websites and working with a team easy [6]. The database used is MySQL.

The application starts by prompting users to either register or login. All authentication and authorization functionalities are handled by Jetstream, which provides the implementation for the application's login, registration, email verification, two-factor authentication and session management.

Once the user is logged in, they are allowed to search for a product. This makes a POST Request to the Product Scraping API. The user is presented with a list of products from Amazon and other details like ratings, price, and a photo.

The user then clicks on the product they want to find about. This makes two requests. This first one is a POST Request to the Review Scraping API. This API returns various reviews left by users on Amazon and removes anonymous reviews. The second request is a POST request made to the sentiment analysis API. This API first calls another API, the official Twitter API, and uses the name of the product to get approximately 1,000 tweets. It then uses a pre-trained sentiment predictor and extracts the sentiments of all the tweets received. Ultimately, it assigns a final sentiment score to the user.

The user, after reviewing the ratings, reviews, and sentiment score, can make a decision on whether to buy the product or not.

The user also has the option to mark products as favourite. Such products are saved in the database so that the user may refer back as needed.

### 3.2. Product Scraper

The Product Scraper API's aim is to scrape the product name, price, ratings, number of reviews given to the product and image of the product from the Amazon search page of the item entered in the search bar by the user.

This API uses two python libraries: BeautifulSoup and Selenium automation testing libraries. Selenium library is used to navigate through amazon using Chrome web browser and BeautifulSoup library is used to scrape required information from the navigated web page. WebDriver uses browser automation APIs provided by browser vendors to control browser and run tests. This is as if a real user is operating the browser. [13]

A Flask-RESTful API accepts POST request sent by the website which sends the item name the user wants to search. This item keyword is sent to selenium Chrome webdriver which silently opens Google Chrome browser and is set to navigate to the Amazon homepage. There, it looks for the search bar component on the site by its HTML tag (#twotabssearchtextbox) and executes the click() function to begin the search.

Once the website yields the search page the link is scraped and is set as a navigation link for the BeautifulSoup scraper function. The scraper function finds all the pre-specified tags under which information of all products on the search result page like product

name, price, ratings, etc. are displayed. The function then extracts the information under the tags and stores it in a list where each list item is a dictionary containing scraped information about each product.[14]

The dictionary items appended in the list are sorted by the maximum ratings and number of reviews. This list is then encoded in JSON format and this JSON object is returned as the output of the API.

### 3.3. Sentiment Analysis API

The Sentiment Analysis model uses Linear Support Vector Machines (SVM) to classify tweets into either positive or negative sentiments. The model first removes stop words from the training data, then vectorises the data, and then the data is passed on for training. The training data used is Sentiment140, made by Alec Go, Richa Bhayani, and Lei Huang. This model achieved an accuracy of 0.78.

Once the model is trained, it is serialized or marshalled using Python’s pickle library. This model is now saved and does not need to be trained again. It can predict the sentiment of any text supplied to it.

A Flask-RESTful API is then built to perform the processing on tweets. This API accepts keywords through POST requests. These keywords are then used as queries for the Twitter API. Twitter API then returns a set of tweets and other metadata such as username, time of tweet, etc. [15]

To access tweets, first a twitter developer account should be created. Twitter allows you to make one app per free account. Twitter also provides you with a API key, API Key Secret, and a Bearer Token. These are the credentials that are required to make requests to the Twitter API. Once Twitter confirms your authentication, it replies with the requested information in JSON format.

The Sentiment Analysis API then loads the pickle into an object. It then iterates through all the retrieved tweets assigning a sentiment score to them. Finally, it calculates a final sentiment score based on the number of positive tweets out of the total number of tweets. This score is returned as the output of the API.

### 3.4. Review Scraper

The Review Scraper API’s aim is to scrape all reviews and customer names from the Amazon product page of the product selected by the user. It uses the python’s beautifulsoup library to extract names and reviews.

This API accepts POST request from the website containing the URL of the product selected by the user. This link is set as navigation for scraping. The scraping function finds and extracts all the information stored under the tags containing customer name and reviews left by the customer.

The dictionary items containing name and the review left by the customer is appended to a list for each customer. This list is encoded in JSON and this object is returned as the output of the API.

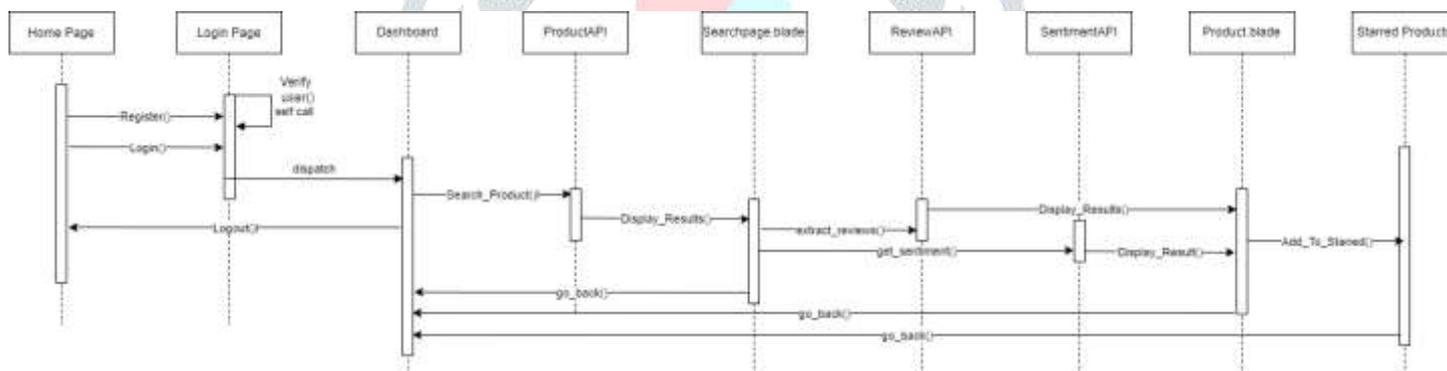


Fig 2. Sequence Diagram

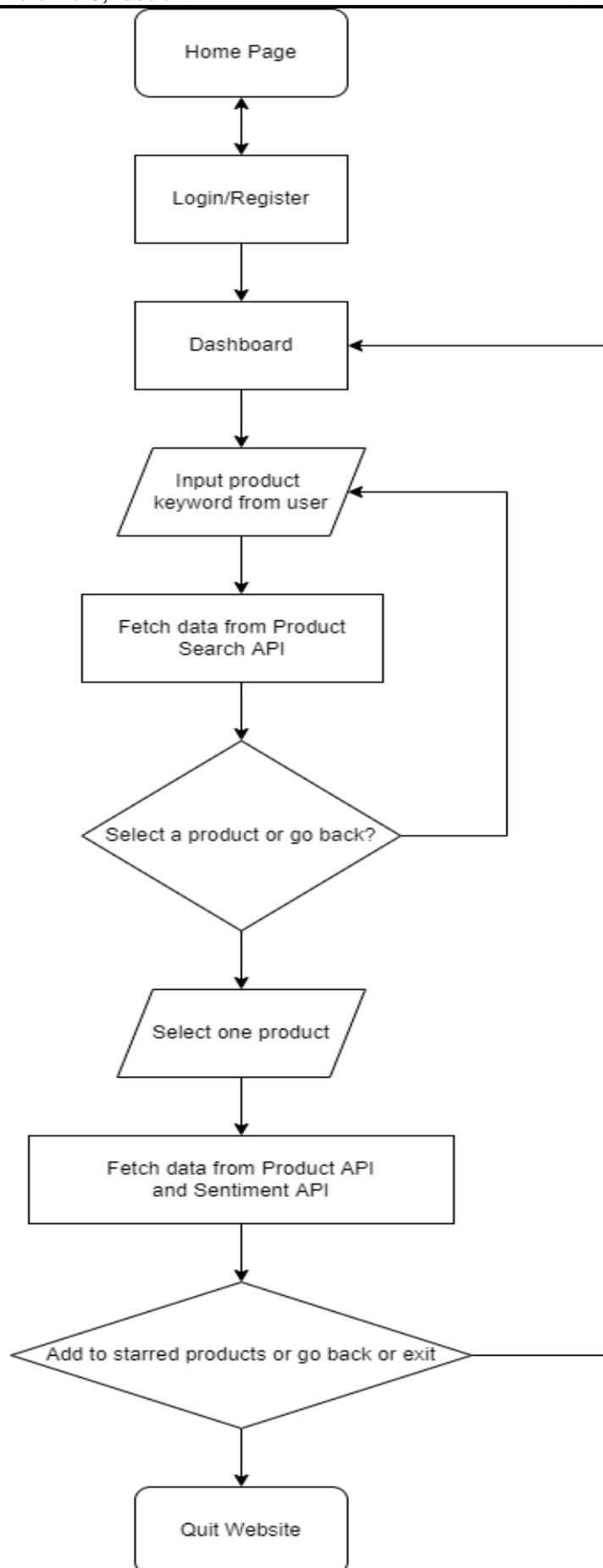


Fig 3. Flowchart

#### IV. CONCLUSION

Thus, we have successfully implemented a website for efficient selecting of a vendor. We used various methods and technologies to provide the user with vendor information, like ratings, reviews and sentiments of tweets about the vendor. The website can thus help businesses make smart decisions with the ease of using a single portal for all information.

## V. REFERENCES

- [1] Lakshay Khanna "Laravel – A Trending PHP Framework" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456- 6470, Volume-4 | Issue-4, June 2020, pp.1374-1377, URL: [www.ijtsrd.com/papers/ijtsrd31260.pdf](http://www.ijtsrd.com/papers/ijtsrd31260.pdf)
- [2] [https://www.tutorialspoint.com/laravel/laravel\\_overview.htm](https://www.tutorialspoint.com/laravel/laravel_overview.htm) International Journal of Trend in Scientific Research and Development (IJTSRD) @ www.ijtsrd.com eISSN: 2456-6470 @ IJTSRD | Unique Paper ID – IJTSRD31260 | Volume – 4 | Issue – 4 | May-June 2020 Page 1377
- [3] <https://www.tutorialspoint.com/php/index.htm>
- [4] <https://www.educba.com/laravel-vs-codeigniter/>
- [5] <https://blog.pusher.com/laravel-mvc-use/>
- [6] [www.laravel.com](http://www.laravel.com)
- [7] Laravel: Up & Running: A Framework for Building Modern PHP Apps
- [8] MySQL Official Documentation
- [9] R. P Padhy, M. R. Patra, S. C. Satapathy, “RDBMS to NoSQL: Reviewing Some Next-Generation Non-Relational Database’s”, International Journal of Advance Engineering Sciences and Technologies, Vol. 11, Issue No. 1, 015-030, 2011.
- [10] Selenium Documentation.[Online].(<http://www.seleniumhq.org>). (Accessed 15 DEC. 2014)
- [11] Rigzin Angmo, Monika Sharma. Selenium Tool:A web based Automation testing Framework. International Journal of Emerging Technologies in Computational and Applied Science,2014.
- [12] Sherry Singla, Harpreet Kaur. Selenium Keyword Driven Automation testing Framework, International Journal of Advance Research in Computer Science and software Engineering, VOL. 4,Issue 6,2014.
- [13] <https://www.selenium.dev/documentation/webdriver/>
- [14] <https://medium.com/analytics-vidhya/scraping-amazon-results-with-selenium-and-python-547fc6be8bfa>
- [15] <https://towardsdatascience.com/deploying-a-machine-learning-model-as-a-rest-api-4a03b865c166>

