

Project Team ID:
CSE20PT40

GIUDE:
SAVITA S
Assistant Professor Dept. Of CSE,
CSE, JSSATEB

SENTIMENTAL ANALYSIS WITH WEB SCRAPING

TEAM MEMBERS

Dhruva V	1JS18CS403
Pruthviraj G	1JS18CS416
Purushothama N	1JS18CS420
Gagan M S	1JS17CS035

ABSTRACT

- E-commerce is getting used more and more these days to purchase products in an online store. A product review is usually used see if the product is worth buying or not.
- This product used by selecting URL a product in amazon.in website opened in a browser, copy the URL first and open this product website and paste it in the input box and press enter, then this product processes the request and shows the score that the selected product has got and a wordcloud image
- The result of this project has shown success by show the result.

INTRODUCTION

- Sentimental analysis
- Machine learning
- Web-scraping
- Flask

AIMS and OBJECTIVES

AIMS:

- Our first aim is by using the web scrapping we generate the unstructured data from the amazon product pages.
- To the scrapped data, we create a machine learning model and it is used for analyzing the sentiments of the cleaned data.
- By doing so, with the help of the model we can classify the sentiment intensity (positive, negative or neutral) of the scrapped data.
- At last the output will be integrated with the help of flask and bootstrap. (extra: bootstrap-->it is the most popular html, CSS and JS library in the world)

OBJECTIVES:

- The main objective behind this project is to provide a platform which will enable users to check the credibility of a retailer/product by scanning reviews.
- Instead of going through potentially hundreds of reviews, our platform offers a one-click result.

Problem Statement & Motivation

- The Primary focus of this project is to find the sentimental scores of a product review by the customers.
- The main motivation behind the project is to understand the reviews submitted online by varying customers for a product. With knowing the product has positive or negative reviews we can have a clear understanding of the reviews submitted.

SYSTEM

EXISTING SYSTEM:

The existing system has many limitations:

- Sentiment model is a separate unit that hasn't been integrated with web scrapping
- It is more complex to implement and its expensive. There is lack of credible user interface.

PROPOSED SYSTEM:

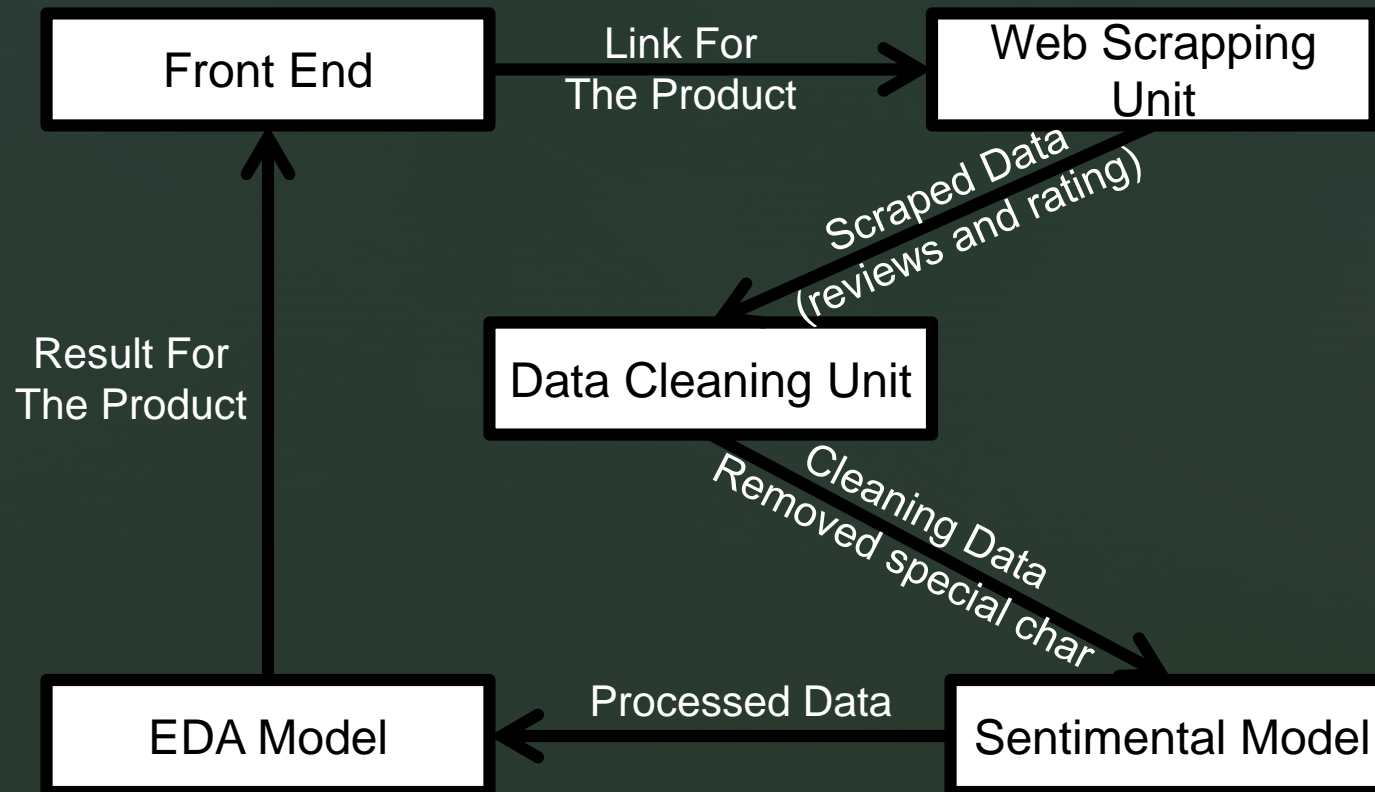
- Integration of both modules (web scrapper and sentiment analyzer) into a single unit.
- Developing a user interface using flask and bootstrap.
- Providing better result on the scrapped Data by implementing VADER intensity analyzer.
- VADER allows us to rate the reviews based on the emotions in the text.
- We generate a word cloud, which is a data visualization technique used for representing text data in which the size of each word indicates its frequency or importance.

SYSTEM REQUIREMENTS

Devices	: Local / Personal Computer.
OS	: Windows(Local Computer)
OS Distro	: Windows 10 (Local Computer)
Storage	: SD Card (8GB minimum)
Battery	: Power Bank / Li-ion(Optional if required)
Application	: python
Browser	: Google chrome or any other.

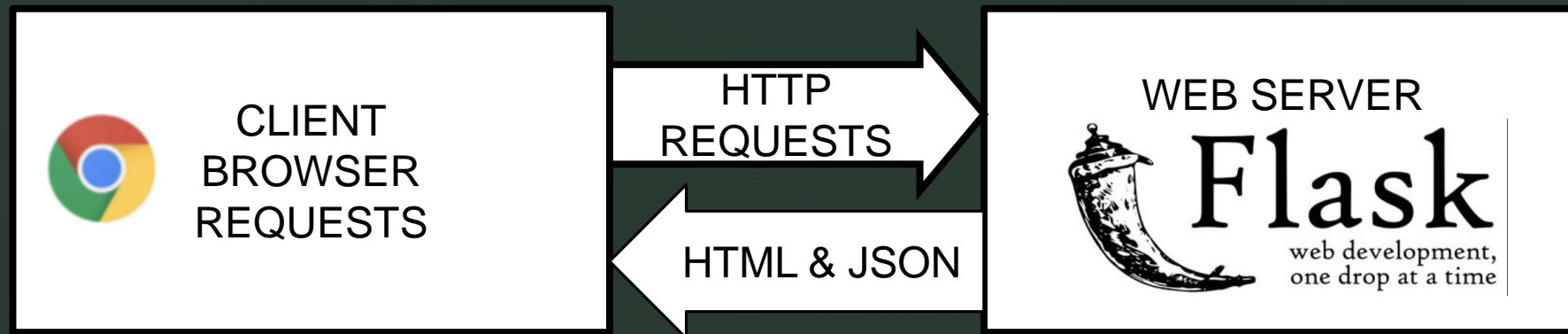
DESIGN AND IT'S IMPLEMENTATION

SYSTEM DESIGN



DESIGN AND IT'S IMPLEMENTATION

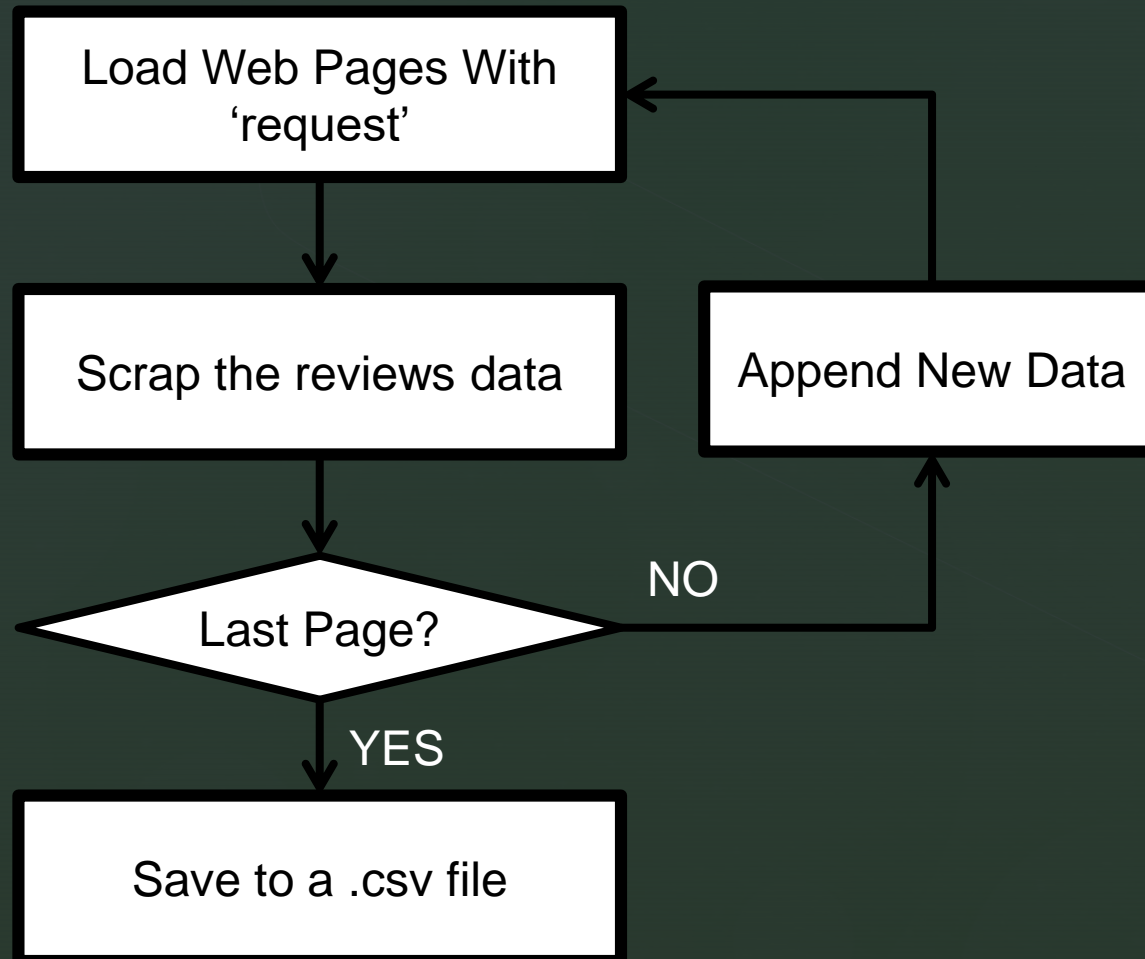
FRONT END



- The user can interact with our product through the web page
- This is implemented using Flask & Bootstrap 5
- Link for the product is sent to server

DESIGN AND IT'S IMPLEMENTATION

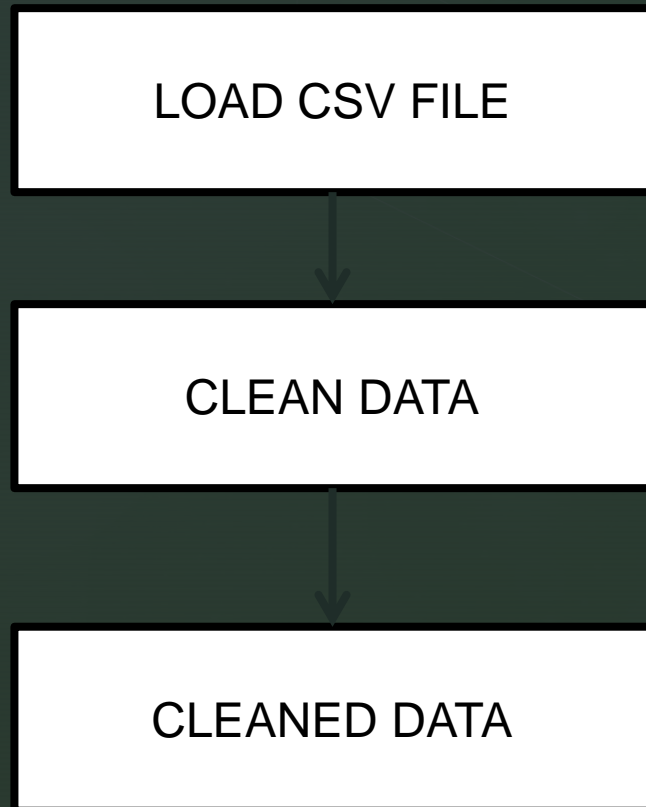
WEB SCRAPING UNIT



- The HTTP request returns a Response Object with all the response data (content, encoding, status, and so on)
- Using a library called BeautifulSoup in Python to do web scraping.
- Do for all review page
- At end saved to a .csv file

DESIGN AND IT'S IMPLEMENTATION

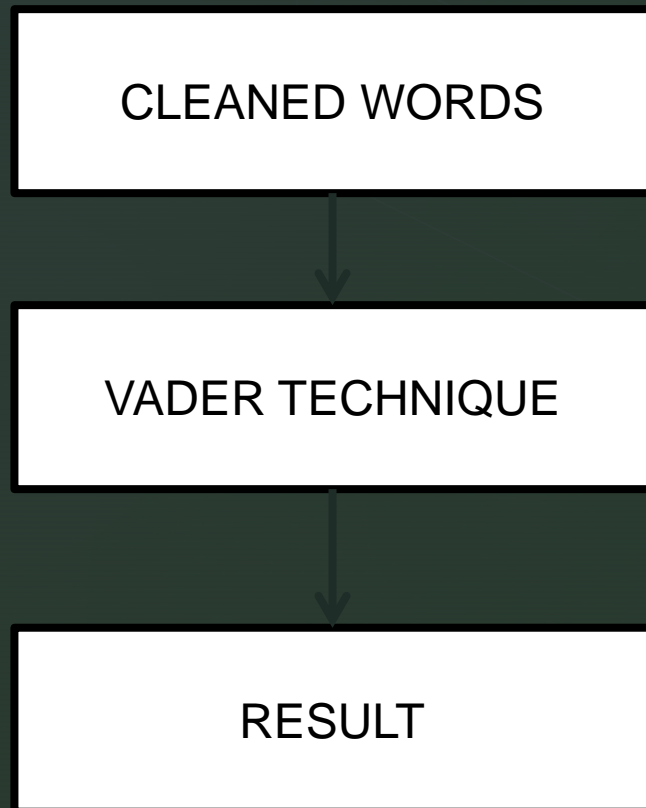
Data Cleaning Unit



- In the cleaning unit we remove the unwanted data from the[i.e. scraped text data] then we convert the text[review column] to lowercase and remove integers or numerical in text.
- Punctuations, null values and extra[spaces as they don't make any sense] and all these things are done by importing regular expressions[import re] Stop words are 'this is in', which does not add much value.
- So, we remove them to decrease the size of data set this process is called normalizing text[with spacy]spacy is most versatile and widely used library in NLP Lemmatization is nothing but normalization of words which means reducing a word into its root form.

DESIGN AND IT'S IMPLEMENTATION

SENTIMENTAL MODEL

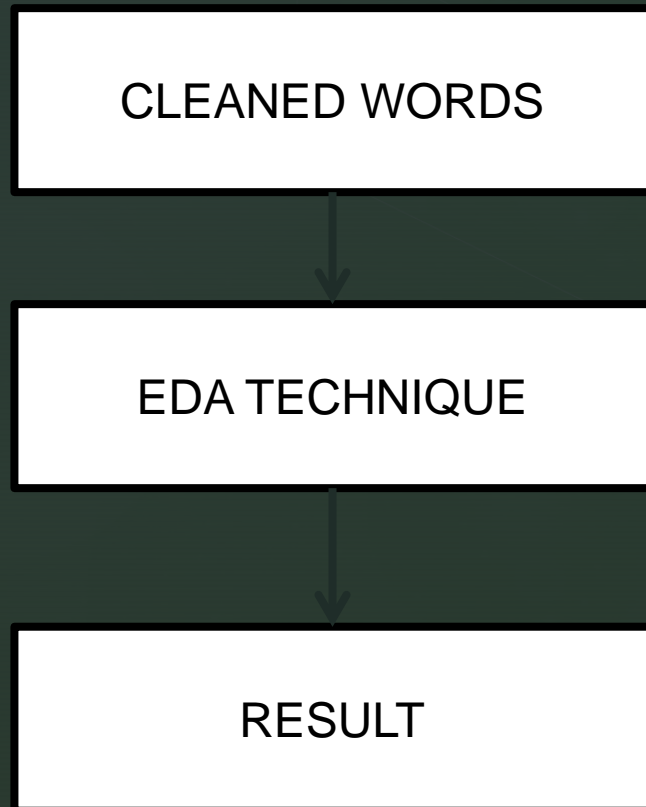


Sentiment analysis is done VADER technique:

- VADER[Valence Aware Dictionary for sential Reasoning] is a Unsupervised model used that is adaptive to interpret emotional [positive/negative] and emotional [strength] text feelings..
- VADER is focused on the lexicons of words related to sentiment. Each of the words in the lexicon is rated as to whether it is positive or negative and assigns scores to them.
- It uses to polarity scores method to get the sentiment metrics for a piece of text.

DESIGN AND IT'S IMPLEMENTATION

EDA MODEL



- EDA stands for Exploratory Data Analysis.
- It is a process of exploring data, generating insights, testing hypotheses, and revealing underlying hidden patterns in the data.
- We will create a Document Terms Matrix that we'll later use as our analysis to get the insights in data and with the help of wordcloud.

Modern tool usage

Here we use different python modules;

- FLASK MODULE : **Flask** is a lightweight WSGI web application framework.
- AMAZON_PRODUCT_REVIEW_SCRAPER MODULE : Uses the Amazon Simple Product API to provide API accessible data.
- PANDAS MODULE : **pandas** is a **Python** package that provides fast, flexible, and expressive data structures designed to make working with structured (tabular, multidimensional, potentially heterogeneous) and time series data both easy and intuitive.
- REQUESTS MODULE : The **requests module** allows you to send HTTP **requests** using **Python**.
- BeautifulSoup MODULE : **Beautiful Soup** is a library that makes it easy to scrape information from web pages.

and many more modules

SNAPSHOTS

The screenshot shows a web browser window with multiple tabs open. The active tab is the Amazon India product page for an AmazonBasics High-Speed HDMI Cable. The browser's address bar shows the URL <https://www.amazon.in/dp/B014I8SP4W>. The Amazon.in header includes the logo, the delivery location (Dhruva, Bangalore 560059), and a search bar with 'Electronics' entered. Below the header is a navigation bar with various categories like Fresh, Gift Cards, Amazon Pay, Kindle eBooks, Pet Supplies, Gift Ideas, dhruva... 's Amazon.in, Toys & Games, Baby, AmazonBasics, Home Improvement, and Car & Motorbike. The main navigation bar lists categories: Electronics, Mobiles & Accessories, Laptops & Accessories, TV & Home Entertainment, Audio, Cameras, and Computer Peripherals. The product page features a large image of the HDMI cable, a vertical sidebar of product thumbnails, and a detailed product description. The product is titled 'AmazonBasics High-Speed HDMI Cable, 3 Feet (2-Pa Supports Ethernet, 3D, 4K video)'. It has a brand of AmazonBasics, a 4.5-star rating from 289,193 ratings, and 873 answered questions. The price is ₹ 469.00, with a crossed-out M.R.P. of ₹ 1,100.00, indicating a 57% discount. The product is eligible for free delivery tomorrow and can be saved extra with 4 offers.

Buy AmazonBasics High-Speed | x | Top Courses to Opt for After Bac | x | Miss Crow with Mr. Lizard - x | +

https://www.amazon.in/dp/B014I8SP4W

test | TL-WR840N | Drama | Web Apps | Typing Lessons | Le... | JSLinux | Practice | Geeksfor... | Coursera | Courses - Dev Colle... | Library | Intro to TensorFlow... | Dashboard | Hacker...

amazon.in prime | Deliver to Dhruva Bangalore 560059 | Electronics

All | Fresh | Gift Cards | Amazon Pay | Kindle eBooks | Pet Supplies | Gift Ideas | dhruva... 's Amazon.in | Toys & Games | Baby | AmazonBasics | Home Improvement | Car & Motorbike

Electronics | Mobiles & Accessories | Laptops & Accessories | TV & Home Entertainment | Audio | Cameras | Computer Peripherals

storite | Storite 2 Pack 50cm High Speed HDMI Male to Female... | 413 reviews | ₹429.00 prime

Electronics > Accessories > Home Audio & Video Accessories > Cables > HDMI Cables > AmazonBasics High-Speed HDMI Cable Supports Ethernet, 3D, 4K and Audio Return

AmazonBasics High-Speed HDMI Cable, 3 Feet (2-Pa Supports Ethernet, 3D, 4K video)

Brand: AmazonBasics

4.5 stars | 289,193 ratings | 873 answered questions

M.R.P.: ₹1,100.00

Price: ₹ 469.00

You Save: ₹ 631.00 (57%)

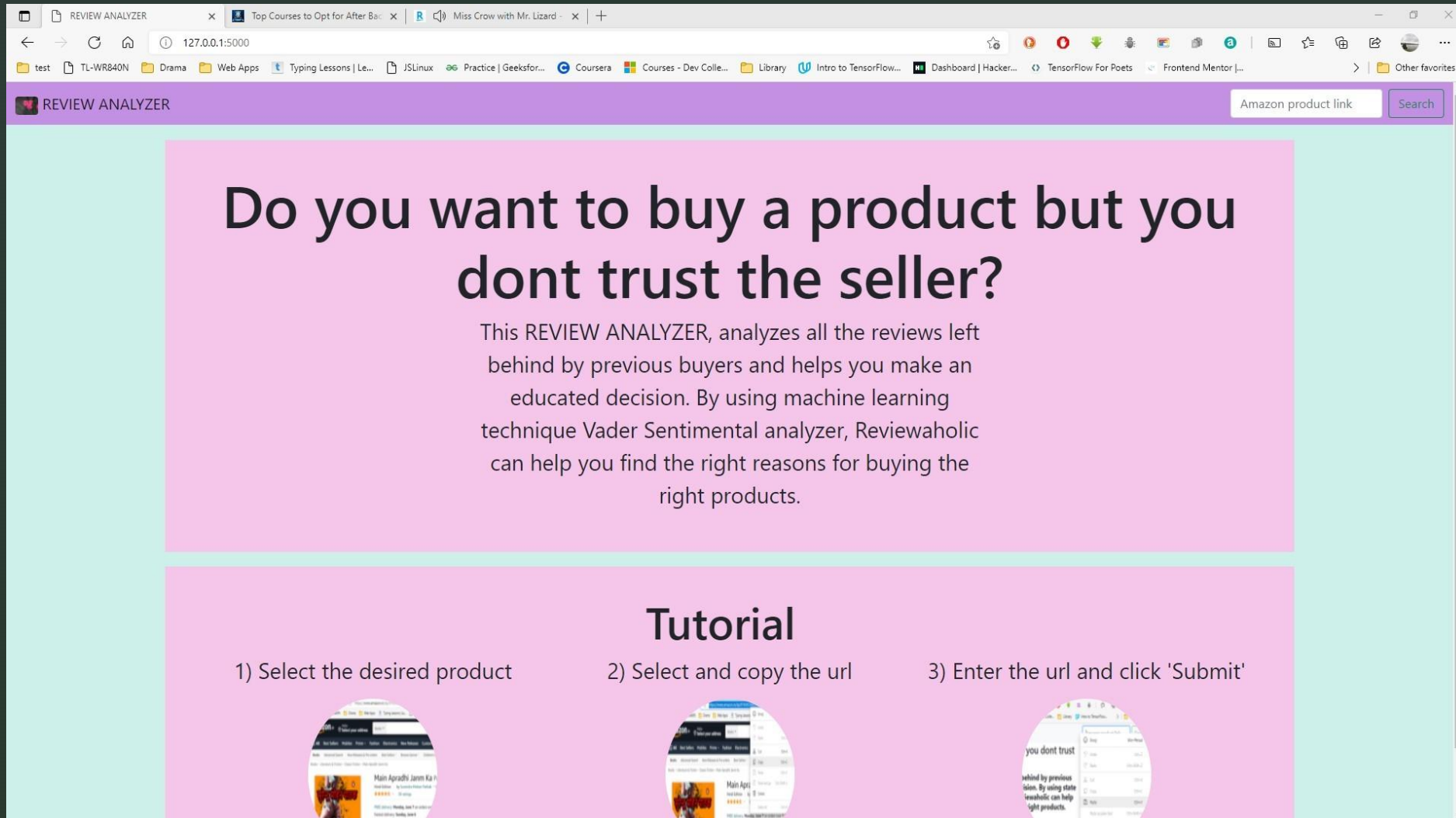
Inclusive of all taxes

FREE delivery: Tomorrow

Order within 8 hrs and 1 min Details

Save Extra with 4 offers

SNAPSHOTS



The screenshot shows a web browser window with the REVIEW ANALYZER application. The browser's address bar shows the URL 127.0.0.1:5000. The application has a purple header bar with the text "REVIEW ANALYZER" and a search bar labeled "Amazon product link" with a "Search" button. The main content area has a light blue background. A large pink box contains the text "Do you want to buy a product but you dont trust the seller?" followed by a paragraph explaining the application's purpose. Below this, a pink box titled "Tutorial" contains three numbered steps: "1) Select the desired product", "2) Select and copy the url", and "3) Enter the url and click 'Submit'". Each step is accompanied by a circular screenshot showing a product page from Amazon.

REVIEW ANALYZER

Amazon product link Search

Do you want to buy a product but you dont trust the seller?

This REVIEW ANALYZER, analyzes all the reviews left behind by previous buyers and helps you make an educated decision. By using machine learning technique Vader Sentimental analyzer, Reviewaholic can help you find the right reasons for buying the right products.

Tutorial

- 1) Select the desired product
- 2) Select and copy the url
- 3) Enter the url and click 'Submit'

■ FUTURE WORKS

- If partnered with amazon.in you do multiple requests without using any round about method.
- If Vader dataset improved it can access more words
- If Vader algorithm is improved we get better result
- Implement continuous learning
- Improve loading time

CONCLUSION

- The task of sentiment analysis especially in the domain of web scrapping is still in developing stage and far from complete. We propose a model which we feel are worth exploring in the future and may result in the further improved performance. In this way, the effects of human confidence can be visualized in sentiment analysis



THANK YOU

