

**SCHOOL OF SCIENCE, ENGINEERING AND TECHNOLOGY**

**DEPARTMENT OF COMPUTER SCIENCE**

**RESEARCH PROPOSAL**

**EVENT FINDER SYSTEM**

This is an original project report submitted to the Department of Computer Science by;

PAUL KIBET GETATA CS/MG/0815/05/18

Supervised by; Mr. Simon Ruoro

**DECLARATION**

This system is an original work developed by me except for where indicated, it has not been presented for any degree or award in any University.

Paul Kibet Getata

………………………………………..

**CERTIFICATION**

The undersigned certifies that he has read and hereby recommends for acceptance of Kabarak University a system entitled Event Finder System.

………………………………….. …………………………………..

Mr. Simon Ruoro Date

Department of Computer Science

Kabarak University

CONTENT

[DEDICATION 6](#_Toc89092573)

[ABSTRACT 7](#_Toc89092574)

[1. INTRODUCTION 8](#_Toc89092575)

[1.1. BACKGROUND STUDY 8](#_Toc89092576)

[1.2. STATEMENT OF THE PROBLEM 9](#_Toc89092577)

[1.3. OBJECTIVE OF THE STUDY 10](#_Toc89092578)

[1.3.1. GENERAL OBJECTIVE 10](#_Toc89092579)

[1.3.2. SPECIFIC OBJECTIVES 10](#_Toc89092580)

[1.4. SIGNIFICANCE OF THE STUDY 10](#_Toc89092581)

[1.5. SCOPE OF THE STUDY 11](#_Toc89092582)

[1.6. LIMITATIONS 11](#_Toc89092583)

[2. LITERATURE REVIEW 12](#_Toc89092584)

[2.1. INTRODUCTION 12](#_Toc89092585)

[2.2. REVIEW OF OBJECTIVE ONE 12](#_Toc89092586)

[2.3. REVIEW OF OBJECTIVE TWO 13](#_Toc89092587)

[2.4. REVIEW OF OBJECTIVE THREE 14](#_Toc89092588)

[2.5. BENEFITS OF SCREEN USAGE TIME MONITOR TO THE USER 14](#_Toc89092589)

[2.6. GENERAL OVERVIEW OF LITERATURE RELATED TO THE MAIN CONCEPTS 15](#_Toc89092590)

[2.6.1. WEB CRAWLER 15](#_Toc89092591)

[2.6.2. RECOMMENDATION ENGINE 15](#_Toc89092592)

[3. METHODOLOGY 16](#_Toc89092593)

[3.1. INTRODUCTION 16](#_Toc89092594)

[3.2. RESEARCH DESIGN METHODS 17](#_Toc89092595)

[3.3. DATA COLLECTION 17](#_Toc89092596)

[3.4. DATA ANALYSIS 17](#_Toc89092597)

[3.5. UML DIAGRAMS 18](#_Toc89092598)

[3.5.1. CONTEXT DIAGRAM 18](#_Toc89092599)

[3.5.2. DATA FLOW DIAGRAM 19](#_Toc89092600)

[3.5.3. USE-CASE DIAGRAM 20](#_Toc89092601)

[4. SYSTEM IMPLEMENTATION AND DEPLOYMENT 21](#_Toc89092602)

[4.1. INTRODUCTION 21](#_Toc89092603)

[4.1.1. SCRAPY 21](#_Toc89092604)

[4.1.2. NATURAL LANGUAGE TOOLKIT (NLTK) 22](#_Toc89092605)

[4.1.3. PANDAS 22](#_Toc89092606)

[4.1.4. SCIKIT-LEARN 22](#_Toc89092607)

[4.2. BACK END DEVELOPMENT 23](#_Toc89092608)

[4.2.1. CODE TESTING 23](#_Toc89092609)

[4.3. CONCLUSION AND FUTURE WORK 26](#_Toc89092610)

[4.3.1. FRONT-END 26](#_Toc89092611)

[4.3.2. IMPROVED STORAGE 26](#_Toc89092612)

[4.3.3. IMPROVED VARIETY 26](#_Toc89092613)

[5. REFERENCES 27](#_Toc89092614)

## DEDICATION

I dedicate this project to our friends, family and the entire school faculty for their continued support.

## ABSTRACT

In today’s day and age, most if not everyone has access to a smart device; their phone, laptop or computer and as such they are continually being bombarded with new information on social media and the while only being rewarded with cheap sense of pleasure. As a result of this a lot of people, particularly those between the ages of 13 to 45, are addicted to their devices with the biggest culprit being the most portable one; the smartphone. Studies have shown that increase in mobile phone addiction has been linked to; rise in loneliness, depression, hindrance in creativity and deep learning, stress, anxiety and insomnia. As much as it is the source of a myriad of problems it still can be a very useful tool and could possibly be the very solution itself. We can make very good use of our devices to not only aid in managing our time, but also provide us with a platform that allows us to explore and find activities that are beneficial to both personal enjoyment and a boost to our overall wellbeing. There are a multitude of software such as inbuilt screen time monitoring software that aim to solve this problem however they all suffer from the same problems; either being too lenient and easily by passable or being too harsh and becoming increasingly annoying to deal with. The system here is one that would employ Machine Learning to help an individual find events that are near them and similarly recommend other events based on their choices.

## INTRODUCTION

### BACKGROUND STUDY

The concept of mobile phone addiction far precedes its relation with social media, however in today’s day and age it manifests itself the most through our smart devices. Our devices have created an amazing new world for us over the last decade; they have created a world of connectivity and convenience like none other. It is simply a matter of a few swipes and taps and we have access to limitless knowledge, information, products, services, statistics and even studies, the convenience of virtual money allowing us to access services from halfway across the world. Truth be told in a lot of ways our phones, laptops and computers have provided us with limitless possibilities but like all good things this comes at a cost, one that is slowly becoming very steep.

As I had mentioned earlier, addiction is not a foreign concept to humanity and is pretty common, from drugs to gambling, they all function in the same way; altering the chemical composition in our brain and releasing our feel good hormone, dopamine. Our devices do the same thing, every time we get a text, a new like, a new follow we get a pleasurable feeling and just like other substances we develop a tolerance and we need more each time; instead of one like, we need ten likes to feel good, instead of five minutes on Facebook we need twenty to get the same effect as before. It may not seem as serious as a cocaine addiction or what not, but it is for this exact reason that it is dangerous, we underestimate our addiction and we’re more likely to spend time on our devices thinking it the norm.

On a global scale, what exactly does this mean? Well, the more people are addicted to their devices the easier it becomes to be manipulated, with constant bombardment of advertisements, information, likes, posts, follows, newer apps, devices etc. It makes people as a whole a lot easier to control. On social media platforms where billions of people worldwide spend their time when they are on their device we have the unfortunate displeasure of viewing people who have a larger notoriety, this is usually accompanied with feelings of inferiority, self-comparison that in large has negatively affected the self-esteem of a lot of people.

In Kenya, it becomes a lot easier to see the extent to which this damage has been done, with our devices we have access to limitless potential; our internet providers from Safaricom to Airtel have provided us with ways through which we can transact on a larger scale from global to regional. With our devices we have been blessed to stay connected with friends and family especially during recent events with the epidemic. A good number of learning institutions were able to transfer onto online platforms, a good extension of what we as people do; we improvise, we adapt and we overcome.

Over the last 10 years less and less children physically go outside to play and socialize. It is more common today to find children who are better are using their parent’s devices than they are at their school work. People these days are more content with making noise on twitter than going out and participating in a charity event. We’re slowly becoming detached from the real world, less and less people know how to do physical activities; simple thing from how to do some basic plumbing, woodwork to more essential skills like changing a tire.

### STATEMENT OF THE PROBLEM

The problem simply put is that we as a society are very quickly becoming addicted to our devices and it is slowly poisoning us as people. Because of this we need to find a viable solution to help curb this problem, allowing us to make the most us of our devices while at the same time maintaining the benefits that our devices promote us in this modern world.

This is a very large problem and as such, a lot of research has gone into this. Institutions such as ReSTART Washington, that is the first internet rehabilitation center, exist to help critically addicted individuals reconnect with the real world and helping socialize them for real world interaction. But in a practical sense not everyone needs to go through an intense rehabilitation program to overcome it let alone simply affording it.

In the market several apps do in fact exist in the market for both mobile devices as well as personal computers and laptops. They however mostly have a flaw, they are able to limit screen usage to a certain degree some being very strong and fully blocking you out of any app.

### OBJECTIVE OF THE STUDY

### GENERAL OBJECTIVE

The main objective of the entire study is to overcome mobile phone addiction in an effective way that intends to not only break but aid in replacing bad habits.

### SPECIFIC OBJECTIVES

The study aims at achieving a universally applicable system whose main objectives are;

* To reduce and limit the screen time of today’s population – the average person spends a minimum of 4 hours on their device daily, with this project we intend to not only reduce that time but provide people with the tools they need to healthier and more productive activities to fill that extra time.
* To increase the physical activity levels in today’s youth – with this project, one of the ways this project intends to help is by suggesting activities in correlation with what interests the individual making it a bit easier to break through and potentially make an impact in an individual’s habits.
* To develop an adaptive system that learns and adapts – one of the biggest objectives of this project, one that intends to overcome the shortcomings of the individual, is having a system that will learn as it moves. Far too many programs and applications exist that are too rigid in their implementation.

### SIGNIFICANCE OF THE STUDY

The study’s finding will no doubt help people overcome or simply just improve their lifestyle from one that is almost entirely revolving around what our devices say to one that gives back the freedom back to the people. Having an unbiased platform that grows in alignment with what you as an individual values and understands what exactly bothers you, can make it easier for people to break through their addictions.

However this does come with the challenge of having a system that can intelligently distinguish between a positive use and a negative use of one’s device.

### SCOPE OF THE STUDY

The study will mainly deal with mobile phone addiction, with the major social media apps that are very addictive such as; Instagram, WhatsApp, TikTok etc. In accordance to that it should also help handle and distinguish other apps such as calendars, to do lists as more positive time expenditure.

In terms of activities, simple rendezvous sporting activities, clubs and societies in the area holding anything interesting within the individual’s specified radius. This could also involve online platforms that are better suited, if an individual is spending time online they might as well spend it doing something productive.

### LIMITATIONS

There are unfortunately going to be a number of limitations during this entire process;

* A system that can effectively distinguish and learn what applications are problematic in itself seem quite problematic. One could be using YouTube tutorials while at the same time could be watching nonsense on YouTube making it a lot harder to distinguish whether a person’s internet usage is positive or negative without making too many assumptions.
* The whole activity recommendation thing though can be very useful in helping people connect and interact, it is ultimately limited especially in Kenya as a lot of areas do not publically post about their events on the internet unless in more urban areas such Nairobi, Nakuru, Naivasha and Mombasa. This provides a smaller target group than is preferred

## LITERATURE REVIEW

### INTRODUCTION

The topic as has been previously been mentioned, is about the effect of mobile phone addiction and how best we can revolutionize the current measures aid in defeating or help people develop better and healthier relationships with their devices.

### REVIEW OF OBJECTIVE ONE

Our first objective is to reduce the screen time usage.

Over the last few years there has been a rise in the number of developers whose intent have been to aid in this reduction. According to WHO, an increase in the use of our devices, particularly our mobile phones, has resulted in a number of conditions both physical and mental; increase in sleep disorders due to the strong blue light emitted by our devices, a number of back problems from poor posture that is as a result of constantly looking down on our phones, narcissism from over use of social media, as well as an over dependency on our devices.

In the United States, reSTART was founded in the state of Washington in the year 2009, and for more than a decade they have had over 90,000 cases of individuals struggling with internet and video game addiction. The latter has by far made up a considerable number of the patients and for good reason, gaming in the modern world is both a blessing and a curse, it is a way millions of computer enthusiasts like myself can make their mark on the world. However on the other side it is a source of despair, yes it can be used as a very good relaxer, it is very good as an indoor social activity but truth be told a lot of people are doing as more than just the occasional relaxer. According to financeonline.com, 2.69 billion gamers exist worldwide and the number is steadily rising. At the top of the food chain mobile users make up 40% of this already enormous number. With 23% of these gamers spending over 11 hours per week on nothing but video games.

This has prompted the creation of a number of applications to aid as tools for individuals such as; Forest, Space, Flipd as well as inbuilt screen time monitoring applications who’s aim to give access to the individual a way to help them keep track and limit their usage. They simply act as reminders to stay below a certain line, Forest for example acts as a reminder for one to keep off of their phone otherwise a virtual tree that has been planted will wilt and die. This obviously has its uses, it adds the element of a bigger picture allowing one to help hone their sense of discipline but it does still have its limits, individuals overtime do tend to lose their sense of appreciation of the app once they do remember there is nothing real at stake and slipping back into old habits becomes real.

### REVIEW OF OBJECTIVE TWO

Our second objective is to aid in increasing mobility in today’s demographic.

According to WHO more than a quarter of the world’s population is insufficiently active. This, in accordance with their studies, is due to an increase in inaction during leisure time and an increase in sedentary behavior with work, as work is becoming more and more digitalized, more and more people spend less time moving. Just as things were during the height of the Corona Virus pandemic, a huge number of institutions relocated their work online; this did allow for a good number businesses and essential services such as; schools, media, restaurant etc. to remain a float but it did impact people’s levels of physical activity. Even simple things such as walking to and from work or school helps, with people cooped up at home, this becomes problematic very quickly.

But similarly a lot of today’s activities can be accessed from an online platform, a good number of apps have already exploited this. Apps such as Hangtime, Meetup, Spotsetter etc. whose main aim is to help people find activities, events, clubs, societies in their area, allowing them to connect with new people who share similar interests. This is a demonstration of how our devices similarly can be a tool for us to reconnect with people and the world in a smarter, more efficient way.

### REVIEW OF OBJECTIVE THREE

To develop an adaptive system that learns and adapts.

This allows for a system that customizes itself to the user allowing for a better service. A version of this exists in the recommendation engine that a lot of social media apps that personalize the content a user consumes to their taste; YouTube, Instagram, TikTok all use this allowing the user to get a personalized feel of the application they are using.

We can adapt this to serve us in a way that we actively search for activities on the internet with our integrated web crawler and then personalize these based on our recommendation engine to suit the user.

### BENEFITS OF SCREEN USAGE TIME MONITOR TO THE USER

* Saves time – a lot of time is lost spending unnecessary scrolling, chatting, liking, playing games when it could be better off utilized doing not only more useful but more gratifying work elsewhere.
* Improved Mental Clarity – less time spent being bombarded with content on one’s device allows for better room to think.
* Improved engagement – apps that are built to suit to the user’s needs tend to have better engagement as they are designed to encourage activity by giving the user what they want.

### GENERAL OVERVIEW OF LITERATURE RELATED TO THE MAIN CONCEPTS

### WEB CRAWLER

This is essential to the system finding and crawling the internet for events and activities and collecting them in one place to form our starting data set. It is from this that we can then parse and decide what content is and isn’t good enough for our system.

### RECOMMENDATION ENGINE

This is one of the central algorithms that provides the personalized content to the user, by employing a matrix factorization technique, it allows it to learn and recommend based on what the user interacts more with as well as a what the user enters as explicit data, the more the individual uses the system, the more their profile is formed allowing better recommendations.

This improves from the use of simple explicit data, that is rigid and that the user has to regularly keep up to date in order for it to remain relevant, as such it improves user interaction.

## METHODOLOGY

### INTRODUCTION

Research methodology guides a researcher in collecting, analyzing and interpreting observed facts. In this chapter we are going to introduce a logical framework to be followed in the study.it is divided into; research design, population and sample, data collection and data analysis. This framework is used to structure, plan and control the process of developing the actual website. There has been a wide variety of frameworks which have evolved over the years each with its own recognized strengths and weakness. Our website methodology is not necessarily suitable for use for all projects since each available methodology is best suited for specific kind of projects based on different technical, organizational and team considerations.

The project was based on answering a number of questions that were crucial to progress;

* What is the most effective way to break a detrimental device usage?
* How effectively can we maintain user engagement?

Efficiency in habit breaking

This is to be achieved by not simply resisting an urge but by regularly reinforcing the idea that there are better activities the user can spend their time doing, ranging from simple stuff such reading a book to more complicated such as attending a nearby event.

Efficiency in maintaining user engagement

Our app aims to be very flexible in dealing with people’s needs and problems and through this we aim to provide a personalized experience to individuals.

### RESEARCH DESIGN METHODS

There are several data mining techniques that have been developed in data mining projects in the past. These techniques include association, classification, clustering, prediction and sequential patterns.

A Content- based filtering algorithm; according to specific density, these algorithms group objects according to functions of a goal. It is more reliant on the product itself, therefore it doesn’t require other user interaction in order for it to work. This makes it ideal for the system.

Knowledge-based – this will make suggestions primarily on the user’s preferences and needs, making it more efficient as time goes on as what the user prefers becomes sharper the more the system is employed.

### DATA COLLECTION

In order for our recommendation engine to be able to work effectively, we need to be able collect this data and then stored on a data file, allowing for its reacquisition and reuse in informing the recommendation engine that scrapes the web looking for different suited activities that are more suited to the individual’s personal taste as well as personal goals.

Python is embedded with a wide variety of packages such as Scrapy that is suited in web scraping allowing for data collection, Pandas that allow for data management and organization and Scikit-Learn that allows us to perform calculations on our data and potentially train our machine accordingly.

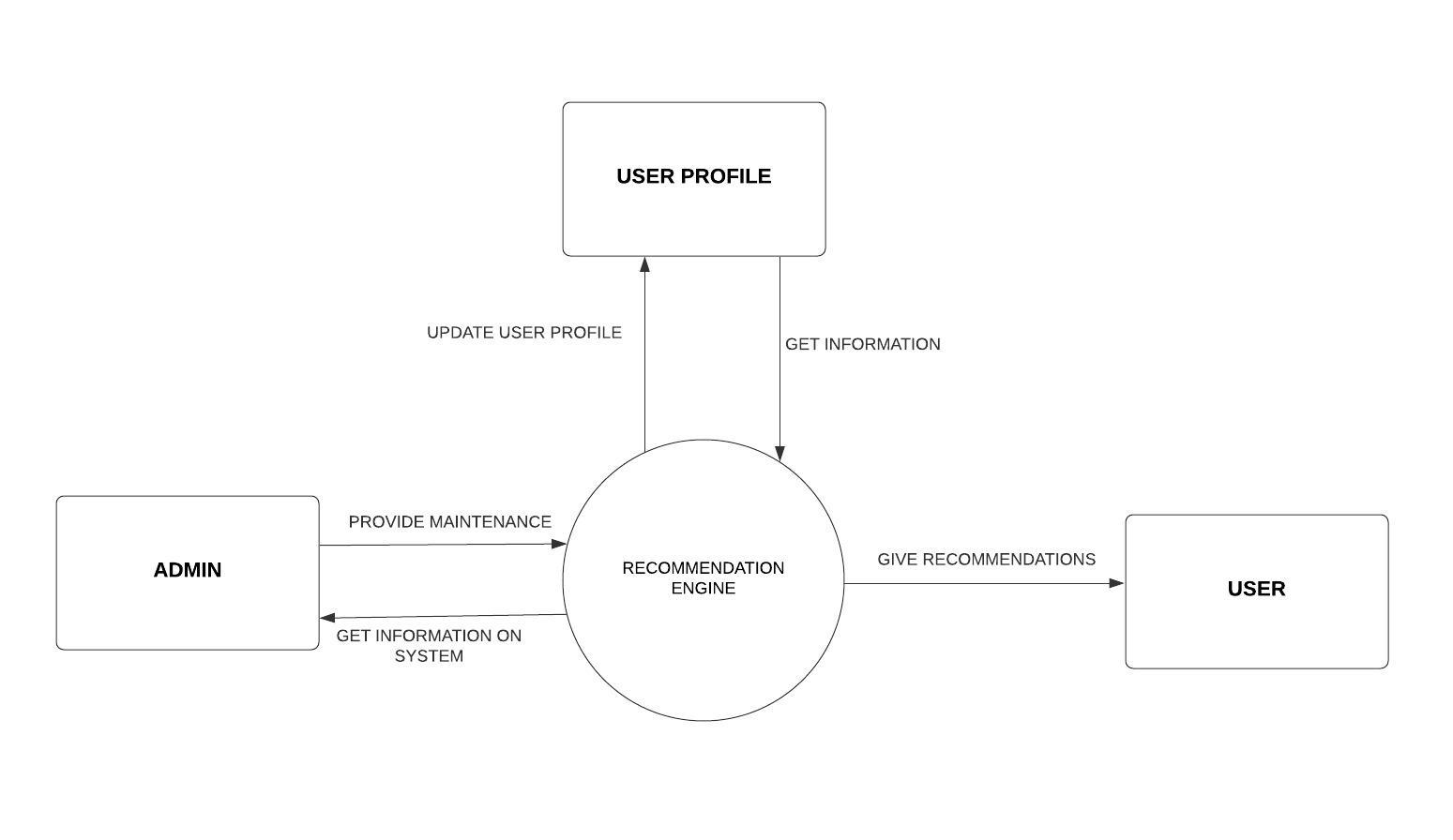
### DATA ANALYSIS

Data is collected, cleaned, parsed and then analyzed by the tools provided to us by the Scikit-Learn library.

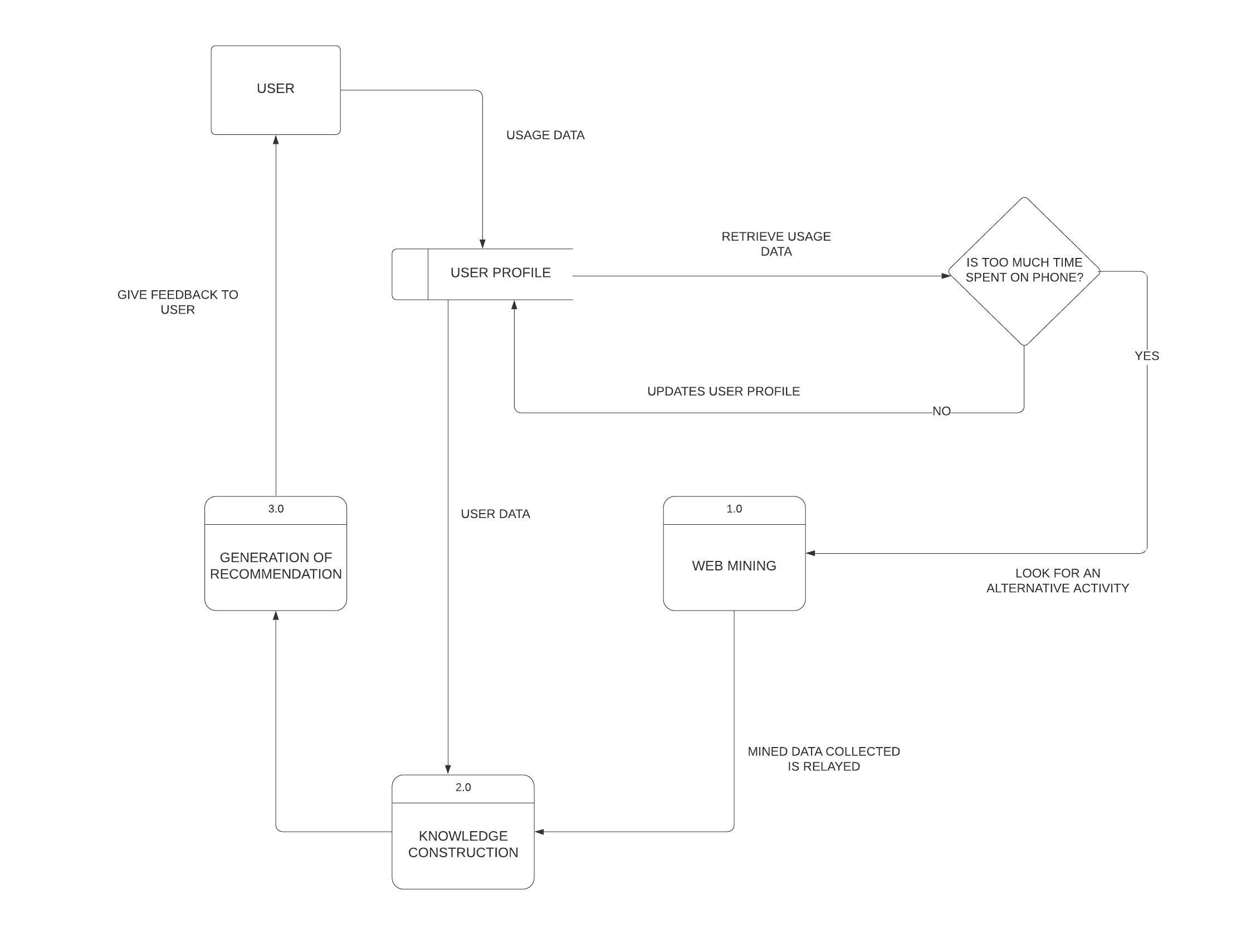
It features a variety of algorithms such as vector machine, random forests and k-neighbours as well as supporting other scientific libraries such as NumPy and SciPy.

## UML DIAGRAMS

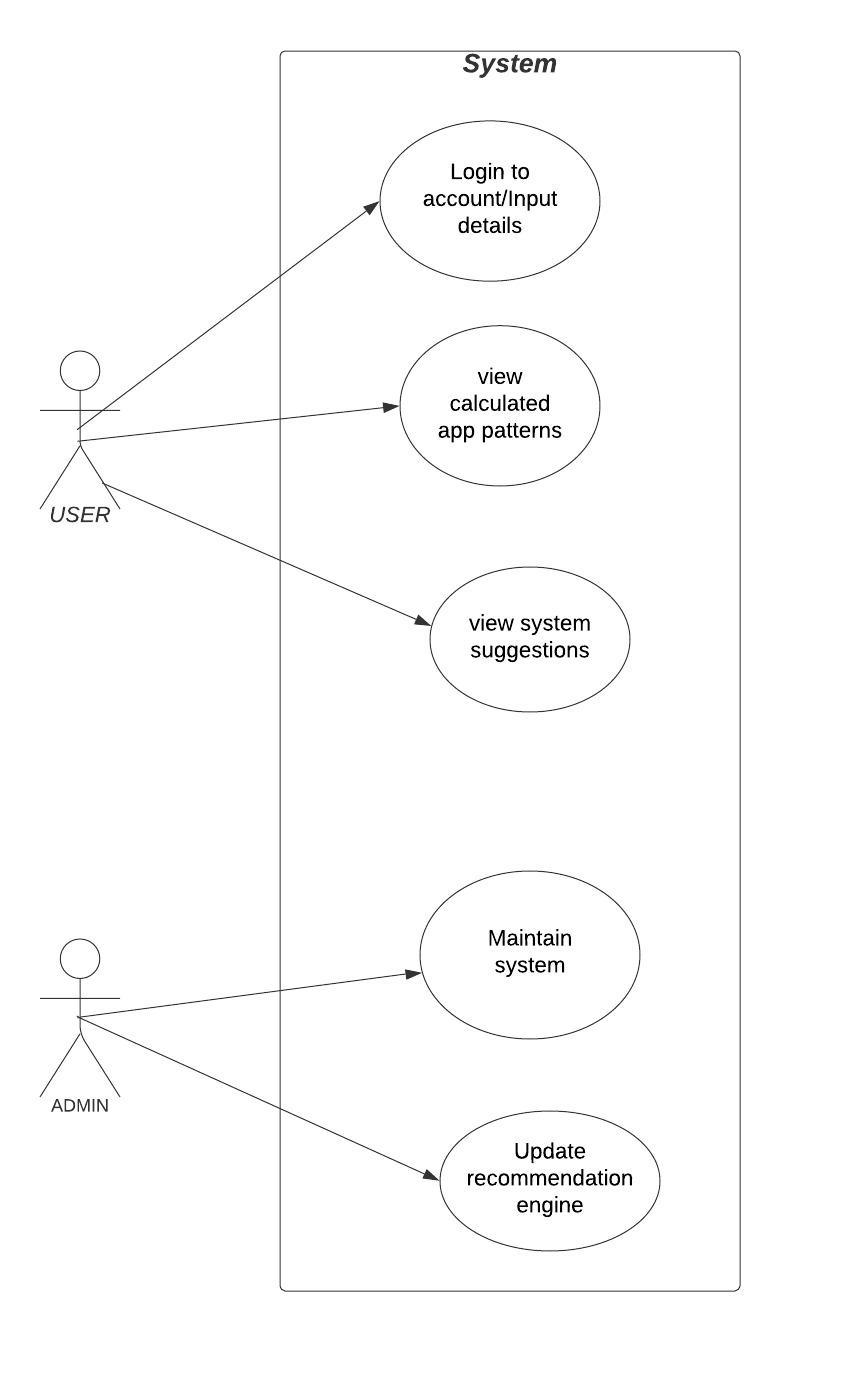
### CONTEXT DIAGRAM



### DATA FLOW DIAGRAM



### USE-CASE DIAGRAM



## SYSTEM IMPLEMENTATION AND DEPLOYMENT

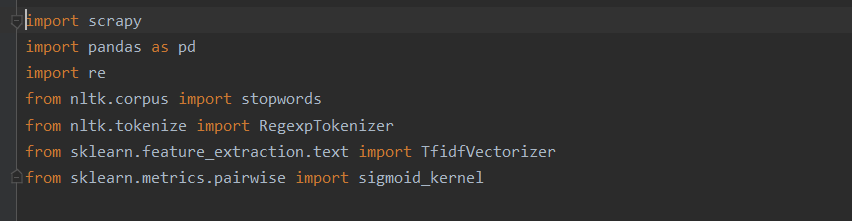
### INTRODUCTION

The system in itself, though not fully robust makes up for this by delivering on what it set out to accomplish. The system effectively runs a web crawler that is used to access the internet and scrape for information, in this case it is events and the information about them.

After which the data is then cleaned and parsed; with unnecessary data removed retaining a very simplistic version of the original that is what is needed by the machine to perform its calculation. All of this was done in preparation for the analysis.

In which the words in the titles are converted into vectors, they are then used to create a sparse matrix for visualization that the machine uses to determine the similarities between different words and titles.

A number of libraries were used to achieve this, the most prominent being;



### SCRAPY

This is a fast and high-level framework designed to scrape and crawl websites, allowing it to be used to extract data from websites and HTML sites. In the system, we use it to data mine, collecting the data and storing it in a CSV file from which we access and use it later.

Scrapy allows us to extract the sought after information, in this case the event titles, dates, location and the event link, if the user is interested in a particular event they can access it accordingly. Here CSS selectors are used to point to the relevant data in the html of the web page.

### NATURAL LANGUAGE TOOLKIT (NLTK)

This is another popular library that finds use in our project. It is an extensive library designed for the interpretation of human language data by machines. It provides a simplistic user interface with a large array of tools such as; text processing libraries, classification, tokenization, stemming, tagging, parsing and even semantics.

In the system, it does not do anything different, as our recommendation system will require a mechanism in order to allow it to understand what is in the titles in order to calculate relations, it comes in handy here. Before we can do anything, NLTK allows us to clean our data, in this case the title, allowing us to remove non English stop words, nonascii characters, make the text lowercase etc.

After which, we develop vectors from the words in the cleaned title, that will then be plugged into the recommendation system in the form of a matrix allowing for similar items to be determined.

### PANDAS

This is a fast and efficient library that provides great tools for reading and writing data between data structures, as we use here to read the CSV file multiply and write back to it.

It provides us with intelligent label-based slicing, fancy indexing and sub setting of data sets. In the system, we index data after its transformation in order to reorder it as some of that is lost during its transformation.

All in all pandas is a great library providing us with great tools to help manage our dataset as we move it along the entire computational process of the system

### SCIKIT-LEARN

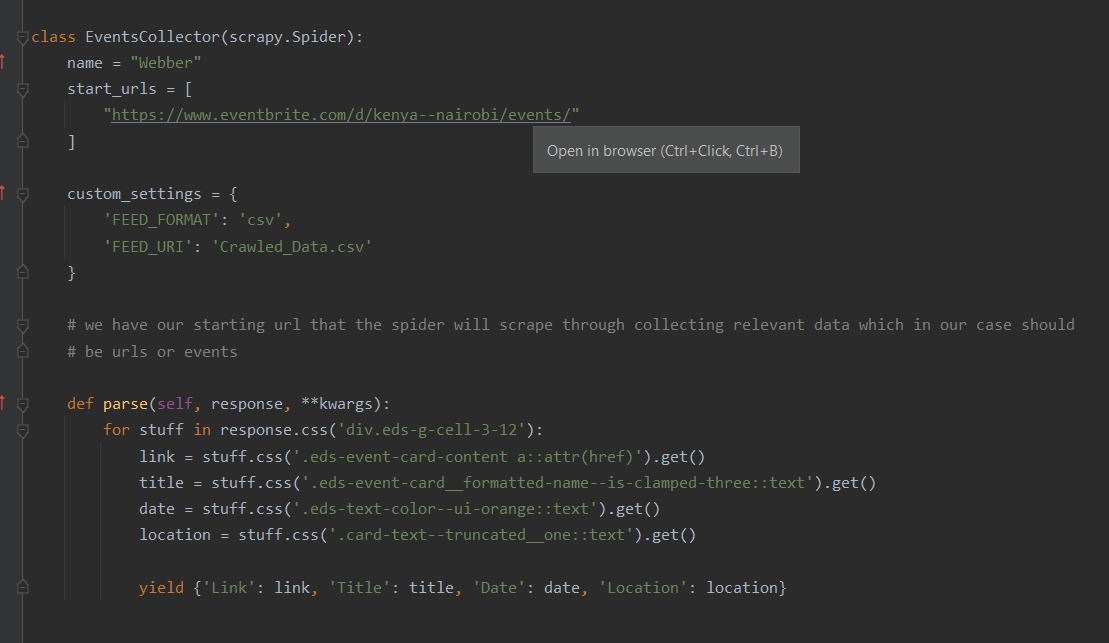
This is the crux of the recommendation engine, scikit-learn is a machine learning based library that provides efficient and simple tools for us to analyze our data. It provides a number of efficient tools such as classification, regression, clustering and even dimensionality reduction.

Scikit learn is used here in the conversion of the words in our titles into vectors and the formation of a sparse matrix, it is through this matrix that we are able to calculate the cosine similarity of the various titles based on the sigmoid curve allowing us to determine similarity of one to the next.

### BACK END DEVELOPMENT

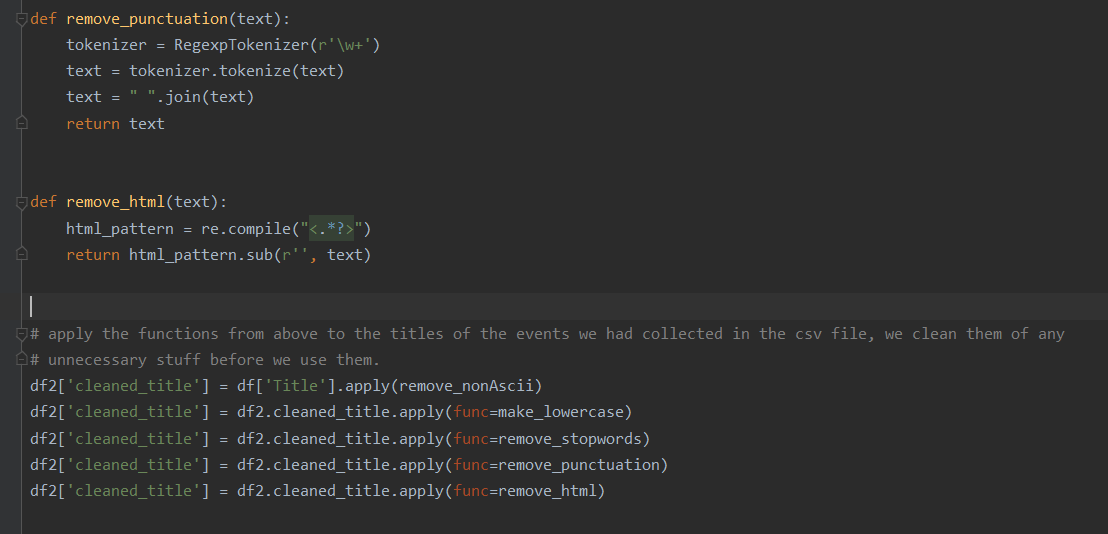
### CODE TESTING

The first section of the code is a web crawler that uses the very robust Scrapy library to crawl the internet and acquire the different URLs, in this case we are scraping the site collecting a number of data; in this we have the events, the location, date and the necessary URL containing that information.

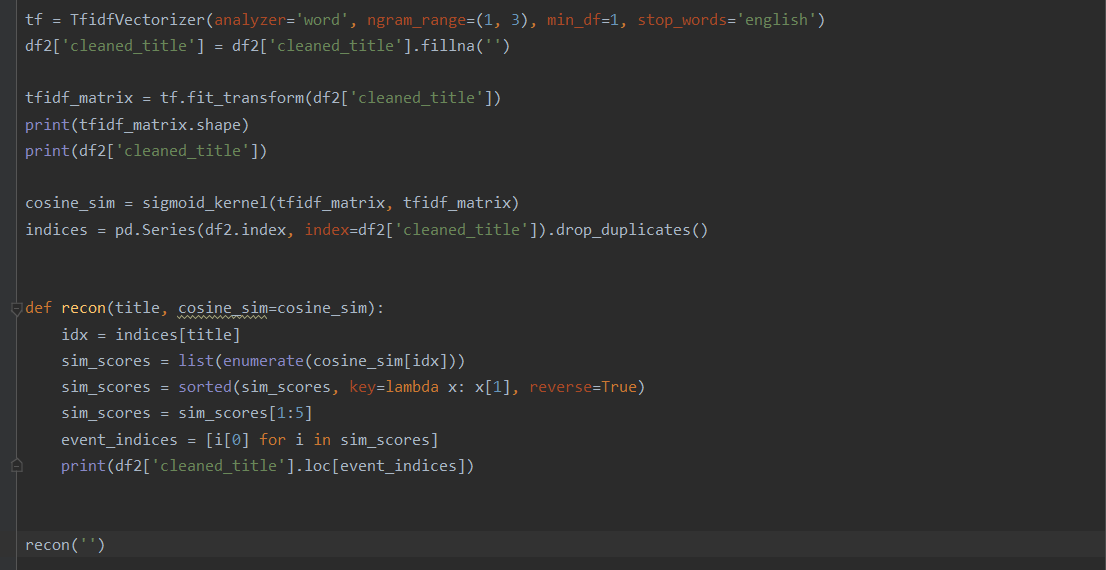


After we have scraped the data we store it into a CSV file, we are interested in the title here in order to make our recommendation but the data lacks a lot of structure; we then need to clean out the title, first removing the redundancy afterwards we clean the title by; removing non-ascii characters, removing all caps, removing stop words etc.





We then need to perform our recommendation, in order to do this we transform our cleaned title into a vector matrix in order for us to be able to visualize and calculate similarity. We use the scikit learn module to perform this converting titles into vectors and using the vectors to develop a sparse matrix, which is a matrix storing a large number of zero-valued elements that is to be used in processing data.



### CONCLUSION AND FUTURE WORK

The system as a whole though in need of work has proven to be a potential solution aiding people find new ways to connect and interact with other people and other more fulfilling and constructive activities outside of the virtual realm.

### FRONT-END

Despite having a Crawler API that provides access to all of the Scrapy core components making it possible to call Scrapy from a script rather than through the terminal in Scrapy Crawl as implemented above, it is still rather complicated to link the spider to a framework such as Flask or Django. However it is implementable allowing a cleaner implementation for a web app and possibly to be scalable down to a mobile application.

### IMPROVED STORAGE

The system’s form of data storage has a lot of room for improvement, due to the lack of a proper connected database a less desirable and robust algorithm that is the content based filtering. With the implementation of a database rather than a data file acting as the data frame, we can ensure longer-term data security.

### IMPROVED VARIETY

With Web scraping we can add a lot more starting URLs, a simple way we can expand the amount of content we have access to, in this case we only have one url limiting the amount of data we have access to.

As well giving access to Scrapy to social media sites, although for a lot of cases this requires verification example of which is Twitter that requires authentication before scraping of the site can properly be done. But it would open up, as a lot of people rather than set up their own websites use social media as their main platform for marketing.

## REFERENCES

H. Liu, J. He, T. Wang, W. Song, and X. Du, “Combining user preferences and user opinions for accurate recommendation,” Electronic Commerce Research and Applications, vol. 12, no. 1, pp. 14–23, 2013

K. Zhang, K. Wang, X. Wang, C. Jin, and A. Zhou, “Hotel recommendation based on user preference analysis,” in Proceedings of the 2015 31st IEEE International Conference on Data Engineering Workshops (ICDEW), pp. 134–138, IEEE, Seoul, South Korea, April 2015.

<https://scrapingrobot.com/blog/data-collection-methods>

<https://www.datacamp.com/community/tutorials/recommender-systems-python>

<https://www.cloudflare.com/learning/bots/what-is-a-web-crawler>

<https://storm.apache.org/releases/current/Tutorial.html>

<https://towardsdatascience.com/why-tiktok-made-its-user-so-obsessive-the-ai-algorithm-that-got-you-hooked>

<https://madasamy.medium.com/introduction-to-recommendation-systems-and-how-to-design-recommendation-system>

<https://www.netaddictionrecovery.com>

<https://emerj.com/ai-sector-overviews/use-cases-recommendation-systems>

<https://medium.com/@m_n_malaeb/recall-and-precision-at-k-for-recommender-systems>