**Visionalyst**

**Outline**

1. INTRODUCTION

1.1 Problem statement

1.2 Background

1.3 System purpose

1.4 System scope

1.5 System objectives and acceptance criteria

1.5.1 Objectives

1.5.2 Success criteria necessary for the project

1.6 Life cycle model

1.7 System methodology

1.8 Solution statement

1.8.1 System overview

1.8.2 Limitations

1.9 Expected results

1.10 Related work PLANNING

2. PLANNING

2.1 Project plan

2.2 Constraints

2.3 Risks ANALYSIS

3. ANALYSIS

3.1 Stakeholder's analysis

3.2 Proposed Solution

3.2.1 Functional requirements

3.2.2 Non-Functional requirements

3.2.2.1 Usability

3.2.2.2 Reliability

3.2.2.3 Performance

3.2.2.4 Supportability

3.2.2.5 Interface

3.3 System models

3.3.1 Scenario

3.3.2 Use case models

3.3.2.1 use case actors

3.3.2.2 use case narratives

3.3.2.3 use case diagrams

3.4 Storyboard

4. DESIGN

4.1 System decomposition

4.2 Object model

4.2.1 Class diagram

4.2.2 Class diagram data dictionary

4.3 Entity relationship diagram

4.4 User interface design

4.4.1 Prototype screenshots

4.4.2 Navigation hierarchy

1. **Introduction**
   1. **Problem statement**

Although most of us use social media platforms to convey our personal feelings and opinions for the world to see, one of the biggest challenges lies in understanding the ‘sentiments’ behind social media posts.

Understanding people’s emotions has become more essential and a challenge for businesses since people express their feelings more openly than ever before.

Translating text (social media conversations, tweets, or posts) allows stakeholders to listen attentively to their customers and meet their needs.

* 1. **Background**

Before sentiment analysis data was unorganized, huge volumes of text data are created everyday (emails, chats, social media conversations, surveys, articles, documents, etc.) and it was very hard to analyze, understand, and sort this data not to mention the time it would take to do all of that and it will be considered an expensive task.

Sentiment analysis, however, help people make sense of all this text by automatically understanding, processing, and then tagging and sorting them accordingly.

* 1. **System Purpose**

**Our system focuses mainly on searching about the needed topics (most important two are):**

* + Business: In marketing field companies use it to develop their strategies, to understand customers’ feelings towards products or brand, how people respond to their campaigns or product launches and why consumers don’t buy some products.
  + Politics: In political field, it is used to keep track of political view, to detect consistency and inconsistency between statements and actions at the government level. It can be used to predict election results as well!

**The main benefits of sentiment analysis include:**

* **Sorting Data at Scale**

**Can you imagine sorting thousands of tweets or conversations?**

**Sentiment analysis helps businesses process these data in an efficient and cost-effective way.**

* **Real-Time Analysis**

Sentiment analysis can identify critical topics in real-time, for example in

the USA election, by providing the candidate with real time tweets, analysis might help one of the candidates to increase popularity on a region, it also might help to know if one candidate is escalating over the other.

* 1. **System scope**
     1. **In scope**

Our scope here is collecting data from social media like twitter, by fetching the tweets from it using twitter API and analyze it in order to provide the users with charts contain summarizations about negative, neutral and positive tweets for what they are searching for..

* + 1. **Out of scope**

Our main focus now is only twitter API as it is the most reliable source for data.

* 1. **System objectives and acceptance criteria**
     1. **Objectives**

**Our objective is to make a sentiment analysis website that will be able to do the following:**

* Search for a product name (ex. iPhone 11) and then analysis will be made on tweets made on twitter and show charts with the customer satisfaction categories, for example Very Bad, Bad, Neutral, Good and Very Good and show the tweets fetched with the sentiment’s status beside.
* Users will be able to register to save all the liked products in order to get back to the reviews and charts or to track the updates made about these products.
* Track the status of the elections and provide real time analysis of the tweets about the elections and candidates.
  + 1. **Success criteria necessary for the project**

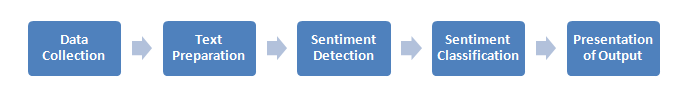
**In order our system to succeed:**

* Our analysis should be at least 80% accurate in order to make our data trusted and valuable for users.
* Retrieving data should be within 15 seconds at maximum.
* Charts provided should be easy to read that normal people can read it with ease, by providing tags and status of the product (very good, good, bad, or very bad with emoji).
  1. **Life cycle model**

Our project is following the **Agile model** for some reasons:

* Using agile model is better as we will be splitting our project into phases(stages) starting from “Data collection“ phase till “presentation of the output” phase, each stage has its own lifecycle that we can’t move from one stage to another before the Testing is accepted.
* The requirements might change or evolve as we move forward into the project without that affecting the project.
* Using agile will make us test our developed model on each stage simultaneously while developing the website instead of testing it just one time at the end of the project like in (waterfall model) so it will help us make sure that there is no bugs or errors on each stage before we move to the next stage.
  1. **System methodology**

**Sentiment analysis passes throw 5 steps:**

****

* **Data collection:**

People usually express their sentiments on social media like Twitter, Facebook, and blogs.

As people express about their feelings differently, with different vocabularies, context of writing, using short forms or even slang, that, in return will make the data huge and unorganized. Manual analysis here would be impossible

There for special programming languages like “python” will be used to help us process and analyze the data.

By using python libraries like “twint” or “tweepy” to help us fetch twitter tweets to analyze them.

* **Text preparation**

Text preparation is nothing but filtering the tweets we have fetched before analysis.

It includes identifying the contents of the tweets and eliminating the non-textual content and the content that is irrelevant so it is only the desired text remains for analysis.

Example, Data cleaning algorithms or regex techniques with python.

* **Sentiment detection**

After filtering the tweet, each sentence of the desired text is examined for subjectivity.

Sentences with subjective expressions are retained, and that which conveys objective expressions are discarded.

And that would be done using common computational tools and libraries like “NLTK”.

* **Sentiment classification**

Here in this stage we classify the sentiments into two groups, Positive and Negative.

Each subjective sentence detected is classified into groups good or bad, like or dislike, positive or negative.

* **Presentation of the output**

At the end we here convert the unstructured text into meaningful information.

After the completion of the analysis phase, the text results are displayed on graphs like “Pie charts”, “Bar charts” and “Line graphs” for the end user to understand the final results.

* 1. **Solution statement**
     1. **System overview**

The proposed system is a website named (Visionalyst), it is a website that analyzes the data of a certain word the user is searching for and give him results in a shape of some charts, categorized into specific thing about it.

For example if the user is searching for iPhone12 then the website will show a sample of tweets of what people think about it and then provide some charts about the iphone12 itself, is it bad or good and will dive deeper into specific parts like battery, camera, screen and so on. Finally, it gives a good insight of the product or the word in general which the user is searching about.

* + 1. **Limitations**

Sometimes it’s hard for humans to understand sentiments accurately imagine for computers.

Sentiment analyses is one of the hardest tasks in natural language processing as it faces limitations and challenges that makes it hard to understands human sentiments.

* + **Subjectivity and tone**

There are two types of text (Subjective and objective)

Objective texts do not contain explicit sentiments, while subjective texts do.

Anything objective sticks to the facts, but anything subjective has feelings.

For example

The package is nice.

The package is red.

Here people would say that sentiment is positive for the first one and neutral for the second but all predicates(adjectives, verbs, and some nouns) should not be treated the same with respect to how they create sentiment. Here nice is more subjective than red.

* + **Context and polarity**

Analyzing sentiment without context is pretty difficult as machine cannot learn about contexts if they are not mentioned explicitly.

One of the problems that arise from context is the change in polarity.

**Everything of it.**

***Absolutely nothing!***

Those two answers can carry different meaning depends on the question

Let us say: **what did you like about the event?**

Here the first answer would be positive, and the second would be

negative.

But what if the question was: **what did you dislike about the event?**

Here the first answer would be negative, but the second would be positive.

The negative in the question will change the analysis.

* + **Irony and sarcasm**

Sometimes people express their negative sentiments using positive words and vice versa, and that can be difficult for computers to detect without a prior understanding of the context of the situation.

Imagine someone saying

"I'm really loving the enormous pool at my hotel!", if this statement is accompanied by a photo of a tiny swimming pool.

The computer would classify this sentence as a positive sentiment while the one wrote this tweet meant the opposite and it was a sarcastic tweet but it is hard for the computer to understand sarcasm and irony without prior understanding of the context.

* 1. **Expected results**

**The client should expect from our website:**

* A sample of tweets of what people says about the topic he is searching for.
* Charts about the tweets itself, is it good, neutral or bad.
* Charts of the percent of each gender said good and bad.
* Deeper insights of the tweets to give the user a clear satisfying overview.
  1. **Related work**
* **[talkwalker.com](https://www.talkwalker.com/)**
  + Talkwalker is AI powered analysis provides real time insights into what's happening on all social channels and online media (Mainly Twitter), across 187 languages. This enables you to quickly identify issues and complaints before a crisis hits.
* **[social-searcher.com](https://www.social-searcher.com/)**
  + Social search is a behavior of retrieving and searching on a social searching engine that mainly searches user-generated content such as news, videos and images related search queries on social media like Facebook, LinkedIn, Twitter, Instagram and Flickr.

1. **Planning**
   1. **Project planning**

**(GANTT CHART)**

* 1. **Constraints**
  + Device with a browser
  + Internet connection
  + Technologies and tools:
    - Python
    - Twitter API
      * Preferred more than one account to avoid rate limits.

| **Resource** | **Endpoint** | **Requests per 15-minute window unless otherwise stated** | |
| --- | --- | --- | --- |
| **Per app** | **Per user** |
| Tweets | Tweet lookup | 300 | 900 |
| Recent search | 450 | 180 |

* 1. **Risks**
  + System failure.
    - Examples:
      * Errors on twitter servers or twitter API.
      * Error on our website server.
  + Twitters users spamming tweets may affect the accuracy of the insights

1. **Analysis**
   1. **Stakeholder’s Analysis**

|  |  |  |  |
| --- | --- | --- | --- |
| **Characteristic** | **Business Stakeholders** | **Normal User** | **Election chef** |
| **Age/ gender** | 50-50% Male and Female,  Median age 40 | 50-50% Male and Female,  Median age 25 | 50-50% Male and Female,  Median age 25 |
| **Education** | Some higher education | Most have degrees | Some higher education |
| **Computer/Web experience** | Med to High | Med to High | Med to High |
| **Domain Expertise** | Med to High | Low to Med | Med to High |
| **Task Knowledge** | Med to High | Low to High | Med to High |
| **Expectation** | Ease of use, Accurate results(charts) & Speed of task | Ease of use, Readable charts & Speed of task | Ease of use, Accurate results(charts) & Speed of task |

* 1. **Proposed Solution**
     1. **Functional requirements:**
* **Sign up**

The user can sign up to our website by setting an email, or phone number and setting password for his account in order to save any desired insights he want to save or share with someone else and to track any updates about the searched topic.

* **Log in**

The user will log in by entering his email (or phone), and his/her password to show his account profile, see updates about his saved insights and reports, save more results.

* **Search**

For the user to find what he is looking for; he/she will have to search for the word he/she needs insights about, and the desired insights will be shown.

* **Save results**

The user can save the topics of what he/she was searching for by clicking on save to my topics button so he can find them later easily from his/her profile page.

* **Retrieve saved results**

The user can ask to retrieve the insights about the topics which is saved earlier from his/her profile such that the insights are directly shown from the saved list instead of searching for it again every time.

* + 1. **Non-functional requirements**
       1. **Usability**

**Our website will help the user to easily use it from the start of when log in until he leaves the website with what he was looking for.**

* Provide keep me logged in checkbox, he will not need to log in every time he visits our website.
* Sign up with social media accounts condition that he will not need to sign up.
* Help completeness while searching for a word or a product
* Provide colorful, simple and easy to read charts for the results so he gets what he is looking for.
  + - 1. **Reliability**
* Our website will be mostly working without system drop as we are going to host it on a highly demanded, secure hosting servers that would lead to high reliability of our website.
* Provide a backup of the user’ data to avoid any data loss whenever a system failure occurs.
* Website traffic up to 10000 user per hour with high performance and fast processing.
  + - 1. **Performance**
* It would take the user 5seconds maximum to be logged in to our website as the use of nowadays technologies made the interaction with the data base easier without the need to reload any pages.
* Apply the best algorithms and techniques in our source code so it won’t take more than 20seconds to retrieve the insights the user was searching for.
* It would take 2-5 seconds to search for the word as we provide word auto completion.
  + - 1. **Supportability**
* We will keep updating our website with new features and functionalities that meets the users’ satisfaction
* Adding more insights graphs and charts like the gender, race, and age.
* Increasing the capacity of the database when the number of the users increases rapidly.
* Improve time efficiency for report generation after searching for a topic.
  1. **System Models**
     1. **Scenario**

**Personas:**

1. **Personal Information.**

* Jhon Mark, Marketing Manager

(Trace customer satisfaction)

* 34 years old
* Married, one child
* Marketing degree
* Creative, Competitive, Strong attention to detail and has commercial awareness.

1. **Things he wants to know:**
   * Brand reputation.
   * Customer reviews.
   * Customer satisfaction rate.
   * Understand consumer behavior.
2. **Things he wants to do:**
   * Improving customer experience.
   * Improving marketing campaigns and product messaging.
3. **Personal Information.**

* Mohamed Magdy, Faculty Student

(Trace customer satisfaction)

* 21 years old.
* Single.
* Has computational skills.
* Love Technology.
* Using Laptops frequently for college.

1. **Things he wants to know:**
   * Information from people has the same laptop he wants to buy.
   * People’s reviews.
2. **Goals:**
   * Track the people’s reviews on products.
   * Making decisions about needed product.
3. **Personal Information.**
   * Peter Mark, President of the election campaign for Donald trump.

(Trace president election’s votes)

* + 51 years old.
  + Married.
  + Politics Degree.

1. **Things he wants to know:**
   * Opinions about Donald trump.
   * States that support or against Donald trump.
2. **Goals:**
   * Focus more on the states that are against Donald trump.
     1. **Use Case Models**
        1. **Use case actors**

* Our main actor is the user searching for a specific topic.
  + - 1. **Use case narratives**
* **Sign up**

|  |  |  |
| --- | --- | --- |
| Use Case ID: | 1 | |
| Use Case Name: | Sign up | |
| Actors: | user | |
| Pre-conditions: | User enters the data needed | |
| Post-conditions: | Forward him to the log in page | |
| Flow of events: | **User Actions** | **System Actions** |
| -User enters the email or phone number | -System validate the email or phone number the user entered  -a true small icon appears if the email or phone is valid  -a false sign icon appears if there is something wrong with the email entered or if there is already account with that email |
| -User enters the password | -system validates the password |
|  | -sends a message to the user if the account creation succeeded |
| Exceptions: | The user chose to sign up with social media account | |

* **Log in**

|  |  |  |
| --- | --- | --- |
| Use Case ID: | 2 | |
| Use Case Name: | Log in | |
| Actors: | user | |
| Pre-conditions: | User enters his data, and clicks on log in button | |
| Post-conditions: | Forward the user to his profile | |
| Flow of events: | **User Actions** | **System Actions** |
| -User enters the email or phone number | -check if the email or phone number exists in the data base  - if exist the system wait for the user to enter the password  -if not, the system warns the user with a message that the email is wrong |
| -User enters the password | -system validates the password  - if wrong, the system warns the user with a message that the password is wrong  - if the password is correct the system waits the user to click on the log in button |
| -user click on log in button | -system forwards the user to his profile page |
| Exceptions: | The user checked the keep me logged in box last time he logged in by himself | |

* **Search for a word**

|  |  |  |
| --- | --- | --- |
| Use Case ID: | 3 | |
| Use Case Name: | Search for word | |
| Actors: | user | |
| Pre-conditions: | User enter the word he wants to search about | |
| Post-conditions: | System forward the user to the Insights page | |
| Flow of events: | **User Actions** | **System Actions** |
| -User enters the word he is looking for its insights | -system shows the analytical results of the desired word in a form of detailed graphs and charts |
| Exceptions: | none | |

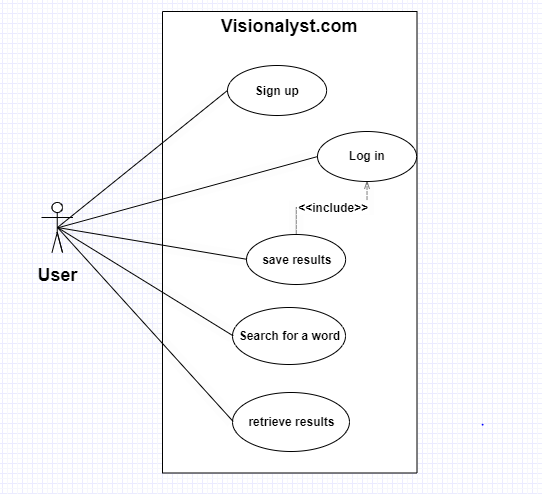
* **Save results**

|  |  |  |
| --- | --- | --- |
| Use Case ID: | 4 | |
| Use Case Name: | Save the results | |
| Actors: | user | |
| Pre-conditions: | User observing the insights of a word he entered | |
| Post-conditions: | Save the results into his profile database | |
| Flow of events: | **User Actions** | **System Actions** |
| -click on save results | - store the results at the data base  -inform the user that the results are saved and can find them in his profile |
| Exceptions: | none | |
| Include: | Log in | |

* **Retrieve results**

|  |  |  |
| --- | --- | --- |
| Use Case ID: | 5 | |
| Use Case Name: | Retrieve results | |
| Actors: | System (the website) | |
| Pre-conditions: | User search for a word | |
| Post-conditions: | Forward the user to the insights page | |
| Flow of events: | **User Actions** | **System Actions** |
| -search for a word | -system shows the analytical results of the desired word in a form of detailed graphs and charts |
| Exceptions: | none | |
| Include: | Search for word | |

* + - 1. **Use case diagrams**



* 1. **Storyboard**