# “I-Detox”

Submitted in partial fulfillment CR of the requirements of the degree

**BACHELOR OF ENGINEERING** IN **COMPUTER ENGINEERING**

By

**Muskan Chhabria 13**

**Divesh Chodda 14**

**Varun Salvi 51**

**Athurva Sawant 63**

Name of the Mentor

**Prof.** GEOCEY SHEJY



**Department of Computer Engineering**

**Vivekanand Education Society’s Institute of Technology**

**Hashu Advani Memorial Complex, Collector’s Colony, Chembur, Mumbai – 400074. India.**

**University of Mumbai**

**(A.Y 2022-23)**

# CERTIFICATE

This is to certify that the Mini Project entitled **“ I-Detox ”** is a bonafide work of **Muskan Chhabria (13). Divesh Chhoda (14), Varun Salvi (51), Athurva Sawant (63)** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **“Bachelor of Engineering”** in **“Computer Engineering” .**

## (Prof. Geocy Shejy)

Mentor

## (Prof. ) (Prof. )

Head of Department Principal

# Mini Project Approval

This Mini Project entitled “**I-Detox”** by **Muskan Chhabria (13). Divesh Chhoda (14), Varun Salvi (51), Athurva Sawant (63)** are approved for the degree of **Bachelor of Engineering** in **Computer Engineering.**

**Examiners**

**1………………………………………**

(Internal Examiner Name & Sign)

## 2…………………………………………

(External Examiner name & Sign)

Date: Place:

# Contents

## Abstract ii

## Acknowledgments iii

## List of Abbreviations iv

## List of Figures v

## List of Tables vi

## Introduction 1

* 1. Introduction
  2. Motivation
  3. Problem Statement & Objectives
  4. Organization of the Report

1. **Literature Survey 11**
   1. Survey of Existing System/**SRS**
   2. Limitation Existing system or Research gap
   3. Mini Project Contribution

## Proposed System 18

* 1. Introduction
  2. Architectural Framework / Conceptual Design
  3. Algorithm and Process Design
  4. Methodology Applied
  5. Hardware & Software Specifications
  6. Experiment and Results for Validation and Verification
  7. Result Analysis and Discussion
  8. Conclusion and Future work.

## References 32

**Abstract**

Internet addiction refers to excessive internet use that interferes with daily life. Due to its negative impact on college students' study and life, discovering students’ internet addiction tendencies and making correct guidance for them timely is necessary. In this system, we provide an approach to estimating a person's internet addiction levels using their behavior data on the system. In detail, we consider a person's addiction to the internet as a hidden variable that affects students’ daily time online together with other behavior. By predicting a person’s daily time online, we will find a person’s internet addiction levels. Finally, extensive experiments are conducted on a real-world dataset. The experimental results show the effectiveness of our method, and it is also consistent with some psychological findings. “I-detox” aims to provide real-time analysis of the exact number of websites visited by the user, giving them the freedom to foresee the probable results given the lack of schedule one follows while browsing the internet. Therefore, this gives a standardized analysis of the web history of the user also inclining the number of times a user visits a particular website. Our solution considers the suitability of terming the user as an addict or not based on the analysis provided upfront. Also, the tool can further be developed to encourage and know the mental state of the user based on his web-browser history and through some questionnaires, thus helping the user as a self or someone concerned with the particular person in question to know about his/her mental health being leaving out an option on seeking professional health or not! With the help of an artificial intelligence-based system, we tend to provide a humble service to the people of our society.

**Acknowledgements**

Efforts have been made on this project. But without the generous assistance and support of our project mentor, **Prof. Geocey Shejy**, it would not have been possible.

We owe a great deal to our project mentor for their direction, ongoing supervision, and provision of important project-related information.

We would like to thank the institution for giving us the opportunity to work on this project; we learned a lot from it.

Additionally, we would like to express our gratitude to everyone who helped us in completing our report efficiently and on schedule.

They provided us with insightful feedback that greatly aided in the preparation of this project.

**List of Abbreviations**

1. **NLP - Natural Language Processing**
2. **NB - Naive Bayes Algorithm**
3. **MultinomialNB**

**List of Figures**

**3.3.1) Flow Chart for Naive Bayes**

**3.3.2) Block Diagram**

**3.3.3) Modular Diagram.**

**List of Tables:**

## 3.6.1 Testing and training accuracy

**3.7.1** **Statistical Representation of Data (CSV) for Website classification**

**INTRODUCTION**

**1.1 Introduction**

Addictive digital experiences are becoming extremely common in the tech industry and the adjective “addictive” is now considered one of the highest compliments paid to an artifact. Existing research has explored what is considered addictive, what are the approaches and strategies performed by tech companies that eventually lead to addiction, and how cultural and societal pressures have contributed to this phenomenon. Addiction to technology has grown into a fully recognized disorder [1]. “I-detox” can be used from the young to the old because, in the age of digitalization, one never knows the extent of addiction one is suffering. Our solution gives the user a clear idea about the number of websites being seen, the type of content being surfed upon, the category it belongs to, and the URL.

**1.2 Motivation**

Our motivation is to aim and target the new age internet addiction. This addiction can spread from the young to old and can be equally dangerous no matter what age group it subsides too. Our application aims to help individuals who want to analyze their browsing pattern, and decide the chances of addiction.

**1.3 Problem Statement and Objective**

“I-detox” hopes to provide the user with an analysis of the number of websites visited by the user, the and no. of times the user visits a website and thus giving him an idea of his browsing method and the probability of being addicted to any of them. It provides a standardized approach for the user to assess his browsing pattern and also classifies the website based on which the user is able to understand the category of content he is browsing and whether it will lead to addiction. Artificial Intelligence and Machine learning-based approaches for the prediction of addiction and to classify the websites into categories. Based upon the data extracted from the user’s browsing history which includes the websites and categories the model would test the data and give the user the exact idea of his browsing pattern. There are some common categories to which websites belong and the ones visited by the user have a view-count associated with them. The solution will focus on both the categories and the view-count. For the categories, the solution will take into consideration the URLs and the bag of words from the websites. For the view-count, the user browsing history will be used which also provides the view-count of each website browsed. “I-detox” would thus indicate to the user if he requires any help to stop himself from being addicted to any harmful website.

The objective of our project is to create a model that studies the users browsing pattern using the provided history in the form of csv. The aim of this project is to classify the website the user is using and letting him know about his browsing information and the lets him know the time spent by the user on each website. Further to see whether the user is addicted to a certain website or if the user is on verge of getting addicted to a certain website by studying the history provided by the browser and also through the answers provided by the user in the questionnaire section.

**1.4 Organization of report**

|  |  |
| --- | --- |
| **Topic Introduction** | Introduction to our project: what is our aim and objective. |
| **Information Gathering** | Information about various pages, journals, documents that we have referred for our project and all the information which we have taken from each document. |
| **System Description** | Basic requirements for what all algorithms and libraries are being used for our project and its basic workflow. |
| **Implementation** | How the system is implemented in various different steps is being explained via flow diagrams. |
| **References** | This section is about the books, journals, papers and various things which we have used for our project. |

**LITERATURE SURVEY**

**2.1 Survey of Existing System**

A study on rough set-aided feature selection for automatic web-page classification by Wakaki et.al gives a basic idea about how the web-pages are classified based on categories.

Keras: The Python Deep Learning library by Son Erişimis a well built python library which makes our thoughts clear on the concepts of deep learning and provides an API for implementing it.

A study on Internet addiction and its relation to psychopathology and self-esteem among college studentsby Kumar Met.al gives basic ideas about Internet use and its relation to psychopathology and self-esteem among college students.

Internet Addiction Facts and Statistics By The Recovery Village. ByMegan Hul. Medically Reviewed By Dr. Andrew Proulx, MD helps us to understand what addiction is and how it is caused.

Internet Addiction: A Handbook and Guide to Evaluation and Treatment by Kimberly S. This book provides cutting-edge coverage by expanding the field to include specific problems such as online gaming, cybersex addiction, and gambling addiction. Its extensive attention to dealing with adolescents is essential, given the rapid rise in media and technology use by both Net Generation young adults and iGeneration teenagers.I am thrilled to have this invaluable, comprehensive, well-written resource for my own work and recommend it to people who need to understand this unique form of addiction.

Internet addiction and problematic Internet use: A systematic review of clinical researchbyDaria J Kuss and Olatz Lopez-Fernandez gives a comprehensive overview of clinical studies on the clinical picture of Internet-use related addictions from a holistic perspective. A literature search was conducted using the database Web of Science.

|  |  |  |
| --- | --- | --- |
| **Research Paper** | **Publication details** | **Description and Research Gap** |
| * **A study on rough set-aided feature selection for automatic web-page classification.**   [**https://www.researchgate.net/publication/220298511\_A\_study\_on\_rough\_set-aided\_feature\_selection\_for\_automatic\_Web-page\_classification**](https://www.researchgate.net/publication/220298511_A_study_on_rough_set-aided_feature_selection_for_automatic_Web-page_classification) | Wakaki, T., Itakura, H., Tamura, M., Motoda, H., & Washio, T. (2006). | A basic idea about how the web-pages are classified based on various  categories.  Research gap : This study is basically focused on the relevant dataset where complexity should be less as possible. This study was focused on Yahoo browser and its accuracy decreases with different browsers. |
| * **Keras: The Python Deep Learning library,**   <https://keras.io/> | Son Erişim: Mayıs 2018. | A well built python library which makes our thoughts clear on the concepts of deep learning and provides API for implementing so  .  Research gap: Tensorflow doesn't allow the training of the models and subsides a limited set of features for Windows users. |
| * **A study on Internet addiction and its relation to psychopathology and self-esteem among college students.**   [**https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6198588/**](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6198588/) | Kumar M, Mondal A.Ind Psychiatry J. 2018 Jan-Jun;27(1):61-66. doi: 10.4103/ipj.ipj\_61\_17. PMID: 30416293; PMCID: PMC6198588. | The main objective of this study is to explore Internet use and its relation to psychopathology and self-esteem among college students.  Research gap: This study is focused only on college going students; whilst internet addiction stays a prominent problem amongst every age group. |
| * **Internet Addiction Facts and Statistics By The Recovery Village.**   [Internet Addiction Facts & Statistics | Serious Social Problem | Learn More (therecoveryvillage.com)](https://www.therecoveryvillage.com/process-addiction/internet-addiction/internet-addiction-statistics/) | Editor Megan Hul. Medically Reviewed By Dr. Andrew Proulx, MD | This helps us to understand what is internet addiction, how it is caused and what are the symptoms for understanding this addiction.  Research Gap: This study is focused on the demographics of the addiction and its effects. Though this is an old study and much factual data is based on assumptions without any concrete proof. |
| * **Internet Addiction: A Handbook and Guide to Evaluation and Treatment. by Kimberly S. Young (Editor), Cristiano Nabuco de Abreu (Editor)**   [**https://books.google.co.in/books?id=C\_omSZQyfYcC&lpg=PA1&pg=PT7#v=onepage&q&f=false**](https://books.google.co.in/books?id=C_omSZQyfYcC&lpg=PA1&pg=PT7#v=onepage&q&f=false) | Dr. Larry Rosen, Past Chair and Professor of Psychology at California State University.  Hilarie Cash, PhD, coauthor of Video Games and Your Kids: How Parents Stay in Control, and cofounder of reSTART: Internet Addiction Recovery Program.  Ran Tao, MD, Professor and Director, and Xiuqin Huang, MD, Associate Professor, Treatment Center for Internet Addiction, General Hospital of Beijing Military Region, China.( 9 Nov 2010). | This book provides cutting-edge coverage by expanding the field to include specific problems such as online gaming, cybersex addiction, and gambling addiction. Its extensive attention to dealing with adolescents is essential, given the rapid rise in media and technology use by both Net Generation young adults and iGeneration teenagers.I am thrilled to have this invaluable, comprehensive, well-written resource for my own work and recommend it to people who need to understand this unique form of addiction.  Research Gap: |
| * **Internet addiction and problematic Internet use: A systematic review of clinical research.**   [**https://pubmed.ncbi.nlm.nih.gov/27014605/**](https://pubmed.ncbi.nlm.nih.gov/27014605/) | Daria J Kuss and Olatz Lopez-Fernandez (2016 March). | Comprehensive overview of clinical studies on the clinical picture of Internet-use related addictions from a holistic perspective. A literature search was conducted using the database Web of Science. |

**2.2 Limitation Existing system or Research gap**

1. On analyzing the current systems existing in the market , we find that the system’s do track their history into a file/text document but don’t go further to categorize it. This may serve as a problem since not every website can be identified by its url.
2. “I - detox” not only provides the history information into a (.csv) file but also categorizes the data into different types of websites adding onto the display of bags of words searched on a particular website.
3. Thus, this makes the system user-friendly providing clarity which website provides what purpose and service and further analyzing how much time was spent on a website.
4. Browsing history/ or what the user accesses the internet for can prove a major aspect to analyze the human-being as a whole and what he/she is thinking. On analyzing one can find if he/she is facing some mental problems.
5. Currently, no system or software goes out of the way to analyze a user’s browsing history and analyze their mental health.
6. That’s what “I-detox” brings to the table mixing both mental health and history tracker together and co-existing a space between the two.
7. Requires user’s consent to access his/her browsing history.
8. Is pretty much dataset dependent.

**2.3 Mini Project Contribution**

1. **Varun Salvi (51)** - Designed Algorithm for I-Detox, Training dataset, implementation of selected classifier algorithm, sending report through email, implementation of parental control.
2. **Divesh Chodda (14)** - Analysis, comparison of different classifier algorithms and choosing one with best accuracy.
3. **Athurva Sawant (63)** - Developing python script with the use of collab to train the dataset with the classifying algorithm.
4. **Muskan Chhabria (13)** - Documentation, analysis of research papers.

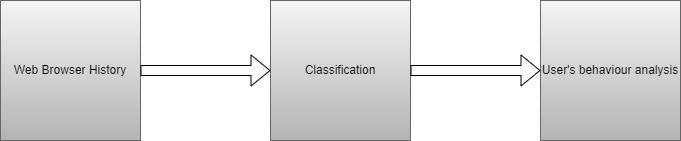
## 

## Proposed System

**3.1 Introduction**

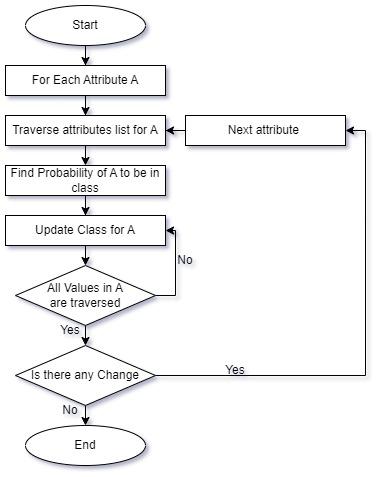
In the IT sector, addictive digital experiences are more prevalent, and the descriptor "addictive" is currently regarded as one of the highest praises given to an artifact. Research that has already been done has looked at what constitutes addiction, what tactics and strategies are used by tech businesses that ultimately result in addiction, and how cultural and societal factors have influenced this phenomena. It is currently believed that activities like binge-watching movies, playing online games, and posting on social networks can be addictive from a health perspective. Technology addiction is now a recognised mental illness. To train a model and assess its correctness, we require a training set and a test set for every machine learning technique. Since we already have text from several websites, we just need to classify it according to the keywords and then apply the results to the data set for the model. Because one never knows the level of their addiction in the age of digitization, "I-detox" can be employed by people of all ages. Our system provides the user with a clear understanding of the number of websites being viewed, the sort of information being browsed, the category it falls under, and the URL.

**3.2 Architectural Framework / Conceptual Design**

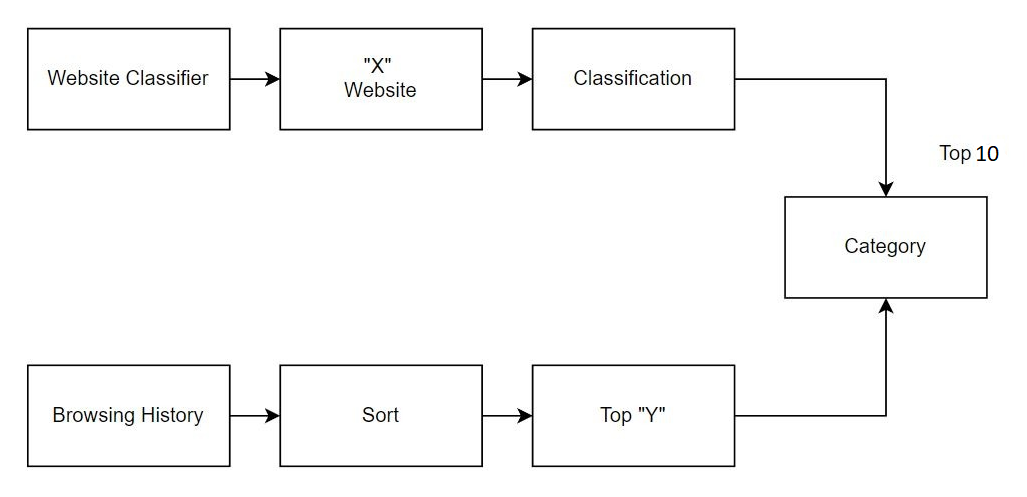
****

3.2.1 Block Diagram

**3.3 Algorithm and Process Design**

****

3.3.1 Flowchart

****

3.3.2 Modular Diagram

**3.4 Methodology Applied**

Artificial Intelligence and Machine learning-based approaches for the prediction of addiction and to classify the websites into categories. Based upon the data extracted from the user’s browsing history which includes the websites, the types of websites, the number of visits or we can say the view count and categories the model would test the data and give the user the exact idea of his browsing pattern. The solution will focus on both the categories and the time-stamp. For the categories, the solution will take into consideration the URLs and the bag of words from the websites. “I-detox” would thus indicate to the user if he requires any help to stop himself from being addicted to any harmful website.

“I-detox” basically operates on the classification of websites. This website classification is based on the categorization of the websites. This categorization takes place on the content of the website, or the bag of words.If theURL is not accessible then we divide the URL into its component strings, create an array of that string and on the basis of this array we classify that particular website.

The metadata of the website is extracted from the website using request and beautifulsoup (bs4) library used for scraping. “Request” will send the request to the website to get the HTML page ; this HTML page will then be used by the bs4 library to extract the metadata from the website. The metadata contains a tag called description, on the basis of the content in the description tag the website is categorized.

Whilst we browse through the internet browsers such as Google Chrome, Mozilla Firefox or Microsoft Edge; the user’s browser history is stored in the root directory of the browser. This generally exists in the Local Disk of the system**(C:\Users\<username>\AppData\Local\Google\Chrome\User Data\Default\History)**. Through this directory one can obtain the user history.

This browsing history contains three tables namely urls, visit\_source and visits. In the urls table, there is a column called visit count for each website. We take this visit count and then we sort the visit count, from that we get the top “y” (where y is the number of specified websites). We take these top y websites and assign the url to these websites using the model which is trained above.

**3.5 Hardware & Software Specifications**

**Software Requirements:**

● **Language**: Python

● **IDE**: Pycharm IDLE

●**Optional Cloud Resources:**

Pythonanywhere Connect

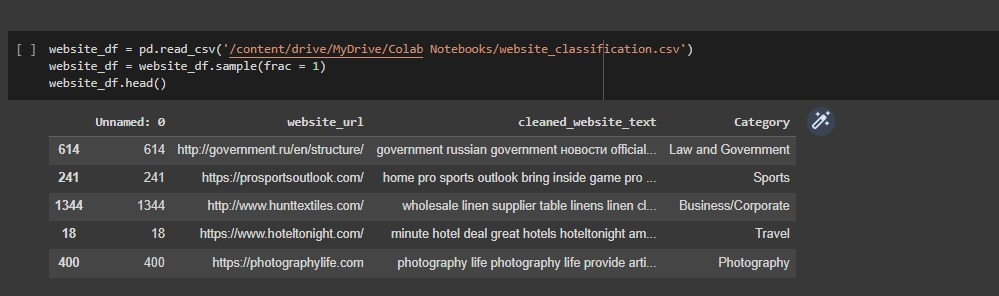
● **Deployment Framework**: Flask

● **To build the application**: Flask, Django

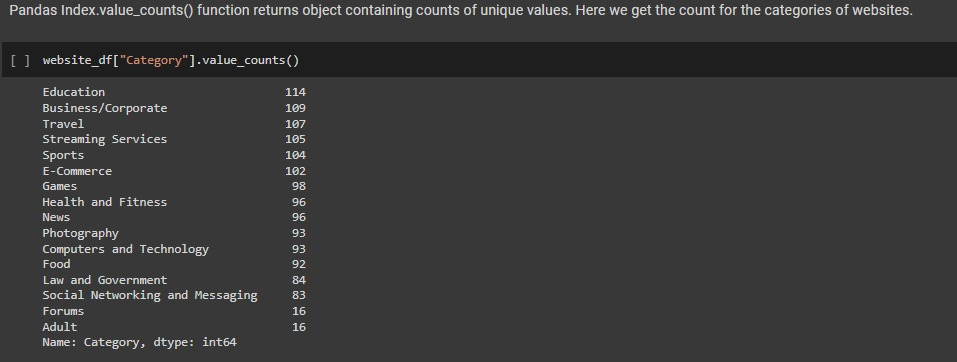
**Hardware Requirements:**

* 2 Gb Ram
* Intel Pentium Gold Processor and above / Ryzen Athlon and above
* 512 Mb Physical Storage and above

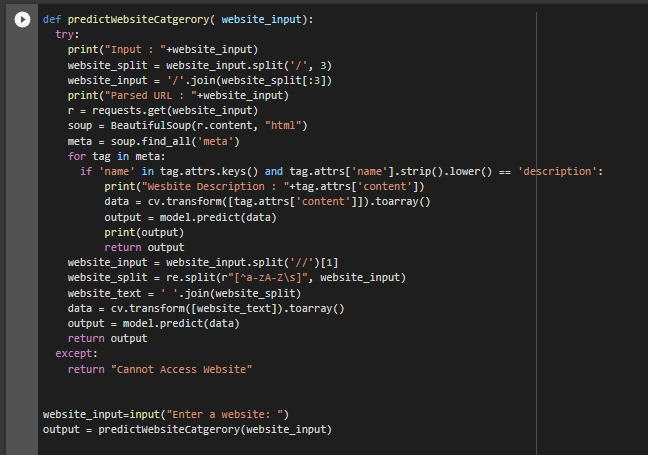
**3.6 Experiment and Results for Validation and Verification**

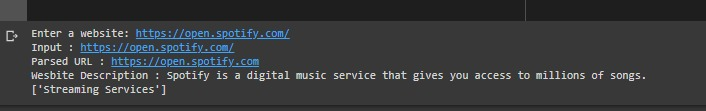
****

**3.6.1: STATISTICAL REPRESENTATION OF DATA(CSV) FOR WEBSITE CLASSIFICATION**

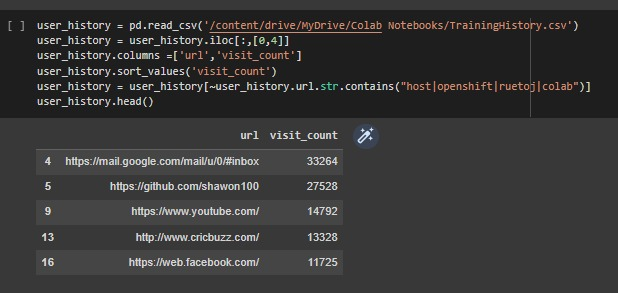
****

**3.6.2 : Categories to be predicted**

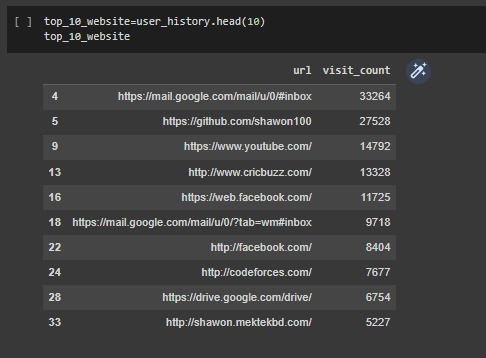
****

****

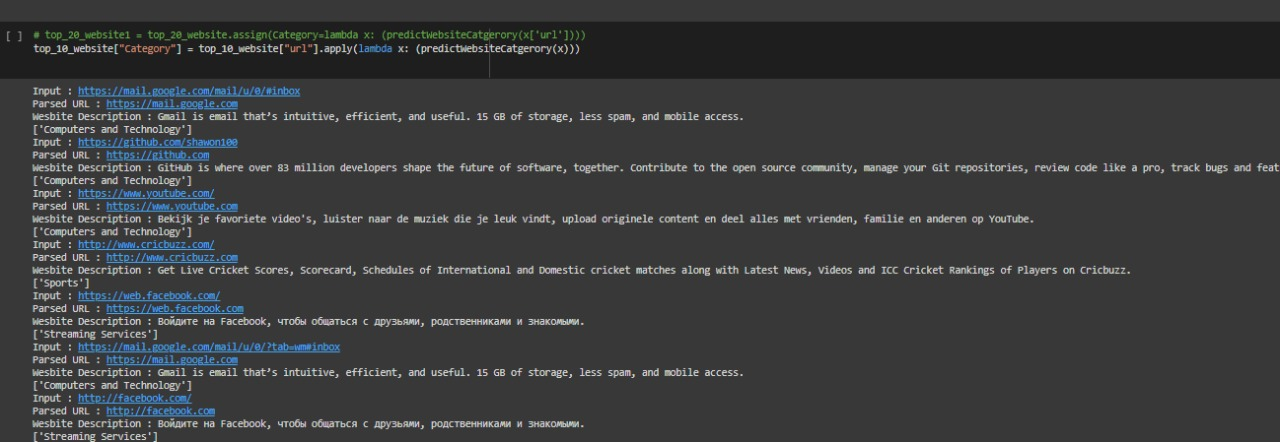
**3.6.3: Classification result**

****

**3.6.4: Browsing history dataset**

****

**3.6.5: Top 10 websites visited**

****

**3.6.6: Assigning categories to top 10 websites**

****

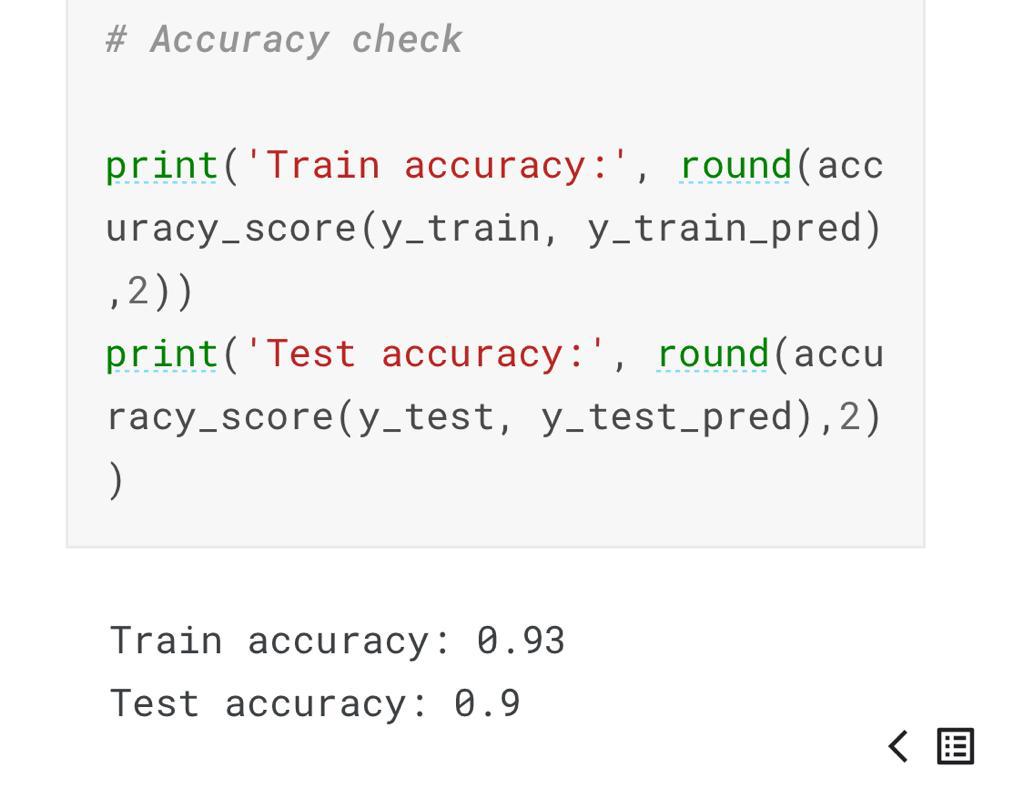
**3.6.7: Classification Output**

**Table

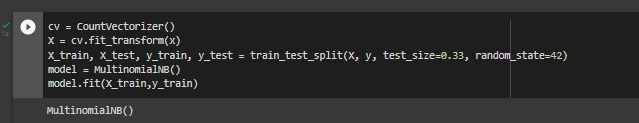
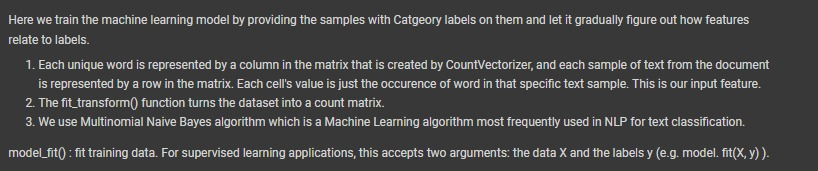
Description automatically generated**

**3.6.8: Final Report as Output**

**3.7 Result Analysis and Discussion**

****

**3.7.1 Accuracy of the NB model**

****

**3.7.2 Multinominal model**

**3.7.4 Why Multinomial Naive Bayes over Naive Bayes algorithm?**

Multinomial Naive Bayes is used when frequency is a feature; this algorithm helps to ignore the non-occurrence of the features.

When categorizing texts based on a statistical examination of their contents, the multinomial naive Bayes algorithm is frequently used. It offers an alternative to "heavy" AI-based semantic analysis and significantly streamlines the classification of textual material. By calculating the likelihood that a document belongs to the class of other papers with the same subject, classification attempts to categorize text fragments (i.e., documents) into different classes.

Each document contains a variety of words (i.e., terms) that help the reader comprehend the information it contains. A class is a tag used to identify one or more papers that are about the same subject. By using statistical analysis to evaluate the idea that a document's phrases have already appeared in other papers from a specific class, documents are assigned to one of the existing classes.

**3.8 Conclusion and Future work.**

Internet and virtual social addiction is second among all types of addiction, with an exaggerated incidence of 16.9% per 100 people and a majority of users in the 16–50 age range who identify as socially uncomfortable and insecure, or what is known scientifically as "nomophobia." In this project, we propose an artificial intelligence-based system that can help the recipient analyze for himself his use of the internet. This system can work with people of all age groups with parents using it to keep a check on their ward’s internet life to the people who find themselves lost in this huge digital world. This system would help the user have a personalized analysis and a fact check about his presence on the internet, the activities he performs, the content he streams to the games he plays because of the availability of view-count recording details of every thin activity. Thus, the question in person with the help of our system can reconcile his use of the internet thus helping him curb and storm out his way through these addiction demons. Our future work revolves around the idea to introduce parental work and introduce a module which would focus and target the mental health of the user based on his/her activity.

**REFERENCES**

1. Wakaki, T., Itakura, H., Tamura, M., Motoda, H., & Washio, T. (2006). A study on rough set-aided feature selection for automatic web-page classification. Web Intelligence and Agent Systems: An International Journal, 4, 431–441.
2. Internet Addiction Facts and Statistics By The Recovery Village | Editor Megan Hull.
3. <https://www.helpguide.org/articles/addictions/smartphone-addiction.htm> (Signs & Symptoms)
4. Kumar M, Mondal A. A study on Internet addiction and its relation to psychopathology and self-esteem among college students. Ind Psychiatry J. 2018 Jan-Jun;27(1):61-66. doi: 10.4103/ipj.ipj\_61\_17. PMID: 30416293; PMCID: PMC6198588.
5. Internet Addiction: Definition, Assessment, Epidemiology and??Clinical Management by Donald Black.
6. <https://comparecamp.com/internet-addiction-statistics/> (STATISTICS: 2020/2021 DATA, FACTS & PREDICTIONS)
7. <https://pypi.org/project/browserhistory/> (This website helps us to understand how we can collect browser history from a pc and convert it into a proper csv format.)
8. [Website Classification using Multinomial Naive Bayes | Kaggle](https://www.kaggle.com/datasets/hetulmehta/website-classification) (Dataset).
9. <https://www.kaggle.com/datasets/shawon10/browser-history>