IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 4, April 2024

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.423

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304278 |

Identifying of Fake Profiles across Online Social Networks using Neural Network

Mr S. Ismail Saheb[1], D. Sumanth Reddy [2], S. Sultan Bee [3], D. Sayara Banu[4], K. Sri Chaithanya [5], C. Supriya [6]

UG Student, Dept. of E.C.E., GATES Institute of Technology, Gooty, Anantapur, Andhra Pradesh, India[2,3,4,5,6]

Assistant Professor, Dept. of E.C.E., GATES Institute of Technology, Gooty, Anantapur, Andhra Pradesh, India [1]

ABSTRACT: These days, there is a noticeable increase in applied sciences. Mobile phones are becoming smarter. Technology is associated with online social networks, which have emerged as a part of everyone's life in terms of making new friends and retaining friends, as well as making their hobbies easier. However, the increase in online networking causes various problems, such as fabricating their profile. — In this research, we employ computing device learning, specifically a synthetic neural community, to identify whether or not a Facebook buddy request is legitimate. We also outline the training and libraries that will be used. We also discuss the sigmoid feature and how weights are determined and used. Finally, we reflect on consideration on the parameters of the social community web page which are utmost necessary in the furnished solution.

KEYWORDS: ESP 32; DS18B20; pH Sensor; Water level sensor; LCD; Relay, AC pump

I. INTRODUCTION

Long range interpersonal communication has end up a notable diversion inside the web as of now, drawing in countless clients, burning through billions of minutes on such administrations. Online Social organization (OSN) administrations assortment from social cooperations based stages like Facebook or My Space, to understanding spread driven stages suggestive of twitter or Google Buzz, to social association trademark brought to introduce frameworks like Flicker. The contrary hand, improving security concerns and safeguarding the OSN privateness actually connote a most significant bottleneck and saw mission. While utilizing social organization's (Sn's), exceptional people share stand-out amounts of their private arrangement. Having our singular expertise totally or to some degree uncovered to the overall population, makes us amazing focuses for exceptional kinds of attacks, the most exceedingly terrible of which could be ID burglary. Data fraud happens when any singular uses character's skill for a private achieve or reason. During the prior years, online recognizable proof burglary has been an essential issue thinking of it as impacted huge number of individuals around the world. Casualties of ID robbery might experience remarkable sorts of punishments; for delineation, they would lose time/cash, get dispatched to reformatory, get their public picture destroyed, or have their associations with partners and friends and family harmed. As of now, by far most of SN's does no longer checks common users‟ obligations and has entirely vulnerable privateness and security approaches. Truth be told, most SN's applications default their settings to negligible privateness; and subsequently, SN's turned into a best stage for misrepresentation and misuse. Person to person communication contributions have worked with data fraud and Impersonation assaults for genuine comparable to innocent assailants. To compound the situation, clients are expected to outfit right comprehension to set up a record in Social Networking sites. Simple observing of what clients share on-line would prompt devastating misfortunes, not to mention, assuming such bills had been hacked. Profile data in web-based organizations will likewise be static or dynamic. The subtleties which can be provided with the guide of the individual on the hour of profile creation is known as static information, the spot as the important part that are described with the guide of the framework inside the organization is called dynamic information. Static information incorporates segment components of an individual and his/her advantages and dynamic information remembers individual runtime propensities and region for the organization. By far most of momentum research relies upon static and dynamic information. Anyway, this isn't applicable to heaps of the informal organizations, where handiest some of static profiles are seen and dynamic profiles as a rule are not clear to the individual organization. In excess of a couple of methods have been proposed by exceptional specialist to understand the phony characters and malignant substance material in web-based informal communities. Each interaction had its own merits and negative marks. The issues

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304278 |

including long range interpersonal communication like security, web-based tormenting, abuse, and savaging and numerous others. Are large numbers of the cases used by misleading profiles on long range interpersonal communication locales. Misleading profiles are the profiles which are not explicit i.e. They're the profiles of people with misleading qualifications. The misleading Facebook profiles all the more generally are enjoyed pernicious and unwanted exercises, bringing on some issues to the social local area clients. People make counterfeit profiles for social designing, online pantomime to stigmatize a man or lady, advancing and lobbying for a person or a horde of people. Facebook has its own security framework to monitor individual certifications from spamming, phishing, etc. Furthermore, the equivalent is regularly called Facebook Immune framework (FIS). The FIS has now not been prepared to notice counterfeit profiles made on Facebook through clients to a greater degree. Online social media is the place each person has a outlook then be able to keep connecting their relations, transfer their updates, join with the people having same likes. Online Social Networks makes use of front end technologies, which permits permanency accounts in accordance with to know each other. Facebook, Twitter are developing along with humans to maintain consultation together with all others. The online accounts welcome people including identical hobbies collectively who makes users easier after perform current friends. Gaming and entertaining web sites which have extra followers unintentionally that means more fan base and supreme ratings.

II. LITERATURE SURVEY

Various fake record recognition methodologies depend on the investigation of individual interpersonal organization profiles, with the point of distinguishing the qualities or a combination thereof that help in recognizing the legitimate and the fake records. In particular, various features are extracted from the profiles and posts, and after that Machine learning algorithms are used so as to construct a classifier equipped for recognizing fake records. For instance, Nazir et al. (2010) [1] describes recognizing and describing phantom profiles in online social gaming applications. The article analyses a Facebook application, the online game “Fighters club”, known to provide incentives and gaming advantage to those users who invite their peers into the game. The authors contend that by giving such impetuses the game motivates its players to make fake profiles. By presenting those fake profiles into the game, the user would increase a motivating force of an incentive for him/herself. Adikari and Dutta (2014) [2] depict recognizable proof of fake profiles on LinkedIn. The paper demonstrates that fake profiles can be recognized with 84% exactness and 2.44% false negative, utilizing constrained profile information as input. Techniques, for example, neural networks, SVMs, and Principal component analysis are applied. Among others, highlights, for example, the number of languages spoken, training, abilities, suggestions, interests, and awards are utilized. Qualities of profiles, known to be fake, posted on uncommon sites are utilized as a ground truth. Chu et al. (2010) [3] go for separating Twitter accounts operated by humans, bots, or cyborgs (i.e., bots and people working in concert). As a part of the detection problem formulation, the Identification of spamming records is acknowledged with the assistance of an Orthogonal Sparse Bigram (OSB) text classifier that uses pairs of words as features. Stringhini et al. (2013) [4] analyze Twitter supporter markets. They describe the qualities of Twitter devotee advertises and group the clients of the business sectors. The authors argue that there are two major kinds of accounts who pursue the “client”: fake accounts(“sybils”), and compromised accounts, proprietors of which don’t presume that their follower’s rundown is expanding. Clients of adherent markets might be famous people or legislators, meaning to give the appearance of having a bigger fan base, or might be cybercriminals, going for making their record look progressively authentic, so they can rapidly spread malware what’s more, spam. Thomas et al. (2013) [5] examine black market accounts utilized for distributing Twitter spam. De Cristofaro et al. (2014) [6] investigate Facebook like cultivates by conveying honeypot pages. Viswanath et al. (2014) [7] identify black market Facebook records based on the examination of anomalies in their like behaviour. Farooqi et al. (2015) [6] explore two black hat online commercial centres, SEO Clerks and My Cheap Jobs. Fayazi et al. (2015) think about manipulation in the online review. [5] Qiang Cao, Michael Sicilianos, Xiaowei Yang, and Tiago Pregueiro. 2012. Aiding the detection of fake accounts in large scale social online services. In Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation (NSDI'12). USENIX Association, Berkeley, CA, USA, 15-15. Users increasingly rely on the trustworthiness of the information exposed on Online Social Networks (OSNs). In addition, OSN providers base their business models on the marketability of this information. However, OSNs suffer from abuse in the form of the creation of fake accounts, which do not correspond to real humans. Fakes can introduce spam, manipulate online rating, or exploit knowledge extracted from the network. OSN operators currently expend significant resources to detect, manually verify, and shut down fake accounts. Tuenti, the largest OSN in Spain, dedicates 14 full-time employees in that task alone, incurring a significant monetary cost. Such a task has yet to be successfully automated because of the difficulty in reliably capturing the diverse behavior of fake and real OSN profiles. We introduce a new tool in the hands of OSN operators, which we call SybilRank. It relies on social graph properties to rank users according to their perceived likelihood of being fake (Sybils). SybilRank is computationally

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304278 |

efficient and can scale to graphs with hundreds of millions of nodes, as demonstrated by our Hadoop prototype. We deployed SybilRank in Tuenti's operation center. We found that ~90% of the 200K accounts that SybilRank designated as most likely to be fake, actually warranted suspension. On the other hand, with Tuenti's current user-report-based approach only ~5% of the inspected accounts are indeed fake. Large scale social online services place immense attention to the experience of their user base, and the marketability of their user profiles and the social graph. In this context, they face a significant challenge by the existence and continuous creation of fake user accounts, which dilutes the advertising value of their network and annoys legitimate users. To this end, we have proposed SybilRank, an effective and efficient fake account inference scheme, which allows OSNs to rank accounts according to their perceived likelihood of being fake. Therefore, this work represents a significant step towards practical Sybil defense: it enables an OSN to focus its expensive manual inspection efforts, as well as to correctly target existing countermeasures, such as CAPTCHAs. The surge in popularity of online social networking (OSN) services such as Facebook, Twitter, Digg, LinkedIn, Google+, and Tuenti has been accompanied by an increased interest in attacking and manipulating them. Due to their open nature, they are particularly vulnerable to the Sybil attack [26], under which a malicious user can create multiple fake OSN accounts. The problem. It has been reported that 1.5 million fake or compromised Facebook accounts were on sale during February 2010 [7]. Fake (Sybil) OSN accounts can be used for various purposes [6, 7, 9]. For instance, they enable spammers to abuse an OSN's messaging system to post spam [28, 51], or waste an OSN advertising † Part of Q. Cao's work was conducted while interning with Telefonica Research. ‡M. Sirivianos is currently with the Cyprus University of Technology. customer's resources by making him pay for online ad clicks or impressions from or to fake profiles. Fake accounts can also be used to acquire users' private contact lists [17]. Sybils can use the "+1" button to manipulate Google search results [11] or to pollute location crowdsourcing results [8]. Furthermore, fake accounts can be used to access personal user information [27] and perform large-scale crawls over social graphs The challenge. Due to the multitude of the reasons behind their creation (§2.1), real OSN Sybils manifest numerous and diverse profile features and activity patterns. Thus, automated Sybil detection (e.g., Machine Learning-based) does not yield the desirable accuracy (§2.2). As a result, adversaries can cheaply create fake accounts that are able to evade detection [19]. At the same time, although the research community has extensively discussed social-graph-based Sybil defences [23, 52–54, 57, 58], there is little evidence of wide industrial adoption due to their shortcomings in terms of effectiveness and efficiency Instead, OSNs employ time consuming manual account verification, driven by user reports on abusive accounts. However, only a small fraction of the inspected accounts are indeed fake, signifying inefficient use of human labour.

III. PROPOSED SYSTEM

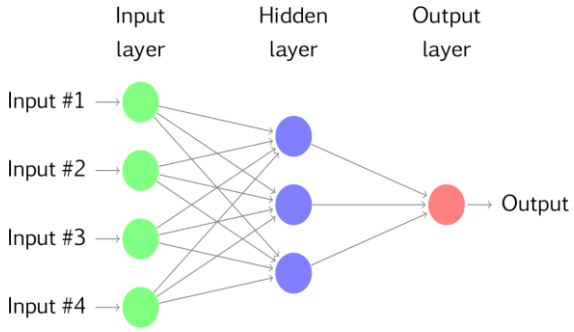
In Our Solution, We Use Machine Learning, Namely An Artificial Neural Network To Determine What Are The Chances That A Friend Request Is Authentic Or Not. We Utilize Microsoft Excel To Store Old And New Fake Data Profiles. The Algorithm Then Stores The Data In A Data Frame. This Collection Of Data Will Be Divided Into A Training Set And A Testing Set. We Would Need A Data Set From The Social Media Sites To Train Our Model. For The Training Set, The Features That We Use To Determine A Fake Profile Are Account Age, Gender, User Age, Link In The Description, Number Of Messages Sent Out, Number Of Friend Requests Sent Out, Entered Location, Location By Ip, Fake Or Not. Each Of These Parameters Is Tested And Assigned A Value. For Example, For The Gender Parameter If The Profile Can Be Determined To Be A Female Or Male A Value Of (1) Is Assigned To The Training Set For Gender. The Same Process Is Applied To Other Parameters. We Also Use The Country Of Origin As A Factor We Then Determine The Number Of Messages Sent Out Parameter By Dividing The Number Of Messages Sent By The Age Of The Account. We Then Determine The Number Of Friend Requests Sent Out Parameter By Dividing The Number Of Friend Computing And Used Primarily For Multi-Dimensional Matrix Multiplication As We Are Dealing With A Large Number Of Numbers That Are Very Dependent On Each Other. In This Paper Using Artificial Neural Networks We Are Identifying Whether Given Account Details Are From Genuine Or Fake Users. Ann Algorithm Will Be Trained With All Previous Users Fake And Genuine Account Data And Then Whenever We Gave New Test Data Then That Ann Train Model Will Be Applied On New Test Data To Identify Whether Given New Account Details Are From Genuine Or Fake Users. Online Social Networks Such As Facebook Or Twitter Contains Users Details And Some Malicious Users Will Hack Social Network Database To Steal Or Breach Users Information, To Protect Users Data We Are Using Ann Algorithm. By Analysing This Features Facebook Will Mark Whether User Profile Is Fake Or Genuine. This Facebook Profile Data We Downloaded From Facebook Website And Using This Data To Train Ann Model. Below Are Some Values From Profile Dataset. Account Age, Gender, User Age, Link_Desc, Status Count, Friend Count, Location, Location, Status 10, 1, 22, 0, 1073, 237, 0, 0, 0 10, 0, 33, 0, 127, 152, 0, 0, 0 10, 1, 46, 0, 1601, 405, 0, 0, 0 10, 0, 25, 0, 704, 380, 0, 0, 0 7, 1, 34, 1, 64, 721, 1, 1, 1 7, 1, 30, 1, 69, 587, 1, 1, 1 7, 1, 36, 1, 61, 782, 1, 1, 1 7, 1, 52, 1, 96, 827, 1, 1, 1 In Above Dataset All Bold Names Are The Dataset Column Names And All Integer

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304278 |

Values Are The Dataset Values. As Ann Will Not Take String Value So We Convert Gender Values To 0 Or 1, If Male Value Is 1 And If Female Value Is 0. In Above Dataset Last Column Give Us Information Of Fake Or Genuine Account If Last Column Contains Value 0 Then Account Is Genuine Otherwise Fake. All Fake Account Will Have Less Number Of Posts As Their Main Intention Is To Send Friend Requests Not Posts, So By Analysing This Features Facebook Mark That Record With Value 1 Which Means It's A Fake Account. We Are Using Above Dataset To Train Ann Model And This Dataset Saved Inside Code 'Dataset' Folder. After Building Train Model We Input Test Data With Account Details And Ann Will Give Result As Fake Or Genuine. Below Are Some Values From Test Data Account_Age, Gender, User_Age, Link_Desc, Status_Count, Friend_Count, Location, Location_Ip 10, 1, 44, 0, 280, 1273, 0, 0 10, 0, 54, 0, 5237, 241, 0, 0 7, 0, 42, 1, 57, 631, 1, 1 7, 1, 56, 1, 66, 623, 1, 1 In Above Test Data Status Column And Its Value Is Not There And Ann Will Predict Status And Give Us Result Whether Above Test Data Is Fake Or Genuine. In Output We Can See Result Of Above Test Data As Genuine Or Fake. Ann Algorithms Details To Demonstrate How To Build A Ann Neural Network Based Image Classifier, We Shall Build A 6 Layer Neural Network That Will Identify And Separate One Image From Other. This Network That We Shall Build Is A Very Small Network That We Can Run On A Cpu As Well. Traditional Neural Networks That Are Very Good At Doing Image Classification Have Many More Parameters And Take A Lot Of Time If Trained On Normal Cpu. However, Our Objective Is To Show How To Build A Real-World Convolutional Neural Network Using Tensorflow. Neural Networks Are Essentially Mathematical Models To Solve An Optimization Problem. They Are Made Of Neurons, The Basic Computation Unit Of Neural Networks. A Neuron Takes An Input (Say X), Do Some Computation On It (Say: Multiply It With A Variable W And Adds Another Variable B) To Produce A Value (Say; Z= Wx + B). This Value Is Passed To A Non-Linear Function Called Activation Function (F) To Produce The Final Output (Activation) Of A Neuron. There Are Many Kinds Of Activation Functions. One Of The Popular Activation Function Is Sigmoid. The Neuron Which Uses Sigmoid Function As An Activation Function Will Be Called Sigmoid Neuron. Depending On The Activation Functions, Neurons Are Named And There Are Many Kinds Of Them Like Relu, Tanh.



IV. IMPLEMENTATION

Guido Van Rossum designed Python, an interpreted, high-level, general-purpose programming language that was initially released in 1991. The Python language has the following features:

• Simple to Understand and Apply Python is a simple language to learn and use. It's a high- level programming language that's helpful to developers.

• Expressive Language Python is more expressive than other languages, making it more intelligible and readable. • Language Interpretation Python is an interpreted language, which means that the interpreter runs the code line by line. This makes debugging simple, making it suitable for novices.

• Language that is cross-platform Python can operate on a variety of platforms, including Windows, Linux, UNIX, and Macintosh, among others. As a result, we may conclude that Python is a portable language.

• Open Source and Free Software Python is a free programming language that can be downloaded from a secure internet address. It's also possible to get the source code. As a result, there is an open supply.

• Object-Oriented Programming Python aids the development of object-oriented languages and concepts such a s classes and gadgets.

• Adaptable It means that other languages, such as C/C++, may be used to put the code together, and as a result, it can be utilized in our Python code.

NUMPY: NumPy is a Python library that includes support for enormous, multi-dimensional arrays and matrices, as well as a large set of high-level mathematical functions that may be applied to those arrays. Numeric, the forerunner to NumPy, was designed by Jim Hugunin with help from a number of other people. Travis Oliphant built NumPy in 2005

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304278 |

by heavily modifying Numeric and combining features from the competitor Numarray. NumPy is a large open-source software project with numerous contributors. The Python programming language was not initially intended for numerical computing, but it quickly caught the attention of the clinical and engineering communities, prompting the formation of a special interest group known as matrix-sig in 1995 with the goal of defining an array computing bundle. Guido van Rossum, a Python clothier and maintainer, was one of its contributors, adding changes to Python's grammar (especially the indexing syntax) to make array computation easier. Jim Fulton completes a matrix package, which is later modified by Jim Hugunin to create Numeric, also known as Numerical Python extensions or NumPy. Hugunin, a PhD student at MIT, joined the Corporation for National Research Initiatives (CNRI) to work on Python in 1997, leaving the maintainer position to Paul Dubois of Lawrence Livermore National Laboratory (LLNL). David Ascher, Konrad Hinsen, and Travis Oliphant were among the early members. NumPy is a non-optimizing byte code interpreter that targets the Python Python reference implementation. Algorithms created for this version of Python are frequently much slower than their compiled counterparts. NumPy tackles the slowness issue in part by providing multidimensional arrays and capabilities and operators that work appropriately on arrays, which necessitates rewriting some code, generally inner loops, in order to use NumPy. NumPy arrays are used to store and perform information in the Python bindings of the widely used computer vision library OpenCV. Because images with more than one channel are really stored as 3-dimensional arrays, indexing, slicing, and protecting with separate arrays are all highly eco-friendly ways to access specific pixels in a picture The NumPy array, which is used as a standard statistical structure in OpenCV for pictures, extracted feature factors, kernel filtering, and other tasks, greatly simplifies the development process and debugging. 6.2.2 PANDAS: Pandas is an open-source, BSD-certified library for the Python programming language that provides high-overall performance, easy-to-use statistics systems, and data analysis tools. A panda is a Python module that provides quick, flexible and expressive facts structures for working with "relational" or "labelled" data in a clean and understandable manner. Its goal is to become the most important high-level building element for undertaking realistic; real-world international records evaluations in Python. It also has the larger goal of being the most powerful and versatile open source data analysis/manipulation device available in any language. It is already well on its way to achieving this goal. A panda is well-suited to a wide range of statistical applications:

• Tabular statistics containing columns of varying types, such as those seen in a SQL database or an Excel spreadsheet
• Time collecting data that are sorted and unordered (but not necessarily at the same frequency).
• Row and column labels for arbitrary matrix information (homogeneously typed or heterogeneous) • Any observational/statistical statistics set of any kind. To be placed into a panda's facts form, the data does not need to be tagged in any way. Pandas' two basic statistics systems, Series (1-dimensional) and DataFrame (2- dimensional), address the vast majority of common use cases in finance, information, social science, and a wide range of engineering disciplines. A panda is built on top of NumPy and is intended to work in conjunction with a variety of different third-party libraries in scientific computing.

MATPLOT LIB: Matplotlib is a Python 2D plotting toolkit that generates book-quality figures in a variety of hardcopy codecs and interactive contexts. Matplotlib is a Python library that may be used in scripts, the Python and IPython shells, the Jupiter notebook, web applications servers, and four graphical user interface toolkits. Matplotlib aims to make both smooth and difficult tasks feasible. With just a few lines of code, you can create graphs, histograms, electrical spectra, bar charts, error charts, scatter plots, and more. See the pattern plots and thumbnail galleries for samples. The pyplot package, when used with IPython, provides a MATLAB-like interface for convenient plotting. Through an object-oriented interface or a set of methods common to MATLAB users, you have complete control over line styles, font houses, axis houses, and so on for the electricity consumer. 6.2.4 SEABORN: Seaborn is a data visualisation package for Python that is mostly based on matplotlib. It provides a high-level interface for creating visually beautiful and useful statistics graphs. Seaborn is a Python module for creating statistical visuals. It's based on matplotlib, and it's tightly integrated with panda's data systems. The goal of Seaborn is to make visualisation a major aspect of information exploration and understanding. Its dataset-oriented charting features operate on facts frames and arrays holding whole datasets, doing the necessary semantic mapping and statistical aggregation internally to provide relevant charts.

Scikit-learn: To put it another way, sci-kit study is a free software system studying library for the Python computer language. It includes support vector machines, random forests, gradient boosting, k-method, and DBSCAN, among other categorization, regression, and clustering algorithms, and is designed to work with the Python numerical and clinical libraries NumPy and SciPy. Scikit- research was created in 2007 as a Google summers of code initiative by David Cornopean. Matthieu Brucher afterwards joined the challenge and began using it in his thesis paintings. INRIA was given consideration in 2010, and the initial public release (v0.1 beta) became released in late January 2010. The

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304278 |

project presently has over 30 active participants and has received financial support from INRIA, Google, Tiny Clues, and the Python Software Foundation. In most cases, solving a problem involves looking at a collection of n samples of facts and then attempting to predict attributes of unknown facts. If a sample has more than one kind, such as a multi-dimensional entry (also known as multivariate information), it is said to have many characteristics or capabilities.

V. SOFTWARE AND HARDWARE REQUIREMENTS

HARDWARE REQUIREMENTS:
• System: Pentium IV 2.4 GHz
• Hard Disk: 40 GB
• Floppy Drive:1.44 Mb
• Monitor :15 VGA Colour
• Mouse: Logitech
• Ram: 512 Mb
SOFTWARE REQUIREMENTS:
- Web technology: HTML, CSS, Java scripts
- Python Web Server: AJAX
- Programming Language: Django
- Algorithm: Python
- Database: MYSQL

VI. RESUL T

This work uses Artificial Neural Networks to identify whether given social network account details are from genuine or fake users. ANN algorithm will be trained with all previous users fake and genuine account dataset and then whenever we gave new test data then that ANN train model will be applied on new test data to identify whether given new account details are from genuine or fake users. Online social networks such as Facebook or Twitter contains user's details and some malicious users will hack social network database to steal or breach users information, To protect users data we are using ANN Algorithm. To train ANN algorithm we are using below details from social networks Account_Age, Gender, User_Age, Link_Desc, Status_Count, Friend_Count, Location, Location_IP, Status All fake users main intention is to send friend request to normal users to hack their machine or to steal their data and never they will have many number of posts or have many following friends and their account age also will have less number of years. By analysing this features Facebook will mark whether user profile is fake or genuine. This Facebook profile data we downloaded from Facebook website and using this data to train ANN model. Below are some values from profile dataset. Account_Age, Gender, User_Age, Link_Desc, Status_Count, Friend_Count, Location, Location_IP, Status

10, 1, 22, 0, 1073, 237, 0, 0, 0
10, 0, 33, 0, 127, 152, 0, 0, 0
10, 1, 46, 0, 1601, 405, 0, 0, 0
10, 0, 25, 0, 704, 380, 0, 0, 0
7, 1, 34, 1, 64, 721, 1, 1, 1
7, 1, 30, 1, 69, 587, 1, 1, 1
7, 1, 36, 1, 61, 782, 1, 1, 1
7, 1, 52, 1, 96, 827, 1, 1, 1

In above dataset all bold names are the dataset column names and all integer values are the dataset values. As ANN will not take string value so we convert gender values to 0 or 1, if male value is 1 and if female value is 0. In above dataset last column give us information of fake or genuine account if last column contains value 0 then account is genuine otherwise fake. All fake account will have less number of posts as their main intention is to send friend requests not posts, so by analysing this features Facebook mark that record with value 1 which means it's a fake account. We are using above dataset to train ANN model and this dataset saved inside code 'dataset' folder. After building train model we input test data with account details and ANN will give result as fake or genuine. Below are some values from test data Account_Age, Gender, User_Age, Link_Desc, Status_Count, Friend_Count, Location, Location_IP

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304278 |

10, 1, 44, 0, 280, 1273, 0, 0
10, 0, 54, 0, 5237, 241, 0, 0
7, 0, 42, 1, 57, 631, 1, 1
7, 1, 56, 1, 66, 623, 1, 1

In above test data STATUS column and its value is not there and ANN will predict status and give us result whether above test data is fake or genuine. In output we can see result of above test data as genuine or fake. Module Details Upload Social Network Profiles Dataset: Using this module we will upload dataset to applicationPre-process Dataset: Using this module we will apply processing technique such as removing missing values and then split dataset into train and test where application use 80% dataset to train ANN and 20% dataset to test ANN prediction accuracy Run ANN Algorithm: Using this module we will train ANN algorithm with train and test data and then train model will be generated and we can use this train model to predict fake accounts from new dataset. ANN Accuracy & Loss Graph: To train ANN model we are taking 200 epoch/iterations and then in graph we will plot accuracy/loss performance of ANN at each epoch/iteration. Predict Fake/Genuine Profile using ANN: using this module we will upload new test data and then apply ANN train model to predict whether test data is genuine or fake.

To run project double click on 'run.bat' file to get below screen

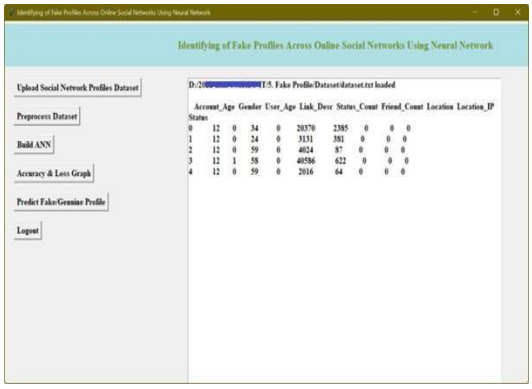


Fig:6.1: Upload Social Network Profiles Dataset

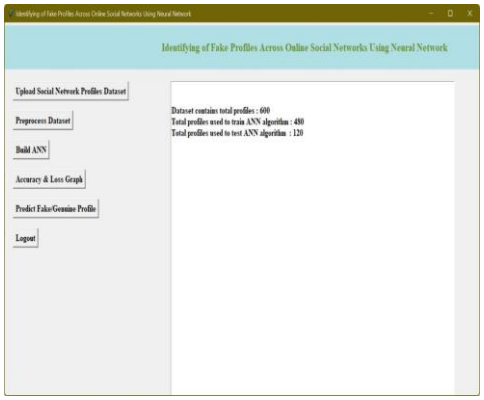


Fig 6.2: Dataset.txt

Fig 6.3: Preprocess Dataset

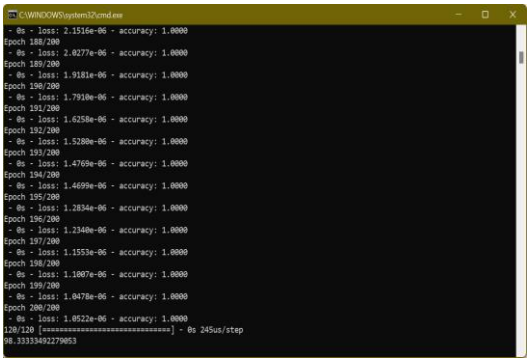


Fig 6.4: Preprocess Database

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304278 |

Fig 6.6: ANN model generated and its prediction

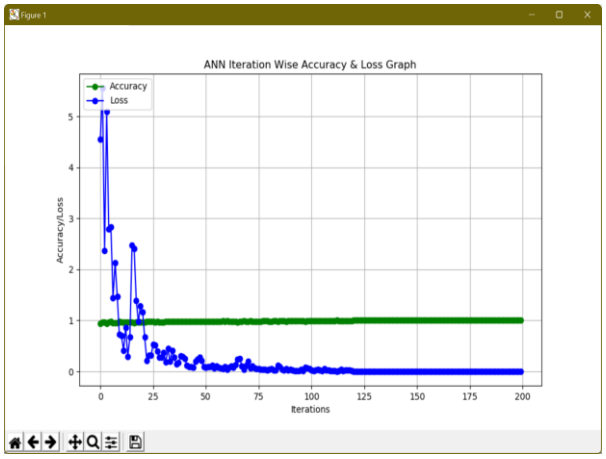


Fig 6.7: ANN Accuracy & Loss Graph

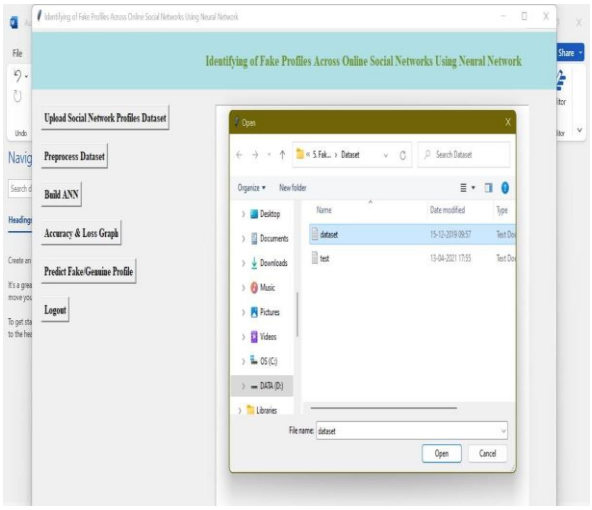


Fig 6.8: Loding test data

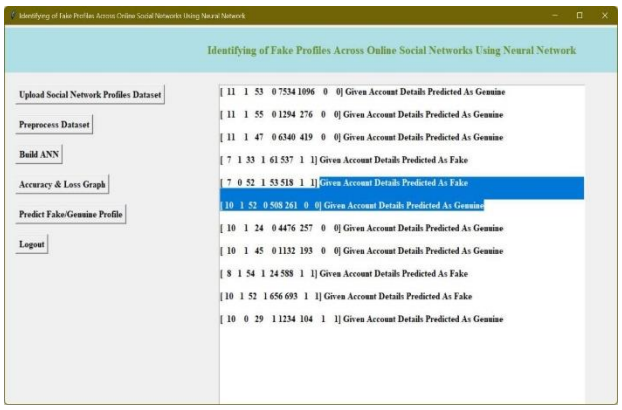


Fig 6.9: ANN Prediction Result

VII. CONCLUSION

In this paper, we use machine learning, namely an artificial neural network to determine what are the chances that a friend request is authentic are or not. Each equation at each neuron (node) is put through a Sigmoid function. We use a training data set by Facebook or other social networks. This would allow the presented deep learning algorithm to learn the patterns of bot behavior by backpropagation, minimizing the final cost function and adjusting each neuron's weight and bias. In this paper, we outline the classes and libraries involved. We also discuss the sigmoid function and how are the weights determined and used. We also consider the parameters of the social network page which are the most important to our solution

VIII. FUTURE WORK

Each input neuron would be a different, previously chosen feature of each profile converted into a numerical value (e.g., gender as a binary number, female 0 and male 1) and if needed, divided by an arbitrary number (e.g., age is always divided by 100) to minimize one feature having more influence on the result than the other. The neurons represent nodes. Each node would be responsible for exactly one decision-making process.

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304278 |

REFERENCES

1. https://www.statista.com/topics/1164/social-networks/https://www.cnbc.com/2018/01/31/f acebook-earnings-q4-2017- arpu.html [3]https://www.cnet.com/news/facebook-breach-affected-50-millionpeople
2. https://www.facebook.com/policy.phpQiang Cao, Michael Sirivianos, Xiaowei Yang, and Tiago Pregueiro. 2012. Aiding the detection of fake accounts in large scale social online services. In Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation (NSDI'12). USENIX Association, Berkeley, CA, USA, 15-15.Akshay J. Sarode and Arun Mishra. 2015. Audit and Analysis of Impostors: An experimental approach to detect fake profile in an online social network. In Proceedings of the Sixth International Conference on Computer and Communication Technology 2015 (ICCCT '15). ACM, New York, NY, USA, 1-8. DOI: https://doi.org/10.1145/2818567.28185 68Devakunchari Ramalingam, Valliyammai Chinnaiah. Fake profile detection techniques in large-scale online social networks: A comprehensive review. Computers & Electrical Engineering,Volume 65, 2018, Pages 165-177, ISSN 0045- 7906,
3. https://doi.org/10.1016/j.compeleceng. 2017.05.020.https://www.enigmasoftware.com/top- 20-countries-the-most-cybercrimepages.cs.wisc.edu/~bolo/shipyard/neur al/local.htmlhttps://stackoverflow.com/questions/40 758562/can-anyone-explainmestandardscaler
4. https://pandas.pydata.org
5. https://www.tutorialspoint.com/python _pandas/index.htm
6. http://www.numpy.org
7. https://www.mathworks.com/products/ matlab.html
8. http://www.deeplearning.net/software/t heano/
9. https://scikit-learn.org/stable/
10. Sai Pooja, G., Raja Rajeswari, P., Yamini Radha, V., Navya Krishna.G., Naga Sri Ram.B., Recognition of fake currency note using convolutional neural networks(2016). International Journal of Innovative Technology and Exploring Engineering, 58-63,8(5). Mohammed Ali Al-Garadi, Mohammad Rashid Hussain, Henry Friday Nweke, Ihsanali,Ghulammujtaba1,Harunachiro Ma,Hasanalikhattak,Andabdullahgani "Predicti- Ngcyber Bullying On Social Networks.Yangzhou, Daewookkim, Junjiezhang,(Member,Ieee),Lili Liu1, Huanjin3, "(IEEE)ProGuard: Detecting Malicious Accounts in Social Network-Based Online Promotions".
11. Mauro Conti University of Padua, Radha Poovendran University of Washington, Marco Secchiero University of Padua,"FakeBook: Detecting Fake Profiles in On- line Social Networks(2012)", ACM /IEEE International Conference on Advances in Social Networks Analysis and Mining.
12. ni .N., Smruthi.M.,"A Hybrid Scheme for Detecting Fake Accounts in Facebook" ISSN: 2277- 3878, (IJRTE)International Journal of Recent Technology and Engineering (2019) , Issue-5S3, Volume-7.
13. NarsimhaGugulothu, Srinivas Rao Pulluri "A Comprehensive Model for Detecting Fake Profiles in Online Social Networks (2016)".Dr.Narsimha.G, Dr.JayadevGyani, P. Srinivas Rao ,"Fake Profiles Identification in Online Social Networks Using Machine Learning and NLP(2018)", International Journal of Applied Engineering Research ISSN 0973-4562, Number 6, Volume 13.
14. Reddy, A. V. N., &Phanikrishna, C. Contour tracking based knowledge extraction and object recognition using deep learning neural networks (2016). Paper presented at the Proceedings on 2nd International Conference on Next Generation Computing Technologies in 2016, NGCT 2016, 352-354. doi:10.1109/NGCT.2016.7877440.
15. V. Rama Krishna, & K.Kanaka Durga. Automatic detection of illegitimate websites with mutual clustering. (2016) International Journal of Electrical and Computer Engineering, 6(3), 995-1001. doi:10.11591/ijece. v6i3.9878
16. feedforward and multi-layer feedforward neural network (2016). Journal of Theoretical and Applied Information Technology, 150-156,84(2). Challa, N., Pasupuleti, S. K, & Chandra, J. V. A practical approach to E-mail spam filters to protect data from advanced persistent threat. (2016) Paper presented at the Proceedings of IEEE International Conference on Circuit, Power and Computing Technologies, ICCPCT 2016, doi:10.1109/ICCPCT.2016.7530239.
17. Rajeswari Rao , & P.Vidyullatha. Machine learning techniques on multidimensional curve fitting data based on r_ square and chi_square methods (2016). International Journal of Electrical and Computer Engineering, . doi:10.11591/ijece.v6i3.91556(3), 974- 979.
18. K. Anand., & Kumar, Anomaly detection in online social network: A survey. Paper presented at the Proceedings of the(2017) International Conference on Inventive Communication and Computational Technologies,ICICCT 2017, 456-459. doi:10.1109/ICICCT.2017.7975239
19. Pradeepini, G., Patil, S. T., & Bangare, S. L (2017). Brain tumor classification using mixed method approach. Paper presented at the International Conference on Information Communication and Embedded Systems, ICICES, doi:10.1109/ICICES.2017.8070748

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304278 |

20. Jaya Lakshmi, R., &Subba Rao, G. V. A re- constructive algorithm to improve image recovery in compressed sensing. Journal of Theoretical and Applied Information Technology (2017), 95(20), 5443- 5453
21. B.V. Babu., & N.S. Rao, Survey on clinical prediction models for diabetes prediction. Journal of Big Data, 4(1) 2017 doi:10.1186/s40537-017-0082-7 Pradeepini, G., Pradeepa, G., Tejanagasri, B., &Gorrepati, S. H. Data classification and personal care management system by machine learning approach. International Journal of Engineering and Technology,2018 (UAE), 7(2.32 Special Issue 32), 219-223. doi: 10.22444/IBVS.6227_old
22. , M., Asha Deepika, R., & Sai Raghavendhar, C. Analysis on prediction of heart disease using data mining techniques. Journal of Advanced Research in Dynamical and Control Systems (2018), 10(2), 126-136.
23. Satish Babu, J., Niveditha, M., Bhavya, V., &Gowthami, K. Data mining techniques for herbs. International Journal of Engineering and Technology,2018(UAE), 7(1.1 Special Issue 1), 406- 410.
24. Shinde, S. A., &Rajeswari, P. R. Intelligent health risk prediction systems using machine learning: A review. International Journal of Engineering and Technology,2018(UAE), 7(3), 1019- 1023. doi:10.14419/ijet.v7i3.12654
25. Sucharitha, G., &Senapati, R. K. Local extreme edge binary patterns for face recognition and image retrieval(2018). Journal of Advanced Research in Dynamical and Control Systems_10, 644-654.
26. V.N. Mandhala, D. Bhattacharyya, &T. Kim . Face detection using image morphology - A review. International Journalof Security and its Applications,2016, 10(4), 89-94. doi:10.14257/ijsia.2016.10.4.10 Satish Babu, J., Niveditha, M., Bhavya, V., &Gowthami, K. Data mining techniques for herbs. International Journal of Engineering and Technology,2018(UAE), 7(1.1 Special Issue 1), 406- 410.
27. Shinde, S. A., &Rajeswari, P. R. Intelligent health risk prediction systems using machine learning: A review. International Journal of Engineering and Technology,2018(UAE), 7(3), 1019- 1023. doi:10.14419/ijet. v7i3.12654
28. Sucharitha, G., &Senapati, R. K. Local extreme edge binary patterns for face recognition and image retrieval (2018). Journal of Advanced Research in Dynamical and Control Systems_10, 644-654.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details