**Use of Artificial Neural Networks to Identify Fake Profiles**

**1. INTRODUCTION**

In 2017 Facebook reached a total population of 2.46 billion users making it the most popular choice of social media [1]. Social media networks make revenues from the data provided by users. The average user does not know that their rights are given up the moment they use the social media network's service. Social media companies have a lot to gain at the expense of the user. Every time a user shares a new location, new photos, likes, dislikes, and tag other users in content posted, Facebook makes revenue via advertisements and data. More specifically, the average American user generates about $26.76 per quarter [2]. That number adds up quickly when millions of users are involved. In today's digital age, the ever-increasing dependency on computer technology has left the average citizen vulnerable to crimes such as data breaches and possible identity theft. These attacks can occur without notice and often without notification to the victims of a data breach. At this time, there is little incentive for social networks to improve their data security. These breaches often target social media networks such as Facebook and Twitter. They can also target banks and other financial institutions. There seems to be a newsworthy issue involving social media networks getting hacked every day. Recently, Facebook had a data breach which affected about 50 million users [3]. Facebook provides a set of clearly defined provisions that explain what they do with the user's data [4]. The policy does very little to prevent the constant exploitation of security and privacy. Fake profiles seem to slip through Facebook's built-in security features.

The other dangers of personal data being obtained for fraudulent purposes is the presence of bots and fake profiles. Bots are programs that can gather information about the user without the user even knowing. This process is known as web scraping. What is worse, is that this action is legal. Bots can be hidden or come in the form of a fake friend request on a social network site to gain access to private information. The solution presented in this paper intends to focus on the dangers of a bot in the form of a fake profile on your social media. This solution would come in the form of an algorithm. The language that we chose to use is Python. The algorithm would be able to determine if a current friend request that a user gets online is an actual person or if it is a bot or it is a fake friend request fishing for information. Our algorithm would work with the help of the social media companies, as we would need a training dataset from them to train our model and later verify if the profiles are fake or not. The algorithm could even work as a traditional layer on the user's web browser as a browser plug-in.

**1.1 Objective of the Project**

Social media networks make revenues from the data provided by users. The average user does not know that their rights are given up the moment they use the social media network's service. Social media companies have a lot to gain at the expense of the user. Every time a user shares a new location, new photos, likes, dislikes, and tag other users in content posted, Facebook makes revenue via advertisements and data. Social media networks make revenues from the data provided by users. The average user does not know that their rights are given up the moment they use the social media network's service. Social media companies have a lot to gain at the expense of the user. Every time a user shares a new location, new photos, likes, dislikes, and tag other users in content posted, Facebook makes revenue via advertisements and data. In this paper, we use machine learning, namely an artificial neural network to determine what the chances that Facebook friend request is authentic are or not.

**2. LITERATURE SURVEY**

**Aiding the detection of fake accounts in large scale social online services**

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 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.

**Audit and Analysis of Impostors: An experimental approach to detect fake profile in online social network**

In the present generation, the social life of every person has become associated with online social networks (OSN). These sites have made drastic changes in the way we socialize. Making friends and keeping in contact with them as well as being updated of their activities, has become easier. But with their rapid growth, problems like fake profiles, online impersonation have also increased. The risk lies in the fact that anybody can create a profile to impersonate a real person on the OSN. The fake profile could be exploited to build online relationship with a targeted person purely through online interactions with the friends of victim. In present work, we have proposed experimental framework with which detection of fake profile is feasible within the friend list, however this framework is restricted to a specific online social networking site namely Facebook. This framework extracts data from the friend list and uses it to classify them as real or fake by using unsupervised and supervised machine learning.

**Fake profile detection techniques in large-scale online social networks: A comprehensive review**

In the present era, online social networks are the most popular and rapid information propagation applications on the Internet. People of all ages spend most of their time on social networking sites. Huge volumes of data are being created and shared through social networks around the world. These interests have given rise to illegitimate users who engage in fraudulent activities against social network users. On social networks, fake profile creation is considered to cause more harm than any other form of cyber crime. This crime has to be detected even before the user is notified about the fake profile creation. Many algorithms and methods, most of which use the huge volume of unstructured data generated from social networks, have been proposed for the detection of fake profiles. This study presents a survey of the existing and latest technical work on fake profile detection.

**Online Social Networks: Threats and Solutions**

Many online social network (OSN) users are unaware of the numerous security risks that exist in these networks, including privacy violations, identity theft, and sexual harassment, just to name a few. According to recent studies, OSN users readily expose personal and private details about themselves, such as relationship status, date of birth, school name, email address, phone number, and even home address. This information, if put into the wrong hands, can be used to harm users both in the virtual world and in the real world. These risks become even more severe when the users are children. In this paper, we present a thorough review of the different security and privacy risks, which threaten the well-being of OSN users in general, and children in particular. In addition, we present an overview of existing solutions that can provide better protection, security, and privacy for OSN users. We also offer simple-to-implement recommendations for OSN users, which can improve their security and privacy when using these platforms. Furthermore, we suggest future research directions.

**Advanced social engineering attacks**

Social engineering has emerged as a serious threat in virtual communities and is an effective means to attack information systems. The services used by today's knowledge workers prepare the ground for sophisticated social engineering attacks. The growing trend towards BYOD (bring your own device) policies and the use of online communication and collaboration tools in private and business environments aggravate the problem. In globally acting companies, teams are no longer geographically co-located, but staffed just-in-time. The decrease in personal interaction combined with a plethora of tools used for communication (e-mail, IM, Skype, Dropbox, LinkedIn, Lync, etc.) create new attack vectors for social engineering attacks. Recent attacks on companies such as the New York Times and RSA have shown that targeted spear-phishing attacks are an effective, evolutionary step of social engineering attacks. Combined with zero-day-exploits, they become a dangerous weapon that is often used by advanced persistent threats. This paper provides a taxonomy of well-known social engineering attacks as well as a comprehensive overview of advanced social engineering attacks on the knowledge worker.

**An analysis of social network-based Sybil defenses**

Recently, there has been much excitement in the research community over using social networks to mitigate multiple identity, or Sybil, attacks. A number of schemes have been proposed, but they differ greatly in the algorithms they use and in the networks upon which they are evaluated. As a result, the research community lacks a clear understanding of how these schemes compare against each other, how well they would work on real-world social networks with different structural properties, or whether there exist other (potentially better) ways of Sybil defense. In this paper, we show that, despite their considerable differences, existing Sybil defense schemes work by detecting local communities (i.e., clusters of nodes more tightly knit than the rest of the graph) around a trusted node. Our finding has important implications for both existing and future designs of Sybil defense schemes. First, we show that there is an opportunity to leverage the substantial amount of prior work on general community detection algorithms in order to defend against Sybils. Second, our analysis reveals the fundamental limits of current social network-based Sybil defenses: We demonstrate that networks with well-defined community structure are inherently more vulnerable to Sybil attacks, and that, in such networks, Sybils can carefully target their links in order make their attacks more effective.

**Discovery of anomalous behaviour in temporal networks**

In this work we consider the problem of detecting anomalous behaviour and present a novel approach that allows ‘behaviour’ to be classified as either to be normal or abnormal by checking the p-value associated with the occurrence of the behaviour which is modelled following a binomial distribution within a discrete time model. We investigate the problem of detecting anomalous behaviour by looking at how communication evolves over time in a social network graph. Under the assumption that some nodes of the network could be labelled qualitatively, we present a novel approach that allows us to infer a subset of nodes of the social network which might share the same qualitative connotation. In other words, assuming one or more members belong to some criminal organisation, we wish to investigate how many other persons belong to the same organisation. We have tested our method in two datasets, VAST2008 and a Twitter Dataset (data collected in 2012), with encouraging results.

**Scalable community detection in massive social networks using MapReduce**

In this paper, we present a community-detection solution for massive-scale social networks using MapReduce, a parallel programming framework. We use a similarity metric to model the community probability, and the model is designed to be parallelizable and scalable in the MapReduce framework. More importantly, we propose a set of degree-based preprocessing and postprocessing techniques named DEPOLD (DElayed Processing of Large Degree nodes) that significantly improve both the community-detection accuracy and performance. With DEPOLD, delaying analysis of 1% of high-degree nodes to the postprocessing stage reduces both processing time and storage space by one order of magnitude. DEPOLD can be applied to other graph-clustering problems. Furthermore, we design and implement two similarity calculation algorithms using MapReduce with different computation and communication characteristics in order to adapt to various system configurations. Finally, we conduct experiments with publicly available datasets. Our evaluation demonstrates the effectiveness, efficiency, and scalability of the proposed solution.

**3. SYSTEM ANALYSIS**

**3.1 Existing System**

In today's digital age, the ever-increasing dependency on computer technology has left the average citizen vulnerable to crimes such as data breaches and possible identity theft. These attacks can occur without notice and often without notification to the victims of a data breach.

**Disadvantages of Existing System:**

1. Security is less.
2. Attacks are more.

**3.2 Proposed System**

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

**Advantages:**

1. Security is more.

**3.3. PROCESS MODEL USED WITH JUSTIFICATION**

**SDLC (Umbrella Model):**

**Umbrella Activity**

**Umbrella Activity**

**Umbrella Activity**

1. Feasibility Study
2. TEAM FORMATION
3. Project Specification PREPARATION

Business Requirement Documentation

ANALYSIS & DESIGN

CODE

UNIT TEST

DOCUMENT CONTROL

ASSESSMENT

TRAINING

INTEGRATION & SYSTEM TESTING

DELIVERY/INSTALLATION

ACCEPTANCE TEST

Requirements Gathering

SDLC is nothing but Software Development Life Cycle. It is a standard which is used by software industry to develop good software.

**Stages in SDLC:**

* Requirement Gathering
* Analysis
* Designing
* Coding
* Testing
* Maintenance

**Requirements Gathering** **stage:**

The requirements gathering process takes as its input the goals identified in the high-level requirements section of the project plan. Each goal will be refined into a set of one or more requirements. These requirements define the major functions of the intended application, define operational data areas and reference data areas, and define the initial data entities. Major functions include critical processes to be managed, as well as mission critical inputs, outputs and reports. A user class hierarchy is developed and associated with these major functions, data areas, and data entities. Each of these definitions is termed a Requirement. Requirements are identified by unique requirement identifiers and, at minimum, contain a requirement title and textual description.



These requirements are fully described in the primary deliverables for this stage: the Requirements Document and the Requirements Traceability Matrix (RTM). The requirements document contains complete descriptions of each requirement, including diagrams and references to external documents as necessary. Note that detailed listings of database tables and fields are *not* included in the requirements document.

The title of each requirement is also placed into the first version of the RTM, along with the title of each goal from the project plan. The purpose of the RTM is to show that the product components developed during each stage of the software development lifecycle are formally connected to the components developed in prior stages.

In the requirements stage, the RTM consists of a list of high-level requirements, or goals, by title, with a listing of associated requirements for each goal, listed by requirement title. In this hierarchical listing, the RTM shows that each requirement developed during this stage is formally linked to a specific product goal. In this format, each requirement can be traced to a specific product goal, hence the term requirements traceability.

The outputs of the requirements definition stage include the requirements document, the RTM, and an updated project plan.

* Feasibility study is all about identification of problems in a project.
* No. of staff required to handle a project is represented as Team Formation, in this case only modules are individual tasks will be assigned to employees who are working for that project.
* Project Specifications are all about representing of various possible inputs submitting to the server and corresponding outputs along with reports maintained by administrator.

**Analysis Stage:**

The planning stage establishes a bird's eye view of the intended software product, and uses this to establish the basic project structure, evaluate feasibility and risks associated with the project, and describe appropriate management and technical approaches.



The most critical section of the project plan is a listing of high-level product requirements, also referred to as goals. All of the software product requirements to be developed during the requirements definition stage flow from one or more of these goals. The minimum information for each goal consists of a title and textual description, although additional information and references to external documents may be included. The outputs of the project planning stage are the configuration management plan, the quality assurance plan, and the project plan and schedule, with a detailed listing of scheduled activities for the upcoming Requirements stage, and high level estimates of effort for the out stages.

**Designing Stage:**

The design stage takes as its initial input the requirements identified in the approved requirements document. For each requirement, a set of one or more design elements will be produced as a result of interviews, workshops, and/or prototype efforts. Design elements describe the desired software features in detail, and generally include functional hierarchy diagrams, screen layout diagrams, tables of business rules, business process diagrams, pseudo code, and a complete entity-relationship diagram with a full data dictionary. These design elements are intended to describe the software in sufficient detail that skilled programmers may develop the software with minimal additional input.

  
When the design document is finalized and accepted, the RTM is updated to show that each design element is formally associated with a specific requirement. The outputs of the design stage are the design document, an updated RTM, and an updated project plan.

**Development (Coding) Stage:**

The development stage takes as its primary input the design elements described in the approved design document. For each design element, a set of one or more software artifacts will be produced. Software artifacts include but are not limited to menus, dialogs, and data management forms, data reporting formats, and specialized procedures and functions. Appropriate test cases will be developed for each set of functionally related software artifacts, and an online help system will be developed to guide users in their interactions with the software.



The RTM will be updated to show that each developed artifact is linked to a specific design element, and that each developed artifact has one or more corresponding test case items. At this point, the RTM is in its final configuration. The outputs of the development stage include a fully functional set of software that satisfies the requirements and design elements previously documented, an online help system that describes the operation of the software, an implementation map that identifies the primary code entry points for all major system functions, a test plan that describes the test cases to be used to validate the correctness and completeness of the software, an updated RTM, and an updated project plan.

**Integration & Test Stage:**

During the integration and test stage, the software artifacts, online help, and test data are migrated from the development environment to a separate test environment. At this point, all test cases are run to verify the correctness and completeness of the software. Successful execution of the test suite confirms a robust and complete migration capability. During this stage, reference data is finalized for production use and production users are identified and linked to their appropriate roles. The final reference data (or links to reference data source files) and production user list are compiled into the Production Initiation Plan.



The outputs of the integration and test stage include an integrated set of software, an online help system, an implementation map, a production initiation plan that describes reference data and production users, an acceptance plan which contains the final suite of test cases, and an updated project plan.

* **Installation & Acceptance Test:**

During the installation and acceptance stage, the software artifacts, online help, and initial production data are loa ded onto the production server. At this point, all test cases are run to verify the correctness and completeness of the software. Successful execution of the test suite is a prerequisite to acceptance of the software by the customer.

After customer personnel have verified that the initial production data load is correct and the test suite has been executed with satisfactory results, the customer formally accepts the delivery of the software.



The primary outputs of the installation and acceptance stage include a production application, a completed acceptance test suite, and a memorandum of customer acceptance of the software. Finally, the PDR enters the last of the actual labor data into the project schedule and locks the project as a permanent project record. At this point the PDR "locks" the project by archiving all software items, the implementation map, the source code, and the documentation for future reference.

**Maintenance:**

Outer rectangle represents maintenance of a project, Maintenance team will start with requirement study, understanding of documentation later employees will be assigned work and they will undergo training on that particular assigned category. For this life cycle there is no end, it will be continued so on like an umbrella (no ending point to umbrella sticks).

**3.4. Software Requirement Specification**

**3.4.1. Overall Description**

A Software Requirements Specification (SRS) – a [requirements specification](http://en.wikipedia.org/wiki/Requirements_specification) for a [software system](http://en.wikipedia.org/wiki/Software_system) is a complete description of the behavior of a system to be developed. It includes a set of [use cases](http://en.wikipedia.org/wiki/Use_case) that describe all the interactions the users will have with the software. In addition to use cases, the SRS also contains non-functional requirements. [Nonfunctional requirements](http://en.wikipedia.org/wiki/Non-functional_requirements) are requirements which impose constraints on the design or implementation (such as [performance engineering](http://en.wikipedia.org/wiki/Performance_engineering) requirements, [quality](http://en.wikipedia.org/wiki/Quality_%28business%29) standards, or design constraints).

System requirements specification: A structured collection of information that embodies the requirements of a system. A [business analyst](http://en.wikipedia.org/wiki/Business_analyst), sometimes titled [system analyst](http://en.wikipedia.org/wiki/System_analyst), is responsible for analyzing the business needs of their clients and stakeholders to help identify business problems and propose solutions. Within the [systems development lifecycle](http://en.wikipedia.org/wiki/Systems_development_life_cycle) domain, the BA typically performs a liaison function between the business side of an enterprise and the information technology department or external service providers. Projects are subject to three sorts of requirements:

* [Business requirements](http://en.wikipedia.org/wiki/Business_requirements) describe in business terms *what* must be delivered or accomplished to provide value.
* Product requirements describe properties of a system or product (which could be one of several ways to accomplish a set of business requirements.)
* Process requirements describe activities performed by the developing organization. For instance, process requirements could specify .Preliminary investigation examine project feasibility, the likelihood the system will be useful to the organization. The main objective of the feasibility study is to test the Technical, Operational and Economical feasibility for adding new modules and debugging old running system. All system is feasible if they are unlimited resources and infinite time. There are aspects in the feasibility study portion of the preliminary investigation:
* **ECONOMIC FEASIBILITY**

A system can be developed technically and that will be used if installed must still be a good investment for the organization. In the economical feasibility, the development cost in creating the system is evaluated against the ultimate benefit derived from the new systems. Financial benefits must equal or exceed the costs. The system is economically feasible. It does not require any addition hardware or software. Since the interface for this system is developed using the existing resources and technologies available at NIC, There is nominal expenditure and economical feasibility for certain.

* **Operational Feasibility**

Proposed projects are beneficial only if they can be turned out into information system. That will meet the organization’s operating requirements. Operational feasibility aspects of the project are to be taken as an important part of the project implementation. This system is targeted to be in accordance with the above-mentioned issues. Beforehand, the management issues and user requirements have been taken into consideration. So there is no question of resistance from the users that can undermine the possible application benefits. The well-planned design would ensure the optimal utilization of the computer resources and would help in the improvement of performance status.

* **TECHNICAL FEASIBILITY**

Earlier no system existed to cater to the needs of ‘Secure Infrastructure Implementation System’. The current system developed is technically feasible. It is a web based user interface for audit workflow at NIC-CSD. Thus it provides an easy access to .the users. The database’s purpose is to create, establish and maintain a workflow among various entities in order to facilitate all concerned users in their various capacities or roles. Permission to the users would be granted based on the roles specified. Therefore, it provides the technical guarantee of accuracy, reliability and security.

**3.4.2. External Interface Requirements**

**User Interface**

The user interface of this system is a user friendly python Graphical User Interface.

**Hardware Interfaces**

The interaction between the user and the console is achieved through python capabilities.

**Software Interfaces**

The required software is python.

**Operating Environment**

Windows XP.

**HARDWARE REQUIREMENTS:**

# Processor - Pentium –IV

* Speed - 1.1 Ghz
* RAM - 256 MB(min)
* Hard Disk - 20 GB
* Key Board - Standard Windows Keyboard
* Mouse - Two or Three Button Mouse
* Monitor - SVGA

**SOFTWARE REQUIREMENTS:**

* Operating System - Windows7/8
* Programming Language - Python

**4. SYSTEM DESIGN**

**UML Diagram:**

**Class Diagram:**

The class diagram is the main building block of object oriented modeling. It is used both for general conceptual modeling of the systematic of the application, and for detailed modeling translating the models into programming code. Class diagrams can also be used for data modeling. The classes in a class diagram represent both the main objects, interactions in the application and the classes to be programmed. In the diagram, classes are represented with boxes which contain three parts:

* The upper part holds the name of the class
* The middle part contains the attributes of the class
* The bottom part gives the methods or operations the class can take or undertake

**Class Diagram:**



**Use case Diagram:**

A **use case diagram** at its simplest is a representation of a user's interaction with the system and depicting the specifications of a use case. A use case diagram can portray the different types of users of a system and the various ways that they interact with the system. This type of diagram is typically used in conjunction with the textual use case and will often be accompanied by other types of diagrams as well.

**Use case Diagram for Admin:**

Generate ANN Train Model

View ANN Train Dataset

Logout

Admin

Login

**Use case Diagram for User:**

Enter User Account

Get Predicted Account Details

User

User Account Check Screen

**Sequence diagram:**

A **sequence diagram** is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. A sequence diagram shows object interactions arranged in time sequence. It depicts the objects and classes involved in the scenario and the sequence of messages exchanged between the objects needed to carry out the functionality of the scenario. Sequence diagrams are typically associated with use case realizations in the Logical View of the system under development. Sequence diagrams are sometimes called **event diagrams**, **event scenarios**, and timing diagrams.



**Collaboration diagram:**

A collaboration diagram describes interactions among objects in terms of sequenced messages. Collaboration diagrams represent a combination of information taken from class, sequence, and use case diagrams describing both the static structure and dynamic behavior of a system.



**Component Diagram:**

In the Unified Modeling Language, a component diagram depicts how components are wired together to form larger components and or software systems. They are used to illustrate the structure of arbitrarily complex systems.

Components are wired together by using an assembly connector to connect the required interface of one component with the provided interface of another component. This illustrates the service consumer - service provider relationship between the two components.

**Component Diagram for Admin:**

Admin

Login

Generate ANN

Train Model

View ANN Train Dataset

Logout

**Component Diagram for User:**

User Account Check Screen

Enter Account Details

User

Get predicted details

**Deployment Diagram:**

A **deployment diagram** in the Unified Modeling Language models the *physical* deployment of artifacts on nodes. To describe a web site, for example, a deployment diagram would show what hardware components ("nodes") exist (e.g., a web server, an application server, and a database server), what software components ("artifacts") run on each node (e.g., web application, database), and how the different pieces are connected (e.g. JDBC, REST, RMI).

The nodes appear as boxes, and the artifacts allocated to each node appear as rectangles within the boxes. Nodes may have sub nodes, which appear as nested boxes. A single node in a deployment diagram may conceptually represent multiple physical nodes, such as a cluster of database servers.

**Deployment Diagram:**

System

Admin

User

**Activity Diagram:**

Activity diagram is another important diagram in UML to describe dynamic aspects of the system. It is basically a flow chart to represent the flow form one activity to another activity. The activity can be described as an operation of the system. So the control flow is drawn from one operation to another. This flow can be sequential, branched or concurrent

Activity Diagram for Admin:

Admin

Login

Generate Train Model

View ANN Train Dataset

Activity Diagram for User:

User Account Check Screen

Enter Account details

Get predicted details

**Data Flow Diagram:**

Data flow diagrams illustrate how data is processed by a system in terms of inputs and outputs. Data flow diagrams can be used to provide a clear representation of any business function. The technique starts with an overall picture of the business and continues by analyzing each of the functional areas of interest. This analysis can be carried out in precisely the level of detail required. The technique exploits a method called top-down expansion to conduct the analysis in a targeted way.

As the name suggests, Data Flow Diagram (DFD) is an illustration that explicates the passage of information in a process. A DFD can be easily drawn using simple symbols. Additionally, complicated processes can be easily automated by creating DFDs using easy-to-use, free downloadable diagramming tools. A DFD is a model for constructing and analyzing information processes. DFD illustrates the flow of information in a process depending upon the inputs and outputs. A DFD can also be referred to as a Process Model. A DFD demonstrates business or technical process with the support of the outside data saved, plus the data flowing from the process to another and the end results.

2. Register Successfully

Admin

4. Accuracy displayed

6. Dataset details displayed

1. Login

3. Generate ANN train model

5. View ANN Train Dataset

7. Logout

8. Enter user account details

9. Get predicted details10. Predicted details are displayed

User

**5. IMPLEMETATION**

**5.1 Python**

Python is a general-purpose language. It has wide range of applications from Web development (like: Django and Bottle), scientific and mathematical computing (Orange, SymPy, NumPy) to desktop graphical user Interfaces (Pygame, Panda3D). The syntax of the language is clean and length of the code is relatively short. It's fun to work in Python because it allows you to think about the problem rather than focusing on the syntax.

**History of Python:**

Python is a fairly old language created by Guido Van Rossum. The design began in the late 1980s and was first released in February 1991.

**Why Python was created?**

In late 1980s, Guido Van Rossum was working on the Amoeba distributed operating system group. He wanted to use an interpreted language like ABC (ABC has simple easy-to-understand syntax) that could access the Amoeba system calls. So, he decided to create a language that was extensible. This led to design of a new language which was later named Python.

**Why the name Python?**

No. It wasn't named after a dangerous snake. Rossum was fan of a comedy series from late seventies. The name "Python" was adopted from the same series "Monty Python's Flying Circus".

**Features of Python:**

**A simple language which is easier to learn**

Python has a very simple and elegant syntax. It's much easier to read and write Python programs compared to other languages like: C++, Java, C#. Python makes programming fun and allows you to focus on the solution rather than syntax.

If you are a newbie, it's a great choice to start your journey with Python.

**Free and open-source**

You can freely use and distribute Python, even for commercial use. Not only can you use and distribute software’s written in it, you can even make changes to the Python's source code.

Python has a large community constantly improving it in each iteration.

**Portability**

You can move Python programs from one platform to another, and run it without any changes.

It runs seamlessly on almost all platforms including Windows, Mac OS X and Linux.

**Extensible and Embeddable**

Suppose an application requires high performance. You can easily combine pieces of C/C++ or other languages with Python code.

This will give your application high performance as well as scripting capabilities which other languages may not provide out of the box.

**A high-level, interpreted language**

Unlike C/C++, you don't have to worry about daunting tasks like memory management, garbage collection and so on.

Likewise, when you run Python code, it automatically converts your code to the language your computer understands. You don't need to worry about any lower-level operations.

**Large standard libraries to solve common tasks**

Python has a number of standard libraries which makes life of a programmer much easier since you don't have to write all the code yourself. For example: Need to connect MySQL database on a Web server? You can use MySQLdb library using import MySQLdb .

Standard libraries in Python are well tested and used by hundreds of people. So you can be sure that it won't break your application.

**Object-oriented**

Everything in Python is an object. Object oriented programming (OOP) helps you solve a complex problem intuitively.

With OOP, you are able to divide these complex problems into smaller sets by creating objects.

**Applications of Python:**

**1. Simple Elegant Syntax**

Programming in Python is fun. It's easier to understand and write Python code. Why? The syntax feels natural. Take this source code for an example:

a = 2

b = 3

sum = a + b

print(sum)

**2. Not overly strict**

You don't need to define the type of a variable in Python. Also, it's not necessary to add semicolon at the end of the statement.

Python enforces you to follow good practices (like proper indentation). These small things can make learning much easier for beginners.

**3. Expressiveness of the language**

Python allows you to write programs having greater functionality with fewer lines of code. Here's a link to the source code of Tic-tac-toe game with a graphical interface and a smart computer opponent in less than 500 lines of code. This is just an example. You will be amazed how much you can do with Python once you learn the basics.

**4. Great Community and Support**

Python has a large supporting community. There are numerous active forums online which can be handy if you are stuck.

**5.2 Sample Code:**

**Views.py:**

from django.shortcuts import render

from django.template import RequestContext

from django.contrib import messages

from django.http import HttpResponse

import pandas as pd

from sklearn.model\_selection import train\_test\_split

from keras.models import Sequential

from keras.layers.core import Dense,Activation,Dropout

from keras.callbacks import EarlyStopping

from sklearn.preprocessing import OneHotEncoder

from keras.optimizers import Adam

global model

def index(request):

if request.method == 'GET':

return render(request, 'index.html', {})

def User(request):

if request.method == 'GET':

return render(request, 'User.html', {})

def Admin(request):

if request.method == 'GET':

return render(request, 'Admin.html', {})

def AdminLogin(request):

if request.method == 'POST':

username = request.POST.get('username', False)

password = request.POST.get('password', False)

if username == 'admin' and password == 'admin':

context= {'data':'welcome '+username}

return render(request, 'AdminScreen.html', context)

else:

context= {'data':'login failed'}

return render(request, 'Admin.html', context)

def importdata():

balance\_data = pd.read\_csv('C:/FakeProfile/Profile/dataset/dataset.txt')

balance\_data = balance\_data.abs()

rows = balance\_data.shape[0] # gives number of row count

cols = balance\_data.shape[1] # gives number of col count

return balance\_data

def splitdataset(balance\_data):

X = balance\_data.values[:, 0:8]

y\_= balance\_data.values[:, 8]

y\_ = y\_.reshape(-1, 1)

encoder = OneHotEncoder(sparse=False)

Y = encoder.fit\_transform(y\_)

print(Y)

train\_x, test\_x, train\_y, test\_y = train\_test\_split(X, Y, test\_size=0.2)

return train\_x, test\_x, train\_y, test\_y

def UserCheck(request):

if request.method == 'POST':

data = request.POST.get('t1', False)

input = 'Account\_Age,Gender,User\_Age,Link\_Desc,Status\_Count,Friend\_Count,Location,Location\_IP\n';

input+=data+"\n"

f = open("C:/FakeProfile/Profile/dataset/test.txt", "w")

f.write(input)

f.close()

test = pd.read\_csv('C:/FakeProfile/Profile/dataset/test.txt')

test = test.values[:, 0:8]

predict = model.predict\_classes(test)

print(predict[0])

msg = ''

if str(predict[0]) == '0':

msg = "Given Account Details Predicted As Genuine"

if str(predict[0]) == '1':

msg = "Given Account Details Predicted As Fake"

context= {'data':msg}

return render(request, 'User.html', context)

def GenerateModel(request):

global model

data = importdata()

train\_x, test\_x, train\_y, test\_y = splitdataset(data)

model = Sequential()

model.add(Dense(200, input\_shape=(8,), activation='relu', name='fc1'))

model.add(Dense(200, activation='relu', name='fc2'))

model.add(Dense(2, activation='softmax', name='output'))

optimizer = Adam(lr=0.001)

model.compile(optimizer, loss='categorical\_crossentropy', metrics=['accuracy'])

print('CNN Neural Network Model Summary: ')

print(model.summary())

model.fit(train\_x, train\_y, verbose=2, batch\_size=5, epochs=200)

results = model.evaluate(test\_x, test\_y)

ann\_acc = results[1] \* 100

context= {'data':'ANN Accuracy : '+str(ann\_acc)}

return render(request, 'AdminScreen.html', context)

def ViewTrain(request):

if request.method == 'GET':

strdata = '<table border=1 align=center width=100%><tr><th><font size=4 color=white>Account Age</th><th><font size=4 color=white>Gender</th><th><font size=4 color=white>User Age</th><th><font size=4 color=white>Link Description</th> <th><font size=4 color=white>Status Count</th><th><font size=4 color=white>Friend Count</th><th><font size=4 color=white>Location</th><th><font size=4 color=white>Location IP</th><th><font size=4 color=white>Profile Status</th></tr><tr>'

data = pd.read\_csv('C:/FakeProfile/Profile/dataset/dataset.txt')

rows = data.shape[0] # gives number of row count

cols = data.shape[1] # gives number of col count

for i in range(rows):

for j in range(cols):

strdata+='<td><font size=3 color=white>'+str(data.iloc[i,j])+'</font></td>'

strdata+='</tr><tr>'

context= {'data':strdata}

return render(request, 'ViewData.html', context)

**6. TESTING**

**Implementation and Testing:**

Implementation is one of the most important tasks in project is the phase in which one has to be cautions because all the efforts undertaken during the project will be very interactive. Implementation is the most crucial stage in achieving successful system and giving the users confidence that the new system is workable and effective. Each program is tested individually at the time of development using the sample data and has verified that these programs link together in the way specified in the program specification. The computer system and its environment are tested to the satisfaction of the user.

## Implementation

## The implementation phase is less creative than system design. It is primarily concerned with user training, and file conversion. The system may be requiring extensive user training. The initial parameters of the system should be modifies as a result of a programming. A simple operating procedure is provided so that the user can understand the different functions clearly and quickly. The different reports can be obtained either on the inkjet or dot matrix printer, which is available at the disposal of the user. The proposed system is very easy to implement. In general implementation is used to mean the process of converting a new or revised system design into an operational one.

## Testing

Testing is the process where the test data is prepared and is used for testing the modules individually and later the validation given for the fields. Then the system testing takes place which makes sure that all components of the system property functions as a unit. The test data should be chosen such that it passed through all possible condition. Actually testing is the state of implementation which aimed at ensuring that the system works accurately and efficiently before the actual operation commence. The following is the description of the testing strategies, which were carried out during the testing period.

### System Testing

Testing has become an integral part of any system or project especially in the field of information technology. The importance of testing is a method of justifying, if one is ready to move further, be it to be check if one is capable to with stand the rigors of a particular situation cannot be underplayed and that is why testing before development is so critical. When the software is developed before it is given to user to user the software must be tested whether it is solving the purpose for which it is developed. This testing involves various types through which one can ensure the software is reliable. The program was tested logically and pattern of execution of the program for a set of data are repeated. Thus the code was exhaustively checked for all possible correct data and the outcomes were also checked.

**Module Testing**

To locate errors, each module is tested individually. This enables us to detect error and correct it without affecting any other modules. Whenever the program is not satisfying the required function, it must be corrected to get the required result. Thus all the modules are individually tested from bottom up starting with the smallest and lowest modules and proceeding to the next level. Each module in the system is tested separately. For example the job classification module is tested separately. This module is tested with different job and its approximate execution time and the result of the test is compared with the results that are prepared manually. The comparison shows that the results proposed system works efficiently than the existing system. Each module in the system is tested separately. In this system the resource classification and job scheduling modules are tested separately and their corresponding results are obtained which reduces the process waiting time.

**Integration Testing**

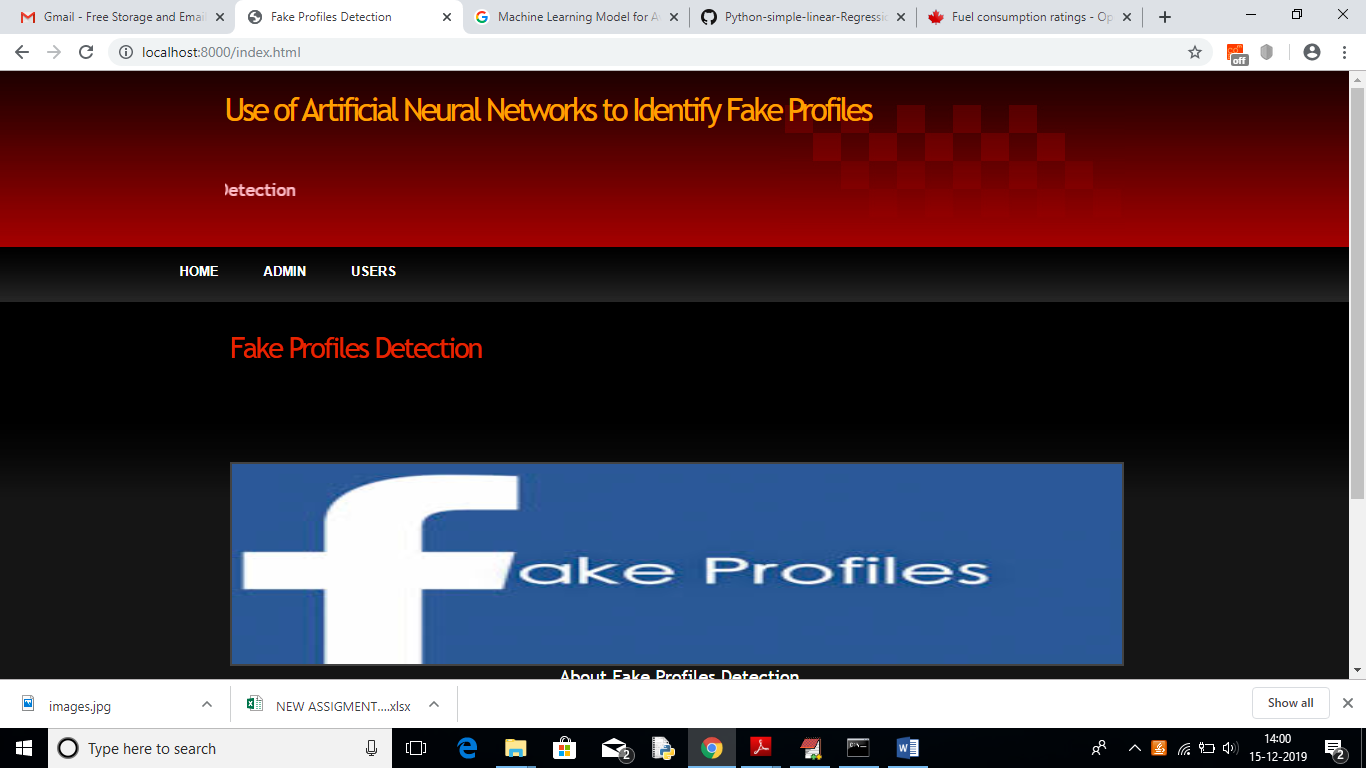
After the module testing, the integration testing is applied. When linking the modules there may be chance for errors to occur, these errors are corrected by using this testing. In this system all modules are connected and tested. The testing results are very correct. Thus the mapping of jobs with resources is done correctly by the system.

**Acceptance Testing**

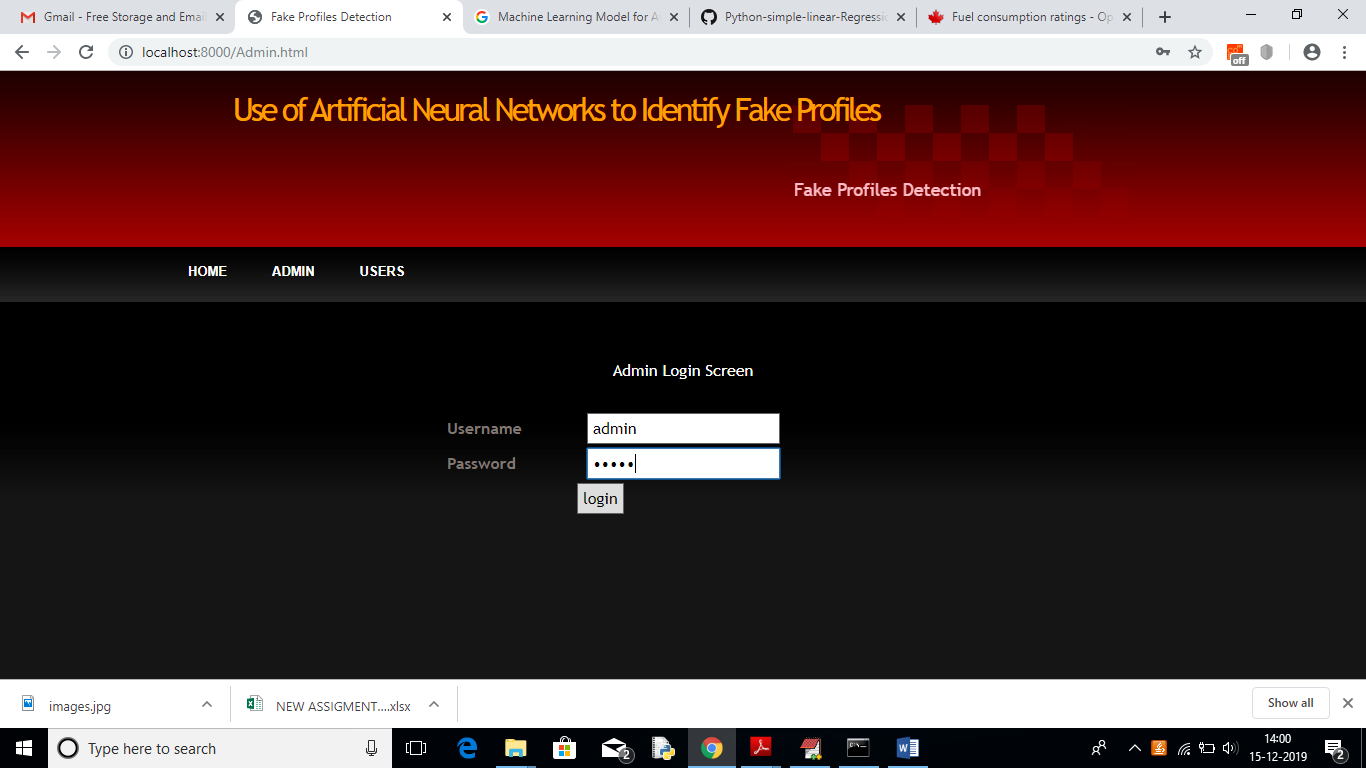
When that user fined no major problems with its accuracy, the system passers through a final acceptance test. This test confirms that the system needs the original goals, objectives and requirements established during analysis without actual execution which elimination wastage of time and money acceptance tests on the shoulders of users and management, it is finally acceptable and ready for the operation.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Test Case Id** | **Test Case Name** | **Test Case Desc.** | **Test Steps** | | | **Test Case Status** | **Test Priority** |
| **Step** | **Expected** | **Actual** |
| 01 | Admin Login | Test whether the admin is login or not into the system | If the Admin cannot enter the valid details | Admin cannot login to the system | Admin can login to the system | High | High |
| 02 | generate ANN train model | Verify the ANN train model generated or not | Without login to the system | cannot generate the ANN train model | ANN train model generated successfully | High | High |
| 03 | view ANN train Dataset | Verify the ANN train Dataset displayed or not | Without generating the training model | the ANN train Dataset details may not displayed | the ANN train Dataset details displayed successfully | High | High |
| 04 | Checking User Accounts | Verify either the user account is fake or genuine | Without entering the user account details | we cannot find that account is fake or genuine | account predicted details are displayed successfully | High | High |

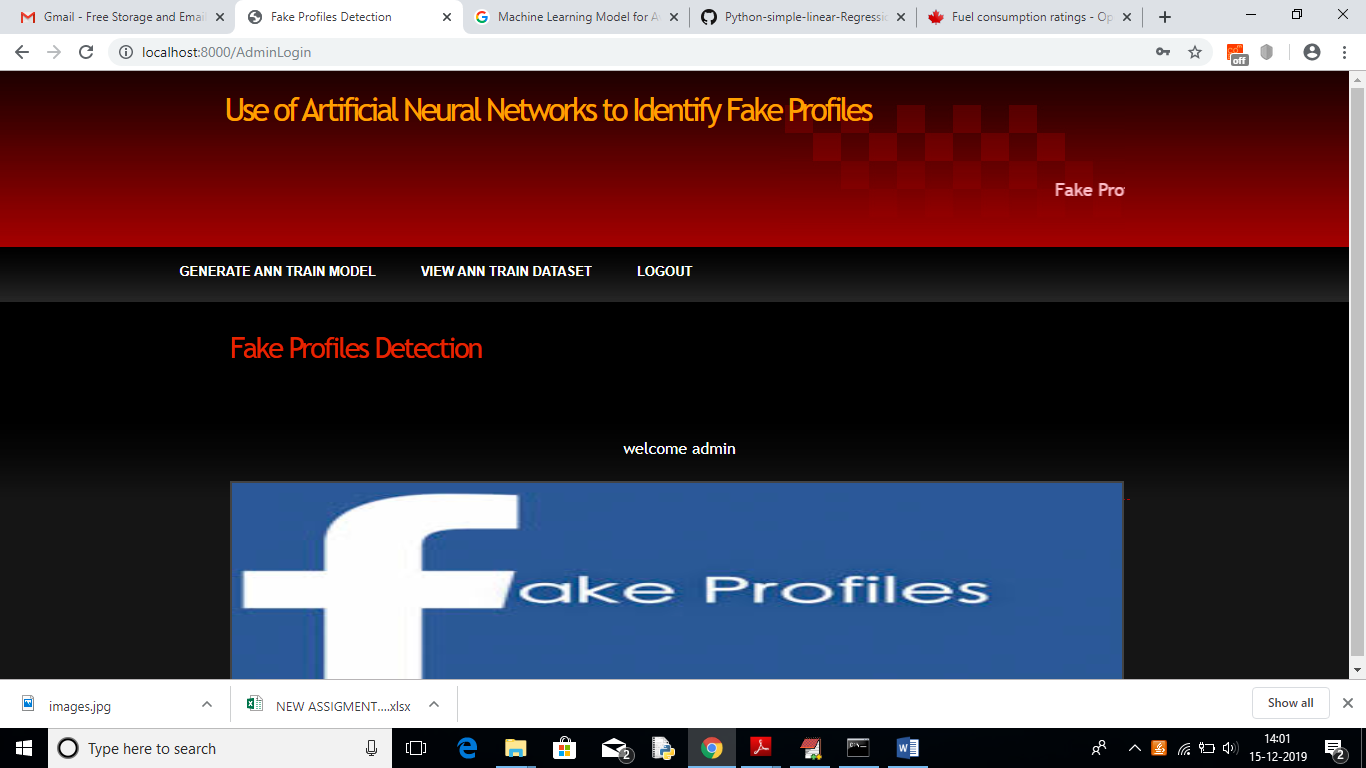
**7. SCREENSHOTS**



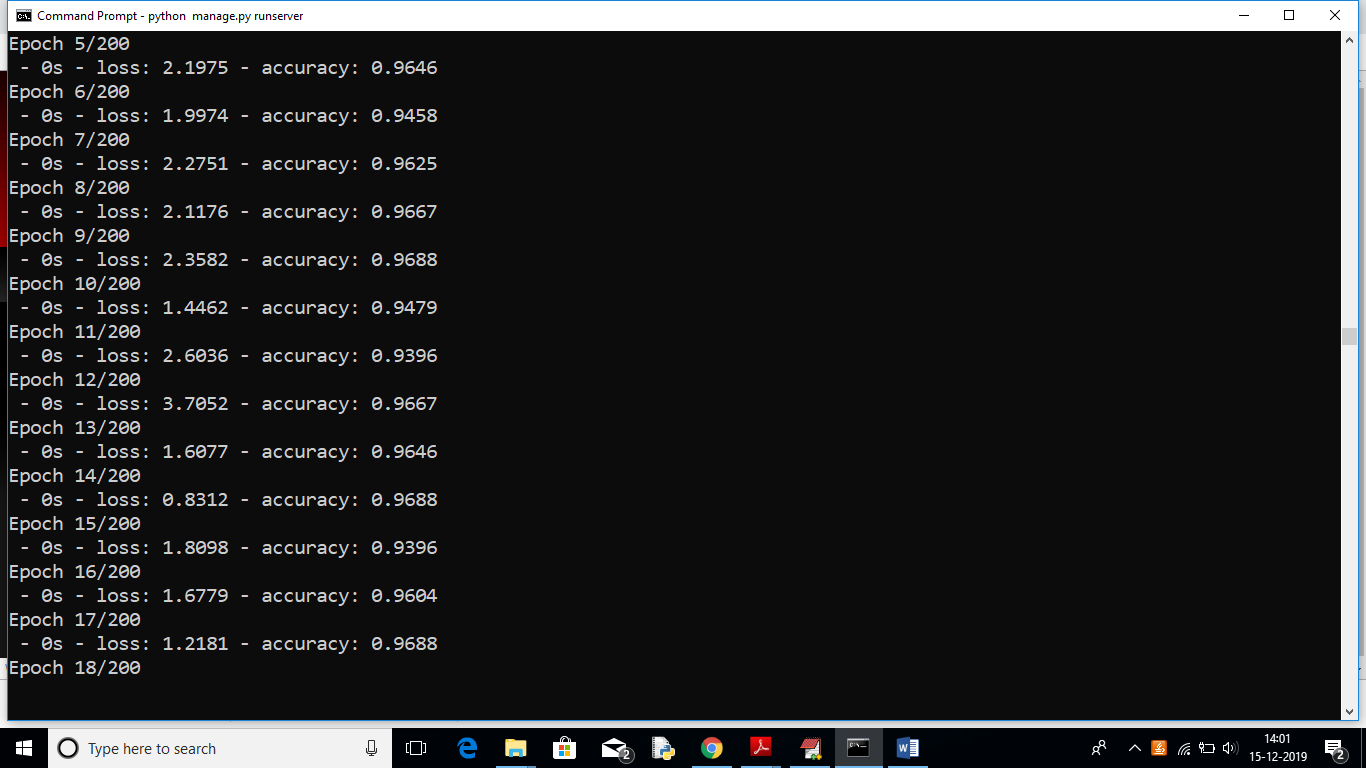
In above screen click on ‘ADMIN’ link to get below login screen



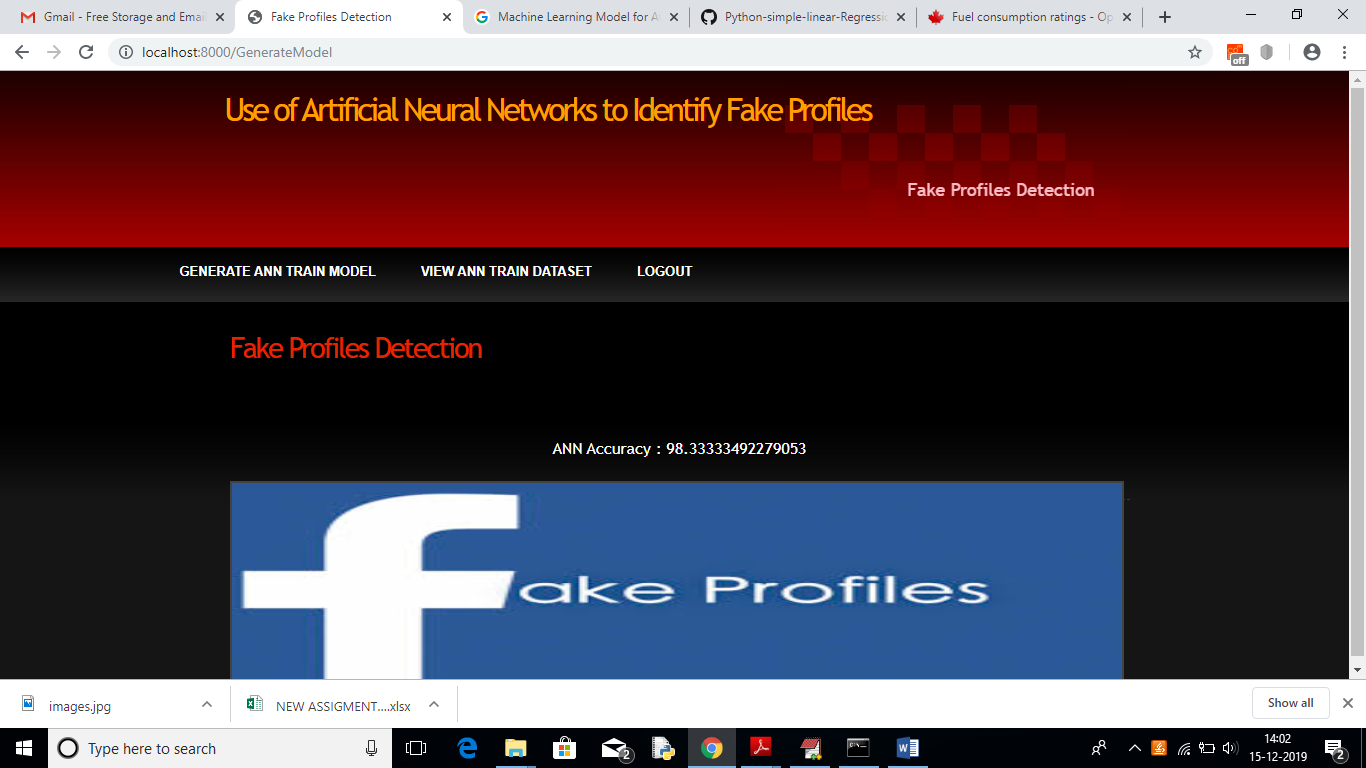
In above screen enter admin and admin as username and password to login as admin. After login will get below screen



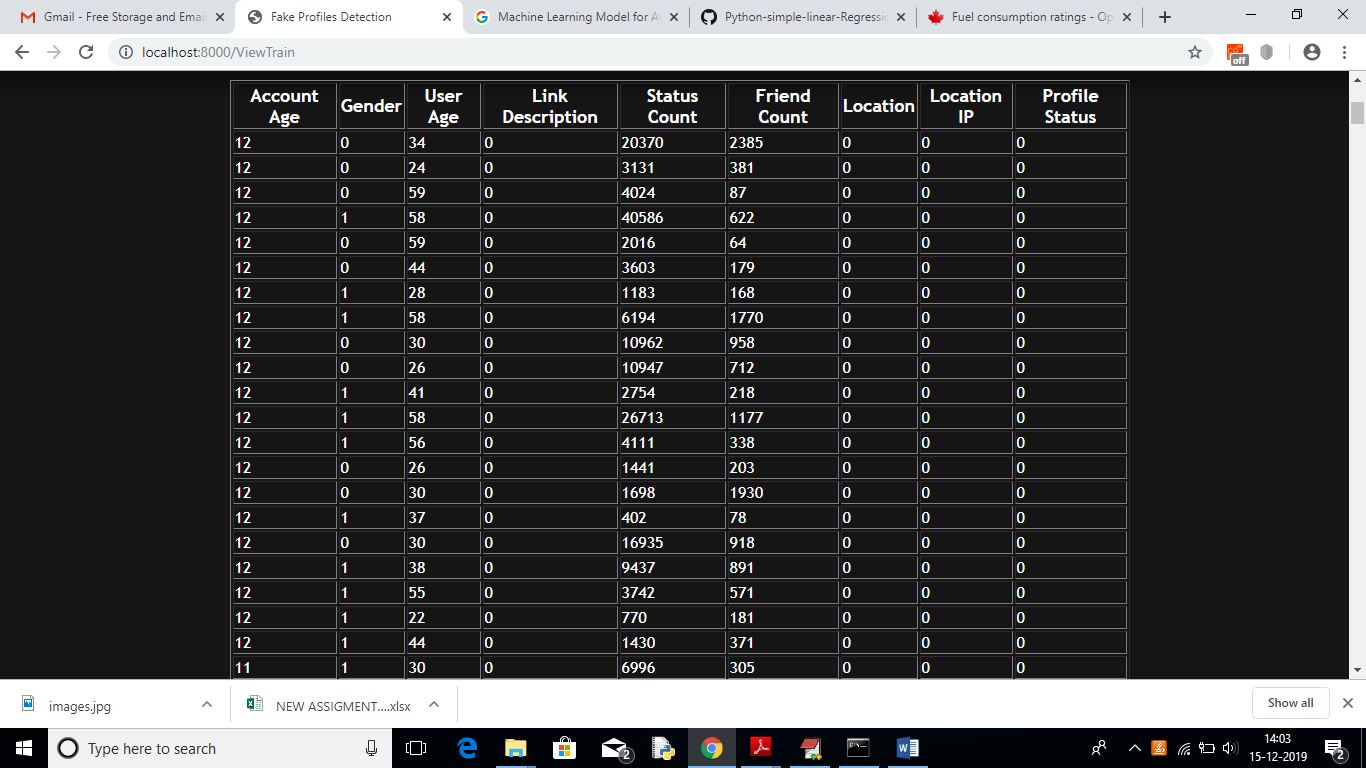
In above screen click on ‘Generate ANN Train Model’ to generate training model on dataset. After clicking on that link you can see server console to check ANN processing details with accuracy



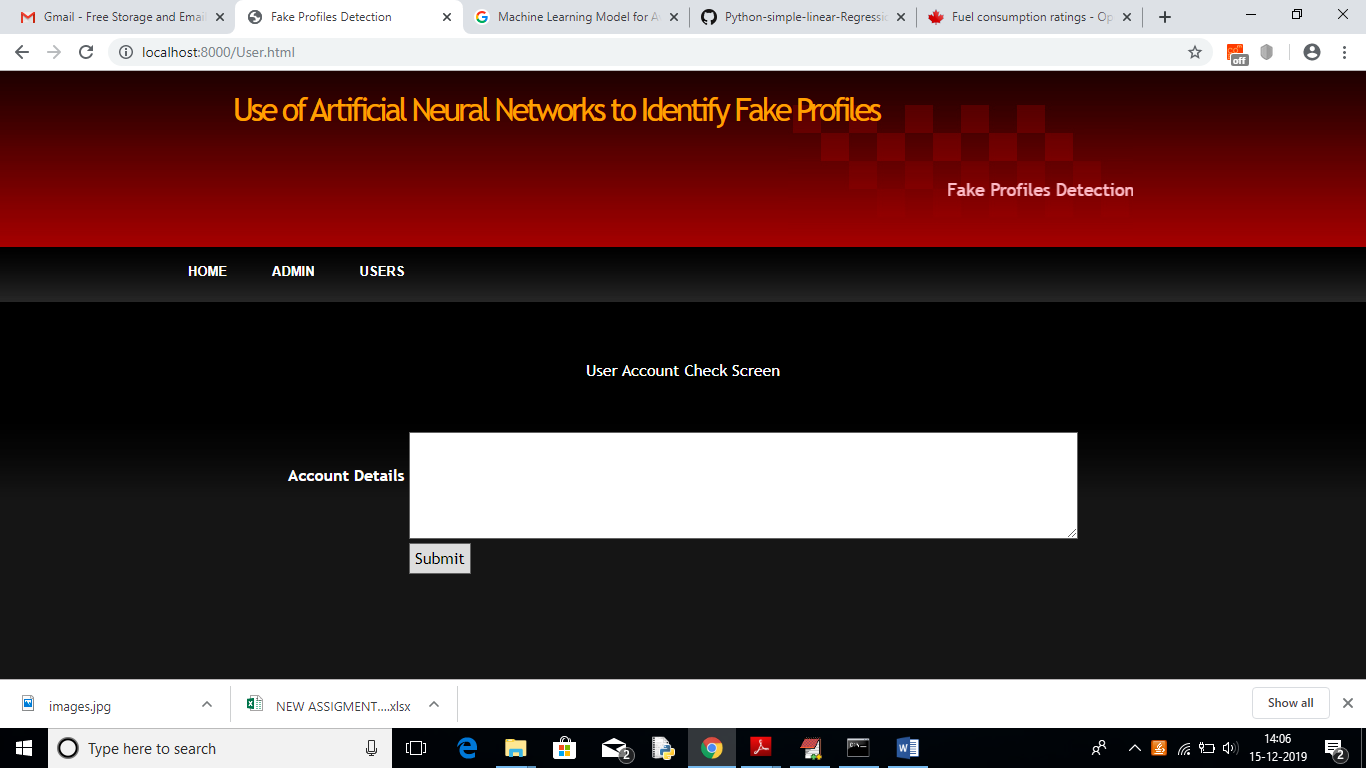
In above black console we can see all ANN details.



In above screen we can see ANN got 98% accuracy to train all Facebook profile. Now click on ‘View Ann Train Dataset’ link to view all dataset details



In above screen we can see all train data and scroll down to view all records. Now ANN train model is ready and you can logout and click on ‘User’ link to get below screen.



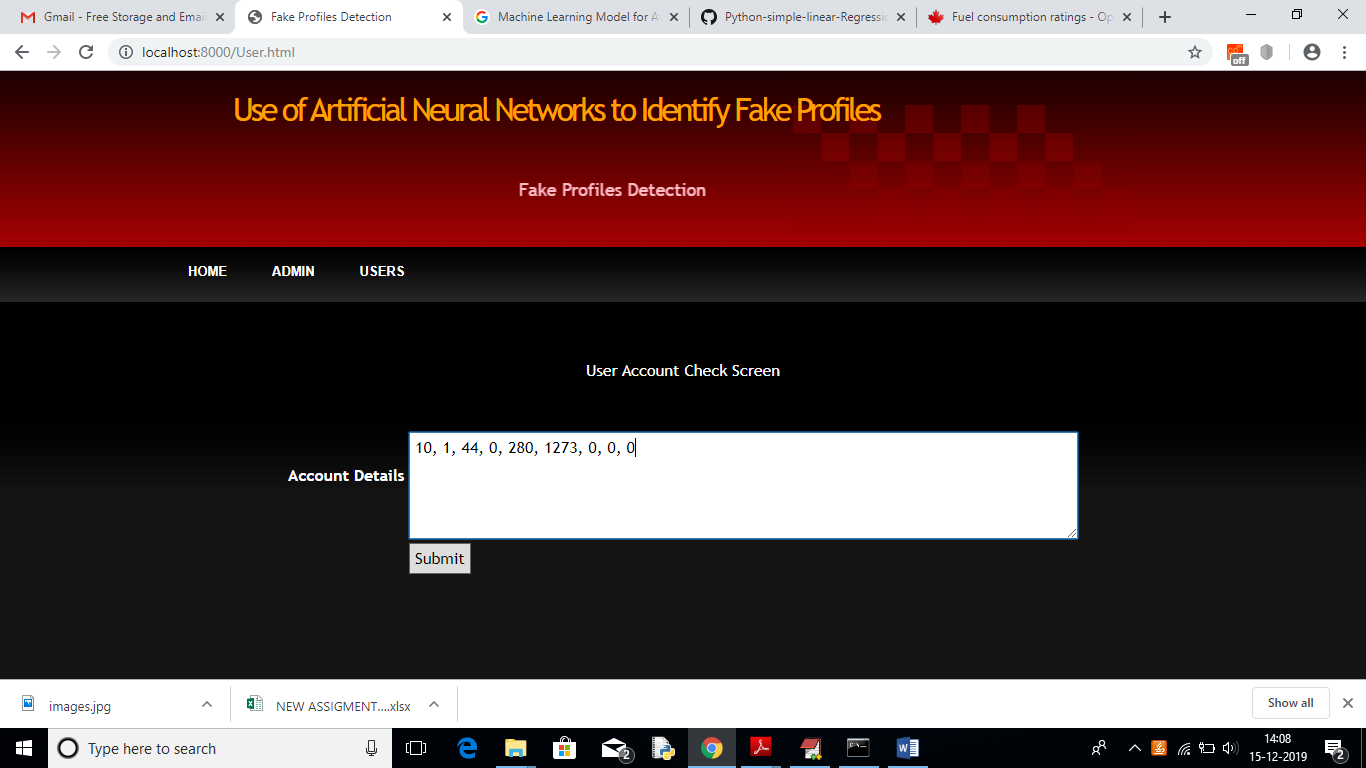
In above screen enter some test account details to get prediction/identification from ANN. You can use below records to check

10,1,44,0,280,1273,0, 0

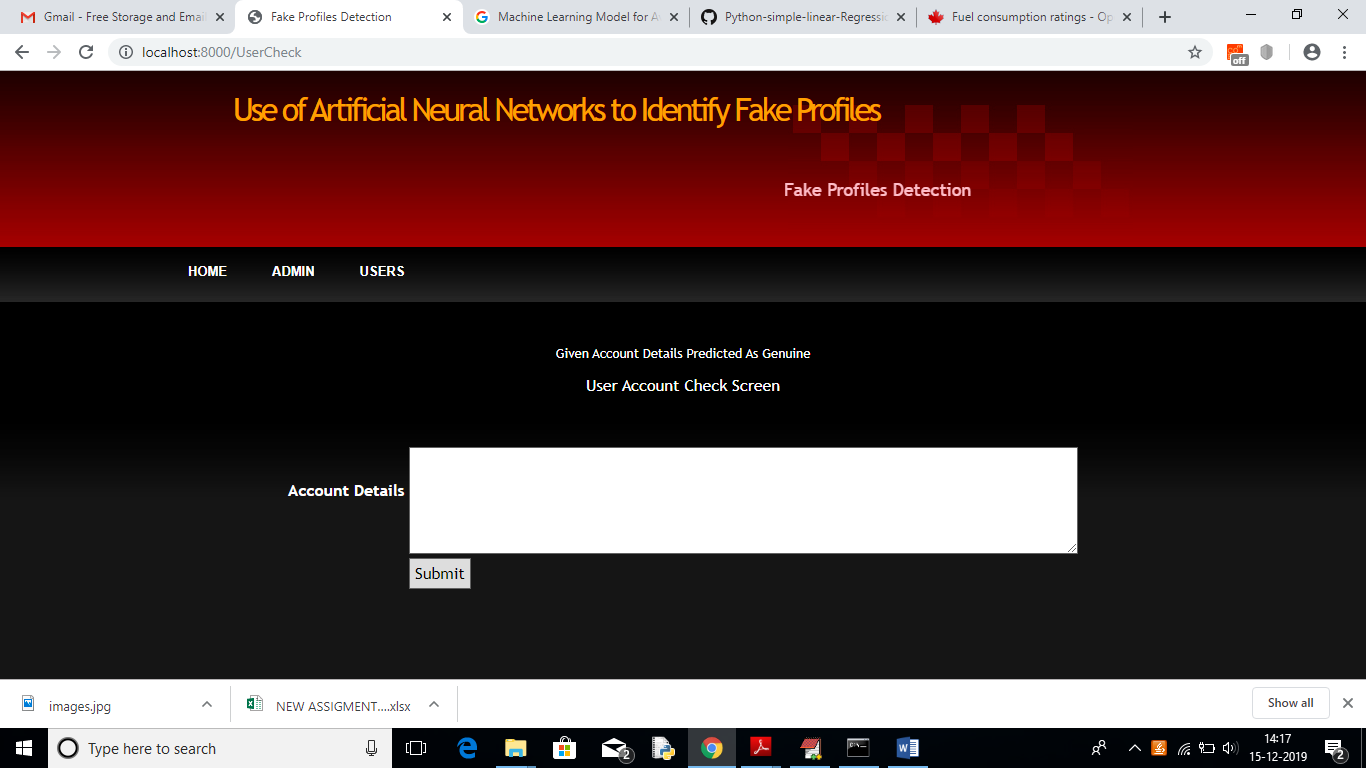
10,0,54,0,5237,241,0,0

7,0,42,1,57,631,1,1

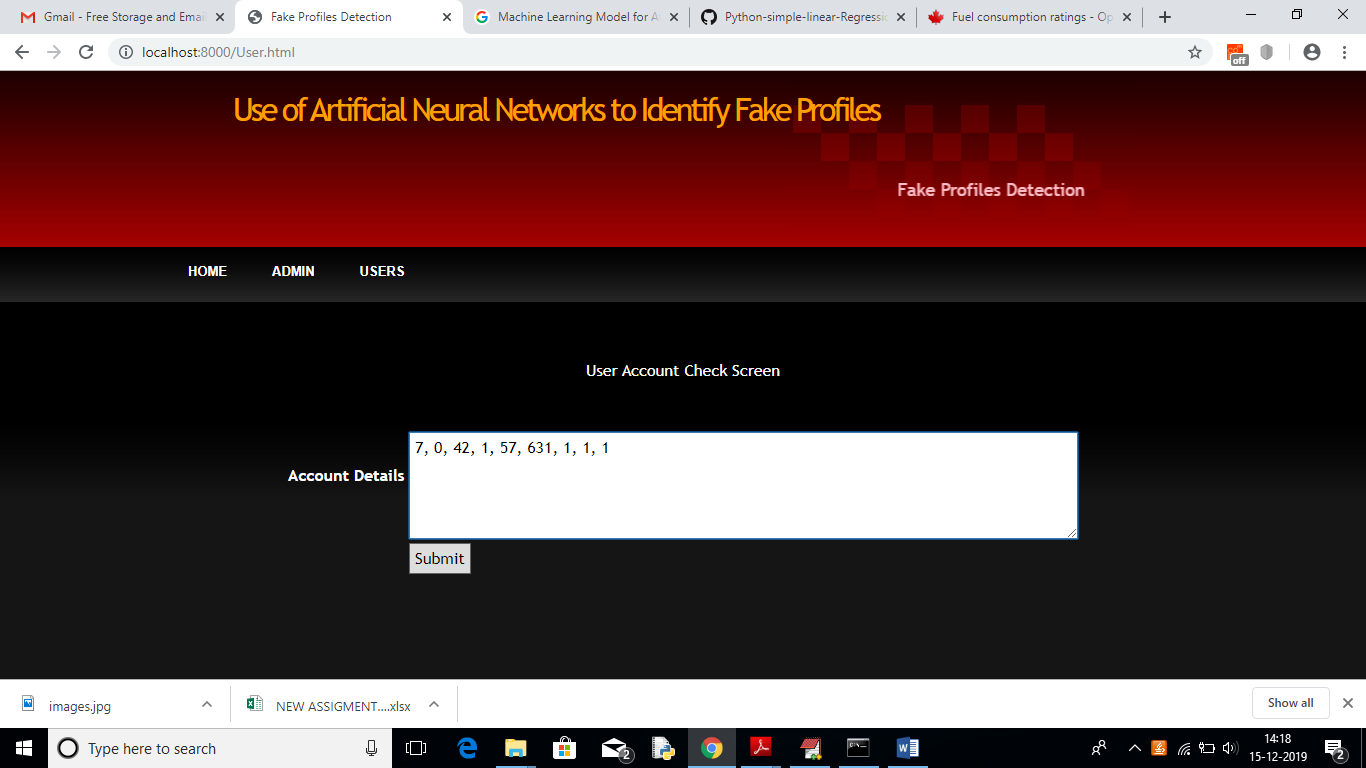
7,1,56,1,66,623,1,1



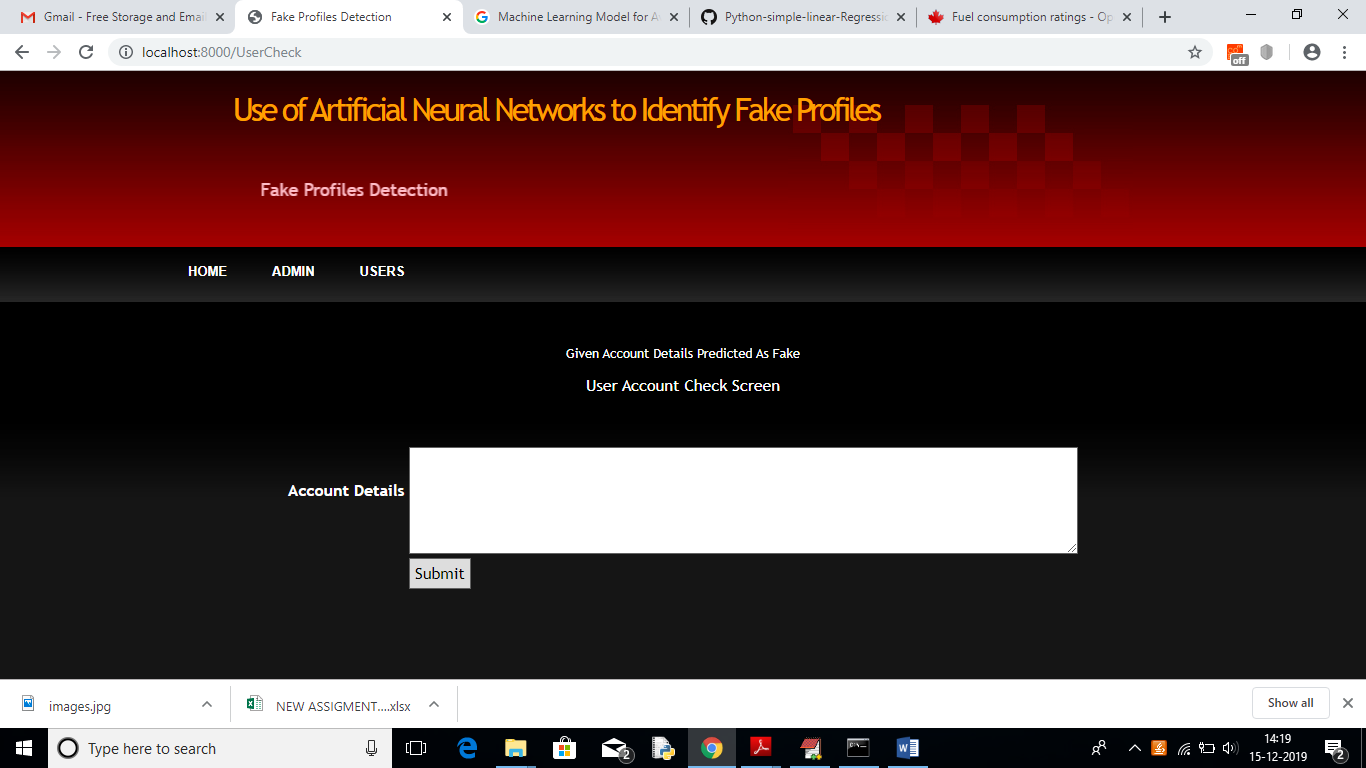
For above input will get below result



In above screen we can see the result predicted as genuine account



For above account details we got below result



In above screen we got result as fake for given account data

**8. 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 back propagation, 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.

**9. REFERENCES**

[1] https://www.statista.com/topics/1164/social-networks/

[2] https://www.cnbc.com/2018/01/31/facebook-earnings-q4-2017- arpu.html

[3] <https://www.cnet.com/news/facebook-breach-affected-50-millionpeople/>

[4] <https://www.facebook.com/policy.php>

[5] Qiang 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.

[6] 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/10.1145/2818567.2818568>

[7] Devakunchari 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, <https://doi.org/10.1016/j.compeleceng.2017.05.020>.

[8] <https://www.enigmasoftware.com/top-20-countries-the-most-cybercrime>

[9] pages.cs.wisc.edu/~bolo/shipyard/neural/local.html

[10] https://stackoverflow.com/questions/40758562/can-anyone-explain-mestandardscaler