GALGOTIAS UNIVERSITY

(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

PREVENTION OF FAKE PROFILE USING FACE RECOGNITION

A Project Report of Capstone Project - 2

Submitted by

ATUL SRIVASTAVA

(18032030061)

in partial fulfilment for the award of the degree

of

MASTERS OF COMPUTER APPLICATION

IN

COMPUTER APPLICATION

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING

Under the Supervision of

Dr. RAJU SHANMUGAM, MTech, PhD

Professor

MAY- 2020

GALGOTIAS
UNIVERSITY

SCHOOL OF COMPUTING AND SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

Certified that this project report **"PREVENTION OF FAKE PROFILE USING FACE RECOGNITION"** is the bonafide work of **"atul srivastava (18032030061)"** who carried out the project work under my supervision.

SIGNATURE OF HEAD	**SIGNATURE OF SUPERVISOR**
Dr. MUNISH SHABARWAL,	Dr. Raju Shanmugam,
PhD (Management), PhD (CS)	MTech, PhD
Professor & Dean,	**Professor**
School of Computing Science & Engineering	**School of Computing Science & Engineering**

GALGOTIAS
UNIVERSITY

SCHOOL OF COMPUTING AND SCIENCE AND ENGINEERING

CERTIFICATE

I hereby certify that the work which is being presented in the project entitled, **"PREVENTION OF FAKE PROFILE USING FACE RECOGNITION"**, in partial fulfilment of the requirements for the award of degree of **Masters of Computer Application** submitted to Galgotias University, Greater Noida, is an authentic record of my own work carried out under the supervision of **Dr. Raju Shanmugam**.

The matter presented in this project has not been submitted for the award of any other degree of this or any other university.

[Signature of Student]

Date :

Atul srivastava

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

SIGNATURE OF HEAD
Dr. MUNISH SHABARWAL,
PhD (Management), PhD (CS)
Professor & Dean,
School of Computing Science & Engineering

SIGNATURE OF SUPERVISOR
Dr. Raju Shanmugam,
MTech, PhD
Professor
School of Computing Science & Engineering

ABSTRACT

In today's world, there is a growing number of people who are intensely dependent on online social networks (OSN). People use social sites to find and make friends, to associate with people who share comparable intrigue, trade news, organize the event, exploring passion in an. According to a Facebook review, 5% of monthly active users had fake accounts, and in the last six months, Facebook has deleted 3 billion accounts. According to the Washington Post, Twitter has suspended over 1 billion suspect accounts over a day in recent months, Detection of a fake profile is one of the critical issues these days as people hold fake ac-counts to slander image, spread fake news, promotes sarcasm that has attracted cybercriminals ' in an There are numerous machine learning methodologies such as supervised learning, SVM-NN, are produced for the effective detection of a fake profile. In this paper, we proposed convolution neural networks with many artificial neural network algorithms like face recognition, prediction, classification and clustering for the efficient identification of account being real or fake and elimination of fake profile account. Furthermore, the study is grounded on the fact of the face-recognizing of the user and performing feature detection and time series prediction. If the user account detected fake it would not be created.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS/ ABBREVIATIONS

1. **CNN –** Convolutional Neural Network
2. **NN –** Neural Network
3. **ANN –** Artificial Neural Network
4. **MP –** Max Pooling
5. **Conv. –** Convolution
6. **Inp. –** Inception
7. **OSN –** Online Social Network

CHAPTER 1

INTRODUCTION AND OVERVIEW

INTRODUCTION

In today’s generation, there are growing numbers of individuals who have online social networks (OSNs) such as Facebook, Twitter, Instagram, LinkedIn, Google +. Social networks permit persons who have common interests or reasons to collaborate. It provides them with access to numerous services for example messaging, posting comments on their cyber-walls which are public, commenting on other users' profiles post and exchanging the masking of identity for malicious purposes has become progressively prevalent over the last few years. People depend intensely on OSNs to stay in touch, trade news, organize events and even run their online business. People rely heavily on online social networks (OSNs) to create and share individual personal profiles, email, images, recordings, audios, and videos, and to find and make friends that have attracted cybercriminals ' interest in carrying out a variety of malignant activities. Government associations use (OSNs) as a forum for effectively providing government driven services to people and educating and informing them about different situations. This heavy utilization of social networks results in immense measures of data being disseminated and Organizations use social networks to promote, advertise, and support their business online. Fake profile accounts show that people do not represent them as a real person. Such an account is manually opened by a person after that actions are automated by bot. Fake profile account is categorized into a Sybil account and duplicate account. A duplicate account applies to a user's account and maintained by the user other than their main account.

Classification based on Social Accounts

Fake accounts are classed into user classified (reported) or unauthorized (unwanted) groups of accounts. User malicious accounts records show individual profiles made by a client from a company or non-human element. Alternatively, undesirable accounts are however user identities that are configured to be used for violation of security and privacy. The Social networking site database for Facebook records a statistic of 4.8% for duplicate accounts, the number of user-misclassified accounts is 2.4% and the number of unauthorized accounts is 1.5 %. In 2019, Facebook announced the deletion of 2.3 billion fake profile accounts. This is almost twice as many as 1,2 billion accounts withdrawn in the first quarter of 2018. The Facebook Compliance Report shows that as much as 5 million of its monthly active users are fake and that there is a growing number of attacks. Facebook is estimated to have more than two billion monthly active users and one billion active users each day in the major online social network reports. Accordingly, only 5 percent of its active monthly users are false in Facebook reports. It is very convenient today to make false accounts. Nowadays, fake profile accounts can be purchased on the web at an extremely cheaper cost furthermore, it can be delivered to the client using publicly supporting Now it is easier to purchase followers online from Twitter and Instagram. The goal behind Sybil account formation is to Defame someone else's image, digital terrorism, terrorist propaganda; fear-based oppressor publicity, campaigns for radicalization, distribution of pornography, fraud and misinformation, popularity shaping, division of opinions, identity insecurity. In this paper, we evaluate the promptly accessible and designed methodologies that are utilized for the fruitful detection of identifying fake accounts on Facebook utilizing AI models, of human-created Detection is accomplished by observing the attitude and the needs of people by examining their experiences. Detection is accomplished by observing the attitude and the needs of people by examining their communication and interaction with each other. Then there is a need to give a fake account to our machine learning model to allow the algorithm to comprehend what a fake account is. We use artificial neural networks and convolution neural network algorithms.

Face Recognition

Face Recognition presents a difficult issue in the field of picture examination and PC vision, and has numerous applications in different sorts of utilizations today .Face Recognition systems can be separated into three classifications: strategies that work on force pictures; techniques that manage video successions; and strategies that require other tactile information, for example, [1]3D data or infra-red symbolism. In continuation of this exploration will be evaluated a few strategies for face Recognition in all classes, their benefits and downsides. Additionally, specific intrigue will be appeared about the Eigenfaces strategy. Face Recognition seems to offer a few favourable circumstances over other biometric strategies, a couple of which are plot here: Almost every one of these advances require some intentional activity by the client, i.e., the client needs to put his hand on a hand-rest for fingerprinting or hand geometry discovery and needs to remain in a fixed situation before a camera for iris or retina distinguishing proof. This is especially helpful for security and observation purposes. Moreover, information procurement when all is said in done is loaded with issues for different biometrics: procedures that depend on all fours can be rendered pointless if the epidermis tissue is harmed somehow or another (i.e., wounded or split). Voice Recognition is vulnerable to foundation clamors out in the open spots and sound-related changes on a telephone line or copying. Marks can be altered or produced. Notwithstanding, facial pictures can be handily gotten with a few modest fixed cameras. Great face Recognition calculations and suitable pre-processing of the pictures can make up for commotion and slight varieties in direction, scale and brightening. At long last, advancements that require various people to utilize a similar hardware to catch their organic qualities possibly open the client to the transmission of germs and pollutions from different clients. Be that as it may, face Recognition is absolutely non-nosy and doesn't convey any such wellbeing dangers.

Purpose

Fake accounts are classed into user classified (reported) or unauthorized (unwanted) groups of accounts. User malicious accounts records show individual profiles made by a client from a company or non-human element. Alternatively, undesirable accounts are however user identities that are configured to be used for violation of security and privacy. The Social networking site database for Facebook records a statistic of 4.8% for duplicate accounts, the number of user-misclassified accounts is 2.4% and the number of unauthorized accounts is 1.5 %. In 2019, Facebook announced the deletion of 2.3 billion fake profile accounts. This is almost twice as many as 1,2 billion accounts withdrawn in the first quarter of 2018. Facebook is estimated to have more than two billion monthly active users and one billion active users each day in the major online social network reports. Accordingly, only 5 percent of its active monthly users are false in Facebook reports. It is very convenient today to make false accounts. Nowadays, fake profile accounts can be purchased on the web at an extremely cheaper cost furthermore, it can be delivered to the client using publicly supporting Now it is easier to purchase followers online from Twitter and Instagram.

Motivation and scope

The goal behind Sybil account formation is to Defame someone else's image, digital terrorism, terrorist propaganda; fear-based oppressor publicity, campaigns for radicalization, distribution of pornography, fraud and misinformation, popularity shaping, division of opinions, identity insecurity. In this project, we evaluate the promptly accessible and designed methodologies that are utilized for the fruitful detection of preventing the creating of fake accounts on social media utilizing AI models, of human face detection and feature detection to uniquely identify each and every social media account hence eliminating existing profiles and prevention of creation of new ones.

PROBLEM IDENTIFICATION

Online Social Network (OSN) There is a growing number of people who hold Online social networks (OSNs), such as Facebook, Twitter, Instagram, LinkedIn, Google+ and hide their identity for the malicious purpose have become increasingly popular over the last few years. People use OSNs to keep in touch with each other’s, share news, organize events, and even run their e business. People are highly dependent on online social networks (OSNs) for creating and sharing personal profiles, text, pictures, audios, and videos, and for finding and making friends which have attracted the interest of cybercriminals for carrying out several malicious activities. Companies use social networks to advertise and promote their business online. Government organizations use OSNs as a platform to deliver government services to citizens in an efficient manner and to educate and inform them about various situations. Such an extensive usage of social networks leads to the dissemination of a massive amount of information.

Fake accounts people establish that aren't representative of them as a real person. Fake accounts can be either human generated, computer-generated (also referred to as "bots"), or cyborgs. A cyborg is a half-human, half-bot account. Such an account is manually opened by a human, but from then onwards the actions are automated by a bot. Fake accounts are categorized into duplicate accounts and false accounts. A duplicate account refers to an account maintained by a user in addition to his/her principal account. False accounts are further broken down into two categories user-misclassified accounts and undesirable accounts User-misclassified accounts represent the personal profiles created by users for a business, organization, or nonhuman entity. On the other hand, undesirable accounts are the user profiles that are intended to be used for purposes that violate security and privacy. Social networking site. Facebook estimates that 4.8 percent are duplicate accounts, 2.4 percent are user misclassified accounts, and 1.5 percent are undesirable accounts. The Facebook company said it removed 2.2

billion accounts in the first quarter of 2019. That’s a jump of nearly double compared to the fourth quarter of 2018 when 1.2 billion accounts were removed. According to Facebook's Enforcement Report, as many as 5% of its monthly active users are fake and the increase is due to the rise in automated attacks. Facebook is one of the largest OSNs in the world with more than 2.2 billion monthly active users and 1.4 billion daily active users. As a result, Facebook estimates that only 5% of its monthly active users are fake. The making of fake accounts is very easy nowadays. Now a day’s fake accounts can be bought online at a very less cost and can be given to the customer via crowdsourcing services. Now a day it is easier to buy Twitter and Instagram followers and likes online. The purpose of creating fake accounts is to defame the character of another person, online extremism; terrorist propaganda; and radicalization campaigns, spreading rumours and false news, influencing popularity, polarizing opinions, identity theft, cyberbully, dissemination of pornography and fraud.

FEASIBILITY STUDY

Preliminary investigation Examine project feasibility, the likelihood the system will be useful to the organization. A Feasibility Study is a formal project document that shows results of the analysis, research and evaluation of a proposed project and determines if this project is technically feasible, cost-effective and profitable. The primary goal of feasibility study is to assess and prove the economic and technical viability of the business idea. The main objective of feasibility study is to test the Technical, Operational and Economical feasibility for adding new modules and debugging old running system.

Different types of feasibility studies that can be undertaken:

- Technical feasibility.
- Economic feasibility.
- Operational feasibility.
- Schedule feasibility.

Technical Feasibility

Technical feasibility is concerned with specifying the equipment and the software to satisfy the user requirements. The technical needs of the system vary considerably but might include:

- The facility to produce outputs in a given time.
- Response time under certain conditions.
- Ability to process a certain volume of transactions at a specified speed.
- Facility to communicate data to a distant location.

In technical feasibility, the configuration of the system is given more importance than the actual hardware. The configuration should provide the complete picture of the system requirements:

- What speeds of input and output should be achieved at particular quality of printing?
- How these units are interconnected so that they would operate smoothly?

Economic Feasibility

Economic analysis or cost/benefit analysis is the most frequently used technique for evaluating the effectiveness of a proposed system. It is the procedure to determine the benefits and saving that are expected from a proposed system and compare them with cost. If benefits outweigh cost, a decision is taken to design and implement the system. Otherwise, further justification or alternative in the proposed system will have to be made if it is to have a chance of being approved. This is an ongoing effort that improves in accuracy at each phase of the system life cycle.

Operational Feasibility

Operational feasibility covers two aspects. One technical performance aspect and the other is the acceptance within the organization. The points to be considered are:

- What changes will be brought with the system?
- What organization structures are disturbed?
- What new skills will be required?
- Do the existing staff members have these skills? If not, can they be trained in due course of time?

Operational feasibility determines how the proposed system will fit within user.

Schedule Feasibility

The duration of time required for the project has been planned appropriately and it is the same as the duration of time expected by the users, therefore the application can be delivered to the users within the expected time duration, satisfying the users. Hence the project is feasible in scheduling.

Feasibility of This Project

According to the definition of technical feasibility the compatibility between the various components of the system is very important. In our project the compatibility of both is very good. The compatibility of face recognition and database server is very good. The speed of output is very good, when we enter the data and click button then the response time is very fast and give result very quickly. The training time are comparably fast to other systems. The system is reliable.

The system can double as a middleware server and prevents fake profile creation providing better online social media experience. Hence, it is quite beneficial way of providing meaningful service. Operational feasibility of the project also exists because in today’s world most of the people are using online social media and many people create sybil accounts. There is nothing complex in the system that cannot used be used by user. Thus, we can say that this project can be integrated with online social media platforms for the intended purpose.

CHAPTER 2

LITERATURE SURVEY

In this section, we summarize some of the related work done in the field of detecting fake profile accounts using machine learning models.

1. **Sarah Khaled, Neamat El-Tazi and Hoda M. O. Mokhtar - "Detecting Fake Accounts on Social Media"**

   They focus on technologies adopted for detect fake accounts and bots in twitter. Classification algorithms of machine learning have been utilized to choose real or fake targ et accounts, those algorithms were support vector machine (SVM), Neural Network (NN). They proposed a new algorithm support vector machine-neural networks for the successful detection of fake accounts. Both the approaches adopt machine learning techniques feature selection and dimension reduction techniques were applied. It was also notices that correlation set records a remarkable accuracy among the other feature selection technique sets as it removes redundancy. The new algorithm classified 98% of the account of training dataset using fewer features.

2. **Mudasir Ahmad wania, Nancy Agarwala & Suraiya Jabinb Syed Zeeshan Hussainb - "Analysing Real and Fake users in Facebook Network based on Emotions"**

   In this paper, the author mainly focuses on fake profile detection using sentiments. The study is done on the post of real account user and fake profile user and similar emotions they use. The experiment is done on Facebook user profile post. Data are trained for 12 emotions, including 8 including Plutchik's eight basic emotions, positive and negativizes, by the use of machine training techniques consequently, outliers are removed using noise removal technique. To train the detection model, many machine learning techniques including

Support Vector Machine (SVM), Naïve Bayes, J Rip and Random Forest have been used. The author concluded that three emotion categories, fear, surprise and trust are found least in the posts of fake users. For all three measures, precision, estimation and AUROC, Random forest provides the best result.

3. **ESTÉE VAN DER WALT and JAN ELOFF - "Using Machine Learning to Detect Fake Identities: Bots vs Humans"**

   This paper focuses on the detection of fake identities of human's vs bots using machine learning models. Numerous fake accounts are enhanced with features used to detect bot accounts, and the collection of features has been extended to different supervised learning models. Human accounts and bot accounts have alike attributes and characteristics. For example: name. It shows, the engineered features used to detect bot accounts have failed for the successful detection of human accounts. The predictive result of the trained machine models was only 49. 75 per cent of the best F1 performance. This is due to the fact that human beings are distinct from both in terms of behavior and characteristic, which cannot be modelled in the same way.

4. **Gayathri A , Radhika S & Mrs. Jayalakshmi S.L - "Detecting Fake Accounts in Media Application Using Machine Learning"**

   Identification of fake accounts in media application by using support vector machines and neural networks. In this paper, they represent a machine learning pipeline to detect fake accounts rather than using prediction for each account. It classifies cluster of Sybil accounts whether they are created by the same individual. The process starts with the selection of profile to be tested then extracting required features and passes them to a trained classifier which classifies account being fake or real along with the feedback. Future work: to use more sophisticated clustering algorithms such as k-means or hierarchical clustering. The other line of research is to predict multi-model utilizing the feature sets used in other spam detection models.

5. **Caruccio L., Desiato D. & Polese G. - "Fake Account Identification in Social Networks"**

   The authors have proposed a method for sleuthing and erasing false records utilizing RFDs (Relaxed Functional Dependency). It grants to feature a few kinds of connections in information. In addition, a RFD can hang on a connection example are if at whatever point two tuple sets (t1, t2) in are, fulfil the similitude imperative on the LHS, they subsequently fulfil the likeness requirement on the RHS. Conditions for which no tuple sets are fulfilling the likeness limitation on the LHS are named key conditions. They are utilizing a substitution strategy to precisely segregate fake records, that abuses calculations for removing RFDs from the data keep inside the in-formal organizations itself. Their trial examination has indicated that it's possible to adequately recognize artificial from human records bolstered the arranged method. This procedure has permitted the U.S. to look out important examples portraying human conduct, which makes it irksome for false records to thoroughly imitate it, in spite of the fact that they appear to be much the same as genuine records.

CHAPTER 3

PROPOSED METHODOLGY

SYSTEM DESCRIPTION

The approach this system takes for fake profile detection is too limiting each user to one and only one account on a particular platform. The system utilized facial recognition using Neural Networks to create a unique facial print id of the account creator to identify him/her. This helps us to uniquely identify the customer by bind the facial print to the account eliminating the possibility for a user to create a new fake account. The facial data is utilized to identify the account creator with existing account holders. This eliminates the possibility of creating new fake accounts by any user. The input data after face detection and preprocessing are supplied to the Inception NN for Face Recognition and Classification. The data processed by the convolutional network by passing through various layers of convolutions, Filters, pooling/sampling and finally classified through a fully connected layer. The various steps of the proposed system are described below:

Data Collection

The data is collected from the user. This is collected from the social media platform and then passed to the system. This step collects two types of data firstly the profile details like name, age and facial data from the sensors for face recognition. This data is used to process and identify whether the user that is trying to create an account is genuine or not.

Data Optimization

The raw data collected in the collection phase is optimized into the format required. The optimization is one of the important steps before the processing of the data as it prepares the data and enhances the data so that during the processing of data the algorithm can produce better results.

Face Detection

Detection involves the detection of the facial data from the collected data and identifies the points required to be processed and specific to the face rather than the whole picture. The detection phase involves facial detection by identifying the points that are of use to us and eliminating the rest unnecessary points. This is done through template matching function. It defines a standard template for all the faces and where the different features can be identified independently like eyes. nose, mouth, contour, etc.

Blink Detection

Detection involves the detection of the eye blink movements of the user. This is implemented as a security feature for the users so that the user cannot verify the profile by holding a image of another user or using a very high resolution poster to mimic another users identity.

Face Recognition

Face Recognition is performed using the Deep Neural Network known as Inception Network. The data is pre-processed before feeding to the convolutional network. CNN utilizes various hidden layers of convolution, and pooling. These layers are arranged in some fashion repeatedly to form the network. After passing through the various hidden layers the output is put to the fully connected layer of the classifier for classification. The data from the output layer is put to comparison by the dataset using SVM and KNN for user detection.

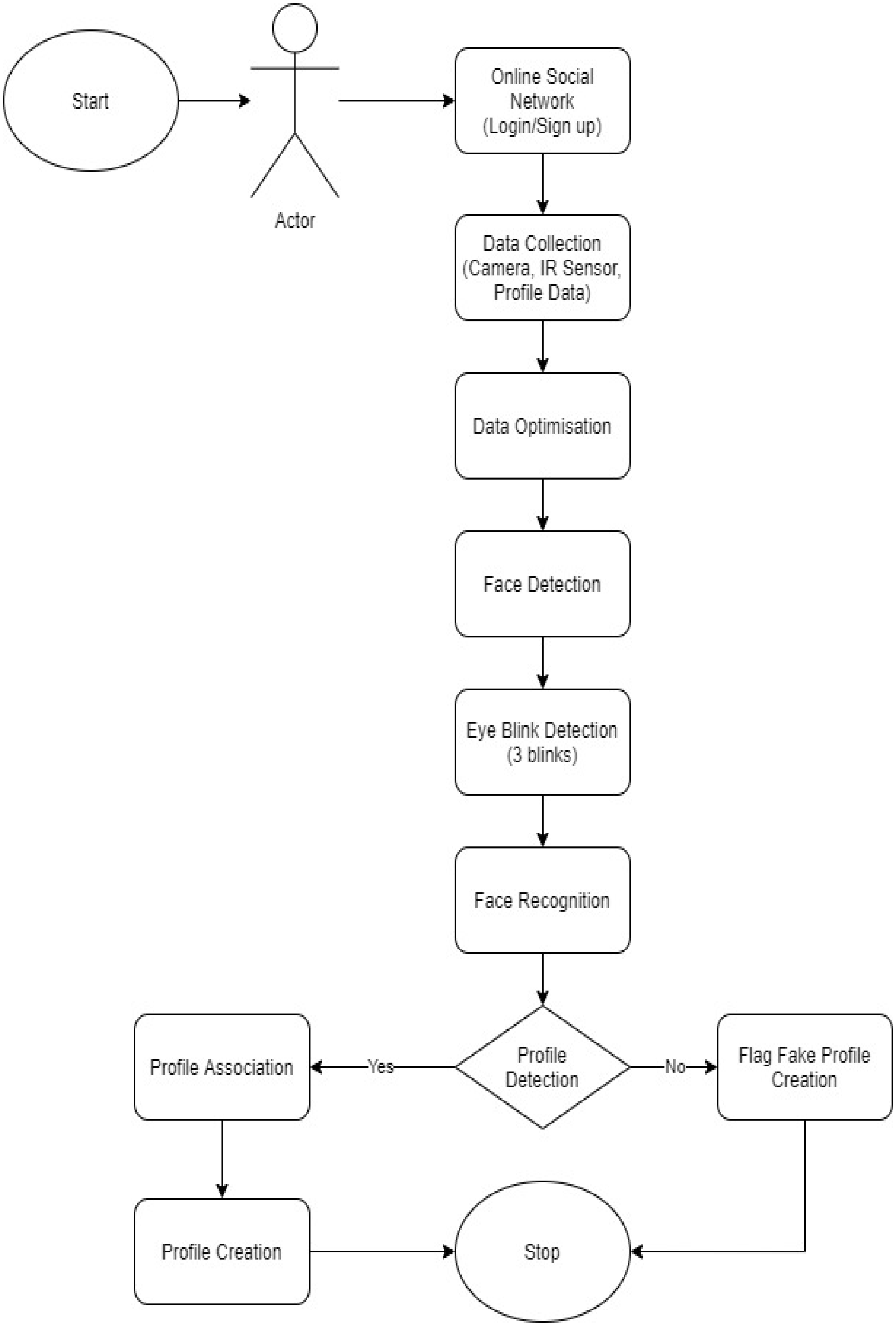
Profile Detection

This involves the main comparison of the data from the data warehouse that utilizes snowflake schema to store the data. The data after the process of classification is passed to this phase. The data is compared with the data in feature data in the data warehouse. If a match is found then the profile is flagged as fake and the account is not created. If the face is unique then the profile is created else it is detected as a fake verification. In the end, if the test is accepted then the profile is created and the facial data is moved to the training.

Profile Association

The data is associated with the user profile once it is created,4. The data is added to the database creating a unique identification of every use and also maintains the data to which all platforms the user is signed up to. This eliminates the possibility of creating a fake profile of the user on a platform he/she is not utilizing.

Figure 1. Flowchart of the Proposed System


Start
Actor
Online Social Network (Login/Sign up)
Data Collection (Camera, IR Sensor, Profile Data)
Data Optimisation
Face Detection
Eye Blink Detection (3 blinks)
Face Recognition
Profile Detection
Yes
No
Profile Association
Flag Fake Profile Creation
Profile Creation
Stop

ALGORITHMS

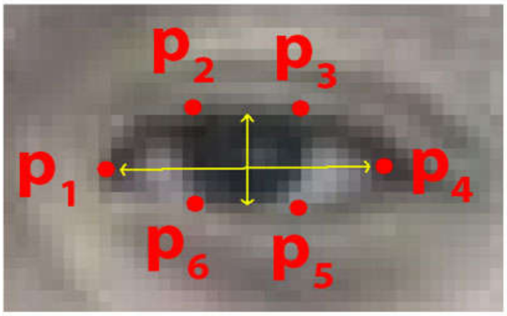
An algorithm is a set of steps for a computer program to accomplish a task. Algorithms put the science in computer science. And finding good algorithms and knowing when to apply them will allow you to write interesting and important programs. The major algorithms involved are:

Face Detection

The algorithm used is Facial landmark detection. Landmark detection starts with face detection, finding faces in the image and their extents (bounding boxes). Facial detection has long been considered a solved problem, and OpenCV contains one of the first robust face detectors freely available to the public. In fact, OpenCV, was primarily known and used for its fast face detection feature, implementing the canonical Viola-Jones boosted cascade classifier algorithm and providing a pre-trained model. shape-predictor: This is the path to dlib's pre-trained facial landmark detector.

Eye Blink Detection

The Histogram of Oriented Gradients method suggested by Dalal and Triggs in their seminal 2005 paper, Histogram of Oriented Gradients for Human Detection demonstrated that the Histogram of Oriented Gradients (HOG) image descriptor and a Linear Support Vector Machine (SVM) could be used to train highly accurate object classifiers — or in their particular study, human detectors.



Each eye is represented by 6 (x, y)-coordinates, starting at the left-corner of the eye and then working clockwise around the remainder of the region.

$$\text{EAR} = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|}$$

Where p1, …, p6 are 2D facial landmark locations. The numerator of this equation computes the distance between the vertical eye landmarks while the denominator computes the distance between horizontal eye landmarks, weighting the denominator appropriately since there is only one set of horizontal points but two sets of vertical points.

Face Recognition

The various model and algorithms used for face recognition are described below.

Triplet Loss Function

Triplet Loss is a loss function for neural networks where a baseline input is compared to a positive input and a negative input. The distance from the baseline input to the positive input is minimized, and the distance from the baseline input to the negative input is maximized. The Triplet Loss is prepared on Triplet {Xa, Xp, Xn} where Xa is a picture of a particular individual (anchor), Xp is the other picture of a similar individual (positive) and Xn is the picture of some other individual (negative). The function ensures that Xa is nearer to Xp than Xn inside a specific edge α. We are utilizing hard triplets for getting hard-negatives and hard-positives as most triplets are trifling to learn and won't give any noteworthy improvement in learning. Photos of a similar individual in various clothing types or postures (hard-positives) or seeing comparable looking yet really changed individuals (hard-negatives) helps in distinguishing the capacity. As the dataset becomes bigger, the quantity of triplets additionally develops along these lines rendering credulous execution of triplet misfortune illogical. In OpenFace, if a negative pair isn't found inside edge, the triplet is discarded. In our methodology, we adjust the triplet misfortune by picking the pair that outcomes in the nearest edge as could be expected under the circumstances α rather than disposing of it. In our triplet loss approach does not consider α as the absolute threshold upon which a negative pair is discarded. The final equation is described in eq 1.

$$\| f(x_i^a) - f\left(x_i^p\right) \|_2^2 < \| f(x_i^a) - f(x_i^n) \|_2^2$$

OpenFace Net

OpenFace implements Triplet Loss which basically arrange input images into triplets and then select certain triplets that falls under the threshold margin. Such implementation will generally discard many triplets from being used in training, in which case those triplets may result in additional performance improvement if utilized properly. OpenFace utilizes Dlib for identifying face district in a picture and results in a case encompassing each face which can be under various postures. This presents a potential issue if quickly utilized as contribution for the neural system and accordingly should be pre-processed. OpenFace utilizes 2D relative change as its pre-processing technique which sets the nose and eye corners moderately near mean areas by resizing and editing pictures to edges of the tourist spots delivered by Dlib face indicator. The aftereffect of this change is a standardized picture in 96 x 96 pixels. The standardised pictures are then taken care of into the system to produce embeddings. These embeddings are mapped into triplets, prepared by triplet misfortune capacity, and produces inclination which is backpropagated through the mapping. The prepared system model a while later can be utilized as a major aspect of the face acknowledgment structure to produce embeddings and afterward arranged. We utilize a similar system model as OpenFace19 does which depends on FaceNet's nn4 model.

The network definition is portrayed in Table 1 underneath. The inception layers com-prise of various convolutions and poolings. The kernel or filter size and stride for each layer are characterized as determined in Table 1. N-Reduction indicates what number of 1x1 convolutions being utilized in every Inception Layer to diminish the dimensionality before getting into the genuine convolutions. N-Convolution and N-Pooling determines what number of layers of convolutions and poolings are being performed. Max Pooling will essentially pick the biggest incentive inside the region determined by the filter, while L2 pooling takes normal of all qualities inside the region.

Table 1. Inception Network Definition

Layers		**Filter Size**	**Stride**	**N Reduction**	**N Convolution**	**N Pooling**
Convolution 1		7	2			
Max Pooling and Normalization		3	2			
Inception 2	Convolution	3	1			
Max Pooling and Normalization		3	2			
Inception 3A	Convolution 1	1	1		64	
	Convolution 2	3	1	96	128	
	Convolution 3	5	1	16	32	
	Max Pooling	3	1			32
Inception 3B	Convolution 1	1	1		64	
	Convolution 2	3	1	96	128	
	Convolution 3	5	1	32	64	
	L2 Pooling	3	1			64
Inception 3C	Convolution 1	3	2	128	256	
	Convolution 2	5	2	32	64	
	Max Pooling	3	2			
Inception 4A	Convolution 1	1	1		256	
	Convolution 2	3	1	96	192	
	Convolution 3	5	1	32	64	
	L2 Pooling	3	1			128
Inception 4E	Convolution 1	3	2	160	256	
	Convolution 2	5	2	64	128	
	Max Pooling	3	2			
Inception 5A	Convolution 1	1	1		256	
	Convolution 2	3	1	96	384	

	L2 Pooling	3	1			96
Inception 5B	Convolution 1	1	3		256	
	Convolution 2	3	1	96	384	
	Max Pooling	3	1			256
Average Pooling		1	1			
Fully Connected		1	1			
L2 Normalization		1	1			

Support Vector Machine (SVM)

Support Vector Machines (SVM) are one of the most valuable strategies in order issues. One clear model is face recognition. Nonetheless, SVM can't be applied when the component vectors characterizing tests have missing sections. An order calculation that has effectively been utilized right now the all-known Support Vector Machines (SVM) , which can be applied to the first appearance space or a subspace of it got in the wake of applying a part extraction methodology The advantage of SVM classifier over customary neural framework is that SVMs can achieve better theory execution.

k- Nearest Neighbour

k-NN is one of the foremost basic classification algorithms in machine learning. It belongs to the supervised learning class of machine learning. k-NN is usually employed in search applications wherever you're looking for “similar” things. The way we measure similarity is by making a vector illustration of the things, and then compare the vectors using an acceptable distance metric (like the geometrician distance, for example). It is typically utilized in data processing, pattern recognition, recommender systems and intrusion detection.

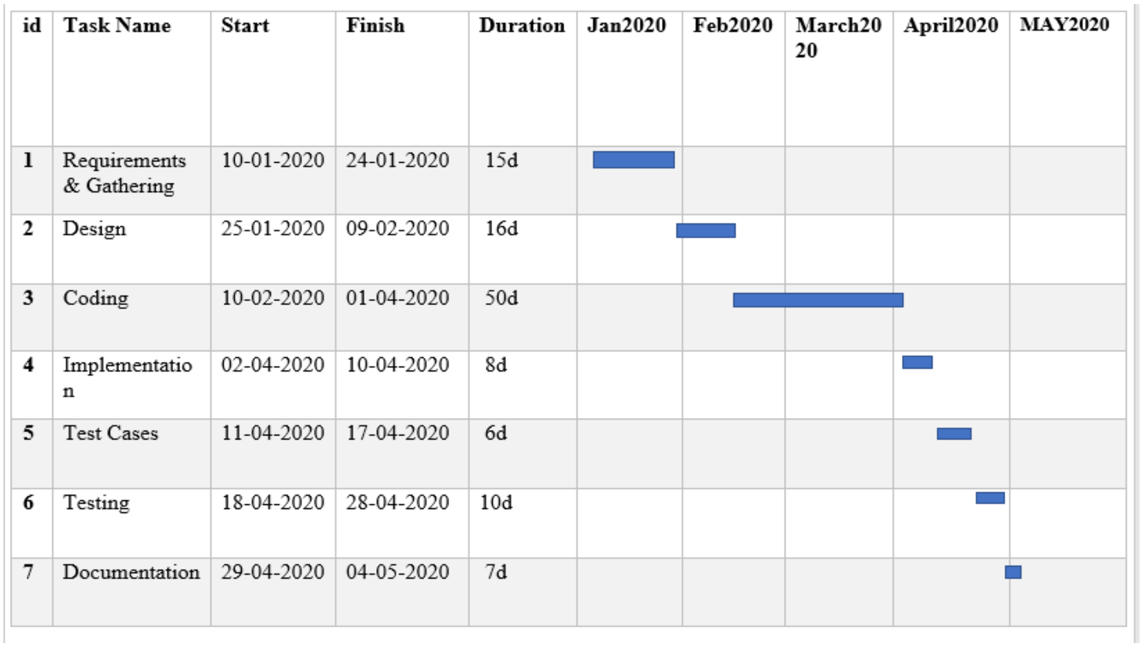
CHAPTER 4

SYSTEM ANALYSIS

GANTT CHART

A Gantt chart is a graphical representation of the duration of tasks against the progression of time. A Gantt chart is a useful tool for planning and scheduling projects. Gantt chart is a project scheduling technique. Progress can be represented easily in a Gantt chart, by colouring each milestone when completed. The project will start in the month of January and end after 4 months at the start of May. A Gantt chart is a horizontal bar chart developed as a production control tool in 1917 by Henry L. Gantt, an American engineer and social scientist. Frequently used in project management, a Gantt chart provides a graphical illustration of a schedule that helps to plan, coordinate, and track specific tasks in a project. A standard technique employed in recent times to keep track of a project's progress is the Gantt chart. They are easy to draw, easy to understand and readily adaptable to other planning approaches

Figure 2. GANTT Chart

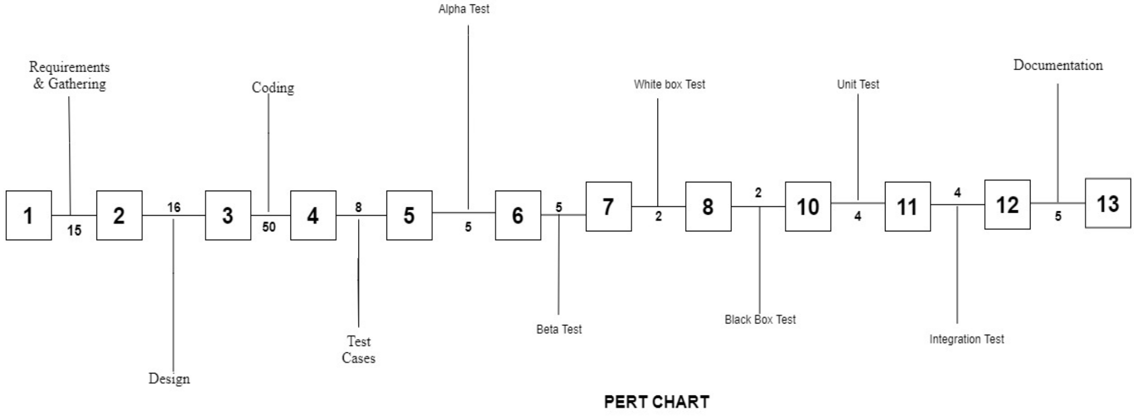


id	Task Name	Start	Finish	Duration	Jan2020	Feb2020	March2020	April2020	MAY2020
1	Requirements & Gathering	10-01-2020	24-01-2020	15d					
2	Design	25-01-2020	09-02-2020	16d					
3	Coding	10-02-2020	01-04-2020	50d					
4	Implementation	02-04-2020	10-04-2020	8d					
5	Test Cases	11-04-2020	17-04-2020	6d					
6	Testing	18-04-2020	28-04-2020	10d					
7	Documentation	29-04-2020	04-05-2020	7d					

PERT CHART

A PERT chart is a project management tool used to schedule, organize, and coordinate tasks within a project. PERT (stands for Program Evaluation Review Technique), a methodology developed by the U.S. Navy in the 1950s to manage the Polaris submarine missile program. A PERT chart presents a graphic illustration of a project as a network diagram consisting of numbered nodes (either circles or rectangles) representing events, or milestones in the project linked by labelled vectors (directional lines) representing tasks in the project. The direction of the arrows on the lines indicates the sequence of tasks. A PERT chart presents a graphic illustration of a project as a network diagram consisting of numbered nodes (either circles or rectangles) representing events, or milestones in the project linked. The PERT chart is sometimes preferred over the Gantt chart, another popular project management charting method, because it clearly illustrates task dependencies. On the other hand, the PERT chart can be much more difficult to interpret, especially on complex projects. Frequently, project managers use both techniques.

Figure 3. PERT Chart



FUNCTIONAL REQUIREMNTS

General Requirements

There are some of the basic requirements which we need to maintain in our system they are listed below:

- The User must not be Registered in the System and should not have account on that social media platform.
- The User should be able to verify itself at the time of account creation.
- Fake Profile Creation Attempts should be flagged
- The social media platform can request for fake profile creation reports.

User-Interface Requirement

- User Interface elements must be easy to understand.
- Image should be easily captured and in well light rooms.
- The user interface should be easy to learn. When the people use the user interface, they should know which element is used for which operations.
- The interface actions and elements should be consistent. When the required person presses any button, required actions should be done by the system.

Software and Hardware Requirements

Software Specification

Software	Minimum Requirements
Operating System	Windows 8, 8.1, 10
Language Requirements	Python 3.6, TensorFlow Framework 1.13rc1
Library Requirements	OpenCv, Numpy, SCIpy, imutils, Dlib, Keras TensorFlow, Scikit, Imageio, Matplotlib
Documentation Tools	MS Word 2019

Hardware Specification

Hardware	Minimum Requirements
Processor	3.4 GHz or more Dual Core Processor
RAM	4 GB or more
Storage Space	20 GB or more
GPU	Nvidia 1060Ti or newer
Camera	sRGB camera 1MP 720p

NON-FUNCTIONAL REQUIREMNETS

Performance Requirements

To achieve good performance the follow in requirements must be satisfied

- *Scalability:* The ease with which a system or component can be modified to fit the problem area.
- *Portability:* The ease with which a system or component can be transferred from one hardware or software environment to another.
- *Security:* It is the ideal state where all information can be communicated across the internet / application secure from unauthorized persons being able to read it and/or manipulate it. It is also the process of preventing and detecting unauthorized use of one's mobile.
- *Maintainability*: The ease with which a software system or component can be modified to correct faults, improve performance, or other attributes, or adapt to a changed environment.
- *Reliability:* The ability of a system or component to perform its required functions under stated conditions for a specified period of time.
- *Reusability:* The degree to which a software module or other work product can be used in more than one computing program or software system.

Safety Requirements

Database is an important aspect of any system. So, it is required to take back up of the database. Special exception handling mechanism should be in place to avoid system error. In case scenarios where data integrity can be compromised, measures should be taken to ensure that all changes are made before the system shuts down.

Software Quality Attribute

- *Functionality:* The capability to provide functions which meet stated and implied needs when the software is used.
- *Reliability:* The capability to be maintains a specified level of performance.
- *Usability:* The capability to be understood, learned and used.
- *Efficiency:* The capability to provide appropriate performance relative to the amount of resources used.
- *Maintainability:* The capability to modified for the purpose of making corrections and improvement
- *Portability:* The capability to adopted for different specified environments without applying actions or means other than those provided for this purpose in the product.

RISK ANALYSIS

Uncertainty, which is constantly present in our daily lives, frequently impacts our decisions and actions. When we talk about risk, we normally mean the chance that some undesirable impact will occur. Hence, we normally seek to avoid or minimize risk. If there is a chance of rain, and we don't want to get wet, we may choose to stay indoors -- avoiding that risk -- or we may take an umbrella to minimize the impact of rain upon us. Uncertainty can impact our decisions and actions in desirable as well as undesirable ways. In risk analysis we usually focus on what can go wrong -- the outcomes that represent loss or damage -- although an effective analysis will also help us understand what can go right as well.

A risk assessment involves evaluating existing physical and environmental security and controls and assessing their adequacy relative to the potential threats of the organization. A business impact analysis involves identifying the critical business functions within the organization and determining the impact of not performing the business function beyond the maximum acceptable outage. Types of criteria that can be used to evaluate the impact include: customer service, internal operations, legal/statutory and financial.

A primary objective of business recovery planning is to protect the organization in the event that all or part of its operations and/or computer services is rendered unusable. Each functional area of the organization should be analyzed to determine the potential risk and impact related to various disaster threats.

Regardless of the prevention techniques employed, possible threats that could arise inside or outside the organization need to be assessed. Although the exact nature of potential disasters or their resulting consequences are difficult to determine, it is beneficial to perform a comprehensive risk assessment of all threats that can realistically occur to the organization. Regardless of the type of threat, the goals of business recovery planning are to ensure the safety of customers, employees and other personnel.

Uncertainty can arise in several ways:

- If the quantity we'd like to know is a competing firm's planned product price, uncertainty arises from our lack of knowledge: The price may be well known to that firm's employees, but it's unknown to us.
- If the quantity is market demand for products like ours, uncertainty arises from the complexity of the process: Demand depends on economic factors, fashions and preferences, and our and other firms' actions -- and even if we knew all of these, we couldn't fully calculate their net impact on final demand.
- If the quantity is a material thickness in nanometres, uncertainty may arise from limits on our ability to measure this physical quantity. We may also have limits on our ability to control fabrication of the material.

Many processes that we want to model -- from the failure rate of an electronic component to the behaviour of a macromolecule -- have inherent randomness for all intents and purposes.

CHAPTER 5

SYSTEM DESIGN

SOFTWARE ENGINEERING PARADIGM

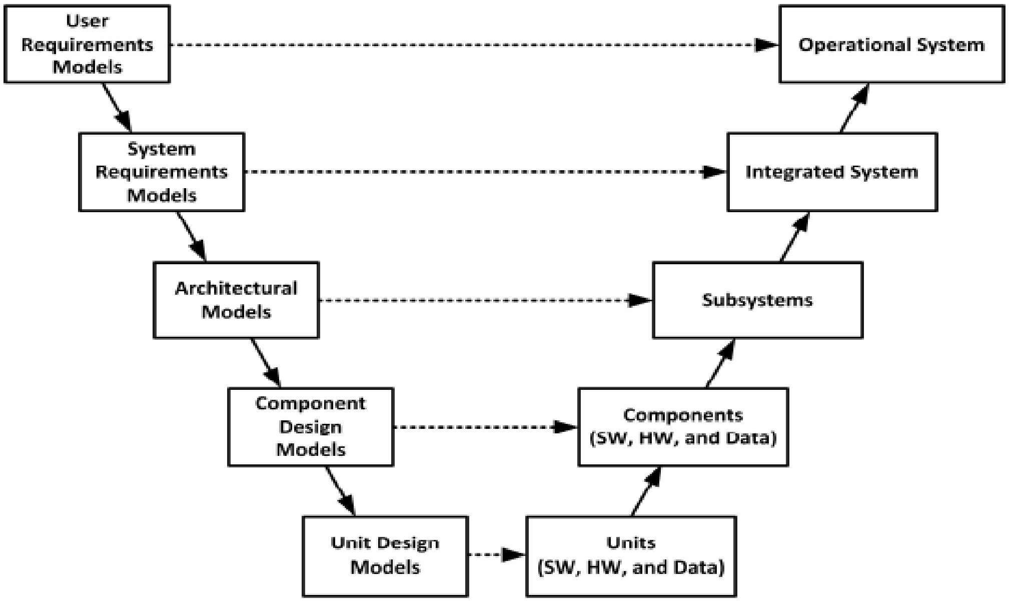
Software engineering is a layered technology. The foundation for software engineering is the process layer. Software engineering processes the glue that holds the technology layers together and enables ratios and timely development of computer software/mobile application. Process defines a framework for a set of key process areas that must be established for the effective delivery of software engineering technology. Software engineering methods provide the technical how-to's for building software. Methods encompass a broad array of tasks that include requirements analysis, design, program construction, testing, and support. Software engineering tools provide automated or semi-automated support for the process and the methods. When tools are integrated so that information created by one tool can be used by another tool, a system for the support of software development, called computer-aided software engineering is established.

System Life Cycle

To solve actual problems in an industry setting, a software engineer or a team of Engineers must incorporate a development strategy that encompasses the process, methods, and tools layers. This strategy is often referred to as a process model or a software engineering paradigm. A process model or a software engineering is chosen based on the nature of the project and application, the methods and tools to be used, and the controls and deliverables that are required.

In this project V-model is used, involved steps given below:

Figure 4. V-MODEL Illustration



In software development, the V-model represents a development process that may be considered an extension of the waterfall model, and is an example of the more general Vmodel. Instead of moving down in a linear way, the process steps are bent upwards after the coding phase, to form the typical V shape. The V-Model demonstrates the relationships between each phase of the development life cycle and its associated phase of testing. The horizontal and vertical axes represent time or project completeness (left-to-right) and level of abstraction (coarsest-grain abstraction uppermost), respectively.

Verification phases

- **Requirements analysis**

  In the requirements analysis phase, the first step in the verification process, the requirements of the system are collected by analysing the needs of the user(s). This phase is concerned with establishing what the ideal system has to perform. However, it does not determine how the software will be designed or built. Usually, the users are interviewed and a document called the user requirements document is generated.

- **System design**

  Systems design is the phase where system engineers analyze and understand the business of the proposed system by studying the user requirements document. They figure out possibilities and techniques by which the user requirements can be implemented. If any of the requirements are not feasible, the user is informed of the issue. A resolution is found and the user requirement document is edited accordingly.

- **Architecture design**

  The phase of the design of computer architecture and software architecture can also be referred to as high-level design. The baseline in selecting the architecture is that it should realize all which typically consists of the list of modules, brief functionality of each module, their interface relationships, dependencies, database tables, architecture diagrams, technology details etc. The integration testing design is carried out in the particular phase.

- **Module design**

  The module design phase can also be referred to as low-level design. The designed system is broken up into smaller units or modules and each of them is explained so that the programmer can start coding directly. The low-level design document or program specifications will contain a detailed functional logic of the module

Validation phases

In the V-model, each stage of verification phase has a corresponding stage in the validation phase. The following are the typical phases of validation in the V-Model, though they may be known by other names.

- **Unit testing**

  In the V-Model, Unit Test Plans (UTPs) are developed during module design phase. These UTPs are executed to eliminate bugs at code level or unit level. A unit is the smallest entity which can independently exist, e.g. a program module. Unit testing verifies that the smallest entity can function correctly when isolated from the rest of the codes/units.

- **Integration testing**

  Integration Test Plans are developed during the Architectural Design Phase. These tests verify that units created and tested independently can coexist and communicate among themselves. Test results are shared with customer's team.

- **System testing**

  System Tests Plans are developed during System Design Phase. Unlike Unit and Integration Test Plans, System Test Plans are composed by client's business team. System Test ensures that expectations from application developed are met. The whole application is tested for its functionality, interdependency and communication. System Testing verifies that functional and non-functional requirements have been met. Load and performance testing, stress testing, regression testing, etc., are subsets of system testing.

- **User acceptance testing**

  User Acceptance Test (UAT) Plans are developed during the Requirements Analysis phase. Test Plans are composed by business users. UAT is performed in a user environment that resembles the production environment, using realistic data. UAT verifies that delivered system meets user's requirement and system is ready for use in real time.

Advantages of V-Model

- This is a highly disciplined model and Phases are completed one at a time.
- V-Model is used for small projects where project requirements are clear.
- Simple and easy to understand and use.
- This model focuses on verification and validation activities early in the life cycle thereby enhancing the probability of building an error-free and good quality product.
- It enables project management to track progress accurately.

Reasons to use V-Model

- It is easy to manage due to the rigidity of the model.
- Each phase of V-Model has specific deliverables and a review process.
- Proactive defect tracking – that is defects are found at early stage.
- The requirements are clearly defined and fixed in this project.
- The project has ample technical resources that are available with technical expertise.

CHAPTER 6

PROJECT DIAGRAMS

DATA FLOW DIAGRAM

The first step is to draw a data flow diagram (DFD). The DFD is a way of expressing system requirements in a graphical form. A DFD is also known as "bubble chart", has the purpose of clarifying system requirements and identifying major transformations that will become programs in system design. So it is the starting point of the design phase that functionally decomposes the requirements specifications down to the lowest level of detail. A DFD consists of a series of bubbles joined by lines. The bubbles represent data transformations and the lines represent data flow in the system.

Data Flow Components

DFDs consist of four basic components that illustrate how data flows in a system:

- *Entity:* An entity is the source or destination of data. Entities either provide data to the system (referred to as a source) or receive data from it (referred to as a sink). Entities are often represented as rectangles (a diagonal line across the right-hand corner means that this entity is represented somewhere else in the DFD). Entities are also referred to as agents, terminators, or source/sink.

- *Process:* The process is the manipulation or work that transforms data, performing computations, making decisions (logic flow), or directing data flows based on business rules. In other words, a process receives input and generates some output. Process names (simple verbs and dataflow names, such as "Submit Payment" or "Get Invoice") usually describe the transformation, which can be performed by people or machines. Processes can be drawn as circles or a segmented rectangle on a DFD, and include a process name and process number.

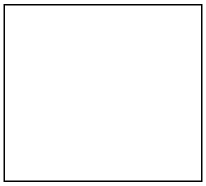
- *Data Store:* A data store is where a process stores data between processes for later retrieval by that same process or another one. Files and tables are considered data stores. Data store names (plural) are simple but meaningful, such as "customers," "orders," and "products." Data stores are usually drawn as a rectangle with the righthand side missing and labelled by the name of the data storage area it represents, though different notations do exist.

- *Data Flow:* Data flow is the movement of data between the entity, the process, and the data store. Data flow portrays the interface between the components of the DFD. The flow of data in a DFD is named to reflect the nature of the data used (these names should also be unique within a specific DFD). Data flow is represented by an arrow, where the arrow is annotated with the data name.

DFD Symbols

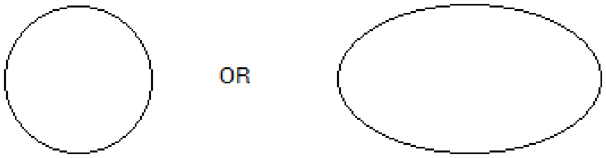
There are four types of symbols that are used to design DFD:

- *SQUARE:* A Square defines a source and the destination of the system Data.



- *ARROW:* An Arrow identifies the Data Flow.

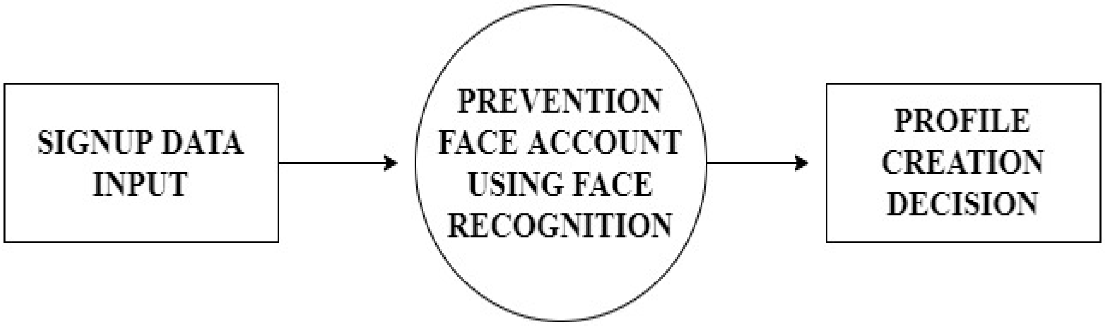
- *CIRCLE or BUBBLE:* A Circle or Bubble represents the process that transforms incoming Data Flow into outgoing Data Flow.



- *OPEN RECTANGLE:* It represents Data Storage.

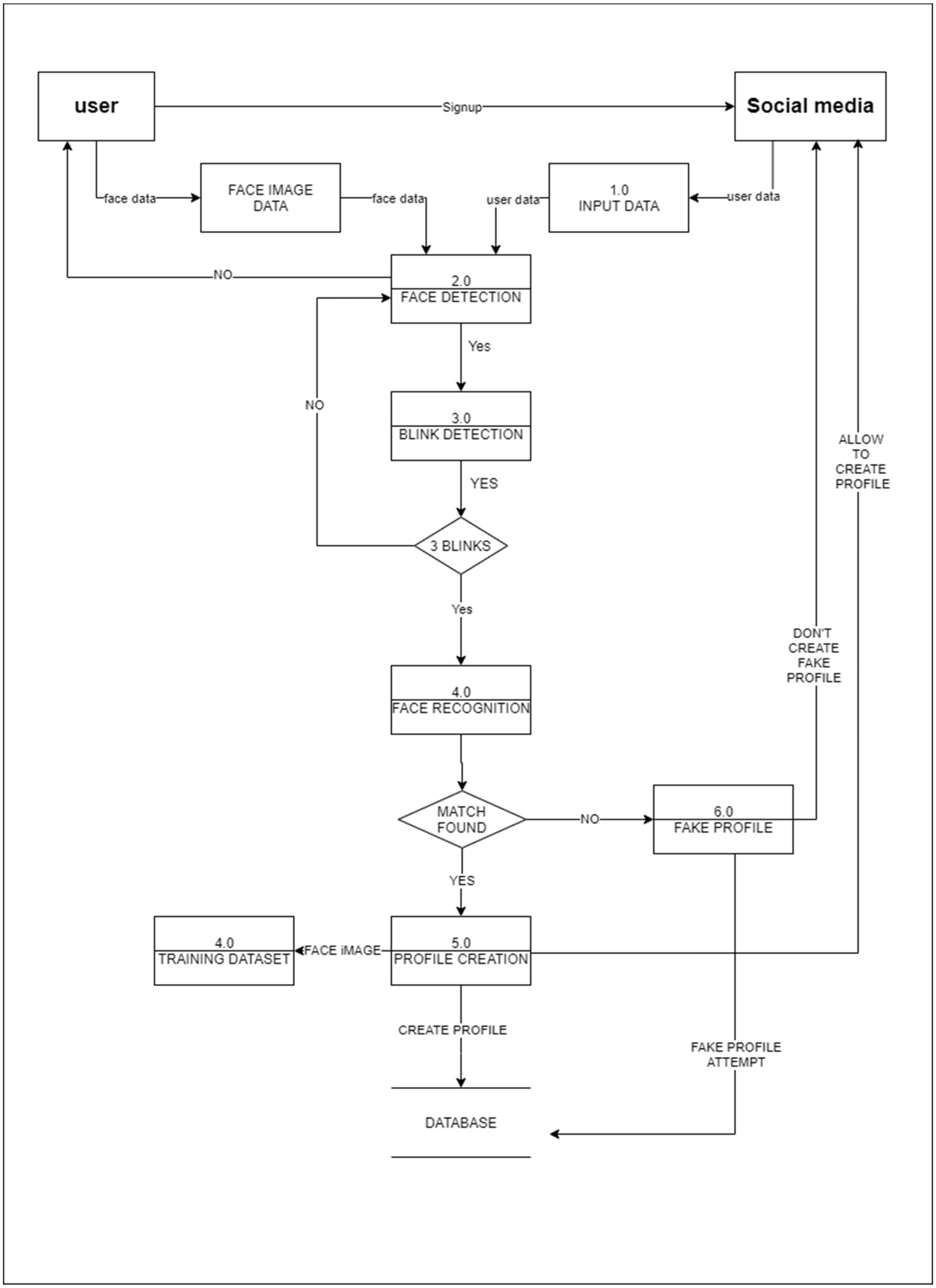
Context Level DFD

Figure 5. LEVEL 0 DFD


SIGNUP DATA INPUT
PREVENTION FACE ACCOUNT USING FACE RECOGNITION
PROFILE CREATION DECISION

Level 1 DFD

Figure 6. LEVEL 1 DFD



ENTITY-RELATIONSHIP DIAGRAM

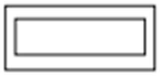
An Entity-Relationship Diagram or ER-Diagram is a photocopy of data structure. ER-Diagram was developed in 1970 by Dr. Peter Chen and others. They specified the representation of large and complex data storage concepts using the ER-Diagrams. An ER-Model is based on a perception of a real world which consists of a collection of basic objects called entities and relationship among these objects. It represents the overall logical structure of a database. An ER-Model is normally expressed as an ER-Diagram which is a graphical representation of a particular database.

Elements of ER-Model are listed below:

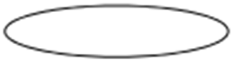
- Entities
- Attributes
- Entity Sets
- Relationships
- Relationship Sets

Entity: An entity can be a person, place, event, or object that is relevant to a given system. For example, a school system may include students, teachers, major courses, subjects, fees, and other items. Entities are represented in ER diagrams by a rectangle and named using singular nouns. The symbol of entity is:

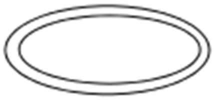
Weak Entity: A weak entity is an entity that depends on the existence of another entity. It uses a foreign key combined with its attribute to form the primary key. The symbol for weak entity is:



Attribute: An attribute is a property, trait, or characteristic of an entity, relationship, or another attribute. Meanwhile, attributes can also have their own specific attributes. The symbol for attribute is:



- *Multi valued Attribute:* If an attribute can have more than one value it is called an multi valued attribute. The symbol for multi valued attribute is:

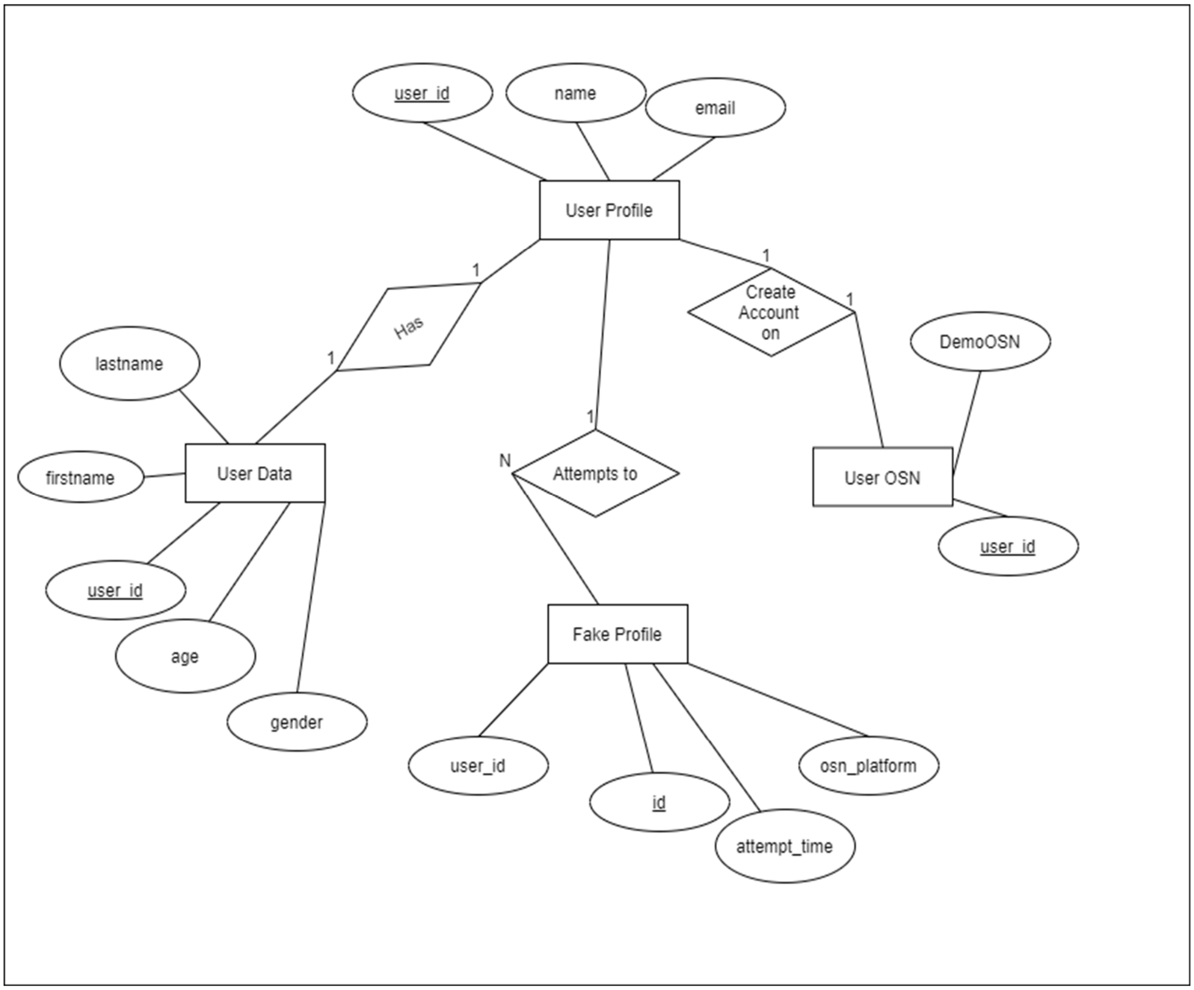


- *Derived Attribute:* An attribute based on another attribute. This is found rarely in ER diagrams. For example, for a circle the area can be derived from the radius.

Relationship: A relationship describes how entities interact. For example, the entity "carpenter" may be related to the entity "table" by the relationship "builds" or "makes". Relationships are represented by diamond shapes and are labelled using verbs.

ER Diagram

Figure 7. ER Diagram



DATABASE NORMALIZATION

Normalization is simply the process of distilling the structure of database to the point where user removes repeated group of data into separate tables. Grouping of data can be done in normalization. It is essentially the process in which a wide table with lots of columns but few rows are taken and redesigned it as several narrow tables with fewer columns but more rows. Normalization is a data analysis method used during the design stage of relational database. The components of normalization are normal forms.

Advantages of normalization

- To improve query retrieval performance.
- It reduces redundancy so problems are minimized.
- It eliminates insertion, deletion and modification problems to great extent.

Disadvantages of normalization

- It degrades query retrieval performance.
- Normalized tables will lose real world meaning.

Various types of normalizations

First normal form (1NF)

The definition of 1NF is there are no repeating groups. All the key attributes are defined. And all the attributes are dependent on the primary key. A relation is said to be in first normal form (1NF) if all the fields are atomic. In other words, only one value is associated with each attribute.

Second normal form (2NF)

A relation schema R is in second normal form (2NF) if, it is in 1NF and all its non-prime attributes are fully functionally dependent on the relation keys.

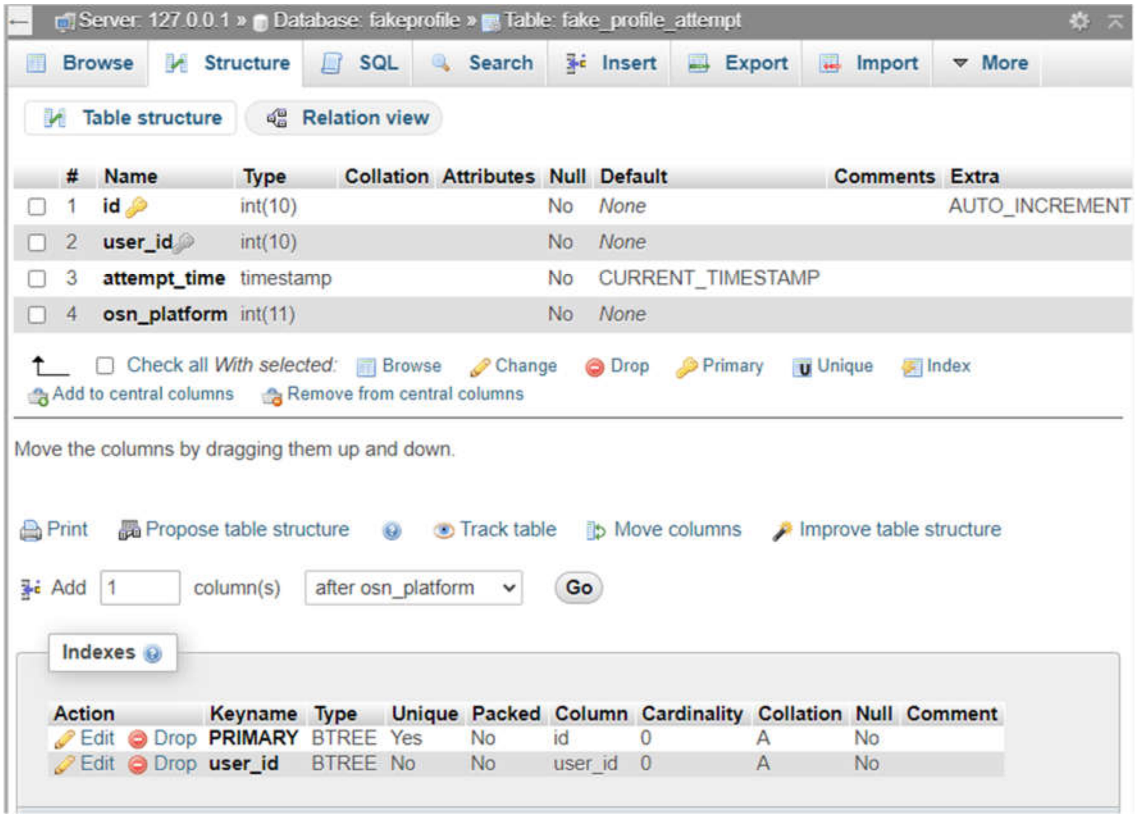
Third normal form (3NF)

The definition of third normal form is, it's in second normal form and it contains no transitive dependencies (where a non-key attribute is dependent on another non-key attribute). A relation schema R is said to be in 3NF if it is in 2NF and if it does not contain any transitive functional dependency i.e. every non-key attribute is non-transitively dependent on the primary key.

DATABASE SCHEMA

Figure 8. Database Tables

Fake Profile


Server: 127.0.0.1 » Database: fakeprofile » Table: fake_profile_attempt
Browse
Structure
SQL
Search
Insert
Export
Import
More
Table structure
Relation view

	#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra
□	1	**id**	int(10)			No	*None*		AUTO_INCREMENT
□	2	**user_id**	int(10)			No	*None*		
□	3	**attempt_time**	timestamp			No	CURRENT_TIMESTAMP		
□	4	**osn_platform**	int(11)			No	*None*		

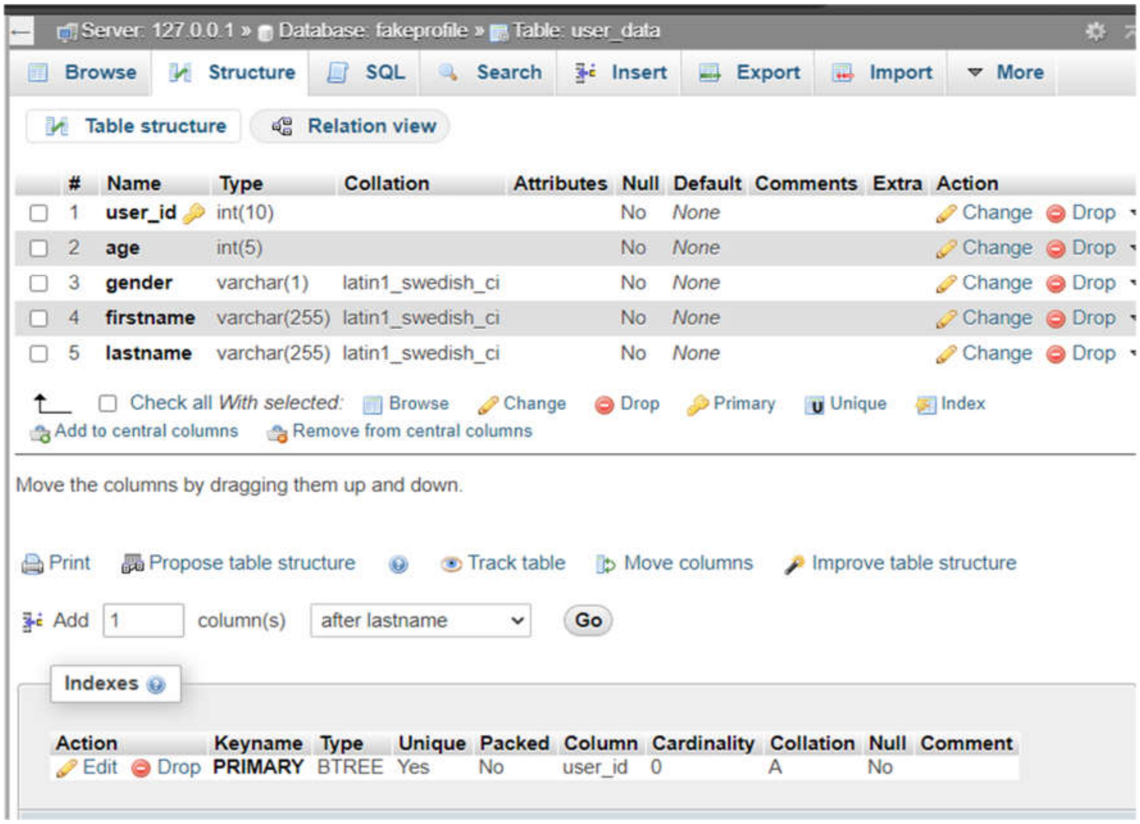
□ Check all With selected: Browse Change Drop Primary Unique Index
Add to central columns
Remove from central columns

Move the columns by dragging them up and down.

Print
Propose table structure
Track table
Move columns
Improve table structure
Add 1 column(s) after osn_platform
Go
Indexes

Action		Keyname	Type	Unique	Packed	Column	Cardinality	Collation	Null	Comment
Edit	Drop	**PRIMARY**	BTREE	Yes	No	id	0	A	No	
Edit	Drop	**user_id**	BTREE	No	No	user_id	0	A	No	

User Data



Server: 127.0.0.1 » Database: fakeprofile » Table: user_data

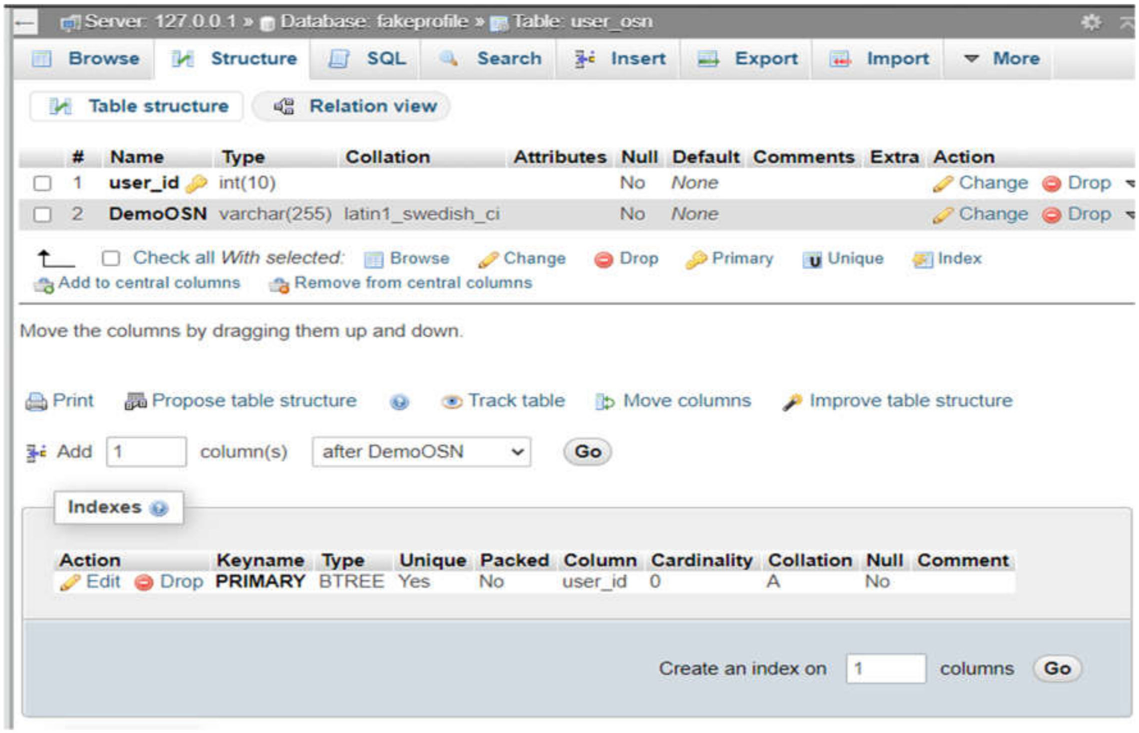
#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra	Action
1	**user_id**	int(10)			No	*None*			Change Drop
2	**age**	int(5)			No	*None*			Change Drop
3	**gender**	varchar(1)	latin1_swedish_ci		No	*None*			Change Drop
4	**firstname**	varchar(255)	latin1_swedish_ci		No	*None*			Change Drop
5	**lastname**	varchar(255)	latin1_swedish_ci		No	*None*			Change Drop

Check all *With selected:* Browse Change Drop Primary Unique Index
Add to central columns
Remove from central columns

Move the columns by dragging them up and down.

Action	Keyname	Type	Unique	Packed	Column	Cardinality	Collation	Null	Comment
Edit Drop	**PRIMARY**	BTREE	Yes	No	user_id	0	A	No	

User OSN



Server: 127.0.0.1 » Database: fakeprofile » Table: user_osn

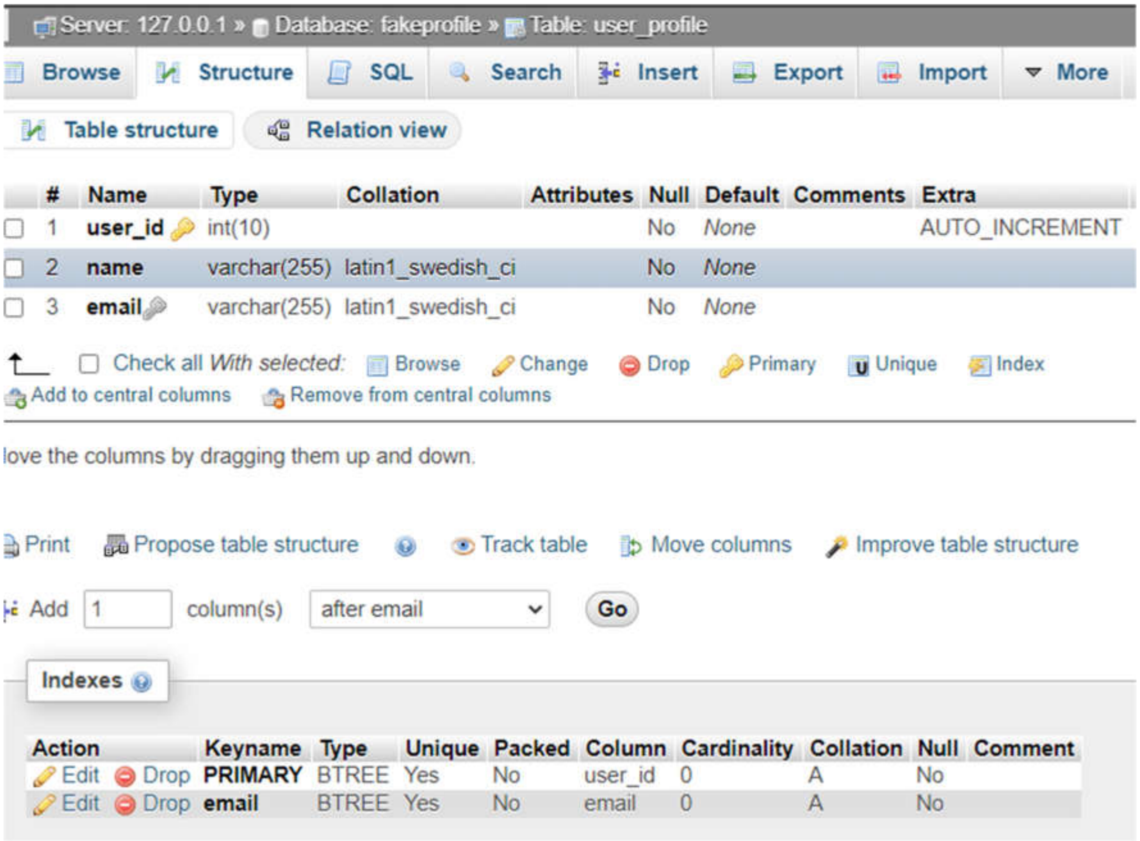
#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra	Action
1	**user_id**	int(10)			No	*None*			Change Drop
2	**DemoOSN**	varchar(255)	latin1_swedish_ci		No	*None*			Change Drop

Check all *With selected:* Browse Change Drop Primary Unique Index
Add to central columns
Remove from central columns

Move the columns by dragging them up and down.

Action	Keyname	Type	Unique	Packed	Column	Cardinality	Collation	Null	Comment
Edit Drop	**PRIMARY**	BTREE	Yes	No	user_id	0	A	No	

User Profile



	#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra
□	1	**user_id**	int(10)			No	*None*		AUTO_INCREMENT
□	2	**name**	varchar(255)	latin1_swedish_ci		No	*None*		
□	3	**email**	varchar(255)	latin1_swedish_ci		No	*None*		

love the columns by dragging them up and down.

Add 1 column(s) after email Go

Action	Keyname	Type	Unique	Packed	Column	Cardinality	Collation	Null	Comment
Edit Drop	**PRIMARY**	BTREE	Yes	No	user_id	0	A	No	
Edit Drop	**email**	BTREE	Yes	No	email	0	A	No	

CHAPTER 7

TESTING, IMPLEMENTATION AND MAINTENANCE

SOFTWARE TESTING

Software testing is a process of verifying and validating a software application or program. The main aim of the testing is to find out the bugs in the developed system before implementing it. It is an important phase of a successful system. After codifying the whole programs of the system, a test is being performed on the system so developed. The output of the testing phase should match the expected results.

Test Cases and Test Criteria

Ideally, we would like to determine a set of test cases such that successful execution of all of them implies that there are no errors in the program. This ideal goal can't usually be achieved due to practical and theoretical constraints. Each test case costs money, as effort is needed to generate the test case, machine time is needed to execute the program for that test case, and more effort is needed to evaluate the results. An ideal test case set is one that succeeds only if there are no errors in the program. One possible ideal set of test cases is one that includes all the possible inputs to the program. This is often called exhaustive testing; however, it is impractical and in feasible. For a given program P and its specification S, a test selection criterion specifies the conditions that must be satisfied by a set of test cases T. for example, if the criterion is that all statements in the program be executed at least once during testing, then a set of test cases T satisfies this criterion for a program P is the execution of P with T ensures that each statement in P is executed at least once. There are two fundamental properties for a testing criterion: reliability and validity. A criterion is reliable if all the sets of test cases that satisfy the criterion detect the same errors. A criterion is valid if for any error in the program there is some set satisfying the criterion that will reveal the error.

Black Box Testing: When we know the specified function that a product has been designed to perform, tests can be conducted that demonstrate each function is fully operational while at the same time searching for errors in each function. A black box test examines some fundamental aspect of a system with little regard for the internal logical structure of the software. Black box testing also called behavioural testing, focuses on the functional requirements of the software. Black box testing attempts to find errors in the following categories:

- Incorrect or missing functions.
- Interface errors.
- Errors in data structures or external database access.
- Behaviour or performance errors.
- Initialization and termination errors.

White Box Testing: When we know the internal workings of a product, tests can be conducted to ensure that internal operations are performed according to specifications and all internal components have been adequately exercised. This testing is sometimes called as glass box testing. Using white box testing methods, the software engineer can derive test cases that guarantee that all independent paths within a module have been exercised at least once exercise all logical decisions on their true and false sides execute all loops at their boundaries and within their operational bounds exercise internal data structures to ensure their validity. In this project our main emphasis is on white box testing. In order to test loops, we used the loop testing technique which is a white box testing technique. Most of the loops used in this project belong to the category of simple loops. We applied the following set of tests to test loops, where n is the maximum number of allowable passes through the loop.

- Skipped the loop entirely.
- Allowed only one pass through the loop.
- Allowed two passed through the loop.

Unit Testing: Unit testing is the testing of an individual unit or group of related units. It falls under the class of white box testing. It is often done by the programmer to test that the unit he/she has implemented is producing expected output against given input.

Integration Testing: Integration testing is testing in which a group of components are combined to produce output. Also, the interaction between software and hardware is tested in integration testing if software and hardware components have any relation. It may fall under both white box testing and black box testing.

Functional Testing: In functional testing, the structure of the program is not considered. Test cases are decided solely on the basis of the requirements or specifications of the program or module, and the internals of the module or the program are not considered for selection of test cases. Due to its nature, functional testing is often called, “black box testing”.

System Testing: System testing is the testing to ensure that by putting the software in different environments (e.g., Operating Systems) it still works. System testing is done with full system implementation and environment. It falls under the class of black box testing.

Stress Testing: Stress testing is the testing to evaluate how system behaves under unfavorable conditions. Testing is conducted at beyond limits of the specifications. It falls under the class of black box testing.

Performance Testing: Performance testing is the testing to assess the speed and effectiveness of the system and to make sure it is generating results within a specified time as in performance requirements. It falls under the class of black box testing.

Usability Testing: Usability testing is performed to the perspective of the client, to evaluate how the GUI is user-friendly? How easily can the client learn? After learning how to use, how proficiently can the client perform? How pleasing is it to use its design? This falls under the class of black box testing.

Acceptance Testing: Acceptance testing is often done by the customer to ensure that the delivered product meets the requirements and works as the customer expected. It falls under the class of black box testing.

Regression Testing: Regression testing is the testing after modification of a system, component, or a group of related units to ensure that the modification is working correctly and is not damaging or imposing other modules to produce unexpected results. It falls under the class of black box testing.

Beta Testing: Beta testing is the testing which is done by end users, a team outside development, or publicly releasing full pre-version of the product which is known as beta version. The aim of beta testing is to cover unexpected errors. It falls under the class of black box testing.

IMPLEMENTATION AND MAINTENANCE

Implementation

Implementation is the stage of a project during which theory is turned into practice. The major steps involved in this phase are:

Acquisition and Installation of Hardware and Software: The hardware and the relevant software required for running the system must be made fully operational before implementation.

Conversion: The conversion is also one of the most critical and expensive activities in the system development life cycle. The data from the old system needs to be converted to operate in the new format of the new system. The database needs to be setup with security and recovery procedures fully defined.

User Training: During this phase, all the programs of the system are loaded onto the user's computer. After loading the system, training of the user starts. Main topics of such type of training are:

- How to execute the forms
- How to enter the data
- How to process the data (processing details)
- How to infer the results

After the users are trained about the computerized system, working has to shift from manual to computerized working. The process is called 'Changeover'.

Maintenance

Maintenance is necessary to eliminate errors in the system during its working life and to tune the system to any variations in its working environments. It has been seen that there are always some errors found in the systems that must be noted and corrected. It also means the review of the system from time to time. The review of the system is done for:

- Knowing the full capabilities of the system
- Knowing the required changes or the additional requirements
- Studying the performance.

COST ANALYSIS

Resource sharing

The main goal is to make all programs, equipment and data available to anyone on the network without regard to the physical location of the resource and the user. Users need to share resources other than files, as a printer. Printers are utilized only a small percentage of the time; therefore, companies don’t want to invest in a printer for each computer. Network can be used in this situation to allow all the users to have any access any of the available printers.

High reliability

The goal of computer network is to provide high reliability by having alternative source of supply. For example, all files could be replicated on two or three machines, so if some of them is unavailable then other copies can be used. In addition, the presence of multiple CPUs means that if one goes down, the other may be available to take over its work, although at reduced performance. For example, applications, the ability to continue in the face of H/W problem is of utmost importance.

Saving money

Small computers have much better price and performance ratio than larger ones. Mainframes are faster roughly by a factor of ten than personal computers but they cost a thousand times more. This imbalance has caused many systems designers to build a system consisting of personal computers, one per user, with data kept on one or more shared file server machines. In this model, the user are called client and the whole arrangement is called the Client-Server model.

Scalability

The ability to increase the system performance gradually as the workload grows just by adding more processes. With centralizes mainframes, adding another processor is very expensive, so user must replace it with Client-Server model. New client and new server can be added as needed.

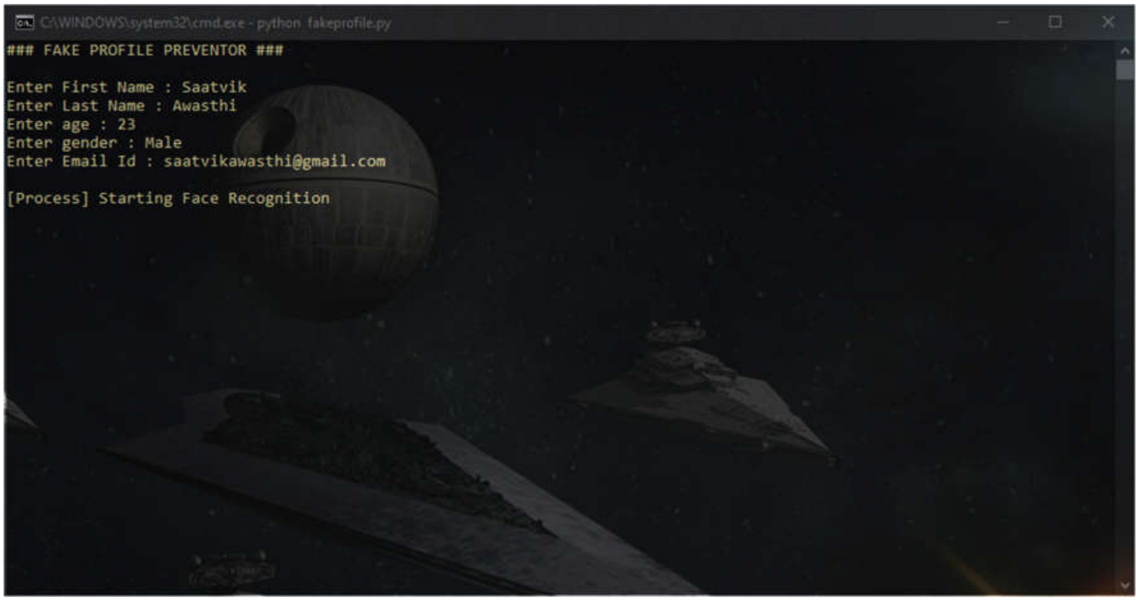
CHAPTER 8

RESULTS AND FUTURE SCOPE

RESULTS

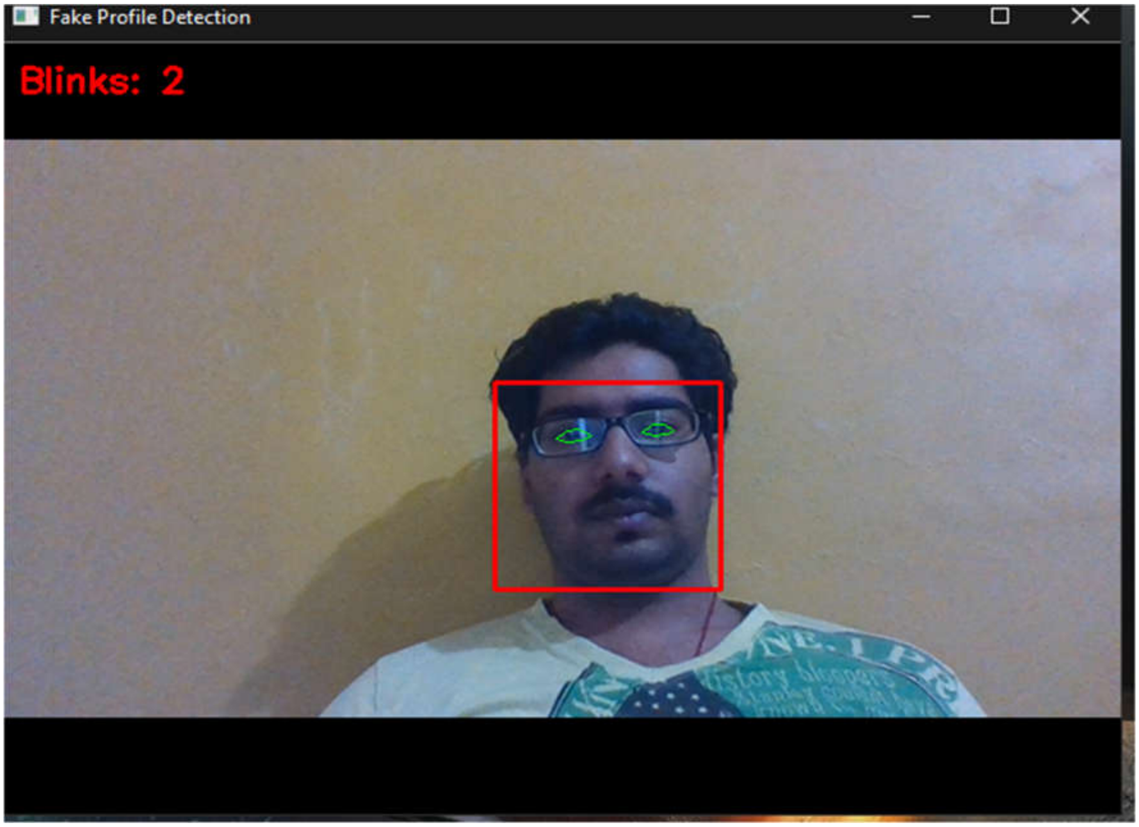
The system implemented Deep Neural Networks (Inception Network) for the identification of facial data. The system achieved an accuracy of 97% on facial recognition. The system is capable enough to identify the fake profile creation activities. The model described above is fully functional and can be improved with by using more sophisticated parameters and utilizing higher quality images.

Input Screen:

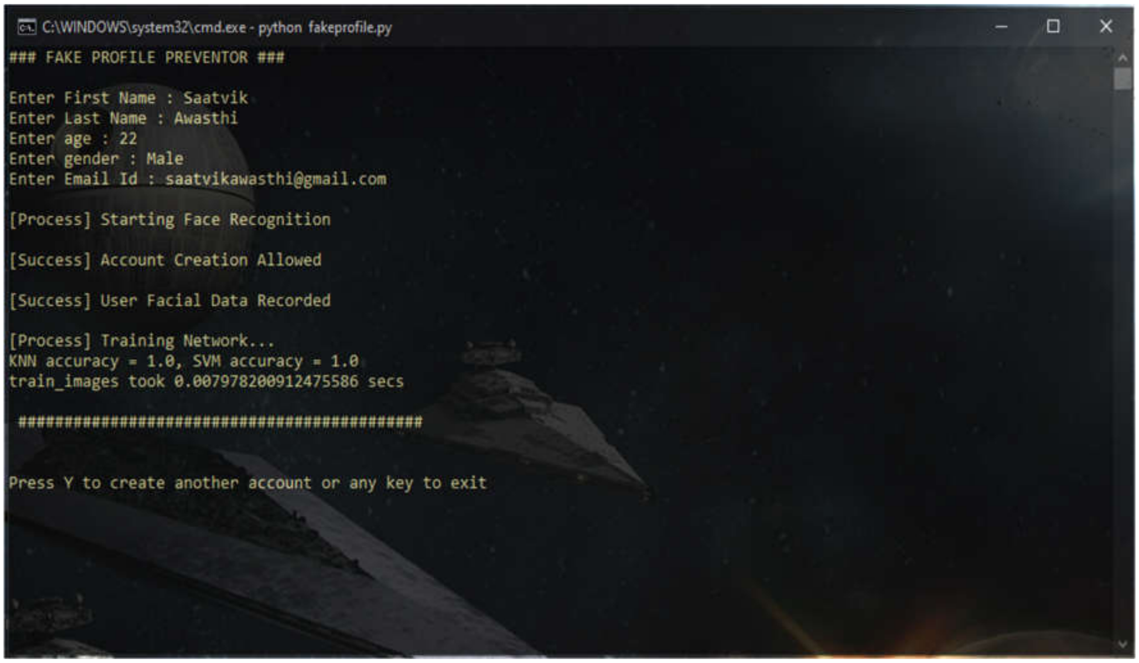


```
### FAKE PROFILE PREVENTOR ###

Enter First Name : Saatvik
Enter Last Name : Awasthi
Enter age : 23
Enter gender : Male
Enter Email Id : saatvikawasthi@gmail.com

[Process] Starting Face Recognition
```

Eye Blink Detection

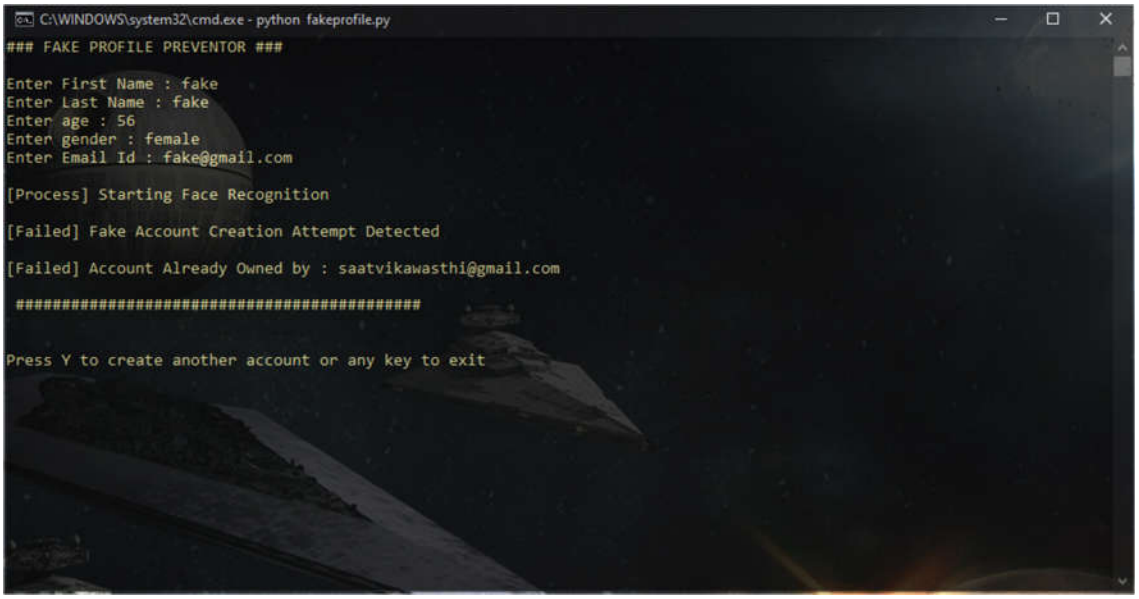


Account Creation allowed



```
### FAKE PROFILE PREVENTOR ###

Enter First Name : Saatvik
Enter Last Name : Awasthi
Enter age : 22
Enter gender : Male
Enter Email Id : saatvikawasthi@gmail.com

[Process] Starting Face Recognition

[Success] Account Creation Allowed

[Success] User Facial Data Recorded

[Process] Training Network...
KNN accuracy = 1.0, SVM accuracy = 1.0
train_images took 0.007978200912475586 secs

 ############################################


Press Y to create another account or any key to exit
```

Fake Profile Attempt



```
### FAKE PROFILE PREVENTOR ###

Enter First Name : fake
Enter Last Name : fake
Enter age : 56
Enter gender : female
Enter Email Id : fake@gmail.com

[Process] Starting Face Recognition

[Failed] Fake Account Creation Attempt Detected

[Failed] Account Already Owned by : saatvikawasthi@gmail.com

 ############################################


Press Y to create another account or any key to exit
```

FUTURE SCOPE

The future scope of this project is to improve the fake profile detection with more sophisticated parameters, develop it as an API for multiple social media platforms to integrate and provide seamless profile validation and fake profile prevention. The project can be equipped with additional functionality to provide like secure logion over face recognition. The project can be improved by adding age and gender prediction technology by using more sophisticated parameters. The age prediction can be used to predict the facial changes over time and use that facial data for verification of the account. Social media accounts across platforms can be combined as one to provide easy login to accounts.

CHAPTER 9

REFRENCES

RESEARCH PAPERS

1. Kevin Santoso, Gede Putra Kusuma; Face Recognition Using Modified OpenFace; 3rd International Conference on Computer Science and Computational Intelligence 2018; doi: 10.1016/j.procs.2018.08.203

2. G. Eason, B. Noble, and I.N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955. (references)

3. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

4. I.S. Jacobs and C.P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

5. B. Hudson, B. R. Voter, “Profile characteristics of fake twitter accounts”, Big Data & Society, 2016.

6. Z. Yang, et al. Uncovering social network sybils in the wild. Transactions on Knowledge Discovery from Data (TKDD) , Vol. 8, No. 1, 2014.

7. Estee Van Der Walt and Jan Eloff,"Using Machine Learning to Detect Fake Identities:Bots vs Humans"IEEE Trans. Emerg.TopicsComput. Intell., vol. 1, no. 1, pp. 61–71 March 2018. [7] K. Elissa, “Title of paper if known,” unpublished.

8. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

9. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

10. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

11. Sarah Khaled, Neamat El-Tazi and Hoda M. O. Mokhtar"Detecting Fake Accounts on Social Media" IEEE International Conference on Big Data.., vol.6 pp 101-110 ,2018 fake profile

ONLINE SOURCES:

1. https://www.google.com
2. https://www.wikipedia.org
3. https://www.codeproject.com
4. https://www.stackoverflow.com