# Mini Project
# Phase-1

~ Sejal Gurkhe (2019130017)

~ Ishita More (2019130039)

~ Under the Guidance of Prof. Kiran Gawande

# "Fraud Web Page Detection using Machine Learning"

# Problem Statement

- The basic idea of this project is to create a website that will detect Fraud Webpage of provided URL.
- The Project is basically based on Machine Learning.
- Using ML Algorithm, we will detect whether the website is safe or not.
- After detection, if the web page is a legitimate web page then use computer networking(optional).

# Literature Survey

| Sr. No. | Title | Publication | Short Description | Link |
|---|---|---|---|---|
| 1 | Phishing Detection using Random Forest, SVM and Neural Network with Backpropagation | 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE) | Random Forest classification method, SVM classification algorithm, and Neural Network classification algorithm were improved. SVM classification algorithm gave better accuracy.SVM is chosen as the final classifier algorithm. | https://ieeexplore.ieee.org/document/9277256 |
| 2 | Phishing Website Detection Based on Machine Learning Algorithm | 2020 International Conference on Computing and Data Science (CDS) | Logistic regression classifier, SVM, Naive Bayes, Decision tree algorithm were discussed. But, Logistic regression classifier is selected. | https://ieeexplore.ieee.org/document/9275957 |
| 3 | Phishing Detection Using Machine Learning Technique | 2020 First International Conference of Smart Systems and Emerging Technologies (SMART TECH) | The proposed technique used SVM which has high accuracy. The proposed technique can detect new temporary phishing sites and reduce the damage caused by phishing attacks. | https://ieeexplore.ieee.org/document/9283771 |
| 4 | Detecting Phishing Website Using Machine Learning | 2020 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA) | This paper discussed the software that is designed to show awareness of the extensive level of its functionality, features that can be displayed in the monitoring. The system fosters many features in comparison to other software. | https://ieeexplore.ieee.org/document/9068728 |

# Literature Survey (contd...)

| Sr. No. | Title | Publication | Short Description | Link |
|---|---|---|---|---|
| 5 | Defending against Phishing Attacks: Taxonomy of Methods, Current Issues, and Future Directions | Springer Science+Business Media, LLC, part of Springer Nature 2020 | The protection strategies talked about in this paper are data mining and heuristics, ML, and deep learning algorithms. The ML procedures give the best outcomes when contrasted with different strategies. | https://link.springer.com/article/10.1007/s11235-020-00733-2 |
| 6 | Defending against Phishing Attacks: Taxonomy of Methods, Current Issues, and Future Directions | Springer Science+Business Media New York 2017 | This paper classified social engineering phishing based on spoofed email attacks and fake websites. paper described various issues and challenges in current solutions defending against phishing attacks. | https://link.springer.com/article/10.1007/s11235-017-0334-z |
| 7 | Intelligent Web-Phishing Detection and Protection Scheme using integrated Features of Images, Frames and Text | an Intelligent System Research Group, Anglia Ruskin Information Technology Institute, Anglia Ruskin University, Chelmsford, the United Kingdom | The primary contribution of this study is the integration of hybrid features that have been extracted from text, images, and frames | https://www.researchgate.net/publication/326832835_Intelligent_Web-Phishing_Detection_and_Protection_Scheme_using_integrated_Features_of_Images_Frames_and_Text |

# Literature Survey (contd...)

| Sr. No. | Title | Publication | Short Description | Link |
|---|---|---|---|---|
| 8 | Defending against Phishing Attacks: Taxonomy of Methods, Current Issues, and Future Directions | 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) | Algorithm- Logistic regression, decision tree, and random forest. | https://ieeexplore.ieee.org/document/8697412 |
| 9 | Intelligent Web-Phishing Detection and Protection Scheme using integrated Features of Images, Frames and Text | ------- | The baseline features perform best when integrated with the Random Forest classifier, achieving competitive accuracy. | DOI: 10.1109/ICCUBEA.2018.8697412 |
| 10 | URL Phishing Analysis using Random Forest | International Journal of Pure and Applied Mathematics | Random Forest classifier is used and compares its classification performance with the SVM algorithm. Random Forest performed better than the SVM algorithm. | https://acadpubl.eu/hub/2018-118-21/articles/21e/49.pdf |

# Objectives:

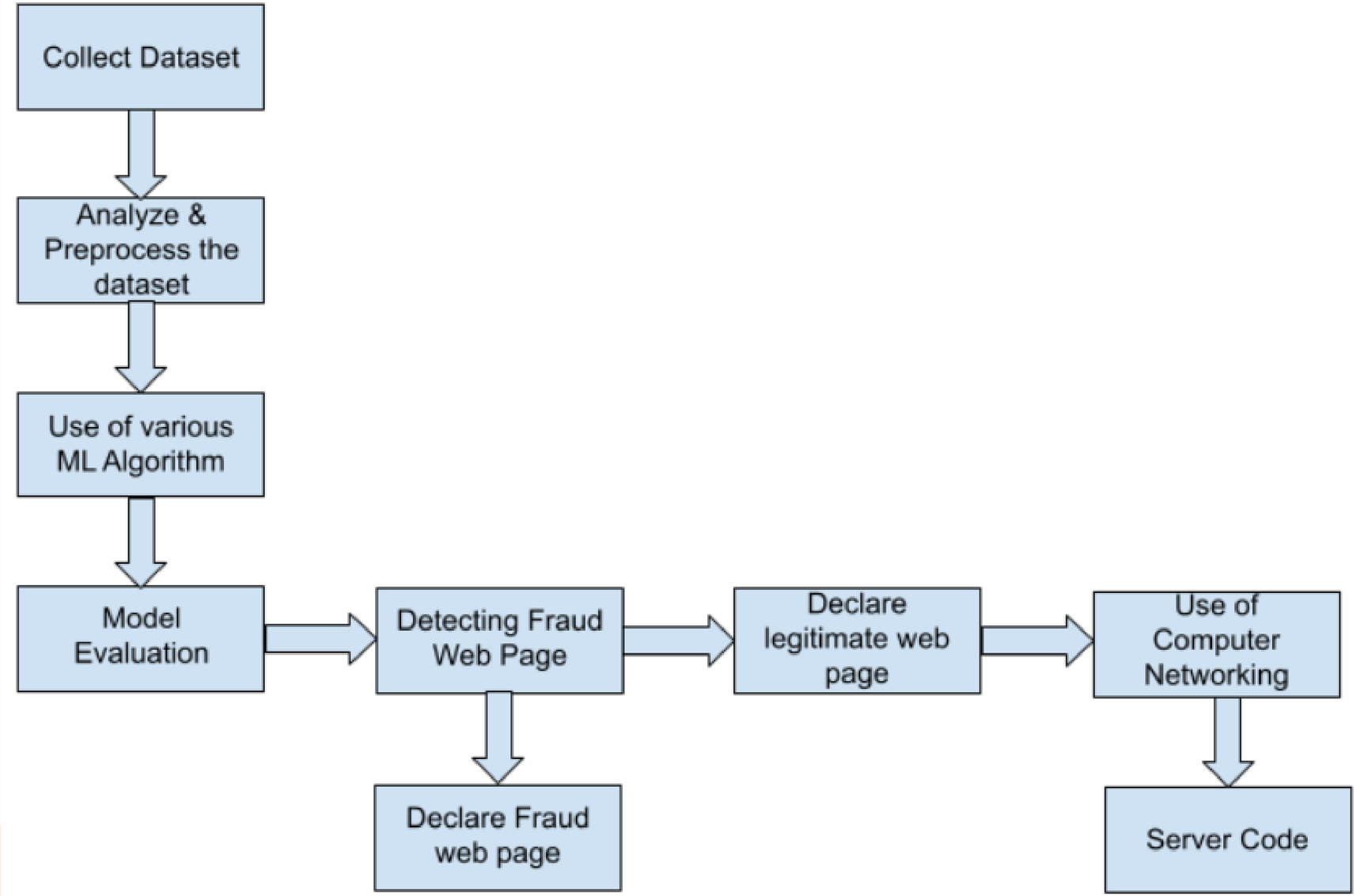
The objective of this project is

- To train machine learning models on the dataset created to predict fraud web pages.
- To use the dataset of fraud websites data and from that data use it to detect Fraud Web pages.
- The performance level of each algorithm is measured and compared.
- If the web page is a legitimate web page then using computer networking, stating whether the web page is more secure or not.

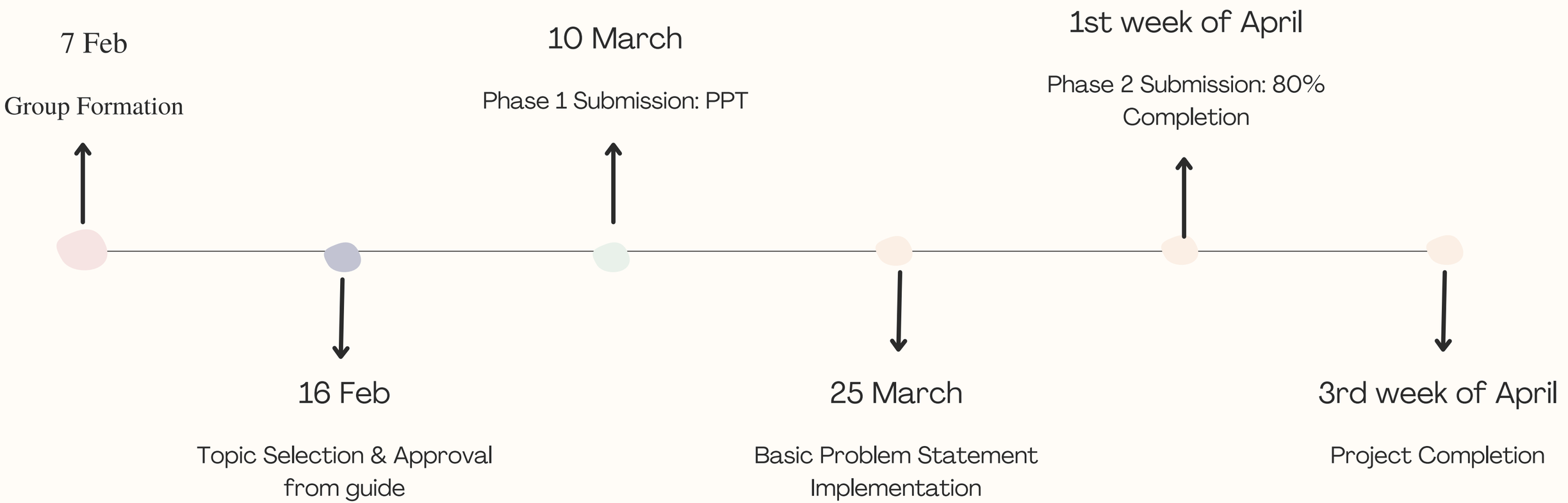
- Scope of Project:

  - A system that can be integrated in order to increase the detection performance in real-time.
  - A classifier that gives a good prediction rate of fraud websites.

- Proposed System:



# Timeline


7 Feb
Group Formation
16 Feb
Topic Selection & Approval from guide
10 March
Phase 1 Submission: PPT
25 March
Basic Problem Statement Implementation
1st week of April
Phase 2 Submission: 80% Completion
3rd week of April
Project Completion

# Reference:


- Weiheng Bai; "Phishing Website Detection Based on Machine Learning Algorithm" https://ieeexplore.ieee.org/document/9275957
- Junaid Rashid; Toqeer Mahmood; Muhammad Wasif Nisar; Tahira Nazir; "Phishing Detection Using Machine Learning Technique" https://ieeexplore.ieee.org/document/9283771
- Mohammed Hazim Alkawaz; Stephanie Joanne Steven; Asif Iqbal Hajamydeen; "Detecting Phishing Website Using Machine Learning" https://ieeexplore.ieee.org/document/9068728

# THANK YOU