Project Design Phase-II
Technology Stack (Architecture & Stack)

Date	03 October 2022
Team ID	PNT2022TMID40333
Project Name	Project - Web Phishing Detection
Maximum Marks	4 Marks

Technical Architecture:

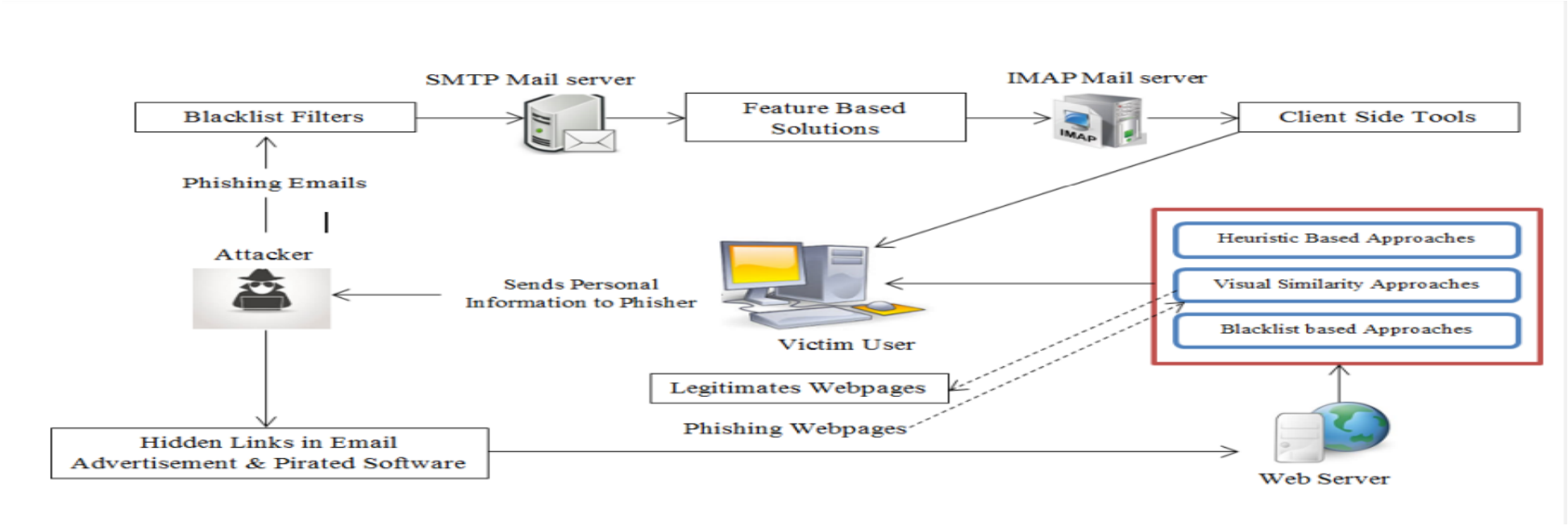
The Deliverable shall include the architectural diagram as below and the information as per the table1 & table 2

Example: Order processing during pandemics for offline mode

Reference: https://developer.ibm.com/patterns/ai-powered-backend-system-for-order-processing-during-pandemics/

Guidelines:

- Include all the processes (As an application logic / Technology Block)
- Provide infrastructural demarcation (Local / Cloud)
- Indicate external interfaces (third party API's etc.)
- Indicate Data Storage components / services
- Indicate interface to machine learning models (if applicable)


SMTP Mail server
IMAP Mail server
Blacklist Filters
Feature Based Solutions
Client Side Tools
IMAP
Phishing Emails
Attacker
Sends Personal Information to Phisher
Victim User
Heuristic Based Approaches
Visual Similarity Approaches
Blacklist based Approaches
Legitimates Webpages
Phishing Webpages
Hidden Links in Email Advertisement & Pirated Software
Web Server

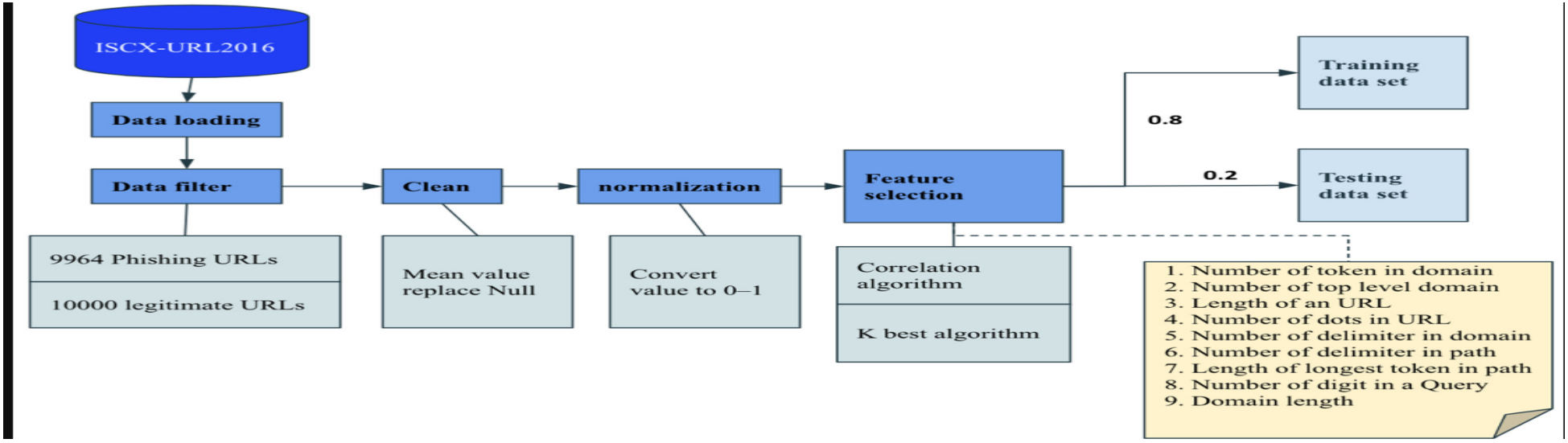

ISCX-URL2016
Data loading
Data filter
Clean
normalization
Feature selection
0.8
0.2
Training data set
Testing data set
9964 Phishing URLs
10000 legitimate URLs
Mean value replace Null
Convert value to 0–1
Correlation algorithm
K best algorithm
1. Number of token in domain
2. Number of top level domain
3. Length of an URL
4. Number of dots in URL
5. Number of delimiter in domain
6. Number of delimiter in path
7. Length of longest token in path
8. Number of digit in a Query
9. Domain length

Table-1 : Components & Technologies:

S.No	Component	Description	Technology
1.	User Interface	How Attackers distribute emails to user e.g. Messages, Mails etc.	HTML, CSS, JavaScript ,python etc.
2.	Logic-1	Users open the email without knowing the tricks.. ● Heuristic based Approaches ● legitimates webpages ● visual similarity Approaches ● Blacklist based Approaches	Java / Python
3.	Logic-2	Users once visit the phishing sites user information will be hacked by hackers.. ● correlation algorithm ● Data processing	IBM Watson STT service,python
4.	Logic-3	User as give the URL as input in the required field and wait for validation.. ● K best algorithm ● ML Classification	IBM Watson,flack API,
5.	Database	Data Type, Configurations etc.	MySQL, NoSQL, etc.
6.	Internal network	Data stolen from the user.. Feature based solutions	IBM DB2, IBM Cloudant etc.
7.	Data	Hacked informations stored in storage	IBM Block Storage or Other Storage Service or Local Filesystem

Table-2: Application Characteristics:

S.No	Characteristics	Description	Technology
1.	Open-Source Frameworks	List the open-source frameworks used ● Flack frame work	Technology of Open source framework ● python libraries
2.	Security Implementations	List all the security / access controls implemented, use of firewalls etc. ● browsers	e.g. Decision tree algorithm, IAM Controls, etc.
3.	Scalable Architecture	Justify the scalability of architecture (3 – tier, Micro-services)	Technology used
4.	Availability	Justify the availability of application (e.g. use of load balancers, distributed servers etc.)	Technology used
5.	Performance	Design consideration for the performance of the application (number of requests per sec, use of Cache, use of CDN's) etc.	Technology used

References:

https://c4model.com/

https://developer.ibm.com/patterns/online-order-processing-system-during-pandemic/

https://www.ibm.com/cloud/architecture

https://aws.amazon.com/architecture

https://medium.com/the-internal-startup/how-to-draw-useful-technical-architecture-diagrams-2d20c9fda90d