**Project Proposal**

**Question/need:**

* What is the framing question of your analysis, or the purpose of the model/system you plan to build?

The goal of the System is phishing features analysis, conducting rapid proof of concept experiments or benchmarking phishing classification models.

* Who benefits from exploring this question or building this model/system?

Anti-phishing researchers

Anti-phishing company

End user

**Data Description:**

* What dataset(s) do you plan to use, and how will you obtain the data?

Phishing Dataset from kaggle website

* What is an individual sample/unit of analysis in this project? What characteristics/features do you expect to work with?

The features :

id

NumDots

SubdomainLevel

PathLevel

UrlLength

NumDash

NumDashInHostname

AtSymbol

TildeSymbol

NumUnderscore

NumPercent

NumQueryComponents

NumAmpersand

NumHash

NumNumericChars

NoHttps

RandomString

IpAddress

DomainInSubdomains

DomainInPaths

HttpsInHostname

HostnameLength

PathLength

QueryLength

DoubleSlashInPath

NumSensitiveWords

EmbeddedBrandName

PctExtHyperlinks

PctExtResourceUrls

ExtFavicon

InsecureForms

RelativeFormAction

ExtFormAction

AbnormalFormAction

PctNullSelfRedirectHyperlinks

FrequentDomainNameMismatch

FakeLinkInStatusBa

RightClickDisabled

PopUpWindow

SubmitInfoToEmail

IframeOrFrame

MissingTitle

ImagesOnlyInForm

SubdomainLevelRT

UrlLengthRT

PctExtResourceUrlsRT

AbnormalExtFormActionR

ExtMetaScriptLinkRT

PctExtNullSelfRedirectHyperlinksRT

CLASS\_LABEL

* If modeling, what will you predict as your target?

Model : Classification and cluster

Target : Predict possible attack