ISSN 2277-2685
IJESR/Sep. 2023/ Vol-13/Issue-3/1-19
Mr.Mrutyunjaya Yalawar *et. al.*, / International Journal of Engineering & Science Research

NetSpam:ANetwork-Based Spam Detection Framework for Reviewsin Online SocialMedia

Mr.Mrutyunjaya Yalawar [1] Muthyam Kavya[2], P Spoorthi Reddy[2], Soma Sai Amaranath[2], S Venkat Sai Kumar[2] [1]Asst. *Professor,Computer Science and Engineering, CMR Engineering College, medchal, T.S, India*[2]*B.Tech, Computer Science and Engineering, CMR Engineering College, medchal, T.S, India*

Abstract—
Nowadays,abigpartofpeoplerelyonavailablecontentinsocialmediaintheirdecisions(e.g.,reviewsandfeedback onatopic orproduct). The possibility thatanybodycan leavea review provides a golden opportunity for spammersto write spam reviews about products and services for

differentinterests.Identifyingthesespammersandthespamcontentisahottopicofresearch,andalthoughaconsiderablenumberofstudies havebeen done recentlytowardthis end, but sofarthe methodologies put forth still barely detect spam reviews, andnone of them show the importance of each extracted feature type.In this paper, we propose a novel framework, named NetSpam,whichutilizesspamfeaturesformodeling reviewdata setsas heterogeneous information networks to map spam detectionprocedure into a classification problem in such networks. Usingthe importance of spam features helps us to obtain better resultsin terms of different metrics experimented on real-world reviewdata sets from Yelp and AmazonWeb sites. The results

showthat*NetSpam*outperformstheexistingmethodsandamongfourcategoriesoffeatures,including review-behavioral,user-behavioral,review-linguistic, anduser-linguistic,

thefirsttypeoffeaturesperformsbetterthantheothercategories.

IndexTerms— Socialmedia, social network, spammer,spam

review,fakereview,heterogeneousinformationnetworks.

I. INTRODUCTION

ONLINESocialMediaportalsplayaninfluentialroleininformationpropagation whichisconsideredasanimportantsourceforproducersintheiradvertisingcampaignsaswellasforcustomersinselectingproductsandservices.Inthepastyears,peoplerelyalotonthewrittenreviewsintheir decision-making

processes, and positive/negative

reviewsencouraging/discouragingthemintheirselectionofproductsandservices.Inaddition,writtenreviewsalsohelpserviceproviders to enhance the quality of their products and

services.Thesereviewsthushavebecomeanimportantfactorinsuccessofabusinesswhilepositivereviewscanbringbenefitsforacompany,negativereviewscanpotentiallyimpactcredibilityandcauseeconomiclosses.Thefact thatanyonewithanyidentitycanleavecommentsasreviewprovidesatempting

opportunity for spammers to write fake reviews designed tomisleadusers'opinion. Thesemisleading reviews arethenmultiplied by the sharing function of social media and prop-agation over the web. The reviews written to change users'perception of how good a product or a service are consideredasspam[11], and are oftenwritteninexchange formoney.As shown in [1], 20% of the reviews in the Yelp website areactuallyspamreviews.

Ontheotherhand,aconsiderableamountofliteraturehas beenpublished onthe techniques used toidentify spamand spammers aswellasdifferent type ofanalysis onthistopic[30], [31]. These techniques can be classified into dif-ferentcategories;someusinglinguisticpatternsintext[2]–[4],which are mostly based on bigram, and unigram, others arebased on behavioral patterns that rely on features extractedfrom patterns in users' behavior which are mostly metadata-based[5-6]–

[8],[9],[34],andevensometechniquesusinggraphsandgraph-basedalgorithmsandclassifiers[10]–[12].

Despite this great deal of efforts, many aspects have beenmissed or remained unsolved. One of them is a classifier thatcan calculate feature weights that show each feature's level ofimportance in determining spam reviews. The general conceptof our proposed framework is to model a given review datasetas a Heterogeneous Information Network (HIN) [19] and tomap the problem of spam detection into a HIN classificationproblem. Inparticular, wemodel reviewdatasetasaHINin which reviews are connected through different node types(such as features and users). A weighting algorithm is thenemployed to calculate each feature's importance (or

weight).Theseweightsareutilizedtocalculatethefinallabelsforreviewsusingbothunsupervisedandsupervisedapproaches.

Toevaluate theproposed solution, weused twosamplereview datasets from Yelpand Amazon websites.Basedonour observations, defining two views for features (review-userand behavioral-linguistic), the classified features as review-behavioral have moreweightsandyieldbetterperformanceon spotting spam reviews in both semi-supervised and unsu-pervised approaches[13-15]. In addition, we demonstrate that usingdifferent supervisions suchas1%,2.5%and5%orusinganunsupervised approach,

makenonoticeablevariationontheperformanceofourapproach.Weobservedthatfea-tureweightscanbeaddedorremovedforlabelingandhence time complexity can be scaled for a specific level ofaccuracy. As the result of this weighting step, we can usefewer features with more weights to obtain better accuracywithlesstimecomplexity.Inaddition,categorizingfeaturesinfourmajorcategories(review-behavioral,user-behavioral,review-linguistic, user-linguistic), helps us to understand

howmucheachcategoryoffeaturesiscontributedtospamdetection.

Insummary,ourmaincontributionsareasfollows:

(i) WeproposeNetSpamframeworkthatisanovelnetwork-basedapproachwhichmodelsreviewnetworksashetero-geneous information

(ii) networks.Theclassificationstepusesdifferent metapath types which areinnovative

(iii) inthespamdetectiondomain.

(iv) Anewweightingmethodforspamfeaturesispro-posedtodetermine therelativeimportance

(v) ofeachfeatureand shows how effective each of features are in identifyingspams from normal reviews. Previous works [16-19], [20] alsoaimedtoaddresstheimportanceoffeaturesmainlyintermof obtained accuracy, but not as a build-in function in theirframework (i.e., their approach isdependent to ground truthfor determining each feature importance). As weexplain inour unsupervised approach, NetSpam is able to find featuresimportance even without ground truth, and only by relying onmetapath definition and based on values calculated for eachreview.

(vi)

(vii) NetSpam improves the accuracy compared to the state-of-

(viii) theartintermsoftimecomplexity,whichhighlydependstothe number of features used to identify a spam review; hence,using features with more weights will resulted in

(ix) detectingfakereviewseasierwithlesstimecomplexity.

ISSN 2277-2685
IJESR/Sep. 2023/ Vol-13/Issue-3/1-19
Mr.Mrutyunjaya Yalawar *et. al.*, / International Journal of Engineering & Science Research

II. RELIMINARIES

Asmentionedearlier,wemodeltheproblemasahet-erogeneous network [21] where nodes areeither real componentsin a dataset (such as reviews, users and products) or

spamfeatures.Tobetterunderstandtheproposedframeworkwefirstpresentanoverviewofsomeoftheconceptsanddefinitionsinheterogeneousinformationnetworks[22]–[24].

A. *Definitions*

*1(HeterogeneousInformationNetwork):*Supposewe have $r(>1)$ types of nodes and $s(>1)$ types of relationlinks between the nodes, then a heterogeneous informationnetworkisdefinedasagraph*G(V, E)* whereeachnode*v* *V*and each link *e* *E*belongs to one particular nodetypeandlinktyperespectively.Iftwolinksbelongtothesametype, the types of starting node and ending node of those linksarethesame.

*Definition2(NetworkSchema):*Givenaheterogeneousinformationnetwork*G(V,E)*,anetworkschema*T* *(A, R)* is a metapath with the object type mapping $\tau VA$and link mapping φER, which is a graph defined overobject type*A*, with links as relations from*R*. The schemadescribesthemetastructureofagivennetwork(i.e.,howmanynodetypesthereareandwherethepossiblelinksexist).

*Definition3(Metapath):*As mentioned above, therearenoedgesbetweentwonodesofthesametype,buttherearepaths. Givenaheterogeneous information network*G=(V,E)*,ametapath*P*isdefinedbyasequenceof relationsinthenetworkschema*T=(A,R)*, denotedinthe form$A_1(R_1)A_2(R_2)...(R_{(l-1)})A_l$,whichdefinesacompositerelation$PR_1oR_2o...oR_{(l1)}$betweentwonodes,where *o*isthecompositionoperatoronrelations.Forconvenience,a metapath can be represented by a sequence of node typeswhen there is no ambiguity, i.e., $PA_1A_2... A_l$. The metapathextends the concept of link types to path types and describesthedifferentrelationsamongnodetypesthroughindirectlinks, i.e.paths,andalsoimpliesdiversesemantics.

*Definition 4 (Classification Problem in Heterogeneous Infor-mation Networks):*Given a heterogeneous information net-work *G(V, E)*, suppose V^r is a subset of V that containsnodesofthetargettype(i.e.,thetypeofnodestobeclassified). *k*denotes the number of the class, and for each class, say$C_1...C_k$,wehave somepre-labeled nodes inV^r associated witha single user. The classification task is to predict the labels foralltheunlabelednodesinV^r.

B. *FeatureTypes*

In this paper, we use an extended definition of the metapathconcept as follows. A metapath is defined as a path betweentwonodes,whichindicatestheconnectionoftwonodesthrough their shared features. When we talk about metadata,werefertoitsgeneraldefinition,whichisdataaboutdata.In our case,the data isthe written review, and by metadatawe mean data about the reviews, including user who wrote thereview, the business that the review is written for, rating valueof the review, date of written review and finally its label asspamorgenuinereview.

In particular, in this work features for users and reviews fallintothecategoriesasfollows(showninTableI):

1) *Review-Behavioral (RB) Based Features:*This featuretype is based on metadata and not the review text itself. TheRB category contains two features; Early time frame

2) (ETF)andThresholdratingdeviationofreview(DEV)[16].

3) *Review-Linguistic (RL) Based Features:* Features in thiscategory are based on the review itself and extracted directlyfrom text of the review. In this work we use two main featuresin RL category; the Ratio of 1st Personal Pronouns (PP1) andtheRatioofexclamationsentencescontaining'!'(RES)[6].

4) *User-Behavioral (UB) Based Features:*These featuresarespecifictoeachindividual userandthey arecalculatedper user, so we can use these features to generalize all of thereviews written by that specific user. This category has twomain features; the Burstiness of reviews written by a singleuser [7], and the average of a users' negative ratio given todifferentbusinesses[25-30].

5) *User-Linguistic(UL)BasedFeatures:*Thesefeaturesare extracted from the users' language and shows how usersaredescribingtheirfeelingoropinionabout whatthey'veexperienced as a customer of a certain business. We use thistype of features to understand how a spammer communicatesin terms of wording. There are two features engaged for ourframeworkinthiscategory;AverageContentSimilarity

TABLEI

FEATURES FOR USERS AND REVIEWS IN FOUR DEFINED CATEGORIES (THE CALCULATED VALUES ARE BASED ON [12,TABLE 2])

Spam Feature	User-based	Review-based
Behavioral-based Features	$$x_{BST}(i) = \begin{cases} 0 & (L_i - F_i) \notin (0, \tau) \\ 1 - \frac{L_i - F_i}{\tau} & (L_i - F_i) \in (0, \tau) \end{cases} \quad (1)$$ where $L_i - F_i$ describes days between last and first review for $\tau = 28$. Users with calculated value greater than 0.5 take value 1 and others take 0.	*Early Time Frame* [16]: Spammers try to write their reviews asap, in order to keep their review in the top reviews which other users visit them sooner. $$x_{ETF}(i) = \begin{cases} 0 & (T_i - F_i) \notin (0, \delta) \\ 1 - \frac{T_i - F_i}{\delta} & (T_i - F_i) \in (0, \delta) \end{cases} \quad (2)$$ *Rate Deviation using threshold* [16]: Spammers, also tend to promote businesses they have contract with, so they rate these businesses with high scores. In result, there is high diversity in their given scores to different businesses which is the reason they have high variance and deviation. (3) where β_1 is some threshold determined by recursive minimal entropy partitioning. Reviews are close to each other based on their calculated value, take same values (in $[0, 1)$).
Linguistic-based Features		*Number of first Person Pronouns, Ratio of Exclamation Sentences containing '!'* [6]: First, studies show that spammers use second personal pronouns much more than first personal pronouns. In addition, spammers put '!' in their sentences as much as they can to increase impression on users and highlight their reviews among other ones. Reviews are close to each other based on their calculated value, take same values (in $[0, 1)$).

III. NetSpam:THEPROPOSEDSOLUTION

Inthissection,weprovidesdetailsoftheproposedsolution whichisshowninAlgorithmIII-1.

A. *PriorKnowledge*

The first step is computing prior knowledge, i.e. the initialprobability of review *u* being spam which denoted as y_u. Theproposed framework works in two versions; semi-supervisedlearning andunsupervised learning. Inthesemi-supervisedmethod,y_u1ifreview*u*islabeledasspaminthepre-labeled reviews, otherwisey_u0. If the label of thisreviewisunknownduetheamountofsupervision,weconsidery_u0 (i.e., we assume *u*as a non-spam review). In theunsupervisedmethod,ourpriorknowledgeisrealizedbyusing

y_u=(1/L) $\sum_{l=1}^{L}$ $f(x_{lu})$where$f(x_{lu})$istheprobabilityof reviewu beingspamaccordingtofeature *l*and*L*isthe

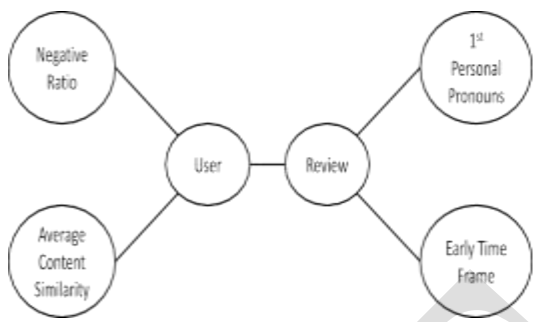

Fig. 1.An example for a network schema generated based on a given spamfeatureslist;NR,ACS,PP1andETF.

the metapaths used in the proposed framework. As shown, thelength of user-based metapaths is 4 and the length of review-basedmetapathsis2.

Formetapathcreation, wedefineanextended versionofthemetapathconceptconsideringdifferentlevelsofspam certainty.Inparticular,two reviewsareconnectedtoeachotheriftheysharesamevalue.Hassan

B. *NetworkSchemaDefinition*

Thenextstepisdefiningnetworkschemabasedonagiven listofspamfeatureswhichdetermines thefeatures engagedinspamdetection.ThisSchemaaregeneraldefinitionsofmeta-

uzzy-based framework and indicate for spam detection, it isbetter to use fuzzy logic for determining a review's label as aspam or non-spam. Indeed, there are different levels of spamcertainty. Weusea stepfunction to determine theselevels.Inparticular,givenareview*u*,thelevelsofspamcertainty for metapathp_l(i.e.,feature*l*)iscalculatedasm^{pl}

$= s \times f(xlu)]$,

C. *MetapathDefinitionandCreation*

AsmentionedinSectionII-A,ametapathisdefinedbyasequenceofrelationsinthenetworkschema.TableIIshowsalldenotedas$m^{pl}_{u,v}$ = m^{pl}_{u}.

Using*s*withahighervaluewillincreasethenumberofeachfeature'smetapathsandhencefewerreviewswouldbe

ISSN 2277-2685
IJESR/Sep. 2023/ Vol-13/Issue-3/1-19
Mr.Mrutyunjaya Yalawar *et. al.*, / International Journal of Engineering & Science Research

AlgorithmIII.1:xNetSpam()

Input: review-dataset,spam-feature-list,pre-labeled-reviews
Output:features-importance(W),spamicity-probability(Pr)
%u,v:review,y_u:spamicityprobabilityofreview$u$
%$f(x_{lu})$:initialprobabilityofreview$u$beingspam%metapathbasedonfeature$l$,L:featuresnumber%n:numberofreviewsconnectedtoareview
%m_u^{pl}:thelevelofspamcertainty
%$m_{u,v}^{pl}$:themetapathvalue
%PriorKnowledge
ifsemi-supervisedmode
if$u \in pre - labeled - reviews$
{y_u=*label(u)*
else
{y_u=0

D. Classification

The classificationpartof*NetSpam*includestwosteps;
(i) weight calculation which determines the importance of eachspam feature in spotting spam reviews, (ii) Labeling whichcalculates the final probability of each review being spam.Nextwedescribethemindetail.

1) *Weight Calculation:*This step computes the weight ofeach metapath. We assume that nodes' classification is donebased on their relations to other nodes in the review network;linked nodes may have a high probability of taking the samelabels. The relations in a heterogeneous information networknot only include the direct link but also the path that can bemeasured by using the metapath concept. Therefore, we needto utilize the metapaths defined in the previous step, whichrepresent heterogeneous relations among nodes. Moreover, thisstep will be able to compute the weight of each relation path(i.e.,theimportanceofthemetapath),whichwillbeusedin thenextstep(Labeling)toestimatethelabelofeachunlabeled

: TABLEII

METAPATHSUSEDINTHENetSpamFRAMEWORK

Row		Type		
		RB	Review-Threshold Rate Deviation-Review	Reviews with same Rate Deviation from average Item rate (based on recursive minimal entropy partitioning)
		UB	Review-User-Negative Ratio-User-Review	Reviews written by different Users with same Negative Ratio
		RB	Review-Early Time Frame-Review	Reviews with same released date related to Item
		UB	Review-User-Burstiness-User-Review	Reviews written by different users in same Burst
		RL		Reviews with same number of Exclamation Sentences containing '!'
		RL	Review-first Person Pronouns-Review	Reviews with same number of first Person Pronouns
		UL	Review-User-Average Content Similarity-User-Review	Reviews written by different Users with same Average Content Similarity using cosine similarity score
		UL		Reviews written by different Users with same Maximum Content Similarity using cosine similarity score

TABLEIII

REVIEW DATASETSUSED IN THIS WORK

	Reviews (spam%)	Users	Business (Resto. & hotels)

260,277
48,121

Main	608,598 (13%)		5,044
Review-based	62,990 (13%)		3,278
Item-based	66,841 (34%)		4,588
User-based	183,963 (19%)		4,568
		7685	243

lots of links with non-spam reviews, it means that it sharesfeatures with other reviews with low spamicity and hence itsprobabilitytobeanon-spamreviewincreases.

IV. EXPERIMENTALEVALUATION

Thissectionpresentstheexperimentalevaluationpartofthisstudyincludingthedatasetsandthedefinedmetricsaswellastheobtainedresults. We used a dataset from Yelp, introduced in [12],which includes almost 608,598 reviews written by customersof restaurants and hotels in NYC.

A. Datasets

Table III includes a summary of the datasets and their char-acteristics. We used a dataset from Yelp, introduced in [31],which includes almost 608,598 reviews written by customersof restaurants and hotels in NYC. The dataset includes thereviewers' impressions and comments about the quality, andother aspects related to a restaurants (or hotels). The datasetalso contains labeled reviews as ground truth (so-called nearground-truth [32]), which indicates whether a reviewis amornot.Yelpdatasetwaslabeledusingfilteringalgorithm

- Item-baseddataset, composes of 10% of the randomlyselected reviews of each item, also based on uniform distrib-ution(aswithReview-baseddataset).

- User-baseddataset, includes randomly selected reviewsusing uniform distribution in which one review is selectedfrom every 10 reviews of single user and if number of reviewswas less than 10, uniform distribution has been changed inordertoatleastonereviewfromeveryusergetselected.

In addition to the presented dataset, we also used anotherreal-worldsetofdatafromAmazon[32-34]toevaluateourworkonunsupervised mode.There isnocrediblelabel intheAmazondataset(asmentionedin[35]),butweusedthis dataset to show how much our idea is viable on otherdatasets beyond Yelp and results for this dataset is presentedonSec.IV-C3.

B. EvaluationMetrics

We have used Average Precision (AP) and Area Under theCurve (AUC) as two metrics in our evaluation. AUC measuresaccuracy of our ranking based onFalsePositive Ratio(FPRas y-axis) against True Positive Ratio (TPR as x-axis) andintegratevalues basedonthesetwomeasured values.Thevalueofthismetricincreasesastheproposedmethodperformswellinranking,andvise-versa.LetAbethelistofsortedspamreviewssothat$A(i)$denotesareviewsortedonthei^{th}indexin A. If the number of spam (non-spam) reviews before reviewin thej^{th}index is equal to $n_j$and the totalengagedbytheYelprecommender,andalthoughnoneofrecommendersareperfect,butaccordingto[36]itproduces

etarerateofreviewers,thedateofthewrittenreview,anddateof actual visit, as well as the user's and the restaurant's id(name).

We created three other datasets from this main dataset asfollow:

-Review-baseddataset,includes10%ofthereviewsintegratetheareaunderthecurveforthecurvethatusestheir values.WeobtainavaluefortheAUCusing:

$$AUC=\sum_{i=2}^{n}(FPR(i)-FPR(i-1))*(TPR(i)) \quad (7)$$

where $n$denotes number ofreviews. ForAP wefirstneedto calculate index of top sorted reviews with spam labels. Letindexes of sorted spam reviews in list A with spam labels

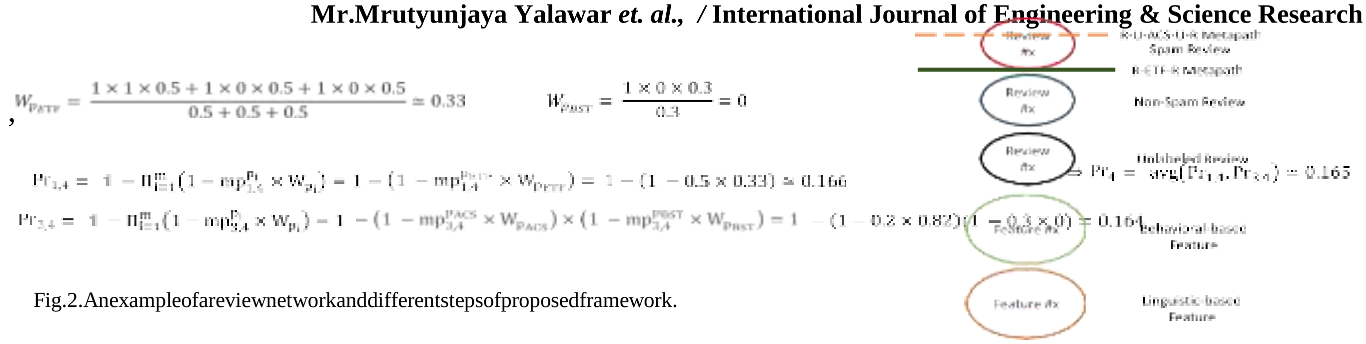

Fig.2.Anexampleofareviewnetworkanddifferentstepsofproposedframework.

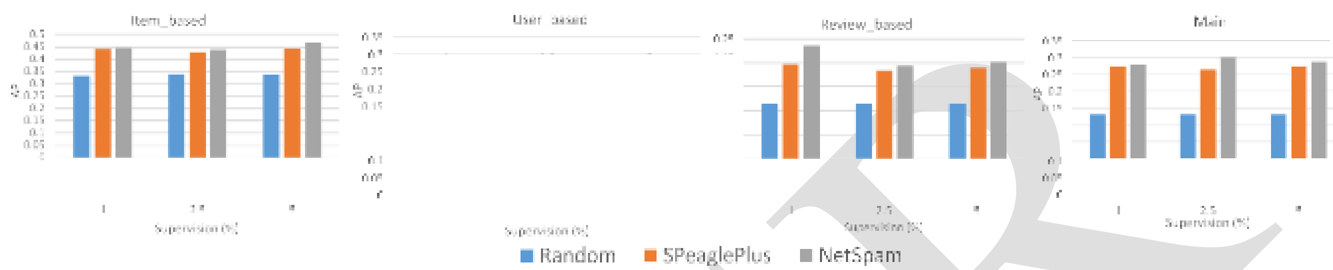

Fig.3.APforRandom,SPeaglePlusandNetSpamapproachesindifferentdatasetsandsupervisions(1%,2.5%and5%).

Fig.4.AUCforRandom,SPeaglePlusandNetSpamapproachesindifferentdatasetsandsupervisions(1%,2.5%and5%).

As the first step, two metrics are rank-based which meanswe can rank the final probabilities. Next we calculate the APand AUC values based on thereviews' ranking in thefinallist.

Inthemostoptimum situation,allofthespamreviewsarerankedontopofsortedlist;Inotherwords,whenwesort spam probabilities for reviews, all of the reviews withspamlabelsarelocatedontopofthelistandrankedasthefirstreviews.WiththisassumptionwecancalculatetheAPandAUCvalues.Theyarebothhighlydependentonthenumberoffeatures.Forthelearningprocess,weusedifferentsupervisionsandwetrainasetforweightcalculation.Wealsoengagethesesupervisionsasfunda-mentallabelsforreviewswhicharechosenasatrainingset.

C. MainResults

In this section, we evaluate NetSpam from different per-spective and compare it with two other approaches, Randomapproach and SPeaglePlus [36]. Tocompare withthefirstone, we have developed a network in which reviews are con-nected to each other randomly. Second approach use a well-known graph-based algorithm called as "LBP" to calculatefinallabels.OurobservationsshowNetSpam,outperformsthese existing methods. Then analysis on our observation isperformedandfinallywewillexamineourframeworkinunsupervised mode. Lastly, we investigate time complexity oftheproposedframeworkandtheimpactofcamouflagestrategyonitsperformance.

*1) Accuracy:*Figures 3 and 4 present the performance intermsoftheAPandAUC.Asit'sshowninallofthefour

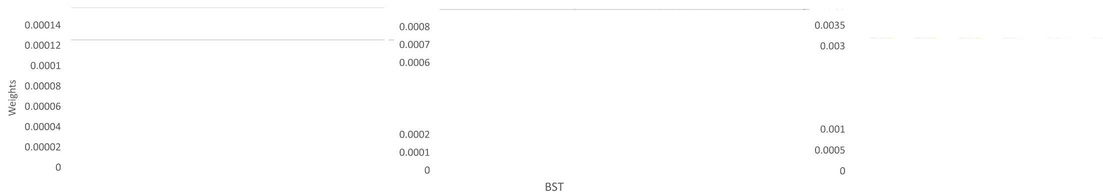

Fig.5.FeaturesweightsforNetSpamframeworkondifferentdatasetsusingdifferentsupervisions(1%,2.5%and5%).

datasetsNetSpamoutperformsSPeaglePlusspeciallywhennumber of features increase. In addition different supervisionshave no considerable effect on the metric values neither onNetSpam nor SPeaglePlus. Results also show the datasets withhigher percentage of spam reviews have better performancebecause when fraction of spam reviews in a certain datasetincreases,probabilityforareviewtobeaspamreviewincreases and as a result more spam reviews will be labeled asspam reviews and in the result of AP measure which is highlydependent on spam percentage in a dataset. On the other hand,AUCmeasuredoesnotfluctuatetoomuch,becausethismetricisnotdependentonspamreviewspercentageindataset,buton the final sorted list which is calculated based on the finalspamprobability.

*2) Feature Weights Analysis:*Next we discuss about fea-tures weights and their involvement to determine spamicity.First we inspect how much APand AUCare dependent onvariable number of features. Then we show these metrics aredifferent for the four feature types explained before (RB, UB,RLandUL).Toshowhowmuchourworkonweightscalculationiseffective, firstwehavesimulatedframeworkon several run with whole features and used most weightedfeatures to find out best combination which gives us the bestresults. Finally, we found which category is most effectivecategoryamongthoselistedinTableI.

*a) Datasetimpressiononspamdetection:*Asweexplained previously, different datasets yield different resultsbased on their contents. For all datasets and most weightedfeatures, there is a certain sequence for features weights. As isshowninFig.5forfourdatasets,inalmostallofthem,features for the Main dataset have more weights and featuresfor Review-based dataset stand in the second position. Thirdposition belongs to User-based dataset and finally Item-baseddataset has the minimum weights (for at least the four featureswithmostweights).

b) Features weights importance: As shown in Table IV,there are couple of features which are more weighted thanothers.Combinationofthesefeaturescanbeagoodhintforobtainingbetterperformance. TheresultsoftheMaindatasetshowallthefour behavioral features areranked asfirst features in the final overall weights. In addition, as shownin the Review-based as well as other two datasets,DEV isthe most weighted feature. This is also same for our secondmost weighted feature, $N\ R$. From the third feature to the lastfeaturetherearedifferentorderforthementionedfeatures.Thethird feature for both datasets User-based and Review-based issame,ETF,whilefortheotherdataset,Item-based,$PP1$is

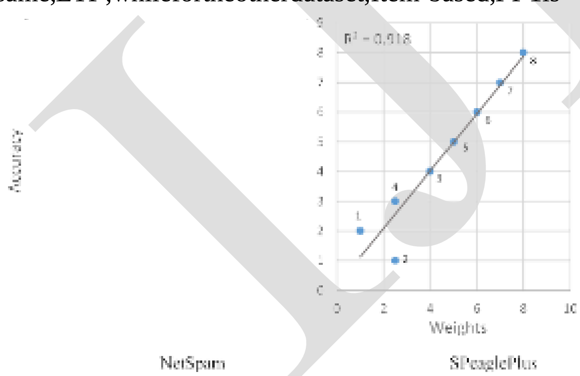

Fig. 6. Regression graph of features vs. accuracy (with **5% data as trainset**)forMaindataset.(seeTableIIfornumbers).

atrank3.Goingfurther,weseeintheReview-baseddatasetall four most weighted features are behavioral-based featureswhichshowshowmuchthistypeoffeaturesareimpor-tant in detecting spams as acknowledged by other works aswell[37],[38].

AswecanseeinFig.6,thereisastrongcorrelationbetweenfeaturesweightsandtheaccuracy.FortheMaindatasetwecansee this correlation is much more obvious and also applicable.Calculating weights using NetSpam help us to understand howmuch a feature is effective in detecting spam reviews; since asmuch as their weights increase two metrics including AP andAUC also increase respectively and therefore our frameworkcan be helpful in detecting spam reviews based on featuresimportance.

The observations indicate larger datasets yield better cor-relationbetweenfeaturesweightsandalsoitsaccuracyinterm of AP. Since we need to know each feature rank andimportance we use Spearman's rank correlation for our work.Inthisexperienceourmaindatasethascorrelationvalueequalto0.838(p-value=0.009),whilethisvalueforournext dataset, User-based one, is equal to0.715 (p-value =0.046). As much as the size of dataset gets smaller in theexperiment,thisvaluedrops.Thisproblemismoreobviousin TABLEIV

WEIGHTSOF ALL FEATURES (WITH **5%DATA AS TRAIN SET**);FEATURES ARE RANKED BASED ON THEIR OVERALL AVERAGE WEIGHTS

Dataset - Weights	DEV	NR	ETF	BST	RES	PPI	ACS	MCS
Main	0.0029	0.0032	0.0015	0.0029	0.0010	0.0011	0.0003	0.0002
Review-based	0.0023	0.0017	0.0017	0.0015	0.0010	0.0009	0.0004	0.0003
Item-based	0.0010	0.0012	0.0009	0.0009	0.0010	0.0010	0.0004	0.0003
User-based	0.0017	0.0014	0.0014	0.0010	0.0010	0.0009	0.0005	0.0004

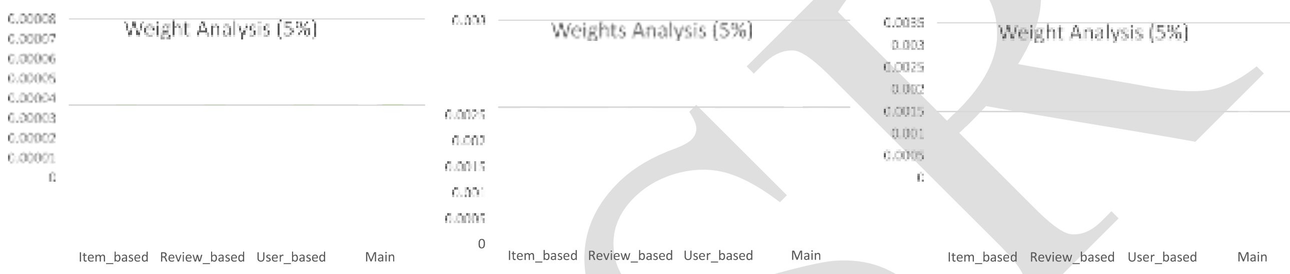

Fig.7.Featuresweightsfordifferentfeaturescategories(RB,UB,RLandUL)with5%supervision,ondifferentdatasets.

ofAP)arecompletelycorrelated.Weobservedvalues0.958(p-value=0.0001),0.764(p=0.0274),0.711(p=0.0481)and 0.874(p=0.0045) fortheMain,User-based,Item-based andReview-baseddatasets,respectively.Thisresultshowsusingweightcalculationmethodandconsideringmetapathconceptcanbeeffectiveindeterminingtheimportanceoffeatures.SimilarresultforSPeaglePlusalsoshowsourweightscalcu-lationmethod canbegeneralizedtoother frameworksandcanbeusedasamaincomponentforfindingeachfeatureweight.Ourresultsalsoindicatefeatureweightsarecompletelydependentondatasets,consideringthisfacttwomostimpor-tantfeaturesinalldatasetsaresamefeatures.Thismeansexceptthefirsttwofeatures,otherfeaturesweightsarehighlyvariableregradingtodatasetusedforextractingweightsof features.

c) Features category analysis: As shown in Fig. 7 thereare four categories with different weights average which isvery important, specially in determining which feature is moreappropriate for spotting spam reviews (refer to Sec. IV-C.2.b).resimilarwehavejustpresentedtheresultsfor5%supervision.Wehaveanalyzedfeaturesbasedontheircategoriesandobtainedresultsinalldatasetsshowthat Behavioralbasedfeatureshavebetterweightsthanlinguisticoneswhichisconfirmedby[39]and [16].Analysisonseparateviewsshowsthatreview-based features have higher weights which leads to bet-ter performance. It is worth to mention that none of previousworkshaveinvestigatedthisbefore.SameanalysisontheMaindataset shows equal importance of both category in findingspams.OntheOtherhand,inthefirstthreedatasetfromTable I, RB has better weights (a bit difference in comparisonwithRU),whichmeansthiscategoryyieldsbetterperformancethan other categories for spotting spam reviews. Differently,for Main dataset UB categories has better weights and has bet-ter performance than RU category and also other categories, inall datasets behavioral-based features yield better performancewithanysupervision.

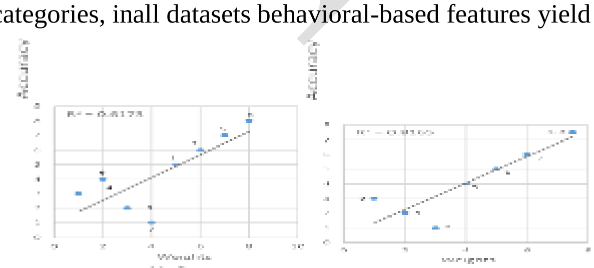

Fig.8.Regressiongraphoffeaturesvs.accuracy(**unsupervised**)forMain dataset.(seeTableIIfornumbers).

ISSN 2277-2685
IJESR/Sep. 2023/ Vol-13/Issue-3/1-19
Mr.Mrutyunjaya Yalawar *et. al.*, / International Journal of Engineering & Science Research

*3) Unsupervised Method:*One of the achievement in thisstudy isthat even without using atrain set,we can stillfindthebest set offeatures which yieldtothebest performance.As itisexplained in Sec. III-A, in unsupervised approachspecialformulation isusedtocalculatefundamental labelsand next these labels are used to calculate the features' weightand finally review labels. As shown in Fig. 8, our observationsshow there is a good correlation in the Main dataset in whichforNetSpamitisequalto0.78(p-value=0.0208) andforSPeaglePlusthisvaluereach0.90(p=0.0021).Asanotherexample for user-baseddatasetthereisacorrelation equalto0.93(p=0.0006)forNetSpam,whileforSPeaglethisvalue is equal to 0.89 (p=0.0024). This observation indicatesNetSpam can prioritize features for both frameworks. Table Vdemonstrates that there is certain sequence in feature weightsanditmeansinspamdetectionproblems,spammersandspam reviews have common behaviors, no matter what socialnetwork they are writing the review for: Amazon or Yelp. Forall of them, DEV is most weighted features, followed by NR,ETFandBST.

*4) Time Complexity:*If we consider the Main dataset asinput to our framework, time complexity with these circum-stancesisequalto$O(e^2m)$where$e$isnumberofedgesin

TABLEV

WEIGHTS OF ALL FEATURES (USING **UNSUPERVISED**APPROACH);FEATURES ARE RANKED BASED ON THEIR OVERALL AVERAGE WEIGHTS

Dataset	DEV	NR	ETF	BST	RES	PPI		MCS
M	0.0029	0.0550	0.0484	0.0445	0.0379	0.0329	(	0.031
Reviev	0.0626	0.0510	0.0477	0.0376	0.0355	0.0346	(	0.034
Item-	0.0638	0.0510	0.0501	0.0395	0.0388	0.0383	(	0.036
User-	0.0630	0.0514	0.0494	0.0380	0.0373	0.0377	(	0.036
Am	0.1102	0.0897	0.0746	0.0689	0.0675	0.0624	(	0.029

creatednetworkorreviewsnumber.Itmeansweneedtocheckif there is a metapath between a certain node (review) withother nodes which is $O(e^2)$ and this checking must be repeatedfor very feature. So, our time complexity for offline mode inwhich we give the Main dataset to framework and calculatespamicity ofwhole reviews, is$O(e^2m)$ where $m$isnumberof features.In online mode, a review is given to NetSpam tosee whether it is spam or not, we need to check if there is ametapathbetweengivenreviewwithotherreviews,whichisin $O(e)$, and like offline mode it has to be repeated for everyfeatureandeveryvalue.Thereforethecomplexityis$O(em)$.

*5) The Impact of Camouflage Strategy:*One of the chal-lenges that spam detection approaches face is that spammersoften write non-spam reviews to hide their true identity knownas camouflage. For example they write positive reviews forgoodrestaurantornegativereviewsforlow-qualityones;henceeveryspamdetectorsystemfailstoidentifythiskindofspammers or at least has some trouble to spot them. In theprevious studies, there are different approaches for handlingthis problem. For example, in [12], the authors assumes thereis always a little probability that a good review written by aspammer and put this assumption in itscompatibility matrix.Inthisstudy,wetriedtohandlethisproblembyusingweightedmetapaths.Inparticular,weassumethatevenifa review has a very little value for a certain feature, itisconsidered infeature weightscalculation. Therefore, insteadofconsideringmetapathsasbinaryconcepts,wetake20valueswhichdenotedas$s$.Indeed,ifthereisacamouflageitsaffection willbe reduced. Asweexplained inSection III-Cin such problems it is better to propose a fuzzy framework,ratherthanusingabipolarvalues$(0,1)$.

V. RELATEDWORKS

In the last decade, a great number of research studies focusontheproblemofspottingspammersandspamreviews.However,sincetheproblemisnon-trivialandchallenging,itremainsfarfromfullysolved.Wecansummarizeourdiscussionaboutpreviousstudiesinthreefollowingcategories.

A. Linguistic-BasedMethods

This approach extract linguistic-based features to find spamreviews. Feng*etal.*[13]use$unigram$, $bigram$ andtheircomposition. Other studies [4], [6], [15] use other features likepairwise features (features between two reviews; e.g. contentsimilarity), percentage of CAPITAL words in a reviews forfinding spam reviews. Lai *et al.* in [33] use a probabilisticlanguage modeling to spot spam. This study demonstrates that2%ofreviewswrittenonbusinesswebsites are actuallyspam.

B. Behavior-BasedMethods

Approachesinthisgroupalmostusereviewsmetadatatoextractfeatures;thosewhicharenormalpatternofareviewer behaviors. Feng*etal.* in[21] focus ondistributionofspammers ratingondifferent products andtracesthem.In[34],Jindal*et.al*extract36behavioralfeaturesandusea supervised method to find spammers on Amazon and [14]indicates behavioral features show spammers' identity betterthanlinguisticones.Xue*etal*.in[32]useratedeviationofa specific user and use a trust-aware model to find the rela-tionship between users for calculating final spamicity score.Minnich *et al.* in [8] use temporal and location features ofusers to find unusual behavior of spammers. Li *et al.* in [10]usesomebasicfeatures(e.gpolarityofreviews)andthenruna HNC (Heterogeneous Network Classifier) to find final labelson Dianpings dataset. Mukherjee *et al.* in [16] almost engagebehavioralfeatureslikeratedeviation,extremityandetc.Xie *et al.* in [17] also use a temporal pattern (time window) tofind singleton reviews (reviews written just once) on Amazon.LucaandZervasin[26]usebehavioralfeaturestoshowincreasing competition between companies leads to very largeexpansionofspamreviewsonproducts.

Crawford *et al.* in [28] indicates using different classifi-cation approach need different number of features to attaindesired performance and propose approaches which use fewerfeatures to attain that performance and hence recommend toimprove their performance while they use fewer features whichleads them tohave better complexity. With this perspectiveour framework is arguable. This study shows using differentapproachesinclassificationyielddifferentperformanceintermsofdifferentmetrics.

C. Graph-BasedMethods

Studies in this group aim to make a graph between users,reviewsanditemsanduseconnectionsinthegraphandalso some network-based algorithms to rank or label reviews(asspamorgenuine)andusers(asspammerorhonest).Akoglu *et al.* in [11] use a network-based algorithm known asLBP (Loopy Belief Propagation) in linearly scalable iterationsrelatedtonumberofedgestofindfinalprobabilitiesfordifferent components innetwork. Fei*etal*.in[7] alsousesame algorithm (LBP), and utilize burstiness of each review tofind spammers and spam reviews on Amazon. Li *et al.* in [10]build a graph of users, reviews, users IP and indicates userswith same IP have same labels, for example if a user withmultiple different account and same IP writes some

reviews,theyaresupposedtohavesamelabel.Wang*etal*.in[18]alsocreateanetworkofusers,reviewsanditemsanduse

basic assumptions(forexampleareviewerismoretrustworthyifhe/shewritesmorehonestreviews)andlabelreviews.Wahyuni and Djunaidy in [37] proposes a hybrid method forspam detection using an algorithm called ICF++ which is anextension to ICFof [18] in which justreview rating are usedto find spam detection. This work use also sentiment

analysistoachievebetteraccuracyinparticular.

Deeper analysis on literature show that behavioral featureswork better than linguistic ones in term of accuracy they yield.Thereisagood explanation forthat;ingeneral, spammerstend to hide their identity for security reasons. Therefore theyare hardly recognized by reviews [38] they write about products,but their behavior is still unusual, no matter what languagethey are writing. In result, researchers combined both featuretypes to increase accuracy of spam detection. The fact thatadding each feature is a time consuming process, this is wherefeature importance isuseful.Basedonour knowledge, thereis no previous method which engage importance of features(knownasweightsinourproposedframework;

NetSpam)intheclassificationstep.Byusingtheseweights,ononehand we involve features importance in calculating final labelsand hence accuracy of NetSpam increase, gradually. On theother hand we can determine which feature can provide betterperformance in term of their involvement in connecting spamreviews(inproposednetwork).

VI. CONCLUSION

This study introduces a novel spam detection frameworknamelyNetSpambasedonametapathconceptaswellasanewgraph-based method to label reviews relying on a rank-basedlabeling approach. The performance of the proposed frame-workisevaluatedbyusingtworeal-worldlabeleddatasetsof Yelp and Amazon websites. Our observations show thatcalculated weights by using this metapath concept can be veryeffective in identifying spam reviews and leads to a betterperformance. In addition, we found that even without a trainset,NetSpam cancalculate the importance

ofeachfeatureandityieldsbetterperformanceinthefeatures'additionprocess, and performs better than previous works, with only asmall number of features. Moreover, after defining four maincategories for features our observations show that the reviews-behavioral category performs better than other categories, interms of AP, AUC as well as in the calculated weights. Theresults alsoconfirm that using different supervisions, similarto the semi-supervised method, have no noticeable effect ondetermining most of the weighted features, just as in differentdatasets.

For future work, metapath concept can be applied to otherproblems in this field. For example, similar framework can beused to find spammer communities. For finding community,reviews can be connected through group spammer features(suchastheproposedfeaturein[39])andreviewswithhighestsimilaritybased onmetapth concept areknown ascommunities. In addition, utilizing the product features is aninteresting future work on this study as we used features morerelated to spotting spammers and spam reviews.

Moreover,whilesinglenetworkshasreceivedconsiderableattentionfromvariousdisciplinesforoveradecade,informationdiffusionand content sharing in multilayer networks is still a youngresearch [37]. Addressing the problem of spam detection insuch networks can beconsidered asanew research lineinthisfield.

REFERENCES

[1] J. Donfro.*A Whopping20% ofYelp Reviewsare Fake*,accessedonJul. 30, 2015. [Online]. Available: http://www.businessinsider.com/20-percent-of-yelp-reviews-fake-2013-9

[2] M. Ott, C. Cardie, and J. T. Hancock, “Estimating the prevalence ofdeception in online review communities,” in *Proc. ACM WWW,* 2012,pp.201–210.

[3] M. Ott, Y. Choi, C. Cardie,and J. T. Hancock,“Finding deceptiveopinion spam by any stretch of the imagination,” in *Proc. ACL,* 2011,pp.309–319.

[4] C. Xu and J. Zhang, “Combating product review spam campaigns viamultipleheterogeneouspairwisefeatures,”in*Proc.SIAMInt.Conf.DataMining*,2014,pp.172–180.

[5] N. Jindal and B. Liu, “Opinion spam and analysis,” in *Proc. WSDM*,2008,pp.219–230.

[6] F. H. Li, M. Huang, Y. Yang, and X. Zhu, “Learning to identify reviewspam,”in*Proc.22ndInt.JointConf.Artif.Intell.(IJCAI)*,2011,pp.1–6.

[7] G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh,“Exploiting burstiness in reviewsfor review spammer detection,”in*Proc.ICWSM*,2013,pp.1–10.

[8] A. J. Minnich, N. Chavoshi, A. Mueen, S. Luan, and M. Faloutsos,“Trueview: Harnessing the power of multiple review sites,” in *Proc.ACMWWW*,2015,pp.787–797.

[9] B. Viswanath *et al.*, “Towards detecting anomalous user behavior inonlinesocialnetworks,”in*Proc.USENIX*,2014,pp.1–16.

[10] H. Li, Z. Chen,B.Liu, X. Wei,andJ. Shao,“Spottingfakereviewsvia collective positive-unlabeled learning,” in *Proc. ICDM,* Dec. 2014,pp.899–904.

[11] L. Akoglu, R. Chandy, and C. Faloutsos, “Opinion fraud detection inonlinereviewsbynetworkeffects,”in*Proc.ICWSM*,2013,pp.1–10.

[12] S. Rayana and L. Akoglu, “Collective opinion spam detection: Bridgingreviewnetworksandmetadata,”in*Proc.ACMKDD*,2015,pp.1–10.

[13] S. Feng, R. Banerjee,and Y. Choi, “Syntacticstylometry for decep-tion detection,” in *Proc. 50th Annu. Meeting Assoc. Comput. Linguis-tics(ACL)*,2012,pp.1–5.

[14] N. Jindal, B. Liu, and E.-P. Lim, “Finding unusual review patterns usingunexpectedrules,”in*Proc.ACMCIKM*,2012,pp.1–4.

[15] E.-P.Lim,V.-A.Nguyen,N.Jindal,B.Liu,andH.W.Lauw,“Detectingproduct review spammers using rating behaviors,” in *Proc. ACM CIKM*,2010,pp.1–10.

[16] A.Mukherjee*etal.*,“Spottingopinionspammersusingbehavioralfootprints,”in*Proc.ACMKDD*,2013,pp.1–9.

[17] S. Xie, G. Wang, S. Lin, and P. S. Yu, “Review spam detection viatemporalpatterndiscovery,”in*Proc.ACMKDD*,2012,pp.823–831.

[18] G. Wang, S. Xie, B. Liu, and P. S. Yu, “Reviewgraph based onlinestorereview spammerdetection,”in*Proc.IEEEICDM*,Dec.2011,pp.1242–1247.

[19] Y.SunandJ.Han,“Miningheterogeneousinformationnetworks:Principlesandmethodologies,”in*Proc.ICCCE*,2012,p.159.

[20] A.Mukherjee,V.Venkataraman,B.Liu,andN.Glance,“Whatyelpfakereviewfiltermightbedoing?”in*Proc.ICWSM*,2013,pp.409–418.

[21] S. Feng, L. Xing, A. Gogar, and Y. Choi, “Distributional footprints ofdeceptiveproductreviews,”in*Proc.ICWSM*,2012,pp.98–105.

[22] Y. Sun, J. Han, X. Yan, P. S. Yu, and T. Wu, “PathSim: Meta path-based top-K similarity search in heterogeneous information networks,”in*Proc.VLDB*,2011,pp.1–12.

[23] Y. Sun, J. Han, P. Zhao, Z. Yin, H. Cheng, and T. Wu, “RankClus: Inte-grating clustering with ranking for heterogeneous information networkanalysis,” in *Proc. 12th Int. Conf. Extending Database Technol., Adv.DatabaseTechnol.*,2009,pp.1–12.

[24] C.Luo,R.Guan,Z.Wang,andC.Lin,“HetPathMine:Anoveltransduc-tive classification algorithm on heterogeneous information networks,” in*Proc.ECIR*,2014,pp.210–221.

[25] R. Hassanzadeh, “Anomaly detection in online social networks: Usingdatamining techniques and fuzzy logic,” School Elect. Eng. Comput.Sci.,QueenslandUniv.Technol.,Brisbane,QLD,Australia,Nov.2014.

[26] M.LucaandG.Zervas,“Fakeittillyoumakeit:Reputation,competition,and yelpreviewfraud,”*Manage.Sci.*, vol.62, no. 16,pp.3412–3427,Jan.2016.

[27] VM Allocation Technique and Optimized Performance Improvement for the Cloud Architecture. Authors: Dr. Md. Rafeeq, N. Navneetha, Dr. N. Subhash Chandra, M. Bhargavi,Dr.KRajeshwarRao

[29] Vempati Krishna , G Sumalatha , A L Sreenivasulu , M Bhargavi “Robust and Effective Spam Filtering From Online Social Media Networks Using Machine Learning Strategies”,Published in Journal of Innovation in Information Technology, Vol. 5(1), Jan–Jun 2021@ ISSN:2581-723X

[28] Teja Sree, N., Sumalatha, G. (2021). Behavioural Analysis Based Risk Assessment in Online Social Networks. In: Kumar, A., Mozar, S. (eds) ICCCE 2020. Lecture Notes in Electrical Engineering, vol 698. Springer, Singapore. https://doi.org/10.1007/978-9815-7961-5_25

[29] Swathi Mattaparthi, Sheo Kumar, Mrutyunjaya S Yalawar, Fake Currency Detection: A Survey on Different Methodologies Using Machine Learning Techniques, 2022/4/29, International Conference on Communications and Cyber Physical Engineering 2018, 463-468, Publisher, Springer Nature Singapore. https://doi.org/10.1007/978-981-19-8086-2_45.

[30] Sheo Kumar, Mrutyunjaya S Yalawar, K-NN SEMANTIC Inquiry on Scrambled Social Information BASE, ICDSMLA 2019: Proceedings of the 1st International Conference on Data Science, Machine Learning and Applications,982-991, 2020,Springer Singapore.

[31] K Vijaya Babu, Mrutyunjaya S Yalawar, G Sumalatha, G Ramesh Babu, Ravi Kumar Chandu, An Overview of Various Security Issues and Application Challenges of the Attacks in Field of Blockchain Technology, 2022/5/16, ICCCE 2021: Proceedings of the 4th International Conference on Communications and Cyber Physical Engineering, 365-374, Springer Nature Singapore.

[32] Mrutyunjaya S Yalawar, K Vijaya Babu, Bairy Mahender, Hareran Singh, A Brain-Inspired Cognitive Control Framework for Artificial Intelligence Dynamic System,2022/4/29,International Conference on Communications and Cyber Physical Engineering 2018,735-745,Springer Nature Singapore. https://doi.org/10.1007/978-981-19-8086-2_70.

[33] Mrutyunjaya S Yalawar, L Umasankar, Mr Bottu Gurunadha Rao, RESIDUE ANALYSIS OF IMAGE QUALITY USING TENSORFLOW, JOURNAL OF CRITICAL REVIEWS, 7,ISSUE 04,1742-1746,2020

[34] Kumar, S., Sai Lavanya, G.V.R. (2020). Face Recognition Using Open CV with Deep Learning. In: Kumar, A., Paprzycki, M., Gunjan, V. (eds) ICDSMLA 2019. Lecture Notes in Electrical Engineering, vol 601. Springer, Singapore. https://doi.org/10.1007/978-981-15-1420-3_122.

[35] L. Pullagura, N. Kittad, G. Diwakar, V. Sathiya, A. Kumar and M. S. Yalawar, "ML based Parkinson's Disease Identification using Gait Parameters," 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, 2022, pp. 561-566, doi: 10.1109/ICACRS55517.2022.10029281.

[36] Dr. C.N. RAVi , Karthikeyan Udaichi1 Miguel Garcia-Torres3,Parameshchari Bidare Divakarachari4 Large-scale system identification using self-adaptive penguin search algorithm, IET Control Theory & Applications,DOI: 10.1049/cth2.12479: , Received: 26 November 2022 Revised: 14, March 2023

[37] Shrivastava, R., Jain, M., Vishwakarma, S.K., Bhagyalakshmi, L., Tiwari, R. (2023), "Cross-Cultural Translation Studies in the Context of Artificial Intelligence: Challenges and Strategies". In: Kumar, A., Mozar, S., Haase, J. (eds) Advances in Cognitive Science and Communications. ICCCE 2022. Cognitive Science and Technology. Springer, Singapore, ISBN: 978-981-19-8086-2_9, pp 91-98. https://doi.org/10.1007/978-981-19-8086-2_9

[38] TarunDharDiwan, Rajesh Tiwari and VivekDubey, "Local Binary Pattern Occurrence Map Method for High Parallel Image Processing", International Conference on Advances in Computing and Communication held at NIT Hamirpur, Himachal Pradesh, India on 8 -10 April 2011, pp 538 – 540, ISBN: 978-81-920874-0-5

[39]Naveen Kumar Sahu and Rajesh Tiwari, "Comparative Analysis of Optimization Algorithms based on Hybrid Soft Computing Algorithm", International Journal for Scientific Research & Development, Vol. 3, Issue 09, 2015, pp 33 – 39, ISSN (online): 2321-0613

ISSN 2277-2685
IJESR/Sep. 2023/ Vol-13/Issue-3/1-19
Mr.Mrutyunjaya Yalawar *et. al.*, / International Journal of Engineering & Science Research