International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-3, January 2020

Detecting Malicious Facebook Applications using LSTM Algorithm

Ayesha Choudhari, Sunil B. Mane

***Abstract*: With twenty million introduces a day [1], outsider applications area unit a big purpose behind the acknowledgment and addictiveness of Facebook. Tragically, programmers have finished the aptitude of exploitation applications for spreading malware and spam. the problem is as of currently important, as we discover that a minimum of thirteen you look after applications in our dataset sq. estimates malignant. Up hitherto, the examination network has fixated on police examination vindictive posts and battles. we tend to tend to point out the issue: Given a Facebook application, can we tend to tend to affirm if it's vindictive? Our key commitment is in making LSTM—Facebook's Rigorous Application Evaluator—ostensibly the essential instrument fixated on police examination malevolent applications on Facebook. to make LSTM, we tend to tend to utilize info assembled by perceptive the posting conduct of 111K Facebook applications seen crosswise over 2.2 million purchasers on Facebook. to start with, we tend to tend to come to a decision plenty of decisions that encourage North yankee nation to acknowledge harmful applications from kind ones. for instance, we discover that vindictive applications normally share names with elective applications, which they for the foremost half demand less authorizations than amiable applications. Second, contributory these recognizing decisions, we tend to show that LSTM can find malignant applications with ninety nine .5% exactness, with no imitative positives and a high obvious positive rate (95.9%). At long last, we tend to tend to research arrange of harmful Facebook applications and distinguish parts that these applications use to unfold. Curiously, we discover that few applications get together and bolster every other; in our dataset, we discover 1584 applications endorsing the being engendering of 3723 choice applications through their posts. Long haul, we tend to tend to examine LSTM as a stage toward creating Associate in Nursing freelance working dog for application appraisal and positioning, during this manner on caution Facebook purchasers before putting in applications.*

Keywords: **Facebook apps, malicious, on-line social networks, spam, LSTM, Machine learning.**

I. INTRODUCTION

ONLINE informal communities (OSNs) modification and energize outsider (applications) to fortify the shopper talent on these stages. Such enhancements encapsulate fascinating or redirecting ways in which of demonstration among on-line companions and numerous exercises like taking part in games or specializing in melodies.

Revised Manuscript Received on January 30, 2020.
***** Correspondence Author
Mrs.Ayesha Choudhari, *, M.Tech in the Department of Computer Engineering at Government College of Engineering, Pune, Maharashtra Email: ayesha2604@gmail.com
Dr.Sunil B. Mane, Associate Professor in the Department of Computer Engineering and Information Technology at Government College of Engineering Pune (An Autonomous Institute of Govt. of Maharashtra), India.

as an example, Facebook provides designers associate degree API [2] that encourages application coordination into Facebook shopper ability. There square measure 500K applications open on Facebook and on the conventional, 20M applications square measure placed in daily. Additionally, a number of applications haven't any inheriting and sustain an enormous shopper base. as an example, Farmville and town Ville applications have 26 .5M and 42.8M shoppers up to currently.

As of late, programmers have begun exploiting the acknowledgment of this outsider application stage and causation harmful applications. harmful applications can supply a paid business for programmers, given the acknowledgment of OSNs, with Facebook driving the methodology with 900M dynamic shoppers. There square measure a number of other ways that programmers can get delight from a harmful application: 1) the applying can reach vast quantities of shoppers and their companions to change surface spam; 2) the applying can acquire clients' on the point of home info like email address, previous neighborhood, and sex; and 3) the applying can "imitate" by creating elective vindictive applications traditional. to create matters a lot of awful, the making ready of malignant applications is disentangled by ready to-utilize toolboxes beginning at $25. In elective words, there is principle and risk, and afterward, there square measure a number of malevolent applications spreading on Facebook day by day [9]. notwithstanding the over-worrying patterns, lately, a shopper has horrendously restricted info at the hour of putting in associate degree application on Facebook. In elective words, the problem is that the accompanying: Given associate degree application's temperament run (the specific image distributed to the applying by Facebook), can we have a tendency to tend to find if the applying is vindictive? By and by, there is not any business administration, in open accessible information,or inquire regarding primarily based instrument to coach a shopper regarding the hazards concerning associate degree application. As we have a tendency to tend to seem in Section III, harmful applications ar way reaching and people they essentially change surface, as A contaminated shopper endangers the protection of each one among its companions. Up till this time, the examination network has given next to no thought to OSN applications expressly. Most examination connected with spam and malware on Facebook has fixated on detection malevolent posts and social spam crusades [10].

At steady time, in a {very} very on the substance of it in reverse advance, Facebook has destroyed its application rating utility as these days.

associate degree current work examines anyway application authorizations and network appraisals correspond to the protection dangers of Facebook applications. At last, there square measure some network primarily based criticism driven endeavors to rank applications, for instance, WhatsApp? These could also be awful unbelievable within the long run, to the current purpose they ought to be gotten nearly no appropriation.

II. LITERATURE SURVEY

Tip spam in area place on social organizations. Recognizing tip spam regarding a thought Brazilian LBSN system, expressly Apontador. In lightweight of a sealed aggregation of tips given by Apontador as crawled data with relation to consumers and zones, we tend to acknowledged shifted attributes able to recognize spam from non-spam tips [1].

S. Ghosh et al depict the Understanding and battling affiliation cultivating within the Twitter social network. Search engines rank locales/pages fixated on chart estimations, for example, PageRank High in-degree gets high Pagerank. affiliation developing in Twitter Spammers seeks when entirely sudden consumers and check out to urge them to hunt when back [2].

Guanjun carver, Nan Sun, Hindu divinity state, Jun Zhang, Yang Xiang, and Houcine Hassan portray the "Measurable Twitter Spam Detection Demystified: Performance, Stability and Scalability" during this paper, they thought of the execution of an honest extent of standard AI estimations, hoping to differentiate those giving satisfactory acknowledgment execution and security enraptured with tons of ground truth info. With the target of achieving steady Twitter spam revealing capability, we tend to any evaluated the figurings as away in lightweight of the power [3].

G. Stringhini, C. Kruegel, and G. genus Vigna portray the police examination spammers on informal communities. facilitate to differentiate spam Profiles nonetheless once they do not contact a nectar profile. The eccentric lead of the buyer profile is perceived and loving therewith the profile is created to acknowledge the transmitter [4].

J. Tune, S. Lee, and J. Kim depict the Spam separating in Twitter abuse sender-beneficiary relationship. A spam separation procedure for social associations mistreatment affiliation info between purchasers. The framework utilizes detachment and accessibility thanks to the options that square measure problematic to manage by spammers and cheap to rearrange spammers [5].

K. Lee, J. Caverlee, and S. Webb depict the Uncovering social spammers: social honeypots and AI. System analyzes anyway spammers World Health Organization target social association goals work. To accumulate the knowledge regarding spamming development, a structure created an incredible course of action of "nectar profiles" on three respectable individual to individual correspondence regions [6].

Nathan Aston, Jacob Liddle, and Wei Hu* depict the Twitter Sentiment in info Streams with Perceptron. The execution feature decline we tend to would possibly create our Perceptron and Voted Perceptron estimations more and cheaper throughout a stream climate. during this paper, manufacture techniques by that twitter assessment are settled every rapidly and exactly on such a prime to bottom scale [7].

K. Thomas, C. Grier, D. Tune, and V. Paxson depict the Suspended records all things considered: Associate in a very Nursing assessment of Twitter spam The acts of spammers on Twitter by separating the tweets sent by suspended customers by and enormous. A rising spam-as-an advantage feature that accompanies sensible and not terribly reliable half programs, restricted time material primarily based shorteners, and Twitter air-con check merchants [8].

K.Thomas, C.Grier, J.Ma, V.Paxson, and D.Song portray the orchestrate Associate in Nursing assessment of a continuing location spam uninflected organization Monarch is AN ongoing system for division stunt, phishing, and malware URLs as they are submitted to web suppliers. A ruler's coming up with summarizes to a number of internet promotion ministrations being centered by address spam, precise arrange depends upon having purpose|some extent|a degree} by point understanding of the Spam campaigns misusing Associate in a nursing organization [9].

X. Jin, C. X. Lin, J. Luo, and J. Han dynasty portray the Social spam protect: an information min-ing primarily {based} usually spam discovery framework for internet based life systems. ordinarily procuring spam practices in a casual network by checking social sensors with clear shopper bases. Presenting every image and substance options and social association options to say spam activities. Integrating with our GAD gathering computation to impact within and out scale data. Presenting Associate in Nursing versatile powerful learning feeling to impact acknowledge existing spams with stressed human undertakings, ANd Perform on-line dynamic figuring out an approach to come to a decision spams incessantly [10].

III. MALICIOUS CONTENT ON FACEBOOK

The prominence and reach of Facebook have moreover force in loads of spam, phishing, malware, and elective types of malevolent movement. Assailants draw unfortunate casualties into sound on vindictive connections advise to outer sources and in skillful their system. These connections may be change surface either through near to home messages (visits), or through divider posts. to acknowledge most permeability, assailants prefer to post connects in public. Regularly, A bad person starts the assault by posting pictures heedfully snatching reviews, that temporary purchasers to like, offer, or treat them so as to require a goose at them.

The activities of feeling, remarking or sharing unfold these pictures into the contused individual's system.

once the cultural assimilation is change surface, the unfortunate casualty is pleased to a pernicious web site, which might further taint her computer, or companions organize through phishing, malware, or spyware.

This phishing page requests that the unfortunate casualty share this video with their companions to require a goose at it. In any case, once the unfortunate casualty shares this video, the page sidetracks to AN irregular advertisement page.

The video appreciates the review/fingernail that appeared within the post does not usually exist. totally different elective sources have referred to such samples of tricks and harmful posts on Facebook within the previous barely any years.

11, twelve still phishing tricks, elective harmful action on Facebook incorporates excluded mass notices, image labeling, post labeling, individual/talk messages then on. Instinctively, a shopper is further conceivable to answer to a message or post from a Facebook companion than from an outsider, therefore creating this social spam a more easy conveyance system than a recent email.

This enlarged weakness to such sanely spam has angry analysts to see, and battle social spam and choice pernicious action on Facebook. we {are going to} normally nowadays analysis the variable assault and discovery procedures that are used within the past to identify and unfold malevolent substance on Facebook severally.

IV. PROPOSED SYSTEM

Proposed system, we evaluate the performance of spam detection in our data set using machine learning algorithms, that is, the LSTM algorithm.

The process of detecting Malicious through the use of machine learning algorithms. Before classification, a classifier containing the knowledge structure must be trained with pre-labeled Posts. Once the classification model wins the knowledge structure of training data, it can be used to predict a new incoming user posts.

The whole process consists of two phases: 1) learning and 2) classification. First of all, the characteristics of the post will be extracted and formatted as a vector. Class tags (spam or non-spam) can be obtained through other approaches (such as manual inspection).

The characteristics and label of the class will be combined as an instance for training. A training post can therefore be represented by a pair that contains a feature vector, which represents a post, and the expected result and the training set is the vector.

The training set is the insertion of the machine learning algorithm; the classification model will be built after the training process. In the classification process, the timely captured user posts will be marked by the trained classification model.

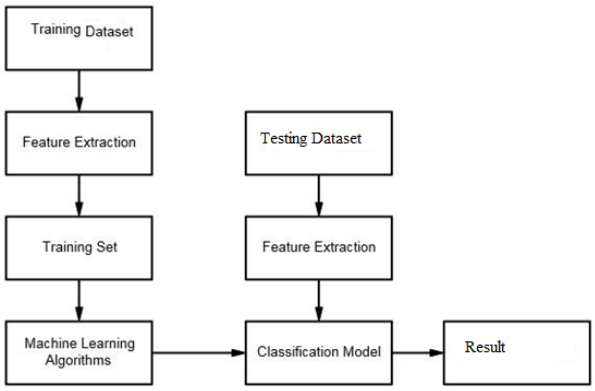


Fig. 1: System Architecture

Data collection

This module depicts the assortment of all Facebook application. The premise of our investigation begins with the assortment of data. it's 2 subcomponents they are: the assortment of Facebook applications with computer address and locomotion for URL redirections. At no matter purpose this half gets a Facebook application with a computer address, it achieves a locomotion string that pursues all redirections of the computer address and appears into the relating science addresses. The slipping string blends these recovered computer address and science chains to the tweet knowledge and pushes it into a line. As we've seen, our crawlers cannot hit malignant landing URLs once they utilize contingent redirections to sidestep crawlers. withal, on the grounds that our location framework does not rely on the highlights of landing URLs, it works solo of such crawler avoidances. 3.3 Feature extraction. we have a tendency to isolate highlights into 2 subsets: on-request highlights and assortment based mostly highlights. we have a tendency to understand that malignant applications ar completely not identical as generous applications. The on-request highlight incorporates 1)App rundown: the pernicious applications, as a rule, have deficient application summaries.2)Requested authorization set: on account of harmful applications, the larger a part of the malignant applications needs simply one consent set that's consent for posting on clients' dividers.

Divert URL

Pernicious applications divert the consumer to space with a poor ill fame.

Customer ID in-application institution computer address: primarily pernicious applications stunt shoppers into introducing completely different applications by utilizing associate alternate client ID within the application institution URL. Post in applications profile: there's no post in malignant applications divider. the full based mostly part incorporates the subsequent.1)App name: harmful applications have associate application name indistinguishable from at any rate one completely different malevolent applications.

Outer association post proportion: essentially this apportion is high for pernicious applications.

Connection coping with

The fundamental capability of this Link taking care of is to acknowledge the skin and within association accessible in your application (URL) and inform you thus on build restorative move. At no matter purpose this application distinguishes such an association issue it'll naturally divert to it space, it's potential that it would be associate inward association or outside association upon your last affirmation. Another vital purpose is that you just will consider the committal to writing phase through the skin association and its one among a sort phishing framework can distinguish the sites that trying|are trying|try} to theft your knowledge or attempting to form your trick.

Preparing

The preparation half incorporates 2 subcomponents: progressing to the record statuses and making ready for the classifier. Since we have a tendency to utilize a disconnected administered learning calculation, the part vectors for making ready are usually a lot of seasoned than embrace vectors for arrangement. to call the preparation vectors, we have a tendency to utilize the record status; URLs from suspended records are viewed as pernicious whereas URLs from dynamic records are viewed as favorable. we have a tendency to over and over update our classifier utilizing marked making ready vectors.

Order and identification

The order half begins our classifier utilizing input embrace vectors to characterize suspicious URLs. The grouping module acknowledges a computer address and also the connected social setting highlights separated within the past advance. These URLs, distinguished as suspicious, are going to be sent to security specialists or progressively advanced distinctive examination conditions for an internal and out examination.

Points of interest Of projected framework

1. Extraction of twelve highlights and classifications as Tag highlights and computer address-based highlights.
2. The framework actualizes a way which can utilize a spot channel element to differentiate whether or not the post is spam or not.
3. The framework actualizes application will likewise be expedited on-line for its utilization and also the info are going to be place away and got from the server.
4. User with most extreme range of spam will be hindered from the framework.
5. Performance assessment done on Dataset by utilizing TPR, FPR, Precision, Re-call and F-measure.

V. ALGORITHMS

LSTM

The LSTM left from traditional neuron-based neural system models and rather conferred the concept of a memory cell. The memory cell can hold its price for a short or very vital time frame as a piece of its knowledge sources, that permits the cell to recall what is necessary and not simply its last processed price. The LSTM memory cell contains three doors that administration anyway data streams into or out of the cell. door controls once new information can stream into the memory. The overlook door controls once partner degree existing little bit of data is unnoticed, permitting the cell to recollect new knowledge. At long last, the yield door controls once the knowledge that's contained within the cell is employed within the yield from the cell. The cell jointly contains masses, that administration every entree. The instructing rule, commonly BPTT, improves these masses bolstered the subsequent system yield blunder.

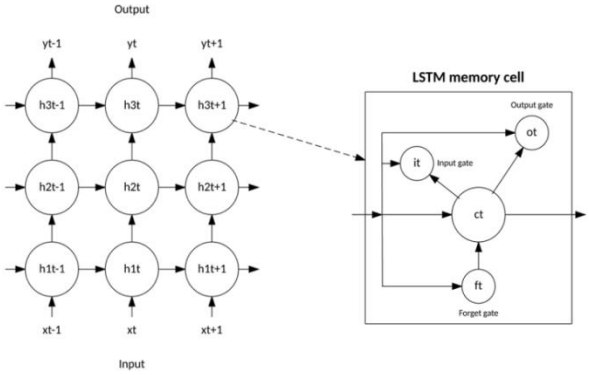


Fig. 3 LSTM Cell Memory

The GRU is a smaller amount advanced than the LSTM, will be ready all the faster, and may be progressively productive in its execution. Be that because it could, the LSTM will be additional} communicative and with more data, will lA improvement of the LSTM was familiar with alluded with because the gated rehashed unit. This model has two entryways, effort forestall the yield door blessing within the LSTM model. For a number of applications, the GRU has contend out appreciate the LSTM, anyway being easier implies that less masses and snappier execution. The GRU incorporates two entranceways: partner degree update door and a reset entryway. The update entranceway shows what extent of the past cell substance to stay up. The reset entranceway characterizes Associate in Nursing approach to fuse the new contribution with the past cell substance. A GRU can show a customary RNN simply by setting the reset entranceway to 1 and what is more the update door to zero.ead to higher outcomes.

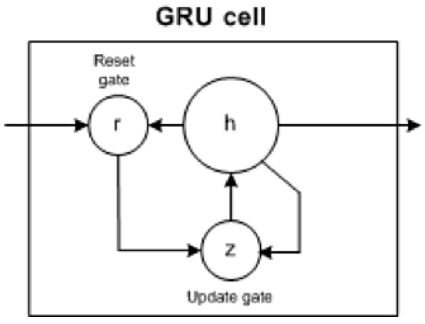


Fig. 2 GRU Cell

The GRU is a lot of easy than the LSTM, are often ready all the faster, and might be more and more effective in its execution. however, the LSTM is often communicatory and with more data, will prompt higher outcomes.

Naive Bayes

It is an associate degree order technique obsessed with mathematician Theorem with suspicion of independence among indicators. In basic terms, a Naïve mathematician categorified expect that the closeness of a particular component in an exceedingly class is inconsequential to the closeness of another element. In our paper Naïve mathematician is employed for checking application is malevolent or not. According to consent set, it contains the number of authorizations needed to urge to any application, however as we tend to understand that pernicious application would like fewer consents to urge to anybody's info once contrasted with generous applications that requirements to satisfy all of the standards to urge to terribly similar things. In our application, we tend to square measure utilizing credulous Bayes algorithm that fills the information in the element table like one or zero as indicated by whether or not the applying requested authorization or not. Bayes hypothesis provides a technique for reckoning back probability P(c—x) from P(c), P(x) and P(x—c). Take a goose at the condition beneath:

1. P(c—x) is that the back probability of sophistication (c, target) given indicator (x, properties).
2.P(c) is that the earlier probability of sophistication.
3.P(x—c) is that the likelihood is that the probability of indicator given category.

C4.5

With regards to AI, an alternative tree could be a tree-like diagram structure, wherever each hub speaks to a check on a property. every branch speaks to the results of the check and also the leaf hubs speak to the category mark non heritable finally decisions created through that branch. The ways that from root to leaf speak to classification rules. the target of this arrange at that time is to talk to the knowledge whereas limiting the unpredictability of the model. a number of calculations for building such increased trees are planned. C4.5, as an example, gets from the notable gap and-overcome procedure and has been generally used in an exceedingly few application fields, being one in every of the foremost acknowledge AI calculations. This arrange constructs alternative trees from plenty of getting ready data utilizing the concept of information entropy. At each hub of the tree, C4.5 picks the property of knowledge} that the majority adequately elements its set agreeing to the standardized data gain (distinction in entropy). it's imperative to require note of that this basic, nonetheless efficient methodology, is supplied for handling missing qualities within the informational indexes and each numerical and pure qualities.

VI. CLASSIFICATION - PREDICTING MALICIOUS

Grouping as a feature of this exploration was done utilizing three calculations - C4.5 Naive Bayes and LSTM. These three calculations were picked as they speak to 2 distinctive ways to cope with grouping. While C4.5 utilizes an alternative tree, Naive Bayes utilizes a probabilistic student. The wood hen [5] data dig programming was utilized for making ready and running the classifiers. The preparation dataset was worked from a known summation of pestilent and favorable sites utilizing the Malware Domain List and Google Safe Browsing API. These pernicious and sort sites were slithered, parsed and handled utilizing MalCrawler [2]

and tweaked python Code thus on build this preparation dataset. Right off the bat, as we tend to hope to rank the qualities obsessed with their capability to foresee pestilent sites, we tend to utilize the Gain magnitude relation strategy for attribute alternative. During this strategy, every property Ai is allowed score obsessed with the info gain among itself and also the category. within the event that C is that the category and associate is that the characteristic, conditions (1) and (2) beneath provide the Entropy H once observation the standard.

$$H(C) = -\sum_{c\epsilon C} p(c) log_2 p(c) \qquad (1)$$

$$H(C/A) = -\sum_{a\epsilon A} p(a) \sum_{c\epsilon C} p(c/a) log_2 p(c/a) \qquad (2)$$

Furthermore, 10 Fold Cross-approval was run, every property successively, to survey the capability of every assign to foresee malevolent sites. The Confusion Matrix created by this 10 Fold Cross-approval is given in Table I. The table clarifies the capability of a credit to foresee deadly sites.

VII. ANALYSIS OF RESULTS OBTAINED

This section describes several classification metrics used by the authors of the papers cited in the next sections. There are a number of different metrics for a model performing a binary classification task. These metrics include accuracy, precision, recall, false positive rate, F1 Score, and area under the curve (AUC) and many of the metrics have more than one name. All of these evaluation metrics are derived from the four values found in the confusion matrix, which is based on the calculated predicted class versus the ground truth.

Table I: Results Obtained During Classification

No	Naive Bayes Classifier				C 4.5 Classifiers				LSTM			
	TN	TN	FP	TP	TN	FN	FP	TP	TN	FN	FP	TP
A1	11%	12%	10%	67%	12%	14%	9%	65%	1%	12%	8%	70%
A2	10%	12%	13%	75%	7%	6%	7%	80%	6%	5%	5%	85%
A3	10%	9%	4%	77%	7%	8%	6%	79%	6%	7%	5%	82%
A4	3%	6%	4%	87%	6%	5%	4%	85%	5%	5%	3%	87%
A5	8%	7%	9%	76%	6%	8%	9%	77%	5%	6%	8%	80%
A6	3%	3%	5%	89%	3%	4%	5%	88%	2%	3%	4%	90%
A7	2%	4%	3%	91%	2%	3%	3%	92%	1%	2%	2%	94%
A8	3%	2%	3%	92%	3%	2%	2%	93%	2%	2%	1%	95%
A9	10%	12%	13%	65%	16%	14%	14%	56%	8%	11%	2%	60%
A10	11%	10%	12%	67%	15%	7%	13%	65%	2%	6%	2%	69%

Accuracy (acc) or Proportion Correct: the ratio of correctly classified examples to all items. The usefulness of accuracy is lower when the classes are unbalanced (i.e., there are a significantly larger number of examples from one class than from another). However, it does provide useful insight when the classes are balanced. acc = TP + TN/ TP + TN + FP + FN . (1)

Positive Predictive Value (PPV) or Precision (p): The ratio of items correctly classified as class X to all items that were classified as class X. p = TP / TP + FP. (2)

Sensitivity or True Positive Rate (TPR) or Probability of Detection (PD) or Recall (r): The ratio of items correctly classified as X to all items that were actually class X. TPR = TP / TP + FN . (3)

Negative Predictive Value (NPV): The ratio of items correctly classified as not X to all items classified as not X. NPV = TN /TN + FN . (4)

Specificity or True Negative Rate (TNR): The ratio of items correctly classified as not X to all items that are not class X. TNR = TN / TN + FP. (5)

False Alarm Rate (FAR) or False Positive Rate (FPR) or Fall-Out: The ratio of items incorrectly classified as class X to all the items that are not class X.

FPR = FP / TN + FP. (6)

F1 Score (F1): The F1 Score is the harmonic mean of the precision (p) and the true positive rate (r). F1 = 2 1 r + 1 p = 2 p ∗ r p + r . (7)

This is a specific version of the F-β function, in which precision and true positive rate are given equal importance.

The outcomes no heritable within the past phase are poor down in succeeding passages to find the foremost affordable arrangement of traits for distinctive Malicious Websites. A. Grouping Accuracy of Attribute The examination of the ten qualities thought-about for identification of vindictive web site utilizing AI is indicated diagrammatically in Figure two. The structured presentation shows the order accuracy for every single characteristic, running 10 Fold Cross approvals, utilizing the three grouping calculations C4.5, Naive mathematician and LSTM.

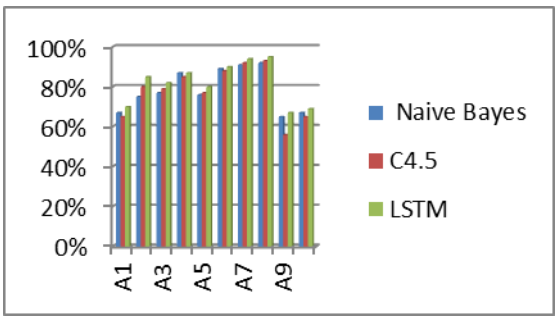
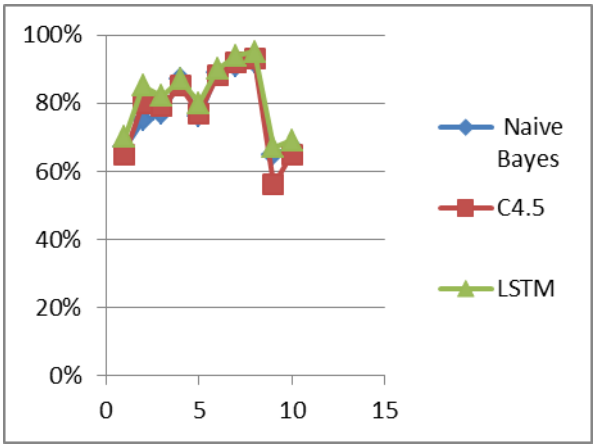


Fig. 2: Classification Accuracy of Attributes.

B. Computational Resources Used

The machine assets (memory and mainframe cycles) used for attribute extraction and pre-handling may be a vital issue to rank the standard. The machine assets used by traits were surveyed utilizing the python, once running the extraction and pre-handling python code. The qualities got were standardized to seem on a size of zero to one hundred. The result got is appeared as an overview in Figure 3.



VIII. CONCLUSION

Applications gift convenient means that for hackers to unfold malicious content on Facebook. However, very little is known about the characteristics of malicious apps and the way they operate. In this paper, employing a massive corpus of malicious Facebook apps determined over a 9-month amount, we tend to showed that malicious apps dissent considerably from benign apps with regard to many features. for instance, malicious apps are much more seemingly to share names with different apps, and that they usually request fewer permissions than benign apps. Investing our observations, associate in nursing correct LSTM classifier for sleuthing malicious Facebook applications. Most curiously, we tend to highlighted the emergence of app-nets—large teams of tightly connected applications that promote one another. We'll still dig deeper into this system of malicious apps on Facebook, and we hope that Facebook can take pleasure in our recommendations for reducing the menace of hackers on their platform.

ACKNOWLEDGMENT

I would like to express my sincere and deepest gratitude to my dissertation guide Dr. Sunil B. Mane for his continuous support and understanding throughout this journey. I would also thank him for providing me enough flexibility for this project at times when I was in real need. I would like to thank all the faculty members of Department of Computer Engineering And IT and the Helpdesk Team for providing me with the necessary help whenever required. I express my gratitude to my colleagues with whom I had valuable discussions on the project. Lastly, but importantly, to my husband, parents and in laws who have supported me all the time.

REFERENCES

1. Guanjun Lin, Nan Sun, Surya Nepal, Jun Zhang,Yang Xiang, and Houcine Hassan"Statistical Twitter Spam Detection Demystified:Performance, Sta-bility and Scalability" IEEE,2017.
2. H. Costa, F. Benevenuto, and L. H. C. Merschmann, "Detecting tip spam in location-based social networks," in Proc. 28th Annu. ACM Symp. Appl. Comput., 2013, pp. 724–729.
3. S. Ghosh et al., "Understanding and combating link farming in the Twitter social network," in Proc. 21st Int. Conf. World Wide Web, 2012, pp. 61–70.
4. G. Stringhini, C. Kruegel, and G. Vigna, "Detecting spammers on social networks," in Proc. 26th Annu. Comput. Sec. Appl. Conf., 2010, pp. 1–9.

5. J. Song, S. Lee, and J. Kim, "Spam filtering in Twitter using sender re-ceiver relationship," in Proc. 14th Int. Conf. Recent Adv. Intrusion Detection, 2011, pp. 301–317.
6. K. Lee, J. Caverlee, and S. Webb, "Uncovering social spammers: social honeypots + machine learning," in Proc. 33rd Int. ACM SIGIR Conf. Res.Develop. Inf. Retrieval, 2010, pp. 435–442.
7. Nathan Aston, Jacob Liddle and Wei Hu*, "Twitter Sentiment in Data Streams with Perceptron," in Journal of Computer and Communications, 2014, Vol-2 No-11.
8. K. Thomas, C. Grier, D. Song, and V. Paxson, "Suspended accounts in ret-rospect: An analysis of Twitter spam," in Proc. ACM SIGCOMM Conf. Internet Meas., 2011, pp. 243–258.
9. K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song, "Design and eval-uation of a real-time URL spam filtering service," in Proc. IEEE Symp. Sec. Privacy, 2011, pp. 447–462.
10. X. Jin, C. X. Lin, J. Luo, and J. Han, "Social spam guard: A data mining based spam detection system for social media networks," PVLDB, vol. 4, no. 12, pp. 1458–1461, 2011 .

AUTHORS PROFILE

Mrs. Ayesha Choudhar, is pursuing M.Tech in the Department of Computer Engineering at Government College of Engineering, Pune, and Maharashtra, India. She received her Bachelor of Engineering Degree in Information Technology from Solapur University, India. Her research interests are in the field of Cyber Security .

Dr. Sunil B. Mane, is currently working as an Associate Professor in the Department of Computer Engineering and Information Technology at Government College of Engineering Pune (An Autonomous Institute of Govt. of Maharashtra), India. He has more than 15 years of teaching experience. He has over 25 research publications in various National/International Journals and Conferences. He is a Board of Studies member in Computer Engineering/Information Technology of various autonomous engineering institutes. He has delivered lectures on information and cyber security domain as invited speaker. He is serving as Co- Chief Investigator for the Information Security Education and Awareness (ISEA) project, Ministry of Information Technology, Govt. of India. His areas of research are Data Privacy and Cyber Security.