IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

PREDICTION OF AN INFULENCER IN A SOCIAL MEDIA NETWORK USING MACHINE LEARNING

Ananya Bhavana D S[1] Dr N. Sivakumar[2]

[1]Research Scholar, Department of Computer Science, Jain (Deemed -to-be-University), Bangalore, Karnataka, India.

[2]Research Supervisor, Department of Computer Science, Jain (Deemed -to-be-University), Bangalore, Karnataka, India.

Given two individuals and their casual association incorporates, our obligation is to expect which one is more influential. In our endeavor, we accumulate planning tests from Kaggle reliant upon human judgment. We use a couple of unmistakable models to make estimates, as Logistic Regression, SVM, Naive Bayes and. We similarly use some partner procedures like cross endorsement, incorporate assurance and data preprocessing. In the results region, we consider the shows of different models and give examination and thoughts to future works. We do different learning models using python

Keywords: Machine learning, SVM, NR, RF algorithm, Social network Dataset Process, Prediction

I. INTRODUCTION:

The assessment of social-based recommender frameworks has pulled in fundamental idea since the progress of Web 2.0. While principal frameworks reliably excused social joint endeavors among clients, the later strategies attempt to consolidate social affiliation data to improve the possibility of the thoughts and make them more changed. As indicated by 42 Essential online media assessments for 20201: There are 3.484 billion incredible electronic media clients. Individuals go through 2 hours and 23 minutes bit by bit through online media investigating and enlightening. In addition, 98.55% of individuals use in any event four online media channels every day. Top online media networks are Facebook (95%), Twitter (84%), Instagram (74%), LinkedIn (62%), and YouTube (61%). Precisely when these individuals experience unmistakable maybe stunning proposals, they will in ordinary go to persuading individuals, regardless of whether they are a piece of the creation association (for example retailers or makers) or are respect added influencers, like industry subject matter experts or expert aides. In light of everything, influencers in an easygoing affiliation are people whose effects are expanded through the affiliation. This effect is generally not escalated and influencers reliably base on restricted subjects to progress. The associate manager figuring everything out the investigation of this creation and favoring it for transport was Xiao Liu. Having the decision to find or see the influencers on social affiliations could be of phenomenal worth since influencers can acknowledge a goliath part in the achievement of different social, political and viral showing tries correspondingly as the redirection occasions. The course in to the persuading clients' obvious confirmation issue is discovering approaches to manage check the spreading limit of influencers in online media by finding the focal qualities of their substance

II. LITERATURE SURVEY

Inclination Learning: A Tutorial Introduction, DS 2011, Espoo, Finland, Oct 2011 by J. Furnkranz and E. Huller Meier[1] clarifies Preference learning issues can be seen several issue assessments, including portrayal of inclinations, sort of propensity model: utility cutoff (ordinal, numeric), propensity affiliation (fragmentary requesting, arranging, …), savvy portrayal, depiction of people/clients and choices/things: identifier, highlight vector, facilitated article, kind of preparing information speedy or circumlocutory investigation, complete or inadequate relations, utilities. "A Practical Guide to Support Vector Classification" by C. W. Hsu, C. C. Chang and C. J. Lin upport vector machine (SVM) is a standard arrangement depiction technique with different application locales. SVM shows its outstanding show in high-dimensional information strategy. During the time spent depiction, SVM part limit setting during the SVM preparing system, nearby the section choice basically impacts the social affair accuracy. This paper proposes two novel dexterous improvement approaches, which at the same time picks the breaking point respects while finding a subset of highlights to broaden SVM depiction exactness. The appraisal bases on two developmental enrolling ways to deal with oversee overhaul the constraints of SVM: molecule swarm streamlining (PSO) and natural calculation (GA). In like manner, we converge over the two adroit progress frameworks with SVM to pick fitting subset highlights and SVM limits, which are named GA-FSSVM (Genetic Algorithm-Feature Selection Support Vector Machines) and PSO-FSSVM (Particle Swarm Optimization-Feature Selection Support Vector Machines) models. Exploratory outcomes show that the depiction exactness by our proposed strategies beats standard design search approach and different ways of thinking. LIBSVM: A library for help vector machines by LIBSVM is a library for Support Vector Machines (SVMs). We have been effectively building up this pack since the year 2000. The objective is to assist clients with enough applying SVM to their applications. LIBSVM has acquired wide all inclusive in AI and different zones. In this article, we present all execution subtleties of LIBSVM. Issues, for example, dealing with SVM improvement issues theoretical mix multiclass demand likelihood assessments and cutoff confirmation are examined in detail.[3] Feng-Jen Yang proposed The need of depiction is essentially referenced, considering everything. As a numerical strategy approach, the Naive Bayes classifier joins a development of probabilistic calculations to track down the best-fitted assembling for a given scrap of information inside an irksome district. In this paper, an execution of Naive Bayes classifier is portrayed. This classifier can be utilized as an overall contraption save and material to different spaces of approaches. To guarantee the rightness of all probabilistic assessments included, a model informational record is picked to test this classifier in An Implementation of Naive Bayes Classifier.

III. SYSTEM DESIGN

First we load the dataset, then it ill move to he different kind of machine learning algorithms. Predict the result with accuracy. Here used logistic regeression, Naïve bayes, random forest,

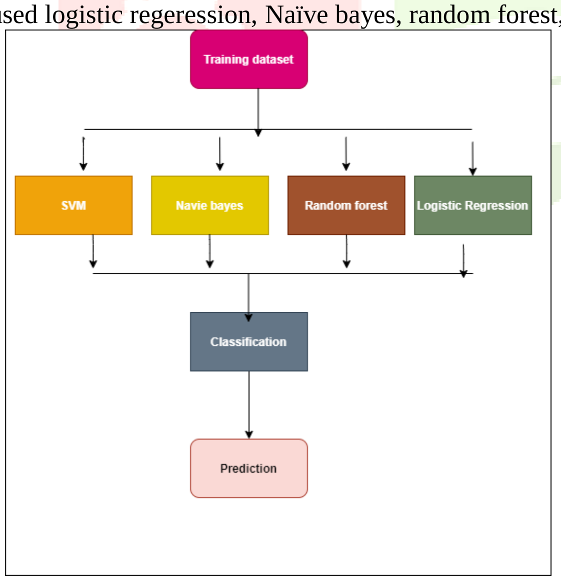


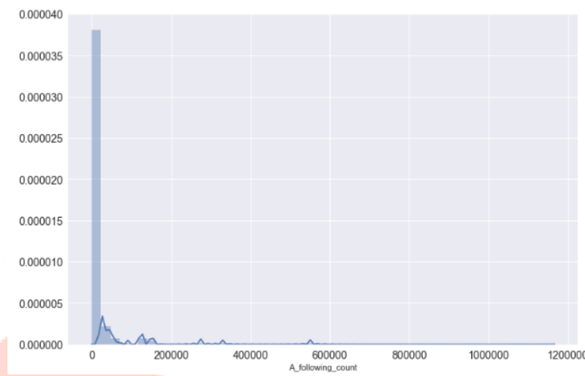
Fig. 1: Design of Project

IV. IMPLMENTATION

1) Dataset: We utilize the dataset from Kaggle given by PeerIndex, including a norm, pair-wise inclination learning task. Each data point depicts two people, An and B. For every individual, there are 11 pre-figured, non-unfriendly numeric highlights dependent upon Twitter advancement provided, which include:

1. # followers	2. # followings	3. #listed
4. # mentions_received	5. # retweets_received	6. # mentions_sent
7. # retweets_sent	8. # posts	9. #network_feature_1
10. # network_feature_2	11. # network_feature_3	

Fig. 2: Data Features



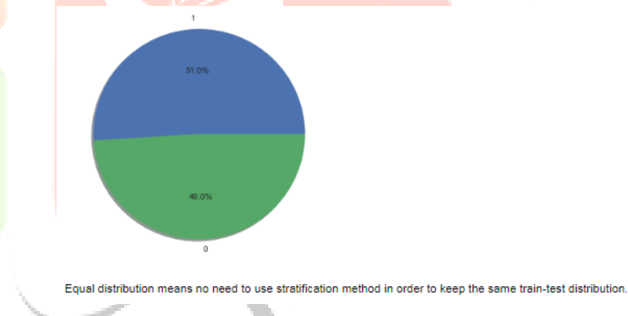


Fig. 3: Visualization

Since we have two people in every information test, we have 22 highlights far and away. Additionally, there is a twofold name watching out for a human judgment about which of the two people is even more amazing in each arranging test. Name 1 strategies An is more persuading than B. Name 0 procedures B is more tempting than A. There are 3000 preparing tests and 2500 testing tests in our dataset. Given a test, our commitment is to expect which individual in this test is seriously persuading.

2) Apply Algorithm: Here applying AI calculation like nb,svm,rf,lr.

3) Data cycle for Better outcome. The chance has shown up to improve it by utilizing highlight scaling or change.

4) Feature Engineering and Feature Selection: highlight arranging methods introducing the open data in really obliging manner so that model could profit by it. This lessened the measure of highlights for more applied way utilizing calculations, for example, PCA or hand making

extra bewildering highlights from suitably open ones utilizing space information.

There is no persuading inspiration to utilize PCA for few highlights. Notwithstanding, I have made 4 additional highlights. The objective of the model is to expect the human judgment about online media influence. One of the vital factor to impact this choice would be various enthusiasts. Since more partner proposes more individuals are investigating that individual's post which will have some impact (it might be positive or adversarial) on individuals' assessment. So joined factors 'A_is_popular' and 'B_is_popular' are made which has a worth of 1 if an individual has in excess of 1000000 partners or 0 notwithstanding.

Second part is made by deviding a partner tally from following check. More prominent the measure of this division, more noteworthy is that individual's after exclude remained from aficionado check. This surmises rather than that individual affecting others, others will an effect on him/her. Highlight affirmation gathers picking the best highlights which are as of now open to us. It very well may be finished by applying different key philosophies on highlights or by utilizing the 'L1' regularization in the model. Here I have utilized Select Best class and Recursive Feature Elimination by RFE class.

RESULT AND CONCLUSION

It was an intriguing issue. Information was in every way that really matters, amazing and information cleaning fundamental was insignificant. Information was amazingly correct slanted which inconceivably afftected the model execution of different non-tree based models. A fundamental log change of the information was figured out some approach to manage this issue and inside and out improved the demonstration of those models. After which I attempted to besides improve that show with include arranging and affirmation. This differently affected various models. The RF, NB, SVM, LR model outcome with definite outcome subject to given information.

From the above outcomes, we can reason that the best accuracy we can accomplish is about 76% utilizing direct models, which isn't unmistakably better compared to our benchmark 70.2%. This recommends that the testing models probably won't be straightly detachable. The test precision of Naive Bayes is near straight models. Notwithstanding, in the event that we can apply the facilitate rising based assessment, we may accomplish much better performance, which is a decent decision for future works.

REFERENCES

[1] J. Furnkranz and E. Hullermeier. Preference Learning: A Tutorial Introducton, DS 2011, Espoo, Finland, Oct 2011.

[2] Hsu C W, Chang C C, Lin C J. A practical guide to support vector classification[J]. 2003.

[3] Chang C C, Lin C J. LIBSVM: a library for support vector machines[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2011, 2(3): 27.

[4] Swingler K. Applying neural networks: a practical guide[M].Morgan Kaufmann, 1996.

[5] G " Python for Informatics: Exploring Information" – Book by Charles Severance

[6] "Practical Data Science Cookbook" – Book by Abhijit Dasgupta, Benjamin Bengfort, Sean Patrick Murphy, and Tony Ojeda.

[7] Stanford WebBase Project. http://www-diglib. stanford.edu/~testbed/doc2/WebBase.

[8] L. A. Adamic. The Small World Web. In Proceedings of the Third European Conference on Research and Advanced Technology for Digital Libraries (ECDL'99), Paris, France, Sep 1999.

[9] L. A. Adamic, O. Buyukkokten, and E. Adar. A social network caught in the Web. First Monday, 8(6), 2003.

[10] Y.-Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong. Analysis of Topological Characteristics of Huge Online Social Networking Services. In Proceedings of the 16th international conference on World Wide Web (WWW'07), Banff, Canada, May 2007.

[11] R. Albert, H. Jeong, and A.-L. B´arab´asi. The Diameter of the World Wide Web. Nature, 401:130, 1999.

[12] L. A. N. Amaral, A. Scala, M. Barth´el´emy, and H. E. Stanley. Classes of small-world networks. Proceedings of the National Academy of Sciences (PNAS), 97:11149–11152, 2000.

[13] A. Awan, R. A. Ferreira, S. Jagannathan, and A. Grama. Distributed uniform sampling in real-world networks. Technical Report CSD-TR-04-029, Purdue University, 2004.

[14] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan. Group Formation in Large Social Networks: Membership, Growth, and Evolution. In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'06), Philadelphia, PA, Aug 2006.

[15] A.-L. B´arab´asi and R. Albert. Emergence of Scaling in Random Networks. Science, 286:509– 512, 1999.

[16] L. Becchetti, C. Castillo, D. Donato, and A. Fazzone. A Comparison of Sampling Techniques for Web Graph Characterization. In Proceedings of the Workshop on Link Analysis (LinkKDD'06), Philadelphia, PA, Aug 2006.

[17] V. Braitenberg and A. Schuz. ¨ Anatomy of a Cortex: Statistics and Geometry. Springer-Verlag, Berlin, 1991.

[18] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph Structure in the Web: Experiments and Models. In Proceedings of the 9th International World Wide Web Conference (WWW'00), Amsterdam, May 2000.

[19] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data, Jun 2007. http://arxiv.org/abs/0706.1062v1.

[10] d. boyd. Friends, Friendsters, and Top 8: Writing community into being on social network sites. First Monday, 11(12), 2006.

[21] P. Erd¨os and A. R´enyi. On Random Graphs I. Publicationes Mathematicae Debrecen, 5:290–297, 1959.