Journal of Nonlinear Analysis and Optimization Vol. 15, Issue. 1, No.15 : 2024 ISSN : **1906-9685** Journal of Wonlinear Analysis and Optimization : Theory 4, Applications ISSN - 200600 Editors-in-Chief : Somoon Ulimporgan Somoon Ulimporgan Somoon Ulimporgan

DETECTION OF FAKE ONLINE REVIEW USING SUPERVISED AND SEMI-SUPERVISED Kornala. Venkatesh¹, Koppula. Jhansi

¹Assistant Professor, PG DEPT, Dantuluri Narayana Raju College, Bhimavaram, Andhara Pradesh. Email:- kornalavenkatesh@gmail.com
²PG Student of M.C.A, Dantuluri Narayana Raju College, Bhimavaram, Andhara Pradesh. Email:- jhansikoppula03@gmail.com

ABSTRACT

Online reviews have great impact on today's business and commerce. Decision making for purchase of online products mostly depends on reviews given by the users. Hence, opportunistic individuals or groups try to manipulate product reviews for their own interests. This paper introduces some semi-supervised and supervised text mining models to detect fake online reviews as well as compares the efficiency of both techniques on data-set containing hotel reviews.

1. INTRODUCTION

Technologies are changing rapidly. Old technologies are continuously being replaced by new and sophisticated ones. These new technologies are enabling people to have their work done efficiently. Such an evolution of technology is online marketplace. We can shop and make reservation using online websites. Almost, everyone of us checks out reviews before purchasing some products or services. Hence, online reviews have become a great source of reputation for the companies. Also, they have large impact on advertisement and promotion of products and services. With the spread of online marketplace, fake online reviews are becoming great matter of concern. People can make false reviews for promotion of their own products that harms the actual users. Also, competitive companies can try to damage each others reputation by providing fake negative reviews.

2. Literature Survey

An extensive survey of the most notable works to date on machine learning-based fake review detection. Firstly, they have reviewed the feature extraction approaches used by many researchers. Then, they detailed the existing data-sets with their construction methods.

Then, they outlined some traditional machine learning models and neural network models applied for fake review detection with summary tables. Traditional statistical machine learning enhances text classification model performance by improving the feature extraction and classifier design. In contrast, deep learning improves performance by enhancing the presentation learning method, algorithm's structure and additional knowledge. They also provided a comparative analysis of some neural network model-based deep learning and transformers that have not been used in fake review detection. The outcomes showed that Roberto achieved the highest accuracy on both datasets. Further, recall, precision, and F1 score proved the efficacy of using Roberto in detecting fake reviews. Finally, they summarized the current gaps in this research area and the possible future direction to get robust outcomes in this domain.

3. IMPLEMENTATION STUDY

EXISTING SYSTEM:

Researchers have been studying various approaches for detecting fake online reviews. These approaches can be categorized into:

Content-Based Methods: Focus on the text of the review.

User Behavior-Based Methods: Focus on the reviewer's behavior, such as country, IP address, and number of posts.

Disadvantages:

- > In the existing work, the system uses only to semi-supervised learning.
- > Only Text Classification as sentiment text and it never finds fake review.

Proposed System & Algorithm:

We implemented both semi-supervised and supervised classification methods:

Semi-Supervised Classification:

Used the Expectation-Maximization (EM) algorithm. The EM algorithm was first proposed by Karimpour et al.

Steps:

- > Derive an initial classifier from the labeled data-set.
- Use this classifier to label the unlabeled data-set, creating a predicted set of labels
 (PU).
- > Combine labeled and unlabeled data-sets, derive a new classifier, and classify the
- unlabeled data-set again.
- > Repeat until PU stabilizes.

3.1 Advantages:

- > The system is very fast and effective due to semi-supervised and supervised learning.
- > Focused on the content of the review based approaches. As feature we have used-word.

IMPLEMENTATION

Modules:

Admin:

In this module, the Admin has to login by using valid user name and password. After login successful he can do some operations such as View All Users and Authorise, Add Movie, View All Movies, View All Movie Reviews, View All Reviews by Rates, View All Movie Recommendations, View All Interesting and Uninteresting Movies, View All Search History, View All Purchased items and booked Movie ,View All Fake Reviews By Movie, View All Fake Reviews By Region, View All Fake Reviews by Rates ,View Movies Rank in chart.

User:

In this module, there are n numbers of users are present. User should register before doing some operations. After registration successful he has to login by using authorized user name and password. Login successful he will do some operations like View My Profile, My Accounts, View All Movies, View All My Purchased Items and Movies, View All Recommended Items and Movies for Me.



4. SYSTEM ARCHITECTURE:

Fig: 3.1 System Architecture

5. RESULTS AND DISCUSSION

Screen-shots:

Home Page:



Fig: 5.1



Admin Login Page:

Fig: 5.2



Admin Home Page:



👻 😹 All Users × + - 0 × ← → C O localhost:8080/Detection%20of%20fake%20online%20reviews%20using%20semi%20supervised%20and%20supervised%20learning/a_all_users.jsp ☆ 🥐 : View All Users and Authorise Log Out User Image User Name User Full Details Status Id omkar Authorized rakesh Authorized Authorized ganesh mahesh Authorized umesh Authorized Manjunath Authorized Q Search

All User Viewed By Admin:

Fig: 5.4

All Movies × +							-	0	×
← → C (O) localhost:8080/Detection%	620of%20fake%20onlin	ne%20reviews%20usin	g%20semi%20supervised	d%20and%20supervised	%20learning/a_all_movies.jsp		\$	0	:
		View A	II Movies						^
Id	Movie Name	Movie Image Pric	et Actor Desc Name (En	ription Upload crypt) Date	Rank Ratings	View Movie Reviews Religion			
1	The Villain	100	aXQga Yw5uY dm11	WRhIG1v 09/09/2019	19 *****	<u>Click</u> <u>here</u> Vijayanagar			
2	Robot2	200	aXQga Y3Rpb aw∪=	24gbW92 //	13 ****	Click here Vijayanagar			ļ
3	Rjakumara	100	Punith Pulith CBtb	XMgYSBm HkgZW50 WlubWVu 3ZpZQ== //		Click here Vijayanagar			l
4	Kaala	150	Rajinikanth FsYSB a2Fhb	GEgYW ~ rbm93 yBLYW LYXJp ~ GFuLG //	4 ★ 🖈	Click here Vijayanagar			l
5	Onti	250	Ramesh T250a BhbiB bWluZ SuYWR dGlvb	SBpcy + 1cGNv yBLYW hIGFj + iBmaW	9 ★★★	Click <u>here</u> Vijayanagar			l
6	Mission_Mangal	3000 HANGAL	SXQga aW9uI ZW50a YmFzZi bQ==	XMgYWN0 GFuZCBz W11bnQg WQgZm1s //	10 * * * *	Click here Vijayanagar			ľ
		A	Qm9sb	H13b29k			THE .	110 414	÷
📮 Q. Search 🛛 🛁	📮 🧿 💆					^ 🐁 🧔 🖽 🕀	ENG (合 中) ● 6	1:18 AM /28/2024	.

View All Movies:



User's Reviews for a Movie Viewed By Admin:





		View all revie	WS			
Id	User Name	Movie Name	Review Details	Date and Time	Region	
3	Omkar	The Villain	nice	09/09/2019 17:35:25	vijayanagar	
4	Omkar	The Villain	villain is bad	09/09/2019 13:03:00	vijayanagar	
6	Omkar	The Villain	villain is bad	09/09/2019 13:03:50	vijayanagar	
7	Manjunath	Onti	It is bad movie	09/09/2019 17:16:29	Malleshwaram	
8	Manjunath	Onti	It is good movie	09/09/2019 17:16:48	Malleshwaram	
9	Maniunath	Robot?	It is good movie	09/09/2019 16:12:50	Malleshwaram	
	Inanjunau	100002	It is totally	07/07/2017 10.12.50	Manconwaran	
10	Manjunath	Robot2	animation based	09/09/2019 16:36:14	Malleshwaram	
11	Ramesh	Mission Mangal	It is hest movie	09/09/2019 18:17:16	Malleshwaram	
12	Ramesh	Mission Mangal	It is bad movie	09/09/2019 18:17:27	Malleshwaram	
13	Manjunath	Mission_Mangal	It is good movie	09/09/2019 18:18:46	Malleshwaram	
14	Manjunath	Mission Mangal	It is beautiful film	09/09/2019 18:19:53	Malleshwaram	
15	Manjunath	Mission_Mangal	It is worst movie	09/09/2019 18:37:21	Malleshwaram	
16	Manjunath	Kaala	It is best movie	09/09/2019 18:45:45	Malleshwaram	
17	Manjunath	Kaala	It is worst movie	09/09/2019 18:46:13	Malleshwaram	
18	omkar	Rjakumara	It is good	09/09/2019 18:47:16	vijaynagar	
19	откаг	Kjakumara	super wonderfull	09/09/2019 18:47:44	vijaynagar	
20	Jhansi	Kalki AD 2898	pan india hit Jai prabhasssss	15/06/2024 12:33:52	Other Andhra Pradesh	
				<u>Back</u>		

View All Movie Reviews:



Purchased Products and Movie: × +							-	0	×
← → C () localhost:8080/Detection%20	of%20fake%20online%20)reviews%20using	%20semi%20sup	ervised%20and%20s	upervised%20le	earning/a_all_purchased.jsp	\$	P	:
	Viev	v All Bo	oked N	Admin Menu Admin Man Log Out			^		
Id	Movie Name	Booked User	No. of Ticket(s)	Booked Date	Booking Price				
1	The Villain	<u>Omkar</u>	2	09/09/2019 17:43:21	200				
2	The Villain	<u>Rakesh</u>	1	09/09/2019 17:12:14	100				
3	The Villain	<u>Ganesh</u>	1	09/09/2019 17:12:48	100				
4	The Villain	<u>Mahesh</u>	1	09/09/2019 17:13:13	100				1
5	Robot2	<u>Mahesh</u>	1	09/09/2019 17:13:48	200				
6	Onti	<u>Manjunath</u>	2	09/09/2019 17:10:50	500				
7	Onti	<u>Manjunath</u>	1	09/09/2019 17:13:48	250				
8	Onti	<u>Manjunath</u>	1	09/09/2019 17:14:09	250				
9	Onti	<u>Manjunath</u>	1	09/09/2019 17:14:29	250				
10	Mission_Mangal	<u>Ramesh</u>	2	09/09/2019 18:16:36	6000				
					Back	<u>k</u>			
									Ŧ
🚆 Q Search 🛛 🚚 📕	9 🔮					- 🤹 🧒 📼 🗗 🕅 🕅	11 6/2	:19 AM :8/2024	٠
				Fig: 5	5.8				

View All Movie Booked:

✓ a_all_rates_fake × +		-	0	×
← → C (① localhost:8080/Detection9	%20of%20fake%20online%20reviews%20using%20semi%20supervised%20and%20supervised%20learning/a_all_rates_fake.jsp	☆	C) :
	REVIEWS •••• View All Fake Rate Details !!			
1	Id User Name Rate Movie Name Date and Time Region			
	1 Manjunath 4 Robot2 09/09/2019 17:14:54 Malleshwaram 2 Manjunath 3 Riakumara 09/09/2019 17:26:35 Malleshwaram			
	3 Manjunath 1 Kaala 09/09/2019 17:31:43 Malleshwaram 6 Ramesh 3 Mission Mangal 09/09/2019 18:18:04 Malleshwaram			
	7 omkar 1 Mission Mangal 09/09/2019 18:48:22 vijaynagar			
🚦 Q Search 🔫	► ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ●	■ ¹ 6/	1:20 At 28/202	м 24

View All Fake Review Details:

Fig: 5.9

🗸 😹 Movies Rank × + - 0 ← → C O localhost:8080/Detection%20of%20fake%20online%20reviews%20using%20se 1%20and%20supervised%20learning/a_chart2.jsp * 0 : . View Movies Rank in Chart Admin Main JS charts Log Out JS Chart Back Q Search 💵 📄 🥥 💕 へ 🧠 😴 🖾 🕀 ENG 🚯 🗇 🚍 11:20 A IN 6(28/202 ۵

Rank Chart For Reviews:

Fig: 5.10

V 🗷 User Register × +		-	0	×				
← → C () localhost.8080/Detection%20of%20fake%20online%20reviews%20using%20semi%20supervised%20and%20supervise	ction%20of%20fake%20online%20reviews%20using%20semi%20supervised%20and%20supervised%20learning/u_register.jsp							
REGISTER	Home Admin User			•				
User Name (required) Password (required)								
Email id (required) Mobile Number (required) Your Address								
Date of Birth (required) Select Gender (required)								
Enter Location (required)								
📲 Q Search 🔫 🂵 📮 🧿 ở	^ 🧌 😌 🗃	ENG 🚯 🔿 🗩	11:26 AM 6/28/2024					

User Registration :



User Login Page:



Fig: 5.12



Recommendations for User:



👻 😹 Review Movie × + ٥ of%20fake%20 rvised%20learning/u_mreview.jsp?pn=The%20Villain מוום שמףכו אושכם וכמווווהם Jupervised HOME PAGE **Review Moview** User Menu Movie Name The Villain Write Review 🚦 Q Search 🛛 🛁 📑 🥥 へ 🔩 🈎 📼 🕞 🛤 🚯 🚭 40) 📼 11:22 AM

User Review Process:



User Book Movie Tickets:



6. CONCLUSION AND FUTURE WORK

We have shown several semi-supervised and supervised text mining techniques for detecting fake online reviews in this research. We have combined features from several research works to create a better feature set. Also we have tried some other classifer that were not used on the previous work. Thus, we have been able to increase the accuracy of previous semisupervised techniques done by Jiten et al. [8]. We have also found out that supervised Naive Bayes classifier gives the highest accuracy. This ensures that our dataset is labeled well as we know semi-supervised model works well when reliable labeling is not available.

In our research work we have worked on just user reviews. In future, user behaviors can be combined with texts to construct a better model for classification. Advanced preprocessing tools for tokenization can be used to make the dataset more precise. Evaluation of the effectiveness of the proposed methodology can be done for a larger data set. This research work is being done only for English reviews. It can be done for Bangla and several other languages.

7. REFRENCES

- Chengai Sun, Qiaolin Du and Gang Tian, "Exploiting Product Related Review Features for Fake Review Detection," Mathematical Problems in Engineering, 2016.
- A. Heydari, M. A. Tavakoli, N. Salim, and Z. Heydari, "Detection of review spam: a survey", Expert Systems with Applications, vol. 42, no. 7, pp. 3634–3642, 2015.
- M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination," in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT), vol. 1, pp. 309–319, Association for Computational Linguistics, Portland, Ore, USA, June 2011.
- J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic Inquiry and Word Count: Liwc," vol. 71, 2001.
- S. Feng, R. Banerjee, and Y. Choi, "Syntactic stylometry for deception detection," in Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers, Vol. 2, 2012.
- J. Li, M. Ott, C. Cardie, and E. Hovy, "Towards a general rule for identifying deceptive opinion spam," in Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL), 2014.
- E. P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, "Detecting product review spammers using rating behaviors," in Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM), 2010.
- J. K. Rout, A. Dalmia, and K.-K. R. Choo, "Revisiting semi-supervised learning for online deceptive review detection," IEEE Access, Vol. 5, pp. 1319–1327, 2017.
- J. Karimpour, A. A. Noroozi, and S. Alizadeh, "Web spam detection by learning from small labeled samples," International Journal of Computer Applications, vol. 50, no. 21, pp. 1–5, July 2012.