Quantile Normalized Neighbor Combinatorial Machine Learning Based Recommendation in Digital Marketing

K.S. Narayanan[a*] and Dr.S. Suganya[b]

[*a]Associate Professor, Rathnavel Subramaniam College of Arts and Science, Coimbatore.
E-mail:narayanan@rvsgroup.com
[b]Associate professor, Rathnavel Subramaniam College of Arts and Science, Coimbatore.
E-mail: suganya_cs@rvsgroup.com

Abstract: The scientific advancement of the contemporary years has set industries on the move. The advancement in marketing has led to the point where reshaping to digital movements is essential. Even though it appears to be a plunge for marketers, as a matter of fact, all mechanized applications and systems that are designed on the basis of artificial intelligence only reduces the complication of conventional targeting and customization procedure. In several applications, the platforms utilized for online promotion carry algorithms for recognizing the best combinations whereas in other cases, the business establishments or institutions involving digital marketing take advantage to design and execute in-house personalized arrangements. As a case study, a method called, Quantile Normalized Neighbor Combinatorial Learning-based Recommendation (QNNCL-R) is applied for generating new leads that will ultimately become customers (i.e., promoting student higher education to admission branding in our scenario) via twitter dataset. The data obtained from the twitter dataset (i.e., higher education) is fed to the recommendationsystem. Then, the relevant set of features and event labels (i.e., tweets) is selected by Quantile Normalized Chi-square Feature selectionand a Neighbor Combinatorial Learning-based Recommendation algorithm with the best performance is selected for the recommendation process for higher education. QNNCL-R method is compared with other algorithms and indicating that QNNCL-R method performs better than other methods.

Keywords: Quantile Normalized, Chi-Square, Neighbor Combinatorial, Machine Learning-based Recommendation

1. Introduction

The internet is a commanding mechanism and it can be utilized to fascinate customers, strengthen credibility and broaden a product or service's brand. Social Media (SM) recommends a plan of action where users or people communicate and cooperate in a virtual manner. Users' opinions are supervised and determined by repeated advertisements that they come across on numerous micro blogging and platforms involving social media. Moreover, business analysts utilize SM for business scrutiny, corporate awareness meeting and product cognizance. The advancements in the business world have handled the purposes of social media to become as one of the essential instruments for marketing plan of action specifically in brand health and brand development.

To obtain uninterrupted incomes and increased dynamic customers or users, crucial business players should realize the actions and purchase inclinations of buyers. To forecast the buying options of purchasers, data pertaining to purchase objectives and inclinations have to be acquired. In (Arasu et al., 2020), Machine Learning integrated Social Media Marketing (ML-SMM) was proposed that involved two steps, Text mining, Machine learning integrated with social media marketing and finally, analysis of ML-SMM analysis via WEKA tool.

To start with the concepts involving social media marketing, application of machine learning was utilized and then combined with the WEKA machine learning tool with the objective of predicting online consumer behavior, therefore ensuring efficient marketing. With this finer reporting communicating potentialities were ensured laying foundations for precision and recall. Despite improvement observed in terms of both precision and recall, the processing time involved in predicting consumer behavior for lead generation was not focused. To address this issue in this work, Quantile Normalized Chi-square Feature selection algorithm is designed usinga preprocessing library that first tokenizes the tweets, followed by which computationally efficient relevant tweets are selected using Quantile Normalized function, therefore contributing to minimum processing time involved in overall lead generation process.

A machine learning method utilizing social media content marketing towards brand health of a company was proposed in (Pappu, 2019) catering to the scope or paint perceptions in the social media convention. Moreover, the important of paints in the social media convention involving social media posts of the company were also discussed and also measure of relevancy followed by discussions made in social media pages were also included.

Finally, the influence of periodic posting of content was identified and means were made to discuss in detail using machine learning techniques, therefore contributing to prediction and forecasting accuracy. Though improvement were observed in terms of prediction and forecasting accuracy, the false positive rate involved in generating lead for social media content marketing was not addressed. To provide solution to this issue, Neighbor Combinatorial Machine Learning Lead Generation algorithm is designed that reduces the false positive rate using combinatorial function.

1.1 Article Contributions

Aiming at improving the traditional recommendation model to obtain good values of recommendation accuracyand at the same time minimizing the overall processing time and false positive rate, QNNCL-R method is applied to distance learning twitter dataset. We propose a Quantile Normalized Chi-square Feature selection model that utilizes quantile normalization function to remove irrelevant tweets, then the Neighbor Combinatorial Learning-based Recommendation to provide accurate recommendations from neighbor users. Experimental results reveal that ourproposed QNNCL-R method approach provides better recommendation results with minimum processing time and false positive rate than the conventional recommendation methods.

1.2 Article Organization

The rest of article is organized as follows. Relevant literature is reviewed in Section 2. QNNCL-R method is derived in Section 3. Experimental settings are provided in Section 4. The detailed discussion with comparison of state-of-the-art methods is provided in Section 5. Section 6 concludes this work.

2. Related Works

A complete literature review of considerable empirical contributions made so far in this research area was handled in (Saura, 2020). However, with the existence of notable challenges from pessimistic electronic word-of-mouth and irritating online brand presence being an issue, an aggregate discernment from various leading experts on issues pertaining to digital and social media marketing was investigated in (Dwivedi et al., 2020).

Digital business platforms (DBPs) like, eBay, Google, and Uber Technologies have perceived immeasurable heightening, the role of marketing in helping DBPs succeed was proposed in (Rangaswamy et al., 2020). Moreover, social media marketing usage in the aspect of small and medium enterprise was designed in (Dahnil et al., 2014).

In (Shen et al., 2020), the value of demand learning via Social Media Exposure (SME) for luxury brand using two-period model was proposed, therefore contributing to cost minimization and accuracy maximization. Customer's attitude towards various brands and their intention towards purchase were analyzed in detail in (Abzari et al., 2014). Yet another detailed focus on the future concerning digital marketing was proposed in (Appel et al., 2020).

An elaborate review of literature on the prominence and impact of a user on social media between September 2010 and September2019 was proposed in (Al-Yazidi et al., 2020). In (Hayat et al., 2019), a keen discussion on Deep Learning architectures via a taxonomy-oriented summary was proposed toward the Social Media Analytics (SMA).A review of literature on the application of twitter across educational domain was investigated in (Malik et al., 2019). A sentiment analysis framework using different deep learning models via activation function with the objective of concentrating on the accuracy aspect was discussed in (Chen et al., 2019). Yet another method using variance-based structural equation model concentrating on social media in higher education institutes was proposed in (Ansari & Khan, 2020).

A systematic literature study on deep learning algorithms was proposed in (Yang et al., 2020). Application of different machine learning algorithms for analyzing the role of sentiment analysis was presented in (Sharma & Jain, 2020). Machine learning and Artificial Intelligence in the area of marketing was analyzed in detail in (Ma & Sun, 2020); (Shah et al., 2020). In (Miklosik et al., 2019) the selection and application of machine learning tools for analyzing the impact made in digital marketing was investigated. Motivated by above methods, QNNCL-R method is proposed.

3. Quantile Normalized Neighbor Combinatorial Learning-based Recommendation (QNNCL-R) Method

The issue of predicting a lead generation method for student population and thereforerecommending a specific higher educational institute is formulated as a Multi Criteria Recommendation problem. At first, the whole data set is divided into eight sub-datasets. Each sub-dataset represents the data of a specific higher educational institute and

consists of different features, some of these features are trivial and other is informative features. The computationally efficient dimensionality reduced tweets are obtained by Quantile Normalized Chi-square Feature selection model. We have proposed an adaptive recommendation system for predicting a specific higher educational institute based on the lead generation via relevant tweets selected. The computationally efficient tweets data selected are fed as input to the proposed method and the recommendationsystem will suggest an institute to the prospective student based on the lead obtained.
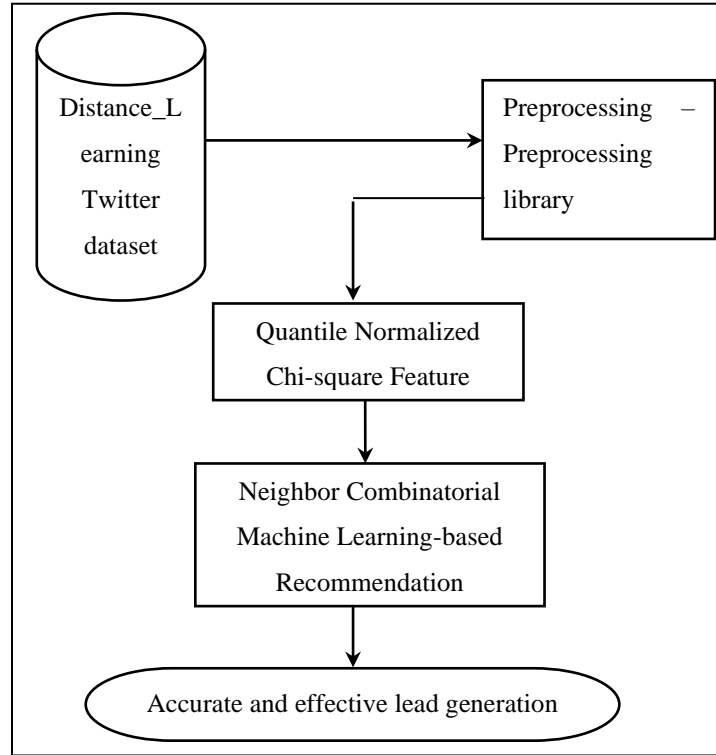

Figure 1. Block Diagram of QNNCL-R Method

Figure 1 illustrates the block diagram of QNNCL-R method. QNNCL-R method is split into three sections. With the distance learning twitter dataset provided as input, first, preprocessing is performed using preprocessing library to significantly tokenize the tweets. Second, feature selection for relevant tweets is obtained by applying the Quantile Normalized Chi-square Feature selection to the preprocessed features and tweets. Significant lead generation mechanism is proposed by Neighbor Combinatorial Machine Learning-based Recommendation.

3.1 Preprocessing

To start with Distance Learning dataset as input (i.e., tweets_raw), with different features '$F = F_1, F_2 \dots, F_n$' involved, the event labels or tweets '$T = T_1, T_2, \dots, T_n$' retrieved from twitter involves combination of URL, non-sentimental data like "#" (hashtags), "@" (posts), and "T" (tweets). Here, the text information tokenized so that 'n' tweets (event labels) are retrieved. With the preprocessing library for tweet data '$T = T_1, T_2, \dots, T_n$' written in Python, the preprocessing is carried out. With this preprocessing library, cleaning or tokenizing tweets are made. The preprocessing library '$PT = PT_1, PT_2, \dots, PT_n$' aids in cleaning or tokenizing URLs, Hashtags, Reserved words, Emojis and smileys.

3.2 Quantile Normalized Chi-Square Feature selection

Feature selection refers to the process of identifying a subset of functional event labels and discarding irrelevant event labels (i.e. irrelevant tweets). Feature selection process estimates accuracy of lead generation for education services and assists eliminating the futile correlation in the data (i.e., tweets) that might reduce the accuracy (i.e., lead generation accuracy). With the apt selection of event labels, insignificant variables are removed, therefore enhancing the accuracy and classification performance involved in lead generation for educational services.

Quantile Normalized Chi-square Feature selection model is applied to the preprocessed tweets, therefore minimizing the training time, finally avoiding the dimensionality curse in lead generation. The Quantile Normalized

Chi-square Feature selection model in our work based on filtering model is applied that employs statistical function to allocate a righteousness scoring value to each feature or data (i.e., tweets). The tweets are processed according to their righteousness score, and then, either selected to be eliminated from the data or retained.
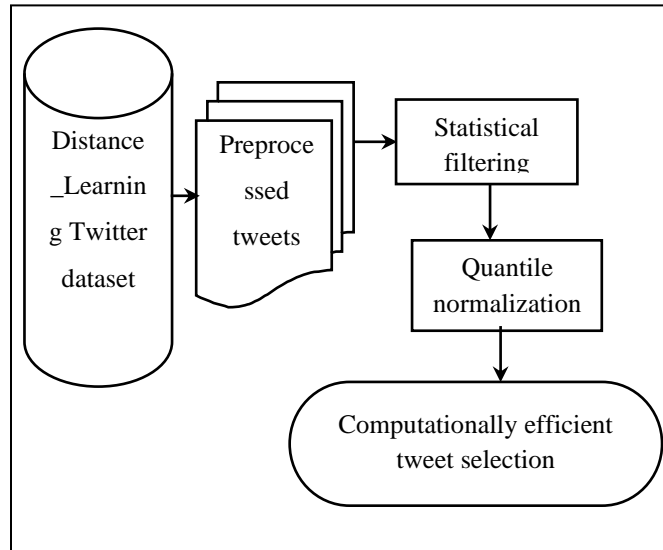

Figure 2. Block Diagram of Quantile Normalized Chi-Square Feature Selection Model

Figure 2 illustrates the block diagram of Quantile Normalized Chi-square Feature selection model. To start with a statistical filtering model based on thechi-squared test is applied to test the independence of two events (i.e. preprocessed tweets), where twoevents 'P' and 'Q' are defined to be independent, if '$Prob\,(PQ) = Prob(P)Prob\,(Q)$' or, '$Prob(P|Q) = Prob(P)\ and\ Prob(Q|P) = Prob\,(Q)$', mathematical formulation is given below.

$$\chi^2_{DF} = \sum \frac{(OPT_i - EPT_i)^2}{EPT_i} \qquad (1)$$

$$DF = (R - 1) * (C - 1) \qquad (2)$$

From (1), subscript 'DF' in chi-squared test 'χ^2_{DF}' symbolizes degree of freedom, 'OPT_i' and 'EPT_i' refers to observed preprocessed tweets and expected preprocessed tweets respectively. Degree of freedom is obtained based on number of magnitudes for one feature (i.e, tweet) 'R' and number of magnitudes for other feature 'C'. After obtaining the chi-square statistic, the 'p' is identified and resultant 'p' value, null hypothesis is either accepted or rejected.

To find reliable measure other than using p-value from chi-squaredtest, we proposed a new measure by adding a Qunatileterm on the 'pv' of feature (i.e., tweets) referred to as quantile normalized p-value 'p_{qnorm}'. Given a set of arrays in a matrix '$X \in PT$', each column of 'X' is sorted to give 'X_{sort}'. Mean across rows of 'X_{sort}' is evaluated and this mean value is assigned to each 'α' element in the row to obtain a matrix 'X'_{sort}'. Finally, the normalized values are obtained by rearranging the order of tweets in each column if 'X'_{sort}' to have 'α' the similar ordering as the original given matrix 'X'. Novel proposed statistical measure,'p_{qnorm}' is formulated as below.

$$p_{qnorm} = \frac{\chi^2_1 - X'_{sort}(pv, DF) - \chi^2_1 - X'_{sort}(\alpha, DF)}{\chi^2_1} \qquad (3)$$

From (3), 'α' refers to the significance measure and 'DF' relating to degrees of freedom.By utilizing this quantity, we are normalizing on the feature (i.e., tweets) with higher cardinality. Simply to say, we are trying to seehow further by percentage the critical value corresponding to 'pv', the '$X'_{sort}(pv, DF)$', is crossingthe critical value '$X'_{sort}(\alpha, DF)$' with respect to a given significance level 'α'.

3.3 Neighbor Combinatorial Machine Learning-based Recommendation Model

With selected computationally efficient tweets, the next step in our work forms a Neighbor Combinatorial Machine Learning-based Recommendation system to provide personalized information by learning user preferences. In this model we insert a Neighbor Combinatorial function into the traditional Machine Learning-based Recommendation system.The model is called asNeighbor Combinatorial as it utilizes both the single common tweets

and set of tweets based on the neighboring users. Neighbor Combinatorial function in Machine Learning-based Recommendation system is utilized to extract those users who have provided similar scores to similar items (i.e., tweets). The objective here remains in extracting those users provided with similar scores to similar tweets by means of Mahalanobis distance. Figure 3 shows the block diagram of Neighbor Combinatorial Machine Learning-based Recommendation model.
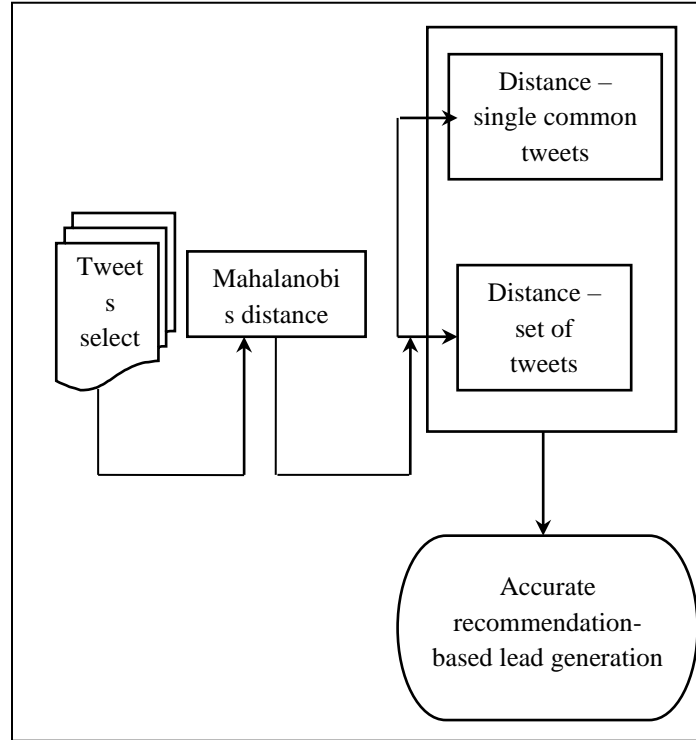

Figure 3. Block Diagram of Neighbor Combinatorial Machine Learning-based Recommendation Model

Assume that there are two users 'a' and 'b', then distance between two users with similar scores for similar tweets is formulated as below.

$$[(a - a')(b - b')]p_{qnorm} = \left[\frac{\sigma_b^2(a-a')-(b-b')\rho_{ab}\sigma_a\sigma_b}{Det(p_{qnorm})} \quad \frac{\sigma_a^2(b-b')-(a-a')\rho_{ab}\sigma_a\sigma_b}{Det(p_{qnorm})} \right] \tag{4}$$

From (4), the distance between two users is estimated based on the variances 'σ_b^2' and 'σ_a^2', covariance between two users '$\rho_{ab}\sigma_a\sigma_b$' and the determinant of quantile normalized value '$Det(p_{qnorm})$' respectively. With this, the distance between two users with similar scores for similar tweets using Mahalanobis distance 'MD' is expressed as given below.

$$MD = \sqrt{\left(\frac{a-a'}{\sigma_a}\right)^2 \left[\left\{ \left[\frac{b-b'}{\sigma_b}\right] - \rho_{ab}\left[\frac{a-a'}{\sigma_a}\right] \right\} \frac{1}{\sqrt{1-\rho_{ab}^2}} \right]^2} \tag{5}$$

From (5), the distance between single common tweets of two users are obtained and in a similar manner, the distance between two users 'a' and 'b' sharing a set of tweets or the polarity rate 'Pol' is mathematically expressed as below.

$$Pol = MD_{ab} = \sqrt{ \frac{\left[\frac{S_{aT}-a'}{\sigma_a}\right]^2 + \left[\left[\left\{ \left(\frac{S_{bT}-b'}{\sigma_b}\right) - \rho_{ab}\left[\frac{S_{at}-a'}{\sigma_a}\right] \right\} \frac{1}{\sqrt{1-\rho_{ab}^2}} \right] \right]^2 }{CR_{ab}} } \tag{6}$$

From the above equation (6), 'S_{aT}' forms the scoring of tweet 'T' given by user 'a' and 'a'' is the mean of the scores given by the user 'a' to all tweets and 'CR_{ab}' corresponds to the set of co-rated tweets of both the users 'a' and 'b' respectively. Finally with the similarity distance estimation, with the objective of obtaining the lead

generation of a tweet for the target user, collective score i.e., the predicted score 'PS_{aT}' of tweet 'T' of a user 'a' or subjectivity value 'Sub' is estimated below.

$$Sub = PS_{aT} = R'_a + \sum_{a' \in NU} Dis(a, a') * (R'_{a'T} - R'_{a'}) \qquad (7)$$

From (7), the predicted score for lead generation based on which the recommendation for certain tweets made is estimated according to the neighboring users 'NU' who have scored tweet 't'. By applying this recommendation with good values of accuracy for lead generation on admission branding is said to be ensured.

4. Experimental Settings

In this section, QNNCL-R method is used to promptly select higher education therefore laying for accurate admission branding based on the sentiment analysis o the tweets about distance learning. First, tweets collected from distance learning dataset are tokenized, followed by which relevant tweets for lead generation are selected in a computational and efficient manner. Then, the historical data of scores for each user's tweets areentered into the recommendation system as input. Finally, the subjectivity of each tweetis classified as a positive, negative or neutral for all tweets. Experimental evaluations are performed in Python by utilizing distance learning (https://github.com/Bhasfe/distance_learning) dataset that consists of three csv files, i.e., raw files, processed files and sentiment files. Table 1, 2 and 3 provides features in raw dataset, processed and sentiment dataset.

Table 1. Raw Data

S. No	Features
1	Numbers
2	Unnamed
3	Content
4	Location
5	User name
6	Retweet count
7	Favorites
8	Created at

Table 2. Processed Data

S. No	Features
1	Numbers
2	Unnamed
3	Content
4	Location
5	User name
6	Retweet count
7	Favorites
8	Created at
9	Processed
10	Length
11	Words
12	Country

Table 3. Sentiment Data

S. No	Features
1	Numbers
2	Unnamed
3	Content
4	Location
5	User name
6	Retweet count
7	Favorites
8	Created at
9	Polarity
10	Subjectivity

11	Label

Comparative analysis of lead generation methods is performed and compared with QNNCL-R, (Arasu et al., 2020) and (Pappu 2019). Analysis is made with different metrics with respect to tweet size.

5. Discussion

5.1 Case 1: Processing Time

A significant amount of time is said to be involved time in predicting student's tweets for arriving at lead generation. This is because of the reason that tweets generated are in fraction of seconds and analyzing them is a time consuming process.

$$PT = \sum_{i=1}^{n} UC_i * Time \ [T_W] \qquad (8)$$

From (8), the processing time 'PT' is measured based on user count 'UC_i' and the time consumed in obtaining the number of words in tweets '$Time \ [T_W]$'. It is measured in milliseconds (ms).
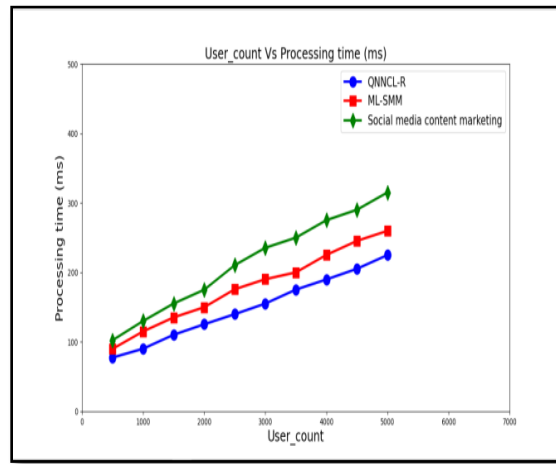

Figure 4. Graphical Representation of Processing Time

Figure 4 shows the processing time involved in obtaining the words in a given tweets. From the figure it is inferred that increasing the user count causes an increase in the number of words in the tweets and subsequently causes an increase in the processing time also. However, with '500' numbers of user counts involved in generating the lead and the time consumed in obtaining the words for single user count being '$0.155ms$' using QNNCL-R, '$0.180ms$' using (Arasu et al., 2020)and '$0.205ms$' using (Pappu 2019), the overall processing time was observed to be '$77.5ms$', '$90ms$' and '$102.5ms$' respectively. From this result processing time using QNNCL-R method is lesser than (Arasu et al., 2020)and (Pappu 2019). The lesser processing time is owing to the application of Quantile Normalized Chi-square Feature selection model. By applying this model, not only the tweets are preprocessed or tokenized followed by which computationally efficient tweets are selected via quantile normalization function where normalization is performed with high cardinality, therefore minimizing the processing time using QNNCL-R by 17% compared to (Arasu et al., 2020)and 30% compared to (Pappu 2019).

5.2 Case 2: Lead Generation Accuracy

Accurate data are required for better lead generation. The dataset used for experimentation probably have large number of user tweets and those tweets may not be accurate. With improper tweets, result in minimizing the lead generation accuracy. Therefore accurate tweets are required to improve lead generation accuracy. This is measured as below.

$$LGen_{acc} = \sum \frac{TRec_{acc}}{T_{size}} * 100 \qquad (9)$$

From (9), lead generation accuracy '$LGen_{acc}$' is measured based on tweet size 'T_{size}' and tweets recommended accurately '$TRec_{acc}$'. It is measured in percentage (%).
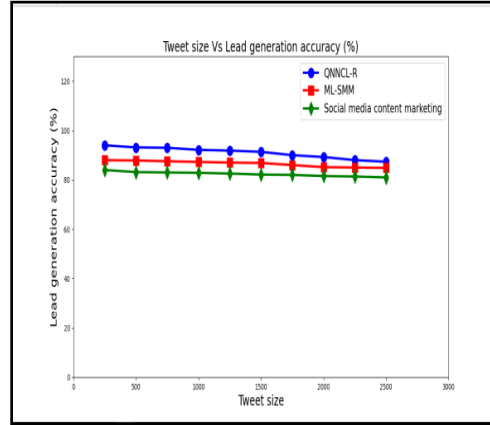
Figure 5. Graphical Representation of Lead Generation Accuracy

Figure 5 shows the lead generation accuracy involved in educational services domain. From the figure the lead generation accuracy is inversely proportion to the tweet size. In other words, increasing the size of tweets causes an increase in the amount of time consumed in processing the tweets for recommendation that in turn results in the minimization of the lead generation accuracy. However, with '250' numbers of tweets taken for simulation and '235' number of tweets correctly recommended using QNNCL-R, '220' number of tweets correctly recommended using (Arasu et al., 2020)and '210' number of tweets correctly recommended using (Pappu 2019), the overall lead generation accuracy was '94%', '88%' and '84%'. From the results, lead generation accuracy is better using QNNCL-R. The reason behind the improvement in lead generation accuracy is due to application of Neighbor Combinatorial Machine Learning Lead Generation algorithm. By applying this algorithm a recommendation model integrating Neighbor Combinatorial function and Machine Learning recommendation is designed. With this combination, first multiple co-rated tweets are estimated and then based on the results, Machine Learning recommendation is applied with aid of distance measure obtains similarity tweets. By combining these two functions, accurate and precise lead generation is obtained using QNNCL-R, therefore improving lead generation accuracy by 5% and 10% compared to (Arasu et al., 2020); (Pappu 2019).

5.3 Case 3: False Positive Rate

False positive rate refers to the ratio of all negative results that still give positive test end results, in other words, the conditional probability of a positive test result given an event that was not present. To be specific, false positive rate refers to the ratio of all negative recommendations made by the students regarding the education institutes and is measured as given below.

$$FPR = \sum_{i=1}^{n} \frac{UC_{WRec}}{UC_i} * 100 \qquad (10)$$

From (10), the false positive rate 'FPR' is measured based on total user counts 'UC_i' and the frequency of user count wrongly recommended 'UC_{WRec}'. It is measured in percentage. The resulting false positive rate is shown in table 4.

Table 4. False Positive Rate Performance Levels using QNNCL-R, ML-SMM (Arasu et al., 2020) and Social Media Content Marketing (Pappu 2019)

User_count	False positive rate (%)		
	QNNCL-R	ML-SMM	Social media content marketing
500	5	8	11
1000	6.35	8	11.35
1500	6.8	8.15	11.55
2000	7.05	8.35	11.85
2500	7.15	8.55	12
3000	7.2	8.85	12.35
3500	7.35	9	12.55
4000	7.55	9.15	13
4500	7.85	9.4	13.15
5000	8	9.35	14

The increasing the user count causes increase in FPR. However, with '500' number of user counts considered for simulation and '25' number of user counts wrongly recommended using QNNCL-R, '40' number of user counts were wrongly recommended using (Arasu et al., 2020) and '55' number of user counts wrongly recommended using (Pappu 2019), FPR were observed to be '5%', '8%' and '11%'. From this analysis, FPR is lesser using QNNCL-R. The reason behind the improvement was applying Quantile Normalized Chi-square Feature selectionalgorithm a statisticaltest measure between every feature variable and label variable was evaluated and existence of relationship were analyzed and eliminating independent feature variable.FPR of QNNCL-R was reduced by 19% and 43% compared to (Arasu et al., 2020); (Pappu 2019).

6. Conclusion

Lead generation via sentiments expressed in twitter messages is important though demanding activities. Most of the current lead generation method via sentiment analysis only identifies the twitter textual detailsand cannot accomplish sufficient performance due to the distinguishing characteristics of information from twitter. Inspired by recent work on machine learning to attain better performance of twitter sentiment analysis for education domain, a method called, QNNCL-R is proposed. First, computationally efficient features are selected by tokenizing the tweets using library function and applying Quantile Normalized function for obtaining relevant features. With relevant features, Neighbor Combinatorial Learning-based Recommendation algorithm is applied by combining the Neighbor Combinatorial and Learning-based recommendation improves lead generation accuracy with minimum processing time and FPR.

References

Arasu, B. S., Seelan, B. J. B., & Thamaraiselvan, N. (2020). A machine learning-based approach to enhancing social media marketing. *Computers & Electrical Engineering*, *86*, 106723.

Pappu A.R. (2019).The Effectiveness of Social Media Content Marketing Towards Brand Health of A Company: Social Media Analytics. *International Journal of Scientific & Technology Research,* 8, (11), pp. 1188-1192.

Saura, J. R. (2020). Using Data Sciences in Digital Marketing: Framework, methods, and performance metrics. *Journal of Innovation & Knowledge*, pp. 92-102.

Dwivedi, Y. K., Ismagilova, E., Hughes, D. L., Carlson, J., Filieri, R., Jacobson, J., & Wang, Y. (2020). Setting the future of digital and social media marketing research: Perspectives and research propositions. *International Journal of Information Management*, pp. 1-37.

Rangaswamy, A., Moch, N., Felten, C., van Bruggen, G., Wieringa, J. E., & Wirtz, J. (2020). The role of marketing in digital business platforms. *Journal of Interactive Marketing*, *51*, 72-90.

Dahnil, M. I., Marzuki, K. M., Langgat, J., & Fabeil, N. F. (2014). Factors influencing SMEs adoption of social media marketing. *Procedia-social and behavioral sciences*, *148*, 119-126.

Shen, B., Xu, X., & Yuan, Q. (2020). Demand learning through social media exposure in the luxury fashion industry: See now buy now versus see now buy later. *IEEE Transactions on Engineering Management*,pp.1-17.

Abzari, M., Ghassemi, R. A., & Vosta, L. N. (2014). Analysing the effect of social media on brand attitude and purchase intention: The case of Iran Khodro Company. *Procedia-Social and Behavioral Sciences*, *143*, 822-826.

Appel, G., Grewal, L., Hadi, R., & Stephen, A. T. (2020). The future of social media in marketing. *Journal of the Academy of Marketing Science*, *48*(1), 79-95.

Al-Yazidi, S., Berri, J., Al-Qurishi, M., & Al-Alrubaian, M. (2020). Measuring Reputation and Influence in Online Social Networks: A Systematic Literature Review. *IEEE Access*, *8*, 105824-105851.

Hayat, M. K., Daud, A., Alshdadi, A. A., Banjar, A., Abbasi, R. A., Bao, Y., & Dawood, H. (2019). Towards deep learning prospects: Insights for social media analytics. *IEEE Access*, *7*, 36958-36979.

Malik, A., Heyman-Schrum, C., & Johri, A. (2019). Use of Twitter across educational settings: a review of the literature. *International Journal of Educational Technology in Higher Education*, *16*(1), 1-22.

Chen, L. C., Lee, C. M., & Chen, M. Y. (2019). Exploration of social media for sentiment analysis using deep learning. *Soft Computing*, 1-11.

Ansari, J. A. N., & Khan, N. A. (2020). Exploring the role of social media in collaborative learning the new domain of learning. *Smart Learning Environments*, *7*(1), 1-16.

Yang, M., Nazir, S., Xu, Q., & Ali, S. (2020). Deep Learning Algorithms and Multicriteria Decision-Making Used in Big Data: A Systematic Literature Review. *Complexity*, *pp. 1-18.*

Sharma, S., & Jain, A. (2020). Role of sentiment analysis in social media security and analytics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(5).

Ma, L., & Sun, B. (2020). Machine learning and AI in marketing–Connecting computing power to human insights. *International Journal of Research in Marketing*, *37*(3), 481-504.

Shah, B., Dalwadi, G., Pandey, A., Shah, H., & Kothari, N. (2020). Online CQI-based optimization using k-means and machine learning approach under sparse system knowledge. *International Journal of Communication Systems*, *33*(3),

Miklosik, A., Kuchta, M., Evans, N., & Zak, S. (2019). Towards the adoption of machine learning-based analytical tools in digital marketing. *IEEE Access*, *7*, 85705-85718.

https://github.com/Bhasfe/distance_learning