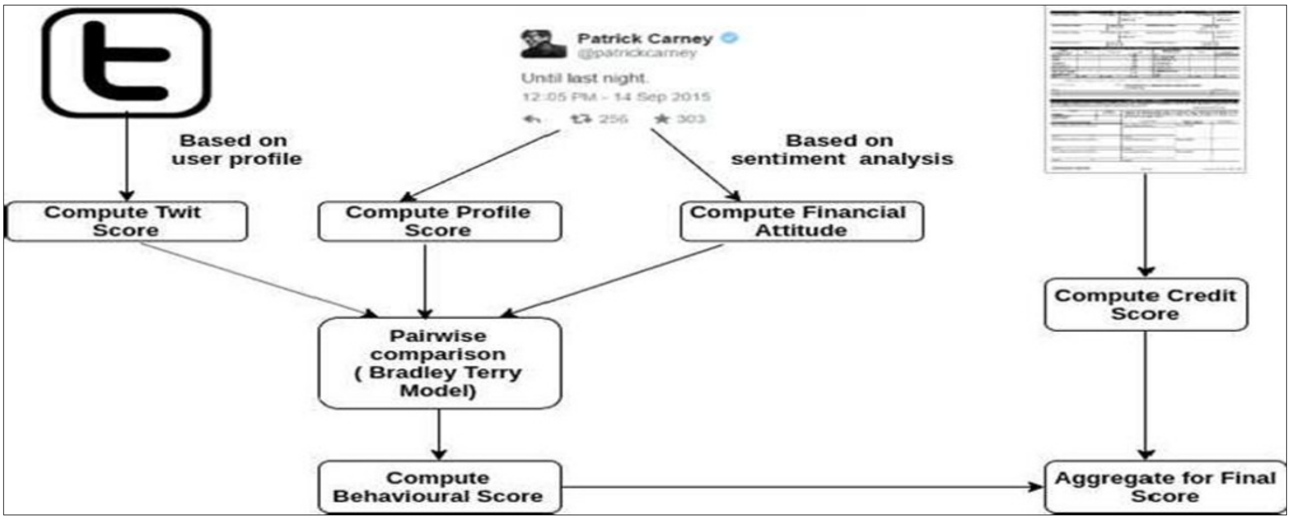


Fig. 1. System Architecture

3.2. Computing the Profile, Financial, Behavioural Score

The Twit Profile, Financial attitude, and Profile Score is computed as stated in Table 1.

Table 1. Score Computation

Score	Definition
Twit Score	The user profile of the person forms the basis of computing the twit score. It is an attempt to rate the quality of Twitter user by various metrics available through the API. A Twitter user with low twit score is more likely to be a sign of a spam account or a less safe user.
Profile Score	The tweets of the user form the basis of this score. They are pre processed by tokenization, lemmatization, stop-word removal etc. Subsequently, sentiment analysis is performed and the percentage of positive tweets is chosen to be the profile score.
Financial Score	The financial tweets of the user form the basis of this score. The financial tweets are identified by checking against a corpus of such terms. A threshold is set on the minimum number of financial tweets that need to be present for the analysis. To classify the tweets as positive or negative, a multi-level voting classifier is implemented and the number of positive tweets is chosen as the financial attitude.

3.3. Bradly Terry Model

Pair wise comparison is pivotal to reveal the importance of the three criteria. Pair wise comparison is utilized to rank the three parameters. The three scores have a priority weight attached to them and are combined based on it. The scores need to be prioritized to reflect their relative importance. Bradley Terry is probabilistic mathematical model that has been chosen to derive the priority weights and the number of times a factor is preferred to others becomes the weight assigned.

3.4. Credit Score

The credit score is based on several factors like payment history, credit exposure, and age of the credit. These parameters, in turn, depend on multiple factors, and weights are assigned for all based on the traditional loan application.

4. Methodology

In this section, the implementation of the scoring system is elucidated upon. The following describes the various methods and modules carried out for computing the Twit, Profile, and Financial Attitude score and executing the Bradley Terry Model. The modules have been carried out in the Python framework along with the necessary external libraries.

4.1. Modelling the Twit Score

The Twit Score is an attempt to rate the quality of Twitter users by various metrics such as friend/follower ratio, profile completeness, and other factors that are all available through the Twitter API. The Twit Score is developed giving equal weightage to elements like FF ratio, relevance, usage, and authenticity. A person with an FF ratio around 4 throws light on the fact that the user is more likely to be followed by all the person he or she is following. This indicates that the user is probably more relevant among his or her friend's circle and has acquired this ratio as a result of people liking what he or she talks about. This factor is further supported by the value of relevance score. A person with FF ratio greater than 10 indicates the person does not have enough amount of followers compared to the number of people he is friends with. This is a reflection of a less socially relevant user and complemented by the usage score. The twit score computation factors are further described in Table 2. The Twit score is initially calculated for a total of 20 and it is later evaluated on a scale of 100.

Table 2. Twit Score Computation

Sl.No.	Score	Description	Credit	Weightage
I	FF Ratio / Friend Follow Ratio	A person's follower count compared to thenumber of people they follow is a good measure of an interesting or integral Twitter user.	5	25% of Twit
II	Relevance Score	This is to verify how influential the useris among his or her followers i.e. the relevance of the user in his circle. This throws light on the impact factor of an user.	5	25% of Twit
II. a.	Listed Ratio	Listed Count / Followers Count. This is calculated as the ratio of followers who have listed them to their total followers. This throws light on how many followers find them impactful.	5	40% of Relevance
II. b.	Re-tweet	Average number of re-tweets shows the impact of an user.	5	60% of Relevance
III	Usage Score	Too much time on social media is not a great quality. In this computation, highly active users have a lower usage score and less active users have higher score.	5 (5+3+1+1)	25% of Twit
III.a.	Tweets Frequency	Scored based on average time diff in terms of day, hour and minutes basis.	5	50% of Usage
III.b.	Media	Amount of status posted.	3	30% of Usage
III.c.	Twitter Bio	Having a bio is seen as a positive factor.	1	10% of Usage
III.d .	Profile Picture	Having a Profile picture is also seen as a positive factor i.e. confidence.	1	10% of Usage
IV	Authenticity Score	This score is used to track the legitimacyof the user and how sound the profile is for consideration.	5	25% of Twit
IV. a.	Duration	Duration is the best way to verify the legitimacy of a person's twitter profile.	3	60% of Twit
IV.b.	FollowersCount	The no. of people they follow can play quite a large part in calculating Twitter handle interest scores. A person who follows a large number of people may not be a perfectly ideal person.	2	40% of Twit
	Twit Score		**20**	100%

4.2. Modelling the Profile and Financial Score

Firstly, the raw tweets are given as an input for applying standard NLP pre-processing technique to extract contextual features. This is fed to multi-level voting ensemble to train the model. The financial tweets are segregated from the list of tweets based on the predefined list of handles and hashtags used.

4.2.1. Feature Extraction

Data Pre-processing. Twitter data may be incomplete, inconsistent and noisy which can produce misleading mining results. Data preprocessing is a proven method of resolving such issues. It transforms raw data into a cleaner and understandable format. It also filters out useless data. Various steps involved in it are data cleaning, integration, feature reduction, and transformation. The list of pre-processing rules applied is shown in Table 3.

Table 3. Score Computation

Rule	Rule Description
Tokenization	Split a Tweet text into smaller words.
RT removal	Remove RT (retweet).
URL removal	Remove URL links.
Lemmatization	Convert words to base form.
Hashtag removal	Removal of '#' words
Convert to lowercase	Transform tweet into small letters.
Stripping whitespaces	Remove additional whitespaces.
Stopword removal	Remove stopwords like a, an provided by the standard NLTK package
Smiley removal	Remove emoticons from tweets.
Replace user mentions	Replace '@' mentions in a tweet
Punctuation removal	Remove punctuation marks and non-english character.

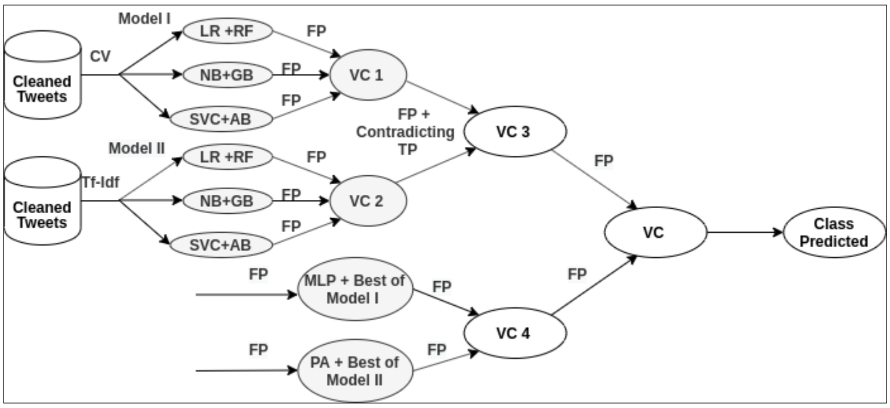


Fig. 2. System Architecture

Count Vector (CV). The vector is a measure of how frequently the word occurs in a document. Count Vector converts a collection of text documents to a matrix of token counts. Consider a Corpus C of D documents {d1,d2…..d_D} (rows) and N unique tokens (columns) extracted out of the corpus C. The size of the Count Vector matrix M will be given by D X N. Example of Count Vector Tokenization is given by

D1: He is a lazy boy. She is also lazy
D2: Neeraj is lazy.

Table 4. Calculation of CV

	He	She	lazy	boy	Neeraj
D1	1	1	2	1	0
D2	0	0	1	0	1

Tf-Idf. This represents Term frequency document frequency. This weight is a statistical measure used to evaluate how important a word is to a document in a collection or corpus. In 'i' is a weight in the document 'j' is calculated as follows

$$TF(i,j) = \frac{Term_i \; frequency \; in \; document_j}{Total \; words \; in \; document_j} \quad (1)$$

$$IDF(i) = \frac{Total \; documents}{Documents \; with \; term_i} \quad (2)$$

4.2.2. Proposed Multi-level Voting Model

Classification Algorithms. The processed dataset retrieved (after preprocessing and feature extraction phase) is fed to the classification phase for the identification of sentiment of the tweets as positive or negative. The machine learning algorithms used are Logistic Regression (LR), Naive Bayes (NB), Support Vector Machine (SVC) along Multilayer Perceptron (MLP). It is built involving a variety of bagging and algorithms such as Random Forest (RF), Gradient Boosting (GB), Adaptive Boosting (AB), and Passive-Aggressive Classifier (PA).

Multi-level Voting Model. In this module, an ensemble classifier is developed and it is implemented as a multi-level voting model that involves various classifiers. The financial corpus built is limited to 925 tweets. Multi-level voting classifiers harnesses the diversity of the individual base models and provide better performance compared to standalone classifiers because they solve the problem posed by a limited volume of data. They are repeatedly combined on the basis of minimum FP ratio and hence overcome the weakness attached with the existing base learning models as illustrated in the results section. Majority voting is implemented by initializing a higher weight to the better performing classifier. Fig.2. illustrates the architecture of the model.

Model I. The inputs are transformed into a CV matrix. The linear machine learning algorithm is combined with bagging or boosting model to form the three classifiers Logistic Regression with Random Forest, Naive Bayes with Gradient Boosting, and Support Vector Machine with Adaptive Boosting. By fusing a bagging or boosting algorithm with a machine learning model, it results in the creation of a stronger classifier.

Model II. The inputs are transformed into a Tf-Idf matrix. The linear machine learning algorithm is combined with bagging or boosting model to form the three classifiers Logistic Regression with Random Forest, Naive Bayes with Gradient Boosting, and Support Vector Machine with Adaptive Boosting. By fusing a bagging or boosting algorithm with a machine learning model, it results in the creation of a stronger classifier.

Voting- Classifier 1. Based on the performance of the set of classifiers in Model I, they are combined with hard voting based on the FP rate. The lower the FP rate, the better the classifier performs, or a greater weight is attached to that classifier.

Voting -Classifier 2. Based on the performance of the set of classifiers in Model II, they are combined with hard voting based on the FP rate. The lower the FP rate, the better the classifier performs or a greater weight is attached to that classifier.

Voting- Classifier 3. This combines the results of both the feature extraction phases. The third level of the voting classifier is retrieved by merging Voting Classifier 1 and 2 based on their FP rate. This classifier combines the FP values of I and II along with the contradicting TP results of the same for testing. This overcomes the weakness attached with the individual learning models of the feature extraction phases and presents a better result.

Voting- Classifier 4. This combines MLP with the best performing model of VC1 and PA with the best performing model of VC 2 based on their FP rate. The best performing model is chosen as the one with the highest accuracy. The Voting Classifier 4 is constructed by retrieving these classifiers based on their minimum FP rate. This strategic combination of these classifiers can reduce the total error rate.

Final Voting- Classifier. The final voting classifier is derived by merging VC 3 and VC 4. Subsequently, the final predicted class is marked on the basis of voting. Hard voting ensures the majority class label predicted at each level is allocated to the tweet. Thus, the classifier is optimized to achieve the highest accuracy of prediction.

4.3. Pairwise Comparison

The Bradley–Terry model [18] is a probability model that can predict the outcome of a paired comparison. Given a pair of individuals i and j drawn from some population, estimates the probability that the pair wise comparison P(i>j) is turns out true if i > j is defined by

$$P(i > j) = \frac{P_i}{P_i + P_j} \tag{3}$$

If the competitions are assumed to be mutually independent as in our model, then the probability is found to satisfy the logit model. The three scores Twit, Profile, and Financial Attitude score are compared as such and the probabilities allocated. The competitions are calculated by using p_{ij} is derived by

$$\frac{p_{ij}}{1 - p_{ij}} = \oint \mathrm{i} - \oint j \tag{4}$$

4.4. Computing the Behavioral Score

The behavioral score of the user is to be computed as follows.

$$Score = (PS \times W1) + (FS \times W2) + (TS\, W3) \tag{5}$$

Where PS, FS, TS stand for Profile Score, Financial Score and Twit Score respectively and weights of W1, W2, W3 represents

$$W1 = \frac{\left(P\ (PS + TS) + P(PS + FS)\right)}{2} \tag{6}$$

$$W2 = \frac{\left(P\ (FS + TS) + P(PS + FS)\right)}{2} \tag{7}$$

$$W3 = \frac{\left(P\ (PS + TS) + P(TS + FS)\right)}{2} \tag{8}$$

All the formulas implemented are original except the default working scores of the mathematical model.

5. Experimental Results

The various experimental results that were recorded during the development of this research work have been presented in the following section. The classification algorithms discussed have been applied to the two different datasets (training and testing) and their results have also been given.

5.1. Data Collection

The tweets and user details required for analysis are not readily available. The required tweet dataset has to be constructed by collecting the tweets, as mentioned, from Twitter API (using Tweepy). Dataset was collected from February and March in 2020. All the pre-processing rules described in Table 3 have been applied and the cleaned tweet is displayed against the original tweet in Table 5.

Table 5. Tweets and their pre-processed text

Tweet	Preprocessed Text
#CanaraBank Shorted 110 PE today for Zero Target, keep tracking. 26th March, 2020 it can be zero. CMP 1.3/-…	today zero target track marchzero
@canarabank i tweet because i cant use mynet banking app due to this problem i m working somewhere i m not taking h…https://t.co/9ivsvJvzDm	tweet cant use net bank app due problem work somwhere take

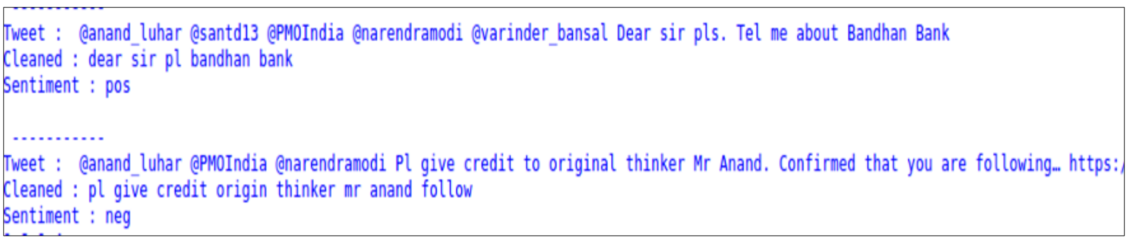
5.2. Profile and Financial Score

5.2.1 Evaluation of metrics of the Multilevel Voting Classifier

The evaluation metrics such as true positive, false positive, true negative, false negative can be found. It has been observed that VC 2 (Tf-Idf) is found to perform better than VC1 (CV) as the features are unigram. Tf-Idf is generally found to cover unigram features better.

Table 6. Evaluation of metrics of the classifier

Model	Accuracy	Precision	Recall
Logistic Regression, Random Forest (CV)	76	72	76
Naive Bayes , Gradient Boosting (CV)	70	76	71
SVC, AdaBoost (CV)	71	71	71
Voting Classifier I (CV Combined)	70	100	71
Logistic Regression, Random Forest (Tf-Idf)	73	75	73
Naive Bayes , Gradient Boosting (Tf-Idf)	68	68	68
SVC, AdaBoost (Tf-Idf)	71	72	71
Voting Classifier II(Tf-Idf Combined)	80	100	80
Voting Classifier III (VC I = VC II)	89	96	89
MLP + Best of Model VC I	89	96	89
PA + Best of Model VC II	63	95	63
Voting Classifier IV (MLP + PA)	89	95	63
Voting Classifier Final (III + IV)	84	100	84


```
Tweet :  @anand_luhar @santd13 @PMOIndia @narendramodi @varinder_bansal Dear sir pls. Tel me about Bandhan Bank
Cleaned : dear sir pl bandhan bank
Sentiment : pos

 ----------
Tweet :  @anand_luhar @PMOIndia @narendramodi Pl give credit to original thinker Mr Anand. Confirmed that you are following… https:/
Cleaned : pl give credit origin thinker mr anand follow
Sentiment : neg
```

Fig. 3. Class for tweets predicted by classifier

This is because adding extra features in the CV may lead to over-fitting. Multi-layer perceptron is found to give the best result when combined with the best performing classifier of model 1. The final voting classifier achieves an accuracy of 84 %. Table 6. illustrates the tweets and their corresponding prediction classification by ensemble. Fig 3 depicts prediction of tweet along with their corresponding class.

5.3. Behavioral Score

Following the method prescribed in the previous sections, the behavioral score of the user is computed as illustrated in Table 7. Pair-wise comparison is implemented on the basis of Bradley-Terry model. The weights have been derived as the average of these probabilistic measures. The threshold value of the behavioral score is aggregated from the Pair-wise comparison. The algorithm has computed this score for 51 relevant users. As mentioned previously, users with null value are a reflection of a negative Profile and Financial Attitude. Hence their score has been 0. The ones whose values are found to exceed the fixed threshold of 170 have a good behavioral score. Further analysis of the scores have been presented as case studies.The behavioral score computation for tweet users is shown in Table 8.

Table 7. Pairwise Comparison

Probability	Weight
P (Profile Score > Twit)	0.13
P (Profile Score < Twit)	0.87
P (Profile Score > Bank)	0.84
P (Profile Score < Bank)	0.16
P (Twit > Bank)	0.44
P (Twit < Bank)	0.56
Average Behavioural Score (for 51 users)	**170**

Table 8. Behavioral Score Computed

Twitter Handle	Profile Score	Twit Score	Financial Attitude	Behavioral Score
@reach4s miles	88.88	55	100	177.66
@saitejapa nems	77.14	75.75	88.88	156.83
@pennews tweet	70	60.5	100	171.24
@Sinha02 4Sinha	64.28	72.5	100	205.37
@Viraraga vanV	0	65	0	0

Table 9. Loans Criteria

Attribute	Factor	Young	Middle-Aged	Senior-Citizen
Other Factors	Educational Qualification	5	2	2
	Nature of Employment	5	3	3
	Experience	2	7	5
	Income	5	8	5
	No. of. dependents	3	5	5
Financial	Loans if any	3	3	3
	Liabilities	2	2	2
	Current Account Balance	6	3	5
	Average credit balance of last 6 months	10	5	10
	Assets purchased on credit if any	4	2	7
Assets	Current Deposit	5	5	5
	Cash Investments	5	5	4
	Movable Assets	3	5	4
	Immovable Assets	2	5	8
Security	Collateral	5	5	7
	Value of collateral	3	3	5
	Guarantor’s Net worth	2	2	10
Credit Exposure	Credit Limit	15	15	10
	Credit card usage	15	10	10
	Total Score	**100**	**100**	**100**

5.4. Evaluation of Credit Score

Credit score can be defined in simple terms as the chances of how likely a person is to pay back the debt. Typical credit scores when summarized over several factors range from 300 to 900. There are generally credit bureaus that calculate these scores. This score not only affects a person’s chances of getting a loan sanctioned but even plays a crucial role in deciding the interest rates. The weights are fixed to the following credit score based on the age group. The factors have been based on a traditional bank application. The weights have been allocated on a scale of 100. On a scale of 900, the threshold is fixed as 500 i.e. a value above that was traditionally granted a loan. Based on that, the threshold is set as 55 for the same. The following weights are assigned as in Table 9.

5.5. Case Studies

The threshold value for behavioural score is fixed as 170 i.e. users having behavioural score greater than the specified limit are considered. Similarly for credit score, on a scale of 100, 55 is the threshold. Due to the non-availability of details or attributes required to compute the credit score, the values presented here are based on assumptions. The total score is computed on a scale of 400 and the values above 226 (thresholds combined) can be sanctioned. The case studies are represented in the following Table 10.

6. Conclusion

Introduction of behavioural score in the banking sector is an advancement that replaces the traditional credit score computation methods and factors. This paper proposes a scoring system that considers both the financial aspects and the social attitude of a person. The first major contribution of this paper is three measures that are used to calculate the behavioural score. The profile score and financial attitude throw light on the personality of the user and calculated by performing sentiment analysis. While traditional sentiment analysis models rely on individual classification algorithms, this system applies a novel multi-level voting classifier. This enhances the accuracy of the results and overcomes the weakness attached with individual prediction models. It can be observed that the model composed of logistic regression and random forest in Level I and along with multi-layer perceptron in Level II reports the best performance. The classifier achieves an overall accuracy of 84%.

Apart from these two scores, there is an emergence of a third unique twit score. The twit score performs a thorough analysis of the profile and scales for factors like authenticity and influence. The second major contribution of this paper is implementation of pairwise comparison (Bradley-Terry Model) which ensures proper aggregation of three scores to compute the behavioural score. A major incentive of this approach is precisely disseminating the magnitude of the three scores while assigning the weights. The behavioural score is summed up with the credit score and the values are evaluated.

Table 10. Case Studies

Sl. No .	Type of User	Profile Score	Financial Score	Twit Score	Behavioral Score	Credit Score	Total Score	Inference
1	Low behavioral score user	50	72.5	76	170	65	235	The user was found to have an average Profile score. A better trend of Twit score and Financial score was observed. This is an example of low risk neutral profile. On analysis of Twit Score, the person has a genuine account but the user's frequency score does not seem convincing. Apart from it, the user enjoys a decent influence in his social circle and has a sound financial attitude. The fact that he has 50 as his profile score serve as litmus to a neutral presence. Since the total score is above the threshold 226, the user is sanctioned a loan
2	Sound behavioral score user	85	90	50	180	70	250	The user was found to have a sound Profile and Financial score. The Twit score was found to be a relatively average value. This case is very similar to the case study 1. In earlier case, the authenticity score was found to be 0. The fact that the person has an authenticity score 4.5 on 5 indicates a strong genuine profile even if slightly pulled down by the FF ratio. A decent influence and frequency score ensure the person is accepted. The combined Behavioral score exceeds the threshold and since the person has a good credit score, the loan can be sanctioned.
3	Good behavioral core user	90.9	87.5	67	221	63	284	It can be reported that user has a very good behavioral score. The person seems to have a good financial attitude as indicated by type of tweets. The Twit score was found to have been a relatively decent value. A strong FF ratio and authenticity score verify the credibility of a person. Though a slightly hampered frequency score, A remarkable relevance score indicates a certain amount of influence in the circle of the user. The computed Behavioral score exceeds the threshold and since the person has a good credit score, the loan can be sanctioned.

This is evaluated against the threshold for loan sanction. This paper assumes the value of credit score due to non-availability of data sources for the same. The third major and main contribution of this paper is that this neoteric multi-linear approach could be incorporated by banks to sanction loans in real time. This would also lower the risk that comes attached with traditional assessment models.

References

[1] A.S Rao, N Pasojevic, Z Li, T Dsouza,. 'Klout score: Measuring influence across multiple social networks', 2015 IEEE International Conference on Big Data (Big Data),(2015)

[2] L. Luceri, T.Braun, S.Giordano, 'Analyzing and inferring human real-life behaviorthrough online social networks with social influence deep learning', Applied Network Scince, 4(1) ,(2019)

[3] Y. Zhang, H. Jia, Y. Diao, M. Hai, Li H,'Research on Credit Scoring by Fusing SocialMedia Information in Online Peer-to-Peer Lending', Procedia Computer Science, 91, pp.168–174, (2016)

[4] S.J.Imtiaz,'A Better Comparison Summary of Credit Scoring Classification.' International Journal of Advanced Computer Science and Applications, 8(7),(2017)

[5] J.Sun, J.Tang, 'A Survey of Models and Algorithms for Social Influence Analysis.' Social Network Data Analytics, pp.177–214, (2011)

[6] A.Bagherjeiran, and R.Parekh, 'Combining Behavioral and Social Network Data for Online Advertising',2008

IEEE International Conference on Data Mining Workshops,(2008)

[7] Y.Wei, P.Yildirim, C.V. Bulte, C.Dellarocas, ‘Credit Scoring with Social Network Data.’ SSRN Electronic Journal, pp 234-258,(2014)

[8] S.Gül,Ö. Kabak, I. Topcu, ‘A multiple criteria credit rating approach utilizing social media data.’Data & Knowledge Engineering, 116, pp.80–99, (2018)

[9] E.Yasa, B.Mucan, “How Turkish Banks benefit from Social Media: Analyzing Banks Formal Links”, International Journal of Strategic Innovative Marketing, Vol. 1, pp. 120-129,(2014)

[10] F.M. Masuabi, L.D. Erasmus, ‘Social media service innovation in South African retail banking: A case study’, 2016 Portland International Conference on Management of Engineering and Technology (PICMET), (2016)

[11] X.Cheng, J.Wu, S.S. Liao, ‘A study ofcontagion in the financial system from theperspective of network analytics’, *Neurocomputing*, 264, pp.42–49,(2017)

[12] J.Bollen, H.Mao,’Twitter Mood as a Stock Market Predictor Computer', 44(10), pp.91– 94,(2011)

[13] S.Rönnqvist, P.Sarlin,'Bank distress in the news: Describing events through deep learning’, *Neurocomputing*, 264, pp.57–70,(2017)

[14] P.Cerchiello,P.Giudici,G.Nicola, ‘Twitter data models for bank risk contagion’, *Neurocomputing*, 264,pp.50–56,(2017)

[15] V.Rao, L. Kupper, “Ties in Paired Comparison Experiments: A Generalization of the Bradley-Terry Model”, *Journal of the American Statistical Association*, vol. 62, no. 317, pp. 194–204, (1967)

[16] Paola Cerchiello, Paolo Giudici, Giancarlo Nicola, "Twitter data models for bank risk contagion",Neurocomputing, Volume 264,Pages 50-56,ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2016.10.101 , (2017)

[17] A.Gupta,P.Kumaraguru,C.Castillo,P.Meier, 'TweetCred: Real-Time Credibility Assessment of Content on Twitter', In: Aiello L.M., McFarland D.(eds) Social Informatics,SocInfo 2014, Lecture Notes in ComputerScience, vol 8851, (2014)

[18] Taylor Francis, 'Extending The Bradley–Terry Model For Paired Comparisons To Accommodate Weights', (2020)