
Using Hidden Markov Model to Monitor Possible Loan Defaults in Banks

Submitted 09/09/20, 1st revision 13/10/20, 2nd revision 15/11/20, accepted 20/12/20

Harish Kamath¹, Noor Firdoos Jahan²

Abstract:

Purpose: Banking business faces a major challenge with defaults. This may not be practical, as there is no control over the borrower's financial situation or their intents to repay. However, if the banks get to know the possible defaults ahead of some actionable time frame, with a certain degree of accuracy in such prediction, Banks may apply any a possible risk mitigation strategy to remediate possible defaults. Willingness to repay the debt and the capability to repay the debt are two primary reasons for the loan default. The subject of this paper is to closely monitor the Facebook activities and check if we can predict if the borrower may become a defaulter any soon, by applying the sentiment analysis on Facebook data and use Hidden Markov model to compute the probabilities of the possible default.

Approach/Methodology/Design: The loan dataset was used for the borrower details and the Facebook data for all those borrowers were gathered. The data from Facebook posts, likes and shares on a borrower were subjected to sentiment analysis, considering income-related information of spend related information on neutral. Hidden Markov model was applied to the polarized data based on the sequence of the sentiment analysis.

Findings: Hidden Markov Model gives the transition probability of state, default or regular, for the observed polarized sentiments from Facebook data for borrowers.

Practical Implications: This mechanism can be integrated into the bank's credit risk management system and could help predict the possibility of a borrower becoming a defaulter. This is very much useful where the tenure of the loan is longer. This research paper fills the gap of active monitoring of the credit risk for long term loans, where the financial status of the borrower could change but the lender doesn't get to know until the borrower stops the repayment.

Keywords: Credit Risk, Financial Risk and Risk Management, Machine Learning, Hidden Markov Model, Sentiment Analysis.

JEL classification: D8, H8, G32.

Paper Type: Research article.

¹Researcher pursuing his PhD in management. He is working as a Senior Vice President for technology division, in a mortgage services based company.
e-mail: bharishkamath@indecmm.net;

²Dr Professor and coordinator of the Department of Marketing and Research centre at R V Institute of Management, Bangalore.
e-mail: noor.firdoos@gmail.com;

1. Introduction

Default is not a new phenomenon. The issue always has been how to reduce it to minimize the default close to Zero. In an ideal scenario, missing the re-payment of the equated monthly instalments can be termed as default. However, Banks have their definitions, governed by their central banks. Some may treat 3 consecutive missed payments as default, some may term it default as soon as the first payment is being missed. The central idea of the paper is not on the definition of the default, but how to identify the possibility of a borrower missing the re-payment in their next instalment. The probabilistic estimation of default ahead of time, it could happen will give the advantage to the Banks to prepare for the possible default and plan the mitigation. In the absence of such warning systems, the mitigation is to face it and live with it with conventional options like issuing demand notice to the borrower, initiate legal proceedings etc., each of these is time-consuming and not desired by the borrower or the banks.

Social media has emerged as a platform for individuals to display their emotions, their status, Hobbies, achievements etc. Individuals want others to know about oneself post the information they want to share. Whatever is shared publicly can be used as a dataset for the analysis about an individual, without invading the data privacy. There are many pieces of research have been done on social media dataset, which analyses an individual based on the post's comments, shares, likes, etc. Facebook seems to be more comprehensive among these. For example, if we consider the twitter, which has the limitation on the text length, or the image/video size restricts the continuous flow of thoughts and needs to be broken up into multiple tweets. In Facebook, an Individual can share or express their feelings without any limitation to their flow of thoughts broken down to multiple posts. This gives an edge to Facebook dataset over the other social media dataset.

There are several mechanisms to analyse this social media dataset and build the interpretation layer to help with the business needs. Machine learning and Artificial intelligence (AI) are the leading options to solve for the business problems from the huge data cache of data. Machine learning is a subset of AI. AI adds up to "learn from mistakes" or learn from/ fix from "false positives" or as some people say, the "adopt the feedback mechanism". Machine learning as a key component of AI has many techniques to solve business problems. Each of such techniques solves for a unique set of problems. The discussion in the paper would be on the Hidden Markov Model (HMM).

In 1906, Andrey Andreyevich Markov invented the Markov chains, when he first created the first theoretical results for stochastic processes that used the term "chain." In 1913 he evaluated Russian-language letter sequences using this model. A stochastic model describes a condition where the data points are ambiguous. Which is to say, it is a design for a process that has some form of unpredictability or randomness. HMM can solve for to understand the probability of the transition

between hidden states using the observed data in a specific sequence. So, we can understand and analyse the sequence of the observed data from social media data set and determine the state, i.e. probability of a borrower becoming a defaulter or continue to be regular. This whole model can be trained and tuned for default prediction and used for continuous monitoring of the probable change in the state of a borrower.

Loans are classified based on the purpose of lending. There are multiple types of loans. They would include Auto loan, personal loan, Home loan, a.k.a. mortgage loan, etc., All of these have certain max tenure. For example, the Auto loan may be extended up to a maximum of 5 years and personal loans can be extended for 2 years. Whereas the mortgage loans have the longest duration, to be re-paid. It would extend up to 30 years. 30 years period is a very long period from the risk management perspective. We can always have a mortgage Insurance to cover when any untoward situation occurs to the borrower. In other cases, the borrower is not able to pay due to the reduced pay or layoff, divorce, addition to family etc., or not willing to pay back. Hence there is a strong case from the business to identify the possible defaults ahead of time by individual borrowers.

Social media datasets are often subjected to sentiment analysis for the purpose of solutioning business problems. Sentiment analysis is the division of words/expressions/ sentences into a various degree of sentiments or various buckets of designated sentiments. General practice is to classify them into three buckets. Positive sentiment, Negative sentiment or Neutral sentiment. The data is put into one of the three-bucket and further analysis will be done. The positive or negative will be determined in the context used for the analysis. There are multiple techniques available for the sentiment analysis. Each of such technologies has its own unique reasoning as to why they are preferred. The general approach is the lexicon-based approach, where the set of words are pre-classified as positive or negative and the words matching them would be considered as positive or negative sentiments, rest of them are treated as neutral.

There are several applications built around the Hidden Markov model for business solutions. Some of them might have been successful and some remained on the academical papers due to limitations of application in the real world. Following related works have helped shape this paper. is a lack of organizations who support the young graduate aspirants with sharia compliance financing.

Problem Statement : Banks cash flow will be impacted by the outstanding of the loan. When a borrower stops repaying the loan for whatever reasons, only then the banks or the lenders get to know that the loan is overdue for repayment or became the default. As an example, the mortgage loans usually have a longer tenure up to 30 years. It would be hard to imagine that the borrower's financial state would remain the same for all those years. Moreover, of the borrower is no more interested in the property they have purchased or no more motivated to repay the loan, even then the loan gets into the default state.

Current credit risk management systems have no way to gather any intelligence on the borrower's capability or willingness to repay. Hence only after the missed payment, the lender would know that it stopped. Epidemics, economic downslides can job loss or economy boom can provide better job to the borrower, which have a negative or positive impact on the repayment.

If banks knew the probability of a borrower becoming a defaulter, before it actually happens, banks can handle the risk better. There could be many borrowers who may be impacted by the economic slowdown, so the impact from such borrower would choke the cashflow of the bank. In order to handle such a situation, they need to know the probability of borrowers becoming the defaulters ahead of time to venture the different strategies to mitigate possible large risk.

This paper attempts to address this problem statement by analysing the Facebook data of the borrowers, whose profiles are publicly available. Posts, Likes and Shares by the borrower will be subjected to sentiment analysis. The polarization of the data will be based on the Income-related or spend related or none of these two(Neutral). Such data will be then run thru Machine learning, especially the hidden Markov Model(HMM), by understanding the sequence of the polarity of posts shares and likes.

Credit risk management and Social media: Credit risk is an unavoidable risk in the Banking business. Credit risk will choke the cash-flow in a banking system. To maintain a healthy cash-flow, one of the important measures is to reduce the defaults. Ensure that the borrowers pay their monthly instalments on time, every time. There are journal articles, research papers on using the Social media dataset as a database for generating the credit score, which could be used to decide on granting the loan. This could be used as a sole decision point or could be clubbed with other decision-making metrics. Some of the microfinance companies, use primitive data from Facebook and decide to lend based on how long the person is on the Facebook, what is the frequency of the posts?, How the borrower is connected socially? Does the borrower have any friends in their network, who has been a defaulter, etc? There are some commercial applications, which generate the score from Facebook, assisting the lending. These are all primarily intended to filter the borrowers, who may possibly become defaulters. These techniques are used in the loan origination phase of the loan lifecycle. There is a need for monitoring the risk at the loan servicing stage of the loan lifecycle.

2. Materials and Methods

The approach includes estimating the potential default/delinquency of the loan lifecycle during the loan servicing period. The risk could have been assessed already at the origination lifecycle by checking the standard set of financial records verifying the ability to repay and willingness to repay the loan. During the service cycle, which is the main duration, which ranges up to 30 years. In this time, there may have

been several changes in the life of the person during the lifetime of the loan, which can have a negative impact on the payment of the loan. It is important to know them in advance, ideally before it happens. The process includes the following steps: First, the loan information is obtained from various financial institutions, with name, (delinquent or regular) loan status. The second set of data is gathered the Facebook details of those borrowers associated with the collected loan data. There have been cases, where the borrower's profiles are private consumption only by restricting public access. There were missing profiles or no profiles available matching the borrower name. Such loans had to be dropped from the corpus.

Loan data collection: About 1400 loans were considered for validation. The dataset contains both the Defaulters (Already missed the onetime payment of the dues) and regulars (Who are paying the dues, on time every time). Using simulations or using a hidden Markov model is often easier to estimate the potential reduction of disposable revenue through posts displayed on Facebook. In this case, an attempt was made over simulated data, to make it more of supervised learning with actual delinquency data, along with regular loan data.

Facebook Data collection: Using social media data grabber tools, the Facebook data on "like," "share" and new "post" were registered for each of the borrower listed, be its normal or defaulter. In the case of defaulters, closely examined the months preceding the actual non-repayment of the loan date. Both these data are carefully collected in case of a re-payment resumed after months.

Pre-processing data for sentiment analysis: Pre-processing of the data from Facebook is critical to the success of the sentiment analysis. The standardization of the data was achieved thru software packages like Natural language tool kit (NLTK) etc., The pre-processing would include the removal of unwanted data like HTML tags, repeated punctuations, removal of URL from the data without impacting the core sentiments. Removal of unnecessary textual expressions like Laughing loudly, etc, which doesn't have bearing on the core sentiments. Splitting the attached word is yet another data cleansing activity.

Using Sentiment Analysis in Social media datasets: Sentiment Analysis is the core activity and the mistakes done here can skew the outcome. Hence utmost care was taken at this stage. Bag of words approach was flowed to split the data into degrees of sentiments. Any events or data points related to income, like a "New Job", "Promotion", "Sales goal reached", "Got appreciation from the company" were considered as positive sentiments, which could positively impact the disposable income. The negative sentiments were those, which would be a possible spend or reduction in disposable income. If the data is nothing to do with income or no impact on income would be treated as neutral sentiments.

Sentiment Analysis is the most common algorithm that analyses a message and tells whether the underlying sentiment is positive, negative or neutral (Amaechi, Dhouioui, and Lopez-Huertas Perez, 2013). Zhao (2013) has a very good, detailed

source of information on using R tool to do the sentiment analysis on the Facebook dataset. Piozzi *et al.* (2016) analyze and explain the research approach and implementation challenges in the Sentiment Analysis space using various techniques such as linked data model, sentiment and emotional analysis, etc. Sentiment Analysis is a methodology where context is set, and any emotions can be measured against the established context as positive, negative, or neutral. This book is addressed at researchers and takes an interdisciplinary approach from many computing fields like the analysis of natural language, machine learning, big data and statistical methodologies. This book has been referred for clarification in understanding the Analysis of emotions and different methods necessary for this study.

In our research anything that could impact the repayment of monthly due is considered as a negative sentiment, they could be anything to do with the reduction in willingness to repay, exhibited explicitly or reduction of disposable income thru exhibited non-regular spends. Sentiment Analysis will help classify the data into negative sentiment, positive sentiment or neutral sentiment. A pool of words will be used for this classification. List of keywords is entered into a text file as a Negative word set or positive word set. Rest is considered neutral sentiment. As per the standard approach, these two text files will be picked up for a set of words and any words matching from the files will be marked and accounted for positive or negative, based on the file where the word is matching.

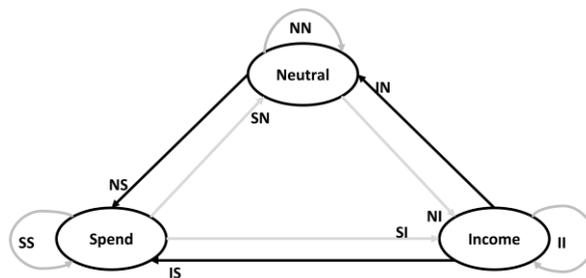
Machine Learning: In Machine learning, there are multiple algorithms for each of the problem's statements. They could be regression, classification, etc., One of the key techniques used in Machine learning. HMM is a graphical probabilistic model type that can assist in predicting a sequence of unknown (hidden) variables from a set of observed variables. A simple example of an HMM is weather prediction (hidden variable) based on the type of clothes that somebody wears (observed). Markov Chain is a random process of its kind. There are many applications built based on HMM, primarily for the probabilistic prediction to expose the probability of hidden states from the observation data. One of the strongest references and previous work on using HMM for credit risk could be seen in Ntwiga's research paper There are several variations in the application of HMM to credit risk such as applying it to low-income earners (Bundi Ntwiga *et al.*, 2018), M-Shwari credit rating using Hidden Markov Model (Davis Bundi and Patrick, 2016). Representation of peer group and individual low-income earners in micro-credit lending by M-Shwari: A hidden Markov model approach (Bundi Ntwiga, Weke, and Kiura Kirumbu, 2016). The data scientist community has recently released the tools to ease HMM model programming using the scikitlearn library in Python. A paper published by the IEEE (Srivastava *et al.*, 2008) addresses HMM mode use.

Applying the Hidden Markov Model to the dataset: Extending the HMM technique used in other researches for credit risk, the HMM model is built. About HMM

(Ghahramani 2002) Credit risk detection using HMM, (Srivastava *et al.*, 2008) are the best sources for the HMM usage in credit risk management.

Hidden Markov Models are a type of stochastic state-space model. HMM assume the existence of "hidden" or "latent" states that are not directly observable. These hidden states influence values which are observable, known as the observations. HMM helps to ascertain the current state from the set of known observations. Hidden Markov Models are a kind of state-space stochastic modelling. HMM assumes that states which are "hidden" or "latent" are not easily observable. Such invisible or hidden states affect observable values, known as observations. HMM helps to assess the current state of the known observations. The HMM model indicates that the future state/event is only dependent on the current state/event and not on any other older state (known as Markov Property). In our case, if we find the pattern of Sentiments (Spend, Earnings, and Neutral), then we can assume that the delinquency will occur based on the current series of sentiments and not on the old ones. The collection of data must, therefore, be current to carry out sentiment analysis.

Figure 1. Understanding the Markov Model



Source: Own study.

Probabilities of the state transition can be defined using a (N x N) matrix. This matrix is known as the Transition Probability Matrix as follows:

$$Z = \begin{bmatrix} II & IS & IN \\ SI & SS & SN \\ NI & NS & NN \end{bmatrix}$$

When the machine transitions one state to another state, the sum of all transition probabilities placed in a row must result in 1. In the above matrix, II+IS +IN = 1. The mathematical formula for the same would be:

$$\sum_{y=1}^N xy = 1$$

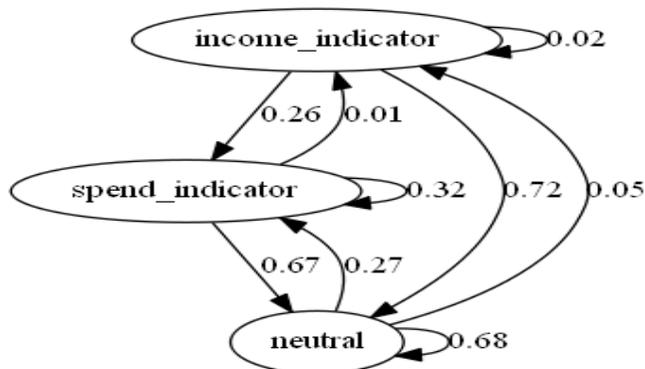
In general applications, the Hidden Markov model will be used for computing the state transition probabilities by the observables. In the case of research, the positive sentiment, Neutral sentiments and Negative sentiments are the observed variables. These sentiments are the result of the sentiment analysis performed on Facebook data. At first, we compute the transition probabilities of each sentiment to the next sentiment, that is, From Income to Spend (IS), Spend to Neutral (SN) and Neutral to Income (NI). All such probabilities are put into and matrix defined above. We can begin the transition from any point and compute the probability of the transition. It is important to note that most of the posts are generally neutral in nature as per the data analysis. There are more negative sentiments than positive sentiments. The events related to positive sentiments are comparatively lower to Negative sentiments in most of the profiles.

In the Facebook post share likes, the time series is being captured to ensure we could enumerate the sequence of polarized messages. Each individual polarized (Sentiment analyzed) messages were tabulated with the sequence number, which will help build the sequence of observations. Figure 2 is a derivation from that tabulated data.

3. Results

The following diagram is the outcome from the HMM Model applied to the dataset after sentiment analysis.

Figure 1. Observed data transition/Sentiment Transition diagram

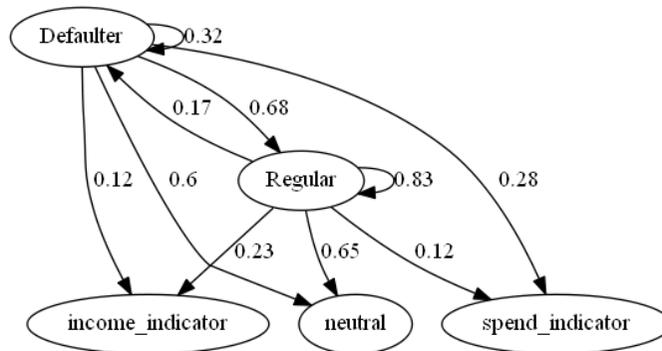


Source: Own study.

From the above diagram, we can infer that there are 2% chances that someone keeps posting about the income indicators back to back whereas 26% of the time who has indicated income would show off the spend also. About 32% of the time who display spend indicators continue to display spend indications. Spend indicators will transition into 67% of the time into Neutral sentiments. The transition probability of Neutral sentiment to Neutral sentiment is about 68%. Similarly, the probability of transition from Neutral to Income indicators is about 5%. And the reverse transition

is about 72%. It is important to note that the sentiments are very high in neutral than the other polarities.

Figure 3. State and Observed data transition data



Source: Own study.

Figure 3 can be interpreted as transition probability between regular to defaulter is 32% and defaulters turning into regular, after missing the payments or receiving the warnings is about 40%. Regulars have a spending transition of about 20% but most of them show (60%) Income indicators being regular. From the above analysis, it can be inferred that the HMM can be used for computing the probability of a borrower becoming defaulter.

Using the above transition matrix with the probabilities, it was verified, on the existing defaulted loans, the match was found to be 95.8%. Accuracy of the current model is sufficient for the possibility of prediction default ahead of time. With this, it can conclude that we can apply the Hidden Markov model on the Facebook data after sentiment Analysis to predict the probability of a borrower becoming defaulter based on the sequence of the posts. However, the accuracy can be fine-tuned by using artificial intelligence or retrain the model with more false positives and false negatives, periodically.

4. Conclusion

Hidden Markov Model is widely used for detecting the hidden state from the observations. With the machine learning gaining the momentum, there is a huge opportunity to look at the possible use cases or implementations, that helps solve the real business needs beyond the lab conditios. Banks struggle to find the possible defaulters ahead of time, that is, before someone really becomes a defaulter. If the Credit Risk Management System (CRMS), can even predict the possible default ahead of 15 days to 3 months, it helps benak on better manage the cashflow.

Long term loans are more riskier than the short term. During the loan servicing period, a lot of changes can happen in a person's life. One can lose the job, one can get promotions. So if we can capture such changes through social media like

Facebook, etc. Then we can apply the sentiment analysis to polarize the data in the context of income or spend. Such data can then be subjected to HMM model and we can make the predictions about a borrower becoming a defaulter, soon. This can be enhanced further and make commercially viable product.

References:

- Akaichi, Jalel, Zeineb Dhouioui, Maria Jose Lopez-Huertas Perez. 2013. Text Mining Facebook Status Updates for Sentiment Classification. In the 17th International Conference on System Theory, Control and Computing, ICSTCC 2013, Joint Conference of SINTES 2013, SACCS 2013, SIMSIS 2013, Volume of Proceedings.
- Bao, Wang, Ning Lianju, Kong Yue. 2019. Integration of Unsupervised and Supervised Machine Learning Algorithms for Credit Risk Assessment. *Expert Systems with Applications*, 128, 301-315.
<https://www.sciencedirect.com/science/article/pii/S0957417419301472>.
- Barboza, Flavio, Herbert Kimura, Edward Altman. 2017. Machine Learning Models and Bankruptcy Prediction. *Expert Systems with Applications*.
- Bhargava, Mandava Geetha, Duvvada Rajeswara Rao. 2018. Sentimental Analysis on Social Media Data Using R Programming. *International Journal of Engineering and Technology (UAE)*.
- Bundi Ntwiga, Davis, Carolyne Ogutu, Michael Kiura Kirumbu, Patrick Weke. 2018. A Hidden Markov Model of Risk Classification among the Low Income Earners. *Journal of Finance and Economics*.
- Bundi Ntwiga, Davis, Patrick Weke, Michael Kiura Kirumbu. 2016. Trust Model for Social Network Using Singular Value Decomposition. *Interdisciplinary Description of Complex Systems*.
- Chen, Qing, Hui Feng Xue, Li Yan. 2011. Credit Risk Assessment Based on Potential Support Vector Machine. In *Proceedings – 2011, 7th International Conference on Natural Computation, ICNC 2011*.
- Cortes, Alvaro, Christopher E. Herbert, Erin Wilson, Elizabeth Clay. 2007. Factors Affecting Hispanic Homeownership: A Review of the Literature. *Cityscape*.
- Cortés, Kristle, Ran Duchin, Denis Sosyura. 2016. Clouded Judgment: The Role of Sentiment in Credit Origination. *Journal of Financial Economics*, 121(2), 392-413.
- Culotta, Aron. 2010. Towards Detecting Influenza Epidemics by Analyzing Twitter Messages. In *SOMA 2010 - Proceedings of the 1st Workshop on Social Media Analytics*.
- Dahiya, Shashi, S.S. Handa, and N.P. Singh. 2015. Credit Scoring Using Ensemble of Various Classifiers on Reduced Feature Set. *Industrija*.
- Dubois, Thomas, Jennifer Golbeck, Aravind Srinivasan. 2011. Predicting Trust and Distrust in Social Networks. In *Proceedings - 2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing, PASSAT/SocialCom 2011*.
- Eddy, Sean R. 2004. What Is a Hidden Markov Model? *Nature Biotechnology*.
- Education, Marwadi *et al.* 2014. An Improved Sentiment Classification Using Lexicon into SVM. *Journal of Chemical Information and Modeling*, 53(1), 37-42.
- Ghahramani, Zoubin. 2002. An Introduction to Hidden Markov Models and Bayesian Networks. *International Journal of Pattern Recognition and Artificial Intelligence*.
- Ghailan, Omar, Hoda M.O., Osman Hegazy. 2016. Improving Credit Scorecard Modeling Through Applying Text Analysis. *International Journal of Advanced Computer*

Science and Applications.

- Kaushik, Chetan, Atul Mishra. 2014. A Scalable, Lexicon Based Technique for Sentiment Analysis. *International Journal in Foundations of Computer Science & Technology*.
- Khadjeh Nassirtoussi, Arman, Saeed Aghabozorgi, Teh Ying Wah, David Chek Ling Ngo. 2014. Text Mining for Market Prediction: A Systematic Review. *Expert Systems with Applications*.
- Kruppa, Jochen, Alexandra Schwarz, Gerhard Arminger, Andreas Ziegler. 2013. Consumer Credit Risk: Individual Probability Estimates Using Machine Learning. *Expert Systems with Applications*.
- Ntwiga, Davis Bundi. 2016. Social Network Analysis for Credit Risk Modeling.
- Pekar, Viktor, Jane Binner. 2018. Forecasting Consumer Spending from Purchase Intentions Expressed on Social Media.
- Pozzi, Federico Alberto, Elisabetta Fersini, Enza Messina, Bing Liu. 2016. Challenges of Sentiment Analysis in Social Networks: An Overview. In *Sentiment Analysis in Social Networks*.
- Risk Management Framework in Banks. 1980. Reserve Bank of India, 1-22.
- Saia, Roberto, Salvatore Carta. 2016. 268 Lecture Notes in Business Information Processing. An Entropy Based Algorithm for Credit Scoring.
- Schoen, Harald *et al.* 2013. The Power of Prediction with Social Media. *Internet Research*, 23(5), 528-543.
- Smales, Lee A. 2016. News Sentiment and Bank Credit Risk. *Journal of Empirical Finance*, 38, 37-61.
- Srivastava, Abhinav, Amlan Kundu, Shamik Sural, Arun K. Majumdar. 2008. Credit Card Fraud Detection Using Hidden Markov Model. In *IEEE Transactions on Dependable and Secure Computing*.
- Tan, T., T.Q. Phan. 2016. Social Media-Driven Credit Scoring: The Predictive Value of Social Structures. *International Conference on Information Systems, ICIS*, 1-11. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85019496658&partnerID=40&md5=71993eabbd3ebab69200a09a22bd87c3>.
- The Basel Committee. 2006. Principles for the Management of Credit Risk. IFAS Extension.
- TransUnion, Cibil. 2018. Quarterly Overview of Consumer Credit Trends Released by TransUnion CIBIL.
- Tumasjan, Andranik, Timm O. Sprenger, Philipp G. Sandner, Isabell M. Welpe. 2010. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. In *ICWSM 2010 - Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*, 178-185.
- Vaidyanathan. 2013. Credit Risk Management for Indian Banks. *South Asia Economic Journal*.
- Wei, Liwei, Jianping Li, Zhenyu Chen. 2007. Credit Risk Evaluation Using Support Vector Machine with Mixture of Kernel. In *Lecture Notes in Computer Science, Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*.
- Wharton. 2014. The Surprising Ways That Social Media Can Be Used for Credit Scoring. *Knowledge@Wharton*.
- Zhao, Yanchang. 2013. R and Data Mining R and Data Mining.
- Zheludev, Ilya, Robert Smith, Tomaso Aste. 2014. When Can Social Media Lead Financial Markets? *Scientific Reports*.