USING SOCIAL MEDIA AS A TOOL FOR MONITORING CREDIT RISK IN BANKING

Mr. Harish Kamath*
Dr. Noor Firdoos Jahan**

Abstract

Credit risk is unavoidable part of the banking business. Credit risk refers to the risk of inability of borrower to repay the principal to meet contractual obligation. This is also termed as default risk. The credit risk is assed at the beginning of the loan origination. The best way to handle such risk is to identify, quantify, monitor and control upfront. Credit risk should be monitored and reported to regulatory authorities, as a part of regulatory compliance. It's an important strategic planning for any financial institution to build a sound Credit Risk Management System (CRMS).

There are multiple risk assessment and mitigation strategies in place to control such risk. It would vary from bank to bank. The primary parameters for any Credit Risk Management System are 'ability to repay' and 'willingness to repay'. The current Credit Risk Management Systems use the static values for the risk parameters. In other words, the values of any parameters are not validated for any change since the data was collected. This study attempts to capture and use the dynamic values for the Credit Risk Management system.

Social Media could be a very good source for collecting the value of 'ability to repay' and 'willingness to repay'. Though the social media is not an authentic data source to base the Credit risk, we can derive a model which can help better quantify the risk, which can then be used for monitoring continually and gets us the most accurate risk index. The Source of data includes but not limited to Twitter, Facebook, News Portals, Weather portals, Public Economic surveys results published in various social portals.

Key words: *Bank, Credit risk, Loan, Risk management, Borrower, Lender.*

Introduction

Credit risk emanates from a bank's dealings with an individual, corporate, institution etc. As per the Reserve Bank of India Guidelines (RBI), Each Banks which are under the purview of RBI, need to have a Credit risk Management system in place. There are similar guidelines by various competent authorities. The primary purpose of this is to reduce the Credit risk of the Banks and ensure reduced Non-Performing assets (NPA) [i]. In the present credit risk framework, Value of parameters is static in

***Mr. Harish Kamath,** Research Scholar, ISBR Research Centre, Bangalore, Email:bharishkamath@gmail.com

****Dr. Noor Firdoos Jahan,** Professor, R V Institute of Management, Bangalore.

nature. In other words the value of parameters used for assessing the credit risk at the beginning of a loan origination may have changed but the risk systems do not track them. Same risk data is used for risk monitoring also. The parameters include periodic Plant visit, Industry study, employment verification etc.

Social media is term used to paraphrase the group of internet based applications or technologies that enable people to create, share or exchange ideas among its members. Social media has its presencein the form of Facebook, blogging and Micro blogging, Twitter, Collaborative projects like Wikipedia, Content communities like YouTube, Social news networking like Digg and Leakernet, Virtual social worlds like SecondLife etc.

Social Media has opened avenues of opportunities in the field of Analytics. There is a huge dataset available for the Analysis of various dimensions of human behaviour. With Social media, the data cannot be verified for its trueness. That's the downside of the dataset. There have been several commercial usages of the Social media dataset. This is used to understand the taste of customer for the future products, Feedback from customers on its current range of products. What change people want to see? How often they want changes to products etc. Fashion Designers, Consumer durables, are the key innovators, leveraged the social media for its product design and feature addition or removal.

Banking and Financial institutions have been the watchers on Social media to leverage the datasets. Due to the credibility of the data being posted on the Social media, the Banking and Financial institutions have been sceptic to leverage such huge treasure of dataset. Social media contains the dataset which can potentially be used to identify any possible credit risk at individual level or at the portfolio level. Social media contains the dynamic and most current data points which can be used to analyse the lifestyle, current status of job, increased expenses etc.

Literature Review

Banks are often limited by static dataset and keep them captured for their credit risk assessment. Credit risk monitoring cannot be driven by static data points like Salary certificate, IT returns statement etc. These data points would vary from month on month. Dynamic nature of lifestyle and social influence would have a positive or negative impact on the credit risk. Credit risk can arise from either capacity to repay or from willingness to repay. Each of these data points is dynamic in nature.

The literatures have been studied in three primary areas.

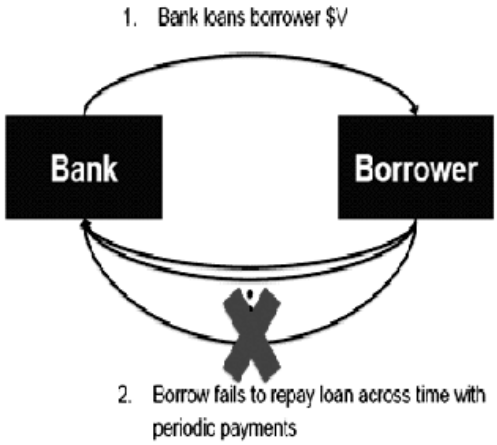
1. Credit Risk – Basic understanding of Credit risk and opportunity for improvement
2. Sentiment Analysis Basic understanding and application
3. Social Media dataset Facts and figures

1.1 Credit risk

- Credit risk is most simply defined as the potential that a bank borrower or counterparty will fail to meet its obligations in accordance with agreed terms. [ii]
- "Credit risk is the risk of loss due to a debtor's non-payment of a loan or other line of credit."

The Risk management journal, by Bank for International Settlement (BIS) says,

- "Principle 3: Banks should identify and manage credit risk inherent in all products and activities. Banks should ensure that the risks of products and activities new to them are subject to adequate risk management procedures and controls before being introduced or undertaken, and approved in advance by the board of directors or its appropriate committee"[iii]
- Credit risk monitoring refers to incessant monitoring of individual credits inclusive of Off-Balance sheet exposures to obligors as well as overall credit portfolio of the bank.[iv]



Source: Global association for Risk professionals. The GARP Risk Series

CREDIT RISK MANAGEMENT, Chapter 1 | Credit Risk Assessment deck Slide 6

Risk Monitoring contains three parts. Risk Identification, Risk measurement and Risk Monitoring

Risk Identification

All types of risk must be identified and their likely effect in the short-run is understood. A bank that has international operations may experience different intensity of credit, market and operational risks in various countries when compared with a pure domestic bank. Also, even within a bank, risks will vary in its domestic operations and its overseas arms.

Risk Measurement is also part of Risk monitoring. Acceptable risk thresholds can be set to generate the alerts, while monitoring the risks.

Risk Measurement

Measurement means weighing the contents and/ or value, intensity, magnitude of any object against a yardstick. The objective should be to find out and understand the exact degree of risk elements in each category in the operational environment. While a very simple qualitative assessment may be sufficient in some cases, sophisticated methodological/statistical models may be necessary in others for a quantitative value. In bank lending activity, a simple one –

page credit rating system along with simple financial analysis may be sufficient for a relatively low credit limit. But for higher credit limits, detailed qualitative and quantitative analysis with sophisticated credit rating models may be required because the size of the credit limit would indicate the size of default risk and its ultimate effect on the institution

Risk Monitoring

Keeping close track of risk identification measurement activities in the light of the risk, principles and policies is a core function in a risk management system. For the success of the system, it is essential that the operating wings perform their activities within the broad contours of the organizations risk perception.

Sentiment analysis

Sentiment analysis (also known as opinion mining) refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. Sentiment analysis is widely applied to reviews and social media for a variety of applications, ranging from marketing to customer service. (Source: Wikipedia)

Sentiment Analysis is now part of commercial applications. Sentiment Analysis has grown beyond the Research papers and now being used in the real-world. It has several applications.

Approach to sentiment analysis, is to, find all positive words in text and increment positive count for each positive word. Similarly find all negative words in text and increment negative count for each word. At the end, which ever count is greater, you can say text is polarized to that sentiment. There are 3 approaches available for performing Sentiment Analysis.

1. **Lexicon based** - considers lexicon dictionary for identifying polarity of the text
2. **Machine learning based approach** - Needs to develop classification model, which is trained using pre-labeled dataset of positive, negative, neutral content.
3. **Combined approach** - Which uses lexicon dictionary along with pre-labelled data set for developing classification model.

1.2 Social media Dataset for Credit Risk Monitoring.

Social media penetration

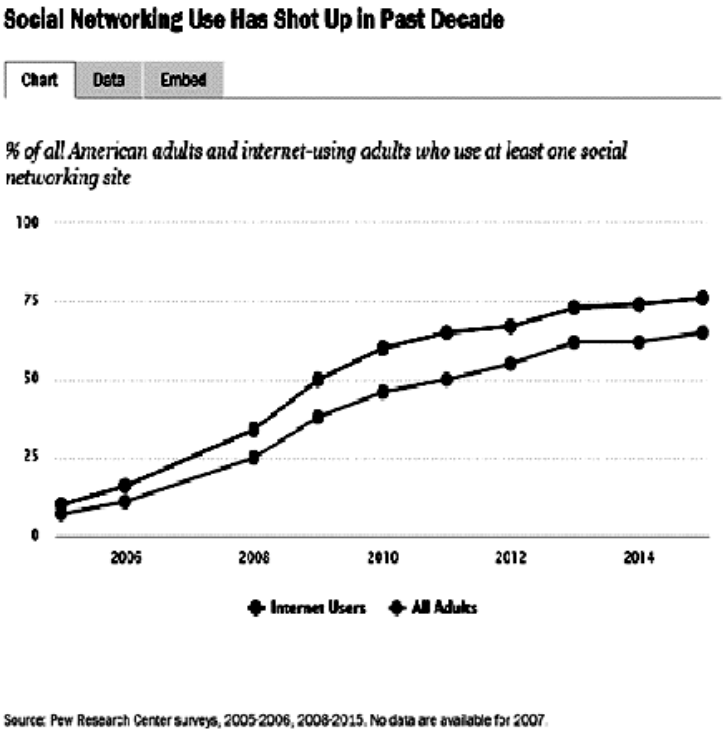
Social media has a large volume of data. The dataset includes Facebook, Twitter, Instagram, Weather Portals, News feeds, National economic Statistics for public consumption etc. Here are the statistics on few of the key Social media components. Source: Pew Research[v]

- There are just over 3 billion active Internet users
- Nearly 2.1 billion people have social media accounts
- 3.65 billion mobile users have access to the internet via smartphones and tablets
- Close to 1.7 billion people have active social media accounts
- There are nearly 1.4 billion Facebook users
- 47% of all Internet users are on Facebook
- 4.5 billion likes are generated daily
- Twitter has 284 million active users at last count

- Google+ chas 363 million users
- Instagram has 300 million users
- 70 million photos and videos are sent daily
- 53% of internet users aged 18-29 use Instagram
- LinkedIn has 347 million registered members
- Viber has over 200 million users
- There are 639 million users on Qzone (China)
- 600 million users on Whatsapp
- Facebook messenger has 500 million users

The usage has gone up drastically over the last decade. 65% of adults now use social networking sites – a nearly tenfold jump in the past decade[vi]

. The following image shows the usage of social media over the decade. This gives us a rich dataset to explore and perform the sentiment Analysis in the context of credit risk.



Source: Pew Research Center surveys, 2005-2006, 2008-2015. No data are available for 2007.

Objective of the study

Banks currently uses a risk model, giving varied weights to different parameters[vii]. These parameters will have different degree of importance. All of these parameters are mostly static in nature, which means the data collected once are not validated at regular intervals for these dynamic parameters. This raises the need for development of dynamic Risk assessment

model, which can perform its risk assessment based on the dynamic values. Banks want to reduce its level of Non-Performing Assets (NPA) in Personal loan or mortgage loans by tightening its credit policy. The development of such framework will standardize the judgment and remove the element of bias and subjectivity.

Method

The Model will leveraging the Sentiment Analysis Methodology to mine the social media and segregate the sentiments into positive or negative in the context of Credit risk

Identify the parameters impacting the credit risk.

Identify the parameters that affect credit risk of scoped loan. The first step in the development of the model is the identification of the various risks and the parameters to be taken into consideration. For this purpose, the various manuals and websites pertaining to the credit appraisal for corporate loans would be carefully studied. Credit risk score done by internal risk monitoring system will be taken into consideration. Both quantitative and qualitative aspects will be taken into consideration while computing the risk levels. Assessment will be undertaken to surface top 20 parameters that need to be taken into account while evaluating a prospect. These parameters would be classified under two heads namely Willingness and capability risks. Weights developed in this study were based on a conceptual understanding and the importance attached by people proficient in this area.

Obtain the Social media dataset

Social media dataset will be the source of Analysis. Data from Facebook, Twitter, LinkedIn, Job Portals, and Weather Portals can be obtained for more accuracy.

Perform Sentiment Analysis

Sentiment Analysis is an application of Natural Language Processing which targets on the identification of the sentiment (positive vs. negative vs. neutral), the subjectivity (objective vs. subjective) and the emotional states of the document. We would label each sentiment which can impact eh credit risk negatively or positively. We will not pay attention to neutral sentiment. There are multiple tools available for Sentiment analysis. These are Python NLTK (Natural Language Toolkit), R, TM (text mining) module, RapidMiner.

Perform Discriminant Analysis

Discriminant Analysis can be used to develop a function to discriminate between three categories of risk (moderate risk clients, high risk clients and very high risk clients). We need to find the factors which contribute the maximum in discrimination. To build the discriminant model the value of the dependent and the independent variables for the set of loans can be entered in the SPSS or any other competent software and the weights/ coefficients for all the parameters can be computed.

Assign Weightage based on recurrence of risk for that parameter

Identify the parameters which are more likely to impact and assign the relative weightage to other parameters. The number of occurrence and the Amount which is at stake, if the risk occurs.

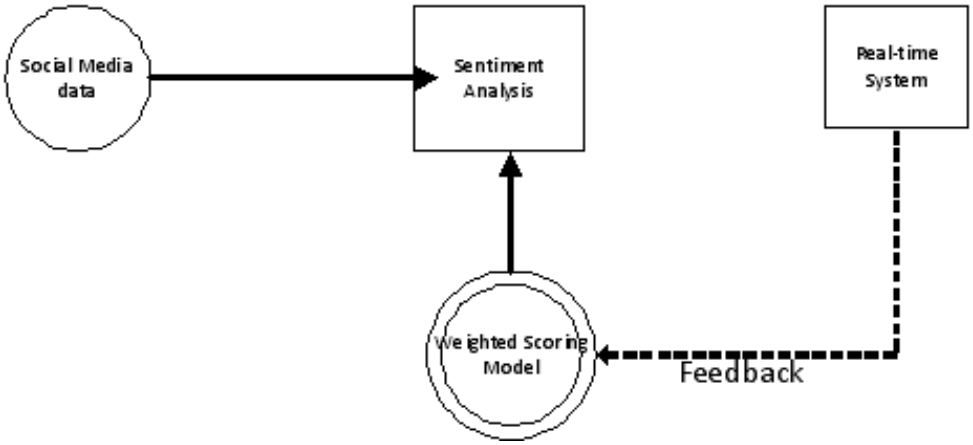
Compute the risk score

Compute the risk score based on the weightage assigned to each of the parameter, derived from the Sentiment Analysis. This will help to generate the alerts for Monitoring system.



Feedback system

Feedback Mechanism is the critical part of this model. The Real-time systems will record the actual default and give the feedback to the system. There are cases where the default has happened from totally un-captured parameter. The Feedback mechanism will help to add that to the scoring system and make the weighted scoring mechanism more robust.



Conclusion

The fundamentals of any Credit risk a.k.a. default risk, lies in multiple factors and omission of the dynamic human factor cannot be ignored. There a huge dataset available in the Internet for capturing the dynamic human factors impacting the Credit risk positively or negatively.

Using the sentiment Analysis, which has been established its credentials over the time. Using the statistical models like Discriminant Model to classify those into Positive or negative in the context of Credit risk.

Weightage can be assigned to each kind of event or label and compute the credit risk score. This will help capture the dynamic risk, and make the model more robust, which otherwise is not factored into risk model.

Literature References

[i]As per latest RBI guidelines vide RBI/2013-14/62 dated July 1, 2013, an asset, including a leased asset, becomes non-performing when it ceases to generate income for the bank.

[ii]BIS bis.org Basel Committee on Banking Supervision (BCBS): https://www.bis.org/publ/bcbsc125.pdf

[iii] Principles for the Assessment of Banks' Management of Credit Risk, it cites this principle about risk management. (https://www.bis.org/publ/bcbsc125.pdf)

[iv]Commercial Credit Risk Monitoring & Control Efforts.Linkedin Article on Credit risk monitoring by Jim RiceManaging Director at Supporting Strategies - Orlando/Rohester

[v]Pew Research CenterInternet, Science & Tech

[vi] Source: Pew Research CenterInternet, Science & Tech. http://www.pewinternet.org/2015/10/08/social-networking-usage-2005-2015/?beta=true&utm_expid=53098246-2.Lly4CFSVQG2lphsg-KopIg.1&utm_referrer=http%3A%2F%2Fwww.pewinternet.org%2Ffact-sheets%2Fsocial-networking-fact-sheet%2F

[vii]**Credit risk Monitoring parameters**

1. Credit files are neatly organized and cross-indexed;
2. Credit files should not be removed from the institution's premises without the requisite
3. approval;

4. Insurance policies are properly assigned to the institution and premiums are current;
5. Credit facilities are disbursed only after all contractual terms and conditions have been
6. met and all required documents received;
7. Security documents are duly executed and properly protected from fire, theft etc;
8. Collateral value is periodically ascertained and monitored;
9. The borrower is making timely repayments on interest, principal and any agreed to fees
10. and commissions;
11. Information on credit provided to the financial institution's management is accurate and
12. timely;
13. Credit administration responsibilities within the financial institution are adequately
14. segregated;
15. Funds disbursed under the credit agreement are used for the purpose for which they were
16. granted;
17. Established policies and procedures, as well as relevant laws and regulations, are being
18. complied with; and
19. Assessments of borrower's business are conducted through regular inspections.

Source: http://www.eccb-centralbank.org/pdf/credit_risk.pdf

Terminology

For the purposes of this proposal, the following descriptions of terms will apply.

'**Origination**' will be used to refer the part of the loan processing where the borrower applies for new loan and a lender process the application. This process generally includes all organization specific steps from applying to loan to disbursal of funds to the borrower. The primary step at this stage is to verify the capability of the borrower to repay the loan. At present CIBIL score is used as one of the tool to verify the capability of an individual to repay along with one's Salary certificate, tenure of one's job or the stream of income as applicable.

'**Capability**' will be used in the context of repayment of the mortgage. Assessment of borrower's stream of income at disposal to ensure that borrower can repay the mortgage beyond meeting his regular monthly expenses.

'**Willingness**' will be referred to depict the willingness of the borrower to repay the loan. Willingness may vary due to various reasons during the life of the loan. Those could be conflicting priorities, Social obligations, slow appreciation of property value or moving to out of current location, causing lose interest in the current property subjected to mortgage

'**Loan Servicing**' is the process by which a company (mortgage bank, servicing firm, etc.) collects interest, principal and escrow payments from a borrower. In India mostly this is done by the lender. The other option for loan servicing could be delegating the loan servicing to third party.

'**Credit risk**' is the Probability of loss from a debtor's default. In banking, credit risk is a major factor in determination of interest rate on a loan, Longer the term of loan, usually higher the interest rate. Credit risk is also called credit exposure. It is the process of identification, measurement, monitoring and control of its activities which would have an adverse impact on the money lent

'**CIBIL score**' Credit Information Bureau (India) Limited (CIBIL). This is India's first Credit Information Company. CIBIL collects and maintains records of an individual's payments pertaining to loans and credit cards. These records are submitted to CIBIL by member banks and credit institutions, on a monthly basis. This dataset is used with custom built analytics and formulas to arrive at credit risk score for an individual. This score plays a vital role to determine the sanction of loan. This plays as a credit risk tool during loan origination.

'**Loan origination**' is the process by which a borrower applies for a new loan, and a lender processes that application. Origination generally includes all the steps from taking a loan application up to disbursal of funds (or declining the application).

'**Loan servicing**' covers everything after disbursing the funds until the loan is fully paid off. Loan origination is a specialized version of new account opening for financial services organizations. Certain people and organizations specialize in loan origination. Mortgage brokers and other mortgage originator companies serve as a prominent example.

Sample Labels determining the sentiments

Itemized indicators	Sentiment type	Source
Family outing - very often	Negative	Facebook, Twitter, Instagram etc.
Addition of family members	Negative	Facebook, Twitter, Instagram etc.
Divorce	Negative	Facebook, Twitter, Instagram etc.
Marriage	Negative	Facebook, Twitter, Instagram etc.
Engagement	Negative	Facebook, Twitter, Instagram etc.
Family functions	Negative	Facebook, Twitter, Instagram etc.
newly found Love interest	Negative	Facebook, Twitter, Instagram etc.
Betting	Negative	Facebook, Twitter, Instagram etc.
Frequent to Gaming sites	Negative	Data from Gaming sites on Social Media
increased visit to parties	Negative	Facebook, Twitter, Instagram etc.
House price is lesser than purchase price	Negative	News, Economic updates
Low appreciation of house rate than interest rate	Negative	News, Economic updates
Real estate market is not doing well	Negative	News, Economic updates
Value of the subject, for which loan was obtained is lower due to market volatility	Negative	News, Economic updates
probable Natural calamity in the borrowers locality	Negative	Weather Portals, News
Transfer to other location within the country	Negative	LinkedIn & Other Social media
Transfer to other location outside the country	Negative	LinkedIn & Other Social media
Moving to a location which has cheaper cost of living	Positive	LinkedIn & Other Social media
probable Natural calamity in the borrowers locality	Negative	Weather Portals, News
Increased expense due to Natural calamity	Negative	Weather Portals, News
The segment in which the borrower is employed is facing downturn	Negative	News, Economic updates
Got a new job	Positive	LinkedIn & Other Social media
Job Loss	Negative	Job Portals, LinkedIn