Understanding the Qualitative and Quantitative Factors Affecting Business Directly and Indirectly

Miss. Madhavi K. Waykole
PG Student
School of Computer Sciences and Engineering,
Sandip University, Nashik (MH), India

Mr. Aakash Ahuja
Managing Partner
ITMTB Technologies

Mr. Umesh B. Pawar
Assistant Professor
School of Computer Sciences and Engineering,
Sandip University, Nashik (MH), India

***ABSTRACT*:** *The aim of this research is to find out the Qualitative and Quantitative factors affecting business directly and indirectly. The highlight of the major outcome is to find out the important factors from the Annual report of the company. The motive of research is helpful in saving a huge amount of time from doing unusual research behind massive pages of documents. The provided solution will give the complex hidden relationships between factors with respect to the multiple companies and multiple news. If relationship exists then prediction is to be particular company will be get affected by the particular news by analysing the different factors responsible for affecting the business. The report generates at the end of the solution will help the investors to decide in which company they should invest by considering the potential impact of news on company. The objective of my project is to develop an automated business solution for stakeholders like Business Analyst, Economist, and Chartered Accountant. This system will be used to generate early signals in real time about events that can qualitatively impact a company's future performance. Emphasize that the business factors are interconnected in a very complex way and it's not easy to understand the direct impact of these features in a precise way. Once this system is built, it will be able to discover these complex hidden relationships and assess them in real time, and hence become a vital tool for analysts to perform portfolio optimization proactively. The key features of our research are the specific use of POS-tagging, Word Distance Matrix and other NLP basic modules.*

Keywords: Factors affecting business, Impact of factor, Factors hidden relationship, Business automated analysis, Corporate reporting, Decision making, Keywords extraction, Natural Language Processing.

I. Introduction

When Fundamental analysis applied on business at that time it is significant to take into consideration both qualitative and quantitative factors. Fundamental analysis is important for any successful trading or investing strategy. All traders should have a basic understanding related to working of fundamental analysis. A trader uses Fundamental analysis for analysing factors in company. Fundamental Analysis method is to examine the security and to find the inherent value for long term investment opportunities. Investors use this method, so they can increase the value of their stocks in the long term. The decisions are based on the information available and the statistics evaluation with the help of financial statements, management processes, etc. These are some of the factors which may have an impact on the company's stock prices in the future needed past and present data for analysis. Future price of the stock decided on the basis of past and present performance and success of the company.

Following are the primary factors to consider when conducting fundamental analysis:

1. Is company making a profit?
2. What are the company turnover rates?
3. Does management take care of employees?
4. What is the company revenue?
5. Is company increasing indebtedness or paying off debt?
6. Is it growing?

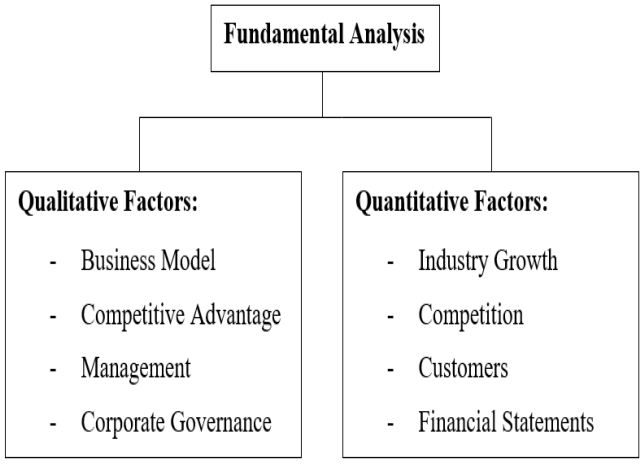

Figure 1: Fundamental Analysis Factors

Figure.1 shows that Fundamental Analysis can be proceeds using Qualitative Factors like Business Model, Competitive Advantage, Management, Corporate Governance and Quantitative factors like Industry Growth, Competition, Customers, Financial Statements [14]

Qualitative Factors: Qualitative factors are outcomes from certain actions that are difficult or impossible to measure in terms of numbers. Basic idea is that Qualitative Factors are consider based upon quality of the factor rather than considering quantity.

1. Business Model:
   Investors should have knowledge about "What exactly does the company do?" This question provides the information about company's business model.
2. Competitive Advantage:
   Competitive advantage is another business consideration for investors. Investor should be asked question like "How does the company differentiate itself from the competition?" Company might be differentiating itself from the competition based on product of that company or strategy of that company. Thus Competitive advantage of a company highly brings long-term success of the company.
3. Management:
   If management of the company is not upto mark then the company will not be able to perform well. Management of the company managed by the expert team of the company. For

investors, management is the most considerable aspect for investing in a company.

4. Corporate Governance:
   Corporate governance describes the policies or rules which are govern by the management of the company. Corporate Governance allows keeping a track of transparency between company and its stakeholders. The need of corporate governance it to undertake security in organization that can be helpful for avoiding illegal and unethical activities.

Quantitative Factors: Quantitative factors are outcomes from certain actions that are measurable in numbers. Quantitative Factors are analysed based on the quantity of factors rather than considering quality.

1. Industry Growth:
   Industry growth is an aspect of an economy which produces higher growth rate as compared to other sectors. Basically Industry Growth means to analyse the overall size of the industry. Industry Growth factor used to find out answers of questions like, "What is the expected revenue from this industry?" "What are the growth prospects?" "Is Industry growing at a faster rate in the future or growth of industry is going to slow down?"
2. Competition:
   Competition factor mostly focuses on to provide good services to the consumer at a lower price. These conditions allow the productive entity to generate more sales or superior margins compared to its market rivals. Competitions add up a lot of challenges in capital market. Investors should know about "How many competitors are in capital market?", "Total number of firms in industry?", "Level of competition?" Thus Competitive Advantage is the most important factor in quantitative analysis.
3. Customers:
   A customer is a facet of business environment that purchases the good services from the companies. Business environment is highly affected by the number of customers because customers are important factor which helps to drive revenues. Target of public-facing businesses is to attract customers by doing aggressive advertisement or lowering prices of their products.
4. Financial Statements.
   Financial Statements are conveying the idea about financial performance and business activities of a company. Financial Statements of a company are written by accountants, government agencies, and firms. Investors used these financial statements for analysing company and make some decisions or make some predictions about future investment. Balance sheet, cash flow statement and income statement are the three most important financial statement reports.

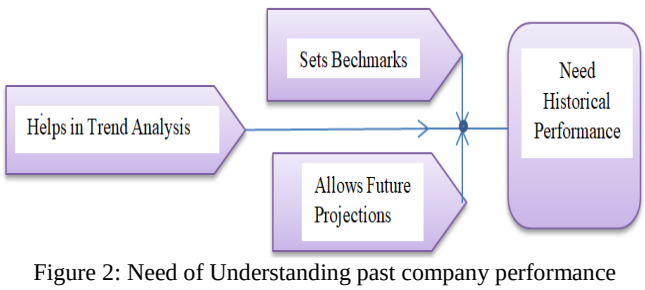

Figure 2: Need of Understanding past company performance

The historical performance frames the reason for understanding the current financial statement of an organization. Financial reports are providing basic overview about performance of company. In the event that they are seen considering comparative articulations from past periods, you can predict that some factors that influences the performance of company. On this premise, one can assess the current time frame's exhibition as well as remark upon the organization's future possibilities. Otherwise analysis of financial statement would not be worthwhile. Thus Investors analyse company by considering trend analysis, deciding some benchmark with respect to the company performance and making some future projections about company.

Figure 3: Important sections of Annual Report

Annual Report is a key which is used to communicate between company and stakeholders. Annual report holds information about the company's achievements, strengths, strategies, goals and objectives. Additionally, Annual Report consists of shareholders letter which is the great opportunity for the CEO and chairman of a company to communicate directly with the investors. Basically annual report is a snapshot of company's past year performance, opportunity to highlight a company's key achievements, and expectations for the coming year. Concerning design, we accept that HTML (appropriately improved for survey on telephones, tablets, and so on.) is favoured by financial specialists and other key partners, and it permits organizations the chance to expand the promoting reach of their informing through video and other intelligent web components and devices. Posting the yearly report as a PDF document additionally keeps on being an adequate configuration.

II. Related work

Research paper [1] depicts that Natural Language Processing is sophisticated technique which can be used for analysis of large-scale documents in corporate environment. Traditional users analyzed the documents manually where as sophisticated users apply NLP technique for analysis of the large-scale documents. In [2] Some Authors conducted research in management level to identify internal and external factors affecting Business Excellence. Researchers conclude that Internal factors like structure, systems, people, strategy, technology, results and External factors like customer forces, cultural forces, social forces, economic forces, partner forces affecting Business Excellence. The study [3] helps to improve economy by preparing financial statements of a company as an information base for taking decisions. The research article [4] explains the different types of preprocessing techniques and tools which are available in Natural Language Processing for handling textual data. Selection of best preprocessing techniques and tools is very important task according to the domain might help to make the text pre-processing easy and efficient.

III. Methodology

Textual analysis is very effective field in natural language processing (NLP). This analysis helpful for finding out the

relation between words. In this paper, Natural Language Processing is used for to find out the important keywords from the documents. A lot of documents need to be processed to find out key information. Keywords define the core information of the documents. Data analysis requires huge amount of pre-processing in order to identify relative information. At present it is almost impossible to keep track of dependency with each other between words. And process all the words in the documents to find out equal importance of the words. Moreover, detecting important words from the documents would be slow and time-consuming process. By extracting keywords from the documents would be a sufficient approach to keep track of important keywords from the documents. Therefore, Keyword Extraction is one of the most important technique in textual data analysis. The main task of important keyword extraction is to extract a specific set of words or keywords that can highlight the main content of the document. Relation between words needs to be captured carefully.

There are various kinds of approaches are available in NLP, including word frequency approach, word collocations and co-occurrences approach, TF-IDF (Term Frequency-Inverse Document Frequency) approach, and RAKE (Rapid Automatic Keyword Extraction) approach. These approaches do not need training data in case of extracting the important keywords from a text. Word Frequency shows the occurrences of each word in a text. Word Frequency determined by utilising Bag-Of-Words (BOG) Approach. Word Collocations and Co-occurrences Approach utilize the N-gram statistics which can be useful for understanding semantic structure in the text. The most commonly used collocations are bi-grams means two terms that appear adjacently (eg. 'customer service', 'video calls' or 'email notification') and tri-grams means a group of three words (eg. 'easy to use' or 'social media channels').

TF-IDF [Term Frequency–Inverse Document Frequency] approach estimates importance of word is to a document with respect to the collection of documents. This approach calculates the occurrences of a word within a text by comparing with its inverse document frequency. Multiplying Term Frequency with Inverse Document Frequency provides the TF-IDF score of a word in a document. This approach proves that if TF-IDF score is higher then the particular word is more relevant to the document. RAKE stands for Rapid Automatic Keyword Extraction algorithm. This algorithm used for extracting key phrases from text by analysing word frequency and its co-occurrences with the other words in text. In Linguistic and Graph-based Approaches, part-of-speech tagging helpful for generating the dependency tree structure and Graph based structure based on grammar representation within the sentences. All used approaches and many possibilities motivate us to work with this topic. We believe that getting better result of extracting keywords from documents is possible using various kinds of methods together and applying it in different fields where a lot of data processing is required.

The aim of this research is to find out the important keywords from the documents to help us save a huge amount of time from doing unusual research behind massive pages of documents. We became successful to achieve our aim and a completely different approach has been presented in the report. This research work can be further extended to identify the keywords which could be more accurate to detect the important factors from a document. POS-tagging might be excellent use of modules to extract important words from the documents. We selected the Nouns as a heart of the sentence but Verbs, Adjectives, adverbs and Business words also plays a big role. The Proposed prototype uses the Word Distance Matrix for finding out the important words from the document. Distance between two words is calculated based on the sequential occurrences of words in the document. Numeric Value represents the distance between two words.

Sentence: "Company will increase its budget for Marketing."In above sentence, "will", "its", "for" are the stop words. After applying preprocessing, stopwords are removed and will get important words list as "company", "increase", "budget", "marketing".

	company	increase	budget	marketing
company	-1	0	1	2
increase	0	-1	0	1
budget	1	0	-1	0
marketing	2	1	0	-1

Table 1: Words Distance Matrix

Figure 4 shows the working of the prototype. Solution of the prototype is applicable on the number of the companies which are the part of Capital Market. Each Company has its own Annual Report and this Annual Report is helpful for finding out the important factors from a particular company. After applying some preprocessing steps and NLP techniques on the annual reports of companies, 3D Word Distance Matrix is created which consists of the hidden relationship between all factors. Basically 3D Word Distance Matrix stores all complex hidden relationships between all companies. Simultaneously Multiple news are preprocessed using NLP techniques. If relationship exist between Companies and News then prediction is that particular company will be get affected by the particular news. Prototype also helps to find out the potential impact of particular news on particular company with the help of Word Distance Matrix.

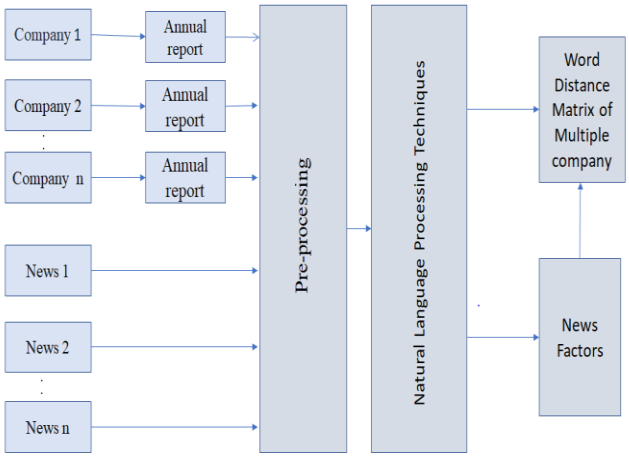

Figure 4: System Architecture

Text preprocessing:

Text preprocessing is an integral part of all text mining for all text documents which are processed under different research areas. Preprocessing removes all unwanted and unnecessary data from text. Poor and dirty text preprocessing lead to have a deteriorate effect on further document text processing operation. Natural Language processing (NLP) has greater contribution in the area of text preprocessing. Different preprocessing steps are require to be perform such as stemming, part-of-speech (POS) tagging, chunking, parsing, information extraction, etc. all these steps can be possible using Natural Language Processing. In NLP each preprocessing step uses different approaches, methods for text preprocessing and also each NLP task has different working

ways with different rules. So there is a need to review and evaluate such NLP task for different NLP steps, so that it will be easier for us to plan for subsequent text mining and text categorization steps in optimize and efficient way.

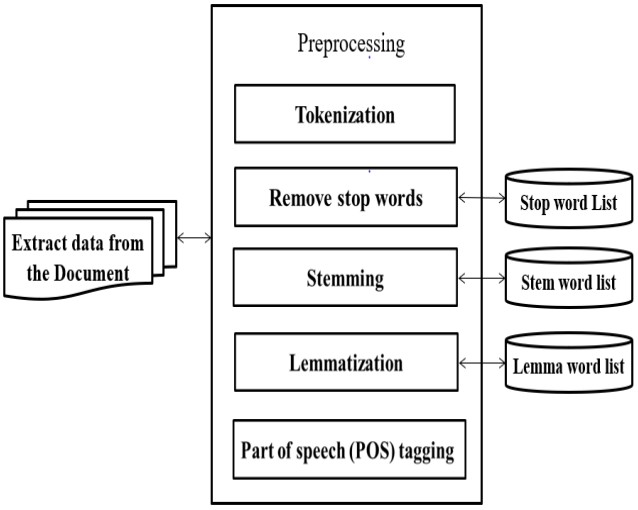

Figure 5: Text Preprocessing Techniques

Figure 5 shows the preprocessing techniques which are used in Natural Language processing for preprocessing of the text.

a. Tokenization:

Tokenization is a method which split text into meaningful tokens or elements. Tokens could be in the form of word, statement, and paragraph. In tokenization step of preprocessing, generated list of tokens is used in further preprocessing steps. Natural Language Toolkit (NLTK) is a text processing tool used in NLP for Tokenization, Parsing, Classification, Stemming, Part-of-speech tagging and semantic reasoning. Most of the Natural Language processing tools uses white spaces as delimiter for splitting tokens in text.

Sentence: “Mr John is a successful businessman.”

Tokenized words: “Mr”, “John”, “is”, “a”, “successful” “businessman”, “.”

b. Stop word Filtering:

Stop words are low informative words in text. These common words are frequently used in text. Example of stop words are “a”, “an”, “the”, “is”, “are”. Main intuition behind removing stop word is to remove unuseful data from text and get valuable, important, informative words from text. In above sentence “is” and “a” are stop words. These words are removed in preprocessing steps.

c. Stemming:

Stemming is the way toward distinguishing a root word that structures into a different representation. For Example, "connected", "connection", "connections", “connects” could all be recognized by the root word or stem word "connect". Difficulties of stemming rely upon the way toward controlling the mistakes. Stemming uses Rule-Based Approach and it removes suffixes like “ing”, “ly”, “s” etc. Thus, in some cases stemming stem the word without understanding the context of the word. Example: “trouble”, “troubled”, “troubles” has stem word is “troubl”.

d. Lemmatization:

Lemmatization is a process of reducing words into their lemma or dictionary. Lemmatization uses Dictionary-Based approach and controls the difficulties occur in stemming process by considering meaning of word. For Example: “trouble”, “troubled”, “troubles” has lemmatized word is “trouble”. Difference between Stemming and Lemmatization is that stemming is the faster process than lemmatization but lemmatization gives more accurate result in case of finding out root word of words because it uses Dictionary-Based Approach.

e. Parts-of-Speech Tagging:

POS improves the “word” and its “context” with large volume of information about itself and its neighbours. The use of a word in a sentence, for example: noun, pronoun, verb, adverb, adjective, article, preposition, and conjunction assists with inducing conceivable information about neighbouring words and syntactic structure weaving around the word. In this way POS Tagging turns into an indivisible part in syntactic parsing.

IV. Result and Analysis

Word Distance Matrix is useful for finding out the complex hidden relationship between factors. If one factor has relation with another factors then these relationships can be find out from word distance matrix with its potential value. Basically Potential Value is a Distance Between two factors. Prototype helpful for finding out exists hidden relationship between factors and also useful for finding out potential impact of particular factor for a particular company. Table 2 Shows the hidden relationship and its potential impact of factors for Reliance company. If someone wants to do searching for word as a “digital” factor then all useful complex hidden relationships with “digital” factor is came out from word distance matrix of Reliance company. Using this scenario investors can be conclude that digital environment of Reliance company might be impacted by considering another factors like “india”, “jios”, “data”, “services”, “mobile”, “wireless”, “subscribers”, “network” etc.

wd	swd	val	company
india	digital	81	Reliance
data	digital	42	Reliance
services	digital	40	Reliance
subscribers	digital	36	Reliance
mobile	subscribers	30	Reliance
mobile	data	30	Reliance
mobile	jios	27	Reliance
mobile	life	27	Reliance
jios	digital	24	Reliance
mobile	services	24	Reliance
mobile	digital	24	Reliance
mobile	embrace	21	Reliance
mobile	evident	21	Reliance
mobile	volume	18	Reliance
mobile	india	18	Reliance
mobile	wireless	15	Reliance
mobile	efficiencies	15	Reliance
improvement	digital	14	Reliance
mobile	network	12	Reliance
mobile	improvement	12	Reliance

Table 2: Hidden Relationship between Factors

The main aim of the prototype is to find out which company will be get affected by particular news. If prototype have the complex hidden relationships of factors with respect to the multiple companies then prototype also able to find out which news has

potential impact on which company by finding relationship between multiple companies and multiple news.

news name: There is a lot of pent-up consumer demand for digital services and India is at an inflection point

	wd	swd	val	company
reliance	reliance	retail	210	Reliance
jio	jio	retail	140	Reliance
retail	retail	reliance	110	Reliance
retail	retail	jio	110	Reliance
retail	retail	terms	99	Reliance
retail	retail	consolidated	99	Reliance
retail	retail	phenomenal	88	Reliance
retail	retail	economic	88	Reliance
retail	retail	categories	77	Reliance
retail	retail	healthy	77	Reliance
retail	retail	impacted	66	Reliance
retail	retail	capacities	66	Reliance
retail	retail	european	55	Reliance
retail	retail	incremental	55	Reliance
retail	retail	oil	44	Reliance
retail	retail	due	44	Reliance

Table 3: Factors Hidden Relationship between Company and News

Above result shows that Reliance company will be get affected by the News. Reliance company has more contribution in digital environment and News also talk about the digital services. Thus In between Reliance company and News some complex hidden relationship is exist with respect to the multiple factors. By using such kind of experimental analysis investors can be easily find out which company will be get affected by particular news and it will be beneficial for investors for taking decisions about investment in terms of multiple companies. Thus prototype will be used to generate early signals in real time about events for investors that can be useful for finding out future performance of companies.

V. Conclusion

There is a bunch of contributing factors which affects the success of the company. If a business hopes to perform smoothly and successfully, they need to take all these elements into consideration before making any decision. Investors spend much of their time to anticipate the next news cycle for investing shares in capital market. Prototype helps to decide which companies are negatively impacted by news and how much it will be impacted. The solution is automated which means there is no manual intervention of Analysts to find out the qualitative and quantitative factors affecting business directly and indirectly. The report generates at the end of the solution will help the investors to decide in which company they should invest. The stakeholders like Analyst, Economist, and Chartered Accountant etc. can use such analysis for decision making and do their investment in a more systematic way. Experimental results demonstrated that the prototype is working significantly well with almost 72% to 75% of perfection. In future, we will consider TF-IDF approach and also working on verb, adjective, adverb and business words along with the present approach to improve result for finding out the most important keywords from the document.

REFERENCES

[1] Lewis, Craig, and Steven Young. "Fad or future? Automated analysis of financial text and its implications for corporate reporting." Accounting and Business Research 49.5 (2019): 587-615.

[2] Vartiak, Lukáš. "Identification Of Internal And External Factors Affecting Business Excellence." (2016)

[3] Osadchy, E. A., et al. "Financial statements of a company as an information base for decision-making in a transforming economy." (2018).

[4] Lourdusamy, Ravi, and Stanislaus Abraham. "A Survey on Text Pre-processing Techniques and Tools." International Journal of Computer Sciences and Engineering Open Access Survey Paper 6 (2018).

[5] Hasan, HM Mahedi, et al. "An empirical study of important keyword extraction techniques from documents." 2017 1st International Conference on Intelligent Systems and Information Management (ICISIM). IEEE, 2017.

[6] Dudhabaware, R. S., & Madankar, M. S. (2014, December). Review on natural language processing tasks for text documents. In 2014 IEEE International Conference on Computational Intelligence and Computing Research (pp. 1-5). IEEE.

[7] Delen, D., Moscato, G., & Toma, I. L. (2018, January). The impact of real-time business intelligence and advanced analytics on the behaviour of business decision makers. In 2018 International Conference on Information Management and Processing (ICIMP) (pp. 49-53). IEEE.

[8] Srinivasan, Padmini, and Ana Cristina Marques. "Narrative analysis of annual reports: A study of communication efficiency." (2017).

[9] Janeska-Iliev, Aleksandra, and Stojan Debarliev. "Factors Affecting Growth Of Small Business: The Case Of A Developing Country Having Experienced Transition." European Scientific Journal 11.28 (2015).

[10] Papagiannopoulou, Eirini, and Grigorios Tsoumakas. "A Review of Keyphrase Extraction." arXiv preprint arXiv:1905.05044 (2019).

[11] Hasan, HM Mahedi, Falguni Sanyal, and Dipankar Chaki. "A Novel Approach to Extract Important Keywords from Documents Applying Latent Semantic Analysis." 2018 10th International Conference on Knowledge and Smart Technology (KST). IEEE.

[12] Tuape, Micheal, and Yirsaw Ayalew. "Factors affecting development process in small software companies." 2019 IEEE/ACM Symposium on Software Engineering in Africa (SEiA). IEEE, 2019.

[13] Niboonkit, Supattra, Worarat Krathu, and Praisan Padungweang. "Automatic discovering success factor relationship entities in articles using named entity recognition." 2017 9th International Conference on Knowledge and Smart Technology (KST). IEEE, 2017.

[14] Madhavi K. Waykole, Aakash Ahuja, Umesh B. Pawar, "A Survey On: Understand The Factors Affecting Business Using Natural Language Processing", IJRAR - International Journal of Research and Analytical Reviews (IJRAR), E-ISSN 2348-1269, P-ISSN 2349-5138, Volume.6, Issue 4, Page No pp.981-983, November-2019