Knowledge-Based Text Mining in Getting Perfect Preferences in Job Finding

Shaziya Islam and Manpreet Kaur

Abstract Although we know that finding the most suitable job using the Internet takes various hours because the job portal gives us the preferences of jobs based on some particular keyword stored in their database but it may not be the preferences you want, so in order to remove wrong relevancy of job preferences and to be appropriate, we have the concept of text mining with knowledge-based in order to filter out most suitable preferences based on our searching criteria, so the problem is to give 100% accuracy using knowledge-based text mining using some technique, so in order to give accurate means 100% results, we have used concept of knowledge-based text mining using R studio by which we get only the job preferences which we want according to our criteria, hence this paper gives 100% accuracy in finding that.

Keywords Text mining · TM · Knowledge-based test mining

1 Introduction

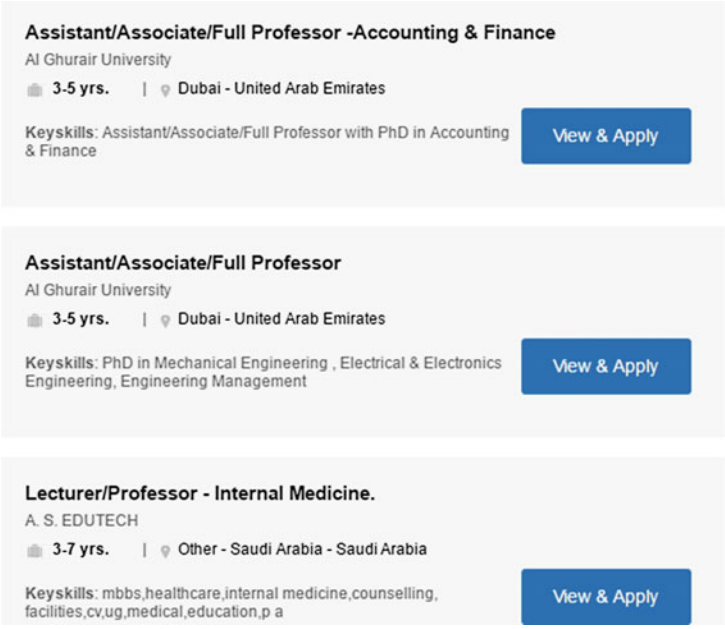
"No one is satisfied with their job in today's scenario" this quotation is based on the survey done on various people working in various organization this means that in every organization, employees want to change their job for betterment of their position, salary, working environment, or the location so each and everyone is registered in some job portals for getting good opportunities and every morning they open their mail for finding some, and they get the mail "10 jobs matching your profile and when they open the mail 9 out of 10 jobs are not the preferences of jobs which they want and 1 out of 10 jobs are the actual job preferences Which we want or to which we can look up for…". And due to this we are wasting our Time in

S. Islam (✉) · M. Kaur
Bhilai, India
e-mail: Shaziya.islam26@gmail.com

M. Kaur
e-mail: Preet.kaur963@gmail.com

P. K. Sa et al. (eds.), *Recent Findings in Intelligent Computing Techniques*,
Advances in Intelligent Systems and Computing 709,
https://doi.org/10.1007/978-981-10-8633-5_5

reading all the preferences and also we are wasting the memory of our mail account like for example, Someone registered on shine job portal for the job of assistant professor in computer science and engineering department and the job preferences are given in below image.



None of the job preferences matches his/her criteria because this suggestion is based on your current position, or it can be based on your profession it is not filtering or refining the appropriate criteria which we want, sometimes it suggests us the job lower than the salary which we are getting presently so that is also not good, hence in order to overcome this problem, there is concept "TEXT MINING" and particularly knowledge-based text mining and in order to proceed with let us see what is text mining and what is knowledge-based text mining.

2 Text Mining

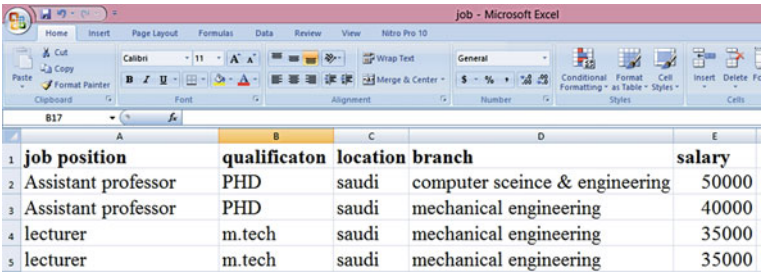
Text mining is also known as t*ext data mining*, which is equivalent to analysis of text in order to have some good or valuable text from the bulk of text, the process of deriving high-quality information from text or extracting the information according to our need [2]. Quality information is typically derived through the patterns such as statistical pattern learning [2]. Text mining usually involves the process of structuring the input text (usually parsing, along with the addition of some derived

linguistic features and the removal of others, and subsequent insertion into a database), deriving patterns within the structured data, and finally evaluation and interpretation of the output. "High quality" in text mining usually refers to some combination of relevance, novelty, and interestingness. Typical text mining tasks include text categorization, text clustering, concept/entity extraction, production of granular taxonomies, sentiment analysis, document summarization, and entity relation modeling (i.e., learning relations between named entities).

Text mining also referred as the text analytics or data text mining helps us in finding some particular words in some particular text or words based on the structure of some particular data, highest quality data can be found from bags of words or big data or from any ordinary document generally text mining includes text categorization, text clustering, and text concept [1–3].

2.1 Knowledge-Based Text Mining

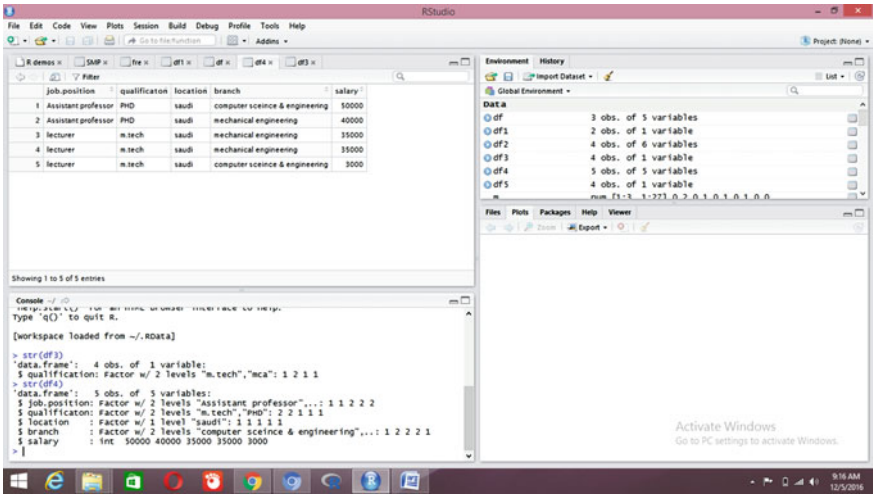
Text mining is the actual technique of refinement of text from bulk of text, but in deep knowledge of text mining, there are types of text mining applied for different criteria for practical examples and in that, there are various techniques and among them there is knowledge-based text mining too and to define this technique, we can divide the word into two words, i.e., **knowledge + based** which means it is dealing with the concept of knowledge so there is some type of requirement of knowledge before processing means if we are dealing with some data online or offline then before processing that data to for getting useful result, we have to go through some knowledge of that data, and consider the data frame having the information about the job finder in below image.



	A	B	C	D	E
1	**job position**	**qualificaton**	**location**	**branch**	**salary**
2	Assistant professor	PHD	saudi	computer sceince & engineering	50000
3	Assistant professor	PHD	saudi	mechanical engineering	40000
4	lecturer	m.tech	saudi	mechanical engineering	35000
5	lecturer	m.tech	saudi	mechanical engineering	35000

So this is the data frame which we have in excel sheet and saved with .csv extension, .csv stands for comma separated values because in computing theory comma separated values is a text having numbers and text separated by comma in plain text so if we can use the csv file, it will be better for our project to mine the text in a better way.

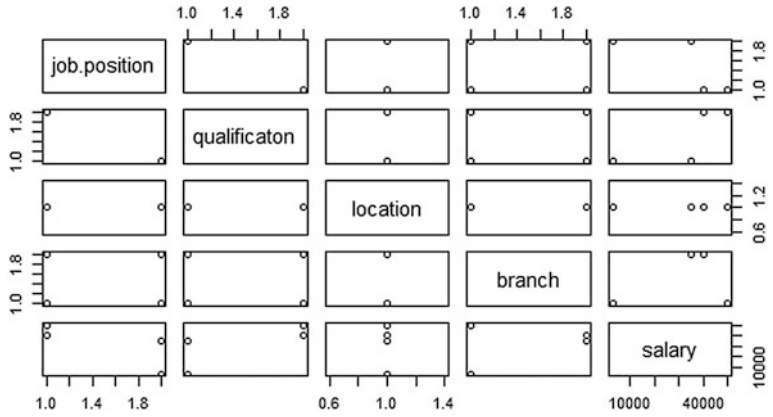
Hence in order to understand the data we have to first load it, in R studio which we are using in mining text in a better way suppose for the above data frame we have to find the details of the data hence we can find it as:



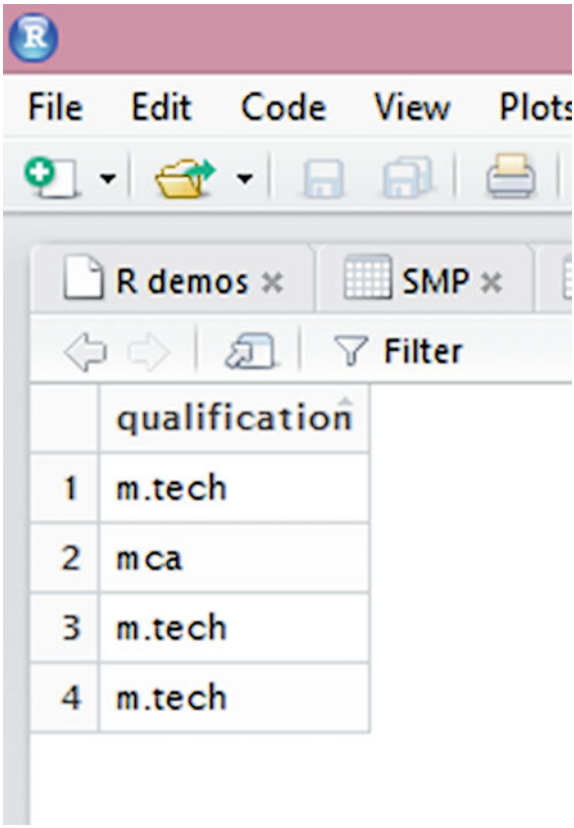
As we can see in the above figure that all the observations made by the **str()** function in order to enquire about the text is made. Hence in order to proceed further, we should find out some knowledge of the data, as this is the small data so it is not required but in the large database, this task can be completed within seconds. We can also use the plot function in R to get the data frame information graphically which can help us to understand the data frame or data set more accurately and for large amount of data. We can use the plot for getting the information about the data frame as easily as writing:

plot(data frame name), like for example in my case the data frame name is df4 so I have written:

plot(df4)and the result is

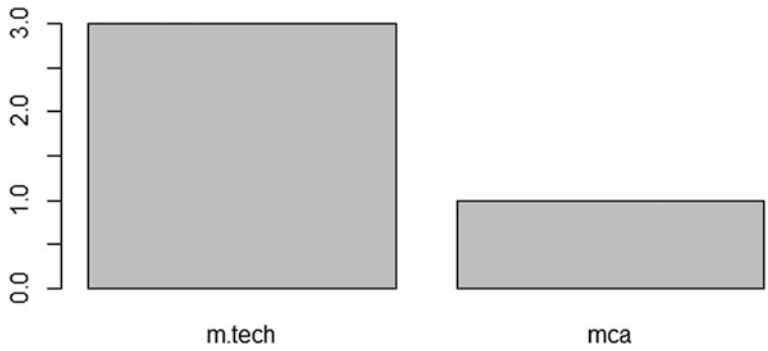


Also, I have another data set like only for educational qualification needed by the companies in my case its educational institutions hence it is data frame as **df3** and the information in that data frame is like:



As seen above in df3, there is only one column and that is for qualification this example is set for the recruiter who only wants the searching criteria to be based on the qualification so in order to inspect this type of data in large quantity, we can again use plot function as:

Plot(df3)-> And the graph is shown as:



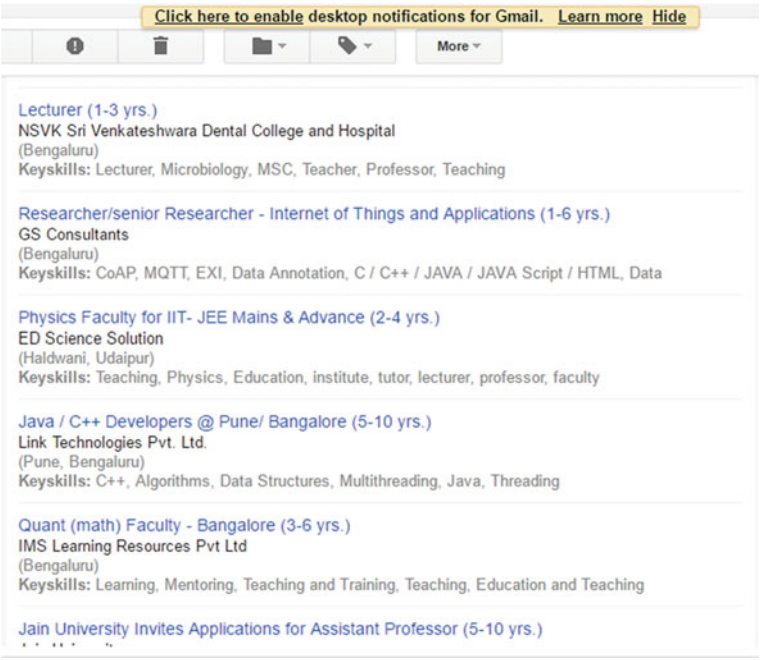
It is clear from the picture that there is more requirement of people having master's degree in technology rather than the people having master's degree in computer application so this will be the great advantage of differentiating the people in terms of their qualification.

2.2 R Tool

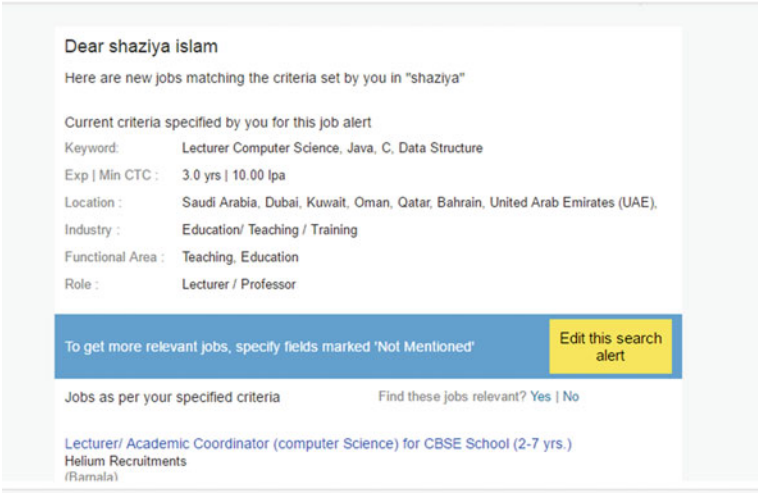
Tool which we have used in mining the text for getting the most appropriate result is the R tool which gives us various functionalities and functions for inspecting and accessing the text, so we have used the R studio for doing all the tasks here taking two data frames namely df2, df4 having the required data and manipulated the data frames using various functions available in R studio.

3 Problem Identification

As said before that no one is satisfied with their job hence 95% people are willing to change their job and that is why they register in some job portal to at least get the job preferences according to the job they want which is based on their qualification, location, and some specific keyword but this is not fulfilled at all, people start their work by opening their mails to have a look for the preferences and they got about 10 jobs preferences or 20 job preferences like shown below:



As shown above, this is the job preferences got for any particular registration in some of the job portal and the criterion was to have the job in Gulf countries in academics for the post of lecturer in computer science and engineering department but as shown in the picture, these are all irrelevant jobs preferences got against the job search criteria, which will not be seen further this is one situation now coming to another situation like shown below:



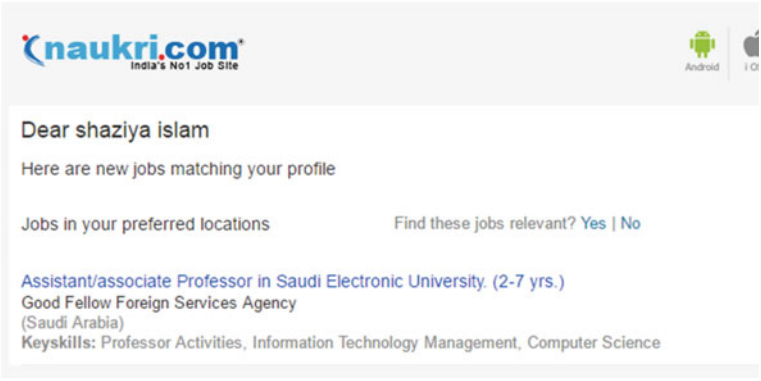
Now see the problem shown in all these screenshots taken on my laptops for the criteria, searching job as shown the criteria is specified clearly in the first image that is keyword: lecturer—computer science, java, c, c++.

Preferences:

Location—Saudi Arabia, Dubai, Oman, Kuwait, and UAE basically all Gulf countries.
Industry—education/teaching
Role—lecturer

But as seen in another image, the preferences are not up to the mark and they give the preferences for lecturer in math's, physics, and chemistry so this is of no use, so this is one of the situations that people get irritated of getting wrong preferences, so what is the use of getting registered on these portals.

Now the most irritating situation shown below:



In above picture, the job search criteria are based on the preferred location, and the criteria but when going through this link the specifications are as follows:

Branch—computer science and engineering
Post—Assistant professor
Qualification—PHD

Now after having a good watch in this criteria again, the eligibility criteria are not fulfilled because the criteria of education were set to master's degree in computer science and engineering and the result is for PhD, so now this is again the most irritating situation because time has been given for registering to the portal and this is the result which is not at all matching the specified criteria, this is only one case if there are 10 cases like this then this is waste of time in accessing these links and knowing about the job, so there should be some criteria for this so that if getting some of the preferences then it should be valid and based on specific criteria only, there should be no other preferences given other than the specific criteria.

4 Solution to the Problem

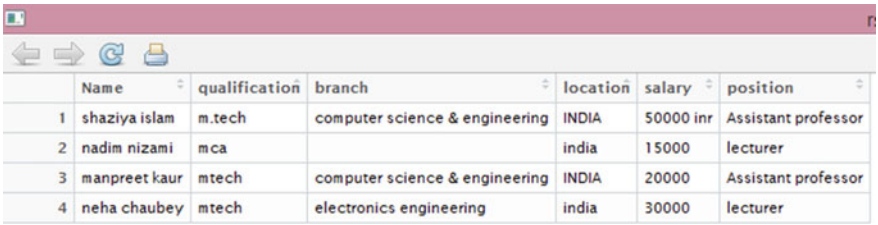
If we use knowledge-based text mining in finding appropriate job preferences, then we can get 100% result of getting only specific criteria-related job preferences this will be made on large amount of data set but for showing the result I have made two data frames:

df2-> data frame having the information of the candidates registered on some of the job portals and it can be loaded to R studio as:

Reading jobdir.csv (discussed above)
df2 <-read.csv("C:/Users/admin/Documents/texts/jobdir.csv")

And now to see whether the data frame is loaded successfully or not, we have to write the data frame name and press enter.

df2



	Name	qualification	branch	location	salary	position
1	shaziya islam	m.tech	computer science & engineering	INDIA	50000 inr	Assistant professor
2	nadim nizami	mca		india	15000	lecturer
3	manpreet kaur	mtech	computer science & engineering	INDIA	20000	Assistant professor
4	neha chaubey	mtech	electronics engineering	india	30000	lecturer

As shown in the above image, there are four candidates with the information—qualification, branch, location, position, and expected salary which are the basic information needed for registering on some job portal.

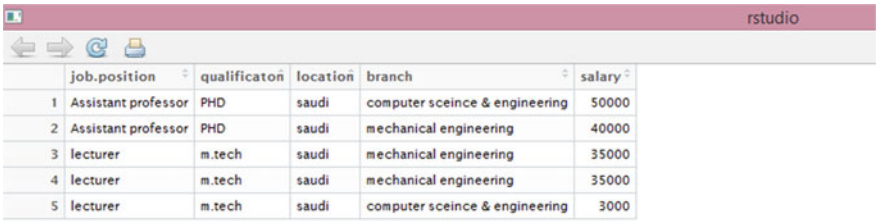
Now coming on the jobs directory coming from some employer to some job portal say:

df4-the data frame having the jobs for the candidates it can be loaded to the R studio as:

df4 <-read.csv ("C:/Users/admin/Documents/texts/job.csv")

And now to see whether the data frame is loaded successfully or not, we have to write the data frame name and press enter.

df4

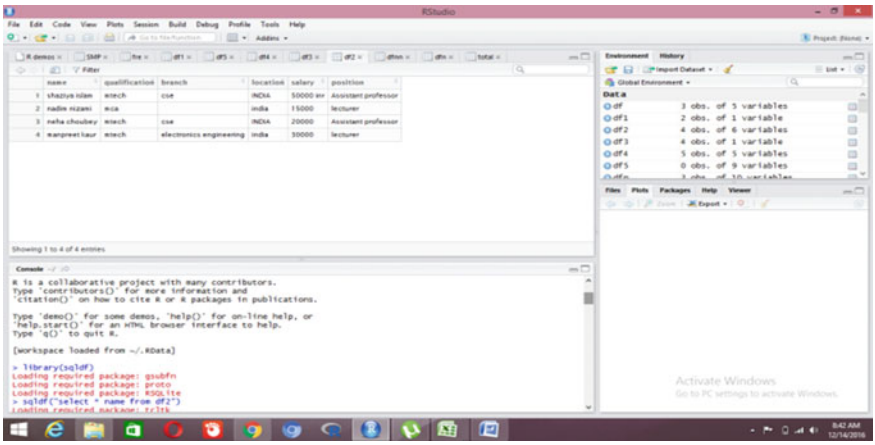


	job.position	qualificaton	location	branch	salary
1	Assistant professor	PHD	saudi	computer sceince & engineering	50000
2	Assistant professor	PHD	saudi	mechanical engineering	40000
3	lecturer	m.tech	saudi	mechanical engineering	35000
4	lecturer	m.tech	saudi	mechanical engineering	35000
5	lecturer	m.tech	saudi	computer sceince & engineering	3000

So here is the jobs directory data frame loaded into df4 having fields job position, qualification, location, branch, and salary so now we want that according to the qualification, branch, and location I should get the accurate result for the position otherwise NA for not applicable position so in R studio we can do it in seconds using the simple sql queries available in R studio but for using these functions, we have to load the sqldf package available in R studio as:

library (sqldf)

And it will be loaded as shown below:



And after loading sqldf package, we can use various queries available in sql to be used in data frames.

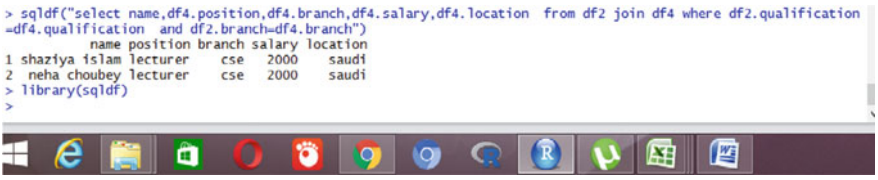
So with our problem as said before we have two data frames df2 and df4, df2 is containing the information of the candidate registered in job portal and df4 is containing the information from the employer regarding the job posting so the query will be

sqldf("select name, df4.position, df4.branch, df4.salary, df4.location from df2 join df4 where df2.qualification = df4.qualification and df2.branch = df4.branch")

and the result will be as shown below:

Name	Position	Branch	Salary	Location
1 Shaziya Islam	Lecturer	Cse	2000	Saudi
2 Neha Choubey	Lecturer	Cse	2000	Saudi

As shown in the screenshot taken from R studio working in windows



As shown in the image, only two names are selected from the two data frames which are exactly suitable for the job position asked by the recruiter hence this removes all other job options not suitable according to our criteria hence this gives the 100% accurate result and this method can be used in bigger data frames available in job sites so that efficient result can be produced which is suitable for both the job searcher and employer.

5 Conclusion

As we have seen that using knowledge-based text mining, we can do various tasks and one of them is getting appropriate jobs preferences from some job portal and we can apply this to various other categories for mining some of the text from which we can get some useful information, this technique can also be used in instigating the comments or tweets of people available in the library of R studio, in short, we can say that all the things are available for our accessing but using knowledge-based text mining in R tool we get accurate results.

References

1. https://www.gettinggeneticsdone.com/2010/05/use-sql-queries-to-manipulate-data.html
2. Shehata, S., Karray, F., Kamel, M.: A concept-based model for enhancing text categorization. In: Proceedings of the 13th International Conference on Knowledge Discovery and Data Mining (KDD '07), pp. 629–637 (2007)
3. Gaikwad, S.V., Chaugule, A., Patil, P.: Text mining methods and techniques. Int. J. Comput. Appl. **85**(17), 0975–8887 (2014)
4. Wu, S.-T., Li, Y., Xu, Y.: Deploying approaches for pattern refinement in text mining. In: Proceedings of the IEEE Sixth International Conference on Data Mining (ICDM '06), pp. 1157–1161 (2006)
5. Mooney, R.J., Nahm, U.Y.: Text mining withi nformation extraction. In: Daelemans, W., du Plessis, T., Snyman, C., Teck, L. (eds.) Proceedings of the 4th International MIDP Colloquium, Sept 2003, Bloemfontein, South Africa, pp. 141–160. Van Schaik Publishers, South Africa (2005)
6. Rathor, A.S., Garg, D.P.: Analysis on text mining techniques. Int. J. Adv. Res. Comput. Sci. Softw. Eng. ISSN 2277 128X
7. Shinde, M.R., Gill, P.C.: Pattern discovery techniques for the text mining and its applications. Int. J. Sci. Res. (IJSR) ISSN (Online) 2319-7064 Impact Factor (2012), 3.358 **3**(5) (2014)
8. https://www.r-bloggers.com/manipulating-data-frames-using-sqldf-a-brief-overview/
9. https://searchbusinessanalytics.techtarget.com/definition/text-mining