

# Predictive Analysis for Risk Reduction in Data Mining

Ridhi Thakur\*<sup>1</sup>

<sup>1</sup>Information Technology, Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India

## ABSTRACT

In this age of artificial Intelligence and machine learning, Business Intelligence is also gaining more significance as the organizations worldwide want to build intelligence into their business processes so they can better understand the customer behavior or patterns and provide insights for business leaders to make the right decisions in the market place to keep them competitive and efficient by reducing the risk optimizing operations or Fighting fraud to the extent possible. Business Intelligence in its crude form always existed before the IT in terms of experience and business expertise with the employees handling certain business process over decades but this process does not guarantee all the factors have been accounted for and no way to prove their analysis out before any decisions to be made by the organization. The Business Intelligence is data driven and has a scientific process behind it to analyze the data and provide models to test the What-If scenarios so Businesses can make less risk prone decisions. That said we cannot make it 100% reliable but it is way far better than making a guess out of one person's perspective. This paper aims to explore Data Mining and Predictive Analysis in the context of business applications and the techniques involved which eventually build the intelligence needed in the Business Intelligence.

**Keywords :** Data Mining, Predictive Analytics, Business Intelligence, Risks, Decisions

## I. INTRODUCTION

In the world of global markets as the organizations are gearing up to cater to customers worldwide there is enormous of data that is being gathered regarding their customers. There is huge amount of data that is being gathered which could be valuable not only for commercial purposes but also for research. For example, hospitals trying to mine the data and apply Predictive Analysis (PA) on the patients will help them identify the patterns and enable the health professionals to make decisions very early and avoid life threatening conditions. On the other hand, the same business analytics when applied to a financial sector or a security organization can help identify patterns of individuals to prevent fraud. This paper

focuses on the mechanics behind the Business Intelligence or BI as we call it. The two building blocks to the Business Intelligence (BI) are Data Mining (DM) and then Predictive Analysis (PA) "Data processing + Domain knowledge => Predictive Analytics = Business Value" [8]. This is a vast subject to get into and something in its organic stage and has a potential to expand and enter into every industry as we speak because decision making is critical to any field and BI is enabling you to take precautionary steps to avoid risk but also improve the operational efficiency. We are trying to look at the basics and various models or techniques involved with Data Mining (DM), Predictive Analysis (PA) and Business Intelligence (BI) and how this all can be or is being applied in various industries.

As research states that almost a billion devices interconnected to the Internet would co-exist by the year 2020, organizations also believe that the sources of information are bound to increase. [11].

Risk of exposure is one of those surface areas that has key beneficiary of advances in predictive analytics (PA) due to its volatile nature to detect and predict vulnerabilities, occurrence of fraud, security breaches and also the excellence of management systems and governance. It is also one area where companies have high intentions, over the next several years, to invest in the technology and increase their usage.” Risks associated with an organization’s workflow are innumerable and inevitable at times. These risks are manageable if businesses have previous data concerning the chance of a disaster. This information helps companies in decision making. Predictive analytics is a process that aids an organization in deciding on adequate precautionary actions to prevent or minimize the losses incurred. Predictive Analysis (PA) is transforming risk management as it helps organizations by informing what is arriving in the future. The goal for data processing here is to make decision support systems that may accurately predict if it is a profitable operation for an organization or not. It is safe to mention that business ought to concentrate on their information if they need to develop, flourish, and succeed. They ought to collect, store, manage, and analyze the information within the most effective ways in to assist the business to run expeditiously and effectively. Skilled analyst ought to be even more involved among the platform; not only they’ll be handy for the corporate, however they’ll conjointly give solid risk management system that’s safe and reliable.

In the below sections the Data Mining (DM), predictive analytics, Business Intelligence (BI) along with their associations, applications and risk management techniques are explained sequentially.

## II. DATA MINING

Data Mining (DM) is an integrated application in the Data Warehouse (DW) and describes a systematic process for pattern recognition sequentially in large data sets of structured and unstructured data to identify and draw conclusions and find the relationships between them using applied math strategies, or genetic algorithms, data files are often searched for statistical anomalies, patterns or rules outlined as “Data Mining (DM) is associate a degree knowledge base subfield of applied science. It’s the statistical method of researching and discovering sequential patterns in enormous data sets involving ways at the intersection of applied computational science, Machine Learning (ML), statistics, and knowledge systems. The goal of the Data Mining (DM) method is to extract information from a data set and remodel it into a visible and vital structure for more use.”

### A. Practical Applications with Data Mining

The real-time applications of Data Mining (DM) are:

- Computerized prediction of future trends and the behaviours
- Computerized detection of unknown models

### B. Data Mining Techniques:

Data Mining (DM) techniques [1] are used in many investigating areas, involving statistics, arithmetic, cybernetics, genetics and marketing. While these Data Mining (DM) techniques are a way to derive efficiencies analyse and predict client behaviour, if it’s used properly, a business will set itself up except for its competition through the application of Predictive Analysis (PA). It is extremely efficient and effective, till now because it attracts these techniques:

- 1) Tracking Patterns: one among the most basic method in Data Mining (DM) is learning to

understand patterns in the given information set. For an instance, consider the sales of a particular product and the difference of price that will appear simply before the vacations, or notice that hotter weather drives more viewers to your web site.

- 2) Classification: It is a very is a complicated Data Mining (DM) technique that forces you to gather various attributes along into visible classes, that you can then use to draw out the future conclusions, or perform some action. For an instance, if you're evaluating data an individual customers' money backgrounds and get his transaction histories, you may be ready to organize them as low to high credit risks. You may then use these classifications to find out even supplementary information regarding those customers.
- 3) Association: This method is related to following the patterns sequentially, however it is extra specific to dependently joined variables. This is usually what is used to publicize the 'people also viewed', 'people also bought' sections of online shopping stores.
- 4) Outlier detection: In a lot of cases, simply recognizing the analyzed pattern can't give you a clear and complete understanding of your data set. You also need to be able to identify changes, or oddity in your analyzed information.
- 5) Clustering: It's like a classification; however, it involves grouping chunks of information according to their similarities. For instance, you would possibly choose to cluster completely different demographics of your audience into dissimilar packages supported to the income they spend, or how usual they are inclined to buy at your store.
- 6) Regression: It is a method one of the methods in data mining (DM)that is used primarily as a style of coming up with and modelling, is employed to spot the chance of a particular variable, given the presence of different variables. For an instance, if

you use it to project a particular value, based on different factors like convenience, shopper demand, and competition. Its main focus is to find a precise relationship between 2 (or more) variables in the particular data.

- 7) Prediction: This is one of the important and valuable data processing techniques, as it can be used the varieties of data you will see in the future. In several cases, simply recognizing and understanding previous inclinations is enough to conclude a somewhat correct prediction of what is going to happen within the future.

### Steps involved in the Data Mining process

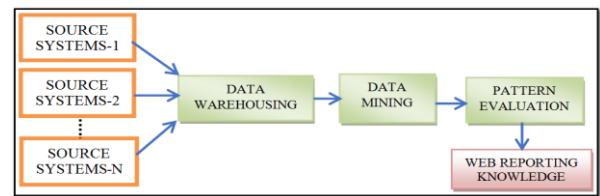


Figure 1. We are discussing the the process of Data Mining process in brief.

### C. Pros and Cons of Datamining

Following are the benefits and limitations [6] provided by Data Mining

TABLE I. PROS AND CONS OF DATA MINING [6]

	PROS	CONS
1	Predict future trends, customer purchase habits	Privacy for security [7]
2	Help with decision making	Amount of data is overwhelming
3	Improve company revenue and lower costs	Great cost at implementation stage
4	Fraud detection	Possible misuse of data

### III.PREDICTIVE ANALYTICS

Predictive analytics(PA) is that the use of information(data sets), mathematical and statistical algorithms and machine learning(ML) to analyse the future events based on the historical data.The most important goal of predictive analytics(PA) is to use the information(data) of what happened to give the best and simplest valuation of what is going to happen.Also,predictive analytics provides an entire knowlege of what is happening and also the information we would want[13].

Due to the to the circulation of text analytics that have created the analysis of unstructured knowledge(data) time saving, Predictive Analysis(PA) is increasing.

Today, we tend to want progressively wanting machines that may take past(historic) and current (present) information to predictict future trends, like sales trends for the approaching months or years, or anticipating client behavior like within the case of deceitful creditcard use.

Predictive Analytics(PA) is that the collection of enormous, partially unstructured knowledge(data) sets from totally different sources. The mixture of various knowledge sources like geographical conditions, traffic and social surfing data, supplemented by internal knowledge is specially important.

Predictive Analytics(PA) processes this knowledge(data sets) using different applied mathematical strategies within the dissimilar data patterns to derive the algorithms. This sequential data is reviewed on the basis of a test data and are then optimized.Additionally note that the increase in the data available will lead to more accurate and developed algorithms. If the optimisation method is finished, the algorithmic rule and also the model may

be practical to knowledge whose classification is unknown.

#### A. How Does Predictive Analytics(PA) Work:

Predictive Analysis(PA) uses varied data models to allocate the data. Using sample data with renowned attributes, the design is trained and is in a position to identify and analyse the new data and also verify its behavior. This information may be useful to predict how the client would possibly behave next. With descriptive models,client data is classified by characteristics and behavior.This information(dataset) is commonly used in selling campaigns to hit a target sets.[9]

Today, a lot of organizations and firms are using predictive analytics (PA) to extend their business and

- To give a better vision
- To be competitive in order to catch up with the modern cometition.
- Having a better time management

#### Predictive Analysis(PA)

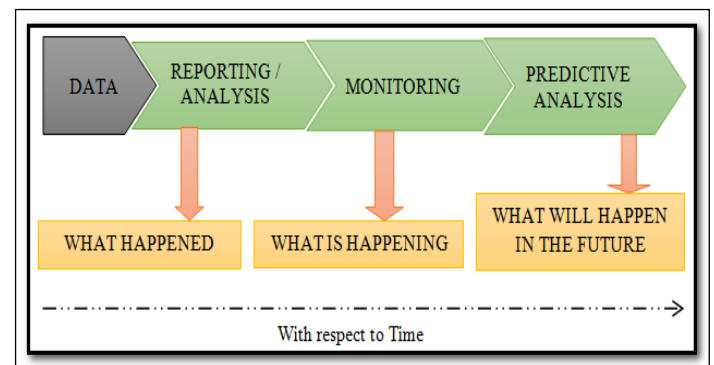


Figure 2. Describes about steps in predictive analysis

#### B. Techniques Used In Predictive Analytics

Listed below are a few techniques used in Predictive Analysis(PA)which make the process easier

- 1) Description: This method summarizes what has happened in the past and tries to analyze and characterize it, with a watch towards predicting similar events within the future. Describing past

behavior and applying predictive models to the resulting data helps to frame opportunities for operational improvement and establish new business opportunities.

- 2) Correlation: Users do correlation analysis to spot relationships and dependencies between totally different data variables to predict how they'll affect one another going forward. Correlations could be positive or negative. Determining that there's no correlation between a set of variables can also be useful in targeting predictive analytics projects which comes in meaningful data.
- 3) Segmentation: This method is a way to analyse enormous collection of entity data, such as a client database, and organize it into smaller teams. All the entities that are collected into the similar subgroup are determined to be identical to each other on the specified characteristics, that lends itself to predicting future behavior or events.
- 4) Classification: Another means of separating different entities in a data set into related groups is to map them into predefined classes based and on relevant characteristics or behaviors. The resulting classification model can be used to categorize new records and also for predictive modelling against the data for the selected subgroups.
- 5) Regression: This method is designed to spot meaningful relationships among data variables, specifically looking at the connections between a dependent variable and alternative factors that may or may not affect it. The information permits analysts to predict future developments related to the dependent variable based on what happens with associated factors.
- 6) Association: another technique for highlighting relationships between data elements for predictive functions is to look for ones that demonstrate affinity. For example, merchandise that usually are purchased along

### C. Applications of Predictive Analysis

Listed below are a few applications of Predictive Analysis [3]:

- 1) Customer Relationship Management (CRM): Predictive Analysis (PA) is used to attain CRM motives such as marketing campaigns, sales and purchaser services. It can be applied throughout the customers life cycle.
- 2) Health Care: Predictive Analysis (PA) can determine the risk of developing certain illnesses and support in taking the right medical cure.
- 3) Fraud Detection: Predictive Analysis (PA) can find inaccurate credit applications, fraudulent transactions both online and offline and identify the thefts.
- 4) Risk Management: Predictive Analysis (PA) does the probabilistic risk assessment to maximize the return to yield accurate forecasts.

### D. Pros and Cons of Predictive Analytics

The benefits and limitations of Predictive Analysis are as in Table .2.

TABLE III. PROS AND CONS OF PREDICTIVE ANALYTICS

	PROS	CONS
1	Improve efficiency in production	Information is not absolute
2	Gain advantage over competitors	Continuous updating required
3	Reduce Risk	It differs for every market
4	Meet customer expectations	Get difficult for large datasets(occasionally)

#### IV. DATA MINING(DM) AND PREDICTIVE ANALYSIS

In this section we have mentioned the association and contrast in brief between Data Mining (DM) and Predictive Analysis (PA)

##### A. Relation between Data Mining and Predictive Analysis:

Knowing what your customers are most likely to do or what they want or how much they are likely to pay to urge it is one of the most effective attainable ways to hit youth audience. For example-think of Netflix binge recommending sci-fi shows this is often a pure example of analytics results. Furthermore, each of the procedure Data Mining (DM) as well as Predictive Analysis (PA) deal with discovering secrets with big data but people often get confused with these methodologies. Data mining uses a software to search for patterns while predictive analytics (PA) uses those patterns to make predictions and direct selections. So, it is safe that Data Mining (DM) turns out to be a stepping stones for Predictive Analysis (PA).

Apart from this Data Mining (DM) is passive while predictive analytics is active and has the potential to supply a transparent image

##### B. Data Mining (DM) Vs Predictive Analytics

Often, Data Mining (DM) and Predictive Analytics (PA) are used interchangeably and also, methods (designs) and tools of Data Mining (DM) play a vital role in predictive analytics solutions. However, Predictive Analytics (PA) goes on the farther phase of the Data Mining (DM). To be competitive in the modest competition corporations have to be able to make the most out of current data information to predict what would possibly happen in the future. Predictive analytics (PA) plays a vital part in being able to capture helpful information (data and use it to

understand and model customer behaviors, sales patterns and substitute trends for the long stretch. Though Predictive Analytics (PA) is sometimes associated to Data Mining (DM) to explain how information (data) is being processed there are substantial differences between these methods. Predictive analytics (PA) and Data Mining (DM) use various kinds of algorithms to discover information (knowledge) and find the most efficient and effective solutions. Data Mining (DM) is a process supported by algorithms to investigate and extract useful information and automatically discovers hidden sequential patterns and relationships from information (data). Instead, predictive analytics (PA) is closely tied to machine learning (ML) as it uses data designs to analyze and make predictions where machines take historical (previous) and current data and apply them to a model to predict future movements [9].

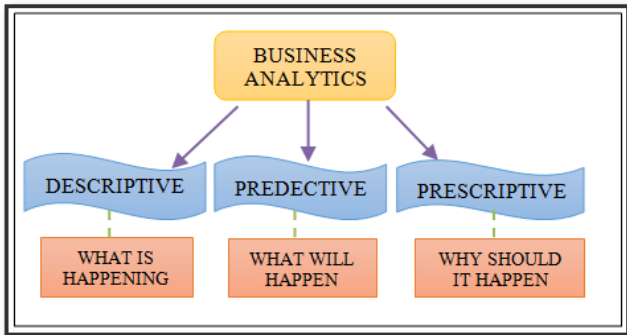
#### V. Business Intelligence (BI)

Business Intelligence (BI) having the ability to use the data (information) you collect is a minimum of as necessary and important as gathering it [16]. Thus, it is necessary to own Business Intelligence (BI) that means Business Intelligence (BI) is the ability to design and remodel the data into cumulative information and this cumulative data into knowledge. It is the most effective way to optimize the choice making method during this sense Business Intelligence (BI) could be a technology and set of methodology applications to gather refine and remodel this information frpm unstructured data internal and external to the corporate in structured information for direct management or for analysis.

Business combines data analysis applications together with unplanned analysis and querying enterprise reporting online analytical processing (OLAP) mobile BI, real-time, BI operational BI, cloud and

software as a service BI, open source BI, collaborative BI and location intelligence.

### Explanation of Business Analytics



**Figure 3.** A brief description of business analytics is explained in three steps

#### A. Relation between Data Mining and Data Intelligence

Data Mining (DM) tools offer higher customers relationship administration too, through mining the data of real habits and numerous patterns. Business Intelligence (BI) acts as a strategic issue for a business, providing information to retort to business problems: getting into new markets, monetary management, value optimization, production designing, analysis of client profiles, gain. Data Mining (DM) and Business Intelligence (BI) have created potential that varied industries, like sales and selling, aid organization or monetary establishments, may have a fast analysis of information and thereby, rising the standard of decision-making method in their industries. Additionally, data processing technologies have bright future in business applications, creating potential new opportunities by machine-driven prediction of trends and behaviors in these businesses. So, however Data Mining (DM) is employed to get Business Intelligence (BI) may be a concept that we are going to hear plenty throughout these years” it's the longer term.” [16]

#### B. Relation between Predictive Analysis and Business Intelligence

Predictive Analysis (PA) modeling has been one in every of the main reasons in dynamically changing the products and services provided by firms in recent years. Yahoo search has drastically overtaken most of its competitors and this has been potential thanks to Yahoo’s investment into Predictive Analysis (PA) because it uses completely different algorithms and varied predictive designs and models to analyze and predict users’ search results and news feeds to raised facilitate the user. Flipkart conjointly depends on predictive models of what reasonable product a user may purchase and the way will they manipulate the user to shop for the merchandise. The advertisements that are typically displayed on user’s screen whereas visiting a web site is principally supported and based on predictive model which helps the corporate to raise popularize amongst people that can be potential customers. The applications of predictive algorithms aren't solely restricted to the net world. Health cares also are transiting towards higher utilizing it to produce quality services to humanity. The predictive designs are supported by the information of a person’s health prices and output gives us the “risk score” that progresses costs and quality of health care system. Predictive analytics (PA) tries to predict behavior in future by finding patterns within the information obtainable with it by applying a variety of different algorithms. [14]

Risk management [4] is often defined as the method or procedure of identification, investigation and either acceptance or qualification of ambiguity in investment decision making model or process. Risk management is regarding managing ambiguity associated with a risk. Ancient risk management focuses on risks stemming from physical or legal causes like natural calamities or accidents, death, fires and lawsuit. Monetary risk management deals with risks which will be managed exploitation listed financial instruments. The foremost recent concept of enterprise risk management provides a tool to

reinforce the worth of systems, each industrial and communal, from a scientific purpose of read. Research (OR) is usually helpful for optimizing risk management.

## VI. RISK REDUCTION

### A. Applications of Predicative Analytics in Risk Management

The following are few general examples of how predictive analytics is used to reduce the risk

- 1) **Breweries:** They must estimate their future sales in contrast to any business. Their revenue predictions rely on the weather. Satellite observation post info will offer weather records all the way down to some sq. meters that permit the danger to be far more accurately foreseen and priced. With additional subtle foretelling, breweries will confiscate tailored insurance merchandise that shield their revenue forecasts from the danger of loss of earnings because of an extraordinarily wet summer.
- 2) **Agriculture:** In this business are more and more moving towards mistreatment more connected machinery. this could feed back higher measuring of soil conditions and chemical use, that successively permits them to maximize crop yields and predict and improve their results 'and then estimate the price of inclemency, and insure against possible future losses.
- 3) **Airlines:** They face high mounted prices with low margins, that the call to ground associate degree heavier-than-air craft because of inclemency will be pricey. In markets wherever regulators provide shoppers a right to assert back the price of tickets if they're left waiting on the runway, the price is even bigger. Once inclemency happens, it's the airline creating the choice of whether or not or to not fly. In basic terms, this call would be the event that triggers associate degree insurance pay-out.

However, no insurance firm will ensure a risk wherever the consumer themselves chooses the trigger, because it creates a perverse incentive.

There are symbols of modern thinking. One chance is for airlines to designate a particular variety of cancellations they expect – like 10 days of precipitation over an outlined amount of winter. Then, if they need to cancel extra airlifts than that variety within the outlined amount, the insurance is activated. The trigger should rely on the particular circumstances of the airline; however, insurance solutions are attainable and imaginative thinking is getting down to come back up with them.

### B. Risk Definition

Risk Assessment permits users to analyse potential issues (risks) related to a given business. The main agenda for Data Mining (DM) and analytics here is to build call support systems that may precisely predict if they are profitable actions for a business firm or not [12].

An example within the banking zone is described following to see how the customers will take over praise. Here we tend to use various type of information (data) so as to pick if an applicant is satisfactory to receive praise or not. More specifically, we tend to assess the chance that a given client does not pay a loan, thus mitigating the impact of default risk.

Input variables

- Socio-defining factors (age, marital status, gender and so on)
- Product purchase details (debit amount. Credit amount...)
- Customer's economical behaviour
- Target variable
- Defaulting



The following table lists a risk assessment test for this application

TABLE IIIII. RISK ASSESSMENT [12]

	Value
Classification accuracy	0.804
error rate	0.196
sensitivity	0.778
specificity	0.84

As we can see, the precision of the predictive design is concerning nearly about 80 percent, which is a smart worth doing this type of applications. Therefore, the given architecture is prepared to assess default risk of recent customers. [12].

### VII. CONCLUSION

The key to Business Intelligence (BI) is data, therefore in areas wherever the information is either not quantified or qualified the decision making is extremely compact. Business Intelligence in areas like Health care or Automotive or Security must be rigorously evaluated with models to prove out the What-If situations before rolling it out for application within the general market place as there's a high impact on the society and to not mention irreversible in some cases. In my opinion Business Intelligence may be a terribly powerful tool if used with the tried models and information sets however because it goes with any tool, we need to apply caution and use heap of alternative parameters like ethics or personal judgment to make sure a right call is taken, therefore I, conclude by stating that this paper presents some important facts and information about Data Mining(DM), Predictive Analysis(PA) and Business Intelligence(BI). Also, the information and examples about risk managements are briefly stated in the previous sections.

### VIII. REFERENCES

- [1] <https://www.datasciencecentral.com/profiles/blogs/the-7-most-important-data-mining-techniques>
- [2] <https://searchsqlserver.techtarget.com/definition/data-mining>
- [3] <https://www.predictiveanalyticstoday.com/what-is-predictive-analytics/>
- [4] Desheng Dash Wu, Shu-heng Chen, David L. Olson, Introduction to the Special Issue on “Business Intelligence (BI) in Risk Management”. Article in Information Sciences · January 2014. DOI: 10.1016/j.ins.2013.10.008
- [5] <https://channels.theinnovationenterprise.com/articles/pros-and-cons-of-datamining-social-interactions>
- [6] <http://bus237datamining.blogspot.com/2012/11/advantages-disadvantages.html>
- [7] <https://www.wideskills.com/data-mining-tutorial/data-mining-techniques>
- [8] <https://www.educba.com/predictive-analytics-vs-data-mining/>
- [9] <https://www.expertsystem.com/data-mining-predictive-analytics-difference/>
- [10] [https://en.wikipedia.org/wiki/Predictive\\_analytics](https://en.wikipedia.org/wiki/Predictive_analytics)
- [11] <https://www.allerin.com/blog/transforming-risk-management-with-predictive-analytics>
- [12] [https://www.neuraldesigner.com/blog/6\\_Applications\\_of\\_predictive\\_analytics\\_in\\_business\\_intelligence](https://www.neuraldesigner.com/blog/6_Applications_of_predictive_analytics_in_business_intelligence)
- [13] <http://www.rmmagazine.com/2014/05/01/the-new-math-bringing-predictive-analytics-into-the-mainstream/>
- [14] Wazurkar, P., Bhadoria, R. S., & Bajpai, D. (2017). Predictive analytics in data science for Business Intelligence (BI) solutions. 2017 7th International Conference on Communication

Systems and Network Technologies (CSNT). Page no: 367-370.doi:10.1109/csnt.2017.8418568

- [15] McCue, C. (2006). Data Mining (DM) and Predictive Analytics in Public Safety and Security. IT Professional, 8(4), page no:12–18. doi:10.1109/mitp.2006.84
- [16] A. Martin, T. Miranda Lakshmi, V. Prasanna Venkatesan. An Analysis on Business Intelligence (BI)Models to Improve Business Performance.IEEE- International Conference on Advances in Engineering, Science and Management (ICAESM -2012) March 30, 31, 2012.Page no:503-508, ISBN: 978-81-909042-2-3 2012 IEEE
- [17] <http://www.blog-geographica.com/2016/11/15/how-data-mining-is-used-to-generate-business-intelligence/>

**Cite this article as :**

Ridhi Thakur, "Predictive Analysis for Risk Reduction in Data Mining ", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 5 Issue 4, pp. 115-124, July-August 2019. Available at doi : <https://doi.org/10.32628/CSEIT195417>  
Journal URL : <http://ijsrcseit.com/CSEIT195417>