

# **DATA PREDICTION MODEL**

**(BASED ON DATA SCIENCE)**

## **MINOR PROJECT SYNOPSIS**

*Submitted in partial fulfillment of the requirements for the award of the degree*

*Of*

## **BACHELOR OF TECHNOLOGY**

*In*

## **MECHANICAL AND AUTOMATION ENGINEERING**

*By*

**MOHD. FAIZAN**

**00696203617**

**F-16**

**AKSHIT GUPTA**

**41296203617**

**F-16**

**DHRUV PARASHER**

**41596203617**

**F-16**

**GAURAV**

**41196203617**

**F-16**

*Guided by*

**Dr. Deepak Bharadwaj**

**(Dept. of Mechanical and Automation Engineering )**



**Department of Mechanical and Automation Engineering**

**DR. AKHILESH DAS GUPTA INSTITUTE OF TECHNOLOGY AND MANAGEMENT**

**(AFFILIATED TO GURU GOBIND SINGH INDRAPRASTHA UNIVERSITY)**

**NEW DELHI – 110078.**

**AUG-DEC, 2020.**

## TABLE OF CONTENTS:-

S NO.	CONTENTS	PAGE NO.
1.	Introduction	3
2.	Objective	4
3.	The Project Plan	4
4.	Advantages of data prediction model	6
5.	Applications of data prediction model	6
6.	Conclusion	7

# INTRODUCTION:

Data science is all about using data to solve problems. The problem could be decision making such as identifying which email is spam and which is not.



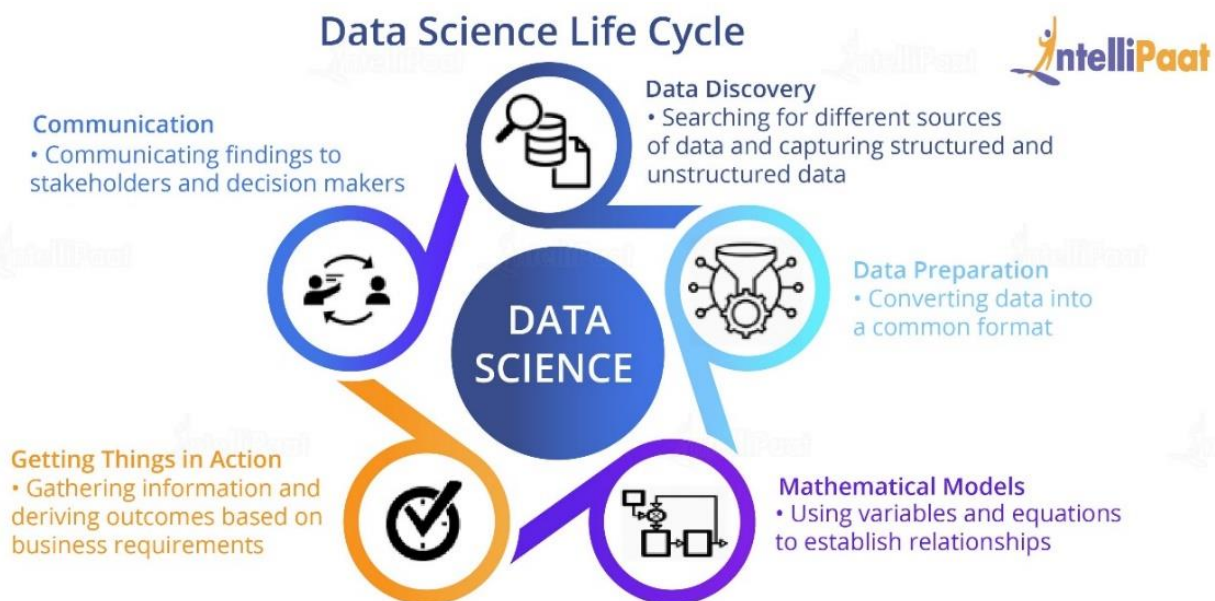
Data Science is a blend of various tools, algorithms, and machine learning principles with the goal to discover hidden patterns from the raw data. Data Science is primarily used to make decisions and predictions making use of predictive causal analytics, prescriptive analytics (predictive plus decision science) and machine learning.

- **Predictive causal analytics** – If you want a model that can predict the possibilities of a particular event in the future, you need to apply predictive causal analytics..
- **Prescriptive analytics:** If you want a model that has the intelligence of taking its own decisions and the ability to modify it with dynamic parameters, you certainly need prescriptive analytics for it. This relatively new field is all about providing advice.
- **Machine learning for making predictions** – If you have transactional data of a finance company and need to build a model to determine the future trend, then machine learning algorithms are the best bet.
- **Machine learning for pattern discovery** – If you don't have the parameters based on which you can make predictions, then you need to find out the hidden patterns within the dataset to be able to make meaningful predictions. This is nothing but the unsupervised model as you don't have any predefined labels for grouping.

# OBJECTIVE:

The main objective of data prediction model is to use current and historical data to make predictions about future customer behaviour, otherwise unknown events, risks, trends and opportunities.

## THE PROJECT PLAN



### **Data Discovery**

The first phase in the Data Science life cycle is data discovery for any Data Science problem. It includes ways to discover data from various sources which could be in an unstructured format like videos or images or in a structured format like in text files, or it could be from relational database systems. Organizations are also peeping into customer social media data, and the like, to understand customer mindset better.

## Data Preparation

Once the data discovery phase is completed, the next stage is data preparation. It includes converting disparate data into a common format in order to work with it seamlessly. This process involves collecting clean data subsets and inserting suitable defaults, and it can also involve more complex methods like identifying missing values by modeling, and so on. Once the data cleaning is done, the next step is to integrate and create a conclusion from the dataset for analysis. This involves the integration of data which includes **merging** two or more tables of the same objects, but storing different information, or summarizing fields in a table using **aggregation**. Here, we would also try to explore and understand what patterns and values our datasets have.

## Mathematical Models

All **Data Science projects** have certain mathematical models driving them. These models are planned and built by the Data Scientists in order to suit the specific need of the business organization. This might involve various areas of the mathematical domain including statistics, logistic and linear regression, differential and integral calculus, etc. Various tools and apparatus used in this regard could be R statistical computing tools, **Python programming language**, **SAS advanced analytical tools**, **SQL**, and various data visualization tools like **Tableau** and **QlikView**.

## Getting Things in Action

Once the data is prepared and the models are built, it is time to get these models working in order to achieve the desired results. There might be various discrepancies and a lot of troubleshooting that might be needed, and thus the model might have to be tweaked. Here, model evaluation explains the performance of the model.

## Communication

Communicating the findings is the last but not the least step in a Data Science endeavor. In this stage, the Data Scientist needs to be a liaison between various teams and should be able to seamlessly communicate his findings to key stakeholders and decision-makers in the organization so that actions can be taken based on the recommendations of the Data Scientist.

# **Data Science Components**

key components of Data Science, which are:

- **Data (and Its Various Types)**

The raw dataset is the foundation of Data Science, and it can be of various types like structured data (mostly in a tabular form) and unstructured data (images, videos, emails, PDF files, etc.)

- **Programming (Python and R)**

Data management and analysis is done by computer programming. In Data Science, two programming languages are most popular: Python and R.

- **Statistics and Probability**

Data is manipulated to extract information out of it. The mathematical foundation of Data Science is statistics and probability. Without having a clear knowledge of statistics and probability, there is a high possibility of misinterpreting data and reaching at incorrect conclusions. That's the reason why statistics and probability play a crucial role in Data Science.

- **Machine Learning**

As a Data Scientist, every day, you will be using Machine Learning algorithms such as regression and classification methods. It is very important for a Data Scientist to know [Machine learning](#) as a part of their job so that they can predict valuable insights from available data.

## ADVANTAGES OF DATA PREDICTION MODEL:

1. Smarter detection
2. Prioritize workloads
3. Monitor progress
4. Detect patterns to initiate action
5. Aggregate and correlate information
6. Optimize processes and performance
7. Identity insights and relationships insights
8. Catch suspicious trends before loss occurs
9. Achieve improved collaboration and control
10. Embed logic into case management systems

## APPLICATIONS OF DATA PREDICTION MODEL:

- Actuarial science,
- Marketing
- Financial services,
- Insurance, Telecommunications,
- Retail,
- Travel,
- Mobility,
- Healthcare,
- Child protection,
- Pharmaceuticals
- Capacity planning,
- Social networking

## CONCLUSION:

This paper has presented the synopsis of our Minor Project, i.e., DATA PREDICTION MODEL. It has shown all the required information about the introduction, objective, plan, advantages and application of the project. It can make a positive contribution to society. Data science can give you some pretty super superpowers. One of them is reshaping industries like healthcare. The amount of data produced about patients and illnesses rises by the second, opening new opportunities for better structured and more informed healthcare. The challenge is to carefully analyze the data in order to be able to recognize problems quickly and accurately – like deepsense.ai did in diagnosing diabetic retinopathy with deep learning.

We would like to extend our sincere gratitude to Dr. Deepak Bharadwaj for guiding us in our project. We would also like to appreciate the immense support provided by our honorable HOD Sir and other faculty members who are always there to guide us and help us in the time of need.