**ANALYSIS OF MEDICAL DATA USING DATA SCIENCE**

By

**Shafaq Ahsan (Roll No)**

**Zumar Noor (18K-1330)**



**Report submitted for the**

**Degree of MS (Data Science)**

**National University of Computer & Emerging Sciences**

Karachi-Pakistan

**2019**

**ANALYSIS OF MEDICAL DATA USING DATA SCIENCE**

**MS Course Project**

By

**Shafaq Ahsan (Roll No)**

**Zumar Noor (18K-1330)**

Course Project Supervisor

**Dr. Muhammad Wasim**

**2019**

**National University of Computer & Emerging Sciences**

Karachi-Pakistan

Pakistan

**Table of Contents**

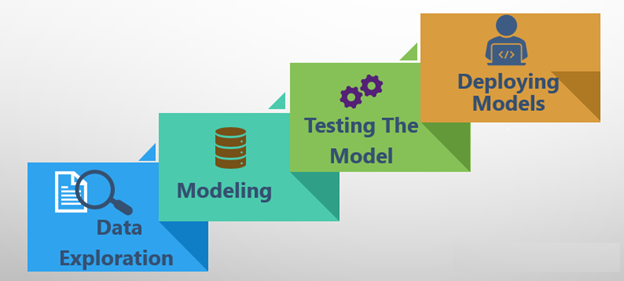
|  |  |  |  |
| --- | --- | --- | --- |
| **1.** | **Introduction** | | 1 |
|  | 1.1. | Problem Statement and Domain | 1 |
|  | 1.2. | List of Requirements | 1 |
|  | 1.3. | Introduction to Data Set | 1 |
|  | 1.4. | Overview of Statistical and Mathematical Methods for Data Science | 1 |
|  |  |  |  |
| **2.** | **Data Analytics** | | 1 |
|  | 2.1. |  | 1 |
|  | 2.2. |  | 1 |
|  | 2.3. |  | 1 |
|  | 2.4. |  | 1 |
|  | 2.5. |  | 1 |
|  | 2.6. |  | 1 |
|  | 2.7. |  | 1 |
|  | 2.8. |  | 1 |
|  | 2.9. |  | 1 |
|  | 2.10 |  | 1 |
|  |  |  |  |
| **3.** | **Results and Conclusion** | | 1 |

**1. Introduction**

Data Science is the area of study which involves extracting insights from vast amounts of data by the use of various scientific methods, algorithms, and processes. It helps you to discover hidden patterns from the raw data.

The term Data Science has emerged because of the evolution of mathematical statistics, data analysis, and big data.

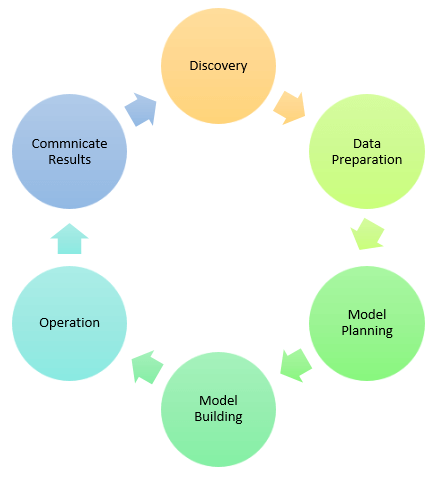
Data Science is an interdisciplinary field that allows you to extract knowledge from structured or unstructured data. Data science enables you to translate a business problem into a research project and then translate it back into a practical solution. Key components of data science are given in figure 1.



*Figure1. Key Components of Data Science*

## Data Science Process

Data science process consists of six key modules ⎯ discovery, data preparation, model planning, model building, operation and communicate results. These processes are given in figure 2.

[](https://www.guru99.com/images/1/030119_1121_WhatisDataS3.png)

*Figure2. Data Science Processes Model*

### i. Discovery

Discovery step involves acquiring data from all the identified internal & external sources which helps you to answer the business question.

The data can be:

* Logs from webservers
* Data gathered from social media
* Census datasets
* Data streamed from online sources using APIs

### ii. Data Preparation

Data can have lots of inconsistencies like missing value, blank columns, incorrect data format which needs to be cleaned. You need to process, explore, and condition data before modeling. The cleaner your data, the better are your predictions.

### iii. Model Planning

In this stage, you need to determine the method and technique to draw the relation between input variables. Planning for a model is performed by using different statistical formulas and visualization tools. SQL analysis services, R, and SAS/access are some of the tools used for this purpose.

### iv. Model Building

In this step, the actual model building process starts. Here, Data scientist distributes datasets for training and testing. Techniques like association, classification, and clustering are applied to the training data set. The model once prepared is tested against the "testing" dataset.

### v. Operationalize

In this stage, you deliver the final baselined model with reports, code, and technical documents. Model is deployed into a real-time production environment after thorough testing.

### vi. Communicate Results

In this stage, the key findings are communicated to all stakeholders. This helps you to decide if the results of the project are a success or a failure based on the inputs from the model.

<We can paste the introduction as it is, which is already there 😉>

|  |  |  |
| --- | --- | --- |
|  | **1.1.** | **Problem Statement and Domain**  <Brief about the problem> |
|  | **1.2.** | **List of Requirements**  <We can add questions asked> |
|  | **1.3.** | **Introduction to Data Set**  <Details about Data Set> |
|  | **1.4.** | **Overview of Statistical and Mathematical Methods for Data Science** |

**<**State what algorithms (regression etc) we have used**>**

**2. Data Analytics**

<We can add code snippet and 1-2 lines of explanation>

**3. Results and Conclusion**

<Results of major questions asked can be included>