**INTERDISCIPLAINARY TRAINING REPORT**

**at**

**Sathyabama Institute of Science and Technology**

**(DEEMED TO BE UNIVERSITY)**

Submitted in partial fulfillment of the requirements for the award of

Bachelor of Engineering Degree in

Computer Science and Engineering

By

**DUMPALI SHARATH CHANDRA REDDY**

**(Reg. No. 39110291)**



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**SCHOOL OF COMPUTING**

**SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY**

**JEPPIAAR NAGAR, RAJIV GANDHI SALAI,**

**CHENNAI, TAMILNADU – 600119.**

**APRIL 2022**

|  |  |  |
| --- | --- | --- |
|  | **SATHYABAMA**  **INSTITUTE OF SCIENCE AND TECHNOLOGY** (DEEMED TO BE UNIVERSITY) **Accredited with Grade “A” by NAAC**  (Established under Section 3 of UGC Act, 1956)  JEPPIAAR NAGAR, RAJIV GANDHI SALAI  CHENNAI– 600119  [www.sathyabama.ac.in](http://www.sathyabama.ac.in) |  |

# DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

**BONAFIDE CERTIFICATE**

This is to certify that this Project Report is the Bonafide work of **DUMPALI SHARATH CHANDRA REDDY (Reg. No. 39110291)** who carried out the project entitled **“ANALYSIS OF SKILLS REQUUIRED FOR A DATA SCIENTIST USING MACHINE LEARNING”** under my supervision during February 2022 - April 2022.

## Internal Guide

## Dr. T. Prem Jacob

**Head of the Department**

**Dr. L. Lakshmanan, M.E., Ph.D.,**



## Submitted for Viva voce Examination held on

**Internal Examiner External Examiner**

**DECLARATION**

I, **DUMPALI SHARATH CHANDRA REDDY** hereby declare that the project report entitled **ANALYSIS OF SKILLS REQUIRED FOR A DATA SCIENTIST USING MACHINE LEARNING** done by me under the guidance of **Dr. T. Prem Jacob** is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering Degree in Computer Science and Engineering

**DATE:**

**PLACE: SIGNATURE OF THE CANDIDATE**

**ACKNOWLEDGEMENT**

I am pleased to acknowledge my sincere thanks to **Board of Management** of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. Sasikala M.E., Ph.D.**, **Dean**, School of Computing, **Dr. S. Vigneshwari, M.E., Ph.D. and Dr. L. Lakshmanan, M.E., Ph.D., Heads of the Department** of **Computer Science and Engineering** for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr. T. Prem Jacob, M.E., Ph.D.,** for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

**TRAINING CERTIFICATE**

**ABSTRACT**

The Data Scientist role demands a person to be a jack of many skills. This information can be used to gain insights such as the key skills and tools required and which degree is preferred the most and the categorization of different roles based on the key skills, etc. This information of 7,000 data scientist jobs around the U.S is collected from Indeed website. The information collected consists of: Company Name, Position Name, Location, Job Description, and Number of Reviews of the Company.

By Analysis Of Skills Required For A Data Scientist Using IBM Watson we will:

* Know fundamental concepts and can work on IBM Watson Studio.
* Gain a broad understanding of Stochastic Gradient Descent Classification.

Artificial Intelligence being the trending technology can end up with the best solution for every typical problem. So here we are going to use the rich set of Machine Learning algorithms to predict the position.

By using the supervised learning algorithms of Machine learning such as Stochastic Gradient Descent classification we can predict the position.

**TABLE OF CONTENTS**

**TITLE PAGE NO**

**CHAPTER 1: INTRODUCTION** 8

* 1. OVERVIEW 8
  2. PURPOSE 8

**CHAPTER 2: LITERATURE SURVEY 8-9**

* 1. EXISTING PROBLEM 8
  2. PROPOSED SOLUTION 9

**CHAPTER 3: THEORTICAL ANALYSIS 9-10**

3.1 BLOCK DIAGRAM 9

3.2 HARDWARE/SOFTWARE DESIGNING 10

**CHAPTER 4: EXPERIMENTAL INVESTIGATIONS 11-12**

**CHAPTER 5: FLOW CHART 13**

**CHAPTER 6: RESULTS 14-15**

**CHAPTER 7: ADVANTAGES AND DISADVANTAGES 16**

**CHAPTER 8: APLLICATIONS 16**

**CHAPTER 9: CONCLUSION 17**

**CHAPTER 10: FUTURE SCOPE 17**

**CHAPTER 11: BIBILOGRAPHY 18**

**CHAPTER 12: APPENDIX 19-24**

**CHAPTER 1**

**INTRODUCTION**

* 1. **Overview:**

The Data Scientist role demands a person to be a jack of many skills. This information can be used to gain insights such as the key skills and tools required and which degree is preferred the most and the categorization of different roles based on the key skills, etc. This information of 7,000 data scientist jobs around the U.S is collected from Indeed website. The information collected consists of: Company Name, Position Name, Location, Job Description, and Number of Reviews of the Company.

* 1. **Purpose:**

By Analysis Of Skills Required For A Data Scientist Using IBM Watson we will:

* Know fundamental concepts and can work on IBM Watson Studio.
* Gain a broad understanding of Stochastic Gradient Descent Classification.

**CHAPTER 2**

**LITERATURE SURVEY**

* 1. **Existing Problem:**

Determining the position of different Data Scientist is a difficult task for a human as he has to consider different parameter like skills, tools, degree and come up with a conclusion. By Using the skills, tools, degree we can predict the position of Data Scientist easily

It is quite difficult for a human to work with such a huge and typical data. So here we are coming up with a solution

* 1. **Proposed Solution:**

Artificial Intelligence being the trending technology can end up with the best solution for every typical problem. So here we are going to use the rich set of Machine Learning algorithms to predict the position.

By using the supervised learning algorithms of Machine learning such as Stochastic Gradient Descent classification we can predict the position.

**CHAPTER 3**

**THEORTICAL ANALYSIS**

**3.1 Block Diagram:**

**Diagram

Description automatically generated**

**3.2 Hardware/software designing:**

**Software specifications:**

|  |  |
| --- | --- |
| **REQUIREMENT** | **SPECIFICATION** |
| Anaconda Navigator | You must have anaconda installed in your device prior to begin. |
| Spyder, Jupyter Notebook, Flask  Frame work | 1. One should have Spyder and Jupyter notebook. 2. One should install flask framework through anaconda prompt for running their web application 3. We need to build the model using Jupyter notebook with all the imported packages. |
| Web browser | For all Web browsers, the following must be enabled:   * cookies * JavaScript |

**Hardware Specifications:**

|  |  |
| --- | --- |
| **REQUIREMENT** | **SPECIFICATIONS** |
| Operating system | Microsoft Windows  UNIX  Linux® |
| Processing | Minimum: 4 CPU cores for one user. For each deployment, a sizing exercise is highly recommended. |
| RAM | Minimum 8 GB. |
| Operating system specifications | File descriptor limit set to 8192 on UNIX and Linux |
| Disk space | A minimum of 7 GB of free space is required to install the software. |

**CHAPTER 4**

**EXPERIMENTAL INVESTIGATIONS**

Analysis or the investigation made while working on the solution:

While working on the solution we investigated on what is IBM cloud, IBM Watson studio, Machine Learning service, Cloud Object Storage. The key role on investigation is collection of the dataset.

**IBM Cloud Account**:

IBM Acquired soft layer, a public cloud platform, to serve as the foundation for its IaaS offering. In October 2016, IBM rolled the soft layer brand under its Blue mix brand of PaaS offerings, giving users to access both IaaS and PaaS resources from a single console. IBM cloud provides a full-stack, public cloud platform with various products in the catalog, including options for compute, storage, networking, end to end developer solutions for app development, testing and deployment, security databases, and cloud native services.

Creating the IBM cloud account by going to the IBM cloud login page and click create on IBM cloud account. Enter our IBM id and an ID is created based on the email that we enter. Completing the remaining fields with our information and click create account by this the account is created.

**Dataset collection**:

The data collection by:

* Articulate the problem early.
* Establish data collection.
* Check our data quickly.
* Format data to make it consistent.
* Reduce data.
* Complete data cleaning.
* Decompose data.
* Take the required fields of data.

**Model Building:**

* Training and testing the model
* Evaluation of Model
* Save the model
* Predicting the output using the model

**CHAPTER 5**

**FLOW CHART**

DATA COLLECTION

DATA PREPROCESSING

IDENTIFY DEPENDENT AND INDEPENDENT VALUES

DATA VISUALISATION

IDENTIFY THE REQUIRED COLUMNS

SPLIT THE DATA INTO TRAIN AND TEST

APPLYING REGRESSION, CLASSIFICATION ALGORITHAM

TEST THE MODEL

SAVE THE MODEL

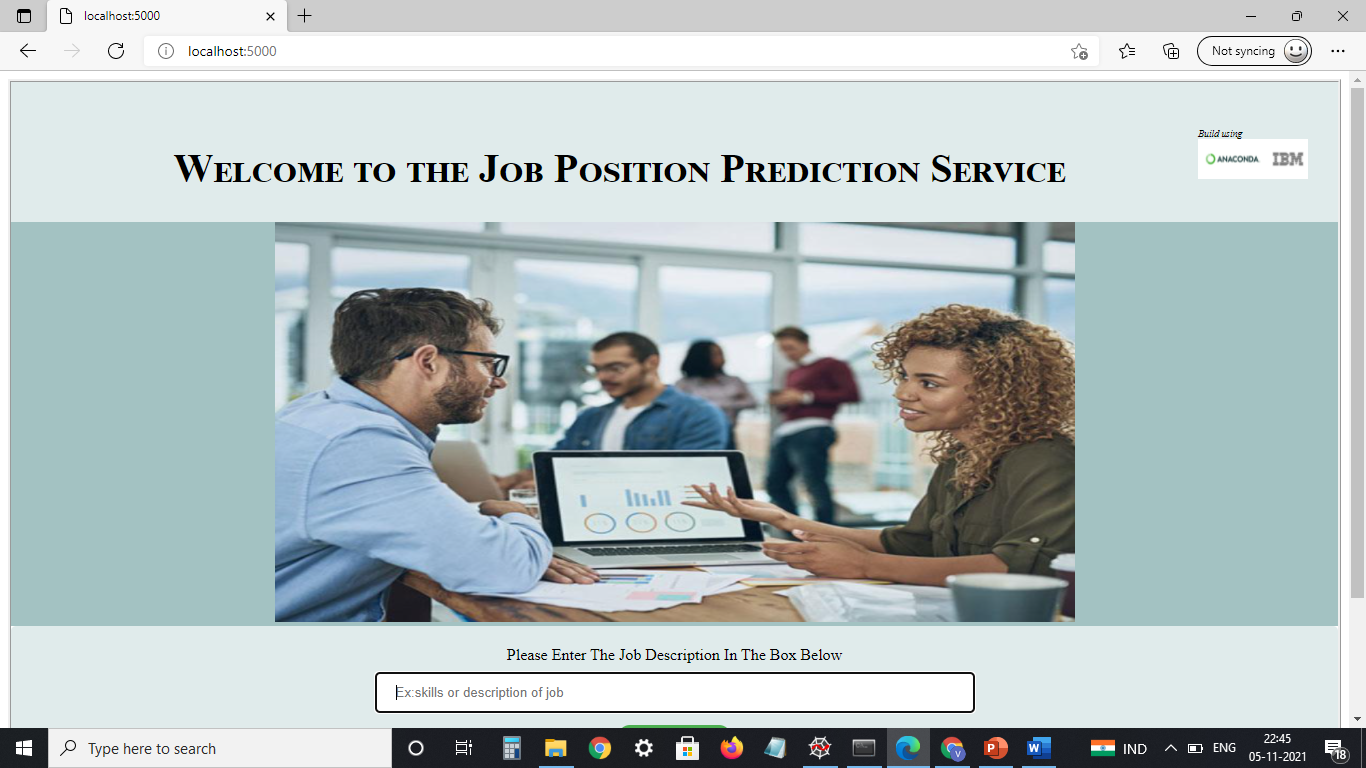
CREATE THE FLASK APPLICATIONAND INTEGRATE WITH MODEL

DEPLOY MODEL ON IBM

**CHAPTER 6**

**RESULTS**

**Final output of the project:**

****

Graphical user interface, application

Description automatically generated



**CHAPTER 7**

**ADVANTAGES AND DISADVANTAGES**

**Advantages:**

* Lower costs - reduces maintenance due to complete report coverage and a zero-footprint environment.
* Faster results - shortens reporting time due to seamless integration and adaptive authoring.
* High performance data access across all source.

**Disadvantages:**

* The permission level for a user cannot be modified
* Data grouping
* Custom visualizations
* Insights in visualization

Lower c Lower costs—reduces maintenance due to complete rep

High performance **data ac CHAPTER 8**

**APLLICATIONS**

**The areas where this solution can be applied:**

* JOB classifieds.
* Area like where can we find job position based on our skills.

**CHAPTER 9**

**CONCLUSION**

**From this entire findings we know fundamental concepts and can work on IBM Watson and machine learning.**

* Gain a board understanding of Stochastic Gradient Descent classification algorithms.
* Learn to build stunning models on IBM cloud.
* To create data visualizations to understand.

**CHAPTER 10**

**FUTURE SCOPE**

**Enhancements that can be made in the future:**

* This model can be further developed to predict the job position among all the possible positions based on skills a person have.
* We can scope the better job in future with easy experience.

**CHAPTER 11**

**BIBILOGRAPHY**

References of previous works or websites visited/books referred for analysis about the project, previous solution findings

https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms

**APPENDIX**

**index.html**

<!DOCTYPE html>

<html>

<head>

<style>

body

{

border-style:ridge

}

div.a

{

float:30px;

padding:30px;

height:80px;

background-color:#e0ebeb

}

#b {

float:right

}

#c {

float:right

}

#d {

float:right

}

#e {

font-size:60%;

text-color:blue;

font-style: italic

}

#f { float:left;

font-style:ceviche one

}

#g {

background-color: #4CAF50;

border: none;

color: white;

padding: 15px 32px;

text-align: center;

text-decoration: none;

font-size: 16px;

margin: 4px 2px;

border-radius: 12px;

}

input[type=text]

{

width: 150%;

padding: 12px 20px;

margin: 8px 0;

display: inline-block;

border: 1px solid #ccc;

border-radius: 4px;

box-sizing: border-box;

}

div.h

{

border-radius: 5px;

background-color:#e0ebeb;

padding: 20px;

}

h3 {

text-transform: uppercase;;

font-size:200%

}

div.i

{

background-color:#a3c2c2

}

div.j

{

font-variant: normal;

font-size:100%

}

p {

font-variant: small-caps

}

h2 {

font-variant: small-caps;s

margin-top:150px;

font-size:250%;

float:center

}

</style>

</head>

<body >

<div class="a">

<dl id="d">

<dt id="e">

Build using

</dt>

<dt> <a href="https://www.ibm.com">

<img id="b"

src="{{url\_for('static',filename='ibm\_img.png')}}"

height="40"

width="40">

</a>

<a href="https://www.anaconda.com/">

<img id="c"

src="{{url\_for('static',filename='anaconda-meta\_img.jpg')}}"

height="40"

width="70" >

</a>

</dt>

</dl>

<h1 id="f"></h1>

<h2 align="center">Welcome to the Job Position Prediction Service</h2>

</div >

<div class="i" style="text-align:center">

<img src="{{url\_for('static', filename='1.png')}}"

height="400"

width="800">

</div>

<div class="h">

<div class="j"

style="width:700px;margin:0 auto;text-align:center;text-style:bold">

Please Enter The Job Description In The Box Below

</div>

<form align="center" method="POST">

<input type="text"

style="width:600px"

name="text"

placeholder="Ex:skills or description of job" >

<br>

<input type="submit" id="g">

</form>

<p align="center">this job is for:</p>

<h3 align="center">{{ text }}</h3>

</div>

<div height="80">

</div>

</body>

</html>