## MINI PROJECT - I

## SYNOPSIS



Department of Computer Science & Application

## Institute of Engineering & Technology

SUBMITTED TO: - SUBMITTED BY: -

Mr. Abhishek Tiwari Shreyash Gaur (201500674)

Devesh Srivastav (201500217)

Shivanshu Agrawal (201500666)

Siddhant Yadav (201500689)

Prakhar Verma (201500495)

# ACKNOWLEDGEMENT

It gives us a great sense of pleasure to present the synopsis of the B.Tech mini project undertaken during B.Tech III Year. This project is going to be an acknowledgement to the inspiration, drive and technical assistance will be contributed to it by many individuals. We owe special debt of gratitude to Mr. Abhishek Tiwari, Technical Trainer, for providing us with an encouraging platform to develop this project, which thus helped us in shaping our abilities towards a constructive goal and for his constant support and guidance to our work.

His sincerity, thoroughness and perseverance has been a constant source of inspiration for us. We believe that he will shower us with all his extensively experienced ideas and insightful comments at different stages of the project & also taught us about the latest industry-oriented technologies. We also do not like miss the opportunity to acknowledge the contribution of all faculty members of the department for their kind guidance and co-operation.

Shreyash Gaur (201500674)

Devesh Srivastav (201500217)

Shivanshu Agrawal (201500666)

Siddhant Yadav (201500689)

Prakhar Verma (201500495)

## ABSTRACT

Company is facing a problem in identifying the right people for promotion (only for manager position and below) and prepare them in time. Currently they are following a long and tiresome process using conventional methods.

The final promotions are only announced after the evaluation and this leads to delay in transition to their new roles. Hence, company needs our help in identifying the eligible candidates at a particular checkpoint so that they can expedite the entire promotion cycle.

Company have provided multiple attributes around Employee's past and current performance along with demographics. Now the task is to predict whether a potential employee at checkpoint in the test set will be promoted or not after the evaluation process.

**CONTENT**

Abstract Declaration Acknowledgement

1. Introduction
   1. Objective
   2. Motivation
   3. Problem Statement
2. Software Requirement
   1. Hardware Requirements
   2. Software Requirements
3. Project Description
4. Working
5. Implementation
6. References

# INTRODUCTION

The HR analytics is revolutionizing the way human resources departments operate, leading to higher efficiency and better results overall. Human resources have been using analytics for years. However, the collection, processing and analysis of data has been largely manual, and given the nature of human resources dynamics and HR KPIs, the approach has been constraining HR. Therefore, it is surprising that HR departments woke up to the utility of machine learning so late in the game. Here is an opportunity to try predictive analytics in identifying the employees most likely to get promoted.

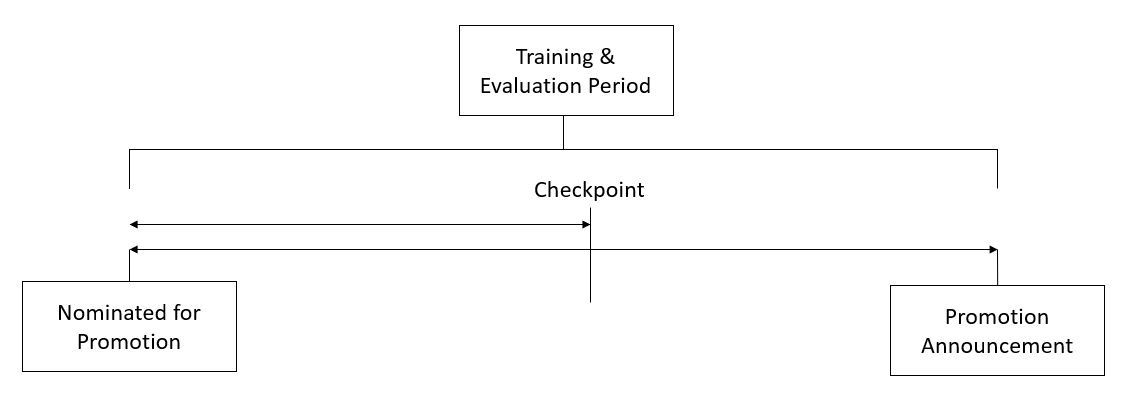
**SOFTWARE AND HARDWARE REQUIREMENTS**

* Python 3.10
* Jupyter notebook
* MS Excel
* VS Code
* Window 11

**PROJECT DESCRIPTION**

Our client is a large MNC and they have 9 broad verticals across the organization. One of the problem your client is facing is around identifying the right people for promotion (only for manager position and below) and prepare them in time. Currently the process, they are following is:

1. They first identify a set of employees based on recommendations/ past performance
2. Selected employees go through the separate training and evaluation program for each vertical. These programs are based on the required skill of each vertical
3. At the end of the program, based on various factors such as training performance, KPI completion (only employees with KPIs completed greater than 60% are considered) etc., employee gets promotion

For above mentioned process, the final promotions are only announced after the evaluation and this leads to delay in transition to their new roles. Hence, company needs your help in identifying the eligible candidates at a particular checkpoint so that they can expedite the entire promotion cycle. 

They have provided multiple attributes around Employee's past and current performance along with demographics. Now, the task is to predict whether a potential promotee at checkpoint in the test set will be promoted or not after the evaluation process.

## Dataset Description

|  |  |
| --- | --- |
| **Variable** | **Definition** |
| employee\_id | Unique ID for employee |
| department | Department of employee |
| region | Region of employment (unordered) |
| education | Education Level |
| gender | Gender of Employee |
| recruitment\_channel | Channel of recruitment for employee |
| no\_of\_trainings | no of other trainings completed in previous year on soft skills, technical skills etc. |
| age | Age of Employee |
| previous\_year\_rating | Employee Rating for the previous year |
| length\_of\_service | Length of service in years |
| KPIs\_met >80% | if Percent of KPIs(Key performance Indicators) >80% then 1 else 0 |
| awards\_won? | if awards won during previous year then 1 else 0 |
| avg\_training\_score | Average score in current training evaluations |
| is\_promoted | (Target) Recommended for promotion |

## Evaluation Metric

The evaluation metric for this competition is F1 Score.

## Public and Private Split

Test data is further randomly divided into Public (40%) and Private (60%) data.

## WORKING

predicting whether a potential employee at checkpoint in the test set will be promoted or not after the evaluation process. We will use a model which will have the best accuracy.

## Evaluation Metric

The evaluation metric for this competition is F1 Score.

## Public and Private Split

Test data is further randomly divided into Public (40%) and Private (60%) data.

**IMPLEMENTATION**

Python is a high-level, general-purpose programming language. Its design philosophy emphasizes code readability with the use of significant indentation. Python is dynamically-typed and garbage-collected. It supports multiple programming paradigms, including structured, object-oriented and functional programming.

### Pandas

It offers data structures and tools for effective data cleaning, manipulation, and analysis. It provides tools to work with different types of data. The primary instrument of Pandas is a two-dimensional table consisting of columns and rows.

NumPy

libraries are based on arrays, enabling you to apply mathematical functions to these arrays. Pandas is actually built on top of NumPy Data visualization methods are a great way to communicate with others and show the meaningful results of analysis. These libraries enable you to create graphs, charts and maps.

Matplotlib

Visualization package is the most well-known library for data visualization, and it’s excellent for making graphs and plots. The graphs are also highly customizable. Another high-level visualization library, Seaborn, is based on matplotlib. Seaborn makes it easy to generate plots like heat maps, time series, and violin plots.

Scikit-learn

The library contains tools for statistical modeling, including regression, classification, clustering and others. It is built on NumPy, SciPy, and matplotlib, and it’s relatively simple to get started. For this high-level approach, you define the model and specify the parameter types you would like to use.

## REFERENCES

**Books:**

* Mitchell Tom M: Machine Learning; McGraw Hill Science/Engineering/Math; 1997
* NG Andrew: Machine Learning Yearning; 2018

## Websites:

* [Analytics Vidya](https://datahack.analyticsvidhya.com/contest/wns-analytics-hackathon-2018-1/#ProblemStatement)
* [Kaggle](https://www.kaggle.com/)
* [Data Science Dojo](https://datasciencedojo.com/)
* [Google](http://www.google.com/)

## Faculty Guidelines:

Mr. Abhishek Tiwari (Technical Trainer in GLA University)

## GitHub Repository link:

<https://github.com/shivanshu989/HR_Analytics_Mini_Project.git>