Employee Absenteeism

<u>Group Members</u>

Raj Ahuja

Yashraj Khokrale

Aniket Patil

Prathamesh Shelar

Introduction:

The goal of this project is to analyse the given Yato Corp data and check the reasons of absenteeism based on the past data. Absenteeism is an employee's intentional or habitual absence from work. While employers expect workers to miss a certain number of workdays each year, excessive absences can equate to decreased productivity and can have a major effect on company finances, morale and other factors. This project looks at the causes of absenteeism, the costs of lost productivity and what employers can do to reduce absenteeism rates in the workplace.

Motivation:

Absenteeism costs companies billions of dollars each year in lost productivity, wages, poor quality of goods/services and excess management time. In addition, the employees who do show up to work are often burdened with extra duties and responsibilities to fill in for absent employees, which can lead to feelings of frustration and a decline in morale.

Occasional absences from work are inevitable - people get sick or injured, have to take care of others, or need time during business hours to handle personal business. It is the habitual absences that are most challenging to employers, and that can have the greatest negative effect on coworkers. Because missed work days have a profound financial effect on a company's bottom line, it is beneficial for most businesses to implement strategies to equitably monitor, reduce and respond to absenteeism.

Methodology:

Problem Statement:

Yato Corp is a fictitious courier company. As we appreciate that human capital plays an important role in collection, transportation and delivery. The company is passing through genuine issue of Absenteeism. The company has shared its dataset and requested to have an answer on what are the reasons for increasing absenteeism.

Data Collection:

The database was created with records of absenteeism at work from July 2007 to July 2010 at a courier company in Brazil. This data is collected from Kaggle. The dataset contains 21 columns of 740 records, which describes different features of the dataset.

Approach:

1. Explore whether a person presenting certain characteristics is expected to be away from work at some point in time or not.

2. We want to know for how many working hours any employee could be away from work based on information like:

   ➢ How far they live from their workplace.
   ➢ How many children and pets they have.
   ➢ Do they have higher education.

Data pre processing:

Data pre processing is an important task to be done prior to analysis to get the data ready for analysis. As good data can only provide better results, data pre processing becomes necessary prior to analysis. In data pre processing, the proposed system performs data cleaning, data imputation, data normalization, and transformation. Data cleaning process removes null values and redundant attributes from the dataset.

Implementation:

On this pre processed data, The proposed system implements the Machine Learning Algorithms to check which algorithm suits better i.e., which algorithm is suitable for prediction. This system also compares the accuracy of algorithms before and after feature selection to select the best algorithm that predicts the absenteeism effectively.

Conclusion:

The main purpose of this project is to classify and predict absenteeism by using machine learning algorithms. This project will help companies to know the reasons for absenteeism and take respective measures accordingly.